Molecular dynamics simulations provide insights into substrate specificity of FAOX family members

Federica Rigoldi[1], Ludovica Spero[1], Andrea Dalle Vedove[2,3], Alberto Redaelli[1], Emilio Parisini[2], Alfonso Gautieri[1]

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano (Italy)

[2]Center for Nano Science and Technology @Polimi, Istituto Italiano di Tecnologia, Via G. Pascoli 70/3, 20133 Milano (Italy)

[3]Dipartimento di Chimica, Materiali e Ingegneria Chimica "G. Natta", Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133, Milano (Italy)

Keywords: Fructosyl Amino Acid oxidase, Amadoriases, Deglycating Enzymes, Molecular Dynamics Simulation, Enzyme specificity, Binding interactions, HbA1c monitoring, Diabetes monitoring, Glycated haemoglobin.

Submitted to Molecular BioSystems

ABSTRACT

Enzymatic assays based on Fructosyl Amino Acid Oxidases (FAOX) represent a potential, rapid and economical strategy to measure glycated hemoglobin (HbA1c), which is in turn a reliable method to monitor the insurgence and the development of diabetes mellitus. However, the engineering of naturally occurring FAOX to specifically recognize fructosyl-valine (the glycated N-terminal residue of HbA1c) has been hindered by the paucity of information on the tridimensional structures and catalytic residues of the different FAOX that exist in nature, and in general on the molecular mechanisms that regulate specificity in this class of enzymes. In this study, we use molecular dynamics simulations and advanced modeling techniques to investigate five different relevant wild-type FAOX (Amadoriase I, Amadoriase II, PnFPOX, FPOX-E and N1-1-FAOD) in order to elucidate the molecular mechanisms that drive their specificity towards polar and nonpolar substrates. Specifically, we compare these five different FAOX in terms of overall folding, ligand entry tunnel, ligand binding residues and ligand binding energies. Our work will contribute to future enzyme structure modifications aimed at the rational design of novel biosensors for the monitoring of blood glucose levels.

INTRODUCTION

The glycated form of haemoglobin (HbA1c) is of particular interest for the diagnosis and the monitoring of diabetes. The hyperglycemia associated with diabetes results in the non-enzymatic glycation of blood proteins, including haemoglobin (which has a half-life of 120 days) and albumin (half-life of 20 days). For this reason, the measurement of the levels of glycated haemoglobin in the blood is a very powerful method for monitoring the insurgence and the development of diabetes[1]. Indeed, while direct measurement of the blood sugar level is affected by daily fluctuations, the long lifetime of haemoglobin combined with the slow, yet irreversible, glycation process makes the detection of HbA1c a good indicator of the average blood glucose concentration over a period of 2-3 months. For this reason, in 2010 the American Diabetes Association designated the level of HbA1c as a powerful indicator for the diagnosis of diabetes.

Since the assessment of glycated haemoglobin is becoming an indispensable part of diabetes diagnosis and control, the HbA1c test demands robustness, high-throughput capacity and cost effectiveness. As a result, several systems have been developed that are used in the clinics to measure HbA1c[2]. Most methods rely on the separation of HbA1c from non-glycated haemoglobin based on the different chemical properties of these two species. These methods include ion exchange chromatography (based on the different isoelectric point), affinity chromatography (based on the different affinity for boronic acid) and capillary electrophoresis (based on the different charge). While meeting the requirement for quality and robustness, these methods are based on relatively complex and expensive techniques that require the intervention of specialized staff and thus fall short for cost-effectiveness and simple delivery at a point-of-care. An alternative method for HbA1c detection exploits the deglycating properties of Amadoriase enzymes, also called Fructosyl Amino Acid Oxidase (FAOD) or Fructosyl Peptide Oxidase (FAOX) enzymes. These enzymes, which react with glycated amino acids and release free amino acid, glucosone and hydrogen peroxide, can be included in a fast, easy-to-use and cost-effective HbA1c monitoring kit.[3]

Such enzyme-based diabetes monitoring systems have been recently proposed[3] and commercialized (e.g., Direct Enzymatic HbA1c Assay™, Diazyme Laboratories). The system is based on a first proteolytic digestion step that releases amino acids, including glycated valines from the N-terminus of glycated haemoglobin. Subsequently, Amadoriase enzymes deglycate valines and produce hydrogen peroxide which, in turn, is measured using horseradish peroxidase and a suitable chromophore. Compared to chromatography- and electrophoresis-based sensing methods, enzymatic assays have the advantage of being simple and less expensive, thus doable at a point-of-care. However, a major drawback of these sensors is that wild-type Amadoriases are unspecific for valines, since they also cleave glycated lysines that are present on haemoglobin, which therefore interfere with the measurement of HbA1c.

In earlier works[4–6], the Sode group has addressed this issue by performing extensive mutagenesis experiments on the N1-1-FAOD from *Pichia* marine yeast, especially at those sites that are putatively involved in ligand binding (e.g., His51 and Asn374). As a result, they managed to provide the enzyme with improved selectivity for fructosyl valine over fructosyl lysine, a favorable feature for the detection of HbA1c. More recent mutagenesis experiments[7,8] were also done by the same group on the fructosyl peptide oxidase from *Phaeospheria nodorum* (PnFPOX). Other Amadoriases that are regarded as promising enzymes for HbA1c sensing include Amadoriase I[9,10] and Amadoriase II[11,12] from *Aspergillus fumigatus*.

These works highlight the need for a precise knowledge of the overall folding architecture, active site conformation and enzyme-substrate interaction details of the different FAOX family members. The lack of such crucial pieces of information has so far been one of the major limiting factors for the development of Amadoriase-like enzymes with improved sensitivity for fructosyl valine and thus useful for diabetes sensing. This limit has been overcome in part by the successful determination of the free and the inhibitor-bound crystal structures of Amadoriase II from *Aspergillus fumigatus*[13], by the more recent crystal structure determination of the free fructosyl peptide oxidase from *Eupenicillium terrenum* (also known as FPOX-E or EtFPOD)[14] and, lately, by the crystal structure of the free and the substrate-bound form of Amadoriase I from *Aspergillus fumigatus*[15].

On the other hand, molecular dynamics (MD) simulations can provide key structural information at the atomistic level on biomolecules[16–19] and biomaterials[20–23] when experimental characterization proves to be difficult or impossible. In this work, we use MD simulations to provide a detailed analysis of the folding, binding pocket conformation and enzyme-substrate dynamic interaction of the five most relevant Amadoriase enzymes for HbA1c sensing (i.e., Amadoriase I, Amadoriase II, FPOX-E, N1-1-FAOD and PnFPOX) in complex with fructosyl-lysine (f-lys) and fructosyl-valine (f-val). This analysis will likely facilitate the development of novel Amadoriase mutants with enhanced selectivity for f-val, thus possibly improving the efficiency of enzyme-based HbA1c sensing systems.

MATERIALS AND METHODS

Homology modelling

Currently, high resolution structural data is available for Amadoriase I (PDB id: 4WCT and 4XWZ), Amadoriase II (PDB id: 3DJD and 3DJE) and FPOX-E (PDB id: 4RSL), but not for N1-1-FAOD and PnFPOX. However, owing to the high sequence homology of these enzymes, reliable initial models of the last two species can be obtained using homology modeling.

Molecular models of N1-1-FAOD (sequence from NCBI id: AY255792) and PnFPOX (sequence form NCBI id: XM_001798659) were generated using the program MODELLER 9v15[24]. Among the available FAOD structures,

the templates were selected based on a balance between: (i) the lowest expectation value (e-value) and (ii) the highest percentage of sequence identity. The e-value is an indicator that provides an estimation of the significance of the alignment (the lower the expectation value the higher the significance). In particular, the e-value describes the number of hits one can expect to see by chance when searching through a database of a given size[25,26]. Fifty homology models for each enzyme were built and ranked by their Discrete Optimized Protein Energy (DOPE), a pairwise atomistic statistical potential[27] optimized for model assessment within MODELLER. Finally, the structure with the lowest DOPE value was chosen. Homology models were further assessed using the Verify3D server[17] (http://services.mbi.ucla.edu/Verify_3D).

Simulated annealing Molecular Dynamics simulations

Further structural optimization of the two homology models thus generated (N1-1-FAOD and PnFPOX) was done by implementing a simulated annealing (SA) procedure in explicit solvent. SA was performed with the ACEMD software using *in-house* TCL scripting. The protocol involved heating and cooling steps followed by a 20 ns NVT simulation at constant physiological temperature. The temperature was first raised from 300 K to 500 K and then lowered back to 300 K, allowing steps of 5 K every 5 ps. We have defined the upper limit temperature in order to allow sampling of flexible loops and, at the same time, preventing large unfolding of the general scaffold obtained from homology modeling[28]. Finally, the models were kept for 20 ns at 300 K to allow structural relaxation. Final conformations were used as the starting points for all the following computational studies.

Classical Molecular Dynamics simulations

We used the high resolution crystal structures of Amadoriase I, Amadoriase II and FPOX-E and the homology models of N1-1-FAOD and PnFPOX to carry out extensive molecular dynamics (MD) simulations and study the enzyme-substrate dynamic interactions and tunnel conformations when each of these enzymes is bound to either f-val or f-lys. Hence, a total of ten molecular systems were generated. The analysis of the dynamic interactions of these substrates within the binding pockets of the five different enzymes was done by means of 200 ns MD simulations in explicit solvent. The proteins were modelled using the AMBER99SBildn force field[29](3) while the FAD co-factor and ligands were modelled using the general amber force field (GAFF)[30]. The complexes were solvated with ≈10,000 TIP3P water molecules and the ionic strength was set to 0.15 M with the addition of Cl^- ions and Na^+ ions. Ions are included using the Autoionize plugin of VMD in order to neutralize the net electric charge of the system (this is a necessary requirement for simulations with periodic boundary conditions), and, secondly, to mimic the ionic strength of the solvent that surrounds the protein. The setup resulted in systems of ≈37,000 atoms in a simulation box of initial dimensions of ≈ 65 × 70 × 90 $Å^3$ (varying slightly depending on the system).

The molecular models were minimized and equilibrated using the NAMD code[31] under constant pressure and temperature (NPT) conditions in order to relax the volume of the periodic box. The pressure was set to 1 atm and the temperature to 300 K, while using a time step of 2 fs, a non-bonded cut-off of 9 Å, rigid bonds and particle-mesh Ewald long-range electrostatics. During minimization and NPT equilibration, the C_α atoms of the protein were restrained by a 10 kcal mol^{-1} $Å^{-2}$ spring constant to prevent protein diffusion. Finally, the production run was performed using ACEMD[32] on a NVIDIA Kepler K20 GPU for a total time of 200 ns. A longer time step of 4 fs was used thanks to the use of the hydrogen mass repartitioning scheme implemented in ACEMD. All other parameters (temperature, non-bonded cut-off, and PME) were kept the same as in the equilibration phase. In order to avoid protein diffusion, a harmonic restraint of 1 kcal mol^{-1} $Å^{-2}$ was applied on the C_α atoms of the protein farther than 30 Å from the binding pocket. The stability of the system was assessed by monitoring the convergence of the root mean square deviation (RMSD) of the protein. Following MD simulations, all analyses were conducted using the VMD software package[33].

