BBA - Proteins and Proteomics 1868 (2020) 140293
https://doi.org/10.1016/j.bbapap.2019.140293

Sequential oxidation of 5-hydroxymethylfurfural to furan-2,5-dicarboxylic acid by an evolved aryl-alcohol oxidase

Javier Viña-Gonzalez[1], Angel T. Martinez[2], Victor Guallar[3,4] and Miguel Alcalde[1*]

[1]Department of Biocatalysis, Institute of Catalysis, CSIC, Cantoblanco, 28049 Madrid, Spain.

[2]Biological Research Center, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain.

[3] Barcelona Supercomputing Center, Jordi Girona 31, 08034 Barcelona, Spain.

[4] ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain.

*Corresponding author: malcalde@icp.csic.es

Keywords: aryl-alcohol oxidase, 5-hydroxymethylfurfural, furan-2,5-dicarboxylic acid, directed evolution.

ABSTRACT

Furan-2,5-dicarboxylic acid (FDCA) is a building block of biodegradable plastics that can be used to replace those derived from fossil carbon sources. In recent years, much interest has focused on the synthesis of FDCA from the bio-based 5-hydroxymethylfurfural (HMF) through a cascade of enzyme reactions. Aryl-alcohol oxidase (AAO) and 5-hydroxymethylfurfural oxidase (HMFO) are glucose-methanol-choline flavoenzymes that may be used to produce FDCA from HMF through three sequential oxidations, and without the assistance of auxiliary enzymes. Such a challenging process is dependent on the degree of hydration of the original aldehyde groups and of those formed, the rate-limiting step lying in the final oxidation of the intermediate 5-formyl-furancarboxylic acid (FFCA) to FDCA. While HMFO accepts FFCA as a final substrate in the HMF reaction pathway, AAO is virtually incapable of oxidizing it. Here, we have engineered AAO to perform the stepwise oxidation of HMF to FDCA through its structural alignment with HMFO and directed evolution. With a 3-fold enhanced catalytic efficiency for HMF and a 6-fold improvement in overall conversion, this evolved AAO is a promising point of departure for further engineering aimed at generating an efficient biocatalyst to synthesize FDCA from HMF.

1. INTRODUCTION

5-Hydroxymethylfurfural (HMF) is a renewable chemical that can be obtained from biomass-derived carbohydrates. Considered as a high value building block, HMF is conventionally produced from the acid-catalyzed dehydration of fructose, ideally from sucrose and inulin hydrolysates [1]. Three consecutive oxidations of HMF produce furan-2,5-dicarboxylic acid (FDCA), one of the most promising value added chemical substrates obtained from biomass for use in the polymer industry [2]. Indeed, biobased molecules are currently in much demand in order to achieve a more sustainable chemical industry, in which the FDCA family of polymers provides of aliphatic and aromatic polyesters and polyamides. Among the most appealing applications of FDCA involves its co-polymerization with diols to synthesize PEFs (poly(ethylene-2,5-furandicarboxylates)), highlighting the possibility of establishing renewable alternatives to the traditional generation of polyesters based on petrochemicals like terephthalic acid, such as poly(ethylene-terephthalates) -PETs. Leaving aside other furan derivatives like 5-(butoxymethyl)furfural and 5-(methoxymethyl)-2-furfural, HMF is the main source for the production of FDCA monomers, which can be synthesized using catalytic nanoparticles like Pt, Pt/Bi or Pd in aqueous HMF solution with oxygen, from a isopropanol dehydration system using an hydrotalcite supported Au catalyst, from HMF derivatives in acetic acid media with a Co/Mn/Br catalytic system, or via enzymatic processes driven by lipases or, as seen recently, dehydrogenases and oxidases [3]. Thus, the preferred environmental-friendly transformation of HMF into FDCA involves enzymatic oxidation by isolated enzymes or whole microorganisms.

Flavooxidase transformation of HMF has hitherto been studied for two members of the glucose-methanol-choline (GMC) oxidoreductase superfamily that share a N-terminal FAD-binding domain: the aryl-alcohol oxidase from *Pleurotus eryngii* (AAO, EC 1.1.3.7) and the 5-hydroxymethylfurfural oxidase from *Methylovorus sp.* strain MP688 (HMFO, EC 1.1.3.47) [4][5][6]. The general catalytic cycle of both enzymes involves dehydrogenative oxidations carried out as two-half reactions, a reductive reaction in which the reducing substrate is 2-

electron oxidized by FAD and an oxidative reaction where H_2O_2 is produced after the reduction of O_2 by the FAD, which in turn returns to its oxidized resting state. In the FDCA synthetic pathway, up to three consecutive 2-electron oxidations can take place from HMF (6 electrons in total), generating two furanic intermediates, DFF (2,5-diformylfuran) and FFCA (5-formyl-furancarboxilyc acid), **Figure 1**. While the physiological roles of AAO and HMFO are distinct, their reaction mechanisms and substrate specificities, ranging from primary benzylic alcohols to conjugated aliphatic systems, mean they are functionally considered as true alcohol oxidases. Accordingly, aldehydes `*sensum stricto*´ are not substrates of these flavoenzymes but only their hydrated *gem*-diols [6][7][8][9]. This is particularly relevant for the FDCA route, in which the poor spontaneous hydration of carbonyl groups diminishes or even abolishes enzyme reactivity, with *gem*-diols forming 29% and 1.8% for DFF and FFCA, respectively [5]. As such, the rate-limiting step of the pathway resides in the oxidation of FFCA, with few *gem*-diols present, indicating the need for this last oxidation to re-accommodate the bulkier substrate to an adequate distance from the FAD co-factor [4]. Together, this makes the final oxidation step from FFCA to FDCA inefficient, generating barely detectable traces of FDCA in the case of AAO and/or weak final conversion yields in the case of native HMFO. Such catalytic bottlenecks in the performance of AAO have so far been overcome by adding fungal peroxygenases to achieve the final conversion of FFCA to FDCA [4] or more recently, by adding catalase to the reaction mixture to avoid a potential inhibitory effect by H_2O_2 [10]. Interestingly, pioneering work has revealed several HMFO variants which show direct improved activity on FFCA [5][11][12]. Indeed, mutations in HMFO that affect the enantioselective oxidation of secondary alcohols notably, confer a striking 1,000-fold improvement in the catalytic efficiency on FFCA. As both secondary alcohols and *gem*-diols have bulky substituents on the α-carbon relative to natural primary alcohols, the introduction of less constraining amino acids helps to relax the catalytic cavity and increase the activity on FFCA [11]. Similarly, we recently performed a directed evolution campaign on AAO for the selective oxidation of secondary benzyl alcohols. Paradoxically, rather

than broadening the active pocket for a more suited binding as in the HMFO variants, our bulkier mutations allowed the substrate to be situated at a better catalytic distance [13]. Given the structural similarities and mechanism of action between AAO and HMFO, it seems reasonable to think that mutations responsible for unlocking the oxidation of secondary alcohols by AAO could be a suitable departure point to enhance FFCA oxidation, as described for HMFO [11].

To date, HMFO is the only GMC oxidase known to perform the three oxidations required to transform HMF into FDCA. Here we show that an evolved AAO can also complete this challenging synthetic route. Structural alignment of HMFO and AAO enabled mutant libraries to be constructed and screened for the oxidation of HMF and FFCA. Subsequently, the best variant for this process was characterized biochemically and the role of the mutations assigned within the FDCA route by computational analysis at the atomic level.

