

# X

## Monte Carlo techniques for Drug Design. The success case of PELE

Joan F. Gilibert<sup>1</sup>, Daniel Lecina<sup>1</sup>, Jorge Estrada<sup>1</sup> and Victor Guallar<sup>1,2</sup>

<sup>1</sup>Barcelona Supercomputing Center (BSC), Jordi Girona 29, E-08034 Barcelona, Spain

<sup>2</sup>ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

### X.1

#### Introduction

The Monte Carlo (MC) method dates back to the 1940s. While recovering from a surgery, Stanislaw Ulam tried to compute the probability of successfully finishing the Canfield solitaire game. After initial attempts looking for the combinatorial solution, he conceived an easier method that consisted on repeatedly simulating different games and counting the number of successful draws [1]. In other words, the idea was using statistical sampling to obtain a numerical approximation for a problem with a difficult analytical solution. Soon after, in the context of the Manhattan project and the first electronic computers, MC techniques were quickly applied to different problems in physics and afterwards became increasingly popular in a multitude of scientific fields, from biology to engineering [2-4].

Nowadays, MC techniques cover a broad spectrum of computational methodologies that make use of random numbers to perform numerical calculations. In drug design, these techniques are typically used to sample the Boltzmann distribution, building a Markov Chain Monte Carlo (MCMC) [5] (mainly by using the Metropolis algorithm [6]), or to find bound structure conformations. Protein-ligand complexes are modeled using an all-atom force field paradigm [7], which yields a highly dimensional phase space that does not allow direct estimates. Luckily, the relative probabilities between any pair of states are known, the ratio of Boltzmann weights, which permits the construction of an MCMC [8]. From this point of view, MC represents an alternative to molecular dynamics (MD)[9]. In the latter, Newton's equations of motion are integrated in small time increments, in the order of the femtosecond, to produce trajectories. One of the main drawbacks of this kind of simulation is the heavy computational cost associated to model the long timescales that correspond to protein-ligand binding and unbinding. On the contrary, MCMC does not produce time trajectories (*i.e.* dynamics); the algorithm draws samples in the energy landscape according to the corresponding statistical ensemble. This opens the door to design strategies that may yield, at least in theory, a more efficient sampling, since subsequent snapshots can be more uncorrelated compared to the MD counterpart, *e.g.* with side-chain rotations, backbone movements or ligand displacements. The main

obstacle of MC simulations, however, is the difficulty of generating uncorrelated poses with a significant statistical weight, given the ruggedness of the energy landscape that is defined by the force field. Protein and ligand flexibility is one of the causes behind. Considering coordinated atom displacements help mitigate the problem, for example with the use of internal coordinates, which take into account coordinated displacements in torsions [10-12]. However, random proposals typically lead to clashes and are usually, but not exclusively, reduced to small variations of the starting structure.

*First applications.* The precursors of MC sampling are condensed matter simulations, such as the study of water carried out by Barker and Watts in 1969 [13]. Several instances of studies involving statistical and thermodynamic properties of water solutions of methane can be found, some examples are the works of Dashevsky and Sarkisov in 1974[14], Owicki and Scheraga in 1977 [15] and by Swaminathan et. al. in 1978 [16]. In a similar fashion, in the late 1960s the first MC simulations of chain molecules were reported [17], including the work of Hagler *et. al.* studying the conformation of oligopeptides in 1978 [18] or the study of bovine pancreatic trypsin by Wako and Scheraga in 1981 [19].

*Free energy calculations.* As computers started growing in power, the possibilities of molecular simulations expanded. Larger and more complex systems could be tackled, and new methods attempted to increase the efficiency of simulations. This permitted obtaining free energy estimations [20, 21], mainly using alchemical methods such as free energy perturbation. An example is the pioneering MCLiquids of Jorgensen's lab, used to study liquid hydrogen fluoride, ammonia or ethanol, which later on evolved into BOSS (Biomolecular and Organic Simulation System) and MCPRO (MC for Proteins) [12]; both programs are still in active development, being used to perform molecular mechanics simulations and in the development of the OPLS force field[22]. Other free energy calculation methods, such as thermodynamic integration, have been used with MC. It can be seen in different works of Essex' group [23], and is included in their software package, ProtoMS [24], which specializes in explicit solvent calculations, for example making use of Grand Canonical MC [25] or JAWS [26]. The Sire software (<http://siremol.org>), a collection of C++/Python building blocks allowing to build custom molecular simulation applications, also emerged from ProtoMS.

*Optimization.* MC methods have been largely used in protein optimization problems, often in combination with enhanced sampling methods such as simulated annealing [27] or replica exchange ideas, where several replicas are kept at slightly different conditions and exchanges between such

copies are attempted periodically. Two examples of the latter are replica exchange with solute tempering (REST) [28] and replica Exchange thermodynamic integration (RETI) [29]. Optimization problems have also been tackled using more sophisticated random schemes, such as the biased probability MC of Abagyan [30], based on generating the random step with an optimal distribution of proposals; the force-biased MC [31], where the acceptance probability becomes one (or nearly one) by using first energy derivatives; or with collective variables motions [32].

A different family of methodologies originated from the MC-minimization technique by Li and Scheraga [33]. This method combined a typical MC random displacement, in this case randomly moving one dihedral angle, followed by an energy minimization. The method identifies at each step a local minimum, and thus the exploration is constrained to different local minima. An important contribution was then developed in 1989 by Still and coworkers: the MacroModel software [34], a simple MC method combining random moves, minimizations and an implicit solvation GB/SA model. This technique is currently included in the Schrödinger suite of programs, being one of the most used molecular MC programs of all times. More recently, these techniques have led to the development of several MC with minimization schemes to perform conformational sampling of proteins, for example the basin-hopping method of Wales [35], which can be used in combination with Wang-Landau sampling [36] to obtain the complete energy landscape, the activation-relaxation technique of Mousseau and coworkers [37] or our Protein Energy Landscape Exploration (PELE), which combines random movements with protein structure prediction algorithms to enhance the exploration [38].

*MC and MD combinations.* Efforts have also been made to combine MC with MD. In Hamiltonian MC (also known as hybrid MC) [39], velocities are added in order to guide proposals and improve the acceptance. Recently, Chodera and coworkers presented Nonequilibrium candidate Monte Carlo (NCCMC)[40], where proposal moves are constructed via nonequilibrium dynamics to generate new configurations. These are built in a series of perturbation mixed with propagation moves, enhancing two orders of magnitude the performance in finding binding modes of toluene in a mutated T4 lysozyme [41].