Substrate clustering analyses

We used the VMD-cluster plugin tool to calculate and visualize the clusters of substrate conformations that were assumed during the 200 ns MD simulations. An RMSD cut-off of 1.3 Å was used to separate and clusterize the different conformations. We determined the representative structures for the four most popular clusters of each substrate in all the five enzymes.

Tunnel conformation

We used the CAVER 3.0.1[34] program to calculate the pathway leading from the buried active site to the bulk solvent. For the analysis of the geometry of the tunnel, the ligand, ions and water molecules included in the MD runs were removed during tunnel calculation. We decided to set the starting point of the tunnel calculation at a position above the flavin plane, and in particular at the same position of the N_ε atom of the ligand, as previously determined in the crystallographic structure of Amadoriase I[15]. We chose probe radius larger than the default value (which is based on the radius of a water molecule) because we are interested in the identification of possible ligand entry tunnels. Thus, we set the probe radius to 1.4 nm, a dimension comparable with that of the cyclic form of glucose. To describe the protein surface correctly, we used a shell depth of 6 Å and the shell radius of 4 Å. We defined these parameters through a trial-and-error optimization procedure starting from default values. During this process we were careful not to overlook important tunnels or select irrelevant branches that are accessible only to molecules that are smaller than the studied ligands. We tried different tunnel clustering thresholds to avoid splitting of similar tunnels (i.e., same path and exit point) into different clusters. The final tunnel clustering threshold was set to 5 Å.

To further characterize the ligand path from bulk solvent towards the catalytic pocket, we classified tunnel-lining residues based on their chemical nature (i.e., charged: LYS, ARG, HIS, GLU, ASP; polar: SER, ASN, THR, GLN, TYR; hydrophobic: ALA, VAL, LEU, ILE, PHE, MET, TRP special: GLY, CYS, PRO). In particular, we used 3.2 Å as cut-off value (the same used for salt bridges interactions) to identify those residues whose side chain contributes to defining the enzyme tunnel. The cut-off value for salt-bridges is the VMD default value, which has also been used in previous work[35–39].

Finally, we decided to perform tunnel calculations on the last 2 ns of our dynamic simulations to capture possible tunnel dynamic changes (using a stable and equilibrated protein conformation). Default values were used for all the other parameters.

Cavity volume measurement

We evaluated the cavity volume of the enzymes active pocket using an *in-house* TCL script that calculates the number of water molecules inside the active pocket (averaged over the last 2 ns of trajectory in the apo forms). We defined pocket borders in the three dimensions as the FAD triple rings plane, the top of the T105-E118 helix (Amadoriase I numbering, while homologous regions are considered for the other enzymes) and the most external residues of the H58-D67 loop (numbering based on the Amadoriase I sequence).

Hydrogen bonds, salt bridges and polar interactions

A first set of analyses involves non-covalent bonds between the protein and the substrates. H-bonds and salt-bridges interactions were calculated using *in-house* TCL scripts on a geometric basis using a donor-acceptor distance < 3.2 Å[35–38] and a donor-hydrogen-acceptor angle < 30°. We also analyzed hydrophobic interactions between the nonpolar moieties of the substrate (i.e. the sidechain of f-val and the aliphatic part of the amino acidic moiety of f-lys) and the sidechain of the nonpolar amino acids within the binding pocket. We calculated hydrophobic interactions using an *in-house* TCL script in VMD whereby two hydrophobic sidechain are considered interacting if the distance between their centers of mass is shorter than 6.5 Å[40].

Binding free energy –MM-PBSA calculation and enthalpy decomposition

The enthalpic component of the binding free energy (herein referred as binding energy for brevity) of the protein-ligand complexes was estimated using the end-state free energy method MM-PBSA (Molecular Mechanics – Poisson Boltzmann Surface Area)[41], as integrated in the AMBER12 software package[42].

The total binding free energy is defined as a differential energy[43]:

$$\Delta G_{binding} = G_{complex} - G_{receptor} - G_{ligand} \qquad \text{[eq.1]}$$

where $G_{complex}$, $G_{receptor}$ and G_{ligand} are the free energy components of the complex, the receptor and the ligand, respectively.

Each term on the right of eq.1 ($G_{complex}$, $G_{receptor}$ and G_{ligand}) is computed as:

$$G = E_{MM} + G_{PBSA} - TS \quad \text{[eq.2]}$$

where E_{MM} is the Molecular Mechanical energy, G_{PBSA} is the solvation free energy calculated by solving the linearized Poisson-Boltzmann equation, and *TS* is the entropic contribution. The entropic contribution is not considered in the MM-PBSA method, hence the name.

The molecular mechanical term (E_{MM}) of eq.2 can be further decomposed and calculated using the AMBER99SBildn and GAFF force fields:

$$E_{MM} = E_{bond} + E_{angle} + E_{torsion} + E_{vdw} + E_{elect} \quad \text{[eq.3]}$$

The solvation energy (G_{PBSA}) in eq. 2 is calculated as the sum of a polar and a non-polar contribution:

$$G_{PBSA} = G_{non-polar} + G_{polar} \quad \text{[eq. 4]}$$

The $G_{non-polar}$ term is determined from the Solvent-Accessible Surface Area (SASA) as follows:

$$G_{non-polar} = \gamma \cdot SASA + \beta \quad \text{[eq.5]}$$

The surface tension proportionality constant γ represents the contribution to the solvation free energy per unit surface area (surface tension). The constant β (offset value) is the free energy of hydration for a point solute (i.e., SASA = 0)[44]. The default values of 0.00542 kcal mol^{-1} Å^{-2} and 0.92 kcal mol^{-1} were used for the surface tension γ and the offset β terms, respectively[41,45–47].

Finally, the G_{polar} term is computed in continuum solvent using a finite Poisson-Boltzmann model. Within this approach it is necessary to define the dielectric constant of the internal cavity (ε_{int}). As proposed by Hou et al[48], this value is determined according to the PSASAD (Protein Solvent Accessible Area Surface Difference, see Table S1) which accounts for the local environment of the cavity (e.g., polar or nonpolar amino acids).

For binding energy calculations, we restricted our analysis to the frames where the sugar moiety of the ligand is correctly positioned within the catalytic pocket. The enthalpic component of the binding free energy was then computed as the average over all the relevant frames of the trajectory.

The full free energy of the system (which includes both the enthalpic and the entropic term) can be assessed using advanced methods such as Free Energy Perturbation (FEP) or Thermodynamic Integration (TI)[49]. However, these methods are computationally expensive and tend to have a large margin of error, which introduces significant uncertainty in the results. Therefore, this term is usually ignored in studies based on the

MM-PBSA method[50–52]. In particular, this approximation is utilized during ligand screening[48] or when evaluating differences among similar ligands or enzymes, with no significant effect on the final results[53,54]. However, for the calculation of absolute affinity values the use of the entropic term is mandatory.

RESULTS AND DISCUSSION

Enzyme structures and sequences comparisons

In this work we investigate the fold of five relevant Fructosyl Amino Acid Oxidases and their binding interaction with two substrates, fructosyl-lysine and fructosyl-valine. These two ligands are particularly relevant because f-lys is the preferred ligand by most FAOX, whereas f-val is of well-known technological importance due to its use to estimate the amount of glycated haemoglobin in blood glucose monitoring tests.

The crystal structures of three of the five enzymes (namely Amadoriase I, Amadoriase II and FPOX-E) are available and were used as the starting point for MD simulations. For the other two enzymes (N1-1-FAOD and PnFPOX), we determined the putative lowest energy structure by homology modelling using the most closely related enzymes as templates. A phylogenetic analysis showed that PnFPOX is most closely related to FPOX-E, while N1-1-FAOD is only marginally similar to any other FAOX (Figure 1). Sequence alignment (Figure 2) and identity matrix (Table 1) analysis show that all the sequences have a high level of strict identity. In particular, the residues that are known to bind the sugar moiety of the ligand, corresponding to E285, G371 and R418 in Amadoriase I[15], are totally conserved.

Based on the phylogenetic tree, identity matrix and sequence alignment, we chose the crystal structure of FPOX-E as template structure for the homology modelling of PnFPOX. For N1-1-FAOD we used a combined template based on the crystal structure of Amadoriase I for the general overall fold (residues from 1 to 49 and from 150 to 427), the crystal structure of Amadoriase II for the region ranging from residue 50 to residue 88, and the crystal structure of FPOX-E for the solvent exposed portion comprised between residues 89 and 149 (N1-1-FAOD numbering). We performed the structural alignment of the five structures and we colored the models thus obtained according to their structural homology and conservation (Figure 3A). As expected, all the five enzymes share a common global fold. In particular, they show the typical *Rossmann fold* in the flavin binding region, where the three residues (E285, G371 and R418) that interact with the sugar moiety of the ligands are totally conserved.

Overall, the main differences among the five enzymes are in the loop regions between amino acids 405 to 420 and 55 to 70 (numbering from Amadoriase I). Interestingly, these regions define the tunnel entry from the bulk solvent to the catalytic pocket and they interact with the (variable) amino acid moiety of the ligand. In particular, the loop 55-70 can assume three different conformations: it is in fact projected outwards towards

the bulk solvent in Amadoriase I, projected inwards towards the catalytic cavity in Amadoriase II and N1-1-FAOD, or it can be much shorter than in these two cases, as observed in PnFPOX and FPOX-E.

Cavity volume estimation and tunnel conformation

We estimated the size of the cavity of the different enzymes by assessing the average number of water molecules that can occupy the tunnel and the binding cavity during the MD simulations. As shown in Table 2, Amadoriase I features a much wider cavity than all the other FAOX. The main reason is that the 55-70 loop of Amadoriase I is folded outwards towards the bulk solvent and not inwards towards the core of the enzyme. This conformation of the loop contributes to the larger cavity of Amadoriase I compared to the other enzymes. As described above, the corresponding loop in the other enzymes either folds inwards or, in the case of FPOX-E, is considerably shorter. Accordingly, N1-1-FAOD and PnFPOX feature the smallest cavities.

In all of the enzymes, the cavity is buried inside the protein structure; in particular, all five enzymes show a similar primary tunnel (see Figure 3B-F) of comparable length, ranging from 11.3 Å (Amadoriase II) to 18.8 Å (N1-1-FAOD), and a comparable bottleneck radius (1.33-1.85 Å), as reported in Table 2.

While Amadoriase I shows only a single tunnel, the other enzymes show also a secondary tunnel. However, in both FPOX-E and PnFPOX the secondary tunnel is rather similar to the primary one, with which it shares the tunnel entry. Interestingly, the second tunnel for PnFPOX is present in only 10% of the analyzed frames. Conversely, Amadoriase II shows a rather different secondary tunnel (green in figure 3C), which is defined by two loop regions that move considerably during the molecular dynamics simulation. Another region that affects tunnel dimensions and exit path is the loop 405-420 (numbering of Amadoriase I). This region creates a wide bend towards the main tunnel of the enzymes, thus contributing to defining its entry and, in the case of N1-1-FAOD, producing an alternative path to reach the catalytic pocket (Figure 3E).

