

Running title: Mapping enzyme-substrate interactions

# ***Mapping enzyme-substrate interactions: Its potential to study the mechanism of enzymes***

Sergi Rodà<sup>1</sup>, Gerard Santiago<sup>1</sup>, Víctor Guallar<sup>1,2,\*</sup>

<sup>1</sup>Barcelona Supercomputing Center (BSC), Barcelona, Spain

<sup>2</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

\* Corresponding author

e-mail: [victor.guallar@bsc.es](mailto:victor.guallar@bsc.es)

phone number: +34 93 413 7727

ORCIDs of authors:

Sergi Roda: 0000-0002-0174-7435

Gerard Santiago: 0000-0002-0506-3049

Victor Guallar: 0000-0003-3274-2482

## **Abstract**

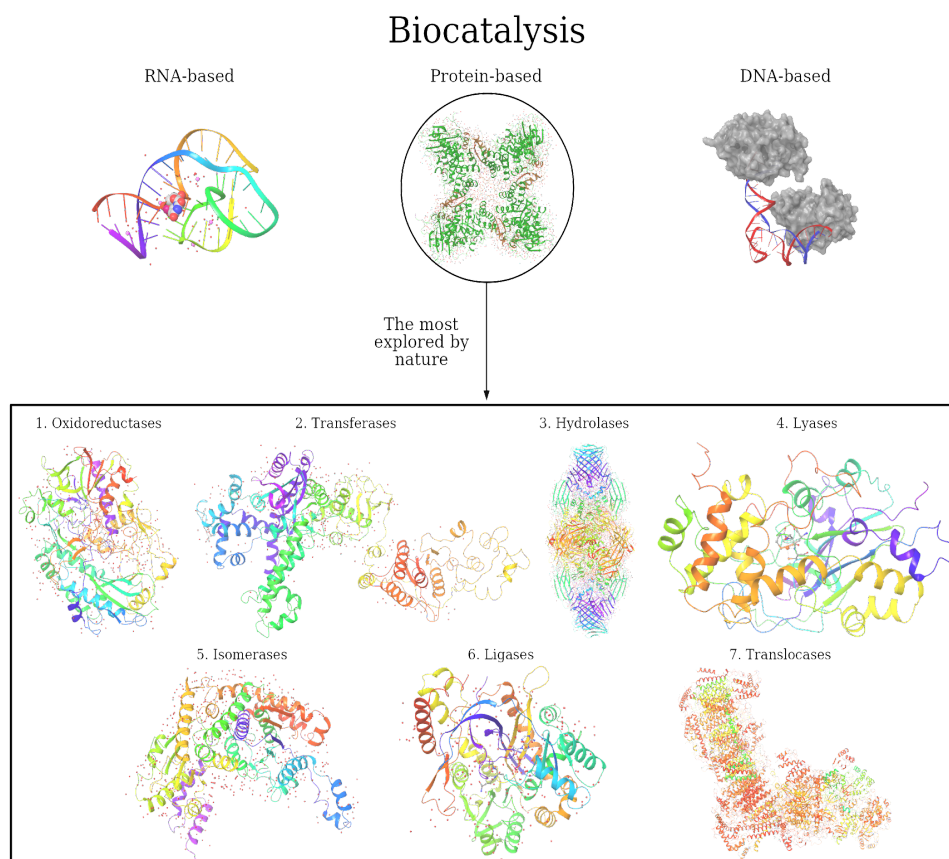
With the increase of the need to use more sustainable processes for the industry in our society, the modeling of enzymes has become crucial to fully comprehend their mechanism of action and use this knowledge to enhance and design their properties. A lot of methods to study enzymes computationally exist and they have been classified on sequence-based, structure-based, and the more new artificial intelligence-based ones. Albeit the abundance of methods to help predict the function of an enzyme, molecular modeling is crucial when trying to understand the enzyme mechanism, as they aim to correlate atomistic information with experimental data. Among them, methods that simulate the system dynamics at a molecular mechanics level of theory (classical force fields) have shown to offer a comprehensive study. In this book chapter, we will analyze these techniques, emphasizing the importance of precise modeling of enzyme-substrate interactions. In the end, a brief explanation of the transference of the information from research studies to the industry is given accompanied with two examples of family enzymes where their modeling has helped their exploitation.

**Keywords:** “enzymology”, “computational chemistry”, “enzyme engineering”, “protein dynamics”, “PELE”, “enzyme-substrate”, “molecular modeling”

## 1) Why modeling enzyme structure and dynamics

To face the current world challenges, human society has to adapt towards an eco-friendly based economy, where industrial production transformation is a milestone. In the field of chemistry, enzymes are becoming the alternative for inorganic catalysts, used in a wide variety of industrial applications (Chapman et al., 2018; Schmid et al., 2001; Rajendra Singh et al., 2016). In organisms, enzymes are the elements that make life kinetically possible. Otherwise, reactions needed for life would occur at rates in which life and the current atmosphere and ecosystems would not exist (Neet, 1998; Richard Wolfenden, 2006; R. Wolfenden & Snider, 2001). For instance, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), the most abundant enzyme on Earth (Dhingra et al., 2004), is the major catalyst involved in the carbon fixation of carbon dioxide.

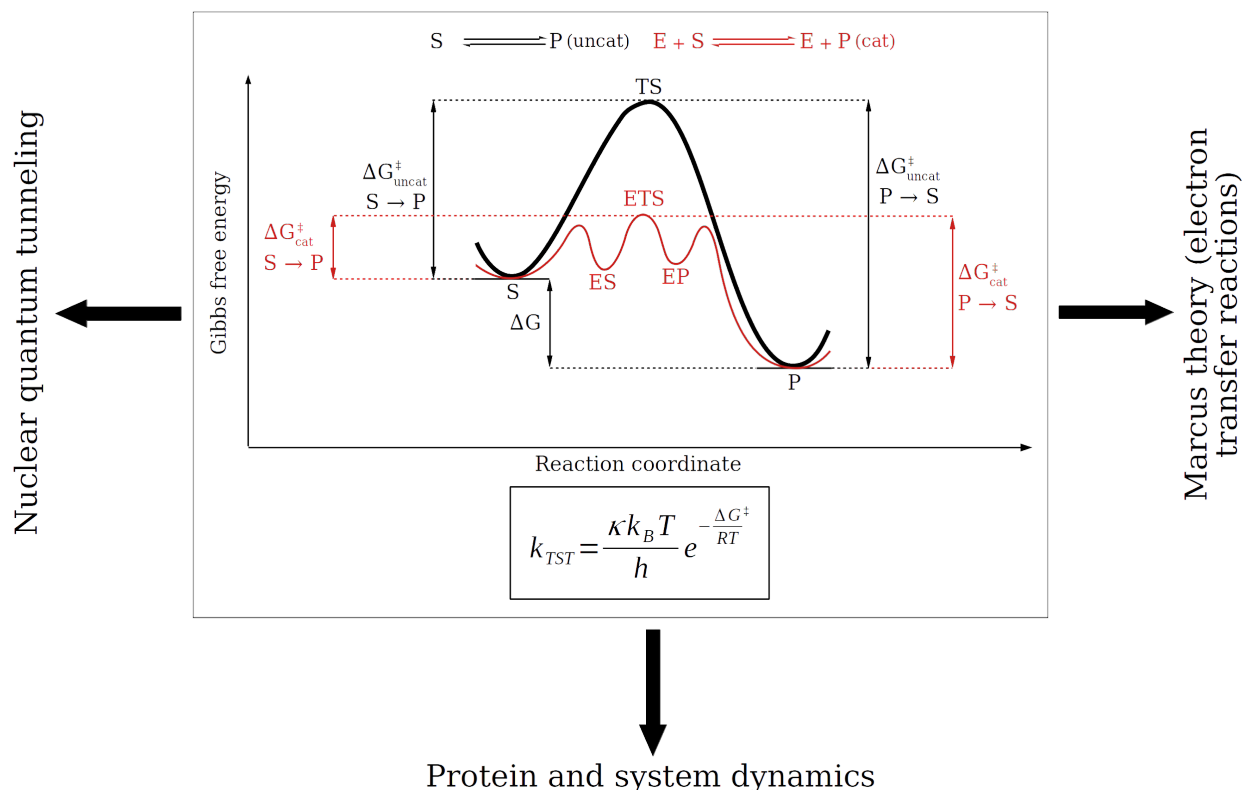
Enzymes can be of other macromolecular nature besides the proteic one, including catalytic RNA and even designed catalytic DNA (Breaker, 1997; Guerrier-Takada et al., 1983; Kruger et al., 1982). However, protein-based catalysts are the most abundant in nature due to the fact that 21 amino acids offer incredibly a lot more combinations in comparison with just 4 nitrogenous bases (Böck et al., 1991; Doudna & Lorsch, 2005). As a result of this high variety of amino acids and of the presence of cofactors in a big part of enzymes, a large number of catalyzed reactions by enzymes exist. This wide variety of protein-based enzymes brings in different types of activities along with many effects that can modulate them. In other words, enzyme activity is affected by pH, temperature, ionic concentration, presence or absence of the cofactor, and by the action of activator or inhibitor molecules (Robinson, 2015).



**Figure 1: Macromolecular nature of biocatalysis.** The figure represents the types of macromolecules where catalysis has been observed or designed, which are DNA, RNA, and proteins. Protein-based catalysis is the most explored by nature, due to the larger options that the number of residues and folding types give. At the bottom, the 7 types of global chemical reactions catalyzed by enzymes (classification of Enzyme Commission number) are represented with a 3D structure of each group. The rainbow labeling of the ribbon in all structures highlights from the N-terminal (or 5' end in RNA) in red to the C-terminal (or 3' end in RNA) in violet. The encircled enzyme is colored in green for the large subunit and in brown for the small one. The DNA of the DNAzyme has each strand stained with a different color. The PDB codes for all structures from the RNA-based to the Translocases are the following: 4OJI, 1GK8, 5XM9, 1CF3, 1TAQ, 6CVM, 1AHJ, 1IAT, 2HGS, and 6RFR.

Notwithstanding, the mechanisms and factors that give enzymes the ability to catalyze reactions are still not fully comprehended (Knowles, 1991; Kraut, 1988). The widely known and most accepted idea to explain enzyme catalysis is the transition state theory (TST) (Eyring, 1935; Truhlar, 2015; Zinovjev & Tuñón, 2017). Enzymes can either decrease the Gibbs free energy of activation ( $\Delta G^\ddagger$ ) or increase the transmission coefficient ( $\kappa$ ), accelerating the rate of the reaction in comparison with an uncatalyzed reaction. Thus, enzymes either stabilize the transition state (TS) of the reaction or they enhance the productive cross of the TS barrier by the reactants (Agarwal, 2006; Truhlar, 2015; Zinovjev & Tuñón, 2017).

Yet, the TST has some limitations, giving birth to other approaches and theories to study the enzyme catalysis of a reaction such as the Marcus theory (Marcus, 1956; Zinovjev & Tuñón, 2017). Moreover, the concrete way in which enzymes decrease  $\Delta G^\ddagger$  or increase  $\kappa$  of the TST remains a highly discussed topic. One of the important factors is the dynamics of proteins, which seems to be the missing piece to understand how enzymes catalyze reactions (Agarwal, 2006; Carvalho et al., 2014; Kamerlin & Warshel, 2010; Kohen, 2015; Petrović et al., 2018). Experimental studies enable to check the activity of an enzyme, its specificity or promiscuity, its optimal conditions, and even get the structure of the system. Still, a full mechanistic knowledge requires the study of the protein dynamics and of the molecular interaction with the substrate, a task that mandates the use of different computational methods (Carvalho et al., 2014).



