

## Human Carboxylesterase 2 in Cocaine Metabolism

Pedro R. Figueiredo<sup>[a,b]</sup>, Ricardo D. González<sup>[a,b]</sup> and Alexandra T. P. Carvalho<sup>[a,c\*]</sup>

<sup>[a]</sup> CNC – Center for Neuroscience and Cell Biology, Institute for Interdisciplinary Research (IIIUC), University of Coimbra, Coimbra, Portugal

<sup>[b]</sup> PhD Programme in Experimental Biology and Biomedicine, Institute for Interdisciplinary Research (IIIUC), University of Coimbra, Casa Costa Alemão, Coimbra, Portugal

<sup>[c]</sup> Almac Sciences, Department of Biocatalysis and Isotope Chemistry, Almac House, 20 Seagoe Industrial Estate, Craigavon, BT63 5QD, Northern Ireland, UK

\*atpcarvalho@uc.pt

### Abstract

Increased hydrolysis of cocaine to non-toxic compounds is a promising way to prevent cocaine-induced toxicity. However, the short half-life of cocaine in the blood and the rapid conversion in the body to the hydrolysis-resistant metabolite benzoylecgonine, limits the therapeutic potential of serum proteins. Therefore, hydrolysis by tissue-specific hydrolases that do not generate benzoylecgonine deserves further investigation.

Here, we report for the first time, the mechanism of cocaine hydrolysis by the human carboxylesterase 2. We have combined conventional and accelerated Molecular Dynamics, which allowed us to identify the structural motions of the  $\alpha 1$  and  $\alpha 10'$  helices that act as a putative lid. Quantum Mechanics/Molecular Mechanics calculations on the full cycle showed that the rate-limiting step is the formation of benzoic acid (deacylation step) with an  $\Delta G$  of 18.3 kcal.mol<sup>-1</sup> (a value in close conformity with the experimental value of 19.7 kcal.mol<sup>-1</sup>).

**Keywords:** human carboxylesterase 2, enzymes, Quantum Mechanics/Molecular Mechanics MD simulations, reaction mechanisms, cocaine hydrolysis, drug metabolism

## 1. Introduction

Carboxylesterases (CEs, E.C. 3.1.1.1) are members of the family of serine hydrolases and are mainly localized within the lumen of the endoplasmic reticulum in different tissues, such as the liver, small intestine, and kidneys.<sup>[1]</sup> Several CE families (CE1-CE5) have been identified to date, but only three are currently characterized in humans (hCE1-hCE3), with most studies focusing on hCE1.<sup>[2]</sup>

hCEs are known to be involved in drug metabolism due to their role in therapeutic drug clearance, but also by their ability to activate prodrugs such as anticancer agents.<sup>[3]</sup> In general, they promiscuously catalyze several hydrolytic reactions with multiple endogenous and exogenous substrates (e.g. esters, thioesters, carbamates, and amides) into the corresponding carboxylic acid and alcohol products.<sup>[4]</sup> These products are usually more polar than the reagent, which results in a water solubility increase. This mechanism aids the elimination of ingested exogenous esters that might be potentially hazardous to the organism.<sup>[5]</sup>

hCE1 and 2 display broad substrate specificity; however, hCE1 is more active towards substrates with a small alcohol group and a large acyl group, although sometimes it can accept structurally distinct alcohol moieties.<sup>[6]</sup> Conversely, the hCE2 enzyme recognizes substrates with large alcohol groups and small acyl groups, translating into higher substrate specificity.<sup>[7]</sup>

Cocaine is the most studied CE substrate, and its metabolism has been extensively analyzed in animal and human cells, both *in vitro* and *in vivo*. The hCE1 and 2 isoenzymes have different ester sites for cocaine hydrolysis. hCE1 catalyzes the hydrolysis of the methyl ester, resulting in benzoylecgonine (BzE) and methanol, while the hCE2 catalyzes the benzoyl hydrolysis to obtain ecgonine methyl ester (EME) and benzoic acid. Although EME and benzoic acid metabolites have a pharmacologically inactive profile,<sup>[8]</sup> the metabolite BzE, converted by hCE1, is a stronger vasoconstrictor than cocaine-like drugs (for example norcocaine and norepinephrine).<sup>[9]</sup> In addition, BzE is a zwitterion extremely difficult to hydrolyze in the body. The largest retention time of cocaine is responsible for its induced long-term toxicity.<sup>[10]</sup>

In the same line of hCE1 and 2 activity, the blood plasma butyrylcholinesterase was found to also catalyze the *in vivo* hydrolysis of cocaine in EME and benzoic acid. Nonetheless, the wild-type metabolic activity is considerably higher towards the biologically inactive enantiomer (+)-cocaine when compared to the naturally occurring (-)-cocaine. The (+)-enantiomer is quickly cleared from plasma (within seconds) before reaching the central nervous system (CNS),<sup>[11]</sup> while the active (-)-form remains in circulation for more than 45 minutes. This time-span is sufficient for the latter to reach the CNS, which has a maximum cocaine response profile within a minutes time frame.<sup>[12]</sup> Recently, this enzyme was been reengineered to perform (-)-cocaine detoxification, and their catalytic efficiency was increased more than 2000-fold.<sup>[13,14]</sup> Nonetheless, the short half-life of cocaine in the blood and the fast conversion to the hydrolysis-resistant metabolite (benzoylecgonine) in the body, limit the therapeutic applications of these enzymes. Therefore, hCE2 has the most desirable metabolic pathway considering that its products are less toxic and more easily excreted.

Mechanistically, the CE hydrolytic reactions follow the classical ping-pong bi-bi reaction mechanism of serine hydrolases, via the formation of two tetrahedral intermediates and an acyl-serine enzyme intermediate. The catalytic triad is usually composed by the Ser-His-Asp/Glu residues, although other variations are known to occur.<sup>[7]</sup> The first hCE1 mechanism was proposed<sup>[15]</sup> as a single-step process, where the reaction moves from the reagents to the acyl monomer with a single first transition-state (TS), and the system resolved into the products with a second TS. Later, a second mechanism was reported<sup>[16]</sup> as a double-step mechanism with the classical ping-pong bi-bi reaction, where the stationary points are connected by four TSs. This mechanistic divergence is due to the different methodology used by these teams. Despite these variations, the first TS was pointed out as the rate-limiting step in both works, with free energy barriers close to the experimental results of 21.5 kcal.mol<sup>-1</sup>.

The hCE2 has been gathering interest for clinical settings applications on account of its ability to selectively activate and eliminate (pro)drugs;<sup>[17]</sup> for instance, a prodrug of gemcitabine (LY2334737) is only activated by hCE2.<sup>[18]</sup> Yet, the heterogeneity of the hCE2 N-glycan structures hinders its structural determination, so that only the hCE1 was structurally characterized by X-ray crystallography studies.<sup>[19,20]</sup> Understanding the detailed enzymatic mechanism of hCE2 can provide important insights for the design of novel pharmaceutical enzymes and their drug analogs activation (like the LY2334737).

Here, we performed conventional and accelerated MD simulations and calculated, using Quantum Mechanics/Molecular Mechanics methods, the one- (1D) and two-dimensional (2D) profiles for the complete catalytic mechanism of cocaine hydrolysis. We identified the structural movements of the  $\alpha 1$  and  $\alpha 10'$  helices that act as a putative lid.