In Table 3 we list, categorized by their chemical nature, all those residues that contribute to tunnel lining in at least 50% of the analyzed frames. Amadoriase I shows the most hydrophobic tunnel, N1-1-FAOD also presents a high number of tunnel-lining hydrophobic residues. Conversely, both FPOX-E and PnFPOX show a high percentage of charged residues facing the tunnel walls. Finally, Amadoriase II shows a mixture of charged and nonpolar residues along the tunnel. We hypothesize that the tunnel-lining residue type could play a role in enzyme affinity for polar or nonpolar substrates. Hydrophobic tunnels (like the one observed for Amadoriase I) could prevent polar ligands to strongly bind within the tunnel along their entry trajectory towards the binding sites, thus favoring the enzyme selectivity for polar substrates (e.g., f-lys). On the other hand, for those enzymes that are mostly active on nonpolar substrates (i.e. PnFPOX and FPOX-E) the high number of charged residues along the tunnel could guide the ligand towards the catalytic pocket and prevent them from remaining stuck in the tunnel.

Molecular Dynamics simulations of enzyme-ligand complexes

To study the molecular bases of ligand recognition of the different FAOX enzymes for different substrates, we ran 200 ns molecular dynamics simulations for each of the five enzymes in complex with either f-lys or f-val, for a total of ten enzyme-substrate complexes. The simulation time that the ligands spend in the catalytic pocket and the conformational stability of the ligands in the cavity during the MD simulations suggests that Amadoriase I, Amadoriase II and FPOX-E present the best environment for the stabilization of f-lys in their active site (Figure 4). These results correlate well with the enzymatic activity of the five selected FAOX (Table 4), which show that the most efficient enzymes for this substrate are Amadoriase I and II. On the other hand, our simulations suggest that the enzymes that provide the most stabilizing environment for f-val are Amadoriase I and N1-1-FAOD, whereas experimental data from enzymatic assays show lower affinity for this substrate. The reason for this difference could be partially attributed to the high variability in the kinetics data that are available in the literature. Furthermore, the experimental enzymatic activity can only be loosely associated with the ligand stability observed in our simulations, since other factors (ligand exit, turnover, catalytic efficiency, etc.) are likely to play a role in defining the observed enzymatic constants.

Substrate clustering analyses

In order to study the configuration of the ligands within the catalytic pocket, we analyzed the most populated conformational clusters found for each ligand within the five investigated FAOX during the 200 ns MD simulations (Figure 5).

In Amadoriase I, the most common conformational cluster for each ligand covers the majority of the simulation time (92% for f-lys and 88% for f-val), indicating that the substrates are highly stabilized within the binding pocket of the enzyme. We observe that the conformational clusters of f-lys differ almost exclusively for the conformation of the tail of the ligand, while f-val clusters show also a slight flexibility in the position of the sugar moiety. These differences can partially explain the higher specificity of this enzyme for the f-lys substrate. In Amadoriase II the most common conformational cluster of f-lys is observed for 54% of the simulation time. Similarly, to Amadoriase I, the f-lys clusters differ mostly for the position of the amino acid moiety of the ligand, albeit with higher mobility, which could explain the lower affinity of Amadoriase II for this substrate compared to Amadoriase I. On the other hand, the main cluster for f-val is observed for 85% of the simulation time, which suggests that this ligand is better stabilized by the local environment.

The two enzymes FPOX-E and PnFPOX show a rather similar behavior: the single major conformational cluster of f-lys is found in 44% and 37% of the simulation time, respectively. The clusters differ mostly for the conformation of the amino acid moiety of the ligand, although the sugar moiety presents a higher degree of mobility within the binding site of FPOX-E. Conversely, f-val shows a primary cluster that is more populated (67% and 73% of the time in FPOX-E and PnFPOX, respectively), which suggests that, overall, these enzymes

have a higher affinity for f-val. This observation correlates well with the experimentally determined enzymatic activity of PnFPOX[55], while the enzymatic characterization of FPOX-E is incomplete[56].

Finally, we observe that both f-lys and f-val are highly unstable in the N1-1-FAOD enzyme, as suggested by the low population of the major conformational cluster (46% for f-lys and 18% for f-val) and by the high mobility of the substrates in the binding pocket, also involving the otherwise usually stable sugar moiety. This result is in accordance with the experimental findings,[4,6] which show that N1-1-FAOD has the worst K_m and V_{max} when compared with all the FAOX considered in this study (see also Table 3).

Polar and Hydrophobic contacts analysis

We analyzed the polar and nonpolar interactions of both ligands (f-lys and f-val) with the wildtype FAOXs in order to define the relevant ligand-binding residues for each enzyme (Figure 6). The stability of each contact (measured as the percentage of the time in which the contact is observed during our MD simulations) is reported in the Supplementary Information (Figure S1). Concerning the polar contacts, the highly conserved catalytic glutamic acid (E285 in Amadoriase I) makes a very stable hydrogen bond with the sugar moiety of the ligands. This interaction is present in all five enzymes when the ligand is in the binding pocket. Another important interaction that stabilizes the sugar moiety of the substrates is provided by a conserved catalytic arginine (R418 in Amadoriase I). This contact is maintained for a smaller share of the simulation time than that mediated by the glutamic acid. Moreover, its strength shows a large variability depending on the enzyme. When the ligand is f-lys, the interaction is maintained for more than 65% of the time, except for FPOX-E where the stability is slightly lower (55% of the time). The interaction shows similar stability when the f-val ligand is in the Amadoriase I, PnFPOX and FPOX-E active site. Conversely, when f-val is in complex with Amadoriase II and N1-1-FAOD, the contact is formed only for 31% and 12% of the time, respectively. A third interaction that is crucial for enzyme activity is the one involving the conserved catalytic glycine (G371 in Amadoriase I). This glycine residue interacts with the glycated amide of the ligands and it is the key factor for enzyme activity due to its interaction with flavin during the oxidation process. The strength of this interaction is lower compared to the aforementioned catalytic contacts and it varies considerably among different complexes. In particular, higher stability is observed in the case of both Amadoriase I and FPOX-E complexed with f-lys, as well as in the complex between N1-1-FAOD and f-val.

The results show also some durable contacts with the amino acid moiety of the ligands, mainly with f-lys owing to its charged amino acid tail. In particular, Amadoriase II shows a very stable contact between R111 and both the sugar and the amino acids moieties of f-lys. N1-1-FAOD, PnFPOX and FPOX-E show a contact between f-lys and a homologue lysine (K357, K376, K379 respectively) for more than 55% of the simulation. Amadoriase II features a lysine residue in the homologous position (K368) which, however, forms negligible contacts with f-

lys. In FPOX-E, a contact between K379 and the C-term of the f-val amino acid moiety is formed for 47% of the time, an interaction that is not observed in the other enzymes.

Amadoriase I shows two stable contacts with the amino acid tail of f-lys, one with E58 (45% of the time) and another with N371. While the former appears to be the major stabilization interaction for polar substrates[13,57], the latter has an analogous interaction in N1-1-FAOD (N354). H-bonds involving this asparagine are even more stable in the MD trajectory when the ligand is f-val (> 64% of the time), suggesting a key role for this residue in stabilizing nonpolar substrates.

We also analyzed the set of hydrophobic contacts between the enzyme and the substrate in each complex. They may, in fact, provide a non-negligible stabilization energy to these systems. This is especially true with f-val, owing to its nonpolar amino acid moiety. All the enzymes except Amadoriase II share an analogue isoleucine (I56 Amadoriase I, I57 in FPOX-E, I49 in N1-1-FAOD, and I58 in PnFPOX) that interacts with the aliphatic chain of f-lys and with the side chain of f-val. In particular, this contact is observed for almost the entire simulation time (more than 97%) in Amadoriase I with both ligands, in N1-1-FAOD in complex with f-val and in FPOX-E and PnFPOX in complex with f-lys. In the homologous position, Amadoriase II features a valine, which does not interact with the ligands. All of the enzymes feature also several phenylalanine or tyrosine residues on the same side of the catalytic pocket, which interact with both ligands. In details, an isostructural phenylalanine shared by all complexes is F268 in Amadoriase I, F263 in Amadoriase II, F266 in FPOX-E, F260 in N1-1-FAOD, F263 Pn-FFPOX. In particular, this residue creates a very stable contact with f-val in both PnFPOX and FPOX-E. FPOX-E presents also a stable contact between a F266 and f-val. Moreover, in both Amadoriase I and Amadoriase II a very firm interaction between both ligands and a homologous phenylalanine (F262 and F257, respectively) is formed. This phenylalanine is replaced by a tyrosine in FPOX-E and PnFPOX (Y260 and Y257, respectively) and the contacts with f-val are highly stable for the entire simulation time. Amadoriase I, Amadoriase II and FPOX-E complexed with f-lys show also contacts with a methionine residue (M374, M97 and M93 respectively), whose positions are, however, not homologous. It is interesting to observe that the nonpolar wall of the binding pocket is larger in N1-1-FAOD (5 Phe and 1 Tyr), while the other enzymes present only with three or four Phe or Tyr residues.

Binding free energy and enthalpy decomposition

Whereas Amadoriase I shows a prevalence of polar and charged residues among those that form its catalytic pocket, Amadoriase II has a more hydrophobic active site.

The enthalpic component of the binding free energy (H) was estimated using the MM-PBSA method. Figure 7 shows the binding energy of f-lys and f-val within each of the wildtype enzymes. For the five enzymes, the binding energy with f-lys range from -40 kcal/mol to -34 kcal/mol, while with f-val it falls in the range between -

33 kcal/mol and -28 kcal/mol. The similarity of these binding energy values is in accordance with the high homology between the five enzymes.

A detailed analysis of the different energetic contributions shows that all five enzymes, in complex with either f-lys or f-val, show similar Van der Waals forces contributions. Regarding electrostatic and polar solvation forces, Amadoriase I shows, with both ligands, lower values than the other four enzymes; the fact that, among the five enzymes, Amadoriase I has the wider catalytic pocket, which increases the distances between the ligand and the catalytic residues, can partially explain this observation. Due to the polarity of the substrate, f-lys binding generally leads to an unfavorable contribution of the solvation energy, which is however balanced by a favorable contribution of the electrostatic interactions, when compared to the binding of f-val. The only exception seems to be represented by Amadoriase I, which exhibits an opposite behavior.

A binding energy decomposition analysis was performed to assess the contribution of the main residues involved in polar contacts with the ligands. As shown in Figure S2 in the Supplementary Information, the contact with the glutamic acid is by far the one that is maintained for the longest time in the course of the simulations and the decomposition analysis shows that it provides an important energetic contribution (always larger than -9.3 kcal/mol), without any relevant differences between different complexes with the two ligands. However, we note that in Amadoriase II the contribution of this residue is smaller compared to all the other enzymes.

