UNIVERSITÀ DI PARMA

DOTTORATO DI RICERCA IN
Scienze del Farmaco, delle Biomolecole e dei Prodotti per la Salute
CICLO XXX

Application of Enhanced Sampling Methods in Drug Design

Coordinatore:
Chiar.mo Prof. Marco Mor

Tutore:
Chiar.mo Prof. Alessio Lodola

Dottoranda:
Donatella Callegari

Anni 2014/2017

Acknowledgements

First and foremost, I want to thank my supervisor Prof. Alessio Lodola: without his encouragement and guidance, this thesis would not have been possible. I would like to thank him for all his contributions of time and ideas to make my Ph.D. experience so productive and stimulating, whilst allowing me to propose my own ideas. His enthusiasm for research was contagious and motivational for me, even during tough and crying times in this Ph.D. quest. One simply could not wish for a better or friendlier supervisor.

I am also grateful to Prof. Marco Mor and Prof. Silvia Rivara for their patience with me and with my attitude. All their constructive comments in these three years were a great help to me for growing up from both a scientific and a human point of view.

I would particularly like to thank Prof. Adrian J. Mulholland for hosting me in Bristol, allowing me to live a lovely and yet stimulating and formative working experience abroad.

Special thanks also to "*le shampiste*" Laura Scalvini and Martina Maccesi, coworkers and especially, friends. I want to thank them for the fun and the time spent together in and out of the office, but also for all the good advices and the reciprocal support during the hard times in the room 116. Without them I would not have survived.

My three years of Ph.D. in Parma were made enjoyable also due to the new friends that became a part of my life. I am grateful for the time, the quests, the company and the beers with the "*wyrm*" buddies: Ralphas, Ippus, Simo, Gatto, Costy, Max.

Finally, a special thanks to my beloved Fabio Pesci (Pes) for his courage to fall in love with a girl at the end of her Ph.D., for all his support and for having changed my life.

Summary

1 Enhanced Sampling Methods in Drug Design 7

1.1 Introduction 9
1.2 Aim of the work 13
1.3 Products of the thesis 14

2 Methods 17

2.1 Molecular Dynamics 19
2.2 Metadynamics 21
  2.2.1 Classic Metadynamics 21
  2.2.2 Well-tempered metadynamics 23
2.3 Umbrella Sampling 24
2.4 QM/MM methods 25

3 Identification of binding modes for small molecules ... 29

3.1 The docking problem 31
3.2 EphA2-EphrinA1 inhibitors 33
3.3 Results and discussion 37
  3.3.1 Identification of LCA “correct” binding mode 37
  3.3.2 Exploiting alternative binding modes 43
  3.3.3 Identification of UniPR129 binding mode 45
3.4 Conclusions 53
3.5 Computational protocol 54

4 Investigating the kinetics of unbinding 57

4.1 Residence time in drug discovery 59
4.2 Dataset of CDK8 inhibitors 62
4.3 Results and discussion 64
4.4 Conclusions 71
4.5 Computational protocol 72

5 Covalent drugs affinity and reactivity 75

5.1 Covalent drugs design and optimization 77
5.2 EGFR inhibitors in NSCLC treatment 79
5.3 Results and discussion...83

5.3.1 Ionization state of Cys797 83
5.3.2 Reaction energetics for Cys797 alkylation. 86
5.3.3 Binding of osimertinib to EGFR mutants 91

5.4 Conclusions 103
5.5 Computational protocol 104

6 Conclusions and future perspectives 111

7 Appendix 115

Chapter 1

Enhanced Sampling Methods in Drug Design

1.1 Introduction

Understanding the mechanism of the recognition process between a ligand and its biological target is becoming increasingly important in the identification of new hits targeting "receptors" of pharmaceutical interest as well as in the optimization of lead compounds. Computational structure-based drug design can accelerate these challenging processes.[1] In the past decade, the rapid development of powerful architectures and better algorithms allowed to perform high-level calculations in reasonable time, allowing computationally driven drug design to increase his impact in medicinal chemistry.[2,3] Computationally driven drug design always needs novel computational approaches that can integrate the amount of information available for small compounds and their pharmaceutically relevant targets. Modern computational approaches aiding drug design require to accurately account for the complexity of drug−target complex formation, considering also the intrinsic flexibility and dynamical behavior of molecular complexes.[4] Things are further complicated by the fact that the awareness of the factors modulating the pharmacological activity of drugs is still evolving. Indeed, the relevance of new parameters such as drug-target *residence time* (i.e. the time a target is occupied by a compound) is being increasingly recognized.[5] It is now well understood that the time-course of drug efficacy depends not only on the drug affinity, but also on the residence time of the drug compound on its molecular target.[6] This parameter can represent an important additional metric to be explored alongside affinity for drug optimization during early-phase drug discovery.[7] In this context, the availability of computational tools able to account also for the kinetics of protein-ligand association and dissociation is required.

---

[1] Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303* (5665), 1813–1818.

[2] Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, 5, 1632−1639.

[3] Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. Accelerating Molecular Modeling Applications with Graphics Processors. *J. Comput. Chem.* **2007**, 28, 2618−2640.

[4] Spyrakis, F.; Bidon Chanal, A.; Barril, X.; Luque, F. J. Protein Flexibility and Ligand Recognition: Challenges for Molecular Modeling. *Curr. Top. Med. Chem.* **2011**, 11, 192−210.

[5] Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Opinion - Drugtarget Residence Time and Its Implications for Lead Optimization. *Nat. Rev. Drug Discovery* **2006**, 5, 730−739.

[6] Tummino, P. J.; Copeland, R. A. Residence Time of Receptor-Ligand Complexes and Its Effect on Biological Function. *Biochemistry* **2008**, 47, 5481−5492.

[7] Copeland, R. A. The Drug-Target Residence Time Model: A 10- Year Retrospective. *Nat. Rev. Drug Discovery* **2016**, 15, 87−95.

In the last decade, several computational methods have been developed to address the challenging issue of ligand-protein recognition. The simplest approach is constituted by molecular docking simulations which face this issue often neglecting conformational flexibility and solvent effects. Ligand-docking software, while able to rapidly find reasonable configurations of a small molecule within a known binding site, fail in correctly estimating the binding free energy, due to several approximations encoded in their scoring function.[8]

Full atomistic methods such as molecular dynamics (MD) could overcome most of the concerns of docking methods, as they describe the time evolution of the ligand-protein complexes (i.e., by integrating the equation of motion starting from the forces acting on the atoms on the base of a classical force field), accounting for structural fluctuation, flexibility and conformational rearrangements.[9] Compared to molecular docking, MD simulations require a much longer computation and while they allow to characterize local minima of protein-ligand complexes, they are not able to sample transitions separated by high barriers (*rare events*). Rare events, such as ligand binding and unbinding processes, require a significant amount of energy to occur, therefore they are expected to take place on long-timescale. Indeed, drug-like molecules with long residence times (more than an hour) are common, and their unbinding cannot be observed by conventional MD calculations even when specialized hardware is used. To overcome the timescale limit, diverse approaches aimed at accelerating conformational transition in the framework of MD simulations have been developed.[10] These advanced techniques are often referred to as *enhance sampling methods* and they include approaches that work at the equilibrium, such as free-energy perturbation[11] and umbrella sampling,[12] and methods which work in condition of non-equilibrium, such as steered MD[13] and metadynamics.[14]

---

[8] Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *J. Am. Chem. Soc*. **2005**, 127, 2600− 2607.

[9] De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem*. **2016**, 59, 4035−4061.

[10] Rocchia, W.; Masetti, M.; Cavalli, A. Enhanced sampling methods in drug design. In *Physico-Chemical and Computational Approaches to Drug Discovery*, 1st ed.; Luque, J., Barril, X., Eds.; Royal Society of Chemistry: Cambridge, U.K., **2012**; 273−301.

[11] Jorgensen WL, Thomas LL. Perspective on free-energy perturbation calculations for chemical equilibria. *J Chem Theory Comput* **2008**, 4:869–876.

[12] Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J Chem Theory Comput* **1977**, 23:187–199.

[13] Isralewitz, B.; Gao, M.; Schulten, K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Curr. Opin. Struct. Biol*. **2001**, 11, 224−230.

[14] Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 12562−12566.

Among these techniques, metadynamics has been shown to be a useful method for docking simulations, as it provides a good estimation of the ligand docking free energy profile, including the effects of solvent reorganization and conformational rearrangement of protein and ligand during the process of binding.[8,15,16] In metadynamics, sampling is accelerated by adding an external potential bias as a sum of repulsive Gaussians, that act on few degrees of freedom, named collective variables (CVs). The external potential added allows the ligand to escape from his local minima (the bound state) and prevents the ligand from revisiting the already sampled minima.[14] Metadynamics is a powerful approach that identifies free-energy minima, allowing to characterize the bound state and alternative binding modes of ligands within the biological target.[8] Moreover, metadynamics has been recently applied also to estimate transition times,[17] with a perspective application in predicting drug-target residence times and dissociation rate constants.[18]

Even more factors should be considered when designing and optimizing covalent drugs. In this case, not only the non-covalent phase of the recognition process of the drug should be explored, but also the reactivity of the warhead held by a drug, responsible for the formation of a covalent bond between the drug and a protein residue, should be investigated and characterized. When studying the reactivity of covalent drugs, one can be interested in the energetics of the reaction that lead to covalent adduct formation. In this case, classical force fields do not help, as covalent bond breaking and formation require the use of quantum mechanics (QM). Again, it would be advantageous to couple a QM Hamiltonian with a dynamical method which, accounting for conformational and solvent reorganization, gives the free energy profile for the reaction of interest.[19] Among available enhanced sampling methods, umbrella sampling emerged as popular approach for modelling chemical reactions as, when wisely used, it allows to reconstruct converged potential of mean force (PMF) profile. In umbrella sampling, a reaction coordinate is selected, then a set of biased-MD simulations are carried out in each window of the reaction coordinate, where a restraining *umbrella* potential is applied to keep the reaction

---

[15] Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. Exploring Complex Protein-Ligand Recognition Mechanisms with Coarse Metadynamics. *J. Phys. Chem. B* **2009**, 113, 4807−4816.

[16] De Vivo, M.; Cavalli, A. Recent advances in dynamic docking for drug discovery. *WIREs Comput Mol Sci* **2017**. doi: 10.1002/wcms.1320

[17] Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys. Rev. Lett.* **2013**, 111, 230602.

[18] Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of Protein-Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, 112, E386−E391.

[19] Branduardi, D.; De Vivo, M.; Rega, N.; Barone, V.; Cavalli, A. Methyl Phosphate Dianion Hydrolysis in Solution Characterized by Path Collective Variables Coupled with DFT-Based Enhanced Sampling Simulations. *J. Chem. Theory Comput.* **2011**, 7, 539−543.

coordinate close to a desired value. The set of chosen values is designed to cover the range from the reactant complex to the product complex.

It is important to point out that computationally demanding high-level calculations in the field of quantum chemistry are required to simulate the bond breaking and formation. In this case, hybrid quantum mechanics/molecular mechanics approaches (QM/MM) offer a good compromise between accuracy and computation speed, as they treat only the reactive portions of the system with QM level, while the rest of the system is simulated at a lower level, allowing a faster calculation.[20]

Calculations of this kind had already proved to give significant information for drug discovery, allowing to elucidate the mechanism of action of covalent inhibitors, predict drug metabolites and understand drug resistance. [21]

The following chapters underline the potential of enhanced sampling methods in drug design and lead optimization campaigns. Ligand docking mechanism and the kinetics of ligand unbinding are explored using metadynamics, while the reactivity of covalent drugs has been studied with Umbrella Sampling coupled with hybrid QM/MM approach.

---

[20] Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, 11, 700− 733.

[21] Lodola, A.; De Vivo, M. The Increasing Role of QM/MM in Drug Discovery. *Adv. Protein Chem. Struct. Biol.* **2012**, 87, 337−362.

1.2 Aim of the work

During my PhD project, I focused my efforts in developing and applying computational approaches aimed to characterize the complex phenomena of molecular recognition, including both kinetic and thermodynamic aspects. In details, I worked on *i)* prediction of the correct binding mode of small molecules within biological targets of pharmaceutical interest, *ii)* prediction of the kinetics of protein-ligand unbinding and *iii)* estimation of covalent drug affinity and reactivity in mutated forms of the targeted protein. Procedures based on the application of *enhanced sampling methods* are described and employed to solve these challenging tasks.
In this work, two enhanced sampling techniques were mainly applied: Metadynamics and Umbrella Sampling, described in detail in chapter 2.

In chapter 3 the use of metadynamics simulations to identify relevant binding modes of small molecules within the binding pocket of biological targets is described. As real-life example, the procedure was applied to the optimization of small molecules inhibitors of EphA2-ephrinA1 protein-protein interaction, explaining how different binding modes were discriminated when more than a pose was found for one ligand. Then, once the correct binding mode according to structure activity relationship (SAR) data was identified, the knowledge gained about EphA2-inhibitors interaction allowed to drive the synthesis of novel compounds.

In chapter 4, a metadynamics-based procedure aimed at ranking a set of compounds according to their experimental residence times for the same target is reported. The protocol was successfully calibrated on the dataset of type-II inhibitors of Cyclin-Dependent Kinase 8.

Finally, in chapter 5, the molecular mechanism of resistance to lung cancer drug osimertinib was explored. The appearance of new mutations (i.e. L718Q) is currently limiting the benefit of osimertinib, an irreversible inhibitor designed to selectively target EGFR T790M and capable of alkylating Cys797. In this chapter, a computational investigation that clarifies the impact of L718Q mutation on *i)* ionization state of Cys797; *ii)* activation energy and reaction energetics for Cys797 alkylation; *iii)* free-energy of binding and *iv)* conformational space of osimertinib, was performed to rationalize the clinical inactivity of osimertinib on EGFR T790M/L718Q variant, giving useful insights for the design of novel inhibitors.

1.3 Products of the thesis

The results discussed in this thesis have been reported in the following publications:

1. **Callegari, D.**; Ranaghan, K. E.; Woods, C. J.; Minari, R.; Tiseo, M.; Mor, M.; Mulholland, A.J.; Lodola, A. L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib. *Submitted.*

2. **Callegari, D**.; Lodola, A.; Pala, D.; Rivara, S.; Mor, M.; Rizzi, A.; Capelli, A. M. Metadynamics Simulations Distinguish Short- and Long-Residence-Time Inhibitors of Cyclin-Dependent Kinase 8. *J Chem Inf Model* **2017**, *57* (2), 159–169.

3. Incerti, M.; Russo, S.; **Callegari, D**.; Pala, D.; Giorgio, C.; Zanotti, I.; Barocelli, E.; Vicini, P.; Vacondio, F.; Rivara, S.; Castelli, R.; Tognolini, M.; Lodola, A. Metadynamics for Perspective Drug Design: Computationally Driven Synthesis of New Protein-Protein Interaction Inhibitors Targeting the EphA2 Receptor. *J. Med. Chem.* **2017**, *60* (2), 787–796.

4. Russo, S.; **Callegari, D**.; Incerti, M.; Pala, D.; Giorgio, C.; Brunetti, J.; Bracci, L.; Vicini, P.; Barocelli, E.; Capoferri, L.; Rivara, S.; Tognolini, M.; Mor, M.; Lodola, A. Exploiting Free-Energy Minima to Design Novel EphA2 Protein-Protein Antagonists: From Simulation to Experiment and Return. *Chemistry* **2016**, *22* (24), 8048–8052.

The results reported in this thesis have also been object to the following communications:

1. Callegari, D. Ranking Short- and Long-Residence-Time Inhibitors of Cyclin-Dependent Kinase 8 with Metadynamics Simulations. *EUROPIN Summer School of Drug Design*, Vienna, September 17-22, **2017**

2. Callegari, D. Ranking Short- and Long-Residence-Time Inhibitors of Cyclin-Dependent Kinase 8 with Metadynamics Simulations. *European School of Medicinal Chemistry ESMEC*, Urbino, July 2-6, **2017**

3. Callegari, D.; Pala, D.; Tognolini, M.; Russo, S.; Incerti, M.; Rivara, S.; Mor, M.; Lodola, A. Exploiting binding minima identified by free-energy simulations to design novel EphA2-ephrinA1 inhibitors. *First congress on the Eph/ephrin system*, Parma, May 5-6, **2016.**

Further results obtained during my PhD and related to this project but not discussed in this thesis, are reported in the following pubblications:

1. **Callegari, D**.; Pala, D.; Scalvini, L.; Tognolini, M.; Incerti, M.; Rivara, S.; Mor, M.; Lodola, A. Comparative Analysis of Virtual Screening Approaches in the Search for Novel EphA2 Receptor Antagonists. *Molecules* **2015**, *20* (9), 17132–17151.

2. Castelli, R.; Tognolini, M.; Vacondio, F.; Incerti, M.; Pala, D.; **Callegari, D**.; Bertoni, S.; Giorgio, C.; Hassan-Mohamed, I.; Zanotti, I.; Bugatti, A.; Rusnati, M.; Festuccia, C.; Rivara, S.; Barocelli, E.; Mor, M.; Lodola, A. Δ(5)-Cholenoyl-Amino Acids as Selective and Orally Available Antagonists of the Eph-Ephrin System. *Eur J Med Chem* **2015**, *103*, 312–324.

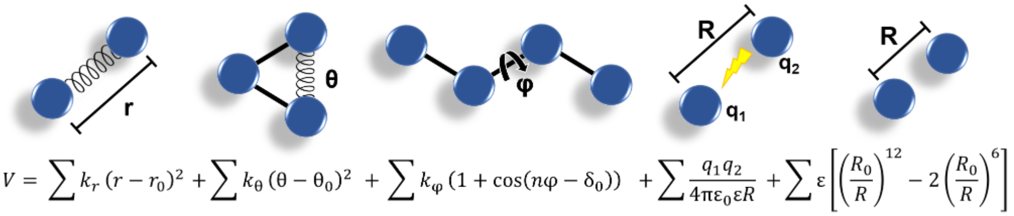
3. Giorgio, C.; Mena, P.; Del Rio, D.; Brighenti, F.; Barocelli, E.; Hassan-Mohamed, I.; **Callegari, D**.; Lodola, A.; Tognolini, M. The Ellagitannin Colonic Metabolite Urolithin D Selectively Inhibits EphA2 Phosphorylation in Prostate Cancer Cells. *Mol Nutr Food Res* **2015**, *59* (11), 2155–2167.

Chapter 2

Methods

2.1 Molecular Dynamics

Molecular dynamics (MD) studies the time-dependent evolution of a molecular system by calculating the interactions and motions of the atoms according to Newton's physics. A molecular mechanics (MM) force field is used to describe the forces between interacting atoms and calculate the overall energy of the system. Figure 2.1 shows graphically the different contributions to the energy of the system.



$$V = \sum k_r\,(r - r_0)^2 + \sum k_\theta\,(\theta - \theta_0)^2 + \sum k_\varphi\,(1 + \cos(n\varphi - \delta_0)) + \sum \frac{q_1 q_2}{4\pi\varepsilon_0 \varepsilon R} + \sum \varepsilon\left[\left(\frac{R_0}{R}\right)^{12} - 2\left(\frac{R_0}{R}\right)^{6}\right]$$

Figure 2.1. Molecular mechanics potential energy function. Atoms represented as blue spheres, bonds as black lines.

The first three terms of the equation in figure 2.1 represent intramolecular interactions of the atoms. They describe variations in potential energy as a function of bond stretching, bending, and torsions between bonded atoms. They are calculated by summations over bond lengths (r), angles (θ), and dihedral angles (φ), respectively. The fourth and fifth terms in the equation represent electrostatic and van der Waals interactions between atoms, respectively, and are denoted as "nonbonded" terms. Both the nonbonded terms are expressed as an inverse power function of the distance R between the considered atoms. The electrostatic contribution is described with the Coulomb potential, where q_1 and q_2 are the partial charges of a pair of atoms, ε_0 stands for the permittivity of free space, and ε is the relative permittivity (or dielectric constant), which takes a value of 1 in vacuum. Finally, the van der Waals interactions are generally treated with a 12−6 Lennard-Jones potential, where ε is a parameter defining the depth of the energy well, whereas R_0 is the minimum energy distance that equals the sum of the van der Waals radii of the two interacting atoms.

During a MD simulation, Newton's laws of motions are integrated, calculating positions and velocities of the atoms at every timestep of MD simulation and thus generating a new configuration of the evolving system at each timestep. The process is repeated in an iterative way for a defined number of timesteps (figure 2.2). In this way, at the end of MD simulation, a trajectory that specifies positions and velocities of all the atoms over time is provided.

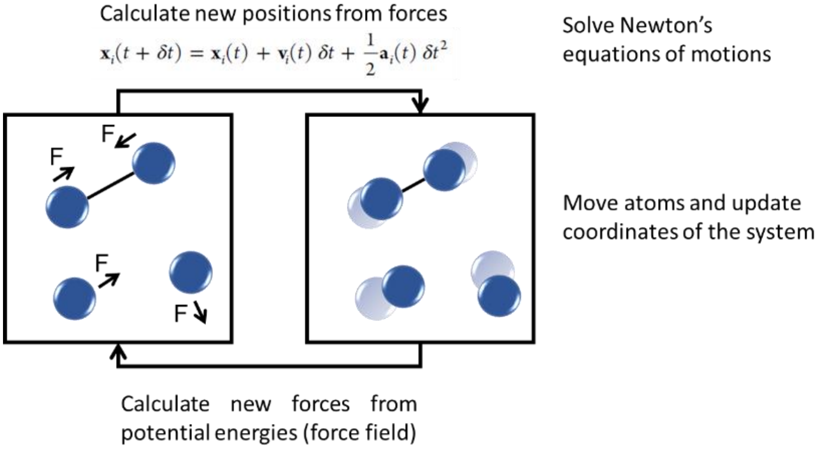


Figure 2.2. scheme of the iterative calculations during a MD simulation.

A typical MD simulation needs to employ a timestep of the order of femtoseconds (i.e., 10^{-15} seconds) to correctly integrate the equations of motions. This is required to properly solve the fastest motion of the system (bond stretching and bending). For this reason, the computational cost of a MD is very high: even with the increased computer power available nowadays, only up to milliseconds of simulations can be performed at a reasonable computational time. Unfortunately, interesting events in protein-ligand recognition often take place on a much longer timescale. Indeed, MD simulation can present some difficulties when sampling complex protein-ligand unbinding processes, which can take place in seconds, minutes or even hours. The results of a MD simulation are meaningful only if the simulation is long enough to allow the system to visit all the energetically relevant configurations. The major difficulty is that some configurations might be separated by high free-energy barriers. Indeed, during a MD simulation, the system may stay trapped for a very long time in one configuration, because it's surrounded by high energy barriers. In such case, obtaining enough statistics of all relevant configurations requires an unreasonable amount of computer time, even with the use of new powerful hardware.

Several enhanced sampling methods have been developed to speed up sampling and overcome this problem. This thesis focuses on the use of two of these approaches: Metadynamics and Umbrella Sampling.

2.2 Metadynamics

Metadynamics (MetaD) is a powerful approach for accelerating sampling and computing multidimensional free-energy surfaces. Since its introduction, MetaD has significantly evolved and the most important improvement is the so-called "well-tempered metadynamics". In the following paragraphs, MetaD in the classic and well-tempered implementations are described focusing on its application to study protein-ligand association and dissociation processes.

2.2.1 Classic Metadynamics[22,23]

In a MetaD simulation, the protein-ligand binding or unbinding process is accelerated by using an external history-dependent potential that allows the system to escape from the local free-energy minimum (e.g. the bound state) where it can be trapped and that prevents the system to revisit configurations already sampled. In this way sampling of the complex free-energy surface (FES) of binding and unbinding processes is significantly accelerated.

The added external potential can be written as a sum of Gaussians added continuously during the simulation in the space defined by a set of coordinates, called *collective variables* (CV):

$$V_G(S,t) = \int_0^t dt' \omega \exp\left(-\sum_{i=1}^{d} \frac{(S_i(R) - S_i(R(t')))^2}{2\sigma_i^2} \right) \tag{1}$$

where ω is an energy rate, S is a set of functions of the microscopic coordinates R of the system and σ_i is the width of the Gaussian for the ith CV. The energy rate ω is constant and usually expressed in terms of a Gaussian height W and a deposition stride τ_G:

$$\omega = \frac{W}{\tau_G}. \tag{2}$$

---

[22] Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 12562−12566.

[23] Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Comput Mol Sci* **2011**, *1* (5), 826–843.

Figure 2.3 shows the effect of the external potential on the evolution of the system. Let assume a very simple case in which free-energy is reported in function of only one collective variable (CV) and only three local minima A, B and C are present. (Figure 2.3a). The system is prepared in minimum A as starting point of the simulation. In a standard MD simulation, the system would remain stuck in this minimum because barriers are larger than thermal fluctuations. Instead, in the MetaD simulation, as time goes by, Gaussians are deposited on CV, causing the underlying bias potential to grow and filling the basin A (Figure 2.3 b, c, d), until eventually the system is pushed out of the basin A into a new local minimum (Figure 2.3e). Once in the new basin, the accumulation of Gaussians starts again (Figure 2.3f). Finally, when all the local minima are completely filled, the system diffuses on the FES.

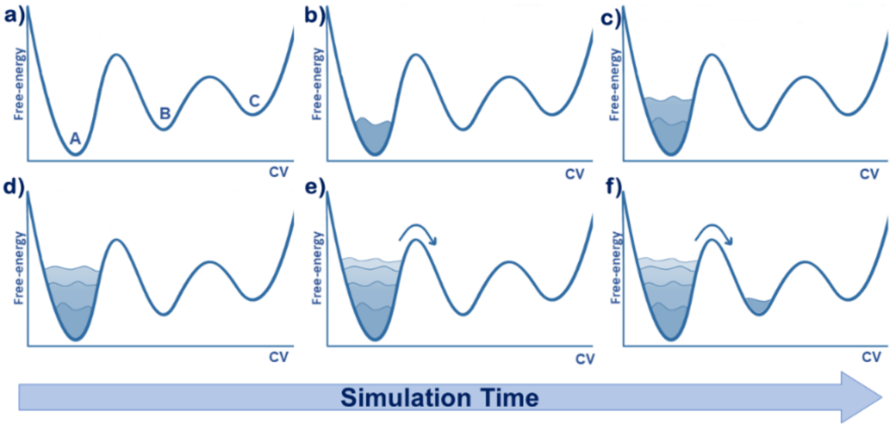


Figure 2.3. Schematic representation of the progressive filling of free-energy minima during a metadynamics simulation.

In this way MetaD can significantly accelerate the sampling of *rare events* such as protein-ligand binding and unbinding by pushing the system out from local free-energy minima. Moreover, MetaD presents other advantages: *i)* No a priori knowledge of the free-energy profile is required, *ii)* it explores low free-energy regions first, *iii)* it provides an unbiased estimate of the free energy surface.

Several parameters need to be set while preparing a simulation. In MetaD, one of the most crucial stage is the choice of an appropriate set of collective variables. The chosen CVs should discriminate between the initial and final state and describe all the relevant intermediates. They should also include all the "slow" motions of the system (that cannot be adequately sampled in the timescale of the simulation), otherwise the bias potential may not converge to the FES in a reasonable simulation time. Moreover, the CVs should be limited in number, because using many CVs means that a high-dimensional space needs to be explored and this may take a large amount of computational time.

Convergence is the major drawback in MetaD. In a single simulation, the bias potential doesn't converge to the free-energy, but oscillates around it. This fact can lead to two consequences: *i)* Risk to overfill the underlying FES and to push the system toward high energy states; *ii)* It is difficult to decide when to stop a simulation.

Generally, if MetaD is used to find the closest saddle point, it should be stopped as soon as the system exits from the minimum. Otherwise, if one is interested in reconstructing a FES, the simulation should be stopped when the motion of the CVs becomes diffusive in the region of interest.

2.2.2 Well-tempered metadynamics[24]

Classic MetaD, as discussed in the previous paragraph, has the well-known problem of convergence. The free-energy doesn't converge, but oscillates around an estimate. For this reason, while it tries to sample all the CV space, it can overfill the FES and push the system towards regions corresponding to unphysically high energy states or thermodynamically non-relevant configurations. To overcome these problems, well-tempered metadynamics (*wt*-MetaD) has been introduced. In *wt*-MetaD, the expression of the bias potential is modified by scaling down the Gaussian height *W* according to equation 3:

$$W = \omega \, \tau_G \, e^{-\frac{V_G(S,t)}{k_B \Delta T}} \tag{3}$$

Where ΔT is an input parameter with the dimension of a temperature.

In this way, the bias deposition rate is decreased over simulation time.

This means that, after initial filling, gaussian of smaller height are added in the same region of CV space. in this way, on top of the basin, when a lot of bias has been accumulated, the additional Gaussians have a smaller height, while when a new region is explored, and the bias is still small, the added Gaussian have large height. In this way, the risk of overfilling and sampling high energy regions is avoided.

Tuning ΔT in a *wt*-MetaD allows to regulate the extent of FES exploration. The ΔT→0 limit corresponds to standard MD, while for ΔT→∞ classic metadynamics is recovered.

---

[24] Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, 100, 020603.

2.3 Umbrella Sampling[25]

While MetaD is one of the most famous among methods working in condition of non-equilibrium, Umbrella sampling (US) is one notable equilibrium-CV-based enhanced sampling method.

Figure 2.4 shows how an US simulation is implemented. In US one or two coordinates are defined. The chosen coordinates are divided into small intervals often called "windows". Then, a bias MD is performed in each window. To do so, an artificial harmonic restraint (the so called "umbrella") is added to the system, with its minimum positioned on the central value of the window, to force the simulation to sample a particular interval of the coordinate of interest (i.e., the range of values around the center of the selected window).

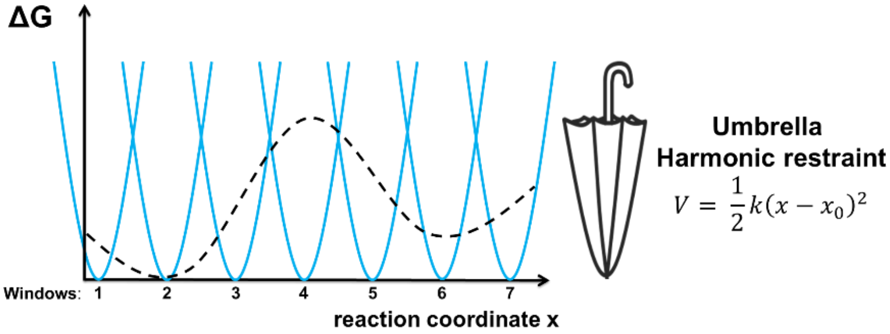


Figure 2.4. graphical representation of US setting. Umbrella harmonic restraints represented as blue lines. Free-energy profile as black dashed line.

