



## OPEN Tailoring FPOX enzymes for enhanced stability and expanded substrate recognition

Hajar Estiri<sup>1,6</sup>, Shapla Bhattacharya<sup>1,2,6</sup>, Jhon Alexander Rodriguez Buitrago<sup>1</sup>, Rossella Castagna<sup>1,3</sup>, Linda Legzdina<sup>1</sup>, Giorgia Casucci<sup>1</sup>, Andrea Ricci<sup>4</sup>, Emilio Parisini<sup>1,5</sup>✉ & Alfonso Gautieri<sup>4</sup>✉

Fructosyl peptide oxidases (FPOX) are deglycating enzymes that find application as key enzymatic components in diabetes monitoring devices. Indeed, their use with blood samples can provide a measurement of the concentration of glycosylated hemoglobin and glycosylated albumin, two well-known diabetes markers. However, the FPOX currently employed in enzymatic assays cannot directly detect whole glycosylated proteins, making it necessary to perform a preliminary proteolytic treatment of the target protein to generate small glycosylated peptides that can act as viable substrates for the enzyme. This is a costly and time consuming step. In this work, we used an *in silico* protein engineering approach to enhance the overall thermal stability of the enzyme and to improve its catalytic activity toward large substrates. The final design shows a marked improvement in thermal stability relative to the wild type enzyme, a distinct widening of its access tunnel and significant enzymatic activity towards a range of glycosylated substrates.

The current clinical standard for monitoring blood glucose levels is glycosylated hemoglobin (HbA1c). Indeed, thanks to its longevity (its half-life is up to 3 months), HbA1c is the marker of election for diabetes diagnosis. To this end, an effective protocol that includes the exploitation of fructosyl peptide oxidases (FPOX) for measuring HbA1c concentration in blood samples has been developed<sup>1–3</sup>. The first step of the protocol involves a proteolytic digestion of HbA1c to release the glycosylated N-terminal dipeptide, fructosyl-valine-histidine (fVH); the dipeptide is then hydrolyzed by a FPOX, a process that results in the production of hydrogen peroxide, which can be measured via a colorimetric assay by coupling horseradish peroxidase and an appropriate chromophore<sup>4</sup>. Likewise, FPOX-based methods can also be used to measure glycosylated albumin (GA), which is a short- and medium-term glycaemic diabetes marker (half-life up to 3 weeks)<sup>5</sup>.

FPOXs are suitable enzymes for diabetes assays thanks to their reactivity on up to six amino acid-long fructosyl peptides<sup>6</sup>. FPOXs belong to the larger family of Fructosyl Amino Acid Oxidases (FAOX)<sup>7–9</sup>, which are enzymes of fungal or bacterial origin that cleave low molecular weight Amadori products, freeing the amine group and generating glucosone and hydrogen peroxide<sup>10–12</sup>.

Other than diabetes detection and surveillance, FPOXs are also considered to be promising therapeutic tools for protein deglycation in biological tissues<sup>13</sup>. Protein glycation is a spontaneous and irreversible non-enzymatic reaction whereby a sugar binds to a protein and generates a covalent adduct named Amadori product<sup>14</sup>. Protein glycation triggers many diabetes-related clinical outcomes, such as for instance arterial stiffening<sup>15</sup>, nephropathy<sup>16</sup>, retinopathy<sup>17</sup> and neuropathy<sup>18</sup>. These complications have indeed been shown to be largely driven by the chemical modification of functional proteins<sup>19,20</sup>.

The inability of wild type FPOXs to act on intact proteins is the most critical issue limiting both their diabetes detection efficiency and their protein deglycation potential<sup>21,22</sup>. This is due to their buried active site and

<sup>1</sup>Department of Biotechnology, Latvian Institute of Organic Synthesis, Aizkraukles 21, Riga 1006, Latvia. <sup>2</sup>Faculty of Materials Science and Applied Chemistry, Riga Technical University, Paula Valdena 3, Riga 1048, Latvia. <sup>3</sup>Dipartimento di Chimica, Materiali e Ingegneria Chimica “Giulio Natta”, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy. <sup>4</sup>Biomolecular Engineering Lab, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy. <sup>5</sup>Department of Chemistry “G. Ciamician”, University of Bologna, Via Selmi 2, 40126 Bologna, Italy. <sup>6</sup>These authors contributed equally: Hajar Estiri and Shapla Bhattacharya. ✉email: emilio.parisini@osi.lv; alfonso.gautieri@polimi.it

narrow access tunnel, as determined in multiple crystal structures<sup>6,12,23</sup>. To the best of our knowledge, there is currently only one report describing an enzyme that shows direct activity on HbA1c without requiring any preliminary proteolytic step. This enzyme has been obtained by introducing a single point mutation (R61G) in a thermostable version of *Aspergillus nidulans*' FPOX (AnFPOX)<sup>24</sup>. Nonetheless, the activity and the stability profile of FPOX enzymes need to be further improved when considering their application in biosensors. Indeed, improving the ability of FPOXs to act on full-length glycated proteins may provide significant advantages in both the development of novel diabetes monitoring tools and in the design of innovative therapies aimed at reducing and/or preventing protein aging (i.e., the accumulation over time of advanced glycation end-products). These improvements may come from engineering an enzyme with a wider access tunnel to the catalytic site; in fact, access tunnel engineering can improve several characteristics of the enzyme, such as activity, specificity, promiscuity, enantioselectivity and stability<sup>25,26</sup>.

Our group has previously reported on a successful engineering campaign aimed at widening the access tunnel to the catalytic site of *Phaeosphaeria nodorum* FPOX (PnFPOX)<sup>27</sup>. In that account, using a rational design approach involving a five-amino acid deletion at the entrance of the active site and a number of further mutations, we described a stable PnFPOX mutant (named L3-35A) that, relative to the wild type enzyme, features a significantly wider and shorter access tunnel. However, the backbone redesign also introduced structural instability and lowered the activity of the engineered enzyme. For this reason, in this work we aimed at improving the L3-35A design and restore the stability and activity of the enzyme, thus providing a FPOX that presents high activity and stability together with a wider access tunnel with respect to the parental enzyme PnFPOX.

In the present study, we set out to enhance the thermal stability of the L3-35A mutant by implementing a methodological approach whereby different thermal stabilization strategies are tested in parallel and their efficiency compared. These various strategies involved mutations to either improve the RMSF of the protein (generating mutants hereafter referred to as D-series mutants), increase the number of salt bridges (C-series mutants) or form disulfide bonds (X-series mutants). Whereas, of all methods, engineered disulfide bond formation usually provides the enzyme with the highest  $T_m$ , the other two methods may in fact be useful when engineering a protein that already contains one or more cysteines. In such cases, the introduction of further cysteine residues may interfere with natural disulfide bond patterns and result in high levels of insoluble protein formation. In essence, we used our PnFPOX mutant L3-35 as a platform to test our design approach on different thermal stabilization strategies, and to provide a systematic comparison of their efficiency relative to the same parent enzyme. Our design process involved the use of extensive molecular dynamics (MD) simulations and targeted optimization using Rosetta-based design<sup>28</sup>, followed by experimental validation of a small number of selected high-scoring mutants. In our validation procedure, we assessed the activity of the selected mutants against several substrates, including small glycated amino acids such as fructosyl-Val (fV) and fructosyl-Lys (fK), the dipeptide fructosyl-Val-His (fVH) and the hexapeptide fructosyl-Val-His-Leu-Thr-Pro-Glu (F6P). Overall, the data shown herein provide evidence that our rational design approach allows the rapid identification of mutants featuring the desired phenotype.

