

M Ű E G Y E T E M 1 7 8 2

BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS  
FACULTY OF CHEMICAL TECHNOLOGY AND BIOTECHNOLOGY

# Computational characterization of covalent enzyme inhibition

Thesis

**Levente Márk Mihalovits**

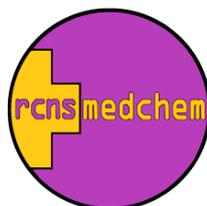
PhD candidate

**György G. Ferenczy**

supervisor

**György Miklós Keserű**

consultant



Medicinal Chemistry Research Group  
Research Centre for Natural Sciences,  
Loránd Eötvös Research Network

2021



## ACKNOWLEDGEMENT

During the years of my PhD a great number of nice and friendly people helped me along the way. I would not have been able to finish my work without their support, therefore I would like to express my gratitude to them.

First of all, I would like to thank my supervisor György G. Ferenczy for his continuous and extensive support throughout these years and for introducing me into the mysterious world of computational chemistry. I am also grateful to György M. Keserű, my consultant for guiding my work with his invaluable advices and providing the required infrastructural background.

I would like to thank László Poppe and Gergely Bánóczy, my former supervisor and consultant, respectively, in the Bioorganic Research Group at BME allowing me to take a glance into the field of biochemistry and related computational methods during my BSc and MSc.

I am grateful to Anita Bajusz-Rácz, Dóra Judit Kiss, Dávid Bajusz, Ádám Kelemen, András Németh and Bálint Varga for helping me in the tiny technical details of calculations and chemistry related problems throughout my PhD years and also for being great friends.

I thank all my colleagues in the Medicinal Chemistry Research Group at RCNS ensuring a healthy and helpful working environment.

A great part of my motivation came from my friends; and I will be always indebted for their encouragement doing my research and always being there for me when I felt a little overwhelmed.

Last, but not least I am unspeakably grateful to my family for supporting me both emotionally and financially to pursue my dream and continue my studies after MSc.



“A person who never made a mistake never tried anything new.”

Albert Einstein



## TABLE OF CONTENTS

Acknowledgement .....	3
Table of contents.....	7
List of abbreviations .....	10
Chapter 1 – Preface.....	11
Chapter 2 – Introduction .....	12
2.1 Aims of the research work included in the thesis .....	12
2.2 Covalent inhibition .....	12
2.2.1 Historical overview .....	12
2.2.2 General aspects of covalent inhibition.....	13
2.2.3 Mechanism and characterization of covalent inhibition .....	13
2.3 Computer aided drug design .....	14
2.3.1 Tools of computational chemistry used in drug discovery .....	14
2.3.2 Tools specific to covalent drug discovery .....	16
2.4 Investigated proteins targeted by covalent inhibitors .....	17
2.4.1 MurA.....	17
2.4.2 KRAS (RAS proteins) .....	18
2.4.3 Tyrosine kinases (EGFR, ITK, BTK, BMX).....	19
2.4.4 Immunoproteasome.....	20
Chapter 3 – Theory .....	22
3.1 Molecular mechanics .....	22
3.2 Quantum mechanics.....	22
3.2.1 Electronic structure based methods .....	23
3.2.2 Density functional theory methods .....	24
3.2.3 Semi-empirical methods .....	25
3.3 Hybrid QM/MM potential .....	25
3.4 Molecular dynamics.....	27
3.4.1 Unbiased molecular dynamics .....	28
3.4.2 Steered molecular dynamics .....	28
3.4.3 Umbrella sampling.....	29
3.4.4 Thermodynamic integration.....	29
Chapter 4 – Materials and methods .....	31
4.1 Molecular mechanics (MM) based methods.....	31
4.1.1 Protein and ligand preparation.....	31

4.1.2 Ligand docking .....	32
4.1.3 Classical MD simulations .....	32
4.1.4 Thermodynamic integration.....	33
4.1.5 pK <sub>a</sub> calculation of protein sidechains.....	34
4.2 Quantum mechanics (QM) based methods .....	35
4.2.1 Model system preparation with cysteine surrogate methyl-thiolate .....	35
4.2.2 Transition state and reaction energy calculations .....	35
4.2.3 Compound pK <sub>a</sub> calculations.....	35
4.3 Hybrid QM/MM based methods.....	35
4.3.1 Preparatory steps.....	35
4.3.2 Steered molecular dynamics (SMD) simulations .....	37
4.3.3 Umbrella Sampling (US) .....	38
4.3.4 Energy correction of the QM region .....	38
4.4 Evaluation and post-procession .....	39
4.4.1 Calculation of binding free energy differences.....	39
4.4.2 Construction of the potential of mean force (PMF) curves and surfaces .....	39
4.4.3 pK <sub>a</sub> estimation derived from thermodynamic integration.....	40
4.4.4 Error estimation .....	40
Chapter 5 – Results and discussion.....	41
5.1 Reactivity prediction of covalent inhibitors.....	41
5.1.1 Results of model reaction calculations .....	41
5.1.2 Correlation with experimental reactivity metrics .....	42
5.2 Catalytic mechanism and covalent inhibition of MurA.....	43
5.2.1 Deprotonation mechanism .....	43
5.2.2 Binding pocket conformations.....	46
5.2.3 Catalytic reaction with natural substrate phosphoenol-piruvate (PEP) .....	47
5.2.4 Covalent inhibition by different covalent warhead chemotypes .....	48
5.2.5 Evaluation and separation of the active and inactive compounds .....	53
5.3. Affinity and selectivity assessment of covalent inhibitors .....	54
5.3.1 KRAS inhibitors .....	54
5.3.2 EGFR inhibitors.....	59
5.3.3 Selectivity studies for covalent kinase inhibitors .....	63
5.4 Selectivity and inhibition mechanism of covalent immunoproteasome inhibitors.....	70
5.4.1 Mechanistic description of the covalent inactivation reaction.....	70
5.4.2 Characterization of the non-covalent binding step for selected inhibitors .....	75

5.4.3 Evaluation of covalent binding step for selected inhibitors.....	75
5.4.4 Selectivity studies between iPS and cPS .....	76
5.4.5 Evaluation of the overall proteasome inhibition.....	78
Chapter 6 – Conclusions .....	81
6.1 Conclusions and thesis points .....	81
6.2 Összefoglalás és tézispontok .....	82
6.3 List of the publications discussed in the thesis .....	85
References.....	86
Appendix A.....	100
Appendix B.....	110

## LIST OF ABBREVIATIONS

<b>ACE</b> – acetamide cap	<b>PEP</b> – phosphoenol pyruvate
<b>ADP(ATP)</b> – adenosine di(tri)phosphate	<b>PES</b> – potential energy surface
<b>BMX</b> – bone marrow tyrosine kinase	<b>PMF</b> – potential of mean force
<b>BTK</b> – Bruton tyrosine kinase	<b>PS</b> – product state
<b>cPS</b> – constitutive proteasome	<b>QM</b> – quantum mechanics
<b>DFT</b> – density functional theory	<b>QSAR</b> – quantitative structure-activity relationship
<b>DFTB</b> – density functional tight binding	<b>RC</b> – reaction coordinate
<b>DM</b> – double mutant	<b>RMSD</b> – root-mean-square deviation
<b>EGFR</b> – epidermal growth factor receptor	<b>RS</b> – reactant state
<b>FDA</b> - U.S. Food and Drug Administration	<b>SCF</b> – self-consistent field
<b>FEP</b> – free energy perturbation	<b>SMD</b> – steered molecular dynamics
<b>FF</b> – force-field	<b>THI</b> – tetrahedral intermediate
<b>GAFF</b> – general amber force-field	<b>TI</b> – thermodynamic integration
<b>GDP(GTP)</b> – guanosine di(tri)phosphate	<b>TS</b> – transition state
<b>GSH</b> – $\gamma$ -glutamyl-cysteinylglycin tripeptide (glutathione)	<b>UNAG</b> – UDP-N-acetylglucosamine
<b>HF</b> – Hartree-Fock	<b>US</b> – umbrella sampling
<b>HOMO</b> – highest occupied molecular orbital	<b>VDW</b> – van der Waals
<b>HTS</b> – high-throughput screening	<b>WHAM</b> – weighted histogram analysis method
<b>IC<sub>50</sub></b> – half-maximal inhibitory concentration	
<b>INT</b> - intermediate	
<b>iPS</b> – immunoproteasome	
<b>ITK</b> – Interleukin 2 Inducible T Cell Kinase	
<b>JAK</b> – Janus kinase	
<b>K<sub>i</sub></b> – inhibition constant	
<b>k<sub>inact</sub></b> – reaction rate constant	
<b>KRAS</b> – Kirsten rat sarcoma viral oncogene homolog	
<b>LUMO</b> – lowest unoccupied molecule orbital	
<b>MD</b> – molecular dynamics	
<b>MM</b> – molecular mechanics	
<b>MUE</b> – mean unsigned error	
<b>MurA</b> – UDP-N-acetylglucosamine-1-carboxyvinyltransferase	
<b>NME</b> – N-methylamide cap	
<b>NMR</b> – nuclear magnetic resonance	
<b>NSCLC</b> – non-small cell lung cancer	
<b>ONIOM</b> - our Own N-layer Integrated molecular Orbital molecular Mechanics	
<b>PDB</b> – protein data bank	

## CHAPTER 1 – PREFACE

Medicine, pharmaceutical, drug. Familiar words for modern society, they are present from the first day of our life. We use them so naturally; we do not even realize how much money, painstaking work and extensive knowledge are behind that pill, we take. Since the dawn of humanity curing diseases and prolonging lifetime have been among the most essential goals, one can achieve. Recently, the outbreak of the coronavirus pandemic emphasized the importance of biology, general medicine, and medicinal chemistry, even more than before. Though, starting from the previous century we have been able to treat several, previously lethal diseases and overall doubling the average lifetime, COVID-19 pandemic indicated that there is still room for improvement. However, the relatively fast response of humanity in the forms of different types of treatments is the result of a global cooperation of researchers, coming from a wide range of scientific fields including the advanced computer technology supporting their work.

As computer performance has been developing continuously since their first appearance in the middle of the 20<sup>th</sup> century, by now, computational capacity reached the level, where the application of theoretical methods, modeling and simulations in scientific research is in pair with experimental work in terms of reliability and feasibility. Chemistry makes no exception; computational chemistry has become an essential tool in drug discovery, aiding and complementing the indispensable experimental work. A good portion of the drug discovery campaigns, which resulted in officially approved drug compound, applied computational methods during the discovery phase [1,2].

The discovery of covalent drugs, a subspecies of inhibitor molecules that forms a covalent bond with the targeted protein, also benefits from computational methods. Although, there are several examples of early covalent drugs such as aspirin or clopidogrel, the purposeful search on covalent inhibitors in the early millennium raised the number of applied covalent drugs and they are now well represented in the treatment of multiple pathological conditions, such as autoimmune diseases [3,4], disorders of the central nervous [5] and cardiovascular systems [6], gastrointestinal illnesses [7], infections [8], and most abundantly they are present in cancer therapy [9,10]. In fact, 10 out of the 14 FDA approved covalent drugs between 2011-2019 were cancer related medicines [11].

In my thesis, I have applied computational chemistry methods to study the covalent inhibition of several enzyme targets by a large variety of compounds. Studying complex systems and processes required the detailed understanding of the inhibition reactions, the characterization of the enzyme-substrate interactions and the enzyme cycle progression, and the application of both protein structure based and ligand-based approaches. Finally, my work concluded with a method that is able to fully address the free energy changes of the complete covalent inhibition process including both the non-covalent and covalent steps.

## CHAPTER 2 – INTRODUCTION

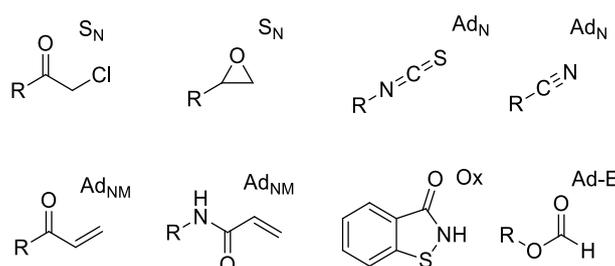
In the current chapter I give a brief overview of the topics discussed in my thesis. More detailed theoretical and practical information about the applied methods is presented in **Chapter 3 – Theory** and **Chapter 4 – Materials and methods**.

### 2.1 Aims of the research work included in the thesis

The primary objective of my PhD thesis was to characterize covalent inhibition applying computational chemistry tools. The calculations consisted of a wide scope of available methods at multiple levels of theory. We aimed to evaluate and predict the binding events and related energy changes of targeted proteins and compounds equipped with different electrophilic warhead chemotypes. Our first goal was to examine model reactions between inhibitor candidates and sidechain substitute molecules, evaluating the reaction barrier and reaction energy. In the next step, we aimed to adapt these model reactions into realistic enzyme environment, taking into account non-covalent binding, as well. Finally, we wanted to develop a general simulation protocol to fully characterize the binding of covalent inhibitors using our previous experiences on model systems and enzyme-ligand complexes.

### 2.2 Covalent inhibition

Enzyme inhibitors can be categorized into two groups, namely non-covalent and covalent inhibitor molecules. Compounds of the former group bind into the protein's proper binding site with secondary non-covalent interactions, while covalent inhibitors additionally form covalent bond with a nucleophilic sidechain of the target protein. The latter process requires a specific moiety on the inhibitor molecule, namely the “covalent warhead” or simply “warhead”, which is an electrophilic center able to receive the nucleophilic attack of the target sidechain (**Figure 1**). The mechanism of covalent reactions covers a wide range, including addition-elimination (Ad-E), nucleophilic substitution ( $S_N$ ), oxidation (Ox), non-Michael nucleophilic addition (Ad<sub>NM</sub>) and most abundantly Michael-addition (Ad<sub>NM</sub>) [12].



**Figure 1.** Examples of electrophilic covalent warheads and their mechanism of action.

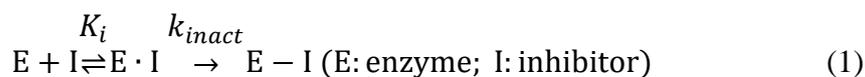
#### 2.2.1 Historical overview

Covalent inhibition is not a recently discovered phenomenon in drug research. One of the earliest compounds with covalent inhibition mechanism is Aspirin, discovered by Bayer in 1899. Interestingly, until 1971, Aspirin's mechanism of action was unknown, though it inhibits the cyclooxygenase enzyme by acylating a serine sidechain, preventing the formation of prostaglandins that cause inflammation [13]. However, during the previous century covalent inhibition was considered as a prohibitive and rejected area of drug discovery owing to the

potential idiosyncratic toxicity associated with the risk of off-target activity. Therefore, the research of covalent inhibitors was brushed aside for decades. A change in this paradigm occurred at the early 2000s, when researchers reconsidered the possible advantages and disadvantages of covalent modification. Since then, a set of effective design principles were set and it has become an established research field in medicinal chemistry, applied successfully in treating pathological conditions, most notably in cancer therapy.

### 2.2.2 General aspects of covalent inhibition

Covalent inhibitors typically target amino acids, equipped with a nucleophilic sidechain, such as serine, tyrosine, threonine, lysine, and most often cysteine [14,15]. Cysteine possesses a prominent role as targeted sidechain, owing to its elevated nucleophilicity and low occurrence in the human proteosome [16,17], which reduces potential off-target reactivity of cysteine specific warheads [18]. The process of covalent inhibition follows a two-step mechanism, namely the non-covalent step, also known as the molecular recognition, in which the inhibitor binds into the binding site and forms non-covalent interactions with the surrounding protein sidechains of the binding site; and the covalent step, which involves the formation of the covalent bond between the ligand and the targeted sidechain (**Eq. 1**). Without reactivity modulation covalent modification is implicated in immunotoxicity and idiosyncratic hypersensitivity reactions [19]. Therefore, the successful design of covalent inhibitors requires not only developing desired enzyme-ligand secondary interactions but also the optimization of the warhead reactivity.



where  $K_i$  denotes the equilibrium constant and  $k_{inact}$  is the rate constant. Once the reactivity problems have been overcome by careful design of the warhead and the non-covalent scaffold primarily responsible to the molecular recognition, covalent inhibitors can advance into highly effective drug compounds. They possess great advantages compared to the non-covalent counterparts, such as increased biochemical efficiency, longer residence time, lower dosage requirements, improved therapeutic index and the potential avoidance of some drug resistance mechanism [20,21].

### 2.2.3 Mechanism and characterization of covalent inhibition

The typical scenario for covalent inhibition involves the fast formation of the non-covalent complex as equilibrium process followed by the covalent reaction possessing higher barrier and either a reversible or irreversible mechanism, based on the free energy profile. As shown in **Eq. 1** the two steps are described by the  $K_i$  equilibrium constant and for irreversible inhibition, the  $k_{inact}$  rate constant. Both binding steps may contribute significantly to the overall binding affinity, hence the complete description of the inhibition requires the evaluation of both constants.  $K_i$  and  $k_{inact}$  can be determined experimentally [22,23], however, the separate determination may be challenging, and in many cases the  $k_{inact}/K_i$  ratio is determined. These constants can be evaluated computationally, as well. The complete process of covalent inhibition can be characterized by modeling the binding event including the non-covalent binding free energy ( $\Delta G$ ) and the transition state free energy ( $\Delta G^\ddagger$ ) of the covalent bond formation. The computational evaluation of the constants is carried out using **Eqs. 2** and **3**.

$$\Delta G = RT \ln(K_i) \quad (2)$$

$$\Delta G^\ddagger = -RT \ln\left(\frac{k_{\text{inact}}}{k_b T/h}\right) \quad (3)$$

where R is the universal gas constant, T is the absolute temperature,  $k_b$  is the Boltzmann constant and h is the Planck constant.

In specific cases,  $IC_{50}$  — measuring the ligand concentration which halves the activity of the inhibited enzyme — serves as the only experimental measure of inhibition. The determination of  $IC_{50}$  is a common and straightforward measurement, however,  $IC_{50}$  value is typically time dependent in covalent inhibition and therefore it is less suitable for comparative computational and experimental analysis.

An ideal covalent inhibitor may be characterized by high target affinity, meaning a low  $K_i$  value, paired with modest reactivity, meaning moderate  $k_{\text{inact}}$  constant to avoid off-target toxicity and maximizing the therapeutic index. The design of covalent modifiers requires the lowering of  $K_i$  and the optimization of  $k_{\text{inact}}$ , therefore, the main objective is improving the  $k_{\text{inact}}/K_i$  ratio. These attempts can be successfully aided by computational methods, offering an effective drug designing strategy for covalent inhibitors.

## 2.3 Computer aided drug design

With the exploding development in computer technology, computational chemistry has been evolved to a commonly used method, aiding a wide scope of chemistry. Material science, biotechnology, analytical chemistry, medicinal chemistry; just a few notable fields, which benefit from computational methods. During my research, calculations were performed in drug design related problems, and a variety of methods was used. The application of such tools is highly fruitful, and the synthesis of computational and experimental approaches results in a valuable and beneficial research protocol.

### 2.3.1 Tools of computational chemistry used in drug discovery

A complete account of computational methods is beyond the scope of this thesis, but a brief description of procedures related to early drug discovery and to my research topic is given below.

At the initial stage of mechanism based drug discovery campaigns the main goal is to identify the target protein or target biological process, whose modification might produce the desired effect, namely treating the selected pathological condition. Aiding this, bioinformatics [24] is a widespread, commonly used set of computational methods, combining biology, computer science, information engineering and statistics; and includes the statistical analysis of large data sets. In addition, modeling of DNA, RNA, protein and biomolecular interactions are also part of the bioinformatics' toolbox. Identifying, characterizing and structurally describing the target protein is essential for structure based drug design [25]. X-ray diffraction, NMR and cryo electron microscopy measurements are common practices to determine protein structures, although, they are not available in some cases. Homology modeling can also produce structural information based on sequence similarity with related proteins with available experimental structure. Recently, AlphaFold, a machine-learning approach was proposed that able to predict three-dimensional structure of proteins derived from physical and biological knowledge of protein structures [26]. The protein models are subjected to refinement processes, involving

geometry optimizations and optionally molecular dynamics simulations. The detailed characterization of the protein function by experimental and computational tools is also essential for an effective drug discovery process.

Once the target and its function have been identified, the next step ahead is to find compounds that are able to bind to the protein and to affect its function. High throughput screening is a frequently applied experimental tool to find chemical starting points, however it requires an enormous amount of resources not typically available in the academia. In contrast, virtual screening offers a less demanding computational alternative to prioritize compounds. Molecular docking — being one of the most fundamental structure-based [25] tool — is the process of “fitting the keys into the lock” using the hypothesis of Emil Fischer. When applied in the context of virtual screening, a vast number of molecules are positioned into the targeted site of the protein to predict their possible binding poses and interactions. The construction of the poses can be performed by different docking algorithms, such as AutoDock, FlexX, Gold or Glide [27]. Some of them are able to take induced fit into account [28], some others are able to handle covalent inhibition [29,30], but the basis of most of them is the optimization of the interactions between the ligand and the enzyme environment. Interactions can be described by force-field (see in **3.1 Molecular mechanics**) or by empirical or knowledge based parameters and scores assigned to the binding poses can be considered as crude estimates of the binding free energy. Finally, at the end of virtual screening, top-scored compounds are selected for further theoretical or experimental evaluation. Molecules selected by virtual screening and confirmed experimentally are hit compounds that can serve as chemical starting points for further stages of drug design campaigns.

To optimize previously identified hit molecules, the enzyme-ligand structures can be examined and manipulated using more sophisticated methods. A docked complex may be subject of geometry optimization to fine tune the initial pose of the substrate or the inhibitor. In case of covalent inhibitors scans can be performed along specific reaction coordinates, simulating the covalent reaction in the active site by structure optimizations with restrained geometric parameters. With molecular dynamics (MD) simulations the dynamic nature of the proteins can be analyzed, too. Loop movements, non-covalent ligand binding-unbinding processes can be investigated by classical force field MD, while the investigation of covalent warhead-sidechain reactions needs MD with quantum chemical potentials [31]. Such simulations also serve as powerful tools for calculating free energy changes in binding events.

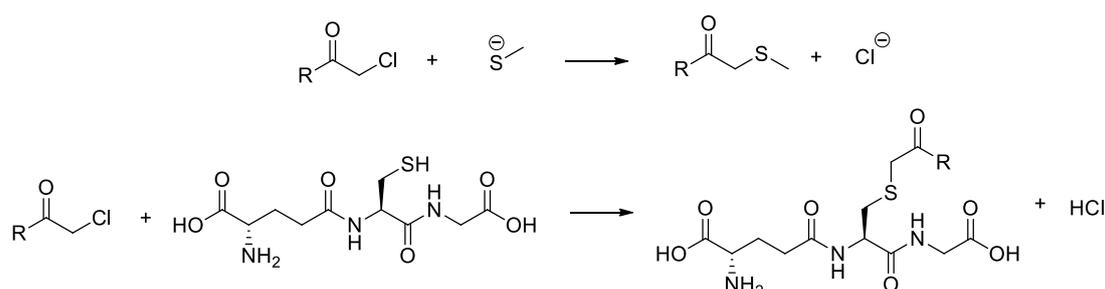
In ligand-based drug discovery [32] the emphasize is on the structure and chemical properties of the ligands only. Quantitative structure-activity relationship (QSAR) [33] calculations consist of the evaluation of different structure related molecular descriptors (electrophilicity, HOMO, LUMO energies, partial charges etc. [34]) with the aim to construct an equation describing a relation between the descriptors and the measured biological activity. Important structural features contributing to protein interaction (pharmacophores) can be evaluated too, and a complement image of the binding site can be constructed based on the extracted information. Virtual screening applications are also available for the ligand-based approach. Once a QSAR model has been built, activity predictions can be made for large compound libraries. Other frequently used techniques include shape screening that uses the shape of the bioactive conformation of the active molecules to filter libraries and to find compounds with a shape similar to the actives. Additionally, pharmacophore screening follows similar ideas, but instead of the shape information, the filtering is based on the 3D orientation of selected pharmacophore elements of the examined molecules.

Methods merging the ligand- and structure-based approaches are available, too [35]. To reduce computational demand, protein-ligand complexes can be simplified to model systems, containing the chemically interesting regions for both the enzyme environment and the ligand. This technique is mostly applied in the high-level assessment of the reactivity of covalent inhibitors, but mechanistic studies may also use the same considerations [36].

### 2.3.2 Tools specific to covalent drug discovery

The formation of the covalent bond involves the rearrangement of the electron structure. Classical force-field potentials typically do not contain the proper terms for the description of the changes in the bonding structure, although attempts to construct reactive force-fields were made previously [37]. Therefore, the modeling of covalent reactions typically requires quantum chemistry. *Ab initio* wave-function methods can achieve high accuracy, however, their computational cost hinders their use for molecules containing large number of heavy atoms. More common alternative is density functional theory calculations, producing sensible precision with moderate resource requirement. Although, performing simulations with large amount of sampling (e.g. molecular dynamics) is more feasible with semiempirical methods (see later in **Chapter 3 – Theory**) having a reduced computational cost.

One of the basic goals of quantum mechanical calculations on covalent inhibitors is the estimation of the covalent reaction's barrier, which accounts for the reaction rate of the covalent step. The calculations are performed for enzyme-ligand complexes, however, model systems are extensively used for preliminary evaluations. Cysteine being the most targeted sidechain, model calculations in which the enzyme and the cysteine is substituted by a single methyl-thiolate compound is used abundantly (**Scheme 1**), but further surrogates are available for nucleophilic sidechains. Reactivity calculations give valuable information on the compounds' off-target activity and aid the optimization of the warhead. Comparison with experimental values is also feasible. The GSH half-life determines the time in which half of the cysteine surrogate tripeptide glutathione is modified covalently by the investigated cysteine targeting inhibitor. The GSH half-life shows good correlation with the calculated barriers of the reaction between the compound and methyl-thiolate [21,36,38].



**Scheme 1.** Example of the model reaction of cysteine targeting inhibitor (top) and example of the reaction between cysteine targeting compounds and the cysteine surrogate GSH (glutathione) measuring the reactivity towards cysteine of covalent compounds (bottom).

Prediction of the transition state free energies ( $\Delta G^\ddagger$ ) is important to estimate the kinetic constant  $k_{\text{inact}}$  using **Eq. 3**. Not only barriers, but complete reaction mechanisms can be evaluated applying the QM potential, giving valuable information on the mechanism of action and supporting the design of improved inhibitors. Including the complete solvated enzyme-ligand system as QM-region is still not feasible computationally. Instead, partitioning into

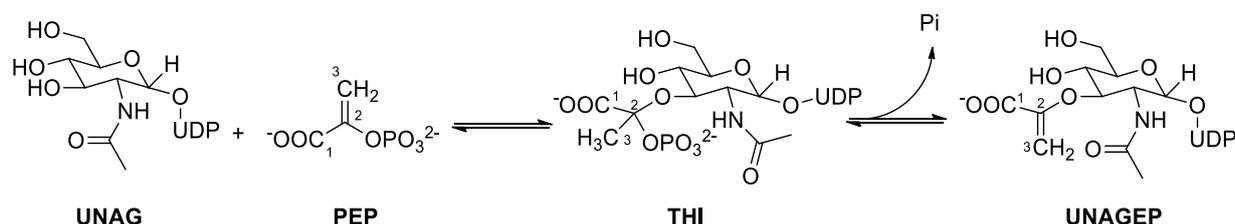
multiple subsystems with varying level of theory (QM/MM, ONIOM) is applied successfully in a vast number of studies [39].

