Chapter 3

Changing the specificity of DIMBOA-glucosidase (maize ß-glucosidase) to that of dhurrinase (sorghum ß-glucosidase) by domain swapping within the C-terminal region

Muzaffer Cicek[a], David Blanchard[a] David R. Bevan[b], and Asim Esen[a]*

[a] *Department of Biology, Virginia Tech, Blacksburg, Virginia 24061-0406*
[b] *Department of Biochemistry, Virginia Tech, Blacksburg, Virginia 24061-0308*

This chapter has been submitted to PNAS

ABSTRACT

The maize ß-glucosidase isozyme Glu1 hydrolyzes a broad spectrum of substrates in addition to its natural substrate DIMBOAGlc, while the sorghum ß-glucosidase isozyme Dhr1 hydrolyzes exclusively its natural substrate dhurrin. To study the mechanism of substrate specificity further, five chimeric ß-glucosidases were constructed by replacing peptide sequences within the C-terminal region of Glu1 with the homologous peptide sequences of Dhr1 where the two enzymes differ by 4 to 22 amino acid substitutions, depending on the length of the swapped regions. Replacement of the extreme 47-amino-acid-long C-terminal domain of Glu1 with the homologous 53-amino-acid-long domain of Dhr1 produced a Glu1/Dhr1 chimeric enzyme, which hydrolyzes substrates that are hydrolyzed by both parental enzymes, including *p*-hydroxy-(S)-mandelonitrile-ß-D-glucoside (dhurrin), which is not hydrolyzed by Glu1. Additional domain-swapping within the 47-amino-acid-long C-terminal domain of Glu1 showed that replacing the peptide ^{466}FAGFTERY^{473} of Glu1 with the homologous peptide ^{462}SSGYTERF^{469} of Dhr1, or replacing the peptide ^{481}NNNCTRYMKE^{490} in Glu1 with the homologous peptide ^{477}ENGCERTMKR^{486} of Dhr1 was sufficient to confer to Glu1 the ability to hydrolyze dhurrin. Data from the activities of various Glu1/Dhr1 chimeras, sequence comparisons, and homology modeling suggest that the Dhr1-specific S462-463, Y465 and F469 substitutions play a key role in dhurrin hydrolysis and the substrate specificity difference between Glu1 and Dhr1.

* To whom correspondence should be addressed

INTRODUCTION

ß-glucosidases (ß-D-glucoside glucohydrolase, EC 3.2.1.21) have been the focus of much research recently because of the key roles in a number of biological processes and potential biotechnological applications. Plant ß-glucosidases have been implicated in defense against pests (1, 2, 3, 4), phytohormone activation (5, 6, 7, 8, 9), lignification (10), and cell wall catabolism (11, 12). ß-glucosidases may have different physiological functions in different plant species, depending on the nature of the aglycone moiety of substrates.

In maize, ß-glucosidase (DIMBOA-glucosidase) occurs in two different forms (isozymes), Glu1 and Glu2. The cDNAs corresponding to both isozymes were cloned and sequenced and the deduced protein products were found to share 90% sequence identity (2). The cDNA corresponding to the sorghum ß-glucosidase isozyme Dhr1 was also cloned and sequenced in our laboratory; it shares ~70% sequence identity with each of the two maize isozymes. The catalytically active form of both maize and sorghum ß-glucosidases is a 120 kD homodimer or its multimers. The primary structures of maize and sorghum ß-glucosidases contain the peptide motifs TFNEP and ITENG (underlined in Fig. 2), which are shown to be highly conserved and make up the catalytic site in all Family 1 ß-glycosidases (13, 14, 15). Furthermore, the 3-D structures of three Family 1 ß-glucosidases (white clover linamarase, white mustard myrosinase, and *Bacillus polymyxa* ß-glucosidase) have recently been solved using crystals of the enzyme-glycosyl complexes (16, 17, 18). In these three cases, the residues of the TFNEP and I/VTENG motifs or their equivalents were found to form a pocket or crater-shaped active site (19). The two catalytic glutamic acids (i.e., the nucleophile and the acid-base catalyst) were positioned within the active site at appropriate distances (2.5-3Å) on opposite sides of the glycosidic bond (Fig. 4).