2. RESULTS AND DISCUSSION

Before starting the directed evolution campaign, we benchmarked a set of AAO mutants from our previous evolutionary campaigns for the heterologous expression and oxidation of secondary alcohols [13][14][15]. The final two candidates for the FDCA route were the FX9 and 15G12 variants. FX9 is a stably expressed variant that carries the H91N-L170M mutations. H91N is a back-to-ancestor/consensus mutation at the si-face of the FAD that stabilizes the co-factor, whereas L170M is situated in a superficial α-helix, far from the active site [15]. The 15G12 mutant harbors the two mutations from FX9 and the I500M-F501W substitutions at the catalytic cavity that are involved in unlocking activity on secondary aromatic alcohols [13], **Figure 2A, B**. In addition, the secretion of both variants by yeast is further promoted by the addition of an evolved chimeric prepropeptide (preαfactor-proKiller) that is cleaved upon exocytosis [14]. The supernatants from microtiter yeast fermentations of the FX9 and 15G12 variants (*i.e.* the high-throughput screening -HTS- format needed for AAO evolution) were assessed for the oxidation of HMF and FFCA using the HRP-ABTS/Amplex red assay (see Methods for details). While neither

of the two mutants produced a reliable response against FFCA in this HTS format, the activity of FX9 on HMF was roughly 33-fold higher (0.3 U/L) than that of 15G12 (0.009 U/L). Since the bulky I500M-F501W substitutions at the catalytic site of 15G12 dramatically affect the activity against HMF, we considered it logical to study those positions in a search for positive epistatic effects. As explained in the introduction, our reasoning was related to the previous design of HMFO mutants for secondary-alcohol oxidation that showed improved activity on FFCA [11]. Accordingly, we constructed a combinatorial saturation mutagenesis (CSM) library for positions 500 and 501 that was probed in a dual HTS assay with HMF and FFCA. No active variants against FFCA were found and ~95% of the clones in the mutagenic landscape were inactive on HMF, indicating that positions 500 and 501 are highly sensitive to modification, **Figure 2C**. Nevertheless, a few clones did show a 2 to 3-fold increase in activity for HMF relative to the FX9 parental type and interestingly, all of them harbored the single F501W mutation (referred to henceforth as the Bantha variant) that was already present in 15G12, while position 500 had reverted to the original Ile500, **Figure 2D**.

The unambiguous relationship between the substrate scopes of AAO and HMFO, along with a similar effect of the mutations involved in secondary alcohol oxidation, particularly regarding activity on HMF, led us to explore other structural determinants of catalysis that could be transferred between these enzymes. Although the sequence identity between AAO and HMFO is rather low (~30%), structural alignment of the catalytic sites revealed other potentially relevant positions for the substrate-enzyme interaction, **Figure 3**. Using the Bantha variant as the template, a new CSM library was constructed that targeted positions Tyr92 and His465. The Tyr92 residue is involved in substrate diffusion and positioning at the active site, whereas His465 contributes to a correct catalytic location of the alcohol in front of the FAD co-factor. These residues correspond to Met103 and Asn511 in HMFO, respectively. Unfortunately, despite these efforts no improved variants were identified in this library.

In the light of these results, the Bantha mutant and the parental FX9 were produced, purified to homogeneity and characterized biochemically, measuring steady state kinetic constants with HMF, **Table 1.** The K_m values were hardly affected by the F501W substitution and therefore, the enhanced catalytic efficiency was exclusively dependent on the 3-fold improvement in the k_{cat}. When comparing the Bantha variant with the wildtype AAO (wtAAO, heterologously expressed in *E. coli* after *in vitro* refolding), the k_{cat} improved 26-fold and the K_m increased ~8-fold, with a resulting 3-fold enhancement in the k_{cat}/K_m. The higher K_m values found in the *S. cerevisiae* variants may reflect the high degree of glycosylation (~50%) of the evolved mutants [15].

Although some promiscuous activity has been detected with FFCA as the substrate [16], the oxidation of HMF into FDCA to complete the whole cascade reaction is yet to be reported for AAO in the absence of ancillary enzymes [4, 10]. The weak equilibrium between the aldehyde group of FFCA and its scarce *gem*-diol, along with the need to re-position the furanic ring of FFCA in the active pocket for the third oxidation step, are thought to explain this lack of activity [4]. However, inhibition of the reaction by the H_2O_2 formed in the HMF and DFF oxidations may also influence this phenomenon [10]. We analyzed the possible production of FDCA from FFCA in a 20 h reaction by HPLC, observing a conspicuous 7-fold improvement in this conversion by Bantha relative to wtAAO, and to a lesser extent for FX9 (roughly half the activity of the Bantha variant), **Table 2.** We then measured the production of FDCA from HMF, recording a ~6-fold improvement in this conversion for the Bantha mutant with low but detectable yields of 3 %. Although the weak conversion of HMF to FDCA by wtAAO (0.5%) was previously attributed to the oxidative power of the H_2O_2 released from consecutive oxidations from HMF through to FFCA [4], we cannot rule out that it actually comes from the latent activity of wtAAO on FFCA that has been now amplified by directed evolution.

It is worth noting that no reaction optimization was performed in our study (*i.e.* supplying FAD and/or catalase to avoid possible enzyme inactivation, increasing substrate

and/or enzyme loads, testing different reaction pHs or temperatures), yet our results unequivocally show that the evolved AAO (without any optimization of the reaction) converts HMF into FDCA via three sequential oxidation steps, using the different reaction intermediates as new substrates to complete the whole route (including FFCA). In a parallel recent work, several AAO mutants and wtAAO were benchmarked to complete the route from HMF to FDCA in the presence of catalase [10]. Such variants were produced after *in vitro* refolding in *E. coli* and when compared with the evolved AAO expressed in yeast, they behaved much less stable (roughly 15°C T_{50} than evolved AAO) [15]. Indeed, the high stability of evolved AAO variants, including the Bantha mutant disclosed in the current work, is determined by the combination of three factors: the glycosylation effect, the presence of the H91N consensus-ancestor mutation, and the natural folding of yeast, all of them lacking in the *E. coli* AAO counterparts [14, 15]. As such, it is hard to find a rational comparison between AAO variants from both expression systems; we can only speculate that insertion of the F501W (see computational analysis below) along with the strong stability of the evolved AAO paved the way for the direct transformation from HMF to FFCA in the absence of catalase.

The F501W and H91N mutations in the Bantha variant correspond to the same residues in native HMFO (Trp466 and Asn102 according to the numbering of HMFO), indicating a strong structure-function relationship between both enzymes. While the H91N ancestor-consensus mutation extends the promiscuous activity on FFCA and stabilized the protein backbone further, the F501W substitution enhanced the HMF activity, **Figure 4**. To rationalize the effect of these substitutions on the unlocked activity for FFCA, we turned to computational analysis by protein energy landscape exploration (PELE), a MonteCarlo based approach to map enzyme-ligand interactions. PELE results on the migration of HMF an FFCA in both AAO and Bantha models are shown in **Figure 5**. Clearly, for wtAAO we only observe the proper positioning of HMF in the active site [17], red profile in the left pannel; the *gem*-diol FFCA cannot reach catalytic positions and is stack into the entrance pathway, blue profile in the left pannel. For the Bantha variant,

however, both substrates diffuse correctly into catalytic positions, finding good proton transfer distances, in the 2-3 Å range, right pannel in **Figure 5**. We further expanded the simulation of wtAAO with FFCA for another 50 epochs but did not find differences. Only when adding bias towards the FAD cofactor (using the epsilon option in AdaptivePELE [18]), we finally found some entrance into the active site. Thus, it seems like Bantha enhances the entrance of FFCA into the active site, which is significantly impeded in wtAAO. **Figure 6** shows a representative structure where FFCA is retained at the entrance pathway in wtAAO, corresponding to an energy of -35 kcal/mol and a distance of ~11Å in the left pannel of **Figure 5**, and compares it with a snapshot from Bantha where the ligand initiates its transition towards the active site. Main differences reside in the shift of Phe397. Importantly, next to the substrate we find Trp501 (Phe501 in wtAAO), which seems to help in pushing Phe397, one of the identified gatekeepers of the active site [19]. Thus, a steric hindrance mechanism might slow down drastically the diffusion of FFCA into the active site of wtAAO. Inclusion of a larger bulky residue, Trp501, seems to slightly open the pathway. While it might look like counterintuitive that inclusion of a larger bulky residue opens the pathway (instead of further blocking it), similar observations have been recently reported in a tyrosinase [20].

