ARTICLE

Successes and Challenges in Multiscale Modelling of Artificial Metalloenzymes: the Case Study of POP-Rh_2 Cyclopropanase

José-Emilio Sánchez-Aparicio,‡[a] Giuseppe Sciortino,‡[a,§] Eric Mates-Torres,[a,+] Agustí Lledós*[a] and Jean-Didier Maréchal*[a]

Received 00th January 20xx, Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Molecular modelling applications in metalloenzyme design are still scarce due to a series of challenges. In top of that, the simulations of metal-mediated binding and the identification of catalytic competent geometries requiring both large conformational exploration and simulation of fine electronic properties. Here, we demonstrate how the incorporation of new tools in multiscale strategies, namely substrate diffusion exploration, allow taking a step further. As a showcase the enantioselective profiles of the most outstanding variants of an artificial Rh_2-based cyclopropanase (GSH, HFF and RFY) developed by Lewis and co-workers (Nat. Commun., 2015, 6, 7789 and Nat. Chem., 2018, 10, 318–324) have been rationalized. DFT calculations on the free-cofactor-mediated process identify the carbene insertion and the cyclopropanoid formation as crucial events, being the latter the enantiodetermining step, which displays up to 8 competitive orientations easily altered by the protein environment. The key intermediates of the reaction were docked into the protein scaffold showing that some mutated residues have direct interaction with the cofactor and/or the co-substrate. These interactions take form of a direct coordination of Rh in GSH and HFF and a strong hydrophobic patch with the carbene moiety in RFY. Posterior molecular dynamics sustain that the cofactor induces global re-arrangements of the protein. Finally, massive exploration of substrate diffusion, based on the GPathFinder approach, defines this event as the origin of the enantioselectivity in GSH and RFY. For HFF, fine molecular dockings suggest that it is likely related to local interactions upon diffusion. This work shows how modelling of long-range mutations on the catalytic profiles of metalloenzymes may be unavoidable and software simulating substrate diffusion should be applied.

Introduction

Molecular modelling has become a major ally in chemical biology with outstanding contributions in drug design and biocatalysis amongst others.[1–3] It has also reached a prominent position in bioinorganics as demonstrated by a plethora of studies dedicated to study the catalytic mechanism of metalloenzymes or their inhibitions.[4–11] Nowadays, modelers count on a wide variety of methods whose applicability depends on the system to treat and the molecular variables to analyse. Accurate representations of electronic properties (e.g., for studying catalysis, any formation of bonds or spectroscopic simulations) are resolved by quantum mechanical based approaches (QM). Large conformational explorations (e.g., conformational transitions of a protein) are generally solved based on classical molecular mechanics methods (MM) that are, for systems of similar size, far more rapid than QM approaches. From a biochemical point of view, though, the counterpart of QM accuracy is to limit the study to systems of small dimensions. Hybrid methods like QM/MM can partially solve the problem by allowing “full size”

[a] *Insilichem, Department of Chemistry, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain. E-mails: agusti@klingon.uab.cat, jeandidier.marechal@uab.cat*
[§] *Current address: Institute of Chemical Research of Catalonia (ICIQ), The Barcelona Institute of Science and Technology, 43007 Tarragona, Spain*
[+] *Current address: School of Chemistry, Trinity College Dublin, College Green, Dublin, 2, Ireland*
‡ These authors have contributed equally to this work
Electronic Supplementary Information (ESI) available: supplementary scheme, supplementary tables, supplementary figures, details of computational methods, Gaudi input files, GPathFinder input files, cartesian coordinates obtained from DFT calculations. See DOI: 10.1039/x0xx00000x

modelling although the conformational sampling remains limited.[12] Modelers need to balance between accuracy, sampling and computational power to select the best method for a given study. In some cases, they cannot rely on one unique method but need to combine several. These multi-level pipelines can be necessary when looking for the origin of the enantioselective profiles of metalloenzymes for which it may need to identify the best arrangement of protein and substrate for the reaction to occur and simulate the transition paths between reactant and products. Prototypical systems with such multi-scale demands are artificial metalloenzymes (ArMs), a family of new-to-Nature biocatalysts obtained by the insertion of homogeneous catalysts into proteic hosts.[13,14]

Artificial metalloenzymes have received increasing attention in the last decades because of several advantages with respect to other *de novo* enzyme designs. This includes: 1) a long list of homogeneous catalysts and protein scaffolds available as demonstrated by more than 40 different artificial metalloenzymes reported to date,[15] 2) the homogeneous catalyst brings a stable and well-defined first coordination sphere of the metal and avoids creating from scratch a metal binding site based purely on natural amino acids. 3). The protein can be handled with the same protein engineering techniques as other enzymes (e.g., mutagenesis, directed evolution, etc.) to control substrate specificity and enantioselectivity. In many aspects, molecular modelling can be particularly interesting to shed light on ArMs mechanisms and help in their design.[13,14] What makes them more challenging though is the lack of structural information on the orientation and/or location of the cofactor in the protein host and how this relates to other aspects of the catalysis (orientation of the substrate, impact on the first coordination sphere of the metal, etc.). In this area, molecular modelling can be particularly informative.

The way our group has been dealing with the computational study of artificial metalloenzymes stands on multi-scale protocols.[16] Those protocols are generally three folded: **1.** They start with QM studies on models with reduced dimensions of the catalytic centre. They tend to include the first coordination sphere of the metal with a prototypical substrate, the cofactor and eventually few candidate amino acids in the vicinity of the putative active site. The presence of metals in these systems ultimately implies that the quantum level should reach at least density functional theory (DFT). **2.** Then, key structures extracted from the step **1** (i.e., pre-catalytic michaelis complexes, transition states, reactive intermediates, etc.) are incorporated in the protein binding site by protein-ligand dockings. This part has required developments of optimized strategies to predict metal-protein interactions.[17–19] **3**. The best models from **2** are consequently refined by classical Molecular Dynamics and/or QM(/MM) approaches to finally identify transition states and reactive paths inside the protein scaffold. Although tedious to set up, these multi-scale approaches have been efficient in decoding the mechanisms and ultimately guide the design of several artificial metalloenzymes.

Molecular modelling has been particularly successful in predicting how mutations of the protein impacts on the enantioselective profiles of several ArMs. However, these mutations are generally close to the metal, and impact on its first or second coordination sphere which in turn can alter the electronic properties of the metal or the binding of the substrate.[20] Nonetheless, like in other enzymatic systems, some mutations further from the active site can also affect the catalytic profile. How long-range mutations impact on the enzyme activity is still controversial and can relate to numerous variables like electrostatic patterns, conformational dynamics of pre-catalytic structures, substrate diffusion, and any global motions of the protein that can influence cofactor and/or substrate binding.[21] These aspects are the most challenging to study with today's computational standards. One excellent example to assess how molecular modelling could deal with this problem is the cyclopropanase that Lewis and co-workers developed recently.[22,23]

The cyclopropane core is one of the most important strained rings in chemical intermediates and part of many natural compounds and drug molecules.[24,25] Its synthesis is intensively investigated with some striking successes obtained by enzyme engineering.[26] The artificial metalloenzymes developed by Lewis and co-workers catalyse the enantioselective cyclopropanation of styrene and result from the inclusion of a dirhodium homogeneous catalyst by covalent linkage (Figure 1C) into *Pyrococcus furiosus* prolyl oligopeptidase (POP). POP is a protein with a two-

domain architecture: a peptidase domain with an α/β-hydrolase fold capped by a seven-bladed β-propeller domain (Figure 1A).[27] In optimizing this scaffold, Lewis *et al.* played both with cofactor anchoring and amino acid mutations. The variants with the highest enantioselective profiles mutations are called (Figure 1B) GSH (with mutations Ser301Gly, Tyr326His and Gly99Ser, with 92% ee for the *S,R* enantiomer), HFF (with mutations Leu328His, Gly99Phe and Gly594Phe, with 92% ee for the *S,R* enantiomer) and RFY (with mutations Gln98Arg, Gly99Phe and Pro239Tyr, with 80% ee for the *R,S* enantiomer). In these variants the cofactor is linked to the protein by replacing Ser477 in GSH and HFF, and Phe413 in RFY. In their work, Lewis *et al.* suggest that the global motion of the protein scaffold could drive the changes in the enantioselective profile between the different mutants by altering the dynamics of the opening of the interdomain region. [27–29] Due to its scientific and methodological interest, we decided to apply our multiscale philosophy on this system, also integrating new approaches that can deal with collective motions and substrate diffusion. We believe this study could provide an interesting material in the context of this Faraday Discussion.

