Modifying the catalytic preference of alpha-amylase towards n-alkanes for bioremediation purposes using *in silico* strategies

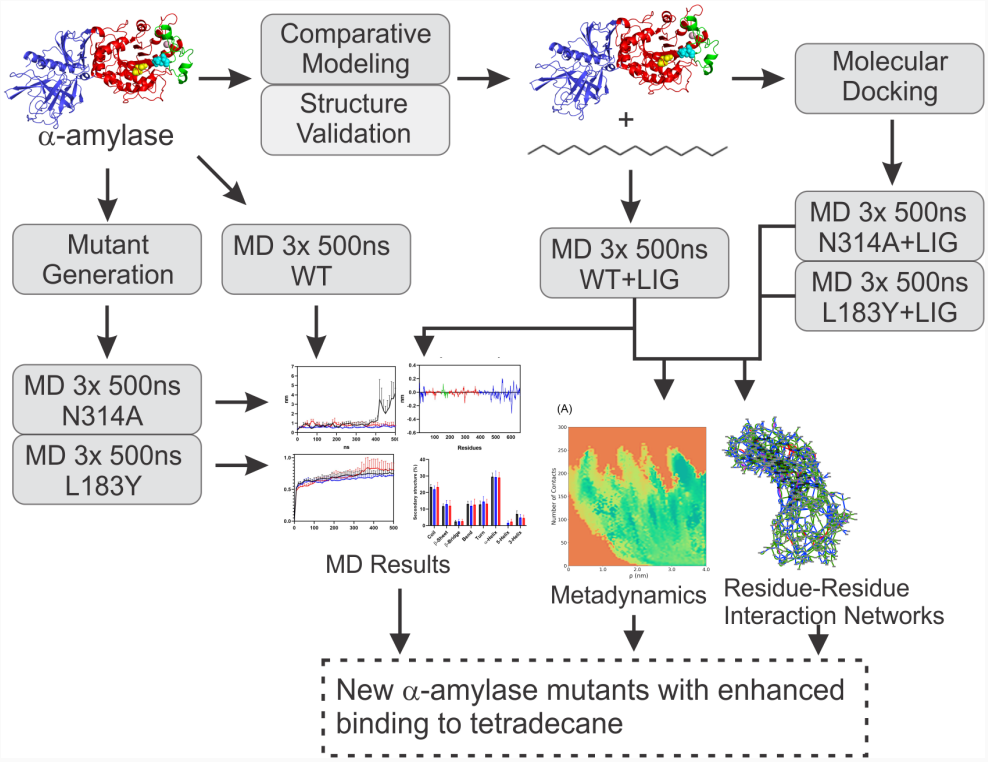
Ederson Sales Moreira Pinto[1,†], Bruno César Feltes[2,3,4,†], Conrado Pedebos[5], Márcio Dorn[1,4,6,*]

[1]Center of Biotechnology, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[2]Department of Genetics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[3]Department of Biophysics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[4]Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[5]School of Chemistry, University of Southampton, Southampton, United Kingdom
[6]National Institute of Science and Technology - Forensic Science, Porto Alegre, Brazil
†These authors contributed equaly to this work
∗Corresponding Author

Abstract

Since the beginning of oil exploration, whole ecosystems have been affected by accidents and bad practices involving petroleum compounds. In this sense, bioremediation stands out as the cheapest and most eco-friendly alternatives to reverse the damage done in oil-impacted areas. However, more efforts must be made to engineer enzymes that could be used in the bioremediation process. Interestingly, a recent work described that α-amylase, one of the most evolutionary conserved enzymes, was able to promiscuously degrade n-alkanes, a class of molecules abundant in the petroleum admixture. Considering that α-amylase is expressed in almost all known organisms, and employed in numerous biotechnological processes, using it can be a great leap towards more efficient applications of enzyme or microorganism-consortia bioremediation approaches. In this work, we employed a strict computational approach to design new α-amylase mutants with potentially enhanced catalytic efficiency towards n-alkanes. Using *in silico* techniques, such as molecular docking, molecular dynamics, metadynamics, and residue-residue interaction networks, we generated mutants potentially more efficient for degrading n-alkanes, L183Y and N314A. Our results indicate that the new mutants have an increased binding rate for tetradecane, the longest n-alkane previously tested, which can reside in the catalytic center for more extended periods. Additionally, molecular dynamics and network analysis showed that the new mutations have no negative impact on protein structure than the WT. Our results aid in solidifying this enzyme as one more tool in the petroleum bioremediation toolbox.

Keywords: Alpha-Amylase, Molecular Dynamics, Bioremediation, n-alkanes, Metadynamics, Residue-Residue Interaction Networks



We describe two new α-amylase mutants that are more efficient for degrading tetradecane and propose an *in silico* protocol for protein engineering. Using molecular docking, molecular dynamics (1.5 µs per system), metadynamics, and residue-residue interaction networks, the two mutants, N314A and L183Y, have an increased binding rate for tetradecane, can reside in the catalytic center for extended periods, and have no adverse effect on protein structure, solidify this enzyme for bioremediation of petroleum-derived compounds.

INTRODUCTION

Despite the ongoing efforts to prevent environmental disasters concerning Oil spills, the consequences of its mismanagement is still a worldwide concern, and continuous efforts must be carried out to design new alternatives to treat affected areas. In this sense, the bioremediation process is currently deemed the most sustainable approach to deal with this issue due to its low-costs and non-invasive, eco-friendly nature[1]. Among the different bioremediation approaches, microorganism consortia are regarded as the most low-cost approach since their employment does not generate wastes and does not necessarily include protein engineering or other high-cost techniques. However, this efficiency is directly associated with exploring new enzymes that can degrade Petroleum Derived Compounds (PDC) or uncover new roles of known enzymes that could be adapted to this reality.

Between the multitude of known PDC, alkanes are the most abundant petroleum admixture compounds, comprising n-alkanes, isoalkanes, and naphthenes[2,3]. Alkanes were already detected in virtually all niches[4–12], and found to be more persistent in specific environments, such as water surface, than Polycyclic Aromatic Hydrocarbons (PAH)[13], which are regarded to be the most toxic PDC[2]. Therefore, the search for enzymes that can efficiently degrade n-alkanes is of utmost importance to quickly recuperate all affected areas.

A new research revealed that endo-1,4-α-D-glucan glucohydrolase (α-amylase), is able to degrade n-alkanes[14] and low-density polyethylene[15] (**Fig.1**A). The study tested two different α-amylases about their capability to degrade n-alkanes: one from *Bacillus subtilis* TB1, a bacteria found to survive in oil-spilled areas; and the other is a commercial one. The experiment briefly consisted of analyzing enzymatic biodegradation of n-alkanes, ranging from 10 to 14 carbons, by a crude enzyme solution. Analysis from gas chromatography showed that after 36h of incubation, C10 and C11 n-alkanes were not detected, C12-C14 n-alkanes diminished in this same period and absent after 72h. This result indicated that α-amylase has the promiscuous ability to degrade n-alkanes and that the catalytic efficiency decreases as the carbon chain length increases.

Additionally, *in silico* docking analysis corroborated that the longer the n-alkane carbon chain was, the larger the total binding energy. This is an exciting discovery, with a direct

influence on the bioremediation process by microbial consortium because α-amylase is a strongly evolutionary conserved enzyme, present in almost all known organisms[16]. As a matter of fact, in previous work, we demonstrated that α-amylase is one of the most well-known enzymes, with decades of structural and physical-chemical knowledge, adapted to a multitude of temperatures, pH, and even salinity[17]. Our study showed that there is enough available α-amylase data to allow it to be virtually adapted to any environment, which is fundamental to the bioremediation process. Therefore, this enzyme's ability to promiscuously degrade n-alkanes is a remarkable trait that should be further explored.

One plausible, low-cost, and potential strategy is to test new α-amylase mutants that could possess a higher binding rate constant to n-alkanes, which, in turn, might increase catalytic efficiency. In this sense, computational methods play a vital role in the continuous endeavors to diminish production and experimental costs. For protein, in particular, simulations by molecular dynamics (MD) and metadynamics (MTD) are grounded as valid and reliable strategies to access a protein's molecular behavior and its relationship to a given molecule.

In this work, we report the identification of two new α-amylase mutants using computational approaches and analyze them with MD and MTD simulations to access their potentially enhanced catalytic efficiency to tetradecane. We observed that both newly generated mutants not only retain the structural stability and folding, presenting no significant changes in the secondary structure profile but can better accommodate the tetradecane in the binding pocket, increasing residence time. Additionally, we investigated the binding energy between each α-amylase variant and the tetradecane, revealing that both mutations conferred an improvement of the affinity for the new non-polar ligand. Moreover, we scrutinized the possibility of structural changes derived from the effect of each mutation and the interaction of each α-amylase variant with tetradecane. Furthermore, the residue-residue interaction network showed no significant variation in the overall topology of the mutants, confirming their stability compared to the WT counterpart, implying that the mutants have no negative impact on protein structure. Lastly, the *in silico* protocol described in this work can be virtually employed in any protein engineering work to reduce the costs of wet-lab procedures and more efficient mutant variants.