## 2.4 Investigated proteins targeted by covalent inhibitors

Basically, every protein, possessing a nucleophilic sidechain can be modified covalently, although, targeting non-catalytic residues are usually avoided, due to off-target toxicity. Unlike non-covalent binders, covalent inhibitors generally modify the targeted proteins irreversibly, resulting in the loss of the enzymatic function. This mechanism of action is undesired in many cases, however, inhibiting oncogenic enzymes gives a valuable tool treating cancer related processes. A great number of potential covalent targets exist, and here I give a short review on the enzymes, which were examined during my work. Note that the protein structures are shown in figures of the corresponding subsections of **Chapter 5 – Results and discussion**.

### 2.4.1 MurA

UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) catalyzes the first committed step of the peptidoglycan (PG) biosynthesis. Peptidoglycan is an essential building block of both Gram-positive and Gram-negative bacteria's cell wall, constructed of repeating units of two sugars, N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) cross-linked by short peptides [40,41]. The inhibition of MurA disrupts the reproduction of bacteria, offering a feasible target for antibiotics. The catalyzed reaction consists of the transfer of an enolpyruvyl group from phosphoenolpyruvate (PEP) to UDP-N-acetylglucosamine (UNAG) forming UDP-N-acetylglucosamine enolpyruvate (UNAGEP) [42] (**Scheme 2**).



**Scheme 2.** Reaction catalyzed by MurA.

The reaction proceeds through a tetrahedral intermediate (THI), where the 3'-OH group of UNAG bounds covalently to C-2 of PEP and the C-3 is converted into a methyl group [43]. During the breakdown of the THI, an inorganic phosphate group is released and the C-2 methyl moiety is deprotonated yielding the product UNAGEP [44].

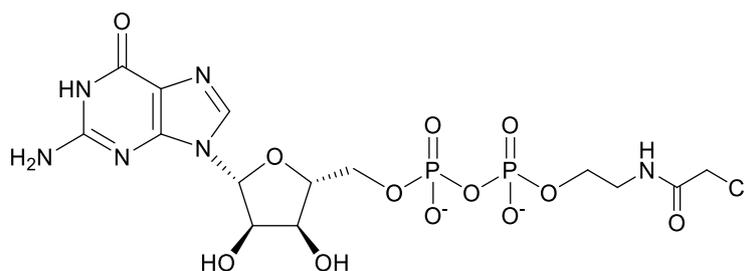
The structure of MurA features two structural domains connected by a double stranded linker and the active site formed at the interface of the two domains. A flexible surface loop containing the active site Cys115 residue is essential in the catalytic cycle and serve as a target for covalent modification [45–48]. (see later at **Figure 11** and **12**) The flexible loop adopts different conformations based on ligand binding. An open conformation was observed in the apo-form [49], while a closed conformation was found upon UNAG and further ligand binding. Interestingly, this conformational change has not been seen in *H. influenzae* MurA, where both the binary and ternary complex are in a half-open conformation [50]. A proposed mechanism [51] for cellular MurA reaction contains MurA in a dormant complex with UDP-N-acetylmuramate (UNAM) and a PEP molecule covalently bound to Cys115. At higher UNAG concentration, the UNAM is displaced by UNAG and the THI is formed. After the departure of

the phosphate group, it is replaced by another PEP, while the product UNAGEP is exchanged by UNAG, completing the catalytic cycle.

Cys115 plays a crucial role in the covalent inhibition of MurA, serving as a target for the clinically approved covalent inhibitor, fosfomycin [52,53]. Fosfomycin is a naturally occurring broad-spectrum antibiotics containing an oxirane warhead, inhibiting MurA in a time-dependent manner and is used to treat uncomplicated urinary tract infections [54]. Since its discovery in 1969, drug designing attempts have failed to produce more effective compound on the inhibition of MurA. Inhibitors discovered by HTS campaigns showed either less efficacy or lower specificity [55–57]. An interesting example of discovered inhibitors is terreic acid which possesses the same oxirane warhead. However, the terreic acid-MurA complex revealed an open loop conformation in contrast to the closed complex of fosfomycin [58].

#### 2.4.2 KRAS (RAS proteins)

RAS proteins are GTPases, acting as molecular switches cycling between the active GTP and inactive GDP bound conformations. They play essential role in regulating cell cycle progression by aiding the signaling of several important pathways [59,60]. There are three major RAS isoforms, namely HRAS, NRAS and KRAS, sharing 92-98% sequence identity in the 1-165 amino terminal residues. The region of the remaining 23-24 (residue number:166-188/189) carboxy-terminal residues is called hypervariable region, due to large differences in these sequences. Structurally four regions, forming the nucleotide-binding site is characterized by the core mechanism of RAS proteins: the phosphate-binding loop, switch I, switch II region and the base binding loops. Conformations of the switch regions differ between the active and inactive forms of the enzymes, after hydrolysis of the GTP in the bound form the phosphate is released and the two regions are able to relax into the GDP bound conformation [61]. Though, hyperactivating mutations in all the RAS isoforms result in oncogenic activity, mutations of KRAS is the most common deficiency in human cancer, leading to tumor proliferation. The most frequent sites of oncogenic mutations are residues 12 and 13 in the phosphate binding loop and residue 61 in the switch II region, however the frequency and the site of mutation is highly cancer dependent [62]. The KRAS-G12C mutation is one out of the three most common oncogenic mutation in KRAS related cancer types, the other two being G12D and G12V. Targeting the highly nucleophilic Cys12 residue of the oncogenic G12C mutant offers compounds with specificity over the wild-type KRAS, as its active site lacks any cysteine residue. The position of residue 12 is highly suitable for inhibitor design as it harbors at the edge of the nucleotide pocket and is adjacent to the flexible switch II region (see later at **Figure 23**). A number of KRAS<sup>G12C</sup> selective covalent compounds were published, containing electrophilic warheads attacking Cys12 [63,64] (**Figure 2**), two of them (sotorasib, adagrasib) are already marketed as cancer drugs [65,66].



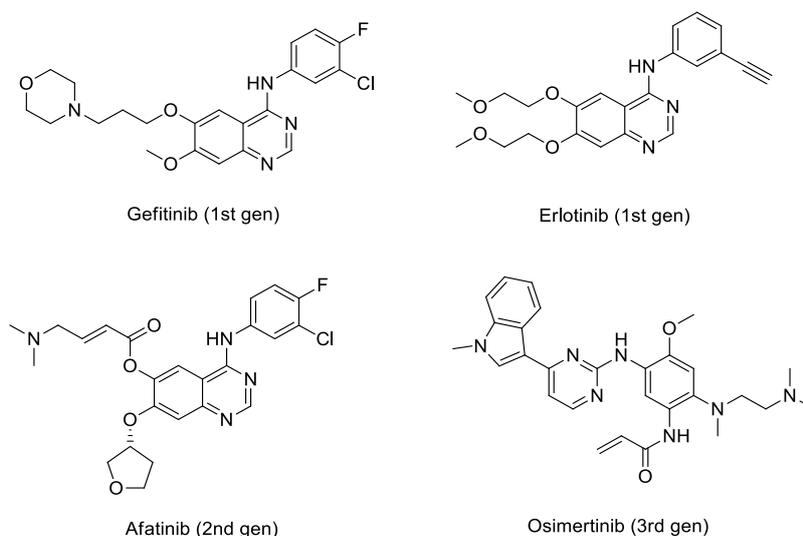
**Figure 2.** A GDP-like covalent inhibitor of KRAS<sup>G12C</sup>.

### 2.4.3 Tyrosine kinases (EGFR, ITK, BTK, BMX)

The following section contains information regarding four examined tyrosine kinases. Coupled with KRAS, they were subject of our methodological studies in which the complete covalent inhibition of different set of compounds were calculated, while less emphasize was put on their mechanism of action.

#### 2.4.3.1 EGFR

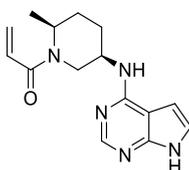
Epidermal growth factor receptor (EGFR) is a single chain transmembrane polypeptide, activated by the epidermal growth factor (EGF) [67]. It is the first member of the ErbB family of receptor tyrosine kinases, also known as HER1 or ErbB-1, taking part in the growth factor signaling. EGFR consists of three domains: the extracellular, ligand binding domain, the hydrophobic, transmembrane domain and the intracellular tyrosine kinase domain [60]. Abnormal signaling related to the malfunction of EGFR leads to uncontrolled cell proliferation, inhibition of apoptosis, angiogenesis, migration, and metastasis, which are the main characteristics of cancer cells [68]. Overexpression of the wild type EGFR and the emergence of specific activating mutants (L858R or G719S) have been found in several type of carcinomas, including lung, breast, head and neck cancers [69,70]. Inhibiting the overexpressed wild-type and mutant proteins in different type of cancer cells raises the opportunity to treat effectively the EGFR related oncogenic events. The first generation of clinically approved EGFR inhibitors are gefitinib [71] and erlotinib [72], two 4-anilinoquinazoline derivatives, binding reversibly and competitively at the ATP binding site of the kinase domain. The insurgence of cancer cell resistance towards gefitinib and erlotinib led to the discovery of second-generation EGFR inhibitors, having a Michael-acceptor type warhead and targeting a non-catalytic cysteine (Cys797) (see later in **Figure 29**) in the ATP binding site of EGFR [73]. These irreversible covalent inhibitors were proved to have high receptor residence time, circumvent ATP competition and be active against T790M mutant EGFR, as well. A notable example of FDA approved second generation EGFR inhibitors is afatinib [74], developed for treating non-small-cell lung cancer (NSCLC). Noteworthy, third-generation of EGFR inhibitors have also been approved [75,76]; they have anilino-pyrimidine structural elements and they show T790M-L858R double mutant selectivity over the wild-type EGFR.



**Figure 3.** FDA approved EGFR inhibitors.

#### 2.4.3.2 TEC-kinases

Interleukin-2 inducible T-cell kinase (ITK) (see later in **Figure 35**), bruton tyrosine kinase (BTK) and bone marrow tyrosine kinase gene on chromosome X (BMX) are non-receptor tyrosine kinases of the TEC-kinase family. Many members of this family are expressed in hematopoietic tissues, where their main function is in the growth, differentiation or both processes of blood cells [77]. Disfunctions of Tec-kinases are implicated in inflammation, immune responses and oncological disorders [78–80]. Members of the TEC-kinase family shares similar domain structure, consisted of an N-terminal pleckstrin-homology (PH) domain, the TEC homology domain (TH), Src homology 3 (SH3) and SH2 domains, finally the catalytic kinase domain (SH1) [81]. The active site of the TEC-kinases along with JAK3 also shows high similarity, with a covalently targetable cysteine (Cys442, ITK numbering) residue. This sidechain has been already targeted by covalent inhibitors [82,83], successfully inactivating the related kinases. Some of them are already approved by FDA [84].



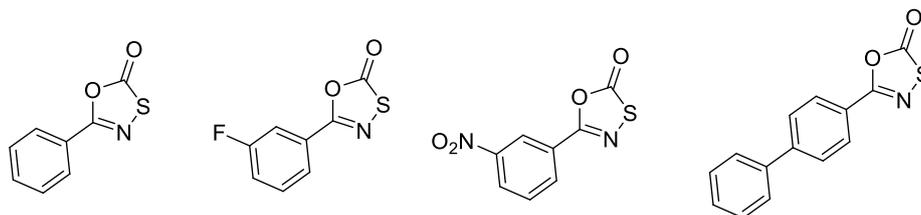
**Figure 4.** Example of a covalent inhibitor of several Tec-kinases and JAK3.

#### 2.4.4 Immunoproteasome

The proteasome is an ATP-dependent N-terminal hydrolase found in eukaryotes and prokaryotes [85]. It is essential in the non-lysosomal protein degradation through the ubiquitin proteasome-system (UPS) [86], taking part in several cellular actions, such as maintaining cell homeostasis, controlling cell cycle progression and regulating the immune system. The deficiencies in proteasome functioning are related to different pathogenic processes such as neoplastic, autoimmune, autoinflammatory, and neurodegenerative disorders [87,88]. Structurally, proteasome consists of 28 subunits forming a barrel shape. Three subunits out of the 28 play essential role in the enzymatic activity and also are the key elements in differentiating the constitutive proteasome (c-20s) and immunoproteasome (i-20s), namely  $\beta 1$ ,  $\beta 2$ ,  $\beta 5$  in cPS and  $\beta 1i$ ,  $\beta 2i$ ,  $\beta 5i$  in iPS. Interestingly, all the catalytically active subunits contain an N-terminal threonine residue which serves as the main target of several covalent inhibitors of proteasomes. The immunoproteasome is an important part of the immune system, expressed in dendritic cells and in lymphocytes. Non-hematopoietic cells can also express immunoproteasome by cytokines such as IFN- $\gamma$  or TNF- $\alpha$  [89–91].

Non-covalent proteasome inhibitors usually target the substrate binding channel [4], while the covalent compounds typically bind to the Thr1 residue. Covalent proteasome inhibitors with various warhead types, such as aldehydes, acrylamides, vinyl-sulfones, boronates,  $\alpha,\beta$ -epoxyketones,  $\alpha$ -ketoaldehydes,  $\beta$ -lactones and oxazoles [4,92] are published previously, including the FDA approved drugs, bortezomib and carfilzomib that target the chymotrypsin-like  $\beta 5$  subunit in both cPS and iPS (see later in **Figure 50** and **51**). In contrast, previous studies showed that selectivity toward the immunoproteasome over the constitutive proteasome in mouse models of various diseases resulted in larger therapeutic index [93–95].  $\beta 5i$  and  $\beta 5$  subunits show high similarity, however, iPS with the sequentially identical but more spacious S1 pocket can be targeted selectively. A series of oxathiazolone containing covalent

inhibitors were reported with inherent iPS selectivity that is modulated by different substituents on the non-warhead region of the compounds [92,96,97] (**Figure 5**).



**Figure 5.** Examples of iPS selective oxathiazolone inhibitors.

## CHAPTER 3 – THEORY

This chapter gives a brief overview of the theoretical background behind the computational methods applied. More detailed descriptions are available in the cited literature.

### 3.1 Molecular mechanics

Molecular mechanics, also known as force-field (FF) method is a simplified, parameter-based tool to calculate the energy of molecular structures. Instead of calculating energies based on the electronic structure, molecular mechanics uses functions of the nuclear coordinates with parameters fitted to experimental results or high-level calculations. Molecular mechanics treats molecules as “ball and spring” models of atoms and bonds. As the solution of the electronic Schrödinger equation is bypassed, every quantum related aspect is ignored during MM calculations.

The basic concept of the parametrizations of force-fields is the observation that molecules are composed of the same set of common structural units. These units have the same physical properties (e.g. bond lengths, vibrations, van der Waals radii), independent of what molecule they are part of. Force-fields are the collections of energy equations parameters of the different di-, tri- and tetra-atomic structural elements, such as bonds, angles, torsions and the functions in which the parameters are applied to describe energy terms of distorting the given unit from the optimal (lowest energy) structure. **Eq. 4** shows the general equation accounting for molecular mechanics energy calculations.

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \quad (4)$$

Here the energy terms are referred to as the stretching, bending, torsional, van der Waals, electrostatic and cross terms energies. The stretching and bonding terms are written as harmonic potentials, where the displacement is described as the difference between the equilibrium and actual distance (stretching) or angle (bending). As the torsional energy must show periodicity, the  $E_{tors}$  term is written as periodic cosine functions. The van der Waals energy is described with a Lennard-Jones potential and accounts for the repulsion or attraction between atoms that are not connected, while the electrostatic energy is calculated as a Coulomb potential between atoms with charge. Finally, cross terms are combinations of the abovementioned fundamental five terms. Applying cross terms in the force-field helps taking into account the molecular structure dependency of the optimal values of specific molecular elements. The exact form of these functions is force-field specific. Notable examples of force-fields, usually applied in the modeling of proteins, carbohydrates and nucleic acids are AMBER (FF14SB) [98], CHARMM (CGenFF) [99], OPLS (OPLS3) [100] and GROMOS [101]. These force-fields basically use the previously mentioned forms of the energy terms; however, the inclusion of cross terms is less frequent in force-fields that are mainly used for protein modeling. Once the force-field is selected, MM calculations consist of the evaluation of the energy at molecular geometries generated by algorithms like energy minimization, Monte-Carlo or molecular dynamics simulations implemented in various program packages.

### 3.2 Quantum mechanics

Quantum mechanics (QM) related energy calculations are more elaborate than the simplified model of molecular mechanics and consider atoms and molecules as a group of

nuclei surrounded by electrons. There are two major groups of QM theories, based on the basic description of the electrons. Wave-function methods are based on solving the Schrödinger equation, while density functional methods are based on the Hohenberg-Kohn theorems for the electron density of many electron systems.

### 3.2.1 Electronic structure based methods

These methods are based on the Schrödinger equation (**Eq. 5**) and often referred to as *ab initio* (from the beginning in latin) methods since the only inputs for such calculations are physical constants.

$$H(r)\psi(r) = E\psi(r) \quad (5)$$

where  $H$  is the Hamilton operator,  $E$  is the energy of the system,  $\psi$  is the wave-function and  $r$  represents particle coordinates.

The form of the Hamilton operator for a general  $N$  particle system contains the kinetic  $T$  and the potential  $V$  energy operator. These can be written as shown in **Eq. 6**.

$$H = T + V = \sum_{i=1}^N \frac{\hbar^2}{2m_i} \nabla^2 + \sum_{i=1}^N \sum_{j>i}^N V_{ij} \quad (6)$$

As the nuclei are significantly heavier than the electrons the Schrödinger equation can be separated into a part which describes the electronic wave function for a fixed nuclear geometry and another part which describes the nuclear wave function. This separation is called the Born-Oppenheimer approximation. Applying this approximation, the electronic Schrödinger equation (**Eq. 7**) can be solved parametrically, based on the coordinates of the nuclei. A series of such calculations with different nuclei geometries yields the Potential Energy Surface (PES) of the examined system.

$$H_e\psi_e(R, r) = E_e\psi_e(R, r) \quad (7)$$

where subscript  $e$  belongs to properties of electrons, while  $R$  and  $r$  refer to coordinates of the nuclei and the electrons, respectively. Further approximation can be made with the introduction of one-electron functions. According to the mean-field model, the movement of a given electron is affected by the potential of the nuclei and the mean-potential of the surrounding electrons. Using a single Slater determinant – built by one-electron functions called molecular orbitals – the Hartree-Fock (HF) (**Eq. 8**) equations can be derived applying the variational principle.

$$F\phi_i = \epsilon_i\phi_i \quad (8)$$

where  $F$  denotes the Fock operator,  $\phi_i$  is a molecular orbital,  $\epsilon_i$  is an expectation value of the Fock operator in the molecular orbital basis and can be interpreted as the energy of the molecular orbital. Molecular orbitals can be built as the linear combination of atomic orbitals. This is a set of functions – usually Gaussian or Slater type – and is referred to as a basis set. Writing the molecular orbitals as the linear-combinations of basis-functions yields the Hartree-Fock-Roothaan method.

With a sufficiently large basis the Hartree-Fock wave function is able to account for 99% of the energy on one hand, on the other hand the residual 1% has the same order of magnitude

as chemical phenomena related energies. The missing electron correlation leads to serious deviations from experimental results. Methods considering the electron correlation (often termed post Hartree-Fock methods), normally use the HF solution as starting point. The most common correlation methods are Configuration Interaction [102], Perturbation theory [103] and Coupled Cluster methods [104].

### 3.2.2 Density functional theory methods

Density functional theory is based on the theorem of Hohenberg and Kohn [105], namely that the ground state energy of molecules is determined completely by the electron density  $\rho$ , the square of the wave function. The theory was transformed into an effective approximate method by Kohn and Sham [106]. Instead of the wave function containing three (or four if spin is also considered) coordinates for each electron, the density consists of three coordinates independent of the size of the system. Though, the relation between the density and energy is known, the functional that connects them is unknown. The general DFT equation is shown in **Eq. 9**.

$$E_{DFT}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \quad (9)$$

where  $T_S$  is the kinetic energy of the non-interacting system,  $E_{ne}$  is the nuclei-electron attraction,  $J$  is the Coulomb and  $E_{xc}$  is the exchange-correlation part of the functional. However, a major problem with the  $E_{xc}$  part is its dependency on the total density of the system.

The exact  $E_{xc}[\rho]$  would provide the exact total energy, including electron correlation; however, evaluating such functional would be as demanding as solving the Schrödinger equation exactly. The main advantage of DFT methods is the application of computationally more feasible, approximate terms for  $E_{xc}[\rho]$ . A usual approach estimating the  $E_{xc}[\rho]$  is to separate it to an exchange ( $E_x[\rho]$ ) and correlation ( $E_c[\rho]$ ) term (**Eq. 10**).

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho] \quad (10)$$

Based on the form of the exchange and correlation terms, various DFT methods have been proposed. Local density approximation (LDA) treats the electron density locally as a density of uniform electron gas. An improvement with respect to LDA is the Local Spin Density Approximation (LSDA), which differentiates between the  $\alpha$  and  $\beta$  (spin) densities. Examples of LDA methods are the Vosko, Wilk, Nusair (VWN) [107] and Perdew, Wang (PW91) [108] functionals. Gradient Corrected Methods (GGA) leave the uniform gas approximation and augment the functional with terms containing the gradient of the electron density. Several gradient corrected exchange terms were proposed including those by Perdew and Wang (PW86) [109] and by Becke (B or B88) [110]. Lee, Yang and Parr introduced a gradient corrected term for the correlation energy (LYP) [111], as well. More elaborate functionals known as hybrid methods were also proposed which take into account the exact exchange energy. Notable examples of hybrid methods are B3LYP, M062X or  $\omega$ B97XD [112]. The inclusion of dispersion correction [113,114] in DFT functionals is a common practice to increase the accuracy of calculations. A notable example of such functionals is the previously mentioned  $\omega$ B97XD.

### 3.2.3 Semi-empirical methods

The main drawback of both electronic structure and DFT methods is the high demand of computational resources. Attempts to reduce the computational cost, while maintaining a reasonable accuracy for the calculations led to the introduction of semi-empirical methods.

Such methods usually consist of two types of simplifications, namely the neglect of certain integrals (mostly, more demanding two-electron integrals) and the introduction of parameters, based on experimental values, substituting remaining integrals. A basic type of semi-empirical electronic structure methods is the Neglect of Diatomic Differential Overlap Approximation (NDDO) [115] whose widely used variants include MNDO [116], AM1 [117] and PM3 [118].

Semi-empirical methods derived from density functional theory were also proposed. The self-consistent-charge density functional tight-binding (SCC-DFTB) [119] method contain neglect, approximation, and parametrization of interaction integrals and is a large step ahead from the original non self-consistent version of DFTB [120]. A more recent semi-empirical version DFTB3 [121] keeps the advantages of the SCC-DFTB and introduces an improved  $\gamma^h$  correction function ameliorating the electron-electron interaction of charge fluctuation, adds a purely empirical parameter improving the description of hydrogen bonding system and extends the DFT exchange-correlation energy function of SCC-DFTB with third order terms of the Taylor series expansion.

Semi-empirical methods have a lot in common with force-field methods, however the main difference is the ability to describe bond formation and disruption. The performance of both approaches is system dependent, although this is less crucial for semi-empirical methods than for force-field methods. Semi-empirical calculations are computationally more demanding than force-field methods, however they are still more economical than *ab initio* or DFT methods.

### 3.3 Hybrid QM/MM potential

Quantum chemistry methods perform well in the description of bond forming and breaking, however biological systems, like solvated enzyme-ligand complexes contain a large number of atoms, and it is not feasible to handle them either with *ab initio* or DFT based methods. In cases, where a large number of energy evaluations is required, like in molecular dynamics, even the application of semi-empirical methods for the whole system is impossible. To overcome this issue and to model covalent reactions in enzymatic environment the coupling of the QM and MM potential is highly desirable. The chemically interesting region is treated with *ab initio*, DFT or semi-empirical QM potential, while the rest of the system is handled by force field equations and parameters [122,123].

QM/MM calculations have several variants and their most important features are summarized in **Table 1** [39] and in **Figure 6**.

**Table 1.** Aspects of the QM/MM calculations. Examples are focusing on reactivity calculations.

<i>Technical detail</i>	<i>Possibilities, examples</i>
<b>Energy calculation scheme</b>	Additive Subtractive
<b>Definition of the QM region</b>	Atoms of the inhibitor molecule + atoms of the reactive sidechain Adding a buffer region (ONIOM)
<b>Subsystem separation</b>	Link atom approach Frozen orbitals Boundary atoms
<b>QM level</b>	DFT functionals (DFTB3, B3LYP, $\omega$ B97XD, PBE) <i>Ab initio</i> (MP2) Semi-empirical (DFTB3, PM3, AM1)
<b>Force-field</b>	AMBER (FF14SB, GLYCAM, GAFF) OPLS3 CHARMM
<b>Interaction scheme</b>	Mechanical embedding (ME) Electrostatic embedding (EE)
<b>Type of calculation</b>	Geometry optimization Single point calculation Adiabatic scan Molecular dynamics
<b>MD method</b>	None Steered MD Umbrella sampling Metadynamics FEP

The total energy of the complete QM/MM system can be handled either by the additive or the subtractive method [124], described by **Eqs. 11** and **12**, respectively.

$$E_{total} = E_{QM(QM)} + E_{MM(MM)} + E_{QM-MM \text{ interaction}} \quad (11)$$

$$E_{total} = E_{MM,total} + E_{QM(QM)} - E_{MM(QM)} \quad (12)$$

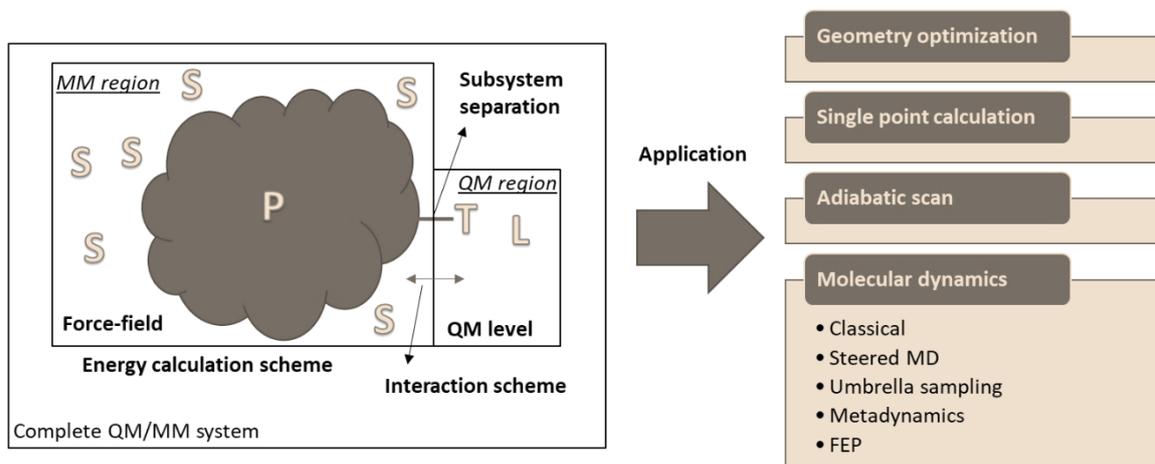
The QM region usually consists of ligand atoms or at least the warhead region of the ligand and either the entire nucleophilic sidechain or the chemically interesting part of it and additionally further relevant sidechain parts. The remaining atoms are treated with force-field parameters. The best practice of separating the two regions is avoiding the cut of chemical bonds, though sometimes this is inevitable. Cutting covalent bonds leads to fragments with dangling bonds. The main approaches to alleviate artificial errors caused by the missing valence are the link atom [125], the frozen localized orbital [126–128] and the boundary atom [129,130] schemes. If the bond cutting is necessary, then the best scenario is to cut an apolar C-C bond, since cutting polar bonds like C-N or C-O bonds may lead to serious errors in the QM/MM calculations. Concerning protein sidechains,  $C\alpha$ - $C\beta$  bonds are suitable candidates for bond cutting yielding a methyl-thiol QM fragments when cysteines are treated with the link atom approach.