**Figure 2: The state of the art and the current issues related to biocatalysis.** The figure presents the typical free energy profile of a reaction uncatalyzed (in black) versus catalyzed (in red) followed by arrows pointing towards the current issues in the TST. Enzymes tend to decrease the value of  $\Delta G^\ddagger$ , increasing the reaction rate. A global mechanism is the formation of the enzyme-substrate (ES) complex, followed by the stabilization of the transition state (TS) of the reaction in the active site (ETS), ending with the releasing of the product (P) from the enzyme (EP). Enzymes can also increase the transmission coefficient ( $\kappa$ , the proportion of reactive trajectories that surpass the TS barrier and end as products). The reaction coordinate is

a metric that represents the progress of the reaction according to some geometric parameters along the biochemical process.

The study of enzymes and their mechanisms has helped in comprehending evolution better, in designing systems with enzymes to diagnose, treat, and cure several diseases, and in synthesizing or refining a wide variety of materials and compounds, which are interesting for different industrial sectors (Chapman et al., 2018; Robinson, 2015; Schmid et al., 2001; Rajendra Singh et al., 2016). Some well-known examples include the use of glucose oxidase to measure and monitor the blood glucose level in patients with diabetes mellitus (Yoo & Lee, 2010), the production of lactose-free milk by using lactase for lactose-intolerant people (Soares et al., 2012), and the invention and success of the PCR technique (Mullis, 1990), which is crucial at any laboratory working with molecular biology, thanks to the discovery of the thermophile *Thermus aquaticus* DNA polymerase (Chien et al., 1976).

More examples are the synthesis of acrylamide, the precursor of the polymer famously used to perform Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) and water treatment (Laemmli, 1970; Smith & Oehme, 1991), is possible due to nitrile hydratase (Yamada & Kobayashi, 1996) and the common features observed in the ATP-grasp enzymes. These features are base residues to stabilize the  $\alpha$  and  $\beta$ -phosphates, an acid residue that forms a hydrogen bond with the amino group in adenine, hydrophobic residues to stabilize adenine ring, three acid residues to coordinate  $Mg^{2+}$  ions, and a shared catalytic mechanism (Galperin & Koonin, 1997, 2012; Thoden et al., 1999).

Moreover, for many systems such a gain of knowledge has been largely promoted by *in silico* studies. For instance, the design of DNA polymerases to accept unnatural nucleoside triphosphates would not have been able without enzyme modeling (F. Chen et al., 2010; Dunn et al., 2016). Thus, addressing the enzymatic mechanisms of action has significantly been revamped through modeling their dynamics and enzyme-substrate interactions, leading to an increased understanding of the system and better design of their potential applications.

## 2) A method for everybody

As the title of the section entails, a large amount of methods used in science can be applied to the study of enzymes. Thus, enzymology requires physicists, mathematicians, and data scientists, besides the chemists and biologists, which are the most associated with the field. This chapter is focused on the computational tools used to study enzymes, many representative methods are summarized in **Table 1**.

Computational methods use either the sequence, the structure or experimental data, through analysis or simulations acting on it, to unveil the catalytic activity of enzymes; different levels might reveal key (different) features of the mechanism of action. Moreover, data extracted from

these methods (along with experimental one) give the option to use artificial intelligence (AI), and more concretely machine learning (ML), to study them as well.

## 2.1 Sequence-based approaches

Even though enzymes tend to be studied using its 3D structure, a lot of information can be extracted from its sequence. Still, all the knowledge is obtained by comparison with one or other multiple sequences (Suplatov et al., 2015). The input of these techniques can be either the DNA coding for the mRNA of the enzyme or the protein primary structure.

Enzyme sequences can be found in several databases, but the main and best curated ones are NCBI and UniProtKB (*National Center for Biotechnology Information*, n.d.; UniProt Consortium, 2019). To search those homologous sequences to the query one in these databases, BLAST (Basic Local Alignment Search Tool) is commonly used (Altschul et al., 1990). Albeit the popularity of this algorithm of sequence comparison, many others exist including SSEARCH (Pearson, 1991), USEARCH (Edgar, 2010), and SANS (Koskinen & Holm, 2012).

Once you have the set of selected homologous sequences, they can be aligned with different multiple sequence alignment (MSA) tools like CLUSTAL (Higgins & Sharp, 1988; Thompson et al., 1994), T-COFFEE (Notredame et al., 2000), and Kalign (Lassmann & Sonnhammer, 2005) in which the progressive alignment approach is used (Feng & Doolittle, 1987). Other approaches exist (Chowdhury & Garai, 2017; Daugelaite et al., 2013) and give room to other popular MSA softwares such as MAFFT (Kato et al., 2002) or MUSCLE (Edgar, 2004).

MSA gives a lot of hidden structural information of the enzyme including correlated residues in the sequence that coevolve, the essential residues of the active site of a particular family of enzymes, the less conserved residues in the structure of the protein, and thus, more prone to be mutable. Thus, the comparison of homologous sequences with the studied one enables to predict substantial information of the enzyme's function. In fact, a lot of softwares used in enzyme characterization and engineering use MSA as the input. Some examples include: HotSpotWizard (Sumbalova et al., 2018), PoPMuSiC (Gilis & Rooman, 2000), or ConSurf server (Ashkenazy et al., 2016) to name a few.

## 2.2 Structure-based approaches

Enzymes structural features are key to decipher the catalytic mechanism. Inside the tools related with the structural rearrangement of the residues of the protein, two main groups have to be made. The first and most global one is molecular modeling, which refers to methods aiming at manipulating the structure of the molecules in a system at a particular level of theory of physics to infer its properties. The other subgroup encompasses other tools that work with structures at the level of analysis and data extraction, without performing large simulations, what

is commonly known as structure-based bioinformatics. Still, the techniques of these two subgroups can be combined to study enzymes.

The use of structure-based bioinformatics to study enzymes and their mechanism of action is largely related with structural alignment, to compare two similar proteins (Godzik, 1996; Ma & Wang, 2014), homology modeling to obtain a 3D structure from a target sequence and a template structure (Chothia & Lesk, 1986; Martí-Renom et al., 2000), detection of cavities and pockets (Le Guilloux et al., 2009), residue flexibility associated with the experimental B-factor (a parameter related with thermal motion and it is obtained in X-ray crystallography) (Debye, 1913), prediction of substrate promiscuity (Martínez-Martínez et al., 2018), use of statistical potentials to score and rank protein-ligand interactions (Fan et al., 2011; Mooij & Verdonk, 2005), and more.

### **2.2.1 Molecular modeling**

Molecular modeling is one of the most widespread techniques to study (bio)chemical and physical systems outside the wet laboratory. The methods can be classified either if they simulate the system with classical physics (referring to the Newton's laws), what it is called molecular mechanics (MM) or if they take quantum physics (referring to the Schrödinger's equation) into consideration, what it is known as quantum mechanics (QM). The treatment of the system in a more simple level of theory enables the computation to be less expensive compared with using the level of theory of quantum physics. Nonetheless, a higher level of theory returns more accurate calculi and better estimations of the user's parameters of interest (Jensen, 2007).

MM simplifies the system by treating atoms as the smallest and indivisible unit without taking into account subatomic particles. Atoms are connected by bonds, which are considered harmonic oscillators. This description of molecules with atoms and their bonds is known as the "ball-and-stick" model. Thus, the potential energy of the system is parametrized according to the atomic coordinates, the type of atoms, and the bonds made between them, what is commonly known as force fields (FF).

The numerical values assigned to the constants of the energy terms, the functional form used, and the number of parameters and terms taken into account is what defines and distinguishes a FF from another. Thus, the accuracy of the energy calculation of the user's system will depend in principle on the applied FF; at a practical level most FF give similar results on standard systems. FF are designed from experimental data or computational data obtained from a higher level of theory, such as QM; a handful of widely used FF exist, including AMBER (Cornell et al., 1995), CHARMM (Brooks et al., 1983), OPLS (Jorgensen et al., 1983), or GROMOS (van Gunsteren & Berendsen, 1987).

QM methods describe the system in the most inner part of particles, electrons and nuclei, and the Schrödinger equation must be solved. Different ways to solve the Schrödinger equation exist and they give room to different levels of accuracy and computational resources, depending on the approximation used to solve it. Semi-empirical methods, for example, use experimental data to reduce the computational cost of solving the Schrödinger equation; methods not using empirical data are called *ab initio* methods. These include Density Functional Theory (DFT) (Kohn & Sham, 1965) and Hartree Fock (HF) methods (Fock, 1930; Hartree, 1928). However, these methods usually only consider the average interaction between electrons. There are more precise methods that do take into account electron correlation, which are Configuration Interaction (Head-Gordon et al., 1994), Møller–Plesset perturbation theory (Møller & Plesset, 1934), or Coupled Cluster (Purvis & Bartlett, 1982).

Therefore, MM enables the study of large systems over a long period of time, while QM gives an accurate description of small systems at a stationary state or over (very) small periods of time. Computational enzymology also takes advantage of both ways, QM/MM techniques, to describe the system and to study its catalytic mechanism (Friesner & Guallar, 2005; Senn & Thiel, 2009; van der Kamp & Mulholland, 2013). The active site residues, (the cofactor), and the substrate/s are treated typically at the QM level and the rest of the system is described at the MM level.

Molecular modeling has led to two Nobel Prizes in Chemistry. One in 1998 for the development of DFT and the computational methods for QM by Walter Kohn and John A. Pople, respectively. The other in 2013 for the development of the QM/MM methods by Martin Karplus, Michael Levitt, and Arieh Warshel.

Typical problems in computational chemistry require the exploration of the potential energy of a system according to its spatial coordinates. The number of possible local minima, and thus, number of possible stable configurations of a molecule increase exponentially as the number of bonds increases (Howard & Kollman, 1988; Maranas & Floudas, 1994). This combinatorial explosion problem gives room to several approaches including molecular dynamics (MD), Monte Carlo (MC) methods, simulated annealing, genetic algorithms, diffusion methods, or distance geometry methods. From these, MD and MC are the best to sample locally, finding a good number of local minima. On the other hand, the remaining methods tend to be used to search for the global minimum of the system.

The concept of these approaches and their application to solve real problems in computational chemistry is a big topic. A lot of books and reviews exist to get further information (Hollingsworth & Dror, 2018; Jensen, 2007; Karplus & McCammon, 2002; Lei & Duan, 2007; Lonsdale et al., 2012; Maximova et al., 2016; Yang et al., 2016).

Thus, the availability of the enzyme structure enables the proper modeling of a conformational sampling and/or dynamics related to the catalyzed reaction, gaining further insights of the key residues involved with the mechanism of action and substrate binding. Later on, this knowledge can be used to design a particular enzyme to enhance its activity against a specific family of



substrates or to extend the conditions in which the enzyme can work (harsher pHs and temperatures), what it is known as enzyme engineering.