## Results and discussion

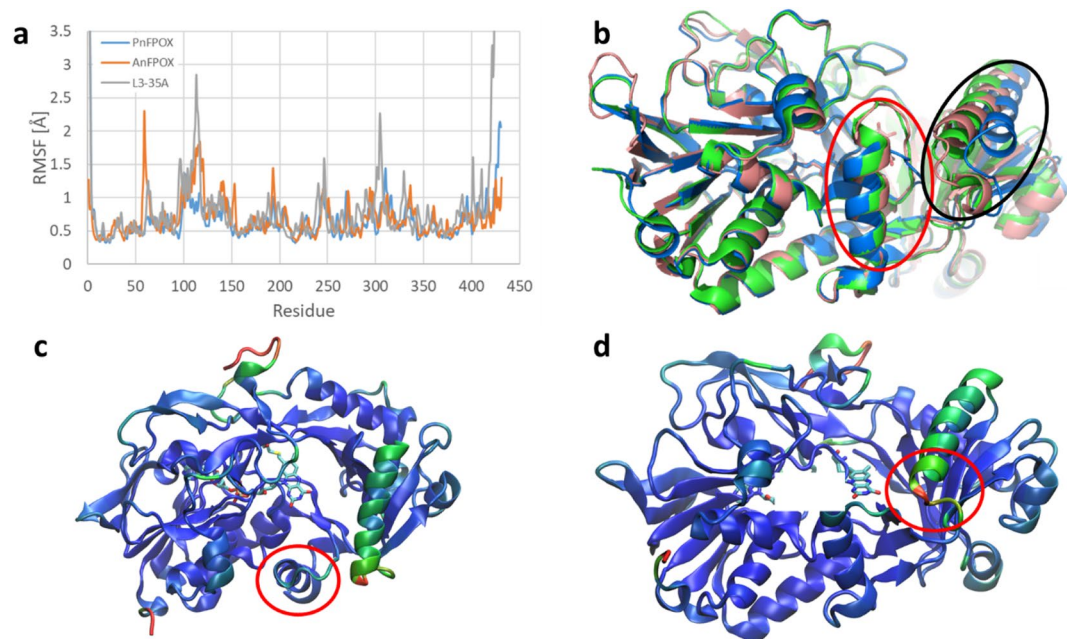
### Stability optimization of the L3-35A PnFPOX mutant

Starting from the crystal structure of the L3-35A PnFPOX mutant (PDB code 6Y4J), we performed MD simulations at variable temperatures, followed by root mean square fluctuation (RMSF) analysis. Our data suggests that, with the exception of the N- and the C-termini, the most unstable regions of the enzyme are helix 96-111 and the surrounding structures (Fig. 1). The relative instability of this helix is likely due to the redesign of residues 58–65 of wild type PnFPOX (red circle in Fig. 1c)<sup>27</sup>. Indeed, while these changes resulted in the widening of the access tunnel to the active site, they also caused the removal of the amino acids facing helix 96-111, thus abrogating previously existing stabilizing interactions. Further regions that present a lower average stability are loops 240–247, 301–307, 401–410.

The use of the Rosetta “Supercharge” method, which aimed at increasing the charge on the surface of the enzyme, led to the design of 1000 enzyme variants, from which those 20 (named C01–C20) that featured the best Rosetta score were selected for further analysis. Similarly, single point mutations screening led to 8095 designs, from which the best 20 were retained (S01–S20). Double point mutations screening generated 271,000 designs (leading to the selection D01–D20), PROSS design produced 9 variants (P01–P09), and disulfide bond design identified 62 enzyme variants (from which we selected the best 8, named X01–X08).

The 77 candidate enzyme variants originating from our different screening methods were then subject to MD simulations and ranked based on their stability at increasing temperatures, as assessed by RMSF analysis. Our MD-based ranking led to the selection of six candidate mutants (C16, D02, X01, X02A, X04, and X07) for experimental production and characterization.

For measuring the thermal stability of the different mutants, we used both thermal shift assay (differential scanning fluorimetry or DSF) and circular dichroism (CD). Both of these methods are sensitive to the thermal stability of the enzymes and can provide information on the structural changes that occur as a function of temperature. As expected, the two methodologies provide very similar results in terms of trends and absolute values and demonstrate that the rational design procedure that we adopted was effective in enhancing the thermal stability ( $T_m$ ) of our reference FPOX enzyme, L3-35A (see Table 1). The D02 variant, which relative to L3-35A features two additional mutations (V110R and D115G), shows an increase in  $T_m$  of  $\approx 1.5$  °C with respect to the parent enzyme. On the other hand, the introduction of extensive surface charges and salt bridges (mutant C16) resulted in a further increase in  $T_m$  of  $\approx 1.5$  °C ( $T_m = 55.2$  °C). Finally, the incorporation of a single disulfide bond in different positions on the L3-35A scaffold led to a significant improvement in thermal stability of all the tested mutants of this series: X01 (54.1 °C), X02 (60.0 °C), X04 (55.2 °C) and X07 (55.3 °C).



**Figure 1.** Stability analysis from MD simulations. The RMSF profile (a) shows that in the L3-35A, the helix 96–111 and the surrounding structures present a high degree of instability compared to the other enzymes. In the structural comparison (b, PnFPOX:green, AnFPOX:pink and L3-35A:blue), this helix (circled black) shows also a different fold compared to the other two enzymes. The instability and different structural arrangement could be due to design operated on L3-35A, which removed part of the helix blocking the tunnel (c, circled red). These removed residues were facing the helix 96–111 in the parental enzymes and were making stabilizing interactions, which could explain why helix 96–111 is unstable in L3-35A, possibly leading to its lower activity (d, color coded by RMSF, where blue represents lower RMSF and green–red the regions with higher RMSF). Comparison of RMSF and structure of PnFPOX (green), AnFPOX (pink) and L3-35A (blue).

Enzyme	$T_m$ (°C) <sup>a</sup>	$T_m$ (°C) <sup>b</sup>
PnFPOX (WT)	53.2 ± 0.2	53.5 ± 0.1
L3-35A	52.3 ± 0.2	52.9 ± 0.1
D02	53.1 ± 0.5	54.8 ± 0.1
C16	55.2 ± 0.1	55.0 ± 0.1
X01	54.1 ± 0.1	54.2 ± 0.1
X02A	60.0 ± 0.4	60.1 ± 0.1
X04	55.2 ± 0.3	55.0 ± 0.1
X07	55.3 ± 0.4	55.4 ± 0.1
X02B	60.2 ± 0.7	60.6 ± 0.1
X02C	64.0 ± 0.2	63.3 ± 0.1

**Table 1.** Melting temperature of the parent enzyme (L3-35A) and its mutants as determined by thermal shift assay (a) and circular dichroism (b). The melting temperature of the wild type enzyme (PnFPOX) is also provided as reference.

To assess the activity of these various enzyme variants, we employed a colorimetric enzymatic assay that allowed us to quantify the release of glucosone probed at 322 nm. We tested the activity of the enzymes on fructosyl-lysine (fK), fructosyl-valine (fV), fructosyl-valine-histidine (fVH) and fructosyl-Val-His-Leu-Thr-Pro-Glu (F6P) (see Table 2). We observed that the activity of the enzyme variants towards fK and fV remained similar to that of the parent enzyme, with a slight increase of activity in the case of D02 and a slight decrease in all the other cases. Overall, we observed that the activity of these mutant enzymes is several folds lower than that of PnFPOX, the naturally occurring enzyme. Also, similarly to the parent enzyme, none of these mutants shows any activity towards the dipeptide or the hexapeptide.