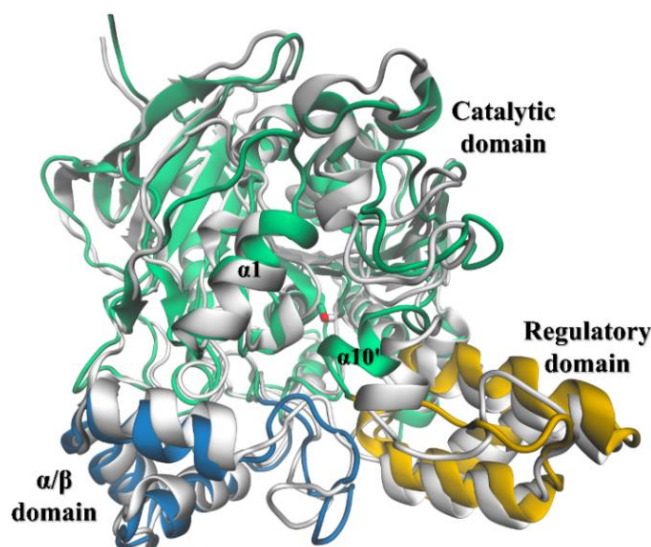
## 2. Results

### 2.1. Structure of the hCE2 enzyme

The hCE2 has the typical  $\alpha/\beta$  hydrolase-fold of the serine hydrolase superfamily. The active site is located in the catalytic domain and surrounded by the  $\alpha/\beta$  and regulatory domains (Figure 1), containing the classical Ser-His-Glu catalytic triad. The serine residue (S201) acts as the nucleophile in the acylation step, while the histidine (H448) function as an acid/base (responsible for proton transfer between the catalytic serine and the substrate) and is stabilized by a nearby glutamate residue (E334). The oxyanion hole region is composed by the backbone amides of G122, G123, and A202.<sup>[21]</sup> This spatial arrangement of hydrogen bond donors is responsible for the stabilization of a negative charge that is developed during the catalytic mechanism on the oxygen atom of the tetrahedral intermediates (**INT-1** and **INT-2**). Similar to the isoenzyme 1, the hCE2 displays two main pockets (acyl and alcohol) that accommodate the substrate in the active site.<sup>[22]</sup> The main residues that line these two pockets are illustrated in Figure S1. Resembling to what has been determined for the hCE1,<sup>[23]</sup> the catalytic triad and oxyanion hole residues of the hCE2 subdivides the active site into two regions with different polarities. The acyl pocket is mainly surrounded by hydrophobic residues (most of them Leucines) that may stabilize the acyl unit of the substrate by Van der Waals (VdW) interactions. On the other hand, the alcohol pocket is less hydrophobic, presenting more aromatic and basic residues.

The active site is capped with two  $\alpha$ -helices ( $\alpha 1$  and  $\alpha 10'$ , Figure 1) that mediate the substrate access to the active site by switching between open/closed conformations.<sup>[24]</sup> We have performed aMD simulations to obtain the open conformation. In this kind of simulation, a potential energy term is added that allows overcoming the free energy barriers in the protein landscape and accessing rare events.<sup>[25]</sup> During the 50 ns aMD simulation performed, we observed that  $\alpha 1$  and  $\alpha 10'$  helices move considerably (Figure 1 and S2). The largest movement was in the  $\alpha 10'$  helix, with a value of up to 4.4 Å, while the  $\alpha 1$  helix only varies up to 2.4 Å (Figure S2). The 50 ns aMD trajectory was clustered into open and closed conformations (Figure S3). The open conformation, which allows the substrate to enter the active site, was selected for docking of the **INT-1** structure.

To evaluate the main movements throughout the simulations we performed PCA analysis. It is based on a mathematical covariance-matrix-based technique capable of capturing the divergence of coordinates during the simulation, taking into account a reference structure. The three main principal components identified (PC1, PC2, and PC3) account for around 74 % of the enzyme's motions (Figure S4). The analysis was performed to the Ca of the enzyme, primarily scanning the whole protein landscape, which resulted in low overall representation.



**Figure 1.** Superimposition of the cMD (after 100 ns, colored) and aMD (active conformation cluster, silver) structures of hCE2. The catalytic domain is shown in green, the regulatory domain in yellow, and the  $\alpha/\beta$  domain in blue. The active site, where the sidechain of the catalytic serine is represented, is capped by two mobile  $\alpha$ -helices ( $\alpha 1$  and  $\alpha 10'$ ) that act as a putative lid.

The modeled nature of the protein could be inducing the N- and C-termini to influence the results. PCA was then adjusted to the residues comprising both  $\alpha 1$  and  $\alpha 10'$  regions alone, corresponding to the above-mentioned value. We considered that the  $\alpha 10'$  movement would lead to movement of the regulatory domain; however, the helices seemed to be the main contributors for the protein movement, as the inclusion of the regulatory domain in the analysis decreased PCA to  $\approx 60\%$ . PCA output was then used to calculate the free energy landscape (FEL) (Figure S5).

To further evaluate the nature of the motions, we analyzed the transition pathways in the webserver eBDIMS, which relies on the essential dynamics-refined elastic network model force field. The platform iteratively performs model evaluations and defines a starting structure for the projection by overlapping the two best-score normal modes (NMs) with the modeled transition. For our system, the NMs represent  $\approx 61\%$  of the movements involved, with an initial RMSD of  $3.40 \text{ \AA}$  and a final RMSD of  $0.34$  and  $0.32 \text{ \AA}$  for forward and reverse transitions, respectively (Figure S6).

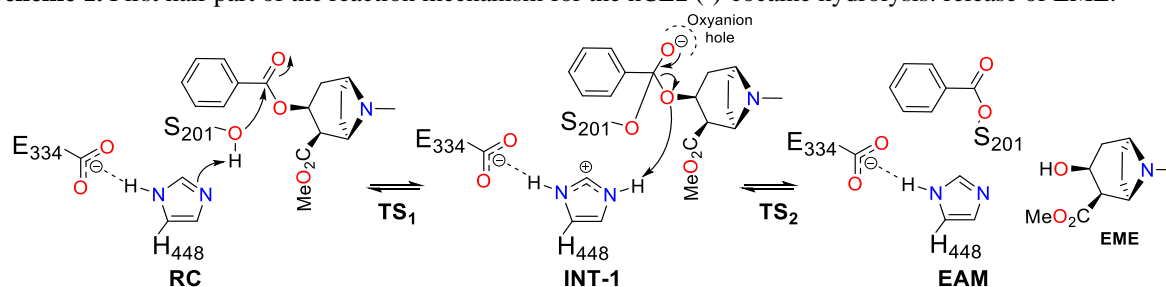
With an open conformation established, we subjected the system to the flexible side-chain covalent docking method. This covalent method differs from the “free” counterpart by anchoring the backbone of the functionalized serine residue. The remaining bonds that are capable to rotate are set free to identify the best position. This placed the benzoyl group in the acyl pocket and the EME group in the alcohol pocket with a binding energy of  $-7.86 \text{ kcal.mol}^{-1}$ . The system was then minimized through cMD, in order to achieve the local minima. The final structure was then selected to perform QM/MM calculations to study the catalytic mechanism of (-)-cocaine hydrolysis.