METHODOLOGY

Protein structure, modeling and validation

The α-amylase WT structure from *B. subtilis*, containing 425 from 659 total amino acids, was retrieved from the Protein Data Bank (PDB)[18] under the code 1BAG[19]. A total of 234 amino acids, related to α-amylase domain C, were missing from the final structure. To model this missing segment, we performed comparative modeling using the Robetta server[20]. The final model was submitted to four different validation tests to guarantee its suitability, using the following tools: (i) Verify3D[21]; (ii) Protein Structure Analysis (ProSA-web)[22]; (iii) PROCHECK[23] and; Qualitative Model Energy Analysis (QMEAN)[24]. The final approved model was named AMY^{WT}.

Mutants generation

Mutants were generated by a semi-rational design approach[25] based on the Rosetta point mutant (pmut) scan application[26,27]. The *pmut* scan is an algorithm designed to predict protein stability by calculating the relative change in the structure energy. The application calculates the structure energy as a function of $\Delta\Delta$G, which is determined as ΔG for mutant structure minus ΔG for WT structure. If the $\Delta\Delta$G in the structure energy is favorable, the tested mutation is selected and stored; otherwise, it is excluded. A command flag is used to generate a PDB file for every hit that is a favorable mutation. Moreover, a combination of PDB files of the same protein in different configurations can be used as input to improve the *pmut* accuracy.Amidst the six WT systems (three of free enzyme and three of α-amylase + tetradecane), we selected PDBs from 0, 250, and 500 ns of each system trajectory and submitted each structure independently to *pmut*. Next, we submitted PDBs of 0, 250, and 500 ns together for each system. In the end, we had a list of thousands of mutations from each *pmut* run. We searched for mutations in regions that would directly affect the enzyme contact with the ligand without compromising the catalytic triad. Additionally, we selected mutations that would increase the binding pocket's affinity for the small, hydrophobic tetradecane. The command line input flags for *pmut* scan were: -ex1 -ex2 -extrachi_cutoff 1

-use_input_sc -ignore_unrecognized_res -no_his_his_pairE -multi_cool_annealer 10 -mute basic core –output_mutant_structures. We analyzed the most consistent mutant candidates, and two mutations were selected, named AMY^{N314A} and AMY^{L183Y}.

Tetradecane and Docking

The tetradecane molecule structure was retrieved from the PDB, and hydrogen atoms were added using the Avogadro software[28]. The new PDB file was submitted to the LigParGen server[29–31], which provides bond, angle, dihedral, and Lennard-Jones parameters for chosen force field (next section), along with the 1.14*CM1A-LBCC option, a tool for generating partial charges parameters for neutrally charged molecules, along with topology files for GROMACS.

Tetradecane was docked into the into the AMY^{WT} structure using Autodock tools 4.2[32] and Autodock Vina 1.1.2[33] software. The grid box was set centered at the α-carbon of the catalytic residue E249 and the box set's dimension with the number of points in x-dimension = y-dimension = z-dimension = 26, and the spacing set as 0.682 A, which covered all the active enzyme center. The ligand was then uploaded to the Autodock software. The exhaustiveness was set as 8. The same parameters were reproduced when docking the ligand to the AMY^{N314A} and AMY^{L183Y} mutants. This generated three new PDB files named AMY$^{WT}_{lig}$, AMY$^{L183Y}_{lig}$, and AMY$^{N314A}_{lig}$.

Molecular Dynamics Simulations

All Molecular Dynamics (MD) simulations were executed using the GROMACS package version 2018.1.[34,35]. The system was set inside a dodecahedron box. The OPLS-AA force field (FF)[36] was selected along to the water model TIP3P[37], under periodic boundary conditions. Na^+ and Cl^- ions were proportionally added to neutralize the system and simulate the physiological condition of 0.15 M. An energy minimization step was conducted employing the Steepest Descent algorithm. After the energy minimization, an equilibration step was applied. Covalent bonds were constrained using the LINCS algorithm[38], and an integration step of 2 fs was applied. The Particle Mesh Ewald method[39] was employed for calculation

of electrostatic interactions, along with the Parrinello-Rahman barostat[40] set with a two ps coupling constant. the V-rescale[41] was employed with a coupling constant of τ= 0.1. In the end there were six simulated systems, each done in triplicates: (i) 3x free AMY^{WT}; (ii) 3x AMY$^{WT}_{lig}$; (iii) 3x free AMY^{N314A}; (iv) 3x AMY$^{N314A}_{lig}$; (v) 3x free AMY^{L183Y} and; (vi) 3x AMY$^{L183Y}_{lig}$. Each MD simulation consisted of 500 ns each, a total of 1.5 μs per system.

Residue-Residue Interaction Networks

Residue-Residue Interaction Networks (RRIN) were created using the Residue Interaction Network Generator (RING 2.0) software[42]. To generate the RRIN, we investigated the time range with the lowest variances in the mean RMSD for each system and extracted 20 PDB frames from those time ranges for each replicate (i.e., MD1, MD2, and MD3), a total of 60 PDBs per system. We selected the most representative structure from the generated PDB pool using the Average Linkage Hierarchical clustering procedure[43] (available at http://www.sbg.bio.ic.ac.uk/ maxcluster) by calculating the protein with the lowest average distance to all other cluster members. This structure was then labeled as the cluster centroid, and the average distance was defined by the cluster's spread (or error).

The final model was submitted to RING 2.0 server taking into consideration the following parameters: (i) Distance thresholds: relaxed; (ii) Network Policy: closest; (iii) Interaction type: multiple; (iv) all other options disabled, except for skip water molecules. Networks were manipulated and analyzed in the Cytoscape v3.6.1. software[44]. The Cytoscape plugin CentiScaPe v2.2.[45] was employed to calculate the most topologically relevant residues. The used centralities were: (i) node degree; (ii) betweenness; and (iii) eigenvector.

Metadynamics

To obtain estimates of the free energy surfaces describing tetradecane binding to the α-amylases, a previously published protocol was followed using volume-based Metadynamics (MTD)[46]. This consists of employing a combination of the well-tempered metadynamics scheme[47] with a restraining potential in the shape of a sphere of finite radius (ρ_s). As collective variables (CV) for applying the history-dependent bias, the spherical coordinates

were used (ρ, θ, ϕ). The repulsive potential at the sphere's border allows for multiple recrossing events and proper binding of both bound and unbound states.

The well-tempered MTD variant was used, applying a bias factor of 20 for all simulations to rescale the Gaussian functions. An initial height of 1.2 $kJ \cdot mol^{-1}$ applied every one ps for the Gaussian hills was used, and a value of 40 Å was applied for ρ_s. Sigmas for CVs ρ, θ, ϕ were set as 1 Å, $\pi/8$ rad, $\pi/16$ rad, respectively. To keep the system within the reference frame and with CVs fixed, a structure alignment was conducted in every step. The free energies were calculated by reweighting as a function of two meaningful CVs, that is, ρ (the distance between the center of masses of the ligand and the catalytic site residues) and the number of contacts (C_n) between ligand and protein[46]. This was useful to discriminate between the bound (ρ between 0.6 and 0.9; C_n between X and Y) and unbound states (ρ higher than 3.0; C_n lower than 10). Statistical errors were obtained by employing the block analysis technique. Entropic correction of the free energy of binding ($\Delta\ G_b^0$) was applied, following equations from Capelli et al.[46], which account for the loss in translational entropy.

Fig.2 illustrates the workflow applied to this work.

RESULTS AND DISCUSSION

α-amylase structure, validation, and docking

The structure of the α-amylase is divided into three distinct domains (**Fig.1**A). Domain A, which is composed of two segments, from residues L42 to I141, and from T193 to G388, is the catalytic domain. It comprises the $(\alpha/\beta)_8$-Barrels, a conserved protein folding of an intern barrel of 8 β-strands surrounded by another barrel of 8 α-helices[48]. Domain B is the smallest one, composed of residues N142 to N192. It is adapted for binding to a structural Ca^{2+}[48]. Domain C, consisting of residues M1 to E41 and Q389 to H659, comprises a typical antiparallel β-sheet structure[48]. Three conserved amino acids constitute the active site: D217, located at the β4 strand of $(\alpha/\beta)_8$-Barrel, which acts as a catalytic nucleophile; E249 positioned at β5 strand, operates as a proton donor and; D310, situated at the β7 strand, which aids to stabilizing the transition-state during catalysis[49] (**Fig.1**B).

We retrieved from the PDB database the same α-amylase employed in the original study that showed its promiscuous activity towards n-alkanes. However, the crystal was missing a large segment, requiring a modeling step. There were two reasons for choosing the selected PDB structure, the obvious one for it being the same one utilized in the original work. The second one is presenting the closest conformation to the enzyme-substrate complex transition state (which is explained in the next section).

Our final α-amylase model (AMY^{WT}) consists of adding the missing segment belonging to domain C, which is absent from the PDB structure 1BAG. The 234 missing residues were modeled into a structure composed of β-sheets and loops, which is consistent with the literature description of this domain[48] (**Fig.1**A). Afterward, we employed four different metrics to evaluate the quality of AMY^{WT} (**Table.1, Figs.S1-S5, see S-Material**). All validation methods approved the model.

Molecular docking was employed to investigate the interaction between tetradecane and α-amylase variants. In the original work by Karimi and Biria, although molecular docking studies using n-alkanes ranging from ten to sixteen carbons were conducted, only n-alkanes ranging from ten to fourteen carbons were tested with actual enzymes[14]. Moreover, there is an inverse correlation between the n-alkanes carbon chain length and α-amylase efficiency. Thus, the tetradecane molecule was chosen as a model for this work due to the wet lab's experimental data to support it as an unmistakable reference.