In table 1 we summarize the most representative software packages that readily allow running MC simulations in relevant scenarios for drug design. Some application examples include several studies using the MCPRO or BOSS tools from the Jorgensen lab. Besides their own applications studies, for example the computationally aided discovery of anti-HIV agents [42], we find additional

studies such as the inhibitor design on tyrosine kinase [43]. In addition, MCPRO+, a version under Schrödinger commercial use, has been also used in the drug design studies for the acetylcholine receptor [44] and the Bcr-Abl kinase [45]. We also find several studies using ProtoMS, more focused in obtaining the calculation of free energies [23] and the role of waters in drug recognition and binding [25, 26]. Commercial software, such as ICM and Macromodel, include several reference and benchmark studies [46-49].

**Table X.1.** Main MC software used in drug development modeling.

Software	Developers	Short description
<b>ProtoMS</b> <a href="http://www.essexgroup.soton.ac.uk/ProtoMS/">http://www.essexgroup.soton.ac.uk/ProtoMS/</a>	Essex lab.	Relative and absolute free energy calculations and water placement with GCMC and JAWS methodology. GNU General Public License.
<b>BOSS</b> <a href="http://zarbi.chem.yale.edu/software.html">http://zarbi.chem.yale.edu/software.html</a>	Jorgensen lab.	General purpose molecular modeling system that performs molecular mechanics calculations, Metropolis MC statistical simulations, and semiempirical quantum mechanics calculations. Free for non-profit/commercial for profit.
<b>MCPRO</b> <a href="http://zarbi.chem.yale.edu/software.html">http://zarbi.chem.yale.edu/software.html</a>	Jorgensen lab.	MC statistical mechanics simulations of peptides, proteins, and nucleic acids; it was derived from BOSS. Free for non-profit/commercial for profit.
<b>PELE</b> <a href="https://pele.bsc.es">https://pele.bsc.es</a> , <a href="http://www.nostrumbodiscovery.com">www.nostrumbodiscovery.com</a>	Guallar lab.	MC specialized in mapping protein-ligand interactions. Free for non profit/ commercial for profit.
<b>Sire</b> <a href="http://siremol.org">http://siremol.org</a>	Christopher Woods, et al.	C++/Python rewrite of ProtoMS. It now includes many more modeling modules. GNU General Public License.
<b>ICM</b> <a href="http://www.molsoft.com/icm_pro.html">http://www.molsoft.com/icm_pro.html</a>	Molsoft, LLC	Biased probability MC. Prediction of the three-dimensional structure of peptides and proteins. Commercial.
<b>Tinker</b> <a href="https://dasher.wustl.edu/tinker">https://dasher.wustl.edu/tinker</a>	Ponder lab.	General molecular modeling program including a MC minimization technique. Free of charge, registered under U.S. Copyright law.
<b>Macromodel</b> <a href="https://www.schrodinger.com/macromodel">https://www.schrodinger.com/macromodel</a>	Schrödinger, LLC	Mixed MC algorithms and other molecular modeling methods. Commercial. Schrodinger also distributes a (commercial) MCPRO+ version.

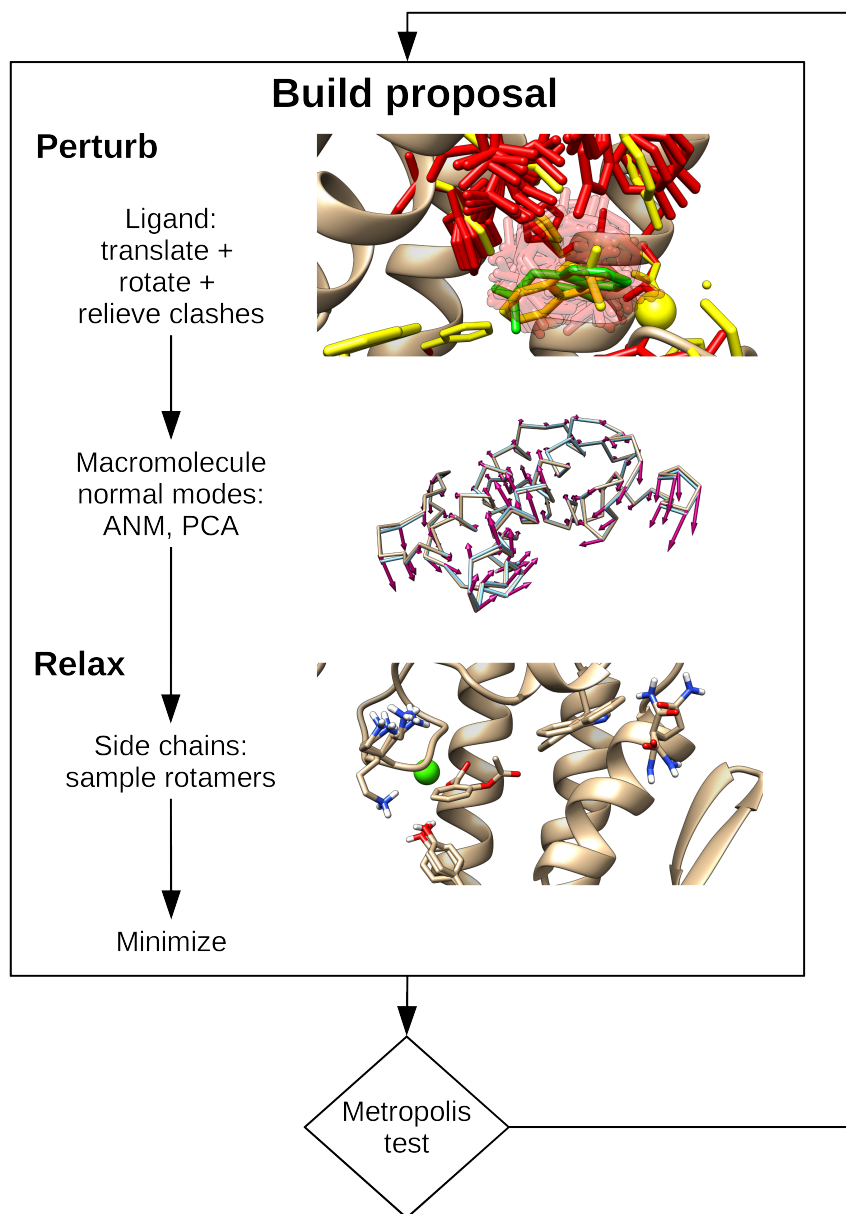
The remaining sections of this chapter explore in detail the PELE program, providing first the main characteristics of the technique, followed by an analysis of the different application studies in mapping protein-ligand interactions.