## 2.2. First reaction step: release of ecgonine methyl ester

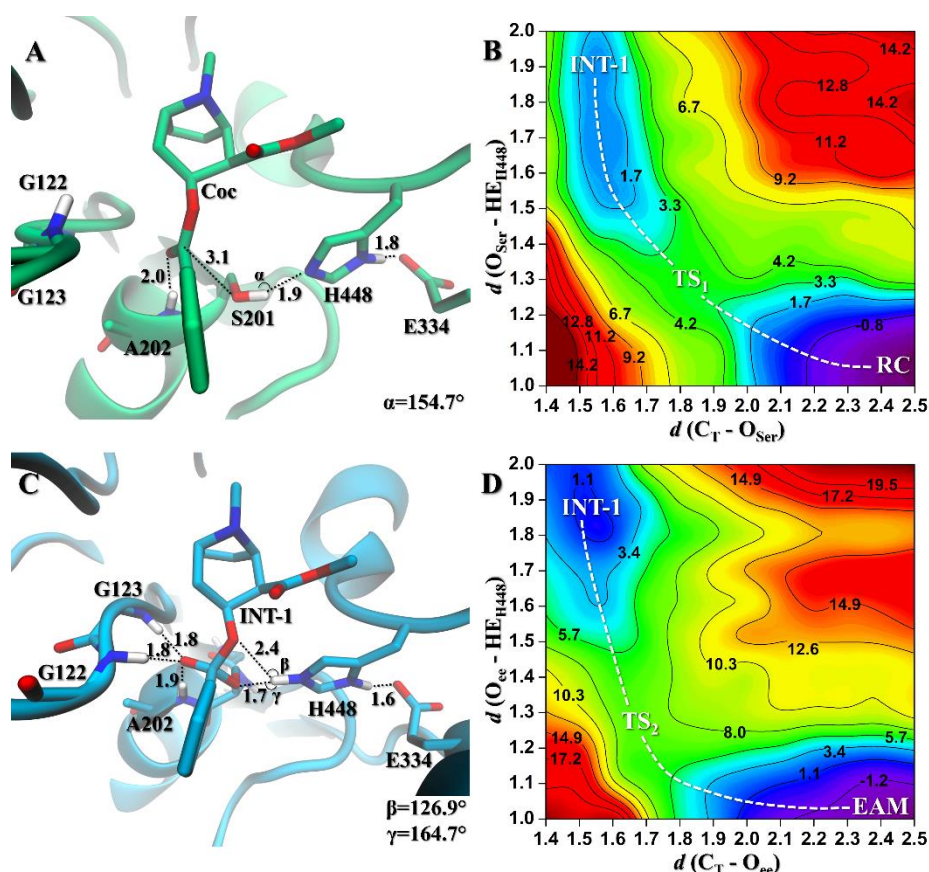
The first half part of the (-)-cocaine hydrolysis catalytic cycle (also known as the acylation step) concerns the release of the EME to generate the enzyme-activated monomer (EAM) structure (Scheme 1).



**Scheme 1.** First half part of the reaction mechanism for the hCE2 (-)-cocaine hydrolysis: release of EME.



The reaction mechanism begins with the entry of a (-)-cocaine molecule to the enzyme's active site pocket. The ecgonine ring and the methoxy group of cocaine make VdW interactions inside the pocket that accommodate this portion of the molecule: hydrophobic residues L77, F81, and L452 interact with the ring; W337 and P441 with the methoxy group (Figure S7). In the **RC** structure (3.1 Å, Figure 2A), a hydrogen bond is formed between the A202 amide hydrogen of the oxyanion hole and the carbonyl oxygen of the substrate (2.0 Å distance). Proton transfer to the H448 residue (1.9 Å and 154.7°, Figure 3A) occurs simultaneously with the nucleophilic attack by the S201 oxygen atom ( $O_{Ser}$ ) on the benzoyl carbonyl carbon atom of cocaine. The first transition state ( $TS_1$ ) has a free energy barrier ( $\Delta G^\ddagger$ ) of 6.8 kcal.mol<sup>-1</sup> and is resolved into the **INT-1** structure, which is 3.4 kcal.mol<sup>-1</sup> below  $TS_1$  (Figure 3B). In the **INT-1** structure (Figure 2C), the histidine residue is positively charged and stabilized by the glutamate residue (E334). The oxyanion hole backbone amides of G122, G123, and A202 are now stabilizing the developed negative charge on the tetrahedral **INT-1** oxygen atom (the H bond distance is 1.8 Å with the glycine residues and 1.9 Å for the alanine residue).



**Figure 2.** Active site pocket reference structures of the stationary points optimized with PM6/parm99SB: **RC** (A) and **INT-1** (C). 2D-PMFs for the first (B) and second (D) steps calculated with B3LYP/6-31G(d,p)<sup>[26]</sup> and Grimme D3 dispersion.<sup>[27]</sup> The key distances are given in Å and the relative energy values in kcal.mol<sup>-1</sup>.

The reaction proceeds with the release of the EME (Scheme 1) to form the **EAM** structure. For that, the positively charged histidine needs to transfer its proton to the alcohol oxygen atom of the EME group ( $O_{ee}$ , Figure 2). The **TS<sub>2</sub>** is 6.5 kcal.mol<sup>-1</sup> above the **INT-1**, and the **EAM** is 10.1 kcal.mol<sup>-1</sup> below **TS<sub>2</sub>** (Figure 2C).

The overall **TS<sub>2</sub>**  $\Delta G^*$  of 9.9 kcal.mol<sup>-1</sup> reflects the higher distance of the HE2 of H448 towards the ecgonine oxygen ( $O_{ee}$ ) when compared to the serine's oxygen ( $O_{Ser}$ ). Also, the angles are far from ideal to move the reaction forward ( $\beta=126.9^\circ$  and  $\gamma=164.7^\circ$ , Figure 2C) and the reverse reaction is favored with a lower free energy barrier (3.4 kcal.mol<sup>-1</sup>, Figure 2B).

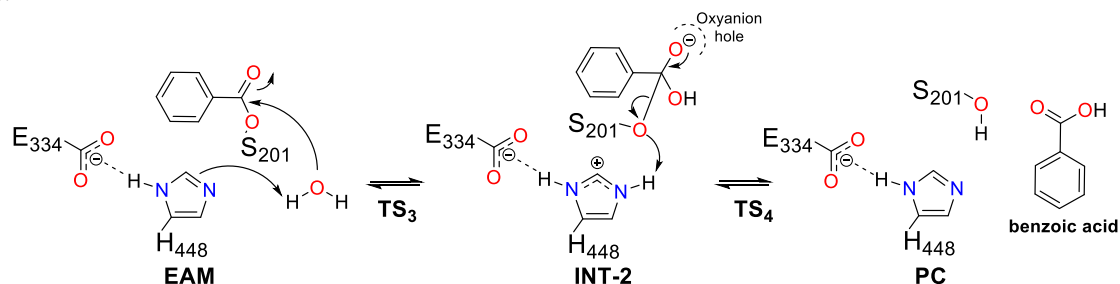
### 2.3. Second reaction step: release of benzoic acid

The second half part of the (-)-cocaine hydrolysis cycle (also known as the deacylation step) concerns the release of benzoic acid to generate the product complex (**PC**, Scheme 2).

The active site pocket with the **EAM** structure contains a large number of water molecules (Figure 3A). The nucleophilic attack of a water molecule to the covalently bonded substrate carbonyl carbon (2.2 Å, Figure 3A) generates the **INT-2** structure. The energy barrier for this reaction (**TS<sub>3</sub>**, Figure 3B) amounts to only 2.1 kcal.mol<sup>-1</sup>.

Alike the **INT-1** structure, the oxyanion hole oxygen atom of the **INT-2** structure makes three hydrogen bonds with the backbone amide groups of G122, G123, and A202 (the hydrogen bond distances are 1.6, 2.0, and 1.7 Å, respectively). These interactions stabilize the negative charge developed in this oxygen atom.