The interaction between the two subsystems is also crucial. There are two main embedding schemes applied in QM/MM calculations. Mechanical embedding (ME) treats the interaction between the two sub-systems at MM level and QM calculations are performed in the absence of the MM environment. Electrostatic embedding (EE) considers the MM

environment and QM/MM interactions by the inclusion of MM charges in the QM Hamiltonian. As force-fields contain a great amount of point charges the application of electrostatic embedding is more common in QM/MM enzyme calculations.

The level of theory used for the QM atoms is another crucial feature of QM/MM calculations. Common *ab initio*, DFT and semi-empirical potentials are used extensively, though one has to consider the accuracy/computational demand ratio before setting up the system. QM/MM MD simulations typically allow less sophisticated potentials like semi-empirical ones, while optimization and single point type calculations are performed with more sophisticated DFT or electronic structure methods. Some common choices for the force field are AMBER [98], OPLS [100], GROMOS [101] and CHARMM [99].

QM/MM potentials were applied successfully for a large number of protein-ligand complexes, including potential energy calculations, mechanistic studies and molecular dynamics simulations evaluating free energy relations of the covalent inhibition reaction [39,123].



**Figure 6.** Graphical representation of the fundamental aspects of QM/MM simulations. (P: protein; T: targeted sidechain; L: ligand; S: solvent)

### 3.4 Molecular dynamics

Classical molecular dynamics (MD) describes the interaction and motion of atoms and molecules according to Newton's physics. The forces between atoms are estimated by the applied force-field and by the QM method applied for the QM region if QM/MM potential is used. The evolving structures over the simulation time passed is calculated by integrating the Newton's laws of motions, providing trajectories with atomic positions and velocities. Molecular dynamics can be performed in various ensembles based on which physical property of the simulation is kept constant. The name of these ensembles derived from the constantly kept properties, like temperature (T), pressure (P), volume (V), the system's particle number (N) and the system's total energy (E). Common ensembles are the NPT, NVT (canonical ensemble) and NVE (microcanonical ensemble). Maintenance of the constant pressure and temperature are done by barostats and thermostats, respectively. Constant volume is achieved by the application of periodic boundary conditions (PBC) meaning that the simulation box is a unit cell repeated in all dimensions infinitely, therefore particles leaving the unit cell enter the

next cell, hence appear in the opposite border of the cell. Applying PBC allows the calculation of nonbonded interactions between unit cells.

Valuable information can be extracted from MD simulations, including structural, kinetic and thermodynamic properties. Additionally, MD simulations, analyzed by visual inspections might also serve with important conformational and mechanistic observations. Special technics such as steered molecular dynamics (SMD), umbrella sampling, metadynamics or free energy perturbations [31] are extensively used and their optimal choice is problem and resource dependent. The following section gives a short overview on MD technics related to my thesis.

### 3.4.1 Unbiased molecular dynamics

MD simulations without any additional biasing potential are called unbiased molecular dynamics. Unbiased simulations coupled with statistical mechanics can be used calculating the equilibrium or association constant ( $K_a$ ) and binding free energies ( $\Delta G$ ) based on the number of bound and unbound conformations during the simulation:

$$K_a = \frac{[PL]_{eq}}{[P]_{eq}[L]_{eq}} \quad (13)$$

$$\Delta G = -k_b T \cdot \ln(K_a C^0) \quad (14)$$

where  $k_b$  is the Boltzmann constant,  $T$  denotes temperature,  $C^0$  is a constant defining the standard 1 M concentration.

Such simulations also serve information about stability of certain conformations, dynamic properties of structure regions or the occurrence of certain events. The main drawback of unbiased simulations is the poor sampling of high energy regions. For stronger binding events unbiased MDs are unable to recover transition rates. These systems require the application of enhanced MD calculations, where an additional biasing potential is introduced attempting to escape the Boltzmann statistics while retaining the correct distribution of statistical ensembles. The following methods are all among the enhanced sampling methods of MD [131].

### 3.4.2 Steered molecular dynamics

Steered molecular dynamics (SMD) [132,133] introduces a time dependent biasing potential, forcing the studied process, e.g. a chemical reaction to proceed during the simulation and enabling the sampling of the high energy regions along the reaction coordinate. The obtained work curve can be used to construct the potential of mean force (PMF) profile using the Jarzynski equation [134,135], however, it requires a series of simulations. It is more common to use SMD simulations to generate starting structures for subsequent umbrella sampling (US) simulations and constructing the PMF from the results of the US. Alternatively, SMD is an excellent tool to study the unbinding of the ligand [136]. The key parameters of SMD simulations are the velocity of the pulling and the associated force constant. Pulling the biasing potential too fast results in highly non-equilibrium sampling, while slow pulling requires higher computational resources. Regarding the force constants, high values might prevent the desired level of sampling during the simulation on one hand, on the other hand too low bias might cause the system unable to overcome energy barriers and to be stuck in a local minimum.

### 3.4.3 Umbrella sampling

Umbrella sampling [137] consists of a series of simulation windows with biasing potentials varying along the studied coordinate. The biasing potential aids the simulations to overcome energy barriers and to reach the desired level of sampling. To construct the PMF of the studied event, post-processing methods are required. One of the most popular methods is the weighted histogram analysis method (WHAM) [138], which reconstructs the PMF from the biased distributions of the simulation trajectories. The main parameters of US simulations are the applied force constant and the distance between the center of the biasing potential in neighboring windows. For reliable results the distributions between neighboring windows should overlap, which requires careful parameter settings.

### 3.4.4 Thermodynamic integration

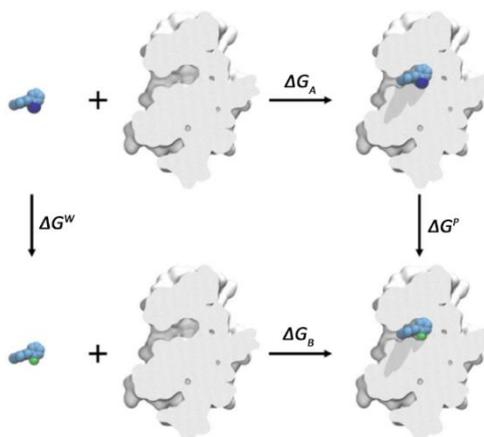
The free energy difference between states A and B can be obtained by thermodynamic integration according to **Eq. 15**.

$$\Delta F(A \rightarrow B) = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda = \int_0^1 \langle V_B(\lambda) - V_A(\lambda) \rangle_{\lambda} d\lambda \quad (15)$$

where  $\langle \rangle_{\lambda}$  denotes ensemble average and the right-hand side is obtained by introducing a hybrid potential as a linear combination of the two states (**Eq. 16**) with a suitably chosen parameter  $\lambda$ .

$$V(\lambda) = (1 - \lambda)V_A + \lambda V_B \quad (16)$$

Free energy differences between ligand related events can be calculated with properly set thermodynamic cycles, applying **Eq. 15**. It can be pictured as ligand A smoothly transforms into ligand B, while the  $\lambda$  value raises from 0 to 1. If this transformation is performed in both bulk solvent and in enzyme environment the binding free energy difference ( $\Delta\Delta G$ ) can be calculated (**Eq. 17**). These alchemical transformations are more efficient than transforming unbound ligand and protein into the bound complex (horizontal arrows in **Figure 7**).



**Figure 7.** Example of a thermodynamic cycle estimating the binding free energy difference between ligand A and B taken from Ref [31].

$$\Delta\Delta G = \Delta G_B - \Delta G_A = \Delta G^P - \Delta G^W \text{ }^1 \quad (17)$$

More elaborate cycles can be constructed depending on the desired level of accuracy or the nature of the transformation. The mutation of ligand A to ligand B is commonly separated to decoupling sub steps. In a typical scenario, the charges of the perturbed atoms are removed, thereafter, the atoms containing only van der Waals radii are transformed into the B state's decharged counterparts. Finally, the charges are reintroduced to the transformed atoms of state B completing the sub cycle. The effect of sidechain mutations to ligand binding can be also studied by thermodynamic integration allowing the computational evaluation of enzyme selectivity towards different ligands. In this scenario state A is related to the enzyme with the original sidechain while state B includes the mutated residue. The complete cycle contains this transformation both with apo and ligand-bound enzymes [139].

I note that free energy perturbation (FEP) [140], introduced by Zwanzig [141] is another technique to calculate binding free energy differences. According to the FEP method, free energy difference between state A and B is calculated using the Zwanzig equation with values derived from MD or Monte-Carlo simulations.

---

<sup>1</sup> Note that in typical applications the free energy difference ( $\Delta\Delta G$ ) for several ligand pairs are evaluated and an arbitrarily selected constant is added to obtain a set of ligand binding free energies ( $\Delta G$ ).

## CHAPTER 4 – MATERIALS AND METHODS

The current section gives technical information on the applied methodologies and calculations. For the theoretical background please refer to the previous two chapters and the references therein. Please note that the structure of the molecules mentioned in this chapter are shown in **Chapter 5 – Results and discussion**.

### 4.1 Molecular mechanics (MM) based methods

#### 4.1.1 Protein and ligand preparation

Protein preparation was performed by Schrödinger's Protein preparation wizard implemented in Maestro [142,143], starting from available PDB structures, unless otherwise stated. The preparation workflow consisted of the filling of missing loops and sidechains, optimizing the hydrogen bond network, removing solvent water molecules and minimizing the remaining atoms, applying the default settings for all the steps. The preparation of the molecules was performed by the LigPrep module of Maestro [144]. Protonation states of compounds were estimated for the target pH of 7±2.

**MurA.** The ternary complex of *E. coli* MurA with fosfomycin and UNAG (PDB: 1UAE) [45] was used as starting structure for MurA related simulations.

**KRAS.** The structure of the initial KRAS<sup>G12C</sup>-compound **1** complex was extracted from PDB: 4NMM [145]. After the protein preparation the non-covalent enzyme-ligand complex was reconstructed through the following steps. The covalent bond between Cys12 and **1** was deleted, while a chlorine atom was introduced to recover the prereactive form of the ligand and the negatively charged cysteine was protonated. The steric clashes were alleviated by restrained minimization [142,143]. Further non-covalent complexes were constructed by modifying compound **1** and minimizing the created ligands in the enzyme environment.

**EGFR.** For activity prediction the EGFR<sup>L858R/T790M</sup> structure complexed with **HKI-272** was extracted from PDB: 3W2Q [146]. After the application of the preparation wizard's protocol the covalent bond between the protein and **HKI-272** was deleted. The bond order between the C $\alpha$  and C $\beta$  of the warhead was increased, while the thiolate was protonated. The obtained complex was energy minimized. Examined EGFR-inhibitor complexes were constructed as described in the KRAS section, modifying **HKI-272**.

The structure of the double mutant EGFR<sup>L858R/T790M</sup> for selectivity prediction was extracted also from PDB: 3W2Q. The binding mode of the selected inhibitors was derived from the cSrc-DM-**RL40** (PDB: 5D12) complex as described by Engel and coworkers [147]. The two enzymes' structures were aligned by Maestro's built-in module, then, **RL40** was extracted into the double mutant EGFR's active site and was modified into compounds **17** and **18** (**Figure 39**). Complexes of the single mutant EGFR were constructed by mutating Met790 into a threonine residue.

**ITK, BTK, BMX.** As these kinases share similar structural characteristics the complexes of ITK, BTK and BMX with ligand **16** (**Figure 34**) were constructed, assuming similar binding mode for all the enzyme-ligand pairs as it was observed for **7KV** complexed with JAK3 (PDB: 5TTU [148]). As ITK showed the lowest K<sub>i</sub> value against **16**, the ITK-**16** complex was built first and was modified into the hybrid structures of BMX and BTK. After protein structure preparation and alignment of ITK (PDB: 3MIY [149]) and JAK3, **7KV** was exported into the active site of ITK and was transformed into **16**, followed by constrained minimization,

alleviating steric clashes. The hybrid structures of BTK and BMX were prepared by mutating specific residues of the ITK structure identified by structural and sequential alignment and visual inspection of the ITK, BTK and BMX complexes. Free hybrid enzymes were constructed by the removal of **16** from the complexes.

**Immunoproteasome.** The starting structure for Immunoproteasome related studies was extracted from PDB: 5M2B [150], the X-ray structure of 10S yeast proteasome including the human  $\beta 5i$  subunit complexed with thiazole-based inhibitor **Ro19**. Only the  $\beta 5i$  subunit was kept and was refined with the Protein preparation wizard, the rest of the PDB structure was deleted. The structure of the constitutive proteasome (cPS) for free energy calculations with ligand **HT1146** and **HT2004** was extracted from PDB: 4R67 [151] and was refined similarly.

#### 4.1.2 Ligand docking

**MurA.** Constrained ligand docking was performed with Glide [152–155] into the active site of the enzyme with the active Cys115-His394 ion pair form, extracted from the QM/MM MD US simulation of the Cys115-His394 proton transfer (see below) at the product state. Positional constraints were applied keeping the warheads of the selected compounds near the nucleophilic negatively charged Cys115 (**Figure A1**). For oxirane derivatives, two of the three atoms of the oxirane ring were constrained to be within 5 Å separation from the middle point of the Cys115-S and His394-N segment. The same constraint was applied for the connected  $sp^2$  carbon atoms of the Michael acceptors and PEP. For halogenated compounds, the carbon bound to the halogen was constrained to be within 4 Å separation from the sulfur of Cys115. A cubic grid with 20 Å edge length was centered on four active site residues: Cys115, Arg120, His394 and Arg397. No side-chain rotation was allowed during docking, the number of poses written was set to 10. Ligands were docked using the OPLS2005 force-field. [156]

**Immunoproteasome.** To obtain a sensible structure for the **HT1146**-iPS complex ligand docking was used. The previously prepared **Ro19**-iPS complex was used to grid creation, selecting **Ro19** as the center of the grid having  $25 \times 25 \times 25 \text{ \AA}^3$  outer and  $10 \times 10 \times 10 \text{ \AA}^3$  inner box volume. **HT1146** was docked into the active site of the  $\beta 5i$  subunit, and the ten best poses were written out. The most feasible pose was selected by visual inspection for further free energy related simulations. Both grid generation and docking were performed with Glide, applying the OPLS2005 force-field. [156]

#### 4.1.3 Classical MD simulations

All of the simulations mentioned in this section were performed with Desmond [157,158] applying the OPLS3 [100] force-field. The simulations were performed using periodic boundary conditions.

**MurA.** To obtain a starting structure for the mechanistic QM/MM studies on the Cys115-His394 proton transfer of MurA, classical molecular dynamic was used. Fosfomycin was deleted from the prepared ternary complex of PDB: 1UAE and the UNAG-MurA complex was immersed into a box of TIP3P [159] waters. The system was relaxed with the default Desmond protocol [158] and 250 ns NVT (constant particle number, volume and temperature) simulation was performed at 300 K providing the starting structure for subsequent simulations on the catalytic mechanism and inhibition of MurA. Snapshots from MD simulations were clustered using Desmond's Trajectory Clustering application. The backbone of the protein was used for calculating the RMSD. 200 frames out of the 2000 snapshots were included in the clustering producing 10 clusters. The center of the main cluster containing 132 frames was used for

investigation of the proton transfer between Cys115 and His394. This structure provided a binding pocket that was able to accommodate the inhibitors and was selected for further studies.

Initial ligand binding modes were obtained by docking (see above) and the stability of the binding modes was confirmed by 20 ns MD simulations starting from five different docking poses selected from the original ten by visual inspection for each ligand. The same Desmond relaxation protocol was applied (TIP3P box, NVT conditions, 300 K) before the productive MDs. Starting complexes for subsequent simulations were obtained from the MD simulations with the docked ligands by selecting snapshots with appropriate orientation of the ligand warhead and cysteine for simulating the covalent reactions.

#### 4.1.4 Thermodynamic integration

All of the simulations related to the thermodynamic integration studies were performed applying periodic boundary conditions with the particle mesh Ewald method and an 8 Å cutoff distance for the nonbonded interactions.

**KRAS, EGFR.** System preparations and thermodynamic integrations were performed using the AMBERtools 18 [160] program package. The systems were constructed by immersing the ligands and the protein-ligand complexes into an octahedral box of TIP3P water molecules followed by neutralization by sodium ions. The constructed systems were relaxed by the following protocol: 1000 steps of steepest descent minimization, 20 ps NVT heating to 300 K, followed by 200 ps of NPT (constant particle number, pressure and temperature) equilibration with the  $\lambda$  coupling parameter set to 0.5. The alchemical transformations were separated into 3 sub steps. The charge removal from the softcore atoms of the reference molecule (decharge step), the modification of the neutral softcore atoms into the perturbed structure (Van der Waals, vdw step) and finally the reintroduction of the charges into the perturbed neutral atoms (recharge step). Every step in both solvent and complex systems contained 11 windows with coupling parameter ranging from 0.0 to 1.0 with 0.1 increments. Every  $\lambda$  window consisted of a 20 ps heating to 300 K in NVT conditions and a 200 ps long NVT productive simulation with 1 fs timestep. Alchemical transformations with non-charge conserving modifications were performed applying the “dummy co-alchemical ion” approach [161] by adding chloride ion into the softcore region of the positively charged ligand to maintain system neutrality throughout the transformation steps. NOSHAKE was applied for the softcore atoms, while SHAKE was applied for the rest of the systems. The general schemes of TI transformations are shown in **Figure A2-A6**.

Calculations of the binding free energy differences of compounds **17** and **18** in the EGFR single and double mutants were carried out as described above. The binding free energy difference of the same ligand in different mutants was calculated as follows. The thermodynamic cycles consisted of the one step (single topology model) modification of the single mutant into the double mutant both with and without the presence of the ligand. The two enzyme structures aligned to each other were immersed into an octahedral box of TIP3P water molecules and were neutralized by sodium ions. The identical residues were merged using ParmEd. The relaxation protocol consisted of a 1000 steps steepest descent minimization, a 20 ps NVT heating to 300 K and a 200 ps NPT equilibration. The Tyr790Met alchemical transformations consisted of 11 windows, with  $\lambda$  values between 0.0 and 1.0 with 0.1 increments. The windows contained 20 ps NVT heating to 300 K and 200 ps NVT productive MDs. Atoms of the perturbed residue were handled as softcore atoms. The simulations were performed with 1 fs timestep, applying the FF14SB force-field. Temperature regulation was

handled by Langevin thermostat, while constant pressure was achieved by the application of the Berendsen barostat.

**Immunoproteasome.** System preparation and thermodynamic integration were carried out using the AMBERtools18 [160] software package. Ligand- $\beta$ 5i complexes were generated assuming the same binding mode for all iPS inhibitors, using the selected docking pose of **HT1146**. Further compounds were generated by modifying **HT1146** into the actual molecules, maintaining the original binding mode. System preparation was carried out with tLEaP. Inhibitor-protein structures were immersed into an octahedral box of TIP3P water molecules and system neutralization was achieved by adding sodium ions. The solvent box preparation was followed by a multistep relaxation protocol, namely 1000 steps of steepest descent minimization, 50 ps of NVT heating to 310.15 K and 1 ns of NPT equilibration. The above steps were performed for the solvated ligand and complex systems including their recharge, discharge and vdw structure at the  $\lambda$  value of 0.5 (**Figure A2-A6**); as the alchemical transformations followed the same protocol described at the previous section. All steps were performed in solvent and enzyme environment, completing the thermodynamic circle. Each step used windows with evenly separated lambda coupling parameter values ranging from 0.0 to 1.0 with 0.1 increment. Every lambda window consisted of a 20 ps long heating to 310.15 K in NVT ensemble and a 1 ns long productive MD run, also in NVT ensemble. SHAKE was enabled for non-softcore atoms only. All of the abovementioned simulations were performed with 1 fs timestep, applying the FF14SB force-field. Temperature regulation was handled by Langevin thermostat with the collision frequency of 2 ps<sup>-1</sup>, while pressure was maintained by the Berendsen barostat with 2 ps relaxation time.

Calculation of the binding free energy difference for **HT1146** and **HT2004** in cPS was performed similarly. Structure of the cPS was extracted from PDB: 4R67. The covalently bound ligand was deleted and the neutral form of Thr1 was rebuilt. Binding poses of **HT1146** and **HT2004** in cPS were imported from the iPS enzyme-ligand complexes after structure alignment of iPS and cPS by Maestro's built-in tool. System preparation and the thermodynamic integration were identical to those described above for iPS.

iPS-cPS selectivity calculation by thermodynamic integration was performed using the "hybrid active site" approach (see below). The binding free energy difference of a selected ligand's binding in two related enzymes is calculated by transforming the first enzyme's active site into the other in the presence of the ligand molecule and without it. Binding free energy differences of **HT1146** and **HT2004** in iPS and cPS were estimated by transforming the active site of iPS into a "hybrid" cPS, namely applying the Gln53Ser mutation to the original iPS structure. The thermodynamic integration followed a single topology model both with and without the inhibitor molecules. The overlapping backbones of the original and mutant proteins were merged with ParmEd. Both systems were immersed in an octahedral box of TIP3P waters and relaxed by the same protocol as described for iPS above. System relaxation and productive MDs was carried out similarly as described previously (**Figure A2-A6**).

#### 4.1.5 pK<sub>a</sub> calculation of protein sidechains

**MurA.** The pK<sub>a</sub> value of MurA's Cys115 and His394 was estimated by thermodynamic integration. Enzyme structure extracted from the classical MD simulations was used as starting conformation. The ligand was deleted, while the identical amino acids of the enzyme structures containing both protonation forms of the studied residue were merged with ParmED. The thermodynamic cycle contained also the examined amino acid capped with an NME and an

ACE group in both states merged to each other. TI with sidechain mutation was performed as described previously following the single topology method.

## 4.2 Quantum mechanics (QM) based methods

### 4.2.1 Model system preparation with cysteine surrogate methyl-thiolate

Initial structures of the reactant, product and TS structures were constructed with the conformational search of Maestro's MacroModel module. OPLS3 was applied for the conformational search, all the atoms were set moving freely, the Minimization and CSearch panels contained default settings. In case of the TS structure searches the C-S bond between the surrogate thiolate and the warhead's carbon was constrained to maintain the initially set value, taken from previous theoretical studies [36,38,162].

### 4.2.2 Transition state and reaction energy calculations

All of the following QM calculations were performed with the Gaussian 16 program package. The previously constructed reactant and product conformations were subjected to geometry optimization applying the B3LYP/6-311G++(d,p) level of theory with the SMD implicit solvation model [163]. The TS structures were optimized by the same level of theory using the opt=modredundant command with the C-S bond frozen. Redundant conformations after the optimization were eliminated, finally, the remaining structures were reoptimized by the opt=TS keyword.

Both the minima and transition state structures were characterized by frequency analysis, the TS structures were verified by having one "negative" frequency belonging to the C-S bond formation.

The total energy of the reactant states was calculated as the sum of the energy of the lowest energy reactant conformation and the energy of an optimized methane-thiolate molecule. Product state and transition state energies were extracted from the lowest energy conformations. Transition barriers were calculated as the difference between the TS and reactant state energies, while the reaction energies as the difference between the product and reactant states.

### 4.2.3 Compound pK<sub>a</sub> calculations

**EGFR.** Some of the studied EGFR inhibitors contain N-atoms that are potentially protonated at the target pH. For ligands **13-15** Schrödinger's Jaguar was used to estimate the pK<sub>a</sub> values of the selected atoms and estimate their protonation states. The quick accuracy level was set for the SCF method, the convergence scheme used the DIIS method [164], while the solvent pK<sub>a</sub> option was set to water with the implicit PBF solvation model [165]. The calculation process consisted of multiple steps including B3LYP/6-31G\*, B3LYP/6-31G\*\*(+), and B3LYP/cc-pVTZ(-f)(+) calculations for geometry optimization, single point solution phase and single point energy determination, respectively.

## 4.3 Hybrid QM/MM based methods

All the QM/MM MD simulations were performed with the AMBERtools18 [160] program package applying periodic boundary conditions with the particle mesh Ewald method and a 8 Å cutoff distance for the nonbonded interactions.

### 4.3.1 Preparatory steps

**MurA.** Starting structures for the simulation of the Cys115-His394 proton transfer and the covalent inhibition reaction – extracted from the classical MD simulations – were immersed

into a box of TIP3P water molecules using tLeAP, with the distance and closeness parameters set to 10.0 Å and 1.0, respectively. The solvated systems were neutralized by sodium ions. Thereafter, the obtained complexes were separated into QM and MM sub regions. The QM region consisted of the ligand molecules and the Cys115 and His394 sidechains cut between the  $C\alpha$ - $C\beta$  bond. The MM region was constructed from the remaining atoms (**Figure A7**). The QM region was treated by the DFTB3 [121] functional, while the MM part was described by the FF14SB force-field [98] and GAFF [166] with 8.0 Å non-bonded cutoff. SHAKE [167] was enabled only for the MM region. The constructed systems were subjected to a relaxation protocol, including energy minimization with 500 steepest descent and 500 conjugate gradient steps, 20 ps heating to 300 K in NVT ensemble using Langevin thermostat and finally, 100 ps equilibration in NPT ensemble applying Berendsen barostat with 5.0 ps pressure relaxation time and Langevin thermostat. The relaxed complexes were used in steered MD and Umbrella sampling simulations.