## **X.2**

### **The PELE Method**

PELE, originally described by Borrelli et al. in 2005 [38] and whose latest version is implemented in C++, follows a heuristic MC approach, generating new conformational proposals using vibration modes of the macromolecule (protein or nucleic acid) with translations and rotations of the ligand, and relaxing the system with structure prediction methods, so that the probability of acceptance in the Metropolis test remains high.

*MC sampling procedure.* The overall MC procedure, shown in Figure 1, involves a ligand plus a receptor perturbation, followed by a relaxation step comprising a side chain sampling and a minimization step. Typical simulations involve from tens to hundreds of processors for thousand MC steps. We describe now briefly each of these steps.



**Figure X.1** Process flow in a PELE simulation: Each simulation step builds a new conformation proposal by perturbing the ligand and the macromolecule, then relaxing it to an energetically favorable conformation by sampling side chain rotamers and by minimizing the whole system; finally, the proposal is either accepted or rejected following the Metropolis test. The top image shows the initial conformation (yellow), all the candidate ligand perturbations obtained in a given step by translating and rotating the ligand (transparent red) and relieving existing clashes by moving side chains (solid red), and the selected perturbation (green). The middle image shows the first ANM mode as magenta vectors placed at the atoms of the  $\alpha$ -carbons trace for the starting protein conformation (brown), and the trace obtained after applying the ANM mode (cyan). The bottom image shows the representative rotamers of the generated clusters for several side chains.

*Ligand Perturbation.* The ligand, formed by a rigid core (at least one atom) and a set of rotatable side chains, is perturbed by translating and rotating it. PELE creates a list of perturbation poses (typically between 1 and 20), and then chooses the one with the lowest system energy. Each pose is generated through a random procedure, where a number of translations and rotations are tried, until a combination is found that doesn't show steric clashes after a clash relieving process; clashes of up to 30% of the van der Waals interatomic distance between atoms are allowed, since the relaxation phase will probably relieve them. When searching for the binding site, large translations are usually selected (around 6 Å), while once in the binding site, small adjustments of  $\sim 1\text{Å}$  (or less) are used; different translation ranges are often applied depending on the ligand solvent accessible surface area (SASA) at each iteration. In addition, to enhance the sampling of rare events, the translational vector may be kept constant in a random direction for a given number of MC simulation steps. Also, for those simulations that require exploring a restricted volume of the search space (as it is done in binding pose refinement), perturbations that place the ligand center of mass outside a predefined box are rejected.

*Receptor Perturbation.* After perturbing the ligand, the macromolecule backbone is perturbed following normal modes calculated using the Anisotropic Network Model (ANM, see [50]) or a Principal Component Analysis (PCA) from different structures (derived from X-ray structures, molecular dynamics, etc.). This phase is typically executed every 4 MC steps [51].

In ANM, if a protein, the macromolecule is modeled as an elastic network of  $\alpha$ -carbons (CAs) nodes using the power contact model of [52]. For nucleic acids, atoms P, C2 and C4' are used as nodes of the network, using the exponential contact model of Setny et al. [53]. A combination of modes is selected to represent the given vibration from which to select the perturbation to try. By default the 6 lowest frequency modes are used, since they are the ones more closely related to conformational movements [50]. It is possible to use a single mode, or to mix several ones randomly, with a preferred mode being given more weight, or weighting each mode according to its frequency. Besides, a given combination of modes can be kept during several simulation steps, changing sense periodically, to allow a more thorough sampling of a given vibrational movement. To apply the movement for all atoms, and not only CAs, harmonic constraints are placed on the latter to their target coordinates, and an all-atom energy minimization is performed, using the multi-scale Truncated Newton algorithm. To avoid reverting the ligand perturbation, a constraint on the center of mass of the ligand's core atoms is typically applied; in this way the ligand perturbation can be coupled to backbone reorganization.

*Side Chain Adjustment.* The side chain readjustment is done for all the side chains in the macromolecule that surround the ligand, and for the ligand itself. In addition one might include those side chains that increased their energy the most during the perturbation phase (regardless of their location in the receptor). The sampling follows the algorithm of Jacobson et al. [54], where the problem becomes manageable by considering only rotamers as possible conformations for the residue side chains and, above all, by placing the best rotamer for each side chain (after clustering) in a residue by residue way, keeping all the other residues fixed at their current rotamers. This is iterated until a new round of placement for all residues does not significantly change the conformation (or a maximum number of iterations is exceeded). If (even with the allowed overlap) there is no rotamer without clashes for some side chain, then this phase fails and the MC step is rejected. Once the side chains have been readjusted, a partial minimization affecting only the side chain atoms is performed.

*Minimization.* The last relaxation step continues through a multi-scale Truncated Newton minimization [55], the same algorithm as used in the minimizations done in previous phases. Along the minimization, the atoms representing the nodes in the ANM phase are (typically) weakly constrained, so that this minimization does not undo the perturbation phase. Convergence criteria is significantly relaxed,  $\sim 0.1 \text{ kcal}/(\text{mol}\cdot\text{\AA})^2$ , allowing to approximate recover the detailed balance condition (lost when introducing the relaxation step) [56].

*Coordinate Exploration.* PELE introduces two further techniques to enhance conformational sampling. In the original one, *the spawning approach* introduced in Borrelli et al. [57], a reaction coordinate (RMSD, distances, etc.) is defined along with a range criteria and a sense (increase or decrease). For a given processor and MC step, if the current reaction coordinate value falls outside of the range from the best achieved value (along all processors and steps), then this processor will abandon its current conformation and receive the one corresponding to the best value, which is constantly updated. This kind of strategy is particularly useful when studying binding or escape trajectories.