Mutants generation, selection, and docking

We concentrated on the residues interacting directly with the tetradecane among the multitude of obtained mutants as possible structure stabilizers. Q104, L182, L183, L185, H221, and N314 are 4Å distant or less from the docked molecule, composing α-amylase binding pocket. The most consistent as well as biochemically relevant results from pmut which were L183Y and N314A.

Table 2 shows the score function from Autodock Vina for each α-amylase variant and the tetradecane molecule. WT α-amylase presented the lowest docking scores amidst the variants, with the lowest score of 3.5 and the highest score of 4.5, in absolute value (**Fig.1C**. The second-lowest scored was α-amylase variant was AMY^{N314A}, which presented the lowest

and highest scores, in absolute value, of 4.5 and 4.6, respectively **Fig.1E**. This mutant has the most consistent scoring outcome, deviating in just 0.1 between the least and best docking results. AMY^{L183Y} showed the best docking results, with absolute values of 4.9 and 5.2 (**Fig.1**D).

Although docking results do not directly indicate the affinity between protein and ligand, they provide initial information about the enzyme-ligand system's interaction. Results show that both mutants can smoother accommodate the tetradecane molecule in their active center, especially AMY^{L183Y}. The considerably higher docking score of AMY^{L183Y} suggests that changing the Leu residue for a Tyr optimizes the intermolecular interactions, facilitating the tetradecane binding process. This mutation characterizes the changing of a small, apolar residue for a larger, polar one. This plurality of differences between the two residues allows different hypothesis of why this substitution improves the docking process. As depicted in (**Fig.1**D), the Tyr residue appears to hinder the binding pocket, keeping the tetradecane molecule thoroughly inside it. In this sense, the α-amylase binding pocket is adapted to interact with starch, a large, polar molecule, which are opposite characteristics of tetradecane. Thus, reducing the active center cavity volume lowers the number of tetradecane possible configurations as it binds to the enzyme. In the case of the AMY^{N314A}, substituting the Asp residue for an Ala increases the hydrophobicity in the active center, becoming a more suitable environment for the hydrophobic tetradecane. The consistent docking results for this α-amylase variant possibly because tetradecane remains stably positioned between the nonpolar residues L183 and A314 (**Fig.1**D-E).

Investigating flexibility and structural changes of α-amylase variants

The structure of α-amylase has different degrees of flexibility for each domain. Domain A is the catalytic one, thus maintaining controlled flexibility is pivotal for catalysis. Domain B is the smallest one, characterized to hold tightly to a Ca^{2+} ion, thus exhibiting the least flexibility of all domains. Domain C, the largest among them, has a core of antiparallel β-sheet along with a large coil structure, assigning a more flexible characteristic to this domain.

Fig.3A shows RMSD values for domains A, B, and C of each α-amylase variant; **Fig.3B**

depicts the difference of RMSF between each mutant and the WT, and; **Fig.S6 (see Supplementary Material** illustrates the radius of gyration (R_g) for each domain of each α-amylase variant. Overall, there is little change in RMSD values for domain A, and only a small decrease is perceived for the mutants in comparison to AMY_{lig}^{WT}. In this sense, AMY_{lig}^{L183Y} presents the lower RMSD value and the most prominent variation between free and ligand-binding systems, suggesting that this mutant might accommodate the tetradecane in its pocket with more efficacy. RMSF analysis of domain A shows no profound difference in the fluctuation between the WT and the mutants systems. However, some subtle changes in the flexibility can be recognized comparing AMY^{WT} - AMY^{N314A}) and (AMY_{lig}^{WT} - AMY_{lig}^{N314A} systems, where the ligand-free system shows prominent positive peaks above the X-axis, especially in the second segment of domain A. This slight increase in the flexibility of AMY_{lig}^{N314A} might be a result of substituting a larger residue (Asn) for a smaller one (Ala), increasing the possibility of movement for the surrounding residues. R_g analysis shows no major difference for domain A among the α-amylase variants, noting that a slight decrease of R_g is observed from AMY^{L183Y} to AMY_{lig}^{L183Y}.

Next, we investigated domain B. RMSD, RMSF, and R_g results show no significant changes amidst the variants or the free and ligand-binding systems. Variations of 0.5 Å or less are observed, which is less than the distance of covalent and hydrogen bonds[50]. These are expected results since domain B is the smallest and structurally adapted for binding and holding a Ca^{2+} ion.

At last, domain C results are evaluated. Amidst the free α-amylase variants, AMY^{N314A} showed the highest value of RMSD. However, the ligand-binding systems present a different behavior. First, AMY_{lig}^{WT} did not only exhibit the highest RMSD amidst the variants, but it was also higher than AMY^{WT}. Furthermore, while AMY_{lig}^{L183Y} kept similar RMSD values throughout the simulation as AMY^{L183Y}, AMY_{lig}^{N314A} system lowered its overall RMSD, presenting values close to AMY_{lig}^{L183Y}. The higher values of RMSD for domain C is expected since it has several coils. The missing segment from the original PDB file belongs to domain C; thus, the peculiar results correspond to the modeled portion of α-amylase. Likewise, RMSF values for domain C displayed more pronounced peaks than the other domains. Nevertheless, it is clear that after bidding to tetradecane, domain C displays less variation, which,

together with R_g results, indicates increased compactness from this region.

To further evaluate the impact of the generated mutations, we performed a DSSP analysis. The results are expressed as the percentage of each secondary structure for each system **Fig.S7 (see Supplementary Material**. One major concern regarding protein engineering studies is how the mutation will affect the overall structure since inserting a mutation might result in a loss of secondary structure. In this sense, no loss of secondary structure was noted in the mutant variants.

Moreover, we wanted to examine the possibility of molecular changes occurring in the mutants on a scale that was not detectable in the previous analysis. Thus, to characterize the generated mutants and AMY^{WT}, we employed an RRIN analysis, which resulted in three different RRIN (**Fig.4**A-C). Each RRIN was then analyzed to identify the most topologically relevant residues. Topological investigations of interaction networks provide valuable information since they often identify significant players in network structure and control of information flow. In this sense, node degree measures the number of connections of a given node. Nodes with above-average node degree values are called "hubs." In protein-protein interaction networks (PPIN), hubs are often regarded as evolutionarily conserved proteins, playing a crucial role in cell survival[45,51]. However, in a RRIN, the residues characterized as hubs will be considered necessary for protein structure, possibly maintaining stability and proper conformation.

Moreover, betweenness measures the number of shortest paths that go through each node. In a PPIN, they are frequently designated as "bridges" between biological processes or biochemical pathways[45,52]. Nodes with above-average betweenness scores are named "bottlenecks." On the other hand, in RRIN, bottlenecks would be residues that possibly coordinate or respond to conformational changes, modulating protein behavior. Finally, the eigenvector centrality measures how regulatory a node could be, based on the node's number of connections and how well connected are their neighbors[45]. We employed the term "switches" to refer to nodes with eigenvector values above the network average[53]. In a PPIN, switches will commonly be regarded as regulatory proteins in clusters, once its neighbors will also have high eigenvector values. In a RRIN, we can translate these nodes as structurally critical residues that could control residue clusters. Thus, nodes possessing all of the above char-

acteristics, the hubs-bottlenecks-switches (HBS), would be the most topologically relevant nodes in the network.

We evaluated all interactions between the residues in all generated RRIN and compared them between the systems. Network analysis showed that there were no significant changes in the interactions types between AMY^{WT} and the mutant variants (**Fig.4**E). This is in agreement with the other analyzes, which showed that the isolated mutant variants had no significant change in structure variance (**Fig.3A**).

Next, we sought to identify the HBS residues in each system (**Fig.4A-C**). Interestingly, although we have found that each α-amylase had a set of exclusive HBS, they all maintained a "core" set of topologically relevant residues in all systems (**Table.3**) - all belong to either domain A or B. This is in agreement with previous observations that α-amylase activity was sustained even after partial removal of domain C[54]. Hence, alterations in the residues listed in Tab.3 would not be advised in terms of protein engineering since their change could significantly impact proper protein behavior - and if they do get mutated, protein stability must be evaluated afterward. In this sense, it is interesting to note that our mutations do not appear as HBS, which would agree with network proprieties expectations. In a PPI network, changes in highly regulatory nodes are typically related to disease phenotypes[51,52] - thus, in RRIN, changes in HBS could lead to protein instability or loss of function. Similarly, the residues belonging to the catalytic triad also do not appear as HBS. Since the HBS reflect topological (structure) significance, it makes sense that the catalytic triad might not appear as a top centrality since its value is associated with enzymatic activity, which HBS does not necessarily detect in RRIN.

Furthermore, RRIN analysis revealed HBS in common only between both α-amylase mutants, implying that the generated mutations also altered the HBS pattern of the WT variant. This suggests that, in the same way that alteration on different nodes in a PPI network can change major regulatory pathways, the same can happen to RRIN. In this case, the network analysis proposes that, although the conventional MD analysis observed no notable structural variance, the mutants do have a distinct topological difference from the WT, further explaining the tetradecane's better accommodation at the catalytic center.