**KRAS.** Starting structure was extracted from PDB: 4NMM, which was refined with the protein preparation wizard and was subjected to restrained minimization. Thereafter, the transition state structure of the covalent bond formation was constructed. The covalent bond between the sulfur and carbon atom was deleted, the torsion angle of O=C-C-H was rotated by 180°. The missing valence of the carbon was filled with chlorine, then, the C-S and C-Cl distances were set to 2.4 Å (RC=0), which belongs to the TS distances observed in preliminary QM/MM simulations at DFTB3/FF14SB level. The constructed system was immersed into an octahedral box of TIP3P waters and was neutralized by sodium ions using tLeAP. The QM region, handled by DFTB3 functional consisted of the ligand's chemically reactive part cut within the linker region (**Figure A7**) and the cysteine residue cut between the  $C\alpha$  and  $C\beta$ . The remaining atoms were part of the MM region, treated by the FF14SB force-field. The dangling bonds at the QM/MM border were filled by hydrogens (link atom approach). The obtained complex was relaxed by the protocol of 500-500 steps of steepest descent and conjugate gradient minimization, 20 ps NVT heating to 300 K and 100 ps NPT equilibration at 1 atm. Restraints with 500 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> force constant was applied to the distances constructing the reaction coordinate to maintain the TS geometry. Further protein-ligand complexes were prepared with the same protocol modifying the ligand in PDB: 4NMM.

**EGFR.** The covalent enzyme-inhibitor complexes were created by transforming the covalently bound acrylamide derivative **HKI-272** (**Figure 28**) in PDB: 3W2Q. The obtained complexes were minimized by Schrödinger's MacroModel [168]. The TS structures were constructed following the proposed deprotonation of Cys797 by Asp800 [169]. The covalent bond between the active site cysteine's sulfur and the ligand's carbon atom was deleted, the Asp800 residue was protonated, while the double bond between the  $C\alpha$  and  $C\beta$  of the ligand was rebuilt. To achieve the TS geometry previously observed in QM/MM MD simulations [170], the distance between the sulfur and reactive carbon of the ligand was set to 2.1 Å. The constructed systems were subjected to the QM/MM relaxation protocol described for KRAS. Atoms of the ligands and the Cys797 residue cut between the  $C\alpha$  and  $C\beta$  formed the QM region, while the remaining atoms were part of the MM part (**Figure A7**).

**ITK, BTK, BMX.** The covalent complexes were generated by creating a covalent bond between ligand **16** and the active site cysteine of the enzymes (ITK, hybrid BTK and BMX used in the TI calculations). After restrained minimization in Maestro, regeneration and relaxation of the TS structures were performed as it is described in the EGFR section. The QM region consisted of the ligand atoms and the cysteine sidechain cut between the  $C\alpha$  and  $C\beta$ , and the rest of the atoms were part of the MM region (**Figure A7**).

**Immunoproteasome.** The docked **HT1146**-iPS structure was immersed into an octahedral box of TIP3P water molecules. The system was relaxed by 1000-1000 steps of steepest descent and conjugate gradient minimization, followed by 100 ps of NVT heating to 310.15 K and 4 ns of NPT equilibration at the same temperature. All these steps were performed applying only the FF14SB potential. QM/MM potential was introduced in a subsequent 100 ps NVT molecular dynamics simulation at 310.15 K. The QM region, treated with DFTB3 method consisted of the ligand molecule, the Thr1 residue and a C $\alpha$ H-NH fragment of the Thr2 residue. The remaining atoms were treated with the FF14SB force-field and GAFF. The dangling bond between the QM and MM regions was treated with the link atom approach. SHAKE was enabled for the MM region only (**Figure A7**).

Starting structures calculating the rate-determining step with further compounds were extracted from the product state of the first step of the carbonate mechanism with **HT1146**. Solvent molecules were deleted and **HT1146** was modified into the compounds to be examined. The constructed enzyme-ligand complexes were subjected to a shortened QM/MM DFTB3/FF14SB relaxation protocol, including 1000 steps of steepest descent minimization, 20 ps of NVT heating to 310.15 K and 100 ps of NPT equilibration at 310.15 K. Interatomic distance constraints were applied for all relaxation steps in order to maintain characteristic distances and arrangements between ligands and the enzyme's active site.

Starting structures for the PMF calculations belonging to the rate determining step in cPS inhibition were constructed similarly as in iPS. The non-covalent complex of cPS (PDB: 4R67), previously used in TI simulations were modified by deprotonating the O $\gamma$  of Thr1 and protonating the terminal amino group, thus, reconstructing the intermediate state before the rate-determining step. The poses of the ligands **HT1146** and **HT2004** were extracted from the iPS-ligand complexes after protein structure alignment. Solvation and relaxation were carried out as described in the previous paragraph.

#### 4.3.2 Steered molecular dynamics (SMD) simulations

**MurA.** QM/MM steered MD (SMD) simulations were carried out along reaction coordinates specific to the reaction type (**Table A1** and **Figure A8**) to obtain starting structures for the subsequent free energy calculations. SMD simulations with the inhibitor compounds were performed along the selected reaction coordinates at 300 K in NVT ensemble applying the Langevin thermostat. The velocity of the pulling was set to 0.1 Å·ps<sup>-1</sup> using a spring constant of 500 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>. The obtained covalent complexes were relaxed in 100 ps unconstrained MD to obtain starting structures for backward SMDs. Backward SMDs were carried out with the same 0.1 Å·ps<sup>-1</sup> velocity and 500 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> force constant. Finally, frames with 0.1 Å increment were extracted from the backward pullings as starting structures for the Umbrella Sampling windows (**Figure A9**). To obtain starting structures for studying the Cys115-His394 proton transfer steered MDs were carried out as described for inhibitor molecules. SMDs with PEP were performed similarly to obtain starting structures for the 2D US. During the simulations, one out of the two reaction coordinates was frozen at varying values to sample the complete 2D conformational space.

**KRAS, EGFR, ITK, BTK, BMX.** Backward and forward QM/MM steered MDs using specific reaction coordinates (**Table A1** and **Figure A8**) were performed starting from the TS structures of the corresponding enzyme-ligand complexes. SMDs were carried out in NVT ensemble at 300 K applying the DFTB3/FF14SB potential (**Figure A9**). The biasing potential along the reaction coordinate was pulled with the velocity of 0.1 Å·ps<sup>-1</sup> applying the force constant of 500 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> starting from the transition state and finishing at the product and

reactant state observed previously for the examined warheads. Frames from the SMD trajectories with 1 ps intervals were extracted as starting structures for the subsequent US simulations.

**Immunoproteasome.** QM/MM steered molecular dynamics simulations along specific reaction coordinates (**Table A1** and **Figure A8**) were carried out to generate starting structures for umbrella sampling simulations belonging to the **HT1146** mechanistic studies and examining the rate-determining step of covalent inhibition in both iPS and cPS. Complex structures after relaxation were subjected to three consecutive 50 ps SMDs (pulling forward, backwards and forward again) at 310.15 K in NVT ensemble. The biasing potential applied to the reaction coordinates during the pullings was treated with the PLUMED patch (ver. 2.6) [171–173] with the force constant set to  $300 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ . The second forward SMDs were used for frame extraction, structures at 0.1 Å increments along the reaction coordinate were taken from the trajectories. For further information please refer to **Figure A9**.

#### 4.3.3 Umbrella Sampling (US)

**MurA.** QM/MM umbrella sampling was performed using the structures extracted from the preceding SMD trajectories. The applied reaction coordinates were warhead and chemistry dependent; further information is shown in **Table A1** and **Figure A8**. Windows, separated by 0.1 Å consisted of 40 ps simulations in NVT ensemble at 300 K applying  $200 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$  force constant for the reaction coordinate. The first 5 ps for each window was discarded and the remaining 35 ps was used to estimate the potential of mean force (PMF).

In case of the two-dimensional US of PEP, windows with 0.1 Å increments for both dimensions were selected and simulations were carried out at NVT ensemble, applying  $200 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$  force constant for both reaction coordinates starting from the structures extracted from SMD simulations. 5 ps out of the total 10 ps simulation of the windows were discarded and the remaining 5 ps was used to evaluate the 2D PMF.

**KRAS, EGFR, ITK, BTK, BMX.** QM/MM MD US simulations starting from structures extracted from the preceding SMD simulations were performed in NVT ensemble at 300 K. Simulation windows contained 40 ps of productive MD with 1 fs timestep, and the force constant of the biasing potential was set to  $400 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ . The number of windows was warhead dependent. The overlap between neighboring windows was inspected by histogram analysis. For PMF construction the last 35 ps of the simulations was used.

**Immunoproteasome.** QM/MM MD US simulations were carried out for both the mechanistic description of the inhibition reaction between iPS and **HT1146** and also calculating the PMF of the rate determining step of the iPS's and cPS's covalent inhibition by selected oxathiazolone compounds. Simulation windows were separated by 0.1 Å reaction coordinate increment and contained 40 ps NVT productive MDs. The biasing potential was handled by the PLUMED patch applying  $300 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$  force constant for the reaction coordinate. The first 5 ps of all productive MDs was discarded, and the remaining 35 ps was used to generate the PMF curves.

#### 4.3.4 Energy correction of the QM region

**MurA.** The QM/MM free energies obtained from DFBT3/FF14SB umbrella sampling were corrected for the Michael-addition reaction in the following manner. The last frame of the US window's trajectory belonging to the intermediate state was extracted and was QM/MM energy minimized with the DFTB3/FF14SB potential. The minimization included 20 steepest descent and 980 conjugate gradient steps without any constraint. Thereafter, atoms outside a

5 Å radius sphere around the QM region were frozen and the rest of the system was energy minimized with a DRMS criterion of 0.0002 Å. Back and forth scans with 0.1 Å increments were performed allowing only the atoms inside the 5 Å shell to move. Once the potential energy curves were converged, the reactant, intermediate and product states were extracted and their geometry was optimized with the relaxation of the scanning constraint, while the outer shell was still frozen. The relaxed structures were then subjected to single point QM/MM energy calculations with DFT using  $\omega$ B97XD functional [174] and aug-cc-pVTZ correlation consistent Dunning basis set [175,176]. Energy corrections were evaluated as the difference between  $\omega$ B97XD and DFTB3 QM/MM energies. These corrections were added to the QM/MM DFTB3 free energies to obtain corrected free energies.

**Immunoproteasome.** To improve the accuracy of the DFTB3 calculated energies QM correction was performed using the Gaussian 16 program package. The reactant, intermediate, product and transition state structures of the carbonate and carbonthioate mechanism were extracted from the last frame of the corresponding windows. The non-QM atoms and further selected QM atoms were deleted keeping the size of the system computationally feasible to handle with more involved methods. The remaining structures with 32 atoms were geometry optimized with B3LYP functional and 6-31G++(d,p) basis set using the implicit PCM solvent model [177]. Stable structures were subjected to a subsequent geometry optimization without any positional constraint using the OPT keyword, while transition states were optimized with the bond lengths forming the reaction coordinate frozen using the OPT=modredundant keyword of Gaussian. This was followed by single point vacuum calculations at the DFTB3 and  $\omega$ B97XD/aug-cc-pVTZ levels of theory, as no PCM model is available for DFTB3 in Gaussian. The difference between the  $\omega$ B97XD and DFTB3 energies was used as the correction of the previously calculated free energies.

## 4.4 Evaluation and post-processing

### 4.4.1 Calculation of binding free energy differences

All of the binding free energy differences derived from the thermodynamic integrations were calculated using the analyse.sh script available at the AMBER website [178] (**Table A2**). The script requires a special form of an output file belonging to the productive MDs, which can be generated by the logdvd1=1 keyword in the input file.

### 4.4.2 Construction of the potential of mean force (PMF) curves and surfaces

**MurA.** The potential of mean force curve of the proton transfer and inhibition reaction and the PMF surface of the reaction with PEP were calculated with the weighted histogram analysis method. The force constant<sup>2</sup> in the metafiles required for the WHAM code [179] was set to 400 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>. The two-dimensional PMF of the MurA-PEP reaction was constructed with the 2D-WHAM implementation, with the force constant set to 400 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> for both sampled coordinates. The number of bins were set to three times the number of windows, and the tolerance parameter was set to 0.0001.

**KRAS, EGFR, ITK, BTK, BMX.** PMF curves were constructed by WHAM. The force constant in the metafile was set to 400 kcal mol<sup>-1</sup>·Å<sup>-2</sup>. Further parameters set for the WHAM procedure were identical to what was described previously.

**Immunoproteasome.** PMF curves derived from the last 35 ps of the productive MD runs were constructed with the WHAM code. As the definition of the force constant is identical in

---

<sup>2</sup> Note that the definition of the force constant in AMBER and in the WHAM code is different.

both WHAM and PLUMED, the force constant in the metafiles was set to  $300 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ . The same bin number and tolerance parameter were set as described above.

#### 4.4.3 pK<sub>a</sub> estimation derived from thermodynamic integration

**Mura.** The pK<sub>a</sub> values of Cys115 and His394 were estimated from thermodynamic integration results of the sidechain mutations. The value of the  $\Delta\Delta G$  was calculated with the `analyse.sh` script, while the pK<sub>a</sub> value was derived from it using **Eq. 18**.

$$pK_{a,prot} = pK_{a,model} + \frac{1}{2.303kT} \Delta\Delta G \quad (18)$$

where  $pK_{a,prot}$  is the pK<sub>a</sub> of the examined residue in the protein environment,  $pK_{a,model}$  is the pK<sub>a</sub> of the single amino acid,  $k$  is the Boltzmann constant,  $T$  is the temperature and  $\Delta\Delta G$  is the calculated free energy difference.

#### 4.4.4 Error estimation

**MurA.** Statistical uncertainties of the calculated PMFs were estimated by the WHAM code's built-in bootstrapping algorithm. We used a 140-step correlation time, estimated from autocorrelation function of several data files of previous US simulations. 10 fake data sets were used for every bootstrap analysis. The standard error is 1.5-2.0  $\text{kcal}\cdot\text{mol}^{-1}$  in the higher energy regions, like transitions states, and it is 0.1-0.4  $\text{kcal}\cdot\text{mol}^{-1}$  in the lower energy regions.

**KRAS, EGFR, ITK, BTK, BMX.** The errors of the TI simulations related to KRAS and EGFR were estimated by performing transformation loops. The theoretical  $\Delta\Delta G$  sum of such cycle is 0  $\text{kcal}\cdot\text{mol}^{-1}$ , therefore deviation from 0  $\text{kcal}\cdot\text{mol}^{-1}$  accounts for the error of thermodynamic integration simulations (**Table A3**).

The statistical uncertainties of the PMF calculations were estimated by the bootstrapping algorithm, described above.

The error of the barrier calculation for a pair of ligands  $i$  and  $j$  was quantified by the difference of the calculated versus experimental barrier free energy difference:

$$error(i,j) = \left( \Delta G_{exp}^{\ddagger}(i) - \Delta G_{exp}^{\ddagger}(j) \right) - \left( \Delta G_{cal}^{\ddagger}(i) - \Delta G_{cal}^{\ddagger}(j) \right) \quad (19)$$

The overall mean unsigned error (MUE) was obtained as the mean of the absolute values of errors obtained for each ligand pair.

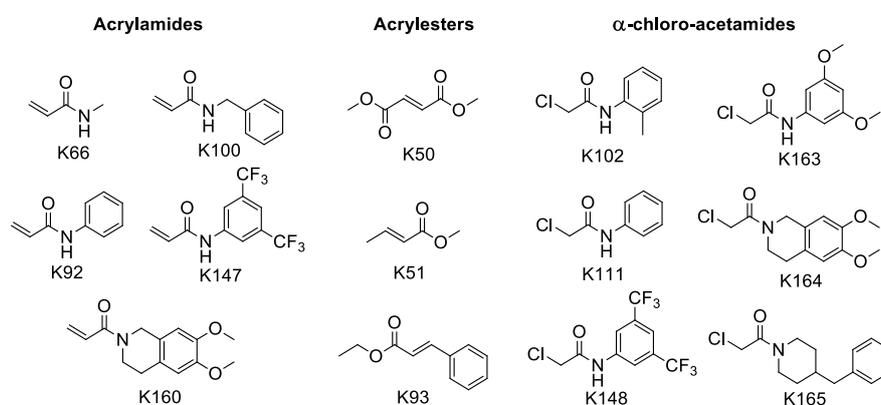
**Immunoproteasome.** Errors of the TI simulations were estimated by the same protocol as described in the previous section, performing transformation loops; starting with **HT1146** and cycling back into it through a sequence of ligand modification TIs (**Table A3**).

The statistical uncertainties of the iPS and cPS PMF calculations were estimated by the bootstrapping algorithm, as it was described above.

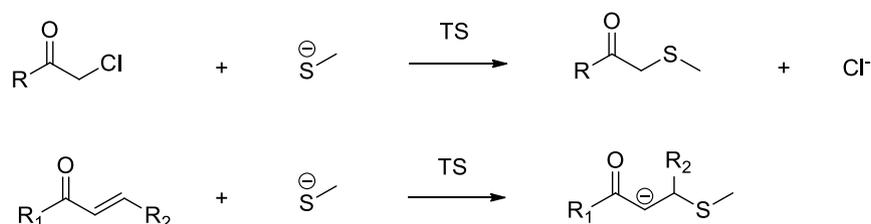
## CHAPTER 5 – RESULTS AND DISCUSSION

### 5.1 Reactivity prediction of covalent inhibitors [39]

According to previous studies [36,38,162] correlation can be observed between experimental reactivity data of electrophilic compounds and the calculated barriers of model reactions between the compound and a nucleophilic fragment. The nucleophilic fragment is a substitute of the targeted nucleophilic sidechain of the studied protein and it is usually a methyl-thiolate in case of cysteine attacking molecules. We selected a couple of compounds (**Figure 8**) with experimental GSH half-lives and calculated the reactant, transition state and product structure of the reaction between the compounds and a methyl-thiolate fragment (**Scheme 3**). Our expectation was to find good correlation between calculated and experimental properties.



**Figure 8.** Selected compounds for QM calculations.



**Scheme 3.** Model reactions.

#### 5.1.1 Results of model reaction calculations

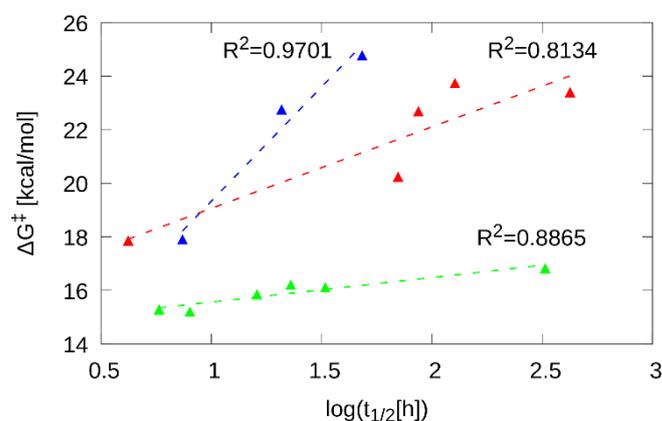
The reaction barriers were determined as the difference between the transition state and reactant state free energies. The geometries were optimized at the B3LYP/6-311G\*\*(d,p) level of theory. Experimental and calculated properties are shown in **Table 2**.

**Table 2.** Results of the reactivity predictions.  $\log(t_{1/2})$  is the experimental GSH half-life and  $\Delta G^\ddagger$  is the calculated barrier of the reaction in **Scheme 3** at B3LYP/6-311G++(d,p) level of theory. The applied temperature for both the experimental measurements and the calculations is 27 °C.

Warhead type	Compound	$\log(t_{1/2}[\text{h}])$	$\Delta G^\ddagger$ [kcal·mol <sup>-1</sup> ]
Acrylamides	<b>K66</b>	2.10	23.7
	<b>K92</b>	1.85	20.2
	<b>K100</b>	2.63	23.4
	<b>K147</b>	0.62	17.9
	<b>K160</b>	1.94	22.7
Acrylesters	<b>K50</b>	0.87	17.9
	<b>K51</b>	1.32	22.8
	<b>K93</b>	1.68	24.8
	<b>K102</b>	2.51	16.8
	<b>K111</b>	1.52	16.1
$\alpha$ -chloro-acetamides	<b>K148</b>	0.76	15.3
	<b>K163</b>	1.36	16.2
	<b>K164</b>	0.90	15.2
	<b>K165</b>	1.21	15.8

### 5.1.2 Correlation with experimental reactivity metrics

The calculated barriers as the function of the logarithm of GSH half-lives are shown in **Figure 9**. We found that better correlation is achieved if we consider molecules with different warheads separately. Therefore, we grouped the examined molecules into three groups, namely acrylesters, acrylamides and chloroacetamides. The  $R^2$  was calculated for these groups and the calculated and experimental values show good agreement. This indicates that the experimental reactivity of covalent inhibitors can be successfully predicted by quantum chemical reaction barrier calculations, however it is recommended to use it for molecules possessing the same warhead. Another solution is to use empirical scaling factors between molecules equipped with different warheads.



**Figure 9.** Calculated reaction barriers of model reactions as a function of experimental GSH half-lives. Different colors indicate compounds with different warheads: acrylamides (red), acrylesters (blue),  $\alpha$ -chloroacetamides (green).

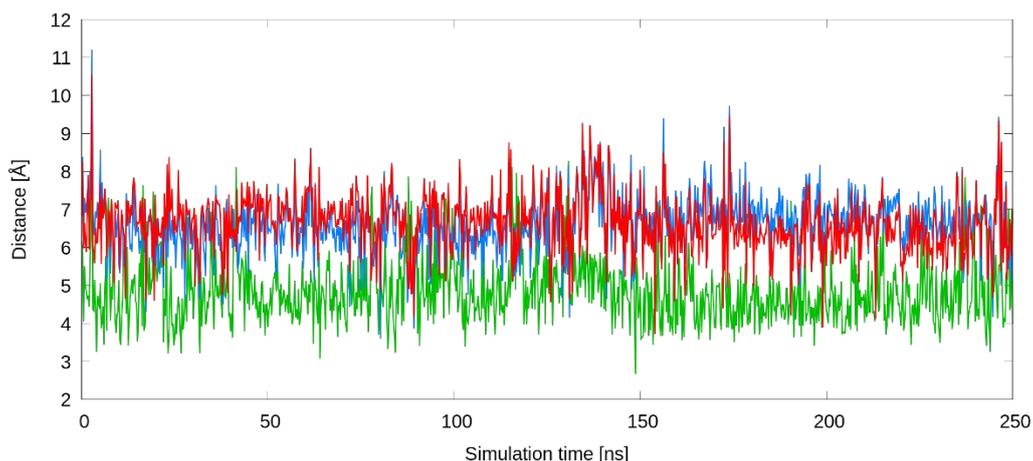
These results confirm that QM barrier calculations for model reactions effectively predict reactivity, thus supporting reactivity tailoring of covalent inhibitors. Reactivity is well described for compounds having the same warhead, and the predictive power is lower when warheads are varied. Note that this highly efficient computational procedure does not include the effect of the enzyme environment. In contrast, studies discussed in the subsequent sections take the enzyme environment into account in the computational description of covalent inhibition.

## 5.2 Catalytic mechanism and covalent inhibition of MurA [170]

Our initial aim in this study was to expand the transition state calculations into enzyme environment and develop a robust method based on QM/MM MD to calculate the covalent inhibition reactions' free energy curves. Our first target was MurA, a well-known bacterial enzyme, which can be targeted covalently at the Cys115 sidechain. Though, the catalytic mechanism of MurA has been studied previously [44,47,51,180,181], the reaction energetics and the role of the active site residues during the catalysis and inhibition have remained elusive. The role of Cys115 found in the highly flexible loop region was formerly thought to be involved in product release, rather than in the reaction [44]. It was later proposed that Cys115 binding to phosphoenol-pyruvate (PEP) precedes the reaction between the PEP and UDP-N-acetylglucosamine (UNAG) [51] (see the reaction catalyzed by MurA in **Scheme 2**). As the deprotonation of the targeted cysteine residue is necessary to increase the nucleophilicity of the sulfur atom to attack the electrophilic warheads [182], the exploration of the deprotonation mechanism of Cys115 is essential to understand the covalent inhibition of MurA. To achieve a better perspective on the mechanism and inhibition of the enzyme we extended our study to the deprotonation mechanism of Cys115, the reaction between PEP and Cys115, the role of the flexible loop containing Cys115 and finally the covalent inhibition of MurA by selected compounds.

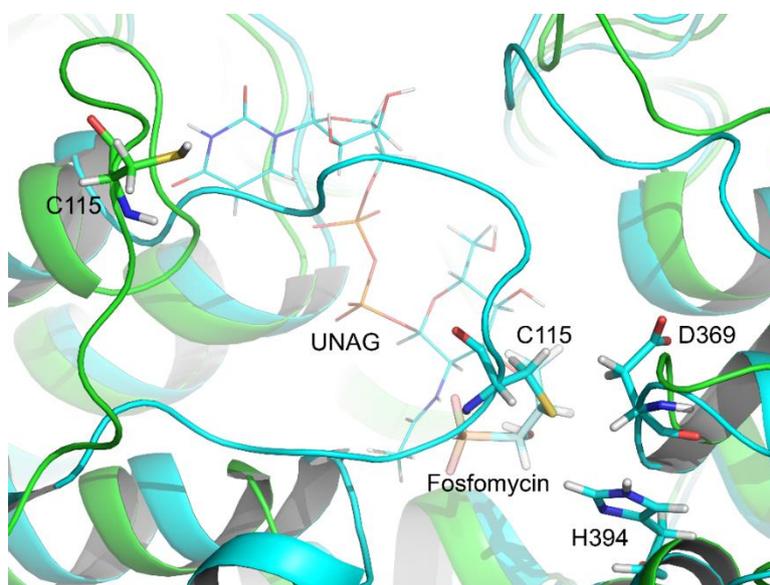
### 5.2.1 Deprotonation mechanism

Deprotonation by water molecules is common for residues at the solvent accessible enzyme surface, however, this is not the case here even though Cys115 is found in a solvent exposed region of MurA. Baum and coworkers [183] observed that  $IC_{50}$  lowers by at least fivefold upon UNAG binding, suggesting a special deprotonation mechanism induced by UNAG. Schönbrunn and coworkers [47] published similar findings; according to their proposal the cysteine might be deprotonated by Asp369 or Arg397. Neither of these hypotheses have been examined previously, therefore we performed classical molecular dynamics simulations to obtain insights into the dynamic nature of the loop and extract information which residues would be able to deprotonate the Cys115. The simulation started from the UNAG-bound loop closed structure created from the fosfomycin-UNAG-MurA ternary complex (PDB: 1UAE) with the deletion of fosfomycin.

