Structure and function of BcpE2, the most promiscuous GH3-family beta-glucosidase for scavenging glucose from heterosides

Benoit Deflandre[1], Cédric Jadot[1], Sören Planckaert[2], Noémie Thiébaut[1], Nudzejma Stulanovic[1], Raphaël Herman[1], Bart Devreese[2], Frédéric Kerff[1,†], and Sébastien Rigali[1,†,*]

[1] InBioS – Center for Protein Engineering, Institut de Chimie B6a, University of Liège, Liège, Belgium

[2] Laboratory for Microbiology, Department of Biochemistry and Microbiology, Ghent University, Ghent, Belgium

[†]These authors jointly supervised this work.

*Corresponding author: srigali@uliege.be

Keywords: Enzyme promiscuity; GH3 family beta-glycosidase; host pathogen interaction; plant heterosides; carbon metabolism

Abstract

Cellulose being the most abundant polysaccharide on earth, beta-glucosidases hydrolyzing cello-oligosaccharides are key enzymes to fuel glycolysis in microorganisms developing on plant material. In *Streptomyces scabiei*, the causative agent of common scab in root and tuber crops, a genetic compensation phenomenon safeguards the loss of the gene encoding the cello-oligosaccharide hydrolase BglC by awakening the expression of alternative beta-glucosidases. Here we reveal that the BglC compensating enzyme BcpE2 is the GH3-family beta-glucosidase that displays the highest reported substrate promiscuity able to release the glucose moiety of all tested types of plant-derived heterosides (aryl β-glucosides, monolignol glucosides, cyanogenic glucosides, anthocyanosides, and coumarin heterosides). BcpE2 structure analysis highlighted a large cavity in the PA14 domain that covers the active site, and the high flexibility of this domain would allow proper adjustment of this cavity for disparate heterosides. The exceptional substrate promiscuity of BcpE2 provides microorganisms a versatile tool for scavenging glucose from plant-derived nutrients that widely vary in size and structure. Importantly, scopolin is the only substrate commonly hydrolyzed by both BglC and BcpE2 thereby generating the potent virulence inhibitor scopoletin. Next to fueling glycolysis, both enzymes thus also interfere with the plant defense mechanisms to fine-tune the strength of virulence.

Introduction

The major source of soil organic carbon derives from plant senescence, with cellulose, xylan, starch, and lignin being the most abundant naturally-occurring carbon-containing polymers. Optimal colonization of carbon rich environments thus mainly relies on the ability of microorganisms to participate and feed on decomposing plant biomass. The same rationale applies to phytopathogenic bacteria for which the energy required for plant colonization must be supported by catabolic pathways biased for carbon sources utilization. Filamentous Gram-positive bacteria of the genus *Streptomyces* are well-known for their role in soil mineralization and their capability to consume diverse poly-, oligo-, and monosaccharides [1–4]. Cellulose being the most abundant polysaccharide on earth, the acquisition of a complete cellulolytic system provides a competitive advantage in organic environments. However, as cellobiose is the main product resulting from cellulolysis [5,6], possessing the CebEFG-MsiK cello-oligosaccharide transporter [7–9] and the beta-glucosidase BglC [10,11] for their subsequent hydrolysis into glucose, would be a sufficient requirement for the survival of streptomycetes within a microbial community consuming plant material [2].

For *Streptomyces scabiei*, the bacterium responsible for common scab in root and tuber crops, cellobiose and cellotriose are of particular importance. Indeed, their import and subsequent catabolism do not only feed glycolysis with glucose, but they are also the environmental triggers of the onset of its pathogenic lifestyle [7,12–17]. This tight link of cellulose byproduct utilization for both primary metabolism and the onset of virulence is perfectly highlighted by the phenotype of the null mutant of gene *scab57721* encoding the β-glucosidase BglC. Indeed, strain *S. scabiei* Δ*bglC* poorly grows when cultured with cellobiose or cellotriose provided as sole carbon sources, and also displays a hypervirulent phenotype due to the constitutive production of thaxtomin A, the main virulence determinant [10]. Surprisingly, we recently showed that BglC is also able to remove the glucose moiety of the scopolin heteroside[18] produced by plants under host colonization thereby generating scopoletin, a potent inhibitor of thaxtomin A production[19]. The hydrolysis of scopolin by BglC displays a substrate inhibition kinetic profile [18] that contrasts with the typical Michaelis–Menten saturation curve observed for the degradation of its natural substrates cellotriose and cellobiose[10]. At low concentration, the scopolin and cellobiose/cellotriose degradation would synergistically reduce thaxtomin A production and or feed glycolysis, while at higher concentrations the activities on both substrates would instead activate the biosynthesis of this main virulence factor. This enzyme

has thus different kinetic properties on either the virulence elicitors emanating from the plant host or a molecule produced by the plant defense mechanisms, thereby occupying a key position to fine-tune the production of thaxtomin A[18].

Surprisingly, we showed that the deletion of *bglC* in *S. scabiei* is not lethal when the mutant strain is inoculated on media with cellobiose supplied as unique carbon source[20]. The unexpected survival of the *bglC* null mutant in this culture condition is due to a genetic compensation phenomenon that awakens the expression of the gene *scab2391* encoding an alternative GH1-family β-glucosidase[20]. This sugar hydrolase was called BcpE1 (SCAB_2391), BcpE standing for BglC compensating enzyme, and is a paralogue of BglC, both enzymes having catalytic properties in the same order of magnitude towards cellobiose[20]. Interestingly, the genetic compensation phenomenon associated with the loss of *bglC* resulted in the overexpression of a second gene (*scab64101*) encoding a GH3-family β-glucosidase called BcpE2 [20]. However, in contrast to BcpE1, BcpE2 cannot hydrolyze cellobiose[20] and therefore the reason why the *bglC*-dependent mechanism of compensation selected the product of *scab64101* as alternative β-glucosidase remained unknown. If the role of BcpE2 is not to provide glucose from cellobiose or cellotriose, how could this enzyme compensate for the loss of function of BglC? In other words, how would *S. scabiei* and other streptomycetes benefit from the activation of BcpE2 in their environmental niche?

In this work we investigated through enzymatic, structural, and expression studies the function of BcpE2. Our results demonstrate that BcpE2 is able to release the glucose moiety of various types of plant-derived heterosides. Thanks to its exceptional substrate promiscuity, BcpE2 safeguards the feeding of glycolysis with glucose hydrolyzed from highly diverse carbon sources. Moreover, BcpE2 also degrades scopolin into scopoletin following a substrate inhibition profile, thereby also compensating for the loss of the function of BglC towards the host defense mechanism.

Results

Structure of BcpE2 of *Streptomyces scabiei*

We obtained the crystallographic structure of BcpE2 at 3.09 Å. The crystal belongs to the $P3_121$ space group with one molecule in the asymmetric unit. The BcpE2 structure is characterized by R_{work} and R_{free} values of 22.3% and 27% respectively (Table S1) and contains residues 10 to 822. Six regions could not be built because of a lack of electron density: the first nine amino acids (AA), the twelve C-terminal residues, which consist in a His_6-Tag and a linker, and four loops (residues 70 to 73, 336 to 339, 436 to 443 and 531 to 536). BcpE2 is monomeric and is made of four domains (Figure 1a, Table S1): an N-terminal $(\beta/\alpha)_8$ TIM barrel domain (residues 10–306), an $(\alpha/\beta)_6$-sandwich domain (residues 316–398 and 551–650), a PA14 domain (residues 404–548), and a C-terminal fibronectin type III-like (fn3) domain (residues 711–821). In addition, a 60 AA linker is present between $(\alpha/\beta)_6$-sandwich and the fibronectin type III-like domains and mostly runs on the $(\beta/\alpha)_8$ TIM barrel domain. This architecture was identified in only 19 sequences among all 199 characterized prokaryotic and eukaryotic GH3s by mining the CAZy database. Only two structures, KmBglI from *Kluyveromyces marxianus* (PDB code 3AC0)[21], and DesR from *Streptomyces venezuelae* (PDB code 4I3G) [22] were found with this architecture. While the $(\beta/\alpha)_8$ TIM barrel, the $(\alpha/\beta)_6$-sandwich, and the fn3 domains are very well superimposed, including the linker preceding the C-terminal domain (rmsd of 0.91 Å over 507 Cα for KmBglI and 0.90 Å over 504 Cα for DesR), the PA14 domain cannot be superimposed simultaneously with the three other domains (Figure 1bc). In BcpE2 and KmBglI, the orientation is roughly similar, and they are characterized by a rmsd of 2.6 Å over 81 Cα when superimposed independently. In DesR, the PA14 domain is 40 AA shorter than in BcpE2 and approximately perpendicular to the orientation in the latter. When superimposed independently, the rmsd between them is 4.7 Å over 88 Cα. The PA14 domain of BcpE2 is also characterized by B factor values significantly higher than the rest of the protein, particularly in the active site vicinity (Figure 1d), indicating a likely flexibility of this domain. This feature was also noted for KmBglI and potentially associated with substrate recognition [21].

