

## Article

# Modulating Glycoside Hydrolase Activity between Hydrolysis and Transfer Reactions Using an Evolutionary Approach

Rodrigo A. Arreola-Barroso, Alexey Llopiz , Leticia Olvera and Gloria Saab-Rincón \* 

Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca 62271, Mexico; rodrigo.arreola@ibt.unam.mx (R.A.A.-B.); alexey.llopiz@ibt.unam.mx (A.L.); leticia.olvera@ibt.unam.mx (L.O.)

\* Correspondence: gloria.saab@ibt.unam.mx

**Abstract:** The proteins within the CAZy glycoside hydrolase family GH13 catalyze the hydrolysis of polysaccharides such as glycogen and starch. Many of these enzymes also perform transglycosylation in various degrees, ranging from secondary to predominant reactions. Identifying structural determinants associated with GH13 family reaction specificity is key to modifying and designing enzymes with increased specificity towards individual reactions for further applications in industrial, chemical, or biomedical fields. This work proposes a computational approach for decoding the determinant structural composition defining the reaction specificity. This method is based on the conservation of coevolving residues in spatial contacts associated with reaction specificity. To evaluate the algorithm, mutants of  $\alpha$ -amylase (*TmAmyA*) and glucanotransferase (*TmGTase*) from *Thermotoga maritima* were constructed to modify the reaction specificity. The K98P/D99A/H222Q variant from *TmAmyA* doubled the transglycosylation/hydrolysis (T/H) ratio while the M279N variant from *TmGTase* increased the hydrolysis/transglycosylation ratio five-fold. Molecular dynamic simulations of the variants indicated changes in flexibility that can account for the modified T/H ratio. An essential contribution of the presented computational approach is its capacity to identify residues outside of the active center that affect the reaction specificity.

**Keywords:** transglycosylation; hydrolysis; contact-residues; amylase; glucanotransferase; coevolution; enrichment-factor; specificity



**Citation:** Arreola-Barroso, R.A.; Llopiz, A.; Olvera, L.; Saab-Rincón, G. Modulating Glycoside Hydrolase Activity between Hydrolysis and Transfer Reactions Using an Evolutionary Approach. *Molecules* **2021**, *26*, 6586. <https://doi.org/10.3390/molecules26216586>

Academic Editor: Stefan Janecek

Received: 23 September 2021

Accepted: 28 October 2021

Published: 30 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Enzymes are accelerators of chemical reactions that occur in living cells, which also work in vitro, making their use in the laboratory, in medical applications, and in industry possible [1–3]. Tailoring an enzyme's ability to carry out specific reactions is one of the greatest challenges that must be met in order to move on to a more sustainable biocatalysis process [4]. In this sense, directed evolution has proven to be a valuable strategy for evolving functions, with the limitation of requiring extensive screening efforts, in order to find an improved biocatalyst [5,6]. De novo design has shown impressive improvements over the last two decades in the development of energy functions for directing the design of proteins [7–9]. However, the subtle changes that confer the necessary dynamics for catalysis have not yet been determined [10,11]. Last year, an enormous breakthrough was made in the implementation of artificial intelligence tools that predict the 3D structure of proteins. Alpha-Fold surpassed the performance obtained so far by any other structure prediction method in the CASP protein modeling competition using this approach. This year, the modeling of protein structures from humans and 20 other genomes through the use of artificial intelligence is in progress using one of the most powerful supercomputers [12,13].

The above-mentioned strategies have highlighted the importance of exploiting the structural and functional information accumulated through thousands of years of evolution. It is acknowledged that residues outside the catalytic site and their contacts can play essential roles in protein structure and function [14,15]. Exploration of these residues

is required to continue improving the functional facet of protein design. Nature has evolved new enzymes for novel reactions from pre-existing ones, giving rise to shared structures and sequence motifs in proteins with distinct functions [16–19]. Functional insights have already been attained by comparing features between evolutionarily related but functionally different enzymes. Although such analysis has focused on the catalytic site, we are starting to understand that the investigation should extend outside the protein core.

Directed evolution and massive gene sequencing of mutational libraries (e.g., deep sequencing) have highlighted the relevance of residues outside the catalytic site for protein function. In the case of deep sequencing, it has also revealed pairs of residues that are in contact, making the reconstruction of 3D protein structures possible by analyzing the effects of concurrent mutations on their activity [20,21]. Structure reconstruction based on deep sequencing data parallels the analysis of residue contacts, in which specific patterns characterize different protein families [22,23]. These patterns are frequent inputs for algorithms predicting 3D protein structures [13,24], identification of members belonging to functional families [25,26], the study of coevolution through protein domains [27], and reconstruction of novel sequences whose structure and function matches an already existing family [28].

A more direct attempt to exploit residue contacts in order to modify function is the analysis of residue covariation in a multiple sequence alignment. The amino acids identified as covariates are targets for mutagenesis, and thus, investigations have identified residues whose substitution impairs function [29,30]. The advent of Alpha-Fold makes structural models available like never before, with similar reliability to those obtained from X-ray crystal diffraction of many proteins for which we only have the amino acid sequence [13]. This fact encourages us to explore currently accessible X-ray structures to compare and identify the information we may be able to draw from Alpha-Fold.

The CAZy database (Carbohydrate Active Enzymes database [31,32]) glycoside hydrolase family 13 of proteins (GH13) shares a  $(\beta/\alpha)_8$  TIM barrel core as a catalytic domain and at least two different domains termed B and C [33]. This enzyme family includes enzymes acting on  $\alpha$ -(1→4) bonds in glucose polymers and oligomers that differ in their reaction specificity. In both subgroups, the enzymes break the O-glycosidic bond of polysaccharide chains in two parts. An intermediate is formed, with one sugar fragment covalently bound to the protein through the anomeric carbon of its reducing end [34]. This double-displacement mechanism ends with the transfer of the sugar moiety to an acceptor molecule, which defines the type of reaction that takes place. Those members with hydrolytic activity transfer the glycoside moiety to water, while others that are predominantly transferases, transfer it to another sugar [35–38].

The balance between transglycosylation and hydrolysis must rely on many factors inherent to the protein, such as its sequence, 3D structure, acid-base properties of its critical residues, hydrophobicity at its active site [39,40], flexibility, and dynamics. The change of the acid-base catalytic residue pKa during the reaction is crucial for both reactions. However, the pKa increase of the acid-base residue may be more relevant for the hydrolysis than for the transglycosylation reaction, since proton removal from a water molecule is energetically more demanding. Recently, Geronimo et al. [41] reported a molecular dynamic simulation at a constant pH for beta-glycosidase (bglc) from *Hypocrea jecorina*. Using this approach, the authors observed that the usual rearrangement of the active center (residues R169, Y204, and W237) is different if acceptor nucleophiles like cellobiose or glucose are present. They found that as a result, the pKa value for the catalytic residue (E441) is lower in the presence of these sugars than in that of water. Their results suggest that protein specificity towards hydrolysis or transglycosylation may be associated with protein dynamics and flexibility, which in turn, can be influenced by the presence or absence of acceptor nucleophiles. Additionally, many loops located in the vicinity of the active site play an essential role in stabilizing the transition state for many amylases [42].

Multiple reports describe the modification of amino acid residues for increasing the transglycosylation/hydrolysis (T/H) ratio of amylolytic enzymes [35,43–45]. The factors considered in the selection of amino acids for this purpose include the location of flexible regions associated with internal water transport [46,47], shifts in acid-base residue dynamics [47], and sidechain conformational changes of residues near the active center [48]. Some other investigations have intended to modify the specificity of the acceptor site toward diverse organic molecules [49] and have succeeded in changing the T/H ratio; nevertheless, the strategies explored have generally focused on punctual mutations near the active site identified by multiple sequence alignments (MSA) of a few glycosidases—presumably with high specificity towards hydrolytic or transfer reactions [43,45,50–52].

The relationship between coevolving residues and protein specificity has been evaluated in the aquaporin family in order to find the residues associated with water or glycerol transport [53]. Additionally, coevolving residues associated with protein specificity were evaluated in twelve families of proteins [54]. However, in many cases, the coevolution of residues is hard to detect because of (1) the presence of changing compensatory network mutations, (2) the significant dependence of covariations on evolutionary distances, (3) the number of proteins in an MSA, and (4) the quality of alignment in the coevolving residues' environment.

Here, we propose an algorithm for the differential analysis of contact patterns among evolutionarily related  $\alpha$ -glycosidases with two distinct reaction specificities. Contact maps, 2D representations of 3D structures, were used to compare the enrichment of each pair of residue contacts. The proposed approach has the advantage of reducing the dimensionality of the system [55,56] in order to identify the elements driving the specificity between hydrolysis and transglycosylation in the glycoside hydrolase family 13. Understanding the molecular determinants of protein specificity could contribute to the development of enzymes for glycosynthesis (e.g., adding a sugar moiety to an organic molecule) and design of enzymes with desired increases in hydrolytic or transglycosidic specificity. We used two model enzymes to validate our predictions:  $\alpha$ -amylase (*TmAmyA*) and glucanotransferase (*TmGTase*), from a hyperthermophilic bacterium *Thermotoga maritima*.

## 2. Results

We analyzed residue-residue contacts in 14 structures (Dataset 1, four transglycosidases and 10 hydrolases belonging to the GH13 family bound to acarbose, Table S1). The selection of structures was based on the availability of structures bound to acarbose, a transition state analog. It is important to mention that all proteins were monomeric to the best of our knowledge. Inclusive *TmGTase*, whose crystal structure suggests a dimeric protein, has been reported as being in an equilibrium of 90% monomers, 10% oligomers when it is in solution [57]. Thus, the possibility that the determinants of specificity detected in this way were due to the oligomeric interphase contribution was ruled out. After comparing the residue contacts of enzymes with transglycosidic activity against these hydrolytic reactions, we identified preferences between the groups for different amino acids when forming pairs in each residue-residue contact. These results agreed with the notion that enzymes work under selective pressure, and that residues coevolve to create the residue-residue contacts that maintain structure and function. We identified contacts in which some amino acids were frequently present in either hydrolases or transferases, and underrepresented in the other group. We measured and expressed these preferences using the enrichment factor ( $\Delta f_{aa}^{ij}$ ) described in the Methods section (Equations (1) and (2)).

### 2.1. Homology Model of *TmAmyA*

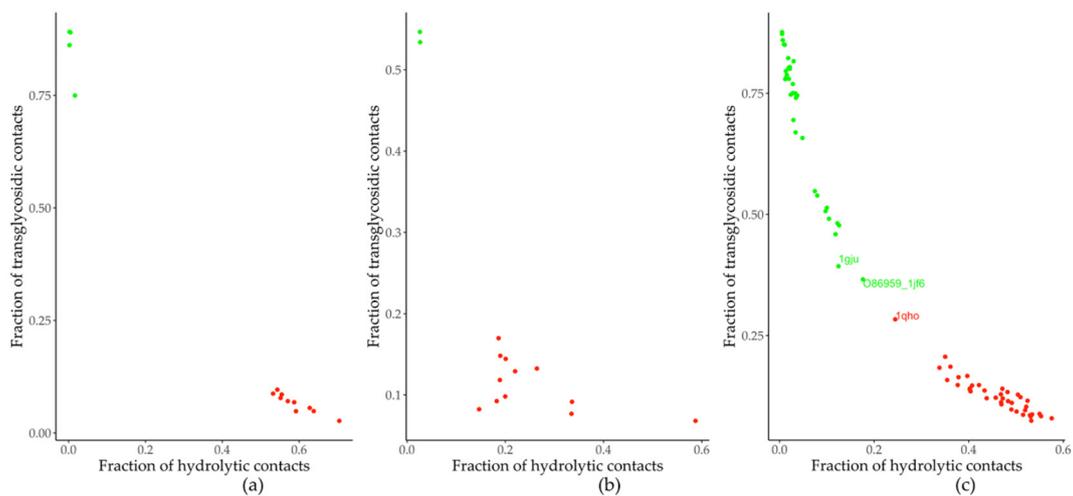
The homology model of *TmAmyA* was constructed using the crystallographic structure of the amylase from *Thermotoga petrophila* (PDB ID 5M99, resolution 1.96 Å) as a template. This model excludes thirty extra residues at the N-terminus of *TmAmyA* not present in the crystallized amylase from *Thermotoga petrophila*, and which do not belong to the core domains of GH13 enzymes. The 504 remaining residues have 98.4 % sequence identity,

showing only six substitutions. The “Structure assessment” tool [58] from the Swiss-Model server was used to validate this model. The QMEAN value was 1.06; a similar value was obtained for the structure used as the template. The MolProbity Score had a value of 2.96, and Ramachandran Favoured was 96.63%. Additionally, the validation with ProSA-web (Z-Score of  $-9.95$ ) had a value similar to other proteins of a similar size. VERIFY3D (99.21% of the residues had an average 3D-1D score  $\geq 0.2$ ) suggests that the model is of adequate quality for the study presented in this work.