3. CONCLUSIONS

Significant efforts are being made to fully develop the biotechnological potential of GMC oxidases. In the case of AAO and HMFO, the production of FDCA is hampered by the chemoselectivity of the alcohol/aldehyde oxidation reactions. Here, the oxidation of FFCA to FDCA by AAO was targeted by laboratory evolution. As a result, the AAO variant Bantha, in which a Trp residue was introduced next to the catalytic base and that carries the H91N consensus-ancestor mutation, displayed enhanced overall activity on the furfural derivatives when converting HMF to FDCA, for the first time showing the ability of this flavoenzyme to perform the entire three consecutive oxidation cascade. Yet preliminary, this study widens the repertoire

of potential biocatalysts for the FDCA cascade, providing a suitable departure point to design an efficient AAO for emergent biobased industries.

4. MATERIAL AND METHODS

4.1 Materials

All chemicals were reagent-grade purity. 5-Hydroxymethylfurfural was obtained from AVABIOCHEM (Zug, Switzerland), 5-formyl-furancarboxilyc acid and furan-2,5-dicarboxylic acid were obtained from TCI chemicals (Zwijndrecht, Belgium). ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)), Horseradish peroxidase (HRP), the Yeast Transformation Kit was purchased from Sigma (Madrid, Spain). Amplex® Red reagent (10-acetyl-3,7-dihydroxyphenoxazine) was obtained from Biogen (Madrid, Spain). Zymoprep Yeast Plasmid Miniprep, Yeast Plasmid Miniprep Kit I and Zymoclean Gel DNA Recovery Kit were from Zymo Research (Orange, CA). Restriction enzymes *Bam*HI, *Xho*I, were from New England Biolabs (Hertfordshire, UK). I-Proof high fidelity DNA polymerase was from Biorad (USA). The episomal shuttle vector pJRoC30 was from the California Institute of Technology (CALTECH, USA). *E. coli* XL2-Blue competent cells were from Stratagene (La Jolla, CA, US. The protease deficient S. *cerevisiae* strain BJ5465 was from LGC Promochem (Barcelona, Spain). Primers were acquired from Isogen Life Science (Barcelona, Spain) and are included in Table S1.

4.2 Culture media

Minimal medium contained 0.67% (w:v) yeast nitrogen base, 1.92 g/L yeast synthetic drop-out medium supplement without uracil, 2% (w:v) raffinose and 25 µg/mL chloramphenicol. SC drop-out plates contained 0.67% (w:v) yeast nitrogen base, 1.92 g/L yeast synthetic drop-out medium supplement without uracil, 2% (w:v) bacto agar, 2% (w:v) D-glucose and 25 µg/mL chloramphenicol. YP medium contained 10 g yeast extract, 20 g peptone and double-distilled H_2O (ddH_2O) to 650 mL. AAO expression medium contained 144 mL YP 1.55x, 13.4 mL 1M KH_2PO_4 pH 6.0 buffer, 22.2 mL 20% galactose (w:v), 0.222 mL 25 µg/mL chloramphenicol and ddH_2O to

200 mL. AAO selective expression medium (SEM) contained 100 mL 6.7% (w:v) sterile yeast nitrogen base, 100 mL 19.2 g/L sterile yeast synthetic drop-out medium supplement without uracil, 100 mL sterile 20% galactose (w:v), 100 mL 1 M KH_2PO_4 pH 6.0, 600 mL sddH_2O and 1 mL 25 g/L chloramphenicol.

4.3 Laboratory evolution of AAO

4.3.1 First generation: Combinatorial saturation mutagenesis at positions Ile500 and Phe501:

The FX9 variant (261 bp) was used as DNA template for saturation mutagenesis technique at position 500 and 501 at the active site of the AAO. Two PCR reactions were carried out in a final volume of 50 µL containing 3% DMSO, 0.8 mM dNTPs (0.2 mM each), 0.03 U/µL iproof DNA polimerase, and 0.2 ng/µL FX9 template and different primers according to the 22-trick protocol [21]. PCR1 contained 0.25 µM RMLN and 0.25 µM mix of reverse primers: 22c1R, 22c2R, 22c3R, 22c4R, 22c5R, 22c6R, 22c7R, 22c8R and 22c9R. PCR2 contained 0.25 µM RMLC and 0.25 µM mix of forward primers: 22c1F, 22c2F, 22c3F, 22c4F, 22c5F, 22c6F, 22c7F, 22c8F and 22c9F. Amplification reactions were carried out in a thermal cycler MycyclerTM (BIO-RAD, USA) with the following PCR program: 98°C for 30 seconds (1 cycle); 98°C for 10 seconds, 50°C for 25 seconds and 72°C for 60 seconds (28 cycles); and 72°C for ten minutes (1 cycle). After purification, PCR products (400 ng each) were mixed with the linearized pJRoC30 (100 ng; ratio PCR product: vector = 4:1) and transformed in yeast for *in vivo* cloning. According to the 22-trick protocol, a library of 1700 individual colonies was screened as described below.

4.3.2 Second generation: Combinatorial saturation mutagenesis at positions Tyr92 and His546:

Three PCR reactions were carried out in a final volume of 50 µL containing 3% DMSO, 0.8 mM dNTPs (0.2 mM each), 0.03 U/µL iproof DNA polimerase, and 0.2 ng/µL Bantha template and different primers according to the 22-trick protocol. PCR1 contained 0.25 µM RMLN and 0.25 µM mix of reverse primers: 92RAHN, 92RCDB and 92RCCA. PCR2 contained 0.25 µM mix of forward primers 92FNDT, 92FVHG and 92FTGG 0.25 µM mix of reverse primers: 546RAHN,

546RCDB and 546RCCA. PCR2 contained 0.25 µM mix of forward primers 546FNDT, 546FVHG and 546FTGG and reverse primer RMLC. Amplification reactions were carried out in a thermal cycler MycyclerTM (BIO-RAD, USA) with the following PCR program: 98°C for 30 seconds (1 cycle); 98°C for 10 seconds, 50°C for 25 seconds and 72°C for 60 seconds (28 cycles); and 72°C for ten minutes (1 cycle). After purification, PCR products (400 ng each) were mixed with the linearized pJRoC30 (100 ng; ratio PCR product: vector = 4:1) and transformed in yeast for *in vivo* cloning. According to the 22-trick protocol, a library of more than 1700 individual colonies was screened as described below.