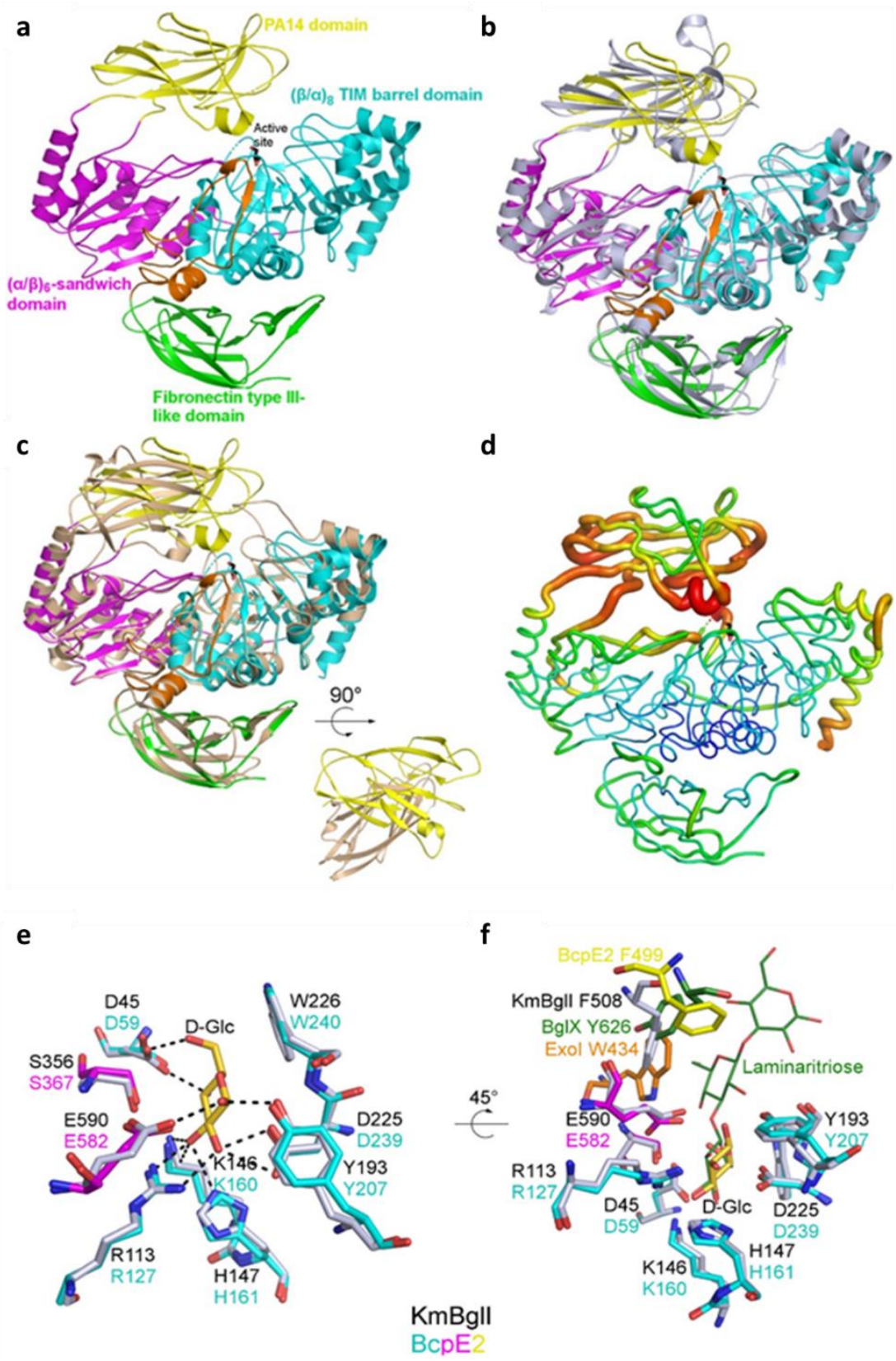

Figure 1. Overall fold and active site description of BcpE2 of *S. scabiei*. (**a**) Cartoon representation of the BcpE2 structure. The glycerol molecule in the active site (black sticks) results from the cryo-protectant solution used for freezing the crystal. (**b**) Superimposition of BcpE2 and KmBglI (light blue). (**c**) Superimposition of BcpE2 and DesR (light orange). Their PA14 are also shown with a 90° rotation to highlight the different orientations. (**d**)

Ribbon representation of BcpE2. The ribbon radius is proportional to the mean B-factor value of the residues and a rainbow coloring scheme (blue low B factor value to red high B-factor value). (**e**) Superimposition of the catalytic site of BcpE2 (residues are colored by domain with the same coloring scheme as in (b) with KmBglI (grey) in complex with D-Glucose (D-Glc in gold) in subsite -1. Hydrogen bonds with D-Glc are shown as black dashed lines. (**f**) Same as (e) with a 45° rotation and the addition of ExoI (orange) and the BglX:Laminaritriose complex (green) superimposed. The aromatic residues of subsite +1 of KmBglI and BglX are also shown as sticks, as well as their likely equivalent in ExoI and BcpE2. The position of F499 in BcpE2 (yellow sticks) comes from the PA14 domain superimposed independently on the PA14 domain of KmBglI.

The catalytic site of BcpE2

Both KmBglI and DesR have been crystallized with a β-D-glucose molecule present in their active site (-1 subsite) [21–23]. This molecule corresponds to the buried end of the substrate, and, in KmBglI, it is found in a pocket made of 9 residues (7 of them being involved in hydrogen bonds) belonging to the $(\alpha/\beta)_8$ TIM barrel and the $(\alpha/\beta)_6$-sandwich domains. These residues, which include the catalytic glutamate and aspartate responsible for the two-step double displacement mechanism [24], are strictly conserved with a very similar orientation in DesR and BcpE2 (rmsd calculated for all non-hydrogen atoms of 0.48 Å and 0.60 Å respectively; Figure 1e). Figure S1 shows the structure-based sequence alignment of BcpE2 with homologous and biochemically characterized proteins (Table S2) that either display the highest structural or primary sequence similarity. The two catalytic residues – Asp239 and Glu582 – are strictly conserved in all homologous GH3s (Figure S1). These residues provide an ideal geometry to stabilize the five hydroxyl groups of glucose with at least one hydrogen bond and facilitate an efficient hydrolysis. The active site of BcpE2 also includes Asp59, Arg127, and three additional aromatic AAs, namely Tyr207, Trp240, and Phe499 which are generally conserved in the closest GH3-family enzymes (Figure S1).

In KmBglI, Phe508, which belongs to the PA14 domain, has been identified by site directed mutagenesis as important for substrate hydrolysis and is part of the subsite (+1)[25]. In BcpE2, Phe499 is equivalent when the two PA14 domains are superimposed independently (Cα 2 Å apart) but is shifted toward the position that the substrate would likely occupy when the entire structures are superimposed (Cα 5 Å apart). This positioning could be the result of the crystal packing and/or due to the flexibility observed. An aromatic residue at this position seems to be a common feature in GH3 enzymes but can come from different structural elements (Figure S1). For example, in the ExoI enzyme form barley[26], it is located on a loop of the $(\alpha/\beta)_6$-sandwich domain, and in the dimer forming BglX from *Pseudomonas aeruginosa* [24], it comes

from an extended loop of the second molecule of the dimer (Figure 1f). In BcpE2, the high B factor values observed in this region and the likely related flexibility make it difficult to extract additional information about substrate specificity in the (+1) subsite at this stage.

Seeking for the natural substrate(s) of BcpE2

Earlier work showed that BcpE2 displayed a strong hydrolytic activity on the synthetic chromogenic substrate pNPβG[20], suggesting that the enzyme should target carbohydrates with a terminal glucose attached by a β-1,4 linkage. According to KEGG pathway, BcpE2 of *S. scabiei* 87-22 is suggested as candidate beta-glucosidase possibly involved in cyanoamino acid metabolism (https://www.genome.jp/kegg-bin/show_pathway?scb00460+SCAB_64101). Two cyanogenic glucosides were thus selected as possible targets of BcpE2, *i.e.*, amygdalin and linamarin (Table 1). To help identifying other putative substrate(s) that could be hydrolyzed by BcpE2, we generated a phylogenetic tree with BcpE2 of *S. scabiei* and the full-length sequences of the fourteen closest characterized bacterial GH3-family β-glucosidases, and with five other characterized bacterial GH3s with lower overall identity but with high query coverage and containing the PA14 domain (Figure S2).

According to these *in silico* analyses combined to a literature survey, 14 candidate natural substrates were selected for BcpE2 (Table 1). Most of them are β-1,4 linked heterosides commonly found in plants, *i.e.*, compounds with glucose (or another carbohydrate moiety) linked by a glycosidic bond to an aglycone. They belong to different types of plant heterosides with extremely variable aglycone moieties, i.e., i) aryl-β-glucosides (arbutin, salicin), ii) monolignol glucosides (p-coumaryl alcohol 4-O-glucoside, syringin, coniferin), iii) antho-cyanosides (cyanin), iv) coumarin heterosides (esculin, scopolin), and cyanogenic glycosides (linamarin, amygdalin).

Table 1

Overview of the activity of BcpE2 and BglC on selected disaccharides and heterosides

Substrate	Category	TLC assays		Kinetic parameters (BcpE2)			
		BcpE2	BglC	K_m (mM)	k_{cat} (s^{-1})	K_m/k_{cat} ($mM^{-1}.s^{-1}$)	K_i (mM)
Cellobiose	Disaccharides (β-1,4 glucose)	-	+	/[20]	/[20]	/[20]	/[20]
Xylobiose	Disaccharides (β-1,4 xylose)	-	-	NT	NT	NT	NT
Laminaribiose	Disaccharides (β-1,3 glucose)	±	+	6.557 ± 1.410	8.74 ± 1.02	**1.33**	NA
Gentiobiose	Disaccharides (β-1,6 glucose)	± / -	± / -	NT	NT	NT	NT

Salicin	Aryl-β-glucosides	+	±	0.142 ± 0.030	53.20 ± 2.37	**375.71**	NA
Arbutin	Aryl-β-glucosides	+	± / -	0.367 ± 0.080	43.17 ± 2.50	**117.49**	NA
***p*-coumaryl alcohol 4-O-glucoside**	Monolignol glucosides	+	±	0.236	77.15	**326.49**	0.593
Syringin	Monolignol glucosides	+	-	0.608 ± 0.166	26.43 ± 3.07	**43.46**	NA
Coniferin	Monolignol glucosides	+	±	0.151 ± 0.043	83.30 ± 5.90	**551.29**	NA
Cyanin (Cyanidin-3,5-di-O-glucoside)	Antho-cyanosides	+	-	NT	NT	NT	NT
Esculin	Coumarin heterosides	+	±	NT	NT	NT	NT
Scopolin	Coumarin heterosides	+	+	0.356	191.90	**539.65**	0.152
Linamarin	Cyanogenic glycosides	+	-	NT	NT	NT	NT
Amygdalin	Cyanogenic glycosides	±	-	NT	NT	NT	NT
4-MUG	Aryl-β-glucosides	+	+	NT	NT	NT	NT

The TLC assays columns summarize the results displayed in Figure 2a by the attribution of qualitative hydrolysis scores for the two enzymes. '+' indicates a complete hydrolysis, '±' indicates an incomplete hydrolysis, '±/-' indicates weak hydrolysis, and '-' indicates the absence of substrate hydrolysis (or glucose release). The kinetic parameters columns summarize the measured by initial velocities plotted as a function of the substrate concentration to obtain Henri-Michaelis-Menten or substrate inhibition curves fitted with GraphPad Prism (9.2.0). The error values of the K_m, k_{cat}, and K_i values indicate the extent of the interval to be considered in order to determine the value with 95% confidence (asymptotic method). Abbreviations: pNPβG: 4-Nitrophenyl β-D-glucopyranoside; 4-MUG: 4-methylumbelliferyl-β-D-glucoside; NT: Not Tested; NA: Not Applicable. /, too weak activity for obtaining kinetic parameters (previously published in[20]).

The candidate substrates listed in Table 1 were tested to determine the ability of BcpE2 to release their glycosidic moiety (Figure 2) β-glucosidase BglC was also included in our enzymatic assays to compare the respective substrate specificities of each enzyme. Both pure six histidine-tagged enzymes were incubated with the candidate substrates and reactions were conducted at their optimal pH (7.5) and temperature (40°C) (optima deduced from results presented in Figure S3 for BcpE2 and described in [10] for BglC). Reaction samples were spotted on thin layer chromatography (TLC) plates and migrated in an elution chamber to separate glucose (or other saccharides) from the remainder moieties of the substrate (Figure 2a).

Cellobiose was first tested as positive and negative control substrate for BglC and BcpE2, respectively. Indeed, as shown in Figure 2a, glucose is only released from cellobiose when this

substrate is incubated with BglC. Other D-glucose disaccharides were tested, namely gentiobiose (D-glucose linked in β(1→6)), and laminaribiose (D-glucose linked in β(1→3)). Surprisingly, BglC could efficiently degrade laminaribiose whereas BcpE2 could only partially degrade this substrate. Both enzymes were equally inefficient on gentiobiose where barely perceptible amounts of glucose were released (Figure 2a). Neither BglC nor BcpE2 was active on xylobiose suggesting that these enzymes cannot properly target D-xylose saccharides.