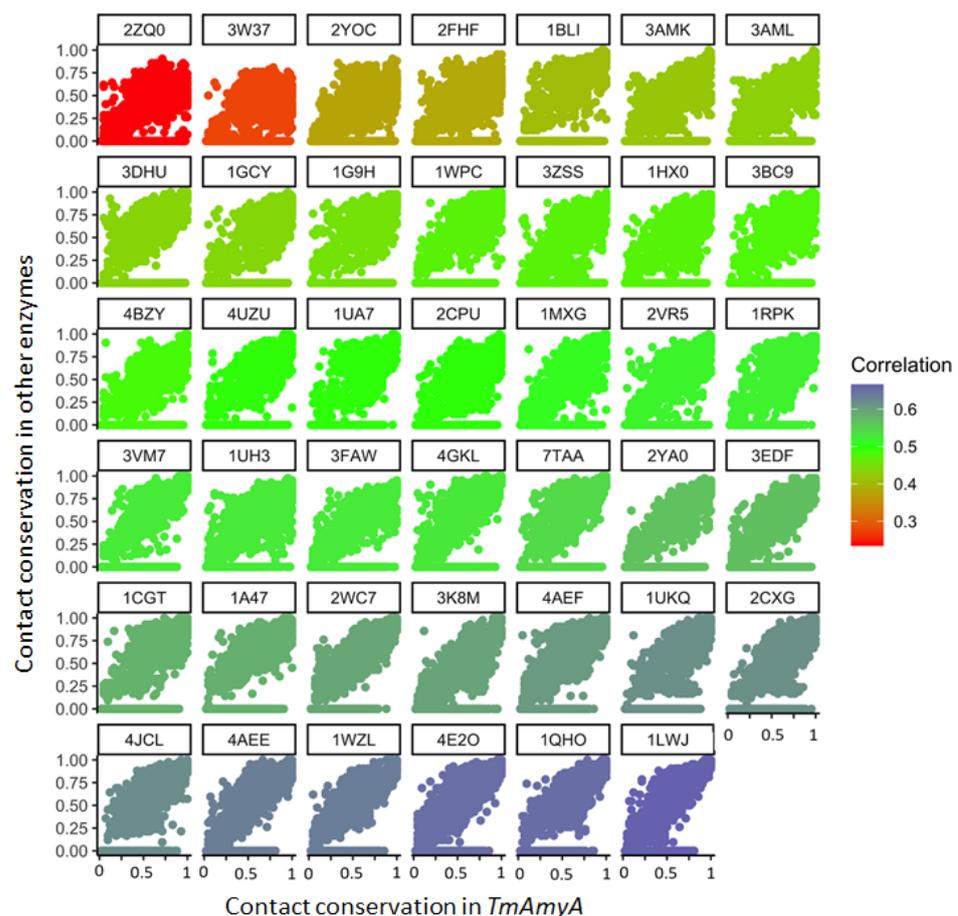
## 2.2. Protein Classification as Hydrolases or Transglycosidases

The residue contacts in *TmGTase* (PDB ID: 1LWJ) were compared with those of the structures within Dataset 1 and Dataset 2 (two transglycosidases and 12 hydrolases of the GH13 family not containing acarbose, different from Dataset 1; Table S2). The amino acids in all residue contacts shared with 1LWJ were qualified as hydrolytic if  $\Delta f_{aa}^{ij} < 0$  (the amino acid was more frequent in hydrolases) and transglycosidic if  $\Delta f_{aa}^{ij} > 0$  (the frequency of the amino acid was higher in the transferases). After qualifying all the amino acids in the residue contacts, contacts formed with a pair of residues deemed transglycosidic were designated transglycosidic; those formed by two hydrolytic residues, as hydrolytic. As expected, the relation of hydrolytic/transglycosidic contacts in each protein classified all the enzymes correctly as hydrolases or transferases in Dataset 1 (Figure 1a), which was used to calculate the  $\Delta f_{aa}^{ij}$  values by the algorithm. The  $\Delta f_{aa}^{ij}$  obtained from Dataset 1 was sufficient to classify all the enzymes in Dataset 2 according to their primary function (Figure 1b), despite its incomplete sampling of the sequence space, as suggested by the residues that remained unclassified using the obtained  $\Delta f_{aa}^{ij}$  (data not shown). The lack of complete information about the reaction specificity of some of the enzymes, and the fact that not all residues identified are involved in determining the reaction specificity, could account for this discrepancy. Nonetheless, the classification of enzymes based on residue-residue contacts seems to indicate the presence of subgroups whose transglycosidic/hydrolytic activity would be appealing to characterize in future studies to see if the classification correlates with their grouping in this graphic (Figure 1a,b). During the process of this work, new members of the GH13 family were added to the CAZy database. Addition of the new structures resulted in Dataset 4, which contained a more significant number of enzymes (31 transglycosidases and 40 hydrolases belonging to the GH13 family, comprising all the characterized enzymes active on 1,4- $\alpha$ -bonds according to the CAZy database, which includes all enzymes in Dataset 1 and Dataset 2), and where it was also possible to discriminate enzyme function using enrichment factors (Figure 1c). Still, the structure, corresponding to an  $\alpha$ -amylase from bacilli (PDB ID:1QHO) was separated from the rest of the  $\alpha$ -amylases, and was closest to the transglycosidases, while the transglycosidases closest to the hydrolases were not from bacilli. As many transglycosidases are of bacilli origin in Dataset 4 (Table S4), this suggests that the enrichment factors in Dataset 4 reflect the specificity and phylogenetic origin of the enzymes.

Measuring contact conservation derived a helpful parameter for phylogenetic analysis. The contact conservation score was calculated as the fractional number of times that each particular contact in a reference protein was present in all the enzymes in the dataset. We computed the contact conservation scores using different enzymes as the reference each time. The residue contact conservation score of each enzyme was plotted against its equivalent contact with other enzymes. The plot for each pair of enzymes produced a correlation coefficient that indicates the degree of evolutionary relatedness between the enzyme pair. Figure 2 shows the correlation between all the proteins in the Dataset 3 (39 enzymes belonging to the GH13 family, one to the GH97 family and one belonging to the GH31) and *TmAmyA*. The lowest correlations correspond to the enzymes identified with the PDB ID: 2ZQ0 (family (GH97) and 3W37 (family GH31)—an expected result since these enzymes did not belong to the GH13 family. Thus, this correlation value could be used as a contact similarity coefficient to discriminate between members of different glycoside hydrolase families.

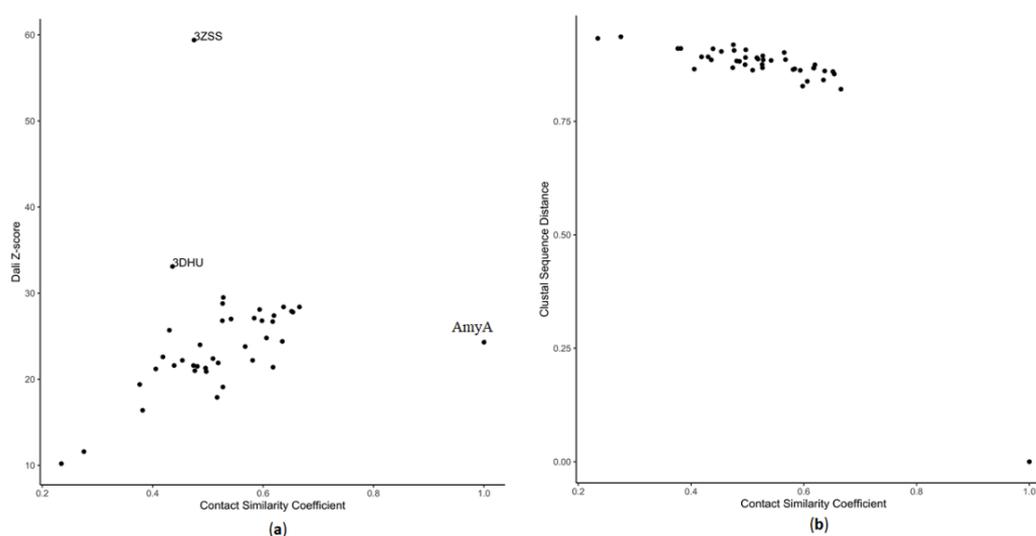


**Figure 1.** Function classification of enzymes in the GH13 family based on the amino acid preference in their residue-residue contacts. (a) Dataset 1 (training group). (b) Dataset 2 (validation group). (c) Dataset 4 (updated dataset). In both cases, enzymes defined as transglycosidic or hydrolytic are shown with green and red dots, respectively.



**Figure 2.** Correlation of the residue-residue contact conservation between all enzymes and those of *TmAmyA*, one of the enzymes modified in this study. The preservation of each contact in a protein is plotted against its conservation in *TmAmyA*. This parameter seems to correlate with phylogenetic relationships. A poor correlation is evident for the enzymes not belonging to the GH13 family (e.g., PDB ID: 2ZQ0, family GH97, and enzyme with the PDB ID: 3W37, family GH31).

The contact similarity coefficient correlated with the structure DALI Z-value (Figure 3a), which showed a relationship with sequence identity [59]. However, its resolution for identifying differences between proteins was greater than that of pairwise sequence distance, for which the values plateaued for the proteins under consideration (Figure 3b). The contact similarity coefficient showed more sensitivity to enzyme similarity than the sequence similarity parameters. Despite the structural and sequence difference between *TmAmyA* and the rest of the enzymes, the contact similarity coefficient was able to identify *Thermotoga maritima* GTase (*TmGTase*) as its closest structure. The contact similarity coefficient also grouped some GH13 subfamily members (subfamily 2: 4JCL, 2CXG, 1UKQ; subfamily 13: 2YOC, 2FHF; subfamily 8: 3AMK, 3AML) and identified outsiders (GH97: 2ZQ0, GH31: 3W37; Table S3). Although it misplaced structures, such as 3K8M, that should have been close to *TmAmyA*, this could be due to additional domains not present in the GH13 family [60]. As the comparison of the dataset with a single structure yielded such identifications, the comparisons between the whole dataset may provide a tool for performing automated classification of enzymes when performing the analysis of domains.

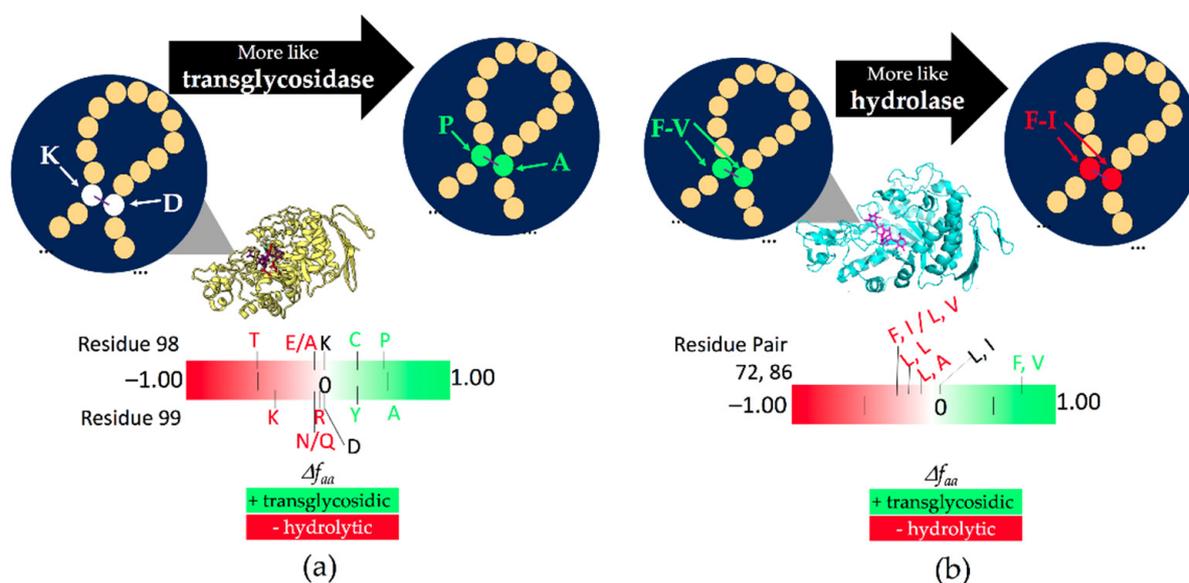


**Figure 3.** Parameters of similarity distance between evaluated enzymes and *TmAmyA* based on their correlation between contact conservation (R-squared) and (a) structure (Dali) and (b) sequence (Clustal) alignments.

### 2.3. Modification of Reaction Specificity in the $\alpha$ -Amylase A from *Thermotoga maritima*

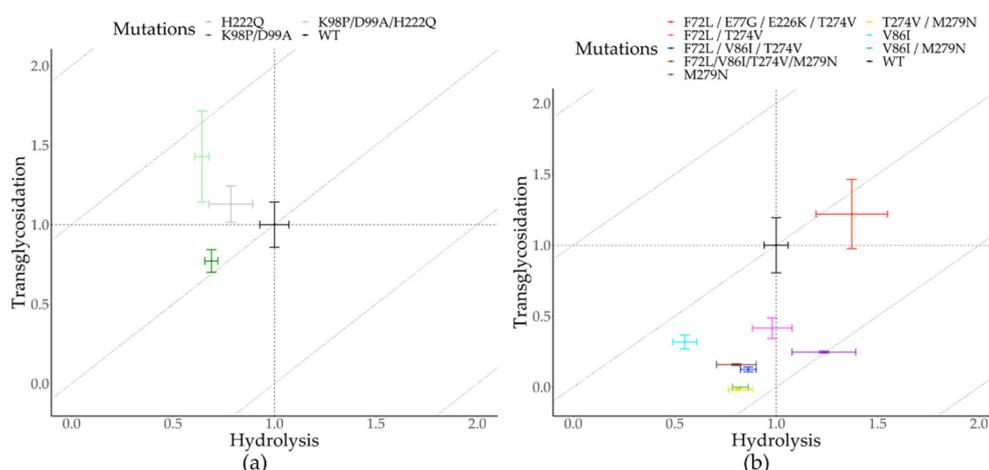
One of the goals for increasing transglycosidic activity in glycoside hydrolases is to create glycosynthetic biocatalysts. We decided to use the *Thermotoga maritima*  $\alpha$ -amylase AmyA (*TmAmyA*) as a model of a hydrolase. This enzyme offers as advantages a high thermal stability, with an optimal temperature above 80 °C, and an efficient saccharifying starch hydrolysis pattern, associated with a considerable transglycosylation activity [43]. Mutations for increasing the transglycosidic activity of *TmAmyA* should substitute residues more frequently found in hydrolases ( $\Delta f_{aa}^{ij} < 0$ ) by others more frequently present in transferases ( $\Delta f_{aa}^{ij} > 0$ ). The criterion for selecting residues for mutagenesis was if the original residue in *TmAmyA* had an enrichment value ( $\Delta f_{aa}^{ij}$ ) below  $-0.2$  (residues enriched in hydrolases) or if the  $\Delta f_{aa}^{ij}$  for potential candidate residues was above 0.2 (enriched in transferases), while the *TmAmyA* residue had a  $\Delta f_{aa}^{ij} \leq 0$ . A valuable result from this approach was the identification of targets for mutagenesis beyond the catalytic site, whose relevance in terms of their specificity and activity has been shown by directed evolution [41]. The exploration of such sites is limited by the availability of selection or high-throughput screening methods; thus, restricting the sequence space that is to be evaluated becomes paramount.

