ACS Catalysis

Perspective

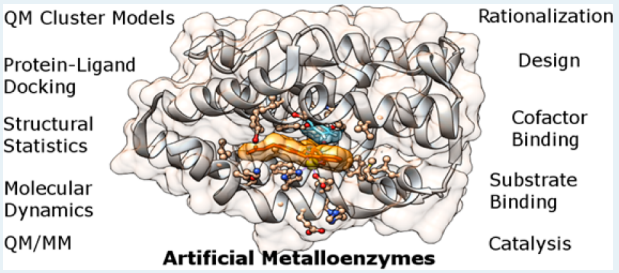
pubs.acs.org/acscatalysis

Toward the Computational Design of Artificial Metalloenzymes: From Protein–Ligand Docking to Multiscale Approaches

Victor Muñoz Robles, Elisabeth Ortega-Carrasco, Lur Alonso-Cotchico, Jaime Rodriguez-Guerra, Agustí Lledós, and Jean-Didier Maréchal*

Departament de Química, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Barcelona, Spain

ABSTRACT: The development of artificial enzymes aims at expanding the scope of biocatalysis. Over recent years, artificial metalloenzymes based on the insertion of homogeneous catalysts in biomolecules have received an increasing amount of attention. Rational or pseudorational design of these composites is a challenging task because of the complexity of the identification of efficient complementarities among the cofactor, the substrate, and the biological partner. Molecular modeling represents an interesting alternative to help in this task. However, little attention has been paid to this field so far. In this manuscript, we aim at reviewing our efforts in developing strategies efficient to computationally drive the design of artificial metalloenzymes. From protein–ligand dockings to multiscale approaches, we intend to demonstrate that modeling could be useful at the different steps of the design. This Perspective ultimately aims at providing computational chemists with illustration of the applications of their tools for artificial metalloenzymes and convincing enzyme designers of the capabilities, qualitative and quantitative, of computational methodologies.



KEYWORDS: *artificial metalloenzymes, biocatalysis, molecular modeling, multiscale approaches, protein–ligand dockings*

INTRODUCTION

Biocatalysis consists of the industrial application of enzymes for the manufacturing of chemical compounds. It is one of the cornerstones for green and sustainable chemistry because enzymes are by nature biodegradable, biocompatible, and easily renewable.[1] Despite being widespread in current industries, most biocatalysts are based on naturally occurring enzymes that, despite their variety, cover only a narrow spectrum of the needs of chemical industries.

During the past century, homogeneous catalysis has been the most prolific chemical field in discovering new chemical reactivities. The award of two recent Nobel Prizes of Chemistry (Chauvin, Grubbs, and Schrock in 2005; Heck, Negishi, and Suzuki in 2010) appears particularly illustrative. However, the transition metal complexes that sustain homogeneous catalysis are in their majority functional under nonenvironmentally friendly conditions, which include apolar solvents and low or high temperatures, among others. Moreover, control over substrate and regio- and enantiospecificities is generally challenging in these complexes; conversely, they are properties inherent to enzymatic activities.

With one-third of naturally occurring biocatalysts containing metal ions, metalloenzymes have been the focus of attention of enzyme designers. One possible framework consists of mutating residues that coordinate the metal in the native biomolecule or simply switch the metal by another. Such approaches have led to interesting outcomes in recent years, although modulating the activity of these scaffolds resides mainly in the biochemical space afforded by the 20 amino acids available in Nature.[2−5]

Another framework consists of physically merging homogeneous catalysts within a biomolecular host. Conceptually mimicking natural hemoenzymes, this strategy is increasingly applied to the development of biocatalysts absent from the biological realm.[6] In the resulting hybrids, also called artificial metalloenzymes, the cofactor (synthetic in this case) provides most of the catalytic specificity of the system. The protein environment protects the homogeneous catalyst from the solvent and generates an asymmetric second coordination sphere that dictates substrate, regio- and enantioselectivities, and specificities (Figure 1). Today, numerous systems developed using this concept have already been reported and include reactivities such as hydration of ketone,[7] transfer hydrogenation,[8] and sulfoxidation.[9] Strategies used to incorporate the cofactor inside the protein include pure host–guest interactions, "Trojan horse" insertion in which the cofactor is covalently bound to the natural ligand of a protein, or covalent anchoring in which peripheral substitutents of the organometallic catalyst chemically bind to the host.[10] A nonexhaustive list of artificial metalloenzymes with their catalytic activities can be found in Table 1.

The successful development of artificial metalloenzymes stands on the quality of the molecular partnership between

Received: January 4, 2015
Revised: February 19, 2015
Published: February 23, 2015

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

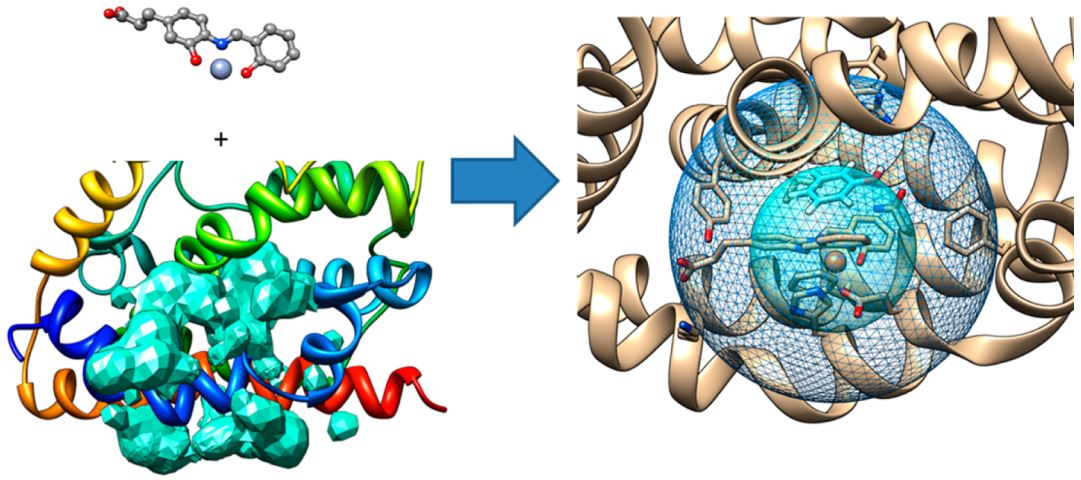


Figure 1. Schematic representation of the process of designing an artificial metalloenzyme. Homogenous catalysts (top left) and a protein host (bottom left) with sufficient vacant sites (solid blue blobs) are merged to provide artificial metalloenzymes (right onside). Their activity is driven for the first coordination sphere of the metal (blue sphere) and substrate (blue stick atoms) binding and orientation defined by the second coordination sphere environment (dark blue mesh sphere).

Table 1. A List of Artificial Metalloenzymes and Their Catalytic Activities

biomolecular scaffold	transition metal	organic cofactor	catalytic activity	substrate	enantioselectivity, %	ee	ref
NikA transport protein	iron(III)	organic ligand L1	oxidation	sulfides	10		13
LmrR	copper(II)	phenanthroline	syn hydration	ketones	84	?	14
		phenanthroline	Diels–Alder reaction	azachalcone	>97	+	15
bovine β-lactoglobulin	rhodium(III)	fatty acid derivatives	hydrogenation	trifluoroacetophenone	26	R	16
β-helical bionanotube	scandium(III)	bipyridine, Ser, Thr	epoxide ring-opening reaction	*cis*-stilbene oxide	17	R	17
streptavidin mutants	osmium(VIII)	quinidine or quinine derivatives	asymmetric dihydroxylation	olefins	95	R	18
	iridium(III)	biotinilated complex	hydrogenation	cyclic imines	96, 78	RS	8
DNA	copper(II)	phenanthroline	syn hydration	enones	72	R	19
DNA	copper(II)	DNA intercalating moiety	Diels–Alder reaction	dienophiles	90	exo	20

substrate, organometallic and biological partners in terms of binding and catalytic control.[11] The actual strategies for their design are time- and resource-consuming and consist mostly of trial-and-error procedures. In general, the main steps involved in the process consist of (1) the identification (by biochemical intuition of the researchers) of one possible protein scaffold able to bind a given homogeneous catalyst, (2) experimental binding assays of the artificial cofactor in this particular host, (3) testing of the catalytic activity for a prototypical substrate, and (4) optimization of the initial hit toward catalytic preferences.[7,12−20]

Molecular information is fundamental in all aspects of the process. However, structural knowledge provided by approaches such as X-ray or NMR are rather scarce for artificial metalloenzymes. These techniques generally fail because the complementarity between subsystems is not optimal especially when dealing with the first candidates of these systems. The interaction among the three partners has not suffered evolutionary pressures, affinity constants are generally low, and substantial protein engineering is needed to stabilize the structure of the hybrid system. Molecular modeling offers an interesting alternative to reach atomic details on the mechanism of artificial metalloenzymes and help in their design; however, the development of synthetic enzymes through in silico approaches is still in its infancy, and only a few attempts have been performed on the particular case of artificial metalloenzymes.