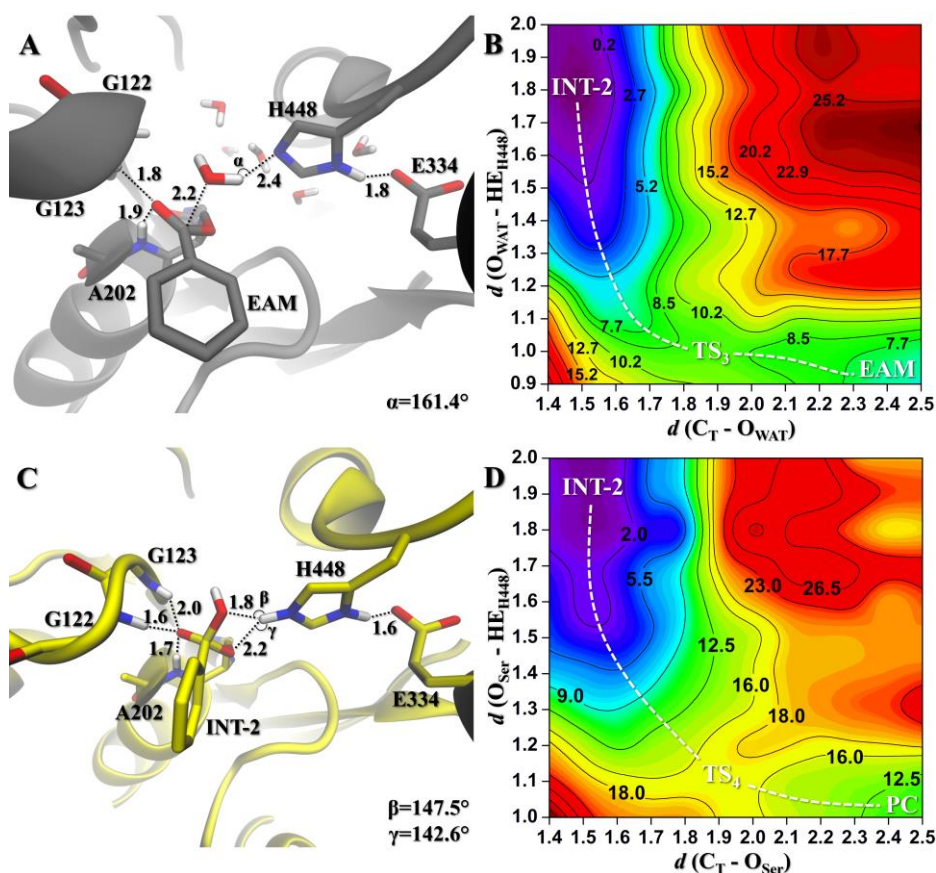
**Scheme 2.** Second half part of the reaction mechanism for the hCE2 (-)-cocaine hydrolysis: release of benzoic acid.



The H448 residue (Figure 3C) is at a similar distance to both oxygen atoms that move the reaction forward ( $O_{Ser}$ , 2.2 Å) or backward ( $O_{WAT}$ , 1.8 Å). The **INT-2** (Figure 4C) is exergonic in relation to the **EAM** with a  $\Delta G$  of -9.3 kcal.mol<sup>-1</sup>.

Then, the system evolves towards the release of the benzoic acid molecule (**PC**, Scheme 2). The final transition state (**TS<sub>4</sub>**, Figure 3D) has a  $\Delta G^*$  of 18.3 kcal.mol<sup>-1</sup> (Figure 3D). The **PC** is 10.0 kcal.mol<sup>-1</sup> above the **INT-2** (Figure 3D) and the enzyme ready for another turnover of cocaine hydrolysis.

The reported turnover number ( $k_{cat}$ ) by Potter and co-workers<sup>[28]</sup> is 0.0889 s<sup>-1</sup> for the hydrolysis of (-)-cocaine to afford EME and benzoic acid. According to Eyring's equation,<sup>[29]</sup> this corresponds to a free energy of about 19.7 kcal.mol<sup>-1</sup> at 37 °C. Our simulations suggest that the rate-limiting step is the formation of benzoic acid (**TS<sub>4</sub>**) with a  $\Delta G^*$  of 18.3 kcal.mol<sup>-1</sup>, which is in excellent agreement with this value. Chao et al.<sup>[30]</sup> also reported that the deacylation step was the rate limiting-step of cocaine hydrolysis by butyrylcholinesterase.



**Figure 3** Active site pocket reference structures of the stationary points optimized with PM6/parm99SB: **EAM** (A) and **INT-2** (C). 2D-PMFs for the third (B) and fourth (D) steps calculated with B3LYP/6-31G(d,p)<sup>[26]</sup> and Grimme D3 dispersion.<sup>[27]</sup> The key distances are given in Å and the relative energy values in kcal.mol<sup>-1</sup>.

### 3. Conclusion

In this work, we have studied the helices movement that cap the active site by performing conventional and accelerated Molecular Dynamic simulations. Our structural analysis indicates that the  $\alpha 1$  and  $\alpha 10'$  helices motion warrants an open conformation of the enzyme (allowing the substrate to enter into the active site). We observe a higher displacement of the  $\alpha 10'$  helix when compared to the  $\alpha 1$ , which is highlighted in the PCA and FEL results, and indicative of the role of both helices in the overall protein movements.

The active site functionalized serine intermediate was placed in the open conformation structure resorting to the flexible side-chain covalent docking method. The benzoyl and the ecgonine groups were placed in the acyl and alcohol pockets, respectively, with -7.86 kcal.mol<sup>-1</sup> of binding energy. We used this structure for the cMD and QM/MM studies. Our calculated 2D-QM/MM profiles on the full catalytic cycle showed that the acylation step requires 9.9 kcal.mol<sup>-1</sup> (**TS<sub>2</sub>**) to move the reaction forward (the **TS<sub>1</sub>** is 6.8 kcal.mol<sup>-1</sup> above the **RC**). This may be related to proton transfer that occurs concomitantly with CO bond formation/breaking: a combination of the higher distance of the HE2 atom of H448 towards the ecgonine oxygen ( $O_{ee}$ ) in comparison to the serine's oxygen ( $O_{Ser}$ ), expressing better positioned angles that favor the reverse reaction (**TS<sub>1</sub>**). In the deacylation step, the **TS<sub>4</sub>** (formation of benzoic acid) was identified as the rate-limiting step of the (-)-cocaine hydrolysis by the hCE2. This step requires 18.3 kcal.mol<sup>-1</sup> to be achieved, being the higher TS in the whole cycle and in agreement with the reported experimental barrier.

Since the hCE1 yields BzE (that is extremely difficult to hydrolyze) and the WT blood plasma butyrylcholinesterase is more active towards the biologically inactive cocaine enantiomer, the hCE2 has the most desirable metabolic pathway. Its metabolites are less toxic



and more easily excreted. These novel mechanistic insights about the hydrolysis of (-)-cocaine by the hCE2 are expected to benefit the discovery of new enhanced enzyme variants that are capable to better hydrolyze (-)-cocaine.

#### 4. Methods

**Modeling.** A homology model of hCE2 was built using Modeller v9.24.<sup>[31]</sup> The crystal structure of hCE1 (PDB code 5A7F, 1.86 Å resolution, monomer)<sup>[32]</sup> was used as template and aligned with the “EST2\_HUMAN” protein sequence (UniprotKB code O00748)<sup>[33]</sup> with the Clustal Omega server.<sup>[34]</sup> The protein sequence for the hCE2 model is presented in Figure S8. The model was then subjected to 100 ns of conventional Molecular Dynamics (cMD) before any analysis, with the settings described in the MD section below.

**Initial Setup.** The protonation states (pH=7.4)<sup>[28]</sup> of the hCE2 model were assigned with MolProbity. The INT-1 and INT-2 residues were geometry optimized in Gaussian09<sup>[35]</sup> using B3LYP<sup>[26]</sup> with the 6-31G(d) basis set and a Polarizable Continuum Model solvent description.<sup>[36]</sup> The atomic partial charges were calculated from the optimized structure resorting to the Restrained Electrostatic Potential (RESP) method<sup>[37]</sup> from the HF/6-31G(d,p) single-point energy calculations.

**Molecular Docking.** The initial position of the INT-1 was obtained by the flexible side-chain covalent docking method<sup>[38]</sup> and performed with the AutoDock4.2 suite of programs using the Lamarckian Genetic Algorithm (LGA).<sup>[39]</sup> A total of 150 runs were carried out. The population was set to 300, the maximum number of generations 27,000, and the maximum number of energy evaluations 2,500,000.