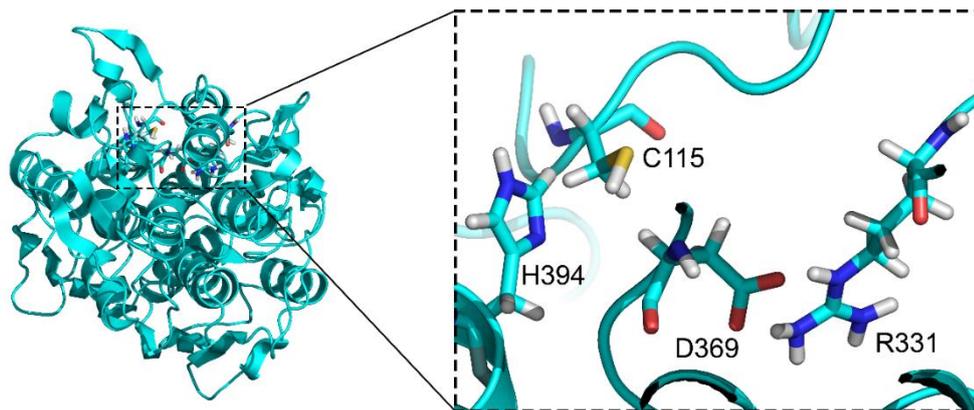


**Figure 10.** Distances between Cys115-S-H and terminal heavy atoms of selected residues: His394-ND1 (green), Asp369 OD1(red), Asp369-OD2 (blue).

The position of Cys115 and its surroundings were monitored throughout the simulation, which revealed two possible deprotonation agents, namely His394 and Asp369. Distance analysis (**Figure 10**) pointed out that both residues approach Cys115 within 5 Å, but visual inspection of the trajectory showed that Asp369 points toward Arg331 during most of the simulation and also forms a salt bridge with it. Interestingly, these residues are separated by over 20 Å from Cys115 in the loop open conformation and Cys115 approaches them when the loop is closed (**Figure 11**). The average distance between Asp396 and Cys115 terminal heteroatoms exceeds 5 Å, while His394 and Cys115 are not only within 4-5 Å separation during the simulation, but they adopt a proper geometry allowing the proton transfer between them (**Figure 12**).



**Figure 11.** Aligned structures of the empty unliganded MurA (PDB: 1EJD; green) in loop open, and the complex with fosfomicin and UNAG (PDB: 1UAE; cyan) in loop closed conformations.



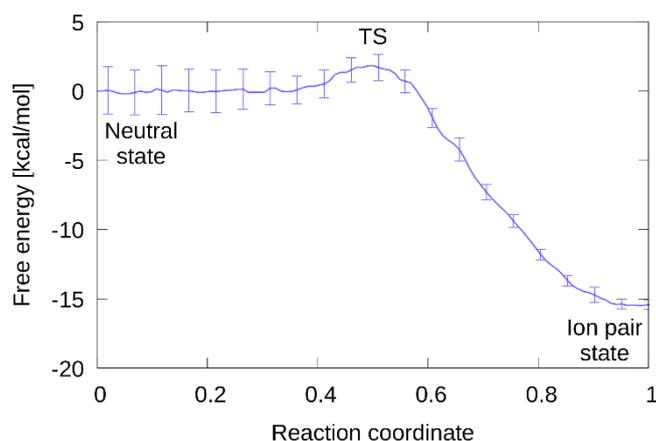
**Figure 12.** A representative snapshot from the loop closed MD simulation of MurA. Cys115, His394, Asp369 and Arg331 are shown in stick representation.

His394 is a highly conserved residue among different MurA species, which underlines its possible role in the activation procedure of Cys115 [184] (**Figure 13**).

Ec	DLRASASLVLAGCIAEGTTVDRIYHIDRGYERIEDKLRALGANIERVKGE-----
Hi	DLRASISLVLAGCIATGETIVDRIYHIDRGYEHIEDKLRGLGAKIERFSGSDEA-----
Aa	DLRASASLVLAGLVAQGETTVVRDVYHIDRGYEKLEEKLLKLGADIERVSEL-----
Ec1	DLRASASLVLAGCIAEGTTVDRIYHIDRGYERIEDKLRALGANIERVKGE-----
Vf	DLRASASLVIAGCIAKGETIVDRIYHIDRGYDKIEDKLTALGANIERVHSDDLLEHHHHH
Vc	DLRASASLVIAGCIAKGETIVDRIYHIDRGYERIEDKLSALGANIERFRD-----
Pa	DLRASASLVIAGLVAEGDTLIDRIYHIDRGYECIEEKQLLQGLGAKIRRVPG-----
Ab	DLRASFSVLVAALVAEGDTLIDRIYHIDRGYEHVEEKQLQGLGAKIKRVS-----
Cj	DLRASSALLAALAAKGTSKVHRIYHIDRGYENLEEKFKDLGAKITRLEE-----
Pp	DLRASASLVLSALVAEGDTLIDRIYHIDRGYECIEEKQLMLGAKIRRVPG-----
Sp	DLRASAALLTGLVAQGETTVGKLVHIDRGYGFHEKLAQLGAKIQRIEASDEDE-----
	***** :*::. * * : : : *:**** ..*: **** * *

**Figure 13.** Multiple sequence alignment of MurA from *E. coli* (Ec), *H. influenzae* (Hi), *A. aeolicus* (Aa), *E. cloacae* (Ec1), *V. fischeri* (Vf), *V. cholerae* (Vc), *P. aeruginosa* (Pa), *A. Baumannii* (Ab), *C. jejuni* (Cj), *P. putida* (Pp) and *S. pneumoniae* (Sp). Residues 369-419 (Ec numbering) are shown. Highly conserved His394 (Ec numbering) is highlighted. Sequence alignment was performed by the T-Coffee multiple sequence alignment package [184].

Further examination on the potential deprotonation mechanism was performed by QM/MM MD free energy calculations applying umbrella sampling. The constructed PMF (**Figure 14**) of the deprotonation of Cys115 by His394 in the unliganded enzyme shows a 2 kcal·mol<sup>-1</sup> reaction barrier and a nearly 15 kcal·mol<sup>-1</sup> energy difference between the reactant and product state in favor of the latter. The calculated reaction energy might be overestimated due to the DFTB3 functional; however, the low barrier indicates that the deprotonation reaction is likely to happen. Thermodynamic integration studies calculating the pK<sub>a</sub> of the two sidechains resulted in pK<sub>a</sub>=5.5 and pK<sub>a</sub>=6.5 for Cys115 and His394, respectively, which also supports the suggested activation mechanism. Visual inspection of the umbrella sampling trajectories revealed that the thiolate form of the Cys115 is stabilized by interactions with the guanidine moiety of Arg397 and with the backbone amide of Thr116.

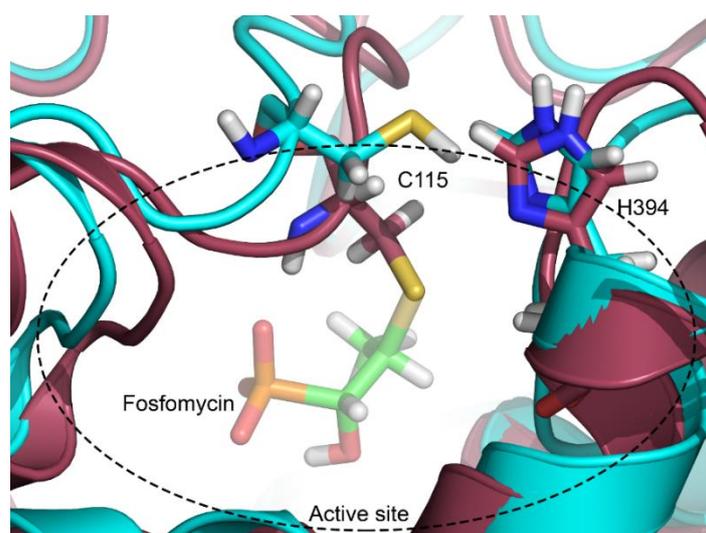


**Figure 14.** Free-energy curve of the proton transfer between Cys115-S and His394-N $\gamma$ . The reaction coordinate changes from the neutral (0) to the ion pair state (1).

These results support the proposed mechanism in which the Cys115 residue is activated upon UNAG binding initiated loop-closure, bringing Cys115 in the proximity of the His394 to facilitate proton transfer from Cys115 sulfur. This activation step is essential in the covalent reaction of Cys115 with both PEP as part of the catalytic cycle and with covalent inhibitors of MurA.

### 5.2.2 Binding pocket conformations

Available MurA X-ray structures contain loop conformations adopting closed, half-open or open states [48]. This underlines the flexible character of the loop, however, none of the structures is well-suited for docking studies used for creating initial binding poses for further MD simulations. The open and half-open structures are possessing too large pockets, where no exact binding site can be defined, while the closed conformations, like the MurA-UNAG-fosfomycin complex, are too tight for docking the covalent ligands. To obtain a suitable active site for ligand docking we inspected the trajectory of the classical MD performed during the deprotonation studies with the MurA-UNAG complex.

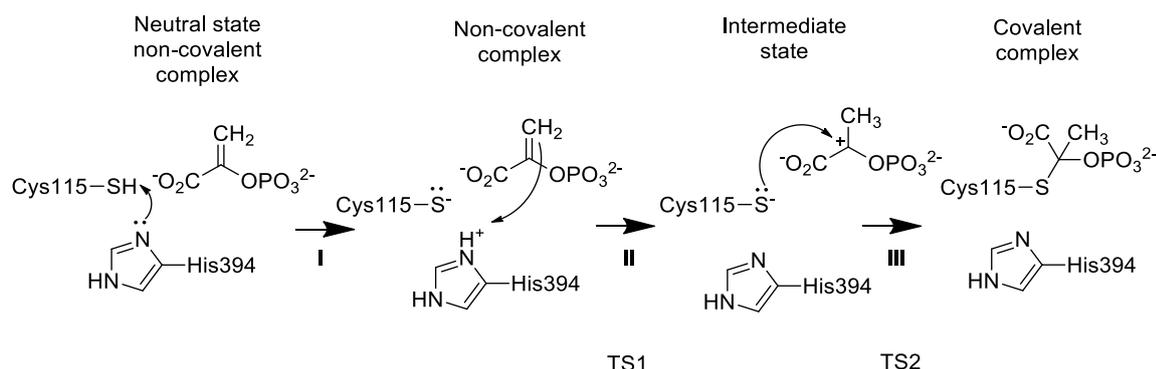


**Figure 15.** Aligned structures of 1UAE (dark red) and the molecular dynamics snapshot with enlarged binding pocket (cyan). The active site is encircled by dashed lines. The covalently bound fosfomycin of 1UAE is faded.

Throughout the MD the size of the active site in the proximity of Cys115 was altering continuously due to the flexible nature of the loop. We performed cluster analysis on the trajectory and the center of the main cluster was selected as the best representation of the empty active site of MurA. This provided a feasible structure containing active site, which is able to accommodate PEP and inhibitor molecules, later to be examined (**Figure 15**).

### 5.2.3 Catalytic reaction with natural substrate phosphoenol-piruvate (PEP)

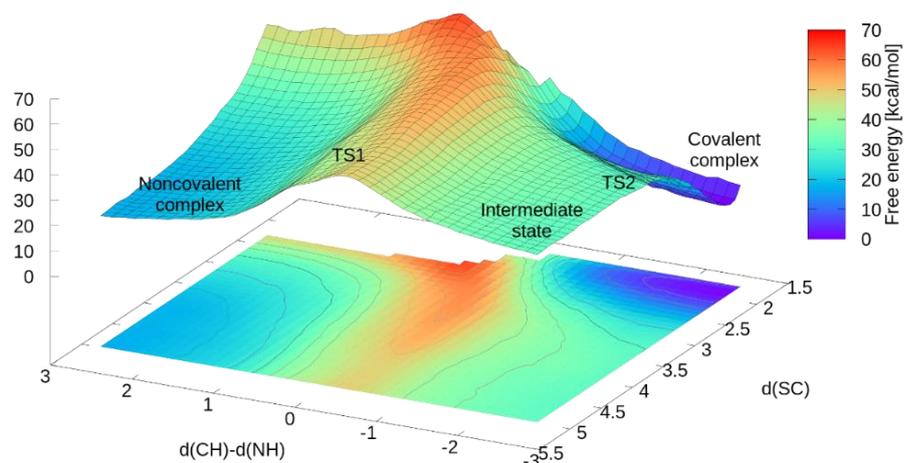
Based on the previous deprotonation studies and the X-ray structure of the MurA-PEP complex [51] containing a covalent bond between the Cys115 and a PEP molecule the following mechanism was proposed for PEP binding (**Scheme 4**).



**Scheme 4.** Proposed reaction mechanism for PEP binding to Cys115.

According to the suggested mechanism the reaction is initiated by the proton transfer between Cys115 S and His394 N $\delta$  (**I**). This step is followed by another proton transfer, in which the proton from His394 is transferred to C3 of PEP, forming a carbocation intermediate (**II**). Finally, the deprotonated S attacks the electrophilic center of the carbocation (**III**) yielding the covalent PEP-MurA complex.

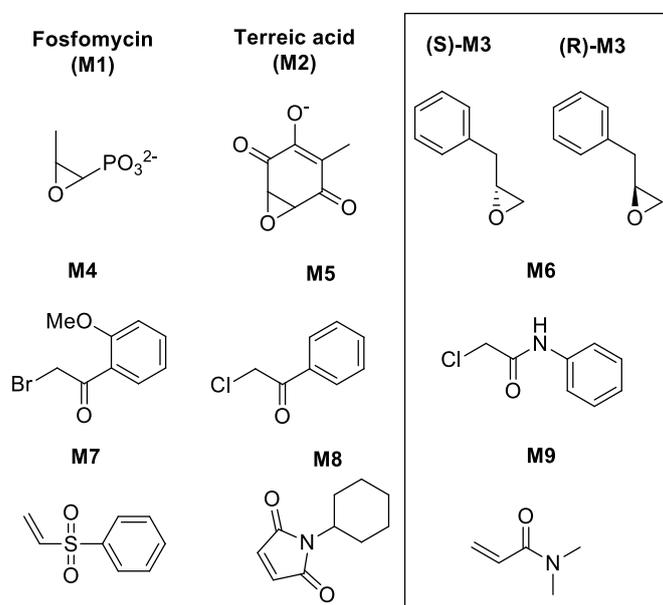
Previously, it has not been clarified that reaction steps **II** and **III** proceed sequentially as depicted or follow a concerted mechanism. To evaluate which way is more likely, we explored the two-dimensional free energy surface of the abovementioned steps (**Figure 16**). The results underlined the depicted mechanism in which the protonation of the double bond precedes the nucleophilic attack of the cysteine residue. The barrier of the first step is found to be 30 kcal·mol<sup>-1</sup>, and the formation of the carbocation is found on a plateau with energy level higher by 12 kcal·mol<sup>-1</sup> related to the reactant (protonated His394) state. Step **III** undergoes with a minimal barrier of 2 kcal·mol<sup>-1</sup>, therefore, the rate limiting step of the reaction is the proton transfer between His394 N $\delta$  and C-3 of PEP. The overall reaction is exothermic, as product state is being 18 kcal·mol<sup>-1</sup> lower than the initial reactant state.



**Figure 16.** Calculated free energy surface for the reaction between PEP and Cys115. Missing data points at high free energies are due to the instability of the highly charged species that prevented sampling.

#### 5.2.4 Covalent inhibition by different covalent warhead chemotypes

In order to evaluate, whether the active-inactive separation of ligands based on their transition barriers against MurA is possible, we selected 3 groups of ligands with various warheads. Each group contained two active and one inactive compounds (**Figure 17**) and their reaction barrier free energies were calculated.



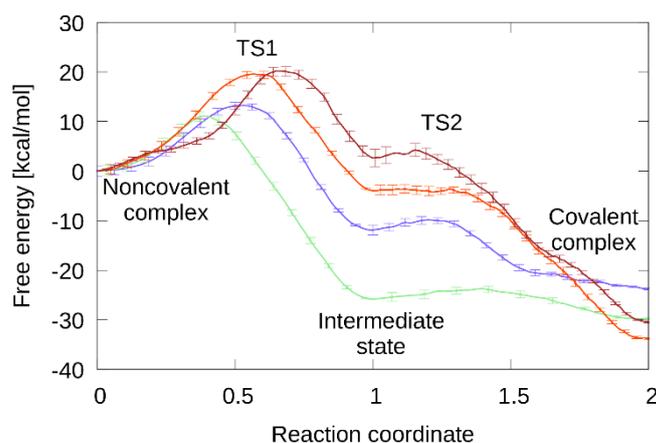
**Figure 17.** Examined compounds against MurA. Framed compounds are inactive molecules.

#### Nucleophilic substitution reaction of oxiranes

Fosfomycin (**M1**) and terreic acid (**M2**) are both oxirane derivatives covalently inhibiting MurA [45,53,58], both having experimental X-ray structures complexed with the protein.



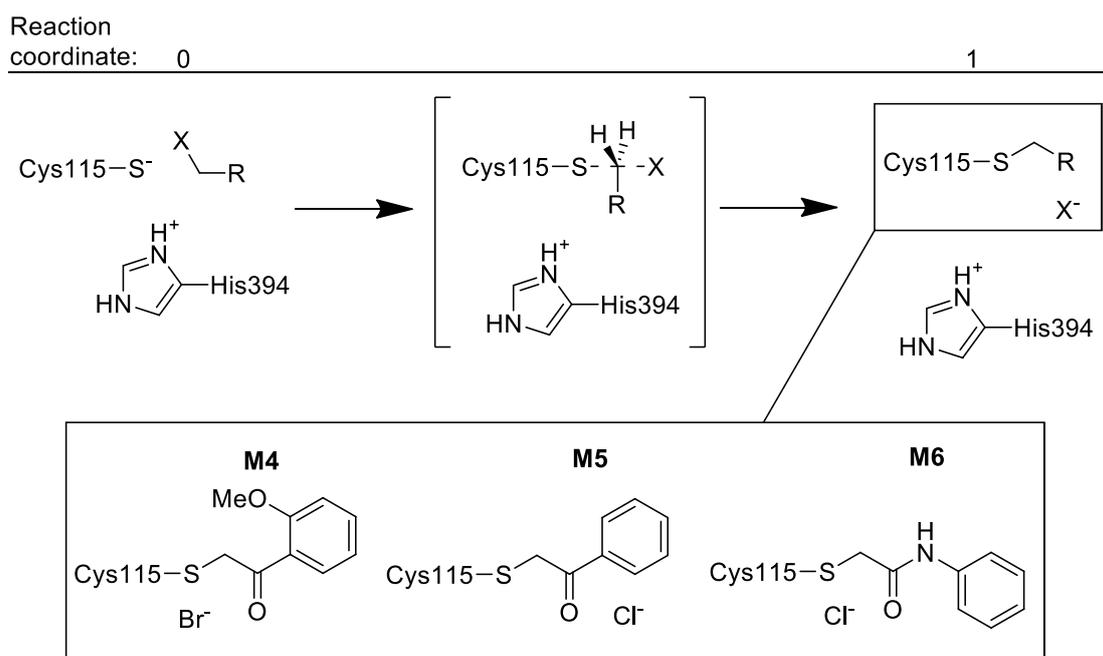
fosfomycin (**M1**) and terreic acid (**M2**), possessing 10-12 kcal·mol<sup>-1</sup> transition state barriers. The second barriers contain only a slight 1-2 kcal·mol<sup>-1</sup> elevation in free energy. The product states are found exothermic for all cases, having over 25-30 kcal·mol<sup>-1</sup> energy gain.



**Figure 18.** PMF curves of the reactions between oxirane derivatives and MurA Cys115. Fosfomycin (**M1** active, blue), terreic acid (**M2** active, light-green), (*S*)-**M3** (inactive red) and (*R*)-**M3** (inactive brown).

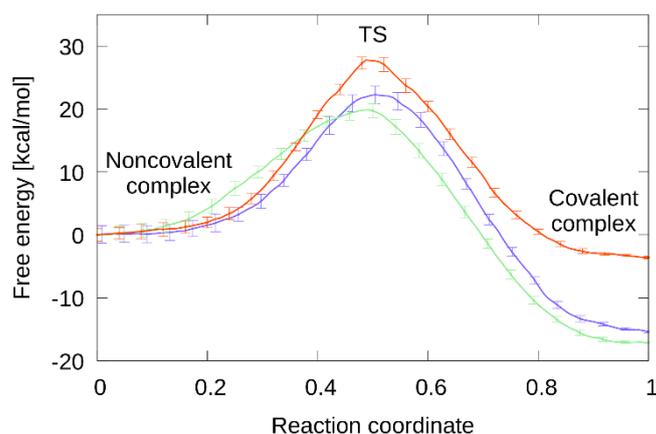
### Nucleophilic substitution reaction of haloketones

Haloketone and haloacetamide warheads are known cysteine attacking inhibitors, which inhibit the targeted proteins via an S<sub>N</sub>2-type one-step mechanism reaction. We evaluated the covalent reaction of two active  $\alpha$ -halogenated acetophenone derivatives (**M4** and **M5**) and an inactive  $\alpha$ -chloroacetamide (**M6**) against the Cys115 residue of MurA. The examined reaction mechanism is shown in **Scheme 6**, and the constructed free energy profiles can be seen in **Figure 19**.



**Scheme 6.** Examined reaction mechanism of the  $\alpha$ -halogenated compounds.

The examined  $S_N2$  reaction consists of one step with a pentacoordinated transition state. During the reaction the distance between the thiolate and the electrophilic carbon atom is decreasing, while the carbon-halogen distance is simultaneously rising, yielding the covalently modified Cys115 residue and a departed halogen ion. Possible subsequent steps are the deprotonation of the histidine by the halogen and finally the diffusion of the HCl or HBr molecules; however, these steps were assumed not to affect the covalent inhibition and were not investigated. The calculated free energy profiles show barriers of 18 and 20 kcal·mol<sup>-1</sup> for **M4** and **M5**, respectively, while the inactive compound **M6** has the highest barrier of 28 kcal·mol<sup>-1</sup>. Additionally, the reaction of the inactive compound is found to be less exothermic than those of the active ones (3 vs. 16 kcal·mol<sup>-1</sup> free energy gain).

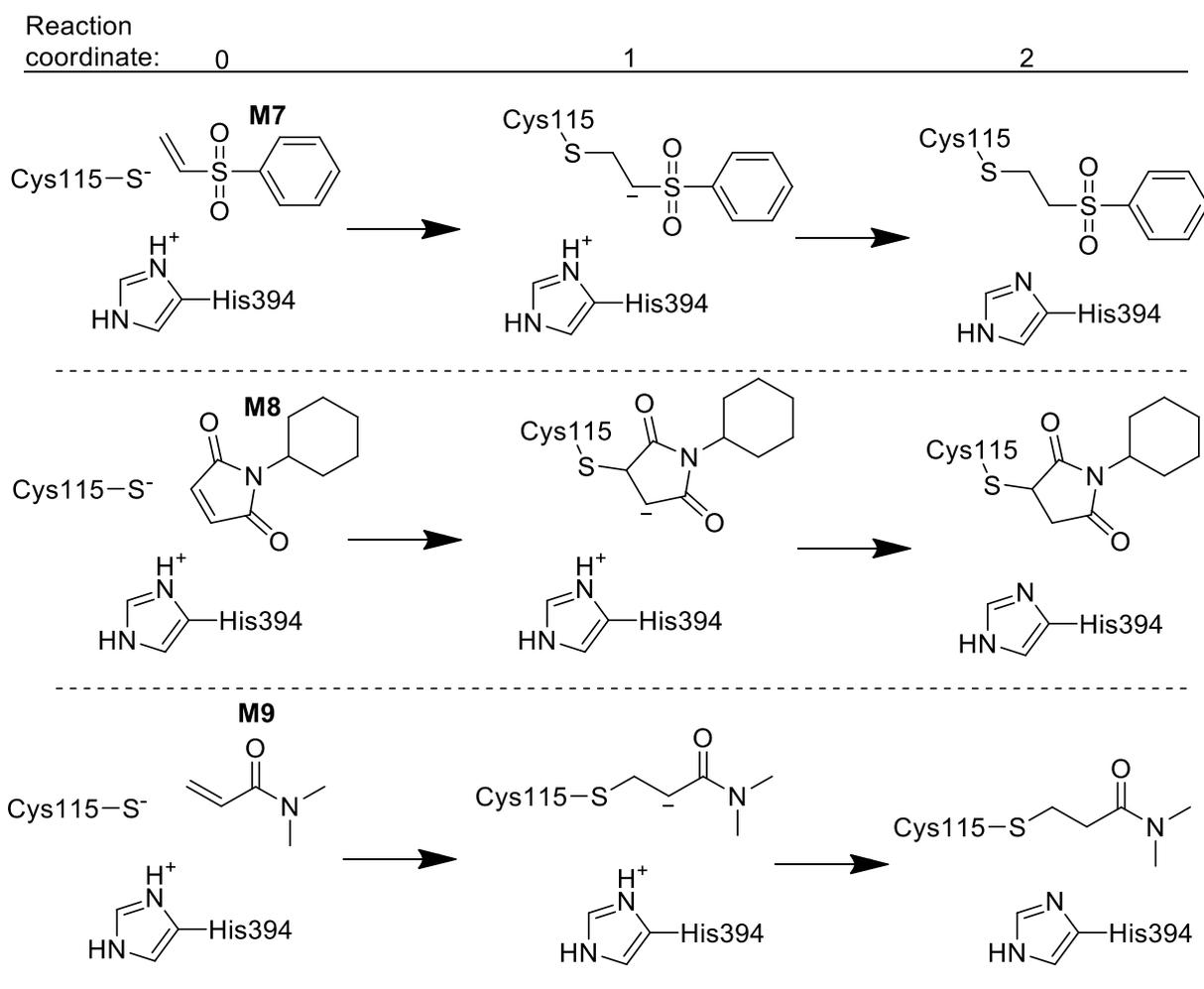


**Figure 19.** PMF curves of the nucleophilic substitution reactions between Cys115 and the  $\alpha$ -halogenated compounds **M4** (active, blue), **M5** (active, light green) and **M6** (inactive, red).

### Nucleophilic conjugation reaction of Michael acceptors

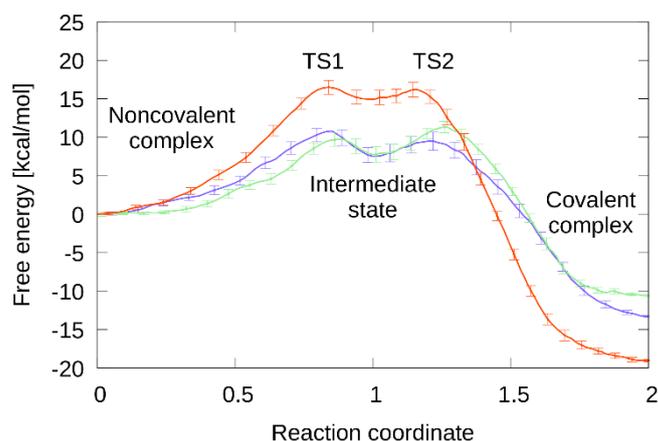
Michael-addition is among the most represented reaction chemistries in covalent inhibition [21,38,185]. We selected two active compounds equipped with a vinyl-sulphonyl (**M7**) and a maleimide (**M8**) warhead, and also an inactive acrylamide derivative (**M9**) and modeled their reaction against Cys115. The first step of the examined reaction consists of the nucleophilic attack of the thiolate on the electrophilic carbon of the double bond, producing the carbanion intermediate. The reaction terminates with the neutralization of the carbanion by proton abstraction from the His394 residue (**Scheme 7**).