This manuscript aims to underline the particularities of artificial metalloenzymes in the area of in-silico-based enzyme design; give an overview of the strategies we have been empowering in the recent years to establish an efficient framework in this field as well as their consequent results; and finally, to focus on what we believed should be the future of modeling-based artificial metalloenzymes. It is a Perspective that intends to motivate computational chemists to consider artificial metalloenzymes as an interesting (but challenging) target as well as present to experimentalists how the variety of computational tools could be relevant for their designs.

A Brief Overview on Molecular Modeling Tools. Molecular modeling is now widespread at the interface between chemistry and biology, with models increasingly accurate, but molecular modeling is also a general term for defining a series of computational methods based on physical models with different degrees of accuracy and computational needs. Methods based on force field approaches, also called molecular mechanics (MM), allow vast geometrical samplings because of the relatively low ratio between the number of atoms and the computational cost. MM approaches are used mainly to study systems of large dimensionality (i.e., an entire protein in a solvated medium) and allow the exploration of large conformational spaces. Generally combined with deterministic (i.e.,

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

molecular dynamics) or stochastic (i.e., Monte Carlo) search algorithms, MM allows the study of changes in the shape of the molecule associated with their motions, extraction of statistical thermodynamics values, or handling of the prediction of the structure of large databases of the compounds, among others. When dealing with the interaction between partners, the additional degrees of freedom associated with translation and rotation increase substantially the geometrical space to explore. In these cases, a common solution consists of using simplified force fields centered on noncovalent terms (called scoring functions) and reducing the number of degrees explored during the conformational search (i.e., all the degrees of freedom of the ligand are considered, but a reduced number of amino acids or, eventually, collective motions are allowed to move during the docking process). Protein−ligand dockings, which aim at predicting the structure of the complexes formed between small molecules and proteins, are based on these premises. In any case, only very specific and nonstandardized MM approaches are able to predict fine electronic effects.[21,22]

Computational methods based on quantum mechanics (QM) accurately reproduce the nature of the electronic properties of the molecules and allow simulating changes in their chemical state. QM approaches are used for very different molecular problems, including spectroscopic and photoelectronic processes and any system in which its coordination or the covalent linkages change during a chemical process. A vast ensemble of QM methods is accessible nowadays. It is likely that those with the wider number of applications are based on the density functional theory (DFT). These methods allow the insertion of fine electronic effects (correlation) for a relatively low additional cost over the typical Hartree−Fock calculations and are particularly relevant in fields such as organometallics. Despite their success, DFT techniques are based on a series of approximations that could substantially limit the reliability of their results (i.e., dealing with changes in spin states of a transition metal is still a challenging task).[23,24] The quest for the best DFT method is a vivid field of research, and still today, DFT capabilities seem system-dependent.[25] Whatever their ground, though, QM approaches are counterbalanced by expansive computational costs that do not allow sampling of large-dimensional problems.

Approaches that combine several methodologies together are increasingly applied in molecular sciences to overcome the limitations of individual methodologies. Generally referred to as multiscale, integrative, or hierarchical methods, their potential has already been widely recognized, including by the Nobel Prize in chemistry awarded to Karplus, Warshel, and Levitt in 2013 "for the development of multiscale models for complex chemical systems". Prototypical multiscale approaches are the hybrid quantum mechanics/molecular mechanics (QM/MM) methods, which considers part of the molecule under a quantum mechanical framework and the remaining part under a molecular mechanics approximation. QM/MM methods are now legion and key in the simulation of biomolecular systems. They differ in the algorithms used in each subset of atoms and how the information is transferred from one to another.[26] Major breakthroughs in metalloenzymes have been reached with these methodologies in decoding enzymatic mechanisms.[27−31]

Other combinations are frequent in biosimulation either under successive steps of different methods or integrated under a unique protocol. Focusing on those related to the study of enzymatic systems, some bridge molecular dynamics and QM/MM calculations. This combination is particularly interesting when an enzyme−substrate complex is relatively well-defined (for example, from an X-ray structure obtained with a substrate analogue) and is aimed to improve the quality of the catalytic path explored under the QM/MM energetic landscape.[32] However, when more complex binding processes need to be modeled, protocols integrating protein−ligand dockings are necessary. Although such combinations are less frequent in the modeling of enzymatic reactions, their use has been particularly relevant in several recent studies, including those related to the study of the reactivity of cytochromes P450s 3A4,[33] the change of specificity of cytochrome P450 2D6,[34] the elucidation of the catalytic mechanism of *Trametopsis cervina* lignin peroxidase,[35] or the promiscuous activity of human carbonic anhydrase against cyanic acid.[36] In all of them, the dockings are used, alone or in combination with MD runs, to provide physically sound complexes between the substrate and the enzyme prior to catalysis.

In-Silico-Based Enzyme Design. Different ways of developing new enzymatic activities are under the scrutiny of designers. Their differences arise from the degree of molecular diversity involved in the biomolecular scaffolds and include the engineering of a few amino acids in the active site, development of catalytically active peptides, or the redesign of a pre-existing scaffold.[37] Computation has been increasingly involved in several of these strategies. We here focus on computer-based designs in which conceptual frameworks best overlap with those that could lead to artificial metalloenzymes.

De novo design of artificial enzymes consists of identifying a protein scaffold and its consequent mutations to catalyze a nonnatural reaction on a given substrate. The combinatorial space to reach an active scaffold is tremendous and not yet achievable by experimental means. Part of de novo enzymes are based on relatively small peptides that could self-assemble. Systems with those dimensions confine the search for activity into a sequential space easier (but still challenging) to handle with respect to large folded proteins.[38−40] For designs considering larger folds, computation is more frequently required.

The most established procedures for computer aided de novo design of artificial enzymes stand on hypothetical transition state structures of a nonnatural reaction that could be embedded in a protein medium. The identification of such geometry is often performed by quantum mechanical calculations on a minimalist active site, which includes the substrate and a series of functional groups representing side chain of amino acids that could stabilize its orientation and participate in the reaction. These cluster models, also referred to as *theozyme*s under the definition of Houk and co-workers,[41] are used as starting points for posterior search algorithms under an explicit protein environment.

Mayo et al. were among the first to generate a novel proteic scaffold from in silico approaches and reached a synthetic $\beta\beta\alpha$ motif designed by screening 1.9×10^{27} possible amino acid sequences.[42] Pursuing their efforts, they computationally identified mutations in the 108-residue *Escherichia coli* thioredoxin, leading to a "protozyme" able to catalyze the histidine-mediated nucleophilic hydrolysis of *p*-nitrophenyl acetate into *p*-nitrophenol and acetate.[43] Subsequently, Mayo and co-workers implemented a new method to place the substrate within the active site of the protein while the designing algorithm is exploring the conformational and chemical space.[44] In their more recent successes, they iteratively

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

execute computational simulations (best designing algorithm and MD simulations) with X-ray crystallography to obtain one of the best Kemp eliminases so far reported (K_{cat}/K_m of 430 M^{-1} s^{-1} rate after three iterations).[45]

The Baker laboratory is another clear example of success in de novo design of artificial enzymes by in silico approaches using a procedure that combines rational design and directed evolution.[46,47] Briefly, their methodology accounts for an extensive search of pre-existing high-resolution protein structures that could accommodate the transition state structure using the RosettaMatch algorithm.[48] A scaffold is considered a match if it satisfies that all the amino acid side chains of the theozyme that can be placed on the protein scaffold. Each match is then optimized using the RosettaDesign methodology for proteins and small molecules.[49,50] Except for catalytic residues of the theozyme, all the remaining amino acids in the vicinity are redesigned so that the final cavity has the maximum shape complementarity with the modeled transition state. All the resulting structures are then screened for compatibility with substrate/product binding and ranked according to the catalytic geometry and the computed transition state binding energy. This way, a handful of different putative new enzymes are selected for experimental characterization. Those that present the final activity will undergo a series of directed evolution steps for further optimization.