4.3.3 High-throughput screening (HTS) assay: Transformed cells were plated in SC drop-out plates and incubated for 3 days at 30°C, individual clones were fermented in sterile 96-well plates containing 200 µL of SEM medium. Plates were sealed and incubated at 30°C, 225 rpm and 80% relative humidity in a humidity shaker (Minitron-INFORS, Biogen, Spain) for 72 hours. Aliquots of 20 µL of yeast supernatants were transferred to a 96-well plate using a robotic station for liquid handling Freedom EVO (Tecan, Männedorf, Switzerland) and 180 µL of HRP-ABTS reagent for 5-hydroxymetylfurfural activity detection or 180 µL of HRP-Amplex red reagent for 5-formyl-furancarboxilyc acid activity detection. The final concentrations for the HRP-ABTS mixture in the well were 5 mM 5-hydroxymethylfurfural, 2.5 mM ABTS, 3 µg/mL HRP in 100 mM phosphate buffer pH 6.0. The final concentrations for the HRP-Amplex red mix in the well were 5 mM 5-formyl-furancarboxilyc acid, 70 µM Amplex Red, 3 µg/mL HRP in 100 mM phosphate buffer pH 6.0. Reagents were dispensed with Multidrop™ Combi Reagent Dispenser (Thermo Scientific, Massachusetts, USA). The plates were incubated at room temperature and activity with the furfural derivative was determined as H_2O_2 production coupled to the oxidation of reagent by the HRP and measured at 418 (ε_{ABTS}. 418 = 36000 M^{-1} cm^{-1}) for ABTS or at 563nm ($\varepsilon_{resorufin}$ 563 = 56000 M^{-1} cm^{-1}) for Amplex Red. One unit of AAO activity is defined as the amount of enzyme that converts 1 µmol of substrate with the stoichiometric formation of H_2O_2 per min

under the reaction conditions. The HTS-assay incorporated two consecutive re-screenings to rule out the selection of false positives as described in previous work [10].

4.4 Protein production and purification of AAO

The FX9 and Bantha variants expressed in *S. cerevisiae* were produced and purified to homogeneity as described in a former work [14]. The native AAO, heterologously expressed in *E. coli* and in vitro refolded ($_{wt}$AAO), was produced and purified as described elsewhere [22].

4.5 Biochemical characterization of AAO

4.5.1 Steady-state kinetic constants: 5-hydroxymethylfurfural oxidation kinetics were measured in 100 mM phosphate buffer pH 6.0. The oxidation 5-hydroxymethylfurfural was follow indirectly coupled with saturated conditions of HRP and Amplex Red substrate (4.5 U/mL HRP and 75 µg/mL Amplex Red final concentrations) following activity at 563nm (ε_{563} = 56000 M^{-1} cm^{-1}).

4.5.2 HPLC analysis: FDCA production for the different AAOs was analysed by HPLC with equipment consisting of a tertiary pump (Varian/Agilent Technologies, CA, USA) coupled to an autosampler (Merck Millipore, MA, USA) and a Ultrabase C8 column (Análisis Vínicos, Ciudad Real, Spain) . For the mobile phase 12 mM phosphate buffer pH 7 (A) and acetonitrile (B) at a flow rate of 1.2 min-1 were used and detection was done at 268 nm. After 1 min with 100% A, B was increased for 3.5 min to 5% and then to 40% in 2.5 min. After maintaining 40% B for 0.5 min, A was returned to 100% in 0.5 min and maintained for 2 min. FFCA and FDCA had a retention time of 2.1 and 1 min respectively. Calibration curve was prepared with 0.05, 0.1, 0.2 and 0.4 mM of FDCA.

4.5.3 Protein modelling: A structural model of the AAO from *P. eryngii* crystal structure at a resolution of 2.55 Å (Protein Data Bank Europe [PDB] accession number 3FIM [23]) was used as scaffold for the wild type AAO model and the homology models for different mutants were made from 3FIM by PyMol (Schrodinger LLC.; http://www.pymol.org). A structural model of the HMFO

from *Methylovorus sp*. (strain MP688) crystal structure at a resolution of 1.9 Å (Protein Data Bank Europe [PDB] accession number 4UDP [11]) was used as scaffold for the wild type HMFO model.

4.5.4 DNA sequencing: All genes were verified by DNA sequencing (BigDye Terminator v3.1 Cycle Sequencing Kit) using the following primers: primers sense, RMLN and AAOsec1F and primers antisense RMLC and AAOsec2R.

4.6 Modelling Methods

The crystal structure with PDB code 3FIM was used to model the wild type AAO and the Bantha mutant. Protein structures were prepared accordingly using Schrodinger's Protein Preparation Wizard. Mutations in Bantha were first introduced manually, followed by a side chain prediction of the three mutations, H91N, L170M and F501W, and of those amino acids within 4 Å with Prime [24]. In both models, histidines 91, 313, 387 and 398 were ε-protonated, 190, 502 and 546 δ-protonated and the remaining histidines were doubly protonated. Acidic residues were deprotonated except Glu389 [17]. HMF and gem-diol FFCA were optimized with Jaguar [25], with density functional M06 and 6-31G** basis set, using the Poisson Boltzmann Finite (PBF) implicit solvent in order to obtain their electrostatic potential derived atomic charges used in PELE simulations. The FAD cofactor was parametrized with the 2005 OPLSAA.

Adaptive-PELE simulations [18], and enhanced sampling version of PELE, were performed using 128 computing cores, 50 epochs and 8 Monte Carlo (MC) steps per epoch. The ligands were placed just outside the entrance pathway [17] and allowed to explore the pathway and the catalytic site. PELE is an all-atom modelling software based on a MC procedure coupled to protein structure prediction algorithms. Its performance in mapping ligand migration and (binding site) induced fit processes has been widely demonstrated [26].

ACKNOWLEDGEMENTS

This research was supported by the EU project H2020-BBI-PPP-2015-2-720297-ENZOX2, by the Spanish Government projects BIO2016-79106-R-Lignolution, and by the Comunidad de Madrid project Y2018/BIO4738-EVOCHIMERA.

REFERENCES

[1] Moreau, C., Belgacem, M. N., & Gandini, A. (2004). Recent catalytic advances in the chemistry of substituted furans from carbohydrates and in the ensuing polymers. *Top. Catal.*, 27(1-4), 11-30.

[2] Werpy, T., & Petersen, G. (2004). Top value added chemicals from biomass: volume I--results of screening for potential candidates from sugars and synthesis gas.

[3] Sousa, A. F., Vilela, C., Fonseca, A. C., Matos, M., Freire, C. S., Gruter, G.-J. M., . . . Silvestre, A. J. (2015). Biobased polyesters and other polymers from 2,5-furandicarboxylic acid: a tribute to furan excellency. *Polym. Chem.*, 6(33), 5961-5983.

[4] Carro, J., Ferreira, P., Rodríguez, L., Prieto, A., Serrano, A., Balcells, B., . . . Ullrich, R. (2015). 5-hydroxymethylfurfural conversion by fungal aryl-alcohol oxidase and unspecific peroxygenase. *FEBS J*, 282(16), 3218-3229.

[5] Dijkman, W. P., & Fraaije, M. W. (2014). Discovery and characterization of a 5-hydroxymethylfurfural oxidase from *Methylovorus* sp. strain MP688. *Appl. Environ. Microbiol.*, 80(3), 1082-1090.

[6] Dijkman, W. P., Groothuis, D. E., & Fraaije, M. W. (2014). Enzyme-catalyzed oxidation of 5-hydroxymethylfurfural to furan-2, 5-dicarboxylic acid. *Angew. Chem. Int.*, 53(25), 6515-6518.

[7] Ferreira, P., Hernández-Ortega, A., Herguedas, B., Rencoret, J., Gutiérrez, A., Martínez, M. J., . . . Martínez, Á. T. (2010). Kinetic and chemical characterization of aldehyde oxidation by fungal aryl-alcohol oxidase. *Biochem. J.*, 425(3), 585-593.

[8] Van Hellemond, E. W., Vermote, L., Koolen, W., Sonke, T., Zandvoort, E., Heuts, D. P., . . . Fraaije, M. W. (2009). Exploring the biocatalytic scope of alditol oxidase from *Streptomyces coelicolor*. *Adv. Synth. Catal.,* 351(10), 1523-1530.