More recently, an enhanced strategy has been introduced, *adaptive-PELE* [58], aiming at eliminating the bias introduced in the spawning approach. In this approach, standard PELE is run for a small number of MC steps ( $\sim 4-20$ ), in what constitutes an epoch. All generated conformations in the epoch (from all processors) are then clustered by using the ligand root mean square distance, where we also consider clusters from previous epochs. Each cluster is assigned a reward value that favors (by default) clusters less explored, that is inversely proportional to the cluster population. Additional



rewards might be defined from physical/structural properties such as reaction coordinates, ligand exposure to solvent, etc. Then, a new epoch is started using a set of selected conformations, based on the reward function, as new seeds. Thus, the overall procedure is designed to enhance the sampling of poorly sampled regions. In addition, for an even faster biased exploration, the spawning and the adaptive techniques can be combined.

*Energy Function.* By default, PELE uses the OPLS 2005 energy function and parameters[59, 60], but also implements the AMBER99sbBSC0 force field[61], more suitable for nucleic acids. For the implicit solvent model, PELE has three options: the OBC model [62] with a non-polar term following the ACE model [63]; the Surface Generalized Born (SGB) model [64], modeling the non-polar term as in [65]; and the Variable Dielectric Generalized Born model [66], using the same non-polar model as SGB. PELE allows placing specific discrete water molecules that are deemed important, which will be moved during the different energy minimizations in a simulation step (the next release will include a water perturbation/relaxation step). As in most molecular mechanics codes, to gain speed negligible contributions to the energy, such as those caused by non-bonded interactions between far apart atoms, are discarded by using a set of distance cutoffs.

## **X.3**

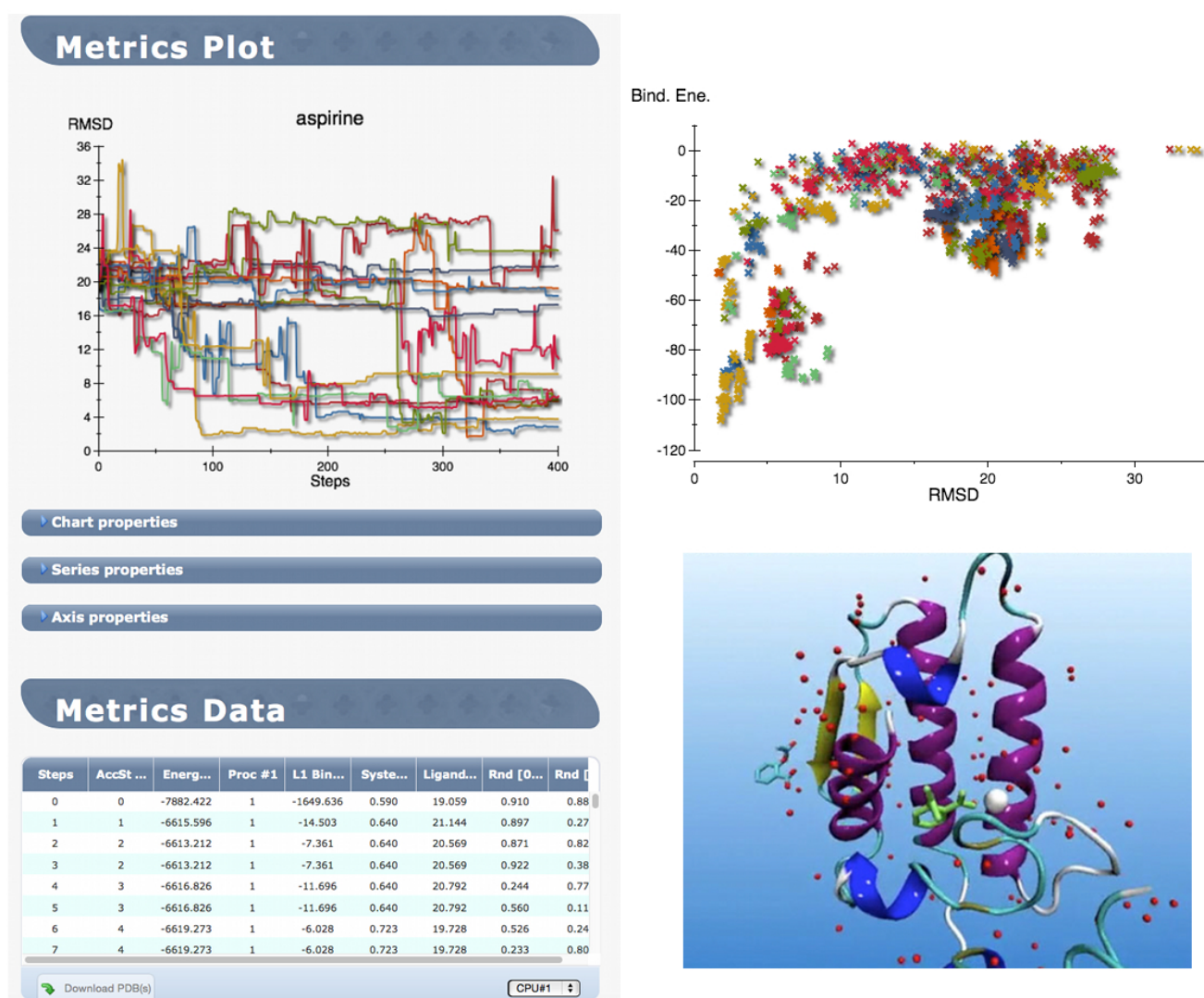
### **Examples of PELE's applications**

We present here examples of application studies using PELE.

*Mapping protein ligand and biomedical studies.* PELE was originally designed to map ligand migration pathways: its first applications aimed at finding exit pathways starting from ligand-bound crystallographic structures. In few hours of a small computational cluster (4 cores), for example, it could map how a palmitate fatty acid molecule escaped the fatty acid binding protein active site, without any additional knowledge nor bias [57]. Many additional initial studies centered on “simple” ligand migration studies involving globin enzymes, following a trend (at that time) using different biased MD studies, such as the random expulsion molecular dynamics or steered MD. Studied systems included truncated hemoglobins TrHbN [67] and TrHBO [68], where the involvement of multiple internal cavities was addressed. In addition, the fast performance of PELE allowed the first migration study on a (tetramer) human hemoglobin, identifying differences in ligand migration between the  $\alpha$  and

$\beta$  subunits and between their tense and relaxed states [69].