Computational involvements in manipulating enzymes are limited not only to de novo designs but also to bioengineering processes, some of them relevant for the present work. For enzymes with proven reactivity, computation can be used to rationalize their mechanism, improve it, or ultimately reorient their activity. An increasing number of studies with this objective have appeared in the literature over the past decade. Both pure quantum mechanical on large models of the active site[51] and hybrid QM/MM calculations[30,35,52,53] are used to this end.

Artificial metalloenzymes constructed by the insertion of homogeneous catalysts into protein have thus far received very little attention from computation.[52,54−56] Conceptually, their design stands on the same premises as pure organic systems: the modeling should identify transition state structures stabilized under a biocompatible host. However, these composites work because of a complementarity of the three different molecular entities and not only on protein−substrate recognition and activation. In this case, molecular modeling needs to handle the cofactor−host−substrate triad as best it can. The need in dealing with metal-mediated recognition processes, the effects they could induce on the structure of the host, and the reactivity of the final composite provide extra complexity for molecular modeling.

Our Computational Framework for the Modeling of Artificial Metalloenzymes. In the field of artificial metalloenzymes, molecular modeling needs to address processes involving large conformational sampling on one side and fine electronic effects on the other. The former are related to the binding of the artificial cofactor in the host and the orientation of the substrate in an efficient manner for the reaction to proceed. In principle, those steps can be achieved by protein−ligand dockings. The latter consists of events related to the identification of stable geometries of the isolated cofactor, the changes in its coordination sphere upon binding, and the characterization of low-energy reactive paths with the emphasis on identifying transition states structures. Pure QM and hybrid QM/MM calculations represent the best candidates for leading this part of the modeling.

In recent years, the objective of our group has been to generate computational protocols efficient for the design of artificial metalloenzymes. As a framework, we decided to use standardized (or lowly tuned) computational chemistry methods as well as multiscale approaches as a function of the problem presented. Regarding protein−ligand dockings, we decided to use the commercial software GOLD, which is one of the few that contain metal parameters, although at the beginning of our work, none were designed for metal-containing ligands.[57,58] It also affords flexible schemes for both receptors and ligands, which we applied in most of our calculations. Quantum simulations are performed using the Gaussian package (Gaussian09)[59] for both pure DFT calculations and QM/MM calculations. The latter are performed using the ONIOM approach using mechanical and electronic embedding.[60] Finally, structural modeling and statistics are performed in the UCSF Chimera[61] platform and include, nonexclusively, the exploration of rotameric conformation of amino acids[62] or the clustering of large sets of geometries using the NMRClust approach.[63] To ease our development of integrated approaches, we also developed a series of interfaces written in Python into the UCSF Chimera environment, which allows rapid input/output exchanges between the different methodologies we use.

1. BINDING OF ORGANOMETALLIC COMPOUNDS TO PROTEIN: THE QUEST FOR RESTING STATE MODELS

The design of artificial metalloenzymes relies, on a first instance, on the identification of structural matches between a biomolecule and a homogeneous catalyst. Only efficient complementarities should lead to a precatalytic state. The availability of 3D models of protein−artificial cofactor complexes is therefore fundamental at this stage. Although protein−ligand dockings represent one of the cornerstones in medicinal chemistry and drug design projects, little attention has been paid to the interaction of organometallic compounds with proteins. Indeed, only a small amount of drug candidates contain transition metal ions. However, metals are considered in several of these techniques for their presence in the active site of metaloproteins and how they influence the binding of organic drugs. In this case, different strategies are available to introduce metal−ligand interactions in the calculation of the energy, ranging from simple electrostatics (hydrogen-bond-donor-like function)[64] to coordination rules.[58,65,66]

For the interaction of organometallic entities with proteins, an accurate computational prediction has not yet been standardized. On the basis of bioinorganic considerations, efficient modeling should take into account (1) changes of the electronic state and geometry of the first coordination sphere of the metal upon binding, (2) geometrical changes on the entire cofactor, and (3) possible induced effects on the protein scaffold. Dealing with all these variables is beyond the scope of standard protein−ligand docking software, and different levels of approximation are mandatory.

A first case scenario consists of the situation in which no ligand exchanges occur on the metal when migrating from solution to its cavity in the host. These so-called "inert scaffold"[67] interactions imply that only subtle rearrangements of the first coordination sphere of the metal happen upon binding but that its overall geometry is little affected. From a

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

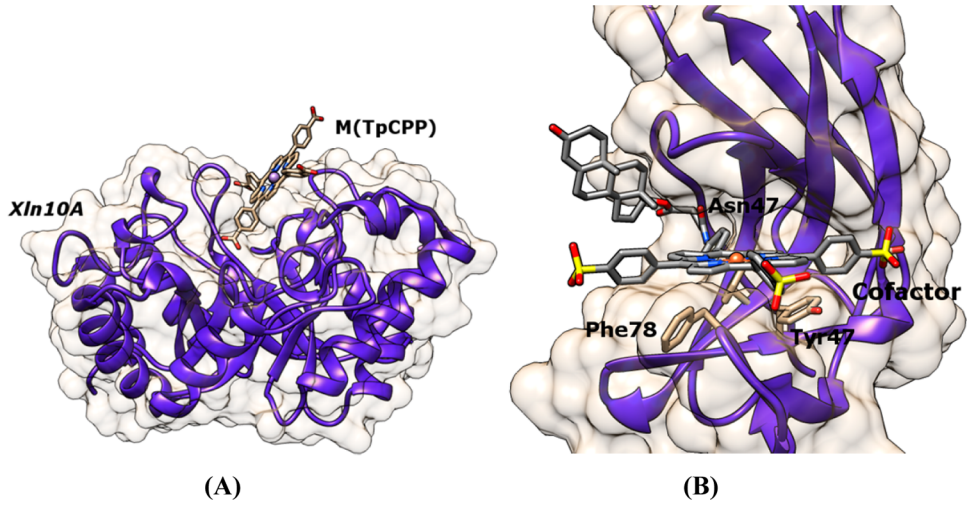


Figure 2. General view of different artificial metalloenzymes for which complementarity between host and cofactor were studied by protein ligand docking: (A) supramolecular interaction between iron porphyrin (concretely Fe(III)TCPP) and the xylanases 10A; (B) NCS bound with a Trojan horse testosterone−porphyrin derivative.

computational point of view, such systems could be simulated by dockings in which the close vicinity of the metal in the ligand remains rigid and the rest of the scaffold is optimized during the conformational search.

On the basis of this hypothesis, we recently showed that protein−ligand docking software behaves for inert scaffolds as well as it does for organic ligands. Limited by the reduced number of crystal structures of organometallics bound to protein available in the Protein Data Bank,[68] we performed a benchmark on structures corresponding to metal-containing inhibitors bound to their kinase targets and designed at Meggers' Laboratory.[67] Using GOLD as the method of choice, we tested different flexible schemes and scoring functions. The results were of very good quality. First, excellent structural matching between calculated low-energy structures and experimental complexes is observed. Between 75 and 94% of the theoretical complexes presented an RMSD lower than 2.5 Å from their experimental counterparts. The scores obtained were also of good quality, with correlations between experiment and theory reaching R^2 values up to 0.8 for those scoring functions that best behave. Of the scoring functions available in Gold, ChemScore[57] appears the most robust for both structural and energetic predictions. Outliers were encountered only when the geometry of the ligand bound to the metal differs substantially between the isolated conditions and the proteic complexes, something happening rarely in this set but that illustrates that the improvement of how dockings can explore the conformational changes related to the first coordination sphere of the metal or the coupling of dockings with accurate electronic methods such as QM/MM would represent a major step forward in those predictions.[69]

The prediction of the binding of synthetic cofactors to proteins under an inert interaction represents an initial step along the quest of 3D models of the resting state of artificial metalloenzymes. Despite an apparent simplicity, such approximation still provides crucial information on the most important features in defining protein−ligand binding: shape, hydrophobic and hydrogen bonding complementarity. Such information is extremely valuable when dealing with a first generation of an artificial enzyme for which structural information is missing. As such, dockings performed under this assumption have been one of the cornerstones in our collaboration with Mahy and co-workers, whose main objective is the development of artificial oxidases, more particularly, peroxidases and cytochromes P450.[70]