Investigating the (un)binding energy between α-amylase variants and tetradecane

The observed promiscuous activity of α-amylase towards degrading n-alkanes is an evident indication of evolutionary innovation. This work explores the relationship between enzyme-substrate affinity and catalytic efficiency to enhance the promiscuous activity efficiency and take advantage of it for bioremediation purposes. There is no computational method that directly investigates the effect of an enzyme mutation on catalytic efficiency. Therefore, evaluating the affinity between enzyme and ligand is an indirect approach to assess the catalytic efficiency, regarding prior considerations are carefully made. While catalytic efficiency is often expressed as K_{cat}/K_M, the affinity presents an inverse correlation with K_M. Therefore, increasing the affinity improves catalytic efficiency regarding the enzyme-substrate complex rather than the free enzyme. The strategy of enhancing catalytic efficiency by strengthening the substrate affinity has shown success in several studies[55–60]. However, it is essential to consider how affinity is assessed, the nuances around this parameter, and how to improve the methodology accuracy. In the past, the main property investigated when selecting amidst mutants generated via *in silico* tools was the binding affinity between enzyme and substrate, relying almost exclusively upon this property[61]. However, assessing the enzyme-substrate affinity alone has shown not to be a reliable parameter. The main problem is that affinity is expressed regarding the dissociation constant (K_d), which in turn is only equal to K_M for genuine Michaelian enzymes. This occurs because K_M is a macroscopic constant that revolves into a combination of diverse microscopic parameters, including K_d. This phenomenon is seen by scrutinizing the mathematical terms of an enzyme-substrate reaction. A typical reaction is described as:

$$E + S \underset{k\text{-}1}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} P + E$$

Where E stands for the free enzyme, S is the substrate; ES is the enzyme-substrate complex; P is the product; k_1, k_{-1}, and k_2 are the velocity of the reaction in the respective direction.

K_d is obtained when a reaction achieves its equilibrium state. Since the formation of

the ES complex is a reversible step of the global reaction, the equilibrium distribution of reactants (E + S) and product (ES) will be achieved when the rates of the forward (k_1) and reverse (k_{-1}) reactions become equal. As the reaction approaches equilibrium, the rates of the opposing reactions become equal, and the concentrations of [E], [S], and [ES] remain constant through time. Therefore, no more variation of any reactant and product concentration is mathematically equal to the general equilibrium constant, in this case K_d. On the other hand, for first order reaction of enzymes that follow Michaelis-Menten kinetics, K_M under rapid equilibrium becomes equal to K_d[62,63]. This way, the rate can be expressed as:

Because of this non-universal aspect about the concept of affinity and its non-precise relationship to catalytic efficiency, recent experiments are relying on parameters such as the binding rate constant (k_{on}) and the unbinding rate constant (k_{off}) to investigate the relationship between an enzyme binding pocket to a specific ligand[64]. Interestingly, K_d can be assessed through these constants as well[65]. The mathematical relationship between these parameters is K_d=k_{off}/k_{on}.

Binding and unbinding rate constants are more appropriate to work with since they are determined directly from the energy of interaction between protein and ligand. There are a variety of computational tools that can manage to assess energy of interaction accurately[66]. In this work, we employed MTD to calculate k_{on}, k_{off}, and derived parameters to study the essence of how each α-amylase variant interacts with tetradecane. Furthermore, regarding α-amylase kinetics, this enzyme has shown to follow the Michaelis-Menten kinetic profile[63,67]. Thus, since assessing K_d through the binding energy (k_{on} and k_{off}) is an exact process, mutants showing decreased K_d implies in decreased K_M, which increases the catalytic efficiency.

Following the previously established rationale, the retrieved structure of α-amylase from PDB is binding to a maltopentaose molecule, a ligand in which this enzyme is evolutionarily adapted to cleavage. Thus, the structural conformation adopted by α-amylase in this crystal is the closest possible to the enzyme-substrate complex transition state when working strictly with molecular mechanics. Moreover, enzymes that display promiscuous activity show an optimized active site for recognition of the native substrate[68], suggesting that specificity

for the ligand and how well it is accommodated in the binding pocket is a decisive factor contributing to the catalytic efficiency of the primary substrate. Additionally, hydrophobic interactions play a significant role in enzyme-substrate recognition, and diverse promiscuous enzymes are known to have a dependence between hydrophobic interactions and catalytic efficiency[68–71]. In this sense, another study showed that the promiscuous binding between α-amylase and B-type procyanidin dimer were predominantly of hydrophobic interactions[72], once again suggesting that hydrophobicity plays a role in α-amylase promiscuous activity. Thus, both mutations increase the possibility of hydrophobic interactions to tetradecane, which would increase the catalytic efficiency.

Moreover, MD analyses were conducted to investigate how each binding pocket accommodates the ligand molecule and how it impacts the residence time. The distance between the tetradecane and the three catalytic residues has been inspected throughout the simulation **(Fig.1F)**. Initial analyses showed that both $\mathrm{AMY}_{lig}^{L183Y}$ and $\mathrm{AMY}_{lig}^{N314A}$ maintain the tetradecane inside their pocket during all simulation. In contrast, in the AMY_{lig}^{WT} system, the tetradecane molecule disengage from the pocket around 400 ns. Amidst the α-amylase mutants, $\mathrm{AMY}_{lig}^{L183Y}$ shows that the ligand is kept closer to the catalytic residues through the whole simulation. The observation that $\mathrm{AMY}_{lig}^{L183Y}$ decreases its (R_g) compared to AMY^{L183Y} and that it holds the ligand molecule closer to its catalytic residues suggests that this mutant encloses its catalytic pocket around the tetradecane tighter than the other variants. All these observations together allow us to recognize that the mutation L183Y stands out among the variants.

To further assess and investigate the unbiased MD simulations results, we performed MTD simulations for each system for 200 ns. From these calculations, an estimate of the ΔG_b^0 can be obtained (**Tab4, Figure S8, see S-Material**), providing a more reliable indication of how much the mutations enhanced the binding of the tetradecane molecule to the α-amylase variants. As explained in the Methods section, CVs corresponding to the spherical coordinates were biased, while the free energy surfaces (FES) were obtained by reweighting and projecting on the ρ and C_n CVs (**Figure 5**). The FES plots (**Fig.5A-C**) show many lower energy regions, providing different paths for the molecule to enter the catalytic site. Despite that, the plot area with the lowest energy basins was the same for all

of the simulated systems - the ligand in the bound state, with a ρ value between 0.6 and 0.9 nm.

The AMY^{L183Y} mutation near the catalytic site promoted an increase in the overall C_n (**Fig.5**) (values between 150-200) while restricting the minimum energy to a more well-defined energy basin. As shown in **Fig.5E** in some conformations sampled during the AMY^{L183Y} metadynamics run, Y183 is capable of trapping the ligand both inside and at the entrance of the binding site. The AMY^{N314A} mutation does not seem to act directly at the binding to the catalytic residues, but it still increases the C_n slightly in the energy minimum (values closer to 150). During the MTD run, its role appeared to be related to a mechanism of trapping and orientation of the ligand by the new interactions provided by the -CH3 group in the Ala residue. Hence, this could facilitate ligand binding and slightly increase its affinity to the enzyme. Both mutations seem to provide not only less conformational variability for the ligand, as mentioned in the previous sections, but also additional interactions due to the presence of bulkier residues (especially in the case of AMY^{L183Y}).

Contrastingly, the WT system displays a lower C_n, with a greater number of less defined energy minima areas (C_n between 100-150). Comparing ligand binding in all systems, the residue in position 183 (either L or Y) seems to interact more frequently to the ligand than residue 314 (either N or A), which might be why L183Y contributed to a higher ΔG_b^0. A possible explanation to the lower free energy of binding for the WT system is due to its lack of the additional interactions provided by the L183Y mutation and, at the same time, by the lack of a small but noticeable contribution, from the Alanine side-chain in N314A (**Fig.5D**). As exemplified in (**Fig.5F**), N314 barely interacts with the ligand, while L183 provides some contacts, but not as many as Y183 (**Fig.5E**), confirmed by the higher C_n of AMY^{L183Y}.

When calculating the ΔG_b^0 for each system, the higher C_n correlates well with a higher affinity. The values obtained follow the same order as seen in the docking results: AMY^{L183Y} shows a slightly higher free energy of binding than AMY^{N314A} and AMY^{WT}, while AMY^{N314A} is higher than AMY^{WT} (ΔG_b^0 = -6.1 kcal/mol, -5.1 kcal/mol, and -3.9 kcal/mol, respectively). Lastly, MTD results showed that AMY^{L183Y} and AMY^{N314A} both present an increased number of contacts with the ligand molecule while the free energy minimum decreases, implying an improvement of affinity for tetradecane.

In summary, we generated two new α-amylase mutants, L183Y and N314A, and investigated pertinent structural parameters through *in silico* strategies. MD analyses showed that both mutants do not compromise their structure by the mutations' effect, which is corroborated with RRIN analysis. Moreover, we observed increased residence time for the tetradecane molecule in both mutants, which agrees with molecular docking and MTD results. MTD analysis showed an increased number of contacts between the tetradecane and the binding pocket of both mutants. Additionally, MTD showed that the binding energy is also increased (in absolute value) for both mutants. The values obtained through this method are considerably close to the results from the molecular docking.

These results' pertinence is meaningful for the overall knowledge about α-amylase structure and the influence it might have for bioremediation purposes. First, α-amylase is highly conserved and found in the three domains of life. Moreover, it has decades of kinetics, thermodynamics studies, structural characterization and identified structural features for adapting in virtually all extreme environments. Thus, α-amylase arises as a potential candidate for bioremediation since it could be easily expressed in most organisms and adapted to perform in different hostile environments.