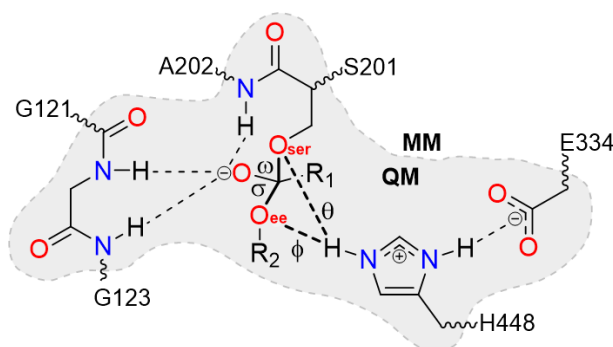
**Molecular Dynamics.** The cMD simulations were performed using the Amber molecular dynamics program (AMBER18)<sup>[40]</sup> with the parm99SB<sup>[41]</sup> and GAFF<sup>[42]</sup> force fields for the protein and ligand, respectively. The structures were placed within an octahedral box of  $\approx$  13,500 TIP3P waters<sup>[40]</sup> and 6 sodium ions were added to neutralize the system. The systems were then subjected to two initial energy minimizations and 500 ps of equilibration in an NVT ensemble using Langevin dynamics with small restraints on the protein of 10 kcal.mol<sup>-1</sup> to heat the systems. Production simulations of 5 ns were carried out at 310.15 K in the NPT ensemble using Langevin dynamics with a collision frequency of 1 ps<sup>-1</sup>. Constant pressure periodic boundary conditions were imposed with an average pressure of 1 atm. Isotropic position scaling was used to maintain pressure with a relaxation time of 2 ps. The time step was set to 2 fs. SHAKE constraints were applied to all bonds involving hydrogen atoms.<sup>[43]</sup> The Particle Mesh Ewald (PME) method<sup>[44]</sup> was used to calculate electrostatic interactions with a cut-off distance of 10 Å. The accelerated Molecular Dynamics (aMD) was conducted for 50 ns and the parameters selected based upon the system size, average dihedral, and potential energy:  $E_{\text{dih}} = 7,561.8$  kcal.mol<sup>-1</sup>,  $\alpha_{\text{dih}} = 373.1$  kcal.mol<sup>-1</sup>,  $E_{\text{tot}} = -118,535.1$  kcal.mol<sup>-1</sup> and  $\alpha_{\text{tot}} = 9,476.0$  kcal.mol<sup>-1</sup>.<sup>[45]</sup>

**Principal Components Analysis and Free Energy Landscape.** Principal Components Analysis (PCA) was used to analyze the data from the aMD simulations, using the Bio3D package<sup>[46]</sup> (R statistical program). The protocol used was described by Gedeon et al.<sup>[47]</sup> The analysis was performed to the C $\alpha$  of the total 8,287 system atoms. The PDB and trajectory structures were fitted and superposed to remove translations and rotations. The variance of the data in each principal component was evaluated by calculating eigenvectors and the corresponding eigenvalues. To evaluate on average how much each residue moves during the trajectory, we calculated the root mean square fluctuations (RMSF). Free energy landscape (FEL) calculations were performed on the CPPTRAJ module from AMBER18,<sup>[40]</sup> followed by elastic-network driven/Brownian Dynamics Importance Sampling (eBDIMS) to evaluate conformation transition pathways, calculated in the eBDIMS server (default parameters).

**Quantum Mechanics/Molecular Mechanics Calculations.** The Quantum Mechanics/Molecular Mechanics (QM/MM) calculations<sup>[48,49]</sup> were performed using the



internal semi-empirical hybrid QM/MM functionality implemented in AMBER18<sup>[40]</sup> with periodic boundary conditions. The last MD structure of each complex was used as the initial structure for the QM/MM calculations. The PM6 semi-empirical method<sup>[50,51]</sup> was employed to describe the QM region while the MM region was described by the parm99SB force field.<sup>[41]</sup> The reactions were conducted at 310.15 K.<sup>[28]</sup> Electrostatic embedding<sup>[52]</sup> was employed and the boundary treated via the link atom approach. Long-range electrostatic interactions were described by an adapted implementation of the PME method<sup>[44]</sup> for QM/MM.<sup>[53]</sup> For the 1D and 2D-PMFs the same set of reaction coordinates was used (Figure 1): i) the distance between the H448 proton HE and the S201 oxygen atom ( $O_{Ser}$ ,  $\theta$ ) and ii) the distance between the H448 proton HE and the substrate oxygen atom ( $O_{ee}$ ,  $\phi$ ).



**Figure 4.** Division of the QM/MM system: all atoms inside the shaded shape belong to the QM subsystem in the QM/MM calculations. All other atoms belong to the MM subsystem. The link atoms are represented with a wavy bond. **INT-1**  $R_1=Ph$ ,  $R_2=EME$ , **INT-2**  $R_1=Ph$ ,  $R_2=H$ .

In the 1D-PMF, one reaction coordinate was incrementally restrained in steps of 0.1 Å using the umbrella sampling method, except near the transition states where smaller steps of 0.02 Å were employed. Then, and in accordance with our previous work,<sup>[54]</sup> the structures from the 1D profiles were used as starting points for the 2D-PMF scans. To keep the reaction coordinates at the requested distances and to ensure enough overlap between windows, an umbrella constraint force of 200 kJ.mol<sup>-1</sup>.Å<sup>-2</sup> was used along the reaction path and incrementally increased to 1,000 kJ.mol<sup>-1</sup>.Å<sup>-2</sup> as the reaction coordinates deflect from the minimum energy path. For every window, a total of 15 ps was simulated with a time step of 1 fs. The distances were scanned from 1.4 to 2.5 Å ( $\omega$  and  $\sigma$ , excepting the third step, where the scanned reaction vary from 0.9 to 2.5 Å) and from 1.0 to 2.0 Å ( $\theta$  and  $\phi$ ) comprising a total of 132 simulations for first, second and fourth steps, and 144 simulations for the third step. The PMFs were computed resorting to the Weighted Histogram Analysis Method (WHAM) with the Monte Carlo bootstrap error analysis<sup>[55,56]</sup> and the minimum energy path traced resorting to the MEPSA v1.4.<sup>[57]</sup>

**High-level Layer Corrections.** For the high-level (HL) layer corrections of the 2D-PMFs, we resorted to the method described in detail in reference<sup>[58]</sup> which is based on the work of Truhlar and co-workers.<sup>[59,60]</sup> Multiple structures were retrieved from the 2D-PMFs and subjected to single-point calculations using the Gaussian09<sup>[35]</sup> program, first with the semi-empirical PM6<sup>[50,51]</sup> and then with B3LYP/6-31G(d,p)<sup>[26]</sup> and Grimme D3 dispersion.<sup>[27]</sup> The corrected energies term ( $E$ ) was interpolated from those structures employing the following equation (1):

$$E = E_{LL} + S(\Delta E_{LL}^{HL}) \quad (1)$$

Where the term  $\Delta E_{LL}^{HL}$  corresponds to the difference between the free energies for the HL layer set, calculated with B3LYP/6-31G(d,p)<sup>[26]</sup> and Grimme D3 dispersion<sup>[27]</sup> and PM6<sup>[50,51]</sup>

(low level (LL) theory), while  $S$  represents the cubic spline function of the difference between the HL and LL theory representing the QM region.

## Acknowledgments

This work was sponsored through grants SFRH/BD/144303/2019, 2020.10114.BD and IF/01272/2015, and project UIDB/04539/2020 financed by Portuguese National funds via FCT – Fundação para a Ciência e a Tecnologia and Regional Operational Program Centro (CENTRO2020) under the Portuguese Partnership Agreement 2020 by the European Regional Development Fund (ERDF). The authors acknowledge the support of the National Distributed Computing Infrastructure (INCD) funded by FCT and FEDER with computational resources, under the projects 01/SAICT/2016 n° 022153 and Advanced Computing Project CPCA/A2/4568/2020.