There are two fundamental questions about ß-glucosidase-catalyzed reactions: (1) how do ß-glucosidases catalyze the hydrolysis of the ß-glycosidic bond between two glycone residues (e.g., cellobiose and other ß-linked oligosaccharides) or that between glucose and an aryl or alkyl aglycone (e.g., many naturally occurring substrates in plants)? (2) What determines substrate specificity, including the site and mechanism of aglycone binding? Much progress has been made in understanding the mechanism of catalysis and defining the roles of specific amino acids (the two catalytic glutamates) within the active site. However, there is virtually no information as to how ß-glucosidases recognize and

interact with the substrates, specifically the aglycone moiety, which is the basis of tremendous diversity in natural substrates and is responsible for subtle substrate specificity differences among ß-glucosidases. The maize ß-glucosidase isozyme Glu1 and its sorghum homologue Dhr1 provide an ideal model system to address questions related to substrate specificity because these enzymes represent extremes in substrate specificity. While Dhr1 hydrolyzes only its natural substrate dhurrin, Glu1 hydrolyzes a broad spectrum of artificial and natural substrates in addition to its natural substrate DIMBOAGlc but does not hydrolyze dhurrin (Fig. 1).

The first attempt to investigate substrate specificity using chimeric ß-glucosidases was made by Singh and Hayashi (20), who exchanged the C-terminal 58-amino-acid-long domain of prokaryotic ß-glucosidase from *Celvibrio gilvus* with the C-terminal 60-amino-acid-long domain of *Agrobacterium tumefaciens* ß-glucosidase. Hayaqshi and Hoal (cited in 20) also showed that the deletion of 70 amino acids from the C-terminal region of *Celvibrio gilvus* ß-glucosidase leads to complete loss of activity. These investigators concluded that the C-terminal region of the enzyme played a major role in determining substrate specificity and catalytic activity. The importance of specific enzyme domains in substrate specificity and catalytic efficiency was also shown in a number of other enzyme chimeras produced from two enzymes that differ with respect to substrate specificity.

In subsequent experiments, it was shown that exchanging a small N-terminal portion between two rice α-amylase isozymes (Amy A and Amy 3D) resulted in a chimeric enzyme (Amy A/3D) that shows high activity on both soluble starch and oligosaccharides while parental enzymes have high activity only on either soluble starch (Amy A) or oligosaccharides (Amy 3D)(21). Cahoon et al. (22) investigated substrate specificity or substrate recognition of plant acyl-ACP desaturases by site-directed mutagenesis. Replacement of specific amino acids resulted in modified acyl-ACP desaturases that show broad specificity towards naturally-occurring fatty acids. Based on the crystallographic structure of active site acyl-ACP desaturases, replacement of two amino acids (L118F/P179I) in the active site converted Δ^{a}-18:0-ACP into16:0-ACP desaturase. Hopfner et al. (23) combined the subdomains from coagulation factor Xa (fXa) and digestive protease trypsin. Even though the overall sequence identity of these proteins is only 40%, the resulting chimeric protein containing the subdomains from both S1 proteases formed a novel active site, exhibiting new substrate specificity.

The purpose of the studies described in this paper was to further characterize the mechanism of substrate (aglycone) recognition and binding in ß-glucosidases using two

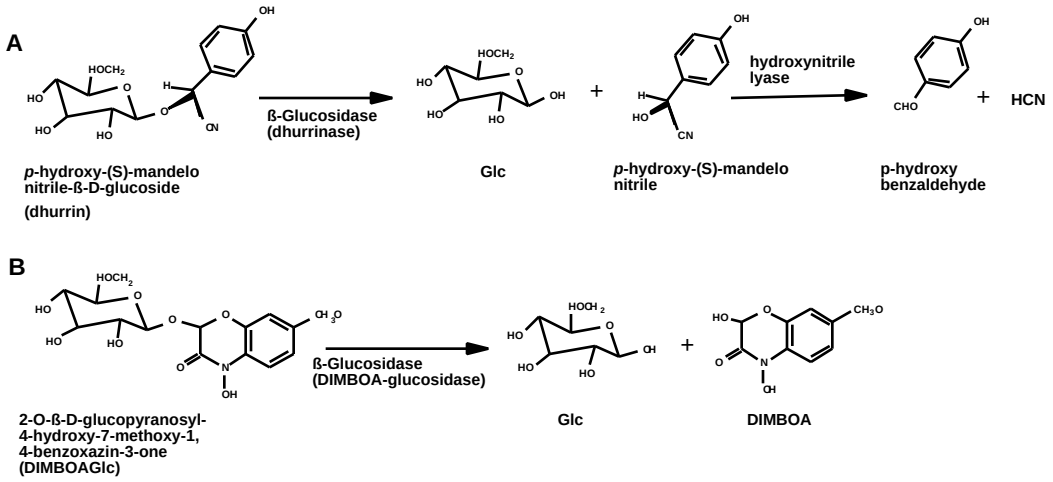

FIG. 1. Hydrolysis of natural substrates by sorghum and maize ß-glucosidases. A, hydrolysis of dhurrin by dhurrinase from sorghum and production of HCN. B, hydrolysis of DIMBOAGlc by DIMBOA-glucosidase from maize.