Based on our thermostability results, we selected the X02A variant for its high  $T_m$ , high expression yield, and detectable activity toward glycosylated amino acids and we set out to increase its activity toward larger substrates (i.e., fVH and F6P). The activity toward these substrates was present in the wild type enzyme PnFPOX but was lost

Enzyme	Specific activity (U/mg)			
	fK	fV	fVH	F6P
PnFPOX	30.18 ± 0.67	29.67 ± 2.56	32.60 ± 1.18	0.78 ± 0.09
AnFPOX-47	–	–	–	0.082 ± 0.002 <sup>a</sup>
L3-35A	0.21 ± 0.02	0.16 ± 0.04	–	–
D02	0.30 ± 0.03	0.32 ± 0.01	–	–
C16	ND	ND	–	–
X01	ND	ND	–	–
X02A	0.15 ± 0.01	0.11 ± 0.01	–	–
X04	ND	0.13 ± 0.01	–	–
X07	–	0.08 ± 0.01	–	–
X02B	2.24 ± 0.07	32.50 ± 0.44	0.87 ± 0.03	0.18 ± 0.06
X02C	1.06 ± 0.01	17.95 ± 2.46	1.62 ± 0.02	0.43 ± 0.06

**Table 2.** Comparison of the enzymatic specific activity (U/mg) on different substrates. Tests are performed in triplicates. ND not detected. <sup>a</sup>Data from reference<sup>24</sup>.

in the L3-35A design. The L3-35A design involved the engineering of the entry tunnel of the wild type PnFPOX enzyme via the shortening of a loop and a helix<sup>27</sup>. In this process, while the amino acids that are directly involved in the catalytic function were not modified, some tunnel-lining residues were replaced. This could explain the reduced activity of both the L3-35A enzyme previously reported and its thermally-stabilized mutants described so far (D02, C16, X01, X02A, X04, X07). Hence, we set out to design mutations to enhance the catalytic activity of X02A, the most thermally-stable mutant of our library (Table 1).

To improve the catalytic activity of the engineered enzymes, the most promising variants resulting from the stabilization steps were sequence- and structurally-aligned with the most active enzymes against the fructosyl-hexapeptide F6P that are available in the literature, namely PnFPOX, FPOX-C and related variants<sup>6</sup> and AnFPOX and related variants<sup>24</sup>. Specifically, tunnel-lining residues associated with improvements in the catalytic activity were selected to be introduced in the engineered enzyme (Fig. 2). Based on this comparison, we identified two designs in which enzyme activity could be partially restored. In the first design (X02B), four mutations were introduced in positions that are directly facing the tunnel in order to match the residues found in PnFPOX (D60G, A61V, D62S, D368H). In the second design (X02C), we also introduced the N56A mutation, which was shown to improve activity in the PnFPOX enzyme<sup>6</sup>, as well as a series of mutations in the loop 61–71. These mutations (A61I, D62S, A63G, D64A, A66L, A67S, D68L, A69E, R71F) were derived from the consensus sequence of PnFPOX and AnFPOX-47, the most active FPOX enzymes currently available.

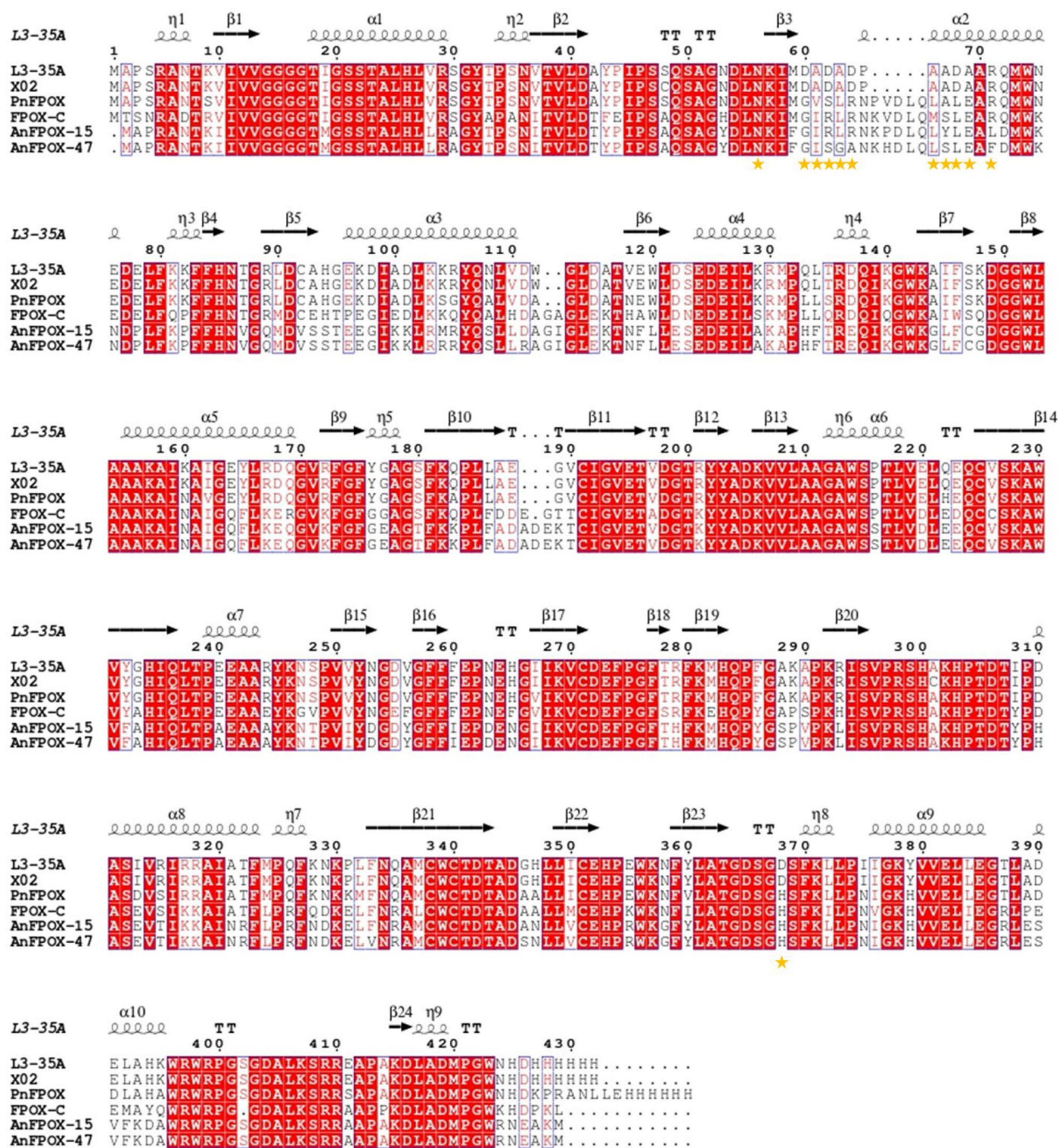
In essence, by reintroducing conserved tunnel-lining residues in the X02A sequence, we generated the X02B and X02C mutants (Fig. 3) and we tested their activity toward the same panel of substrates that were tested with the parent enzyme (Table 2). It is worth noting that the X02B and X02C enzymes present limited activity toward the dipeptide ( $\approx 0.87$  and  $1.62$  U/mg, respectively) compared to the WT ( $32.60$  U/mg). However, while the WT significantly shows little, if any, ability to catalyze the oxidative reaction on the F6P hexapeptide ( $0.78$  U/mg, which is equivalent to 2.4% of the activity on fVH), the X02B and X02C mutants feature significant activity also on the hexapeptide ( $0.18$  and  $0.43$  U/mg, which are equivalent to 20.7% and 26.5% of the activity observed on the fVH). We measured the kinetic parameters (Table 3) for the WT and the latest engineered enzyme (X02C) against the fructosyl-valine substrate, which are found consistent with those reported in the literature for the enzyme family<sup>7,10,27,29</sup>. For the substrate fructosyl-hexapeptide, a comparison of the  $K_m$  of our mutant X02C ( $0.03$  mM) with the wild type PnFPOX ( $0.81$  mM) shows higher affinity for the substrate.