Computational methods

The methodology used in this work is a multi-scale computational pipeline. It is inspired by our previous work on artificial metalloenzymes,[16,30–33] although adapted for the particular problem of the dirhodium cyclopropanase. Up to 5 different methods are used in this workflow. All computational details are presented in the Supplementary information of this manuscript. Here, we provide the most important information and the general scheme (Figure 2). The overall workflow consists in:

- **Determination of the lowest energetic pathways** for the cyclopropanation of (4-methoxyphenyl)methyldiazoacetate (**1**) and styrene (**2**) mediated by the artificial cofactor of the enzyme Rh_2-$(OAc)_2$(esp-CO_2Ac). Calculations are performed in water model in absence of any contribution of the protein environment. This model consists of the dirhodium cofactor (Figure 1C) and the two substrates. Calculations were performed at the DFT level of theory with the Gaussian09 software, using the B3LYP functional combined with the Grimme's D3 correction for dispersion and the SMD continuum solvent model.[34–43] This study identifies that the determining steps of the cyclopropanation that could be influenced by the protein scaffold are two intermediate structures (namely **II** and **III**).

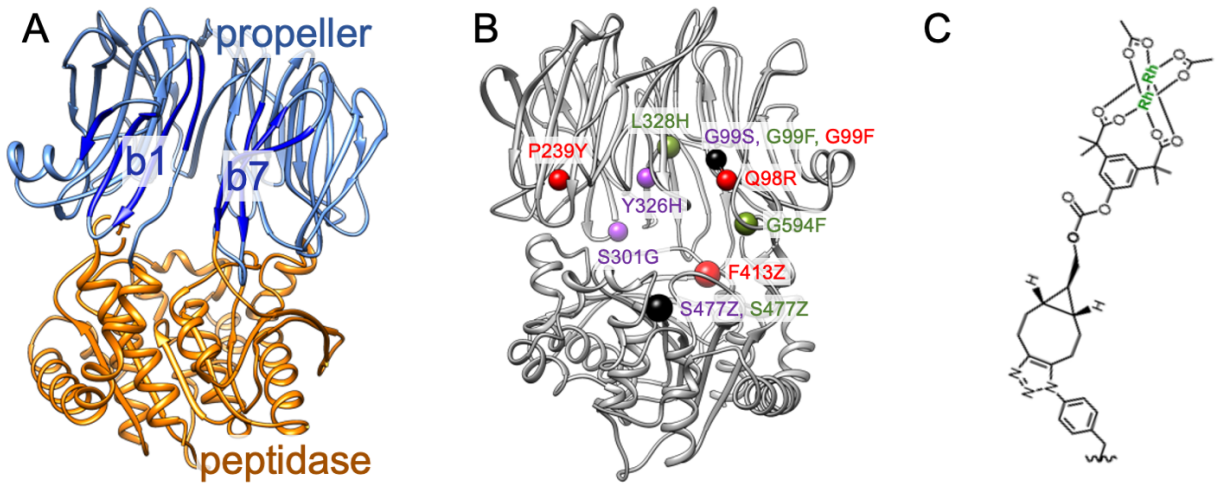


Figure 1. (A) POP protein scaffold (PDB code 6can[27]), propeller domain shown in blue and peptidase domain shown in orange. First and last blades of the propeller domain are highlighted in darker blue and labelled as b1 and b7, respectively. (B) Location of the mutations of the three variants studied here: GSH in purple (Ser301Gly, Tyr326His, Gly99Ser, Ser477Z), HFF in green (Leu328His, Gly99Phe, Gly594Phe, Ser477Z) and RFY in red (Gln98Arg, Gly99Phe, Pro239Tyr, Phe413Z); mutations in the pore of the propeller domain are not shown for clarity. (C) Structure of the dirhodium cofactor (named Z in the mutation scheme).

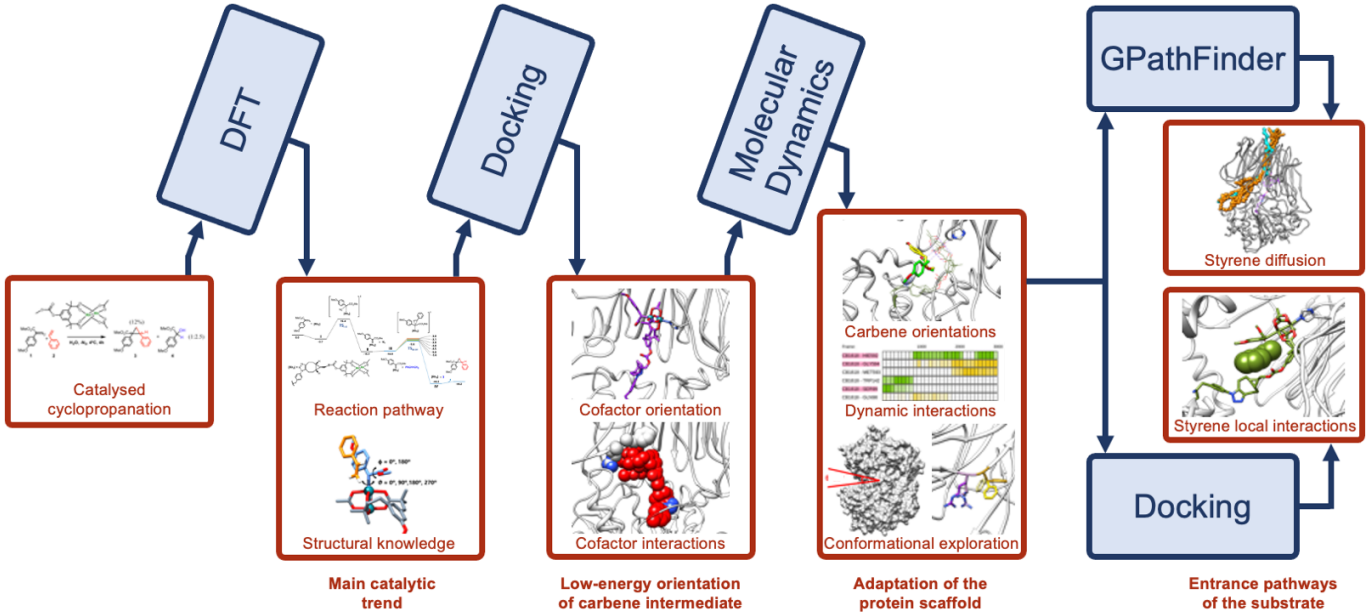


Figure 2. Computational pipeline employed in the study of the POP-Rh2 cyclopropanase. Methods are shown in blue, while inputs and outputs of each stage are shown in dark red.

- **Docking of the intermediates to POP**. First, we carried out an exploratory docking with the multi-objective genetic algorithm GaudiMM,[17,19] to assess whether the dirhodium cofactor has the possibility to form a coordination bond with an amino acid donor in the vacant site of the Rh cofactor. Dockings were performed on a series of POP structures with wide global conformational changes as extracted from the Protein Data Bank (codes 5t88 and 6can[27]) and displacements along low-energy normal modes. This allowed us to efficiently explore the large conformational space of the protein.[44–47] Second, a refinement docking was performed on the best-scored structures to evaluate the affinities, relative orientations and interactions of the Rh_2-cofactor into the biohybrid binding pockets. For this stage, we used the Gold5.8 software[48] with our recent version of GoldScore accounting for binding with coordination.[18,49] As a result, the best structures of each variant were selected as initial frames for the further Molecular Dynamics (MD) simulations.
- **Molecular dynamics simulations on the carbene intermediate.** The system was solvated with a cubic box of pre-equilibrated water molecules. The AMBER14SB force field[46] was used for the standard residues, while the GAFF force field was adopted for the remaining atoms. Rh-bonding force constants and equilibrium parameters were obtained through the Seminario's method.[34,50–52] The convergence of the trajectories was assessed by RMSD, all-to-all RMSD and cluster counting analyses,[53,54] considering all the atoms of the cofactor and the carbene (Figures S1-S3). In light of the analyses, it was decided to extend the MD simulations under periodic boundary conditions[55] up to 300 ns, 400 ns and 200 ns for GSH, HFF and RFY variants, respectively. This step allowed us to obtain a pool of representative orientations of the carbene for each POP variant, as well as analyses[56–59] of the relevant carbene-protein interactions and the POP scaffold dynamics.
- **Simulation of the styrene diffusion pathways.** Calculations to assess the binding route of the styrene and the possibility to arrive either to *Re* or *Si* faces of the carbene were carried out with GPathFinder.[60] Each representative orientation of the carbene (obtained in the previous MD section) was tested for the possibility to bind the styrene at its *Re* and *Si* faces by ubicating the final point of the binding route at the corresponding coordinates. The resulting binding pathways, further validated with Non-Covalent Interaction (NCI) analyses,[61,62] allowed to rationalize the influence of the styrene binding route in the enantioselectivity of the different POP variants.