When selecting a residue to mutate, preference was given to residues not in the catalytic site that were conserved among the enzymes used (in more than 10 out of 14 enzymes). As a proof of principle, we selected a contact complying with these restrictions to produce the variant K98P/D99A double mutant (K98 with  $\Delta f_{aa}^{ij} = 0$  against P98 with  $\Delta f_{aa}^{ij} = 0.45$  on one side, and D99 with  $\Delta f_{aa}^{ij} = 0.00$  against A99 with  $\Delta f_{aa}^{ij} = 0.4$  on the other side; Figure 4a). These residues were located in a very long loop joining the  $\beta$ 2-strand and helix-2 in the TIM barrel, which is not part of any of the highly conserved regions in the GH13 family. K98P/D99A substitutions represent a drastic change of physicochemical properties (Table S5). However, a Pro residue would favor the turn observed in the structural model at the targeted position, and substituting both residues would change the  $\Delta f_{aa}^{ij}$  to a positive value, which is desirable in transglycosidases (Figure 4a; for more  $\Delta f_{aa}^{ij}$  values at this position see Table S5). These mutations were evaluated in the wild-type background and combined with the H222Q mutation near to the catalytic site previously reported to increase the transglycosylation/hydrolysis ratio [28].



**Figure 4.** Schematic representation of substitutions in glycosidases based on enrichment factors. (a) K98P/D99A in TmAmyA (b) F72L/V86I for TmGTase. The residues more frequently found in hydrolases and transglycosylases are shown in the lower part of the figure for each pair in red and green, respectively.

In addition to its impact in the synthesis of alkyl-glycosides, the alcoholysis reaction (i.e., reaction of a sugar with an alcohol to produce an alkyl-glycoside) can be used as a proxy for the transfer reaction to other sugars [30]. This approximation was used because *TmAmyA* transglycosylation products do not significantly accumulate during the predominant hydrolysis reaction. We thus performed the depolymerization of starch in the presence of 10% 1-butanol to obtain butyl-glycosides as products and evaluated the alcoholysis yield in the different *TmAmyA* variants, as previously described [43]. As a result, the double mutant K98P/D99A succeeded in increasing the T/H ratio by a factor of 1.17 (Figure 5a). The mutations produced by the enrichment factors had a more significant impact on hydrolytic activity, with an approximately 25% reduction (Table 1). On the other hand, the alcoholysis yield was 17% lower over the wild-type background, while having a near 10% increase compared to the H222Q variant.



**Figure 5.** Transglycosylation/Hydrolysis (T/H) ratio of glycosidases. (a) TmAmyA and its variants; (b) TmGTase and its variants. Transglycosylation and hydrolysis values were normalized to each wild-type enzyme, which has the coordinates (1,1). The dashed lines indicate equivalent T/H ratios. The central diagonal corresponds to the wild-type enzyme. Enzymes with a better T/H ratio are over the central diagonal line, and enzymes with a better H/T ratio are under the central diagonal line.

**Table 1.** TmAmyA variants production after 12 h of reaction.

TmAmyA Variant	Hydrolysis (mEq Dextrose/ $\mu$ g Protein $\times 10^{-2}$ )	Transglycosylation (mEq Butyl Glucoside/ $\mu$ g Protein $\times 10^{-4}$ )	Transglycosylation/Hydrolysis (T/H) Ratio $\times 10^{-2}$
Wild type	$2.8 \pm 0.2$	$7 \pm 1$	$2.5 \pm 0.6$
K98A/D99P	$1.93 \pm 0.09$	$5.4 \pm 0.5$	$2.8 \pm 0.4$
H222Q	$2.2 \pm 0.3$	$7.9 \pm 0.8$	$3.6 \pm 0.4$
H222Q/K98A/D99P	$1.8 \pm 0.1$	$10 \pm 2$	$5.6 \pm 0.4$

#### 2.4. Increasing Hydrolase Activity in the 1,4- $\alpha$ -Glucanotransferase from *Thermotoga maritima*

As a proof of concept, we wanted to know if the enrichment factor could be used in the other direction—to turn an enzyme which is mainly a transferase into a hydrolase. For this purpose, we selected the GTase of *T. maritima* (TmGTase). In this case, we analyzed residue pairs instead of residues within a contact pair, as the analysis of residue pairs was an efficient parameter for classifying GH13 enzymes according to their function (Figure 1). The contact pairs were better than the individual residues for classifying enzymes according to their reaction specificity. For this reason, the use of contact pair enrichment, instead of the enrichment of individual amino acids within each pair, should increase the chance of selecting substitutions that transform a transglycosidase into a hydrolase. This would also ensure the selection of pairs more representative of those found in the transglycosidic and hydrolytic enzymes. Additionally, we included the parameter of betweenness centrality—a measurement of the role of a node in transferring information within a network [61]—to restrict our search of mutation sites further. This centrality parameter is calculated as the sum of the fraction of the paths between all pair nodes  $i$  and  $j$  containing the node  $v$ , distinct from  $i$  and  $j$  [62]. This parameter is reported as a measure of the importance of specific amino acid residues for the structure and function of proteins [63].

We identified pairs of residues present in all transferases that differed from those in the hydrolases. We focused on the pairs of residues in the top 10% of the central betweenness values. We detected pairs of residues differentially enriched in glycoside hydrolase clusters around F72 and F273 and selected F72/V86 from the first cluster and residues T274/M279 from the second cluster as mutagenic targets (for the enrichment values of these clusters see Tables S6 and S7). It is worth mentioning that position 279 is part of the fourth highly conserved sequence region in the GH13 family. We then investigated the role of the single

and combined substitutions F72L, V86I, T274V, M279N in transglycosylation and hydrolysis reactions (Figure 4b, Tables S6 and S7). Although a decrease in hydrolytic activity was observed for many mutants tested in *TmGTase*, they produced a sharper reduction of transglycosidic activity in all of the variants, entirely impairing their activity in some of them (Table 2). As a result, the goal of increasing the H/T ratio was achieved. Of particular interest is the variant M279N, which, besides decreasing the transglycosidic activity by four-fold, it increased the desired hydrolytic activity by 25%, yielding a five-fold increment in the H/T ratio. It is worth mentioning that during the characterization of the variants, we detected a variant with a higher overall activity, which, besides the engineered F72L/T274V mutations, contained two unintended modifications: E77G/E226K. This variant does not change the H/T ratio because both reactions were favored by the extra substitutions, with an increase of about 40% overall activity relative to the wild-type protein (Figure 5b).

**Table 2.** *TmGTase* variants specific activity.

<i>TmGTase</i> Variant	Hydrolytic Activity ( $\times 10^{-5}$ mg Starch/ $\mu$ g Protein/min)	Transglycosidic Activity ( $\times 10^{-3}$ mg Starch/ $\mu$ g Protein/min)	Hydrolysis/Transglycosidation (H/T) Ratio ( $\times 10^{-2}$ )
Wild type	5.1 $\pm$ 0.3	4.0 $\pm$ 0.8	1.3 $\pm$ 0.3
V86I	2.9 $\pm$ 0.3	1.3 $\pm$ 0.2	2.2 $\pm$ 0.6
M279N	6.3 $\pm$ 0.8	1.01 $\pm$ 0.03	6 $\pm$ 1
V86I/M279N	4.2 $\pm$ 0.2	ND	NA
T274V/M279N	4.2 $\pm$ 0.3	ND	NA
F72L/V86I/T274V	4.4 $\pm$ 0.2	0.51 $\pm$ 0.07	9 $\pm$ 2
F72L/V86I/T274V/M297N	4.1 $\pm$ 0.5	0.65 $\pm$ 0.03	6 $\pm$ 1
F72L/E77G/E226K/T274V	7.0 $\pm$ 0.9	5 $\pm$ 1	1.4 $\pm$ 0.5
F72L/T274V	4.9 $\pm$ 0.5	1.7 $\pm$ 0.2	2.9 $\pm$ 0.6

ND: non-detectable activity; NA: not available.

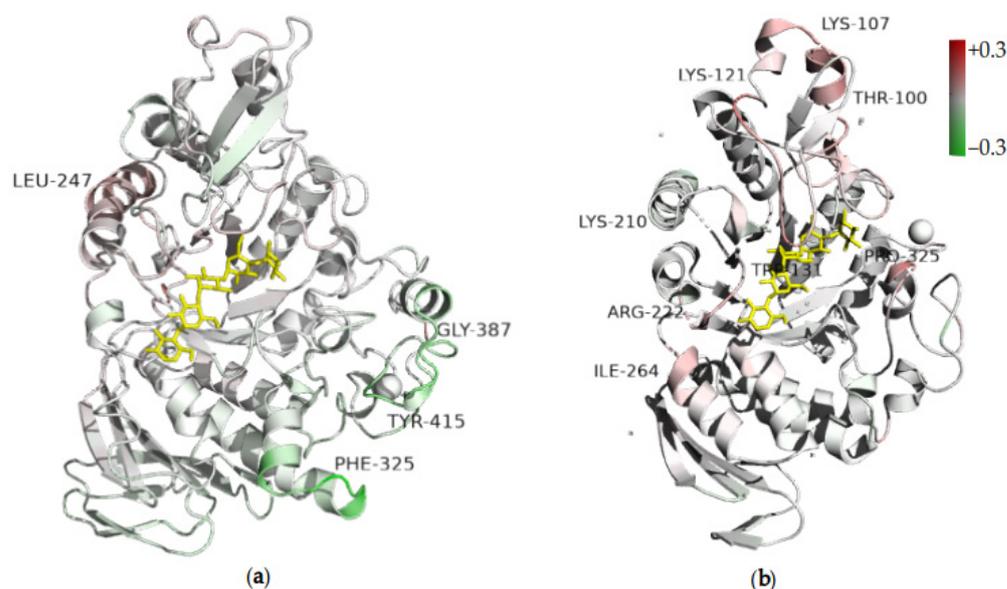
### 2.5. Molecular Dynamic Simulations

The role of protein dynamics in enzymatic catalysis is well recognized [64,65]. Thus, we carried out molecular dynamics simulations for some protein variants to explain the effect of mutations on the activities of the enzymes in structural terms. In the case of *TmAmyA*, the root mean square fluctuation (RMSF) difference between the triple mutant (K98P/D99A/H222Q) and wild-type protein was discreet (Figure 6). The  $\alpha$ 4 helix (catalytic domain), marked with residue 247, was more flexible in K98P/D99A/H222Q than in the wild-type *TmAmyA* (Figures S1 and 6a). This helix is near the loops containing the catalytic aspartate (nucleophile) and glutamate (acid-base). Contrarily, the K98P/D99A/H222Q mutant was rigidified at the loops and helices comprising residues 325, 387, and 415.

The changes in flexibility were also studied for wild-type *TmGTase* against M279N single and T274V/M279N double mutants. An augmentation in the RMSF of the double mutant T274V/M279N was evident when compared to those in the wild type *TmGTase*, around residues 100, 107, 121 (B-domain), 131 (+2 subsite, B domain), 210 (loop  $\alpha$ 4 $\beta$ 5), 222 (helix  $\alpha$ 5), 264 (helix  $\alpha$ 6), and 325 (loop  $\alpha$ 4 $\beta$ 5, near subsite -3; Figure 6b and S3). While RMS fluctuations for the single mutant M279N were similar to the double mutant around residues 100, 107, 121, and 131, they were increased relative to the other proteins near residue 222. Thus, the B domain, which is important for substrate specificity in other GH13 enzymes, is more mobile in both mutants [57].

The dihedral angles ( $\chi_{1-3}$ ) of the catalytic acid-base residue of wild type and mutants from *TmAmyA* and *TmGTase* were analyzed, as elsewhere [47], and are shown in Figures S2 and S4. We observed that contrary to David et al., the  $\chi_3$  angle showed a more diffuse distribution in the more hydrolytic variants of both enzymes (wild-type *TmAmyA* vs. K98P/D99A/H222Q, and wild-type *TmGTase* vs. M279N). Furthermore, we observed a larger conformational sampling of the catalytic acid-base in the more hydrolytic variant of each pair (Figure S5). As Lundemo et al. has pointed out, the residue chain's mobility and

orientation could be better described by the  $\chi_1$  and  $\chi_2$  angles. Moreover, when studying the cyclodextrin glucosyltransferase from *Bacillus stearotherophilus* NO2, a GH13 enzyme, Kong et al. defined a new angle for analyzing E253 in this bacillus CGTase—finding that it is more flexible in mutant L277M, which is less hydrolytic than the wild-type protein [66].



**Figure 6.** RMSF difference (in nanometers) along the structure with high and low transglycosidic variants of glycosidases. (a) TmAmyA K98P/D99A/H222Q triple mutant (less hydrolytic than the wild-type enzyme). (b) wild-type TmGTase (high transglycosidic) vs. T274V/M279N. The active center is delimited by acarbose (yellow stick) and is located away from the zone with the modified fluctuations. No differences in RMSF are shown in white, while positive or negative RMSF changes are red and green, respectively.

In *TmAmyA*, the  $\chi_3$  angle mainly occupies two conformations in the triple mutant, while it seems not to have any preference in the wild-type enzyme (Figure S2a–c). Further studies are required to elucidate the effect of this amino acid mobility and orientation on the reaction specificity.