During the simulation, the value of the coordinate is tracked and registered at every step of simulation in each window. To speed up the process, usually, the minimum of the biasing harmonic potential is repeatedly moved to different windows of the coordinate of interest, in order to carry out the subsequent simulations. In practice, the final configuration in a window is usually employed as starting configuration in the following one. The different windows must have some overlap: any particular value of the coordinate must be sampled to a significant extent in more than one window. In this way, all the reaction coordinate is adequately sampled, and histograms of the distribution of values along the coordinate (from all the values registered in each window) can be calculated. Finally, the effect of the biasing

---

[25] Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in monte carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

potential can be removed using the so-called "weighted histogram" or WHAM method,[26] to reconstruct a potential of mean force (PMF) profile.

Although US is one of the most accurate techniques for free-energy calculations, its practical application in drug design is mostly limited by its elevated computational cost. Indeed, many overlapping windows are needed to examine the free energy of an event, where each window must be prepared, sufficiently equilibrated, and finally sampled. For this reason, US is rarely applied to study protein-ligand binding/unbinding, while it's widely used to dynamically sample shorter events, such as chemical reactions, where the coordinates to be sampled are smaller compared to protein-ligand binding. This application of US to study chemical reactions can give useful advantages in drug discovery, such as elucidating the mechanism of action of covalent inhibitors, predicting drug metabolites, or studying enzyme catalysis.

2.4 QM/MM methods[27]

As mentioned above, US can be employed to calculate the free-energy profile of chemical reactions. To simulate a chemical reaction, modelling the breaking and formation of chemical bonds is required. MM force fields can deal with protein structure and dynamics, allowing to study protein conformational changes. The simplicity of MM force fields allows relatively long timescale (now up to millisecond) simulations of large systems, but cannot be applied to model the breaking and making of bonds (and electronic reorganization) typical of a chemical reaction. Indeed, MM describes molecules using simple functional forms (harmonic terms for bond stretching), treats electrostatic interactions using fixed atom-centered point charges and does not model changes in electronic polarization. Moreover, MM force field parameters are developed based on the properties of stable molecules, and so are not applicable to model transition states and intermediates of a chemical reaction.

---

[26] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free- Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

[27] (a) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700− 733. (b) Warshel, A. Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, 32, 425−443. (c) Lodola, A.; De Vivo, M. The Increasing Role of QM/MM in Drug Discovery. *Adv. Protein Chem. Struct. Biol.* **2012**, *87*, 337−362. (d) Lodola, A.; Mulholland, A. J. Computational enzymology. *Methods Mol. Biol.* **2013**, *924*, 67-89.

On the other hand, quantum-chemical methods calculate computationally the distribution of electrons in molecules, using the principles of quantum mechanics (QM), which treats the electrons as waves, with their distribution described mathematically by wavefunctions. QM can be used to calculate what happens in a chemical reaction, by calculating the changes in the distribution of electrons in the reaction as bonds are broken and formed. There are different QM methods available, many of them focused on finding the solution to the Schrodinger equation. The Schrodinger equation can be solved exactly only for molecular systems containing one electron. For more complex systems approximations are required. According to the approximations made, QM methods can be divided into three main types: *ab initio*, density functional theory (DFT) and semi-empirical approaches. QM calculations can give excellent results for reactions of small molecules, but require very large computational resources, limiting the size of the system that can be treated (up to hundreds of atoms for the more approximate methods). Biochemical systems (such as substrate-enzymes and protein-inhibitor complexes) are too large (thousands of atoms) to be described at any level of QM theory.

"Hybrid" (QM/MM) approaches that combine QM methods with MM allow more extensive calculations than is possible with purely QM techniques. In the QM/MM approach only a small part of the system, where the bond breaking/forming occurs, is treated quantum mechanically (along with protein side chains that may be involved in the reaction), while the rest of the protein (and surrounding solvent) is treated by MM (figure 2.5).

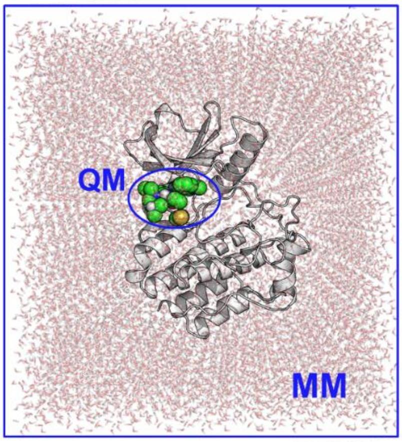


Figure 2.5. In a QM/MM calculation the system is split into two regions: a small region encapsulating the reaction (shown here as sphere) is modelled with a QM method, while the rest of the protein (white cartoon) and surrounding solvent (thin sticks) are modelled using MM.

In typical QM/MM calculations, the energy of the whole system (E_{add}^{total}), can be written as the sum of four contributions:

$$E_{add}^{total} = E_{QMregion}^{QM(MM)} + \mathrm{E}_{MMregion}^{MM} + E^{QM/MM} + E^{boundary} \tag{3}$$

The first term, the energy of the QM region in the presence of the MM atoms ($E_{QMregion}^{QM(MM)}$), is given by QM methods. In *ab initio* methods, the MM atomic partial charges are included directly through one-electron integrals, while in semi-empirical approaches, the MM point charges are treated as "cores" (core electrons are treated together with nucleus as an atomic core and only valence electrons are treated directly). The second term of equation 3, the energy of the atoms in the MM region ($E_{MMregion}^{MM}$), is calculated by a MM force field.
The QM/MM interaction energy ($E^{QM/MM}$), consists of the non-electrostatic interactions between the QM and MM regions, thus including van der Waals and bonded interaction. QM/MM bonded interactions involving at least one MM atom are treated with MM force field (bond stretching, angle bending, torsion, etc.). QM/MM van der Waals interactions (representing dispersion attraction and exchange repulsion interactions between QM and MM atoms) are usually calculated by MM (e.g. through Lennard-Jones terms). To do so, MM van der Waals parameters are assigned to each QM atom, and for convenience, they are typically the same as those for equivalent MM atoms in the force field. Finally, the boundary energy, $E_{Boundary}$, is added because the simulation system can only include a finite number of atoms, so terms to reproduce the effects of the surroundings must be included.

QM/MM methods simultaneously exploit the benefits of both the QM and MM methods, allowing to obtain a compromise between the speed and simplicity of MM methods, and the accuracy and computational costs of QM approaches. Therefore, QM/MM approach has the advantage that chemical reactions can be investigated at QM level within large system with a reasonable computation time. Indeed, Hybrid QM/MM potentials have been widely applied to enzyme catalysis and reported to give fair descriptions of reaction geometries and energetics.

Chapter 3

Identification of binding modes for small molecules

3.1 The docking problem

As anticipated in the first chapter, studying the recognition between a small molecule and the biological target is an important step in structure-based drug design. In this context, the simplest method that has been developed to study drug-target complementarity is molecular docking. Molecular docking searches for favorable interactions between one ligand within the binding site of the target receptor, providing a set of possible binding poses. Typical docking programs commonly employ a two-phase approach: *i) Posing:* several configurations of the ligand are generated within the binding site of the receptor by means of different searching techniques; *ii) Scoring:* the stability of the ligand-protein binding modes generated at the previous step is evaluated, then stable configurations identified are ranked according to the result of a scoring function.[28] This approach allows to obtain a collection of putative binding modes in short time, rather to provide a univocal solution. Even though docking programs are usually able to find the binding pose which at best reproduce the experimental configuration, not always it is located among the top scored ones.[29] For this reason, it is sometimes difficult to choose the "right" binding pose (the one that returns the best ligand-target affinity) among the set of possible solutions proposed by docking programs.[30] This behavior is often noted as *docking problem*.

The docking problem is mainly related to approximations encoded in docking programs to achieve a fast calculation: *i)* limited sampling of the conformational space, which neglects protein flexibility; *ii)* partial or total neglection of solvent effects *iii)* approximate estimation of the binding free energy. Therefore, scoring functions have a limited reliability, especially when entropic effects (solvation/desolvation) play a significant role in stabilizing the protein-ligand complex. In this context, MD-based *dynamic docking*, may overcome the docking problem.[31] Indeed, MD takes into account factors that are usually approximated or neglected in standard static docking protocols for drug design: *i)* the full flexibility of the protein-ligand complex, which is therefore free to move and evolve over the simulation time; *ii)* the full solvation of the drug-target system, which is critical to

---

[28] Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151-166.

[29] Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759-4767.

[30] Bottegoni, G.; Cavalli, A.; Recanatini, M. A comparative study on the application of hierarchical-agglomerative clustering approaches to organize outputs of reiterated docking runs. *J Chem Inf Model,* **2006**, *46*, 58 – 65.

[31] De Vivo, M.; Cavalli, A. Recent Advances in Dynamic Docking for Drug Discovery. *WIREs Comput Mol Sci*, **2017**, e1320.

assess for water-mediated interactions, allowing also to evaluate solvation and desolvation phenomena during ligand binding and unbinding; *iii)* the system temperature, which is usually maintained at 300 K; It is important to point out that dynamic docking should consider several different trajectories to acquire enough statistics and provide reliable predictions. As mentioned in chapter 2, statistics may not be sufficient in MD because of high energy barriers that limit sampling. In this case, enhanced sampling methods have been applied to accelerate the formation and/or disruption of the drug-target complex. Among them, metadynamics has proved to be a valuable post-docking tool to be used in place of scoring functions.[32] Metadynamics allows to reconstruct the FES of protein-ligand binding and characterize free-energy minima, where the global minimum corresponds to the most favorable configuration of protein-ligand complex and other higher free-energy minima may represent alternative binding modes.[33,34]

In this chapter, metadynamics is applied to reconstruct the FES of binding for Epha2-EphrinA1 inhibitors, with the aim to recover the correct binding mode of the compounds among the ones proposed by traditional static docking and then drive the synthesis of new compounds according to the proposed binding pose.

---

[32] Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. Exploring Complex Protein-Ligand Recognition Mechanisms with Coarse Metadynamics. *J Phys Chem B* **2009**, *113* (14), 4807–4816.

[33] Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *J. Am. Chem. Soc.* **2005**, 127, 2600− 2607.

[34] Limongelli, V.; Bonomi, M.; Marinelli, L.; Gervasio, F. L.; Cavalli, A.; Novellino, E.; Parrinello, M. Molecular Basis of Cyclooxygenase Enzymes (COXs) Selective Inhibition. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (12), 5411–5416.

3.2 EphA2-EphrinA1 inhibitors

The EphA2 receptor is a member of the Erythropoietin-producing hepatocellular carcinoma (Eph) receptors,[35] a family of tyrosine kinase receptors that are activated by specific membrane-anchored peptides, called ephrins.[36] The EphA2 receptor has been found overexpressed in many cancer types, including melanoma, prostate, breast, colon, lung, pancreatic, and lung cancers.[37] It has been reported that both EphA2 and its ephrin-A1 ligand are involved in carcinogenesis,[38] promoting the self-renewal of tumor propagating cells, the acquisition of a migratory phenotype, and the formation of new blood vessels.[39] Since EphA2 is expressed in the adult in tumor cells and vasculature but not in normal vasculature,[40] this receptor emerges as a promising target for the development of new antiangiogenic therapies.[41] Two main strategies can be used to inhibit the tumorigenic activity of the EphA2 receptor. The first one is based on the use of classical ATP-mimicking agents directed at the intracellular kinase domain of Eph receptor,[42] while the second one is based on molecular agents targeting the extracellular ligand binding domain of EphA2 (Figure 3.1), thus preventing its interaction with ephrin ligands.[43,44]

---

[35] Himanen, J. P.; Saha, N.; Nikolov, D. B. Cell-Cell Signaling via Eph Receptors and Ephrins. *Curr. Opin. Cell Biol.* **2007**, *19*, 534−542.

[36] Pasquale, E. B. Eph-ephrin Bidirectional Signaling in Physiology and Disease. *Cell* **2008**, *133*, 38−52.

[37] Pasquale, E. B. Eph Receptors and Ephrins in Cancer: Bidirectional Signalling and Beyond. *Nat. Rev. Cancer* **2010**, *10*, 165−180.

[38] Wykosky, J.; Debinski, W. The EphA2 receptor and ephrinA1 ligand in solid tumors: function and therapeutic targeting. *Mol. Cancer Res.* **2008**, *6*, 1795–1806.

[39] Boyd, A. W.; Bartlett, P. F.; Lackmann, M. Therapeutic targeting of EPH receptors and their ligands. *Nat. Rev. Drug Discov.* **2014**, *13*, 39–62.

[40] Ogawa, K.; Pasqualini, R.; Lindberg, R. A.; Kain, R.; Freeman, A. L.; Pasquale, E. B. The ephrin-A1 ligand and its receptor, EphA2, are expressed during tumor neovascularization. *Oncogene*, **2000**;19, 6043–6052.

[41] Tandon, M.; Vemula, S. V.; Mittal, S. K. Emerging strategies for EphA2 receptor targeting for cancer therapeutics. *Expert Opin. Ther. Targets* **2011**, *15*, 31–51.

[42] Unzue, A.; Lafleur, K.; Zhao, H.; Zhou, T.; Dong, J.; Kolb, P.; Liebl, J.; Zahler, S.; Caflisch, A.; Nevado, C. Three stories on Eph kinase inhibitors: from in silico discovery to in vivo validation. *Eur. J. Med. Chem.* **2016**, *112*, 347−366.

[43] Riedl, S. J.; Pasquale, E. B. Targeting the Eph System with Peptides and Peptide Conjugates. *Curr. Drug Targets.* **2015**, *16*, 1031-1047.

[44] Lamminmaki U, Nikolov D, Himanen J. Eph Receptors as Drug Targets: Single-Chain Antibodies and Beyond. *Curr. Drug Targets.* **2015**, *16*, 1021-1030.

The interaction between EphA2 and ephrinA1 ligand has been recently resolved by X-ray crystallography.[45] The interaction between these two proteins is primarily mediated by the amino-terminal ligand binding domain (LBD) of EphA2, which forms a large hydrophobic cavity that accommodates the protruding G-H loop from ephrinA1 (Figure 3.1).[46] The binding interface is dominated by van der Waals contacts between two predominantly hydrophobic surfaces and is reinforced by two salt bridges, the first one engaging EphA2 Arg103 and ephrinA1 Glu119 and the second one involving EphA2 Arg159 and ephrinA1 Asp86 (Figure 3.1). In the last years, small molecules able to inhibit the activation of the EphA2 receptor by disrupting the EphA2-ephrinA1 protein-protein interaction have been reported in the literature.[47,48]

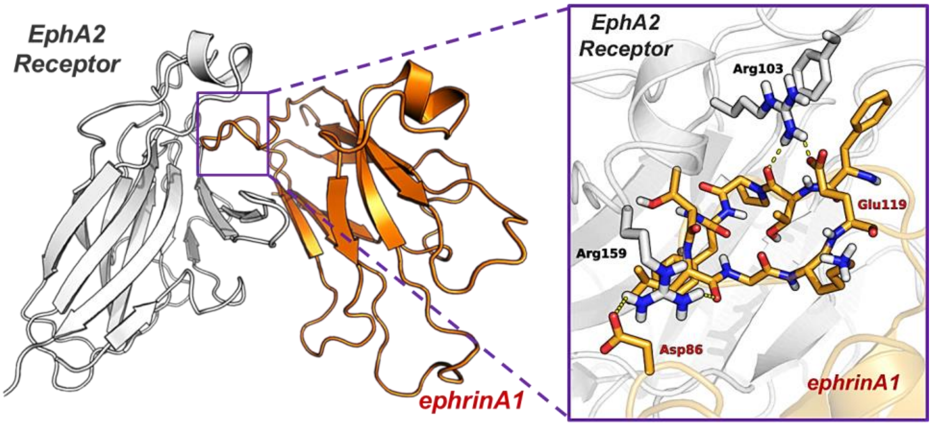


Figure 3.1. EphA2 extracellular ligand binding domain (white cartoon and white carbons) in complex with ephrinA1 ligand (orange cartoons and carbons). Salt bridges and h-bond represented as yellow dashed lines.

---

[45] Himanen, J. P.; Goldgur, Y.; Miao, H.; Myshkin, E.; Guo, H.; Buck, M.; Nguyen, M.; Rajashankar, K. R.; Wang, B.; Nikolov, D. B. Ligand Recognition by A-class Eph Receptors: Crystal Structures of the EphA2 Ligand-binding Domain and the EphA2/ephrin-A1 Complex. *EMBO reports* **2009**, *10* (7), 722–728.

[46] Himanen, J. P. Ectodomain Structures of Eph Receptors. *Semin. Cell Dev. Biol.* **2012**, *23* (1), 35–42.

[47] Tognolini, M.; Hassan-Mohamed, I.; Giorgio, C.; Zanotti, I.; Lodola, A. Therapeutic perspectives of Eph-ephrin system modulation. *Drug Discov. Today*. **2014,** *19*, 661-669.

[48] Barquilla, A.; Pasquale, E. B. Eph receptors and ephrins: therapeutic opportunities. *Annu Rev. Pharmacol. Toxicol.* **2015**, *55*, 465-487.

Among them, lithocholic acid[49] (LCA, Figure 3.2) and its analogues were identified as competitive and reversible antagonists of the EphA2 receptor, active in prostate cancer cells (PC3) at non-cytotoxic concentration.[50] Starting from LCA, a set of amino acid conjugates able to interfere with the EphA2–ephrin-A1 interaction with potency in the low micromolar range was recently synthesized.[51] Among them, the *N*-(3α-hydroxy-5β-cholan-24-oyl)-L-β-homotryptophan (UniPR129, Figure 3.2) emerged as a promising EphA2 antagonist (i.e. K*i* = 380 nM),[52] able to suppress angiogenesis in human umbilical vein endothelial cells (HUVECs) through selective inhibition of EphA2 activity. Despite its promising pharmacodynamics profile, UniPR129 presents poor physicochemical properties, which have limited the *in vivo* employment of this compound.

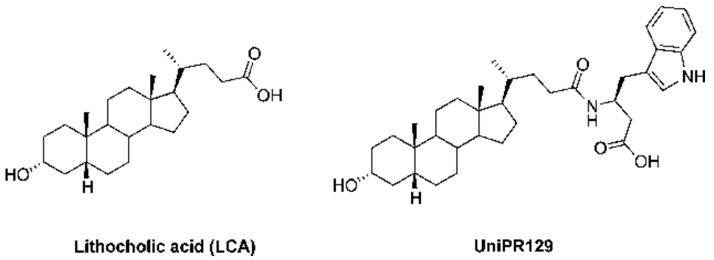


Figure 3.2 Structures of Reference EphA2 Receptor Antagonists

To improve the physicochemical properties of the class without hampering the affinity for the EphA2 receptor, the identification of the productive binding mode of EphA2 inhibitors becomes essential. Unfortunately, the X-ray structure of the EphA2-LCA complex is not available yet and the low solubility of UniPR129 hampers the possibility to use X-ray and NMR spectroscopy to recover the structure of EphA2-UniPR129 complex. In this context, molecular simulations could help to

---

[49] Giorgio, C.; Hassan Mohamed, I.; Flammini, L.; Barocelli, E.; Incerti, M.; Lodola, A.; Tognolini, M. Lithocholic acid Is an Eph-Ephrin ligand interfering with Eph-kinase activation. *PloS One* **2011**, *6*, e18128.

[50] Tognolini, M.; Incerti, M.; Hassan-Mohamed, I.; Giorgio, C.; Russo, S.; Bruni, R.; Lelli, B.; Bracci, L.; Noberini, R.; Pasquale, E. B.; Barocelli, E.; Vicini, P.; Mor, M.; Lodola, A. Structure-activity relationships and mechanism of action of Eph-Ephrin antagonists: interaction of cholanic acid with the EphA2 receptor. *ChemMedChem* **2012**, *7*, 1071–1083.

[51] Incerti, M.; Tognolini, M.; Russo, S.; Pala, D.; Giorgio, C.; Hassan-Mohamed, I.; Noberini, R.; Pasquale, E. B.; Vicini, P.; Piersanti, S.; Rivara, S.; Barocelli, E.; Mor, M.; Lodola, A. Amino acid conjugates of lithocholic acid as antagonists of the EphA2 receptor. *J. Med. Chem.* **2013**, *56*, 2936–2947.

[52] Hassan-Mohamed, I.; Giorgio, C.; Incerti, M.; Russo, S.; Pala, D.; Pasquale, E. B.; Vicini, P.; Barocelli, E.; Rivara, S.; Mor, M.; Lodola, A.; Tognolini, M. UniPR129 is a competitive small molecule Eph-Ephrin antagonist blocking in vitro angiogenesis at low micromolar concentrations. *Br. J. Pharmacol.* **2014**, *171*, 5195-5208.

clarify the binding mode of these ligands to EphA2 receptor providing information useful for the design of new derivatives. In particular, combining molecular simulations, together with structure-activity relationship (SAR) analysis[53,54] could provide an alternative method to propose a model of interaction useful for prospective drug design. As mentioned in the previous paragraph, the inclusion of both conformational flexibility and solvent reorganization are often fundamental to get a reasonable model of protein-ligand binding. For this reason, in this chapter, *dynamic docking* by means of MD and MetaD simulations was performed for both LCA and UniPR129, starting from the top-ranked solutions obtained with Glide software. FES of EphA2-LCA and EphA2-UniPR129 were reconstructed. The analysis of the FESs of binding allowed to identify binding modes consistent with SAR data for both compounds. Moreover, exploiting an alternative binding mode, corresponding to a second free-energy minimum in the FES of EphA-LCA binding, allowed to synthetize a new compound with increased potency compared to LCA. The retrieved binding mode of UniPR129 allowed to drive the synthesis of novel EphA2 receptor antagonists, equally or slightly more potent than UniPR129, featured by better physicochemical properties or improved metabolic stability.

---

[53] Kuhnert, M.; Köster, H.; Bartholomäus, R.; Park, A. Y.; Shahim, A.; Heine, A.; Steuber, H.; Klebe, G.; Diederich, W. E. Tracing binding modes in hit-to-lead optimization: chameleon-like poses of aspartic protease inhibitors. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 2849-2853.

[54] Castelli, R.; Tognolini, M.; Vacondio, F.; Incerti, M.; Pala, D.; Callegari, D.; Bertoni, S.; Giorgio, C.; Hassan-Mohamed, I.; Zanotti, I.; Bugatti, A.; Rusnati, M.; Festuccia, C.; Rivara, S.; Barocelli, E.; Mor, M.; Lodola, A. Δ^5-Cholenoyl-amino acids as selective and orally available antagonists of the Eph-ephrin system. *Eur J Med Chem.* **2015**, *103*, 312-324.

3.3 Results and discussion

As introduced in the previous paragraphs, even if docking protocols can provide reliable binding geometries, their inability to account for relevant aspects, such as conformational protein flexibility and solvent reorganization, could limit the identification of all the accessible binding modes. For this reason, well-tempered metadynamics (*wt*-MetaD) simulations were applied to characterize the FES of binding of EphA2-LCA and EphA2-UniPR129 to evaluate the existence and probability of binding geometries alternative to those originally proposed by docking studies. The retrieved binding poses were then applied to drive the synthesis of new compounds.

3.3.1 Identification of LCA "correct" binding mode

LCA was initially docked in the ligand-binding domain of EphA2.[13] Docking simulations performed with Glide 6.3 yielded two equally scored families of docking poses of LCA within the ligand binding domain of EphA2 receptor. The first docking pose (binding mode A, docking score: -5.4 kcal/mol, figure 3.3A) is similar to the one already reported for LCA.[50] In this first docking pose, LCA occupies the same space as the ephrinA1 G–H loop, inserting its cyclopenta[a]perhydrophenanthrene scaffold into the hydrophobic EphA2 receptor channel. The pentanoic acid fragment, emerging from position 17 of the LCA core, forms a salt bridge with Arg103, mimicking the interaction with Glu119 from ephrinA1 and the 3α-hydroxy group of LCA weakly interacts with the backbone carbonyl group of Phe156. The second binding mode obtained with docking (binding mode B, docking score: -4.7 kcal/mol, Figure 3.3B) has a complete reversed accommodation of LCA within the hydrophobic EphA2 ligand binding domain. The carboxylate group of LCA forms a salt bridge with Arg159 (which is usually engaged in a salt bridge with Asp 86 of ephrinA1) and a h-bond with the NH backbone group of Asn60, while its 3α-hydroxyl group interacts with the side chain of Ser68.

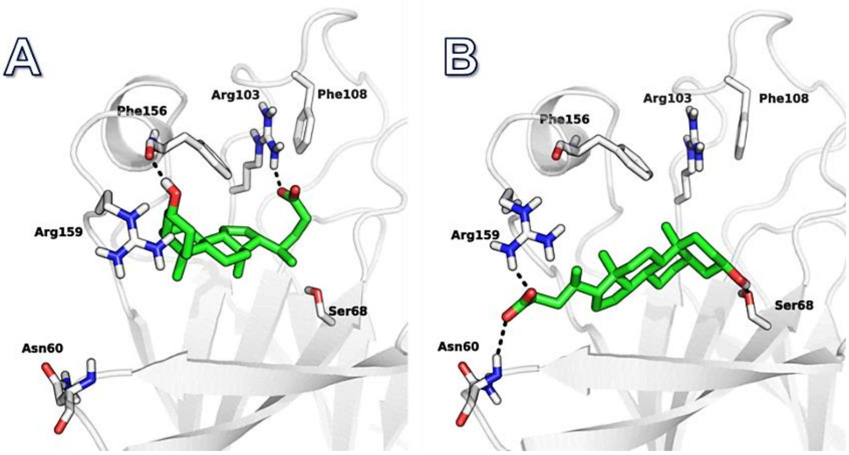
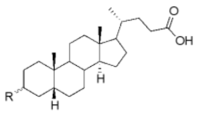


Figure 3.3. Binding modes of LCA (green carbon atoms) within the EphA2 receptor (white ribbons and carbon atoms) as obtain with docking simulations.

Unfortunately, both binding modes proposed by traditional docking were not consistent with SAR observed for decoration at the C3 position of the LCA (Table 3.1). Indeed, the interactions undertaken by 3α-hydroxy group of LCA in both binding poses cannot explain why cholanic acid (the LCA analogue lacking the 3α-hydroxyl group) is more potent than LCA itself.

Table 3.1. Structure–activity relationship data for LCA derivatives at position 3 obtained from EphA2-ephrinA1 displacement experiments.

	
R	**K_i (μM)[a]**
-αOH (LCA)	49 ± 3.0
-βOH	25 ± 4.0
-H	5.1 ± 1.4

[a] Values are the mean ±SE from a minimum of three independent experiments.

Molecular dynamics simulations were then performed starting from both binding modes to briefly assess their stability. For each pose, a short 30ns-long MD was simulated. In both cases, the proposed binding mode resulted quite stable, with the LCA hydrophobic core stably oscillating around its position until the end of the simulation (Figure 3.4).

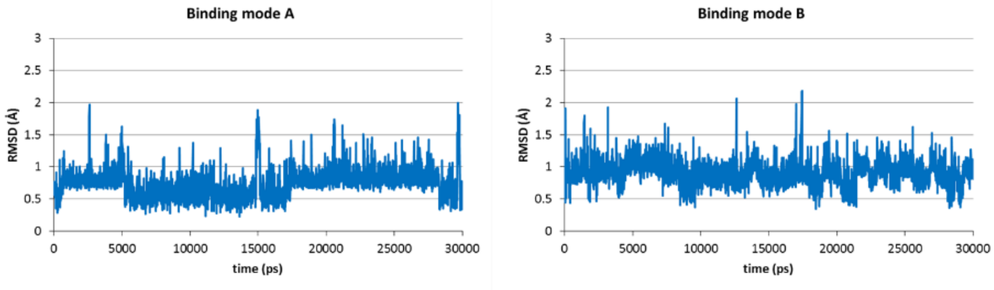


Figure 3.4. RMSD time series of LCA heavy atoms during a 30ns-long MD simulation, starting from binding mode A (left) or binding mode B (right).

wt-MetaD simulations were then performed to reconstruct the free-energy surface (FES) of binding, in order to exploit free-energy minima and characterize the most probable binding configuration for LCA. To this aim, some preliminary metadynamics simulations were run to identify CVs able to describe the full process of ligand binding/unbinding and discriminate between different binding modes visited by the ligand (see Appendix). Two CVs were chosen to drive LCA out of EphA2 binding site and to explore different binding modes. These two variables were **CV1**, the distance between the center of mass (COM) of the EphA2 ligand binding domain and the COM of the ligand steroid moiety, and **CV2**, the angle vector taken between the same protein reference point and the major inertia axis of the steroid moiety (Figure 3.5). These two CVs have already shown good results when used to perform flexible docking with metadynamics.[33]

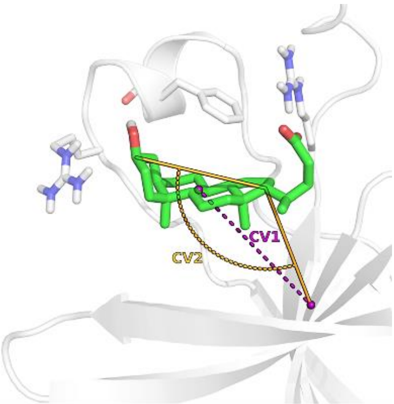


Figure 3.5. Graphical representations of the CVs employed for the MetaD simulations. Center of mass points are depicted in purple, distance **CV1** is represented as purple dashed lines, angle **CV2** is represented with yellow lines.

Distance **CV1** could drive LCA out from EphA2 binding site. Moreover, recrossing events between bound (CV1<15Å) and unbound (CV1>15Å) states were observed (Figure 3.6), also because of the use of a repulsive wall along **CV1**, which pushes the ligand back to the binding site once it reached the unbound state (see Computational Protocol for details).

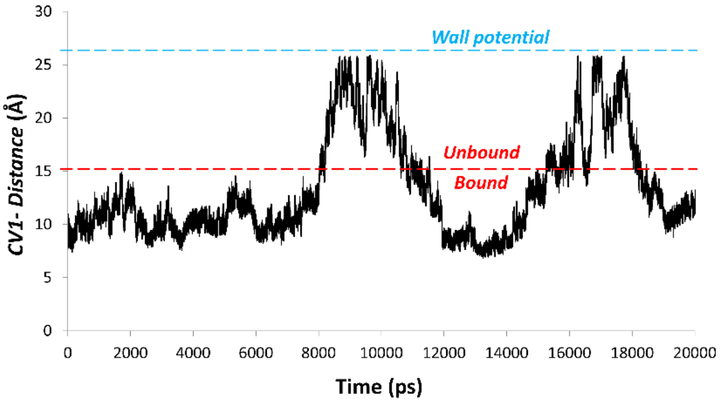


Figure 3.6. time series of the distance collective variable **CV1** for a representative metadynamics simulation of EphA2–LCA binding. Multiple recrossing events can be detected.