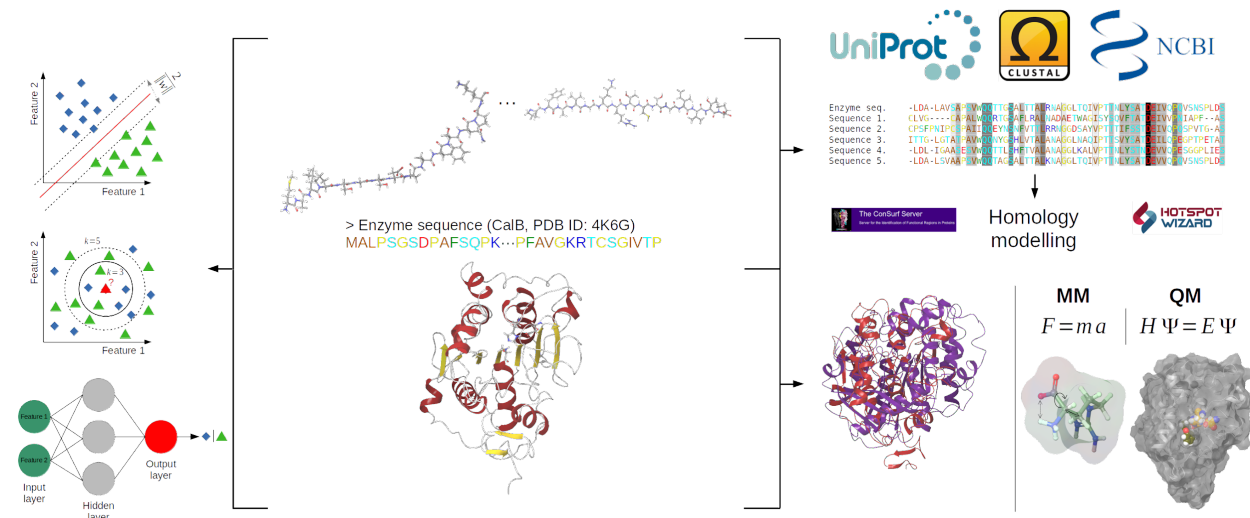
### **2.3 AI-based approaches**

In spite of the fact that AI is still entering the field of enzymology, several number of applications to study protein properties in general have been developed (Larrañaga et al., 2006; Paladino et al., 2017). One of the newest and most interesting ones refers to the AlphaFold built by Google to predict the 3D structure of a protein with its sequence as the only input (Senior et al., 2020). This method uses a convolutional neural network that predicts distance between residues and torsions of the backbone of the protein structure from its sequence. With these predictions, a protein-specific potential is created, and then, is further minimized by gradient descent, obtaining the final predicted structure. Another related method refers to the one published by AlQuraishi, which uses a recurrent neural network and parametrizes the local protein structure into torsional angles that take into account covalent chemistry. Subsequently, the predicted local protein structures are coupled into the overall protein structure with recurrent geometric units. Finally, the model uses a differentiable loss function to score the predicted structure over the experimental one and train the model. Thus, this model does not use co-evolution information and shares similar results with other protein structure prediction algorithms (AlQuraishi, 2019). The reach of predictions close to experimental structures would be a rather important aspect for the field of computational enzymology, since it would enable to perform molecular modeling studies of any enzyme, whether the experimental 3D structure is available or not.

The prediction of protein functions and the functional relevance of their amino acids in the 3D structure is crucial to understand and design enzymes. In that regard, DeeProtein software has been created, which also takes the sequence as input and infers the regions of the protein sequence that are important for the biological activity (Belzen et al., 2019). Another model regarding function has been created based on the available enzymes in the RCSB PDB to predict the Enzyme Commission number (Amidi et al., 2018).

The prediction of the optimal pH range has also been accomplished with the use of ML (Khan et al., 2015), which could be really useful in the field of enzyme engineering. Another model consists in the prediction of the solubility of proteins based on their structure (Hou et al., 2020), which could be useful to avoid the trial of computationally promising variants but that would actually make the protein aggregate.

However, these techniques still need substantial improvement in order to really be used in the main projects of research centers and enterprises working with enzymes. These needs are often associated with the lack of enough high-quality datasets with no imbalances nor biases.



**Figure 3: Computational approaches to study an enzyme.** The figure summarizes the current methods used in computational enzymology to study catalytic mechanisms and more. The input of all methods consists in one of the levels of protein structure, from the primary level to the quaternary one. Thus, there are computational methods that work with the amino acid sequence of the enzyme to search for homologous sequences, perform a MSA, and use its information to infer properties of evolution/function (top right). Concerning the tertiary and quaternary structure of proteins, there are mainly two types of methods, structural bioinformatics and molecular modeling (bottom right). Lastly, the data obtained from computational and experimental studies related with enzymes and proteins can be used to train ML models to predict important features related with the mechanism of action of the enzyme (left). The sequence and structure can be found in the following PDB code: 4K6G. The aligned structure in the figure refers to this PDB code: 3GUU.

### 3) Enzymes mechanism: the potential of mapping enzyme-substrate interactions

As stated in the previous section, there are a plethora of different techniques to study and predict the relation between the sequence and/or structure with the function of enzymes. Moreover, these methods are often used in combination with amino acid mutations, in order to rationalize mechanistic data and, importantly, to design new variants with optimized (improved) properties. Still, an exhaustive analysis of the enzymatic mechanism demands typically the use of molecular modeling. Such study might be necessary when addressing substrate specificity, activity, etc.; sequence tools typically reduce to classification and conservation analysis.

A comprehensive molecular analysis should include three main aspects: enzyme dynamics, enzyme-substrate interactions, and electronic effects. The former aims at understanding the global dynamics of the enzyme, thus the flexibility associated with the active site, opening cavities, allosteric effects, etc. It is mainly studied through the use of MD and MC techniques

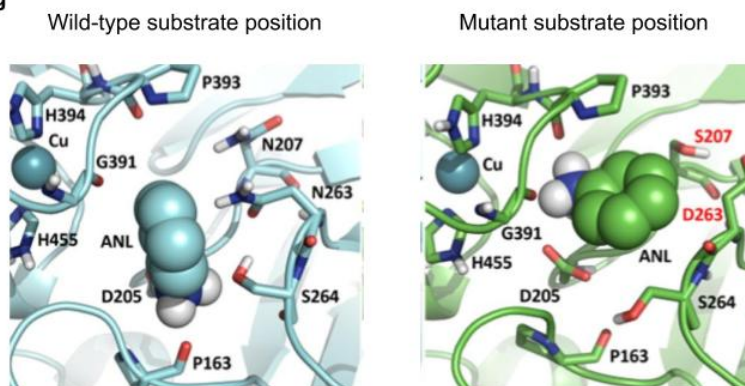
(Cossins et al., 2012); (simpler) coarse grain techniques, such as ANM, might offer an initial quick view. The importance of an exhaustive dynamical study has been discussed in multiple studies and reviews (Kamerlin & Warshel, 2010; Kohen, 2015). Electronic effects are obviously addressed with QM techniques and are important when studying the chemical reactive step and if addressing possible (non-reactive) polarization aspects. However, in this section we want to underline the importance of an accurate analysis of enzyme-substrate interactions. Since we developed the Protein Energy Landscape Exploration (PELE) software, a specialized protein-ligand heuristic MC technique first introduced in 2005, we have applied it now to more than 100 systems, with more than half of them being enzymes (Gilabert et al., 2018). Our experience indicates that many properties can be studied with a robust analysis of how the substrate migrates and interacts with the enzyme. Obviously this includes aspects such as substrate specificity but also reactivity studies, typically studied with more expensive QM techniques. In most of the attempted studies, we obtain a quantitative correlation of the enzymatic efficiency constants from the (statistical) description of the active site enzyme-substrate induced fit. Examples include comparing different substrates against the same enzyme (and variations of it), as in a fungal aryl-alcohol oxidase (Serrano et al., 2019), laccases (Monza et al., 2015; Pardo et al., 2016) and unspecific peroxidases (Molina-Espeja et al., 2016)), or comparing the same substrate to different enzymes, as in unspecific peroxidases (Aranda et al., 2019) and peroxygenases (F. Lucas et al., 2016). These studies all seem to indicate the role of preorganization and near attack conformation (Hur & Bruice, 2003).

The sampling potential of PELE, capable of probing protein-ligand induced fit while mapping the entire protein surface, also allows for advanced mechanistic studies. When studying two different manganese peroxidases, for example, we could find that the ABTS active one binds the substrate in an unexpected (and fully solvated) surface site, using two histidines as electrostatic anchors. The inactive enzyme, however, did not show any significant predicted binding sites, presenting an asparagine and a glycine at those surface positions. Introducing these two solvent exposed mutations, we obtained *in vitro* ABTS activity in a similar fashion (similar  $k_{cat}$  and  $K_m$ ) to the active enzyme (Acebes et al., 2016). Global (full enzyme) substrate binding explorations have also allowed adding artificial active sites in enzymes: the design of PluriZymes. By identifying alternative substrate binding sites (complementary to the active site), and turning them into additional active centers, after insertion of catalytic triads, we could introduce a second hydrolase activity on several esterases ([Santiago et al. 2018](#); [Alonso et al. 2020](#)).

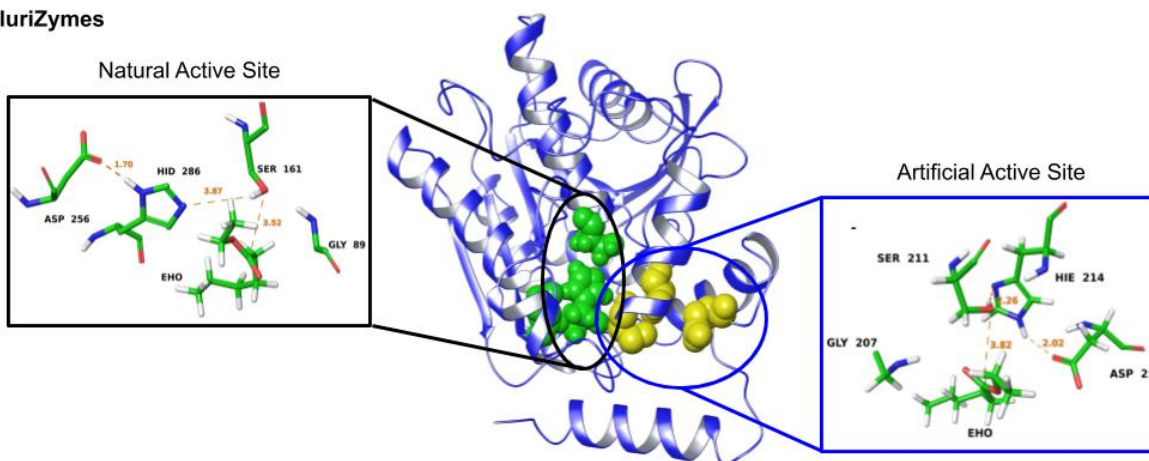
Thus, it seems like obtaining an exhaustive structural sampling of the enzyme-substrate interaction, even at the MM level of theory, seems to be a key aspect when addressing enzymatic catalysis. Obviously, adding more sophisticated levels of theory, such as an electronic description through mixed QM/MM techniques, might turn into a more robust analysis. We have implemented such an approach when aiming at correlation between redox substrate potential with enzymatic activity (M. F. Lucas et al., 2017), or when increasing a laccase oxidation on a difficult to oxidize substrate, the aniline cation (Santiago et al., 2016). Still, we are

turning more and more to exclusively using classical FF approaches when addressing enzyme-substrate interactions for mechanistic and engineering studies (Figure 4).

### Substrate Positioning



### PluriZymes



**Figure 4:** Examples of engineering success cases based on substrate positioning studies using classical force-fields. Introduced mutations are affecting the substrate position in the active site, creating a more favorable environment (substrate positioning) or to find a binding site where to create activity (PluriZymes). The substrate positioning panel has the substrate (aniline) displayed with the CPK model and the active site residues are displayed with the thick tube representation. The PluriZymes panel shows the protein scaffold and the zoomed areas highlight the natural and added active sites with the catalytic triad, the oxyanion hole, and the bound substrate predicted with PELE displayed with the thick tube representation. The PDB codes used are 5ANH (to perform homology modeling) and 5JD4, respectively.