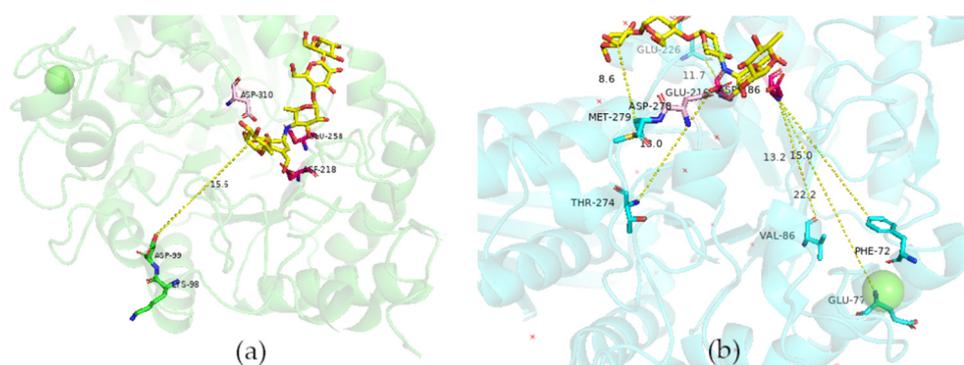
The change in structural dynamics modifies the pKa of E216 in *TmGTase*. PROPKA calculations [67] of this residue were performed on the structures corresponding to three different times in the simulation: 200, 300, and 400 ns, when the RMSD values plateau. The pKa of the catalytic acid-base residue was higher for both variants than for the wild-type *TmGTase*. For this parameter, the average was  $3.0 \pm 0.97$  for the wild-type, and  $6.1 \pm 0.54$ , and  $4.8 \pm 0.43$  for T274V/M279N and M279N, respectively. These results agree with the notion that hydrolysis requires a more basic residue than transglycosylation [41]. Although the pKa of the acid-base residue changes drastically during the reaction, this analysis suggests that the enzyme is tuned to increase its pKa. Despite considering only the free enzyme in the simulation, it is interesting to notice a shift in pKa—although the enzyme has not yet formed the covalently bound sugar-enzyme intermediate. Thus, these values have to be taken with care since they do not reflect the acid-base residue's environment during the second step of the reaction. In the case of residue E258 of *TmAmyA*, the K98P/D99A/H222Q triple mutant has a pKa value similar to the wild type for its catalytic acid-base residue (around 4.8 for both). This result suggests a different mechanism for the change in reaction specificity, one exclusively affecting the hydrolysis reaction.

Additionally, the average distance between D278 and E216 was 1 Å closer in both mutants than in the wild-type enzyme *TmGTase* (Figure S6). As these two acid groups influence each other pKas, reducing the space increases the pKa of at least one of the participating amino acids, in order to avoid electrostatic repulsion. Consistently with

PROPKA calculation, the average distance fluctuations between D310 and E258 are similar for *TmAmyA* proteins. These residues are closer in the wild-type *TmAmyA* than in its triple mutant, shifting to a further distance after 350 ns of simulation, and averaging the same for both variants (Figure S7). These results suggest a different dynamic for both glycosidases at the region close to D278, a residue that functions as a transition state stabilizer. Additionally, according to MD simulations, for the *TmGTase* T274V/M279N variant, D278 established a stable hydrogen bond with K324 that was not observed in other *TmGTase* variants (Figure S8). For *TmAmyA*, the equivalent residue to K324 was a Gly, unfit for forming a hydrogen bond with a corresponding catalytic aspartate.

### 3. Discussion

We implemented a methodology for identifying mutagenic target sites to modulate the transglycosylation/hydrolysis (T/H) ratio in two members of the GH13 family. This methodology selected target residues far from the active site (for this work between 11.1 and 22.2 Å away) that modified the reaction specificity (Figure 7a,b). It was interesting to note that it also could select residues close to the catalytic site, such as residue 279 of *TmCGTase*, which is next to the catalytic residue D278 (2.7 Å; Figure 7b). This reaction specificity modulation seemed to work in both directions; however, the more significant contribution was the reduction in the undesired reaction, and to a lesser extent, the increase in the desired one, in most cases.



**Figure 7.** Distances in Angstrom between the mutation sites (green and cyan) and the catalytic residues (red) (a) *TmAmyA*. (b) *TmGTase*. The active center is delimited by the competitive inhibitor acarbose (yellow). In pink is shown the D310 residue that acts as a transition state stabilizer.

Conservation of residue-residue contacts can be used to recapitulate the clustering of some of the GH13 subfamilies (Figure 2), as well as the separation of the enzymes according to their reported specificity (Figure 1). In this classification, the number of transglycosidic residue-residue contacts (formed by two residues enriched in transferases) seemed to play a more prominent role in distinguishing between functions. Hydrolytic residue-residue contacts (constituted by two residues enriched in hydrolases) are low for some hydrolases. In contrast, transglycosidic residue-residue contacts were always abundant in transglycosidic enzymes and underrepresented in the hydrolytic ones. These results are consistent with previous works where patterns of contacts have been used to distinguish protein families [26] and add to their use in functional and evolutionary classification of subfamilies.

Our analysis detected many evolutionary relationships. It indicated that *TmGTase* and *TmAmyA* resulted from a gene duplication event after speciation, as these were the closest structures in the analysis, but the enzymes have different functions. The structural analysis also pointed in this direction. It was striking that *TmAmyA* was the nearest to the Firmicutes' enzymes than to those of other Archaeas—a branch to which *T. maritima* belongs [68]. This advocates for a horizontal transfer between the two groups, resulting in a close relationship between AmyA's GH13 subfamily 36 and Firmicutes' subfamily 2,

whose members include both bacterial and archaeal enzymes [69]. This relationship is also conveyed by the similarity of the B domains between the *TmGTase* structure (PDB ID: 1LWJ) and the other enzymes of the Firmicutes group. Nonetheless, domain B varied the most out of the three core domains of the GH13 family; residues 148–161 in 1LWJ could be readily aligned with 180–191 of the *Geobacillus thermoleovorans* CCB\_US3\_UF5 GTase (PDB ID: 4E2O), and many features were shared throughout the structures. It is noteworthy that this *Geobacilli* enzyme has not been assigned yet to a subfamily in the CAZy database. These results commend a more extended use of residue contacts in the classification of enzymes, complementary to sequence and structure analysis, as their analysis is more sensitive than sequence analysis and can be automated.

Results from directed evolution, especially in CGTases, have emphasized the relevance that residues at subsites +1 and +2 have on determining reaction specificity [50,51]. In these works, however, as the authors recognize, a completely random mutagenic scheme mainly allows the exploration of single mutants. The opportunity to guide the mutagenic process with structural or mechanistic information would accelerate the evolution of reaction specificity in these proteins. The successful combination of mutations has required several rounds of mutagenesis and screening in order to find them. Still, most of the work reported on alpha-amylases explores only the active site. The role that dynamics play in stabilizing the transition state of reactions is much subtler but no less important. This contribution is encoded in the protein sequence, and decoding it through the analysis of contact conservation patterns expands the exploration to positions beyond the active site, in order to influence reaction specificity. Thus, our analysis of contact conservation patterns detects residues that have already been explored and that change reaction specificity. For example, it predicts the substitution A230V in CGTase from *B.circulans* [51], and suggests the substitution of other residues in contact with A230 that have not yet been investigated to improve the hydrolysis reaction even more.

As part of the results obtained with the implemented algorithm, we evaluated the rationality of other reported mutations in glycoside hydrolase enzymes. The mutation A289F, that introduces transglycosylation activity in *B. stearothermophilus*  $\alpha$ -amylase, could have been predicted by a change in enrichment factors from  $-0.03$  to  $+0.1$  [45]. It also occurs with the mutation V286F in *B. licheniformis*  $\alpha$ -amylase (BLA) [44], with a similar change in enrichment factor values. Interesting mutants on which our group are now working are V286F/T329M (*B. licheniformis*) and A289F/T335M (*B. stearothermophilus*), which include a new mutation suggested by the enrichment factor change from  $-0.3$  (T) to  $+0.7$  (M). Although a mutation in H222Q in *TmAmyA* would not have been chosen, the substitution of histidine at that position (enrichment factor  $-0.05$ ) to leucine (enrichment factor:  $0.25$ ) is indicated by the analysis of enrichment factors to increase transglycosylation.

As for the mutations characterized in this work, the equivalent residue to *TmGTase* T274 in *Aspergillus oryzae*  $\alpha$ -amylase was occupied by V293, rendering a functional  $\alpha$ -amylase (Figure S10). This loss of function might have been provoked by a need to remodel the network to which T274 belongs. (Figure S10), emphasizing again the necessity of a methodology for identifying groups of residues that interact together. In conjunction with residue 274, position 279 is involved in a hydrogen bond network comprising D314, R281, F311, T274, S275, N276, K244, F273, T274, M279, and S280 (Figure S9).

The mutations identified in this work seemed to be connected to the catalytic site, probably influencing the catalytic site in addition to the dynamic changes described earlier. One of such elements was linked to the calcium ion to which *TmAmyA* D99 was bound, which might reflect once again the importance of metal ions in  $\alpha$ -amylase structure and function [36,70] (Figures S11 and S12). As observed in this work for the *TmGTase* variants, mutations modifying the transglycosylation/hydrolysis (T/H) ratio change the dynamics of the loops surrounding the protein's active center, including the B domain. This behavior agrees with the GH51 retaining  $\alpha$ -L-arabinofuranosidase from *Thermobacillus xy-lanilyticus* [71] and the almond  $\beta$ -glucosidase, in which the movements of four strategically located loops act as a lid for the active center, controlling the catalytic activity [72].

Additionally, in the GH13 family, a chimeric amylosucrase from *Deinococcus geothermalis* (DGAS) and *Neisseria polysaccharea* (NPAS) had a differential fluctuation in loops 4, 7, and 8 between both variants, associated with T/H changes [73]. The mutation A226N in DGAS also modified the T/H ratio from 0.59 (wild type) to 0.9 and diminished the flexibility in loops 2, 3, 4, 7, and 8 [74]. Additionally, it was reported for the *T. kodakarensis* glycogen branching enzyme (family GH57) that a loop extending from G227 to P248 contained a Tyr (residue 233) whose modification changed the branching/hydrolysis ratio from 41 (wild-type) to 16.2 without affecting branching activity [75].

For the helix extending from residue 221 to 231, we observed a change in mobility and T/H ratio for *TmGTase* M279N. Changes in specificity have been observed for mutations in the equivalent helix. In *Bacillus stearothermophilus* amylase (BStA; from I270–K279), the mutation I277F increases the specific activity of this hydrolase [76]. This result might also explain why F72L/E77G/E226K/T274V has a higher hydrolytic activity than the F72L/T274V variant as residue E226 interacts with two charged residues (D225, R22), producing a kink in the helix. Substituting residue E226 for a positively charged residue (E226K) might change the inclination and dynamic of this helix, thus affecting the catalytic acid-base. Together, these results suggest the importance of the dynamic reconfiguration of the helix comprising residues 221–231 in hydrolytic activity.

The comparison of residue contacts between groups might find applications outside of function, as suggested by a work in which the stability of a GH13 enzyme was modified by mutagenesis at hotspots identified by direct coupling analysis [77], creating and removing hydrogen bonds between residues. It also could identify sites where fluorescent probes can be inserted while minimizing functional impairment. This methodology finds its parallel in some coevolution analysis methodologies that have been reported [29,30]. Still, it required less data to identify sites crucial for function, as contacts were directly observed from the structure while the coevolution analysis inferred them. Both perspectives highlight the importance of analyzing residue contacts when a function is studied, even away from the catalytic site. Previous works aiming to classify enzymes of unknown function and predict and verify 3D models have also suggested the importance of residue contacts [25,26], where contact map prediction has played an important role.

The method presented in this work is limited by the availability of 3D structures, which are usually less abundant than the sequences employed in coevolutionary studies. Both strategies benefit by including many proteins in the study. Still, as mentioned earlier, the contact map analysis requires fewer proteins, as contacts are obtained directly from structures. In contrast, coevolution studies infer this information via correlation analysis. This limitation could be overcome by including sequences in the analysis aligned to the closest structure, but it should be considered that the number of enzymes experimentally characterized as 1,4- $\alpha$ -glycoside hydrolases is much higher than those of the 1,4- $\alpha$ -glycosyltransferases (only 31), at least as reported in the CAZy database.

Another limitation is the incomplete information available on the reaction specificity of each enzyme in the glycoside hydrolase family. This analysis could also be enriched if characterization of both functions was available for all enzymes. This characterization would also determine if there is a functional meaning to the grouping of GH13 enzymes based on their hydrolytic or transglycosidic contacts. This methodology should also be extended to include more simultaneous mutation sites than the maximum of four that was evaluated in this work. The residues explored were in contact with others forming a more complex network. Therefore, exploring mutants comprising a complete network of interacting residues should give a more noticeable change in reaction specificity. Despite these limitations, contact maps could be an additional auxiliary in the quest to predict critical functional residues as they are already being used to predict structures.

Our methodology aimed to create a parameter to guide the modulation between transglycosidation and hydrolysis specificity. As Healp and Blouin pointed out [78] in their study of the evolvability of the GH13 family stability, a quantitative prediction of specificity requires catalytic information with a single substrate for all the members of

the family under analysis. In contrast, the GH13 family has been characterized with a great range of substrates for both the hydrolytic and transglycosidic activities, and in many cases, just for their most prominent activity. Even when the most common substrate, starch, has been used, variations in 1,6-ramification might preclude this characterization. We surmounted this obstacle by performing a differential analysis of enzymes with a distinct specificity, thus creating a parameter (the enrichment factor) that offers a qualitative guide to transform reaction specificity that might require more experimental data in order to become quantitative. Nevertheless, this guide might reduce the sequence space that needs to be explored in order to achieve a specificity change, both in rational and directed evolution studies.