Conversely, the energetic contribution of the catalytic arginine varies among the different complexes and the contacts it forms are less maintained over time. For Amadoriase I and PnFPOX the interaction energy between R417 and R415 with f-lys and f-val is similar (≈ -5 kcal/mol). Amadoriase II and N1-1-FAOD show a significant energetic contribution of this catalytic arginine only in the complex with f-lys, while the homologous arginine in FPOX-E show a significant energetic contribution only in the complex with f-val.

The H-bonding interactions of the ligands with the catalytic glycine are generally limited in terms of both persistence in time and energetic contribution. However, we observe that the contact of the ligand with this residue is often water-mediated (as found also in the Amadoriase I crystal structure[15]), although water-mediated contact are not computed in this analysis.

Regarding the energetic contributions of those residues that interact with the amino acid moiety of the ligand, our results show highly energetic contacts mostly with f-lys, which features a charged tail. In the complex of Amadoriase II and f-lys, the most relevant energetic contribution is the one involving R111. The large vibration of this residue during the simulations is reflected in the high standard deviation values of the binding energy. In Amadoriase I, f-lys interacts mostly with E58 (-1.6 kcal/mol). FPOX-E, N1-1-FAOD and PnFPOX share a homologous lysine residue inside the catalytic pocket (K379, K375, K376), which in complex with f-lys provides a significant energetic contribution and it is present for a significant portion of time in all simulations (>55% of

the entire simulation time). FPOX-E shows this interaction also with f-val, although the binding energy involved is lower. In Amadoriase I and N1-1-FAOD, another important contact involving the amino acid tail of the ligand is provided by asparagine N371 and N354, respectively, especially when binding f-val (-3.7 and -3.3 kcal/mol, respectively). In FPOX-E we observe that R61 forms a stabilizing contact with f-lys (-6.8 kcal/mol) that is not present in the other enzymes, due to the different conformation of the loop 55-70 in this enzyme. The N1-1-FAOD enzyme shows a stabilizing interaction involving f-lys and D53 (-10.6 kcal/mol), which is missing in the other four enzymes due to the different conformation of this region.

The energetic contribution of the hydrophobic interactions is rather limited (less than -2.7 kcal/mol) compared to the polar contacts. Therefore, the relevance of a specific hydrophobic contact can be evaluated only relative to the other existing hydrophobic contacts. The energetic contributions of the isostructural phenylalanines F268 in Amadoriase I, F263 in Amadoriase II, F266 in FPOX-E, F260 in N1-1-FAOD and F263 PnFPOX is non negligible for all the enzymes except for N1-1-FAOD in complex with f-val. These residues provide a similar stabilization energy for Amadoriase II, FPOX-E and PnFPOX in complex with f-lys. Furthermore, Amadoriase I and Amadoriase II feature a stable interaction between their homologous phenylalanines F262 and F257, respectively, and both ligands. The homologous tyrosines in FPOX-E and PnFPOX (Y260 and Y257, respectively) provide a stabilizing energy contribution when the enzymes are in complex with f-val. Finally, the highly conserved isoleucine residues I56 in Amadoriase I, I57 in FPOX-E, and I49 in N1-1-FAOD and I58 in PnFPOX provide a stabilizing interaction with both ligands in the Amadoriase I and the PnFPOX complexes and in the FPOX-E complex with f-lys.

CONCLUSIONS

A detailed knowledge of the overall folding, catalytic residues type and configuration, tunnel size and conformation together with their enzymatic activity towards different substrates (size and type of glycated amino acids or peptides) of wild-type FAOX is of paramount importance for a knowledge-based design of novel deglycating enzymes with expanded functionality.

In this work, we first predicted the structure of the N1-1-FAOD and PnFPOX enzymes using homology modelling. Then, we used these models and the experimental crystal structures of Amadoriase I, Amadoriase II and FPOX-E to run extensive molecular dynamics simulations in order to compare the structures of these FAOX enzymes and assess their relevant interactions with two relevant ligands, f-val and f-lys. We showed that all five enzymes share a common global fold and that the main differences among them are located in the loop regions that define the entry of the tunnel that leads the substrate from the bulk solvent to the catalytic pocket. Interestingly, these loop regions interact with the (variable) amino acid moiety of the ligand, thus contributing to the specificity of the substrate recognition process. By studying the stability of the ligands

within the catalytic pocket of the different enzymes, we observed that Amadoriase I, Amadoriase II and FPOX-E feature the best environment for the stabilization of f-lys, whereas f-val is highly stabilized in Amadoriase I and N1-1-FAOD. Finally, we analyzed the polar and nonpolar interactions as well as the binding energy of f-lys and f-val with the five wildtype FAOXs and we provided a detailed description of the relevant ligand-binding residues for each enzyme.

Our results will likely facilitate the rational design of novel deglycating enzymes for HbA1c sensing. The measurement of glycated haemoglobin is becoming a standard in the diagnosis of diabetes and, in general, in blood glucose monitoring. At present in Europe there are approximately 60 million people living with diabetes (387 million worldwide) and the prevalence of this costly disease is increasing, which highlights the need for inexpensive, fast and easy-to-use diabetes monitoring tests. The successful design of novel enzymes with increased affinity for f-val will provide a significant improvement in enzyme-based glycaemia monitoring systems, and thus is expected to substantially impact on fields such as diabetes prevention, diagnosis and management.

ACKNOWLEDGEMENTS

This work has been supported by Fondazione Cariplo, grant n° 2013-0766.

REFERENCES

1 C. Weykamp, W. G. John and A. Mosca, *J. Diabetes Sci. Technol.*, 2009, **3**, 439–45.

2 W. G. John, *Clin Chem Lab Med*, 2003, **41**, 1199–1212.

3 L. Liu, S. Hood, Y. Wang, R. Bezverkov, C. Dou, A. Datta and C. Yuan, *Clin. Biochem.*, 2008, **41**, 576–83.

4 S. Miura, S. Ferri, W. Tsugawa, S. Kim and K. Sode, *Protein Eng. Des. Sel.*, 2008, **21**, 233–239.

5 S. Kim, S. Miura, S. Ferri, W. Tsugawa and K. Sode, *Enzyme Microb. Technol.*, 2009, **44**, 52–56.

6 S. Miura, S. Ferri, W. Tsugawa, S. Kim and K. Sode, *Biotechnol. Lett.*, 2006, **28**, 1895–1900.

7 S. Kim, E. Nibe, S. Ferri, W. Tsugawa and K. Sode, *Biotechnol. Lett.*, 2010, **32**, 1123–1129.

8 S. Kim, E. Nibe, W. Tsugawa, K. Kojima, S. Ferri and K. Sode, *Biotechnol. Lett.*, 2012, **34**, 491–497.

9 C. Mennella, R. C. Borrelli, F. Vinale, M. Ruocco and V. Fogliano, *Ann. N. Y. Acad. Sci.*, 2005, **1043**, 837–44.

10 X. Wu, B. a Palfey, V. V Mossine and V. M. Monnier, *Biochemistry*, 2001, **40**, 12886–95.

11 Y. Qian, J. Zheng and Z. Lin, *Appl. Microbiol. Biotechnol.*, 2013, **97**, 8599–607.

12 J. Zheng, H. Guan, L. Xu, R. Yang and Z. Lin, *Appl. Microbiol. Biotechnol.*, 2010, **86**, 607–13.

13 F. Collard, J. Zhang, I. Nemet, K. R. Qanungo, V. M. Monnier and V. C. Yee, *J. Biol. Chem.*, 2008, **283**,

27007–16.

14 W. Gan, F. Gao, K. Xing, M. Jia, H. Liu and W. Gong, *Acta Crystallogr. Sect. F, Struct. Biol. Commun.*, 2015, **71**, 381–387.

15 F. Rigoldi, A. Gautieri, A. Dalle Vedove, A. P. Lucarelli, S. Vesentini and E. Parisini, *Proteins*, 2016, **84**, 744–758.

16 A. Masic, L. Bertinetti, R. Schuetz, S. Chang, T. H. Metzger, M. J. Buehler and P. Fratzl, *Nat. Commun.*, 2015, **6**, 1–8.

17 Z. Qin and M. J. Buehler, *Nat. Commun.*, 2013, **4**, 2187.

18 A. Gautieri, A. Redaelli, M. J. Buehler and S. Vesentini, *Matrix Biol.*, 2013, **34**, 89–95.

19 A. Gautieri, S. Vesentini, A. Redaelli and R. Ballarini, *Int. J. Non. Linear. Mech.*, 2013, **56**, 25–33.

20 A. Gautieri, S. Vesentini and A. Redaelli, *J. Mol. Model.*, 2010, **16**, 1845–51.

21 O. Tokareva, M. Jacobsen, M. Buehler, J. Wong and D. L. Kaplan, *Acta Biomater.*, 2014, 10, 1612–1626.

22 M. Takeuchi, M. Iwaki, J. Takino, H. Shirai, M. Kawakami, R. Bucala and S. Yamagishi, *Lab. Invest.*, 2010, **90**, 1117–27.

23 M. I. Solar and M. J. Buehler, *Nat. Nanotechnol.*, 2012, **7**, 417–419.

24 A. Fiser and A. Sali, *Macromol. Crystallogr. Pt D*, 2003, **374**, 461–91.

25 W. R. Pearson, *J. Mol. Biol.*, 1998, **276**, 71–84.

26 B. Rost, *Protein Eng.*, 1999, **12**, 85–94.

27 M. Y. Shen and A. Sali, *Protein Sci.*, 2006, **15**, 2507–2524.

28 R. Sankararamakrishnan, K. Konvicka, E. L. Mehler and H. Weinstein, *Int. J. Quantum Chem.*, 2000, **77**, 174–186.

29 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins-Structure Funct. Bioinforma.*, 2010, **78**, 1950–1958.

30 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

31 M. T. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. V Kale, R. D. Skeel and K. Schulten, *Int. J. Supercomput. Appl. High Perform. Comput.*, 1996, **10**, 251–268.

32 M. J. Harvey, G. Giupponi and G. De Fabritiis, *J. Chem. Theory Comput.*, 2009, **5**, 1632–1639.

33 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33.

34 E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor and J. Damborsky, *Plos Comput. Biol.*, 2012, **8**.

35 J. Guo, Y. M. Cheng, J. P. Lees-Miller, L. L. Perissinotti, T. W. Claydon, C. M. Hull, S. Thouta, D. E. Roach, S. Durdagi, S. Y. Noskov and H. J. Duff, *Biophys. J.*, 2015, **108**, 1400–1413.

36 D. M. A. Smith, T. P. Straatsma and T. C. Squier, *Biophys. J.*, 2012, **103**, 1576–1584.

37 E. D. Mazumder, C. Jardin, B. Vogel, E. Heck, B. Scholz, D. Lengenfelder, H. Sticht and A. Ensser, *PLoS One*, 2012, **7**.