- **Docking of the styrene.** For the HFF variant, we had to extend the work to evaluate the interactions of the intermediate **III** in the catalytic site. We employed a procedure analogous to that for the intermediate **II** docking. This two-step docking study (exploratory with GaudiMM[17,19,63] and refinement with Gold5.8[18,48,49]) allowed to rationalize the enantiomeric preference of the HFF variant.

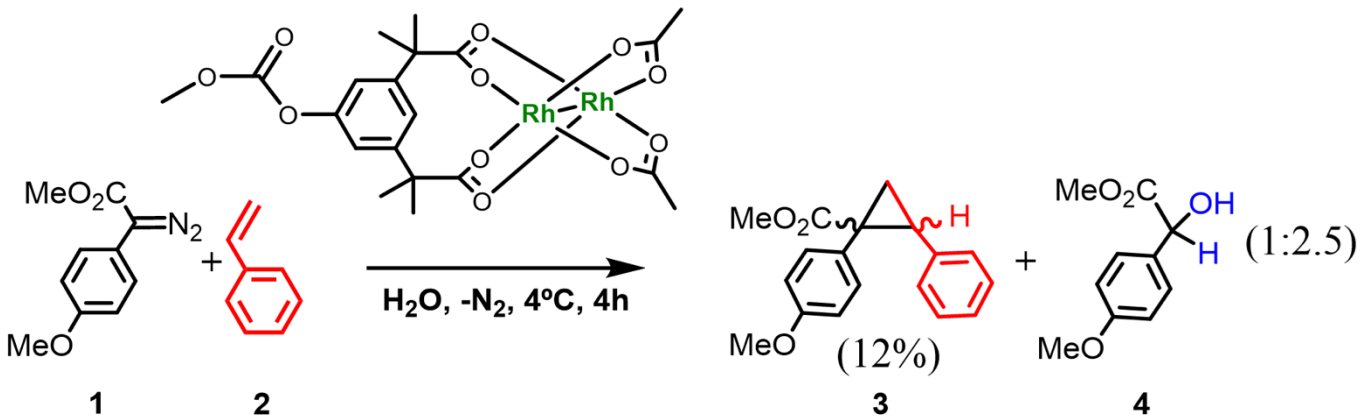
Results and discussion

Determinant steps of the cyclopropanation

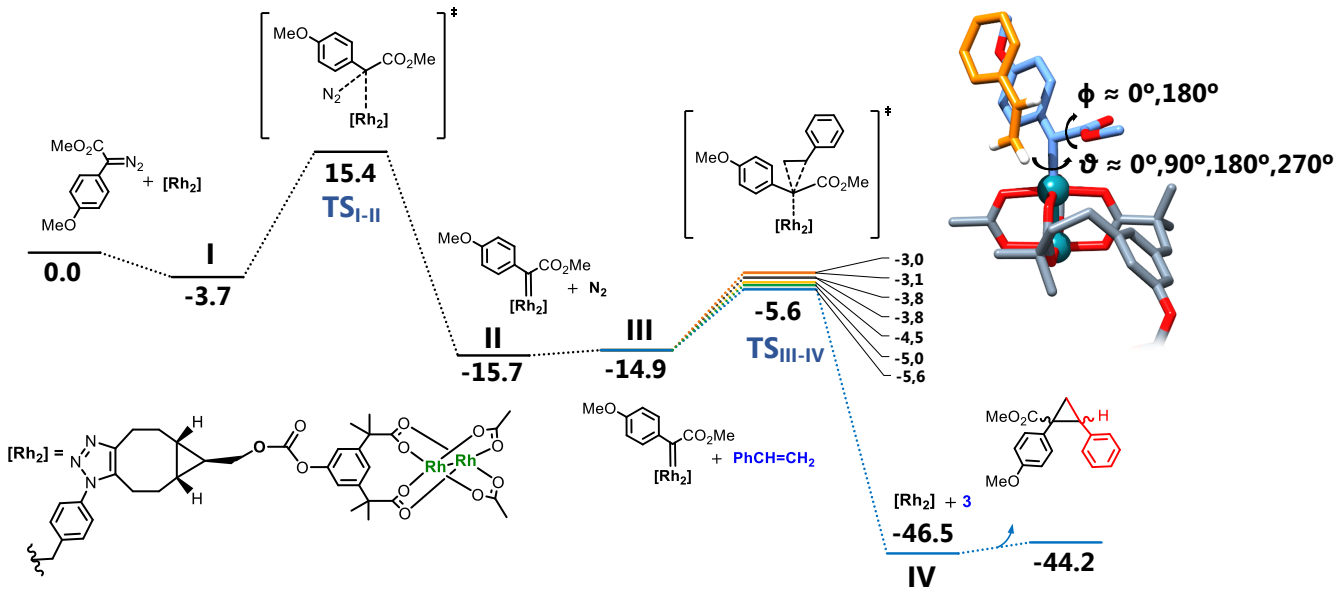
With the aim to define the main features of the catalytic events, a full DFT (B3LYP-D3) analysis of the potential energy surface has been carried out for the isolated Rh_2-cofactor mediated cyclopropanation of (4-methoxyphenyl)methyldiazoacetate (**1**) and styrene (**2**) in water (Scheme 1).

The general Rh_2-cofactor catalysed cyclopropanation mechanism is the result of two subsequent events: i) carbene formation through the Rh_2-mediated N_2 extrusion from the co-substrate **1**; and ii) approach of the styrene to the carbene intermediate with concomitant cyclopropanation of its double bond. The second step, depending on the face which the styrene approaches the co-substrate, is responsible for the stereo-selectivity of the process. The complete profile is depicted in Scheme 2 and it is in general agreement with the previous work by Autschbach and co-workers.[64]

The carbene formation and concerted N_2 extrusion display an activation barrier of 19.1 kcal·mol^{-1} (**TS_{I-II}**), which results as the rate determining step of the reaction. The reaction proceeds falling at -15.7 kcal·mol^{-1} to intermediate **II** (rhodium carbenoid). The styrene approach requires ~1.0 kcal·mol^{-1} forming the supramolecular aggregate **III** (rhodium carbenoid-styrene)**.** It must be highlighted that intermediate **II** shows eight different energetically accessible conformations deriving from the C_2 axis rotation of the carbene and on the methylester flip giving four and two orientations, respectively (Scheme 2, angles ϑ and ϕ). The enantiodetermining and final step is the CC bond formation by cyclopropanation upon approach of the styrene to one of the faces of the plane defined by the (4-methoxyphenyl)methyldiazoacetate. In the case of the isolated achiral cofactor, the symmetric environment prevents any enantiomeric preference. The eight possible transition states display energy barriers from intermediate **II** + separated styrene ranging from 10.1 up to 12.7 kcal·mol^{-1} (Scheme 2).



Scheme 1. Rh_2-$(OAc)_2$(esp-CO_2Ac)-mediated cyclopropanation of (4-methoxyphenyl)methyldiazoacetate (**1**) and styrene (**2**) in water.[23]



Scheme 2. Gibbs energy profile for the reaction of the complete cofactor in water. All Gibbs energy values are given in kcal·mol^{-1}. For intermediates **II** and **III** only the lower energy conformations are reported for clarity.