To further compare our final design with similar FPOX, we compared it with a recently published engineered FPOX, named AnFPOX-47, which is reported to be the most active on intact glycated hemoglobin among the reported designs that present a wider tunnel<sup>24</sup>. The comparison shows that X02B and X02C have significantly higher activity than AnFPOX-47 toward F6P. Moreover, X02B and X02C feature a much higher  $T_m$  relative to all the other L3-35A mutants, i.e.  $60.17$  °C and  $64.02$  °C respectively (Table 1). Overall, within our panel of mutated enzymes the thermal stability of X02C increased significantly (more than  $10$  °C) with respect to the wild type PnFPOX. Interestingly, thermal stabilization also resulted in a significant increase in the production yield, the X02C variant featuring a nearly three-fold increase in expression yield with respect to the WT.

### Structural correlation analysis

To identify the possible molecular mechanism underlying the differences in enzymatic activity observed in our study, we determined the crystal structures of some of our best performing variants and we compared their overall fold as well as their tunnel and catalytic site. At the structural level, the tridimensional architecture of all mutants reported herein is maintained, confirming that our mutagenesis campaign did not impair the conformational stability of the enzyme. Indeed, when comparing the overall RMSD values of all the crystallized mutants to the wild type PnFPOX (PDB code 5T1E), the differences are all within a maximum of  $1.8$  Å (see Table 4).

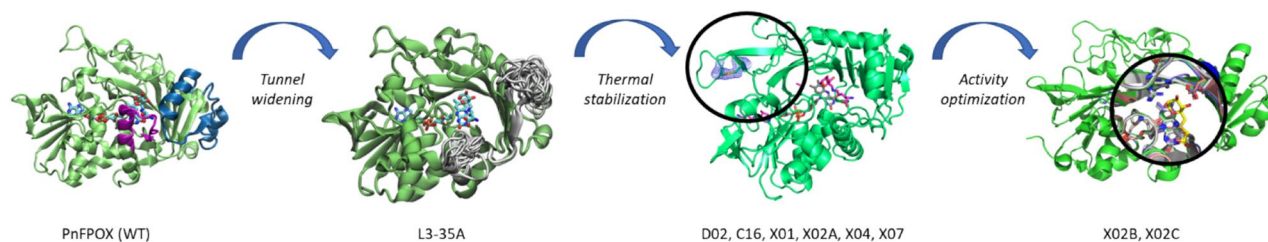
We then compared the residues in the catalytic site (W235, E278, G372, R415) of our mutants with those of the WT enzyme. These residues are highly conserved in all FPOX enzymes and bind to the sugar moiety of



**Figure 2.** Sequence alignment of homologous enzymes used to design mutations for enzyme activity enhancement. The L3-35A and X02A enzymes were sequence- and structurally-aligned against the most active enzymes on the fructosyl-hexapeptide F6P that are available in the literature, namely PnFPOX and FPOX-C<sup>6</sup>, AnFPOX-15 and AnFPOX-47<sup>24</sup>. Yellow stars represent positions considered for mutagenesis.

the substrates. We observed that the orientation of these residues is maintained (see Fig. 4), suggesting that the differences in the catalytic activity are not due to differences in the core catalytic site.

These results could be explained by the modification of the tunnel size and of its lining residues (see Fig. 5 and Table 4). In the WT, the small and long tunnel (bottleneck radius 2.2 Å, length 13.3 Å) allows greater stabilization of small substrates while limiting the entrance of larger substrates, hence explaining the marked decrease of activity when moving from glycosylated amino acids to the glycosylated hexapeptide. On the other hand, the AnFPOX-47 developed by Ogawa and coworkers<sup>24</sup> presents a slightly larger tunnel (bottleneck radius of 2.7 Å) with comparable length (13.2 Å), which may explain its activity towards larger substrates. Our previously developed L3-35A enzyme variant<sup>27</sup> exhibits a much wider and shorter tunnel (bottleneck radius of 3.7 Å and length of 6.4 Å) but shows limited activity on both small and large substrates, which could be due to the limited



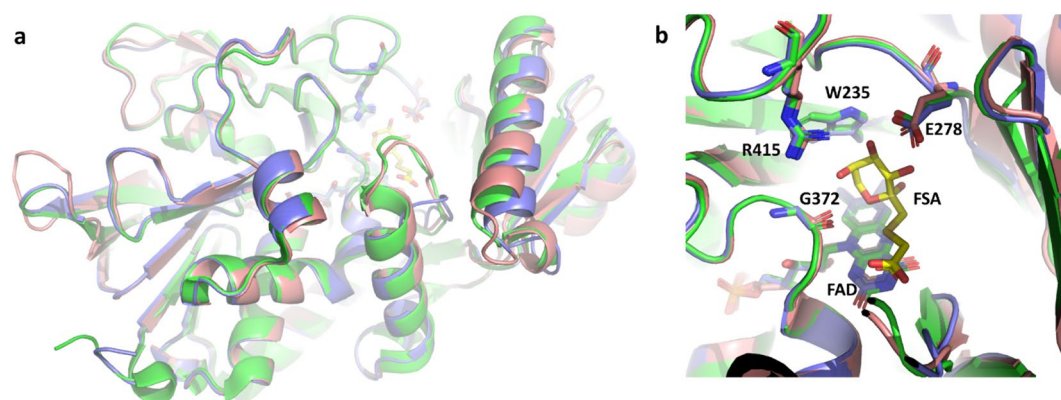
**Figure 3.** Schematic of the rational design strategy, showing the names of the relevant enzyme variants obtained at each stage.