38 J. M. Sanders, M. E. Wampole, M. L. Thakur and E. Wickstrom, *PLoS One*, 2013, **8**.

39 A. K. Nair, A. Gautieri and M. J. Buehler, *Biomacromolecules*, 2014, **15**, 2494–500.

40 A. Krokhotin and N. V. Dokholyan, *Methods Enzymol.*, 2015, 553, 65–89.

41 B. R. Miller, T. D. Mcgee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.

42 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, 26, 1668–1688.

43 W. Xue, D. Pan, Y. Yang, H. Liu and X. Yao, *Antiviral Res.*, 2012, **93**, 126–137.

44 R. M. Levy, L. Y. Zhang, E. Gallicchio and A. K. Felts, *J. Am. Chem. Soc.*, 2003, **125**, 9523–9530.

45 D. Sitkoff, K. A. Sharp and B. Honig, *Biophys. Chem.*, 1994, **51**, 397–409.

46 L. T. Chong, Y. Duan, L. Wang, I. Massova and P. a Kollman, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 14330–14335.

47 H. Tzoupis, G. Leonis, T. Mavromoustakos and M. G. Papadopoulos, *J. Chem. Theory Comput.*, 2013, **9**, 1754–1764.

48 T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.

49 M. De Vivo, M. Masetti, G. Bottegoni and A. Cavalli, *J. Med. Chem.*, 2016, **59**, 4035–4061.

50 N. Foloppe and R. Hubbard, *Curr. Med. Chem.*, 2006, **13**, 3583–3608.

51 J. Wang, T. Hou and X. Xu, *Curr. Comput. Aided. Drug Des.*, 2006, **2**, 287–306.

52 N. Homeyer and H. Gohlke, *Mol. Inform.*, 2012, **31**, 114–122.

53 T. Yang, J. C. Wu, C. Yan, Y. Wang, R. Luo, M. B. Gonzales, K. N. Dalby and P. Ren, *Proteins Struct. Funct. Bioinforma.*, 2011, **79**, 1940–1951.

54 N. Spackova, T. E. Cheatham III, F. Ryjacek, F. Lankas, L. van Meervelt, P. Hobza and J. Sponer, *J. Am. Chem. Soc.*, 2003, **125**, 1759–1769.

55 S. Kim, S. Ferri, W. Tsugawa, K. Mori and K. Sode, *Biotechnol. Bioeng.*, 2010, **106**, 358–366.

56 K. Hirokawa, K. Gomi and N. Kajiyama, *Biochem. Biophys. Res. Commun.*, 2003, **311**, 104–111.

57 Z. Lin and J. Zheng, *Appl. Microbiol. Biotechnol.*, 2010, **86**, 1613–9.

58 M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. Mcgettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, *Bioinformatics*, 2007, **23**, 2947–2948.

59 N. Saitou and M. Nei, *Mol. Biol. Evol.*, 1987, **4**, 406–425.

60 M. Kimura, *J. Mol. Evol.*, 1980, **16**, 111–120.

61 M. Takahashi, M. Pischetsrieder and V. M. Monnier, *J. Biol. Chem.*, 1997, **272**, 12505–12507.

62 X. L. Wu, M. Takahashi, S. G. Chen and V. M. Monnier, *Biochemistry*, 2000, **39**, 1515–1521.

63 M. Takahashi, M. Pischetsrieder and V. M. Monnier, *J. Biol. Chem.*, 1997, **272**, 3437–3443.

64 S. Ferri, Y. Miyamoto, A. Sakaguchi-Mikami, W. Tsugawa and K. Sode, *Mol. Biotechnol.*, 2013, **54**, 939–943.

65 F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Z. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, D. G. Higgins, J. Söding, J. D. Thompson and D. G. Higgins, *Mol. Syst. Biol.*, 2011, **7**, 539.

66 H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley and R. Lopez, *Nucleic Acids Res.*, 2013, **41**, 597–600.

67 W. Li, A. Cowley, M. Uludag, T. Gur, H. McWilliam, S. Squizzato, Y. M. Park, N. Buso and R. Lopez, *Nucleic Acids Res.*, 2015, **43**, W580–4.

68 X. Robert and P. Gouet, *Nucleic Acids Res.*, 2014, **42**, W320–W324.

69 S. Ferri, S. Kim, W. Tsugawa and K. Sode, *J. Diabetes Sci. Technol.*, 2009, **3**, 585–592.

70 K. Hirokawa, K. Gomi, M. Bakke and N. Kajiyama, *Arch. Microbiol.*, 2003, **180**, 227–231.

FIGURES AND FIGURE LEGENDS

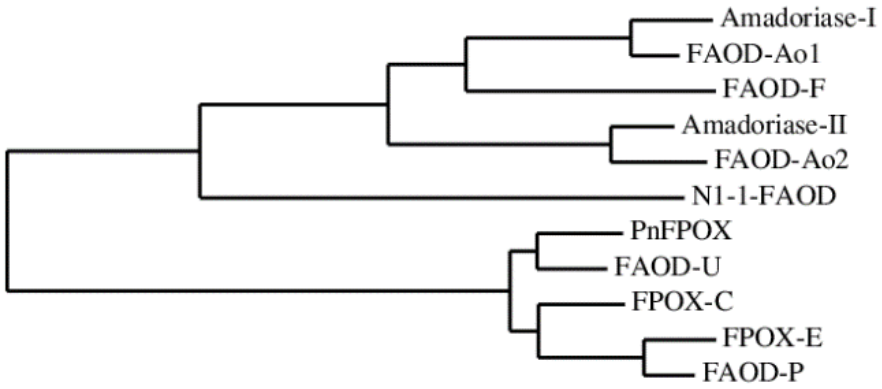


Figure 1 - Phylogenetic tree. Phylogenetic tree of fructosyl amino acid oxidase enzymes generated using the Clustal Omega program with the neighbor-joining algorithm[58–60].

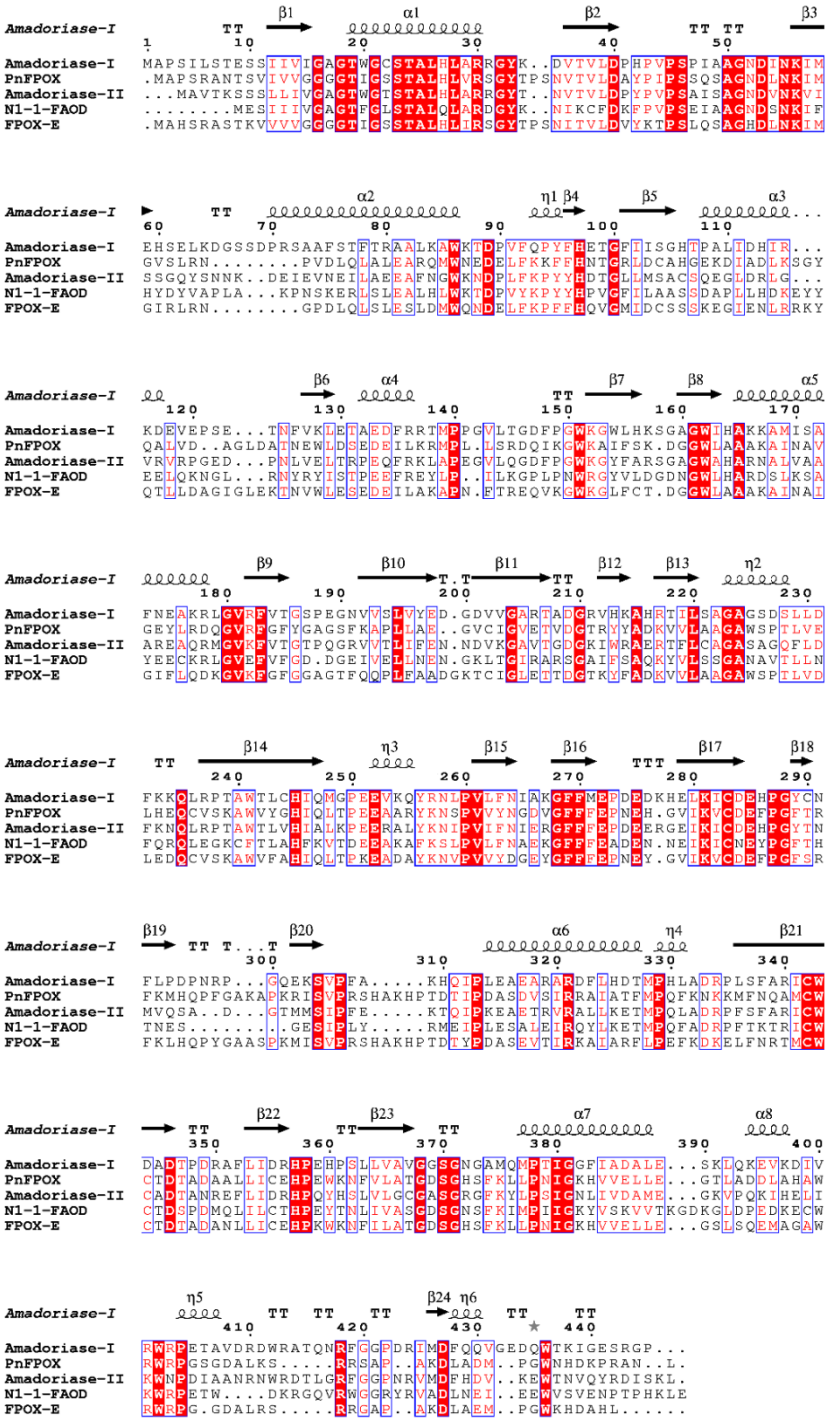


Figure 2 - Sequence alignment. Structure-based sequence alignment of the five selected FAOD enzymes. Amadoriase I and Amadoriase II sequences from *Aspergillus Fumigatus*[61–63], PnFPOX from *Phaeospheria nodorum*[64], and N1-1- FAOD from *Pichia* species N11[4,7] were aligned using ClustalW 1.82[65–67] and colored using ESPript 3.0[68] White in a red box shows strict identity, while red in a white box indicate similarity. Three residues are conserved at the catalytic positions: E285, G371 and R418 (Amadoriase I numbering).