Considering the low yield of the reaction in aqueous media (Scheme 1), we further expanded our analysis to the competitive reaction (carbene insertion into the O-H bond of water), where a water molecule acts as a nucleophile toward the coordinated carbene forming the α-hydroxyester byproduct (**4**). From intermediate **III'** -the supramolecular adduct analogue to **III** with H_2O instead of styrene- the water nucleophilic attack gives rise to intermediate **IV'**. The following 1,3 proton migration leads to the 1-methoxy-2-phenyl-1,2-ethandiol, evolving into the **4** byproduct throughout a keto-enolic tautomerism (Scheme S1). The nucleophilic attack was particularly sensitive to microsolvation effects. Increasing the number of explicit H_2O molecules from one to three, the energy barriers decrease by 3.2 kcal·mol^{-1}. Although the side reaction barrier results higher by 1.7 kcal·mol^{-1} compared to the main process, a full solvated model would be necessary to properly reproduce the experimental yield. Nevertheless, this is out of the scope of this work for which the obtained qualitative trend is enough to define the role of microsolvation: a less microsolvated environment would prevent the formation of the byproduct.

Carbene intermediate inside the POP protein

As seen in the DFT study, the formation of the intermediate **II** (cofactor + carbene adduct) is the rate determining step and slowest part of the reaction. Therefore, most of the reaction time is required for the carbene formation, whereas the cyclopropanation occurs almost instantaneously. With that in mind, we modelled the structure of the POP protein + intermediate **II**. First, by employing docking techniques, we focused on the possibility of finding a coordination bond between a Rh atom of the cofactor and an amino acid of the POP protein. The existence of a Rh-His coordination bond was previously hypothesised in the case of the GSH and HFF variants.[22,23,29] Its presence would confer greater stability to the structure and could be a factor to justify the enantiospecificity of the different variants of the enzyme.

The POP protein is known to explore a wide range of the interdomain angle between the propeller and peptidase domains (named θ),[27] making it unavoidable to take this into consideration in the modelling protocol. Also, the cofactor flexibility and the different intermediate **II** conformations should be considered. To achieve so, we applied a two-step docking protocol. In a first exploratory step with the GaudiMM software,[19] we considered several structures of the POP protein covering the whole range of θ (Figure 3), together with allowing full flexibility to the intermediate **II**, the candidate sidechains to form a coordination bond and the mutated amino acids of each variant (see Computational methods for further details). This first exploration discarded the possibility of a Rh-POP

coordination bond for the RFY variant, whereas suggested the feasibility of His326-Rh for the GSH variant and His328-Rh for the HFF variant (noticeably in both cases are implicated one of the mutated amino acids of the variant). In addition, we observed that coordination in the GSH variant is only possible for intermediate values of θ (29º and 32º, see Figure 3), whereas for the HFF case the "closed" conformation is also viable. This suggests that the presence of the His-Rh bond constrains the conformations allowed for the protein scaffold.

The best solutions obtained from the exploratory docking were submitted to a further refinement docking procedure using our recently updated version of GoldScore and providing fine evaluation of metal-amino acid interactions.[18,48] While the previous exploratory step allowed to test mainly the geometric feasibility, the main objective in this refinement stage was to obtain the best possible structures in terms of an energy-based score. The His326-Rh and His328-Rh coordination were confirmed for the GSH and HFF variants, whereas interactions between the intermediate **II** and Arg98/Phe99 (two of the mutated amino acids) were found for the RFY case among the best-scored solutions (Figure 4).

The location and relative orientation of intermediate **II** inside POP could influence the enantiodetermining step of the reaction. We therefore performed first classical Molecular Dynamics simulations on the three variants loaded with intermediate **II**. The simulations started from the best structures obtained in the previous docking stage (GSH

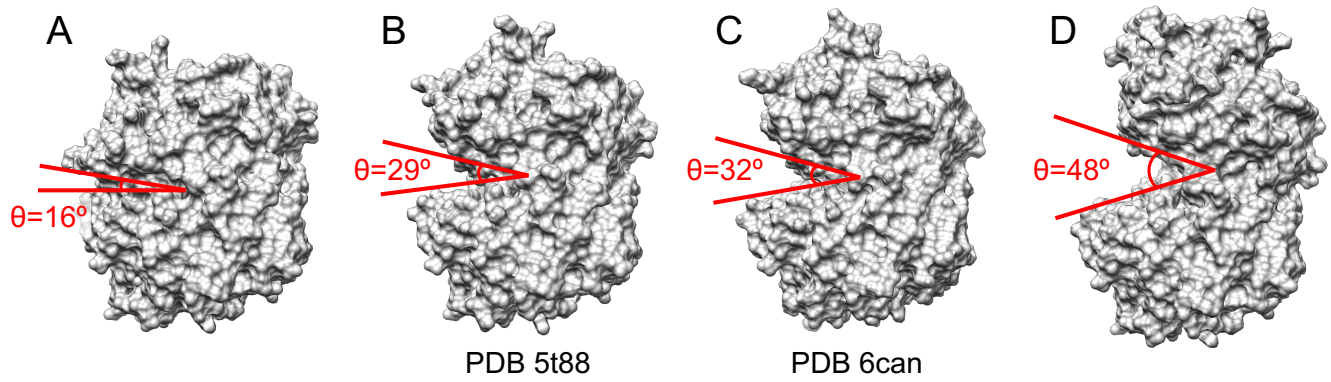


Figure 3. Different conformations of the POP protein (shown in surface representation) used in the first exploratory docking experiment, allowing to cover the whole range of POP interdomain angles. Interdomain angle θ (as defined in [27]) is indicated in each case. (A) Closed conformation (θ = 16º), generated by displacements along low-energy normal modes on the PDB code 5t88.[27] (B) θ = 29º, PDB code 5t88. [27] (C) θ = 32º, PDB code 6can.[27] (D) Open conformation (θ = 48º), generated by displacements along low-energy normal modes on the PDB code 5t88.[27]

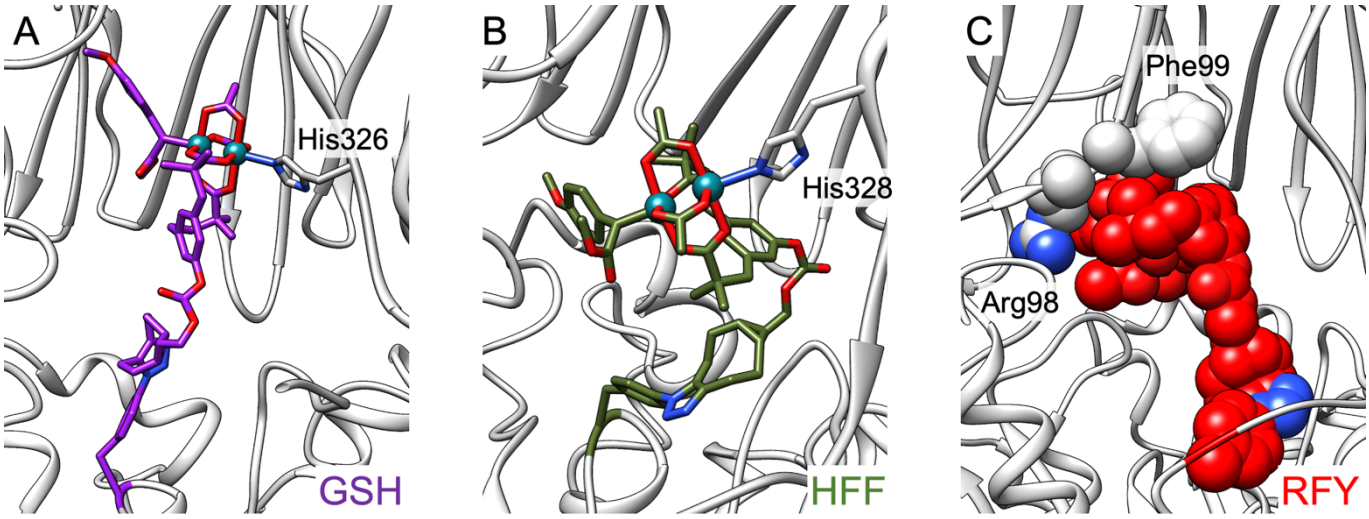


Figure 4. Best-scored structures obtained from the refinement docking stage. (A) GSH variant. Intermediate **II** is shown in purple sticks, while His326 is highlighted in grey sticks. (B) HFF variant.

and HFF with Rh-His coordination and RFY without Rh coordination bond). Along the simulations, changes in the relative orientation of the carbene to the cofactor are observed in all three variants. A clustering study taking into consideration the carbene atoms revealed three representative orientations of the carbene for the GSH variant, whereas each HFF and RFY variants account for two representative structures (Figure 5). The residues interacting with the carbene along the simulations are also dependent on the variant (Tables S1-S3). In the GSH case, the carbene mainly interacts with amino acids ubicated at blades b1, b2 and b3 of the propeller domain, including the mutated Ser99, as well as with some residues of the flexible loop at the peptidase/barrel interface. For the HFF, however, the residues interacting with the carbene are mainly located at the loop linking the third and fourth beta strands of the peptidase domain. Interestingly, in the GSH and HFF variants, the cofactor is anchored at the same residue 477 but the Rh ions coordinate different histidine residues (His326 for GSH and His328 for HFF). This drastically impacts on location of the carbene in the binding site. Although the cofactor is at position 413 in RFY, in this variant, the carbene interacts with residues ubicated at blades b7, b1 and b2 of the propeller domain; in a quite similar manner than for the GSH variant. This shows the importance of the second coordination sphere of the metal to define cofactor and substrate location and orientation. In agreement with the docking results, the interaction of the carbene with Arg98 and Phe99 is maintained, suggesting that these two mutations of the RFY variant are mainly involved in stabilizing the intermediate **II**.