## 4) From modeling to engineering

In the previous section, we already explored the potential of using a dynamical enzyme-substrate analysis to engineer enzymes. Here, we want to give a more broad perspective of how modeling can help and ease the enhancement of enzyme properties. In fact, different industrial sectors use the *in silico* modeling of enzymes combined with *in vitro* methods to

optimize their biocatalysts to real bioprocessing (Bornscheuer, U. T., & Pohl, M., 2001; Jemli et al., 2016; Kirk et al., 2002; Lutz, 2010). For instance, a wide list of engineered and newly characterized enzymes involved in the manufacturing of products of interest for the society exist, including different type of hydrolases for the detergent industry, hydrolases and some oxidoreductases for the food and beverage industry, and oxidoreductases, transferases, hydrolases, lyases, and isomerases for the pharmaceutical industry, the organic synthesis, and the waste management to name a few (Chapman et al., 2018; Rajendra Singh et al., 2016).

From the crystal structure (or homology model), key residues for substrate recognition can be identified. These identified residues are mutagenesis targets with beneficial effects towards enzyme activity, with some web tools already available for the *in silico* identification of these hotspots, like Swiss-Model (Schwede et al., 2003), Rosetta (Rohl et al., 2004), IntFOLD (McGuffin et al., 2015), Threading ASSEmblyRefinement (I-TASSER) (Roy et al., 2010), or Protein Homology/AnalogY Recognition Engine (Phyre) (Kelley & Sternberg, 2009). This approach is the first attempt to connect modeling with engineering.

The next step is to merge this structural information with the overall dynamic of the system. With this combination, modeling has become a powerful tool for enzyme engineering, as we will see in some examples focusing on cytochrome P450 (CYP) and transaminases (TA).

CYPs consist in a large superfamily of enzymes which can oxidize a broad range of substrates aiming at different physiological roles such as biosynthesis of sterols and other secondary metabolites or the clearance of toxic substances in the liver. Moreover, CYPs are enzymes that have been extensively studied with well-known catalytic mechanism and fold, including the heme cofactor (Denisov et al., 2005; Shaik et al., 2005). Thus, the knowledge of these biocatalysts and their capacity to catalyze oxidations of structurally complex molecules regioselectively and stereoselectively make them encouraging options for engineering of new chemical reactions using computational-based methods. Although directed evolution has had great success in the engineering of new chemistry in CYPs (Bajaj et al., 2016; K. Chen et al., 2018; Farwell et al., 2015; Hernandez et al., 2016; Kan et al., 2016; Prier et al., 2017; Ritesh Singh et al., 2015; Zhang et al., 2019), several studies of rational design aided with *in silico* tools have been performed (Ba et al., 2013; Dodani et al., 2016; Hayashi et al., 2008; Jung et al., 2018; Khatri et al., 2018; Rühlmann et al., 2017; Steck et al., 2020; Syed et al., 2013; Toporkova et al., 2013).

The complex nature of these enzymes has given researchers to modify and engineer different parts of the system including the substrate-binding active site, the substrate access channel, the residues involved in binding the heme cofactor, and more. Besides, these studies retail more insights of the structure-function relationships and have given substantial knowledge of the mechanism of action of CYPs. One remarkable example is the one where MD simulations revealed a key (unsolved crystallographically) loop that controlled its regioselectivity and by analyzing the different possible states of the loop, they obtained a mutant that switched the enzyme's regioselectivity (Dodani et al., 2016). A recent paper used a mechanism-guided

approach to design CYPs to substantially enhance their C-H amination (the best nitrene transfer biocatalysts reported) by disrupting conserved structural elements related with the heme cofactor (Steck et al., 2020). Another study uses a particular CYP and the binding poses of progesterone predicted by molecular docking to engineer the regioselective hydroxylation of the substrate with the products having different medical applications (Khatri et al., 2018). Many other studies in CYPs use the information of the enzyme-substrate interactions obtained with molecular docking to select those residues to mutate in order to enhance the selectivity or activity towards a specific family of substrates (Ba et al., 2013; Hayashi et al., 2008; Jung et al., 2018; Syed et al., 2013).

Despite the many industrial applications of CYPs, the electron source constant requirement is a major limitation for full implementation. Constant amounts of expensive cofactors have to be introduced in the reaction, just to keep CYP's operating. But recently discovered unspecific peroxidases (UPOs), might overcome CYPs. UPOs can speed up similar chemical reactions as CYPs, but using hydrogen peroxide as an electron source, is cheaper and more accessible compared to NADP (Freakley et al., 2019; Hofrichter & Ullrich, 2014; Hrycak & Bandiera, 2012).

Like in CYPs, UPOs have been recently studied using *in silico* methods to enhance their reactivity, where their application towards fatty acids has been discussed. The study of substrate positing has proved to be of extreme importance for UPOs, with a crucial role defining the catalyzed reaction. The global dynamic behaviour of the enzyme and the enzyme-substrate complex formation will be also relevant for UPOs engineering (Aranda et al., 2019).

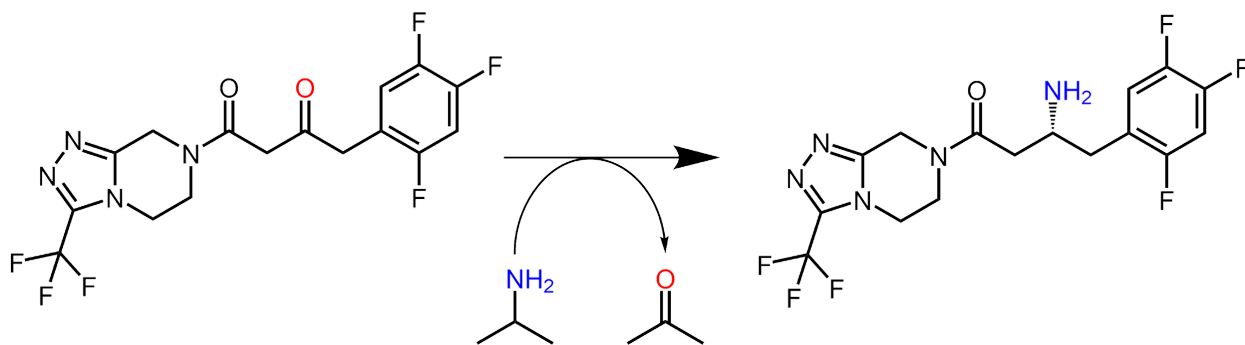
Other families have been the target of similar studies, like TAs. TAs are a superfamily of enzymes that catalyze the transfer of an amine group from a donor to an acceptor (usually a ketone or aldehyde) by using pyridoxal 5'-phosphate (PLP) to perform the oxidative deamination of the donor and the reductive amination of the acceptor (Ghislieri & Turner, 2014; Kelly et al., 2018; Malik et al., 2012). TAs are classified according to the relative position of the amine group to the carboxyl group of the substrate.  $\alpha$ TAs require the presence of a carboxyl group next to the C atom bound to the amine group, while  $\omega$ TAs does not have this requirement, making them more attractive for industrial bioprocessing due to the acceptance of a wider range of substrates. The active site geometry of  $\omega$ TAs is fully comprehended with a two-site binding model with a large (L) and small (S) pocket (Shin & Kim, 2002). There is a concept of linking the S pocket with the place involved in substrate binding but not with the catalysis, meaning that this region could be redesigned to enhance the acceptance of bulkier residues by the TA (Park & Shin, 2011). Still, the L pocket can be modified too (Dourado et al., 2016; Savile et al., 2010).

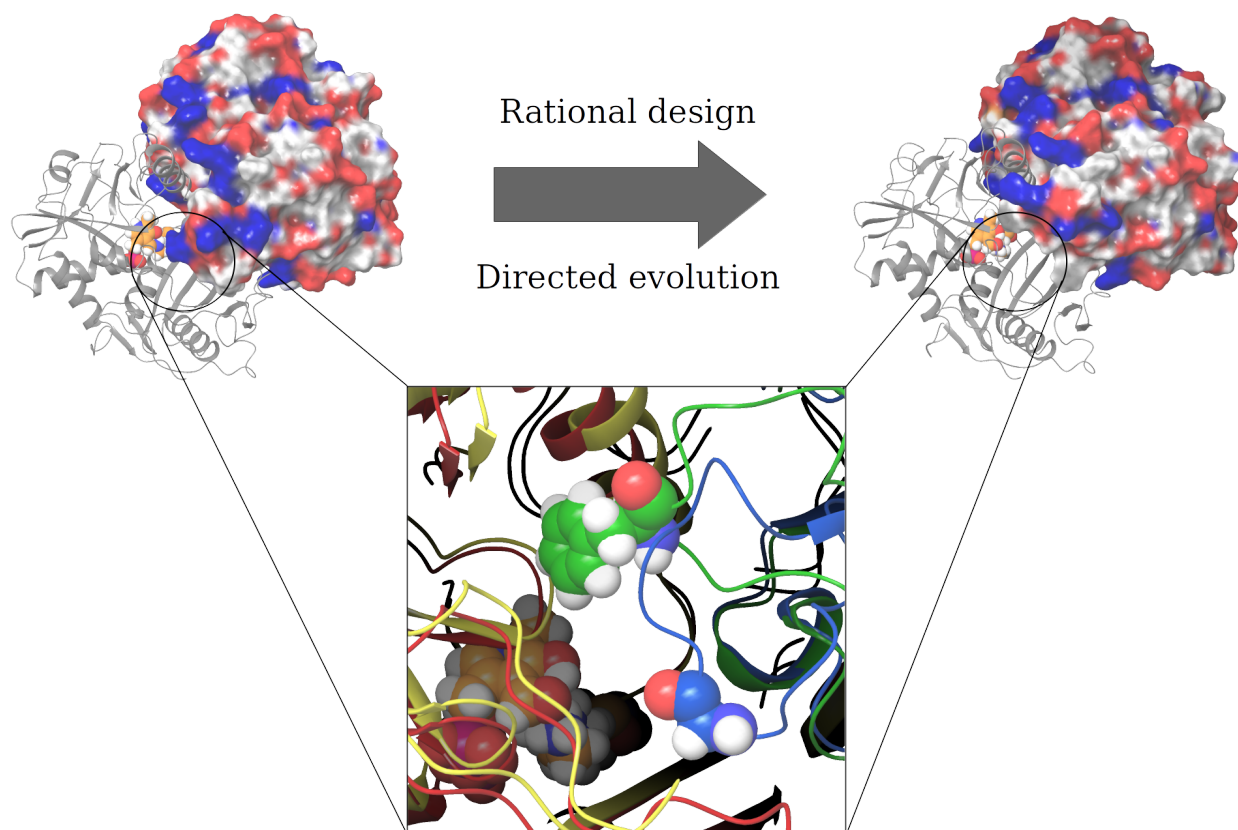
The *in silico* rational design of  $\omega$ TAs to enhance chiral amine synthesis has thrived with several examples. One of these successful cases refers to the engineering of a (R)-selective  $\omega$ TA binding pocket to accept pro-sitagliptin (the precursor of sitagliptin, an antidiabetic drug). The key mutation, G136F, changes the conformation of a loop near the active site, giving a broader binding site with bigger volume (it enlarges the L pocket) (Guan et al., 2015). Then, it was further evolved to gain activity against it. The result was a 10-13% increase in overall yield, 53%

increase in productivity, and 19% reduction in total waste, making the biocatalytic process replace the former rhodium-catalyzed asymmetric enamine hydrogenation, which needed toxic heavy metals and implied more expensive costs (Savile et al., 2010).

In the case of (S)-selective  $\omega$ TAs, a study has also accomplished their rational design to accept bulky ketones with mutations aiming at enlarging the L pocket without compromising it, other mutations decreasing the charges near the active site (due to the nonpolar nature of the used ketone in the study), and a final mutation to improve the overall thermostability (Dourado et al., 2016). The design of the variants was based on molecular docking, MD simulations, and further structural analysis, highlighting the importance of modeling the interactions between the enzyme and a specific substrate.