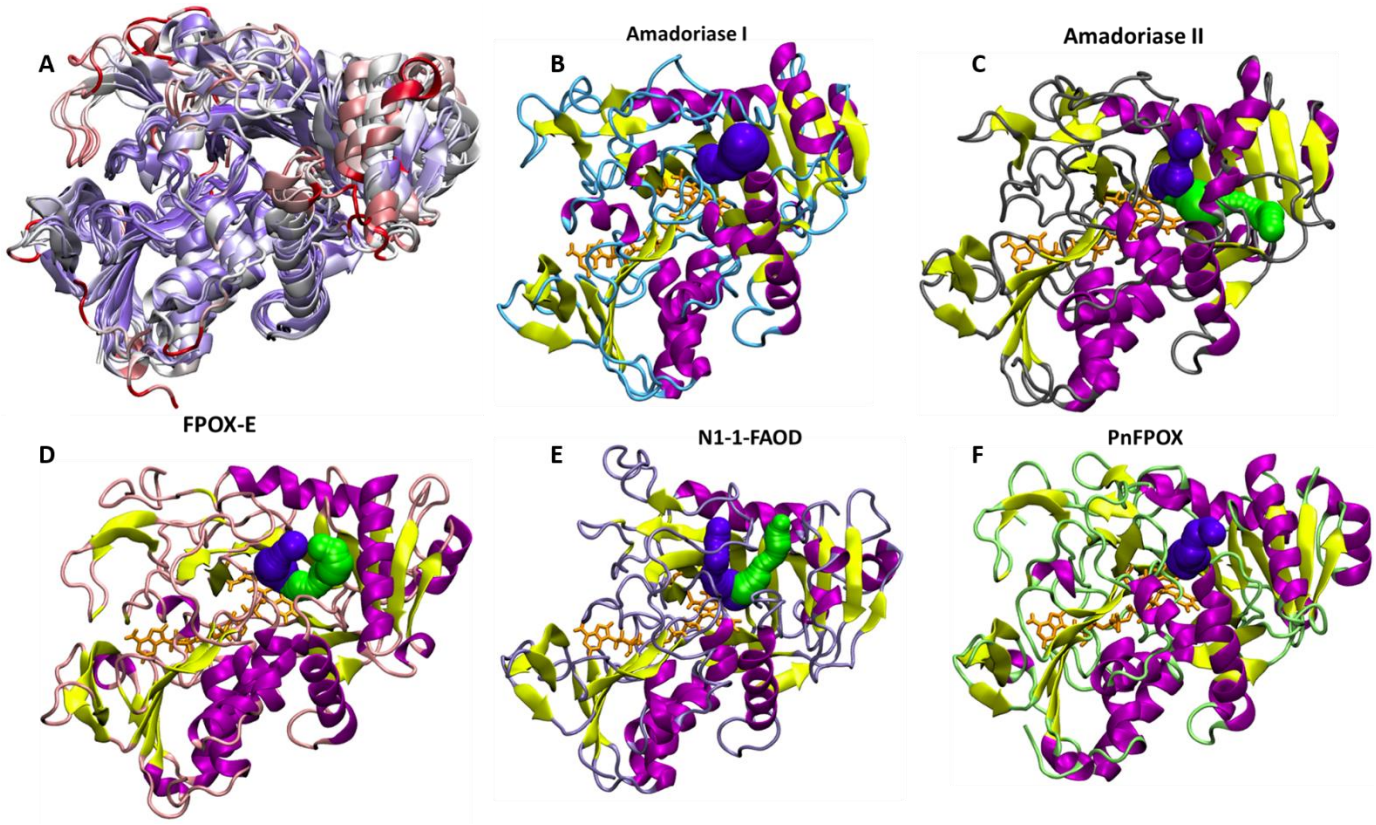


Figure 3 – FAOX structures. (A) Superimposition of five wildtype enzyme structures colored by conserved regions. Blue, white and red indicate high, medium and low structure conservation regions, respectively. The remaining panels (B-F) show the overall folding and tunnel of the five FAOX. The main tunnel (shown in blue) is of similar length in all FAOX enzymes (14 Å).

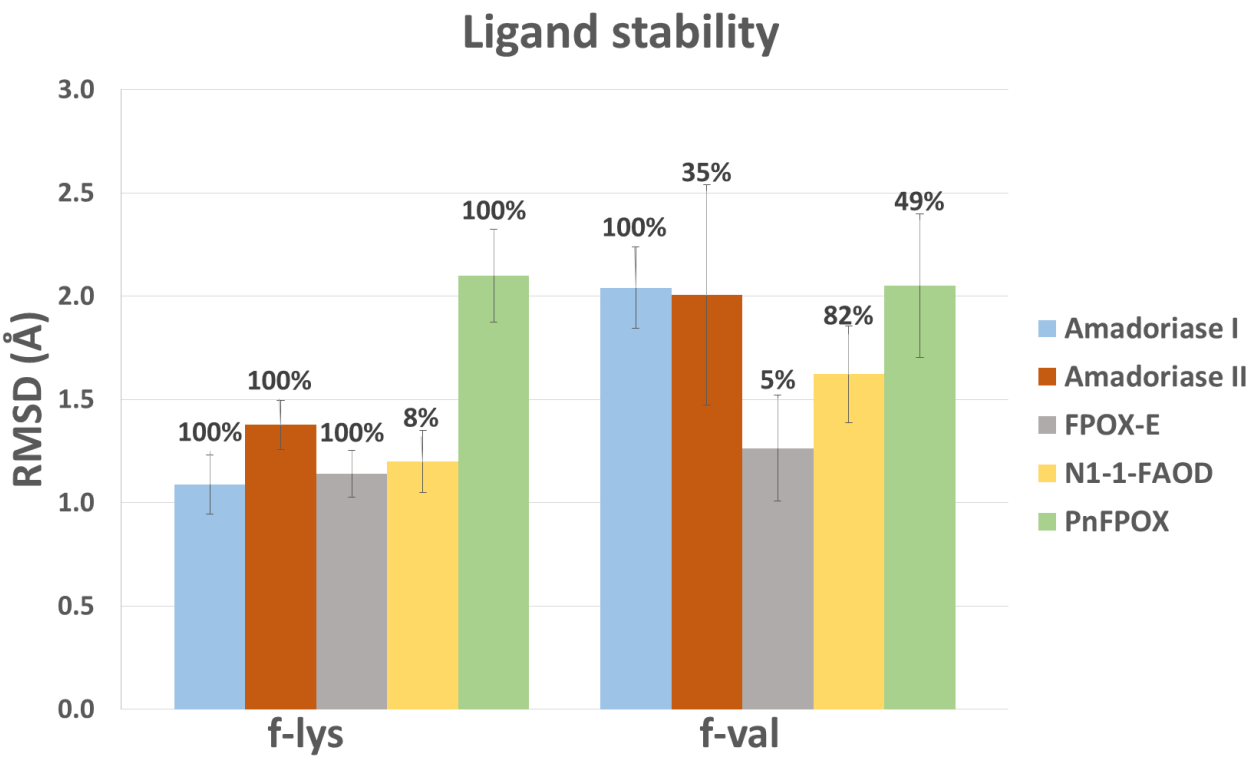


Figure 4 – Ligand stability. Root mean square deviation plots of the two ligands, measured over the time of the binding trajectory in which each ligand stays within the catalytic pocket (indicated above each bar).

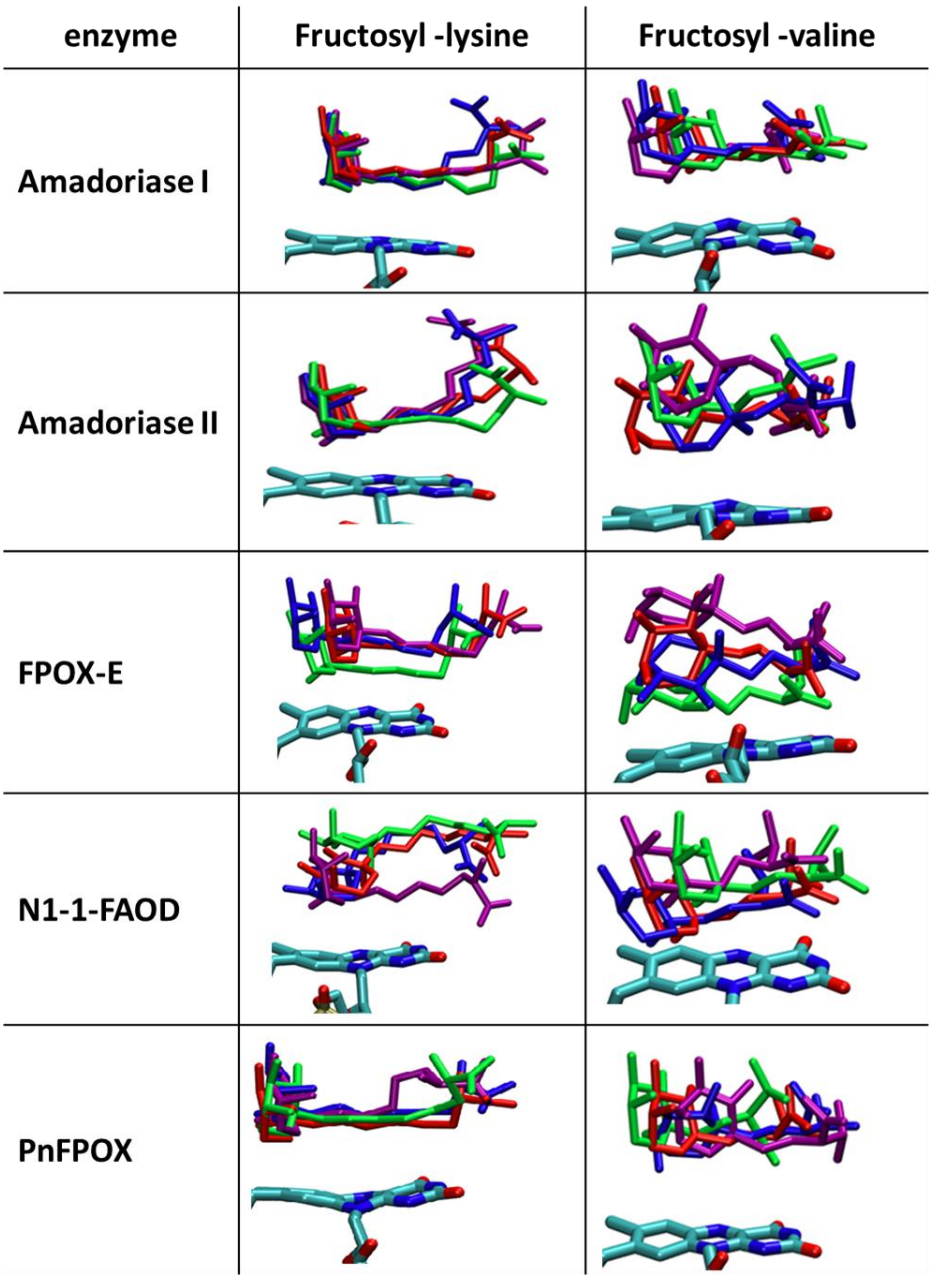


Figure 5 – Ligand conformational clusters. Each panel shows the four most common conformation clusters of f-lys and f-val inside the catalytic pocket of the five enzymes. The most populated cluster is shown in red, the second in blue and the others in green and purple, respectively. The flavin ring is shown for reference.