[9] Li, Y.-S., Ho, J.-Y., Huang, C.-C., Lyu, S.-Y., Lee, C.-Y., Huang, Y.-T., . . . Li, T.-L. (2007). A Unique Flavin Mononucleotide-Linked Primary Alcohol Oxidase for Glycopeptide A40926 Maturation. *J. Am. Chem. Soc.*, 129(44), 13384-13385.

[10] Serrano, A., E. Calviño, J. Carro, M. I. Sánchez-Ruiz, F. J. Cañada and A. T. Martínez (2019). Complete oxidation of hydroxymethylfurfural to furandicarboxylic acid by aryl-alcohol oxidase. *Biotechnol. Biofuels,* 12(1): 217.

[11] Dijkman, W. P., Binda, C., Fraaije, M. W., & Mattevi, A. (2015). Structure-based enzyme tailoring of 5-hydroxymethylfurfural oxidase. *ACS catal.*, 5(3), 1833-1839.

[12] Martin, C., Ovalle Maqueo, A., Wijma, H. J., & Fraaije, M. W. (2018). Creating a more robust 5-hydroxymethylfurfural oxidase by combining computational predictions with a novel effective library design. *Biotechnol. Biofuels*, 11, 56-56.

[13] Viña-Gonzalez, J., Jimenez-Lalana, D., Sancho, F., Serrano, A., Martinez, A. T., Guallar, V., & Alcalde, M. (2019). Structure-Guided Evolution of Aryl Alcohol Oxidase from *Pleurotus eryngii* for the Selective Oxidation of Secondary Benzyl Alcohols. *Adv. Synth. Catal.,* 361(11), 2514-2525.

[14] Viña-Gonzalez, J., Gonzalez-Perez, D., Ferreira, P., Martinez, A. T., & Alcalde, M. (2015). Focused Directed Evolution of Aryl-Alcohol Oxidase in *Saccharomyces cerevisiae* by Using Chimeric Signal Peptides. *Appl. Environ. Microbiol.*, 81(18), 6451-6462.

[15] Viña-Gonzalez, J., Elbl, K., Ponte, X., Valero, F., & Alcalde, M. (2018). Functional expression of aryl-alcohol oxidase in *Saccharomyces cerevisiae* and *Pichia pastoris* by directed evolution. *Biotechnol. Bioeng.*, 115(7), 1666-1674.

[16] Karich, A., Kleeberg, S. B., Ullrich, R., & Hofrichter, M. (2018). Enzymatic preparation of 2, 5-furandicarboxylic acid (FDCA)—a substitute of terephthalic acid—by the joined action of three fungal enzymes. *Microorganisms*, 6(1), 5.

[17] Hernández-Ortega, A., Borrelli, K., Ferreira, P., Medina, M., Martínez, A. T., & Guallar, V. (2011). Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J*, 436(2), 341-350.

[18] Lecina, D., Gilabert, J. F., & Guallar, V. (2017). Adaptive simulations, towards interactive protein-ligand modeling . *Sci Rep*, 7(1), 8466.

[19] Carro, J., Amengual-Rigo, P., Sancho, F., Medina, M., Guallar, V., Ferreira, P., & Martínez, A. T. (2018). Multiple implications of an active site phenylalanine in the catalysis of aryl-alcohol oxidase . *Sci Rep*, 8(1), 8121.

[20] Deri, B., Kanteev, M., Goldfeder, M., Lecina, D., Guallar, V., Adir, N., & Fishman, A. (2016). The unravelling of the complex pattern of tyrosinase inhibition. *Sci Rep*, 6(1), 34993.

[21] Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z.-G., Opperman, D. J., Reetz, M. T., & Acevedo, J. P. (2012). Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.*, 2(2), 83-92.

[22] Ruiz-Dueñas, F. J., Ferreira, P., Martínez, M. J., & Martínez, A. T. (2006). In vitro activation, purification, and characterization of *Escherichia coli* expressed aryl-alcohol oxidase, a unique H_2O_2-producing enzyme. *Protein Expr. Purif.*, 45(1), 191-199.

[23] Fernandez I.S., Ruiz-Dueñas F.J., Santillana E., Ferreira P., Martinez M.J., Martinez A.T., & Romero A. (2009). Novel structural features in the GMC family of oxidoreductases revealed by the crystal structure of fungal aryl-alcohol oxidase. *Acta Crystallogr. D* 65, 1196–1205.

[24] Schrödinger Release 2019-1: Prime, Schrödinger, LLC, New York, NY, 2019.

[25] Schrödinger Release 2019-1: Jaguar, Schrödinger, LLC, New York, NY, 2019.

[26] Gilabert, J. F., Lecina, D., Estrada, J. and Guallar, V. (2019). Monte Carlo Techniques for Drug Design: The Success Case of PELE. In Biomolecular Simulations in Structure-Based Drug Discovery (eds F. L. Gervasio and V. Spiwok).

FIGURE LEGENDS:

Figure 1. AAO/FDCA route via alcohol/gem-diol oxidation.

Figure 2. Mutations in the evolved AAO variants. After several cycles of evolution for functional expression, the FX9 variant (**A**) maintained the same residues as the wildtype AAO (wtAAO) at the re-side of FAD in the catalytic pocket, plus the H91N mutation (in pink) and substitution L170M. The 15G12 mutant (**B**) acquired the two substitutions for secondary alcohol oxidation presenting mutations H91N-L170M-I500M-F501W and weaker activity on HMF (inset). The Bantha variant (**D**), with H91N-L170M-F501W substitutions, had increased activity on HMF (inset). (**C**) Mutational landscape of the combinatorial library. Clone activity is plotted in descending order, with the horizontal line showing the activity of the parental type on HMF. The insets in A, B, D show the activity in the HTS format on HMF.

Figure 3. Structural alignment of the catalytic pockets of AAO and HMFO. Residues from AAO (PDB accession number 3FIM) are depicted in dark blue, while the residues in HMFO (PDB accession number 4UDP) are depicted in light blue.

Figure 4. (A) Structural details of the wtAAO, FX9 and Bantha mutants. (B) Sequence alignment of the Bantha variant with HMFO, highlighting the mutations acquired in AAO corresponding to native residues in HMFO.

Figure 5. PELE's interaction energy along the catalytic proton transfer coordinate to His502 for HMF (in red) and FFCA (in blue) for the wtAAO (left pannel) and Bantha (right pannel) models. Energies units are in kcal/mol and distances in Å.

Figure 6. Two representative snapshot along the binding of FFCA in wtAAO (grey protein and substrate) and Bantha (blue protein and substrate). FFCA and FAD shown in ball and stick representation. Residues 501 and 397 shown in stick representation.

Table 1. Kinetic parameters for the Bantha and FX9 variants.

	Kinetic constant	Bantha	FX9	wtAAO*
HMF	K_m (mM)	12.4 ± 0.8	13 ± 2	1.6 ± 0.2
	k_{cat} (s^{-1})	18.8 ± 0.5	6.8 ± 0.3	0.67 ± 0.001
	k_{cat}/K_m (mM^{-1} s^{-1})	1.51	0.52	0.42

AAO kinetic constants were measured at 25 °C in 100 mM phosphate buffer (pH 6.0). All the reactions were performed in triplicate: * From [4]. Values considered the formation of one equivalent of H_2O_2 for the oxidation of HMF into DFF (originally [4] had a correction considering two H_2O_2 equivalents for two consecutive oxidations of HMF into DFF and into FFCA).

Table 2. Production of FDCA by wtAAO and the Bantha mutant.