Angle **CV2** was able to discriminate between the two binding modes proposed by docking, as elucidated by the different ranges of values assumed by **CV2** during the 30 ns-long MD simulations of both binding mode A and B (Figure 3.7). Indeed, in the MD simulation of binding mode A, angle **CV2** stably oscillates in the range of values 80°-140°, while in the case of binding mode B, **CV2** fluctuates in the range 20°-60°.

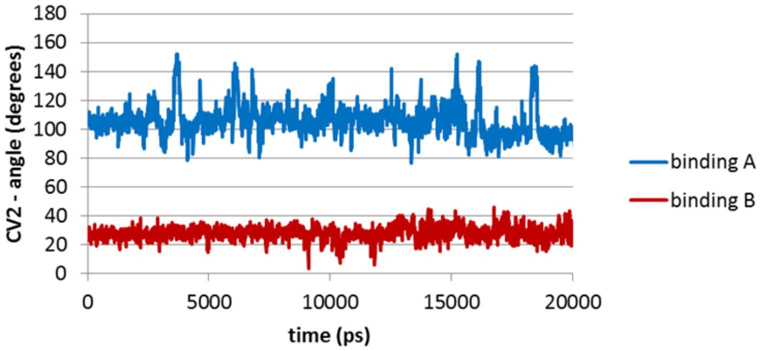


Figure 3.7. Angle **CV2** time series during a 30ns-long MD simulation, starting from binding mode A (cyan line) or binding mode B (red line).

To evaluate reproducibility, the FES of EphA2-LCA binding process was calculated for four independent replicas of *wt*-MetaD. Calculated FESs resulted reproducible when the starting velocities assigned to system atoms were randomly changed (see Computational Protocol). Each simulation was considered ended when the ligand visited all the identified binding minima at least once. The final FES of EphA2-LCA binding process was finally calculated averaging those obtained by the four independent simulations, for both binding mode A and B. The averaged FESs resulted reproducible also when the starting point of simulation was changed (Figure 3.8).

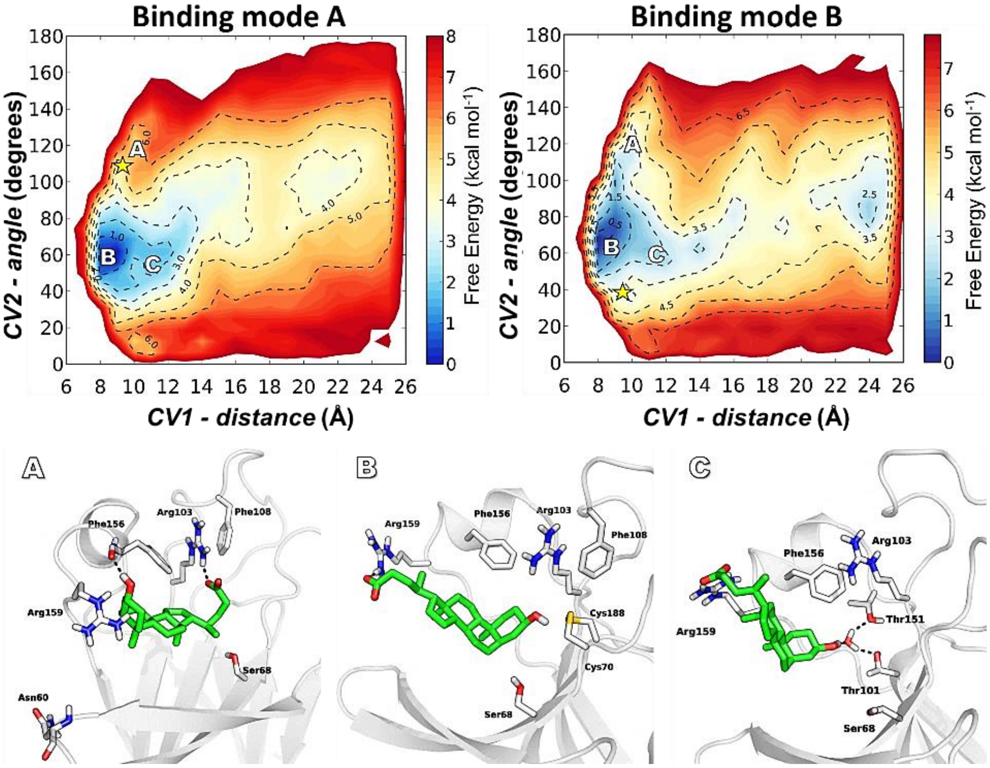


Figure 3.8. Averaged free-energy surfaces in the **CV1** and **CV2** space of EphA2-LCA unbinding, calculated from four independent simulations starting from binding mode A (upper left panel) and binding mode B (upper right panel). Starting configurations are indicated with a yellow star on the FESs. Relevant geometries identified along the simulations are also reported (bottom panel). EphA2 is depicted in white cartoons and carbons atoms, while LCA carbon atoms are coloured in green.

Figure 3.8 reports the FESs of binding for the LCA-EphA2 system in the **CV1** and **CV2** space, starting from both binding modes, together with a representation of relevant geometries. Visual inspection of the surfaces identifies three free-energy minima (**A**-**C**), in which LCA is accommodated within EphA2 adopting different binding orientations, and a rather wide transition state (TS) region separating minima **A**-**C** from the basin where LCA lives in its unbound state (**CV1** > 20Å). The free-energy area corresponding to binding mode A reported in Figure 3.3 (basin **A, CV1** = 10 ± 0.5 Å, and **CV2** = 120 ± 5 °) is significantly less stable than the deepest minima of the FES (basins **B** and **C**) by nearly 5.5 kcal/mol.

In free-energy minimum **B** (**CV1** = 8 ± 0.5 Å; **CV2** = 60 ± 10 °), the carboxylate group of LCA forms a salt bridge with Arg159, similarly to binding mode B proposed by traditional docking, but its 3α-hydroxyl does not undertake polar inarctions and lays in proximity of a disulfur bridge formed by Cys70 and Cys188. In contrast with the binding modes retrieved by docking, the geometry assumed in basin B of the FES is consistent with SAR data for position 3 of the steroid.[6] In fact, the lack of polar interactions between the 3α-hydroxyl of LCA and the protein (the closest polar groups of Arg103 and Ser68 are ≈ 6 Å away) could explain why cholanic acid is more potent than LCA itself in a EphA2-ephrin-A1 displacement assay.

Basin **B** is in strict contact with another minimum (basin **C, CV1** = 11 ± 0.5 Å; **CV2** = 55 ± 5 °), in which the 3α-hydroxyl of LCA forms water-mediated H-bonds with Thr101 and Thr151 located at the bottom of EphA2 binding site, while the carboxylate forms a salt bridge with Arg159, as seen in minimum **B**. The area of FES in the interval 17-19 Å for **CV1** and 50-110° for **CV2** represents the TS region, which corresponds to a set of high energy configurations, where the steroidal nucleus is already in the solvent but the carboxylate group of LCA still interacts with Arg159. Once overcome the TS region (**CV1** > 20 Å), LCA reaches the bulk solvent (basin ≈ 3.5 kcal mol^{-1} higher than **B**) where both receptor and ligand are in their unbound state.

Overall, this computational investigation indicates that LCA interacts with EphA2 assuming preferably a completely different binding mode compared to the one previously reported in the literature. This finding is consistent with the SAR reported for LCA.

3.3.2 Exploiting alternative binding modes

FES can suggest the existence of free-energy minima corresponding to binding geometries distinct from those observed in X-ray structures or NMR experiments,[9,10] that could be exploited to chemically modify an active compound to improve its potency.

In particular, the binding mode assumed by LCA in basin **C** of Figure 3.8 suggests that the introduction of bulky polar substituents at the 3α position, able to directly interact with Thr101 and Th151, would lead to more potent EphA2 antagonists. A 3α-carbamoyloxy-5β-cholan-24-oic acid derivative was therefore designed (UniPR413, Fig. 3.9) and its FES of binding with EphA2 was reconstructed applying the same protocol used for LCA.

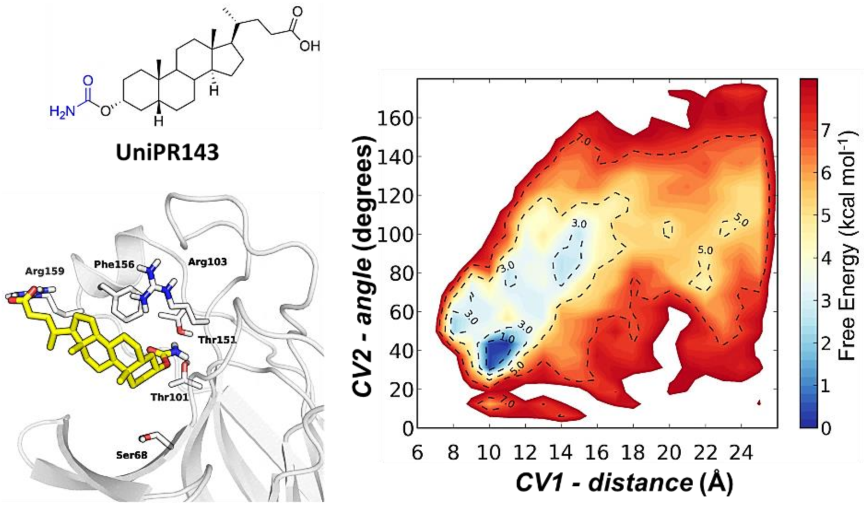


Figure 3.9. structure of UniPR143 (upper left panel); representative geometry of the free-energy minimum UniPR143 binding to EphA2 (bottom left panel, yellow carbon atoms), FES of EphA2-UniPR143 binding the **CV1** and **CV2** space (right panel).

The resulting FES (Figure 3.9) shows the presence of a deep and well-defined free-energy basin (**CV1** = 11 ± 0.5 Å; **CV2** = 35 ± 5°), in which the 3α-carbamoyloxy of UniPR413 forms H-bonds with the two threonine residues (Figure 3.9, right panel). Calculations showed that the ΔG_{bind} of UniPR413 (obtained subtracting the free-energy of the basin corresponding to the unbound state from that of minimum **B)** is ≈ 1.5 kcal mol^{-1} lower than that found for LCA, thus indicating that UniPR413 should be a better EphA2 binder than LCA.

Encouraged by these results, UniPR413 was synthetized and tested for its ability to disrupt EphA2-ephrinA1 binding. The displacement assay (Figure 3.10) indicates that UniPR413 (pIC_{50} = 4.7 ± 0.1) resulted 4-fold more potent than its cognate ligand

LCA (pIC_{50} = 4.1 ± 0.1) in preventing ephrinA1 binding to EphA2, indicating that the described computational approach has a reasonable predictive power. Notably, a new compound (UniPR413) with inhibitory potency comparable to that of amino acid conjugates of LCA was obtained, following a chemical intuition that the original docking model would have discouraged.

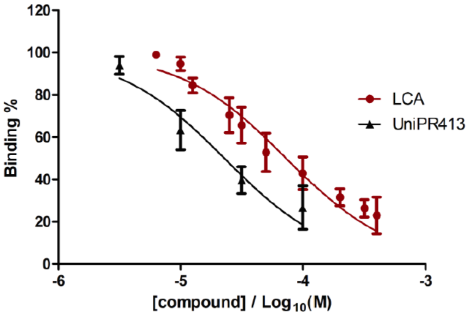


Figure 3.10. Dose–response curves for LCA (red curve) and UniPR413 (black curve). Both compounds dose-dependently prevent binding of ephrin-A1 to immobilized EphA2-Fc receptor. Experiments were performed in triplicate. Standard errors are also reported.

3.3.3 Identification of UniPR129 binding mode

A similar approach to that used to study EphA2-LCA interaction was applied to study the binding mode of the potent inhibitor UniPR129.

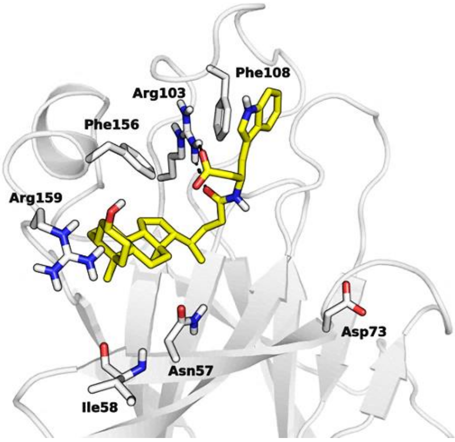


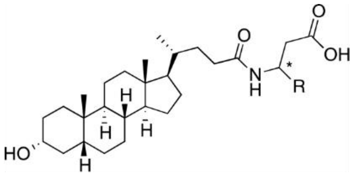
Figure 3.11. Binding mode of UniPR129 (yellow carbon atoms) within EphA2 ligand binding domain (white cartoons and white carbon atoms) as reported in reference 52.

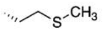
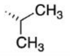
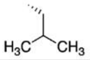
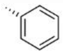
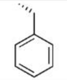
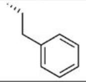
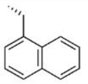
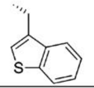
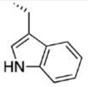
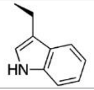
In the binding pose reported in the literature (Figure 3.11),[52] UniPR129 interacts with EphA2 by inserting its steroidal scaffold in the hydrophobic channel of EhA2 ligand binding domain. The carboxylate group forms a salt bridge with Arg103 and the indole ring, inserted in an accessory hydrophobic pocket, tightly interacts with Phe108, a conserved residue responsible for the recognition of one of the two aromatic residues (Phe111) of the ϕ-x-x-ϕ binding motif of ephrin ligands.[55] The reported binding mode, however, only partially explained the SAR data of amino acid derivatives of LCA.

---

[55] Lema Tome, C. M.; Palma, E.; Ferluga, S.; Lowther, W. T.; Hantgan, R.; Wykosky, J.; Debinski, W. Structural and functional characterization of monomeric EphrinA1 binding site to EphA2 receptor. *J. Biol. Chem*. **2012**, *287*, 14012−14022.

Table 3.2. IC_{50} Values of β-Substituted β-Alanine Conjugates of LCA Obtained from EphA2−ephrin-A1 Displacement Experiments (as reported in Incerti et al. *J. Med. Chem.* **2017**, *60*, 787–796.)



Cpd.	R-	IC_{50} (μM)[a,b]	PSA ($Å^2$)[c]
1	H	29 [23-36]	111
2		83 [50-138]	104
3		10 [6.6-16]	96
4		inactive[d]	105
5		14 [10-20]	104
6		18 [12-26]	104
7		inactive[d]	98
8		3.9 [3.0-5.0]	104
9		1.8 [1.5-2.1]	102
10 **UniPR129**		0.91 [0.80-1.1]	120
11		26 [18-37]	123

[a] Values are the mean from at least three independent experiments. [b] Numbers in brackets represent the 95% confidence interval for IC_{50}. [c] Polar surface area calculated with QikProp obtained from minimized 3D-structures of the listed compounds. [d] No signal detected up to 100 μM.

SAR Analysis of Amino Acid Conjugates of LCA

Table 3.2 reports the SAR of decoration at the β position of β-Alanine conjugates of LCA obtained from EphA2−ephrinA1 displacement experiments. Aliphatic side chains, either linear (compound **2)** or branched (compounds **3** and **4**), when inserted at the β position gave unsatisfactory results, considering that only β-homovaline derivative **3** resulted slightly more potent than **1** at preventing ephrin-A1 binding to EphA2. Similarly, the introduction of a phenyl (**5**), a benzyl **(6**), or a phenylethyl (**7**) substituent gave poor results. A significant improvement in the inhibitory potency was instead obtained introducing larger aryl side chains into the structure of **1**. Both in the case of the α-naphthylmethyl derivative **8** and the benzo[*b*]thiophen-3-ylmethyl one **9**, single digit micromolar IC_{50} values were observed. This improvement in the inhibitory potency could be ascribed to the ability of the napthyl and benzothiophene nucleus to undertake productive steric or electrostatic interactions with an accessory pocket of the EphA2 binding site. The best improvement in the activity was obtained with the indol-3-ylmethyl derivative **10** (UniPR129), having inhibitory potency in the sub-micromolar range (IC_{50} = 0.9 μM). The higher potency of UniPR129 compared to **8** and **9** could be attributed to the ability of the indole NH group to form a key H-bond within the EphA2 receptor. Compound **11** may support this hypothesis. Compound **11** is the diastereoisomer of UniPR129, in which the chirality of the β carbon was inverted. Compound **11**, likely projects its indol-3-ylmethyl chain in another region of EphA2 binding site thus failing to undertake a productive H-bond with the receptor, and resulted indeed 25-fold less potent than UniPR129.

However, the reported docking mode failed to satisfactorily explain this SAR data. In particular, in the reported pose was not possible to identify a specific interaction of the indole ring of UniPR129 that could explain the gain in the inhibitory potency of this compound when compared to analogues **8**, **9** and **11**.

As the inclusion of both conformational flexibility and solvent reorganization are often fundamental to get a reasonable guess of a protein-ligand binding scheme,[56,57] MD simulation was performed starting from the above described docking pose of UniPR129. A 30 ns-long simulation showed that the docked conformation was not a stable configuration of UniPR129 within EphA2, as revealed by its high root-mean-square deviation (RMSD) registered after only 5 ns of simulation (Figure 3.12, left panel). On the other hand, the same MD identified another accommodation of UniPR129**,** maintained for nearly 25 ns of simulation.

---

[56] Zhao, H.; Caflisch, A. Molecular Dynamics in Drug Design. Eur. *J. Med. Chem.* **2015**, *91*, 4−14

[57] Spyrakis, F.; Bidon Chanal, A.; Barril, X.; Luque, F. J. Protein Flexibility and Ligand Recognition: Challenges for Molecular Modeling. *Curr. Top. Med. Chem.* **2011**, *11*, 192−210

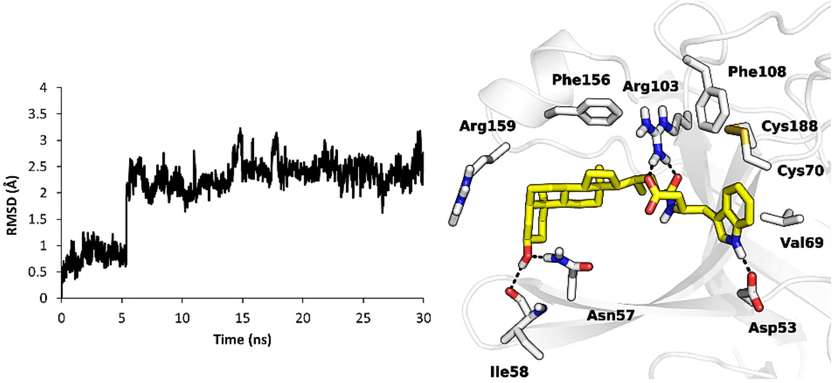


Figure 3.12. RMSD time series of UniPR129 calculated from MD simulation trajectory using the docked conformation of the ligand as reference structure (left). Most stable binding conformation of the EphA2-UniPR129 complex observed in MD simulation (right).

In this new model of interaction (Figure 3.12, right panel), the indole ring of UniPR129 is inserted in a solvent-exposed sub-pocket of EphA2 delimited by delimited by Phe108, Cys70, Val69, and Asp53. The proximity with Phe108, Cys70 and Val69 could explain the higher activity of compounds presenting side chains of large size versus those of small size (i.e. compounds **8-10** *vs* compounds **2-7**). The NH group of the indole ring of UniPR129 undertakes a H-bond with the carboxylate group of Asp53, which may explain the negative effect of replacing the indolyl substituent with the α-napthyl group of **8** or the benzo[*b*]thiophen-3-yl group of **9**.

Free-energy calculations

To verify this model of interaction, the free-energy surface (FES) of binding/unbinding for the EphA2-UniPR129 complex was reconstructed by means of *wt*-MetaD. The unbinding process was simulated using the same computational protocol previously described for LCA, employing as CVs the distance **CV1** and the angle **CV2** defined above. Figure 3.14 shows the resulting FES of unbinding of UniPR129 from EphA2 receptor. The surface identifies a deep free-energy minimum region containing protein-ligand geometries (**CV1** = 9.5 Å; **CV2** = 100°) similar to the binding mode observed during the plain MD simulation and reported in Figure 3.12. On the same FES, configurations corresponding to the binding pose obtained by docking were observed in a high free-energy region (**CV1** = 10 Å; **CV2** = 130°), in agreement with the poor structural stability of this arrangement observed during the plain-MD simulation. No other free-energy binding minima were detected. A well-defined region corresponding to the unbound state of the ligand was identified (**CV1** > 20 Å; 60 < **CV2** < 130°), which allowed to estimate the Helmholtz free-energy of binding (ΔA_{bind}). The calculated ΔA_{bind} for the EphA2-UniPR129 complex (-8.35 ± 0.5 kcal/mol) was in reasonable agreement with the experimental one of -8.80

kcal/mol, deduced from the previously reported K_i,[52] supporting the validity of the proposed model of interaction.

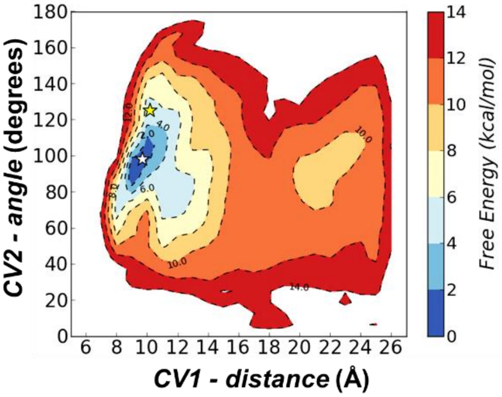
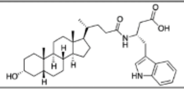
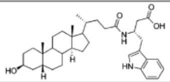
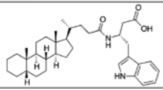
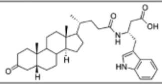
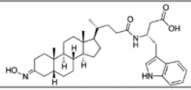
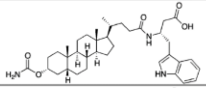
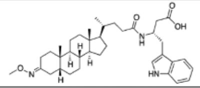


Figure 3.14. FES of EphA2-UniPR129 unbinding in the **CV1** and **CV2** space. The yellow star indicates the position of binding mode identified by docking, while white stars indicate the position of the most stable binding mode identified by plain MD simulation.

Validation of the binding model

To validate the new binding mode reported in figure 3.12, novel EphA2 antagonists were designed and tested. The new proposed binding mode emphasizes the importance of the 3α-hydroxyl group of UniPR129 for receptor binding. Indeed, this group undertakes H-bonds with both Asn57 side chain and Ile58 backbone. To experimentally validate that the presence of a substituent in the 3α position with H-bond capability is important to maintain high inhibitory potency, derivatives of compound UniPR129 were designed and synthetized, replacing the 3α-hydroxyl with a 3β-hydroxyl (**12**) group, a simple hydrogen atom (**13)**, or a 3-keto group (**14**). In all these cases, a significant reduction of the inhibitory potency was indeed observed (Table 3.3). Starting from the model of Figure 3.12, substituents able to form a H-bond network similar to that of the 3α-hydroxyl group were then designed. Supported by MetaD simulations (*vide infra*), the 3α-hydroxyl group of UniPR129 was replaced with a 3-hydroxyimino (**15**) or a 3α-carbamoyloxy (**16**) group.

Table 3.3. IC_{50} of 3-Substituted derivatives of UniPR129 from EphA2−ephrinA1 displacement experiments (as reported in Incerti et al. *J. Med. Chem.* **2017**, *60*, 787–796.)

Cpd.	Structure	IC_{50} (μM)[a,b]	PSA (Å^2)[c]
10 **UniPR129**		0.91 [0.80-1.1]	120
12		17 [13-24]	118
13		28 [22-35]	96
14		28 [22-36]	127
15		3.1 [2.8-3.6]	141
16		0.80 [0.51-0.98]	158
17		13 [8-20]	127

[a]Values are mean from at least three independent experiments. [b]Numbers in brackets denote the 95% confidence interval for IC50. [c]Polar surface area calculated with QikProp obtained from minimized 3D-structures of the listed compounds.

Simulations indicated that both the hydroxyimino and the carbamoyloxy groups were able to form productive interactions with Asn57 and Ile58 (Figure 3.15) and that estimated free-energy of binding ΔA_{bind} for the EphA2 receptor of **15** and **16** was comparable with the reference compound UniPR129. These new compounds were then tested in wet assay, and showed an inhibitory potency comparable or even slightly higher than the lead compound UniPR129. Indeed, **15** and **16** inhibited ephrin-A1 binding to EphA2 with IC_{50} values of 3.1 μM and 0.8 μM, respectively.

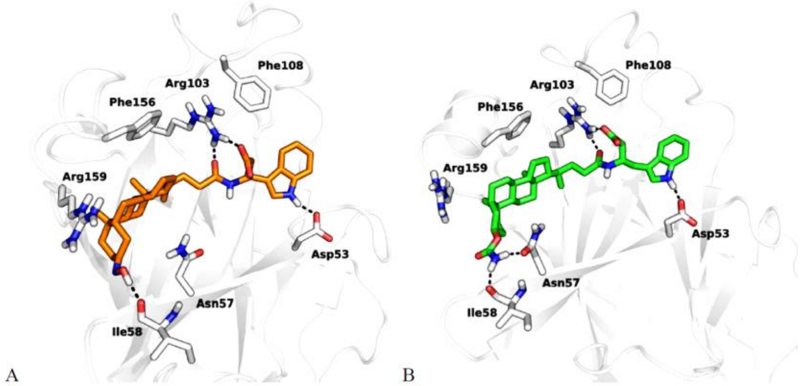


Figure 3.15. A) Binding mode of compound **15** (orange carbon atoms) within the EphA2 receptor (white cartoons, white carbon atoms) as obtained by META-D simulations. B) Binding mode of compound **16** (green carbon atoms) within EphA2 receptor (white cartoons, white carbon atoms) as obtained by META-D simulations.

Again, the removal of the H-bond donor group as in the case of **17**, led to a reduction in the inhibitory potency confirming that position 3 of the steroidal nucleus is accommodated in a hydrophilic pocket of EphA2 with strict stereo-electronic requirements (Table 3.3).

As a conclusive step of validation of the proposed binding mode, the free-energy of binding of a small set of selected EphA2 inhibitors by means of *wt*-MetaD was calculated and their agreement with the experimental data was analyzed. Table 3.4 reports the calculated ΔA_{bind} for UniPR129, compound **1**, **9, 13** and **16** along with their experimental $pIC_{50.}$ The computed ΔA well parallels inhibitory data on ephrin-A1 binding to EphA2 obtained in the wet assay. Remarkably, the applied computational methodology was able to reproduce the gain in the inhibitory potency obtained introducing the benzo[*b*]thiophen-3-ylmethyl chain of **9** in the side chain of the β-alanine conjugate **1**, or replacing the sulfur atom of **9** in the NH group of UniPR129. The effect on the potency obtained through modification of the position 3 of UniPR129 was also well captured by *wt*-MetaD simulations, as indicated by the ΔA_{bind} values calculated for compound **13** and **16**, with the former higher and the latter substantially equal to that of UniPR129.

Table 3.4. pIC_{50} values and calculated binding free-energy (ΔA) for selected EphA2 antagonists (as reported in Incerti et al. *J. Med. Chem.* **2017**, *60*, 787–796.)

Cpd	**pIC_{50}[a]**	$\Delta A_{calc.}$(kcal/mol)
1	4.54	- 4.50
9	5.74	- 7.65
10 (UniPR129)	6.01	- 8.35
13	4.55	- 3.35
16	6.15	- 8.71

[a]pIC_{50} values obtained from data reported in Tables 3.2 and 3.3.

Physicochemical properties of selected EphA2 antagonists

Computer-aided exploration of the SAR around UniPR129, allows to identify other effective EphA2 antagonists (compounds **15** and **16**) with comparable or slightly higher potency compared to UniPR129. It was then evaluated if these compounds were characterized by better in vitro ADME properties, according to their slightly higher polar surface area (PSA, Table 3.3).[58] compound **10** (UniPR129) reached barely detectable levels in plasma at 1 h after oral administration in mice at 30 mg/kg**.** A further analysis of its properties revealed that while UniPR129 possesses both a fair kinetic solubility and plasma stability, it suffers from rapid degradation in mouse liver microsomes (MLM), mainly for the oxidation of the carbon at position 3. Compounds **15** and **16,** being not susceptible of a direct oxidation at position 3 displayed a higher metabolic stability in MLM and reached a high plasma concentration in vivo after oral administration at 30 mg/kg, indicating that the introduction of polar groups diverse than a 3α-hydroxyl group is an effective strategy to maintain good inhibitory potency on EphA2 while gaining a higher oral bioavailability in mice.

---

[58] Valsami, G.; Macheras, P. Computational-Regulatory Developments in the Prediction of Oral Drug Absorption. *Mol. Inf.* **2011**, *30*, 112−121

Table 3.4. Physicochemical properties, in vitro metabolic stability and in vivo concentration of selected compounds (as reported in Incerti et al. *J. Med. Chem.* **2017**, *60*, 787–796.)

Cpd	**pIC$_{50}$**	**log$D_{oct,7.4}$**	**Solubility[a] (µM)**	**% compound in plasma[b]**	**$t_{1/2\ MLM}$ (min)[c]**	**In vivo C_{1h} (µM)**
10	6.01±0.07	4.90±0.15	31.8±4.2	98.3 ± 9.5	16.8±1.5	0.05±0.05
15	5.50±0.06	4.23±0.11	51.9±4.4	93.7 ± 11.3	60.4±9.3	0.40±0.10
16	6.15±0.15	5.01±0.20	18.7±3.4	98.2 ± 2.5	156.3±10.1	0.85±0.25

[a]From DMSO stock solution. Final DMSO concentration in 50 mM MOPS buffer pH 7.4:1%. [b]Mouse liver microsomes. [c]Compounds (30 mg/kg) were orally administered to at least four fasted mice.