By expanding the exploration to few more cores (tens to hundreds) we realized that we could perform a complete non-biased exploration, mapping the entire surface active site search (from the solvent) and binding; similar studies, at that time, could only be performed by the special purpose (and 10M \$ cost) Anton computer (see, for example, the pioneer study by the Shaw lab. [70]). Figure 2 shows an example of such an exhaustive search using PELE's web server on the aspirin-phospholipase A2 complex. Additional studies involved quite complex systems such as prolyl oligopeptidase [71] and the binding of porphyrin into Gun4 [72]. Analogous simulations have been also produced for a DNA receptor, where the non-covalent binding of cisplatin was modeled with PELE's MC techniques [73].



**Figure X.2.** Example of active site search and binding on the aspirin-phospholipase A2 complex performed by the PELE web server (using 16 processors and 4 hours). The left panel shows a screen

capture of the server with the evolution of the ligand RMSD (all heavy atoms) to the crystal bound-complex (PDB ID: [1OXR](#)) along the MC simulation. Crystal-like models, with RMSD  $<2 \text{ \AA}$ , are found in  $\sim 90$  MC steps, requiring approximately 1 hour of wall clock simulation. The top right panel, also extracted from the server, shows the molecular mechanics protein-ligand interaction energy with respect to the same RMSD, indicating that the modeled bound complexes ( $<2.0 \text{ \AA}$ ) can clearly be identified as the best affinity models; we should emphasize that no crystallographic information is used along the simulation but only to analyze the results. The bottom right panel shows the initial molecular model, with the ligand on the bulk solvent ( $\sim 21 \text{ \AA}$  RMSD); notice that we have also superposed in green the ligand active site bound structure (underlying the presence of a calcium ion). For more information, please visit <https://pele.bsc.es/pele.wt/examples>.

Such modeling capabilities shifted our attention to biomedical/pharmaceutical studies, attracting, in addition, the interest from some pharmaceutical companies. Studies included research on cancer targets such as mTOR [74] and BCL-2 [75], on the glycosylation disorder through the study of the human phosphomannomutase2 receptor [76], and on diverse nuclear hormone receptors (NHRs), in collaboration with AstraZeneca [77, 78]. To enhance ligand screening, we have recently coupled PELE with advanced analysis techniques such as Markov State Models (MSM), approximating absolute binding free energies in a fast manner [78, 79].

In most cases, however, modeling the receptor-drug interactions does not require such extensive study; rearrangements of the ligand at the already known binding site suffice. Thus, we placed particular emphasis on protocols mapping the active site induced fit. Our initial benchmark study showed significant improvement from that of state of the art commercial software, providing accurate cross docking and apo docking [80]. Applied biomedical studies in this line, for example, included the prediction of drug resistance in the HIV-1 protease receptor, where we were able to blindly identify high resistance patients using only their viruses' sequence data [81]. We should underline that this effort involved modeling 15-30 mutations in each receptor of 42 different patients.

PELE's position at the forefront of the induced fit modeling techniques is clear if we look at the results of the latest Community Structure-Activity Resource (CSAR) competition (<http://www.csardock.org/>). The CSAR competition is a community-wide blind evaluation of docking and scoring methods performed by the University of Michigan (and participated by GlaxoSmithKline). In the summary article from the organizers [82], we find that PELE is the software most highlighted: "...In particular, submission J (Glide-XP2/PELE7) deserves special recognition for docking well to all targets. This group cross-docked all the ligands against all the given crystal structures. To see robust

performance across an ensemble of protein structures for multiple targets is an impressive accomplishment...”

As stated, a recent enhanced conformational sampling has been introduced by means of an adaptive sampling [58]. This procedure improves about one order of magnitude the exploration performed by PELE. Such enhancement, for example, allowed to map complex ligand binding mechanisms, such as in G protein–coupled receptors (GPCRs) or in NHRs in only three hours using a commodity computing cluster (~32 computing cores), achieving interactive rates (~5 minutes) in difficult induced fit cases.

These advances motivated PELE’s transfer of the technology to Nostrum Biodiscovery ([www.nostrumbiodiscovery.com](http://www.nostrumbiodiscovery.com)) a recent SME dedicated to drug development by means of supercomputation, which currently performs the commercial applications of the technique; academic (non-profit) usage, however, is still possible through the public web site <https://pele.bsc.es>, where hardware and software resources are available [83].

*Enzyme Characterization.* Additional applied studies have centered on modeling enzymatic mechanisms and engineering; the same techniques applied in mapping protein-drug interactions can be used in the study of substrate recognition by enzymes. While these studies often require combining PELE with mixed quantum mechanics and molecular mechanics (QM/MM) techniques, biochemical activity can be (largely) addressed by predicting the enzyme-substrate complex with simpler molecular mechanics techniques. A substrate-bound model for the aryl alcohol oxidase with *p*-anisyl alcohol, for example, was developed in 2011 by mapping the substrate entrance in the apo crystallographic structure, which, due to the small size cavity, did not allow direct docking studies [84]. This model, which has recently been confirmed by crystallographic studies [85], allowed to select mutations for mapping the enzymatic mechanism and, more importantly, to improve the oxidation of secondary alcohols [86]. Similar simulations on toluene 4-Monooxygenase allowed designing an improved (2.1x) variant for 2-phenylethanol oxidation [87].

A complete substrate migration study exploring the entire enzyme surface was used to compare two manganese peroxidases: one active towards ABTS and another one with no activity. Simulations identified the oxidation active site and allowed to introduce two solvent exposed mutations activating the ABTS-inactive manganese peroxidase achieving high turnovers rates [88].

In another set of enzymatic studies on laccases, PELE’s induced fit procedure was coupled with QM/MM spin density single point calculations, achieving excellent prediction of oxidation rates on different substrates[89]. This protocol was later expanded using an *in silico* directed evolution protocol,

allowing to improve activity and to modify the specificity on laccases [90].

In summary, in this book chapter we aimed at providing first a quick overview of the development and applications of MC techniques in drug design. Then, we centered on delivering a deeper analysis of the techniques and capabilities in PELE, our own MC code that provides a quick and accurate manner to map protein ligand interactions.

## Acknowledgments

Along the development of PELE in the last years, we gratefully acknowledge financial support from the European Union (in particular from the ERC program) and from the Catalan and Spanish Governments. In addition we want to thank all present and past members from the EAPM lab. at BSC for their dedication and work.