The development of computational approaches to resolve biological problems is a growing research area [79,80]. Many aspects of the protein structure–function relationship are of particular interest in engineering enzymes for biocatalysts. Additionally, artificial intelligence could be an approach for improving and complementing other strategies that are already employed for mining functional data. This tool could help overcome the effects of epistasis, which occludes the selection of combined mutations, to improve the desired function. Recently, Timonina et al. [81] reported a method based on artificial intelligence, named Zebra3D, which centers its analysis on specific structure regions (SSRs) of the protein family after aligning their 3D structures. It classifies proteins into subfamilies, with distinct structural elements for each enzyme associated with substrate specificity in human aldose reductase and catalytic activities for  $\alpha/\beta$ -hydrolases. Analyses such as Zebra3D would benefit by including the study of residue contacts. Residue contact analysis not only classifies enzymes functionally, but also identifies distant interacting regions—especially when networks of contacts are considered. The investigation of more extensive networks is underway.

## 4. Materials and Methods

### 4.1. Bioinformatic Analysis

#### 4.1.1. Analysis 1

CMView 1.0 [55] was used to create the contact map (cut-off distance 5 Å, all atoms) of 14 members of the CAZy family GH13 (Dataset 1, Table S1) and were aligned against the contact map of the *Thermotoga maritima* 1,4- $\alpha$ -glucanotransferase *TmGTase* (PDB: 1LWJ). All selected structures were bound to a transition state analog, 12 were hydrolases and 4 were transferases. We wrote an R program to obtain the *TmGTase* residue-residue contacts shared with the rest of the structures and their corresponding amino acids. The algorithm also obtained the frequency of each amino acid at each position for all contacts ( $f_{aa}^{ij}$ ) within a group (hydrolase,  $f_{aa}^{ij,H}$ , or transferase,  $f_{aa}^{ij,T}$ ) and an enrichment factor ( $\Delta f_{aa}^{ij}$ ) defined as:

$$\Delta f_{aa}^{ij} = f_{aa}^{ij,T} - f_{aa}^{ij,H} \quad (1)$$

We qualified the 14 structures used, and an additional set of 14 enzymes (twelve hydrolases and two transferases with structures without any ligand bound; Dataset 2, Table S2) as transferases or hydrolases based on the enrichment factor scored by both residues at each contact. To do this, the residue contacts were compared with those of *TmGTase* (PDB ID: 1LWJ). The amino acids in all residue contacts shared with 1LWJ were qualified as hydrolytic if  $\Delta f_{aa}^{ij} < 0$  (the amino acid was more frequent in hydrolases) and transglycosidic if  $\Delta f_{aa}^{ij} > 0$  (the frequency of the amino acid was higher in the transferases). After qualifying all the amino acids in the residue contacts, contacts formed with a pair of residues deemed transglycosidic were designated transglycosidic; those formed by two hydrolytic residues, as hydrolytic.

During the development of this work, new members of family GH13 were included in the CAZy database, so that we created Dataset 4, which consists of 71 members—40 classified as hydrolases and 31 as transferases. Dataset 4 was used to calculate the

enrichment factors, and then to classify hydrolases or transglycosidases based on their contacts, which were based on the enrichment factors.

Dataset 3 was created to compare contact conservation as a classification criterion with other probed methods such as sequence alignment [82] or structure conservation [83]. This data set contained proteins from Dataset 1 and Dataset 2 and eleven additional proteins classified as members of the GH13 family in CAZy—a GH97 enzyme (PDB: 2ZQ0), and a GH31 enzyme (PDB:3W37)—all acting on 1,4- $\alpha$ -glycosidic bonds either as a hydrolase or transferase (Table S3). All of the sequences were aligned with ClustalW [82]. The structures were aligned using the DALI server [83]. In the case of the contact maps, all the contacts for every protein were identified. The fraction of enzymes sharing each contact was calculated (contact conservation score). All enzymes were used as a reference to calculate the conservation of all the contacts, and then compared against those of every other protein in the dataset. The contact conservation score sets of all the enzymes were plotted against each other.

#### 4.1.2. Analysis 2

Enrichment factors of pairs of amino acids were calculated from Dataset 1. The contact maps and alignments obtained from CMView were analyzed using 6 Å as a cut-off for all the atoms. The R program used for analysis calculated the frequency of all amino acid pairs for all contacts ( $f_{aap}^i$ ) and an enrichment factor for the pair:

$$\Delta f_{aap}^i = f_{aap}^{i,T} - f_{aap}^{i,H} \quad (2)$$

The contact map of *TmGTase* was transformed into a 3D network, whose nodes were the  $\alpha$ -carbon of each residue and the edges of the residue contacts. For this network, the central betweenness of its nodes was calculated. The betweenness centrality (BC) of a node  $x$  that is part of a network  $V$  was calculated through the following equation:

$$BC_x = \sum_{u,v \in V} \frac{\sigma_{u,v}(x)}{\sigma_{u,v}} \quad (3)$$

where  $\sigma_{u,v}$  is the number of paths between the nodes  $u$  and  $v$ , and  $\sigma_{u,v}(x)$  is the number of times these paths contain the node  $x$  [62].

#### 4.1.3. Analysis 3

Enrichment factors for all the contacts in the homology model of the *Thermotoga maritima*  $\alpha$ -amylase AmyA (*TmAmyA\_2*) were calculated. *TmAmyA\_2* was compared against 71 structures of GH13 enzymes (31 transferases and 40 hydrolases, Dataset 4, Table S5). Contact maps and alignments were produced with CMView using a cut-off distance of 5 Å between all atoms.

#### 4.1.4. Selection of Mutation Sites

The results from Analysis 1 were used to identify positions in the homology model *TmAmyA\_1*. Mutation sites were selected based on their conservation score, enrichment factor and their distance from the catalytic site. The conservation score of the sites considered was above 10 (out of 14). The occupying amino acid should have scored an enrichment factor below or equal to 0, with the alternative to be substituted with an amino acid having an enrichment factor above 0.2. Residues not in contact with the transition state analog were preferred.

Additionally, results from Analysis 2 were used to identify positions in *TmGTase* as targets for mutagenesis. Mutation sites were selected based on their conservation score (above 10), enrichment factor (top and lowest 1%), and their central betweenness score (top 20%).

The code for the enrichment factor calculation is available at: <https://github.com/rhodbacter/ContactMaps-Catalysis/> (accessed on 26 July 2021).

#### 4.1.5. Molecular Dynamic Simulations

The molecular dynamic (MD) simulation and analysis were accomplished using GRO-MACS (GRoningen MACHine for Chemical Simulation) version 2020.4 [84], as reported [85], with slight modifications. The atom coordinates for each protein were obtained from the Protein Data Bank [86]. All modification or visualization of PDB files was carried out with UCSF Chimera [87] or Pymol (Schrödinger, Inc., New York, NY, USA). Incomplete structures were completed with Modeller [88]. Hydration was performed using “Simple Point Charge Extended” (SPCE) in a cubic box of 1 nm × 1 nm × 1 nm. The forcefield was OPLS-AA/L (optimized potentials for liquid simulations). The system was neutralized with Na<sup>+</sup> and Cl<sup>−</sup> ions. Later, minimization by steepest descent minimization was performed until the energy decreased to <1000 kJ/mol/nm. Equilibration was run using a leap-frog integrator and with a modified Berendsen thermostat for 1000 ps, with a step of 500,000—first for a NVT ensemble, followed by one with an NPT ensemble. Finally, the simulations were performed for 500 ns with an NPT ensemble using a leap-frog integrator and a modified Berendsen thermostat. Analysis of results were done with GROMACS tools, and visualization was accomplished with Xmgrace [89]. The pK<sub>a</sub>s of aspartic and glutamic acids were calculated with PROPKA 3.4.0 [67].

#### 4.1.6. Homology Models

All homology models were created using the Swiss Model [90–98] with the best matching template. The quality of the models was assured through their QMEAN values (Table S4). Additionally, the validation of the model obtained for *TmAmyA* was done with the structure assessment tools (QMEAN, Ramachandran favored, and MolProbity Score) from the Swiss-Model server [58], VERIFY 3D [99,100] and ProSA-web [101,102].

### 4.2. Experimental

All reagents were purchased from Sigma-Aldrich (St Louis, MO, USA) unless stated otherwise. All enzymes were purchased from New England Biolabs (Ipswich, MA, USA) unless otherwise stated.

#### 4.2.1. Construction of the TmGTase Gene Vector

The Tm0364 (KEGG ID) gene was amplified from a *T. maritima* cDNA library by PCR reaction using *pfu* DNA polymerase. The PCR product was purified using agarose gel electrophoresis and Roche PCR purification kit (Roche Diagnostics GmbH, Mannheim, Germany). Vector pET22a and the PCR product were digested using the restriction enzymes *NdeI* and *XhoI*. The resulting fragments were purified through agarose gel electrophoresis and the Roche PCR purification kit (Roche Diagnostics GmbH, Mannheim, Germany) and finally ligated using *T4* DNA ligase (ThermoFisher, Waltham, MA, USA). The ligation product was used to transform the genes into electrocompetent *Escherichia coli* MC1061 cells. The *TmAmyA* gene was cloned in the same way, and it has been previously reported [43].

#### 4.2.2. Construction of TmAmyA and TmGTase Variants

Mutations were constructed either through the Mega Primer method (using *pfu* DNA polymerase) [103] or Quick Change (with *DpnI* and ThermoFisher Scientific *phusion* high fidelity enzyme) using oligonucleotides (Table S8) purchased from Unidad de Síntesis y Secuenciación de DNA of Instituto de Biotecnología, UNAM.

#### 4.2.3. TmAmyA and TmGTase Variants Expression

All the pET22a plasmids containing the genes were transformed into calcium competent *E. coli* K12 ER2738 cells. Induction was performed with 0.5 mM IPTG (*TmAmyA* variants) or 0.1 mM IPTG (*TmGTase* variants). In all cases, purification was performed over two steps after sonication cell lysis: (1) heating at 70 °C for 1 h, and (2) affinity chromatography of the heated supernatant with a Ni-NTA agarose column (ThermoFisher Scientific, Waltham, MA, USA) following the supplier’s protocol. Protein concentration

was determined based on the Bradford method [104] using a Pierce™ Coomassie Protein Assay Kit, following the instructions of the manufacturer.

#### 4.2.4. Characterization of TmAmyA Variants

An alcoholysis reaction was started by adding 20 U ( $\mu\text{eq}$  of dextrose produced per min) of enzyme to 1 mL of 100 mg/mL starch in 10% butanol, 50 mM Tris, 150 mM NaCl, and 2 mM  $\text{CaCl}_2$  at pH 7.0, and measured after 12 h of reaction at 85 °C. Hydrolysis and transglycosylation products were measured as reducing sugars by DNS reagent [105] and as the formation of butyl glycoside by HPLC after digestion of the reaction products with glucoamylase (Sigma-Aldrich, St. Louis, MO, USA), respectively. HPLC analysis was performed in a Waters-Millipore 510 HPLC system equipped with an automatic sampler (model 717 Plus, Waters Corp., Milford, MA, USA), a refractive-index detector (Waters 410, Waters Corp., Milford, MA, USA) and a Hypersil GOLD™ Amino column (Thermo Scientific, Wilford, UK), using acetonitrile:water (80:20) as the mobile phase at a flow rate of 1.0 mL/min.

#### 4.2.5. Characterization of TmGTase Variants

Tranglycosylation activity was measured in the reactions, started by the addition of 12  $\mu\text{g}$  of enzyme to 1 mL of 5 mg/mL maltoheptaose in 50 mM Tris, 150 mM NaCl, 2 mM  $\text{CaCl}_2$  pH 7.0 and kept up at 70 °C. Finally, quantification of sugar complex with iodine reagent [106–109] after 1, 2, 5, 10, 15, 30, 60 and 90 min was measured at 580 nm and reported as the starch equivalents produced.

In other independent experiment, hydrolytic reactions were started by the addition of 60  $\mu\text{g}$  of enzyme to 1 mL of 10 mg/mL starch in 50 mM Tris, 150 mM NaCl, 2 mM  $\text{CaCl}_2$  pH 7.0. Hydrolysis at 70 °C was measured by the reducing sugars produced using DNS reagent [105] after 6, 12, 18 and 24 h.

## 5. Conclusions

We present a method based on contact maps that reveals structural features important to function, even if they are not part of the protein sequence—such as the ions identified in this study. The study results showcase the ability to analyze residue contacts in order to bring new insights to our understanding of protein function beyond than the catalytic site. We believe this process can be extended beyond protein function to other properties such as thermal stability, tolerance to salinity, pH, or pressure. As in other studies employing contacts between residues, we observed that residue contacts are a tool that permits the identification of protein families that can be complemented by employing enrichment factors, as presented in this work, allowing the identification of subgroups. This could be used as a tool complementary to analysis through coevolution of residues. Both techniques overcome their limitations by interchanging information and using such as artificial intelligence that are more readily available every day.