3.4 Conclusions

Metadynamics-based methods show great promise in elucidating the mode of action and in computing the change in free energy associated with the binding of small molecules to their receptor. In the present chapter, FESs of the binding processes of EphA2-LCA and EphA2-UniPR129 complexes were reconstructed and characterized. The analysis of the FESs allowed to identify binding modes consistent with SAR data for both compounds. Moreover, exploiting an alternative binding mode, corresponding to a second free-energy minimum in the FES of EphA-LCA binding, allowed to synthetize a new compound (UniPR143) with increased potency compared to LCA, following a chemical intuition that the traditional docking model would have discouraged. The retrieved binding mode of UniPR129, combined with the exploration of the SAR around 5β-cholan-24-oyl-L-β-homotryptophan derivatives, allowed to drive the synthesis of novel EphA2 receptor antagonists, with inhibitory potency in the low micromolar range and, more importantly, featured by better physicochemical properties or improved metabolic stability, which ensured their oral bioavailability in mice. These results highlight the power of enhanced sampling methods in the rational design of novel active compounds with improved pharmacological properties.

3.5 Computational protocol

Docking simulations were performed starting from the crystal structure of the EphA2–ephrin-A1 complex (3HEI.pdb).[45] The EphA2–ephrin-A1 complex was submitted to a protein preparation procedure using Maestro 10.5 software.[59] This approach includes addition of missing side chains and hydrogen atoms, assignment of the tautomeric state of histidine residues to maximize the number of hydrogen bonds, and geometric optimization of the whole system to a root-mean-square displacement (RMSD) value of 0.3 Å. At the end of this procedure, ephrin-A1 ligand and solvent molecules were deleted from the EphA2 active site and the resulting structure was employed to build a docking grid with Glide 7.0.[60] The grid was centred on the ligand binding domain of EphA2, in a region delimited by Arg103, Phe156 and Arg159, using enclosing and bounding boxes of 20 and 14 Å on each side. The atomistic model of LCA was built with Maestro, and its geometry was optimized by energy minimization using OPLS2005 force field[61] to a gradient of 0.01 kcal mol^{-1} Å^{-2}. Docking simulations were then performed using Glide 7.0, starting from minimized structures of LCA. Two different top-ranked solutions were selected to be submitted to MD simulation.

EphA2-UniPR129 complex was generated starting from the previous reported binding mode of UniPR129 within the same X-ray structure of EphA2 receptor. Then, both EphA2-LCA complexes and EphA2-UniPR129 complex were solvated by TIP3P water molecules and its total charge neutralized by adding 6 Na$^+$ ions. All bond lengths to hydrogen atoms were constrained using M-SHAKE. Short-range electrostatic interactions were cut off at 9 Å, whereas long-range electrostatic interactions were computed using the Particle Mesh Ewald method.[62] A RESPA integrator was used with a time-step of 2 fs, and long-range electrostatics were computed every 6 fs. The solvated complexes were equilibrated using 5-ns long molecular dynamics simulations in the NPT ensemble at 1 atm and 300 K followed by 5 ns of simulations in the NVT ensemble at 300 K using the Langevin

---

[59] Maestro, version 10.5; Schrödinger, LLC: New York, 2016.

[60] Glide, version 7.0; Schrödinger, LLC: New York, 2016.

[61] Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). J. Comput. Chem. 2005, 26, 1752−1780

[62] Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N Log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

thermostat.[63] A production run of 30 ns in NVT condition was then performed at 300 K again with the Langevin thermostat. All the MD simulations were performed using Desmond 4.5 software[64] in combination with the OPLS2005 force field.

Molecular models of EphA2 in complex with compounds **1**, **9**, **13** and **16** were build, starting from the conformation of EphA2 obtained at the end of the plain MD simulation of the EphA2-UniPR129 complex. A new docking grid was build, using the same protocol previously described and the minimized structures of compounds **1**, **9**, **13** and **16** were docked into the new grid using Glide 7.0. The best docked poses were selected, and the resulting EphA2-inhibitor systems were solvated by TIP3P water molecules, neutralized by adding 6 Na^+ ions, and equilibrated by plain MD applying the same simulation schedule employed for the EphA2-LCA and EphA2-UniPR129 complexes.

Well-tempered MetaD simulations of protein-ligand unbinding

wt-MetaD simulations in the NVT ensemble were performed starting from the five equilibrated EphA2-inhibitor complexes. Also in this case, the simulations were performed with Desmond 4.5 software in combination with the OPLS2005 force field. The MetaD biasing potentials were added on two distinct CVs relevant for ligand unbinding, namely the distance between a COM of the EphA2 ligand binding domain and the COM of the ligand steroid moiety (**CV1**), and on the angle vector taken between the same protein reference point and the major inertia axis of the steroid moiety (**CV2**). Gaussians were deposited every 0.25 ps, with a starting height of 0.15 kcal/mol and gradually decreased on the basis of adaptive bias with a ΔT of 1200 K. The width of the Gaussians was 0.5 Å for the distance and 2.5 degrees for the angle. During the simulation, a repulsive wall was placed along the distance (**CV1**), at the value of 27 Å. Thus, when the compounds reached the edge of the wall, they felt a repulsive bias which pushed them back toward the EphA2 binding site. The form of the wall is reported in the following equation:

$$\frac{h_{wall}}{1 + exp\left(\frac{r_{wall} - r}{w_{wall}}\right)}$$

where h_{wall} is the wall height (expressed in kcal mol-1), r_{wall} is the wall location on **CV1** and w_{wall} is the width of the wall. The well-tempered MetaD simulations were considered ended when the height of the deposited Gaussian in the space of the

---

[63] Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-acetylalanyl-N′-methylamide. *Biopolymers* **1992**, *32*, 523−535.

[64] (a) Desmond Molecular Dynamics System, version 4.5; D. E. Shaw Research: New York, 2016. (b) Maestro−Desmond Interoperability Tools, version 4.5; Schrödinger: New York, 2016.

employed CVs decreases close to zero.[65] For practical reasons, a cut-off of 0.01 kcal/mol, corresponding to a residual height of the hill smaller than the 10% of the initial height, was applied to define the end of the *wt*-MetaD simulation. Convergence of the *wt*-MetaD simulations was verified by comparing the FES at different times of simulation once the cut-off in the Gaussian height was reached. In these conditions, the error associated to the estimation of ΔA_{bind} was lower than 0.5 kcal/mol. All the *wt*-MetaD simulations converged in less than 100 ns. FESs of EphA2-LCA binding/unbinding were reconstructed as average of four independent *wt*-MetaD simulations (Figure 3.16).

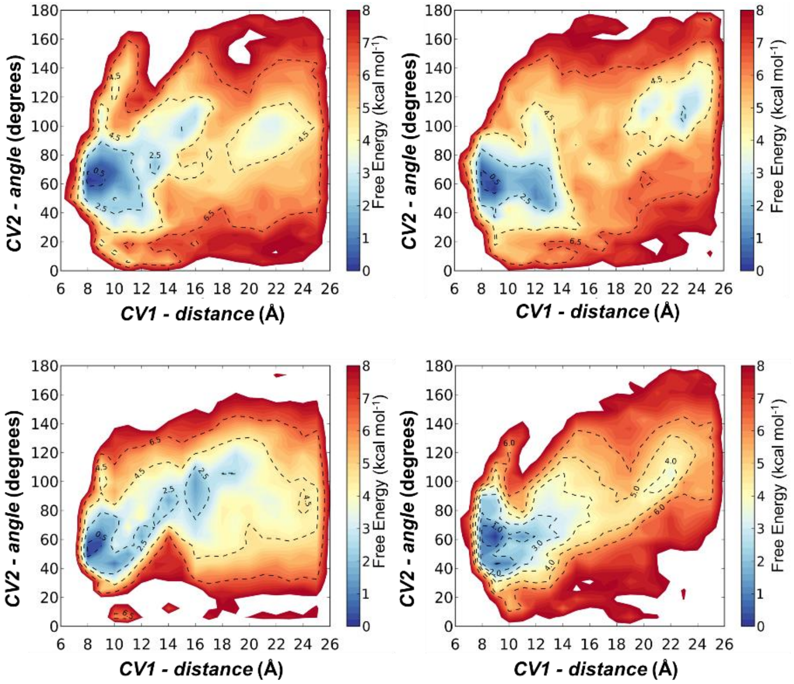


Figure 3.16. FESs of four independent simulations of EphA2-LCA binding computed in the **CV1** and **CV2** space, from simulations starting from binding mode A.

---

[65] Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered Metadynamics: A smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

Chapter 4

Investigating the kinetics of unbinding

4.1 Residence time in drug discovery

The early-phases of drug discovery are traditionally focused on the optimization of the affinity of a drug for its target with the final aim of increasing the drug potency in vivo. Recently, the relevance of binding kinetics in the lead optimization phase of drug discovery has been increasingly recognized. Drug-target binding *in vivo* is far from equilibrium conditions, due to the concurrence of complex physiological processes such as absorption, metabolism, distribution, and elimination as well as cellular-related events (e.g., receptor internalization and degradation). In these conditions, the lifetime of the drug−target complex represents the main determinant of the pharmacological response.[66] In this context, it is now well understood that the time-course of drug efficacy depends not only on the drug affinity, but also on the residence time of the drug compound on its molecular target. *Residence time* (RT), defined as the period for which a target is occupied by a compound, is related to the dissociation rate constant k_{off} (RT = $1/k_{\mathrm{off}}$) of the receptor–ligand complex.[67]

Residence time represents now an important metric for drug optimization, that can help improving pharmaceutically relevant drug properties, such as duration of action,[68] in vivo efficacy,[69]

selectivity,[70] and safety.[71] Indeed, increasing residence time by minimization of the k_{off}, can be a potential advantage for an extended pharmacological efficacy, which can be maintained even when systemic drug concentration has decreased. Moreover, a long residence time gives to the drug potentially significant advantages in term of target selectivity and safety, as protracted residence time for the primary target along with short residence times for off-targets would increase selectivity and eventually reduce the risk of unwanted side-activities. Residence time and the associated k_{off}, are therefore currently considered to be a key success factor for lead optimization during early-phase drug discovery and perhaps as important as the

---

[66] Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Opinion – Drugtarget Residence Time and Its Implications for Lead Optimization. *Nat. Rev. Drug Discovery* **2006**, 5, 730−739.

[67] Tummino, P. J.; Copeland, R. A. Residence Time of Receptor-Ligand Complexes and Its Effect on Biological Function. *Biochemistry* **2008**, 47, 5481−5492

[68] Vanderheyden, P. M.; Fierens, F. L.; Vauquelin, G. Angiotensin II Type 1 Receptor Antagonists. Why Do Some of Them Produce Insurmountable Inhibition? *Biochem. Pharmacol.* **2000**, 60, 1557−1563.

[69] Sykes, D. A.; Dowling, M. R.; Charlton, S. J. Exploring the Mechanism of Agonist Efficacy: A Relationship between Efficacy and Agonist Dissociation Rate at the Muscarinic M3 Receptor. *Mol. Pharmacol.* **2009**, 76, 543−551.

[70] Guo, D.; Dijksteel, G. S.; van Duijl, T.; Heezen, M.; Heitman, L. H.; IJzerman, A. P. Equilibrium and Kinetic Selectivity Profiling on the Human Adenosine Receptors. *Biochem. Pharmacol.* **2016**, 105, 34−41.

[71] Guo, D.; Hillger, J. M.; IJzerman, A. P.; Heitman, L. H. Drug-Target Residence Time–a Case for G Protein-Coupled Receptors. *Med. Res. Rev.* **2014**, 34, 856−892.

apparent affinity such as half-inhibitory concentration (IC_{50}) or the dissociation constant (K_D).[72]

In this context, the availability of computational methods able to predict k_{off} values or accurately rank different compounds binding the same target would accelerate drug optimization. In principle, either the k_{off} value or the residence time ($1/k_{off}$) can be estimated by performing molecular dynamics (MD) simulations.[73] However, as described in chapter 2, MD suffers from the inability to exhaustively explore drug−target conformational space within a reasonable time scale.[74] In recent years, different computational approaches aimed at estimating residence times have been reported in the literature,[75] including methods based on conformational flooding,[76] steered-MD simulations,[77] adiabatic-bias molecular dynamics,[78] and smoothed-potential MD.[79] Recently, Tiwary and Parrinello readopted the conformational flooding algorithm to *wt*-MetaD and were able to accurately calculate the rate of conformational transition of the so-called alanine dipeptide,[80] and to fairly estimate the residence time of benzamidine within trypsin enzyme.[81] To get consistent information from this approach, the following conditions should be met: *i)* the free-energy barrier separating the two basins involved in the transition should be meaningfully larger than RT (0.6 kcal/mol), *ii)* the employed CVs should discriminate the basins separated by the free-energy barrier, *iii)* no bias has to be deposited in the transition state area, in a way that the spontaneous well-to-well transition is unaffected. To satisfy the last condition, an infrequent deposition of

---

[72] Copeland, R. A. The Drug-Target Residence Time Model: A 10-Year Retrospective. *Nat. Rev. Drug Discovery* **2016**, 15, 87−95.

[73] Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular Determinants of Drug-Receptor Binding Kinetics. *Drug Discovery Today* **2013**, 18, 667−673.

[74] Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys*. **2004**, 120, 11919.

[75] Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem*. **2016**, 67, 159−184.

[76] Grubmüller, H. Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Phys. Rev. E* **1995**, *52*, 2893−2906.

[77] Capelli, A. M.; Bruno, A.; Entrena Guadix, A.; Costantino, G. Unbinding Pathways from the Glucocorticoid Receptor Shed Light on the Reduced Sensitivity of Glucocorticoid Ligands to a Naturally Occurring, Clinically Relevant Mutant Receptor. *J. Med. Chem.* **2013**, 56, 7003−7014.

[78] Bortolato, A.; Deflorian, F.; Weiss, D. R.; Mason, J. S. Decoding the Role of Water Dynamics in Ligand-Protein Unbinding: CRF1R as a Test Case. *J. Chem. Inf. Model.* **2015**, 55, 1857−1866.

[79] Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of Protein-Ligand Unbinding via Smoothed Potential Molecular Dynamics Simulations. *Sci. Rep*. **2015**, 5, 11539.

[80] Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys. Rev. Lett.* **2013**, *111*, 230602.

[81] Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of Protein-Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E386−E391.

Gaussians is required during the MetaD simulation, with a consequent significant slowdown of the calculation. Moreover, when all the conditions are fulfilled, a distribution of transition times registered for the same molecular system must be built to assess the reliability of the estimation. Indeed, the distribution of calculated residence times should follow a Poisson distribution,[82]

Despite the useful information it provides, the practical effectiveness of this approach is limited by the high amount of computational resources required to evaluate residence time for only one protein-ligand complex. Despite being promising, this method is still too computationally demanding for any high-throughput screening purpose.

For this reason, the aim of the work presented in this chapter was to calibrate an alternative and yet new MetaD-based protocol that could reproduce the variation of experimental residence times, starting from Meta-D simulations, accounting for the full unbinding of a ligand from its binding site. To be applied to the drug design process, such a method should be simple and provide a clear indication of the expected residence time at least in a qualitative way, allowing to rank or classify compounds. In this context, it is not required to give an absolute estimation of residence times, but it is sufficient that compounds having high structural similarity and significantly different RTs within the series are correctly classified.

In this chapter, the calibration of a new MetaD-based protocol aimed at ranking a set of congeneric compounds binding the same target, according to the experimental residence time is described.

---

[82] Salvalaglio, M.; Tiwary, P.; Parrinello, M. Assessing the Reliability of the dynamics reconstructed from metadynamics. *J. Chem. Theory Comput.* **2014**, *10*, 1420−1425.

4.2 Dataset of CDK8 inhibitors

The chosen dataset for calibration is the recently reported set of arylpyrazole inhibitors of cyclin-dependent kinase 8 (CDK8).[83] For this dataset both experimental residence times and X-ray structures are available. These compounds, reported in Figure 4.1, show a common pyrazol-5-yl-urea scaffold and can be divided into three different classes according to their experimental residence time: *i)* short residence time (SRT) inhibitors **1**−**3** (RT < 1.4 min), *ii)* medium residence time (MRT) inhibitors **4** and **5**, which exhibit residence times of 14 and 57 min respectively, *iii)* long residence time (LRT) inhibitors **6** and **7**, which show extremely long residence times of 1626 and 1944 min, respectively. Compound **1** has no substituent to the common scaffold. Compounds **2** and **3** present small hydrophilic substituents (hydroxyethyl and morpholinoethyl respectively), compounds **4** and **5** carry longer hydrophilic substituent such as a morpholinopropyl or a hydroxypentyl chain, while compounds **6** and **7** which have a 1-(2-(4-(3-*tert*-butyl-1-*p*-tolyl-1H-pyrazol-5-ylcarbamoyl)-piperazin-1-yl)ethyl) or *tert*-butoxycarbonylaminopropyl chain present long lipophilic chains.

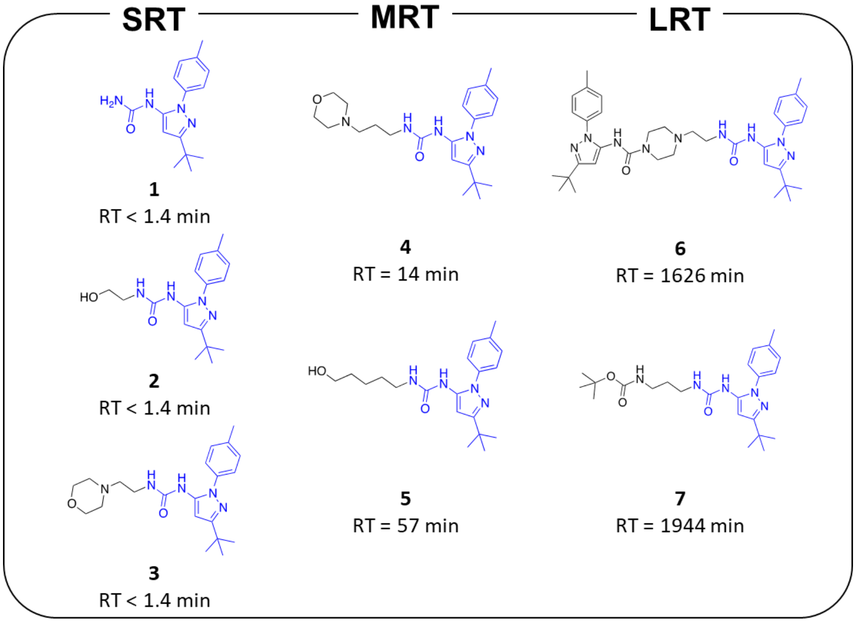


Figure 4.1. Chemical structures of the CDK8 inhibitors reported in reference 83 initially considered in this study and their experimental residence times. The common 1-(3-tert-butyl-1-p-tolyl-1H-pyrazol-5-yl)-urea scaffold is depicted in blue

---

[83] Schneider, E. V.; Böttcher, J.; Huber, R.; Maskos, K.; Neumann, L. Structure-Kinetic Relationship Study of CDK8/CycC Specific Compounds. *Proc. Natl. Acad. Sci. U. S. A*. **2013**, 110, 8081−8086.

These compounds bind CDK8 with the same binding mode in an allosteric pocket which is accessible only when the kinase is in the so-called "DMG-out" inactive conformation (Figure 4.2). Their common 1-(3-*tert*-butyl-1-*p*-tolyl-1H-pyrazol-5-yl)-urea scaffold lies near the DMG motif of the activation loop (corresponding to the DFG sequence of other kinases), with the urea donating two H-bonds to the carboxylate group of Glu66 and accepting a H-bond from the backbone NH of Asp173, and the substituents projected toward the hinge region. While compounds **1**−**3** do not establish direct contacts with the hinge region, compounds **4** and **5** are able to interact with the backbone NH of Ala100 of the hinge region, using their morpholinopropyl and hydroxypentyl chains. In contrast, the chains of compounds **6** and **7** do not interact with the hinge region but protrude into the front pocket of CDK8.

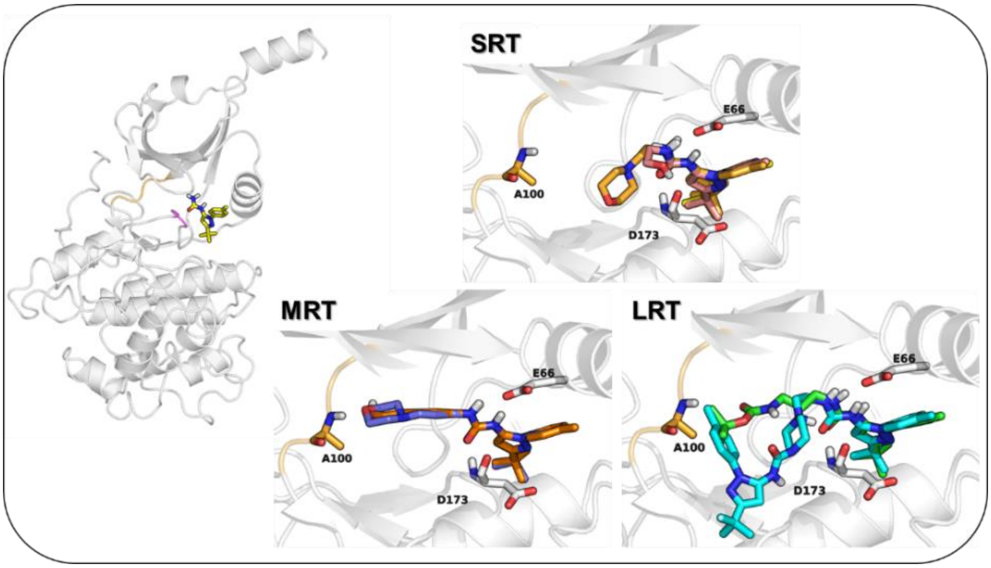


Figure 4.2 Binding mode of compounds **1**−**7** within CDK8 active site. Binding mode of compounds **1**-**3** (SRT panel), compounds **4** and **5** (MRT panel), compounds **6** and **7** (LRT panel). Color code: compound **1** (yellow carbon atoms C), **2** (pink C), **3** (orange C), **4** (blue C), **5** (red C), **6** (cyan C), and **7** (green C). Protein carbon atoms are represented in white. The protein backbone is represented as a white cartoon, except for the hinge region, represented as orange cartoon.

The chosen dataset can be considered an ideal set for calibrating a MetaD-based protocol to predict binding kinetics for three main reasons: *i)* the uncertainty of the target variable is under control, *ii)* collected residence times range over three orders of magnitude, *iii)* the arylpyrazole inhibitors possess a comparable binding mode and share the same mechanism of action.

4.3 Results and discussion

As mentioned in chapter 2, one critical stage in performing MetaD simulations is the choice of a suitable set of CVs. Preliminary simulations of unbinding were carried out to identify the appropriate set of CVs. Seven CVs were chosen to describe ligand roto-translational movements and conformational changes during the protein-ligand unbinding process (Figure 4.3), providing reliable results for this protein−ligand complex (see Appendix for details). The first three CVs describe the translational movements of the ligand center of mass (COM) with respect to the CDK8 active site (Figure 4.3, panels A-C). In particular, **CV1** (A) is the distance between the COM of compound **1** and the COM of the protein hinge region (amino acids 98−103); **CV2** (B) is the angle defined by the COM of the ligand, the COM of the protein hinge region, and the COM of β-strand 6 of the kinase active site (amino acids 92−98); **CV3** (C) is the dihedral angle defined by these three COMs and a fourth COM taken on β-strand 5 of the kinase active site (amino acids 80−87). A second set of three CVs (Figure 4.3, panels D−F) describe the rotation of the whole ligand with respect to the CDK8 active site. **CV4** (D) is the angle formed by the major axis of the ligand (approximately defined by selecting two atoms of the rigid pyrazolylurea scaffold) and the COM of the protein hinge region; **CV5** (E) is the dihedral angle defined by the major axis of the ligand, the COM of the protein hinge region, and the COM of β-strand 1 of the kinase active site; and **CV6** (F) is the angle formed by the minor axis of inertia of the ligand (approximately defined by selecting two atoms of the rigid phenylpyrazole scaffold) and the COM of the protein hinge region. Finally, **CV7** is the RMSD of the ligand with respect to the initial experimental orientation, included to account for degrees of freedom orthogonal to those explored by the previous CVs.

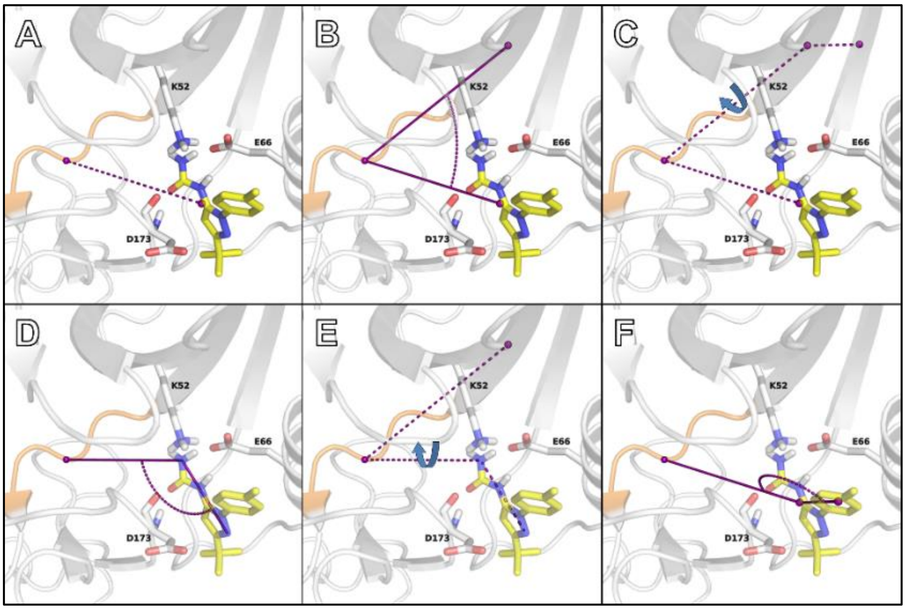


Figure 4.3. Graphical representations of the first six CVs employed for the metadynamics simulations involving the CDK8−compound **1** system. Center of mass points employed to define the CVs are depicted in purple, while axes are described by plain purple lines. (A) **CV1** distance, (B) **CV2** angle, (C) **CV3** dihedral angle, (D) **CV4** angle, (E) **CV5** dihedral angle, (F) **CV6** angle.

To simulate the process of protein-ligand unbinding classic MetaD approach was applied, avoiding well-tempering conditions. In this way, since in classic MetaD simulations the biasing potential is deposited as Gaussian functions with constant height, the simulation time spent in the bound state is proportional to the integral of the deposited potential in the CV space. Thus, it's possible to discriminate compounds according to the time they spent in the bound state, assuming that ligands that need to cross higher or multiple energy barriers require more energy from the MetaD protocol to exit from their binding sites. For this reason, the simulation time required to move from the crystallographic pose of the ligand to the completely unbound state (t_{MetaD}) was registered at the end of the unbinding simulations and used to compare and classify all the compounds. As a ligand could follow different unbinding paths during MetaD simulations, values of t_{MetaD} were collected from replicated simulations and averaged. Moreover, to do so, an objective criterion to define of the unbound state is required to register t_{MetaD}. While recent investigations mainly used geometrical criteria,[84,85] in this work an energy-based definition that does not require time-consuming analysis of MetaD trajectories, but

---

[84] Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of Protein-Ligand Unbinding via Smoothed Potential Molecular Dynamics Simulations. *Sci. Rep*. **2015**, 5, 11539.

[85] Tiwary, P.; Mondal, J.; Morrone, J. A.; Berne, B. J. Role of Water and Steric Constraints in the Kinetics of Cavity-Ligand Unbinding. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 12015−12019.

could be extracted from the time course of deposited META-D potentials was applied. In a MetaD simulation, the biasing potential accumulated in the CV space grows proportionally to the simulation time. On the other hand, when a basin is flat, as in the case of a small molecule free to move in the bulk solvent, no accumulation of deposited potential is expected, and it is unlikely that the same region of the CV space is explored twice or more. To test that, a MetaD simulation of compound **1** free to move in a box of bulk solvent was performed (including a portion of CDK8 for the definition of the CVs). In these conditions the average value of the registered potential was approximately constant at 0.8 kcal/mol (figure 4.4).

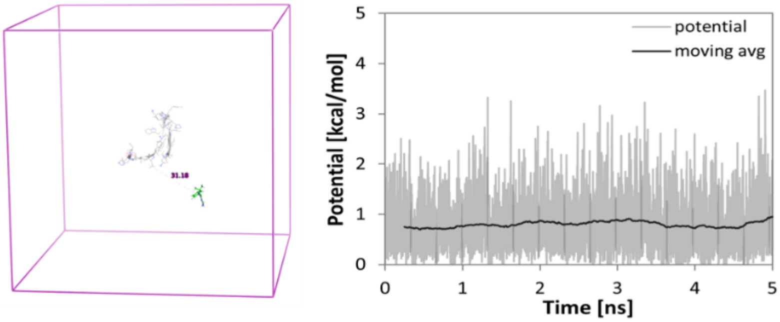


Figure 4.4 Representative configuration of compound **1** in the bulk solvent with the portion of CDK8 required to define the seven CVs (left). Biasing potential accumulated during the MetaD simulation in the explored regions of the space of seven CVs for compound **1** freely moving in a bulk solvent (right). The moving average (black line) was calculated for the preceding 500 ps of the simulation.

This value of biasing potential in the unbound state could be used as a cut-off value to identify when the ligand has reached the completely unbound and solvated state. This new approach was then tested by simulating the full unbinding process of compound **1** from CDK8. Figure 4.5 reports the time course of the 500 ps-averaged potential deposited during the simulation (left panel) along with the time course of **CV1** (right panel), which is the distance between the ligand and the binding site COM and offers a perspective of the ligand−protein distance. At early simulation times (0−1 ns), a significant growth of the average potential deposited is registered, and the ligand is still in its starting basin (**CV1** < 15 Å). When compound **1** exits from the first basin, moving outside the crystallographic binding pocket, but still interacting with the surface of CDK8, several fluctuations in the average potential are observed (1−4 ns, 15 Å < **CV1** < 30 Å). At longer simulation times (>4.45 ns, **CV1** > 30 Å), compound **1** is no longer in contact with the CKD8 surface and is completely solvated in its unbound state. According to the test reported in figure 4.4, the average potential decreases below 0.8 kcal/mol.

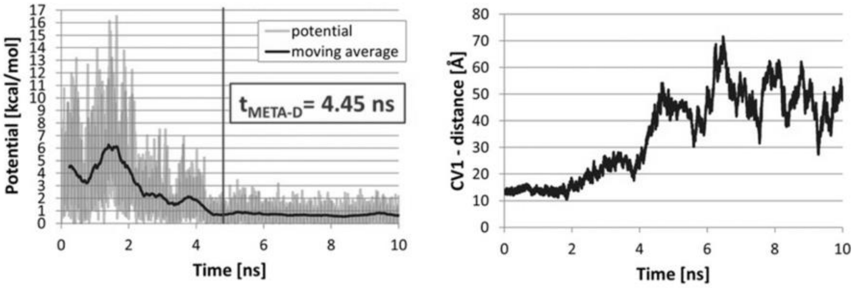


Figure 4.5. META-D simulation of unbinding of compound 1 from CDK8. Time course of the average potential deposited during the simulation (left panel). Time course of CV1 (right panel).