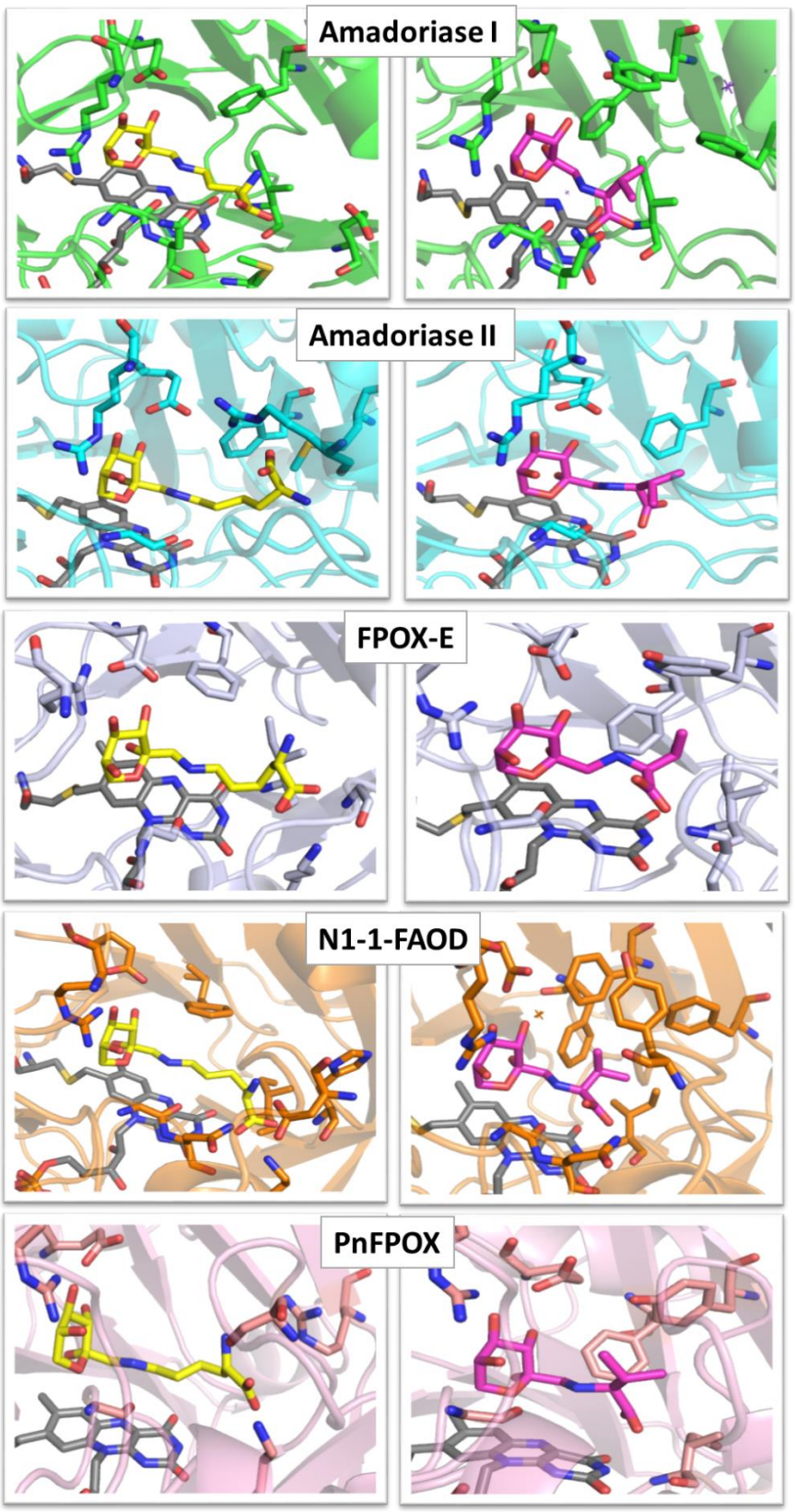


Figure 6 – Ligand binding residues. The residues forming relevant interactions (polar or hydrophobic) with f-lys (yellow) and f-val (pink) within the catalytic pocket are shown for each enzyme.

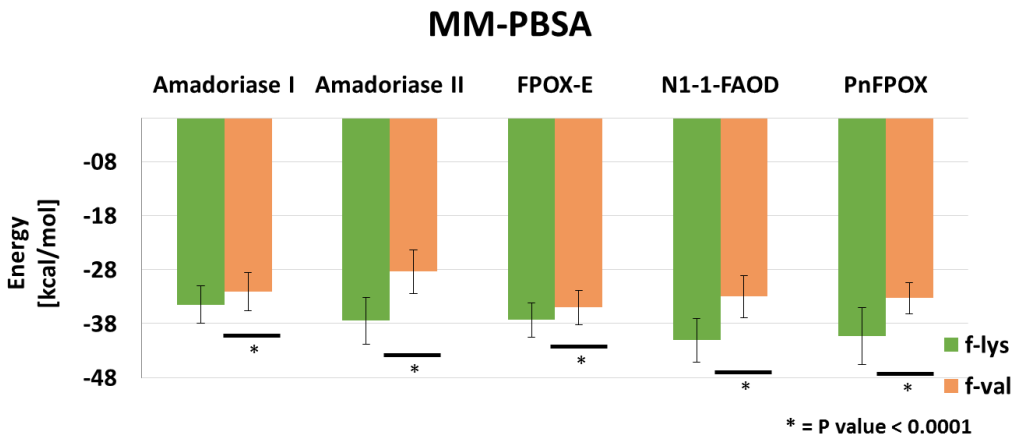


Figure 7. MM-PBSA calculations. For each enzyme, the values of the enthalpic component of the binding free energy are reported for the two different ligands. All enzymes show similar binding energies when in complex with the same ligand. Moreover, they all show a lower binding energy for f-lys compared to f-val.

TABLES AND TABLES LEGENDS

Enzyme	Amadoriase I	Amadoriase II	N1-1-FAOD	PnFPOX	FPOX-E
Amadoriase I	-	51.15	40.91	33.49	32.46
Amadoriase II	51.15	-	40.67	35.18	34.38
N1-1-FAOD	40.91	40.67	-	33.33	34.24
PnFPOX	33.49	35.18	33.33	-	72.45
FPOX-E	32.46	34.38	34.24	72.45	-

Table 1. Strict identity matrix. PnFPOX shows a high degree of identity with FPOX-E, whose structure is used as a template to build the PnFPOX structural model. Conversely, there is not an equivalent template for N1-1-FAOD, which shows less than 41% sequence identity with all the other available FAOXs structures.

Enzyme	Cavity size	Id	Average Bottleneck [Å]	Maximum Bottleneck [Å]	Average Length [Å]	Average Curvature [Å]
Amadoriase I	170.2 ± 5.2	1	1.83 ± 0.17	2.27	15.73 ± 1.63	1.22 ± 0.06
Amadoriase II	111.2 ± 5.3	1	1.83 ± 0.17	2.04	11.30 ± 1.13	1.24 ± 0.06
		2	1.56 ± 0.03	1.59	22.16 ± 1.24	1.24 ± 0.05
FPOX- E	104.4 ± 2.9	1	1.66 ± 0.22	2.25	12.92 ± 3.29	1.26 ± 0.07
		2	1.61 ± 0.21	2.15	17.95 ± 4.9	1.54 ± 0.22
N1-1-FAOD	73.2 ± 4.1	1	1.45 ± 0.12	1.82	18.88 ± 1.50	1.35 ± 0.10
		2	1.48 ± 0.12	1.63	19.27 ± 1.4	1.21 ± 0.05
PnFPOX	77 ± 2	1	$1.^{33}$ ± 0.01	1.34	13.99 ± 0.48	1.16 ± 0.03
		2	1.64 ± 0.22	2.13	19.49 ± 1.75	1.37 ± 0.13

Table 2. Cavity and tunnels. Cavity size is measured as the average number of water molecules that can be accommodated within the cavity. We report the main features for each tunnel (id 1 and 2, where suitable) in terms of average and maximum radius of the tunnels' bottleneck, average length of the tunnels and their curvatures.

Residue Type	Amadoriase I	Amadoriase II	FPOX-E	N-1-1 FAOD	PnFPOX
Charged	32	39	45	28	41
Polar	14	11	9	4	18
Hydrophobic	50	39	25	48	22
Special	5	11	21	20	19

Table 3. Tunnel-lining residue types. All the residues that are found within 3.2 Å from the tunnel borders for at least 25 ns are reported, classified by chemical nature (Charged: R,K,E,D,H. Polar: S, T,N,Q. Hydrophobic: A, V, I, L, M, F, W, Y. Special: C,G,P). For FPOX-E and PnFPOX, results are mediated for the two tunnels because of their similar pathway and identical exit gates. For N1-1-FAOD and Amadoriase II only the results for the principal tunnel are reported.

Enzyme	f-lys		f-val	
	K_m [mM]	V_{max} [mmol/min/mg]	K_m [mM]	V_{max} [mmol/min/mg]
Amadoriase I				
Collard et al, 2008	0.08	6.61	5.20	4.96
Amadoriase II				
Collard et al, 2008	2.4	1.51	0.43	3.16
Kim et al, 2010	-	-	1.9	13
N1-1-FAOD				
Miura et al, 2006	9.8	0.36	5.6	0.38
Miura et al, 2007	10	0.88	5.6	0.93
Ferri et al, 2009	0.9	9.5	3.9	6.5
Kim et al, 2010	-	-	4.8	9.2
PnFPOX				
Kim et al, 2010 (crude)	25.2	0.203	0.48	2.34
Kim et al, 2010 (purified)	12.2	2.56	0.45	24.2
Kim et al, 2012	-	-	0.46	24
FPOX-E				
Hirokawa et al, 2003	n/a	0.42	0.318	20.6

Table 4. Summary of the kinetic parameters of wild-type Amadoriase I[13] and II[7,13], N1-1-FAOD[4,6,7,69], PnFPOX[8,55], FPOX-E[56,70].