## References

- [1] T. Satoh, M. Hosokawa, *Chem.-Biol. Interac.* **2006**, *162*, 195–211.
- [2] T. Satoh, P. Taylor, W. F. Bosron, S. P. Sanghani, M. Hosokawa, B. N. L. Du, *Drug. Metab. Dispos.* **2002**, *30*, 488–493.
- [3] M. K. Ross, J. A. Crow, *J. Biochem. Mol. Toxicol.* **2007**, *21*, 187–196.
- [4] S. Sonal P., S. Paresh C., S. Marissa A., B. William F., *Protein Pept. Lett.* **2009**, *16*, 1207–1214.
- [5] S. Casey Laizure, V. Herring, Z. Hu, K. Witbrodt, R. B. Parker, *Pharmacotherapy* **2013**, *33*, 210–222.
- [6] M. Hosokawa, *Molecules* **2008**, *13*, 412–431.
- [7] D. Wang, L. Zou, Q. Jin, J. Hou, G. Ge, L. Yang, *Acta Pharm. Sin. B* **2018**, *8*, 699–712.
- [8] E. T. Shimomura, G. F. Jackson, B. D. Paul, in *Critical Issues in Alcohol and Drugs of Abuse Testing*, Elsevier, **2019**, pp. 215–224.
- [9] N. J. Connors, R. S. Hoffman, *J. Pharmacol. Exp. Ther.* **2013**, *347*, 251–257.
- [10] R. Pal, M. Megharaj, K. P. Kirkbride, R. Naidu, *Sci. Total Environ.* **2013**, *463–464*, 1079–1092.
- [11] S. J. Gatley, R. R. MacGregor, J. S. Fowler, A. P. Wolf, S. L. Dewey, D. J. Schlyer, *J. Neurochem.* **1990**, *54*, 720–733.
- [12] S. J. Gatley, *Biochem. Pharmacol.* **1991**, *41*, 1249–1254.
- [13] F. Zheng, L. Xue, S. Hou, J. Liu, M. Zhan, W. Yang, C.-G. Zhan, *Nat. Commun.* **2014**, *5*, 3457.
- [14] K. E. Larrimore, I. C. Kazan, L. Kannan, R. P. Kendle, T. Jamal, M. Barcus, A. Bolia, S. Brimijoin, C.-G. Zhan, S. B. Ozkan, T. S. Mor, *Sci Rep* **2017**, *7*, 10419.
- [15] J. Yao, X. Chen, F. Zheng, C.-G. Zhan, *Mol. Pharmaceutics* **2018**, *15*, 3871–3880.
- [16] M. Yan, Z. Zhang, Z. Liu, C. Zhang, J. Zhang, S. Fan, Z. Yang, *Molecules* **2019**, *24*, 4057.
- [17] J. Lamego, B. Cunha, C. Peixoto, M. F. Sousa, P. M. Alves, A. L. Simplicio, A. S. Coroadinha, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 1161–1173.
- [18] S. E. Pratt, S. Durland-Busbice, R. L. Shepard, K. Heinz-Taheny, P. W. Iversen, A. H. Dantzig, *Clin. Cancer. Res.* **2013**, *19*, 1159–1168.
- [19] S. Bencharit, C. L. Morton, J. L. Hyatt, P. Kuhn, M. K. Danks, P. M. Potter, M. R. Redinbo, *Chem. Biol.* **2003**, *10*, 341–349.
- [20] C. D. Fleming, C. C. Edwards, S. D. Kirby, D. M. Maxwell, P. M. Potter, D. M. Cerasoli, M. R. Redinbo, *Biochemistry* **2007**, *46*, 5063–5071.
- [21] M. Alam, D. E. Vance, R. Lehner, *Biochemistry* **2002**, *41*, 6679–6687.
- [22] G. Vistoli, A. Pedretti, A. Mazzolari, B. Testa, *J. Comput. Aided. Mol. Des.* **2010**, *24*, 771–787.
- [23] G. Vistoli, A. Pedretti, A. Mazzolari, B. Testa, *Bioorg. Med. Chem.* **2010**, *18*, 320–329.
- [24] J. Lian, R. Nelson, R. Lehner, *Protein Cell* **2018**, *9*, 178–195.
- [25] L. C. T. Pierce, R. Salomon-Ferrer, C. Augusto F. de Oliveira, J. A. McCammon, R. C. Walker, *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.
- [26] C. S. Ashvar, F. J. Devlin, K. L. Bak, P. R. Taylor, P. J. Stephens, *J. Phys. Chem.* **1996**, *100*, 9262–9270.
- [27] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- [28] M. Hatfield, L. Tsurkan, J. Hyatt, X. Yu, C. Edwards, L. Hicks, R. Wadkins, P. Potter, *Br. J. Pharmacol.* **2010**, *160*, 1916–1928.
- [29] M. G. Evans, M. Polanyi, *Trans. Faraday Soc.* **1935**, *31*, 875.
- [30] C.-G. Zhan, D. Gao, *Biophys. J.* **2005**, *89*, 3863–3872.
- [31] B. Webb, A. Sali, *Curr. Protoc. Bioinformatics* **2016**, *54*, 5.6.1–5.6.37.
- [32] V. A. de Souza, D. J. Scott, J. E. Nettleship, N. Rahman, M. H. Charlton, M. A. Walsh, R. J. Owens, *PLOS ONE* **2015**, *10*.

- [33] E. V. Pindel, N. Y. Kedishvili, T. L. Abraham, M. R. Brzezinski, J. Zhang, R. A. Dean, W. F. Bosron, *J. Biol. Chem.* **1997**, 272, 14769–14775.
- [34] F. Madeira, Y. mi Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn, R. Lopez, *Nucleic Acids Res.* **2019**, 47, 636–641.
- [35] M. J. Frisch, G. W. Trucks, J. R. Cheeseman, G. Scalmani, M. Caricato, H. P. Hratchian, X. Li, V. Barone, J. Bloino, G. Zheng, T. Vreven, J. A. Montgomery, G. A. Petersson, G. E. Scuseria, H. B. Schlegel, H. Nakatsuji, A. F. Izmaylov, R. L. Martin, J. L. Sonnenberg, J. E. Peralta, J. J. Heyd, E. Brothers, F. Ogliaro, M. Bearpark, M. A. Robb, B. Mennucci, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, A. Rendell, R. Gomperts, V. G. Zakrzewski, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, Gaussian 09, Revision B.01. Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford CT, **2009**.
- [36] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, 105, 2999–3094.
- [37] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, 97, 10269–10280.
- [38] G. Bianco, S. Forli, D. S. Goodsell, A. J. Olson, *Protein Sci.* **2016**, 25, 295–301.
- [39] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J. Comput. Chem.* **2009**, 30, 2785–2791.
- [40] R. Salomon-Ferrer, D. A. Case, R. C. Walker, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, 3, 198–210.
- [41] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, 65, 712–725.
- [42] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, 25, 1157–1174.
- [43] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, 23, 327–341.
- [44] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, 98, 10089–10092.
- [45] J. Wereszczynski, J. A. McCammon, in *Computational Drug Discovery and Design* (Ed.: R. Baron), Springer New York, New York, NY, **2012**, 515–524.
- [46] B. J. Grant, L. Skjærven, X. Yao, *Protein Sci.* **2021**, 30, 20–30.
- [47] P. C. Gedeon, J. R. Thomas, J. D. Madura, in *Molecular Modeling of Proteins* (Ed.: A. Kukol), Springer New York, New York, NY, **2015**, 253–287.
- [48] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, 103, 227–249.
- [49] A. T. P. Carvalho, A. Barrozo, D. Doron, A. V. Kilshtain, D. T. Major, S. C. L. Kamerlin, *J. Mol. Graph. Model.* **2014**, 54, 62–79.
- [50] G. Jindal, A. Warshel, *J. Phys. Chem. B* **2016**, 120, 9913–9921.
- [51] J. J. P. Stewart, *J. Mol. Model.* **2007**, 13, 1173–1213.
- [52] D. Bakowies, W. Thiel, *J. Phys. Chem.* **1996**, 100, 10580–10594.
- [53] K. Nam, J. Gao, D. M. York, *J. Chem. Theory Comput.* **2005**, 1, 2–13.
- [54] P. R. Figueiredo, B. C. Almeida, D. F. A. R. Dourado, A. F. Sousa, A. J. D. Silvestre, A. T. P. Carvalho, *ChemCatChem* **2020**, 12, 4845–4852.
- [55] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *J. Comput. Chem.* **1992**, 13, 1011–1021.
- [56] A. Grossfield, An Implementation of WHAM: The Weighted Histogram Analysis Method, v2.0.9, [http://membrane.urmc.rochester.edu/wordpress/?page\\_id=126](http://membrane.urmc.rochester.edu/wordpress/?page_id=126), **2018**.
- [57] I. Marcos-Alcalde, J. Setoain, J. I. Mendieta-Moreno, J. Mendieta, P. Gómez-Puertas, *Bioinformatics* **2015**.
- [58] J. J. Ruiz-Pernía, E. Silla, I. Tuñón, S. Martí, V. Moliner, *J. Phys. Chem. B* **2004**, 108, 8427–8433.
- [59] Y.-Y. Chuang, J. C. Corchado, D. G. Truhlar, *J. Phys. Chem. A* **1999**, 103, 1140–1149.
- [60] K. A. Nguyen, I. Rossi, D. G. Truhlar, *J. Chem. Phys.* **1995**, 103, 5522–5530.