The approach was then extended to the whole set of crystallized CDK8−inhibitor complexes, performing 14 independent MetaD simulations for each system complex. Following the time course of the deposited biasing potential, each simulation was stopped when the ligand was completely solvated (when the average potential reached the cut-off value of 0.8 kcal/mol).

In Table 4.1 the average t_{MetaD} values for compounds **1**−**7** and their standard errors of the mean are reported. The approach could correctly discriminate LRT inhibitors from all the others. While compounds **6** and **7** require more than 15 ns of MetaD to reach the solvated state, compounds **1**−**5** took less than 12 ns of MetaD to completely unbound from CDK8. Moreover, a separation can be seen between MRT inhibitors (compounds **4** and **5**) and SRT inhibitors (compounds **1** and **2**), even though the SRT morpholino derivative **3** has a relatively high t_{MetaD}.

Table 4.1. Experimental Residence Times (RT) and Calculated t_{MetaD} Values (mean ± SEM, n = 14) for CDK8 Inhibitors **1**−**7**. Color code: SRT (red), MRT (yellow), LRT (green)

Compound	RT (min)	t_{MetaD} (ns)
1	<1.4	4.87 ± 0.41
2	<1.4	7.6 ± 0.61
3	<1.4	9.62 ± 0.76
4	14	10.85 ± 1.28
5	57	11.2 ± 0.94
6	1626	>25
7	1944	17.98 ± 1.04

Taking into account that the pKa of alkylmorpholines is about 7.4,[86] the unclear classification of compound **3**, modeled in its neutral state, could be related to its ionization state. Therefore, 14 MetaD simulations of compound **3** in its protonated

---

[86] Hall, H. K., Jr Potentiometric Determination of the Base Strength of Amines in Non-protolytic Solvents. *J. Phys. Chem.* **1956**, *60*, 63−70.

state were performed. For consistency, also MetaD simulations for compound **4** were repeated with compound **4** in its protonated form. The new t_{MetaD} values obtained are only slightly affected by the protonation state suggesting that the morpholine amino groups of compound **3** (t_{MetaD} = 10.39 ± 1.39 ns) and compound **4** (t_{MetaD} = 11.48 ± 1.42 ns) do not significantly interact with the CDK8 active site. Molecular simulations with fixed protonation states for titrable groups are poor approximations of complex proton-exchange equilibria. Constant pH Molecular dynamics algorithms[87],[88] may be better suited to treat cases with ligands and protein residues having weakly basic or acidic groups.

Then, MetaD unbinding trajectories were analyzed to investigate if they could provide useful insights about how structural fragments can influence ligand unbinding. In contrast to the results of Schneider et al.,[89] who suggested based on an analysis of the X-ray structures that the occupation of the front pocket can lead to long residence times (compounds **6** and **7**), MetaD simulations indicate that either the ureido (**6**) or carbamoyl (**7**) fragment in the pendant chain undertake H-bond interactions with Lys52, a key polar residue proximal to the DMG motif. This interaction delays ligand unbinding, accounting for the extended occupancy of the binding site (Figure 4.6). Considering that, it might be possible to rationally design LRT inhibitors that specifically target this Lys52.

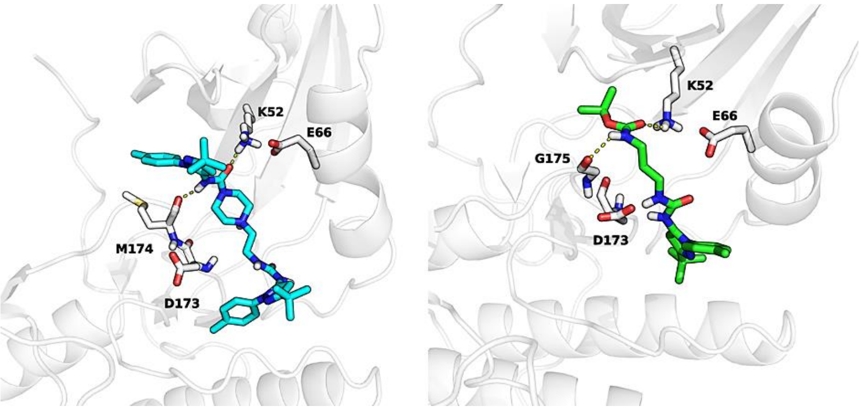


Figure 4.6. Representative snapshots of the MetaD simulations of CDK8 in complex with compound **6** (left, cyan carbon atoms) or with compound **7** (right, green carbon atoms), in which either the ureido (**6**) or carbamoyl (**7**) fragment in the pendant chain establishes H-bond interactions with Lys 52

---

[87] Swails, J. M.; York, D. M.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10*, 1341−1352.

[88] Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmuller, H. Constant pH Molecular Dynamics in Explicit Solvent with λ-Dynamics. *J. Chem. Theory Comput.* **2011**, *7*, 1962−1978.

[89] Schneider, E. V.; Böttcher, J.; Huber, R.; Maskos, K.; Neumann, L. Structure-Kinetic Relationship Study of CDK8/CycC Specific Compounds. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, 110, 8081−8086.

To validate the proposed MetaD protocol, other three-arylpyrazole CDK8 inhibitors reported by Schneider et al., were included in the dataset. The methyl derivative **8**, the hydroxypropyl derivative **9**, and the hydroxybutyl derivative **10** are represented in Figure 4.7. For these compounds, the experimental residence times are reported but X-ray structures are lacking. Compounds **8** and **9** belong to the SRT class, having very short residence times (<1.4 min), whereas compound **10** is classified as an MRT inhibitor, having a residence time of 7 min.

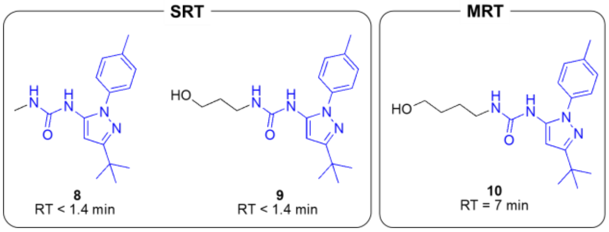


Figure 4.7 chemical structures and experimental RT of compounds **8**−**10**

Compounds **8**−**10** were docked into the CDK8 kinase binding site taken from the X-ray structure of the complex with compound **1** using Glide software[90] (details in Computational Protocol). Figure 4.8 shows the top-ranked docking solutions of the modeled compounds **8-10.** Their arylpyrazole moiety has the same binding mode of compound **1**, with the urea group undertaking H-bonding interactions with Glu66 and Asp173. The hydroxybutyl chain of compound **10** was long enough to undertake a H-bond with the hinge region of CDK8.

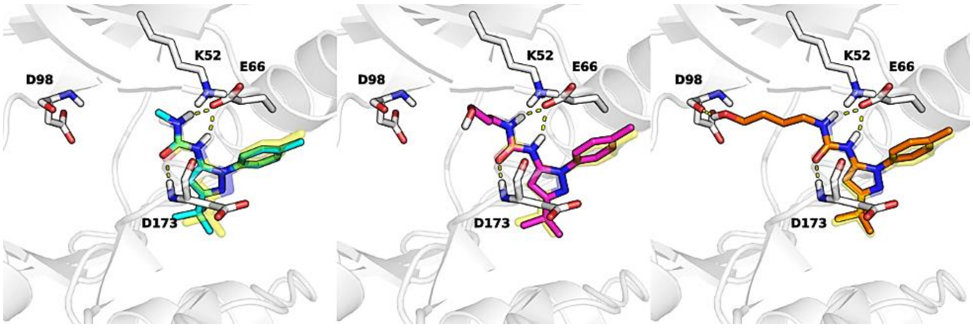


Figure 4.8. Top-ranked docking solutions of compound **8** (left, green carbon atoms), **9** (middle, pink carbon atoms) and **10** (right, orange carbon atoms) within CDK8. Structure of compound **1** in its crystallographic binding pose represented in pale yellow.

---

[90] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

Then, these complexes were submitted to the same MetaD protocol described for the previous other CDK8 inhibitors. The resulting t_{MetaD} values for compounds **8-10** are reported in Table 4.2 along with those obtained for compounds **1**−**7**. As shown, compound

8 and **9** were correctly ranked as SRT inhibitors (t_{MetaD} = 6.67 ± 0.54 and 7.43 ± 0.57 ns, respectively), whereas compound **10** was well assigned to the MRT class (t_{MetaD} = 10.75 ± 0.99 ns).

Table 4.2. Experimental Residence Times (RT) and Calculated t_{MetaD} Values (mean ± SEM, n = 14) for CDK8 Inhibitors **1**−**10**. Color code: SRT (red), MRT (yellow), LRT (green), compounds **8-10** in bold.

Compound	RT (min)	t_{MetaD} (ns)
1	<1.4	4.87 ± 0.41
8	**<1.4**	**6.67 ± 0.54**
9	**<1.4**	**7.43 ± 0.57**
2	<1.4	7.6 ± 0.61
3	<1.4	9.62 ± 0.76
10	**7**	**10.75 ± 0.99**
4	14	10.85 ± 1.28
5	57	11.2 ± 0.94
6	1626	>25
7	1944	17.98 ± 1.04

These results indicate that the method has a reasonable predictive power for this class of inhibitors and could be used for prospective drug discovery of newly designed analogues.

4.4 Conclusions

The residence time of a protein−ligand complex is a critical parameter which affects biological effects *in vivo*. Even with the availability of powerful computer resources and computational techniques, the prediction of residence time is still a challenging task. In this chapter, it is described how a new MetaD-based approach was set up and validated to rank a set of arylpyrazole CDK8 inhibitors for which both the X-ray structures of the enzyme−inhibitor complexes and the experimental RTs of the inhibitors were available. This approach is based on two properties directly resulting from the classical MetaD algorithm: *i)* the lack of accumulation of the deposited potential in the unbound state, employed to define when to stop the simulation, and *ii)* the MetaD simulation time (t_{MetaD}), used to perform comparison among ligands binding the same target.

Even though more investigations are required to support the general applicability of the method, the proper ranking of residence times of arylpyrazole CDK8 inhibitors spanning more than 3 orders of magnitude suggests that this approach can be applied in drug design projects aimed to optimize the experimental residence time of a congeneric series of ligands.

4.5 Computational protocol

Model Building and equilibration

CDK8−inhibitor complexes were prepared starting from X-ray strucures downloaded from the Protein Data Bank (PDB IDs: 4F6S, 4F6U,4F6W, 4F7J, 4F7L, 4F7O, and 4F7N).[91] The complexes were prepared using the Protein Preparation Wizard in Maestro version 10.4.[92] Part of the activation loop (residues 177−194) was missing in all the X-ray structures and was built in Prime 4.2[93] using the homologue CDK6 (PDB ID 1BI8,[94] similarity score 42%, crystallized with its DFG motif in the out conformation) as a template structure (sequence alignment in Figure 4.9); Compounds **1**−**5** and **7**-**10** were modeled in their neutral state, whereas compound **6**, carrying a basic nitrogen in a piperidine ring, was modeled in its protonated form. All the complexes were energy-minimized using the OPLS2005 force field[95] as implemented in Macromodel 11.[96] The prepared and minimized CDK8−inhibitor complexes were solvated by TIP3P water molecules in a simulation box 12 Å distant from the protein in every direction, neutralized by addition of five Cl^- ions, and equilibrated for 20 ns of MD simulation (5 ns in the NPT ensemble and 15 ns in the NVT ensemble) at 300 K using the Langevin thermostat.[97] All bond lengths to hydrogen atoms were constrained using M-SHAKE. Short-range electrostatic interactions were cut off at 9 Å, whereas long-range electrostatic interactions were computed using the PME method.[98] A RESPA integrator was used with a time step of 2 fs. MD simulations were performed using the OPLS2005 force field in Desmond 4.4 software.[99]

---

[91] Schneider, E. V.; Böttcher, J.; Huber, R.; Maskos, K.; Neumann, L. Structure-Kinetic Relationship Study of CDK8/CycC Specific Compounds. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 8081−8086.

[92] *Maestro*, version 10.4; Schrödinger, LLC: New York, **2015**.

[93] *Prime*, version 4.2; Schrödinger, LLC: New York, **2015**.

[94] Russo, A. A.; Tong, L.; Lee, J.-O.; Jeffrey, P. D.; Pavletich, N. P. Structural Basis for Inhibition of the Cyclin-Dependent Kinase Cdk6 the Tumor Suppressor p16ink4a. *Nature* **1998**, *395*, 237−243.

[95] Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005**, *26*, 1752−1780.

[96] *MacroModel*, version 11; Schrödinger, LLC: New York, **2015**.

[97] Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-acetylalanyl-N′-methylamide. *Biopolymers* **1992**, 32, 523−535.

[98] Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N Log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

[99] (a) *Desmond Molecular Dynamics System*, version 4.4; D. E. Shaw Research: New York, **2015**. (b) *Maestro−Desmond Interoperability Tools*, version 4.4; Schrödinger, LLC: New York, **2015**.

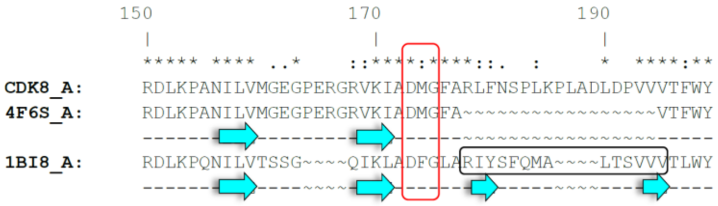


Figure 4.9. Sequence alignment of a common segment between CDK8 and CDK6 enzymes. The sequence of CDK8 (marked as CDK8_A in the first line) reported in the Uniprot website (code: P49336) is aligned with the CDK8 and CDK6 sequences extracted from PDB structures 4F6S and 1BI8, respectively. The red box underlines the DM(F)G motif of the activation loop, the black box indicates the sequence of CDK6 used to build the missing part of the activation loop in CDK8.

META-D Simulations of unbinding

Equilibrated CDK8−inhibitor complexes were submitted to classic MetaD simulations with no well-tempering conditions using the Desmond 4.4 software (GPU implementation) on NVDIA780 graphic cards. using the previously described set of seven collective variables. Gaussians were deposited every 0.5 ps to obtain the full unbinding of the ligand in reasonable times of simulation (25 ns). The width of the Gaussians was defined based on a series of preliminary MetaD simulations of the CDK8−compound **1** complex. As a rule of thumb, the width of the Gaussians was set to one-fifth of the variation of the CV required to drive the ligand out from the crystallographic free energy minimum. In detail, 0.6 Å was used for the distance CV1, 6° for the angle CV2, 5° for the torsion angle CV3, 15° for the angle CV4, 15° for the torsion angle CV5, 15° for the angle CV6, and 0.3 Å for the ligand root-mean-square deviation (RMSD) (CV7).

*MetaD simulation of compound **1** in bulk solvent*

To measure the parameter t_{MetaD}, which corresponds to the simulation time necessary for the ligand to unbind, a criterion for the detection of the unbound state was developed as follows. Compound **1** and a portion of CDK8 surrounding the ligand (residues 92−103) were extracted from the X-ray structure (PDB ID: 4F6S) and prepared for simulations in a water box. The N- and C-terminus of the truncated protein were capped using the Protein Preparation Wizard implemented in Maestro. Compound **1** was translated 30Å apart from the center of mass of the CDK8-derived peptide. The system was solvated by a cubic box of TIP3P water molecules with a length of 100Å and equilibrated by MD for 2.5 ns under NPT conditions followed by 2.5 ns under NVT conditions at 300 K. All the MD parameters were the same as employed for the equilibration of the whole CDK8−inhibitor complexes. The equilibrated complex was submitted to a MetaD simulation in the space of the seven CVs described above with the same protocol of Gaussian deposition. The ligand was free to explore the solvent bulk without any interaction with the protein atoms. Under these conditions, corresponding to an unbound state, no accumulation of the potential deposited by the MetaD protocol was observed, and the average value of the potential showed a constant value of 0.8 kcal/mol for all points of the CV space explored throughout the entire simulation (5 ns).

*Docking of compounds **8-10***

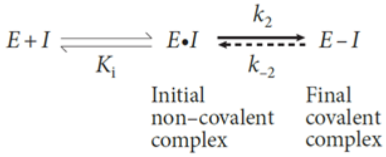
The CDK8-compound **1** complex, prepared as descibed above, was employed to build a docking grid centered on inhibitor **1**, applying enclosing and bounding boxes of 20 and 10 Å, respectively. Molecular models of compounds **8-10** were built with Maestro, and their geometry was optimized by energy minimization using OPLS2005 force field to a gradient of 0.01 kcal mol^{-1} Å^{-2}. Docking simulations were then performed using Glide 6.9, starting from minimized structures of inhibitors **8-10**. The best scored docking solutions of **8-10** in which the common (3-(tert-butyl)-1-(p-tolyl)-1H-pyrazol-5-yl)urea scaffold assumes the same interactions previously described for compound **1** were selected and the resulting complexes with CDK8 were equilibrated and used for MetaD simulations with the same parameters defined above.

Chapter 5

Covalent drugs affinity and reactivity

5.1 Covalent drugs design and optimization

Despite the many examples of successful covalent drugs, principles for the rational design of these molecules have only recently emerged, thus enabling the expansion of this therapeutic class.[100] It is now clear that structural bioinformatics approaches, coupled with structure-based drug design, may help the design of selective covalent drugs. We should keep in mind, that in the case of covalent inhibition, there are at least two steps that we should consider for aiding the design of new compounds by means of computational methods:



i) *Recognition step*: the compound first binds in a non-covalently way to the target protein, positioning its weakly reactive electrophile group (*warhead*) close to a specific nucleophile on the protein;
ii) Chemical step: the compound then forms a specific covalent adduct with the targeted protein.[101] In this context, both recognition and chemical steps of covalent inhibition should be exploited to optimize a covalent drug. In cases in which bond formation is effectively irreversible, k_{-2} will be zero. Non-covalent affinity (K_i) of the drug must be high enough to guarantee that the compound binds selectively to the targeted protein and reaches an adequate residence time for a covalent reaction. Similarly, the reaction rate (k_2) of the chemical step must be high enough to allow the reaction to occur within the lifetime of the non-covalent complex formed in the recognition step of the inhibition. However, as highly reactive electrophiles should be avoided to circumvent non-specific reactions, the reaction rate must be achieved primarily by the optimal positioning of the warhead relative to the nucleophile residue on the target. Modelling both recognition and chemical steps would thus enable medicinal chemist to study covalent reaction mechanisms and fine-tune covalent drugs with appropriate affinity, reactivity and selectivity.

---

[100] De Cesco, S.; Kurian, J.; Dufresne, C.; Mittermaier, A. K.; Moitessier, N. Covalent Inhibitors Design and Discovery. *European Journal of Medicinal Chemistry* **2017**, *138* (Supplement C), 96–114.

[101] Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The Resurgence of Covalent Drugs. *Nat Rev Drug Discov* **2011**, *10* (4), 307–317.

A major challenge for the treatment with covalent inhibitors in cancer therapies is the emergence of drug resistance.[102] Even though targeting non-catalytic amino acids is a strategy to confer selectivity of covalent inhibitors, mutations within the binding site of the receptor may circumvent inhibition. The rarity of the targeted residue may indicate that it is not essential to the function of the protein, thus offering an easy way to cancer to overcome inhibition. Moreover, mutations other than the targeted residue in the binding site may have a role in drug resistance, by affecting drug-target recognition, drug reactivity or the positioning of the warhead with respect to the targeted amino acid. Also in this context, exploiting both recognition and chemical steps of covalent inhibition could give hints to understand drug resistance.

In the present chapter, the molecular mechanism of resistance to lung cancer drug osimertinib is explored. The appearance of new mutations (i.e. L718Q) is currently limiting the benefit of osimertinib, an irreversible inhibitor designed to selectively target EGFR T790M and capable of alkylating Cys797. In this chapter, a computational investigation elucidating the impact of L718Q mutation on *i)* ionization state of Cys797; *ii)* activation energy and reaction energetics for Cys797 alkylation; *iii)* free-energy of binding and *iv)* conformational space of osimertinib, was performed to rationalize the clinical inactivity of osimertinib on EGFR T790M/L718Q variant, giving useful insights for the design of novel inhibitors.

---

[102] Engel, J.; Lategahn, J.; Rauh, D. Hope and Disappointment: Covalent Inhibitors to Overcome Drug Resistance in Non-Small Cell Lung Cancer. *ACS Med. Chem. Lett*. **2016**, *7*, 2–5.

5.2 EGFR inhibitors in NSCLC treatment

Epidermal growth factor receptor (EGFR) is a trans-membrane protein which possesses an extracellular EGF binding domain and an intracellular tyrosine kinase domain.[103] EGFR activation by its physiological ligand EGF leads to receptor dimerization and phosphorylation, two events that trigger the activation of signal transduction cascades promoting cell proliferation.[104] In non-small cell lung cancer (NSCLC), overexpression of EGFR or mutations in its kinase domain have been observed in at least 50% of the cases.[105] Of the known kinase mutations, more than the 90% occurs as short in-frame deletions in exon 19 or as point mutation in exon 21, the latter resulting in arginine replacing leucine at codon 858 (L858R).[106] These mutations are critical for NSCLC insurgence and progression as their presence results in constitutive activation of EGFR regardless of the presence of EGF.[107] The first-generation of EGFR inhibitors (i.e., gefitinib, **1**, Figure 5.1) represents the first-line therapy for NSCLC that harbors activating EGFR mutations (i.e. L858R substitutions in exon 21).[108] Although patients achieve good responses to the therapy, most of them acquired drug resistance within 1 year treatment, which is driven in about 60% of cases by a EGFR T790M point mutation.[109] Visual inspection of the X-ray structures showed that while the presence of a methionine at the gatekeeper position did not block binding of inhibitors within the ATP binding site,[110] it prevents the formation of a water mediated H-bond involving N3 nitrogen the quinazoline nucleus and the threonine present in the native form of EGFR. The change in the steric and lipophilic property of the gatekeeper is likely responsible

---

[103] Hynes, N. E.; Lane, H. A. ErbB receptors and cancer: the com-plexity of targeted inhibitors. *Nat. Rev. Cancer* **2005**, *5*, 341−354.

[104] Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell*, **2000**, *103*, 211−225.

[105] Normanno, N.; De Luca, A.; Bianco, C.; Strizzi, L.; Mancino, M.; Maiello, M. R.; Carotenuto, A.; De Feo, G.; Caponigro, F.; Salomon, D. S. Epidermal Growth Factor Receptor (EGFR) Signaling in Cancer. *Gene* **2006**, 366, 2−16.

[106] Sharma, S. V.; Bell, D. W.; Settleman, J.; Haber, D. A. Epidermal Growth Factor Receptor Mutations in Lung Cancer. *Nat. Rev. Cancer* **2007**, *7*, 169−181

[107] Pao, W.; Chmielecki, J. Rational, Biologically Based Treatment of EGFR-Mutant Non-Small-Cell Lung Cancer. *Nat. Rev. Cancer* **2010**, *10*, 760−774.

[108] Carmi, C.; Lodola, A.; Rivara, S.; Vacondio, F.; Cavazzoni, A.; Alfieri, R. R.; Ardizzoni, A.; Petronini, P. G.; Mor, M. Epidermal Growth Factor Receptor Irreversible Inhibitors: Chemical Exploration of the Cysteine-Trap Portion. *Mini-Rev. Med. Chem*. **2011**, *11*, 1019− 1030.

[109] Engelman, J. A.; Jaⁿne, P. A. Mechanisms of Acquired Resistance to Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors in Non-Small Cell Lung Cancer. *Clin. Cancer Res.* **2008**, *14*, 2895−2899.

[110] Michalczyk, A.; Klu̇ter, S.; Rode, H. B.; Simard, J. R.; Gru̇tter, C.; Rabiller, M.; Rauh, D. Structural insights into how irreversible inhibitors can overcome drug resistance in EGFR. *Bioorg. Med. Chem.* **2008**, *16*, 3482−3488.

for the reduced inhibitory potency of first-generation EGFR inhibitors.[111] Second-generation of EGFR inhibitors, such as afatinib (**2**, Figure 5.1), demonstrated a promising activity against T790M in preclinical models.[112] Thanks to presence of an acrylamide warhead capable of alkylating Cys797, afatinib is able to circumvent ATP competition and thus to overcome the unfavorable effect on the potency caused by the presence of methionine at the gatekeeper position. [113] Nevertheless, afatinib failed to overcome T790M-mediated resistance in the clinic due to dose-limiting toxicity resulting from inhibition of the wild-type (*WT*) form of EGFR.[114] Osimertinib (**3**, Figure 5.1)[115] is a third-generation EGFR inhibitor approved for patients affected by metastatic EGFR T790M mutation-positive NSCLC who have progressed on or after the first and second generation of EGFR TKI therapy.[116] As for other third-generation inhibitors WZ4002 (**4**) and Rociletinib (**5**), Osimbertinib possesses a 2-aminopirimidine scaffold which confers selectivity for the oncogenic forms of EGFR versus the *WT*, and an acrylamide group that alkylates Cys797 ensuring to it the ability to potently inhibit EGFR also in the presence of the T790M mutation.[117]

---

[111] Yun, C.-H.; Mengwasser, K. E.; Toms, A. V.; Woo, M. S.; Greulich, H.; Wong, K.-K.; Meyerson, M.; Eck, M. J. The T790M Mutation in EGFR Kinase Causes Drug Resistance by Increasing the Affinity for ATP. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2070−2075.

[112] Hirsh, V. Afatinib (BIBW 2992) development in non-small-cell lung cancer. *Future Oncol.* **2011**, *7*, 817−825.

[113] Carmi, C.; Mor, M.; Petronini, P. G.; Alfieri, R. Clinical Perspectives for Irreversible Tyrosine Kinase Inhibitors in cancer. *Biochem. Pharmacol.* **2012**, *84*, 1388−1399.

[114] Katakami, N.; Atagi, S.; Goto, K.; Hida, T.; Horai, T.; Inoue, A.; Ichinose, Y.; Koboyashi, K.; Takeda, K.; Kiura, K.; Nishio, K.; Seki, Y.; Ebisawa, R.; Shahidi, M.; Yamamoto, N. LUX-Lung 4: a phase II trial of afatinib in patients with advanced non-small-cell lung cancer who progressed during prior treatment with erlotinib, gefitinib, or both. *J. Clin. Oncol.* **2013**, *31*, 3335−3341.

[115] Finlay, M. R. V.; Anderton, M.; Ashton, S.; Ballard, P.; Bethel, P. A.; Box, M. R.; Bradbury, R. H.; Brown, S. J.; Butterworth, S.; Campbell, A.; Chorley, C.; Colclough, N.; Cross, D. A. E.; Currie, G. S.; Grist, M.; Hassall, L.; Hill, G. B.; James, D.; James, M.; Kemmitt, P.; Klinowska, T.; Lamont, G.; Lamont, S. G.; Martin, N.; McFarland, H. L.; Mellor, M. J.; Orme, J. P.; Perkins, D.; Perkins, P.; Richmond, G.; Smith, P.; Ward, R. A.; Waring, M. J.; Whittaker, D.; Wells, S.; Wrigley, G. L. Discovery of a Potent and Selective EGFR Inhibitor (AZD9291) of Both Sensitizing and T790M Resistance Mutations That Spares the Wild Type Form of the Receptor. *J. Med. Chem.* **2014**, *57*, 8249−8267.

[116] Ramalingam, S. S.; Yang, J. C.; Lee, C. K.; Kurata, T.; Kim, D. W.; John, T.; Nogami, N.; Ohe, Y.; Mann, H.; Rukazenkov, Y.; Ghiorghiu, S.; Stetson, D.; Markovets, A.; Barrett, J. C.; Thress, K. S.; Jänne, P. A. Osimertinib As First-Line Treatment of EGFR Mutation-Positive Advanced Non-Small-Cell Lung Cancer. *J. Clin. Oncol.* **2017**

[117] Cross, D. A.; Ashton, S. E.; Ghiorghiu, S.; Eberlein, C.; Nebhan, C. A.; Spitzler, P. J.; Orme, J. P.; Finlay, M. R.; Ward, R. A.; Mellor, M. J.; Hughes, G.; Rahi, A.; Jacobs, V. N.; Red Brewer, M.; Ichihara, E.; Sun, J.; Jin, H.; Ballard, P.; Al-Kadhimi, K.; Rowlinson, R.; Klinowska, T.; Richmond, G. H.; Cantarini, M.; Kim, D. W.; Ranson, M. R.; Pao, W. AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discovery* **2014**, *4*, 1046−1061.

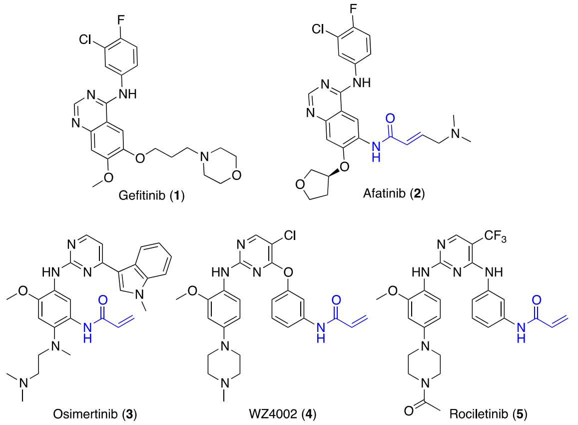


Figure 5.1. Structure of relevant EGFR inhibitors.

Development of novel forms of resistance is currently limiting the therapeutic benefit of osimertinib into the clinic.[118] C797S mutation, which replaces the cysteine with a significantly less nucleophilic serine, emerged as the main determinant of resistance to third generation EGFR inhibitors.[119] Other mutations, i.e., L718Q occurring in the P-loop and L844V, occurring in the ATP binding site, were initially reported to confer resistance to pyrimidine-based inhibitors, including compound **4** and rociletinib.[120] In the case of osimertinib, a moderate activity on cells harboring these mutations was maintained, suggesting that this inhibitor might be still clinically effective in patients harboring L718Q or L844V mutation. However, a cell proliferation assay conducted in Ba/F3 cells expressing L858R/T790M/L718Q EGFR mutant, showed that osimertinib resulted considerably less potent (~100-fold) at inhibiting cell growth than in Ba/F3 cells expressing either L858R/T790M double mutant or L858R/T790M/L844V triple mutant, suggesting that L718Q mutation likely affects the therapeutic activity of osimertinib. As a matter of fact, clinical resistance to osimertinib has recently emerged in a NSCLC patient expressing EGFR L858R/T790M double mutant who acquired L718Q mutation.[121]

---

[118] Engel, J.; Lategahn, J.; Rauh, D. Hope and Disappointment: Covalent Inhibitors to Overcome Drug Resistance in Non-Small Cell Lung Cancer. *ACS Med. Chem. Lett*. **2016**, *7*, 2–5.