**Supplementary Materials:** The following are available online, Table S1. Dataset S1. Internal group: collection of PDB structures employed to determine the enrichment factors [110–123]. Table S2. Dataset S2. External group: set of 3D structures used to test the ability of the enrichment factor to classify functionally enzymes in the GH13 family [36,124–136]. Table S3. Dataset S3. Enzymes to evaluate contact conservation and its correlation between enzymes. Table S4. Dataset S4: proteins reported as characterized by CAZy database [31]. For transferases, all sequences were considered and modelled using the Swiss model if needed. For the hydrolase only, sequences with reported structures were used. The ID for the structures is the PDB ID, while for models it is the Uniprot or GenBank ID, followed by the PDB ID of the template used. For the models, the Q-mean value is reported as resolution. All models were minimized using Rosetta. Table S5. Enrichment factors for contact 64–65 (D98-K99 in TmAmyA). These suggested the mutations D98P/K99A to make TmAmyA (a hydrolase) more like a transglycosidase. Residues that are not in the table have an enrichment factor of zero. Table S6. Enrichment values for the residues around residue 72 (F72 in TmGTase). While searching for a pair to mutate in TmGTase to augment its resemblance with a hydrolase, we found

this cluster of residues with high enrichment values, suggesting that these residues are important for residue function. The pair of residues to be mutated has an asterisk (\*). Table S7. Enrichment values for the residues around residue 273 (F273 in TmGTase). This trio of residues should be important for switching function as its residues have been selected both in hydrolases and transglycosidases. We mutated residues 274 and 279 (T274 and M279 in TmGTase), which are not in direct contact with the catalytic site. Figure S1. RMSF obtained from Molecular Dynamic (MD) simulation during 500 ns for TmAmyA wild type (red line) and K98P/D99A/H222Q mutant (black line). Here, the residue numbers are displaced by  $-29$  relative to Liebl et al. [36]. Figure S2. Conformational analysis of  $\chi$  dihedral angles of acid-baseresidue (Glu258) of TmAmyA for wild type (continue line) and K98P/D99A/H222Q (dotted lines). (a) dihedral angles  $\chi_3$  (b) dihedral angles  $\chi_2$  (c) dihedral angles  $\chi_1$ . Figure S3. RMSF obtained from molecular dynamic (MD) simulation during 500 ns for TmGTase wild type (red line), M279N (black line) and T274V/M279N mutant (blue line). Figure S4. Conformational analysis of  $\chi$  dihedral angles of acid-baseresidue (Glu216) of TmGTase for wild type (continue line), M279N (dashed line), and T274V/M279N (dotted lines). (a) dihedral angles  $\chi_3$  (b) dihedral angles  $\chi_2$  (c) dihedral angles  $\chi_1$ . Figure S5. Representation of frames around 0.01, 100, 200, 300, and 400 ns for glycosidases. The increase in the intensity of color corresponds with the increment of frames number (a) Comparison of TmAmyA wild type (blue) with mutant D98P/K99A/H222Q (gray) (b) Comparison of TmGTase wild type (blue) with mutant T274V/M279N (red). (c) Comparison of TmGTase wild type (blue) with mutant M279N (green). Figure S6. Change of average distance of D278 and E216 in TmGTase distance during MD simulation for wild type (blue line) M279N (red line) and T274V/M279N (gray line). Figure S7. Change of average distance of D310 and E258 in TmAmyA distance during MD simulation for wild type (red line) and D98P/D99A/H222Q (red line). Figure S8. Number of hydrogen bonds during MD simulation of TmGTase for K324 and D278: (a) wild type, (b) M279N, (c) T274V/M279N. Table S8. Primers used to create the mutants used in this study. Figure S9. Structure of TmGTase (PDB ID 1LWJ), highlighting the connection between the mutation sites and the catalytic residues (pink) including the binding subsites demarcated by acarbose (yellow). Residues T274 and M279 (orange sticks spheres) participate in a H-bond network (residues represented as cyan sticks). Dotted lines (red) indicate the distance between these residues. Only in T274V/M279N, D278 (green stick) and K324 (white stick) was a hydrogen bond detected during molecular dynamic analysis to form a hydrogen bond. Residue E226 (orange sticks) is part of a helix (residues 221–231) connected to the loop that contains the catalytic acid-baseresidue E216 (pink stick). Figure S10 Contact network of residue T274. (a) TmGTase and (b) its equivalent in *Aspergillus oryzae*  $\alpha$ -amylase where T274 corresponds to V293 (PDB ID: 7taa). Figure S11. Residues F72 and V86 (orange sticks) affect the mobility and inclination of a  $\beta$ -strand (green cartoon) reaching the catalytic site of TmGTase. Positions 72 and 86 were additionally mutated in TmGTase. Modifying these residues far from the active site had a detrimental effect on activity, disfavoring the transglycosidic activity preferentially. These residues interact indirectly with the active center through a  $\beta$ -strand constituted by residues 85 to 90, which form a super-secondary structure with the  $\beta$ -strands from 182–185 and 211–215, being in the last the acid-baseresidue. Figure S12. Residues K98 and D99 mediate the interaction of a calcium ion with the active site. A calcium ion (blueish green sphere) interacts with residues K98 and D99 (orange stick) in the TmAmyA 3D-structural model. Residues K98 and D99 connect the calcium ion to the +1 and +2 sites (H86, Y88) through a loop (green sticks). These sites are the acceptor binding positions during the transglycosidation reaction. The inhibitor acarbose is shown by yellow sticks to show the binding subsites. The catalytic residues D218 and E258 (red and pink, respectively) delimit the enzyme's active center.

**Author Contributions:** Conceptualization, R.A.A.-B. and G.S.-R.; methodology, R.A., L.O. and A.L.; software, R.A.A.-B.; validation, R.A.A.-B. and L.O.; formal analysis, R.A.A.-B.; investigation, R.A.A.-B.; resources, G.S.-R.; data curation, R.A.A.-B. and A.L.; writing—original draft preparation, R.A.A.-B., A.L. and G.S.; writing—review and editing, R.A.A.-B., A.L. and G.S.-R.; supervision, G.S.-R.; project administration, G.S.-R.; funding acquisition, G.S.-R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by DGAPA-UNAM through PAPIIT Grant No. IN211020 to G.S.R. Additionally, the project was accomplished thanks to the scholarships Num 443241 to R.A. and 958307 to A.L. supported by Consejo Nacional de Ciencia y Tecnología (CONACYT), México.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors are indebted to Jorge Arturo Yañez and Paul Gaytan for oligonucleotide synthesis and DNA sequencing, Jerome Verleyen, for computer cluster supervision, to Juan Manuel Hurtado-Ramírez and David Santiago Castañeda-Carreón for computer technical support, as well as Shirley Ainsworth and Omar Arriaga, for librarian assistance.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

**Samples Availability:** Samples of the plasmids are available from the authors.

## References

1. Chapman, J.; Ismail, A.E.; Dinu, C.Z. Industrial applications of enzymes: Recent advances, techniques, and outlooks. *Catalysts* **2018**, *8*, 238. [[CrossRef](#)]
2. Farooq, M.A.; Ali, S.; Hassan, A.; Tahir, H.M.; Mumtaz, S.; Mumtaz, S. Biosynthesis and industrial applications of  $\alpha$ -amylase: A review. *Arch. Microbiol.* **2021**, *203*, 1281–1292. [[CrossRef](#)] [[PubMed](#)]
3. Bell, E.L.; Finnigan, W.; France, S.P.; Green, A.P.; Hayes, M.A.; Hepworth, L.J.; Lovelock, S.L.; Niikura, H.; Osuna, S.; Romero, E.; et al. Biocatalysis. *Nat. Rev. Methods Prim.* **2021**, *1*, 46. [[CrossRef](#)]
4. Sheldon, R.A.; Brady, D. Streamlining design, engineering, and applications of enzymes for sustainable biocatalysis. *ACS Sustain. Chem. Eng.* **2021**, *9*, 8032–8052. [[CrossRef](#)]
5. Gargiulo, S.; Soumillion, P. Directed evolution for enzyme development in biocatalysis. *Curr. Opin. Chem. Biol.* **2021**, *61*, 107–113. [[CrossRef](#)]
6. de Pina Mariz, B.; Carvalho, S.; Batalha, I.L.; Pina, A.S. Artificial enzymes bringing together computational design and directed evolution. *Org. Biomol. Chem.* **2021**, *19*, 1915–1925. [[CrossRef](#)] [[PubMed](#)]
7. Alford, R.F.; Leaver-Fay, A.; Jeliakov, J.R.; O’Meara, M.J.; DiMaio, F.P.; Park, H.; Shapovalov, M.V.; Renfrew, P.D.; Mulligan, V.K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048. [[CrossRef](#)] [[PubMed](#)]
8. O’Meara, M.J.; Leaver-Fay, A.; Tyka, M.D.; Stein, A.; Houlihan, K.; Dimaio, F.; Bradley, P.; Kortemme, T.; Baker, D.; Snoeyink, J.; et al. Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J. Chem. Theory Comput.* **2015**, *11*, 609–622. [[CrossRef](#)] [[PubMed](#)]
9. Park, H.; Bradley, P.; Greisen, P.; Liu, Y.; Mulligan, V.K.; Kim, D.E.; Baker, D.; Dimaio, F. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **2016**, *12*, 6201–6212. [[CrossRef](#)] [[PubMed](#)]
10. Schmitz, S.; Ertelt, M.; Merkl, R.; Meiler, J. Rosetta design with co-evolutionary information retains protein function. *PLoS Comput. Biol.* **2021**, *17*, e1008568. [[CrossRef](#)]
11. Petrovic, D.; Risso, V.A.; Kamerlin, S.C.L.; Sanchez-Ruiz, J.M. Conformational dynamics and enzyme evolution. *J. R. Soc. Interface* **2018**, *15*, 20180330. [[CrossRef](#)] [[PubMed](#)]
12. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Židek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596. [[CrossRef](#)]
13. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
14. Crean, R.M.; Gardner, J.M.; Kamerlin, S.C.L. Harnessing Conformational Plasticity to Generate Designer Enzymes. *J. Am. Chem. Soc.* **2020**, *142*, 11324–11342. [[CrossRef](#)] [[PubMed](#)]
15. Bunzel, H.A.; Anderson, J.L.R.; Mulholland, A.J. Designing better enzymes: Insights from directed evolution. *Curr. Opin. Struct. Biol.* **2021**, *67*, 212–218. [[CrossRef](#)] [[PubMed](#)]
16. Goldman, A.D.; Beatty, J.T.; Landweber, L.F. The TIM Barrel Architecture Facilitated the Early Evolution of Protein-Mediated Metabolism. *J. Mol. Evol.* **2016**, *82*, 17–26. [[CrossRef](#)]
17. Hughes, A.L. Adaptive evolution after gene duplication. *Trends Genet.* **2002**, *18*, 433–434. [[CrossRef](#)]
18. Hughes, A.L. The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lon. B* **1994**, *256*, 119–124. [[CrossRef](#)]
19. Wierenga, R.K. The TIM-barrel fold: A versatile framework for efficient enzymes. *FEBS Lett.* **2001**, *492*, 193–198. [[CrossRef](#)]
20. Schmiedel, J.M.; Lehner, B. Determining protein structures using deep mutagenesis. *Nat. Genet.* **2019**, *51*, 1177–1186. [[CrossRef](#)]
21. Rollins, N.J.; Brock, K.P.; Poelwijk, F.J.; Stiffler, M.A.; Gauthier, N.P.; Sander, C.; Marks, D.S. Inferring protein 3D structure from deep mutation scans. *Nat. Genet.* **2019**, *51*, 1170–1176. [[CrossRef](#)] [[PubMed](#)]
22. Salinas, V.H.; Ranganathan, R. Coevolution-based inference of amino acid interactions underlying protein function. *Elife* **2018**, *7*, 1–20. [[CrossRef](#)]
23. Baussand, J.; Carbone, A. A combinatorial approach to detect coevolved amino acid networks in protein families of variable divergence. *PLoS Comput. Biol.* **2009**, *5*, e1000488. [[CrossRef](#)] [[PubMed](#)]