The first of our studies allows rationalization of the difference in activity of iron(III)-tetra-α_4-*ortho*-carboxyphenylporphyrin (Fe(ToCPP)) and iron(III)-tetra-*para*-carboxyphenylporphyrin (Fe(TpCPP)) systems embedded into xylanase A (Xln10A) from *Streptomyces lividans*. Xln10A is a glycoside hydrolase that hydrolyzes β-1,4 bonds in the main chain of xylan and is available at low cost and in large quantities.[71] Of the most important results, the protein−ligand dockings showed that Fe(TpCPP) enters deeper into the large Xln10A cleft than its Fe(ToCpp) counterpart. This better complementarity is due to a major part of the porphyrin ring anchored into the binding site as well as a substantial hydrogen-bonding network between the peripheral carboxylates of consecutive aromatic substituents and two polar patches of the receptor (Figure 2A). Moreover, the calculated Fe(TpCPP)-Xln10A complex shows the cofactor with one of its faces slightly packed on the surface of the binding side protein and the other accessible to the solvent. This orientation is in agreement with the experimental observation that only one imidazole could coordinate the iron of the porphyrin. Similar approaches on the same target also concluded that a sulfoxinated tetra-*para*-phenylporphyrin shows different binding modes with regard to its carboxylic counterparts with wider variability in interacting with polar patches of the receptor.[72] A final study with Xln10A as a receptor for porphyrin complexes showed that metallic Schiff base cofactors displayed very limited complementarities to the Xln10A binding site but that Mn(TpCPP) afforded a cavity vast enough to accommodate a substrate for the epoxidation of the series of aromatic styrenes. Interestingly, one of the residues of the receptor (Arg139) is identified to control the access of the substrates.[73]

More recently, we focused on another receptor: an engineered mutant of neocarzinostatin (NCS). NCS is a 113 amino acid chromoprotein secreted by *Streptomyces* that binds a nine-membered enediyne "chromophore" responsible for the cytotoxic and antibiotic activities of the protein−ligand complex.[74] The NCS 3.24 mutant allows the binding of two

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

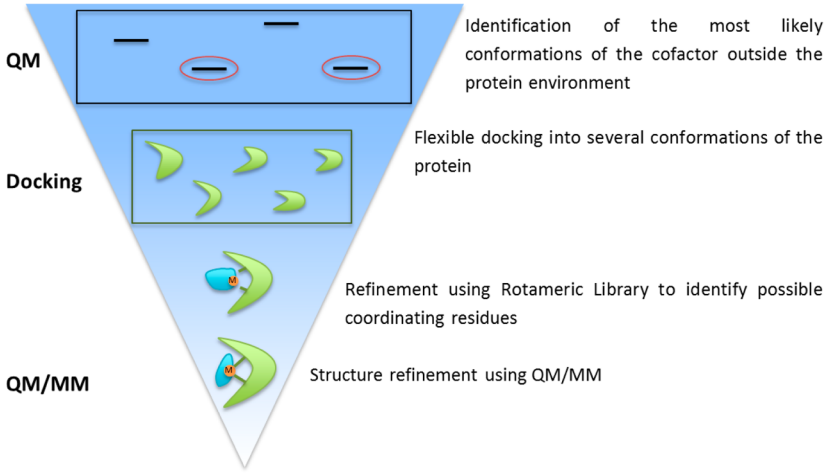


Figure 3. Steps of our integrative procedure combining docking, structural statistics and quantum mechanical based calculations.

testosterone molecules in its hydrophobic binding site in place of the natural chromophore. Using a so-called "Trojan horse" strategy, Mahy and co-workers synthesized an iron(III)−porphyrin−testosterone derivative able to bind to NCS-3.24. The resulting hybrid is able to catalyze the chemoselective and slightly enantioselective (ee = 13%) sulfoxidation of thioanisole by H_2O_2.

To increase the yield and the enantioselectivity of the construct, protein−ligand dockings were applied to look for improvement of the interaction between the cofactor and the protein. The molecular modeling showed that the porphyrin macrocycle fits perfectly into the protein binding site and is well sandwiched between the two subdomains of the protein (Figure 2B). However, the metal ion remains exposed to the solvent, which could explain the moderate enantioselectivity observed. The study also gave hints on possible improvements in the "Trojan horse" strategy because the artificial cofactor has filled up the two testosterone sites entirely and displaced its conjugated scavenger out to the solvent. Smaller cofactors are therefore expected to better fit inside the binding site of the enzyme and provide a wider asymmetric environment for enantioselective reactions.[75,76] A final scaffold we studied is a family of porphyrin-binding catalytic antibodies that are able to perform peroxidase activities. In conjunction with X-ray structures that were not conclusive on the geometry of the cofactor in the hapten recognition site, we could qualitatively rationalize both activity and binding.[77]

Although the binding of organometallics to their host in an inert fashion is frequent with drug compounds, for homogeneous catalysts, this hypothesis is valuable only as a "first shoot" for structural knowledge. In a wider context, results obtained under this assumption have to be nuanced. On one side, the absence of coordination changes during binding needs to be compared with experimental data, mainly spectroscopic, to validate such an approximation. Moreover, this approximation is interesting mainly for resting state structures. Indeed, either prior to or during the catalysis, one or several groups bound to the metal are likely to be displaced from its isolated situation in solvent to its binding to the protein cavity. When disposing of the clear idea on which groups could leave the cofactor (i.e., labile water on the top of the iron in a heme like complex), a possible strategy consists of mixing dockings and quantum-based approaches to identify correct resting states of the artificial metalloenzymes.

To this end, we developed an integrative procedure that combines docking, structural statistics, and quantum mechanical-based calculations (Figure 3). In this process, stable structures of the isolated cofactor obtained either from pure quantum mechanical calculations or from a database of small molecules (considering, if necessary, spin and oxidation states) are initially docked into the receptor cavity.

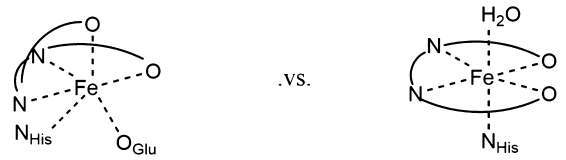
During the docking, we simulate the formation of possible coordination bonds between the metal and atoms of the protein by removing the most likely leaving group from its first coordination sphere and using a pseudometal atom type. In Gold, our program of choice, a hydrogen-like function is located at the vacant coordination site with directions that respect the coordination rules of the metal (i.e., octahedral, square planar, etc.) and can interact with Lewis basis atoms. The resulting binding modes are further analyzed to identify additional residues that could reach the metal ion. On the basis of statistics of metaloprotein three-dimensional structures, any residue with the C_α under the cut-off of 9 Å from the metal could display one of its rotameric states coordinating the ion. Once those amino acids are identified, the final step along the process consists of generating the different coordination modes of a given docking solution by rotameric refinement and pursuing with QM/MM calculations of the resulting complex. QM/MM calculations are generated with an initial minimization constraining the coordination bond to a reasonable distance and subsequently releasing the constraint to avoid artifacts along the optimization. The potential energies of the final models are compared together, and those with the lowest energy are compared and discussed and could eventually be used for further designs.

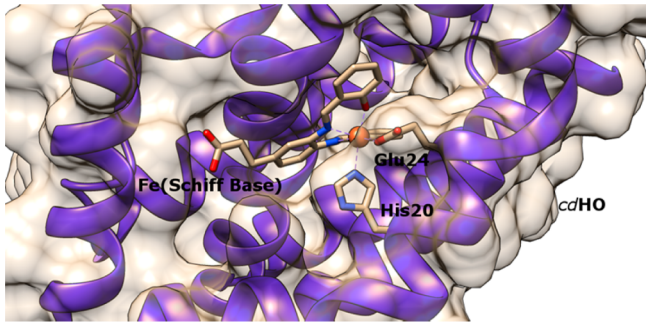
We tested this approach for the first time in 2010 on the structure of an artificial metalloenzyme obtained by the substitution of the heme by a Fe(Schiff base) salophen in *Corynebacterium diphtheria* heme oxygenase (*cd*HO).[78] *cd*HO is a small all-α enzyme that performs the first step of the oxidation of the heme.[54] The Fe(Schiff base)−*cd*HO resulted as a superoxidase able to work thanks to successive reductions performed by the electron partner of the natural enzyme. Importantly, the crystal structure of Fe(Schiff base)-*cd*HO shows major differences from other salophen and heme-bound enzymes (Scheme 1 and Figure 4A).