[119] Thress, K. S.; Paweletz, C. P.; Felip, E.; Cho, B. C.; Stetson, D.; Dougherty, B.; Lai, Z.; Markovets, A.; Vivancos, A.; Kuang, Y.; Ercan, D.;Matthews, S.E.; Cantarini, M.; Barrett, J.C.; Jan̈ne, P. A.; Oxnard, G. R. Acquired EGFR C797S Mutation Mediates Resistance to AZD9291 in Non-Small Cell Lung Cancer Harboring EGFR T790M. *Nat. Med.* **2015**, *21*, 560−562

[120] Ercan, D.; Choi, H. G.; Yun, C. H.; Capelletti, M.; Xie, T.; Eck, M. J.; Gray, N. S.; Jänne, P. A. EGFR Mutations and Resistance to Irreversible Pyrimidine-Based EGFR Inhibitors. *Clin. Cancer Res.* **2015**, *21*, 3913-3923.

[121] Bersanelli, M.; Minari, R.; Bordi, P.; Gnetti, L.; Bozzetti, C.; Squadrilli, A.; Lagrasta, C. A.; Bottarelli, L.; Osipova, G.; Capelletto, E.; Mor, M.; Tiseo, M. L718Q Mutation as New Mechanism of Acquired Resistance to AZD9291 in EGFR-Mutated NSCLC. *J. Thorac. Oncol.* **2016**, *11*, e121-e123

L718Q thus emerged as novel mutation able to reduce osimertinib potency in vitro and to confer to NSCLC resistance in vivo. Starting from the visual inspection of the X-ray structure of EGFR-osimertinib complex,[122] it has been proposed that the replacement of a leucine with a glutamine at the position 718 of EGFR could reduce the affinity of EGFR for osimertinib and/or hinders Cys797 alkylation by acrylamide warhead. Considering that Gln718 occupies a peripheral position of the ATP binding site of EGFR (Figure 5.2), the precise role on how Gln718 hampers osimertinib activity remains largely unexplained.

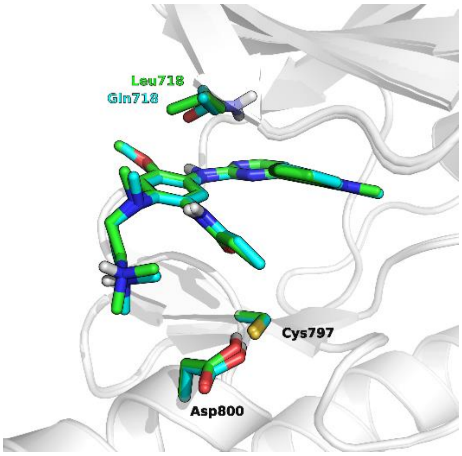


Figure 5.2. Superposition of X-ray structure derived models of osimbertinib in non-covalent complex with EGFR T790M (green carbon atoms) and EGFR T790M/L718Q mutant (cyan carbon atoms).

The availability of a computational protocol able to rationalize the influence of a mutation in the binding site of EGFR is of pivotal interest in cancer drug discovery,[123] as this may help the design of novel inhibitors able to circumvent these mutations.

In this chapter, the impact of L718Q mutation on the inhibitory activity of osimertinib by modelling both the *chemical step* (i.e., Cys797 alkylation) and the *recognition step* (i.e. formation of the non-covalent complex) was investigated, starting from X-ray[122] derived models of osimertinib in complex with EGFR T790M and EGFR T790M/L718Q mutants. The influence of the L718Q mutation was assessed estimating for both molecular systems *i)* the preeminent ionization state

---

[122] Yosaatmadja, Y.; Silva, S.; Dickson, J. M.; Patterson, A. V.; Smaill, J. B.; Flanagan, J. U.; McKeage, M. J.; Squire, C. J. Binding mode of the breakthrough inhibitor AZD9291 to epidermal growth factor receptor revealed. *J. Struct. Biol.* **2015**, *192*, 539-544.

[123] (a) Sutto, L.; Gervasio, F. L. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10616−10621. (b) Park, J.; McDonald, J. J.; Petter, R. C.; Houk, K. N. Molecular Dynamics Analysis of Binding of Kinase Inhibitors to WT EGFR and the T790M Mutant. *J. Chem. Theory Comput.* **2016**, *12*, 2066−2078.

for Cys797; *ii)* the energetics for Cys797 alkylation; *iii)* the free-energy of binding for the formation of the non-covalent complex, *iv)* changes in the conformational space of EGFR-osimertinib complexes. A wide range of computational approaches was applied to model all these phenomena, combining classic molecular dynamics (MD) simulations with umbrella sampling (US) and hybrid quantum mechanics molecular mechanics (QM/MM) calculations.

5.3 Results and discussion

5.3.1 Ionization state of Cys797

The reactivity of cysteines versus electrophilic compounds depends on the protonation state of their thiol group as indicated by the pH dependency of the reaction rate for covalent bond formation.[124] EGFR has a solvent exposed cysteine (Cys797) that is the target of covalent modification of second and third generation EGFR inhibitors. Cys797 thiol not only is readily alkylated by Michael acceptors but it easily transformed in sulfenic acid (i.e., Cys-S-OH) in presence of oxidative stimuli.[125] These evidences indicate that the Cys797 thiol may live in anionic form at physiological pH. This could be due to the presence of an organized microenvironment in EGFR able to stabilize its negative charge.[126] Furthermore, Cys797 side chain takes contacts with Asp800, with the sulfur atom placed at 3.3 Å from one of the oxygen of the carboxylic group in the EGFR-osimertinib structure.[122] Due to the proximity of these electronegative atoms, it is unlikely that both the –SH and the –COOH lives in anionic form at physiological pH, thus proposing Asp800-COO$^-$ as the natural acceptor of the Cys-SH proton (Figure 5.3).

---

[124] (a) Berteotti, A.; Vacondio, F.; Lodola, A.; Bassi, M.; Silva, C.; Mor, M.; Cavalli, A. Predicting the Reactivity of Nitrile-Carrying Compounds with Cysteine: a Combined Computational and Experimental study. *ACS Med. Chem. Lett.* **2014**, *5*, 501−505; (b) Flanagan, M. E.; Abramite, J. A.; Anderson, D. P.; Aulabaugh, A.; Dahal, U. P.; Gilbert, A. M.; Li, C.; Montgomery, J.; Oppenheimer, S. R.; Ryder, T.; Schuff, B. P.; Uccello, D. P.; Walker, G. S.; Wu, Y.; Brown, M. F.; Chen, J. M.; Hayward, M. M.; Noe, M. C.; Obach, R. S.; Philippe, L.; Shanmugasundaram, V.; Shapiro, M. J.; Starr, J.; Stroh, J.; Che, Y. Chemical and Computational Methods for the Characterization of Covalent Reactive Groups for the Prospective Design of Irreversible Inhibitors. *J. Med. Chem.* **2014**, *57*, 10072−10079.

[125] Schwartz, P. A.; Kuzmic, P.; Solowiej, J.; Bergqvist, S.; Bolanos, B.; Almaden, C.; Nagata, A.; Ryan, K.; Feng, J.; Dalvie, D.; Kath, J. C.; Xu, M.; Wani, R.; Murray, B. W. Covalent EGFR Inhibitor Analysis Reveals Importance of Reversible Interactions to Potency and Mechanisms of Drug Resistance. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 173−178.

[126] Truong, T. H.; Carroll, K. S. Redox Regulation of Epidermal Growth Factor Receptor Signaling Through Cysteine Oxidation. *Biochemistry.* **2012**, 51, 9954-9965.

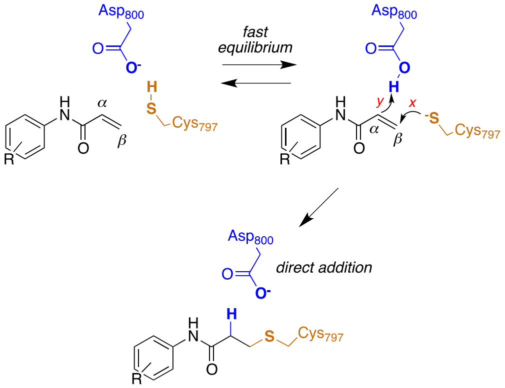


Figure 5.3. Scheme of the reaction mechanism of Cys797 alkylation for acrylamide-based inhibitors of EGFR.

On the base of these considerations, recently my research group proposed that the most abundant protonation state for this cysteine/aspartate pair of EGFR is Cys797−S⁻/Asp800−COOH. This hypothesis was supported by free-energy simulations showing that Cys797−S⁻/Asp800−COOH pair is significantly more stable than Cys797−SH/Asp800−COO⁻ one, at in least in the presence of a *N*-(4-anilinoquinazolin-6-yl) acrylamide inhibitor.[127] This proposal has been verified by titration experiments on *WT*-EGFR showing that Cys797 thiol has pKa of 5.5,[128] a value significantly lower compared to a free thiol in solution, and approaching the pKa of a carboxylic acid. I started my computational analysis evaluating the impact of the L718Q mutation on the pKa of Cys797. To this end, using the X-ray structure of osimertinib in non-covalent complex with *WT*-EGFR, molecular models of EGFR T790M and T790M/L718Q mutants in complex with this inhibitor were built, equilibrated by molecular dynamics (MD) simulations, and employed for computational studies. pKa calculations performed with PropKa (which use an empirical scoring function)[129] and H++ (which solves the Poisson-Boltzmann equation)[130] on different snapshots taken from MD trajectory indicated that L718Q mutation has a little effect on the acidity of Cys797 thiol group, as the pKa shift was

---

[127] Capoferri, L.; Lodola, A.; Rivara, S.; Mor, M. Quantum Mechanics/Molecular Mechanics Modeling of Covalent Addition between EGFR–Cysteine 797 and N-(4-Anilinoquinazolin-6-yl) Acrylamide. *J. Chem. Inf. Model.* **2015**, *55*, 589–599.

[128] Truong, T. H. Ung, P. M.; Palde, P. B.; Paulsen, C. E.; Schlessinger, A.; Carroll, K. S. Molecular Basis for Redox Activation of Epidermal Growth Factor Receptor Kinase. *Cell Chem. Biol.* **2016**, *23*, 837-848.

[129] Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.

[130] Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating pK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40*, W537−W541.

negligible with both methods (Table 5.1). Then, to verify if the L718Q mutation could somehow affect the prototropic equilibrium of Cys797/Asp800 pair, the free-energy difference (ΔΔG) between Cys797−S⁻/Asp800−COOH state and Cys797−SH/Asp800−COO⁻ was calculated. To do so, SCC-DFTB/AMBER QM/MM potential[131] was combined with umbrella sampling (US), starting from the molecular models of osimertinib in complex with EGFR T790M or EGFR T790M/L718Q variant and progressively moving the proton present on the carboxylic group of Asp800 to the thiolate group of Cys797 (details in the Computational protocol paragraph).

Table 5.1. ΔpKa for Cys797 calculated as difference between pKa computed in EGFR T790M and pKa computed in EGFR T790M/L718Q

	Propka	**H++**
frame1	-0.25	-0.25
frame2	-0.09	0.03
frame3	-1.77	-1.19
frame4	0.81	1.84
frame5	0.99	1.31
frame6	-1.41	-1.06
frame7	-0.13	0.82
frame8	-1.16	-1.15
frame9	1.76	2.14
frame10	1.44	2.21
average	**0.02 ± 1.22**	**0.47 ± 1.37**

Convergence of the US simulations was assessed calculating the difference in the free-energy between Cys797−S−/Asp800−COOH and Cys797−SH/Asp800−COO⁻ states as function of the time of simulation. This difference reached a stable value after 30 ps of simulation (Figure 5.4). US simulations show that the Cys797−S⁻/Asp800−COOH state resulted as stable as Cys797−SH/Asp800−COO⁻one for EGFR T790M (estimated difference of 0.5 kcal/mol) while in the case T790M/L718Q the Cys797−S⁻/Asp800−COOH state was more stable by nearly 2 kcal/mol. It thus appears that L718Q mutation stabilizes further the thiolate form of Cys797.

---

[131] Seabra, G. M.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E. Implementation of the SCC-DFTB Method for Hybrid QM/ MM Simulations within the Amber Molecular Dynamics Package. *J. Phys. Chem. A* **2007**, *111*, 5655−5664.

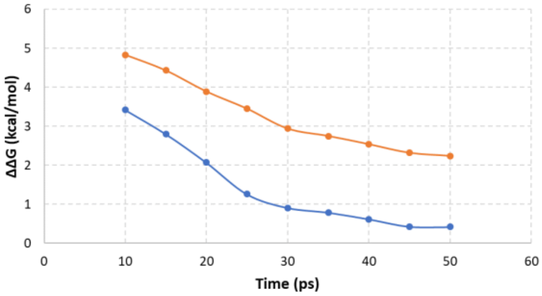


Figure 5.4. Free-energy difference (kcal/mol) between Cys797−S−/Asp800−COOH and Cys797−SH/Asp800−COO− states as function of the time of US simulation from data collected every 5ps for EGFR T790M (blue line) and EGFR T790M/L718Q (green line).

5.3.2 Chemical Step: reaction energetics for Cys797 alkylation.

Recent computational studies[127,132] indicate that acrylamide-based inhibitors of EGFR likely alkylate Cys797 following a *direct addition* mechanism in which the thiolate group of Cys797 attacks β carbon (Cβ) of the warhead, while the carboxylic acid group of Asp800 protonates the α carbon (Cα) leading to a stable 3-(methylsulfanyl)propanamide adduct (Figure 5.3). Using this reaction scheme as reference mechanism of alkylation also for the Michael-type addition involving osimertinib and Cys797, this key chemical process was modelled for both EGFR T790M and EGFR T790M/L718Q forms. To obtain useful information on the reaction energetics, the free-energy surface (FES) of the *direct addition* mechanism was reconstructed at SCC-DFTB/AMBER level coupled with US using intuitive and simple distances among reactive atoms as reaction coordinates, i.e. $S_{Cys797}{-}C\beta_{acrylamide}$ distance for the nucleophilic attack (event x), and $[(H_{asp800}{-}C\alpha_{acrylamide}) - (H_{asp800}{-}O_{asp800})]$ difference of distances for the protonation the Cα of the olefinic bond of acrylamide (event y). While SCC-DFTB is known to underestimate barriers, this approach has been satisfactorily applied to elucidate the effects of mutations on several enzyme-catalyzed reactions.[133]

---

[132] Paasche, A.; Schiller, M.; Schirmeister, T.; Engels, B. Mechanistic Study of the Reaction of Thiol-Containing Enzymes with α,β-Unsaturated Carbonyl Substrates by Computation and Chemoassays. *ChemMedChem* **2010**, *5*, 869−880.

[133] (a) Xu, D. G.; Guo, H. Quantum Mechanical/Molecular Mechanical and Density Functional Theory Studies of a Prototypical Zinc Peptidase (Carboxypeptidase A) Suggest a General Acid-General Base Mechanism. J. *Am. Chem. Soc*. **2009**, *131*, 9780−9788. (b) Lodola, A.; Branduardi, D.; Capoferri, L.; De Vivo, M.; Mor, M.; Piomelli, D. A Catalytic Mechanism for Cysteine N-terminal Nucleophile Hydrolases, as Revealed by Free energy simulations. *PLoS One* **2012**, *7*, e32397. (c) Pan, L. L.; Song, L. F.; Miao, Y.; Yang, Y.; Merz, K. M. Mechanism of Formation of the Nonstandard Product in the Prenyltransferase Reaction of the G115T Mutant of FtmPT1: A Case of Reaction Dynamics Calling the Shots? *Biochemistry* **2017**, 56, 2995-3007.

Analysis of the FES of a chemical process is useful to explain how changes in the microenvironment of the reaction could affect position and geometry of the transition state (TS), and thus the energy barrier of the investigated process.[134]
Figure 5.5A reports SCC-DFTB/AMBER FES describing alkylation of Cys797 by osimertinib for EGFR T790M while Figure 5.5B reports the SCC-DFTB/AMBER FES for the same biochemical process in presence of EGFR T790M/L718Q. In both cases, the minimum free-energy path connecting reactants (***R***, upper right corner) and the products (***P***, bottom left corner) follows a diagonal pathway indicating that nucleophilic attack and protonation of the incipient carbanion species are tightly coupled events. No carbanion/enolate species was indeed identified as a stable minimum during QM/MM-US simulation, consistently with previously reported DFT calculations describing addition of thiolate derivative on acrylamide performed on similar systems.[127,132] The activation free-energy (ΔA_{act}) at SCC-DFTB/AMBER level of theory resulted equal to 7.8 ± 0.2 kcal/mol for alkylation of EGFR T790M and of 8.1 ± 0.2 kcal/mol for alkylation of EGFR T790M/L718Q. The reaction free-energy (ΔA_{reac}) resulted highly negative, i.e., of -10.3 ± 0.2 and of -12.2 ± 0.2 kcal/mol for EGFR T790M and EGFR T790M/L718Q, respectively, indicating that Cys797 alkylation by osimertinib is highly exergonic and thus a spontaneous process.

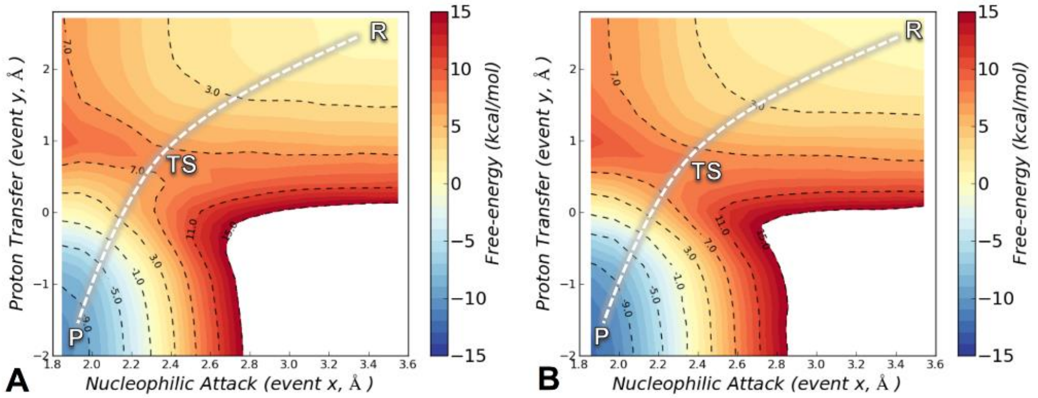


Figure 5.5 SCC-DFTB/AMBER FES for the Cys797 alkylation by osimertinib in the presence of EGFR 790M (panel A) or EGFR T790M/L718Q (panel B). The reaction coordinates (nucleophilic attack and proton transfer), are given in angstroms. Free energies are given in kcal/mol, and the contour levels are set at 2 kcal/mol.

Overall, these results indicate that L718Q mutation does not affect the energetics of the alkylation. Consistently with this finding, analysis of the geometries identified

---

[134] Jencks, W. P. General acid-base catalysis of complex reactions in water. *Chem. Rev.* **1972**, 72, 705–718.

along the minimum free-energy path connecting ***R*** to ***P*** indicates that L718Q mutation has a negligible impact on the transition state (TS) geometries obtained by US simulation (Table 5.2).

A first commonality shared by the reaction path followed by EGFR T790M and EGFR T790M/L718Q was represented by the conformation assumed by the two conjugated-double bonds of the acrylamide group which remained in s-*cis* during the entire alkylation process. A further analysis of the minimum free-energy path (Figure 5.6 and Figure 5.7), revealed that for both systems, the key event of the reaction was the nucleophilic attack of the Cys797 sulfur atom on the acrylamide Cβ, which required the whole desolvation of the thiolate anion to be completed.

Table 5.2. geometries of TSs for Cys797 alkylation identified with QM/MM calculations.

	EGFR T790M	**EGFR T790M/L718Q**
S-Cβ distance (Å)	2.41 ± 0.05	2.24 ± 0.07
S-Cβ-Cα angle (degrees)	118.60 ± 4.41	114.10 ± 3.74
H-Cα distance (Å)	1.85 ± 0.08	1.48 ± 0.04
n WAT	1.64 ±0.48	0.77 ± 0.58

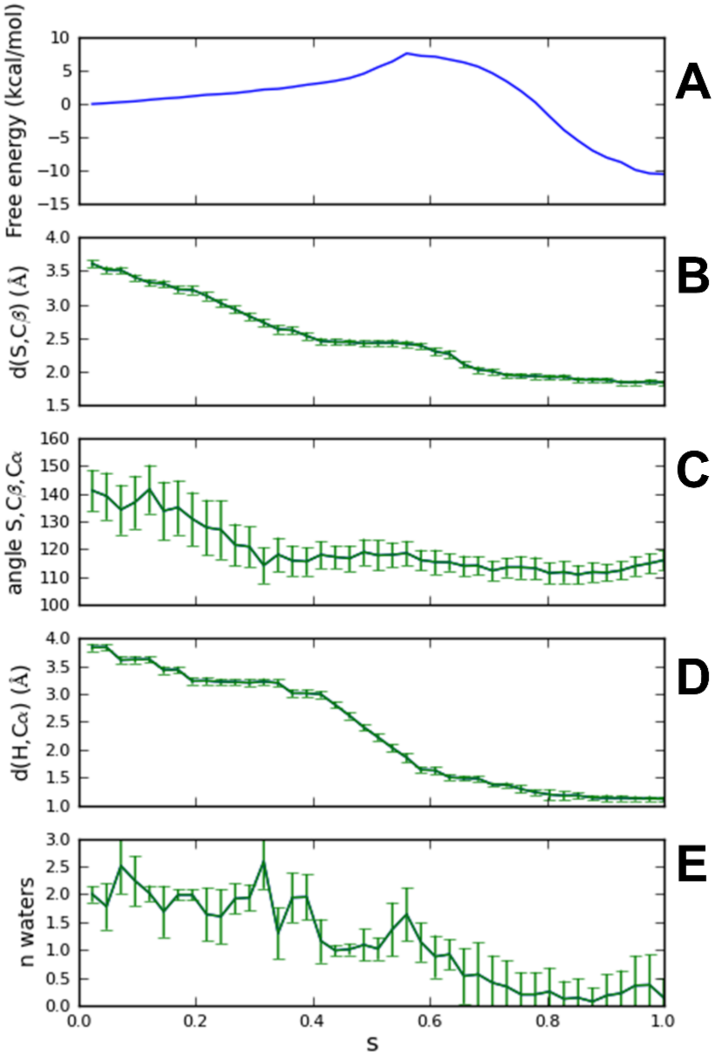


Figure 5.6. Analysis of the minimum free-energy path for Cys797 alkylation of EGFR T790M. Free energy profile of the reaction (Panel A). Evolution of the distance between S_{Cys797} and acrylamide Cβ along the path S (Panel B). Evolution of the angle between S_{Cys797}, acrylamide Cβ and acrylamide Cα along S (Panel C). Evolution of the distance between H_{Asp800} and acrylamide Cα along S (Panel D). Number of waters within 3.5 Å of the thiolate S_{Cys797} along S (Panel E). Values are represented as average with error bars representing the standard deviations.

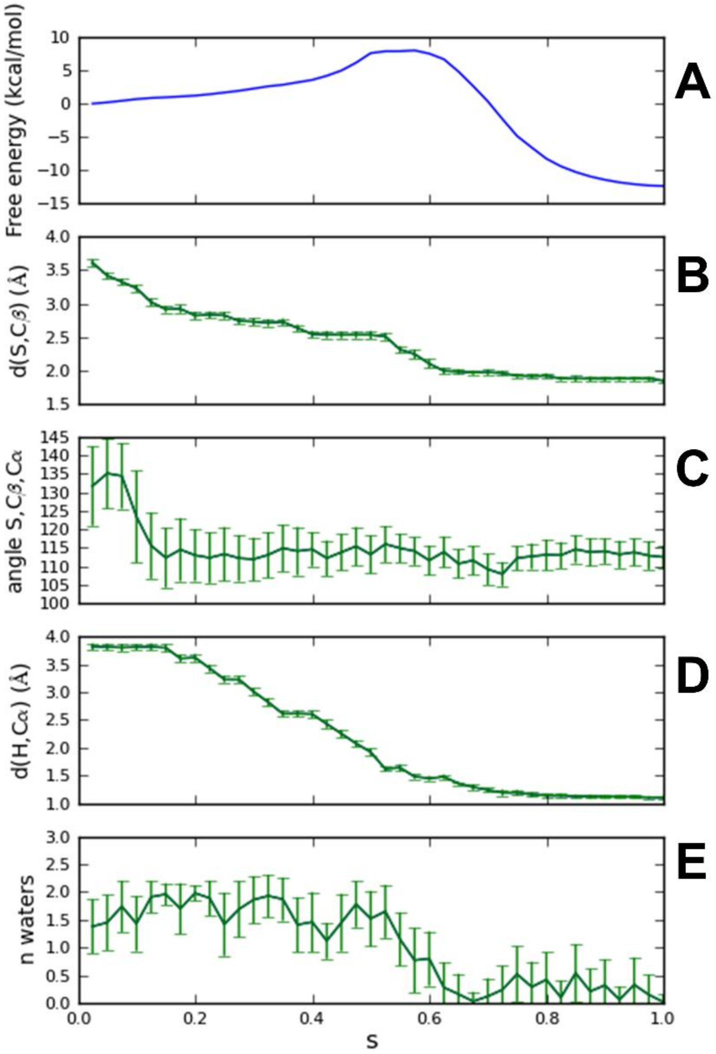


Figure 5.7. Analysis of the minimum free-energy path for Cys797 alkylation of EGFR T790M/L718Q. Free energy profile of the reaction (Panel A). Evolution of the distance between S_{Cys797} and acrylamide Cβ along the path S (Panel B). Evolution of the angle between $S_{Cys797,}$ acrylamide Cβ and acrylamide Cα along S (Panel C). Evolution of the distance between H_{Asp800} and acrylamide Cα along S (Panel D). Number of waters within 3.5 Å of the thiolate S_{Cys797} along S (Panel E). Values are represented as average with error bars representing the standard deviations.

In agreement with this finding, in the TS structures (Table 5.2) for Cys797 alkylation of EGFR T790M, the formation of S−Cβ bond was found to be the dominating event of the process considering its average distance of 2.41 ± 0.05 Å. Protonation of the Cα by Asp800 was instead only at its beginning at the TS, considering that the H−Cα distance resulted of 1.85 ± 0.08 Å (Figure 5.8A) at this stage of the reaction. In the case of EGFR T790M/L718Q, the TS structures were slightly more advanced toward the product as indicated by the S−Cβ and H−Cα distances of 2.25 ± 0.07 Å and 1.48 ± 0.04 Å, respectively (Figure 5.8B). These minor differences in the average geometries of the TS for the *direct addition* mechanism make reasons of the negligible difference (0.3 kcal/mol) in the computed activation free-energy. Similar geometries were found also for products of the reaction. The thioether adduct ***P*** was characterized by S−Cβ and H−Cα distances of 1.83 ± 0.03 Å and 1.11 ± 0.04 Å, respectively for EGFR T790M and of 1.84 ± 0.02 Å and 1.10 ± 0.03 Å, respectively.

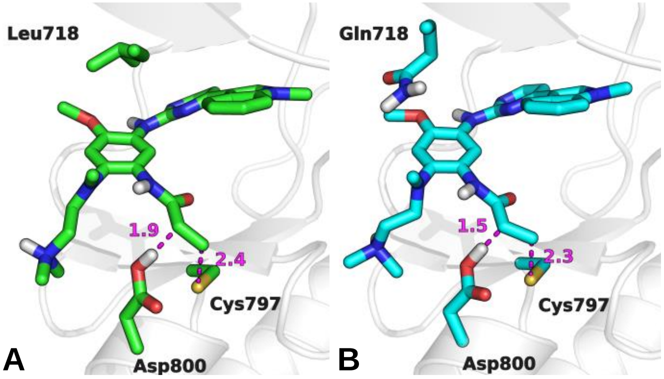


Figure 5.8. Representative geometry of the TS structure for Cys797 alkylation in the case of EGFR T790M (panel A, green carbon atoms) and EGFR T790M/L718Q (panel B, cyan carbon atoms). The secondary structure of EGFR is represented with white cartoons. Key distances are depicted with a dashed magenta line.

5.3.3 Recognition Step: Binding of osimertinib to EGFR mutants

It has been recently proposed that the main effect of the replacement of Leu718 with Gln718 is to disrupt the beneficial hydrophobic interactions involving the methoxyphenyl moiety[135] present in third generation EGFR inhibitors (Figure 5.2). I then tested this hypothesis by calculating the binding free-energy (ΔA_{bind}) of osimertinib for EGFR T790M and EGFR T790M/L718Q, using the recently reported

---

[135] Yan, X. E.; Zhu, S. J.; Liang, L.; Zhao, P.; Choi, H. G.; Yun C. H. Structural basis of mutant-selectivity and drug-resistance related to CO-1686. *Oncotarget* **2017**, *8*, 53508-53517.

WaterSwap methodology[136] in combination with classic MD simulations performed using AMBER99SB force field.[137] WaterSwap estimates ΔA_{bind} using a reaction coordinate that switches the ligand of interest with an equivalent volume of water in the protein binding site.[136] While this methodology tends to overestimate the ΔA_{bind} value, it is able to accurately capture the effect of mutations occurring near or in the binding site.[138] To obtain a robust estimation of the binding free-energy, five independent WaterSwap calculations were performed for each system, employing five different protein-ligand configurations taken from a 300 ns-long MD simulations, as starting point for the simulation. WaterSwap simulations (Table 5.3) indicate that osimertinib has similar ΔA_{bind} for EGFR T790M (-34.07 ± 2.79 kcal/mol) and EGFR T790M/L718Q (-36.85 ± 3.03 kcal/mol), suggesting that L718Q mutation does not affect the inhibitor affinity.