Regarding the analysis of the protein scaffold dynamics, the interdomain angle θ remains stable in the GSH variant (standard deviation of 1.5º), whereas is more flexible for the HFF and RFY variants with a standard deviation of 3.7º and 4.3º, respectively (Table 1, Figure S4). In all cases, the range of θ observed in the simulations is narrower than in a previous computational study[27] on the *apo* POP, where a range from ~10º to ~40º was reported. Our findings are supported by a very recent study[29] on a novel enzyme derived from the GSH variant that catalyses diazo cross-coupling. In this study, the MD simulations show a wider range of θ for the *apo* (i.e. without cofactor) structure than for the *holo* (i.e. with the cofactor) ones (Table 1). Altogether both studies show that the metallic cofactor constrains the POP scaffold and favours tighter conformations of the protein.

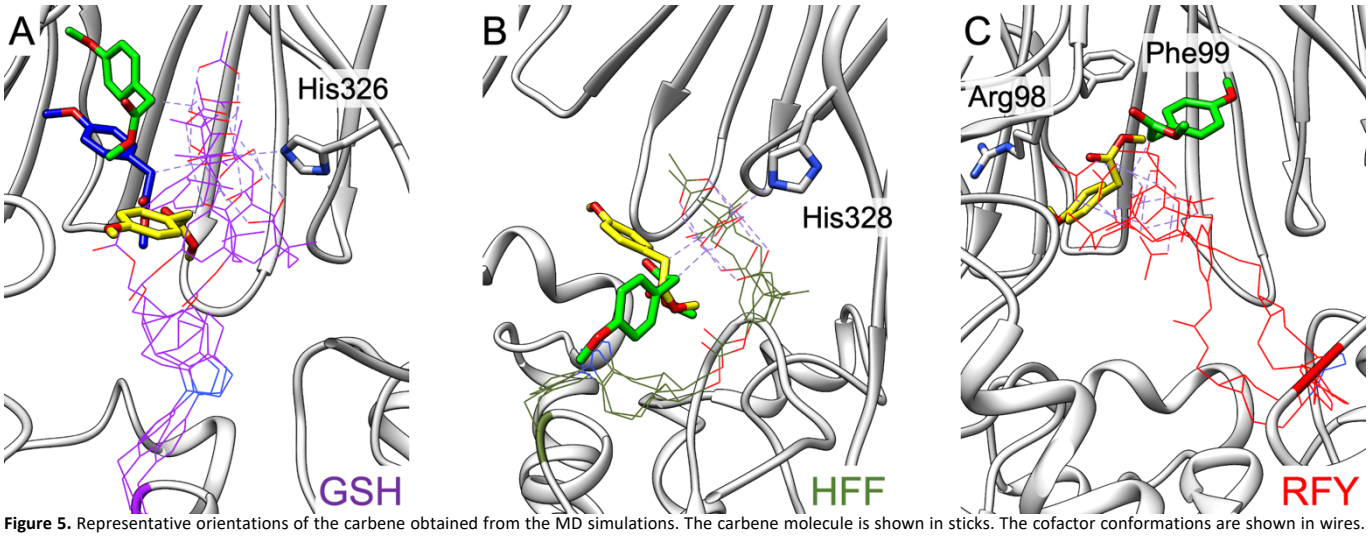


Figure 5. Representative orientations of the carbene obtained from the MD simulations. The carbene molecule is shown in sticks. The cofactor conformations are shown in wires. Only one conformation of the POP protein backbone is shown in grey ribbon for clarity. (A) GSH variant, three representative carbene orientations, His326 is highlighted in grey sticks. (B) HFF variant, two representative carbene orientations, His328 is highlighted in grey sticks. (C) RFY variant, two representative carbene orientations, Arg98 and Phe99 are highlighted in grey sticks.

Table 1. Values of the interdomain angle θ along the MD simulations of this study and the ones reported in [29]

Variant	Mean and standard deviation of θ (degrees)	Maximum of θ (degrees)	Minimum of θ (degrees)
GSH with Rh-His	24.6 ± 1.5	30.1	19.8
HFF with Rh-His	23.9 ± 3.7	30.0	11.3
RFY no Rh-His	25.7 ± 4.3	36.3	17.4
5-G apo[29, a]	Not reported	~39	~10
5-G no Rh-His[29, b]	Not reported	~30	~17
5-G with Rh-His[29, c]	Not reported	~25	~13

[a] Enzyme without the dirhodium cofactor

[b] Enzyme with the dirhodium cofactor and without a Rh-His coordination bond

[c] Enzyme with the dirhodium cofactor and a Rh-His coordination bond

Catalytically competent geometries of styrene

Diffusion of the styrene (2) is key in GSH and RFY variants. Once the intermediate **II** is formed, the next step in the reaction pathway is the approach of the styrene to the carbene to form the intermediate **III**. To this step to success, the styrene should find a binding route from the solvent to carbene's site. What faces of the carbene (*Re*, *Si* or both) are accessible for the styrene would ultimately determine the chirality of the product. Two factors may influence the binding routes of the styrene. First, the conformations and orientations of the intermediate **II**, which limit the space available inside the binding pocket of the enzyme. Second, the POP itself, which presents two possible entrance channels,[27] the first one through a pore located at the propeller domain and the other through the interdomain region (Figure S5). A preference for the interdomain entrance has been reported for non-modified POPs,[65] but the access through the pore cannot be discarded depending on the substrate or the presence of the cofactor. With the uncertainty about the entrance channel and conformational variability of POP in mind, we opted for an extensive sampling with our ligand pathway simulator GPathFinder.[60] GPathFinder is based in our multi-objective genetic algorithm (GA) GaudiMM and models the entrance of chemicals into proteins using a conformational selection approach. At the starting of the simulation, the GA is fed with representative structures of the protein that are extracted from a molecular dynamics run or displacements along collective motions (e.g., from Normal Mode analysis). During the evaluation process, the best solutions of successive displacements of the ligand are collected and classified to form continuous diffusion paths. Because of the multi-objective philosophy of GPathFinder, different evaluation functions can be used, very wide sampling can be performed, and channelling pathways found without any previous knowledge. It is a very interesting alternative or complementary tool to steered- or meta- dynamics approaches that require previous knowledge about the channel and much higher computational resources.