Enzyme	Substrate	$K_m$ (mM)	$V_{max}$ (mM/min)	$k_{cat}$ ( $s^{-1}$ )	$k_{cat}/K_m$ ( $mM^{-1} s^{-1}$ )
PnFPOX	fV	$5.93 \pm 1.61$	$41.67 \pm 8.34$	$99.33 \pm 19.86$	16.75
X02C	fV	$0.94 \pm 0.06$	$10.56 \pm 0.56$	$20.83 \pm 1.10$	22.16
PnFPOX	F6P	$0.81 \pm 0.30$	$15.48 \pm 1.20$	$36.83 \pm 2.83$	45.47
X02C	F6P	$0.03 \pm 0.01$	$0.26 \pm 0.01$	$0.52 \pm 0.02$	17.22

**Table 3.** Kinetic parameters of wild-type PnFPOX and the engineered X02C enzymes.

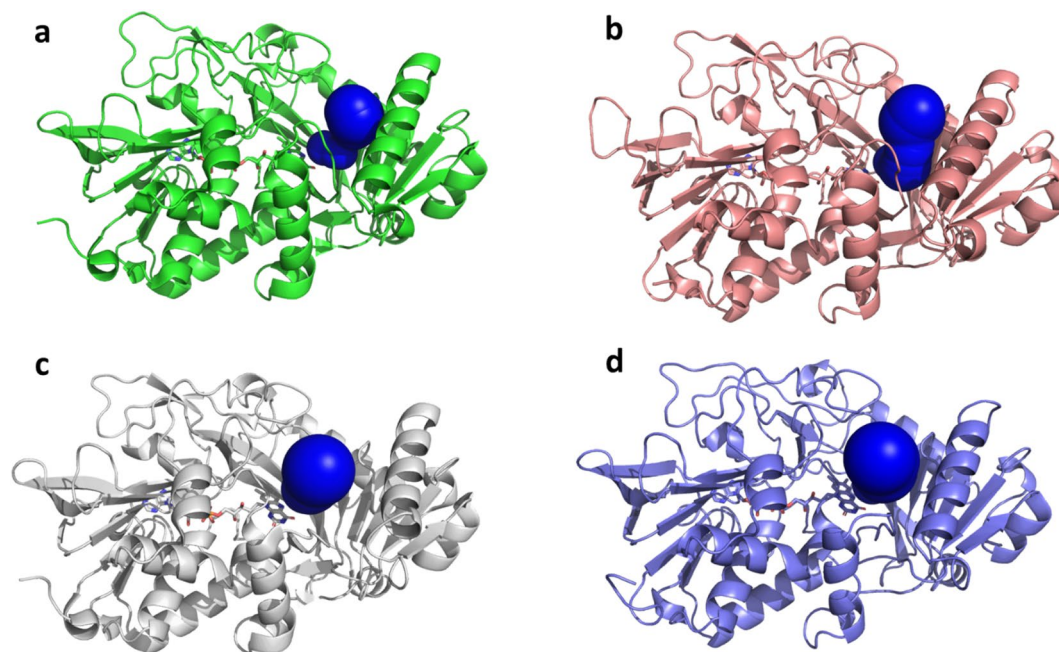
Enzyme	RMSD ( $\text{\AA}$ )	Tunnel bottleneck radius ( $\text{\AA}$ )	Tunnel length ( $\text{\AA}$ )
PnFPOX	–	2.2	13.3
L3-35A	1.8	3.7	6.4
D02	1.4	2.9	11.7
X02A	1.7	3.1	9.4
X04	1.8	3.0	10.9
X02B	1.5	3.0	10.2

**Table 4.** Calculated RMSDs between the WT enzyme PnFPOX and the different mutants, and tunnel geometry. For selected enzymes, it was not possible to obtain the crystal structure.



**Figure 4.** Structural comparison of relevant FPOX enzyme: PnFPOX (green), AnFPOX-47 (pink) and X02B (violet). The FSA inhibitor is depicted in yellow. The overall structural alignment shows no significant differences between the enzymes, except for the entrance tunnel (a). The position of the catalytic site is also highly conserved (b).

constraints of bound substrates due to the wider catalytic pocket, as well as to the removal of tunnel lining residues that contribute to the specific binding and stabilization of substrates. In an effort to revert the latter effect, we re-engineered the tunnel-lining residues in the X02B and X02C enzymes. The X02B enzyme features a wide and short tunnel (bottleneck radius of 3.0  $\text{\AA}$  and length of 10.2  $\text{\AA}$ ), in between the WT and its parent L3-35A, while at the same time showing a similar activity on both small and larger substrates. Hence, the wider tunnel of



**Figure 5.** Comparison of the binding tunnel of PnFPOX (a), AnFPOX-47 (b), L3-35A (c) and X02B (d).

X02B, while reducing the stabilization of small substrates (and, as a consequence, the catalytic activity towards them) allows larger substrates to access the catalytic site more easily, leading to comparable enzymatic activity.

## Conclusions

As part of an ongoing investigation aimed at engineering an FPOX enzyme to provide it with both enhanced thermostability and a wider access tunnel to the catalytic site, we set out to design a number of novel FPOX mutants starting from our previously characterized mutant L3-35A of PnFPOX, which featured a much larger active site relative to the wild type enzyme. To increase enzyme thermostability, we tested several strategies, namely introducing mutations to either improve the RMSF of the protein, increase the number of salt bridges or form single disulfide bonds. Our thermostability findings align with the expectation that the improved RMSF, salt bridges formation, and disulfide bonds formation approaches can all effectively improve the thermal stability of enzymes, albeit to a different extent. In particular, we have shown that the introduction of a disulfide bond allowed us to produce a PnFPOX variant that shows remarkably higher thermal stability compared to the other mutants. Furthermore, the catalytic activity, which was significantly lost during the design process, was reintroduced by engineering ligand-binding and tunnel-lining position, which led to a new, thermally stabilized and active FPOX enzyme that features a wider access tunnel to the catalytic pocket. Ongoing efforts are aimed at testing the enzymes on whole glycosylated proteins (i.e., albumin and hemoglobin) to test whether the wider access tunnel of X02C, together with the restored catalytic activity, can actually provide catalytic activity directly on whole proteins. This may represent a significant step toward the use of these enzymes on larger substrates and on full proteins, thus possibly allowing the development of *in vitro* HbA1c diagnostic devices based on a direct enzymatic oxidation system. Indeed, there is a rapidly growing demand for cheap, efficient and rapid diabetes monitoring tests. This could be met by developing enzymatic assays for glycosylated hemoglobin and albumin that do not require a preliminary digestion of the proteins. The elimination of this step would provide several advantages to the test, such as the shortening of its measuring times, a simplification of its overall procedure and a substantial reduction of its costs.

From the perspective of protein engineering, the use of rational design methods like the one described in this work offers a promising strategy for the design of thermostable enzymes with improved performance in various industrial and biomedical applications. This study also highlights the potential of computational methods and targeted mutations to improve the activity and substrate range of enzymes, and it paves the way for the design of thermostable enzymes with improved performance toward different targets and applications. We believe that this rational approach can accelerate the production and the use of biocatalysts in different industrial sectors, thus in turn allowing the improvement of chemical processes through the application of green chemistry concepts.

## Experimental procedures

### Molecular dynamics simulations

To identify the most suitable regions for structural modifications in the L3-35A enzyme, we ran 100 ns MD simulations following previously adopted protocols<sup>23,27,30</sup>. Specifically, we used the AMBER19SB force field<sup>31</sup> for protein, water (TIP3P) and ions, while we used the general amber force field (GAFF)<sup>32</sup> for modeling the FAD cofactor. We solvated the molecular model with a 15 Å pad of TIP3P water and we introduced counter ions to

neutralize the system charge, resulting in a final simulation box of  $\approx 50,000$  atoms. Hydrogen mass repartitioning was applied to allow a time step of 4 fs<sup>33</sup>. The system was subjected to 1000 energy minimization steps and it was equilibrated for 100 ps at a pressure of 1 atm and at a temperature of 300 K with NAMD software<sup>34</sup>, using a non-bonded cut-off of 12 Å, rigid bonds and particle-mesh Ewald long-range electrostatics. During the equilibration simulation, the Ca atoms of the protein were restrained by a 10 kcal mol<sup>-1</sup> Å<sup>-2</sup> spring constant. The 100 ns production runs were performed using a NVT ensemble whereby all the parameters (non-bonded cut-off, and PME) were the same as in the equilibration phase. Root mean square deviations (RMSD) were monitored after each MD run to assess structural convergence, while root mean square fluctuations (RMSF) were calculated to identify unstable regions for subsequent design. The tunnel dimension was analyzed with CAVER 3.0.3 software<sup>35</sup> using a probe radius of 1.4 Å, a shell depth of 5 Å, and a shell radius of 2 Å.