## Supporting Information

### Human Carboxylesterase 2 in Cocaine Metabolism

Pedro R. Figueiredo<sup>[a,b]</sup>, Ricardo D. González<sup>[a,b]</sup> and Alexandra T. P. Carvalho<sup>[a\*]</sup>

<sup>[a]</sup>CNC – Center for Neuroscience and Cell Biology, Institute for Interdisciplinary Research (IIIUC), University of Coimbra, Coimbra, Portugal

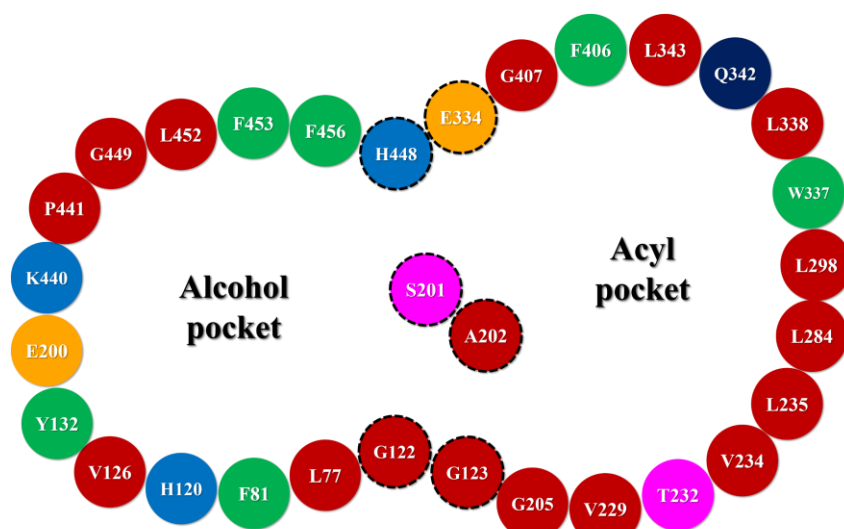
<sup>[b]</sup>PhD Programme in Experimental Biology and Biomedicine, Institute for Interdisciplinary Research (IIIUC), University of Coimbra, Casa Costa Alemão, Coimbra, Portugal

\*atpcarvalho@uc.pt

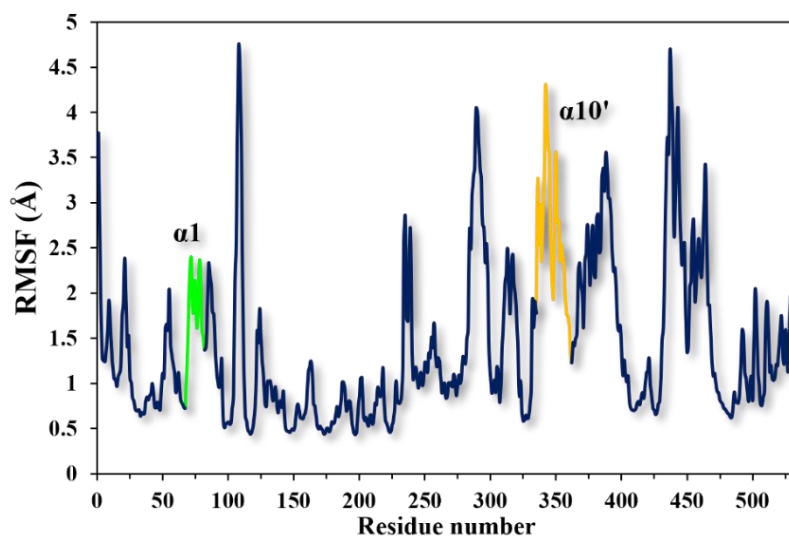
### Table of Contents

<b>Figure S1.</b> Schematic illustration of the hCE2 pockets.....	S2
<b>Figure S2.</b> RMSF for the simulated 50 ns of aMD.....	S2
<b>Figure S3.</b> Cluster structures from the 50 ns aMD trajectory. ....	S3
<b>Figure S4.</b> Bio3D PCA plot for the 50 ns aMD trajectory. ....	S3
<b>Figure S5.</b> Free energy landscape (FEL) plots .....	S4
<b>Figure S6.</b> Normal Mode and RMSD (initial and final points).....	S4
<b>Figure S7.</b> VdW interactions in the <b>RC</b> structure. ....	S4
<b>Figure S8.</b> Protein FASTA sequence for the hCE2 protein “EST2_HUMAN”.....	S5

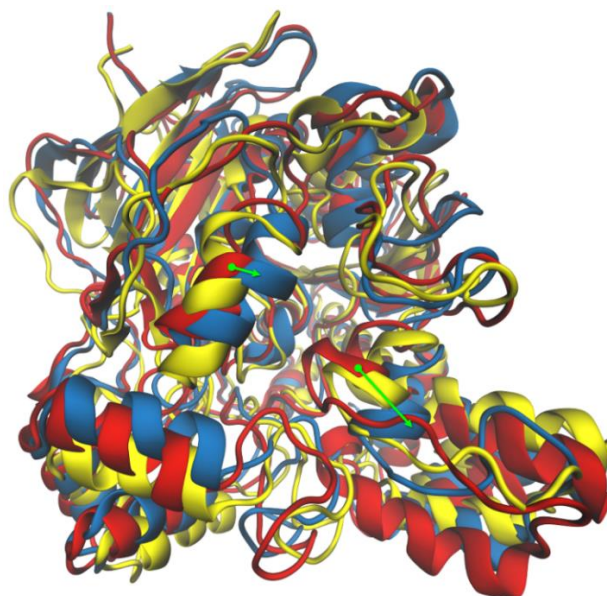




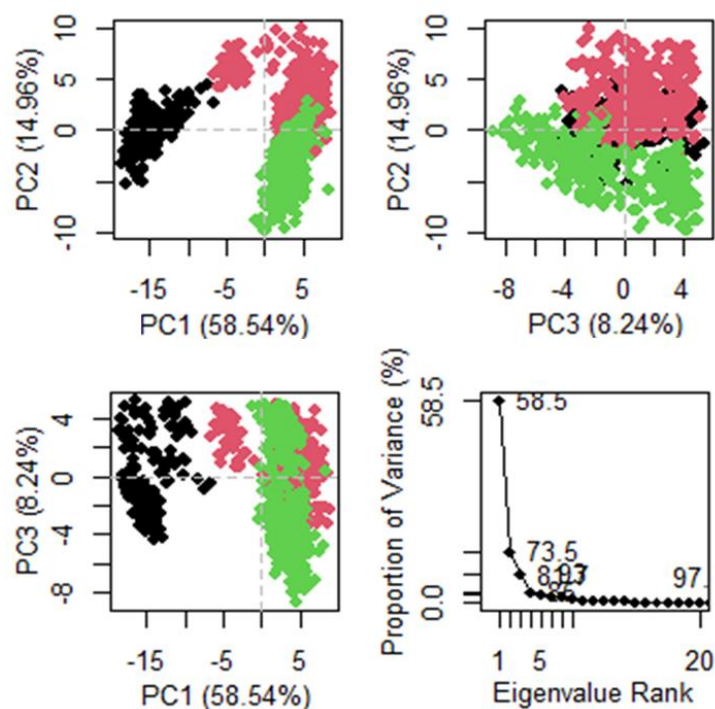
**Figure S1.** Schematic illustration representing the main residues that line the catalytic pockets of the hCE2. The different types of residues are color-coded: aliphatic (red), aromatic (green), acidic (orange), basic (blue), hydroxylic (pink), and amidic (dark blue). The catalytic triad and oxyanion hole residues are black dashed.