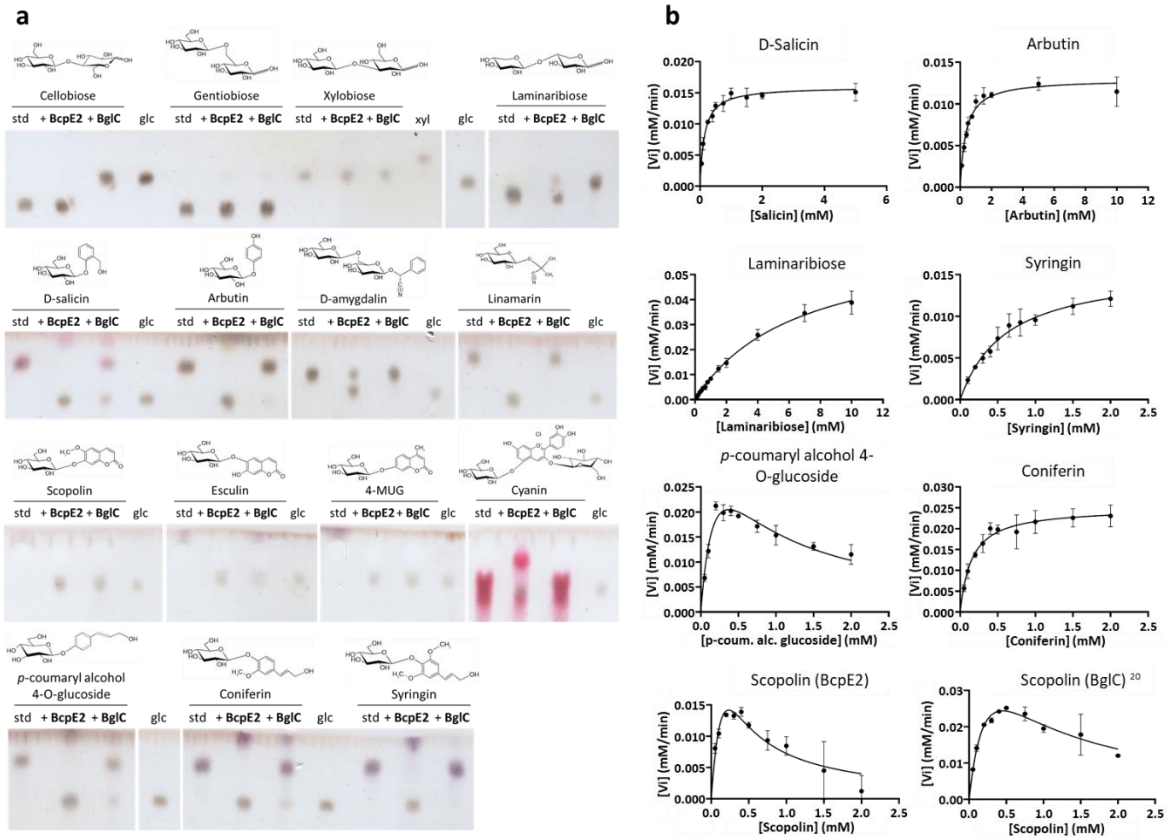

Figure 2. Substrate specificities of BcpE2 and BglC. (**a**) TLC plates revealing the release of glucose (glc) after incubation of a variety of substrates (5 mM) with BcpE2 or BglC (1 µM) compared to the intact substrate (standard (std)). The chemical structure is displayed above each substrate. (**b**) Non-linear regressions of the kinetic analyses of BcpE2 towards seven substrates and one for BglC. Plots of the initial velocity (Vi, mM/min) estimated by the rate of glucose released by the enzyme as a function of substrate concentrations (in mM). Individual values were entered into the GraphPad Prism software (9.2.0) which fitted the data to the Henri-Michaelis-Menten model by a non-linear regression. In the case of a decrease of the Vi at high substrate concentrations, the data were fitted to the Substrate Inhibition model. Error bars display the standard deviation values determined for the Vi by three replicates at each substrate concentration.

Strikingly, BcpE2 revealed to be active on all tested heterosides (Figure 2a). Complete hydrolysis was observed for i) the two aryl-β-glucosides salicin and arbutin, ii) the cyanogenic

glucoside linamarin, iii) the pink/purple anthocyanoside Cyanidin-3,5-di-O-glucoside chloride (Cyanin), iv) all three monolignol glucosides syringin, coniferin and *p*-coumaryl alcohol 4-O-glucoside, v) the coumarin heteroside esculin, and vi) the synthetic substrate 4-MUG (Figure 2a). Significant yet incomplete hydrolysis by BcpE2 was also observed for the cyanogenic glucoside amygdalin. In contrast, BglC was inactive on most tested heterosides except the synthetic substrate 4-MUG, and scopolin as previously described[18]. Only partial substrate hydrolysis by BglC could be observed for esculin, coniferin, *p*-coumaryl alcohol 4-O-glucoside, and salicin (Figure 2a). Overall, we observed that the substrate specificity of BcpE2 is broad and often complementary to that of BglC (Table 1). Surprisingly and despite an extensive variability in the aglycone parts of the tested compounds, BcpE2 managed to generate glucose from all the heteroside substrates considered in this study.

After determining the best candidate substrates of BcpE2 by preliminary enzymatic assays on TLC, a subset of them was used to evaluate the kinetic parameters of BcpE2. The values of the Michaelis constant (K_m), catalytic rate constant (or turnover; k_{cat}) and catalytic efficiency (k_{cat}/K_m) were determined for BcpE2 towards the seven following substrates: salicin, arbutin, laminaribiose, syringin, *p*-coumaryl alcohol glucoside, coniferin, and scopolin (Table 1) based on the non-linear regressions displayed in Figure 2b.

The best affinity of BcpE2 was – as indicated by the lowest K_m values – observed towards salicin and coniferin with a K_m of about 0.15 mM. The other substrates displayed values on the same order of magnitude, except laminaribiose for which the 6.557 mM estimated K_m value indicates a low affinity of the β-glucosidase for this substrate (as already anticipated from assays on TLC plates Figure 2a). Regarding the turnover parameter, the monolignol glucoside coniferin was the most efficiently hydrolyzed substrate with a k_{cat} value of 83.3 s^{-1}. This heteroside thus had the highest catalytic efficiency at about 550 $mM^{-1}.s^{-1}$, surpassing salicin and *p*-coumaryl alcohol glucoside by some margin, and is thereby the best reported substrate for BcpE2.

While most of the tested substrates exhibited conventional Henri-Michaelis-Menten behavior upon hydrolysis by BcpE2, two heterosides revealed substrate inhibition, namely the *p*-coumaryl alcohol glucoside and scopolin (Figure 2b). Indeed, the monolignol glucoside *p*-coumaryl alcohol glucoside showed a decrease in the initial velocity at high substrate concentration. The calculated inhibition constant (K_i) value for *p*-coumaryl alcohol glucoside was estimated at about 0.6 mM and the theoretical K_m and k_{cat} values without the inhibition phenomenon would be of 0.236 mM and 77.15 s^{-1}, respectively. This suggests that this

monolignol glucoside is also a good substrate for BcpE2, yet at relatively low concentrations. Interestingly, scopolin is the only natural substrate to be hydrolyzed by both BcpE2 and BglC in the TLC experiment (Figure 2a). We therefore decided to evaluate their respective kinetic parameters to evaluate how efficiently they degrade this substrate. In order to obtain similar initial velocity values at low substrate concentrations, a concentration 20-times higher of BglC compared to BcpE2 was required, suggesting a much better k_{cat} for the latter. Also, as indicated by the aspect of the non-linear regressions in Figure 2b, both enzymes are subjected to substrate inhibition. However, the K_i value for BcpE2 appears to be lower thus indicating a stronger inhibition compared to BglC (Table 1).

Production of BcpE2 is induced by the aryl-β-glucoside salicin

To validate the role of BcpE2 in heteroside degradation *in vivo*, it is mandatory to show that BcpE2 is produced when *S. scabiei* encounters these types of molecules in its environment. From all tested substrates that are best hydrolyzed by BcpE2, we chose salicin as a putative natural elicitor of BcpE2 production (also due to its availability in terms of cost and quantity) for our *in vivo* production assays. *S. scabiei* was cultivated under conditions that allow BglC production (minimal medium containing cellobiose) and/or with salicin as putative trigger for BcpE2 production. The different intracellular crude extracts were separated by anion exchange chromatography and the fractions obtained were first tested against pNPβG as substrate to detect those containing β-glucosidases, and then subjected to targeted proteomics for the identification and quantification of BglC and BcpE2. The semi-quantitative abundances of BglC and BcpE2 under the three tested culture conditions are presented in Figure 3.

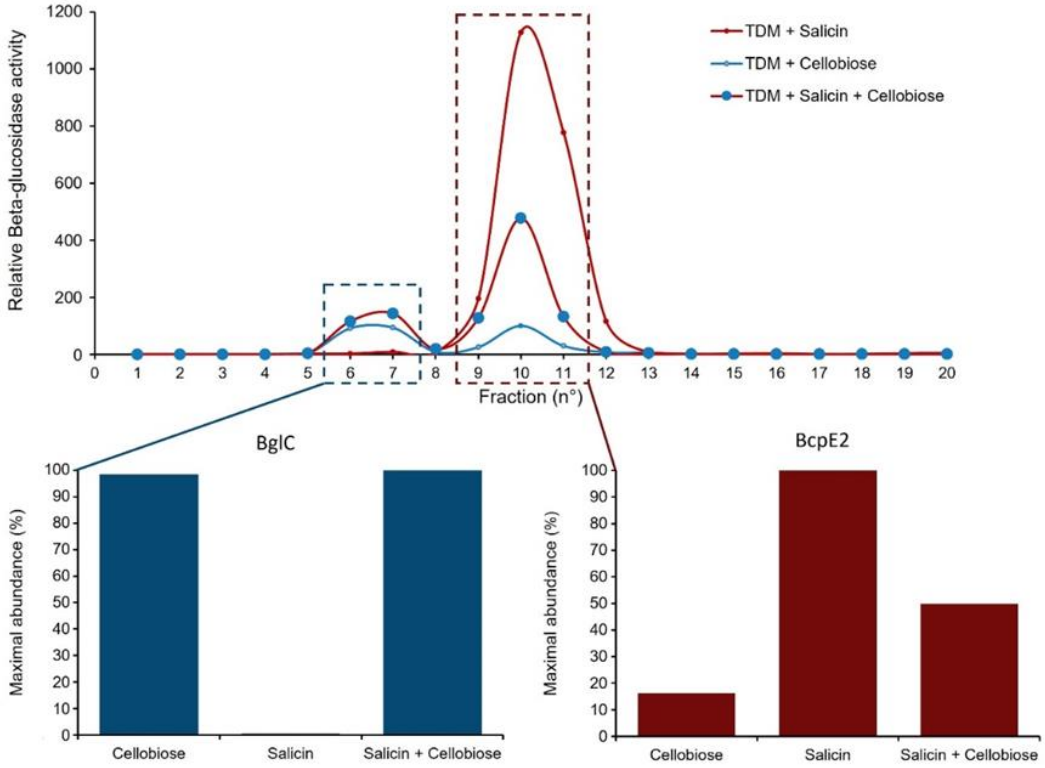

Figure 3. Induction of the respective production of BglC and BcpE2 by cellobiose and salicin. (Top pannel) Relative β-glucosidase activity in anion exchange chromatography fractions obtained from the full protein extracts of *S. scabiei* cultured in TDM medium supplemented with salicin (red trait), cellobiose (blue trait), or both substrates (red trait with blue circles). (Bottom pannel) Relative abundance of BglC in the first active peak (fractions 6-7) and of BcpE2 in the second active peak (fractions 9-11) determined by targeted proteomics (LC-MRM (Liquid chromatography multiple reaction monitoring) after tryptic digestion of the protein fractions). In each culture condition, the relative protein abundance was reported to the maximal abundance measured for the given protein.