First, the iron displays an octahedral configuration with a distorted cofactor and diverges from the planar geometry

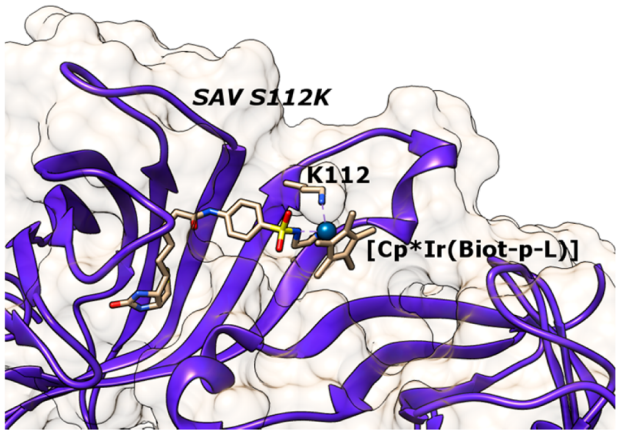
DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

Scheme 1. General Geometries of X-ray Geometry of the Fe(III)Schiff Base·*cd*HO System (left) and Resting States of Naturally Occurring Hemoenzymes (right)





(A)



(B)

Figure 4. Examples of artificial metalloenzymes in which the binding of the cofactor occurs under an active coordination sphere: (A) predicted geometries of the Fe(III)(Schiff Base)-*cd*HO resting state and (B) structure of the binding site of the [Cp*Ir(Biot-p-L)Cl] ⊂ S112K Sav system with a coordinated lysine 112 to the iridium complex, as predicted by the docking procedure.

observed with salophen and porphyrin systems (Figure 4A). In this geometry, one of the oxygen atoms of the catalyst migrated from the equatorial position to the axial position. Second, two residues coordinate the iron, the His20 that occupies one of the axial positions, and the glutamate 24 that fulfills the fourth position of the equatorial plane. This geometry does not follow the general trend of heme-bound enzymes, including the natural substrate-bound heme oxygenases. In those systems, the iron coordinates the four atoms of the macrocycle in equatorial position and binds His20 as a unique protein ligand at the proximal axial site. Moreover, the resting state of hemoenzymes remains either square planar or octahedral by the binding of an additional labile ligand to the metal, generally a water molecule. Finally, the *cd*HO structure also presents a displacement of the helix A, to which both His20 and Glu24 are bound, that has been unreported before in X-ray structures of heme oxygenases. As a whole, the geometry presents all the molecular features that challenge computational prediction of the binding of organometallics to protein.

Our study consisted of a blank experiment in which we started with a set of structures of the heme oxygenase (none of them corresponding to the crystal structure of Fe(Schiff base)-*cd*HO) and the chemical structure of the cofactor. Applying the protocol described previously, we obtained two low-energy structures: one with excellent structural similarity to the crystal structure of the artificial metalloenzyme, the second with a square pyramidal geometry reminiscent of the heme enzymes and in which glutamate 24 was eliminated from the first coordination of the metal.[54]

The first structure clearly illustrates that the combination of protein−ligand dockings and QM/MM approaches would lead to excellent predictions for active binding of homogeneous catalysts to a protein host, even with changes of the coordination sphere of the metal. The second apparently suggests a failure of the simulation in discriminating between different binding modes but also that a possible equilibrium between hexacoordinated and pentacoordinated geometries exists in solution. The crystal structure snapshot could have somehow trapped an intermediate out of the catalytic path of the enzyme.

We further studied this aspect by investigating the transition between both structures considering all the spin and oxidation states conceivable in the initiation step of the catalysis by QM/MM calculations. We showed that the X-ray structure corresponds to the real resting state of the enzyme in Fe(III) state, and the square pyramidal one corresponds to the reduced Fe(II) form of the enzyme. Energy decomposition using different QM/MM partitions allowed identifying that the first coordination sphere of the metal is the most important factor in dictating the geometry of the final complex. In addition, this study clearly demonstrated that the transition between both structures is energetically feasible only when the reduction has occurred, hence providing additional evidence of the divergence in the mechanism of action of artificial and natural heme-like enzymes. The transition state vector also shows that the reorganization of the cofactor, the displacement of the glutamate out of the first coordination sphere of the metal and the entire rearrangement of the helix A, are intrinsically related.[79]

More recently, we applied the same procedure to artificial imine reductases designed by Ward and co-workers and resulting from the incorporation of a biotinylated Cp*Ir Noyori's-like catalyst (Cp* = C_5Me_5−) within different mutants of the homotetrameric streptavidin (Sav) (referred to as Cp*Ir(Biot-p-L)Cl] ⊂ Sav). Mutants at position S112 reveal major differences in both the Ir/streptavidin ratio and the enantioselectivity for the production of salsolidine. For [Cp*Ir(Biot-p-L)Cl] ⊂ S112A Sav, the reaction rate and the enantioselectivity (which reach up to 96% ee for (*R*)-salsolidine) decrease upon saturating all biotin binding sites, whereas for [Cp*Ir(Biot-p-L)Cl] ⊂ S112K Sav, the rate and the ee remain almost constant as a function of the ratio Ir/streptavidin (ee near 78% for (*S*)-salsolidine). Our docking complemented the X-ray structures that only partially resolve the location and the orientation of the cofactor into the cavity of the hosts. In collaboration with Ward's group, our calculations verify that the S112A and S112K Sav mutants prefer binding the S_{Ir} and R_{Ir} enantiomeric forms of the cofactor, respectively, a phenomenon not observed on natural enzymes binding organometallic cofactors. Moreover, it shows

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

that the binding in the S112K mutant could be stabilized by a coordination of the metal with the N_ε atom of lysine 112, forming a resting state structure stabilizing the orientation of the cofactor and differing from those of the S112A mutant in which no additional coordination is observed between the metal and the protein (Figure 4B).[80]

2. MODELING THE CATALYTIC ACTIVITY OF ARTIFICIAL METALLOENZYMES

The binding of the synthetic cofactor is part of the molecular events that condition the design of artificial metalozymes, but their activity can be understood only from the binding of the substrate and its activation. Protein−ligand dockings on its own can provide some relevant insights by allowing the identification of substrate binding modes that are catalytically consistent.[81−83]

The composite resulting from the insertion of Mn(III)-meso-tetrakis(*p*-carboxyphenyl)porphyrin (Mn(TpCPP)) into Xln10A displays epoxidative activity on a series of styrenes. Subtle enantioselective preferences toward *S* products are observed in most substrate but the most remarkable ee is observed with *p*-methoxystyrene, featuring a stereoselectivity of 80% in favor of the *R* isomer. The docking of the different substrates into the binding site of a model of the artificial metalloenzyme previously generated showed that the predicted orientations of the substrate in the active site of Mn(TpCPP)-Xln10A consistent with the formation of *S*-epoxide are slightly more stable than those for *R* epoxide ones. However, for *p*-methoxystyrene, the trend is inverted, with orientations consistent with *R* epoxide formation being more stable. Such inversion is associated with an additional H-bond between tyrosine 172 and the oxygen atom of the *p*-methoxy substituent (Figure 5).[73]

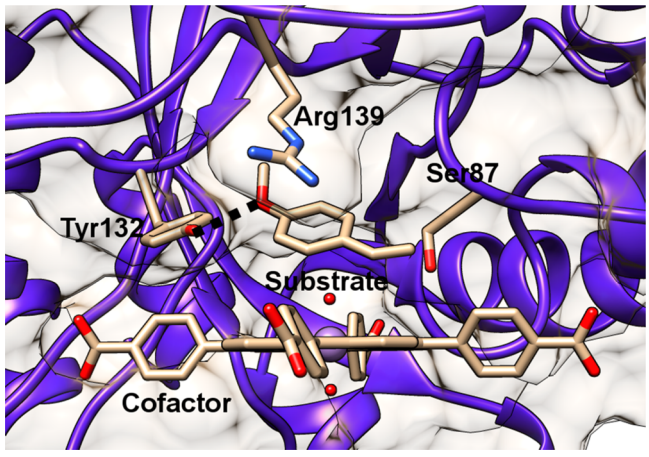


Figure 5. Predicted low-energy complexes of the *p*-methoxystyrene inside the cavity of the Mn(TpCPP))-Xyl10A model, corresponding to the orientation of the substrate consistent with the formation of the *R* epoxide product.