CONCLUSIONS

We discussed and investigated each appropriate step to achieve an enhanced promiscuous activity of α-amylase over the original evolutionary adapted activity from the theoretical knowledge behind protein architecture to structural mechanics and dynamics. In this sense, we focused on increasing the binding affinity of tetradecane, the largest n-alkane experimentally observed to be degraded by α-amylase. Regarding the affinity, the mutants generated in this work, L183Y and N314A, showed higher binding free energy (in absolute value) in both molecular docking and MTD investigations, along with an increased residence time.

Moreover, generating α-amylase mutants with an increased affinity towards the tetradecane is only practicable if the overall protein structure and native dynamics are not compromised. We carefully inspected each mutation's effect on the α-amylase structure, applying several different approaches. The results showed that the mutants had no negative varia-

tion from the WT and presented increased stability when accommodating the tetradecane molecule inside the binding pocket.

The protocol employed in this work, based on comparative modeling, mutant generation, molecular docking, MD, MTD, and RRI networks, is an integrated set of tools and analyses for protein and biocatalyst engineering. Since the field of computational protein design still has space for improvement and development of more accurate tools, this protocol can be used for future studies and be applied in diverse biotechnological fields, such as environmental, industrial, or pharmaceutical.

Lastly, bioremediation is the cheapest and most environmentally friendly remediation alternative. Each year new organisms, enzymes, and metabolic pathways are described to act on a different pollutant. However, as we discover new enzymes already adapted to degrade a specific molecule, we should also make an effort to explore the maximum potential of those enzymes. In this sense, two strategies may suffice: (i) apply engineering techniques to enhance the evolutionarily adapted activity or; (ii) investigate promiscuous enzymes' activities, which could be propitious to widen bioremediation tool options since enzymes are often highly dependent on specific conditions to achieve their full catalytic efficiency. Moreover, utilizing well-known conserved enzymes makes it easier to adapt them to the host's metabolism, as an α-amylase mutant should be, making this an attractive strategy yet to be adequately explored.

ACKNOWLEDGMENTS

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) [19/2551-0001906-8], Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [*311611 / 2018-4*], and was financed, in part, by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [[DAAD/CAPES PROBRAL 88881.198766/2018-0; *Finance Code 001*] - Brazil.

Additional Supporting Information may be found in the online version of this article.

Conflict of interest statement.

The authors declare no conflicts of interest of any kind.

Data Availability Statement.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. A. Dzionek, D. Wojcieszyńska, and U. Guzik, Electronic Journal of Biotechnology **19**, 28 (2016).

2. E. B. Overton, W. Sharp, and P. Roberts, Toxicity of petroleum (CRC Press: Boca Raton, FL, 1994).

3. J. H. Norman, Nontechnical Guide to Petroleum Geology, Exploration, Drilling, and Production (PennWell, 2001).

4. C. Riccardi, P. Di Filippo, D. Pomata, F. Incoronato, M. Di Basilio, M. P. Papini, and S. Spicaglia, Sci Total Environ **393**, 50 (2008).

5. J. Guo, J. Fang, and J. Cao, Chem Cent J **6**, 92 (2012).

6. H. Zou, G. Sheng, C. Sun, and O. Xu, Water Res **30**, 2003 (1996).

7. S. Wang, G. Liu, Z. Yuan, and C. Da, Ecotoxicol Enviro. Saf **150**, 199 (2018).

8. K. Tran, C. Y. Charlie, and E. Y. Zeng, Enviro Toxicol Chem **16**, 189 (1997).

9. S. Barathi and N. Vasudevan, Environ Int **26**, 413 (2001).

10. H. Baoune, J. D. Aparicio, G. Pucci, A. O. El Hadj-Khelil, and M. A. Polti, J of Soils Sediments **19**, 1 (2019).

11. S.-P. Wu, S. Tao, Z.-H. Zhang, T. Lan, and Q. Zuo, Environ Pollut **147**, 203 (2007).

12. N. Afshar-Mohajer, M. A. Fox, and K. Koehler, Sci Total Environ **654**, 924 (2019).

13. E. Yergeau, C. Maynard, S. Sanschagrin, J. Champagne, D. Juck, K. Lee, and C. W. Greer, Appl Environ Microbiol **81**, 5855 (2015).

14. M. Karimi and D. Biria, Chemosphere **152**, 166 (2016).

15. M. Karimi and D. Biria, Scientific reports **9**, 1 (2019).

16. S. Singh and L. Guruprasad, Protein and peptide letters **21**, 948 (2014).

17. É. S. M. Pinto, M. Dorn, and B. C. Feltes, Chemosphere p. 126202 (2020).

18. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Research **28**, 235 (2000).

19. Z. Fujimoto, K. Takase, N. Doui, M. Momma, T. Matsumoto, and H. Mizuno, Journal of molecular biology **277**, 393 (1998).

20. Y. Song, F. DiMaio, R. Y.-R. Wang, D. Kim, C. Miles, T. Brunette, J. Thompson, and D. Baker, Structure **21**, 1735 (2013).

21. D. Eisenberg, R. Lüthy, and J. U. Bowie, Methods in enzymology **277**, 396 (1997).

22. M. Wiederstein and M. J. Sippl, Nucleic acids research **35**, W407 (2007).

23. R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, Journal of applied crystallography **26**, 283 (1993).

24. P. Benkert, M. Biasini, and T. Schwede, Bioinformatics **27**, 343 (2011).

25. Q. Liu, G. Xun, and Y. Feng, Biotechnology advances **37**, 530 (2019).

26. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, et al., Methods Enzymol **487**, 545 (2011).

27. B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker, science **302**, 1364 (2003).

28. M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, Journal of cheminformatics **4**, 17 (2012).

29. W. L. Jorgensen and J. Tirado-Rives, Proceedings of the National Academy of Sciences **102**, 6665 (2005).

30. L. S. Dodda, J. Z. Vilseck, J. Tirado-Rives, and W. L. Jorgensen, The Journal of Physical Chemistry B **121**, 3864 (2017).

31. L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, and W. L. Jorgensen, Nucleic acids research **45**, W331 (2017).

32. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, Journal of computational chemistry **30**, 2785 (2009).

33. O. Trott and A. J. Olson, Journal of computational chemistry **31**, 455 (2010).

34. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1**, 19 (2015).

35. E. L. B. H. M.J. Abraham, D. van der Spoel and the GROMACS development team, GROMACS User Manual version 2018.1 (2018), URL `www.gromacs.org`.

36. W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, Journal of the American Chemical Society **118**, 11225 (1996).

37. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, The Journal of chemical physics **79**, 926 (1983).

38. B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, Journal of computational chemistry **18**, 1463 (1997).

39. T. Darden, D. York, and L. Pedersen, The Journal of chemical physics **98**, 10089 (1993).

40. M. Parrinello and A. Rahman, Journal of Applied physics **52**, 7182 (1981).

41. G. Bussi, D. Donadio, and M. Parrinello, The Journal of chemical physics **126**, 014101 (2007).

42. D. Piovesan, G. Minervini, and S. C. Tosatto, Nucleic acids research **44**, W367 (2016).

43. M. de Hoon, S. Imoto, J. Nolan, and S. Miyano, Bioinformatics **20**, 1453 (2004).

44. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, Genome Res **13**, 2498 (2003).

45. G. Scardoni, M. Petterlini, and C. Laudanna, Bioinformatics **25**, 2857 (2009).

46. R. Capelli, P. Carloni, and M. Parrinello, The Journal of Physical Chemistry Letters **10**, 3495 (2019).

47. A. Barducci, G. Bussi, and M. Parrinello, Phys. Rev. Lett. **100**, 020603 (2008).

48. S. Janecek, B. Svensson, and B. Henrissat, Journal of molecular evolution **45**, 322 (1997).

49. Š. Janeček, B. Svensson, and E. A. MacGregor, Cell Mol Life Sci **71**, 1149 (2014).

50. T. Steiner, Acta Crystallographica Section B: Structural Science **54**, 456 (1998).

51. M. Vidal, M. E. Cusick, and A.-L. Barabási, Cell **144**, 986 (2011).

52. J. Loscalzo and A.-L. Barabasi, Wiley Interdiscip Rev: Syst Biol and Med **3**, 619 (2011).

53. B. C. Feltes, J. de Faria Poloni, I. J. G. Nunes, S. S. Faria, and M. Dorn, Frontiers in genetics **11** (2020).

54. H.-F. Lo, L.-L. Lin, W.-Y. Chiang, M.-C. Chie, W.-H. Hsu, and C.-T. Chang, Archives of microbiology **178**, 115 (2002).

55. Y. Xia, W. Cui, Z. Cheng, L. Peplowski, Z. Liu, M. Kobayashi, and Z. Zhou, ChemCatChem **10**, 1370 (2018).

56. A. Li, L. Ye, X. Yang, B. Wang, C. Yang, J. Gu, and H. Yu, ChemCatChem **8**, 3229 (2016).

57. E. Durmaz, S. Kuyucak, and U. O. Sezerman, Protein Engineering, Design & Selection **26**, 325 (2013).

58. H.-W. Jiang, Q. Chen, J. Pan, G.-W. Zheng, and J.-H. Xu, Applied biochemistry and biotechnology (2020).

59. K. Fan, H. Wang, J. Xi, Q. Liu, X. Meng, D. Duan, L. Gao, and X. Yan, Chemical Communications **53**, 424 (2017).

60. X. Wang, Y. Nie, X. Mu, Y. Xu, and R. Xiao, Scientific reports **6**, 24574 (2016).

61. S. M. Marques, D. Bednar, and J. Damborsky, Frontiers in chemistry **6**, 650 (2019).

62. D. Voet, J. G. Voet, and C. W. Pratt, Fundamentals of biochemistry: life at the molecular level (John Wiley & Sons, 2016).