Table 5.3. ΔA_{bind} (kcal/mol) of Osimertinib calculated with WaterSwap

	EGFR T790M	**EGFR T790M/L718Q**
frame1	-30.59	-39.33
frame2	-36.62	-38.86
frame3	-35.99	-33.62
frame4	-31.24	-38.83
frame5	-35.33	-33.65
frame6	-34.08	-39.98
frame7	-35.80	-41.94
frame8	-35.67	-29.61
frame9	-35.77	-24.13
frame10	-37.01	-27.87
average	-34.87 ± 0.68	-34.78 ± 1.89

Finally, the effect of L718Q mutation on the conformational space of the EGFR-osimertinib complex was investigated. In first place, the influence of the mutation was analized on *i)* the binding of the 2-aminopirimidine scaffold of osimbertinib to EGFR hinge region within the ATP binding pocket, *ii*) the accommodation of the

---

[136] Woods, C. J.; Malaisree, M.; Hannongbua, S.; Mulholland, A. J. A water-swap reaction coordinate for the calculation of absolute protein-ligand binding free energies. *J. Chem. Phys.* **2011**, *134*, 054114.

[137] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712−725.

[138] Woods, C. J.; Malaisree, M.; Long, B.; McIntosh-Smith, S.; Mulholland, A. J. Computational assay of H7N9 influenza neuraminidase reveals R292K mutation reduces drug binding affinity. *Sci. Rep.* **2013**, *3*, 3561.

acrylamide warhead. To this end, four independent extensive 300ns-long MD simulations of osimertinib in non-covalent complex with EGFR T790M or with EGFR T790M/L718Q were performed, for a total simulation time of 1.2 µs for each molecular system, using AMBER99SB force field. The analysis of the trajectories of a first set of 300 ns of MD simulation (Figure 5.9) indicated that L718Q mutation, has only a small effect on the overall stability of EGFR. Indeed, the root-mean square deviation (RMSD) of protein backbone was of 1.60 ± 0. 49 Å for EGFR T790M and 2.20 ± 0.31 Å for EGFR T790M/L718Q. On the other hand, a significant increment in the flexibility of the P-loop (i.e. the glycine-rich stretch in which the mutation is located) was observed when Leu718 was replaced by the Gln residue, as indicated by the P-loop backbone RMSD values of 1.99 ± 0. 49 Å and 3.50 ± 0. 85 Å for EGFR T790M and EGFR T790M/L718Q, respectively. The increment of flexibility of the P-loop did not affect the overall stability of osimertinib. In the MD simulation, the RMSD for osimertinib heterocycle core resulted equal to 1.13 ± 0. 32 Å for EGFR T790M and 1.62 ± 0. 49 Å for EGFR T790M/L718Q double mutant. Also, the distance of the key H-bond undertaken by the N3-pyrimidine nitrogen of osimertinib and backbone N-H group of Met793 remains essentially the same i.e., 2.19 ± 0. 21 Å for EGFR T790M and 2.11 ± 0. 15 Å for EGFR T790M/L718Q double mutant. Similar conclusions were drawn analyzing RMSD time series for all the other replicas (Figure 5.10, 5.11, 5.12).

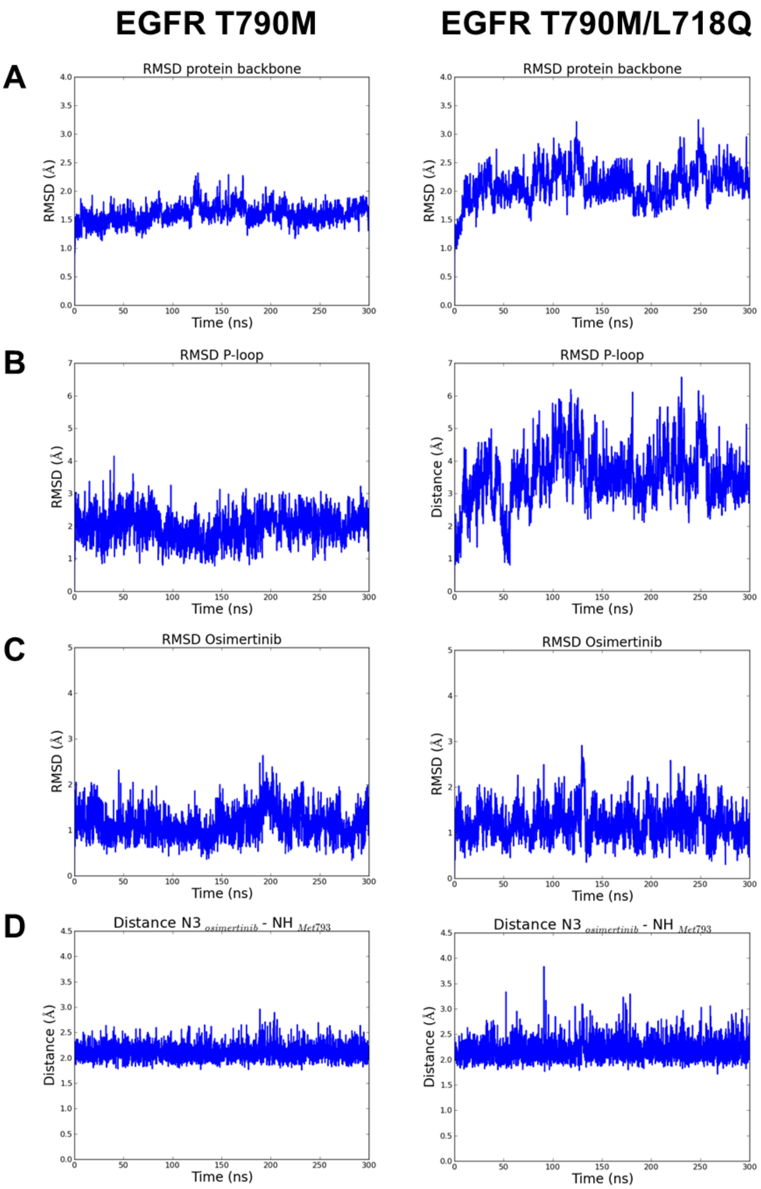


Figure 5.9. Analysis of MD trajectory of replica 1. RMSD time series of protein backbone (panel A), RMSD time serie of P-loop backbone (panel B), RMSD time series of osimertinib heterocycle core (panel C), time series of the distance between the N3-pyrimidine nitrogen of osimertinib and backbone N-H group of Met793 (panel D).

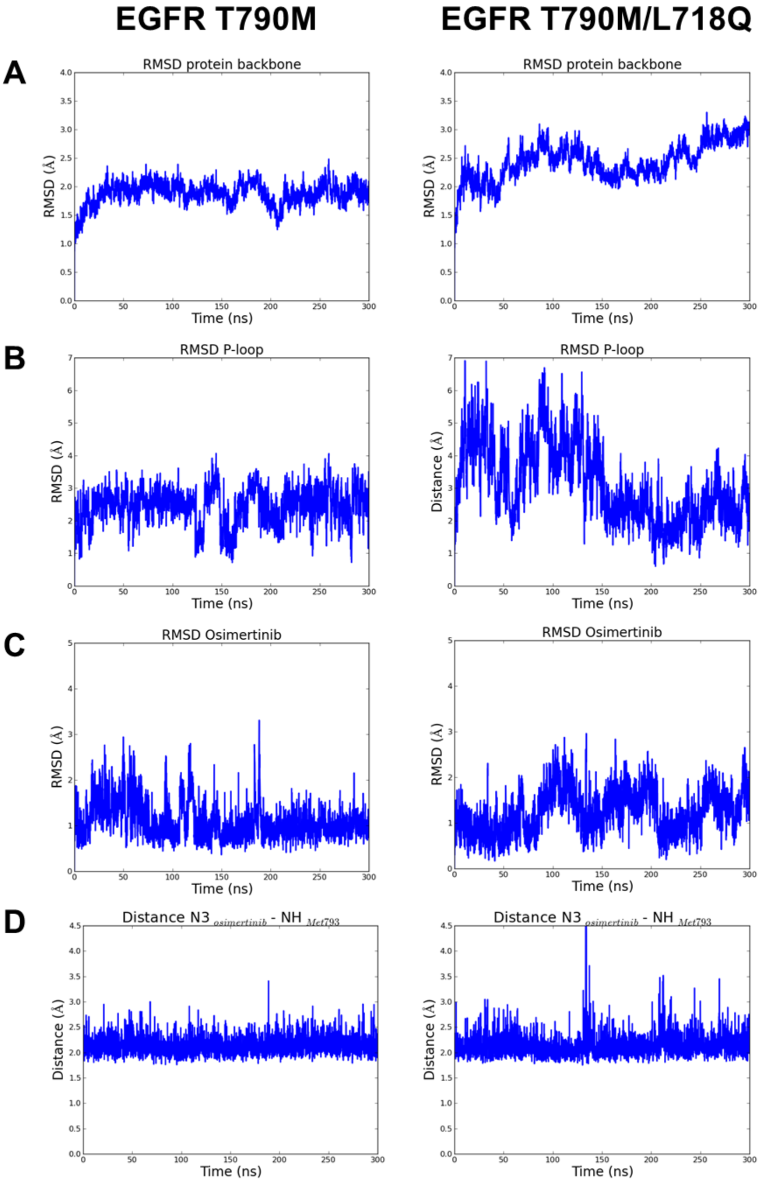


Figure 5.10. Analysis of MD trajectory of replica 1. RMSD time series of protein backbone (panel A), RMSD time serie of P-loop backbone (panel B), RMSD time series of osimertinib heterocycle core (panel C), time series of the distance between the N3-pyrimidine nitrogen of osimertinib and backbone N-H group of Met793 (panel D).

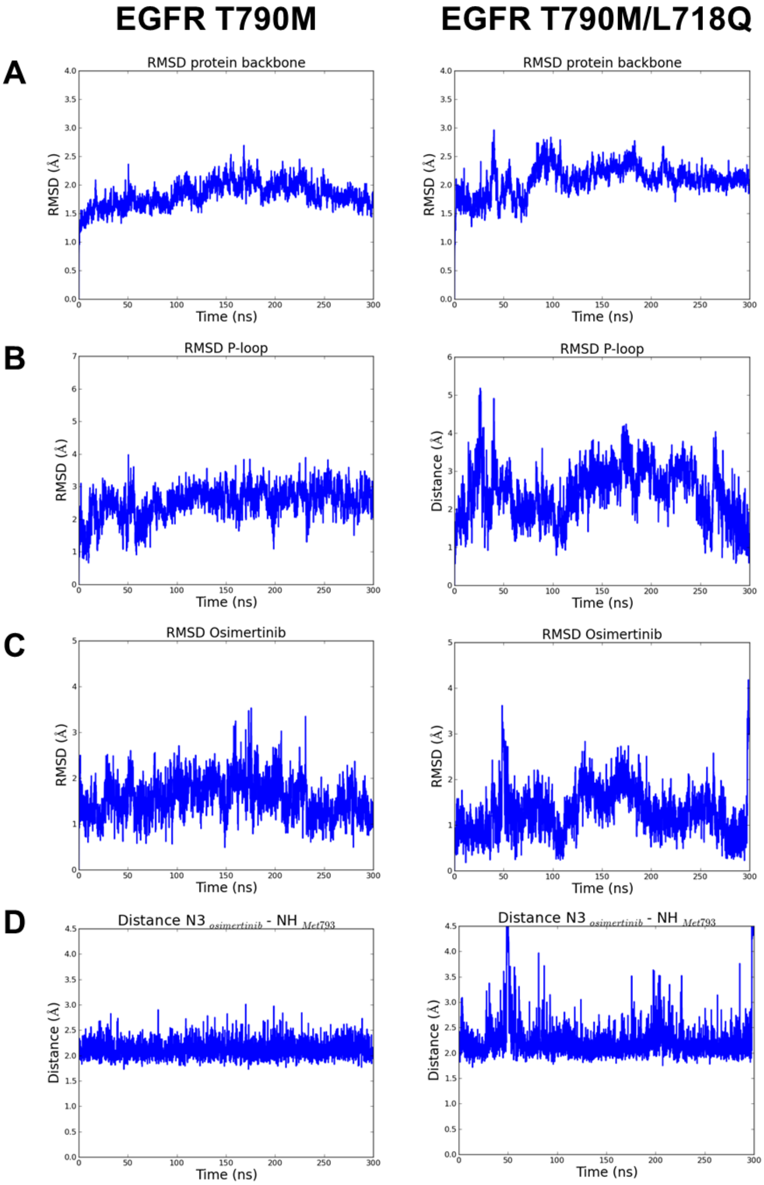


Figure 5.11. Analysis of MD trajectory of replica 1. RMSD time series of protein backbone (panel A), RMSD time serie of P-loop backbone (panel B), RMSD time series of osimertinib heavy atoms (panel C), time series of the distance between the N3-pyrimidine nitrogen of osimertinib and backbone N-H group of Met793 (panel D).

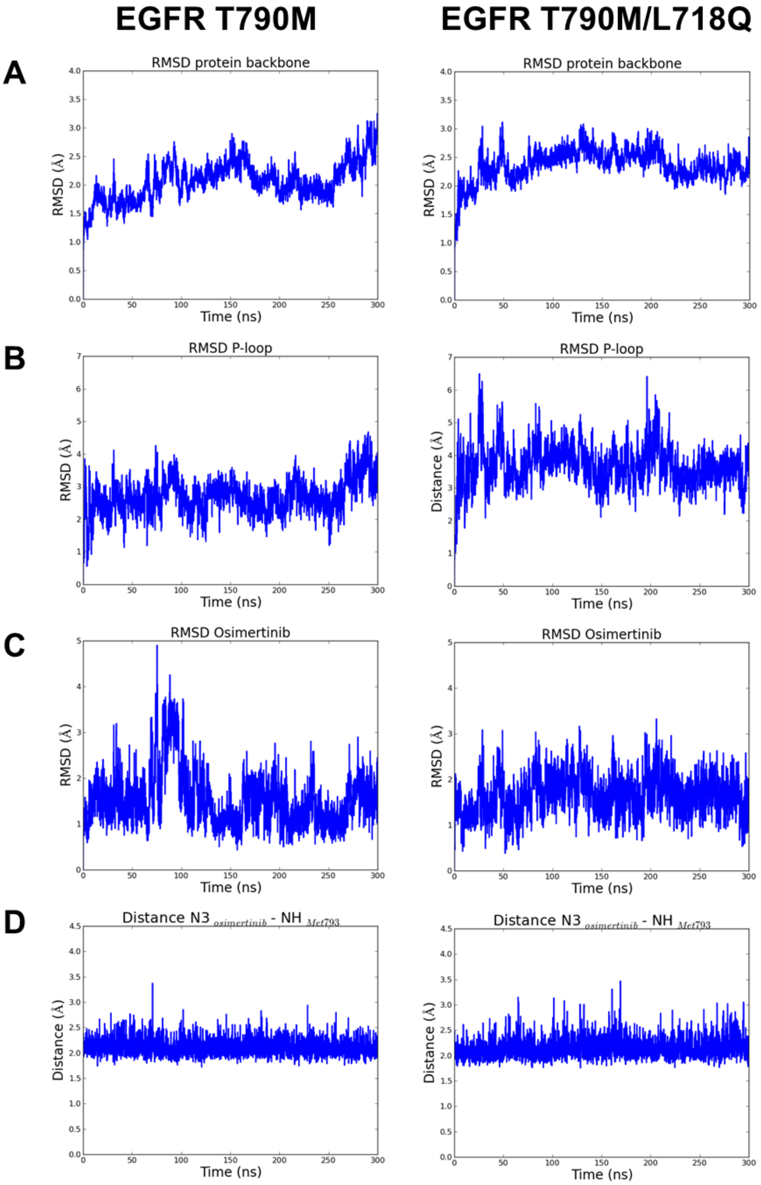


Figure 5.12. Analysis of MD trajectory of replica 1. RMSD time series of protein backbone (panel A), RMSD time serie of P-loop backbone (panel B), RMSD time series of osimertinib heavy atoms (panel C), time series of the distance between the N3-pyrimidine nitrogen of osimertinib and backbone N-H group of Met793 (panel D).

A further analysis of the MD simulations showed that the presence of a Gln instead of Leu in the P-loop had a barely significant effect on the average distance between Cys797 sulfur atom and the Cβ of the acrylamide, which moves from 4.9 ± 0.90 Å for EGFR T790M to 6.0 ± 1.0 Å for EGFR T790M /L718Q. However, a detailed analysis of the MD trajectories indicated that while for EGFR T790M the fraction of "reactive snapshots", i.e. configurations in which the nucleophile (Cys797 S) and electrophile (acrylamide Cβ) are separated by a distance smaller than the sum of the Van der Waals radius[42] (3.9 Å with AMBER99SB force field), corresponds to the 20.5 % of the overall populations of collected states, while in the case of EGFR T790M/L718Q double mutant, only the 0.7 % of collected snapshots places Cys797 sulfur and acrylamide Cβ at a distance shorter than 3.9 Å. Similar results were obtained in other MD replicas (Table 5.4).

Table 5.4. Percentage of configurations in which the nucleophile (S_{Cys797}) and electrophile ($C\beta_{acrylamide}$) are separated by a distance smaller than 3.9 Å in MD trajectories.

	EGFR T790M	**EGFR T790M/L718Q**
replica1	20.47	0.73
replica2	14.37	0.07
replica3	12.07	0.07
replica4	5.9	0.73

Additional analysis of the simulations, indicates that the distribution of the S-Cβ distance depends on the rotatable bond connecting the nitrogen of the acrylamide to the position 6 carbon of the central ring of osimertinib (Figure 5.13). In details, the analysis of the time series of S-Cβ distance and the C_1-C_2-N_1-C_3 dihedral angle, describing the rotation along the above-mentioned C-N bond, indicates that these two geometrical descriptors are highly correlated (Figure 5.14).

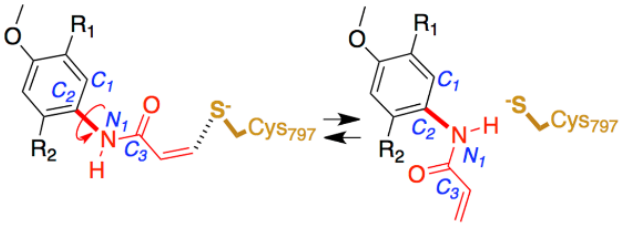


Figure 5.13. Definition of C_1-C_2-N_1-C_3 dihedral angle.

Thus, when the S-Cβ distance approached the critical value of 3.9 Å the dihedral angle assumed an average value -60° ± 10, close to the value observed in the X-ray structure, while when S-C_β distance overcame the value of 5.0 Å, a nearly 180° shift in the C_1-C_2-N_1-C_3 dihedral angle was observed. This trend was observed both in

EGFR T790M and in EGFR T790M/L718Q although with some crucial differences between the two systems. The time series reported in Figure 5.14 suggest the existence of at least two minima for EGFR-osimertinib non-covalent complexes.

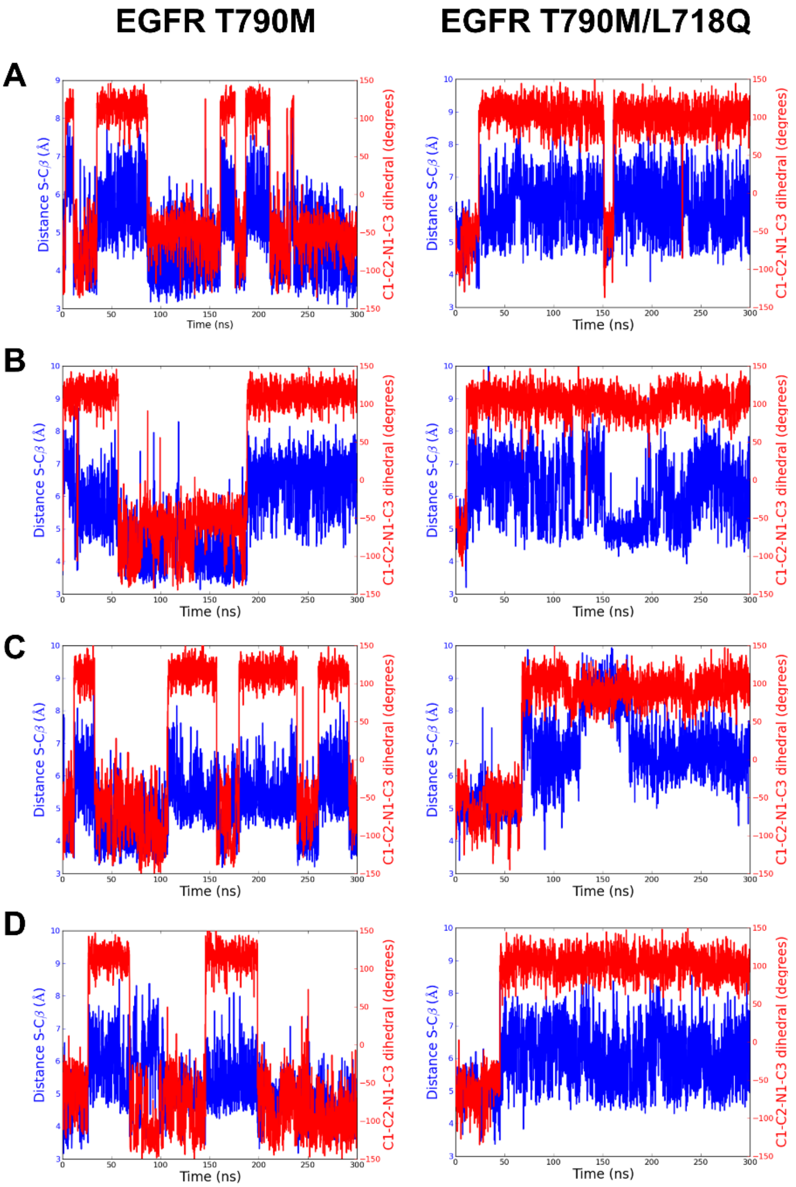


Figure 5.14. Time series for the S-Cβ distance (blu line) and the C_1-C_2-N_1-C_3 dihedral (red line) for EGFR T790M and EGFR T790M/L718Q in replica 1 (panel A), replica 2 (panel B), replica 3 (panel C) and replica 4 (panel D).

Using data from 4 independent 300ns-long MD simulations (for a total of 1.2 μs of simulation), conformational 2D FESs in the space of S-Cβ distance and C_1-C_2-N_1-C_3 dihedral angle were built. The resulting FESs indicate the existence of two alternative basins (labelled as ***a*** and ***b***) for both EGFR T790M (Figure 5.15 A) and EGFR 790M/L718Q (Figure 5.15 B) mutants.

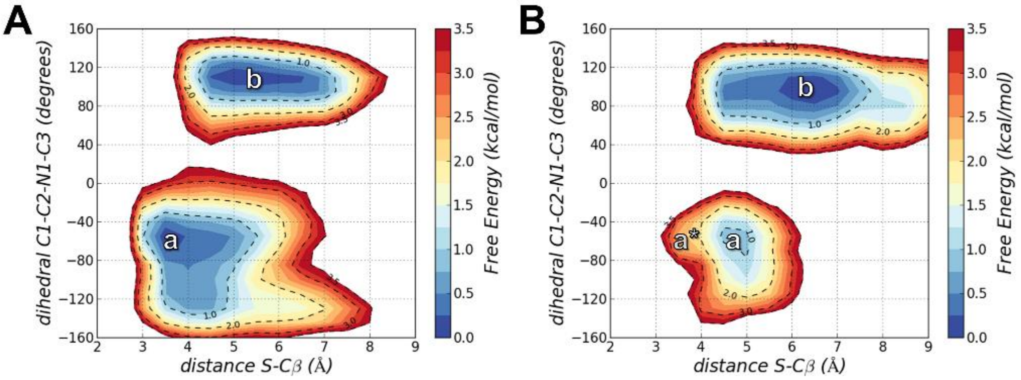


Figure 5.15. SCC-DFTB/AMBER FES for the Cys797 alkylation by osimertinib in the presence of EGFR 790M (panel A) or EGFR T790M/L718Q (panel B). The reaction coordinates (nucleophilic attack and proton transfer), are given in angstroms. Free energies are given in kcal/mol, and the contour levels are set at 2 kcal/mol.

In the case of EGFR T790M, the region of the FES described by ***a*** identifies a well-defined minimum centered at a S-Cβ distance of 3.6 Å and a C_1-C_2-N_1-C_3 dihedral angle of -60°. This specific arrangement of the reactants well superposes to the QM/MM geometries of the EGFR T790M-osimertinib complex belonging to basin ***R*** of Figure 5.5. Region ***a*** thus contains "reactive conformations", i.e., conformations extending from the ground state of the complex that lie on the transition path leading to the alkylation of Cys797. Region ***b*** identified a wider minimum corresponding to unreactive conformations of the EGFR-osimertinib complex, in which the S-Cβ distance was dispersed over a 4.5-6.5 Å interval due to a dramatic change in C_1-C_2-N_1-C_3 dihedral angle which has assumed values ranging from +100° to +120°. According to the calculated FES, ***a*** and ***b*** possess nearly the same free-energy and thus they can be regarded as equally populated conformational states.

In the case of EGFR T790M/L718Q, dramatic changes were observed in the position of minima ***a*** and ***b*** and in their relative free-energy and population. Firstly, minimum ***a*** does not anymore describe reactive conformations. Configurations laying in this area of the FES are featured by a S-Cβ distance approaching the value of 5.0 Å and thus hardly capable of leading to Cys797 alkylation. Minimum ***b*** describes unreactive conformations where the S-Cβ distance assumed even higher values (5.5-7.0 Å range) while the C_1-C_2-N_1-C_3 dihedral angle remained closed to the interval ranging from +90° to +110°. Basin ***b*** represents the global minimum of the computed FES for the EGFR T790M/L718Q system, resulting more stable than basin ***a*** by ~ 1 kcal/mol. The group of reactive geometries identified on the FES of Figure 5B by ***a**** are higher in energy than basin **b** by ~ 3 kcal/mol and can be regarded as a low populated state. The pronounced stabilization of basin ***b*** in the case of the EGFR T790M/L718Q mutant arises from the formation of a key H-bond undertaken by the polar side chain of Gln718 and the carbonyl oxygen of the acrylamide group of osimertinib (Figure 5.16) which is maintained for a significant period of the MD simulation.

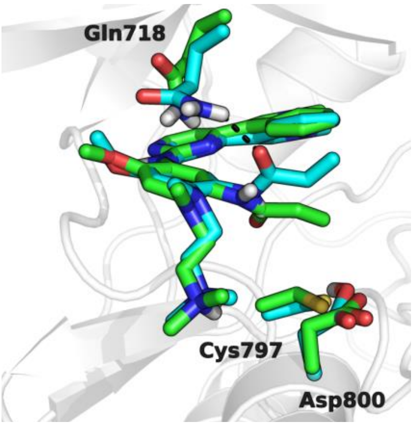


Figure 5.16. Representative minimum geometries structure for EGFR T790M/L718Q – osimertinib complexes taken from basin ***a**** (green carbon atoms) and ***b*** (cyan carbon atoms) of Figure 5.5B. The H-bond involving Gln718 and the acrylamide warhead is depicted with a dashed black line. The secondary structure of EGFR is represented with white cartoons.

Analysis of the time series comparing the evolution of the H-bond distance with that of the C_1-C_2-N_1-C_3 dihedral angle (Figure 5.17) and the S-Cβ distance (Figure 5.18), indicates that the formation of the H-bond is the key driver leading to the stabilization of the unreactive states for EGFR T790M/L718Q – osimertinib complex. The electrostatic attraction generated by Gln718 sidechain and felt by the carbonyl oxygen of acrylamide drives the transition along C_1-C_2-N_1-C_3 dihedral angle and places the electrophile Cβ far away from Cys797 thiolate. This H-bond mediated stabilization of a non-reactive formation emerged as a straightforward strategy to hamper the formation of a novel covalent bond thus to escape irreversible inhibition.

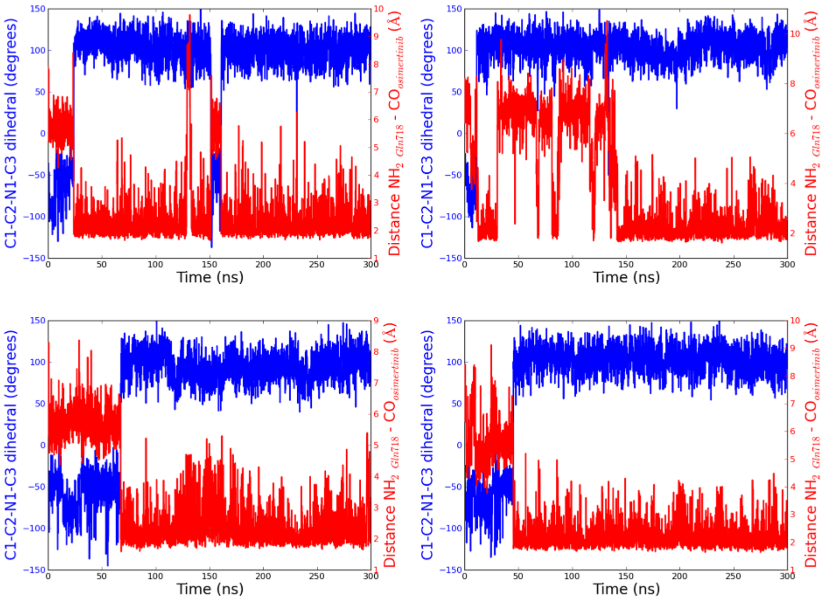


Figure 5.17. Time series comparing the evolution of the C1-C2-N1-C3 dihedral (blue line) with the evolution of the H-bond distance undertaken by the polar side chain of Gln718 and the carbonyl oxygen of the acrylamide group of osimertinib (red line) for each of the four MD replicas.

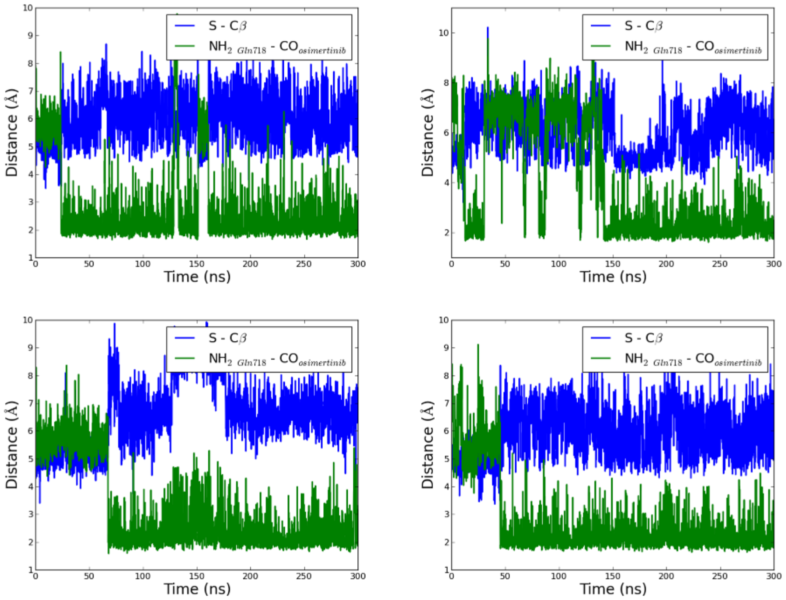


Figure 5.18. Time series comparing the evolution of the S-Cβ distance (blue line) with the evolution H-bond distance undertaken by the polar side chain of Gln718 and the carbonyl oxygen of the acrylamide group of osimertinib (green line) for each of the four MD simulations.