Other studies of rational engineering in  $\omega$ TAs exist (Midelfort et al., 2013; Nobili et al., 2015; Svedendahl et al., 2010) and all they share the fact that the computational analyses are based on modeling the enzyme-substrate interactions with MM methods. Thus, the information these methods provide have been key to develop these applications of  $\omega$ TAs in the pharmaceutical industry.





**Figure 5: Engineering of a  $\omega$ TA to enable the synthesis of sitagliptin with the ketone precursor.** The top panel shows the asymmetric synthesis of sitagliptin with the use of isopropylamine as amine donor. The bottom panel showcases both the WT and mutant dimer structures with one chain (chain A) being represented with the surface labeled as residues charge (blue being positively charged and red being negatively charged) and the other chain (chain B) with ribbon representation, and the modified lysine and PLP are displayed with the CPK model and the C atoms stained in orange. The zoomed region is one of the binding pockets in both the WT and mutant enzymes; the critical residue (G136) that leads to the major change is displayed with the CPK model and with C atoms stained in blue on the WT and in green on the mutant. The color of the ribbon in chain A is red and blue in chain B in the WT enzyme, while in the mutant the color of the ribbon in chain A is yellow and green in chain B. The PDB codes for the WT and mutant enzymes are 3WWH and 3WWJ, respectively.

Finally, computational studies can also be applied when there is no enzyme as a starting point, but still imposing an enzyme-substrate organization (Welborn & Head-Gordon, 2019); through a *de novo* design. The transition states are used to build the active site (or theozyme) to be completed with a protein scaffold. Next steps will be improving the binding mode of the active site and further characterization and optimization of the new enzyme (Malisi et al., 2009). One of the first attempts was a hydrolytic enzyme by Mayo in 2001 (Bolon & Mayo, 2001). More complex catalysis were engineered, consisting in, for example, a nucleophilic cysteine–histidine dyad with amides groups (backbones preferably) for oxyanion stabilization. The resulting



enzymes were found to cleave activated ester substrates by a two-step acylation/deacylation reaction (Richter et al., 2012). In spite of these success cases, the main computational *de novo* design limitation still is the development of the required protein folding from scratch, typically resulting in low activity enzymes.

Overall, we have attempted to give a quick overview of different *in silico* options providing biocatalytic understanding and facilitating enzyme engineering, where we focused on underlying the potential of methods mapping the enzyme-substrate interaction. Still these are mostly based on molecular modeling and bioinformatic techniques. In the (possibly very) near future, however, with the increase of experimental data and most likely with data augmentation from modeling, we expect to see the rise of machine learning techniques. In one form or the other, one thing appears clear to us, computational modeling will most likely guide most future engineering efforts.

## References

- Acebes, S., Fernandez-Fueyo, E., Monza, E., Fatima Lucas, M., Almendral, D., Ruiz-Dueñas, F. J., Lund, H., Martinez, A. T., & Guallar, V. (2016). Rational Enzyme Engineering Through Biophysical and Biochemical Modeling. In *ACS Catalysis* (Vol. 6, Issue 3, pp. 1624–1629). <https://doi.org/10.1021/acscatal.6b00028>
- Agarwal, P. K. (2006). Enzymes: An integrated view of structure, dynamics and function. *Microbial Cell Factories*, 5, 2.
- Alonso, S., Santiago, G., Cea-Rama, I., Fernandez-Lopez, L., Coscolín, C., Modregger, J., Ressmann, A. K., Martínez-Martínez, M., Marrero, H., Bargiela, R., Pita, M., Gonzalez-Alfonso, J. L., Briand, M. L., Rojo, D., Barbas, C., Plou, F. J., Golyshin, P. N., Shahgaldian, P., Sanz-Aparicio, J., ... Ferrer, M. (2020). Genetically engineered proteins with two active sites for enhanced biocatalysis and synergistic chemo- and biocatalysis. In *Nature Catalysis* (Vol. 3, Issue 3, pp. 319–328). <https://doi.org/10.1038/s41929-019-0394-4>
- AlQuraishi, M. (2019). End-to-End Differentiable Learning of Protein Structure. *Cell Systems*, 8(4), 292–301.e3.

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.
- Amidi, A., Amidi, S., Vlachakis, D., Megalooikonomou, V., Paragios, N., & Zacharaki, E. I. (2018). EnzyNet: enzyme classification using 3D convolutional neural networks on spatial representation. *PeerJ*, *6*, e4750.
- Aranda, C., Municoy, M., Guallar, V., Kiebist, J., Scheibner, K., Ullrich, R., del Río, J. C., Hofrichter, M., Martínez, A. T., & Gutiérrez, A. (2019). Selective synthesis of 4-hydroxyisophorone and 4-ketoisophorone by fungal peroxygenases. In *Catalysis Science & Technology* (Vol. 9, Issue 6, pp. 1398–1405). <https://doi.org/10.1039/c8cy02114g>
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, *44*(W1), W344–W350.
- Bajaj, P., Sreenilayam, G., Tyagi, V., & Fasan, R. (2016). Gram-Scale Synthesis of Chiral Cyclopropane-Containing Drugs and Drug Precursors with Engineered Myoglobin Catalysts Featuring Complementary Stereoselectivity. *Angewandte Chemie*, *55*(52), 16110–16114.
- Ba, L., Li, P., Zhang, H., Duan, Y., & Lin, Z. (2013). Semi-rational engineering of cytochrome P450sca-2 in a hybrid system for enhanced catalytic activity: Insights into the important role of electron transfer. In *Biotechnology and Bioengineering* (Vol. 110, Issue 11, pp. 2815–2825). <https://doi.org/10.1002/bit.24960>
- Belzen, J. U. zu, Bürgel, T., Holderbach, S., Bubeck, F., Adam, L., Gandor, C., Klein, M., Mathony, J., Pfuderer, P., Platz, L., Przybilla, M., Schwendemann, M., Heid, D., Hoffmann, M. D., Jendrusch, M., Schmelas, C., Waldhauer, M., Lehmann, I., Niopek, D., & Eils, R. (2019). Leveraging implicit knowledge in neural networks for functional dissection and engineering of proteins. In *Nature Machine Intelligence* (Vol. 1, Issue 5, pp. 225–235).

<https://doi.org/10.1038/s42256-019-0049-9>

- Binkley, J. S., Whiteside, R., Hariharan, P. C., Seeger, R., Hehre, W. J., Lathan, W. A., Newton, M. D., Ditchfield, R., & Pople, J. A. (1978). *GAUSSIAN 76: an ab initio molecular orbital program*. <https://doi.org/10.2172/6738000>
- Böck, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B., & Zinoni, F. (1991). Selenocysteine: the 21st amino acid. In *Molecular Microbiology* (Vol. 5, Issue 3, pp. 515–520). <https://doi.org/10.1111/j.1365-2958.1991.tb00722.x>
- Bolon, D. N., & Mayo, S. L. (2001). Enzyme-like proteins by computational design. *Proceedings of the National Academy of Sciences of the United States of America*, 98(25), 14274–14279.
- Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., Broglia, R. A., & Parrinello, M. (2009). PLUMED: A portable plugin for free-energy calculations with molecular dynamics. In *Computer Physics Communications* (Vol. 180, Issue 10, pp. 1961–1972). <https://doi.org/10.1016/j.cpc.2009.05.011>
- Bornscheuer, U. T., & Pohl, M. (2001). Improved biocatalysts by directed evolution and rational protein design. *Current Opinion in Chemical Biology*, 5(2), 137–143.
- Borrelli, K. W., Vitalis, A., Alcantara, R., & Guallar, V. (2005). PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Journal of Chemical Theory and Computation*, 1(6), 1304–1311.
- Bowers, K. J., Chow, D. E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., & Shaw, D. E. (2006). Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *ACM/IEEE SC 2006 Conference (SC'06)*. <https://doi.org/10.1109/sc.2006.54>

- Breaker, R. R. (1997). DNA enzymes. *Nature Biotechnology*, 15(5), 427–431.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. In *Journal of Computational Chemistry* (Vol. 4, Issue 2, pp. 187–217).  
<https://doi.org/10.1002/jcc.540040211>
- Car, R., & Parrinello, M. (1985). Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters*, 55(22), 2471–2474.
- Carvalho, A. T. P., Barrozo, A., Doron, D., Kilshtain, A. V., Major, D. T., & Kamerlin, S. C. L. (2014). Challenges in computational studies of enzyme structure, function and dynamics. In *Journal of Molecular Graphics and Modelling* (Vol. 54, pp. 62–79).  
<https://doi.org/10.1016/j.jmgm.2014.09.003>
- Chapman, J., Ismail, A., & Dinu, C. (2018). Industrial Applications of Enzymes: Recent Advances, Techniques, and Outlooks. In *Catalysts* (Vol. 8, Issue 6, p. 238).  
<https://doi.org/10.3390/catal8060238>
- Chen, F., Gaucher, E. A., Leal, N. A., Hutter, D., Havemann, S. A., Govindarajan, S., Ortlund, E. A., & Benner, S. A. (2010). Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proceedings of the National Academy of Sciences of the United States of America*, 107(5), 1948–1953.
- Chen, K., Huang, X., Kan, S. B. J., Zhang, R. K., & Arnold, F. H. (2018). Enzymatic construction of highly strained carbocycles. *Science*, 360(6384), 71–75.
- Chien, A., Edgar, D. B., & Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *Journal of Bacteriology*, 127(3), 1550–1557.
- Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. In *The EMBO Journal* (Vol. 5, Issue 4, pp. 823–826).

<https://doi.org/10.1002/j.1460-2075.1986.tb04288.x>

- Chowdhury, B., & Garai, G. (2017). A review on multiple sequence alignment from the perspective of genetic algorithm. *Genomics*, *109*(5-6), 419–431.
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. In *Journal of the American Chemical Society* (Vol. 117, Issue 19, pp. 5179–5197).
- <https://doi.org/10.1021/ja00124a002>
- Cossins, B. P., Hosseini, A., & Guallar, V. (2012). Exploration of Protein Conformational Change with PELE and Meta-Dynamics. *Journal of Chemical Theory and Computation*, *8*(3), 959–965.
- Daugelaite, J., O' Driscoll, A., & Sleator, R. D. (2013). An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics. *ISRN Biomathematics*, *2013*, 1–14.
- Debye, P. (1913). Interferenz von Röntgenstrahlen und Wärmebewegung. *Annalen Der Physik*, *348*(1), 49–92.
- Denisov, I. G., Makris, T. M., Sligar, S. G., & Schlichting, I. (2005). Structure and chemistry of cytochrome P450. *Chemical Reviews*, *105*(6), 2253–2277.
- Dhingra, A., Portis, A. R., Jr, & Daniell, H. (2004). Enhanced translation of a chloroplast-expressed RbcS gene restores small subunit levels and photosynthesis in nuclear RbcS antisense plants. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(16), 6315–6320.
- Dias, S. E. D., Martins, A. M., Nguyen, Q. T., & Gomes, A. J. P. (2017). GPU-based detection of protein cavities using Gaussian surfaces. *BMC Bioinformatics*, *18*(1), 493.
- Dodani, S. C., Kiss, G., Cahn, J. K. B., Su, Y., Pande, V. S., & Arnold, F. H. (2016). Discovery of

a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models. *Nature Chemistry*, 8(5), 419–425.

Dominguez, C., Boelens, R., & Alexandre M J. (2003). HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information. In *Journal of the American Chemical Society* (Vol. 125, Issue 7, pp. 1731–1737). <https://doi.org/10.1021/ja026939x>

Doudna, J. A., & Lorsch, J. R. (2005). Ribozyme catalysis: not different, just worse. In *Nature Structural & Molecular Biology* (Vol. 12, Issue 5, pp. 395–402).