It has been shown previously that the semiempirical DFTB3 functional is unable to predict stable carbanion intermediate during adiabatic scanning, showing no minimum on the potential energy curve of the  $C\beta-S^-$  bond formation [186–188]. In contrast, our QM/MM free energy calculations with DFTB3/FF14SB potential were able to identify the carbanion intermediate, showing a minimum in the PMF curves. In addition, the capability of DFTB3 to predict the carbanion intermediate was shown for several Michael-addition reactions with EGFR inhibitors (see below) [139], a vinylsulfone based Cruzain inhibitor [189] and small molecules in water [139].



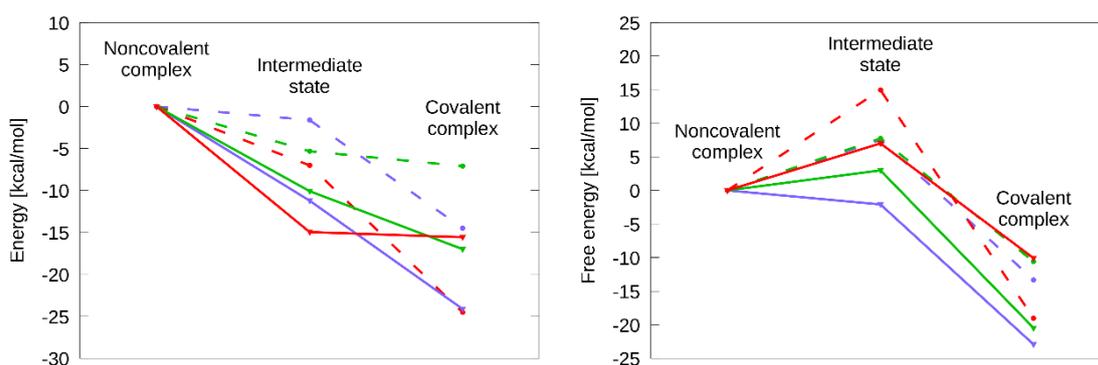
**Scheme 7.** Examined mechanisms for the Michael-addition reactions.

The constructed free-energy curves for the reaction of compounds **M7**, **M8** and **M9** with Cys115 are shown in **Figure 20**. Highly similar reaction profiles belong to the two active compounds **M7** and **M8** with about 9 and 12 kcal·mol<sup>-1</sup> barriers for the first step, respectively, and with about 6 kcal·mol<sup>-1</sup> reaction energy for the intermediate state. The intermediate state is connected with the product state by a low barrier of 2-4 kcal·mol<sup>-1</sup>. In line with the experimental results, the inactive compound (**M9**) has the largest barrier reaching nearly 16 kcal·mol<sup>-1</sup>, which leads to a high energy intermediate state. The following step contains a 1-2 kcal·mol<sup>-1</sup> barrier, which proceeds to a lower energy product state than it is found for the active compounds, namely (-19) vs. (-14)-(-11) kcal·mol<sup>-1</sup>. This unexpected observation prompted us to investigate, if this is an artefact of the semiempirical DFTB3 functional and whether a correction with a more involved QM potential is feasible. We selected the  $\omega$ B97XD functional [174] and aug-cc-pVTZ basis set [175,176] as it has been shown to describe thio Michael-additions with similar accuracy as the CCSD(T) method [186]. Corrections to the DFTB3 free energies of the prominent reaction states were performed as described at **Chapter 4 – Materials and methods**.



**Figure 20.** PMF curves of the examined Michael-addition reaction between Cys115 and **M7** (active, blue), **M8** (active, light-green) and **M9** (inactive, red).

The results of the QM correction calculations are shown in **Figure 21**. The overall trend of the three states ( $E_{\text{ncov}}^{\text{QM/MM}} > E_{\text{int}}^{\text{QM/MM}} > E_{\text{cov}}^{\text{QM/MM}}$ ) calculated by DFTB3 agrees with the  $\omega$ B97XD results. However, the stabilization of the reactant state for the inactive compound is overestimated by DFTB3, and the correction shifts the free energy of **M9** above those of **M7** and **M8** in the product state.



**Figure 21.** QM/MM energies with DFTB3/FF14SB (dashed lines) and  $\omega$ B97XD/aug-cc-pVTZ/FF14SB (straight lines) methods for compounds **M7** (blue), **M8** (green) and **M9** (red) [left panel]. Uncorrected (DFTB3/FF14SB) (dashed lines) and corrected ( $\omega$ B97XD/aug-cc-pVTZ/FF14SB) (straight lines) QM/MM free energies for compounds **M7** (blue), **M8** (green) and **M9** (red) [right panel].

### 5.2.5 Evaluation and separation of the active and inactive compounds

The summary of the calculated results can be seen in **Table 3**. In all three examined reaction chemistries the highest calculated transition state barrier belongs to the inactive compounds. This justifies the application of the developed method to separate active and inactive ligands based on their transition state barriers.

**Table 3.** Experimental (MurA and GSH) and calculated (DFTB3/FF14SB) properties of MurA targeting compounds. The applied temperature for both the experimental measurements and the calculations is 27 °C. (Experimental data were taken from ref. [12])

Compound	<i>MurA</i>				<i>GSH</i>	<i>Calc.</i>
	Residual activity [%]	IC <sub>50</sub> ( <i>E. coli</i> ) [μM]	Type	Reversibility	Half-life [h]	ΔG <sup>‡</sup> [kcal·mol <sup>-1</sup> ]
<b>M1</b> (Fosfomycin)	NA <sup>a</sup>	8.8 <sup>b</sup>	Active	Irreversible	NA <sup>a</sup>	13.3
<b>M2</b> (Terreic acid)	NA <sup>a</sup>	14 <sup>c</sup>	Active	Irreversible	NA <sup>a</sup>	11.1
<b>M3</b> <sup>d</sup>	93	-	Inactive	-	77.5	19.6 <sup>e</sup> , 20.2 <sup>f</sup>
<b>M4</b>	3	0.38	Active	Irreversible	0.0	22.3
<b>M5</b>	3	2.25	Active	Irreversible	0.0	19.9
<b>M6</b>	110	-	Inactive	-	32.8	27.8
<b>M7</b>	12	15	Active	Irreversible	0.2	10.8
<b>M8</b>	1	0.55	Active	Reversible	0.0	11.3
<b>M9</b>	95	-	Inactive	-	127.0	16.5

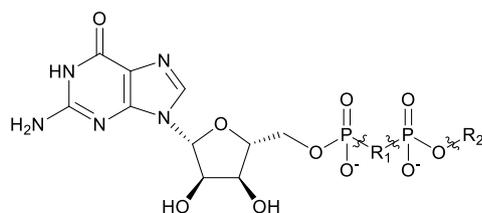
<sup>a</sup>NA – not available, <sup>b</sup>Ref. [183], <sup>c</sup>*E. Cloacae* [58], <sup>d</sup>Racemic, <sup>e</sup>R-enantiomer, <sup>f</sup>S-enantiomer

### 5.3. Affinity and selectivity assessment of covalent inhibitors [139]

The next step ahead in the course of our work was taking into account the complete covalent inhibition including both the non-covalent and covalent step. In case of MurA we were able to differentiate between active and inactive compounds by modeling the covalent reaction between the inhibitors and MurA. However, this approach is based on the assumption that the energetics of the covalent inhibition solely depends on the covalent step. To examine both steps of the covalent inhibition, the evaluation of  $K_i$  and  $k_{inact}$  constants are required. Therefore, we developed a coupled MD based method including MM thermodynamic integration for the non-covalent and QM/MM umbrella sampling for the covalent step. These simulations are able to provide us with free energies that can be used to derive  $K_i$  and  $k_{inact}$  according to **Eqs. 2 and 3**. We selected several covalently targetable enzymes, for which data of covalent inhibition reactions with experimental  $K_i$  and  $k_{inact}$  values, or at least with experimental  $k_{inact}/K_i$  ratio is available in the literature. The experimental values were extracted from the work of Xiong et al. (KRAS) [64] and Schwartz et al. [190] (EGFR), while data for the selectivity studies were taken from the publications of Thorarensen [148] (ITK, BTK, BMX) and Engel [147] (EGFR). The non-covalent binding free energies were calculated as free energy differences ( $\Delta\Delta G$  values), which were transformed into an absolute scale (see later). The activation free energies were derived from constructed PMF curves as in the case of MurA.

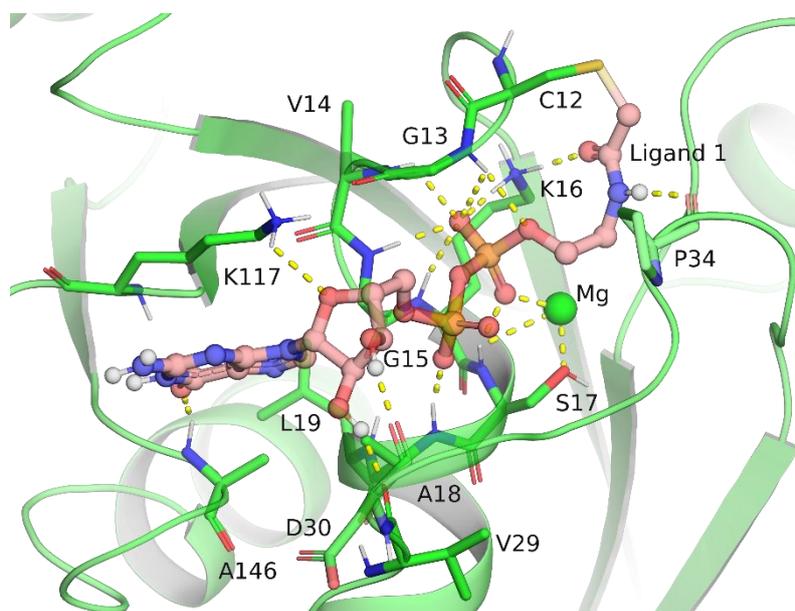
#### 5.3.1 KRAS inhibitors

We selected a series of GTP-like  $\alpha$ -chloroacetamides (**Figure 22**) inhibiting the oncogenic KRAS<sup>G12C</sup> mutant. For the binding conformation of the protein, we used the crystal structure of the **1**-KRAS<sup>G12C</sup> (PDB: 4NMM) complex. The selected structure shows the characteristic properties of the inactive, GDP bound state of the active site of KRAS (**Figure 23**). The GDP-like ligand binding mode was conserved during all the simulations.



Code	R <sub>1</sub>	R <sub>2</sub>	Code	R <sub>1</sub>	R <sub>2</sub>
1			6		
2			7		
3			8		
4			9		
5			10		

**Figure 22.** Examined inhibitors of KRAS<sup>G12C</sup>.

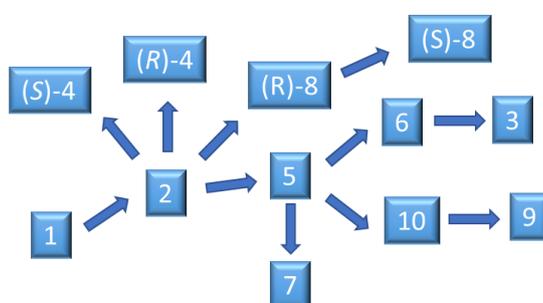


**Figure 23.** Characteristic interactions of compound **1** and KRAS<sup>G12C</sup>.

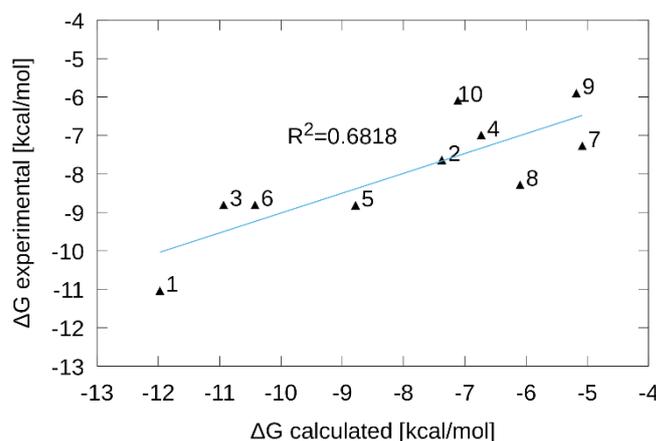
### Non-covalent step

The tree-plots of the alchemical transformations and the results of the TI simulations are shown in **Figure 24** and in **Table 4**, respectively. Free energy differences obtained by the TI simulations were shifted to an absolute scale to minimize the root-mean-square deviation (RMSD) between the calculated and experimental binding free energies. Note that the shift does not affect the correlation between the experimental and calculated properties and does not

change compound ranking, either. According to the results, sensible correlation is found between the calculated and experimental binding free energies (**Figure 25**) with an error of  $1.8 \text{ kcal}\cdot\text{mol}^{-1}$  (see the error estimation in **Chapter 4 – Materials and methods**) that is comparable with state of the art FEP calculations [191,192]. The RMSD value was found to be  $1.4 \text{ kcal}\cdot\text{mol}^{-1}$  (**Table A3**) and the coefficient of determination ( $R^2$ ) to be 0.68. Trajectory analysis highlighted a few characteristic secondary interactions between the protein and the ligands. Ligands with the highest binding affinities interact with the amide hydrogen of Gly15 through the linker between the two phosphate groups both when it is an oxygen and a fluorine containing group. Interestingly, similar to **2** the fluorine atom of (*S*)-**4** points away from Gly15, therefore the hydrogen bond cannot be formed. Hydrogen bonds connecting warhead amides to Lys16-NH<sub>3</sub> and Pro34 amide carbonyl are present in the X-ray structure and were preserved during the simulations. These interactions play an essential role in both steps of the covalent reaction, recognizing the inhibitor molecule and positioning the warhead toward the targeted sidechain.



**Figure 24.** Tree-plot of the alchemical transformations during TI simulations.

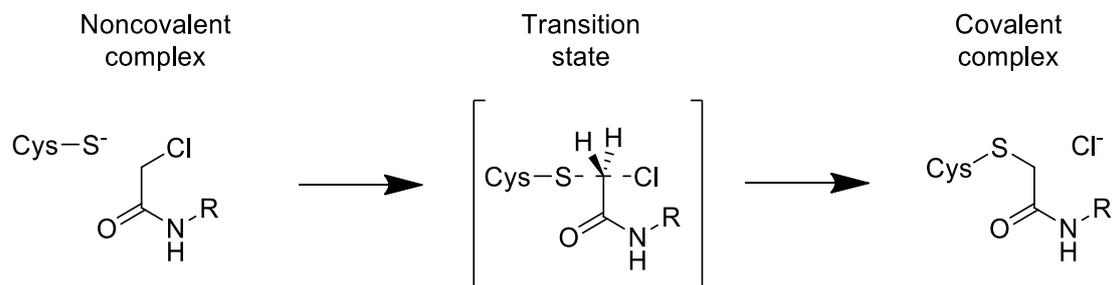


**Figure 25.** Comparison of the experimental and calculated binding free energy values for KRAS<sup>G12C</sup> inhibitors **1-10**.

### Covalent step

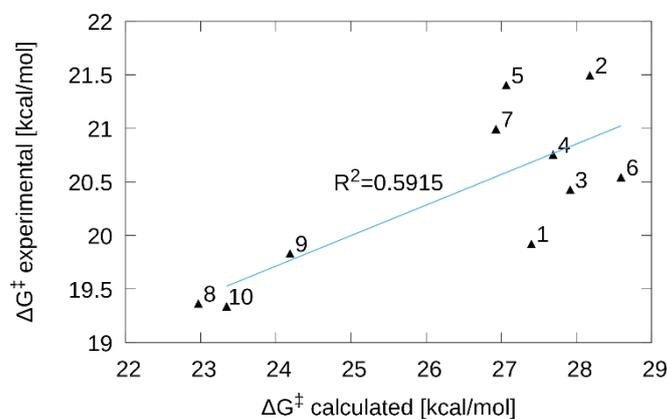
The covalent bond formation of the KRAS inhibitors was modeled as in the case of MurA. The free energy profiles of the inhibition reactions were evaluated and the free energy barriers related to the experimentally determined  $k_{\text{inact}}$  reaction rates were investigated. The PMF curves (**Figure A10**) were derived from QM/MM MD simulations coupled with umbrella sampling.

The modeled S<sub>N</sub>2 reaction mechanism between the  $\alpha$ -chloroacetamides and Cys12 is shown in **Scheme 8**.



**Scheme 8.** Mechanism of the modeled S<sub>N</sub>2 reactions.

The reaction follows a one-step mechanism in which the thiolate attacks the electrophilic carbon; the S-C distance is decreasing, while the C-Cl bond is lengthening, simultaneously. The transition state belongs to the formation of a pentacoordinated complex. Finally, the bond between the inhibitor and Cys12 is formed, and the chloride ion departs the complex. The thiol deprotonation was assumed to be ligand independent as no structural motif of the ligand appears to be able to accept the thiol proton. The presence of deprotonated Cys12 is suggested by the results of pK<sub>a</sub> calculations by thermodynamic integration [193] and by a Poisson-Boltzmann model [194], estimating the Cys12 pK<sub>a</sub> 8.6 and 8.3, respectively. These correspond to the presence of 7 % and 14 % thiolate form that reacts according to the proposed reaction scheme. Experimental and calculated quantities, describing the covalent step of the inhibition reaction are collected in **Table 4**, while the correlation between the calculated and experimental free energy barriers are shown in **Figure 26**.



**Figure 26.** Comparison of the experimental and calculated reaction barriers for the covalent step of KRAS<sup>G12C</sup> inhibitors **1-10**.

The calculated barriers for the S<sub>N</sub>2 reactions were found to be overestimated by the DFTB3/FF14SB QM/MM potential that agrees with earlier observations of chemistry dependent systematic errors of DFTB3 [195–197]. Analysis of **Figure 26** revealed two separated groups of molecules. Compounds with a ring moiety in the proximity of the warhead (**8-10**) are located in the bottom left, while the rest of the molecules are found in the upper right part of the plot. Cyclic amides calculated as the most reactive compounds is in line with the

experimental results, although the differences are slightly overestimated. The overall error of the barrier calculations described by the mean unsigned error (MUE) is 1.89 kcal·mol<sup>-1</sup>. Visual inspection of the simulation trajectories revealed two significant hydrogen bonds between ligands and KRAS<sup>G12C</sup>. One of them is found between the amide hydrogen of the warhead and the backbone oxygen of Pro34, which interaction was present only in case of ligands with two carbon linker region (**1-4**). The other interaction can be found between Lys16-NH<sub>3</sub> and the carboxyl oxygen of the acetamide group of the warhead. Interestingly, the latter hydrogen bond might serve as a warhead positioning constraint [198], keeping the warhead region of the inhibitor in the vicinity of the targeted cysteine residue (**Figure 23**). This interaction was observed in almost every examined complex with two exceptions, namely (**S**)-**8** and **9**. The carbonyl oxygen of these compounds' warhead points away from the Lys16 sidechain.

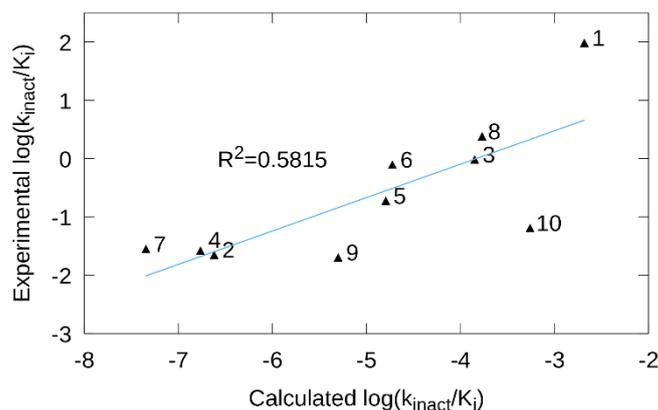
### Evaluation of the complete covalent inhibition:

As both steps of the covalent inhibition might play essential role in the observed activity, both  $K_i$  and  $k_{inact}$  may give important contribution to the observed inhibition [199]. It is also customary to characterize the covalent inhibition with the  $k_{inact}/K_i$  ratio. **Figure 27** showing the relation between the experimental and calculated  $\log(k_{inact}/K_i)$  values underlines a proper correlation between the experimental and computed properties. Though, the scale of the values differs, it is due to the systematic overestimation of the computed  $k_{inact}$  rate constants. Note that the computed  $K_i$  equilibrium constants are derived from  $\Delta G$  values shifted by a constant to minimize the RMSD (see above), however the applied shift does not affect the correlation with experimental  $\log(k_{inact}/K_i)$ . To conclude, **Figure 27** confirms that the developed method, based on the calculation of  $K_i$  and  $k_{inact}$  is well suited for compound prioritization.

**Table 4.** Experimental inhibition constants for the non-covalent binding ( $K_i$ ), corresponding binding free energies ( $\Delta G_{exp}$ ), calculated (FF14SB) binding free energies ( $\Delta G_{calc}$ ), experimental inhibition rate constant for the covalent reaction ( $k_{inact}$ ), corresponding free energy barrier ( $\Delta G_{exp}^\ddagger$ ), calculated (DFTB3/FF14SB) free energy barrier ( $\Delta G_{calc}^\ddagger$ ) and experimental and calculated  $\log(k_{inact}/K_i)$  values for KRAS inhibitors. The applied temperature for both the experimental measurements and the calculations is 27 °C. (Energies are in kcal·mol<sup>-1</sup>. Experimental data were taken from ref. [64])

Compound	$K_i$ [ $\mu$ M]	$\Delta G_{exp}$	$\Delta G_{calc}$	$k_{inact}$ [min <sup>-1</sup> ]	$\Delta G_{exp}^\ddagger$	$\Delta G_{calc}^\ddagger$	$\log(k_{inact}/K_i)_{exp}$	$\log(k_{inact}/K_i)_{calc}$
1	0.009	-11.0	-12.0	0.860	19.9	27.4	1.98	-2.68
2	2.700	-7.6	-7.4	0.060	21.5	28.2	-1.65	-6.62
3	0.380	-8.8	-10.9	0.365	20.4	27.9	-0.02	-3.85
4	8.000 <sup>a</sup>	-7.0	-6.7 <sup>b</sup>	0.210 <sup>a</sup>	20.8	27.7 <sup>b</sup>	-1.58	-6.76
5	0.374	-8.8	-8.8	0.070	21.4	27.1	-0.73	-4.80
6	0.382	-8.8	-10.4	0.300	20.5	28.6	-0.10	-4.72
7	5.000	-7.3	-5.1	0.140	21.0	26.9	-1.55	-7.35
8	0.920 <sup>a</sup>	-8.3	-6.1 <sup>b</sup>	2.200 <sup>a</sup>	19.4	23.0 <sup>b</sup>	0.38	-3.77
9	50.000	-5.9	-5.2	1.000	19.8	24.2	-1.70	-5.30
10	36.000	-6.1	-7.1	2.300	19.3	23.3	-1.19	-3.26

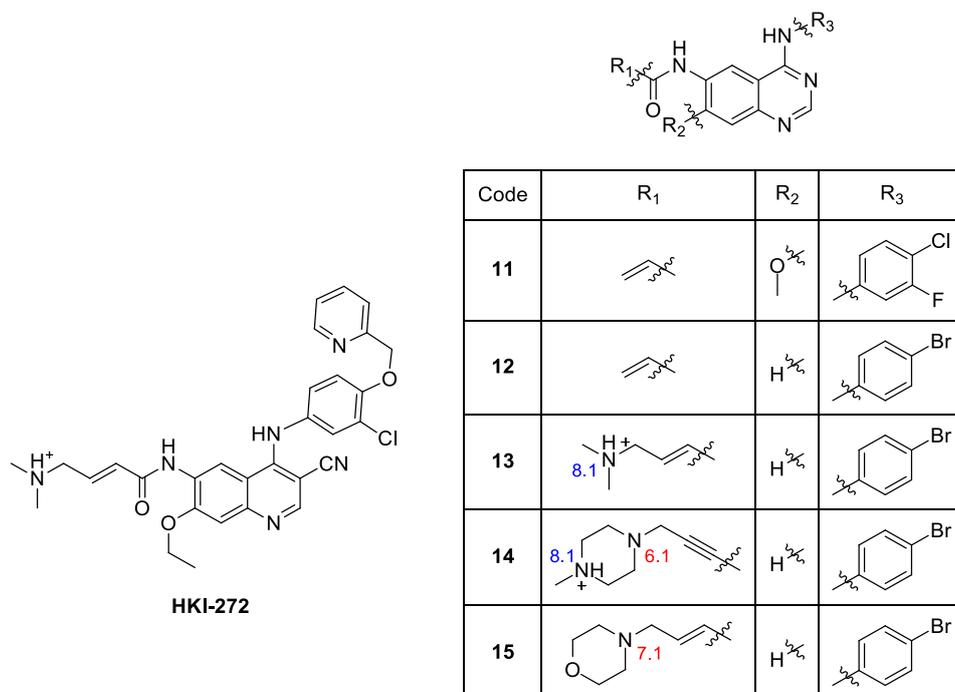
<sup>a</sup>racemic <sup>b</sup>average of the results of the two diastereomers



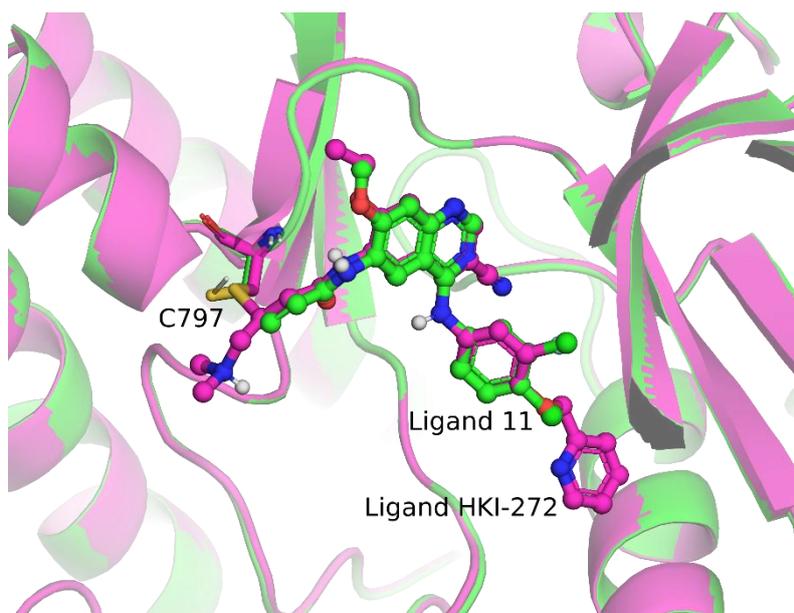
**Figure 27.** Comparison of the experimental and calculated  $\log(k_{inact}/K_i)$  for KRAS<sup>G12C</sup> inhibitors **1-10**.