5.4 Conclusions

The clinical use of third generation of EGFR inhibitors is hampered by the insurgence of novel mutations which confers lung cancer cells the ability to escape EGFR inhibition and thus death. Among the recently reported mutations, L718Q emerged as an intriguing mutation considering that this occurs in a region of EGFR peripheral to the ATP binding site. Previous proposal based on X-ray visual inspection hypothesized that Gln718 somehow hinders Cys797 alkylation.[120,121,135] In this chapter, through a multilevel approach, computational evidences show that L718Q mutation does not affect the ionization state of the Cys797 nor the activation free-energy for Cys797 alkylation. Free-energy calculations indicate that L718Q mutation does not affect either the affinity of osimertinib for EGFR. Simulations instead show the Gln718 through a simple H-bond stabilizes a specific and yet "non-reactive" conformation of osimertinib in which the electrophilic Cβ of its acrylamide substituent is kept away from Cys797 thiolate. This stabilization likely hampers Cys797 alkylation and may thus favor the displacement of osimertinib from EGFR by ATP, which in cells is present at mM concentration.[11] This global picture makes reason of the poor inhibitory activity of osimertinib in Ba/F3 cells harboring L718Q mutation[120] and the lack of therapeutic activity of osimertinib in patients who have acquired this mutation.[121] The protocol designed in this chapter (i.e. *i.* evaluation of the ionization state of the nucleophile, *ii.* estimation of the free-energy barrier for covalent adduct formation, *iii.* binding affinity for the formation of the non-covalent complex, *iv.* mutual orientation of the reactant within the target active site), may represent a viable and generalizable strategy to investigate the effects of mutations on covalent inhibition *in silico*. Information gathered here could be exploited to design a new generation of EGFR inhibitors overcoming L718Q mutation, i.e., analogues of osimertinib in which the rotation along the C-N bond, connecting the warhead to the central phenyl ring, is hindered are expected to inhibit also this EGFR variant.

5.5 Computational protocol

Model Building and equilibration of the systems.

EGFR T790M and EGFR T790M/L718Q in complex with osimertinib were built starting from *WT* EGFR-osimertinib crystal structure (4ZAU.pdb)[122] using t-leap available in AMBER 16 (AMBER 2016, University of California, San Francisco, CA). The resulting complexes were immersed in a box of TIP3P water molecules[139] and neutralized with 3 Cl^- ions by using the t-leap tool implemented in AMBER16. The total system size amounted to 46936 atoms (14122 TIP3P waters) in the case of EGFR T790M – osimertinib complex and 46283 atoms in the case of EGFR T790M/L718Q – osimertinib complex (13905 waters). The solvated complexes were energy-minimized and gradually heated to 300 K in NVT ensemble and equilibrated at pressure of 1 atm in NPT ensemble. These systems were then submitted to a short MD simulation (10 ns) in the NVT ensemble. In all these simulations, the AMBER99SB force field was applied to describe protein atoms while the generalized Amber force field (GAFF)[140] was employed to describe osimertinib atoms. The *pmemd.cuda* module of AMBER16 was used to perform these simulations on NVIDIA K20 GPU cards. Full electrostatic and van der Waals interactions were computed within a cutoff of 10 A° and long- range electrostatic interactions were treated using the particle mesh Ewald (PME). The covalent bonds involving hydrogens were constrained using SHAKE function allowing the use of a time-step equal to 2 fs. The final structures were employed to estimate pKa for Cys797, for QM/MM simulations, and to perform longer MD simulations.

pKa calculations.

Ten equally spaced snapshots were taken from the 10 ns-long MD trajectories for EGFR T790M–osimertinib and EGFR T790M/L718Q–osimertinib complexes. The structures of the complexes were exported as PDB file and submitted to pKa calculations with H^{++} and Propka3.0 software using default settings.

---

[139] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

[140] Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

*Application of the QM/MM Potential***.**

In this chapter, the self-consistent charge - density functional tight binding (SCC-DFTB)[141] model was applied to describe the QM region and the AMBER99SB force field to describe the MM region. The SCC-DFTB approach is based on the second-order expansion of the total DFT energy with respect to the charge-density variation. Hybrid QM/MM potentials derived from the SCC-DFTB theory have been widely applied to enzyme catalysis and reported to give fair descriptions of reaction geometries and energetics.[142] In the EGFR-osimertinib complexes, side chain atoms of Cys797, Asp800 and the acrylamide portion of the inhbitor were treated with the SCC-DFTB method with dispersion correction. All the other atoms of the system were described with AMBER99SB force field. The resulting QM system was composed of 23 atoms including three link atoms, placed along the C−C bond connecting Cβ of Cys797 and Asp800 to their backbone Cα and along C-N bond connecting the -NH(CO)=CH_2 acrylamide group to central phenyl ring of osimertinib. The *adjust_q* function was applied to conserve the charge of the systems.

US simulation for Cys797 deprotonation by Asp800.

A set of US simulation was performed by adding a spring constant of 100 kcal mol^{-1} Å^{-2} along [(H_{asp800}− O_{asp800}) - (H_{asp800}− S_{cys797})] difference of distances as reaction coordinate ***r***. ***r*** was sampled from 1 to -1 with a step-size of 0.1 Å for a total of 21 windows. For each window, 50 ps of QM/MM MD simulation at 300 K in NVT conditions were performed for a total of 1 ns of simulation, using the *sander.MPI* module of AMBER16. All the atoms of the system (including hydrogens) were allowed to move during the simulations, and a time step of 0.2 fs was used to integrate the equation of motion. PME approach was used to treat the QM/MM long-range electrostatic interactions. The unbiased FES was calculated by using the weighted histogram analysis method (WHAM).[143] Convergence of the simulations was evaluated calculating the difference in the free-energy between Cys797−S−/Asp800−COOH and Cys797−SH/Asp800−COO^- states as function of the time of simulation.

---

[141] Elstner, M. The SCC-DFTB Method and its Application to Biological systems. *Theor. Chem. Acc.* **2006**, *116*, 316−325.

[142] Gruden, M.; Andjeklović, L.; Jissy, A. K.; Stepanović, S.; Zlatar, M.; Cui, Q.; Elstner M. Benchmarking Density Functional Tight Binding Models for Barrier Heights and Reaction Energetics of Organic Molecules. *J. Comput. Chem.* **2017**, *38*, 2171-2185.

[143] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free- Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992,** *13*, 1011−1021.

US simulation for Cys797 alkylation.

US simulations were carried out adding a spring constant of 100 kcal mol^{-1} Å^{-2} along the reaction coordinate x, described by (S_{Cys797}–$C\beta_{acrylamide}$) distance and reaction coordinate y described by [(H_{asp800}–$C\alpha_{acrylamide}$) - (H_{asp800}–O_{asp800})] difference of distance. Coordinate x was sampled from 3.6 to 1.8 Å with a step-size of 0.1 Å for a total of 19 windows while coordinate y was sampled from 2.8 to -2.0 Å with a step-size of 0.2 Å for a total of 26 windows. Overall the simulated grid is composed by 494 windows. For each one, 50 ps of QM/MM MD simulation at 300 K in NVT conditions were performed for a total 24.7 ns. The same settings described for Cys797-Asp800 proton transfer were applied. The presence of a normal distribution of the sampled data along x and y was verified within each simulated window. The employed spring constant allowed a satisfactory overlapping of the sampled windows, which is critical to obtain a reliable and continuous FES performed for each umbrella. The unbiased FES was calculated by using the weighted histogram analysis method (WHAM). Convergence of the simulations was evaluated calculating the ΔA_{act} and ΔA_{react}.as function of the time of simulations. These two observables reached stable values within 30 ps of QM/MM MD simulation for each window. Average values and standard deviations of ΔA_{act} and ΔA_{react} reported in the previous paragraphs were calculated from data collected at 30, 40 and 50 ps of simulations for each window for both EGFR-osimertinib complexes (Figure 5.18).

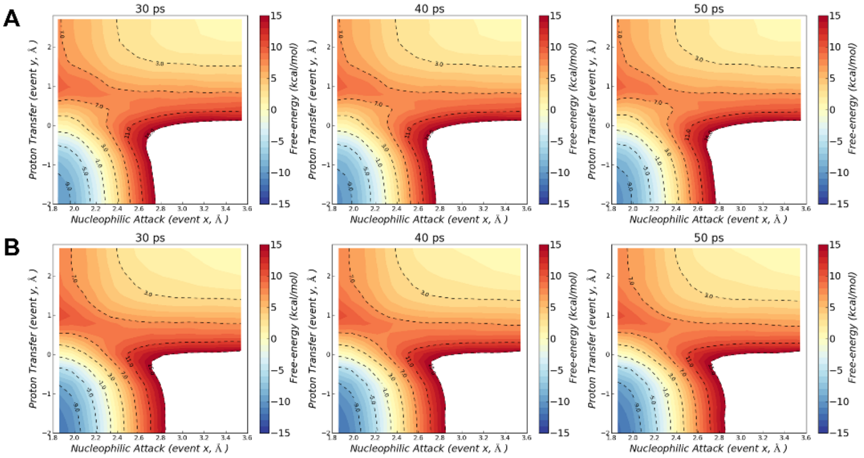


Figure 5.18. Free-energy surfaces calculated from data collected at 30, 40 and 50 ps of simulations for each window for Osimertinib-EGFR T790M complex (panel A) and for Osimertinib-EGFR T790M/L718Q complex (panel B).

The US simulations were then replicated to evaluate reproducibility (Figure 5.19).

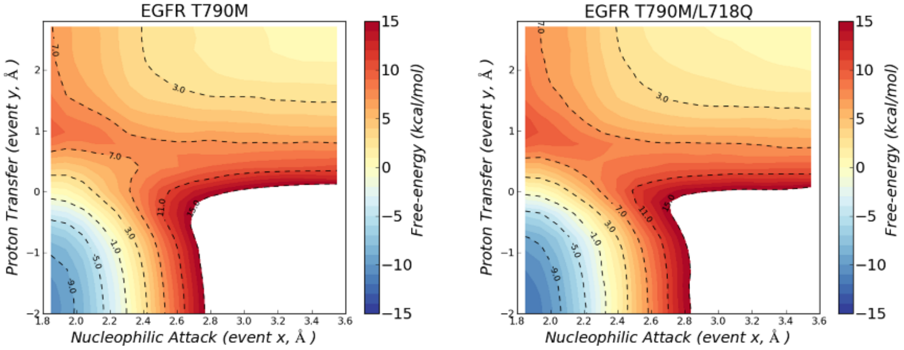


Figure 5.19. SCC-DFTB/AMBER FES of the replica of Cys797 alkylation by osimertinib in the presence of EGFR 790M (left panel) or EGFR T790M/L718Q (right panel). The reaction coordinates (nucleophilic attack and proton transfer), are given in angstroms. Free energies are given in kcal/mol, and the contour levels are set at 2 kcal/mol.

WaterSwap calculations.

From a 300 ns-long trajectory for EGFR T790M – osimertinib and EGFR T790M/L718Q – osimertinib complexes, 5 equally spaced snapshots were selected and employed for water-swap calculations. Binding free energies were calculated using replica-exchange thermodynamic integration over 16 λ windows (0.005, 0.071, 0.137, 0.203, 0.269, 0.335, 0.401, 0.467, 0.533, 0.599, 0.665, 0.731, 0.797, 0.863, 0.929, 0.995) over the water-swap reaction coordinate. 30 million Monte Carlo moves were performed for each window, with the free-energy gradient averaged over the last 20 million steps. Simulations were performed with a 15 Å coulomb and Lennard Jones non-bonded cutoff. A reflection sphere was also used to constrain sampling to within a 15 Å radius of osimertinib.

Unbiased MD simulations and free energy calculations.

Starting from equilibrated systems of EGFR T790M – osimertinib and EGFR T790M/L718Q – osimertinib complexes 4 independent runs of 300 ns-long MD simulations at 300 K in NVT conditions were performed, using the setting described above. 1.2 μs of MD simulations were performed for each complex, for total of 2.4 μs of simulation time. Snapshots collected from MD trajectories (30,000 for each replica, for a total of 120,000 for each system) were employed to build a 2D-free-energy in the space of S-Cβ distance and C_1-C_2-N_1-C_3 dihedral angle. In details, the free-energy along these variables was computed applying the following equation:

$$A(r) = - k_B T \ln P(r)$$

where P is the probability distribution, k_B is the Boltzmann constant, and T is the simulation temperature. The 2D distribution function P(r) was obtained performing a histogram analysis with the *cpptraj* module[144] implemented in AMBER16. Snapshots were separated in classes based on the values of the S-Cβ distance and C_1-C_2-N_1-C_3 dihedral angle. S-Cβ distance ranged from 2.5 to 9.5 Å with a class width of 0.1 Å while the C_1-C_2-N_1-C_3 dihedral angle ranged from -160° to +160° with a class width of 15°. 2D-free-energy surfaces in the space of S-$C_β$ distance and C_1-C_2-N_1-C_3 dihedral angle for each independent replica are reported for EGFR T790M – osimertinib and EGFR T790M/L718Q – osimertinib complexes in Figure 5.20 and figure 5.21 respectively.

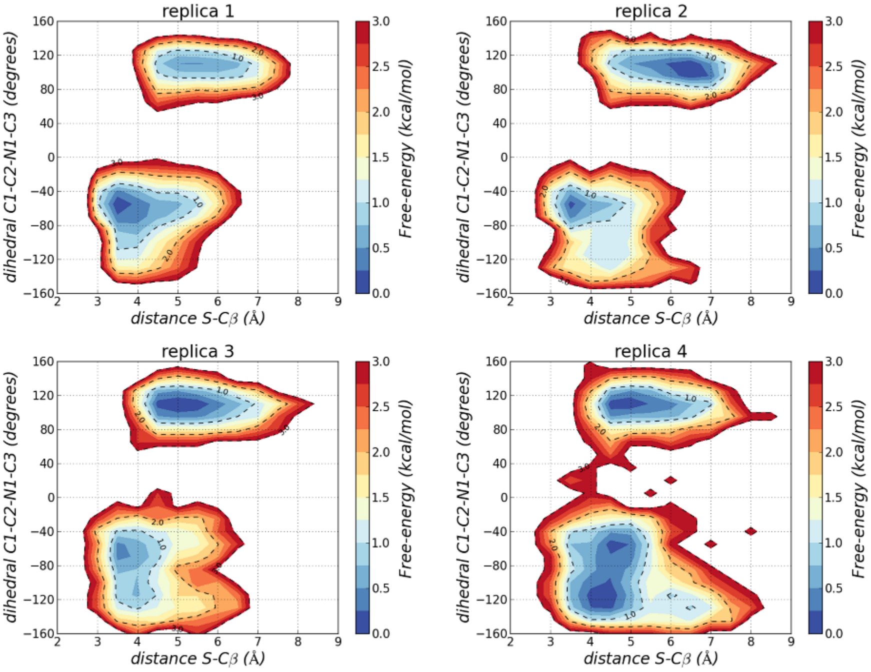


Figure 5.20. Free energy surface build from frequency distribution of conformations obtained from four independent 300-ns long MD simulations for EGFR T790M in non-covalent complex with osimertinib.

---

[144] Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.

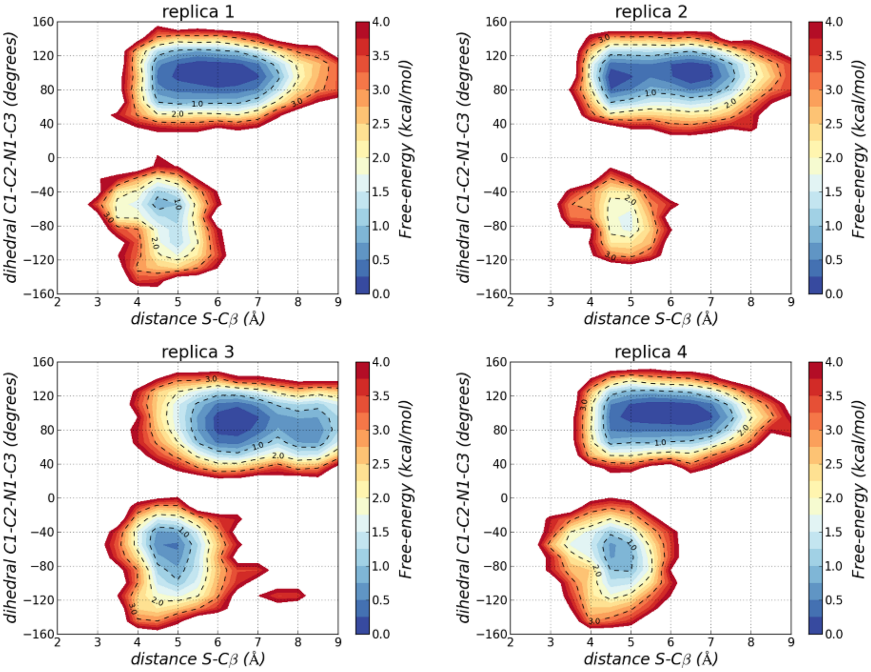


Figure 5.21. Free energy surface build from frequency distribution of conformations obtained from four independent 300-ns long MD simulations for EGFR T790M/L718Q in non-covalent complex with osimertinib.

Conclusions and future perspectives

Computational drug design can accelerate the challenging process of designing and optimizing a new drug candidate.

Mainly because of the advent of powerful hardware architectures, it is nowadays possible to perform molecular dynamics (MD) simulations in the time frame of microseconds up to a few milliseconds, allowing a thorough sampling of the conformational space of large biomolecules and the investigation of the protein-ligand binding process. These long simulations can then be coupled to free-energy methods to provide the free-energy profile of protein−ligand binding, which may be employed to estimate the thermodynamics and kinetics of the binding process. The ability of estimating thermodynamics and kinetics of protein-ligand binding through computational methods has the potential to impact both drug design and lead optimization phases of drug discovery, indicating the most favorable compounds and modifications.

Even though MD can be quite powerful, it is still computationally demanding. This limits its practical use for drug discovery to just a small number of ligands. In the past few decades, researchers have reported different strategies for overcoming this limit by enhancing the sampling of relevant regions of the free-energy surface.

In this thesis, for example, I investigated different applications of two enhanced sampling methods (i.e., metadynamics and umbrella sampling) in drug design and lead optimization. In particular, I explored ligand-protein recognition and kinetics of ligand unbinding using metadynamics, and I studied the reactivity of a covalent drug with umbrella sampling using a QM/MM approach.

Metadynamics showed great promise in predicting the structure of EphA2 receptor in complex with small molecules inhibitors. Indeed, the free-energy profile of protein-ligand binding reconstructed with metadynamics in this work allowed to identify free-energy minima corresponding to binding modes consistent with available SAR data, which allowed to drive the synthesis of novel EphA2 receptor antagonists. Moreover, metadynamics showed promising results in estimating the kinetics of protein-ligand unbinding. Indeed, the metadynamics-based protocol here proposed was able to correctly rank congeneric CDK8 inhibitors in terms of residence times, suggesting that this approach could be applied in drug design projects aimed to optimize the experimental residence time of a congeneric series of ligands.

Finally, covalent inhibition was investigated as well with enhanced sampling, designing a protocol that evaluates different features of covalent binding. In particular, ionization state of the nucleophile and estimation of the free-energy barrier for covalent adduct formation are evaluated with umbrella sampling, while mutual orientation of the reactants within the target active site is evaluated with extensive MD. This proposed protocol may represent a viable and generalizable strategy to investigate covalent binding and/or the effects of mutations on covalent inhibition *in silico*.

Taken together, the results of this thesis highlight the potentials of enhanced sampling methods in drug design and lead optimization campaigns and suggest computational protocols that may be generalized and employed to design and optimize compounds targeting other receptors of pharmaceutical interest.

Appendix

choice of CVs for metadynamics

11.1 CVs for EphA2-antagonist unbinding

As mentioned in chapter 2, one critical stage in performing MetaD simulations is the choice of a suitable set of Collective variables (CVs). To this aim, some preliminary simulations starting from binding mode A were run to identify CVs able to describe the full process of EphA2-LCA binding/unbinding and discriminate between different binding poses of the ligand within the binding site. The first set of CVs tested to explore EphA2-LCA unbinding included the distance between the center of mass (COM) of the EphA2 ligand binding domain and the COM of the ligand steroid moiety as CV1, and the distance between carbonilic carbon of LCA and Arg103 as CV2. Figure 7.1 shows the evolution of this first set of CVs during a representative metadynamics of unbinding of the EphA2-LCA complex.

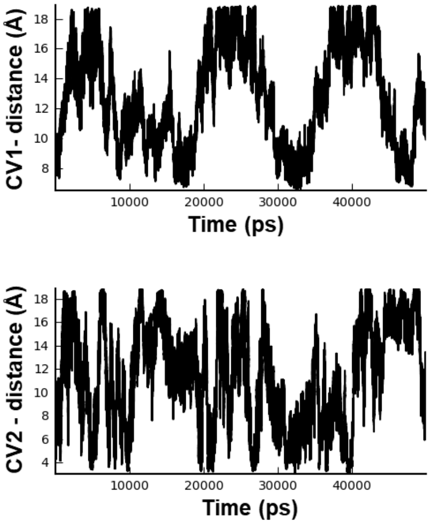


Figure 7.1. time course of CV1 (upper panel) and CV2 (lower panel) in a representative preliminary simulation of EphA2-LCA unbinding.

CV1 appears promising in pushing the ligand out of the bound state and is able to discriminate between bound and unbound states. Moreover, with the use of an additional potential placed at 20Å of CV1, recrossing events between bound and unbound state are simulated. Conversely, CV2 is useful to identify when the ligand is bound to Arg103, but fails to clearly discriminate if LCA is bound to Arg159. Indeed, for medium-high values of CV2, the ligand could be bound to Arg159 or it could be in proximity to exit from the binding site either. That said, as the chosen CVs should discriminate between the initial and final state and describe all the relevant intermediates, with this set of CVs it was not possible to obtain an exhaustive and converged free-energy profile (FES) of the unbinding process of interest. Then, a new set of CVs was tested. These CVs, proposed by Gervasio et al., already showed good results when used to perform flexible docking with metadynamics.[145] In details, CV1 was the same described above, while CV2 was the angle vector taken between the same protein reference point and the major inertia axis of the steroid moiety (see figure 3.5 in the main text). This CV2 was promising as it assumed different values in MD simulations when starting the simulation from different binding poses (Figure 3.7 in the main text).

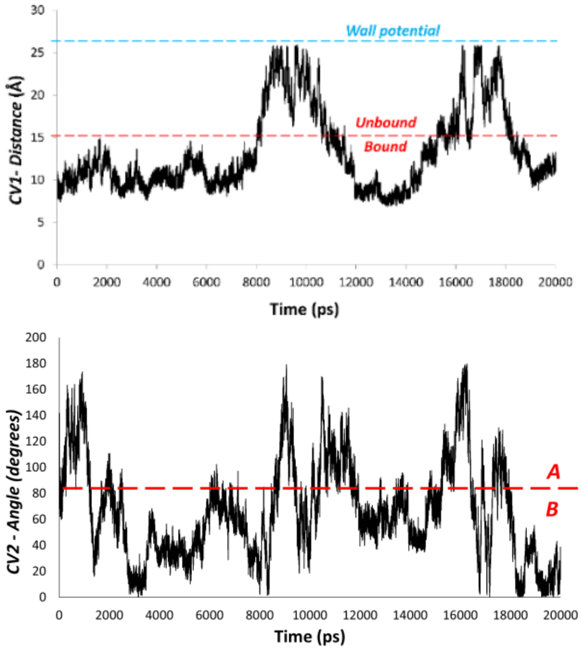


Figure 7.2. time course of CV1 (upper panel) and CV2 (lower panel) in a representative preliminary simulation of EphA2-LCA unbinding.

Figure 7.2 shows the evolution of CVs during a representative MetaD of unbinding of the EphA2-LCA complex with the new set of CVs. The angle defined as CV2 was able to discriminate different orientations assumed by LCA within EphA2 binding site during the binding/unbinding process. Indeed, CV2 was greater than 80° when LCA binds EphA2 with binding mode A, while CV2 was lower than 80° when LCA was in the orientation B. This set of CVs was therefore employed to characterize the FES of unbinding of EphA2-LCA complex.

---

[145] Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *J. Am. Chem. Soc.* **2005**, 127, 2600− 2607.

11.2 CVs for CDK8-inhibitor unbinding

It's important to point out that in chapter 4 the aim of MetaD simulations was not to characterize the FES of unbinding, but only to accelerate the dynamic of the unbinding process of CDK8-inibitor complexes, in order to rank different inhibitors according to their residence time. For this reason, the characterization and the convergence of the FES are not an issue for this calculation. In this case, the ideal set of CVs should be able to pull the ligand from the bound to the unbound state (without the need of recrossing events along the CVs) in a reasonable computational time, allowing to compare the unbinding time of a large set of compounds.

Moreover, to this aim, the distributions of calculated unbinding times for compounds having different experimental residence times should be separated enough to significantly discriminate the compounds. To identify the appropriate set of CVs, preliminary MetaD simulations were performed for the CDK8-inhibitor complexes having the shortest (CDK8-compound **1**) and the longest (CDK8-compound **7**) dissociation rates, using the same MetaD parameters described in the computational protocol in chapter 4.5. The first set of CVs tested was the same set proposed by Gervasio et al. and already employed to simulate EphA2-LCA unbinding in chapter 3. In details, they were defined as: CV1, the distance between the center of mass (COM) of the ligand and the COM of the protein hinge region (amino acids 98−103); CV2 is the angle formed by the major axis of the ligand (approximately defined by selecting two atoms of the rigid pyrazolylurea scaffold) and the COM of the protein hinge region. (Figure 7.3)

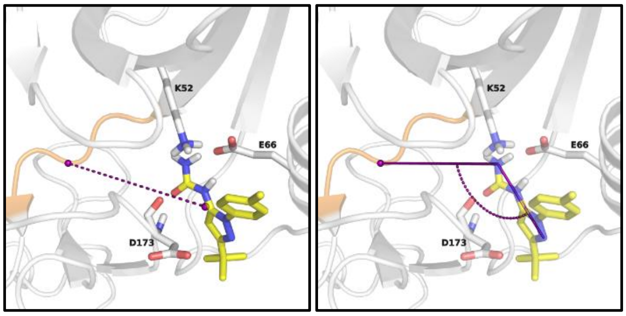


Figure 7.3. Graphical representations of distance CV1 (left panel) and angle CV2 (right panel) on CDK8-compound 1 complex. Center of mass points employed to define the CVs are depicted in purple, while axes are described by plain purple lines.

With this set of CVs, the unbinding processes were very fast for both complexes, with simulated unbinding times of 0.82 ± 0.09 ns and 1.53 ± 0.09 ns for CDK8-compound **1** and CDK8-compound **7** respectively. The range of unbinding times was too narrow to significantly discriminate compounds having residence times in the range between the residence time of compound **1** and the one of compound **7**. Then, a dihedral angle proposed by Masetti et al.[146] was also included in these preliminary simulations to add the exploration of the rotational movements of the ligand during the unbinding process (Figure 7.4). This third CV was the dihedral angle defined by the major axis of the ligand, the COM of the protein hinge region, and the COM of β-strand 1 of the kinase active site.

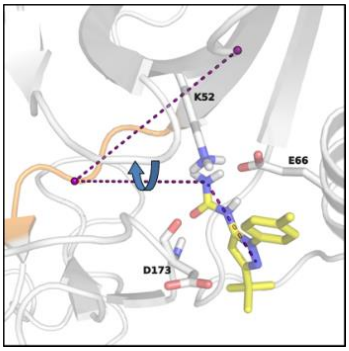


Figure 7.4. Graphical representations of dihedral CV3.

In this case, the addition of a third CV allowed to obtain a slightly slower unbinding: MetaD unbinding times were slightly longer, but the total range of data was still too restricted (1.92 ± 0.18 ns and 3.86 ± 0.19 ns for CDK8-compound **1** and CDK8-compound **7**, respectively).

To improve the distribution of calculated residence times, the RMSD of the ligand together with other 3 CVs describing the roto-translational movement of the ligand were added. In details, they were the angle defined by the COM of the ligand, the COM of the protein hinge region, and the COM of β-strand 6 of the kinase active site (amino acids 92−98); the dihedral angle defined by these three COMs and a fourth COM taken on β-strand 5 of the kinase active site (amino acids 80−87) and last, the angle formed by the minor axis of inertia of the ligand (approximately defined by selecting two atoms of the rigid phenylpyrazole scaffold) and the COM of the protein hinge region.

---

[146] Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. Exploring Complex Protein-Ligand Recognition Mechanisms with Coarse Metadynamics. *J. Phys. Chem. B* **2009**, 113, 4807−4816.

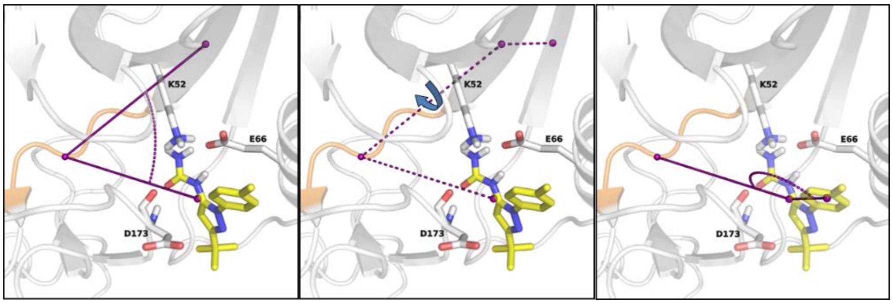


Figure 7.5. Graphical representations of the additional 3 CVs added to simulate the unbinding process.

Finally, when a suitable set of variables is used, (e.g., as in the case of the 7 CV simulations) the sampling of the unbinding process was significantly slowed down: the unbinding MetaD times were 4.87 ± 0.41 ns for CDK8-compound 1 and 17.98 ± 1.04 ns for CDK8-compound 7. The simulations with 7CVs can allow therefore to compare different compounds having residence times in the range between the residence time of compound **1** and the one of compound **7**.