### Design and selection of stabilizing mutations

The identification of stabilizing mutations was performed using five different tools. The design focused on the most unstable region of the enzyme as identified in the previous step, with the exclusion of the catalytic amino acids, the amino acids interacting with the FAD cofactor, and the residues lining the tunnel. The first design was performed using the Rosetta Supercharge tool<sup>36</sup>. This tool was employed to reengineer the protein surface with amino acids that present a high net charge, which is reported to prevent aggregation of partially unfolded states. The second design was performed using the Rosetta Point Mutant “pmut” scan application<sup>37</sup>, which helped to identify stabilizing single point mutations. In the third approach, the *pmut* tool was used to identify stabilizing double point mutations, where only pair of residues that are close enough to interact with each other are tested (i.e., the distance of at least one of their atoms being closer than 4.5 Å). In the fourth approach, the PROSS web server<sup>38</sup> was used to generate 9 different enzyme designs with an incremental number of mutations. Disulfide bond engineering was performed by employing the disulfide-by-design web server<sup>39</sup>. For each of the 5 design strategies, the best 20 enzyme variants based on the internal scoring were retained (except for PROSS design, where all 9 variants were retained, and for disulfide design, where 8 variants were retained) for further analysis with MD simulations. Overall, library generation and ranking using the five different methods resulted in the selection of 77 different enzyme variants. Each model was simulated for 100 ns, as described for the reference L3-35A enzyme, and the best 6 variants in terms of RMSF stabilization<sup>27,30</sup> were selected for experimental validation.

### Protein expression and purification

The DNA sequences of the different FPOX variants were cloned in either pET15(b) or pET17(b) expression vectors (Novagen), to generate C-terminal 6His-tagged proteins. Reference 5T1E and L3\_35 DNA constructs, as well as the D02 and the C16 variants, were transformed and expressed in *E. coli* BL21 Star (DE3) cells (Invitrogen). Variants featuring cysteine mutations to allow disulfide bond formation (X02A, X04, X07, X02B, X02C) were transformed in SHuffle T7 *E. coli* cells (New England Biolabs). For each variant, 20 mL of an overnight culture was inoculated into 2L of Luria–Bertani (LB) broth supplemented with 100 mg/L ampicillin and grown at 37 °C until OD<sub>600</sub> = 0.6. Protein expression was induced using isopropyl 1-thio- $\beta$ -D-galactopyranoside (IPTG) at a final concentration of either 0.4 mM or 0.1 mM, followed by overnight culture at either 25 °C or 18 °C respectively, with shaking speed of 220 rpm. Cells were collected by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5 mM FAD, and 5% glycerol), supplemented with 0.2 mM protease phenylmethylsulfonyl fluoride (PMSF) and DNase and then lysed by sonication on ice. After centrifugation, the soluble fraction of the cell lysate was passed through a 0.45- $\mu$ m filter and then loaded onto a HisTrap HP 5 mL (GE Healthcare) column equilibrated with buffer A (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 5% glycerol, 20 mM imidazole). The Ni affinity column was washed using ten column volumes of buffer A, then the bound C-terminal His-tagged enzyme was eluted with the same buffer supplemented with 400 mM imidazole. The eluted fraction was loaded onto a HiPrep 26/60 Sephacryl S-100 size exclusion column (GE Healthcare) pre-equilibrated with buffer C (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 5% glycerol). The desired protein was concentrated using an Amicon 20 centrifugal filter with a molecular cutoff of 10 kDa to  $\approx 2$  mg/mL, and stored at  $-80$  °C. Protein concentration was assessed using a NanoDrop (Thermo Scientific). Sample purity was evaluated by 12% SDS-PAGE.

### Protein crystallization

Crystals of different mutants (D02, X02A, X04, X07, X02B) were grown at room temperature using the vapor diffusion method by mixing a 2  $\mu$ L drop of a 10–15 mg/mL protein sample with a 2  $\mu$ L drop of a solution containing 0.1 M MES monohydrate pH 6.5, 16–26% of different Polyethylene glycol (PEG) including 20 K (for D02, X02A and X02B) and 8 K (for X04). Crystals, which appeared after 1–4 days, were frozen using 25% (v/v) glycerol as cryoprotectant prior to X-ray diffraction data collection.

### Data collection and processing

For the FPOX D02 crystals, diffraction data were collected using a Dectris PILATUS3 6 M detector and a radiation of  $\lambda = 0.918$  Å on Bessy beamline 14.1 of the Deutsches Elektronen- Synchrotron (BESY II Light Source), Berlin, Germany. Reflection-image processing was performed using DIALS<sup>40</sup> and AIMLESS<sup>41</sup> from the CCP4 suite<sup>42</sup>. For FPOX X02A and X04, X-ray diffraction data were collected using radiation of  $\lambda = 1.0$  Å and a PSI PILATUS 6 M detector on the X06DA beamline at Swiss Light Source (SLS). Data was processed using autoproc. For X02B, diffraction data were collected using a DECTRIS EIGER2 XE 16 M detector and a radiation of  $\lambda = 0.9537$  Å on the I04 beamline at Diamond Light Source (Oxfordshire, United Kingdom). Data was processed using the autoproc package. Data collection and refinement statistics are shown in Table 5.



	D02	X02A	X02B	X04
PDB code	8BLZ	8BLX	8BJY	8BMU
Wavelength (Å)	0.918	1.000	0.9537	1.000
Space group	C 2 2 2 <sub>1</sub>	C 2 2 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 4 <sub>2</sub> 2 <sub>1</sub> 2
Unit cell (Å)	a = 91.10, b = 129.49, c = 86.78	a = 90.52, b = 130.26, c = 87.62	a = 63.11, b = 88.49, c = 90.94	a = 90.24, b = 90.24, c = 131.44
Resolution range (Å)	64.75–1.70 (1.73–1.70)	56.68–1.71 (1.74–1.71)	63.42–1.48 (1.50–1.48)	74.39–1.645 (1.673–1.645)
Total reflections	707,191 (38,534)	641,243 (29,830)	1,053,583 (48,251)	1,012,704 (25,719)
Unique reflections	55,830 (2946)	55,592 (2685)	82,805 (4255)	62,950 (2416)
Multiplicity	12.70 (13.10)	11.5 (11.10)	12.7 (11.3)	16.1 (10.6)
Completeness (%)	98.70 (99.30)	98.9 (96.5)	96.0 (100.0)	94.7 (74.1)
mean(I)/sig(I)	12.60 (2.20)	12.7 (1.0)	16.7 (0.4)	20.5 (1.1)
Wilson B-factor	12.08	21.4	27.04	11.12
Rmerge	0.17 (1.40)	0.152 (2.416)	0.067 (4.526)	0.106 (2.013)
Rmeas	0.17 (1.45)	0.159 (2.532)	0.070 (4.743)	0.110 (2.104)
Rpim	0.05 (0.39)	0.046 (0.743)	0.020 (1.406)	0.026 (0.589)
CC1/2 (%)	1.00 (0.75)	0.998 (0.461)	1.000 (0.419)	0.999 (0.458)
Reflections in refinement	55,772 (3879)	55,537 (5360)	80,561 (7099)	60,833 (939)
Reflections in free set	2729 (209)	2755 (264)	3995 (332)	3043 (44)
Rwork	0.177	0.1812	0.19	0.1766
Rfree	0.199	0.2103	0.215	0.2009
RMSD bonds (Å)	0.0144	0.011	0.012	0.012
RMSD angles (Å)	2.214	1.74	1.75	1.84
Ramachandran favoured (%)	97.2	95.74	96.71	95.52
Ramachandran allowed (%)	2.8	3.78	3.06	4.01
Ramachandran outliers (%)	0	0.47	0.24	0.47
Rotamer outliers (%)	1.1	3.68	0.85	2.53
Clash score	1.2	5.38	4.28	6.23
Overall number of atoms (non-H)	3745	3788	3758	3917
In macromolecules	3350	3354	3364	3376
In ligands	99	119	65	113
In solvent	296	315	329	428
Average B-factor (Å <sup>2</sup> )	21.7	30.07	30.39	22.03
For macromolecules	21	29.46	29.53	20.32
For ligands	23.9	37.65	32.17	34.43
For solvent	28.3	33.68	38.91	32.23