### 5.3.2 EGFR inhibitors

Further validation of the developed method was performed for the oncogenic EGFR T790M/L858R double mutant. We selected four acrylamides and a propargylamide derivative – equipped with the proper experimental data on EGFR inhibition – to examine against the double mutant (**Figure 28**). Binding poses of the corresponding molecules were built based on the X-ray structure of the **HKI-272-EGFR**<sup>T790M/L858R</sup> (PDB: 3W2Q) covalent complex (**Figure 29**). The protonation state of **13**, **14** and **15** was estimated by  $pK_a$  calculations. As the  $pK_a$  of the tertiary amine of **13** and **14** was found to be higher than the  $pH=7.5$  value of the inhibition measurement [147], their protonated form was used. The  $pK_a$  of **15** was calculated as 7.1, near to the 7.5 value, therefore both protonation states were considered for **15**. The calculated  $pK_a$ s are shown in **Figure 28**.



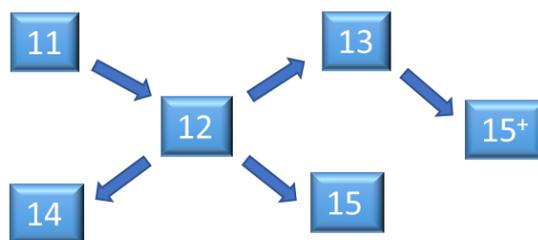
**Figure 28.** Investigated EGFR inhibitors **11-15** and the bound inhibitor in PDB 3W2Q (HKI-272). The calculated  $pK_a$  values of the protonated and deprotonated amine groups of **13-15** are shown in blue and red numbers, respectively. Both protonated and deprotonated **15** were used during the calculations.



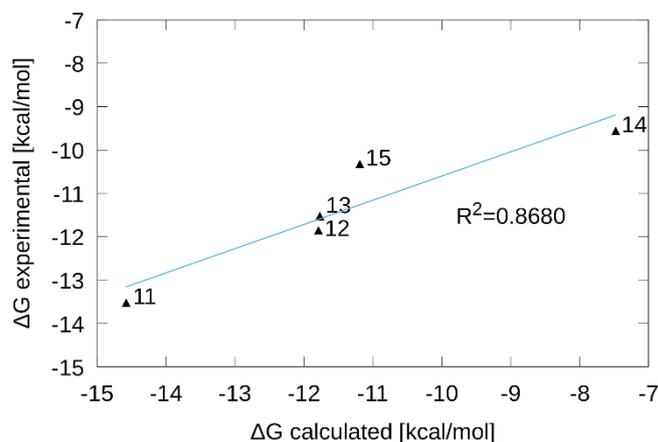
**Figure 29.** Experimental binding mode of **HKI-272** in EGFR double mutant (PDB: 3W2Q, pink) and superimposed modeled binding mode of ligand **11** (green).

#### **Non-covalent step:**

The tree-plot of the alchemical transformations and the binding free energies derived from TI calculations are shown in **Figure 30** and **Table 5**, respectively. The calculated  $\Delta G$  values show good agreement with the experimental ones (**Figure 31**). The error is given as  $1.18 \text{ kcal}\cdot\text{mol}^{-1}$  (see error estimation in **Chapter 4 – Materials and methods**), while the RMSD was found to be  $1.12 \text{ kcal}\cdot\text{mol}^{-1}$ , again within the error margin of high quality FEP calculations. The largest deviation from the experimental binding free energies belongs to **14**, the only compound with the propargylamide warhead. Trajectory analysis was used to identify secondary interactions between the ligands and enzyme environment and to clarify the differences between the binding free energies of the examined compounds. An intramolecular hydrogen bond between the methoxy oxygen and the amide hydrogen of the warhead of **11** was found, that stiffens the molecule and keeps the warhead near to Cys797. Molecules **12**, **13** and **14** show similar binding modes, in which the warhead of the compounds is positioned parallel with the backbone of the targeted cysteine. Finally, a hydrogen bond between the ether oxygen of **15** and the guanidine moiety of Arg803 was present during a significant part of the simulations. These subtle variations in the examined inhibitors' binding mode account for their binding free energy differences. Note that results in **Table 5** refer to the deprotonated **15**. Binding free energy for protonated **15** was found to be  $-12.3 \text{ kcal}\cdot\text{mol}^{-1}$ . This gives a small contribution to the affinity of the system, since the difference between the  $\Delta G$ s of the two forms is only  $1.1 \text{ kcal}\cdot\text{mol}^{-1}$  and the deprotonated form is present in more than 70 %, according to the  $pK_a$  calculation. ( $pK_a=7.1$ ,  $pH=7.5$ ; see **Figure 28**)



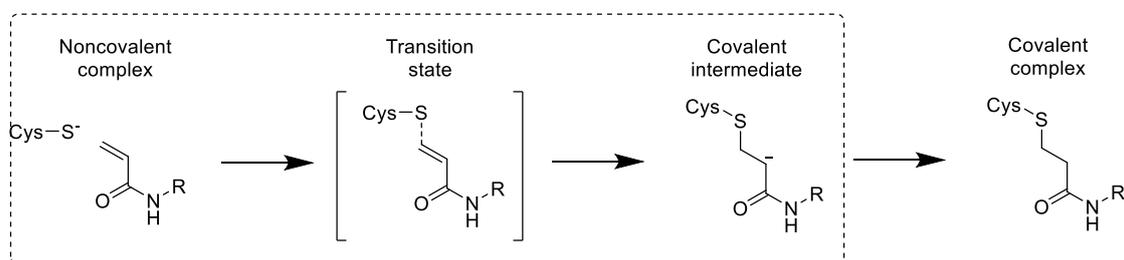
**Figure 30.** Tree-plot of the alchemical transformations during the EGFR TI calculations. The plus sign denotes the protonated form of **15**.



**Figure 31.** Comparison of the experimental and calculated binding free energies of EGFR inhibitors **11-15**.

### Covalent step:

The barriers of the covalent reaction steps were determined by QM/MM umbrella sampling simulations. The selected compounds inhibit Cys797 through Michael-addition reaction. The reaction is initiated by the deprotonation of Cys797 [169], however this step was considered ligand independent and was not simulated during our MDs. Thereafter the reaction proceeds with the nucleophilic attack of the thiolate and the protonation of the carbanion. These steps might occur in a concerted or in a consecutive manner.

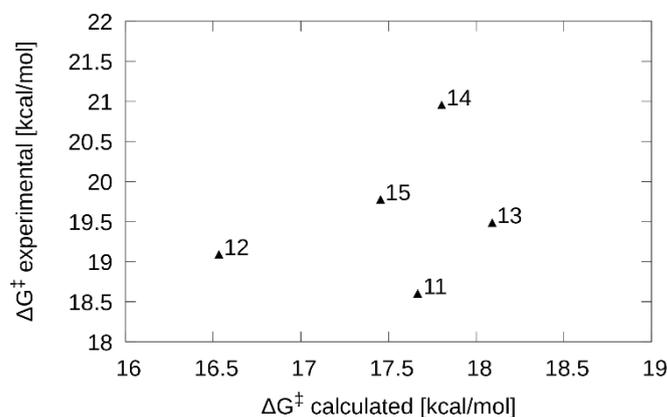


**Scheme 9.** Mechanism of the Michael-addition reactions. Only the steps within the dashed line were investigated by calculations.

The preference of consecutive versus concerted reaction mechanism is affected by the stability of the carbanion, as a reduced stability carbanion shifts the preference toward the concerted mechanism. Therefore, we studied the Michael-addition reactions of **11** on Cys797 of EGFR with steered molecular dynamics simulations using several reaction coordinates. The

coordinates were formed as various linear combinations of distances characterizing the nucleophilic attack and the protonation steps (**Figure A13**). Based on these calculations the consecutive reaction with  $C\beta$ -S bond formation followed by  $C\alpha$ -H bond formation is preferred for **11** using the DFTB3/FF14SB potential. A free energy minimum belonging to the carbanion intermediate was observed for all the examined reactions of Michael-acceptor molecules inhibiting EGFR. In addition, the nucleophilic attack producing the carbanion intermediate was found to be the rate determining step based on previous studies [36,38,200].

Considering the abovementioned observations, only the barrier of the nucleophilic attack of the thiolate towards the warhead was evaluated and compared to the experimental barrier derived from the measured  $k_{\text{inact}}$ . The protonation step was omitted from the MD simulations. Experimental and calculated data for the covalent reaction of EGFR inhibition are shown in **Table 5**. The comparison of the experimental and calculated barriers is shown in **Figure 32**, while the calculated PMFs are shown in **Figure A11**. Note that **Table 5** and **Figure 32** show data for neutral **15**. The difference between the calculated reaction barrier for protonated and neutral **15** is only  $0.7 \text{ kcal}\cdot\text{mol}^{-1}$ , and the majority of compound **15** can be found in its neutral form at  $\text{pH}=7.5$  (see above).



**Figure 32.** Comparison of the experimental and calculated reaction barriers for the covalent step of EGFR inhibitors **11-15**.

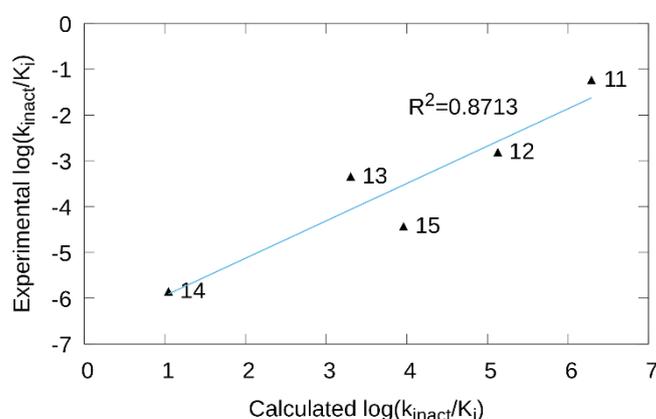
A reasonable agreement between the calculated and experimental barriers can be found (**Table 5**). The overall error quantified by the MUE is  $1.16 \text{ kcal}\cdot\text{mol}^{-1}$ . A fairly rigid binding mode with no significant change during chemical reactions was observed for all the examined compounds. The quinazoline ring moiety preserves its position in all the simulations. The only notable movement was observed for molecular parts connected to  $C\beta$  of the warhead in molecules **13-15**. Terminal acrylamides **11** and **12** adopt both *s*-cis and *s*-trans conformations in their non-covalent complexes, with the *s*-cis conformation becoming the dominant as the reaction proceeds with decreasing  $C\beta$ -S distance.

### Evaluation of the complete covalent inhibition:

The correlation between the calculated and experimental  $k_{\text{inact}}/K_i$  values describing the complete covalent inhibition is shown in **Figure 33**. The coefficient of determination is found to be  $0.87 (R^2)$  indicating that the computed  $k_{\text{inact}}/K_i$  values well characterize inhibitory activity, and are suitable for compound prioritization.

**Table 5.** Experimental inhibition constants for the non-covalent binding ( $K_i$ ), corresponding binding free energies ( $\Delta G_{\text{exp}}$ ), calculated (FF14SB) binding free energies ( $\Delta G_{\text{calc}}$ ), experimental inhibition rate constant for the covalent reaction ( $k_{\text{inact}}$ ) corresponding free energy barrier ( $\Delta G_{\text{exp}}^\ddagger$ ), calculated (DFTB3/FF14SB) free energy barrier ( $\Delta G_{\text{calc}}^\ddagger$ ) and experimental and calculated  $\log(k_{\text{inact}}/K_i)$  values for EGFR inhibitors. The applied temperature for both the experimental measurements and the calculations is 27 °C. (Free energies are in kcal·mol<sup>-1</sup>. Experimental data were taken from ref. [190])

Compound	$K_i$ [nM]	$\Delta G_{\text{exp}}$	$\Delta G_{\text{calc}}$	$k_{\text{inact}}$ [min <sup>-1</sup> ]	$\Delta G_{\text{exp}}^\ddagger$	$\Delta G_{\text{calc}}^\ddagger$	$\log(k_{\text{inact}}/K_i)_{\text{exp}}$	$\log(k_{\text{inact}}/K_i)_{\text{calc}}$
<b>11</b>	0.14	-13.5	-14.6	8	18.6	17.7	-1.24	6.29
<b>12</b>	2.3	-11.9	-11.8	3.5	19.1	16.5	-2.82	5.13
<b>13</b>	4	-11.5	-11.8	1.8	19.5	18.1	-3.35	3.30
<b>14</b>	108	-9.6	-7.5	0.15	21.0	17.8	-5.86	1.05
<b>15</b>	30	-10.3	-11.2	1.1	19.8	17.5	-4.44	3.96



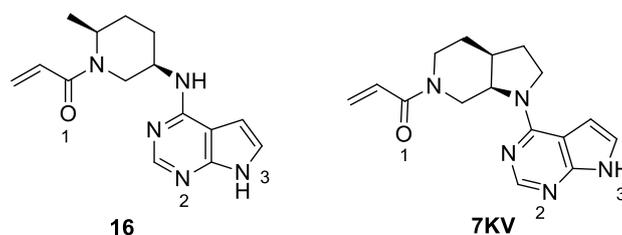
**Figure 33.** Comparison of the experimental and calculated  $\log(k_{\text{inact}}/K_i)$  for EGFR inhibitors **11-15**.

### 5.3.3 Selectivity studies for covalent kinase inhibitors

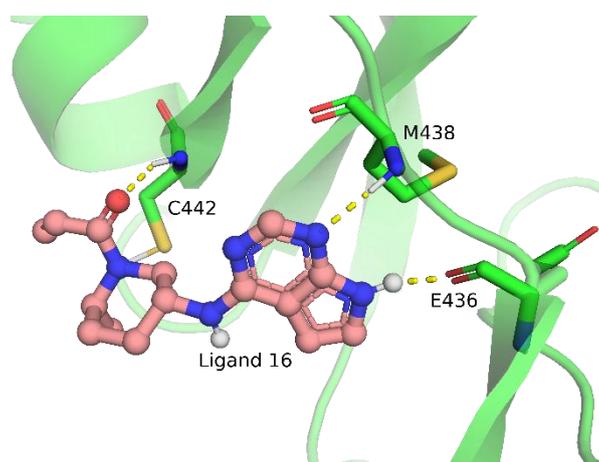
#### Selectivity study of ITK, BTK and BMX:

Selectivity is a prominent attribute of protein inhibitors. Our aim was to expand the developed MD based free energy calculations able to prioritize covalent inhibitors against the same target for studying the selectivity of these inhibitors. We selected compounds with the proper experimental data for the inhibition reaction against related proteins. Firstly, we selected an acrylamide compound **16** (Figure 34) with known  $K_i$  and  $k_{\text{inact}}$  constants against three similar kinases, namely, ITK, BTK and BMX, all of them having good quality X-ray structures complexed with different ligands. Our objective was to reproduce the binding affinity and the reaction rate differences between the kinase complexes of **16**. This protocol might be a useful tool evaluating and predicting covalent compound selectivity over structurally similar proteins. Note that calculating non-covalent binding affinity ( $K_i$ ) differences of various ligands against a protein is an established tool in non-covalent drug design. To evaluate the binding free energy differences of the same ligand in different enzymes requires an alternative approach creating thermodynamic cycles for TI calculations with alchemical transformation of protein sidechains. Our method generates the related proteins as the mutants of each other, focusing on the amino acid variations in the binding sites. We selected the ITK structure as reference since the lowest

$K_i$  value, therefore, the best binding affinity was observed for **16** in ITK. Consequently, similar protein conformations were assumed for the related kinases, BTK and BMX. As experimental X-ray structure for ITK-**16** complex is unavailable in the PDB, the X-ray structure of a related ligand (**7KV**) in complex with JAK3 (PDB: 5TTU), a tyrosine kinase with similar active site was used to obtain a binding mode for the ITK-**16** complex (**Figure 34**). After protein structural alignment with an available ITK (PDB: 3MIY) X-ray structure the binding mode was converted into ITK. Finally, **7KV** was modified into ligand **16**, producing the initial ITK-**16** protein-ligand complex for the TI and US calculations (**Figure 35**).



**Figure 34.** Investigated compound **16** inhibiting ITK, BTK and BMX, and compound **7KV** whose complex with JAK3 (PDB: 5TTU) served as template for the binding mode of **16**.



**Figure 35.** Modelled binding mode of ligand **16** in ITK derived from the X-ray structures of ITK (PDB: 3MIY) and the **7KV**-JAK3 complex (PDB: 5TTU).

The “hybrid” BTK and BMX structures were constructed by mutating specific active site residues of ITK. The selection of the residues to be mutated was based on the structural alignment of ITK, BTK and BMX enzymes (**Figure A14**). Residues within a 3 Å radius of the bound ligand with sidechains pointing towards the ligand were selected to be modified, constructing the hybrid proteins. This procedure requires 5 and 6 mutations for generating hybrid BTK and BMX, respectively, from ITK, while generating the hybrid BMX from the hybrid BTK requires only one mutation (**Figure 36**).

ITK and BTK					
Query	358	IDPSELTFVQEI	IGSGQFGLVHLGYWLNKDKVAIKTIREGAMSEEDFIEEA	EVMMLKLSHPK	417
Sbjct	397	...KD...LK.	L.T....V.KY.K.RGQYD....M.K.S...DE.....K...N...E.		456
Query	418	LVQLYGVLCLEQAPICLV	FEFMEHGCLSDYLRTQRGLFAAETLLGMCLDVCEGMAYLEEAS		477
Sbjct	457	.....TK.R..FIIT	.Y.AN...LN...EM.HR.QTQQ...E..K....A.E...SKQ		516
Query	478	VIHRDLAARNCLVGENQVIKVS	DFGMTRFVLDQYTSSTGTFKFPVKWASPEVFSFSRYSS		537
Sbjct	517	FL.....NDQG.V.....LS.Y....E....V.S....R.SP...LMY.KF..			576

ITK and BMX					
Query	362	ELTFVQEI	IGSGQFGLVHLGYWLNKDKVAIKTIREGAMSEEDFIEEA	EVMMLKLSHPKLVQL	421
Sbjct	416	.I.LLK.	L.....V.K..K.KGQYD..V.M.K.S...DE.FQ..QT.....KF		475
Query	422	YGVLCLEQAPICLV	FEFMEHGCLSDYLRTQRGLFAAETLLGMCLDVCEGMAYLEEASVIHR		481
Sbjct	476	...SKEY..YI.	T.YISN...LN...SHGKGLEPSQ..E..Y.....F..SHQF...		535
Query	482	DLAARNCLVGENQVIKVS	DFGMTRFVLDQYTSSTGTFKFPVKWASPEVFSFSRYSSKSDV		541
Sbjct	536	.....DRDLCV.....Y.....V..V.....SA....HYFK.....			595

BTK and BMX					
Query	394	GSWEIDPKDLTFLKELGTGQFGVV	KYKWRGQYDVAIKMIKEGSMSEDEFIEEA	KVMMNL	453
Sbjct	408	.HM.LKREEI.L.....S.....L...K.....V.....FQ..QT..K.			467
Query	454	SHEKLVQLYGVCTKQRPIFIIT	EYMANGCLLNLYLREMRHRFQTQQLLEMCKDVCEAMEYL		513
Sbjct	468	..P...KF....S.EY..Y.V...	IS.....SHGKGLEPS.....Y....G.AF.		527
Query	514	ESKQFLHRDLAARNCLVNDQGVVKVS	DFGLSRYVLDDEYTSSVGSKFPVRWSPPEVLMYS		573
Sbjct	528	..H..I.....DRDLC.....MT.....Q.V....T....K..A...FH.F			587

**Figure 36.** Alignments of ITK, BTK and BMX near the active site (dots represent identical amino acids, active site residues are highlighted with grey, the residues considered for active site mutations are highlighted with orange).

Therefore, the binding free energy difference of **16** binding to hybrid BTK and to hybrid BMX is feasible to calculate by thermodynamic integration. However, the transformation of ITK to either hybrid BTK or BMX would be beyond the applicability domain of TI due to poor sampling of the significant conformational changes.

The binding free energy difference between **16** binding to BTK and BMX was evaluated by sidechain mutational TI, namely mutating Met439 of ITK to an isoleucine. The calculated result for the  $\Delta\Delta G$  is  $2.3 \text{ kcal}\cdot\text{mol}^{-1}$  which is close to the experimental  $2.8 \text{ kcal}\cdot\text{mol}^{-1}$  derived from the  $K_i$  values. We also performed the ITK to BTK, and ITK to BMX transformations, but no sensible results were obtained, owing to the abovementioned sampling issues. The collection of the experimental and calculated properties is shown in **Table 6**. The binding mode and characteristic interactions, particularly the hydrogen bonds between atoms 1, 2 and 3 of **16** (**Figure 34**) and residues Cys442 (NH), Met438 (NH) and Glu436 (O), respectively, were found to be intact during the complete mutation (**Figure 35**).

**Table 6.** Experimental inhibition constants for the non-covalent binding ( $K_i$ ), corresponding binding free energies ( $\Delta G_{\text{exp}}$ ) and their difference ( $\Delta\Delta G_{\text{exp}}$ ), calculated (FF14SB) binding free energy difference ( $\Delta\Delta G_{\text{calc}}$ ), experimental inhibition rate constant for the covalent reaction ( $k_{\text{inact}}$ ) corresponding free energy barrier ( $\Delta G_{\text{exp}}^\ddagger$ ), calculated (DFTB3/FF14SB) free energy barrier ( $\Delta G_{\text{calc}}^\ddagger$ ) for the binding of **16** to different kinases. The applied temperature for both the experimental measurements and the calculations is 27 °C. (Free energies are in kcal·mol<sup>-1</sup>. Experimental data were taken from ref. [148])

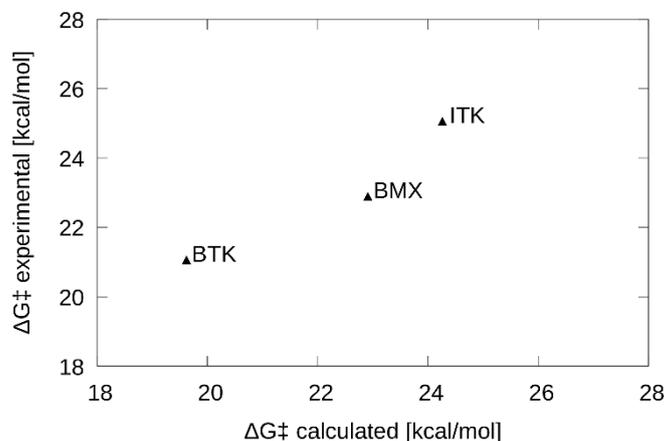
Enzyme	$K_i$ [ $\mu\text{M}$ ]	$\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$ [kcal·mol <sup>-1</sup> ]	$k_{\text{inact}}$ [min <sup>-1</sup> ]	$\Delta G_{\text{exp}}^\ddagger$	$\Delta G_{\text{calc}}^\ddagger$
ITK	0.0269	-10.4	-	-	0.000144	25.1	24.3
BTK	62.3	-5.8	-	-	0.124	21.1	19.6
BMX	0.545	-8.6	-2.8 <sup>a</sup>	-2.3 <sup>a</sup>	0.00564	22.9	22.9

<sup>a</sup>BTK to BMX mutation

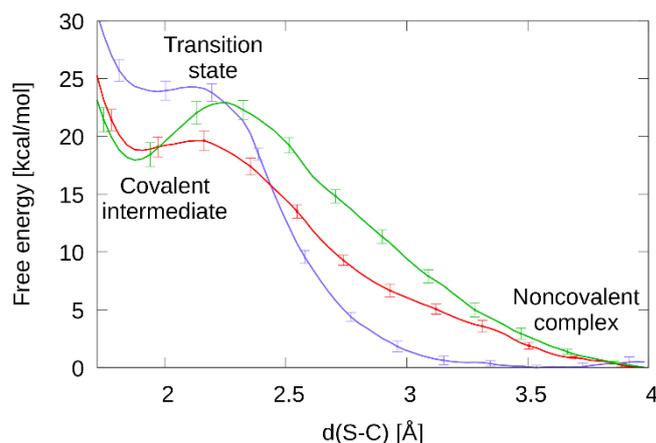
Ligand **16** contains an acrylamide warhead moiety, therefore reacts with the targeted cysteines of the three kinases through Michael-addition reaction. The covalent reactions were investigated by PMF calculations with MD using umbrella sampling. The reaction proceeds as described for the EGFR ligands studied, following the consecutive mechanism. Therefore, only the rate determining nucleophilic attack of the cysteine thiolate on the  $\beta$ -carbon of the acrylamide moiety was studied. We assumed the targeted cysteine residue is deprotonated, although, the Asp445 residue of ITK is found to increase the pK<sub>a</sub> of Cys442 [82]. We proposed that the nearby aspartate is able to deprotonate the targeted cysteine providing the active thiolate form, consequently, the simulation of the reaction between **16** and ITK included a protonated Asp445 residue. The experimental and calculated values characterizing the covalent binding step are shown in **Table 6**.

The experimental barrier of BMX was reproduced well in the calculated PMF of the **16**-hybrid BMX reaction. The barriers of **16** against the hybrid BTK and ITK reactions were slightly underestimated (**Table 6** and **Figure 37**) compared to the experimental values. Trajectory analysis revealed subtle differences in the enzyme-ligand complexes. The negatively charged thiolate of ITK interacts with the Ser444 hydroxyl group, while the protonated Asp445 carboxyl group and the warhead does not adopt a position similar to that in the hybrid BTK and BMX complexes (**Figure A15**). In BTK and BMX, a hydrogen bond can be found between the carbonyl oxygen of the warhead and the amide hydrogen of Cys442 (ITK numbering), while this hydrogen bond is absent in the ITK complex due to steric reasons. The presence or absence of this hydrogen bond affects the reaction energy of the nucleophilic attack as the stabilization of the carbanion species results in lowered transition state energy, as it is observed in the BTK and BMX reactions (potential well at the reaction coordinate below 2 Å in **Figure 38**). The covalent intermediate state of the ITK-**16** complex with 5 kcal·mol<sup>-1</sup> higher free energy indicates the absence of the mentioned interaction. Additionally, the negatively charged thiolate is able to form further hydrogen bonds with nearby donors at the binding site. The analysis of the MD simulations of ITK shows that the thiolate interacts with the sidechains of Ser444 and the protonated Asp445. Interaction with the backbone amide hydrogen of the same Asp445 residue was also observed. The former two stabilize the negatively charged sulfur, resulting in reduced nucleophilicity of the thiolate, thus higher transition barrier compared to the other kinases. The previously discussed interactions are partly missing in simulations of BTK and BMX. These complexes show the analogue of the latter interaction only; specifically, the sulfur accepts a hydrogen bond from the backbone amide nitrogen of the asparagine replacing Asp445

in these enzymes. Similarly to ITK, this interaction reduces the nucleophilicity of the thiolate in BMX and BTK, although it is more frequent in case of BMX, explaining the elevated barrier compared to BTK. To conclude, the experimental trend of the increasing barrier heights of BTK, BMX and ITK (**Figure 37**) was reproduced well in our MD based free energy calculations, underlining the usefulness of the method in potential off-target ranking.