Substrate	[Mm]	Enzyme	[μM]	Time (h)	Conversion (%)	FDCA yield (%)
HMF	2	Bantha	3.3	48	100	3
HMF	2	wtAAO	3.3	48	100	0.5
FFCA	2.5	Bantha	3.3	20	7	7
FFCA	2.5	wtAAO	3.3	20	1	1

Reactions were performed at room temperature in phosphate buffer (100 mM, pH 8.0) and analyzed by HPLC-PDA.

Supplementary Table 1. List of primers

Oligo	Sequence
RMLN	5’-CCTCTATACTTTAACGTCAAGG-3’
RMLC	5’-GGGAGGGCGTGAATGTAAGC-3’
22c1F	5’-GAGACAACGCCAACACG*NDTNDT*CACCCAGTTGGAAC-3’
22c1R	5’-GTTCCAACTGGGTG*AHNAHN*CGTGTTGGCGTTGTCTC-3’
22c2F	5’-GAGACAACGCCAACACG*NDTVHG*CACCCAGTTGGAAC-3’
22c2R	5’-GTTCCAACTGGGTG*CDBAHN*CGTGTTGGCGTTGTCTC-3’
22c3F	5’-GAGACAACGCCAACACG*NDTTGG*CACCCAGTTGGAAC-3’
22c3R	5’-GTTCCAACTGGGTG*CCAAHN*CGTGTTGGCGTTGTCTC-3’
22c4F	5’-GAGACAACGCCAACACG*VHGNDT*CACCCAGTTGGAAC-3’
22c4R	5’-GTTCCAACTGGGTG*AHNCDB*CGTGTTGGCGTTGTCTC-3’
22c5F	5’-GAGACAACGCCAACACG*VHGVHG*CACCCAGTTGGAAC-3’
22c5R	5’-GTTCCAACTGGGTG*CDBCDBC*GTGTTGGCGTTGTCTC-3’
22c6F	5’-GAGACAACGCCAACACG*VHGTGG*CACCCAGTTGGAAC-3’
22c6R	5’-GTTCCAACTGGGTG*CCACDB*CGTGTTGGCGTTGTCTC-3’
22c7F	5’-GAGACAACGCCAACACG*TGGNDT*CACCCAGTTGGAAC-3’
22c7R	5’-GTTCCAACTGGGTG*AHNCCA*CGTGTTGGCGTTGTCTC-3’
22c8F	5’-GAGACAACGCCAACACG*TGGVHG*CACCCAGTTGGAAC-3’
22c8R	5’-GTTCCAACTGGGTG*CDBCCA*CGTGTTGGCGTTGTCTC-3’
22c9F	5’-GAGACAACGCCAACACG*TGGTGG*CACCCAGTTGGAAC-3’
22c9R	5’-GTTCCAACTGGGTG*CCACCA*CGTGTTGGCGTTGTCTC-3’
92FNDT	5’-GGTCTAGCTCTGTTAAC*NDT*ATGGTCATGATGCGTG-3’
92FVHG	5’-GGTCTAGCTCTGTTAAC*VHG*ATGGTCATGATGCGTG-3’
92FTGG	5’-GGTCTAGCTCTGTTAAC*TGG*ATGGTCATGATGCGTG-3’
92RAHN	5’-CACGCATCATGACCAT*AHN*GTTAACAGAGCTAGACC-3’
92RCDB	5’-CACGCATCATGACCAT*CDB*GTTAACAGAGCTAGACC-3’
92RCCA	5’-CACGCATCATGACCAT*CCA*GTTAACAGAGCTAGACC-3’
546FNDT	5’-CCTTCGCGCCCAACGCA*NDT*ACCCAAGGACCGATATA-3’
546FVHG	5’-CCTTCGCGCCCAACGCA*VHG*ACCCAAGGACCGATATA-3’
546FTGG	5’-CCTTCGCGCCCAACGCA*TGG*ACCCAAGGACCGATATA-3’
546RAHN	5’-TATATCGGTCCTTGGGT*AHN*TGCGTTGGGCGCGAAGG-3’

546RCDB	5’-TATATCGGTCCTTGGGT*CDB*TGCGTTGGGCGCGAAGG-3’
546RCCA	5’-TATATCGGTCCTTGGGT*CCA*TGCGTTGGGCGCGAAGG-3’
AAOsec1F	5’-GTGGATCAACAGAAGATTTCGATCG-3’
AAOsec2R	5’-GTGGTTAGCAATGAGCGCGG-3’

Codon substitutions are shown in italics (where N = A/T/C/G; D = no C; V =no T, H = no G; B = no A; R = A/G; Y = C,T).