63. N. P. Guerra and L. Pastrana Castro, The Scientific World Journal **2012** (2012).

64. M. Bernetti, A. Cavalli, and L. Mollica, MedChemComm **8**, 534 (2017).

65. A. C. Pan, H. Xu, T. Palpant, and D. E. Shaw, Journal of chemical theory and computation **13**, 3372 (2017).

66. T. Siebenmorgen and M. Zacharias, Wiley Interdisciplinary Reviews: Computational Molecular Science **10**, e1448 (2020).

67. N. Sakač and M. Sak-Bosnar, Int. J. Electrochem. Sci **7**, 3008 (2012).

68. A. Babtie, N. Tokuriki, and F. Hollfelder, Current opinion in chemical biology **14**, 200 (2010).

69. O. Khersonsky and D. S. Tawfik, Biochemistry **44**, 6371 (2005).

70. B. R. Villiers and F. Hollfelder, ChemBioChem **10**, 671 (2009).

71. L. Afriat, C. Roodveldt, G. Manco, and D. S. Tawfik, Biochemistry **45**, 13677 (2006).

72. T. Dai, J. Chen, Q. Li, P. Li, P. Hu, C. Liu, and T. Li, International journal of biological macromolecules **113**, 427 (2018).

Figure 1: **α-amylase structure, docking, and atomic distances.** A) Full α-amylase structure, showing each domain, the calcium ion, and the location of both generated mutants (see Results); B) Catalytic triad, responsible for coordinating α-amylase interaction to tetradecane (see Results); C) Docking results illustrating the accommodation of tetradecane at the AMY^{WT} catalytic center (see Results); D) Docking results illustrating the accommodation of tetradecane at the AMY^{L183Y} catalytic center (see Results); E) Docking results illustrating the accommodation of tetradecane at the AMY^{N314A} catalytic center (see Results); F) Atomic distance investigation between each catalytic residue and tetradecane (see Results). AMY$^{WT}_{lig}$ is represented by a black line; AMY$^{L183Y}_{lig}$ mutant is represented by a blue line, and; a red line represents AMY$^{N314A}_{lig}$.

Figure 2: **Workflow employed in this work.**. After generating and validating the model, a total of six systems were simulated, each consisting of triplicates of 500ns. A MTD analysis was performed on the three systems containing the tetradecane molecule.

Figure 3: **RMSD and RMSF investigation of α-amylase.** A) RMSD is shown for each domain separately; B) RMSF is expressed as a function of each mutant system subtracted from the WT system. Positive values indicate higher fluctuation in the WT structure, while negative values illustrate higher mutant structure fluctuation.

Figure 4: **RRIN for AMY^{WT}, AMY^{L183Y}, and AMY^{N314A}.** A) RRIN for AMY^{WT}; B)RRIN for AMY^{L183Y}; C) RRIN for AMY^{N314A}. The subnetworks containing the HBS identified in the centrality analysis are found within the squares. Green edges represent van der Waals interactions, blue edges depict H-Bonds, purple edges illustrate ionic interactions, and red edges represent π-π stacking and cation-π interactions; D) Venn diagram showing the HBS residues in common and unique to each α-amylase; D) Pie charts depicting the percentage of interaction types belonging to each α-amylase. The difference in network layouts is just an artifact from the visualization.

Figure 5: **Free energy surfaces (FES) and tetradecane binding modes for AMY^{WT}, AMY^{L183Y}, and AMY^{N314A}.** A) FES of AMY^{WT}; B) FES of AMY^{L183Y}; C) FES of AMY^{N314A}; All FES are plotted on two CVs, namely distance between the ligand and the centre of mass of the catalytic site (ρ) and the number of contacts (C_n). D) Binding pose for the tetradecane to AMY^{WT} showing fewer interactions at the catalytic site; E) Binding pose for the tetradecane to AMY^{L183Y} showing the role of Y183 (cyan) trapping the ligand in the catalytic site; F) Binding pose for the tetradecane to AMY^{N314A} showing the role of A314 (yellow) which provides a few more interactions in the binding site. The same color scheme from Figure 1C-E was followed here.

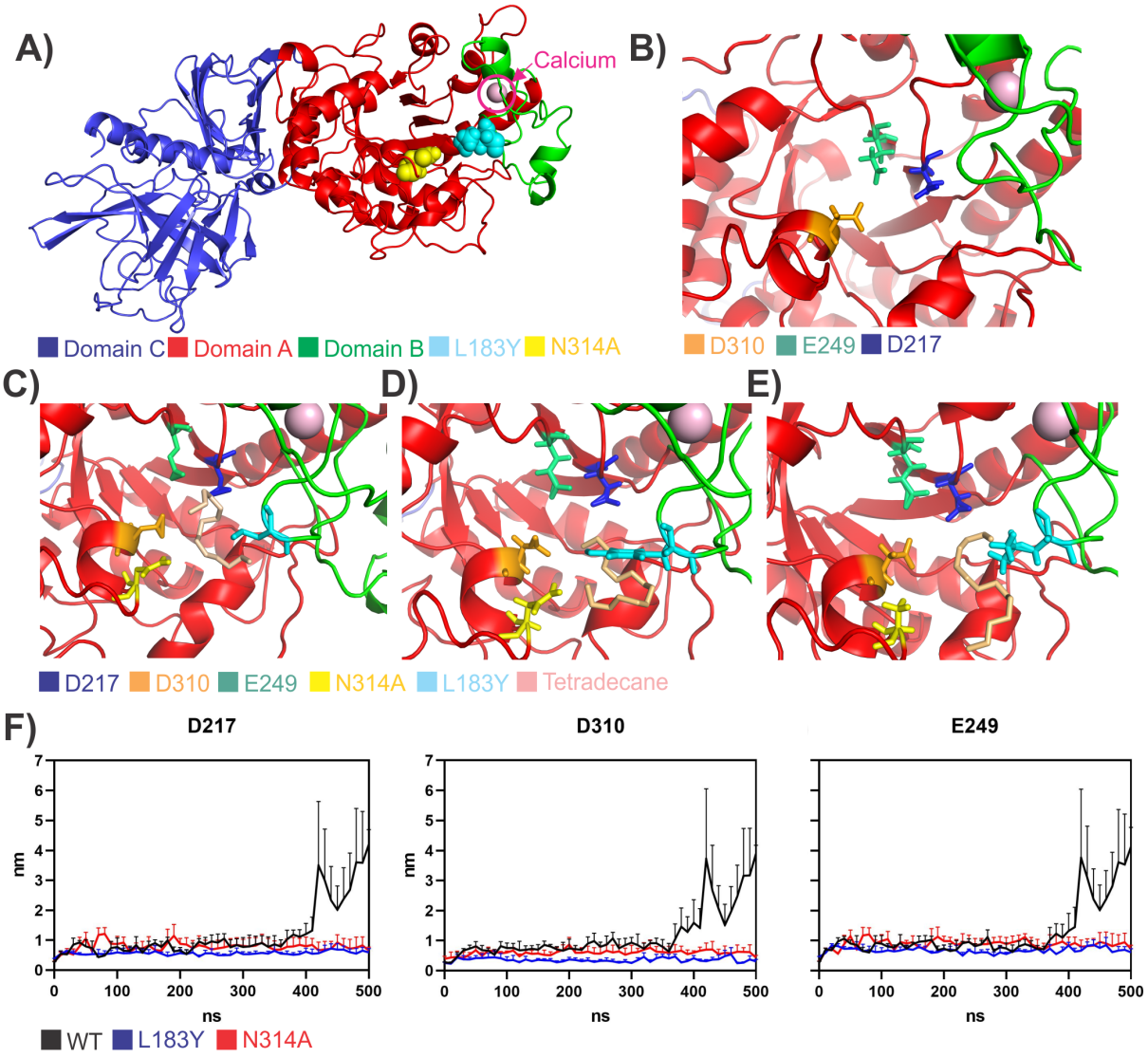


Figure 1: **α-amylase structure, docking, and atomic distances.** A) Full α-amylase structure, showing each domain, the calcium ion, and the location of both generated mutants (see Results); B) Catalytic triad, responsible for coordinating α-amylase interaction to tetradecane (see Results); C) Docking results illustrating the accommodation of tetradecane at the AMY^{WT} catalytic center (see Results); D) Docking results illustrating the accommodation of tetradecane at the AMY^{L183Y} catalytic center (see Results); E) Docking results illustrating the accommodation of tetradecane at the AMY^{N314A} catalytic center (see Results); F) Atomic distance investigation between each catalytic residue and tetradecane (see Results). AMY$^{WT}_{lig}$ is represented by a black line; AMY$^{L183Y}_{lig}$ mutant is represented by a blue line, and; a red line represents AMY$^{N314A}_{lig}$.