<https://doi.org/10.1038/nsmb932>

Dourado, D. F. A. R., Pohle, S., Carvalho, A. T. P., Dheeman, D. S., Caswell, J. M., Skvortsov, T., Miskelly, I., Brown, R. T., Quinn, D. J., Allen, C. C. R., Kulakov, L., Huang, M., & Moody, T. S. (2016). *Rational Design of a (S)-Selective-Transaminase for Asymmetric Synthesis of (1S)-1-(1,1'-biphenyl-2-yl)ethanamine* [Data set]. <https://doi.org/10.1021/acscatal.6b02380>

Dunn, M. R., Otto, C., Fenton, K. E., & Chaput, J. C. (2016). Improving Polymerase Activity with Unnatural Substrates by Sampling Mutations in Homologous Protein Architectures. *ACS Chemical Biology*, 11(5), 1210–1219.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST.

*Bioinformatics* , 26(19), 2460–2461.

Eyring, H. (1935). The Activated Complex and the Absolute Rate of Chemical Reactions. In *Chemical Reviews* (Vol. 17, Issue 1, pp. 65–77). <https://doi.org/10.1021/cr60056a006>

Fan, H., Schneidman-Duhovny, D., Irwin, J. J., Dong, G., Shoichet, B. K., & Sali, A. (2011). Statistical Potential for Modeling and Ranking of Protein–Ligand Interactions. In *Journal of Chemical Information and Modeling* (Vol. 51, Issue 12, pp. 3078–3092).

<https://doi.org/10.1021/ci200377u>

Farwell, C. C., Zhang, R. K., McIntosh, J. A., Hyster, T. K., & Arnold, F. H. (2015).

Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Central Science*, 1(2), 89–93.

Feng, D.-F., & Doolittle, R. F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. In *Journal of Molecular Evolution* (Vol. 25, Issue 4, pp.

351–360). <https://doi.org/10.1007/bf02603120>

Fock, V. (1930). Näherungsmethode zur Lösung des quantenmechanischen

Mehrkörperproblems. In *Zeitschrift für Physik* (Vol. 61, Issues 1-2, pp. 126–148).

<https://doi.org/10.1007/bf01340294>

Freakley, S. J., Kochius, S., van Marwijk, J., Fenner, C., Lewis, R. J., Baldenius, K., Marais, S.

S., Opperman, D. J., Harrison, S. T. L., Alcalde, M., Smit, M. S., & Hutchings, G. J. (2019).

A chemo-enzymatic oxidation cascade to activate C–H bonds with in situ generated H<sub>2</sub>O<sub>2</sub>.

In *Nature Communications* (Vol. 10, Issue 1). <https://doi.org/10.1038/s41467-019-12120-w>

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. (2004).

Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47(7), 1739–1749.

Friesner, R. A., & Guallar, V. (2005). Ab initio quantum chemical and mixed quantum

mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis.

*Annual Review of Physical Chemistry*, 56, 389–427.

Galperin, M. Y., & Koonin, E. V. (1997). A diverse superfamily of enzymes with ATP-dependent

carboxylate-amine/thiol ligase activity. *Protein Science: A Publication of the Protein Society*, 6(12), 2639–2643.

- Galperin, M. Y., & Koonin, E. V. (2012). Divergence and convergence in enzyme evolution. *The Journal of Biological Chemistry*, 287(1), 21–28.
- Ghislieri, D., & Turner, N. J. (2014). Biocatalytic Approaches to the Synthesis of Enantiomerically Pure Chiral Amines. In *Topics in Catalysis* (Vol. 57, Issue 5, pp. 284–300). <https://doi.org/10.1007/s11244-013-0184-1>
- Gilabert, J. F., Lecina, D., Estrada, J., & Guallar, V. (2018). Monte Carlo Techniques for Drug Design: The Success Case of PELE. In *Biomolecular Simulations in Structure-Based Drug Discovery* (pp. 87–103). <https://doi.org/10.1002/9783527806836.ch5>
- Gillis, D., & Rومان, M. (2000). PoPMuSiC, an algorithm for predicting protein mutant stability changes: application to prion proteins. *Protein Engineering*, 13(12), 849–856.
- Godzik, A. (1996). The structural alignment between two proteins: Is there a unique answer? In *Protein Science* (Vol. 5, Issue 7, pp. 1325–1338). <https://doi.org/10.1002/pro.5560050711>
- Grosdidier, A., Zoete, V., & Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, 39(Web Server issue), W270–W277.
- Guan, L.-J., Ohtsuka, J., Okai, M., Miyakawa, T., Mase, T., Zhi, Y., Hou, F., Ito, N., Iwasaki, A., Yasohara, Y., & Tanokura, M. (2015). A new target region for changing the substrate specificity of amine transaminases. *Scientific Reports*, 5, 10753.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3 Pt 2), 849–857.
- Halgren, T. (2007). New method for fast and accurate binding-site identification and analysis. *Chemical Biology & Drug Design*, 69(2), 146–148.
- Hartree, D. R. (1928). The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. In *Mathematical Proceedings of the Cambridge Philosophical*



- Society* (Vol. 24, Issue 1, pp. 89–110). <https://doi.org/10.1017/s0305004100011919>
- Hayashi, K., Sugimoto, H., Shinkyo, R., Yamada, M., Ikeda, S., Ikushiro, S., Kamakura, M., Shiro, Y., & Sakaki, T. (2008). Structure-Based Design of a Highly Active Vitamin D Hydroxylase from *Streptomyces griseolus* CYP105A1†‡. In *Biochemistry* (Vol. 47, Issue 46, pp. 11964–11972). <https://doi.org/10.1021/bi801222d>
- Head-Gordon, M., Rico, R. J., Oumi, M., & Lee, T. J. (1994). A double correction to electronic excited states from configuration interaction in the space of single substitutions. In *Chemical Physics Letters* (Vol. 219, Issues 1-2, pp. 21–29). [https://doi.org/10.1016/0009-2614\(94\)00070-0](https://doi.org/10.1016/0009-2614(94)00070-0)
- Hernandez, K. E., Renata, H., Lewis, R. D., Jennifer Kan, S. B., Zhang, C., Forte, J., Rozzell, D., McIntosh, J. A., & Arnold, F. H. (2016). Highly Stereoselective Biocatalytic Synthesis of Key Cyclopropane Intermediate to Ticagrelor. *ACS Catalysis*, 6(11), 7810–7813.
- Higgins, D. G., & Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, 73(1), 237–244.
- Hofrichter, M., & Ullrich, R. (2014). Oxidations catalyzed by fungal peroxygenases. *Current Opinion in Chemical Biology*, 19, 116–125.
- Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, 99(6), 1129–1143.
- Holm, L., & Sander, C. (1996). Mapping the Protein Universe. In *Science* (Vol. 273, Issue 5275, pp. 595–602). <https://doi.org/10.1126/science.273.5275.595>
- Hou, Q., Kwasigroch, J. M., Rومان, M., & Pucci, F. (2020). SOLart: a structure-based method to predict protein solubility and aggregation. *Bioinformatics*, 36(5), 1445–1452.
- Howard, A. E., & Kollman, P. A. (1988). An analysis of current methodologies for conformational searching of complex molecules. *Journal of Medicinal Chemistry*, 31(9), 1669–1675.

- Hrycay, E. G., & Bandiera, S. M. (2012). The monooxygenase, peroxidase, and peroxygenase properties of cytochrome P450. *Archives of Biochemistry and Biophysics*, 522(2), 71–89.
- Hur, S., & Bruice, T. C. (2003). The near attack conformation approach to the study of the chorismate to prephenate reaction. In *Proceedings of the National Academy of Sciences* (Vol. 100, Issue 21, pp. 12015–12020). <https://doi.org/10.1073/pnas.1534873100>
- Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., & Friesner, R. A. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins*, 55(2), 351–367.
- Jemli, S., Ayadi-Zouari, D., Hlima, H. B., & Bejar, S. (2016). Biocatalysts: application and engineering for industrial purposes. *Critical Reviews in Biotechnology*, 36(2), 246–258.
- Jensen, F. (2007). *Introduction to Computational Chemistry*. John Wiley & Sons.
- Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, 267(3), 727–748.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. In *The Journal of Chemical Physics* (Vol. 79, Issue 2, pp. 926–935). <https://doi.org/10.1063/1.445869>
- Jung, E., Park, B. G., Yoo, H.-W., Kim, J., Choi, K.-Y., & Kim, B.-G. (2018). Semi-rational engineering of CYP153A35 to enhance  $\omega$ -hydroxylation activity toward palmitic acid. *Applied Microbiology and Biotechnology*, 102(1), 269–277.
- Kamerlin, S. C. L., & Warshel, A. (2010). At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins*, 78(6), 1339–1375.
- Kan, S. B. J., Lewis, R. D., Chen, K., & Arnold, F. H. (2016). Directed evolution of cytochrome c for carbon-silicon bond formation: Bringing silicon to life. *Science*, 354(6315), 1048–1051.
- Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules.

*Nature Structural Biology*, 9(9), 646–652.

Katoh, K., Misawa, K., Kuma, K.-I., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066.

Kawabata, T. (2010). Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins*, 78(5), 1195–1211.

Kelley, L. A., & Sternberg, M. J. E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. In *Nature Protocols* (Vol. 4, Issue 3, pp. 363–371).

<https://doi.org/10.1038/nprot.2009.2>

Kelly, S. A., Pohle, S., Wharry, S., Mix, S., Allen, C. C. R., Moody, T. S., & Gilmore, B. F. (2018). Application of  $\omega$ -Transaminases in the Pharmaceutical Industry. *Chemical Reviews*, 118(1), 349–367.

Khan, Z. U., Hayat, M., & Khan, M. A. (2015). Discrimination of acidic and alkaline enzyme using Chou's pseudo amino acid composition in conjunction with probabilistic neural

network model. In *Journal of Theoretical Biology* (Vol. 365, pp. 197–203).

<https://doi.org/10.1016/j.jtbi.2014.10.014>

Khatri, Y., Jóźwik, I. K., Ringle, M., Ionescu, I. A., Litzenburger, M., Hutter, M. C., Thunnissen, A.-M. W. H., & Bernhardt, R. (2018). Structure-Based Engineering of Steroidogenic CYP260A1 for Stereo- and Regioselective Hydroxylation of Progesterone. In *ACS Chemical Biology* (Vol. 13, Issue 4, pp. 1021–1028). <https://doi.org/10.1021/acscchembio.8b00026>

Kirk, O., Borchert, T. V., & Fuglsang, C. C. (2002). Industrial enzyme applications. *Current Opinion in Biotechnology*, 13(4), 345–351.

Knowles, J. R. (1991). Enzyme catalysis: not different, just better. In *Nature* (Vol. 350, Issue 6314, pp. 121–124). <https://doi.org/10.1038/350121a0>

- Kohen, A. (2015). Role of dynamics in enzyme catalysis: substantial versus semantic controversies. *Accounts of Chemical Research*, 48(2), 466–473.
- Kohn, W., & Sham, L. J. (1965). Self-Consistent Equations Including Exchange and Correlation Effects. In *Physical Review* (Vol. 140, Issue 4A, pp. A1133–A1138).  
<https://doi.org/10.1103/physrev.140.a1133>
- Koskinen, J. P., & Holm, L. (2012). SANS: high-throughput retrieval of protein sequences allowing 50% mismatches. *Bioinformatics*, 28(18), i438–i443.
- Kraut, J. (1988). How do enzymes work? In *Science* (Vol. 242, Issue 4878, pp. 533–540).  
<https://doi.org/10.1126/science.3051385>
- Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31(1), 147–157.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259), 680–685.
- Lafayette, L., & Wiebelt, B. (2017). Spartan and NEMO: Two HPC-Cloud Hybrid Implementations. In *2017 IEEE 13th International Conference on e-Science (e-Science)*.  
<https://doi.org/10.1109/escience.2017.70>
- Larrañaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., Lozano, J. A., Armañanzas, R., Santafé, G., Pérez, A., & Robles, V. (2006). Machine learning in bioinformatics. *Briefings in Bioinformatics*, 7(1), 86–112.
- Lassmann, T., & Sonnhammer, E. L. L. (2005). Kalign--an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics*, 6, 298.
- Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics*, 10, 168.