## References

1. Special Issue, Stanislaw Ulam 1909-1984. (1987) In Los Alamos Sci.
2. Vitalis, A., and R. V. Pappu. (2009) Methods for Monte Carlo simulations of biomacromolecules. *Annu. Rep. Comput. Chem.* **5**,49-76.
3. Robert, C. P. (2014) Monte Carlo Methods. In Wiley StatsRef: Statistics Reference Online. John Wiley & Sons, Ltd.
4. Liu, J. S. (2004) Monte Carlo Strategies for Scientific Computing. Springer, New York.
5. Hastings, W. K. (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**,97-109.
6. Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. (1953) Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21**,1087-1092.
7. Ponder, J. W., and D. A. Case. (2003) Force fields for protein simulations. In Protein Simulations. 27-+.
8. Frenkel, D., and B. Smit. (2002) Understanding Molecular Simulation. From Algorithms to Applications. Academic Press, San Diego.
9. Paquet, E., and H. L. Viktor. (2015) Molecular dynamics, monte carlo simulations, and langevin dynamics: a computational review. *Biomed Res Int* **2015**,183918.
10. Favrin, G., A. Irbäck, and F. Sjunnesson. (2001) Monte Carlo update for chain molecules: Biased Gaussian steps in torsional space. *The Journal of Chemical Physics* **114**,8154-8158.
11. Yun, M. R., R. Lavery, N. Mousseau, K. Zakrzewska, and P. Derreumaux. (2006) ARTIST: an activated method in internal coordinate space for sampling protein energy landscapes. *Proteins* **63**,967-975.
12. Jorgensen, W. L., and J. Tirado-Rives. (2005) Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **26**,1689-1700.
13. Barker, J. A., and R. O. Watts. (1969) Structure of water; A Monte Carlo calculation. *Chem. Phys. Lett.* **3**,144-145.
14. Dashevsky, V. G., and G. N. Sarkisov. (1974) The solvation and hydrophobic interaction of

non-polar molecules in water in the approximation of interatomic potentials: The Monte Carlo method. *Mol. Phys.* **27**,1271-1290.

15. Owicki, J. C., and H. A. Scheraga. (1977) Monte Carlo calculations in the isothermal-isobaric ensemble. 2. Dilute aqueous solution of methane. *JACS* **99**,7413-7418.
16. Swaminathan, S., S. W. Harrison, and D. L. Beveridge. (1978) Monte Carlo studies on the structure of a dilute aqueous solution of methane. *JACS* **100**,5705-5712.
17. Lal, M. (1969) 'Monte Carlo' computer simulation of chain molecules. I. *Mol. Phys.* **17**,57-64.
18. Hagler, A. T., P. S. Stern, R. Sharon, J. M. Becker, and F. Naider. (1979) Computer simulation of the conformational properties of oligopeptides. Comparison of theoretical methods and analysis of experimental results. *JACS* **101**,6842-6852.
19. Wako, H., and H. A. Scheraga. (1981) Use of distance constraints to fold a protein. *Macromolecules* **14**,961-969.
20. Lee, W.-G., R. Gallardo-Macias, K. M. Frey, K. A. Spasov, M. Bollini, K. S. Anderson, and W. L. Jorgensen. (2013) Picomolar Inhibitors of HIV Reverse Transcriptase Featuring Bicyclic Replacement of a Cyanovinylphenyl Group. *JACS* **135**,16705-16713.
21. Michel, J., and J. W. Essex. (2008) Hit Identification and Binding Mode Predictions by Rigorous Free Energy Simulations. *J. Med. Chem.* **51**,6654-6664.
22. Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *JACS* **118**,11225-11236.
23. Michel, J., and J. W. Essex. (2010) Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput. Aided Mol. Des.* **24**,639-658.
24. C J Woods, J. M., M Bodnarchuk, S Genheden, R Bradshaw, G A Ross, C Cave-Ayland, H Bruce-Macdonald, A I Cabedo Martinez and J Graham. 2017. ProtoMS 3.3. <http://protoms.org>.
25. Ross, G. A., M. S. Bodnarchuk, and J. W. Essex. (2015) Water Sites, Networks, And Free Energies with Grand Canonical Monte Carlo. *JACS* **137**,14930-14943.
26. Bodnarchuk, M. S., R. Viner, J. Michel, and J. W. Essex. (2014) Strategies to Calculate Water Binding Free Energies in Protein–Ligand Complexes. *J. Chem. Inf. Model.*
27. Kirkpatrick, S., C. D. Gelatt, Jr., and M. P. Vecchi. (1983) Optimization by simulated annealing. *Science* **220**,671-680.
28. Cole, D. J., J. Tirado-Rives, and W. L. Jorgensen. (2014) Enhanced Monte Carlo Sampling through Replica Exchange with Solute Tempering. *J. Chem. Theory Comput.* **10**,565-571.
29. Woods, C. J., J. W. Essex, and M. A. King. (2003) The Development of Replica-Exchange-Based Free-Energy Methods. *The Journal of Physical Chemistry B* **107**,13703-13710.
30. Abagyan, R., and M. Totrov. (1994) Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **235**,983-1002.
31. Neyts, E. C., B. J. Thijsse, M. J. Mees, K. M. Bal, and G. Pourtois. (2012) Establishing Uniform Acceptance in Force Biased Monte Carlo Simulations. *J. Chem. Theory Comput.* **8**,1865-1869.
32. Noguti, T., and N. Gō. (1985) Efficient monte carlo method for simulation of fluctuating conformations of native proteins. *Biopolymers* **24**,527-546.
33. Li, Z. Q., and H. A. Scheraga. (1987) Monte Carlo -minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **84**,2633-2636.
34. Chang, G., W. C. Guida, and W. C. Still. (1989) An internal-coordinate Monte Carlo method for searching conformational space. *JACS* **111**,4379-4386.
35. Evans, D. A., and D. J. Wales. (2003) Free energy landscapes of model peptides and proteins. *The Journal of Chemical Physics* **118**,3891-3897.
36. Wang, F., and D. P. Landau. (2001) Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86**,2050-2053.