24. Ovchinnikov, S.; Kim, D.E.; Wang, R.Y.R.; Liu, Y.; Dimaio, F.; Baker, D. Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Proteins Struct. Funct. Bioinform.* **2016**, *84*, 67–75. [[CrossRef](#)]
25. Bandyopadhyay, D.; Huan, J.; Prins, J.; Snoeyink, J.; Wang, W.; Tropsha, A. Identification of family-specific residue packing motifs and their use for structure-based protein function prediction: II. Case studies and applications. *J. Comput. Aided. Mol. Des.* **2009**, *23*, 785–797. [[CrossRef](#)]
26. Bandyopadhyay, D.; Huan, J.; Prins, J.; Snoeyink, J.; Wang, W.; Tropsha, A. Identification of family-specific residue packing motifs and their use for structure-based protein function prediction: I. Method development. *J. Comput. Aided. Mol. Des.* **2009**, *23*, 773–784. [[CrossRef](#)]
27. Yeang, C.H.; Haussler, D. Detecting coevolution in and among protein domains. *PLoS Comput. Biol.* **2007**, *3*, 2122–2134. [[CrossRef](#)] [[PubMed](#)]
28. Russ, W.P.; Figliuzzi, M.; Stocker, C.; Barrat-Charlaix, P.; Socolich, M.; Kast, P.; Hilvert, D.; Monasson, R.; Cocco, S.; Weigt, M.; et al. An evolution-based model for designing chorisate mutase enzymes. *Science* **2020**, *369*, 440–445. [[CrossRef](#)]
29. Wang, X.; Jing, X.; Deng, Y.; Nie, Y.; Xu, F.; Xu, Y.; Zhao, Y.; Hunt, J.F.; Montelione, G.T.; Szyperski, T. Evolutionary coupling saturation mutagenesis: Coevolution-guided identification of distant sites influencing *Bacillus naganensis* pullulanase activity. *FEBS Lett.* **2020**, *594*, 799–812. [[CrossRef](#)]
30. Tondnevis, F.; Dudenhausen, E.E.; Miller, A.M.; McKenna, R.; Altschul, S.F.; Bloom, L.B.; Neuwald, A.F. Deep Analysis of Residue Constraints (DARC): Identifying determinants of protein functional specificity. *Sci. Rep.* **2020**, *10*, 1691. [[CrossRef](#)]
31. Glycogenomics Group at AFMB Carbohydrate-Active enZymes Database. Available online: <http://www.cazy.org/> (accessed on 25 October 2019).
32. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [[CrossRef](#)] [[PubMed](#)]
33. Møller, M.S.; Svensson, B. Structural biology of starch-degrading enzymes and their regulation. *Curr. Opin. Struct. Biol.* **2016**, *40*, 33–42. [[CrossRef](#)] [[PubMed](#)]
34. Uitdehaag, J.C.; Mosi, R.; Kalk, K.H.; van der Veen, B.A.; Dijkhuizen, L.; Withers, S.G.; Dijkstra, B.W. X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the alpha-amylase family. *Nat. Struct. Biol.* **1999**, *6*, 432–436. [[CrossRef](#)]
35. Tran, P.L.; Cha, H.J.; Lee, J.S.; Park, S.H.; Woo, E.J.; Park, K.H. Introducing transglycosylation activity in *Bacillus licheniformis*  $\alpha$ -amylase by replacement of His235 with Glu. *Biochem. Biophys. Res. Commun.* **2014**, *451*, 541–547. [[CrossRef](#)]
36. Liebl, W.; Stemplinger, I.; Ruile, P. Properties and gene structure of the *Thermotoga maritima* alpha-amylase AmyA, a putative lipoprotein of a hyperthermophilic bacterium. *J. Bacteriol.* **1997**, *179*, 941–948. [[CrossRef](#)]
37. Roujeinikova, A.; Raasch, C.; Sedelnikova, S.; Liebl, W.; Rice, D.W. Crystal structure of *Thermotoga maritima* 4- $\alpha$ -glucanotransferase and its acarbose complex: Implications for substrate specificity and catalysis. *J. Mol. Biol.* **2002**, *321*, 149–162. [[CrossRef](#)]
38. Blakeney, A.B.; Stone, B.A. Activity and action pattern of *Bacillus licheniformis*  $\alpha$ -amylase in aqueous ethanol. *FEBS Lett.* **1985**, *186*, 229–232. [[CrossRef](#)]
39. Kuriki, T.; Kaneko, H.; Yanase, M.; Takata, H.; Shimada, J.; Handa, S.; Takada, T.; Umeyama, H.; Okada, S. Controlling substrate preference and transglycosylation activity of neopullulanase by manipulating steric constraint and hydrophobicity in active center. *J. Biol. Chem.* **1996**, *271*, 17321–17329. [[CrossRef](#)]
40. Cha, H.-J.; Yoon, H.-G.; Kim, Y.-W.; Lee, H.-S.; Kim, J.-W.; Kweon, K.-S.; Oh, B.-H.; Park, K.-H. Molecular and enzymatic characterization of a maltogenic amylase that hydrolyzes and transglycosylates acarbose. *Eur. J. Biochem.* **1998**, *253*, 251–262. [[CrossRef](#)] [[PubMed](#)]
41. Geronimo, I.; Payne, C.M.; Sandgren, M. The role of catalytic residue pKa on the hydrolysis/transglycosylation partition in family 3  $\beta$ -glucosidases. *Org. Biomol. Chem.* **2018**, *16*, 316–324. [[CrossRef](#)]
42. Kosugi, T.; Hayashi, S. Crucial Role of Protein Flexibility in Formation of a Stable Reaction Transition State in an  $\alpha$ -Amylase Catalysis. *J. Am. Chem. Soc.* **2012**, *134*, 7045–7055. [[CrossRef](#)] [[PubMed](#)]
43. Damián-Almazo, J.Y.; Moreno, A.; López-Munguía, A.; Soberón, X.; González-Muñoz, F.; Saab-Rincón, G. Enhancement of the alcoholytic activity of  $\alpha$ -amylase AmyA from *Thermotoga maritima* MSB8 (DSM 3109) by site-directed mutagenesis. *Appl. Environ. Microbiol.* **2008**, *74*, 5168–5177. [[CrossRef](#)]
44. Rivera, M.H.; López-Munguía, A.; Soberón, X.; Saab-Rincón, G. A-Amylase from *Bacillus licheniformis* mutants near to the catalytic site: Effects on hydrolytic and transglycosylation activity. *Protein Eng.* **2003**, *16*, 505–514. [[CrossRef](#)] [[PubMed](#)]
45. Saab-Rincón, G.; Del-Río, G.; Santamaría, R.I.; López-Munguía, A.; Soberón, X. Introducing transglycosylation activity in a liquefying  $\alpha$ -amylase. *FEBS Lett.* **1999**, *453*, 100–106. [[CrossRef](#)]
46. Teze, D.; Hendrickx, J.; Dion, M.; Tellier, C.; Woods, V.L.; Tran, V.; Sanejouand, Y.H. Conserved water molecules in family 1 glycosidases: A DXMS and molecular dynamics study. *Biochemistry* **2013**, *52*, 5900–5910. [[CrossRef](#)]
47. David, B.; Arnaud, P.; Tellier, C.; Sanejouand, Y.-H. Toward the design of efficient transglycosidases: The case of the GH1 of *Thermus thermophilus*. *Protein Eng. Des. Sel.* **2019**, *32*, 309–316. [[CrossRef](#)]
48. Lundemo, P.; Karlsson, E.N.; Adlercreutz, P. Eliminating hydrolytic activity without affecting the transglycosylation of a GH1  $\beta$ -glucosidase. *Appl. Microbiol. Biotechnol.* **2017**, *101*, 1121–1131. [[CrossRef](#)]

49. Tran, L.T.; Blay, V.; Luang, S.; Eurtivong, C.; Choknud, S.; González-Díaz, H.; Ketudat Cairns, J.R. Engineering faster transglycosidases and their acceptor specificity. *Green Chem.* **2019**, *21*, 2823–2836. [CrossRef]
50. Kelly, R.M.; Leemhuis, H.; Dijkhuizen, L. Conversion of a Cyclodextrin Glucanotransferase into an  $\alpha$ -Amylase: Assessment of Directed Evolution Strategies. *Biochemistry* **2007**, *46*, 11216–11222. [CrossRef]
51. Leemhuis, H.; Rozeboom, H.J.; Wilbrink, M.; Euverink, G.-J.W.; Dijkstra, B.W.; Dijkhuizen, L. Conversion of Cyclodextrin Glycosyltransferase into a Starch Hydrolase by Directed Evolution: The Role of Alanine 230 in Acceptor Subsite +1. *Biochemistry* **2003**, *42*, 7518–7526. [CrossRef]
52. Kim, T.-J.; Park, C.-S.; Cho, H.-Y.; Cha, S.-S.; Kim, J.-S.; Lee, S.-B.; Moon, T.-W.; Kim, J.-W.; Oh, B.-H.; Park, K.-H. Role of the Glutamate 332 Residue in the Transglycosylation Activity of Thermus Maltogenic Amylase. *Biochemistry* **2000**, *39*, 6773–6780. [CrossRef]
53. Lin, X.; Hong, T.; Mu, Y.; Torres, J. Identification of residues involved in water versus glycerol selectivity in aquaporins by differential residue pair coevolution. *Biochim. Biophys. Acta Biomembr.* **2012**, *1818*, 907–914. [CrossRef] [PubMed]
54. Chakrabarti, S.; Panchenko, A.R. Coevolution in defining the functional specificity. *Proteins Struct. Funct. Bioinform.* **2009**, *75*, 231–240. [CrossRef] [PubMed]
55. Vehlow, C.; Stehr, H.; Winkelmann, M.; Duarte, J.M.; Petzold, L.; Dinse, J.; Lappe, M. CMView: Interactive contact map visualization and analysis. *Bioinformatics* **2011**, *27*, 1573–1574. [CrossRef]
56. Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Giuliani, A. Protein Contact Networks: An Emerging Paradigm in Chemistry. *Chem. Rev.* **2013**, *113*, 1598–1613. [CrossRef]
57. Roujeinikova, A.; Raasch, C.; Sedelnikova, S.; Liebl, W.; Rice, D.W. Crystallization and preliminary X-ray crystallographic studies on 4- $\alpha$ -glucanotransferase from *Thermotoga maritima*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2001**, *57*, 1046–1047. [CrossRef]
58. Available online: <https://swissmodel.expasy.org/assess> (accessed on 14 October 2019).
59. Holmes, J.B.; Tsai, J. Characterizing Conserved Structural Contacts by Pairwise Relative Contacts and Relative Packing Groups. *J. Mol. Biol.* **2005**, *354*, 706–721. [CrossRef]
60. Koropatkin, N.M.; Smith, T.J. Article SusG: A Unique Cell-Membrane-Associated  $\alpha$ -Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. *Struct. Des.* **2010**, *18*, 200–215. [CrossRef] [PubMed]
61. Koutrouli, M.; Karatzas, E.; Paez-Espino, D.; Pavlopoulos, G.A. A Guide to Conquer the Biological Network Era Using Graph Theory. *Front. Bioeng. Biotechnol.* **2020**, *8*, 34. [CrossRef]
62. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1978**, *1*, 215–239. [CrossRef]
63. Doshi, U.; Holliday, M.J.; Eisenmesser, E.Z.; Hamelberg, D. Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 4735–4740. [CrossRef] [PubMed]
64. Campbell, E.; Kaltenbach, M.; Correy, G.J.; Carr, P.D.; Porebski, B.T.; Livingstone, E.K.; Afriat-Jurnou, L.; Buckle, A.M.; Weik, M.; Hollfelder, F.; et al. The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **2016**, *12*, 944–950. [CrossRef]
65. Van Der Kamp, M.W.; Prentice, E.J.; Kraakman, K.L.; Connolly, M.; Mulholland, A.J.; Arcus, V.L. Dynamical origins of heat capacity changes in enzyme-catalysed reactions. *Nat. Commun.* **2018**, *9*, 1–7. [CrossRef]
66. Kong, D.; Wang, L.; Su, L.; Wu, J. Effect of Leu 277 on Disproportionation and Hydrolysis Activity in *Bacillus stearothermophilus* NO2 Cyclodextrin Glycosyltransferase. *Appl. Environ. Microbiol.* **2021**, *87*, 1–13. [CrossRef] [PubMed]
67. Li, H.; Robertson, A.D.; Jensen, J.H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins Struct. Funct. Bioinform.* **2005**, *61*, 704–721. [CrossRef] [PubMed]
68. Zhu, Q.; Mai, U.; Pfeiffer, W.; Janssen, S.; Asnicar, F.; Sanders, J.G.; Belda-ferre, P.; Al-ghalith, G.A.; Kopylova, E.; McDonald, D.; et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun.* **2019**, *10*, 5477. [CrossRef]
69. Stam, M.R.; Danchin, E.G.J.; Rancurel, C.; Coutinho, P.M.; Henrissat, B. Dividing the large glycoside hydrolase family 13 into subfamilies: Towards improved functional annotations of  $\alpha$ -amylase-related proteins. *Protein Eng. Des. Sel.* **2006**, *19*, 555–562. [CrossRef]
70. Janeček, S.  $\alpha$ -amylase family: Molecular biology and evolution. *Prog. Biophys. Mol. Biol.* **1997**, *67*, 67–97. [CrossRef]
71. Zhao, J.; Tandrup, T.; Bissaro, B.; Barbe, S.; Poulsen, J.N.; Andr, I.; Dumon, C.; Lo, L.; Donohue, M.J.O. Probing the determinants of the transglycosylation / hydrolysis partition in a retaining  $\alpha$ -L-arabinofuranosidase. *New Biotechnol. J.* **2021**, *62*, 68–78. [CrossRef]
72. Biswas, P.; Adhikari, A.; Pal, U.; Singh, P.; Das, M.; Saha-Dasgupta, T.; Choudhury, S.S.; Das, R.; Pal, S.K. Flexibility modulates the catalytic activity of a thermostable enzyme: Key information from optical spectroscopy and molecular dynamics simulation. *Soft Matter* **2020**, *16*, 3050–3062. [CrossRef] [PubMed]
73. Seo, D.-H.; Jung, J.-H.; Jung, D.-H.; Park, S.; Yoo, S.-H.; Kim, Y.-R.; Park, C.-S. An unusual chimeric amylosucrase generated by domain-swapping mutagenesis. *Enzyme Microb. Technol.* **2016**, *86*, 7–16. [CrossRef]
74. Seo, D.-H.; Jung, J.-H.; Park, C.-S. Improved polymerization activity of *Deinococcus geothermalis* amylosucrase by semi-rational design: Effect of loop flexibility on the polymerization reaction. *Int. J. Biol. Macromol.* **2019**, *130*, 177–185. [CrossRef] [PubMed]
75. Xiang, G.; Leemhuis, H.; van der Maarel, M.J.E.C. Structural elements determining the transglycosylating activity of glycoside hydrolase family 57 glycogen branching enzymes. *Proteins Struct. Funct. Bioinform.* **2021**, prot.26200. [CrossRef] [PubMed]
76. Holm, L.; Koivula, A.K.; Lehtovaara, P.M.; Hemminki, A.; Knowles, J.K.C. Random mutagenesis used to probe the structure and function of *Bacillus stearothermophilus* alpha-amylase. *Protein Eng. Des. Sel.* **1990**, *3*, 181–191. [CrossRef]