As previously reported, the production of BglC is triggered by the presence of cellobiose. Salicin was neither able to induce (when provided as unique carbon source) nor repress (when in combination with cellobiose) the production of BglC suggesting that the expression of *bglC* is not under the control of this aryl β-glucoside. By contrast, BcpE2 was instead maximally produced when salicin was supplied as unique carbon source and the supply of cellobiose in addition to salicin reduced BcpE2 production to half of this level. When salicin was not supplied in the culture media, the production levels of BcpE2 dropped to about 16% of its maximal production level. Our results show that the production of BcpE2 is indeed triggered upon sensing the presence of salicin, one of the substrates for which the enzyme displayed the most efficient catalytic properties.

Discussion

In this work, we report the structural and biochemical characterization of BcpE2, a GH3-family β-glucosidase of the common scab phytopathogen *S. scabiei*. This protein displays low similarity compared to biochemically characterized enzymes of this family, indicating that BcpE2 could have novel functional specificities. The crystal structure of BcpE2 revealed the presence of four domains – including a rather uncommon PA14 domain predicted to be involved in substrate specificity – organized around a catalytic pocket, which can accommodate D-Glucose as buried residue. BcpE2 was highly active against a wide variety of plant heterosides mostly containing glucose as carbohydrate residue, with salicin and coniferin as the most efficiently hydrolyzed substrates.

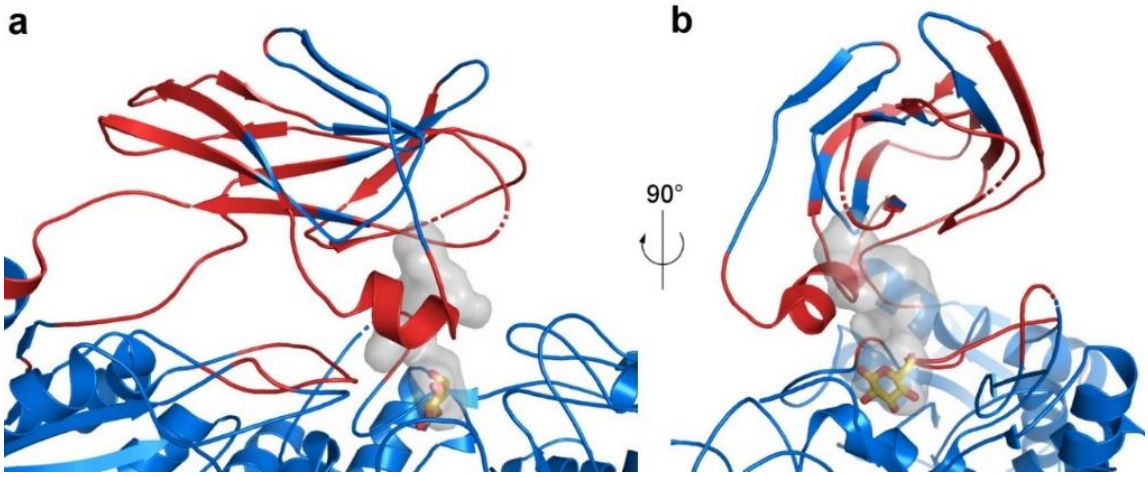

Figure 4. Active site flexibility of BcpE2. (**a**) Cartoon representation of the BcpE2 active site and the PA14 domain. Loops with missing amino acids are shown as dashed lines. Residues with a B factor of the Cα above 100 Å^2 are in red and the others in blue. The active site pocket is represented with a transparent surface with the D-glucose molecule in the (+1) subsite from the superimposed KmBglII structure as yellow sticks. (**b**) Same as in (a) with a 90° rotation.

The BcpE2 structure reveals two features likely contributing to its broad substrate specificity, *i.e.*, i) a wide cavity in the PA14 domain connected to the D-glucose specific subsite (-1), and ii) the high flexibility of this domain, especially the structure elements defining this cavity (Figure 4). Three of the four loops not fully defined in the electron density are indeed adjacent to the cavity, which is also surrounded by the residues with the highest B factors. This will therefore provide sufficient plasticity to accommodate the various substrates and an easy access to efficiently load the substrates and expel the products of the hydrolysis. Except for the

aromatic feature of Phe499, which seems to be a characteristic feature of GH3 enzymes, the residues of PA14 defining the cavity are not conserved even among closely related proteins (Figure S1). This could indicate that the main purpose of the cavity would be to hold the substrate shielded in the active site just long time enough for hydrolysis to take place, without contributing to the specificity except for setting a size limit.

At this stage, the perhaps most difficult question to answer is "which heteroside could escape hydrolysis by BcpE2", in other words, *to what extent can BcpE2 tolerate substrate promiscuity?* Glycosylated phytochemicals include phenylpropanoids, cyanogenic glucosides, coumarin heterosides, quinones, mono- or triterpenes, polyphenols, flavonoids (anthocyanosides, flavanols, isoflavonoids, flavonols and flavones), monolignols, amongst many others. Assessing the efficiency of BcpE2 to hydrolyze other substrates with an even wider spectrum of aglycone moieties will likely reveal the extent of its catalytic potential. A single heteroside, amygdalin, was not thoroughly hydrolyzed by BcpE2 (see TLC assay, Figure 2a), but it is also the only tested compound bearing gentiobiose instead of a glucose molecule as glycone. The gentiobiose disaccharide was poorly degraded by BcpE2 and it was thus not surprising that amygdalin was not thoroughly hydrolyzed. Despite this, the presence of the aromatic residue in the structure of this cyanogenic glucoside appears to enhance the activity of BcpE2 when compared to the hydrolysis of gentiobiose alone. The presence of at least one aromatic cycle in the chemical structure of the heterosides degraded by BcpE2 appears to be a common feature except for linamarin (Figure 2a). The fact that this cyanogenic glucoside is also efficiently hydrolyzed by BcpE2 suggests that the presence of an aromatic residue in the aglycone is not a mandatory feature to be accommodated as a substrate.

Why is BcpE2 selected for compensating the loss of BglC?

As cello-oligosaccharides are not natural substrates of BcpE2, the reason why the *bglC*-dependent mechanism of genetic compensation selected the product of *scab64101* as alternative β-glucosidase was a complete mystery[20]. In light of the discovery of the substrate and enzymatic specificities of BcpE2, it is now easier to understand why the product of this gene was selected to compensate for the loss of *bglC*/BglC. Indeed, we now know that BcpE2 can compensate the impaired cello-oligosaccharide consumption by feeding glycolysis with glucose hydrolyzed from a multitude of plant-derived heterosides including monolignol glycosides which are some of the most ubiquitous molecules (Figure 5). BcpE2 would therefore act as a glucose scavenging enzyme, able to provide the most readily metabolized carbon source from a plethora of

compounds that *S. scabiei* would encounter during host colonization. For soil-dwelling saprophytic streptomycetes, BcpE2 would be equally important as most of the organic carbon are molecules generated from plant senescence. Indeed, genome mining of streptomycetes revealed that most of them possess an orthologue of BcpE2. It is also striking that BglC and BcpE2 possess complementary substrate ranges on the tested molecules as the compounds properly hydrolyzed by one are generally badly degraded by the other (Table 1). Overall, the multiplicity of possible substrates for BcpE2, and to a lesser extent for BglC, suggests that these enzymes would be part of a "Swiss Army Knife" for providing the bacteria the most readily mobilizable sugar from carbon rich environments. Our results demonstrate that the addition of salicin, one of the best substrates hydrolyzed by BcpE2, as supplement nutrient in the culture medium strongly induces the production of BcpE2 (Figure 3). In *S. venezuelae*, a similar situation has been reported where the production of two distinct β-glucosidases responds to either cellobiose or salicin. The enzyme induced by salicin hydrolyzes aryl β-glucosides but is poorly active on cellobiose while the other enzyme is highly active on – and induced by – cellobiose[31]. Due to their similarity in terms of catalytic specificity and responsiveness to either cellobiose or salicin, we speculate that these two enzymes of *S. venezuelae* are the orthologues of BglC (VNZ_32510) and BcpE2 (VNZ_10820), which share 70%, and 75% of identity with the proteins of *S. scabiei*, respectively. This finding further suggests that this type of GH3 β-glucosidase is not exclusive to *Streptomyces* species that are phytopathogenic

However, the orthologue of BcpE2 found in *S. scabiei* is most likely involved in additional functions related to pathogenicity as it is the case for BglC[10]. Indeed, importantly, BcpE2 could also compensate for the loss of a recently discovered function of BglC, i.e., the glucosidase activity on the phytoalexin scopolin[18]. *in vitro* TLC assay revealed that scopolin is the only natural plant compound to be degraded in common by BcpE2 and BglC. Interestingly, the aglycone moiety of scopolin is scopoletin, which has been described as a strong inhibitor of thaxtomin A biosynthesis in *S. scabiei*[19]. Since scopoletin and scopolin have been reported to be produced by plant roots or tubers especially in response to the application of thaxtomin A or under stress conditions such as pathogen infection [19,27–30], *S. scabiei* is very likely to encounter these molecules upon host colonization. It is therefore tempting to speculate that BglC and BcpE2 are both involved in the management of these compounds and that BcpE2 would be overproduced in the absence of BglC to take over the contribution of the latter.