Figure1

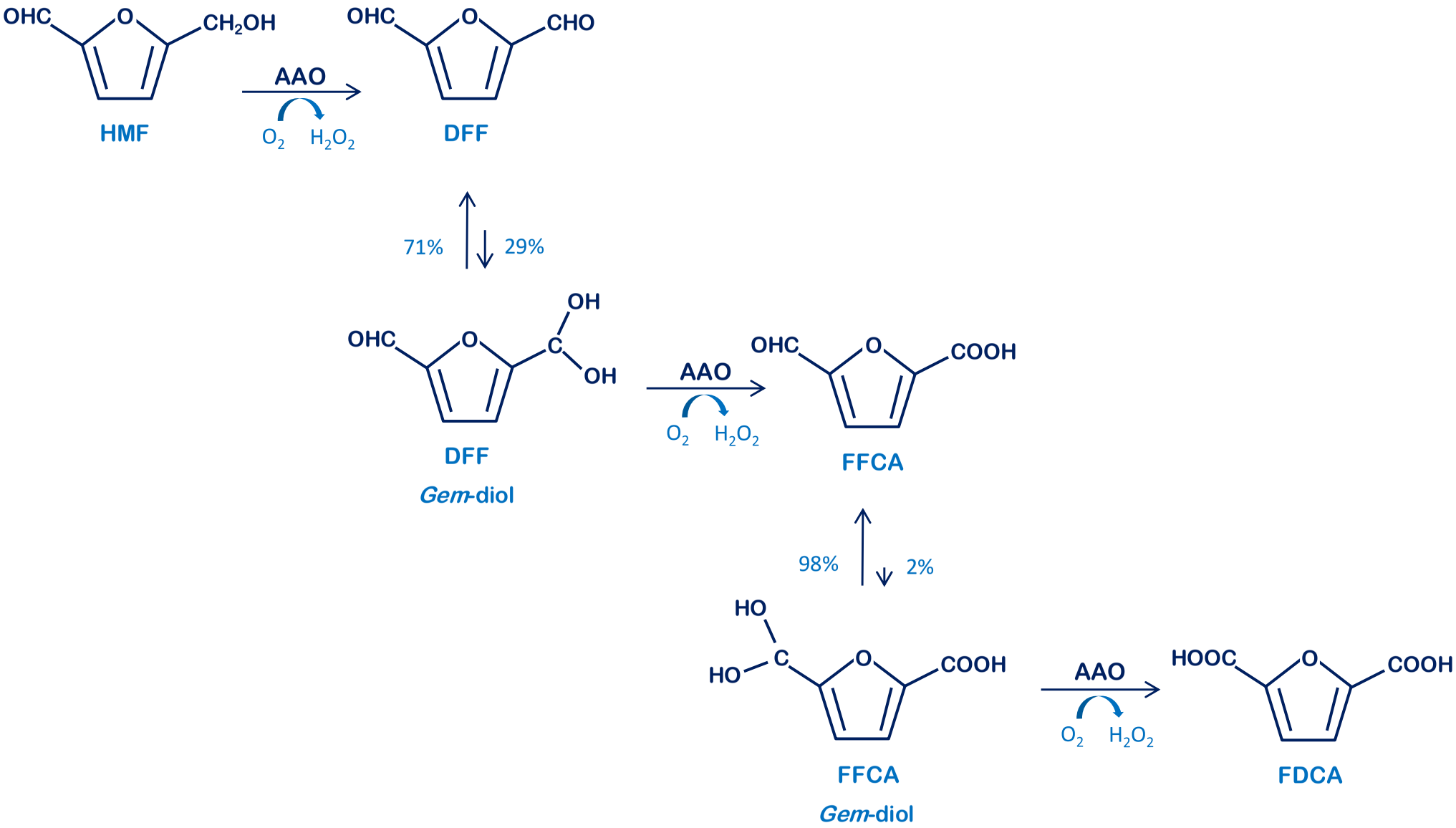
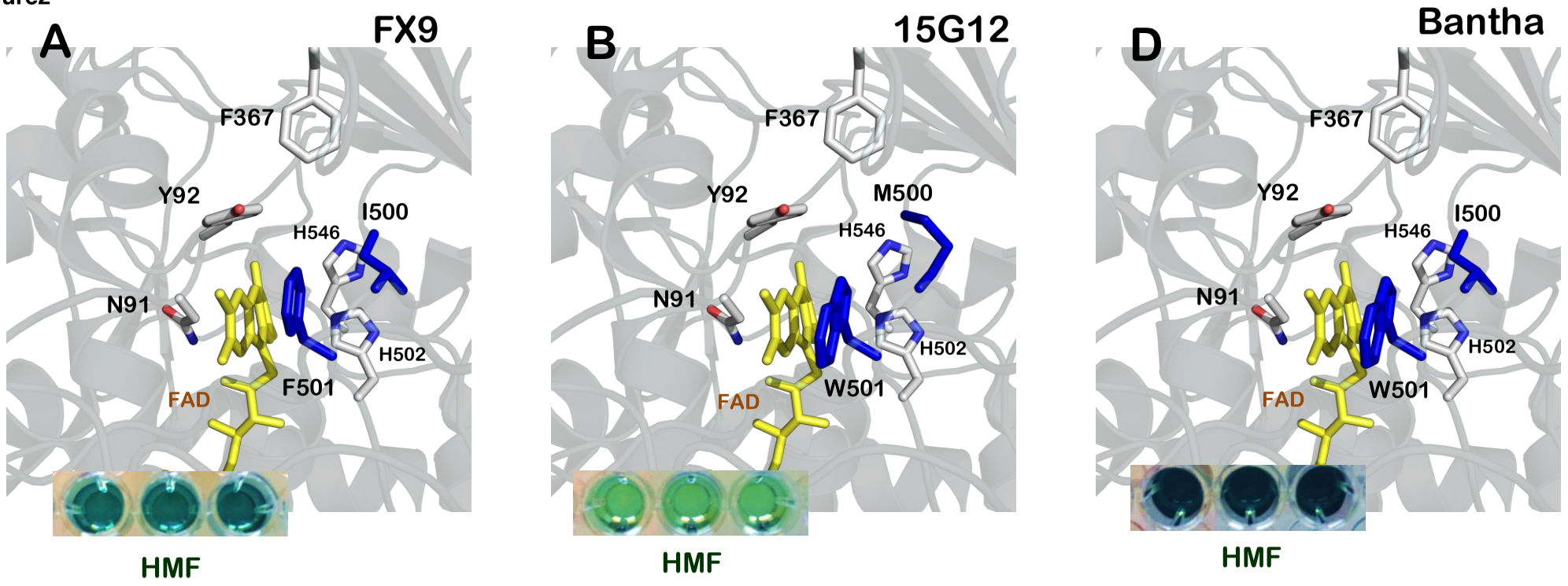

OHC
O
CH2OH
HMF
AAO
O2 H2O2
OHC
O
CHO
DFF
71%
29%
OHC
O
OH
C
OH
DFF
Gem-diol
AAO
O2 H2O2
OHC
O
COOH
FFCA
98%
2%
HO
HO
C
O
COOH
FFCA
Gem-diol
AAO
O2 H2O2
HOOC
O
COOH
FDCA

Figure2


A
FX9
F367
Y92
I500
H546
N91
H502
FAD
F501
HMF
B
15G12
F367
Y92
M500
H546
N91
H502
FAD
W501
HMF
D
Bantha
F367
Y92
I500
H546
N91
H502
FAD
W501
HMF