77. Wang, C.; Huang, R.; He, B.; Du, Q. Improving the thermostability of alpha-amylase by combinatorial coevolving-site saturation mutagenesis. *BMC Bioinform.* **2012**, *13*, 263. [CrossRef]
78. Hleap, J.S.; Blouin, C. The response to selection in Glycoside Hydrolase Family 13 structures: A comparative quantitative genetics approach. *PLoS ONE* **2018**, *13*, e0196135. [CrossRef] [PubMed]
79. Xu, Y.; Verma, D.; Sheridan, R.P.; Liaw, A.; Ma, J.; Marshall, N.M.; McIntosh, J.; Sherer, E.C.; Svetnik, V.; Johnston, J.M. Deep Dive into Machine Learning Models for Protein Engineering. *J. Chem. Inf. Model.* **2020**, *60*, 2773–2790. [CrossRef]
80. Narayanan, H.; Dingfelder, F.; Butté, A.; Lorenzen, N.; Sokolov, M.; Arosio, P. Machine Learning for Biologics: Opportunities for Protein Engineering, Developability, and Formulation. *Trends Pharmacol. Sci.* **2021**, *42*, 151–165. [CrossRef]
81. Timonina, D.; Sharapova, Y.; Švedas, V.; Suplatov, D. Bioinformatic analysis of subfamily-specific regions in 3D-structures of homologs to study functional diversity and conformational plasticity in protein superfamilies. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1302–1311. [CrossRef]
82. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [CrossRef]
83. Holm, L. Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **2010**, *38*, W545–W549. [CrossRef]
84. James, M.; Murtola, T.; Schulz, R.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Softw. X* **2015**, *2*, 19–25. [CrossRef]
85. Sefidbakht, Y.; Siadat, O.R.; Taheri, F. Homology modeling and molecular dynamics study on Schwannomyces occidentalis alpha- amylase. *J. Biomol. Struct. Dyn.* **2016**, *35*, 574–584. [CrossRef]
86. Protein Data Bank. Available online: <https://www.rcsb.org/> (accessed on 26 July 2021).
87. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef]
88. Fiser, A.; Kinh, R.; Do, G.; Andrej, S. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753–1773. [CrossRef]
89. Grace. Available online: <http://plasma-gate.weizmann.ac.il/Grace/> (accessed on 10 July 2021).
90. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [CrossRef]
91. Bienert, S.; Waterhouse, A.; de Beer, T.A.P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository—New features and functionality. *Nucleic Acids Res.* **2017**, *45*, D313–D319. [CrossRef]
92. Guex, N.; Peitsch, M.C.; Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **2009**, *30*, S162–S173. [CrossRef]
93. Studer, G.; Tauriello, G.; Bienert, S.; Biasini, M.; Johner, N.; Schwede, T. ProMod3—A versatile homology modelling toolbox. *PLOS Comput. Biol.* **2021**, *17*, e1008667. [CrossRef]
94. Studer, G.; Rempfer, C.; Waterhouse, A.M.; Gumienny, R.; Haas, J.; Schwede, T. QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics* **2020**, *36*, 1765–1771. [CrossRef]
95. Studer, G.; Biasini, M.; Schwede, T. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics* **2014**, *30*, i505–i511. [CrossRef] [PubMed]
96. Benkert, P.; Biasini, M.; Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **2011**, *27*, 343–350. [CrossRef]
97. Bertoni, M.; Kiefer, F.; Biasini, M.; Bordoli, L.; Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci. Rep.* **2017**, *7*, 10480. [CrossRef]
98. Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. IDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **2013**, *29*, 2722–2728. [CrossRef] [PubMed]
99. Bowie, J.U.; Lüthy, R.; Eisenberg, D. A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure. *Science* **1991**, *253*, 164–170. [CrossRef]
100. Lüthy, R.; Bowie, J.U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **1992**, *356*, 83–85. [CrossRef]
101. Wiederstein, M.; Sippl, M.J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, *35*, W407–W410. [CrossRef]
102. Sippl, M.J. Recognition of errors in three-dimensional structures of proteins. *Proteins Struct. Funct. Genet.* **1993**, *17*, 355–362. [CrossRef]
103. Sarkar, G.; Sommer, S.S. The “megaprimer” method of site-directed mutagenesis. *Biotechniques* **1990**, *8*, 404–407.
104. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248–254. [CrossRef]
105. Miller, G.L. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.* **1959**, *31*, 426–428. [CrossRef]
106. Xiao, Z.; Storms, R.; Tsang, A. A quantitative starch-iodine method for measuring alpha-amylase and glucoamylase activities. *Anal. Biochem.* **2006**, *351*, 146–148. [CrossRef]
107. Zeeman, S.C.; Tiessen, A.; Pilling, E.; Kato, K.L.; Donald, A.M.; Smith, A.M. Starch Synthesis in Arabidopsis. Granule Synthesis, Composition, and Structure. *Plant Physiol.* **2002**, *129*, 516–529. [CrossRef] [PubMed]

108. Mould, D.L.; Syngde, R.L.M. Separations of polysaccharides related to starch by electrokinetic ultrafiltration in collodion membranes. *Biochem. J.* **1954**, *58*, 571–585. [[CrossRef](#)]
109. Mould, D.L. Potentiometric and spectrophotometric studies of complexes of hydrolysis products of amylose with iodine and potassium iodide. *Biochem. J.* **1954**, *58*, 593–600. [[CrossRef](#)]
110. Wind, R.D.; Liebl, W.; Buitelaar, R.M.; Penninga, D.; Spreinat, A.; Dijkhuizen, L.; Bahl, H. Cyclodextrin formation by the thermostable alpha-amylase of *Thermoanaerobacterium thermosulfurigenes* EM1 and reclassification of the enzyme as a cyclodextrin glycosyltransferase. *Appl. Environ. Microbiol.* **1995**, *61*, 1257–1265. [[CrossRef](#)]
111. Nakamura, A.; Haga, K.; Yamane, K. Four aromatic residues in the active center of cyclodextrin glucanotransferase from alkalophilic *Bacillus* sp. 1011: Effects of replacements on substrate binding and cyclization characteristics. *Biochemistry* **1994**, *33*, 9929–9936. [[CrossRef](#)] [[PubMed](#)]
112. van der Veen, B.A.; van Alebeek, G.J.; Uitdehaag, J.C.; Dijkstra, B.W.; Dijkhuizen, L. The three transglycosylation reactions catalyzed by cyclodextrin glycosyltransferase from *Bacillus circulans* (strain 251) proceed via different kinetic mechanisms. *Eur. J. Biochem.* **2000**, *267*, 658–665. [[CrossRef](#)] [[PubMed](#)]
113. Park, H.-S.; Park, J.-T.; Kang, H.-K.; Cha, H.; Kim, D.-S.; Kim, J.-W.; Park, K.-H. TreX from *Sulfolobus solfataricus* ATCC 35092 displays isoamylase and 4-alpha-glucanotransferase activities. *Biosci. Biotechnol. Biochem.* **2007**, *71*, 1348–1352. [[CrossRef](#)]
114. Jørgensen, S.; Vorgias, C.E.; Antranikian, G. Cloning, Sequencing, Characterization, and Expression of an Extracellular  $\alpha$ -Amylase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* in *Escherichia coli* and *Bacillus subtilis*. *J. Biol. Chem.* **1997**, *272*, 16335–16342. [[CrossRef](#)]
115. Robyt, J.F.; French, D. The action pattern of porcine pancreatic alpha-amylase in relationship to the substrate binding site of the enzyme. *J. Biol. Chem.* **1970**, *245*, 3917–3927. [[CrossRef](#)]
116. Kramhøft, B.; Bak-Jensen, K.S.; Mori, H.; Juge, N.; Nøhr, J.; Svensson, B. Involvement of individual subsites and secondary substrate binding sites in multiple attack on amylose by barley alpha-amylase. *Biochemistry* **2005**, *44*, 1824–1832. [[CrossRef](#)]
117. Tonozuka, T.; Ohtsuka, M.; Mogi, S.; Sakai, H.; Ohta, T.; Sakano, Y. A neopullulanase-type alpha-amylase gene from *Thermoactinomyces vulgaris* R-47. *Biosci. Biotechnol. Biochem.* **1993**, *57*, 395–401. [[CrossRef](#)]
118. Kanai, R.; Haga, K.; Akiba, T.; Yamane, K.; Harata, K. Biochemical and crystallographic analyses of maltohexaose-producing amylase from alkalophilic *Bacillus* sp. 707. *Biochemistry* **2004**, *43*, 14047–14056. [[CrossRef](#)]
119. Brayer, G.D.; Sidhu, G.; Maurus, R.; Rydberg, E.H.; Braun, C.; Wang, Y.; Nguyen, N.T.; Overall, C.M.; Withers, S.G. Subsite mapping of the human pancreatic alpha-amylase active site through structural, kinetic, and mutagenesis techniques. *Biochemistry* **2000**, *39*, 4778–4791. [[CrossRef](#)]
120. Choi, J.-H.; Lee, H.; Kim, Y.-W.; Park, J.-T.; Woo, E.-J.; Kim, M.-J.; Lee, B.-H.; Park, K.-H. Characterization of a novel debranching enzyme from *Nostoc punctiforme* possessing a high specificity for long branched chains. *Biochem. Biophys. Res. Commun.* **2009**, *378*, 224–229. [[CrossRef](#)]
121. Tan, T.-C.; Mijts, B.N.; Swaminathan, K.; Patel, B.K.C.; Divne, C. Crystal structure of the polyextremophilic alpha-amylase AmyB from *Halothermothrix orenii*: Details of a productive enzyme-substrate complex and an N domain with a role in binding raw starch. *J. Mol. Biol.* **2008**, *378*, 852–870. [[CrossRef](#)]
122. Shipman, J.A.; Cho, K.H.; Siegel, H.A.; Salyers, A.A. Physiological characterization of SusG, an outer membrane protein essential for starch utilization by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **1999**, *181*, 7206–7211. [[CrossRef](#)]
123. Nitta, Y.; Mizushima, M.; Hiromi, K.; Ono, S. Influence of molecular structures of substrates and analogues on Taka-amylase A catalyzed hydrolyses. I. Effect of chain length of linear substrates. *J. Biochem.* **1971**, *69*, 567–576.
124. Nitschke, L.; Heeger, K.; Bender, H.; Schulz, G.E. Molecular cloning, nucleotide sequence and expression in *Escherichia coli* of the beta-cyclodextrin glycosyltransferase gene from *Bacillus circulans* strain no. 8. *Appl. Microbiol. Biotechnol.* **1990**, *33*, 542–546. [[CrossRef](#)]
125. Li, Z.; Li, B.; Gu, Z.; Du, G.; Wu, J.; Chen, J. Extracellular expression and biochemical characterization of alpha-cyclodextrin glycosyltransferase from *Paenibacillus macerans*. *Carbohydr. Res.* **2010**, *345*, 886–892. [[CrossRef](#)]
126. Violet, M.; Meunier, J.C. Kinetic study of the irreversible thermal denaturation of *Bacillus licheniformis* alpha-amylase. *Biochem. J.* **1989**, *263*, 665–670. [[CrossRef](#)]
127. Nakada, T.; Kubota, M.; Sakai, S.; Tsujisaka, Y. Purification and characterization of two forms of maltotetraose-forming amylase from *Pseudomonas stutzeri*. *Agric. Biol. Chem.* **1990**, *54*, 737–743. [[CrossRef](#)]
128. Tonozuka, T.; Mogi, S.; Shimura, Y.; Ibuka, A.; Sakai, H.; Matsuzawa, H.; Sakano, Y.; Ohta, T. Comparison of primary structures and substrate specificities of two pullulan-hydrolyzing alpha-amylases, TVA I and TVA II, from *Thermoactinomyces vulgaris* R-47. *Biochim. Biophys. Acta* **1995**, *1252*, 35–42. [[CrossRef](#)]
129. Mijts, B.N.; Patel, B.K.C. Cloning, sequencing and expression of an alpha-amylase gene, amyA, from the thermophilic halophile *Halothermothrix orenii* and purification and biochemical characterization of the recombinant enzyme. *Microbiology* **2002**, *148*, 2343–2349. [[CrossRef](#)]
130. Plaza-Vinuesa, L.; Hernandez-Hernandez, O.; Moreno, F.J.; de Las Rivas, B.; Muñoz, R. Unravelling the diversity of glycoside hydrolase family 13  $\alpha$ -amylases from *Lactobacillus plantarum* WCFS1. *Microb. Cell Fact.* **2019**, *18*, 183. [[CrossRef](#)]
131. Buedenbender, S.; Schulz, G.E. Structural base for enzymatic cyclodextrin hydrolysis. *J. Mol. Biol.* **2009**, *385*, 606–617. [[CrossRef](#)] [[PubMed](#)]

132. Han, P.; Zhou, P.; Hu, S.; Yang, S.; Yan, Q.; Jiang, Z. A novel multifunctional  $\alpha$ -amylase from the thermophilic fungus *Malbranchea cinnamomea*: Biochemical characterization and three-dimensional structure. *Appl. Biochem. Biotechnol.* **2013**, *170*, 420–435. [[CrossRef](#)] [[PubMed](#)]
133. Li, D.; Park, J.-T.; Li, X.; Kim, S.; Lee, S.; Shim, J.-H.; Park, S.-H.; Cha, J.; Lee, B.-H.; Kim, J.-W.; et al. Overexpression and characterization of an extremely thermostable maltogenic amylase, with an optimal temperature of 100 degrees C, from the hyperthermophilic archaeon *Staphylothermus marinus*. *N. Biotechnol.* **2010**, *27*, 300–307. [[CrossRef](#)]
134. Park, J.-T.; Song, H.-N.; Jung, T.-Y.; Lee, M.-H.; Park, S.-G.; Woo, E.-J.; Park, K.-H. A novel domain arrangement in a monomeric cyclodextrin-hydrolyzing enzyme from the hyperthermophile *Pyrococcus furiosus*. *Biochim. Biophys. Acta* **2013**, *1834*, 380–386. [[CrossRef](#)] [[PubMed](#)]
135. Jun, S.-Y.; Kim, J.-S.; Choi, K.-H.; Cha, J.; Ha, N.-C. Structure of a novel  $\alpha$ -amylase AmyB from *Thermotoga neapolitana* that produces maltose from the nonreducing end of polysaccharides. *Acta Crystallogr. D. Biol. Crystallogr.* **2013**, *69*, 442–450. [[CrossRef](#)]
136. Tomazic, S.J.; Klibanov, A.M. Mechanisms of irreversible thermal inactivation of *Bacillus* alpha-amylases. *J. Biol. Chem.* **1988**, *263*, 3086–3091. [[CrossRef](#)]