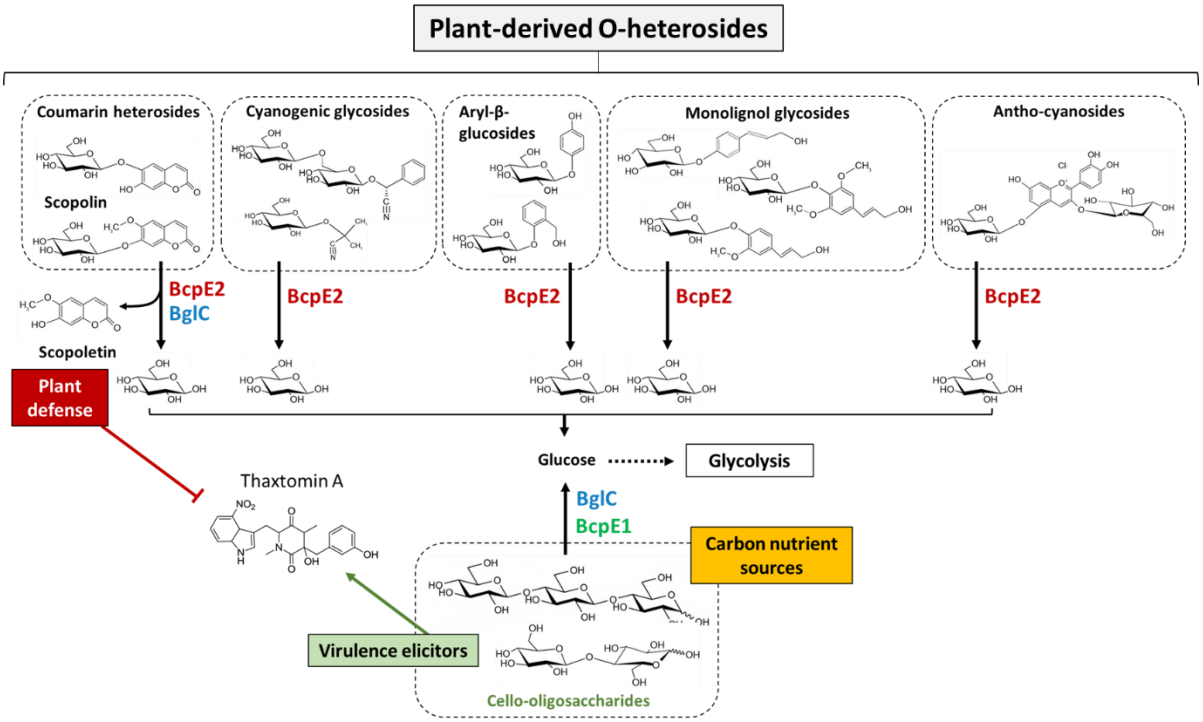

Figure 5. BcpE1 and BcpE2-mediated enzymatic compensation for the loss of BglC. BcpE1 (green) can compensate the activity of BglC (blue) by generating glucose from the hydrolysis of cello-oligosaccharides cellobiose and cellotriose. BcpE2 (red) can also fuel glycolysis by removing glucose from multiple plant heterosides. In addition, BcpE2 can also compensate the role of BglC in plant defense mechanism by displaying a substrate inhibition kinetic profile on scopolin, thereby generating the potent thaxtomin A production inhibitor scopoletin.

Perspectives

An important question that remains to be answered is "*what are the environmental triggers that induce the expression of bcpE2"*. In other words, "*does the substrate promiscuity of BcpE2 correlate with a mechanism of expression control of bcpE2 sensitive to multiple and dissimilar compounds?"*. We are currently seeking for the transcription factor that controls the expression of *bcpE2* orthologues in streptomycetes. Will this transcription factor be able to sense the presence of multiple substrates of BcpE2 or instead only few structurally similar substrates that would somehow witness for the possible presence of plant-heterosides? The answer to this question is crucial for properly understanding the role of this versatile enzyme. In addition, inactivation of orthologues of *bcpE2* in other model streptomycetes should provide further insight into the importance of this promiscuous enzyme and explain the success of these filamentous bacteria in colonizing plant-derived organic soils.

Materials and methods

Strains, chemicals, and culture conditions

Two strains of *Escherichia coli* were used in the present work: (i) DH5α for routine molecular biology applications, and (ii) BL21(DE3) Rosetta™ (Novagen) for heterologous proteins production. Both *E. coli* strains were cultured in LB (BD Difco LB broth) medium supplemented with the appropriate antibiotics (kanamycin (50 µg/mL), chloramphenicol (25 µg/mL)). *Streptomyces scabiei* 87-22 was routinely cultured at 28°C. Tryptic Soy Broth (TSB, Sigma-Aldrich, 30 g/L) was used for liquid pre-cultures. The modified TDM (thaxtomin defined medium [14], (Johnson et al., 2007)) minimal medium was prepared as described in (Jourdan et al., 2018) and after autoclaving were supplemented with filter-sterilized carbon sources. The substrates used in this study were purchased from Carbosynth (Cellobiose, Amygdalin, Linamarin, Xylobiose, Laminaribiose, Gentiobiose, Salicin, Arbutin, and Syringin), or from Sigma-Aldrich (4-Nitrophenyl-β-D-glucopyranoside (pNPβG), Esculin, Cyanin chloride, 4-Methylumbelliferyl β-D-glucopyranoside (4-MUG), Coniferin (Abietin), and p-Coumaryl alcohol 4-O-glucoside).

Heterologous production of His_6-tagged proteins and purification

BcpE2-His_6 and His_6-BglC were produced in *E. coli* BL21(DE3) Rosetta™ transformed with plasmids pBDF004 and pSAJ022, respectively, and purified by nickel affinity chromatography as already described in [10,20]. The pure proteins were stored at -20 °C and used in HEPES buffer (50 mM, pH 7.5).

Determination of the pH and temperature optima of BcpE2-His_6

The β-glucosidase activity was typically determined by the degradation of 4-Nitrophenyl-β-D-glucopyranoside (pNPβG). 95 µL of a determined BcpE2-His_6 concentration diluted in HEPES buffer (50 mM, pH 7.5) were mixed with 5 µL of pNPβG (20 mM). After incubation at 25°C, the reaction was stopped by the addition of 100 µL of Na_2CO_3 (2 M). The release of *para*-nitrophenol was monitored by measuring the absorbance at 405 nm with a TECAN Infinite® 200 PRO. The temperature optimum was determined by varying the incubation temperature from 5 to 60 degrees Celsius with 5°C increments. The pH optimum was determined by varying the pH of the reaction with the use of 3 distinct buffers, *i.e.*, (i) MES buffer (50 mM) for pH ranging between 5.0 and 6.5, (ii) HEPES buffer (50 mM) for pH ranging between 7.0 and 8.5, and (iii) CHES buffer (50 mM) for pH ranging between 9.0 and 10.0. The measured activity

was reported to the maximal value obtained in each experiment which was set to 100%. The results are presented in supplementary Figure S3.

TLC for hydrolysis of cello-oligosaccharides

Semi-quantitative substrate degradation was assessed by thin layer chromatography (TLC). Reactions were carried out with the BcpE2-His_6 and His_6-BglC enzymes (1 µM) and the substrates (5 mM) in HEPES 50 mM pH 7.5 at 40°C for 10 min. At the end of the reaction, the mixture was incubated for 5 min in a boiling water bath to inactivate the enzyme. 1-µL samples of the inactivated reaction mixtures were spotted next to undigested standards on aluminum-backed TLC plates (Silica gel Matrix, Sigma-Aldrich) and thoroughly dried. The protocol, adapted from [32], consisted in eluting the loaded TLC plate in a TLC chamber filled with an elution buffer (Chloroform – Methanol – Acetic acid – Water (50:50:15:5 (v/v))). After air-drying the eluted plate, sulfuric acid (5%) in ethanol was sprayed onto the TLC plate and the excess liquid was drained. The revelation was conducted by heating the TLC plate on a hot plate.

Determination of kinetic parameters for BcpE2-His_6

The hydrolysis of non-chromogenic substrates for β-glucosidases was followed by glucose quantification, either by HPLC (see HPLC quantification of glucose) or with the D-Glucose HK Assay kit (Megazyme) following the microplate procedure. BcpE2-His_6 was mixed with the substrate at variable concentrations in HEPES 50 mM pH 7.5, and the incubation was conducted at 40°C for 4 min. The reaction was terminated by a 5-min incubation in a boiling water bath. At least 10 concentrations – if possible distributed around the K_m value – were tested in triplicate for each substrate to estimate initial velocity values. The obtained data – initial velocity (V_i, mM/min) in function of substrate concentration ([S], mM) – were fitted to the Henri-Michaelis-Menten equation $V_i = (V_{max}*[S])/(K_m + [S])$ using the GraphPad Prism (version 9.2.0) software. K_m (mM), V_{max} (mM/min), k_{cat} (s^{-1}) and the specificity constant (k_{cat}/K_m ($mM^{-1}s^{-1}$)) were determined for each substrate. Substrate inhibition constants were also determined with GraphPad Prism following the equation $V_i =(V_{max}*[S])/(K_m + [S]*(1+[S]/K_i))$.

HPLC quantification of glucose

Glucose quantification was performed on a Waters HPLC device composed of a Separation Module (e2695) and a Refractive Index (RI) Detector (2414) set at 50°C. 35 µL of glucose-containing samples from terminated reactions were injected on a Aminex HPX-87P (Bio-Rad) column (300 x 7.8 mm) placed in an oven at 80°C. An isocratic flow of milli-Q water was

conducted for 20 min at a flow rate of 0.6 mL/min. The RI detector was set on channel 410 and sensitivity 256, and the measurements were expressed in RI Units (RIU). The peak areas associated with glucose (R_t = 11.6 min) were integrated and converted into glucose concentrations based on a linear standard curve ranging from 0.1 ng/µL to 400 ng/µL following the equation: y = 644.09x – 3170.4 (y being the Peak area (µV*sec) and x being the glucose amount (ng) / 10 µL injected).

Crystallization and structure determination of BcpE2-His$_6$

BcpE2 was concentrated to 17.4 mg/mL in HEPES 50 mM pH 7.5 and crystallized using the sitting-drop vapor diffusion method. 0.2 µL of protein was mixed with 0.2 µL of precipitant solution (methylpentane-2,4-diol (MPD) 45%, Tris-HCl 0.1M pH 8.5 and 0.2M ammonium acetate) and crystals grew at room temperature. The crystals were transferred into a cryoprotectant solution containing 50% MPD and 50% polyethylene glycol 400 before flash-freezing in a liquid nitrogen bath. Diffraction data were collected at the Soleil Synchrotron Proxima 2a beamline (Paris). Data were integrated and scaled using XDS (X-ray Detector Software [33]). Initial phases were obtained by molecular replacement using the structure of DesR from *S. venezuelae* as a search model (PDB code 4I3G [22]) using Phaser[34]. The structure was built with Coot (Crystallographic object-oriented toolkit [35]) and refined with BUSTER refine[36]). The figures were prepared using PyMOL (The PyMOL Molecular Graphics System, Version 2.4.1 Enhanced for Mac OS X, Schrödinger, LLC.).

Computational tools

The structure-based alignment was built using the MultAlin-ESPript (3.0) combined tool [37,38] using the structure of BcpE2 (PDB code: 7PPJ) as reference for positioning secondary structure elements. The sequences of seven additional characterized GH3 enzymes were included in the alignment (Table S2).

The phylogenetic tree was constructed using the phylogeny.fr tool with the "One Click" mode[39]. The repertoire of characterized bacterial and eukaryotic GH3 proteins was obtained from the CAZy database (accessed on August 27th 2021) and the amino acid sequences obtained were subsequently used in a BLASTp analysis against BcpE2 to select the appropriate proteins for the phylogenetic analysis (Table S2).

The search for PA14 domains was carried out using the MOTIF Search tool (Genome.jp) on the characterized bacterial and eukaryotic GH3 proteins from the CAZy database. The scan for motifs included the Pfam and NCBI-CDD databases in which the pfam07691 or 400161 and

214807 PSSM-Ids were searched for, respectively. In addition, the ScanProsite tool (Expasy) was used on the same amino acid sequences searching for the PA14 (PROSITE entry: PS51820) motif. A manual inspection was conducted to search for the presence of a PA14 domain in the closest GH3s (compared to BcpE2) that were not selected by the search tool. This inspection consisted in a comparison of the predicted secondary structures in the appropriate region of the proteins.