SUPPLEMENTARY INFORMATION

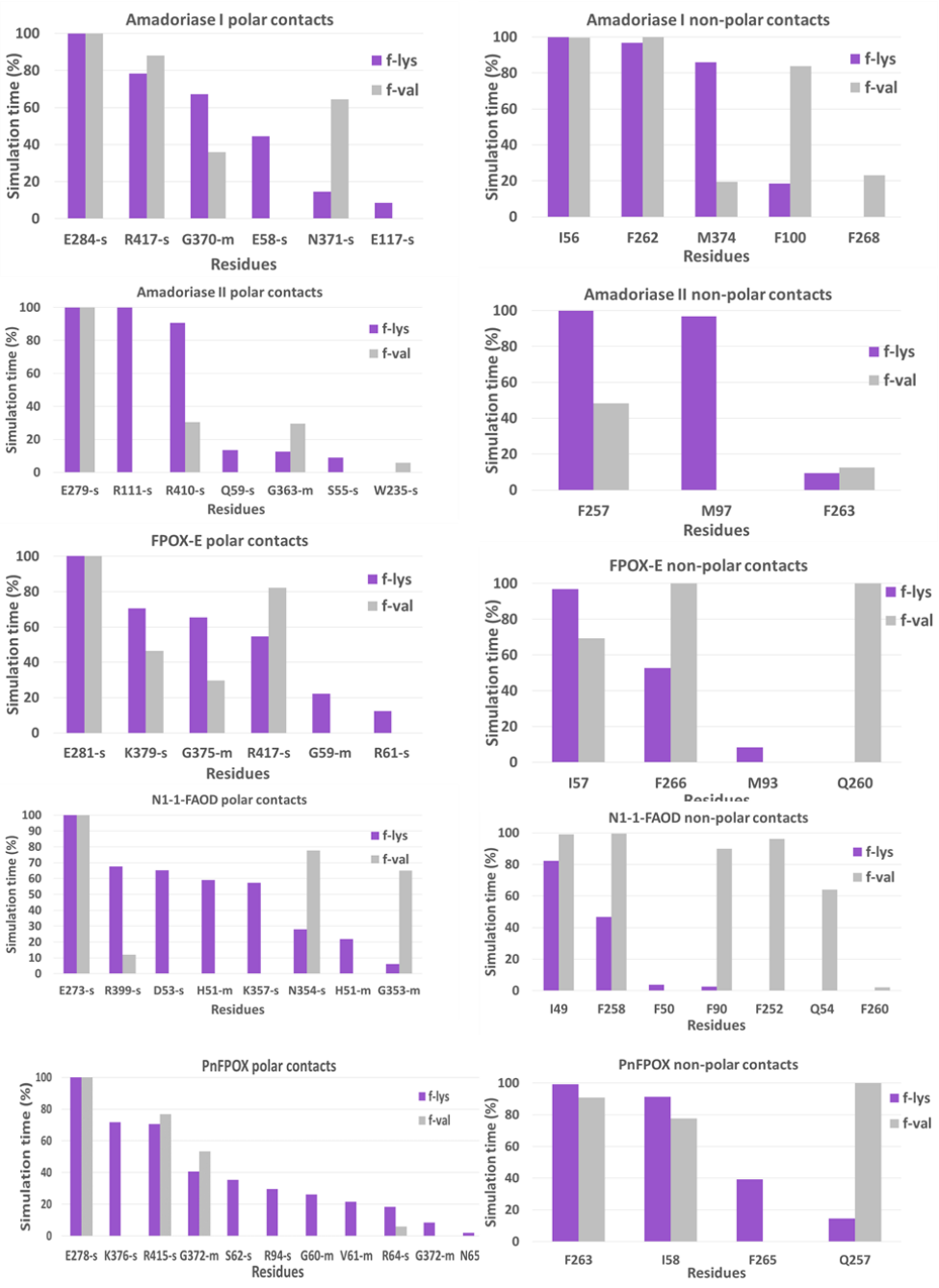
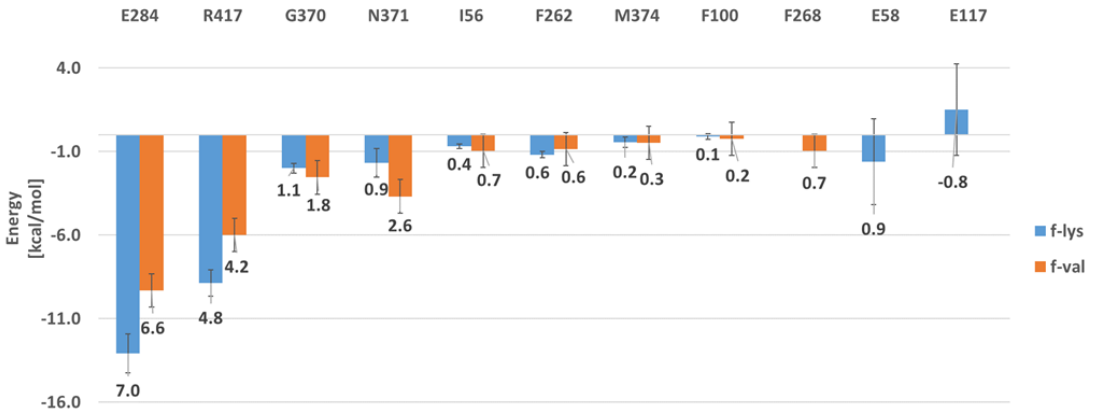


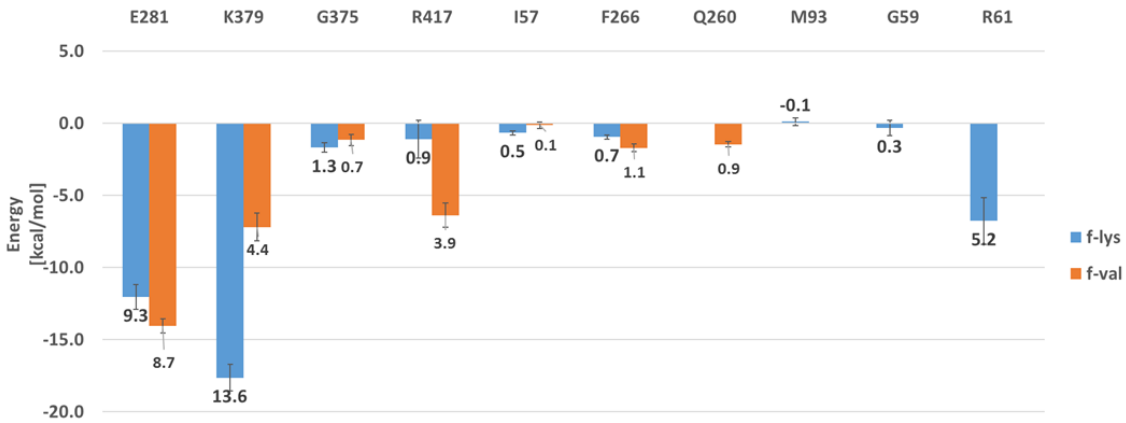
Figure S1. Lists of polar and nonpolar interactions of each enzyme with both ligands. A threshold of 2 % of persistence during simulation time is applied to select interactions for analysis. The durability of the contacts is evaluated along the MD trajectory by analyzing the frames where the ligand-sugar moiety is stable and correctly positioned in the catalytic pocket. Therefore, the percentage reported in the table are relative to the time of complex stability.

Amadoriase I



Amadoriase II

FPOX-E



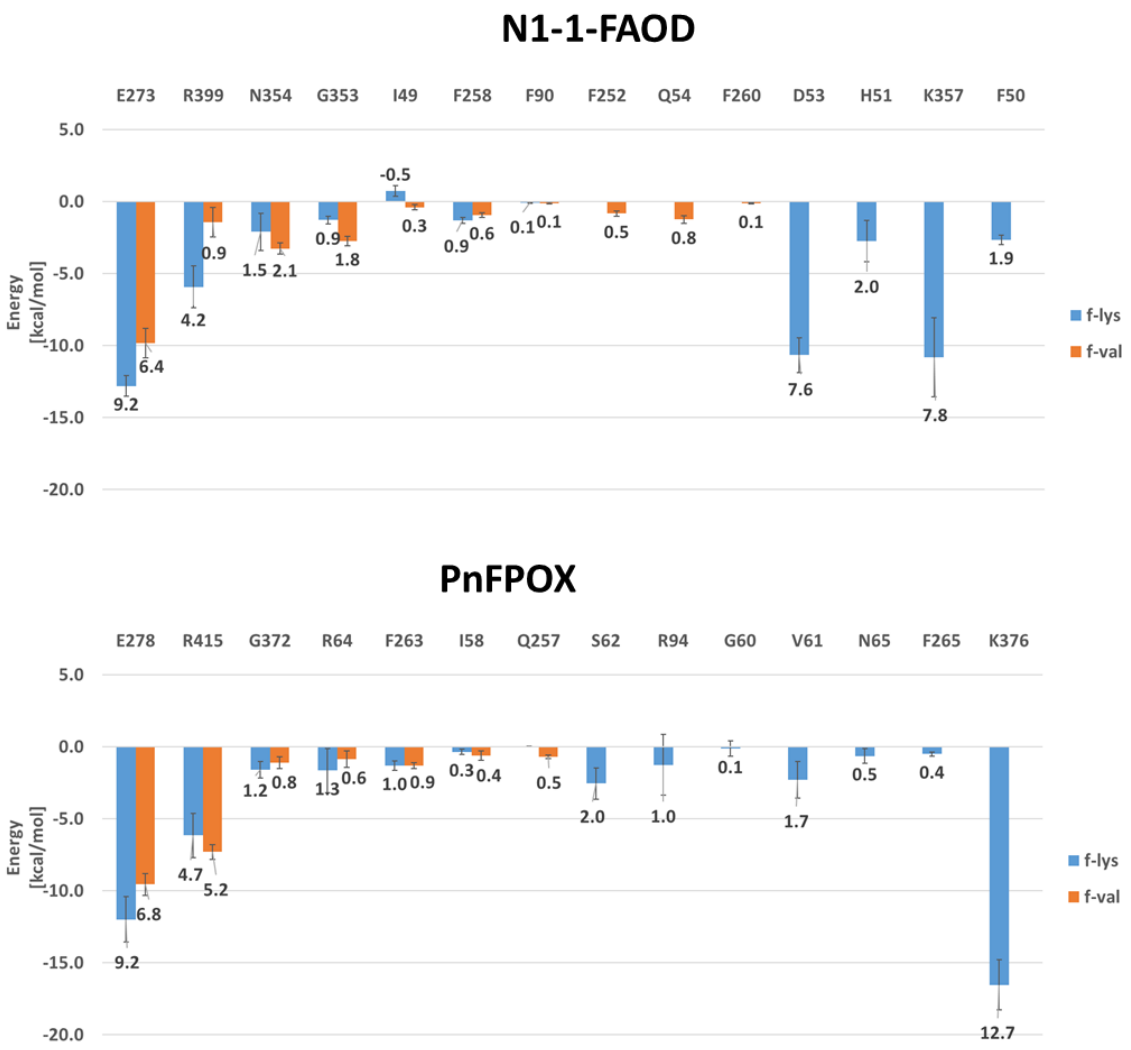


Figure S2. Decomposition of the enthalpic component of the binding free energy contribution on the enzyme residues that mostly interact with the ligands. The labels at the top of each histogram are the percentage of relative contribution compared to the total enzyme-ligand interactions.

Enzyme	f-lys		f-val	
	PSASAD	ε_{int}	PSASAD	ε_{int}
Amadoriase I	71	4	27	2
Amadoriase II	43	3	30	2
FPOX-E	61	3	47	3
N1-1-FAOD	51	3	34	2
PnFPOX	52	3	33	2

Table S1. The PSASAD is defined as the Solvent Accessible Surface Area Difference (SASAD) between ligand-bonded and apo states of the protein polar atoms that strongly interact with the ligand. To calculate PSASAD, we chose polar atoms within a 5 Å distance from f-lys or f-val for each enzyme and we estimated SASAD with and without ligands. We excluded the f-val hydrophobic tail from the SASA calculation owing to its inability to interact with polar atoms. Due to the charged nature of the binding interface for all the enzymes, we limited our choice of internal dielectric constant (ε_{int}) to values greater than 2. Based on the PSASAD calculation, all enzymes show a ε_{int} in the range between 2.0 and 4.0. For binding energy calculations, we restricted our analysis to those frames where the sugar moiety of the ligand is correctly positioned within the catalytic pocket. The binding energy was then computed as the average over all the relevant frames of the trajectory.