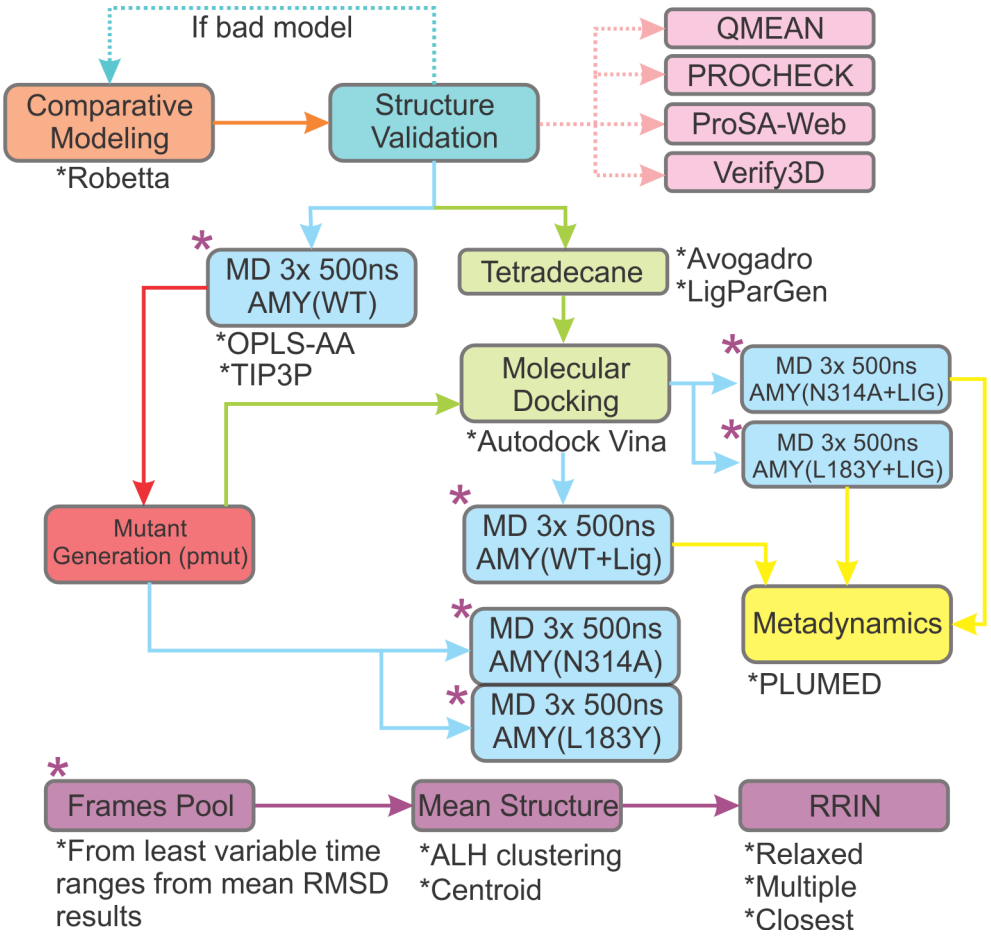


Figure 2: **Workflow employed in this work.**. After generating and validating the model, a total of six systems were simulated, each consisting of triplicates of 500ns. A MTD analysis was performed on the three systems containing the tetradecane molecule.

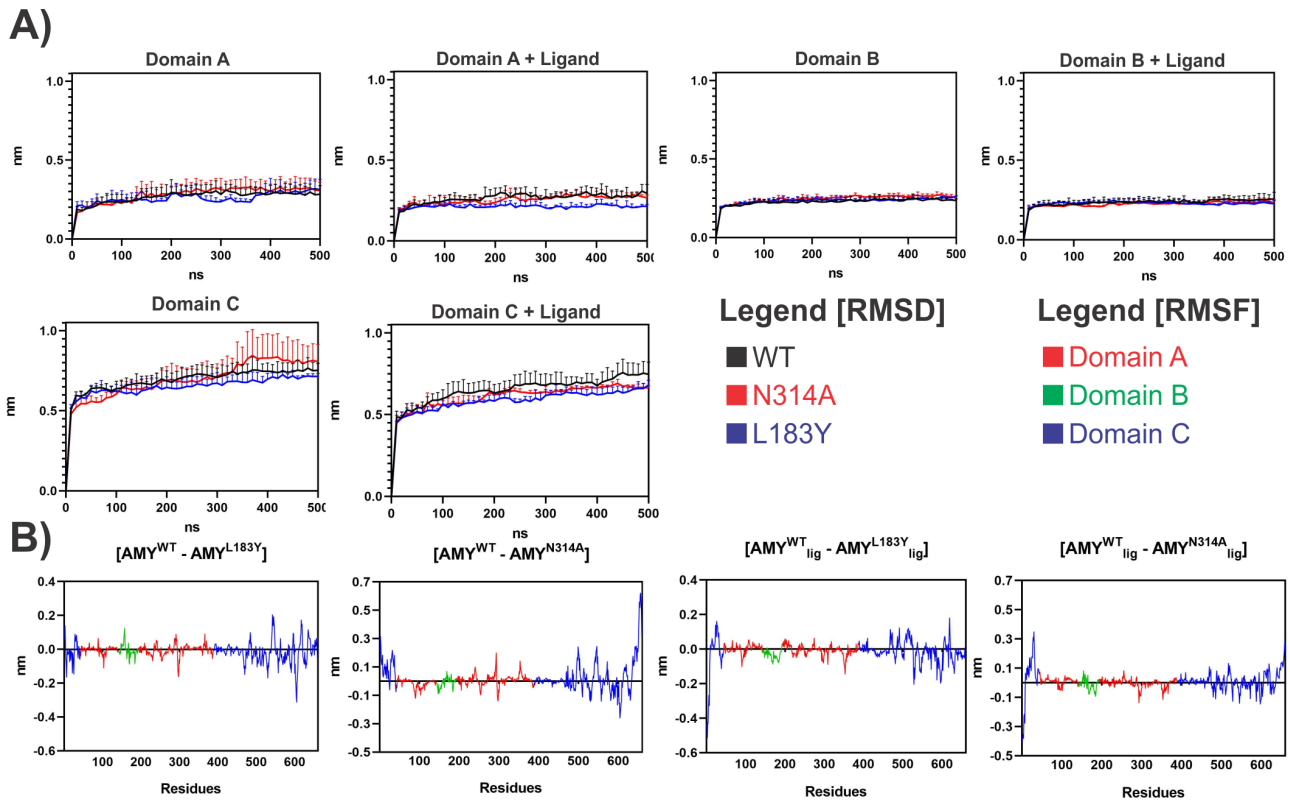


Figure 3: **RMSD and RMSF investigation of α-amylase.** A) RMSD is shown for each domain separately; B) RMSF is expressed as a function of each mutant system subtracted from the WT system. Positive values indicate higher fluctuation in the WT structure, while negative values illustrate higher mutant structure fluctuation.