37. El-Mellouhi, F., N. Mousseau, and L. J. Lewis. (2008) Kinetic activation-relaxation technique: An off-lattice self-learning kinetic Monte Carlo algorithm. *Physical Review B* **78**,153202.
38. Borrelli, K. W., A. Vitalis, R. Alcantara, and V. Guallar. (2005) PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput.* **1**,1304-1311.
39. Duane, S., A. D. Kennedy, B. J. Pendleton, and D. Roweth. (1987) Hybrid Monte Carlo. *Phys. Lett. B* **195**,216-222.
40. Nilmeier, J. P., G. E. Crooks, D. D. L. Minh, and J. D. Chodera. (2011) Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation. *Proceedings of the National Academy of Sciences* **108**,E1009-E1018.
41. Samuel, G., L. Nathan M., G. Patrick, R. Ariën S., F. Josh, R. Gregory, C. John D., and M. David. 2017. Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes Using Nonequilibrium Candidate Monte Carlo.
42. Jorgensen, W. L. (2016) Computer-aided discovery of anti-HIV agents. *Bioorg. Med. Chem.* **24**,4768-4778.
43. Lovering, F., C. Aevazelis, J. Chang, C. Dehnhardt, L. Fitz, S. Han, K. Janz, J. Lee, N. Kaila, J. McDonald, W. Moore, A. Moretto, N. Papaioannou, D. Richard, M. S. Ryan, Z. K. Wan, and A. Thorarensen. (2016) Imidazotriazines: Spleen Tyrosine Kinase (Syk) Inhibitors Identified by Free-Energy Perturbation (FEP). *ChemMedChem* **11**,217-233.
44. Wang, N. X., G. B. Watson, M. R. Loso, and T. C. Sparks. Molecular modeling of sulfoxaflor and neonicotinoid binding in insect nicotinic acetylcholine receptors: impact of the Myzus beta1 R81T mutation.
45. Peng, Z., D. S. Maxwell, D. Sun, B. A. Bhanu Prasad, A. Pal, S. Wang, J. Balatoni, P. Ghosh, S. T. Lim, A. Volgin, A. Shavrin, M. M. Alauddin, J. G. Gelovani, and W. G. Bornmann. Imatinib analogs as potential agents for PET imaging of Bcr-Abl and c-KIT expression at a kinase level.
46. Neves, M. A., M. Totrov, and R. Abagyan. (2012) Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J. Comput. Aided Mol. Des.* **26**,675-686.
47. Cross, J. B., D. C. Thompson, B. K. Rai, J. C. Baber, K. Y. Fan, Y. Hu, and C. Humblet. (2009) Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **49**,1455-1474.
48. Tubert-Brohman, I., W. Sherman, M. Repasky, and T. Beuming. (2013) Improved Docking of Polypeptides with Glide. *J. Chem. Inf. Model.* **53**,1689-1699.
49. Salameh, B. A., I. Cumpstey, A. Sundin, H. Leffler, and U. J. Nilsson. (2010) 1H-1,2,3-Triazol-1-yl thiodigalactoside derivatives as high affinity galectin-3 inhibitors. *Bioorg. Med. Chem.* **18**,5367-5378.
50. Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. (2001) Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.* **80**,505-515.
51. Cossins, B. P., A. Hosseini, and V. Guallar. (2012) Exploration of Protein Conformational Change with PELE and Meta-Dynamics. *J. Chem. Theory Comput.*
52. Eyal, E., L. W. Yang, and I. Bahar. (2006) Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* **22**,2619-2627.
53. Setny, P., and M. Zacharias. (2013) Elastic Network Models of Nucleic Acids Flexibility. *J. Chem. Theory Comput.* **9**,5460-5470.
54. Jacobson, M. P., G. A. Kaminski, R. A. Friesner, and C. S. Rapp. (2002) Force field validation using protein side chain prediction. *J. Phys. Chem. B* **106**,11673-11680.
55. Zhu, K., M. R. Shirts, R. A. Friesner, and M. P. Jacobson. (2007) Multiscale optimization of a truncated Newton minimization algorithm and application to proteins and protein-ligand

- complexes. *J. Chem. Theory Comput.* **3**,640-648.
56. Lecina, D. 2017. A study on protein-ligand interactions using a Monte Carlo procedure. In *Physics*. Universitat de Barcelona, Barcelona.
  57. Borrelli, K. W., A. Vitalis, R. Alcantara, and V. Guallar. (2005) PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput.* **1**,1304–1311.
  58. Lecina, D., J. F. Gilabert, and V. Guallar. (2017) Adaptive simulations, towards interactive protein-ligand modeling. *Sci. Rep.* **7**,8466.
  59. Banks, J. L., H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio, and R. M. Levy. (2005) Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **26**,1752-1780.
  60. Kaminski, G. A., R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. (2001) Evaluation and Reparametrization of the Opls-Aa Force Field for Proteins Via Comparison With Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **105**,6474-6487.
  61. Pérez, A., I. Marchán, D. Svozil, J. Spöner, T. E. Cheatham, C. A. Laughton, and M. Orozco. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of  $\alpha/\gamma$  conformers. *Biophys. J.* **92**,3817-3829.
  62. Onufriev, A., D. Bashford, and D. A. Case. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **55**,383-394.
  63. Schaefer, M., C. Bartels, and M. Karplus. (1998) Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* **284**,835-848.
  64. Ghosh, A., C. S. Rapp, and R. A. Friesner. (1998) Generalized Born Model Based on a Surface Integral Formulation. *J. Phys. Chem. B* **102**,10983-10990.
  65. Gallicchio, E., L. Y. Zhang, and R. M. Levy. (2002) The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.* **23**,517-529.
  66. Zhu, K., M. R. Shirts, and R. A. Friesner. (2007) Improved Methods for Side Chain and Loop Predictions via the Protein Local Optimization Program: Variable Dielectric Model for Implicitly Improving the Treatment of Polarization Effects. *J. Chem. Theory Comput.* **3**,2108-2119.
  67. Marti, M. A., A. Bidon-Chanal, A. Crespo, S. R. Yeh, V. Guallar, F. J. Luque, and D. A. Estrin. (2008) Mechanism of product release in NO detoxification from Mycobacterium tuberculosis truncated hemoglobin N. *J. Am. Chem. Soc.* **130**,1688-1693.
  68. Guallar, V., C. Lu, K. Borrelli, T. Egawa, and S.-R. Yeh. (2009) Ligand Migration in the Truncated Hemoglobin-II from Mycobacterium tuberculosis: The Role of G8 tryptophan. *J. Biol. Chem.* **284**,3106-3116.
  69. Lucas, M. F., and V. Guallar. (2012) An Atomistic View on Human Hemoglobin Carbon Monoxide Migration Processes. *Biophys. J.* **102**,887-896.
  70. Shan, Y., E. T. Kim, M. P. Eastwood, R. O. Dror, M. A. Seeliger, and D. E. Shaw. (2011) How Does a Drug Molecule Find Its Target Binding Site? *JACS* **133**,9181-9183.
  71. Kotev, M., D. Lecina, T. Tarragó, E. Giralt, and V. Guallar. (2015) Unveiling Prolyl Oligopeptidase Ligand Migration by Comprehensive Computational Techniques. *Biophys. J.* **108**,116-125.
  72. Kopečná, J., I. Cabeza de Vaca, N. B. P. Adams, P. A. Davison, A. A. Brindley, C. N. Hunter, V. Guallar, and R. Sobotka. (2015) Porphyrin Binding to Gun4 Protein, Facilitated by a