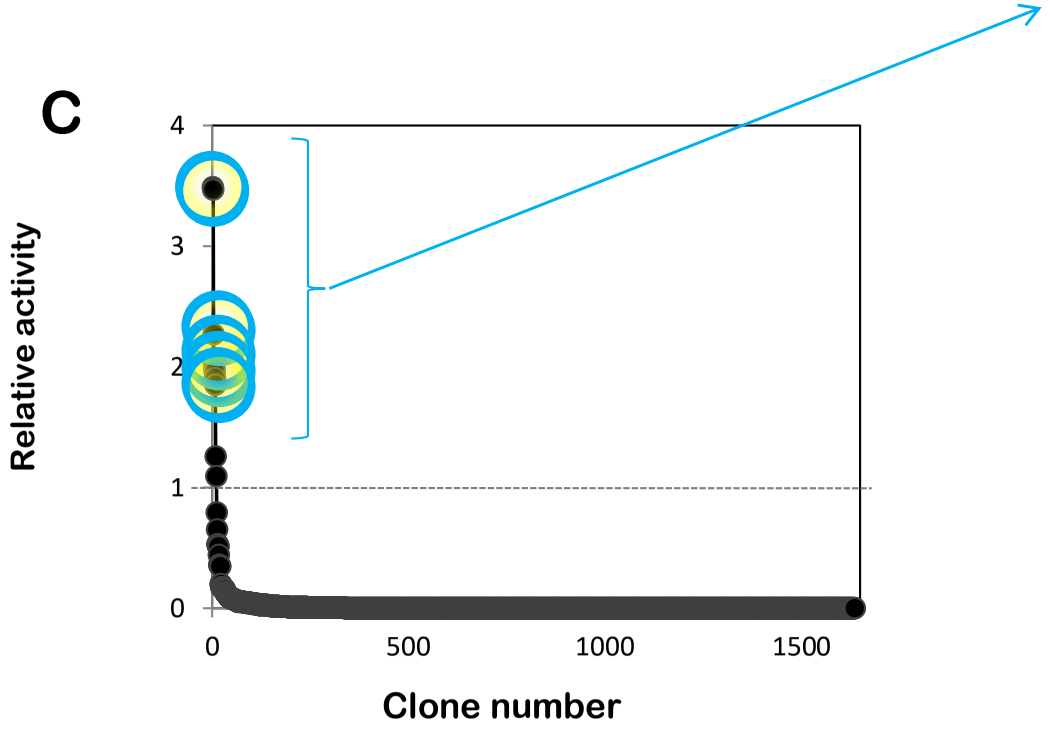

C
Relative activity
4
3
2
1
0
0
500
1000
1500
Clone number

Figure3

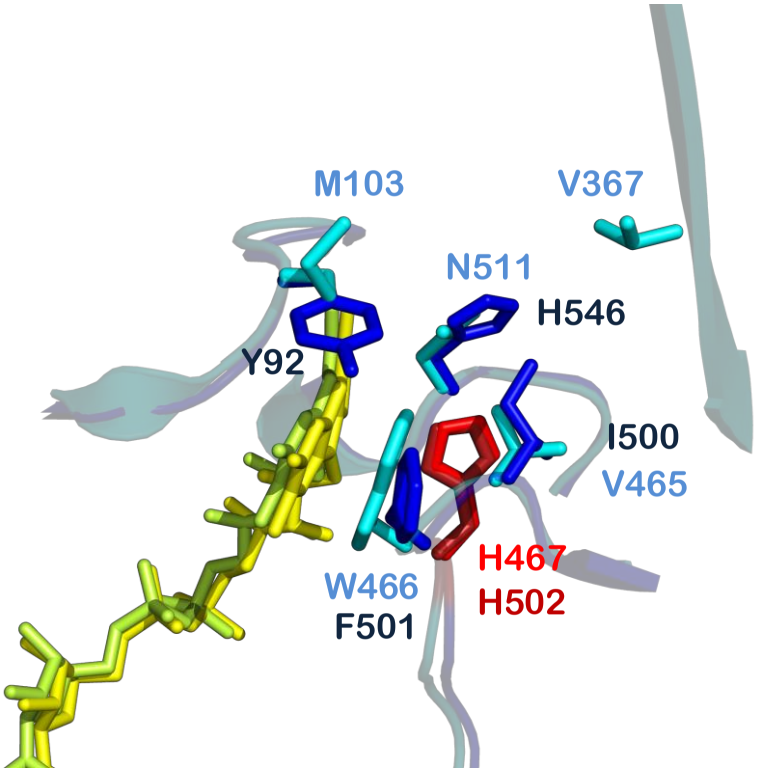
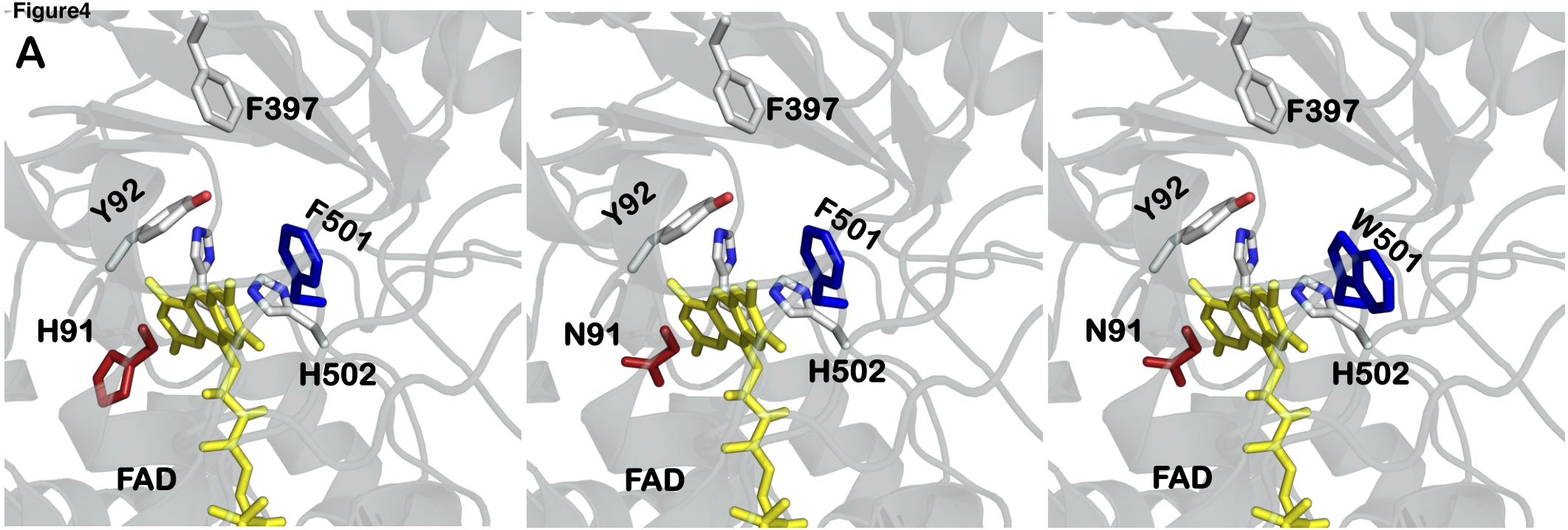


Figure4



B

Protein sequence alignment

Bantha	$_{89}$SVNYMV$_{94}$	$_{499}$TIWHPV$_{504}$
HMFO	$_{100}$SVNMVV$_{105}$	$_{464}$GVWHAS$_{469}$

Figure5

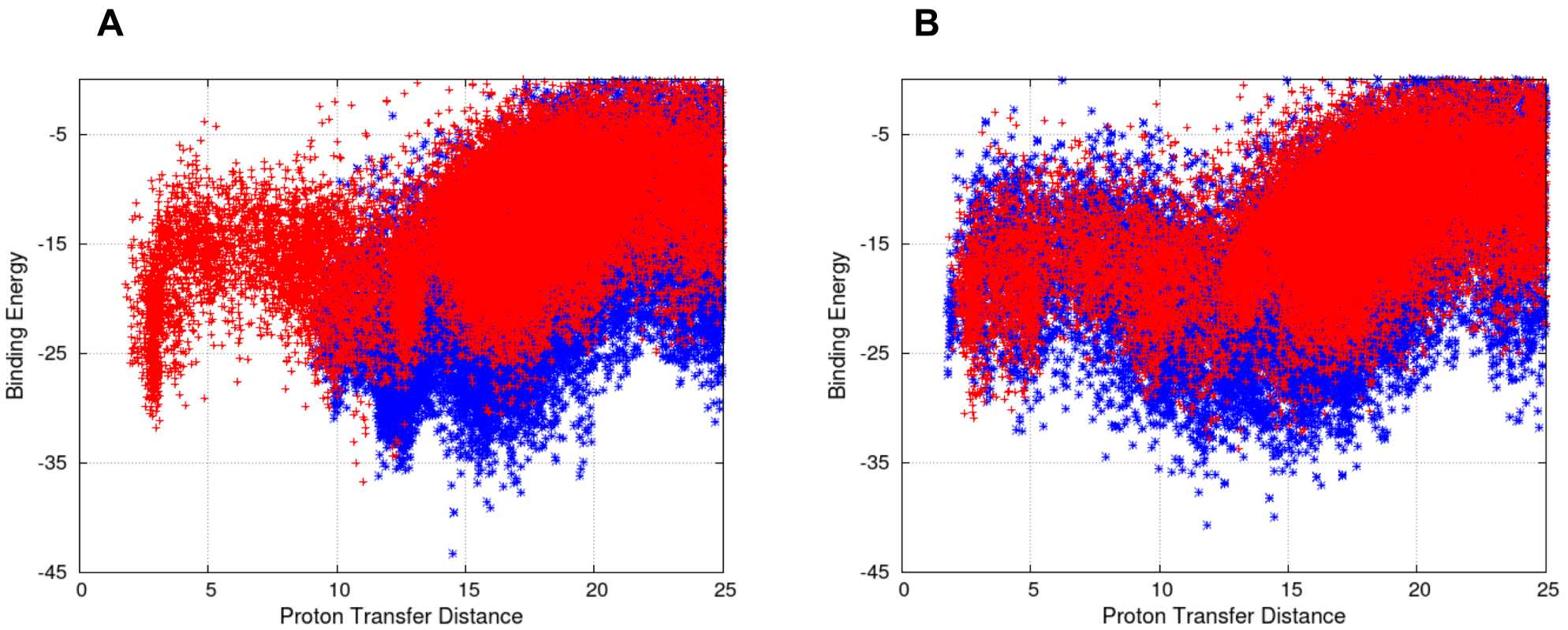

A
B
Binding Energy
Proton Transfer Distance
-5
-15
-25
-35
-45
0
5
10
15
20
25

Figure6