Determination of the intracellular β-glucosidase activity

Anion exchange chromatography (AXC) to obtain fractions of intracellular β-glucosidases was performed similarly to the method described in [20] with protein extracts prepared from cultures of *S. scabiei* 87-22 in TDM supplemented with salicin (0.1%) and/or cellobiose (0.1%). Briefly, 48-hours pre-cultures in TSB were washed twice in TDM without carbon source. After resuspension of the mycelium in the conditions described above, the culture was carried out for 7.5 hours at 28°C. After centrifugation, the mycelium pellet was resuspended in HEPES buffer (50 mM, pH 7.5) and disrupted with an Avestin Emulsiflex C3 homogenizer (3 lysis cycles). The soluble fraction was obtained by centrifugation of the lysed cell suspension and filtering (0.22 µm cut-off) of the supernatant. Using an NGC Quest 10 (Bio-rad) and a HiTrap™ Q HP column (GE healthcare), protein fractions were generated by elution with a linear NaCl gradient (0 to 1 M). The β-glucosidase activity of each fraction was determined by the standard assay using pNPβG as substrate (as described in [10,20]) and reported to the estimated protein content (Abs_{280nm}) of the fraction. These relative activities were then normalized to the maximal activity observed in the second peak (corresponding to BcpE2) of the TDM + cellobiose condition.

Targeted proteomics analysis

Collection of fractions by anion exchange chromatography and subsequent liquid chromatography – multiple reaction monitoring (LC-MRM) to monitor the relative abundance of BglC and BcpE2 in protein fractions was performed as previously described [20,40], and detailed in the supplementary file Table S3.

Acknowledgements

The work of Be.D. was supported by an aspirant grant from the FNRS (grant 1.A618.18), and a FNRS grant "Crédit de recherche" (grant CDR/OL J.0158.21) to S.R. Ba.D. was supported by a Bijzonder Onderzoeksfonds (BOF, grant 01B08915)-basic equipment from the Ghent University special research funds. F.K. and S.R. are research and senior-research associates of

522 the FRS-FNRS (Brussels, Belgium), respectively. We are very grateful to Sebastien Santini
523 (CNRS/AMU IGS UMR7256) and the PACA Bioinfo platform (supported by IBISA) for the
524 availability and management of the phylogeny.fr website, and to the assistance and support of
525 the team of beamline proxima 2a at the Soleil synchrotron.

526

References

1. Hodgson, D. A. Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv Microb Physiol* **42**, 47–238 (2000).

2. Book, A. J. *et al.* Evolution of High Cellulolytic Activity in Symbiotic *Streptomyces* through Selection of Expanded Gene Content and Coordinated Gene Expression. *PLoS Biol* **14**, e1002475 (2016).

3. Lewin, G. R. *et al.* Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. *Annu Rev Microbiol* **70**, 235–254 (2016).

4. Chater, K. F., Biró, S., Lee, K. J., Palmer, T. & Schrempf, H. The complex extracellular biology of *Streptomyces*. *FEMS Microbiol Rev* **34**, 171–198 (2010).

5. Lakhundi, S., Siddiqui, R. & Khan, N. A. Cellulose degradation: a therapeutic strategy in the improved treatment of Acanthamoeba infections. *Parasit Vectors* **8**, 23 (2015).

6. Takasuka, T. E., Book, A. J., Lewin, G. R., Currie, C. R. & Fox, B. G. Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*. *Sci Rep* **3**, 1030 (2013).

7. Jourdan, S. *et al.* The CebE/MsiK Transporter is a Doorway to the Cello-oligosaccharide-mediated Induction of *Streptomyces scabies* Pathogenicity. *Sci Rep* **6**, 27144 (2016).

8. Schlösser, A., Jantos, J., Hackmann, K. & Schrempf, H. Characterization of the binding protein-dependent cellobiose and cellotriose transport system of the cellulose degrader *Streptomyces reticuli*. *Appl Environ Microbiol* **65**, 2636–2643 (1999).

9. Schlösser, A., Kampers, T. & Schrempf, H. The *Streptomyces* ATP-binding component MsiK assists in cellobiose and maltose transport. *J Bacteriol* **179**, 2092–2095 (1997).

10. Jourdan, S. *et al.* Contribution of the β-glucosidase BglC to the onset of the pathogenic lifestyle of *Streptomyces scabies*. *Mol Plant Pathol* **19**, 1480–1490 (2018).

11. Spiridonov, N. A. & Wilson, D. B. Cloning and biochemical characterization of BglC, a beta-glucosidase from the cellulolytic actinomycete *Thermobifida fusca*. *Curr Microbiol* **42**, 295–301 (2001).

12. Wach, M. J., Krasnoff, S. B., Loria, R. & Gibson, D. M. Effect of carbohydrates on the production of thaxtomin A by *Streptomyces acidiscabies*. *Arch Microbiol* **188**, 81–88 (2007).

13. Jourdan, S., Francis, I. M., Deflandre, B., Loria, R. & Rigali, S. Tracking the Subtle Mutations Driving Host Sensing by the Plant Pathogen *Streptomyces scabies*. *mSphere* **2**, e00367-16 (2017).

14. Johnson, E. G., Joshi, M. V., Gibson, D. M. & Loria, R. Cello-oligosaccharides released from host plants induce pathogenicity in scab-causing *Streptomyces* species. *Physiological and Molecular Plant Pathology* **71**, 18–25 (2007).

15. Francis, I. M., Jourdan, S., Fanara, S., Loria, R. & Rigali, S. The cellobiose sensor CebR is the gatekeeper of *Streptomyces scabies* pathogenicity. *mBio* **6**, e02018 (2015).

16. Planckaert, S. *et al.* Proteomic Response to Thaxtomin Phytotoxin Elicitor Cellobiose and to Deletion of Cellulose Utilization Regulator CebR in *Streptomyces scabies*. *J Proteome Res* **17**, 3837–3852 (2018).

17. Deflandre, B. *et al.* The virulome of *Streptomyces scabiei* in response to cello-oligosaccharide elicitors. *Microbial Genomics* **8**, 000760.

18. Deflandre, B. & Rigali, S. Old Enzyme, New Role: The β-Glucosidase BglC of *Streptomyces scabiei* Interferes with the Plant Defense Mechanism by Hydrolyzing Scopolin. *Biophysica* **2**, 1–7 (2021).

19. Lerat, S. *et al. Streptomyces scabiei* and its toxin thaxtomin A induce scopoletin biosynthesis in tobacco and Arabidopsis thaliana. *Plant Cell Rep* **28**, 1895–1903 (2009).

20. Deflandre, B. *et al.* Deletion of *bglC* triggers a genetic compensation response by awakening the expression of alternative beta-glucosidase. *Biochim Biophys Acta Gene Regul Mech* **1863**, 194615 (2020).

21. Yoshida, E. *et al.* Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β-glucosidase from *Kluyveromyces marxianus*. *Biochem J* **431**, 39–49 (2010).

22. Zmudka, M. W., Thoden, J. B. & Holden, H. M. The structure of DesR from *Streptomyces venezuelae*, a β-glucosidase involved in macrolide activation. *Protein Sci* **22**, 883–892 (2013).

23. Yoshida, E. *et al.* Purification, crystallization and preliminary X-ray analysis of beta-glucosidase from *Kluyveromyces marxianus* NBRC1777. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **65**, 1190–1192 (2009).

24. Mahasenan, K. V. *et al.* Catalytic Cycle of Glycoside Hydrolase BglX from *Pseudomonas aeruginosa* and Its Implications for Biofilm Formation. *ACS Chem Biol* **15**, 189–196 (2020).

25. Yoshida, E. *et al.* Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β-glucosidase from *Kluyveromyces marxianus*. *Biochemical Journal* **431**, 39–49 (2010).

26. Hrmova, M. *et al.* Structural Rationale for Low-Nanomolar Binding of Transition State Mimics to a Family GH3 β-d-Glucan Glucohydrolase from Barley. *Biochemistry* **44**, 16529–16539 (2005).

27. Bednarek, P., Schneider, B., Svatos, A., Oldham, N. J. & Hahlbrock, K. Structural complexity, differential response to infection, and tissue specificity of indolic and phenylpropanoid secondary metabolism in Arabidopsis roots. *Plant Physiol* **138**, 1058–1070 (2005).

28. Clarke, D. D. The accumulation of scopolin in potato tissue in response to infection. *Physiological Plant Pathology* **3**, 347–358 (1973).

29. Nolte, P., Secor, G. A., Gudmestad, N. C. & Henningson, P. J. Detection and identification of fluorescent compounds in potato tuber tissue with corky patch syndrome. *American Potato Journal* **70**, 649–666 (1993).

30. Ward, J. L., Baker, J. M., Llewellyn, A. M., Hawkins, N. D. & Beale, M. H. Metabolomic analysis of Arabidopsis reveals hemiterpenoid glycosides as products of a nitrate ion-regulated, carbon flux overflow. *Proc Natl Acad Sci U S A* **108**, 10762–10767 (2011).

31. Chatterjee, S. & Vining, L. C. Glucose suppression of beta-glucosidase activity in a chloramphenicol-producing strain of *Streptomyces venezuelae*. *Can J Microbiol* **28**, 593–599 (1982).

32. Gao, J. & Wakarchuk, W. Characterization of Five β-Glycoside Hydrolases from *Cellulomonas fimi* ATCC 484. *Journal of Bacteriology* **196**, 4103–4110 (2014).

33. Kabsch, W. XDS. *Acta Crystallographica Section D Biological Crystallography* **66**, 125–132 (2010).

34. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Cryst* **40**, 658–674 (2007).

35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).

36. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr* **60**, 2210–2221 (2004).

37. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881–10890 (1988).

38. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324 (2014).

39. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**, W465-469 (2008).

40. Planckaert, S. *et al.* Identification of Novel Rotihibin Analogues in *Streptomyces scabies*, Including Discovery of Its Biosynthetic Gene Cluster. *Microbiol Spectr* **9**, e0057121 (2021).

Supplementary data

Supplementary Table S1

Table S1 - Data collection and refinement statistics

	BcpE2 (PDB code7PJJ)
Data Collection:	
Wavelength	0.98010
Space group	P 3_1 2 1
a, b, c (Å)	109.62, 109.62, 164.18
α, β, γ (°)	90, 90, 120
Resolution range (Å)[a]	47.4 – 3.09 (3.27 – 3.09)
Rmerge (%)[a]	27.3 (222)
<I>/<σI> [a]	6.9 (0.9)
Completeness (%)[a]	99.3 (96.6)
Redundancy [a]	6.7 (6.4)
CC 1/2[a]	0.991 (0.301)
Refinement:	
Resolution range (Å)[a]	47.4 – 3.09 (3.2 – 3.09)
No. of unique reflections[a]	21414 (1981)
R_{work} (%)[a]	22.3 (33.6)
R_{free} (%)[a]	27.0 (35.5)
No. Atoms	
Protein	5904
Solvent	14
RMS deviations from ideal stereochemistry	
Bond lengths (Å)	0.009
Bond angles (°)	1.00
Mean B factor (Å^2)	
Protein	90.9
Solvent	94.3
Ramachandran plot:	
Favored region (%)	92.7
Allowed regions (%)	6.4
Outlier regions (%)	0.9

[a] Numbers in parenthesis refer to the highest resolution shell.