- Lei, H., & Duan, Y. (2007). Improved sampling methods for molecular simulation. In *Current Opinion in Structural Biology* (Vol. 17, Issue 2, pp. 187–191).  
<https://doi.org/10.1016/j.sbi.2007.03.003>
- Lonsdale, R., Harvey, J. N., & Mulholland, A. J. (2012). A practical guide to modelling enzyme-catalysed reactions. *Chemical Society Reviews*, 41(8), 3025–3038.
- Lucas, F., Babot, E. D., Cañellas, M., del Río, J. C., Kalum, L., Ullrich, R., Hofrichter, M., Guallar, V., Martínez, A. T., & Gutiérrez, A. (2016). Molecular determinants for selective C25-hydroxylation of vitamins D2 and D3 by fungal peroxygenases. In *Catalysis Science & Technology* (Vol. 6, Issue 1, pp. 288–295). <https://doi.org/10.1039/c5cy00427f>
- Lucas, M. F., Monza, E., Jørgensen, L. J., Ernst, H. A., Piontek, K., Bjerrum, M. J., Martínez, Á. T., Camarero, S., & Guallar, V. (2017). Simulating Substrate Recognition and Oxidation in Laccases: From Description to Design. *Journal of Chemical Theory and Computation*, 13(3), 1462–1467.
- Lutz, S. (2010). Beyond directed evolution—semi-rational protein engineering and design. *Current Opinion in Biotechnology*, 21(6), 734–743.
- Ma, J., & Wang, S. (2014). Algorithms, Applications, and Challenges of Protein Structure Alignment. In *Advances in Protein Chemistry and Structural Biology* (pp. 121–175).  
<https://doi.org/10.1016/b978-0-12-800168-4.00005-6>
- Malik, M. S., Park, E.-S., & Shin, J.-S. (2012). Features and technical applications of  $\omega$ -transaminases. *Applied Microbiology and Biotechnology*, 94(5), 1163–1171.
- Malisi, C., Kohlbacher, O., & Höcker, B. (2009). Automated scaffold selection for enzyme design. In *Proteins: Structure, Function, and Bioinformatics* (Vol. 77, Issue 1, pp. 74–83).  
<https://doi.org/10.1002/prot.22418>
- Maranas, C. D., & Floudas, C. A. (1994). A deterministic global optimization approach for

molecular structure determination. In *The Journal of Chemical Physics* (Vol. 100, Issue 2, pp. 1247–1261). <https://doi.org/10.1063/1.467236>

Marcus, R. A. (1956). On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. I. In *The Journal of Chemical Physics* (Vol. 24, Issue 5, pp. 966–978). <https://doi.org/10.1063/1.1742723>

Martínez-Martínez, M., Coscolín, C., Santiago, G., Chow, J., Stogios, P. J., Bargiela, R., Gertler, C., Navarro-Fernández, J., Bollinger, A., Thies, S., Méndez-García, C., Popovic, A., Brown, G., Chernikova, T. N., García-Moyano, A., Bjerga, G. E. K., Pérez-García, P., Hai, T., Del Pozo, M. V., ... The Inmare Consortium. (2018). Determinants and Prediction of Esterase Substrate Promiscuity Patterns. *ACS Chemical Biology*, *13*(1), 225–234.

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Šali, A. (2000). Comparative Protein Structure Modeling of Genes and Genomes. In *Annual Review of Biophysics and Biomolecular Structure* (Vol. 29, Issue 1, pp. 291–325). <https://doi.org/10.1146/annurev.biophys.29.1.291>

Maximova, T., Moffatt, R., Ma, B., Nussinov, R., & Shehu, A. (2016). Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics. In *PLOS Computational Biology* (Vol. 12, Issue 4, p. e1004619). <https://doi.org/10.1371/journal.pcbi.1004619>

McGuffin, L. J., Atkins, J. D., Salehe, B. R., Shuid, A. N., & Roche, D. B. (2015). IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences. *Nucleic Acids Research*, *43*(W1), W169–W173.

Midelfort, K. S., Kumar, R., Han, S., Karmilowicz, M. J., McConnell, K., Gehlhaar, D. K., Mistry, A., Chang, J. S., Anderson, M., Villalobos, A., Minshull, J., Govindarajan, S., & Wong, J. W. (2013). Redesigning and characterizing the substrate specificity and activity of *Vibrio*

- fluvialis aminotransferase for the synthesis of imagabalin. *Protein Engineering, Design & Selection: PEDS*, 26(1), 25–33.
- Molina-Espeja, P., Cañellas, M., Plou, F. J., Hofrichter, M., Lucas, F., Guallar, V., & Alcalde, M. (2016). Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution. *Chembiochem: A European Journal of Chemical Biology*, 17(4), 341–349.
- Møller, C., & Plesset, M. S. (1934). Note on an Approximation Treatment for Many-Electron Systems. In *Physical Review* (Vol. 46, Issue 7, pp. 618–622).  
<https://doi.org/10.1103/physrev.46.618>
- Monza, E., Fatima Lucas, M., Camarero, S., Alejaldre, L. C., Martínez, A. T., & Guallar, V. (2015). Insights into Laccase Engineering from Molecular Simulations: Toward a Binding-Focused Strategy. In *The Journal of Physical Chemistry Letters* (Vol. 6, Issue 8, pp. 1447–1453). <https://doi.org/10.1021/acs.jpcllett.5b00225>
- Mooij, W. T. M., & Verdonk, M. L. (2005). General and targeted statistical potentials for protein-ligand interactions. *Proteins*, 61(2), 272–287.
- Mullis, K. B. (1990). The Unusual Origin of the Polymerase Chain Reaction. In *Scientific American* (Vol. 262, Issue 4, pp. 56–65). <https://doi.org/10.1038/scientificamerican0490-56>  
*National Center for Biotechnology Information*. (n.d.). Retrieved February 24, 2020, from <https://www.ncbi.nlm.nih.gov/>
- Neese, F. (2012). The ORCA program system. *WIREs Computational Molecular Science*, 2(1), 73–78.
- Neet, K. E. (1998). Enzyme catalytic power minireview series. *The Journal of Biological Chemistry*, 273(40), 25527–25528.
- Nobili, A., Steffen-Munsberg, F., Kohls, H., Trentin, I., Schulzke, C., Höhne, M., & Bornscheuer, U. T. (2015). Engineering the Active Site of the Amine Transaminase from *Vibrio fluvialis* for

- the Asymmetric Synthesis of Aryl-Alkyl Amines and Amino Alcohols. In *ChemCatChem* (Vol. 7, Issue 5, pp. 757–760). <https://doi.org/10.1002/cctc.201403010>
- Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, *302*(1), 205–217.
- Oliveira, S. H. P., Ferraz, F. A. N., Honorato, R. V., Xavier-Neto, J., Sobreira, T. J. P., & de Oliveira, P. S. L. (2014). KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinformatics*, *15*, 197.
- Paladino, A., Marchetti, F., Rinaldi, S., & Colombo, G. (2017). Protein design: from computer models to artificial intelligence. In *Wiley Interdisciplinary Reviews: Computational Molecular Science* (Vol. 7, Issue 5, p. e1318). <https://doi.org/10.1002/wcms.1318>
- Pardo, I., Santiago, G., Gentili, P., Lucas, F., Monza, E., Medrano, F. J., Galli, C., Martínez, A. T., Guallar, V., & Camarero, S. (2016). Re-designing the substrate binding pocket of laccase for enhanced oxidation of sinapic acid. In *Catalysis Science & Technology* (Vol. 6, Issue 11, pp. 3900–3910). <https://doi.org/10.1039/c5cy01725d>
- Park, E.-S., & Shin, J.-S. (2011). Free energy analysis of  $\omega$ -transaminase reactions to dissect how the enzyme controls the substrate selectivity. *Enzyme and Microbial Technology*, *49*(4), 380–387.
- Pearson, W. R. (1991). Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics*, *11*(3), 635–650.
- Petrović, D., Risso, V. A., Kamerlin, S. C. L., & Sanchez-Ruiz, J. M. (2018). Conformational dynamics and enzyme evolution. In *Journal of The Royal Society Interface* (Vol. 15, Issue 144, p. 20180330). <https://doi.org/10.1098/rsif.2018.0330>
- Philipp, D. M., & Friesner, R. A. (1999). Mixedab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. In *Journal of Computational Chemistry*



(Vol. 20, Issue 14, pp. 1468–1494).

[https://doi.org/10.1002/\(sici\)1096-987x\(19991115\)20:14<1468::aid-jcc2>3.0.co;2-0](https://doi.org/10.1002/(sici)1096-987x(19991115)20:14<1468::aid-jcc2>3.0.co;2-0)

Prier, C. K., Zhang, R. K., Buller, A. R., Brinkmann-Chen, S., & Arnold, F. H. (2017).

Enantioselective, intermolecular benzylic C–H amination catalysed by an engineered iron-haem enzyme. In *Nature Chemistry* (Vol. 9, Issue 7, pp. 629–634).

<https://doi.org/10.1038/nchem.2783>

Purvis, G. D., & Bartlett, R. J. (1982). A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. In *The Journal of Chemical Physics* (Vol. 76, Issue 4, pp. 1910–1918). <https://doi.org/10.1063/1.443164>

Richter, F., Blomberg, R., Khare, S. D., Kiss, G., Kuzin, A. P., Smith, A. J. T., Gallaher, J., Pianowski, Z., Helgeson, R. C., Grjasnow, A., Xiao, R., Seetharaman, J., Su, M., Vorobiev, S., Lew, S., Forouhar, F., Kornhaber, G. J., Hunt, J. F., Montelione, G. T., ... Baker, D. (2012). Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *Journal of the American Chemical Society*, *134*(39), 16197–16206.

Robinson, P. K. (2015). Enzymes: principles and biotechnological applications. *Essays in Biochemistry*, *59*, 1–41.

Rohl, C. A., Strauss, C. E. M., Misura, K. M. S., & Baker, D. (2004). Protein Structure Prediction Using Rosetta. In *Numerical Computer Methods, Part D* (Vol. 383, pp. 66–93). Elsevier.

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725–738.

Rühlmann, A., Antovic, D., Müller, T. J. J., & Urlacher, V. B. (2017). Regioselective Hydroxylation of Stilbenes by Engineered Cytochrome P450 from *Thermobifida fusca* YX. In *Advanced Synthesis & Catalysis* (Vol. 359, Issue 6, pp. 984–994).

<https://doi.org/10.1002/adsc.201601168>

Santiago, G., de Salas, F., Fátima Lucas, M., Monza, E., Acebes, S., Martínez, Á. T., Camarero, S., & Guallar, V. (2016). Computer-Aided Laccase Engineering: Toward Biological Oxidation of Arylamines. In *ACS Catalysis* (Vol. 6, Issue 8, pp. 5415–5423).