plant enzymes, the maize isozyme Glu1 and sorghum isozyme Dhr1, as model systems. To this end, we have constructed five chimeric enzymes between maize and sorghum ß-glucosidases by domain-swapping. Target domains were selected based on amino acid sequence comparisons and analysis of modeled 3D structures. We demonstrate that the maize Glu1 isozyme gains the ability to hydrolyze dhurrin when the C-terminal 47 amino acid-long region or specific subdomains within this region are replaced by corresponding Dhr1 region.

MATERIAL AND METHODS

Construction of Chimeric ß-Glucosidases. The first step towards understanding the basis of substrate specificity in maize and sorghum ß-glucosidases was to construct cDNAs encoding chimeric enzymes by domain swapping. Wild type and chimeric enzymes were expressed in *E. coli* and characterized with special emphasis on substrate specificity. Since we had already cloned and expressed cDNAs encoding maize Glu1 and Glu2 isozymes as well as the sorghum Dhr1 isozyme in *E. coli* (24), the construction and expression of chimeric cDNAs using these wild type parental templates were straightforward. The criteria for chimeric constructs was that the swapped region includes one or more amino acid substitutions within the C-terminal domain that map to or around the active center in the modeled 3-D structures of Glu1 and Dhr1. Chimeric cDNAs were constructed by using the PCR based recombination technique of overlap extension and the high fidelity thermostable Turbo® *pfu* polymerase (Stratagene, CA) as described (24, 25). Sequences of oligonucleotide primers used in PCR and the peptide regions from which they are derived are shown in Table 1 and Fig. 2A and B, respectively. The constructs were made so as to encode chimeric enzymes Glu1/Dhr1, where the enzyme before the slash contributed the N-terminal region and the one after the slash contributed the C-terminal region (Fig. 2). As a first step in defining the domain determining substrate specificity of Glu1 and Dhr1, Chimera 2 (abbreviated hereafter as Chim 2) was constructed by replacing the extreme 47-amino-acid-long C-terminal region (amino acids 466-512) of Glu1 with the corresponding 53-amino acid-long region (amino acids 462-514) of Dhr1. The 5'-portion of chimeric cDNA was amplified on the *glu1* template using the primer pair T7 (sense) and ß-glu-101 (antisense), while the 3'-portion was amplified on the *dhr1*