Supplementary Figure S1. Structure-based sequence alignment of BcpE2 and its closest characterized GH3 enzymes.

The two catalytic residues – Asp239 and Glu582 – are strictly conserved in all homologous GH3s. The active site of BcpE2 also includes Asp59, Arg127, and three additional aromatic AAs, namely Tyr207, Trp240, and Phe499 which are generally conserved. Asp59 and Arg127 show perfect conservation among all GH3 enzymes displayed in Figure S1, and Trp240 was systematically found next to the catalytic aspartate. Tyr207 was found in all sequences except in BglI of *Schwanniomyces etchellsii* which presented a leucine residue instead. The important Phe499 residue was shared by most GH3-family enzymes considered but sometimes aligned at an adjacent position as in BglI of *S. etchellsii* and KmBglI of *K. marxianus*. Only DesR from *S. venezuelae* and Bgl3B from *Cellulomonas fimi* did not display any phenylalanine residue in this subregion.

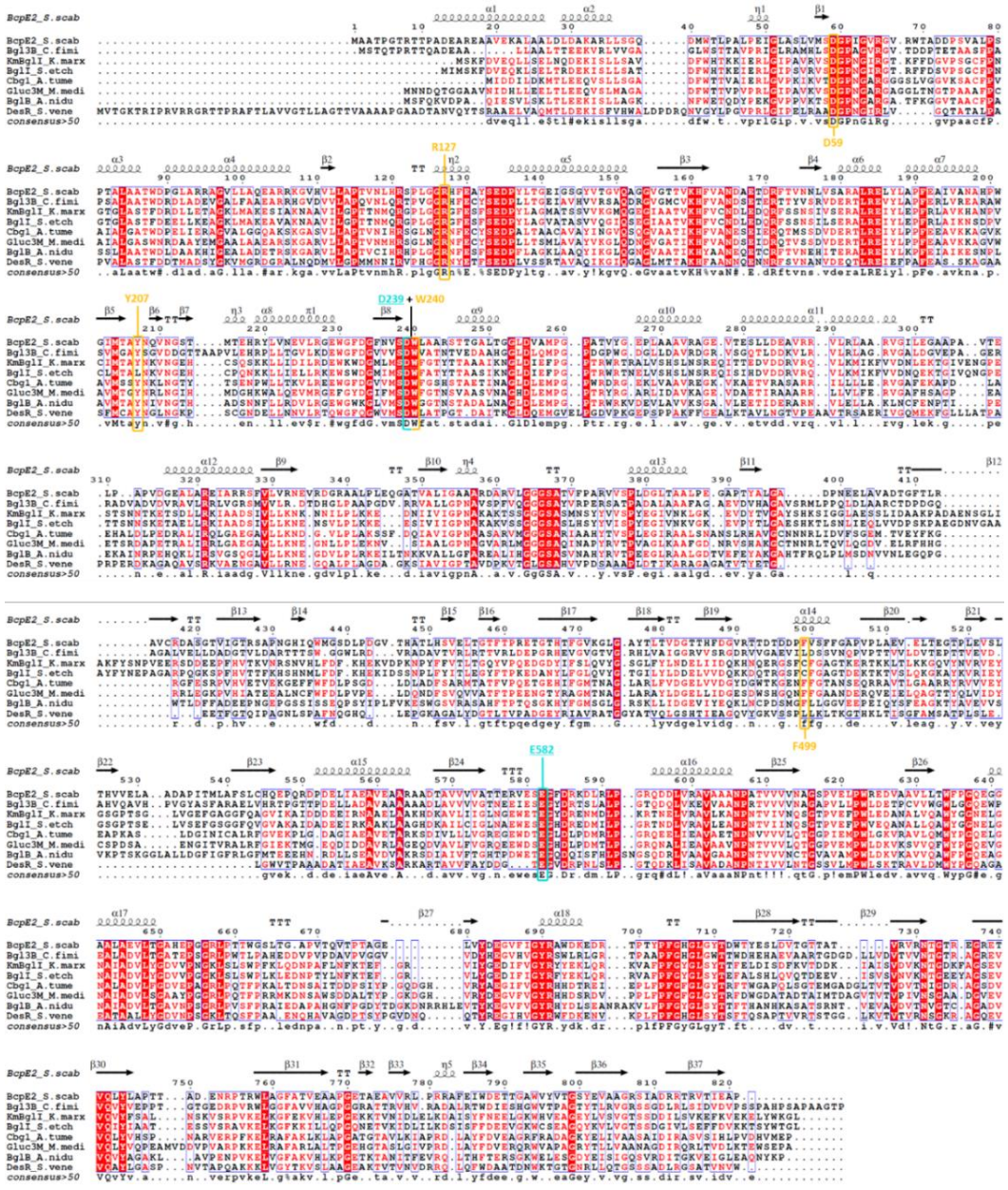

Figure S1. Structure-based sequence alignment of BcpE2 and its closest characterized GH3 enzymes. The residues involved in the active site are indicated in yellow and the catalytic residues, also in the active site, are highlighted in cyan. The red color on residues either indicate perfect conservation (red background) or biochemical similarity (red letter). Blue frames highlight conserved regions for which a consensus motif is found. The "!", "#", and "%" symbols in the consensus sequence shows the conservation of branched-chain, acidic or amide, and hydrophobic amino acids (AAs), respectively.

Supplementary Figure S2. Phylogeny of BcpE2 and its closest characterized GH3-family β-glucosidases

BcpE2 of *S. scabiei* is part of the clade that includes Bgl3B of *Cellulomonas fimi*, Cba from *Cellulomonas biazotea*, Gluc3M of *Martelella mediterranea*, and Cbg1 from *Agrobacterium tumefaciens*. Bgl3B of *Cellulomonas fimi*, and Cba from *Cellulomonas biazotea* are the closest partially characterized non-*Streptomyces* actinobacterial GH3s and have been reported to be active on gentiobiose and cellobiose as natural substrates, respectively (Gao & Wakarchuk, 2014; W. K. R. Wong et al., 1998). BcpE2 also shares a common ancestor with Gluc3M, a cold-active and alkali-stable β-Glucosidase from the Gram-negative rhizobiaceae *Martelella mediterranea*, and with the β-Glucosidase Cbg1 from *Agrobacterium tumefaciens*. Cbg1 is the second closest described homologue though it only has 39% and 55% of AA identity and similarity, respectively, compared to BcpE2. Interestingly, Cbg1 is involved in the virulence induction of some *A. tumefaciens* strains targeting Douglas fir trees (*Pseudotsuga menziesii*) by hydrolyzing the monolignol glucoside coniferin (Castle et al., 1992; Morris & Morris, 1990). Moreover, Cbg1, like BcpE2, is not properly active on cellobiose (Castle et al., 1992; Deflandre et al., 2020) but was instead reported to be active on salicin and arbutin. Beyond the synthetic substrates pNPβG and 4-Nitrophenyl β-D-galactopyranoside (pNPβGal), Gluc3M also exhibited significant activities toward salicin, and konjac powder (glucomannan) (Mao et al., 2010). As Cbg1 hydrolyzes the monolignol glucoside coniferin (Castle et al., 1992), the two other monolignol glucosides syringin and p-coumaryl alcohol 4-O-glucoside were also included as candidate substrates. The coumarin heteroside esculin, which is hydrolyzed by the fungal enzyme - KmBglI GH3 enzyme (Yoshida et al., 2010) - belonging to the phylogenetic clade adjacent to the clade of BcpE2 (Figure S2) was selected as well. The coumarin heteroside scopoline was included because it is a new substrate hydrolyzed by BglC of *S. scabiei* (Deflandre et al. 2022). BcpE2 of *S. scabiei* 87-22 was also suggested by KEGG pathway as candidate beta-glucosidase possibly involved in cyanoamino acid metabolism (https://www.genome.jp/kegg-bin/show_pathway?scb00460+SCAB_64101). Two cyanogenic glucosides were thus also tested as possible targets of BcpE2, *i.e.*, amygdalin and linamarin. The glycone moiety of amygdalin is the disaccharide gentiobiose and its aglycone part is mandelonitrile, the cyanohydrin of benzaldehyde; linamarin is a glucoside of acetone cyanohydrin. These cyanide-bearing heterosides are plant phytoanticipins whose activation requires the action of a β-glucosidase to release the toxic aglycone moiety from the glycosidic residue. Finally, the natural plant anthocyanidin pigment cyanin, and the synthetic aryl-β-glucoside substrate 4-methylumbelliferyl-β-D-glucoside (4-MUG) were also tested.

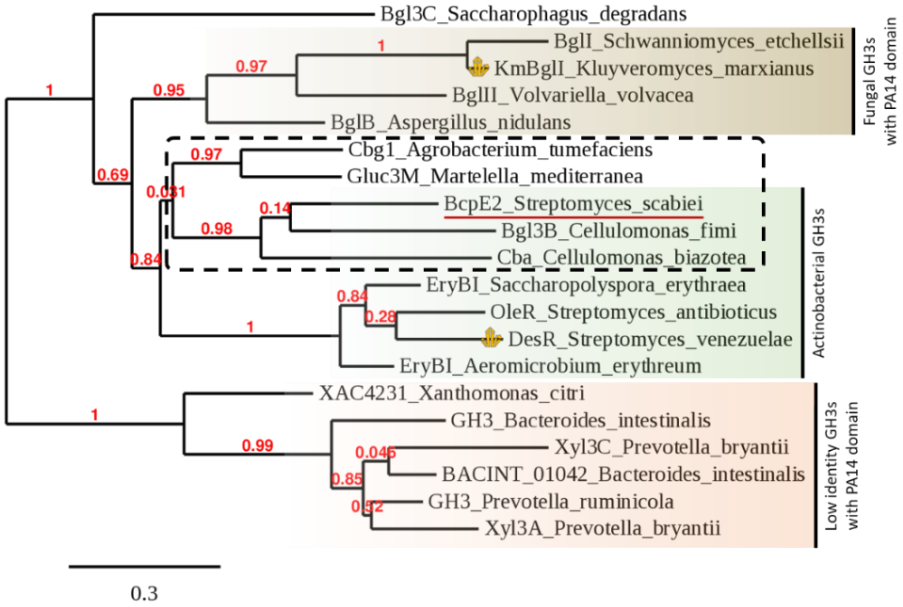

Figure S2. Phylogeny of BcpE2 and its closest characterized GH3-family β-glucosidases in order to identify possible candidate substrates of BcpE2. The 19 closest characterized bacterial and fungal GH3-family beta-glucosidase have been selected based on BLASTp score and/or high query coverage. The dotted square delineates the clade with the closest characterized homologues of BcpE2.

Supplementary Figure S3. Determination of the pH and temperature optima of BcpE2.