**Table 5.** Diffraction data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

### Structure determination and refinement

For all four structures, the initial phases were obtained by molecular replacement using Phaser<sup>43</sup> and the atomic coordinates of the L3-35A FPOX mutant (PDB code 6Y4)<sup>27</sup> as a search model. Refinement was performed by alternating rounds of REFMAC5<sup>44</sup> and manual adjustments in Coot<sup>42</sup>. Water molecules were added both manually and automatically using the Coot\_refine tool from the CCP4 cloud package<sup>45</sup>.

### Thermal shift assay

The melting temperature of each variant was measured using differential scanning fluorimetry (DSF) with Sypro-Orange dye (ThermoFisher Scientific) on an Applied Biosystems 7500 Real-Time PCR system (ThermoFisher Scientific). For each mutant, 12.5 µL of enzyme at the starting concentration of 10 µM was added to 12.5 µL of a 10× Sypro-Orange dye solution diluted from a 5000× stock in the same buffer that was used for the protein: 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 5% glycerol) to reach the final volume of 25 µL, a 5× final concentration of the dye and a 5 µM concentration of the enzyme. All the T<sub>m</sub> measurements were performed in 4 replicates. Fluorescence was monitored during the thermal denaturation occurring to the protein upon increasing the temperature from 15 to 95.3 °C.

### Circular dichroism

Circular dichroism (CD) measurements were performed on a Jasco J-1500 spectrophotometer at 20 °C. CD was performed to detect the secondary structure of different variants and to measure the thermostability of the enzyme. The spectra were the average of five scans, recorded using a 1.0 mm path length quartz cuvette on 5 µM samples of the different FPOX variants in a 50 mM Tris pH 8.0, 150 mM NaCl, 5% glycerol buffer

solution. Thermal denaturation curves were measured in 1.0 mm path length cuvettes closed with a parafilm on 5  $\mu\text{M}$  samples. The increasing temperature-induced denaturation at a rate of 1  $^{\circ}\text{C}/\text{min}$  from 5 to 95  $^{\circ}\text{C}$  and the ellipticity at 222 nm were recorded at 0.2  $^{\circ}\text{C}$  intervals using a 1-nm bandwidth and a response of 10 s. Using the Boltzmann function, the midpoints of the thermal-denaturation curves ( $T_m$ ) were determined by fitting the data to a sigmoidal transition curve. The secondary structure of different variants was measured at 200–250 nm wavelengths at 20  $^{\circ}\text{C}$ . Mean residue ellipticity  $[\theta]_{\text{MR}}$  was calculated and plotted versus wavelength. Each spectrum was determined as the average of three scans. The CD spectra of all the enzymes described here are shown in Figs. S1–S7 in the Supplementary Information.

### Enzymatic activity

The enzymatic activity of all variants was assessed at room temperature by measuring the amount of glucosone released from the substrate over time, as described previously<sup>27</sup>. The increase in absorbance at 322 nm (glucosone  $\epsilon_{322} = 149.25 \text{ M}^{-1} \text{ cm}^{-1}$ ) was monitored in an Infinite M1000 (Tecan) at 25  $^{\circ}\text{C}$ . The 200  $\mu\text{L}$  reaction mixture contained 20 mM Tris HCl pH 7.4, 20 mM o-phenylenediamine, 2 mM fructosyl-lysine, or fructosyl-valine or dipeptide (fructosyl-valine histidine) or hexapeptide (1-Deoxyfructosyl-Val-His-Leu-Thr-Pro-Glu) as a substrate. After 1 min of pre-incubation, the reaction was started by adding the enzyme at a final concentration of 0.04–1 mg/mL depending on the activity of the variant. One unit (U) is defined as the amount of enzyme required to produce 1  $\mu\text{mol}$  of glucosone per minute, and specific activity is expressed as U/mg of the enzyme. Enzyme kinetic constants ( $V_{\text{max}}$ ,  $K_m$ ,  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_m$ ) were determined using different substrate concentration in the assay (fructosyl-valine from 0.05 to 10 mM, fructosyl hexapeptide from 0.05 to 1 mM). Data points are the result of two independent experiments and the estimation of Micheaelis-Menten parameters was derived by Lineweaver-Burk plot fitting.

### Data availability

The final crystallographic coordinates of the crystal structures shown here are available in the RCSB PDB (accession codes: 8BJY, 8BLZ, 8BLX, 8BMU).