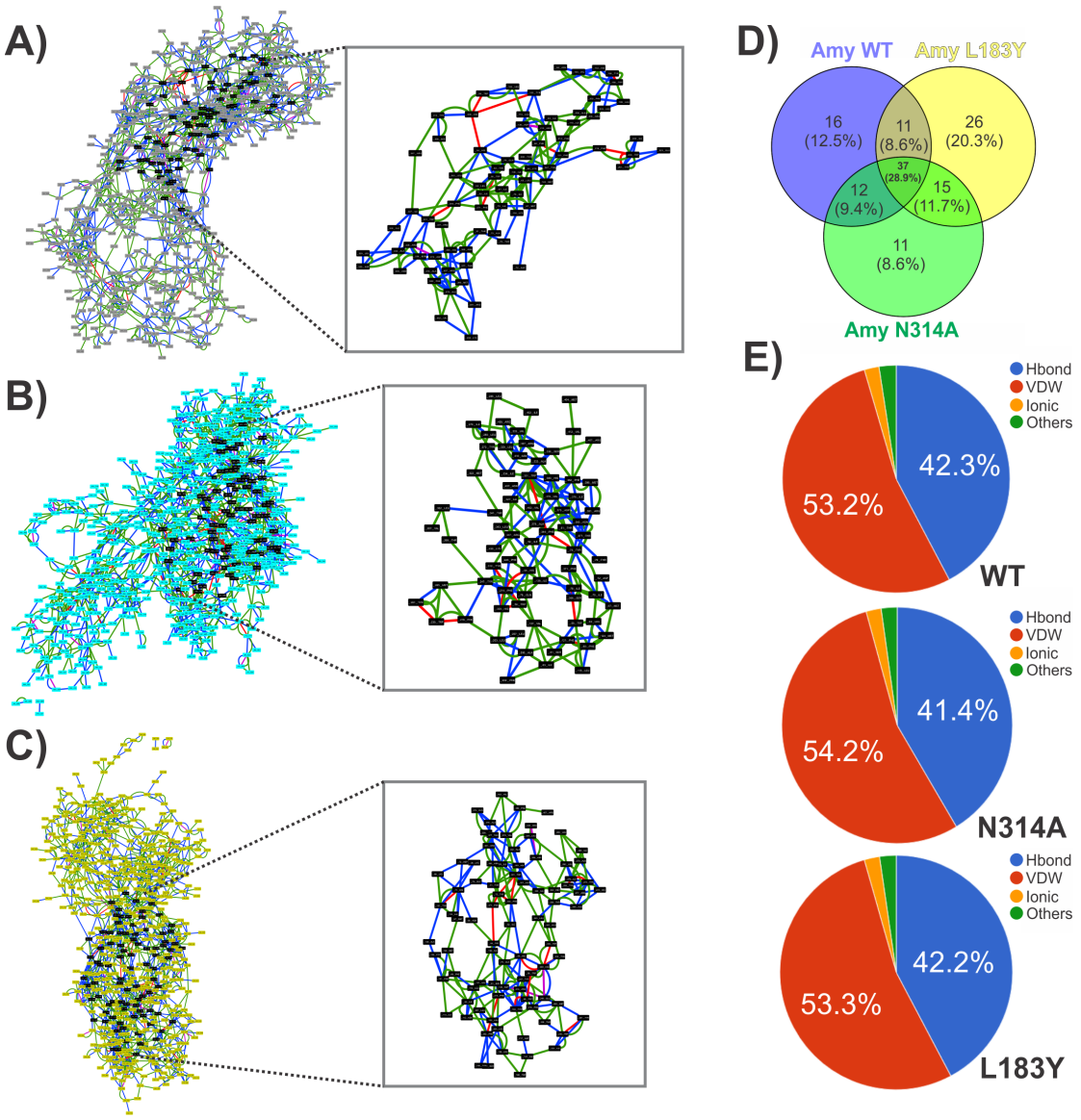


Figure 4: **RRIN for AMY^{WT}, AMY^{L183Y}, and AMY^{N314A}.** A) RRIN for AMY^{WT}; B)RRIN for AMY^{L183Y}; C) RRIN for AMY^{N314A}. The subnetworks containing the HBS identified in the centrality analysis are found within the squares. Green edges represent van der Waals interactions, blue edges depict H-Bonds, purple edges illustrate ionic interactions, and red edges represent π-π stacking and cation-π interactions; D) Venn diagram showing the HBS residues in common and unique to each α-amylase; D) Pie charts depicting the percentage of interaction types belonging to each α-amylase. The difference in network layouts is just an artifact from the visualization.

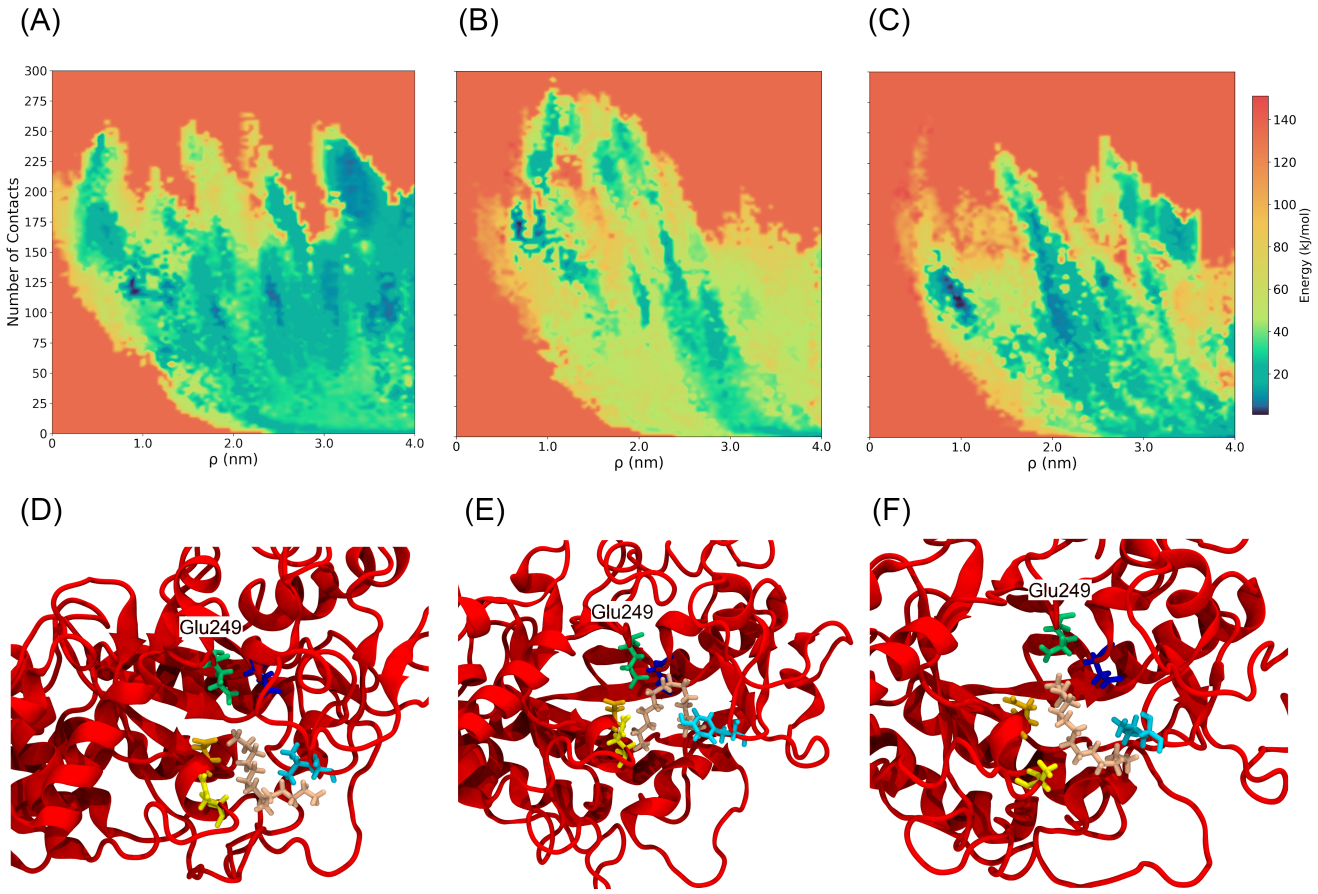


Figure 5: **Free energy surfaces (FES) and tetradecane binding modes for AMY^{WT}, AMY^{L183Y}, and AMY^{N314A}.** A) FES of AMY^{WT}; B) FES of AMY^{L183Y}; C) FES of AMY^{N314A}; All FES are plotted on two CVs, namely distance between the ligand and the centre of mass of the catalytic site (ρ) and the number of contacts (C_n). D) Binding pose for the tetradecane to AMY^{WT} showing fewer interactions at the catalytic site; E) Binding pose for the tetradecane to AMY^{L183Y} showing the role of Y183 (cyan) trapping the ligand in the catalytic site; F) Binding pose for the tetradecane to AMY^{N314A} showing the role of A314 (yellow) which provides a few more interactions in the binding site. The same color scheme from Figure 1C-E was followed here.

Table 1: Summary of validation test results

Validation	**Metric**	**Results**
4*Ramachandran	Res in most favoured regions	**88.3%**
	Res in additional allowed regions	**11.1%**
	Res in generously allowed regions	**0.2%**
	Res in disallowed regions	**0.3%**
ProSA-Web	Z-Score	**-10.18**
QMEAN	Z-Score	**>1**
Verify 3D	3D-1D score	**PASS**

Table 2: Molecular docking scoring results

Enzyme	**Least score**	**Best score**
$\mathbf{AMY}_{lig}^{WT}$	-3.5	-4.5
$\mathbf{AMY}_{lig}^{N314A}$	-4.5	-4.6
$\mathbf{AMY}_{lig}^{L183Y}$	-4.9	-5.2

Table 3: HBS's in common between each system.

System	**HBS Residues**
AMY^{WT} vs AMY^{L183Y} vs AMY^{N314A}	Thr51, Ile52, His54, Trp58, Phe60, Met67, Asp72, Gln79, Ile83, Leu115, Phe121, Met124, Ile136, Phe202, Leu203, Phe214, Phe216, Trp235, Tyr247, Tyr261, His309, Tyr312, Ile327, Leu329, Trp331, Ala332, Leu343, Phe344, Phe371, Thr377, Ala378, Asn380, Arg381,Phe382, His383, Arg407, Val462
AMY^{L183Y} vs AMY^{N314A}	Trp56, Tyr75, Ser81, Val137, Leu207, Met265, Tyr272, Ile276, Leu280, Leu285, Met322, Ile334, Ile376, Val379, Arg458
AMY$^{WT}_{lig}$ vs AMY$^{N314A}_{lig}$ vs AMY$^{L183Y}_{lig}$	Ile52, His54, Trp188, Trp58, Phe60, Ile327, Arg328, Leu63, Tyr198, Trp331, Met67, Ile334, Phe202, Ile70, Tyr75, Leu343, Ile78, Phe344, Gln79, Phe214, Arg215, Ile83, Phe216, Trp235, Phe371, Ile376, Thr377, Tyr247, His383, Phe121, Met124, Tyr261, Ile136
AMY$^{N314A}_{lig}$ vs AMY$^{L183Y}_{lig}$	Thr51, Thr76, Ala128, Val267, Asn142, Leu302, Val303, His309
HBS in Common (All)	Ile52, His54, Trp58, Phe60, Met67, Gln79, Ile83, Phe121, Met124, Ile136, Phe202, Phe214, Phe216, Trp235, Tyr247, Tyr261, Ile327, Trp331, Leu343, Phe344, Phe371, Thr377, His38

Table 4: Free energy differences (ΔG^0_b) obtained by using the Volume metadynamics technique for AMY^{WT} and its mutants. Δ_{Metad} corresponds to the value found in the metadynamics simulations and ΔG^0_b corresponds to the final value after applying the entropic correction.

System	Δ_{Metad}	ΔG^0_b
AMY^{WT}	-3.1 ± 0.5	-3.9 ± 0.5
AMY^{L183Y}	-5.3 ± 0.3	-6.1 ± 0.3
AMY^{N314A}	-4.3 ± 0.4	-5.1 ± 0.4