The enzyme was obtained as BcpE2-His_6 by heterologous production in *E. coli* and subsequent purification by Nickel affinity chromatography as previously described (Deflandre et al., 2020). BcpE2-His_6 shows neutral and mesophilic pH and temperature parameters, with optimal activities displayed around pH 6.5-7.5 and 35-40°C, respectively Figure S3). The optimal pH window was relatively narrow, since the enzyme displayed about 60% of its maximal activity at pH values only 0.5 above or below the 6.5-7.5 range (Figure S3, left panel). This pH range is in line with the cytoplasmic compartmentalization of BcpE2 in *S. scabiei*. Below 20°C and above 45°C, the activity rapidly dropped under 50% of the optimal activity (Figure S3, right panel). See the materials and methods section for the detailed protocol.

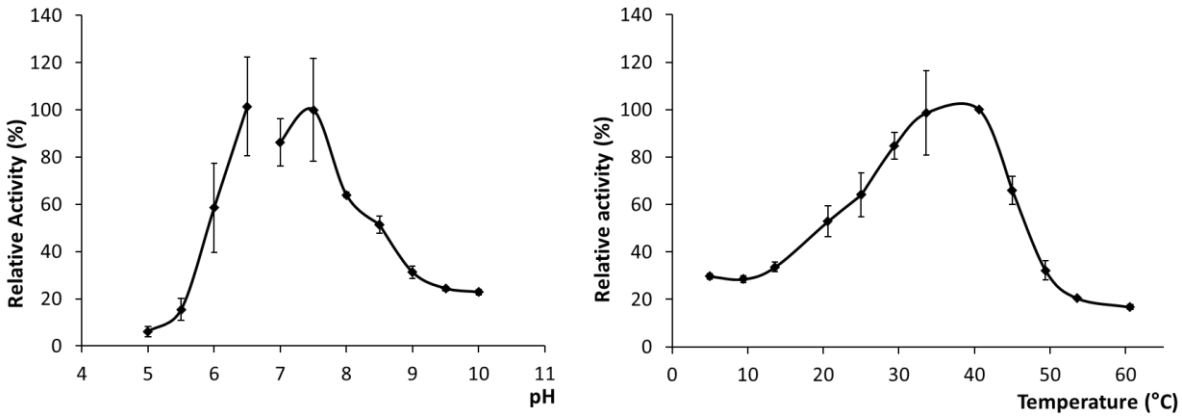

Figure S3. Determination of the pH and temperature optima of BcpE2. Relative activity assays with pNPβG as substrate, normalized to the maximal value measured in each assay. The influence of the pH (left panel) was assessed by increments of 0.5 from pH 5 to pH 10, and the temperature (right panel) was assessed by increments of 5°C from 5 to 60°C.

Supplementary Table S2

Table S2 - Characteristics of GH3-family proteins homologous to BcpE2 (used for structure-based alignment and phylogenetic tree construction)

Organism	GH3 name	Query cov. (%)	Identity (%)	PA14 domain	NCBI protein ID	Ref.
Streptomyces Phaeolivaceus	BcpE2-like	99	86	Yes	WP_152168420.1	-
Streptomyces diastatochromogenes	BcpE2-like	99	80	Yes	WP_094217992.1	-
Streptomyces venezuelae	BcpE2-like	99	75	Yes	WP_150183929.1	-
Streptomyces rimosus	BcpE2-like	96	66	Yes	WP_033029078.1	-
Streptomyces coelicolor	BcpE2-like?	96	47	Not found	WP_011031033.1	-
Cellulomonas fimi	Bgl3B	96	44	Not found	AEE44608.1	(Gao & Wakarchuk, 2014)
Agrobacterium tumefaciens	Cbg1	97	39	Yes	AAA22082.1	(Castle et al., 1992)
Martelella mediterranea	Gluc3M	95	39	Yes	ADC53302.1	(Mao et al., 2010)
Aspergillus nidulans	BglB	97	36	Yes	EAA65189.1	(Bauer et al., 2006)
Schwanniomyces etchellsii	BglI	96	34	Yes	ACF93471.1	(Pandey & Mishra, 1997)
Cellulomonas biazotea	Cba	96	41	Not found	AAC38196.1	(W. K. R. Wong et al., 1998)
Kluyveromyces marxianus	KmBglI	96	33	Yes	ACY95404.1	(Yoshida et al., 2010)
Volvariella volvacaea	BglII	97	35	Yes	AAG59831.1	(X. Li et al., 2005)
Saccharophagus degradans	Bgl3C	98	32	Yes	ABD81934.1	(H. Zhang et al., 2011)
Saccharopolyspora erythraea	EryBI	97	35	Yes	CAA74702.1	(Jakeman & Sadeghi-Khomami, 2011)
Streptomyces antibioticus	OleR	92	37	Yes	AAC12650.1	(Quiros et al., 1998)
Aeromicrobium erythreum	EryBI	94	37	Yes	AAU93797.1	(Reeves et al., 2008)
Streptomyces venezuelae	DesR	95	35	Yes	ACR54627.1	(Zmudka et al., 2013)
Spirochaeta thermophila	STHERM_c14600	76	44	Not found	ADN02400.1	(Angelov et al., 2011)
Dictyoglomus turgidum	Dtur_0219	75	41	Not found	ACK41548.1	(Kim et al., 2011)
Herpetosiphon aurantiacus	HaGH03	76	43	Not found	ABX04075.1	(R.-F. Wang et al., 2015)
Hungateiclostridium thermocellum	BglB	74	40	Not found	ABN52488.1	(Romaniec et al., 1993)
Paenibacillus xylanilyticus	BglA	79	38	Not found	AFC68969.1	(D.-J. Park et al., 2013)
Thermoclostridium stercorarium	Bgl3Z	78	39	Not found	CAB08072.1	(Adelsberger et al., 2004)
Acetivibrio thermocellus	Bgl3B	74	39	Not found	CAA33665.1	(Gräbnitz et al., 1989)
Xhantomonas citri	XAC4231	89	32	Yes	AAM39066.1	(Vieira et al., 2021)
Cellulomonas fimi	Bgl3C	77	44	Not found	AEE47485.1	(Gao & Wakarchuk, 2014)
Paenibacillus sp. TS12	GH3	78	38	Not found	BAC16750.1	(Sumida et al., 2002)
Uncultured bacterium	AS-Esc6	77	41	Not found	AHG23300.1	(Biver et al., 2014)

Prevotella bryantii	Xyl3A	97	26	Yes	ADD92014.1	(Dodd et al., 2010)
Prevotella ruminicola	GH3	98	27	Yes	ACN78955.1	(Dodd et al., 2009)
Bacteroides intestinalis	GH3	95	26	Yes	EDV05842.1	(Hong et al., 2014)
Bacteroides intestinalis	BACINT_01042	92	27	Yes	ZP_03013483.1	(Pereira et al., 2021)
Prevotella bryantii	Xyl3C	89	27	Yes	ADD92016.1	(Dodd et al., 2010)

695

696

697

Supplementary Table S3

Transitions selected for MRM relative quantification of BglC/BcpE2 β-glucosidases

Peptide	Precursor ion mass (m/z)	CE (V)	Fragment y-ion mass (m/z)
BglC			
LVDELLAK	450.7737^{++}	16	787.456^{+} 688.3876^{+} 573.3606^{+}
TDPVASLR	429.7376^{++}	15	642.3933^{+} 545.3406^{+} 446.2722^{+}
BcpE2			
AGVLLAQEAR	514.2984^{++}	18	800.4625^{+} 687.3784^{+} 574.2944^{+}
DASGTVIGTR	488.7565^{++}	17	703.4097^{+} 646.3883^{+} 545.3406^{+}
AADTAVVVVATTER	701.8805^{++}	25	973.5677^{+} 874.4993^{+} 775.4308^{+}
BSA			
AEFVEVTK	461.7477^{++}	16	722.4083^{+} 575.3399^{+} 476.2715^{+}
QTALVELLK	507.8133^{++}	18	785.5131^{+} 714.476^{+} 601.3919^{+}
Phos B			
VFADYEEYVK	631.8006^{++}	22	945.42^{+} 830.3931^{+} 667.3297^{+}
LLSYVDDEAFIR	720.8721^{++}	26	964.4734^{+} 865.405^{+} 750.3781^{+}
VLYPNDNFFEGK	721.8512^{++}	26	1067.479^{+} 970.4265^{+} 856.3836^{+} 741.3566^{+}

Detailed protocol of the targeted proteomic approach.

Fractions collected from the anion exchange chromatography (AXC) were subjected to trichloroacetic acid (TCA) precipitation. The AXC fractions were mixed with 100% (w/v) TCA solution (Sigma-Aldrich) (4:1), the proteins were precipitated overnight at 4°C and collected by centrifugation (16,000g; 30 min, 4°C). The protein pellet was washed twice with ice-cold acetone and solubilized in 50 mM ammonium bicarbonate containing 2 M urea. Protein concentrations were estimated by Bradford's method (Coomassie Plus Protein Assay kit, Pierce). Dried protein (10 µg) was solubilized in 2 M urea/50mM NH_4HCO_3 spiked with bovine serum albumin (MS-grade protein standard) (1:250) (Thermo-Scientific), and denatured by heating to 80°C for 10 min. The solution was subsequently reduced with 5 mM dithiothreitol (Sigma-Aldrich) for 10 min at 60°C and alkylated with 15 mM iodoacetamide for 20 min at RT in the dark before digestion with trypsin (Promega, Madison, USA)

712 overnight at 37°C (1:50 w/w). Acidified digested samples were desalted using OMIX C18 pipette tips
713 (Agilent). The desalted peptides were dried under vacuum and dissolved in 0.1% formic acid, 3% ACN
714 and 10 fmol/µl phosphorylase B (Hi3 Phos B Standard, Waters).

715 For BglC and BcpE2 identification by MRM, samples (0.5 µg) were injected onto an ultraperformance
716 liquid chromatography (UPLC) M-class system (Waters) and trapped on a 300 µm x 50 mm, 5 µm, 100
717 Å Acquity UPLC M-Class Symmetry C18 Trap Colum (Waters). The washing step on the trap column
718 was performed for 2 min with 3% B at a flow rate of 15 µL/min, with solvent A 0.1% HCOOH in H_2O
719 (Biosolve) and solvent B 0.1% HCOOH in ACN (Biosolve). Subsequently, the peptides were separated
720 on a 150 µm x 100 mm, 1.8 µm HSS T3, iKey separation device in 10 min at a flow rate of 2 µL/min
721 using a linear solvent B gradient (3-50%). The separated peptides were introduced into the IonKey
722 source coupled to a Waters Xevo TQ-S triple-quadrupole mass spectrometer for detection of the analytes
723 in the positive-ion mode (ESI+). The MRM mode with transitions of selected proteotypic peptides at a
724 set cone voltage and different collision energies for each precursor, was used for detection,
725 normalization (BSA) and MS performance check (Phos B) (Figure S3). Capillary voltage was set at
726 3.5kV, the cone voltage at 35V and the source temperature at 120°C. In the collision cell, argon was
727 introduced at a flow rate of 0.15 mL/min. Data were acquired with the developed MRM mode
728 (MassLynx 4.1), subsequently uploaded into Skyline (Pino et al., 2020) for data analysis, and, subjected
729 to a Savitsky-Golay Smoothing, the total area under the curve (AUC) for each peptide was calculated
730 and normalized to BSA.

731

732

733