<https://doi.org/10.1021/acscatal.6b01460>

Santiago, G., Martínez-Martínez, M., Alonso, S., Bargiela, R., Coscolín, C., Golyshin, P. N., Guallar, V., & Ferrer, M. (2018). Rational Engineering of Multiple Active Sites in an Ester Hydrolase. In *Biochemistry* (Vol. 57, Issue 15, pp. 2245–2255).

<https://doi.org/10.1021/acs.biochem.8b00274>

Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., & Hughes, G. J. (2010). Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*, 329(5989), 305–309.

Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., & Witholt, B. (2001). Industrial biocatalysis today and tomorrow. *Nature*, 409(6817), 258–268.

Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research*, 31(13), 3381–3385.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792), 706–710.

Senn, H. M., & Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angewandte Chemie*, 48(7), 1198–1229.

Serrano, A., Sancho, F., Viña-González, J., Carro, J., Alcalde, M., Guallar, V., & Martínez, A. T.

- (2019). Switching the substrate preference of fungal aryl-alcohol oxidase: towards stereoselective oxidation of secondary benzyl alcohols. In *Catalysis Science & Technology* (Vol. 9, Issue 3, pp. 833–841). <https://doi.org/10.1039/c8cy02447b>
- Shah, J. K., Marin-Rimoldi, E., Mullen, R. G., Keene, B. P., Khan, S., Paluch, A. S., Rai, N., Romanielo, L. L., Rosch, T. W., Yoo, B., & Maginn, E. J. (2017). Cassandra: An open source Monte Carlo package for molecular simulation. *Journal of Computational Chemistry*, 38(19), 1727–1739.
- Shaik, S., Kumar, D., de Visser, S. P., Altun, A., & Thiel, W. (2005). Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chemical Reviews*, 105(6), 2279–2328.
- Shindyalov, I. N., & Bourne, P. E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering*, 11(9), 739–747.
- Shin, J.-S., & Kim, B.-G. (2002). Exploring the active site of amine:pyruvate aminotransferase on the basis of the substrate structure-reactivity relationship: how the enzyme controls substrate specificity and stereoselectivity. *The Journal of Organic Chemistry*, 67(9), 2848–2853.
- Singh, R., Kolev, J. N., Sutera, P. A., & Fasan, R. (2015). Enzymatic C(sp)-H Amination: P450-Catalyzed Conversion of Carbonazidates into Oxazolidinones. *ACS Catalysis*, 5(3), 1685–1691.
- Singh, R., Kumar, M., Mittal, A., & Mehta, P. K. (2016). Microbial enzymes: industrial progress in 21st century. *3 Biotech*, 6(2), 174.
- Smith, E. A., & Oehme, F. W. (1991). Acrylamide and polyacrylamide: a review of production, use, environmental fate and neurotoxicity. *Reviews on Environmental Health*, 9(4), 215–228.

- Soares, I., Tavora, Z., Patera, R., & Baroni, S. (2012). Microorganism-Produced Enzymes in the Food Industry. In *Scientific, Health and Social Aspects of the Food Industry*.  
<https://doi.org/10.5772/31256>
- Steck, V., Kolev, J. N., Ren, X., & Fasan, R. (2020). Mechanism-Guided Design and Discovery of Efficient P450-Derived C—H Amination Biocatalysts. In *Journal of the American Chemical Society*. <https://doi.org/10.1021/jacs.9b12859>
- Sumbalova, L., Stourac, J., Martinek, T., Bednar, D., & Damborsky, J. (2018). HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Research*, *46*(W1), W356–W362.
- Suplatov, D., Voevodin, V., & Švedas, V. (2015). Robust enzyme design: bioinformatic tools for improved protein stability. *Biotechnology Journal*, *10*(3), 344–355.
- Svedendahl, M., Branneby, C., Lindberg, L., & Berglund, P. (2010). Reversed Enantioselectivity of an  $\omega$ -Transaminase by a Single-Point Mutation. In *ChemCatChem* (Vol. 2, Issue 8, pp. 976–980). <https://doi.org/10.1002/cctc.201000107>
- Syed, K., Porollo, A., Miller, D., & Yadav, J. S. (2013). Rational engineering of the fungal P450 monooxygenase CYP5136A3 to improve its oxidizing activity toward polycyclic aromatic hydrocarbons. *Protein Engineering, Design & Selection: PEDS*, *26*(9), 553–557.
- Taylor, W. R., & Orengo, C. A. (1989). Protein structure alignment. *Journal of Molecular Biology*, *208*(1), 1–22.
- Thoden, J. B., Kappock, T. J., Stubbe, J., & Holden, H. M. (1999). Three-dimensional structure of N5-carboxyaminoimidazole ribonucleotide synthetase: a member of the ATP grasp protein superfamily. *Biochemistry*, *38*(47), 15480–15492.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific

- gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673–4680.
- Toporkova, Y. Y., Ermilova, V. S., Gorina, S. S., Mukhtarova, L. S., Osipova, E. V., Gogolev, Y. V., & Grechkin, A. N. (2013). Structure-function relationship in the CYP74 family: conversion of divinyl ether synthases into allene oxide synthases by site-directed mutagenesis. *FEBS Letters*, 587(16), 2552–2558.
- Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455–461.
- Truhlar, D. G. (2015). Transition state theory for enzyme kinetics. In *Archives of Biochemistry and Biophysics* (Vol. 582, pp. 10–17). <https://doi.org/10.1016/j.abb.2015.05.004>
- UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1), D506–D515.
- van der Kamp, M. W., & Mulholland, A. J. (2013). Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry*, 52(16), 2708–2728.
- van Gunsteren, W. F., & Berendsen, H. J. (1987). Groningen molecular simulation (GROMOS) library manual. *Biomos, Groningen*, 24(682704), 13.
- Welborn, V. V., & Head-Gordon, T. (2019). Computational Design of Synthetic Enzymes. In *Chemical Reviews* (Vol. 119, Issue 11, pp. 6613–6630). <https://doi.org/10.1021/acs.chemrev.8b00399>
- Wolfenden, R. (2006). Degrees of difficulty of water-consuming reactions in the absence of enzymes. *Chemical Reviews*, 106(8), 3379–3396.
- Wolfenden, R., & Snider, M. J. (2001). The depth of chemical time and the power of enzymes as catalysts. *Accounts of Chemical Research*, 34(12), 938–945.

- Yamada, H., & Kobayashi, M. (1996). Nitrile hydratase and its application to industrial production of acrylamide. *Bioscience, Biotechnology, and Biochemistry*, 60(9), 1391–1400.
- Yang, Y., Pan, L., Lightstone, F. C., & Merz, K. M., Jr. (2016). The Role of Molecular Dynamics Potential of Mean Force Calculations in the Investigation of Enzyme Catalysis. *Methods in Enzymology*, 577, 1–29.
- Yoo, E.-H., & Lee, S.-Y. (2010). Glucose biosensors: an overview of use in clinical practice. *Sensors*, 10(5), 4558–4576.
- Zhang, R. K., Chen, K., Huang, X., Wohlschlager, L., Renata, H., & Arnold, F. H. (2019). Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp<sup>3</sup> C–H functionalization. In *Nature* (Vol. 565, Issue 7737, pp. 67–72). <https://doi.org/10.1038/s41586-018-0808-5>
- Zinovjev, K., & Tuñón, I. (2017). Quantifying the limits of transition state theory in enzymatic catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, 114(47), 12390–12395.

## Tables

**Table 1: Recompilation of computational tools used to study enzyme mechanisms and dynamics.**

Method	Function	System size	Computational time	Group	Softwares / Examples
Sequence alignment	To align protein/DNA/RNA sequences Phylogenetic analysis.	Any	Seconds	Sequence-based bioinformatics	CLUSTAL (Higgins & Sharp, 1988; Thompson et al., 1994), T-COFFEE (Notredame et al., 2000),

					Kalign (Lassmann & Sonnhammer, 2005), MAFFT (Kato et al., 2002) MUSCLE (Edgar, 2004).
Database searching of sequences	To find homologous sequences to the user's target one.	Any	Seconds to hours	Sequence-based bioinformatics	BLAST (Altschul et al., 1990), SSEARCH (Pearson, 1991), USEARCH (Edgar, 2010), SANS (Koskinen & Holm, 2012).
Structural alignment	To align 3D structures.	Any	Seconds	Structure-based bioinformatics	DALI (Holm & Sander, 1996), combinatorial extension (Shindyalov & Bourne, 1998), sequential structure alignment program (Taylor & Orengo, 1989).
Homology modeling	To build the 3D structure from the target sequence and a template	Any	Minutes	Structure-based bioinformatics	MODELLER (Martí-Renom et al., 2000), Prime (Jacobson et al., 2004), SWISS-MODEL (Schwede et

	as a model.				al., 2003).
Cavity & pocket detection	To find cavities and pockets in 3D structures.	Any	Minutes	Structure-based bioinformatics	Fpocket (Le Guilloux et al., 2009), GHECOM (Kawabata, 2010), KVFinder (Oliveira et al., 2014), GaussianFinder (Dias et al., 2017), SiteMap (Halgren, 2007).
MM minimization	To obtain the minimized/relaxed 3D structure of the system based on classical physics.	$10^6 - 10^3$ atoms	Minutes	Molecular modeling (MM)	AMBER (Cornell et al., 1995), CHARMM (Brooks et al., 1983), OPLS (Jorgensen et al., 1983), GROMOS (van Gunsteren & Berendsen, 1987).
QM minimization	To obtain the minimized/relaxed 3D structure of the system based on quantum physics.	$10^3 - 10^2$ atoms (semiempirical). $10^2 - 10$ atoms (DFT, HF). $\leq 10$ atoms (Post-HF)	Minutes to hours (semiempirical). Hours to days (DFT, HF, Post-HF)	Molecular modeling (QM)	Gaussian (Binkley et al., 1978), ORCA (Neese, 2012), Qsite (Philipp & Friesner, 1999).
QM/MM	To obtain	$10^6 - 10^3$	Hours to days	Molecular	MM: AMBER



minimization	the minimized/relaxed 3D structure of the system based on classical physics for the vast majority of atoms and on quantum physics for the active site.	atoms (MM region). $10^3 - 1$ atoms (QM region)		modeling	(Cornell et al., 1995), CHARMM (Brooks et al., 1983), OPLS (Jorgensen et al., 1983), GROMOS (van Gunsteren & Berendsen, 1987). QM: Gaussian (Binkley et al., 1978), ORCA (Neese, 2012), Qsite (Philipp & Friesner, 1999).
MD	To simulate the dynamical behaviour of the system in a particular environment over time.	MM or QM/MM size limits	Days to weeks	Molecular modeling	AMBER (Cornell et al., 1995), CHARMM (Brooks et al., 1983), GROMOS (van Gunsteren & Berendsen, 1987), Desmond (Bowers et al., 2006), PLUMED (Bonomi et al., 2009), Car-Parrinello MD (Car & Parrinello, 1985).
MC	To explore the conformational space	MM, QM, or QM/MM size limits	Hours to days	Molecular modeling	PELE (Borrelli et al., 2005; Car & Parrinello,

	of the system independently of time.				1985), Spartan (Lafayette & Wiebelt, 2017), Cassandra (Shah et al., 2017)
Molecular docking	To see the interaction mode of a molecule bound to another molecule (the system).	MM size limit	Seconds to minutes	Molecular modeling	AutoDock Vina (Trott & Olson, 2010), GOLD (Jones et al., 1997), Glide (Friesner et al., 2004), Haddock (Dominguez et al., 2003), SwissDock (Grosdidier et al., 2011).
Protein folding prediction from sequence	To infer the 3D structure of a protein from its sequence of residues.	50 - 300 residues	From minutes to days (Training). From seconds to days (Prediction)	AI-based approaches	AlphaFold (Senior et al., 2020)