Protein−ligand dockings are not able to identify true transition state structures along an enzymatic reaction. They are limited to providing substrate binding modes that are to be contextualized in terms of prereactive orientations. The characterization of real, true transition state structures can be performed only by means of QM/MM calculations. However, for most artificial metalloenzymes, this task cannot be dissociated to a wide conformational sampling because little molecular information is available for the location and orientation of the cofactor and its interaction with the substrate.

We developed a protocol combining docking, QM, and QM/MM calculations in which both substrate and cofactor are taken into account during the geometrical search. This methodology, reminiscent of the work of Houk, Mayo, and Baker, consists of three successive steps: (1) Study of the catalytic mechanism in a cluster model of the enzyme by DFT calculations. The reduced model consists generally of the cofactor, the substrate, and amino acids likely to interact on the reaction center. (2) Docking of the geometries of the transition state structures obtained in step 1 into the binding site of the artificial metalloenzyme. Those pseudotransition states are generated while imposing few geometry variables extracted from the structure of the transition state models. (3) Refinement by QM/MM calculations of the pseudotransition state structures obtained in step 2 and identification of true transition state structures on the full potential energy surface. From this step, the lowest energy paths can be identified and compared with experiment. At each step along the process, the models with the substantially highest energies are neglected for the next step forward. A scheme of the protocol employed, showing the sequential steps and its application to an example, is depicted in Figure 6.

A first application of this methodology has been the study of the catalytic mechanism of the artificial [Cp*Ir(Biot-p-L)Cl] ⊂ S112A transfer hydrogenase mutant mentioned earlier.[8] Although the mechanism of reduction of ketones by Noyori's-like complexes is now widely accepted, the one leading to the reduction of imine has not yet reached a consensus. Although the metal center is well-known to provide the transfer of hydride, there is still discussion on the source of the proton. As a consequence, our model system considers several mechanistic hypotheses for this step, including the organometallic moiety itself, a hydronium from the medium, or a positively charged lysine that the active site could contain. Calculations were performed for processes leading to *R* and *S* chiral reduced imine.

The first step of our protocol allowed discarding mechanisms in which hydride and a proton are transferred from the homogeneous catalysts. All of those pathways are systematically 15 kcal mol^{-1} higher than any other ones, a magnitude difficult to imagine that a protein scaffold could counterbalance. After docking the remaining pseudotransition state structures in the streptavidine vestibule, QM/MM refinements led to the identification of eight different reaction paths. Those involving proton transfer from the lysine residues located in the binding site are the less favored. The lowest-energy mechanism implies the transfer of the hydride on the substrate that was protonated in solution prior its access to the SAV112A site. The lowest-energy *R* and *S* paths clearly indicate preference toward the formation of the *R* product. The corresponding ee calculated on the difference in energy of the transition states reaches 80%, a magnitude in good agreement with the 98% reported experimentally. Interestingly, the geometries of the predicted transition state structures of *proR* and *proS* mechanisms clearly show a major drift of the cofactor and the substrate into the SAV cavity, something absent from natural hemoenzymes in which the location of the cofactor is well stabilized (Figure 7).

The relevance of a correct identification of the location of the cofactor is consistent with the first QM/MM study reported on an artificial metalloenzyme by Morokuma et al.[52] They analyzed the reaction mechanism of the polymerization of phenyl-

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

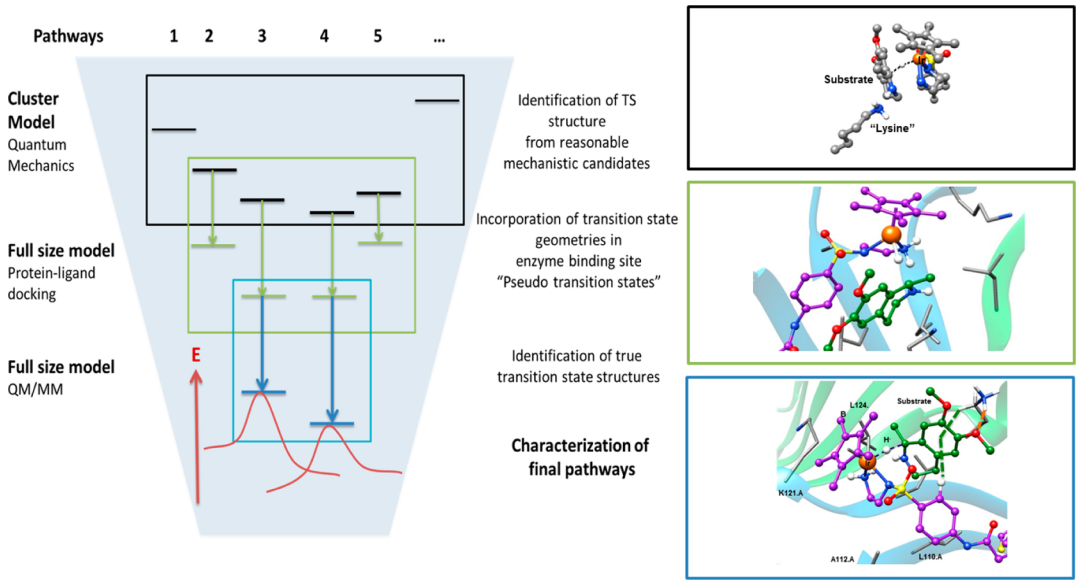


Figure 6. Schematic procedure for the identification of catalytic mechanism inside a proteic scaffold of artificial metalloenzymes considering the uncertainty of the location of the cofactor (left) and its application to mechanistic study of an artificial transfer hydrogenase (right).

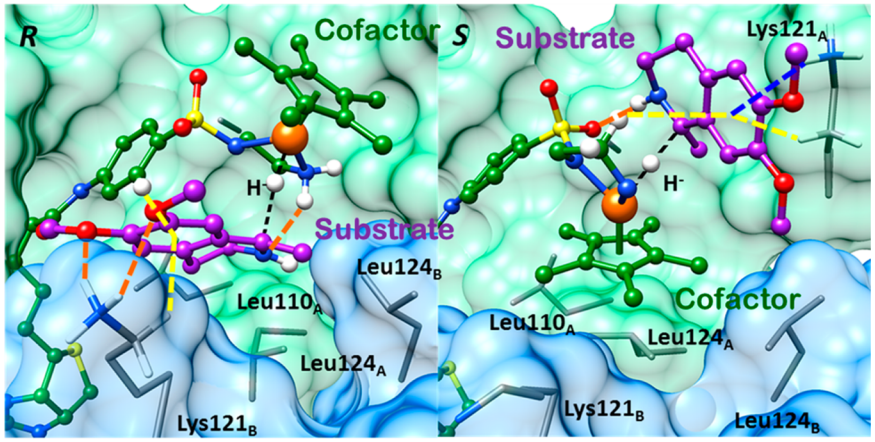


Figure 7. Calculated transition state structures of the lowest energy paths leading to the formation of *R* (left) and *S* (right) salsolidine.

acetylene by an artificial metalloenzyme developed by Ueno and co-workers. This biometallic hybrid was obtained through the insertion of a $[Rh(norbornadiene)Cl]_2$ catalyst inside a horse L-chain *apo*-ferritin. After investigating different reaction mechanisms, the authors were able to characterize the most likely cavity in the *apo*-ferritin structure that could shelter the inorganic cofactor and favor the polymerization reaction.

CONCLUSION AND PERSPECTIVE

The development of artificial metalloenzymes is becoming a major field of investigation. To date, designers have based most of their work on (bio)chemical intuitions, with limited structural information. Molecular modeling can be useful in this field, although dealing with all the energetic and structural aspects that need to be considered represents a real tour de force. Among them, metal-mediated recognition processes involved in cofactor binding are fundamental but out of the scope of current state-of-the-art protein−ligand docking methodologies. Another fundamental aspect is to determine three-dimensional models of catalytically active conformations of the cofactor−substrate−receptor triad.

A few years ago and based on our experience on organometallics, drug design, and bioinorganics, we embarked on establishing and benchmarking procedures convenient for artificial metalloenzyme design. The fruits of the first steps in this venture are summarized in this manuscript. Here, we try to show which approximation we had to contemplate, the evolution of our approaches, the most relevant elements of our achievements, and what accuracy is to be expected.

Regarding the binding of the homogeneous catalyst to a protein, we validated standard protein−ligand docking procedures in generating accurate 3D models if no chemical changes of the first coordination sphere of the metal occur upon binding. Although scoring functions and parameters to deal specifically with metal ions in ligands still leave room for improvement, calculations performed under this hypothesis are extremely instructive. Such dockings are particularly relevant for composites that represent the first line of candidates for artificial metalloenzymes. This benchmark also suggests that high-throughput virtual screening of large databases of organometallics and proteins to detect novel frameworks is achievable.