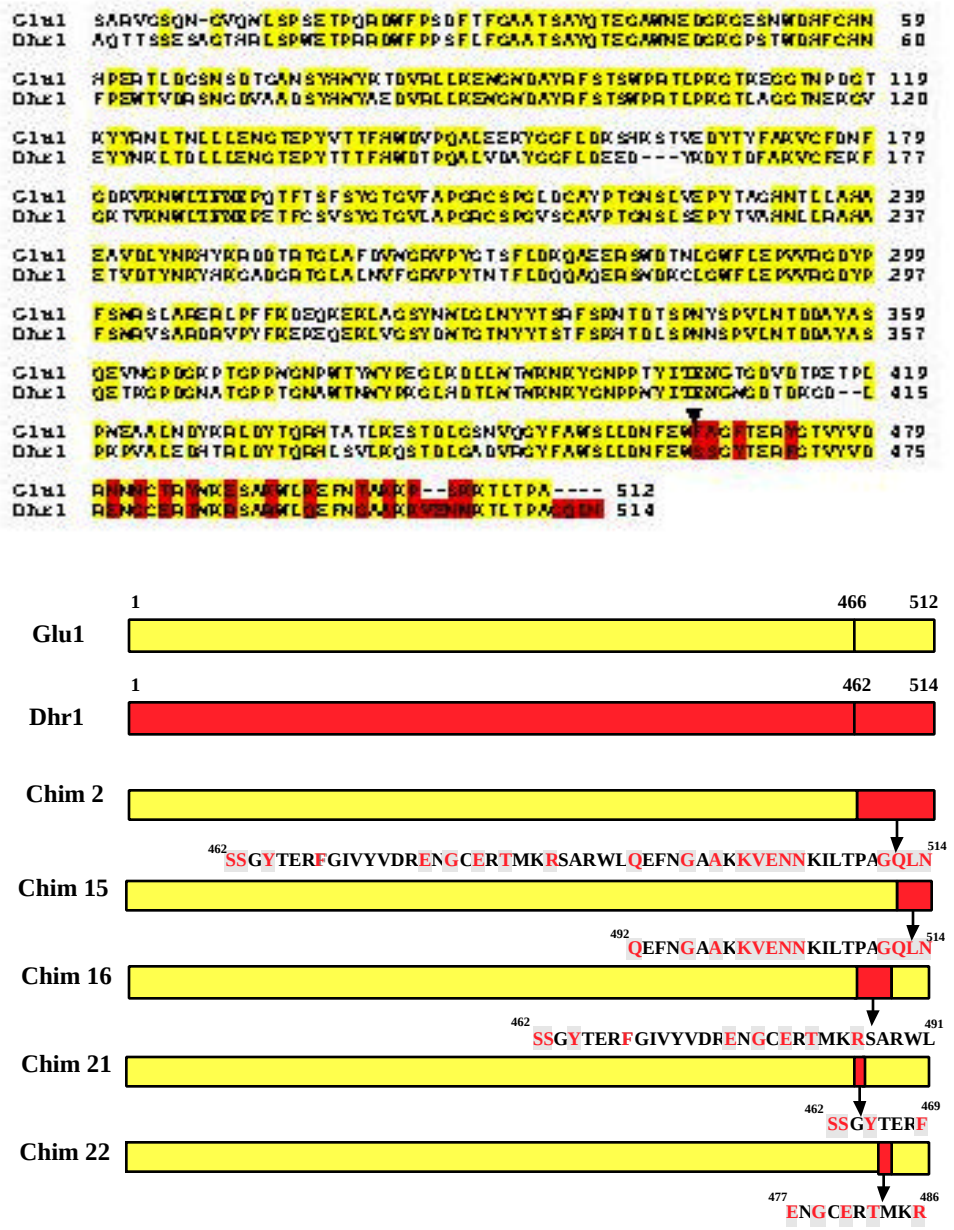

FIG. 2. **A**, Alignment of the sequences of maize ß-glucosidase (Glu1) and sorghum dhurrinase (Dhr1). The identical regions shown in yellow background. The arrow indicates the swapped C-terminal region in Chim 2 and invariant amino acids between Glu1 and Dhr1 are shown in red background. The bolded peptide motifs TFNEP and ITENG are highly conserved in all Family 1 ß-glycosidases; they contain the two key catalytic glutamic acids and also form the glycone-binding site within the active site. The invariant GYFAWSL peptide region is junction site to construct the Chim 2. **B**, diagrammatic representation of structures of wild type parental Glu1 and Dhr1 and five Glu1/Dhr1 chimeric ß-glucosidases generated by domain-swapping between Glu1 and Dhr1 isozymes. The lengths of exchanged domains ranged from 8 (Chim 21) to 53 (Chim 2), and they were all expressed in *E. coli* and characterized with respect to substrate specificity. The sites at which Glu1 and Dhr1 sequences differ are indicated in red color.

template using the primer pair ß-glu-100 (sense) and T7term (antisense). Junction primers ß-glu-100 (sense) and ß-glu-101 (antisense) were derived from the region of cDNA encoding the peptide sequence GWFAWSL, which is invariant in Glu1 and Dhr1 (Fig. 2A). The two PCR fragments were gel-purified, denatured, mixed, annealed (by overlapping the complementary ends that contain primers ß-glu-100 and ß-glu-101 sequences) and extended to obtain the full-length chimeric Glu1/Dhr1 cDNA coding sequence. The full-length chimeric cDNA was amplified by using the vector specific primer pair of T7 (sense) and $T7_{ter}$ (antisense) on the overlap extended template. The resulting PCR product was purified, digested with *Nhe*1 and *Xho*1, and cloned into the expression plasmid pET21a. The construction of other chimeric cDNAs (*chim 15, chim 16, chim 21,* and *chim 22*) followed the procedure described for *chim 2,* except the primer pairs used for PCR (Table 1). Among the five chimeric enzymes produced, the sizes of the swapped peptides varied from 8 to 53, all coming from the C-terminal fragment 462-514 of Dhr1 (Table 1).