Here, we sampled the styrene binding routes from the solvent to the *Re* and *Si* faces of the carbene for each orientation of the carbene obtained in the MD simulations of the previous step. GPathFinder was instructed to optimise two objectives along the styrene binding process; the steric clashes and vina score.[66] For every representative orientation of the intermediate **III,** 240 possible binding routes were obtained. Their quality was analysed based on both objectives. The analysis for the clashes was done with standard filtering (threshold of 100 $Å^3$[3,19,60]) and post-processing the best pathways with the QM-based method to analyse noncovalent interactions in biomolecules NCIPLOT4[62] (Figure 6). Interestingly, in all three variants, routes for the styrene were found through the interdomain entrance and the pore channel indicating that both routes are possible. However, striking differences are observed between variants. For the GSH, GPathFinder shows a preference for routes leading to *S,R*

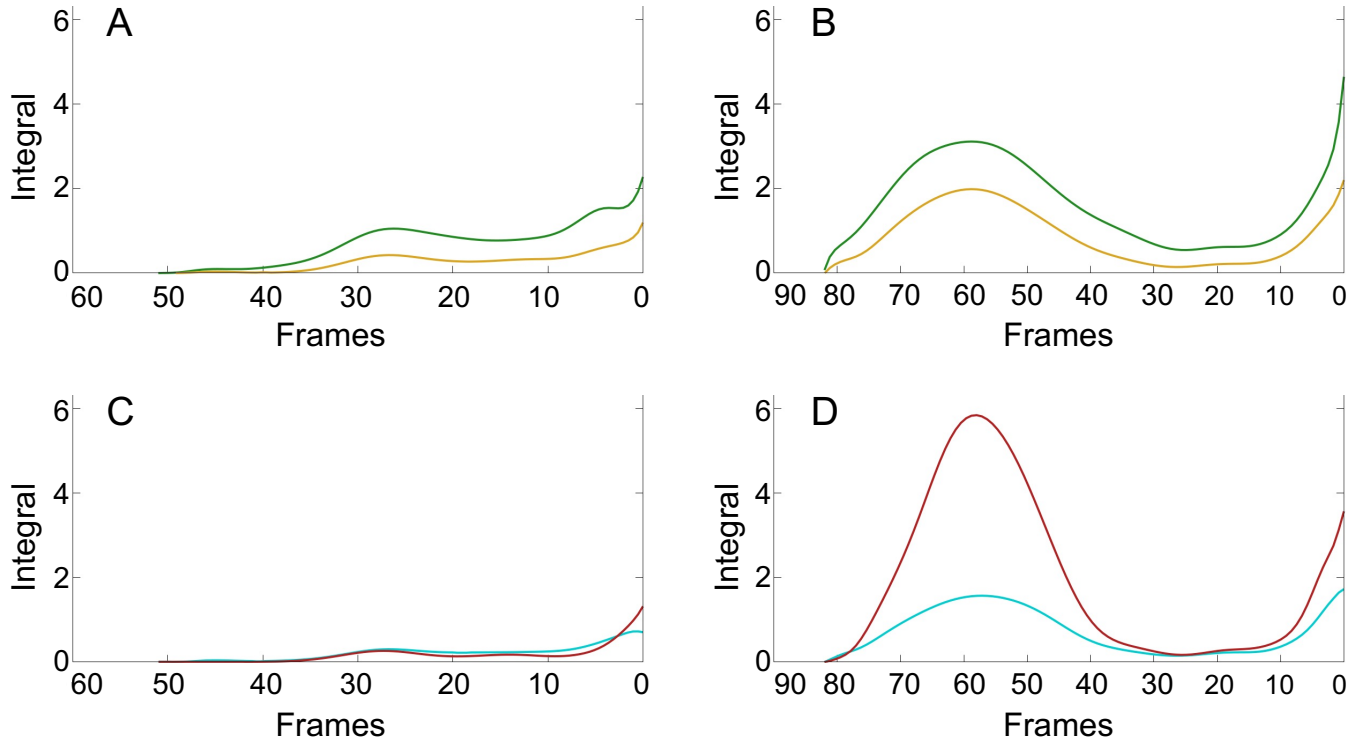


Figure 6. Example of NCIPLOT analyses over a valid and an invalid binding route for the styrene. All plots show the integral of the interactions for the whole trajectory of the styrene, from the solvent (highest frame number) to the binding site (frame number 0). (A) Example of valid route. The module and sign of Van der Waals interactions are shown in green and yellow, respectively. (B) Example of invalid route. The module and sign of Van der Waals interactions are shown in green and yellow, respectively. (C) Example of valid route. Attractive and repulsive interactions are shown in blue and red, respectively. (D) Example of invalid route. Attractive and repulsive interactions are shown in blue and red, respectively.

enantiomer catalytically competent geometries (64% for *S,R* and 36% for *R,S*), in agreement with the experimental observation (Figure 7A). For the RFY variant, the trend is almost completely inverted with 66% leading to catalytically competent geometries with the *R,S* product and 34% for the *S,R* enantiomer, also in agreement with experiments (Figure 7C). However, no such clear tendency was observed for the HFF variant, with much more equilibrated values (54% for *S,R* and 46% for *R,S*; Figure 7B). Similar predictions are reached with the analyses of the vina score objective. Considering the maximum vina score of every binding route, which can be understood as a raw estimation of the barrier that the styrene must overcome in its binding process, (Table 2) it appears that: i) in the GSH variant, routes leading to *S,R* enantiomer are favoured by a difference of ~17 kcal/mol; ii) in the RFY variant, routes leading to *R,S* enantiomer are favoured by a difference of ~12 kcal/mol; and iii) in the HFF variant, the average of the vina-score barriers is similar for the *S,R* and *R,S* enantiomers (difference of ~1 kcal/mol). Looking at the number of best-scored routes (those with vina score < 0 kcal/mol along all the trajectory) instead of the average scores of the whole pool of solutions, the same conclusion is obtained.

Altogether, the modelling of the styrene pathways shows that the GSH and RFY mutants, both coming from a directed evolution approach, apparently gain their enantiospecificity by controlling substrate channelling and consequently the pre-organization of the binding site towards different isomeric catalysis. However, the HFF mutant, which comes from a former series of structure-based optimization, does not produce such differential behaviour. Other molecular reasons have to be responsible for the catalytic preference in this case.

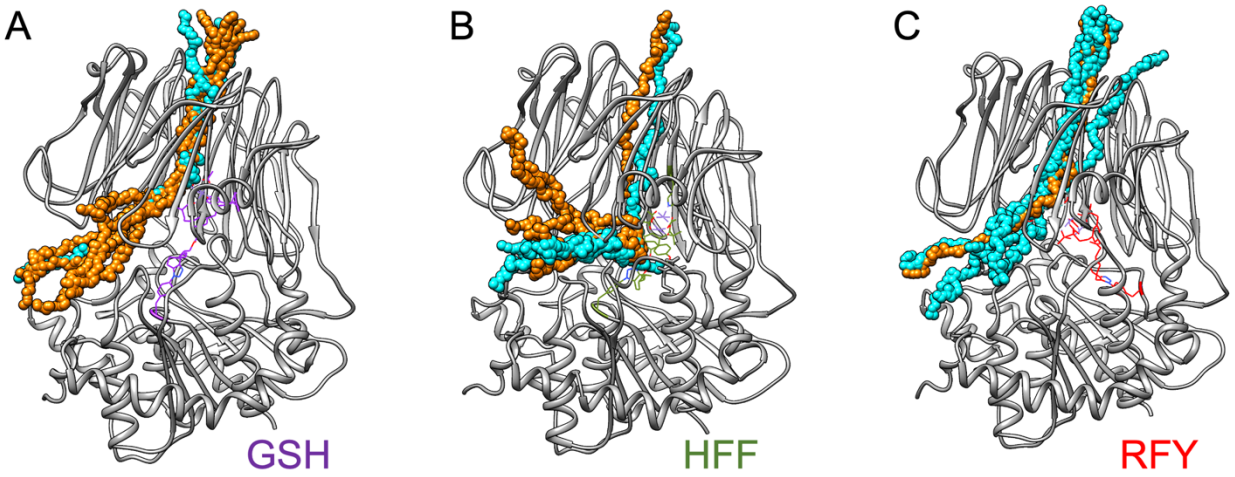


Figure 7. Schematic representation of the best-scored styrene binding routes. Trajectories favouring the *S,R* enantiomer are shown in orange, whereas trajectories favouring the *R,S* enantiomer are shown in cyan. Only one conformation of the POP protein backbone is shown in grey ribbon for clarity. (A) GSH variant. Intermediate **III** and His326 are highlighted in purple wires. (B) HFF variant. Intermediate **III** and His328 are highlighted in green wires. (C) RFY variant. Intermediate **III** is highlighted in red wires.

Table 2. Average of vina-score barriers and number of best solutions for the styrene binding routes in the three studied POP variants.