Received: 13 July 2023; Accepted: 19 October 2023

Published online: 30 October 2023

### References

- Miura, S., Ferri, S., Tsugawa, W., Kim, S. & Sode, K. Development of fructosyl amine oxidase specific to fructosyl valine by site-directed mutagenesis. *Protein Eng. Des. Sel.* **21**, 233–239 (2008).
- Kim, S., Miura, S., Ferri, S., Tsugawa, W. & Sode, K. Cumulative effect of amino acid substitution for the development of fructosyl valine-specific fructosyl amine oxidase. *Enzyme Microb. Technol.* **44**, 52–56 (2009).
- Miura, S., Ferri, S., Tsugawa, W., Kim, S. & Sode, K. Active site analysis of fructosyl amine oxidase using homology modeling and site-directed mutagenesis. *Biotechnol. Lett.* **28**, 1895–1900 (2006).
- Collard, F. *et al.* Crystal structure of the deglycating enzyme fructosamine oxidase (amadoriase II). *J. Biol. Chem.* **283**, 27007–27016 (2008).
- Hatada, M. *et al.* Development of a screen-printed carbon electrode based disposable enzyme sensor strip for the measurement of glycated albumin. *Biosens. Bioelectron.* **88**, 167–173 (2017).
- Ferri, S., Miyamoto, Y., Sakaguchi-Mikami, A., Tsugawa, W. & Sode, K. Engineering fructosyl peptide oxidase to improve activity toward the fructosyl hexapeptide standard for HbA1c measurement. *Mol. Biotechnol.* **54**, 939–943 (2013).
- Takahashi, M., Pischetsrieder, M. & Monnier, V. M. Isolation, purification, and characterization of amadoriase isoenzymes (fructosyl amine-oxygen oxidoreductase EC 1.5.3) from *Aspergillus* sp. *J. Biol. Chem.* **272**, 3437–3443 (1997).
- Ferri, S., Kim, S., Tsugawa, W. & Sode, K. Review of fructosyl amino acid oxidase engineering research: A glimpse into the future of hemoglobin A1c biosensing. *J. Diabetes Sci. Technol.* **3**, 585–592 (2009).
- Lin, Z. & Zheng, J. Occurrence, characteristics, and applications of fructosyl amine oxidases (amadoriases). *Appl. Microbiol. Biotechnol.* **86**, 1613–1619 (2010).
- Wu, X., Palfey, B. A., Mossine, V. V. & Monnier, V. M. Kinetic studies, mechanism, and substrate specificity of amadoriase I from *Aspergillus* sp. *Biochemistry* **40**, 12886–12895 (2001).
- Gan, W. *et al.* Structural basis of the substrate specificity of the FPOD/FAOD family revealed by fructosyl peptide oxidase from *Eupenicillium terrenum*. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **71**, 381–387 (2015).
- Rigoldi, F. *et al.* Molecular dynamics simulations provide insights into the substrate specificity of FAOX family members. *Mol. Biosyst.* **12**, 2622–2633 (2016).
- Monnier, V. M. & Wu, X. Enzymatic deglycation with amadoriase enzymes from *Aspergillus* sp. as a potential strategy against the complications of diabetes and aging. *Biochem. Soc. Trans.* **31**, 1349–1353 (2003).
- Paul, R. G. & Bailey, A. J. Glycation of collagen: The basis of its central role in the late complications of ageing and diabetes. *Int. J. Biochem. Cell Biol.* **28**, 1297–1310 (1996).
- Sell, D. R. & Monnier, V. M. Molecular basis of arterial stiffening: Role of glycation—A mini-review. *Gerontology* **58**, 227–237 (2012).
- Goh, S. Y. & Cooper, M. E. Clinical review: The role of advanced glycation end products in progression and complications of diabetes. *J. Clin. Endocrinol. Metab.* **93**, 1143–1152 (2008).
- Nagaraj, R. H., Linetsky, M. & Stitt, A. W. The pathogenic role of Maillard reaction in the aging eye. *Amino Acids* **42**, 1205–1220 (2012).
- Vincent, A. M., Russell, J. W., Low, P. & Feldman, E. L. Oxidative stress in the pathogenesis of diabetic neuropathy. *Endocr. Rev.* **25**, 612–628 (2004).
- Gautieri, A., Redaelli, A., Buehler, M. J. & Vesentini, S. Age- and diabetes-related nonenzymatic crosslinks in collagen fibrils: Candidate amino acids involved in advanced glycation end-products. *Matrix Biol.* **34**, 89–95 (2014).
- Snedeker, J. G. & Gautieri, A. The role of collagen crosslinks in ageing and diabetes—The good, the bad, and the ugly. *Muscles Ligaments Tendons J.* **4**, 303–308 (2014).
- Capuano, E. *et al.* Studies on the effect of Amadoriase from *Aspergillus fumigatus* on peptide and protein glycation in vitro. *J. Agric. Food Chem.* **55**, 4189–4195 (2007).
- Qian, Y., Zheng, J. & Lin, Z. Loop engineering of amadoriase II and mutational cooperativity. *Appl. Microbiol. Biotechnol.* **97**, 8599–8607 (2013).

23. Rigoldi, F. *et al.* Crystal structure of the deglycating enzyme Amadoriase I in its free form and substrate-bound complex. *Proteins Struct. Funct. Bioinform.* **84**, 744–758 (2016).
24. Ogawa, N. *et al.* Creation of haemoglobin A1c direct oxidase from fructosyl peptide oxidase by combined structure-based site specific mutagenesis and random mutagenesis. *Sci. Rep.* **9**, 942 (2019).
25. Kokkonen, P., Bednar, D., Pinto, G., Prokop, Z. & Damborsky, J. Engineering enzyme access tunnels. *Biotechnol. Adv.* **37**, 107386 (2019).
26. Gautieri, A., Rigoldi, F., Torretta, A., Redaelli, A. & Parisini, E. In silico engineering of enzyme access tunnels. *Methods Mol. Biol.* **2397**, 203–225 (2022).
27. Rigoldi, F. *et al.* Rational backbone redesign of a fructosyl peptide oxidase to widen its active site access tunnel. *Biotechnol. Bioeng.* <https://doi.org/10.1002/bit.27535> (2020).
28. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. & Meiler, J. Practically useful: What the Rosetta protein modeling suite can do for you. *Biochemistry* **49**, 2987–2998 (2010).
29. Kim, S., Ferri, S., Tsugawa, W., Mori, K. & Sode, K. Motif-based search for a novel fructosyl peptide oxidase from genome databases. *Biotechnol. Bioeng.* **106**, 358–366 (2010).
30. Rigoldi, F. *et al.* Thermal stabilization of the deglycating enzyme Amadoriase I by rational design. *Sci. Rep.* **8**, 3042 (2018).
31. Tian, C. *et al.* Ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **16**, 528–552 (2020).
32. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
33. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass reparameterization. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).
34. Humphrey, W. V. M. D. Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
35. Chovancova, E. *et al.* CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **8**, e1002708 (2012).
36. Der, B. S. *et al.* Alternative computational protocols for supercharging protein surfaces for reversible unfolding and retention of stability. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0064363> (2013).
37. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science (80-)* **302**, 1364–1368 (2003).
38. Goldenzweig, A. *et al.* Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell* **63**, 337–346 (2016).
39. Craig, D. B. & Dombkowski, A. A. Disulfide by design 2.0: A web-based tool for disulfide engineering in proteins. *BMC Bioinform.* <https://doi.org/10.1186/1471-2105-14-346> (2013).
40. Winter, G. *et al.* DIALS: Implementation and evaluation of a new integration package. *Acta Crystallogr. D Struct. Biol.* **74**, 85–97 (2018).
41. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution?. *Acta Crystallogr. D Struct. Biol. Crystallogr.* **69**, 1204–1214 (2013).
42. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).
43. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* <https://doi.org/10.1107/S0021889807021206> (2007).
44. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
45. Kovalevskiy, O., Lebedev, A. & Krissinel, E. Helping researchers to solve their structures: automation and user guidance in CCP4 Cloud. *Acta Crystallogr. Sect. A Found. Adv.* **77**, C766–C766 (2021).

## Author contributions

Conceptualization: E.P., A.G.; methodology: H.E., J.A.R.B., R.C., E.P., A.G.; formal analysis: H.E., S.B., J.A.R.B., L.L., G.C., A.R.; investigation: H.E., S.B., J.A.R.B., L.L., G.C., A.R.; resources: R.C., E.P., A.G.; data curation: H.E., J.A.R.B., R.C., E.P., A.G.; writing, review and editing: H.E., S.B., J.A.R.B., R.C., L.L., G.C., A.R., E.P., A.G.; supervision: H.E., R.C., E.P., A.G.; funding acquisition: E.P., A.G. All authors have read and agreed to the published version of the manuscript.

## Funding

The authors wish to thank the personnel of the Swiss Light Source (Proposal Number 20220753), the Diamond Light Source (proposal number mx32544) and the BESSY II Light Source (proposal number MX-202-00211-ST) for help with the data collection. E.P. thanks the ERDF project BioDrug (No. 1.1.1.5/19/A/004) and the Latvian Council of Science (Grant No. lzp-2020/2-0013) for financial support. We acknowledge the CINECA award under the ISCRA initiative (Grant code IsB26\_W2EB and IsCa2\_REZYME) for the availability of high-performance computing resources and support.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-45428-1>.

**Correspondence** and requests for materials should be addressed to E.P. or A.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023