Expression and Purification. Wild type and chimeric ß-glucosidases were produced in *E. coli* pLyS cells (F*ompT hsdSP* r_B-m_B- *gal dcm*) under the control of the T7 RNA polymerase promoter in the expression plasmid pET-21a (Novagene) as described by Cicek and Esen (24). The cell lysis and protein extraction procedure was performed as described (24). For purification, ß-glucosidase was precipitated from crude cell extracts with a 35 to 65 % ammonium sulfate (AS) cut. The precipitate was dissolved in 50 mM Na-acetate buffer, pH 5.0 and centrifuged at 18 000xg for 30 min. The supernatant was adjusted to a final concentration of 0.5 M AS and centrifuged at 18 000xg for 30 min. Then the supernatant was applied to a ToyoPearl-butyl 650M (TasoHaas, Montgomeryville, PA) hydrophobic interaction chromatography column (1.5 X 14 cm). The column was washed to base line absorbance with 0.5 M AS in buffer and eluted with approximately five bed volumes of a reverse salt gradient of 0.5 M to 0.1 M AS in 50 mM Na-Acetate buffer, pH 5.0. The resulting fractions were assayed for ß-glucosidase activity using the artificial substrate *p*NPG or the natural substrates DIMBOAGlc or dhurrin. The fractions with activity were pooled based on purity as judged by SDS-PAGE. The pooled fractions were adjusted to 0.5 M AS and were rechromatographed on a Toyopearl-phenyl 650M column as described above. The purification protocol was the same for both Glu1 and chimeras except slight changes of the AS concentration in the binding and elution steps of hydrophobic interaction chromatography. For Dhr1 purification, 0.05 M phosphate

buffer, pH 7.0, was used in all steps because Dhr1 was not stable at pH 5.0 (acetate buffer). Again, the fractions that had ß-glucosidase activity were checked for purity by SDS-PAGE, pooled and concentrated approximately 10-fold using a 30K cut-off spin column (Gelman Sciences). The concentrated enzymes were then used for kinetic analysis with special emphasis on substrate specificity.

Enzyme Assays. For activity assays in native PAGE gels, the purified parental and chimeric enzymes were electrophoresed into 6% alkaline gels to obtain zymograms using the fluorogenic substrate 4-methylumbelliferyl-ß-D-glucoside (4MUGlc) and the chromogenic substrate 6-bromo-2-naphthyl-ß-D-glucoside (6BNGlc) as described (26, 27).

Kinetic parameters, K_m and k_{cat} (V_{max}/E_t) for parental and chimeric ß-glucosidases were determined with varying substrate concentration from 0.098 to 16.66 mM in citrate-phosphate buffer, pH 5.8, for the artificial substrates *p*NPG and *o*NPG. Protein content was adjusted to appropriate concentration according to the Bradford assay (BioRad) for all activity assays. K_m and k_{cat} values were determined based on the amount of *p*-nitrophenol and *o*-nitrophenol released from *p*NPG and *o*NPG, respectively. Each assay was performed in quadruplicate in a microtiter plate in a total volume of 140 µl containing 70 µl substrate and 70 µl enzyme solution. The reaction mixture was incubated at room temperature (~23 ˚C) for 10 min and the reaction was stopped by adding 70 µl 0.4 M Na_2CO_3. The absorbance was read in a Dynatech microplate reader at 410 nm. The K_m and k_{cat} values for the natural substrates DIMBOAGlc and dhurrin were determined by the Peroxidase-Glucose-Oxidase (PGO) coupled reaction (28). Fifty µl aliquots containing 0.031 to 2.0 mM DIMBOAGlc or dhurrin in phosphate-citrate buffer, pH 5.8, were placed in quadruplicate in wells of a microtiter plate followed by 50 µl of diluted ß-glucosidase solution, 50 µl PGO enzymes and 50 µL ABTS (2,2'-azinobis-3-ethylbenzthiazolinesulfonic acid). The reaction mixture was incubated at 37 °C for 30 min, and the absorbance was read in the microplate reader at 410 nm.

Inhibition experiments were performed at the substrate concentration range of 1 to 8 K_m in the case of *p*NPG for Glu1, between 1 and 10 K_m in the case of dhurrin for Dhr1. Inhibitors were applied at four different concentration, 0.5-2 K_m. The type of inhibition was determined and calculated by Line-Weaver-Burk linearization using Enzyme Kinetics ® software (Trinity Software, Fl).