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

For the prediction of 3D models involving the modification of the first coordination of the metal upon binding, we thought that the best chances of success were with allying protein−ligand docking with QM and QM/MM approaches. This strategy allows reproducing complex processes and gives highlights on major induced effects, such as the distortion of the cofactor and large scale motions of the receptor, and eventually complement crystallographic observations. However, its success depends on hypothesizing which coordinated ligand(s) leave(s) the cofactor and is (are) replaced by protein residues. This is relatively straightforward for labile water molecules in heme-like systems but still challenging for organometallic compounds for which interaction with biological scaffolds are less documented. Generating multiple coordination sphere candidates during the docking run is therefore primordial for more advanced designs and a methodological challenge we are now exploring.

Finally, we showed that bridging pure quantum mechanics calculations on model systems, protein−ligand dockings, and QM/MM calculations allow identification of true transition state structures in artificial metalloenzymes. This approach is efficient enough to characterize reaction paths even when the location of the cofactor is uncertain. It also provides information on fine structural events, such as those that control the enantioselective profile of artificial metalloenzymes. Importantly, this procedure is computationally far less demanding than other deterministic protocols, such as those performed with stirred molecular dynamics or metadynamics.

Although most of the tools for computer-aided design of artificial metalloenzymes are now part of the toolbox of computational chemists, their success in this field will also be dependent on our ability to deal with the fine-tuning between simulation of binding processes and catalytic mechanisms. In particular, better sampling protein−ligand docking techniques are of the most important aspects to incorporate into an integrative framework. As such, methods than allow fast introduction of large-scale (collective) motions[84] and enhanced sampling in docking will be a major asset.[85]

With few years dedicated to this field, we believe that our experience illustrates the potential of molecular modeling tools for the rationalization of the reactivity of existing artificial metalloenzymes. Decoding their molecular mechanism at the atomic level first provides useful information for further optimization steps (i.e., control over regio- and enantioselectivities and specificities) and also affords conceptual knowledge on nonnatural bioinorganic interactions. With the lack of molecular information on these composites, we hope this could serve in the development of the entire field of artificial metalloenzymes.

The challenge in the years ahead consists of expanding our modeling framework so that computation could become an interesting tool for design purposes. To reach such in silico designs, focus should be given, among others, on strategies that allow the identification of protein scaffolds that could host the artificial cofactors and satisfy the chemical requirements for the reactivity to occur as well as predict suitable redesign of the protein−substrate−cofactor interface. To this end, combinations and adaptations of the approaches described in this manuscript with those already established in enzyme design are among the most interesting. Still, the main barrier to overcome consists of the simultaneous exploration of both the biological (i.e., mutations of the protein scaffold) and chemical (i.e., nature of the cofactor and its substituents) spaces so that calculations could ascertain the most promising complementarity of the protein−substrate−cofactor triad. In this, there is no doubt that the integration of different methodologies will be crucial to success.

AUTHOR INFORMATION

Corresponding Author

*E-mail: JeanDidier.Marechal@uab.cat.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support from the Spanish Ministerio de Economia y Competitividad (Project CTQ2014-54071-P) is acknowledged. V.M.R. is grateful to the Spanish MINECO for a FPI Fellowship. J.D.M. also deeply thanks Profs. Mahy and Ward and Dr. Ricoux for their introduction to the world of artificial metalloenzymes.

REFERENCES

(1) Golynskiy, M. V.; Seelig, B. *Trends Biotechnol.* **2010**, *28*, 340−345.
(2) Okrasa, K.; Kazlauskas, R. J. *Chem.—Eur. J.* **2006**, *12*, 1587−1596.
(3) Lin, Y.-W.; Yeung, N.; Gao, Y.-G.; Miner, K. D.; Lei, L.; Robinson, H.; Lu, Y. *J. Am. Chem. Soc.* **2010**, *132*, 9970−9972.
(4) Jing, Q.; Okrasa, K.; Kazlauskas, R. J. *Chem.—Eur. J.* **2009**, *15*, 1370−1376.
(5) Podtetenieff, J.; Taglieber, A.; Bill, E.; Reijerse, E. J.; Reetz, M. T. *Angew. Chem., Int. Ed.* **2010**, *122*, 5277−5281.
(6) Köhler, V.; Wilson, Y. M.; Lo, C.; Sardo, A.; Ward, T. R. *Curr. Opin. Biotechnol.* **2010**, *21*, 744−752.
(7) Rosati, F.; Roelfes, G. *ChemCatChem* **2010**, *2*, 916−927.
(8) Dürrenberger, M.; Heinisch, T.; Wilson, Y. M.; Rossel, T.; Nogueira, E.; Knörr, L.; Mutschler, A.; Kersten, K.; Zimbron, M. J.; Pierron, J.; Schirmer, T.; Ward, T. R. *Angew. Chem., Int. Ed.* **2011**, *123*, 3082−3085.
(9) Pordea, A.; Creus, M.; Panek, J.; Duboc, C.; Mathis, D.; Novic, M.; Ward, T. R. *J. Am. Chem. Soc.* **2008**, *130*, 8085−8088.
(10) Heinisch, T.; Ward, T. R. *Curr. Opin. Chem. Biol.* **2010**, *14*, 184−199.
(11) Williams, R. J. P. *Chem. Commun.* **2003**, 1109−1113.
(12) Ringenberg, M. R.; Ward, T. R. *Chem. Commun.* **2011**, *47*, 8470−8476.
(13) Esmieu, C.; Cherrier, M. V.; Amara, P.; Girgenti, E.; Marchi-Delapierre, C.; Oddon, F.; Iannello, M.; Jorge-Robin, A.; Cavazza, C.; Ménage, S. *Angew. Chem., Int. Ed.* **2013**, *52*, 3922−3925.
(14) Bos, J.; García-Herraiz, A.; Roelfes, G. *Chem. Sci.* **2013**, *4*, 3578−3582.
(15) Bos, J.; Fusetti, F.; Driessen, A. J. M.; Roelfes, G. *Angew. Chem., Int. Ed.* **2012**, *51*, 7472−7475.
(16) Chevalley, A.; Salmain, M. *Chem. Commun.* **2012**, *48*, 11984−11986.
(17) Inaba, H.; Kanamaru, S.; Arisaka, F.; Kitagawa, S.; Ueno, T. *Dalton Trans.* **2012**, *41*, 11424−11427.
(18) Köhler, V.; Mao, J.; Heinisch, T.; Pordea, A.; Sardo, A.; Wilson, Y. M.; Knörr, L.; Creus, M.; Prost, J.-C.; Schirmer, T.; Ward, T. R. *Angew. Chem., Int. Ed.* **2011**, *50*, 10863−10866.
(19) Boersma, A. J.; Coquière, D.; Geerdink, D.; Rosati, F.; Feringa, B. L.; Roelfes, G. *Nat. Chem.* **2010**, *2*, 991−995.
(20) Roelfes, G.; Feringa, B. L. *Angew. Chem., Int. Ed.* **2005**, *44*, 3230−3232.
(21) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Chem. Theor* **2007**, *3*, 1960−1986.
(22) Deeth, R. J.; Anastasi, A.; Diedrich, C.; Randell, K. *Coord. Chem. Rev.* **2009**, *253*, 795−816.
(23) Harvey, J. N. *Struct. Bond.* **2004**, *112*, 151−184.