Variant	Mean and standard deviation (all solutions[a]) of vina-score barrier for *S,R* enantiomer (kcal/mol)	Mean and standard deviation (all solutions[b]) of vina-score barrier for *R,S* enantiomer (kcal/mol)	Difference of vina-score barriers (*S,R* - *R,S* for the mean of all solutions[b]) (kcal/mol)	Number of best-scored solutions[c] for *S,R* enantiomer	Number of best-scored solutions[c] for *R,S* enantiomer
GSH	24.1 ± 21.7	41.0 ± 25.9	-16.9	32	0
HFF	37.7 ± 18.8	38.8 ± 22.2	-1.1	0	0
RFY	36.4 ± 19.5	24.1 ± 18.5	12.4	0	14

[a] Average and standard deviation calculated for 720 solutions.

[b] Average and standard deviation calculated for 480 solutions.

[b] Number of solutions with a vina score < 0 kcal/mol along all the route.

Local interactions in the active site are key in HFF variant. We hypothesized that a more local network of interactions could explain the enantioselective profile of the HFF variant. To check this assumption, we used a two-step docking protocol similar to the procedure employed with the intermediate **II**: a first exploratory docking considering all the flexibility shown in the MD study, followed by GOLD[48] calculations on the most promising combinations to obtain reliable energetic scores with GoldScore for metals.[18,49] The results of the docking experiments (Table S4) show a preference for catalytically competent geometries for the *S,R* enantiomer, with a difference of GoldScore of ~4, accounting for the experimental enantiomeric excess. Interactions between the styrene and amino acids of the blades b3 and b4 (Asn143, Phe156, Ser196) were found in the case of best-scored pose leading to the *S,R* enantiomer, whereas interactions between the styrene and the cofactor were dominant in the best-scored pose leading to the *R,S* enantiomer. As both structures were found almost equally accessible for the styrene from the solvent, the enantiospecificity in the case of the HFF mutant is apparently more related to local interactions in the active site which makes the substrate to preferentially adopt a good orientation to produce the *S,R* enantiomer.

As a final remark regarding the search of catalytically competent geometries of the styrene, our findings show that the study of interactions away from the catalytic site is sometimes necessary to ascertain key aspects of the reaction (in this case, the determinants of enantioselectivity). An analysis of the substrate binding process could become unavoidable in those cases. However, different methodologies could be used depending on whether the binding channel is known beforehand or not. Indeed, massive conformational exploration is required when no clear route has been identified while more expensive (*ergo* accurate) methods could be used for a specific channel and, in the best-case scenario, lead to energetic magnitudes.

Conclusions

Mutagenic profiles can influence biocatalytic activities in a variety of ways. In metalloenzymes, mutations can impact on the first coordination sphere of the metal and alter, for example, its electronic features. They can also modify substrate binding affinity or how the substrate and cofactors organize for pre-catalytic geometries. The study of the effects of protein mutations on the first and second coordination spheres of the metal is something quite standard in nowadays computational bioinorganics. They can be analysed by the combination of computational methods like QM or classical MD. However, the organization of the pre-catalytic geometries may not be only due to the vicinity of the metal but the result of a cascade of more global events. Such behaviours are still challenging to identify by computational means.

Here, we choose the di-rhodium POP cyclopropanase from Lewis and co-workers as a case study because the most interesting variants in terms of enantioselective profiles, namely GSH (ee 92% *S,R*), HFF (ee. 92% *S,R*) and RFY (ee 80% *R,S*) have mutations far from the metallic site. We got a better picture of their mechanisms thanks to the incorporation of a ligand pathway modelling tool into a multi-scale framework and could hypothesize on the molecular role of long-range mutations.

From the start, cluster models studied with DFT calculations show that the carbene and the cyclopropanoid formations have barriers about ~ 19 and ~ 10 kcal/mol respectively. The DFT intermediate structures were docked *a posteriori* into the protein scaffold and refined by classical molecular dynamics. This part of the modelling process shows that some residues mutated in the variants have a direct interaction with the cofactor and/or the substrate either in form of a coordination bond (His326 and His328 for GSH and HFF, respectively) or a strong hydrophobic patch between Phe99 and Arg98 and the carbene in RFY. Differences between variants also highlight global re-arrangements of the protein scaffold, in particular regarding the interdomain region. Based on the hypothesis that those motions could affect substrate (styrene) binding, we performed GPathFinder calculations in the three variants and analysed their propensity to lead to *S,R* and *R,S* pre-reactive geometries. This hypothesis appears valid for GSH and RFY, where the enantioselective trends agree with those observed experimentally. For HFF, though, styrene binding pathways are not explicative and additional dockings were necessary to show that the experimental ee may relate to the organization of the second coordination sphere.

Multi-scale modelling applied to ArM has shown massive successes in the last decade although some limitations still exist. These procedures work by sustaining that the best solutions at one level are the most appropriate for the next level. By combining different methods, it is not possible to provide a unique energetic framework for the entire study. It could, moreover, in the worst-case scenario, stay blind when the protein scaffold alters drastically the mechanism of the first coordination sphere of the metal. Another challenge, common to other enzymatic systems, is to understand the impact of certain kind of mutations, especially those that are not in direct contact with the first coordination sphere of the metal. In this work, in the absence of clear identification of the substrate diffusion paths, we incorporated an atomistic model (GPathFinder) to explore the large conformational space related to ligand channelling. Strikingly, this approach shows that each variant provides better routes for a given enantiomeric pre-catalytic geometry although this may not be a general case (as shown for the HFF variant). Two challenges remain at least. One is to be able to assess the more efficiently possible those systems where substrate channels may be

affected by mutations from those who are not. Another relates to the modelling of substrate diffusion. Indeed, when no routes are known beforehand, massive conformational explorations are needed hence requiring methods that calculate rapidly the energy (i.e., those based on docking approaches). It is only with clear knowledge of the location of the diffusion path that one would eventually embark in accurate energetic methods although still expensive by today's standards. Researchers still have room for improvement in balancing between time and accuracy in their studies on how long-range mutations could impact catalytic profiles.

Author Contributions

Conceptualization J.-D. M. and A. L.; funding acquisition J.-D. M. and A. L.; supervision J.-D. M. and A. L; investigation J.-E. S.-A., G. S., E. M.-T., J.-D. M. and A. L.; visualization J.-E. S.-A., G. S. and E. M.-T., writing J.-E. S.-A., G. S., J.-D. M. and A. L.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank Prof. Jared C. Lewis for helpful discussions. Financial support from Spanish Ministry of Science and Innovation is acknowledged (MICINN project PID-2020-116861GB-I00).