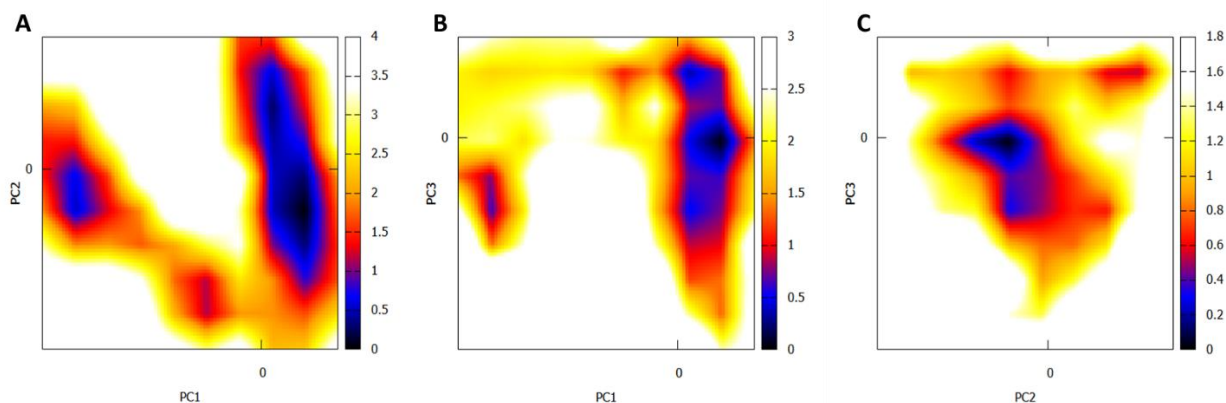
**Figure S2.** RMSF for the simulated 50 ns of aMD. Residues of the  $\alpha 1$  helix are represented in green and residues from the  $\alpha 10'$  helix represented in yellow.



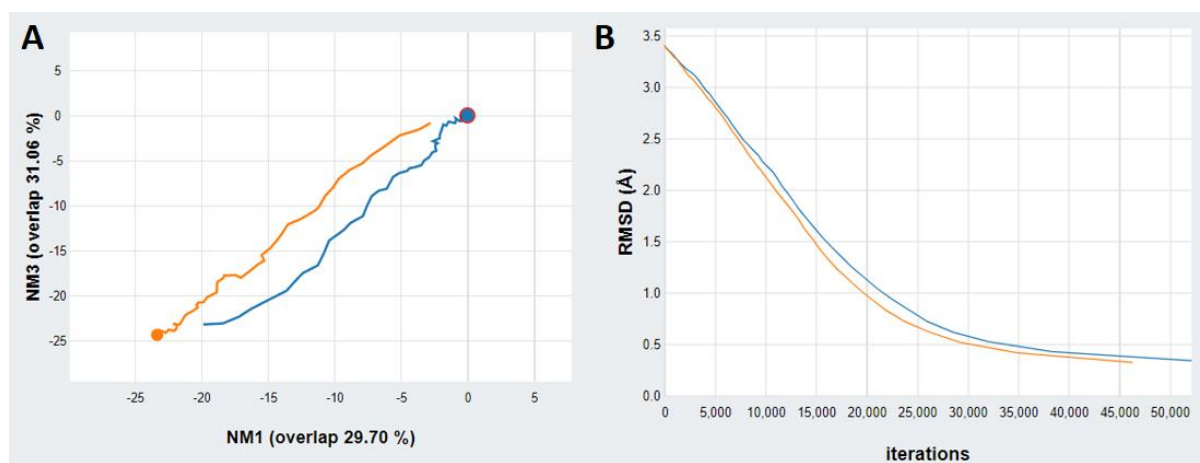
**Figure S3.** Cluster structures from the 50 ns aMD trajectory. The first frame of the trajectory is represented in red, and the yellow and blue structures concern the clustered structures. The green arrows correspond to the movement between the open and closed conformations.



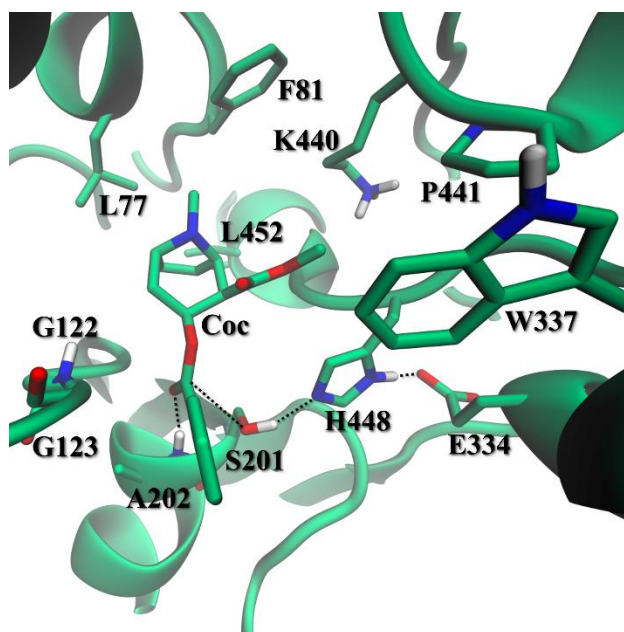
**Figure S4.** Bio3D PCA plot for the 50 ns aMD trajectory. The two-dimensional projection of each simulated frame reflects the product of the  $C\alpha$  coordinate in that specific frame with the eigenvector of the principal component. The results are time-related and color-represented (blue to green to pink). The fourth graph is a scree plot of the amount of variance of the data in each calculated principal component.



**Figure S5.** Free energy landscape (FEL) plots for the three combinations of principal components: PC1-PC2 (A), PC1-PC3 (B), and PC2-PC3 (C).



**Figure S6.** Normal Mode of the transition pathway as calculated in eBDIMS (A) and initial and final RMSD for forward (in blue) and reverse (in orange) transitions (B).



**Figure S7.** VdW interactions between the hCE2 active site and cocaine in the **RC** structure.

MRLHRLRARLSAVACGLLLLLVRGQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQ  
TFLGIPFAKPPLGPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTFP  
SDSMSEDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMMLAALENVV  
VVIIQYRLGVLGFFSTGDKHATGNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGE  
SAGGTSVSSLVVSPISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSE  
ALVGCLRGKSKEEILAINKPFKMIPGVVDGVFLPRHPQELLASADFQPVPSIVGVNNN  
EFGWLIPKVMRIYDTQKEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQ  
TLQAQFQEMMADSMFVIPALQVAHFQCSRAPVYFYEFQHQP SWLKNIRPPHMKADH  
GDELPFVFRSFFGGNYIKFTEEEEQLSRKMMKYWANFARNGNPNNGEGLPHWPLFDQ  
EEQYLQLNLQPAVGRALKAHRLQFWKKALPQKIQELEEPEERHTEL

**Figure S8.** Protein FASTA sequence for the hCE2 protein “EST2\_HUMAN” UniProtKB code O00748.