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

(24) Swart, M.; Groenhof, R.; Ehlers, A. W.; Lammertsma, K. *J. Phys. Chem. A* **2004**, *108*, 5479−5483.
(25) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157−167.
(26) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(27) Bathelt, C. M.; Mulholland, A. J.; Harvey, J. N. *J. Phys. Chem. A* **2008**, *112*, 13149−13156.
(28) Oláh, J.; Mulholland, A. J.; Harvey, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6050−6055.
(29) Piazzetta, P.; Marino, T.; Russo, N. *Inorg. Chem.* **2014**, *53*, 3488−3493.
(30) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185−199.
(31) Bathelt, C. M.; Zurek, J.; Mulholland, A. J.; Harvey, J. N. *J. Am. Chem. Soc.* **2005**, *127*, 12900−12908.
(32) Martí, S.; Andrés, J.; Silla, E.; Moliner, V.; Tuñón, I.; Bertrán, J. *Angew. Chem., Int. Ed.* **2007**, *46*, 286−290.
(33) Lonsdale, R.; Rouse, S. L.; Sansom, M. S. P.; Mulholland, A. J. *PLOS Comput. Biol.* **2014**, *10*, e1003714.
(34) Shi, R.; Li, W.; Liu, G.; Tang, Y. *Chin. J. Chem.* **2013**, *31*, 1219−1227.
(35) Miki, Y.; Pogni, R.; Acebes, S.; Lucas, F.; Fernández-Fueyo, E.; Baratto, M. C.; Fernández, M. I.; de los Ríos, V.; Ruiz-Dueñas, F. J.; Sinicropi, A.; Basosi, R.; Hammel, K. E.; Guallar, V.; Martínez, A. T. *Biochem. J.* **2013**, *452*, 575−584.
(36) Piazzetta, P.; Marino, T.; Russo, N. *Phys. Chem. Chem. Phys.* **2014**, *16*, 16671−16676.
(37) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. *Angew. Chem., Int. Ed.* **2013**, *52*, 5700−5725.
(38) Koder, R. L.; Anderson, J. L. R.; Solomon, L.; Reddy, K. S.; Moser, C. C.; Dutton, P. L. *Nature* **2009**, *458*, 305−9.
(39) Dieckmann, G. R.; McRorie, D. K.; Tierney, D. L.; Utschig, L. M.; Singer, C. P.; O'Halloran, T. V.; Penner-Hahn, J. E.; DeGrado, W. F.; Pecoraro, V. L. *J. Am. Chem. Soc.* **1997**, *119*, 6195−6196.
(40) Ghosh, D.; Pecoraro, V. L. *Curr. Opin. Chem. Biol.* **2005**, *9*, 97−103.
(41) Tantillo, D. J.; Houk, K. N. *Curr. Opin. Chem. Biol.* **1998**, *2*, 743−750.
(42) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82−87.
(43) Bolon, D. N.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14274−14279.
(44) Lassila, J. K.; Privett, H. K.; Allen, B. D.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16710−16715.
(45) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 3790−3795.
(46) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. a; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* **2008**, *453*, 190−195.
(47) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St. Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. *Science* **2010**, *329*, 309−313.
(48) Zanghellini, A.; Jiang, L. I. N.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Ro, D. *Protein Sci.* **2006**, *15*, 2785−2794.
(49) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364−1368.
(50) Meiler, J.; Baker, D. *Proteins* **2006**, *65*, 538−548.
(51) Lind, M. E. S.; Himo, F. *Angew. Chem., Int. Ed.* **2013**, *125*, 4661−4665.
(52) Ke, Z.; Abe, S.; Ueno, T.; Morokuma, K. *J. Am. Chem. Soc.* **2012**, *134*, 15418−15429.
(53) Frushicheva, M. P.; Warshel, A. *ChemBioChem* **2012**, *13*, 215−223.
(54) Robles, V. M.; Ortega-Carrasco, E.; Fuentes, E. G.; Lledós, A.; Maréchal, J.-D. *Faraday Discuss.* **2011**, *148*, 137−159.
(55) Muñoz Robles, V.; Vidossich, P.; Lledós, A.; Ward, T. R.; Maréchal, J.-D. *ACS Catal.* **2014**, *4*, 833−842.
(56) Allard, M.; Dupont, C.; Muñoz Robles, V.; Doucet, N.; Lledós, A.; Maréchal, J.-D.; Urvoas, A.; Mahy, J.-P.; Ricoux, R. *ChemBioChem* **2012**, *13*, 240−251.
(57) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins* **2003**, *52*, 609−623.
(58) Kirton, S. B.; Murray, C. W.; Verdonk, M. L.; Taylor, R. D. *Proteins* **2005**, *58*, 836−844.
(59) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T., Jr.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; C, G.; Pople, J. A. *Gaussian 09*; Revision D.01, Gaussian, Inc.: Wallingford, CT, 2009.
(60) Dapprich, S.; Komaromi, I.; Byun, S. K.; Morokuma, K. S.; Frisch, M. J. *J. Mol. Struct. THEOCHEM* **1999**, *461−462*, 1−21.
(61) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(62) Chase, F.; Avenue, B. *Curr. Opin. Struct. Biol.* **2002**, *12*, 431−440.
(63) Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. J. *Protein Eng.* **1996**, *9*, 1063−1065.
(64) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2009**, *31*, 455−461.
(65) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(66) Seebeck, B.; Reulecke, I.; Kämper, A.; Rarey, M. *Proteins* **2008**, *71*, 1237−54.
(67) Meggers, E. *Chem. Commun.* **2009**, 1001−1010.
(68) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(69) Ortega-Carrasco, E.; Lledós, A.; Maréchal, J.-D. *J. Comput. Chem.* **2014**, *35*, 192−198.
(70) Mahy, J.-P.; Maréchal, J.-D.; Ricoux, R. *Chem. Commun.* **2015**, *51*, 2476−2494.
(71) Ricoux, R.; Dubuc, R.; Dupont, C.; Maréchal, J.-D.; Martin, A.; Sellier, M. *Bioconjugate Chem.* **2008**, *19*, 899−910.
(72) Ricoux, R.; Allard, M.; Dubuc, R.; Dupont, C.; Maréchal, J.-D.; Mahy, J. P. *Org. Biomol. Chem.* **2009**, *7*, 3208−3211.
(73) Allard, M.; Dupont, C.; Muñoz Robles, V.; Doucet, N.; Lledós, A.; Maréchal, J.-D.; Urvoas, A.; Mahy, J.-P.; Ricoux, R. *ChemBioChem* **2012**, *13*, 240−251.
(74) Drevelle, A.; Graille, M.; Heyd, B.; Sorel, I.; Ulryck, N.; Pecorari, F.; Desmadril, M.; van Tilbeurgh, H.; Minard, P. *J. Mol. Biol.* **2006**, *358*, 455−471.
(75) Sansiaume-Dagousset, E.; Urvoas, A.; Chelly, K.; Ghattas, W.; Maréchal, J.-D.; Mahy, J.-P.; Ricoux, R. *Dalt. Trans.* **2014**, *43*, 8344−8354.
(76) Urvoas, A.; Ghattas, W.; Maréchal, J.-D.; Avenier, F.; Bellande, F.; Mao, W.; Ricoux, R.; Mahy, J.-P. *Bioorgan. Med. Chem.* **2014**, *22*, 5678−5686.
(77) Muñoz Robles, V.; Maréchal, J.-D.; Bahloul, A.; Sari, M.-A.; Mahy, J.-P.; Golinelli-Pimpaneau, B. *PLoS One* **2012**, *7*, e51128.

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480

(78) Ueno, T.; Yokoi, N.; Unno, M.; Matsui, T.; Tokita, Y.; Yamada, M.; Ikeda-Saito, M.; Nakajima, H.; Watanabe, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9416−9421.
(79) Ortega-Carrasco, E.; Lledós, A.; Maréchal, J. *J. R. Soc. Interface* **2014**, *11*, 20140090.
(80) Muñoz Robles, V.; Dürrenberger, M.; Heinisch, T.; Lledós, A.; Schirmeister, T.; Ward, T. R.; Maréchal, J.-D. *J. Am. Chem. Soc.* **2014**, *136*, 15676−15683.
(81) Toledo, L.; Masgrau, L.; Maréchal, J.-D.; Lluch, J. M.; González-Lafont, A. *J. Phys. Chem. B* **2010**, *114*, 7037−7046.
(82) Maréchal, J.; Yu, J.; Brown, S.; Kapelioukh, I.; Rankin, E. M.; Wolf, C. R.; Roberts, G. C. K.; Paine, M. J. I.; Sutcliffe, M. J. *Drug Metab. Dispos.* **2006**, *34*, 534−538.
(83) McLaughlin, L.; Paine, M. J. I.; Kemp, C.; Maréchal, J.-D.; Flanagan, J. U.; Ward, C. J.; Sutcliffe, M. J.; Roberts, G. C. K.; Wolf, C. R. *J. Biol. Chem.* **2005**, *280*, 38617−38624.
(84) Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. *J. Chem. Theory Comput.* **2005**, *1*, 1304−1311.
(85) Amaro, R. E.; Baron, R.; McCammon, J. A. *J. Comput. Aided Mol. Des.* **2008**, *22*, 693−705.

DOI: 10.1021/acscatal.5b00010
ACS Catal. 2015, 5, 2469−2480