- Flexible Loop, Controls Metabolite Flow through the Chlorophyll Biosynthetic Pathway. *J. Biol. Chem.* **290**,28477-28488.
73. Cabeza de Vaca, I., M. F. Lucas, and V. Guallar. (2015) New Monte Carlo Based Technique To Study DNA–Ligand Interactions. *J. Chem. Theory Comput.* **11**,5598-5605.
74. Espona-Fiedler, M., V. Soto-Cerrato, A. Hosseini, J. M. Lizcano, V. Guallar, R. Quesada, T. Gao, and R. Perez-Tomas. (2012) Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: prodigiosin vs. obatoclax. *Biochem. Pharmacol.* **83**,489-496.
75. Hosseini, A., M. Espona-Fiedler, V. Soto-Cerrato, R. Quesada, R. Pérez-Tomás, and V. Guallar. (2013) Molecular interactions of prodiginines with the BH3 domain of anti-apoptotic Bcl-2 family members. *PLoS One* **8**.
76. Andreotti, G., I. Cabeza de Vaca, A. Poziello, M. C. Monti, V. Guallar, and M. V. Cubellis. (2014) Conformational response to ligand binding in phosphomannomutase2: insights into inborn glycosylation disorder. *J. Biol. Chem.* **289**,34900-34910.
77. Edman, K., A. Hosseini, M. K. Bjursell, A. Aagaard, L. Wissler, A. Gunnarsson, T. Kaminski, C. Köhler, S. Bäckström, T. J. Jensen, A. Cavallin, U. Karlsson, E. Nilsson, D. Lecina, R. Takahashi, C. Grebner, M. Lepisto, A. C. Hogner, and V. Guallar. (2015) Ligand binding mechanism in steroid receptors; from conserved plasticity to differential evolutionary constraints. *Structure* **23**,2280-2290.
78. Grebner, C., D. Lecina, V. A. Gil, J. Ulander, P. Hansson, C. Dellsen, C. Tyrchan, K. Edman, Anders C. Hogner, and V. Guallar. (2017) Exploration of binding mechanisms in nuclear hormone receptors by Monte Carlo simulations and X-ray derived motion modes. *Biophysical Journal* **ASAP**.
79. Takahashi, R., V. A. Gil, and V. Guallar. (2014) Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations. *J. Chem. Theory Comput.* **10**,282-288.
80. Borrelli, K. W., B. Cossins, and V. Guallar. (2010) Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J. Comput. Chem.* **31**,1224-1235.
81. Hosseini, A., A. Alibés, M. Noguera-Julian, V. A. Gil, R. Paredes, R. Soliva, M. Orozco, and V. Guallar. (2016) Computational Prediction of HIV-1 Resistance to Protease Inhibitors. *J. Chem. Inf. Model.* **56**,915–923.
82. Carlson, H. A., R. D. Smith, K. L. Damm-Ganamet, J. A. Stuckey, A. Ahmed, M. A. Convery, D. O. Somers, M. Kranz, P. A. Elkins, G. Cui, C. E. Peishoff, M. H. Lambert, and J. B. Dunbar. (2016) CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. *J. Chem. Inf. Model.* **56**,1063.
83. Madadkar-Sobhani, A., and V. Guallar. (2013) PELE web server: atomistic study of biomolecular systems at your fingertips. *Nucleic Acids Res.* **41**,W322-W328.
84. Hernandez-Ortega, A., K. Borrelli, P. Ferreira, M. Medina, A. T. Martinez, and V. Guallar. (2011) Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J* **436**,341-350.
85. Carro, J., M. Martinez, M. Medina, A. T. Martinez, and P. Ferreira. (2017) Protein dynamics promote hydride tunnelling in substrate oxidation by aryl-alcohol oxidase. *Phys. Chem. Chem. Phys.*
86. Hernandez-Ortega, A., F. Lucas, P. Ferreira, M. Medina, V. Guallar, and A. T. Martinez. (2012) Role of active site histidines in the two half-reactions of the aryl-alcohol oxidase catalytic cycle. *Biochemistry* **51**,6595-6608.
87. Hosseini, A., M. Brouk, M. F. Lucas, F. Glaser, A. Fishman, and V. Guallar. (2014) Atomic picture of ligand migration in toluene 4-monooxygenase. *The Journal of Physical Chemistry B* **119**,671-678.

88. Acebes, S., E. Fernandez-Fueyo, E. Monza, M. F. Lucas, D. Almendral, F. J. Ruiz-Dueñas, H. Lund, A. T. Martinez, and V. Guallar. (2016) Rational Enzyme Engineering Through Biophysical and Biochemical Modeling. *ACS Catalysis* **6**,1624-1629.
89. Lucas, M. F., E. Monza, L. J. Jørgensen, H. A. Ernst, K. Piontek, M. J. Bjerrum, Á. T. Martinez, S. Camarero, and V. Guallar. (2017) Simulating Substrate Recognition and Oxidation in Laccases: From Description to Design. *J. Chem. Theory Comput.* **13**,1462-1467.
90. Giacobelli, V. G., E. Monza, M. Fatima Lucas, C. Pezzella, A. Piscitelli, V. Guallar, and G. Sannia. (2017) Repurposing designed mutants: a valuable strategy for computer-aided laccase engineering - the case of POXA1b. *Catalysis Science & Technology* **7**,515-523.