Notes and references

1 M. Aminpour, C. Montemagno and J. A. Tuszynski, *Molecules*, 2019, **24**, 1693.
2 A. Romero-Rivera, M. Garcia-Borràs and S. Osuna, *Chem. Commun.*, 2017, **53**, 284–297.
3 M. Orio and D. A. Pantazis, *Chem. Commun.*, 2021, **57**, 3952–3974.
4 K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu and E. A. Pidko, *Chem. Rev.*, 2019, **119**, 2453–2523.
5 G. Sciortino, J.-D. Maréchal and E. Garribba, *Inorg. Chem. Front.*, 2021, **8**, 1951–1974.
6 W.-J. Wei, H.-X. Qian, W.-J. Wang and R.-Z. Liao, *Front. Chem.*, 2018, **6**, 638.
7 M. Q. E. Mubarak, E. F. Gérard, C. F. Blanford, S. Hay and S. P. de Visser, *ACS Catal.*, 2020, **10**, 14067–14079.
8 X. Sheng, M. Kazemi, F. Planas and F. Himo, *ACS Catal.*, 2020, **10**, 6430–6449.
9 K. Hiraga, P. Mejzlik, M. Marcisin, N. Vostrosablin, A. Gromek, J. Arnold, S. Wiewiora, R. Svarba, D. Prihoda, K. Clarova, O. Klempir, J. Navratil, O. Tupa, A. Vazquez-Otero, M. W. Walas, L. Holy, M. Spale, J. Kotowski, D. Dzamba, G. Temesi, J. H. Russell, N. M. Marshall, G. S. Murphy and D. A. Bitton, *ACS Synth. Biol.*, 2021, **10**, 357–370.
10 M. Kazemi, X. Sheng and F. Himo, *Chem Eur J*, 2019, **25**, 11945–11954.
11 S. Ahmadi, L. Barrios Herrera, M. Chehelamirani, J. Hostaš, S. Jalife and D. R. Salahub, *Int. J. Quantum Chem.*, 2018, **118**, e25558.
12 H. M. Senn and W. Thiel, *Angew. Chem. Int. Ed.*, 2009, **48**, 1198–1229.
13 V. Muñoz Robles, P. Vidossich, A. Lledós, T. R. Ward and J.-D. Maréchal, *ACS Catal.*, 2014, **4**, 833–842.
14 I. Drienovská, L. Alonso-Cotchico, P. Vidossich, A. Lledós, J.-D. Maréchal and G. Roelfes, *Chem. Sci.*, 2017, **8**, 7228–7235.
15 F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis and T. R. Ward, *Chem. Rev.*, 2018, **118**, 142–231.
16 L. Alonso-Cotchico, J. Rodríguez-Guerra, A. Lledós and J.-D. Maréchal, *Acc. Chem. Res.*, 2020, **53**, 896–905.
17 G. Sciortino, E. Garribba, J. Rodríguez-Guerra Pedregal and J.-D. Maréchal, *ACS Omega*, 2019, **4**, 3726–3731.
18 G. Sciortino, J. Rodríguez-Guerra Pedregal, A. Lledós, E. Garribba and J.-D. Maréchal, *J. Comput. Chem.*, 2018, **39**, 42–51.
19 J. Rodríguez-Guerra Pedregal, G. Sciortino, J. Guasp, M. Municoy and J.-D. Maréchal, *J. Comput. Chem.*, 2017, **38**, 2118–2126.
20 V. M. Robles, M. Dürrenberger, T. Heinisch, A. Lledós, T. Schirmer, T. R. Ward and J.-D. Maréchal, *J. Am. Chem. Soc.*, 2014, **136**, 15676–15683.
21 C. G. Acevedo-Rocha, A. Li, L. D'Amore, S. Hoebenreich, J. Sanchis, P. Lubrano, M. P. Ferla, M. Garcia-Borràs, S. Osuna and M. T. Reetz, *Nat. Commun.*, 2021, **12**, 1621.
22 P. Srivastava, H. Yang, K. Ellis-Guardiola and J. C. Lewis, *Nat. Commun.*, 2015, **6**, 7789.
23 H. Yang, A. M. Swartz, H. J. Park, P. Srivastava, K. Ellis-Guardiola, D. M. Upp, G. Lee, K. Belsare, Y. Gu, C. Zhang, R. E. Moellering and J. C. Lewis, *Nat. Chem.*, 2018, **10**, 318–324.
24 D. Y.-K. Chen, R. H. Pouwer and J.-A. Richard, *Chem. Soc. Rev.*, 2012, **41**, 4631.
25 T. T. Talele, *J. Med. Chem.*, 2016, **59**, 8712–8756.
26 P. S. Coelho, E. M. Brustad, A. Kannan and F. H. Arnold, *Science*, 2013, **339**, 307–310.
27 K. Ellis-Guardiola, H. Rui, R. L. Beckner, P. Srivastava, N. Sukumar, B. Roux and J. C. Lewis, *Biochemistry*, 2019, **58**, 1616–1626.
28 J. C. Lewis, *Acc. Chem. Res.*, 2019, **52**, 576–584.
29 D. M. Upp, R. Huang, Y. Li, M. J. Bultman, B. Roux and J. C. Lewis, *Angew. Chem. Int. Ed.*, 2021, anie.202107982.

30 V. Muñoz Robles, E. Ortega-Carrasco, L. Alonso-Cotchico, J. Rodriguez-Guerra, A. Lledós and J.-D. Maréchal, *ACS Catal.*, 2015, **5**, 2469–2480.
31 L. Villarino, K. E. Splan, E. Reddem, L. Alonso-Cotchico, C. Gutiérrez de Souza, A. Lledós, J.-D. Maréchal, A.-M. W. H. Thunnissen and G. Roelfes, *Angew. Chem. Int. Ed.*, 2018, **57**, 7785–7789.
32 L. Villarino, S. Chordia, L. Alonso-Cotchico, E. Reddem, Z. Zhou, A. M. W. H. Thunnissen, J.-D. Maréchal and G. Roelfes, *ACS Catal.*, 2020, **10**, 11783–11790.
33 F. Christoffel, N. V. Igareta, M. M. Pellizzoni, L. Tiessler-Sala, B. Lozhkin, D. C. Spiess, A. Lledós, J.-D. Maréchal, R. L. Peterson and T. R. Ward, *Nat. Catal.*, 2021, **4**, 643–653.
34 Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian Inc.*, Wallingford CT, 2011.
35 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
36 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.
37 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
38 M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654–3665.
39 D. Andrae, U. Häußermann, M. Dolg, H. Stoll and H. Preuß, *Theor. Chim. Acta*, 1990, **77**, 123–141.
40 A. W. Ehlers, M. Böhme, S. Dapprich, A. Gobbi, A. Höllwarth, V. Jonas, K. F. Köhler, R. Stegmann, A. Veldkamp and G. Frenking, *Chem. Phys. Lett.*, 1993, **208**, 111–114.
41 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
42 V. S. Bryantsev, M. S. Diallo and W. A. Goddard III, *J. Phys. Chem. B*, 2008, **112**, 9709–9719.
43 J. Rodríguez-Guerra Pedregal, P. Gómez-Orellana and J.-D. Maréchal, *J. Chem. Inf. Model.*, 2018, **58**, 561–564.
44 A. Bakan, L. M. Meireles and I. Bahar, *Bioinformatics*, 2011, **27**, 1575–1577.
45 P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L. P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks and V. S. Pande, *PLoS Comput. Biol.*, 2017, **13**, 1–17.
46 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins Struct. Funct. Bioinforma.*, 2006, **65**, 712–725.
47 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
48 M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor, *Proteins Struct. Funct. Bioinforma.*, 2003, **52**, 609–623.
49 G. Sciortino, E. Garribba and J.-D. Maréchal, *Inorg. Chem.*, 2019, **58**, 294–306.
50 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
51 P. Li and K. M. Merz, *J. Chem. Inf. Model.*, 2016, **56**, 599–604.
52 J. M. Seminario, *Int. J. Quantum Chem.*, 1996, **60**, 1271–1277.
53 G. Sciortino, J.-E. Sánchez-Aparicio, J. Rodríguez-Guerra Pedregal, E. Garribba and J.-D. Maréchal, *Metallomics*, 2019, **11**, 765–773.
54 A. Grossfield and D. M. Zuckerman, *Annu. Rep. Comput. Chem.*, 2009, **5**, 23–48.
55 J. Rodríguez-Guerra Pedregal, L. Alonso-Cotchico, L. Velasco-Carneros and J.-D. Maréchal, *ChemRxiv*, DOI:10.26434/chemrxiv.7059263.v1.
56 I. Y. B.-S. D.A. Case S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, *Amber 2018*, University of California, 2018.
57 M. Scheurer, P. Rodenkirch, M. Siggel, R. C. Bernardi, K. Schulten, E. Tajkhorshid and T. Rudack, *Biophys. J.*, 2018, **114**, 577–583.
58 R. González-Alemán, D. Hernández-Castillo, J. Caballero and L. A. Montero-Cabrera, *J. Chem. Inf. Model.*, 2020, **60**, 467–472.
59 Y. Haddad, V. Adam and Z. Heger, *Biophys. J.*, 2019, **116**, 2062–2072.
60 J.-E. Sánchez-Aparicio, G. Sciortino, D. V. Herrmannsdoerfer, P. O. Chueca, J. R.-G. Pedregal and J.-D. Maréchal, *Int. J. Mol. Sci.*, 2019, **20**, 3155.
61 R. A. Boto, F. Peccati, R. Laplaza, C. Quan, A. Carbone, J.-P. Piquemal, Y. Maday and J. Contreras-García, *J. Chem. Theory Comput.*, 2020, **16**, 4150–4158.
62 F. Peccati, *J. Chem. Inf. Model.*, 2020, **60**, 6–10.
63 A. Krammer, P. D. Kirchhoff, X. Jiang, C. M. Venkatachalam and M. Waldman, *J. Mol. Graph. Model.*, 2005, **23**, 395–407.
64 J. Hansen, J. Autschbach and H. M. L. Davies, *J. Org. Chem.*, 2009, **74**, 6555–6563.
65 M. Kotev, D. Lecina, T. Tarragó, E. Giralt and V. Guallar, *Biophys. J.*, 2015, **108**, 116–125.
66 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455-461.