**Figure 37.** Comparison of the experimental and calculated reaction barriers for the covalent binding of **16** to ITK, BTK and BMX.

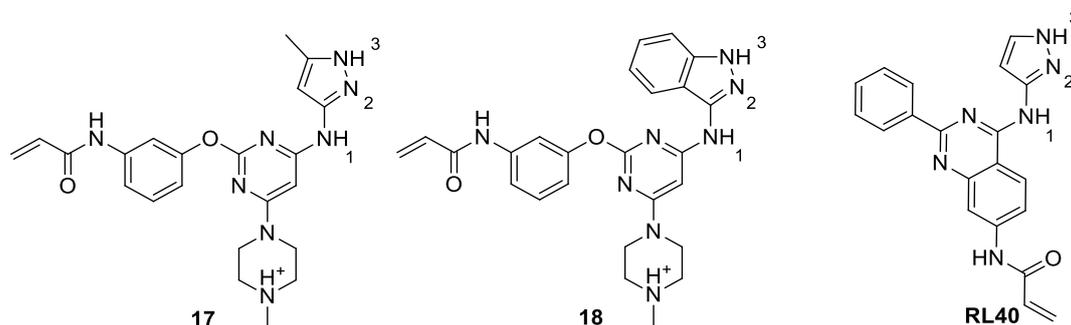


**Figure 38.** PMFs of the covalent reaction between **16** and ITK (blue), **16** and BTK (red), **16** and BMX (green).

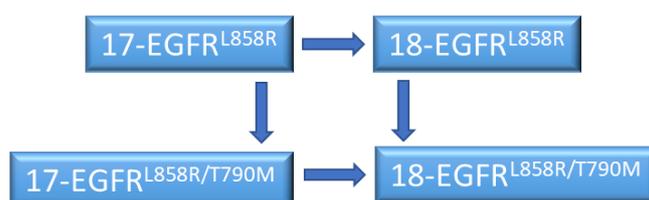
### Selectivity predictions for covalent inhibitors of EGFR mutants:

Finally, we selected two acrylamides **17** and **18** inhibiting both the L858R and L858R/T790M double oncogenic mutants of EGFR (**Figure 39**). The selection was made due to the vastly different non-covalent affinity against the targeted proteins as shown by the over two and three order of magnitude difference in  $K_i$  for **17** and **18**, respectively (**Table 7**). However, only modest differences can be observed in their  $k_{inact}$  values characterizing the chemical reaction of the covalent step. Therefore, the main focus was on the non-covalent binding of these compounds, which was investigated by specific sidechain mutational and ligand transformational TI calculations (**Figure 40**). We used the X-ray crystal structure of the

**RL40**-cSrc-DM complex (PDB: 5D12) to estimate the initial binding mode of the molecules in both the single and double mutant EGFRs (**Figure 41**).



**Figure 39.** Examined molecules (**17** and **18**) for EGFR mutations in their predominant protonation state and the covalently bonded molecule in PDB 5D12 (**RL40**).



**Figure 40.** Applied transformations during the TI calculations of EGFR mutants examining selectivity differences.

Calculated  $\Delta\Delta G$  of the T790M mutations for both **17** and **18** show negative values indicating better affinity to the double mutant enzyme that is in accordance with the experiments (**Table 8**); however, the binding affinity differences are slightly underestimated in both cases. Though, the more pronounced  $\Delta\Delta G$  difference of **18** compared to **17** (larger by  $1.9 \text{ kcal}\cdot\text{mol}^{-1}$  in experiment and by  $2.6 \text{ kcal}\cdot\text{mol}^{-1}$  in calculation) was correctly reproduced. While sidechain mutational TI calculations show the effect of the changes in the protein environment on the non-covalent binding, the alchemical transformations of ligands binding to the same EGFR variant clarifies the effect of ligand structural variation on binding free energies. Consequently, we carried out ligand transformational TI calculations, modifying **17** into **18** in both EGFR mutants. The results show slightly worse binding affinity for **18** in both the single and double mutants with  $\Delta\Delta G$  of  $0.5 \text{ kcal}\cdot\text{mol}^{-1}$  and  $0.3 \text{ kcal}\cdot\text{mol}^{-1}$ , respectively. In contrast, the experimental data show higher affinity of **18** towards the double mutant compared to **17**. However, the difference between the calculated and experimental  $\Delta\Delta G$ s does not exceed  $1 \text{ kcal}\cdot\text{mol}^{-1}$  in either mutant.

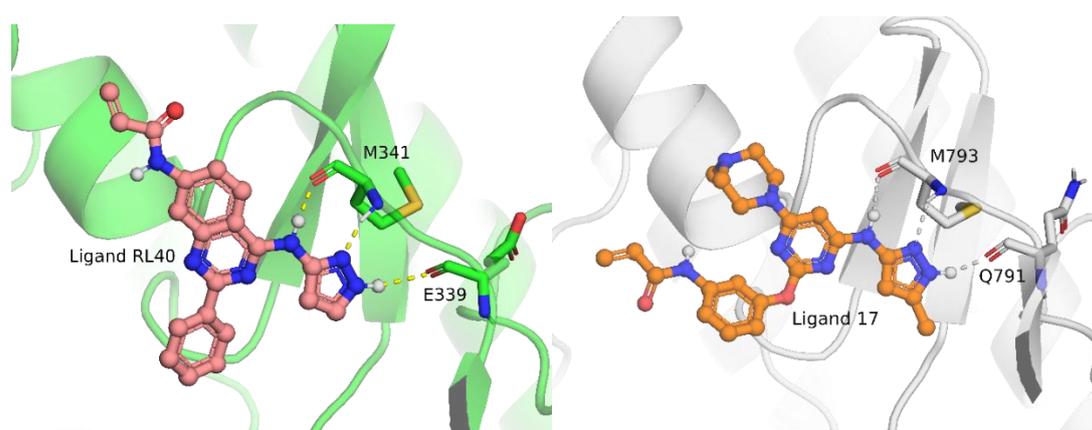
**Table 7.** Experimental non-covalent binding affinity ( $K_i$ ), corresponding binding free energy ( $\Delta G_{\text{exp}}$ ) and the experimental inhibition rate constant ( $k_{\text{inact}}$ ) with the corresponding experimental activation free energy ( $\Delta G_{\text{exp}}^\ddagger$ ) of the covalent binding step for inhibitors of the EGFR mutants. (Experimental data were taken from ref. [147].)

Ligand	EGFR mutant	$K_i$ [nM]	$\Delta G_{\text{exp}}$ [kcal·mol <sup>-1</sup> ]	$k_{\text{inact}}$ [min <sup>-1</sup> ]	$\Delta G_{\text{exp}}^\ddagger$ [kcal·mol <sup>-1</sup> ]
<b>17</b>	L858R	70.2	-9.8	0.017	22.2
<b>17</b>	L858R/T790M	0.64	-12.6	0.116	21.1
<b>18</b>	L858R	833	-8.3	0.055	21.5
<b>18</b>	L858R/T790M	0.32	-13.0	0.137	21.0

**Table 8.** Experimentally derived ( $\Delta\Delta G_{\text{exp}}$ ) and calculated (FF14SB) ( $\Delta\Delta G_{\text{calc}}$ ) binding free energy differences for the sidechain and ligand mutations. The applied temperature for both the experimental measurements and the calculations is 27 °C. (Experimental data were taken from ref. [147].)

Ligand	Mutation	$\Delta\Delta G_{\text{exp}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc}}$ [kcal·mol <sup>-1</sup> ]
<b>17</b>	T790 > M790	-2.8	-0.7
<b>18</b>	T790 > M790	-4.7	-3.3
Enzyme	Transformation	$\Delta\Delta G_{\text{exp}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc}}$ [kcal·mol <sup>-1</sup> ]
L858R	<b>17</b> > <b>18</b>	1.5	0.5
L858R/T790M	<b>17</b> > <b>18</b>	-0.4	0.3

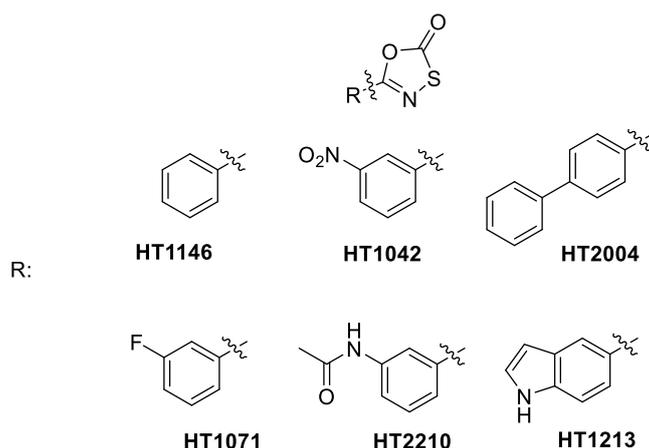
After the trajectory analysis of the corresponding simulations a hydrogen bond system was revealed that is formed by the Gln791 carbonyl oxygen, the Met793 carbonyl oxygen and the Met793 amide hydrogen with atoms 1, 2 and 3 (**Figure 39** numbering) of the inhibitor molecules (**Figure 39** and **Figure 41**). These hydrogen bonds remained intact throughout the simulations. Additionally, rotation around the bond between the ether oxygen and the carbon atom of the heteroaromatic ring of the warhead was observed enabling a binding mode similar to that found previously for EGFR inhibitors **11-15**.



**Figure 41.** left: Binding mode of ligand **RL40** in cSrc double mutant (PDB: 5D12); right: Modelled binding mode of ligand **17** in EGFR double mutant (PDB: 3W2Q) based on the cSrcDM-**RL40** complex.

## 5.4 Selectivity and inhibition mechanism of covalent immunoproteasome inhibitors [201]

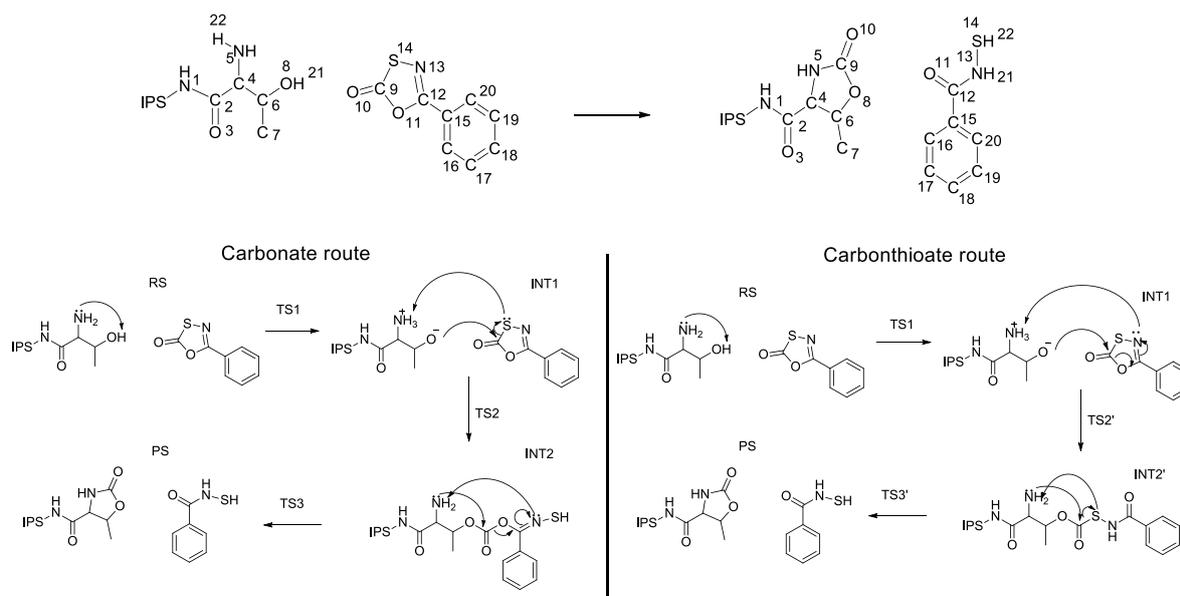
Finally, we studied the covalent inhibition of immunoproteasome (iPS) by a set of oxathiazolone derivatives experimentally characterized in ref [97] (**Figure 42**). The main goal was to validate the developed free energy calculations also on complex binding reactions, as oxathiazolone compounds inhibit iPS through a multistep, yet not clarified mechanism [202]. In addition, we further validated the applicability of the sidechain mutational selectivity calculations, based on the “hybrid” protein approach.



**Figure 42.** Investigated iPS inhibitors.

### 5.4.1 Mechanistic description of the covalent inactivation reaction

Fan et. al [97] proposed two possible reaction mechanisms for the oxathiazolone inhibition reaction. Both mechanisms (**Figure 43**) are initiated by the activation of the Thr1 residue by a proton transfer between the hydroxy and terminal amino group. The two mechanisms branch at the nucleophilic attack of the alcoholate with the difference found in the disruption of the oxathiazolone ring. The carbonate route contains the breaking of the bond between the carbon and the sulfur atom, while the carbonthioate pathway proceeds through the splitting of the carbon-oxygen bond. The following step consists of a proton transfer from the Thr1 terminal  $\text{NH}_3^+$  group in both mechanisms. Thereafter, a second nucleophilic attack occurs in both cases, in which the amino group attacks the electrophilic carbon atom, originated from the oxathiazolone ring. The reaction terminates with a final proton transfer from the heterocyclic amino group yielding the same product in both the carbonate and carbonthioate pathways. After the formation of the oxazolidinone ring, the product compound decomposes into an amide and leaves the active site; however, these events have no effect on the preceding covalent binding. In our work we focused solely on the inhibition mechanism (**Figure 43**), though, it also has been proposed [202] that both INT2 and INT2' can decompose by hydrolysis, resulting in the reformation of Thr1 and preserved enzyme activity.



**Figure 43.** Examined reaction mechanisms of **HT1146** with IPS. Atom numbering is shown in the overall reaction scheme on the top.

### Activation of the Thr1 residue:

Varying protonation states of residues forming the active site of iPS have been proposed, and the reaction of the Thr1 residue of the proteasome with several ligands has been also investigated [203]. A number of studies proposed that the nucleophilic attack of the hydroxy oxygen of Thr1 is accompanied by proton transfer either to the N $\epsilon$  of Lys33 [204,205] or to the terminal amine of Thr1 [206,207]. The activation of the Thr1 residue, namely the proton transfer between the O $\gamma$ H and NH $_2$  groups of Thr1 can proceed either directly or via a water molecule. Structure analysis revealed that the position of the O $\gamma$ H and NH $_2$  groups is compatible with both proton transfer mechanisms. A ligand-dependent proton transfer mechanism is also proposed in previous theoretical studies of the human proteasome. According to these investigations the peptide hydrolysis starts with water-mediated proton transfer [48], while the first step in covalent inhibition by syringolin A [47] and epoxomicin [46] proceeds through direct transfer.

Analysis of the 4 ns relaxation MD trajectory of the **HT1146**-iPS complex revealed that the proximity of a water molecule to the Thr1 residue is a rare event, therefore we excluded the water assisted transfer and investigated only the direct proton transfer between the O $\gamma$ H and NH $_2$  groups of Thr1. This assumption agrees with the proton transfer proposed in the presence of other covalent inhibitors [206,207]. The obtained activated oxalate is found to be stabilized via interactions with the NH $_3^+$  groups of Thr1 and Lys33 when the DFTB3/FF14SB potential was applied. Note that the free energy profile shown in **Figure 44** shows the activation step with a single curve for both mechanism as they are identical in both pathways. Interestingly, the obtained free energy of 14.4 kcal $\cdot$ mol $^{-1}$  related to TS1 is in line with the 12.3 kcal $\cdot$ mol $^{-1}$  and 12.0 kcal $\cdot$ mol $^{-1}$  energy values obtained with B3LYP/6-31++G\*\*/FF14SB potential for the reaction barrier of the same proton transfer [206,207].

### **Carbonate route:**

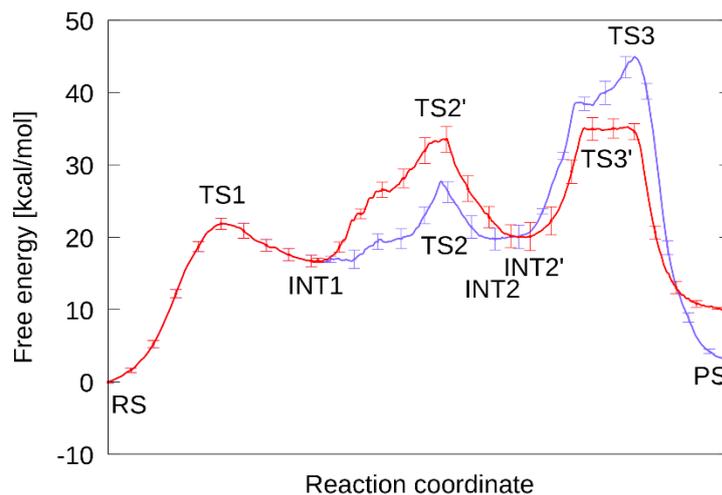
According to the carbonate pathway, the activation step is followed by the O8-C9 bond formation and a second proton transfer from N5 to S14 yielding the INT2 state (**Figure 43**). We performed a series of SMD simulations for this step with various combinations of the relevant atom distances as reaction coordinates. The simulations yielded work curves corresponding to an asynchronously concerted process in which the formation of the O8-C9 bond is fairly advanced when the proton moves from N5 to S14. Subsequent umbrella sampling simulations showed similar results using structures from two-dimensional SMD simulations along the conformational space defined by the (C9-S14)–(O8-C9) and (N5-H22)–(S14-H22) distance combinations (**Figure A16**). The second nucleophilic attack in which N5 attacks C9 is followed by the final proton transfer from N5 to N13, producing the product (PS) structure. Once the N5-C9 bond was formed and the C9-O11 bond was broken, a rotation along the C12-C15 bond was observed that brings N5 and N13 into a relative position that facilitates the proton transfer between them. The rotation around the C15-C12 bond is accompanied by the interaction between the O11 carbonyl oxygen and the positively charged amine group of Lys33, which interaction stabilizes the excess negative charge on the leaving group during the ring formation process.

### **Carbenthioate route:**

The carbenthioate pathway differs from the carbonate route at the oxathiazolone ring disruption, namely the transition between intermediate INT1 and INT2'. The deprotonated O8 attacks the carboxylic C9 similarly to the analogous step in the carbonate route, but during the formation of INT2' the C9-O11 bond breaks, while the proton transfer occurs asynchronously between N5 to N13 (**Figure 43**). In the following nucleophilic attack N5 binds to C9 and the remaining C-S bond between the ligand and Thr1 breaks. Finally, the last proton transfer from N5 to the leaving group, producing PS proceeds without conformational rearrangement, as the S14 atom is in a proper position to accept the abstracted proton. As observed in the previous simulations, the bond formation between the attacking agent and the acceptor atom is fairly advanced during the proton transfer, highlighting the asynchronous nature of the studied transitions.

### **Free energy profile of the studied reactions and QM correction:**

After the exploration of the possible reaction pathways with steered molecular dynamics (SMD) simulations, we were able to calculate the potential of mean force curves of the aforementioned reaction steps with QM/MM umbrella sampling simulations. The PMF curves based on the results of the MD simulations were constructed by WHAM and are shown in **Figure 44**.



**Figure 44.** Calculated PMF curves for the carbonate (blue) and carbonthioate (red) pathways of **HT1146**. Structure labels are identical to those in **Figure 43**.

As previously mentioned, the activation of the Thr1 residue is identical in the mechanisms, therefore they share the same free energy profile. In the next step both the carbonate (INT1→INT2) and carbonthioate (INT1→INT2') pathways contain a nucleophilic attack coupled with a proton transfer from N5. Although the two steps occur consecutively, no intermediate was found between them according to the constructed PMF. The second barriers (TS2, TS2') were found to be higher than TS1, with more pronounced difference for TS2', while the second intermediates (INT2, INT2') show slightly higher free energies than INT1. The final step of the two routes, INT2→PS in the carbonate and INT2'→PS in the carbonthioate pathway proceeds with the largest barriers. However, TS3' is observed at the same energy level as TS2', while the difference between TS3 and TS2 shows larger gap with almost 20 kcal·mol<sup>-1</sup> in favor of TS2. The calculated free energies of the overall pathways (free energy of PS) are slightly different, though, theoretically they must be identical as the same product is formed in both routes. The factor that is responsible for the reaction free energy difference is the structural difference at the product states; the alternative pathways contain different product conformations accompanied by differing enzyme-solvent environment, which might explain the modest numerical variation in the product state energies.

The inaccuracy of the obtained reaction free energy profiles is suggested by two observations. Firstly, we obtained endothermic reaction free energies for both routes, and secondly, the reaction barrier of 22.4 kcal·mol<sup>-1</sup> corresponding to the experimental  $k_{\text{inact}}$  of **HT1146** ( $9.3 \cdot 10^{-4} \text{ s}^{-1}$ ) [97] is largely overestimated by our calculations. These shortcomings can be attributed to the inaccuracies of the approximate, semiempirical nature of DFTB3. However, corrections applied to the QM region with more involved potentials are shown to be an efficient method to improve the results and to obtain better correlation with the experimental data [122].

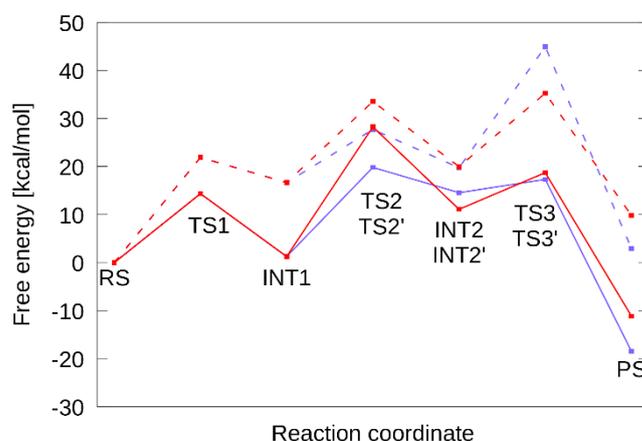
Therefore, we performed QM correction applying the  $\omega$ B97XD potential. We collected the free energies of the reaction, transition and intermediate states of the examined pathways and corrected them with the difference between the  $\omega$ B97XD and DFTB3 single point energies obtained from QM calculations. (see the details at **Chapter 4 – Materials and methods**). The results of the corrections and the corrected PMF are shown in **Table 9** and in **Figure 45**.

**Table 9.** Results of the QM correction calculations. All entries are energies in kcal·mol<sup>-1</sup> units. The amount of correction is the difference between the  $\omega$ B97XD/aug-cc-pVTZ and DFTB3 energies.

<i>Carbonate route</i>							
Structure	RS	TS1	INT1	TS2	INT2	TS3	PS
$\omega$ B97XD SP	0.0	27.1	45.7	12.0	3.6	9.3	-22.9
DFTB3 SP	0.0	34.7	61.1	19.9	8.8	36.9	-1.5
Difference	0.0	-7.5	-15.4	-7.9	-5.2	-27.6	-21.4
PMF	0.0	21.9	16.7	27.7	19.8	44.9	2.9
Corr. PMF	0.0	14.4	1.3	19.8	14.6	17.3	-18.5

<i>Carbonythioate route</i>							
Structure	RS	TS1	INT1	TS2'	INT2'	TS3'	PS
$\omega$ B97XD SP	0.0	27.1	45.7	33.2	-9.1	-12.7	-21.4
DFTB3 SP	0.0	34.7	61.1	38.5	-0.2	3.8	-0.4
Difference	0.0	-7.5	-15.4	-5.3	-8.9	-16.5	-21.0
PMF	0.0	21.9	16.7	33.6	20.0	35.3	9.8
Corr. PMF	0.0	14.4	1.3	28.3	11.1	18.7	-11.1

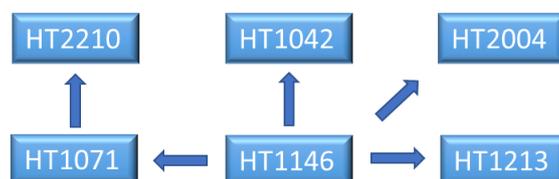


**Figure 45.** Original (dashed line) and QM corrected (straight line) free energy profiles for the reaction of HT1146 with IPS calculated on the carbonate (blue) and carbonythioate (red) pathways.

After the application of high-level QM correction, the free energy profiles of the reactions were significantly changed. According to the corrected PMFs the carbonate pathway seems to be the favored mechanism over the carbonythioate route. Noteworthy, in both cases the height of the third barrier has been lowered, therefore the second step – the nucleophilic attack of the deprotonated oxygen and the accompanied N5-S14 proton transfer – turned out to be the rate-determining step. The corrected barrier for the carbonate route is 19.8 kcal·mol<sup>-1</sup>, which is lower than the experimental value of 22.4 kcal·mol<sup>-1</sup> obtained from  $k_{\text{inact}}$ . However, it indicates sensible reaction free energy profiles after the QM correction.

#### 5.4.2 Characterization of the non-covalent binding step for selected inhibitors

The non-covalent binding step of the examined inhibitors was evaluated by thermodynamic integration. We used the previously docked structure of **HT1146** as reference for the iPS-ligand complexes. The starting complexes were built by modifying **HT1146** inside the enzyme pocket (**Figure 46**). The complete description of the transformations can be found in **Chapter 4 – Materials and methods** and in the **Appendix**.



**Figure 46.** Applied transformations during the TI calculations

The calculated and experimental non-covalent binding free energies are shown in **Table 10**. The  $\Delta\Delta G$  values derived from the thermodynamic integration were converted into binding free energies by minimizing the root-mean-square deviation (RMSD) from the experimental results. Again, this shift does not affect the correlation between the calculated and experimental data and does not distort the compound ranking. As only the experimental  $k_{\text{inact}}/K_i$  ratio is available for **HT1071** and **HT2210**, the correlation cannot be shown either for the individual  $\Delta G$  (derived from  $K_i$ ), or for the  $\Delta G^\ddagger$  (derived from  $k_{\text{inact}}$ ) results.

#### 5.4.3 Evaluation of covalent binding step for selected inhibitors

Once the binding free energies were calculated for the selected compounds, we performed QM/MM MD umbrella samplings for the previously found rate determining step of the oxathiazolone compounds. The calculated free energy barriers were transformed into  $k_{\text{inact}}$  values, using **Eq. 3**. The selection of the compounds was made to maximize the range of inhibitory activity expressed with the  $k_{\text{inact}}/K_i$  ratio. We considered the activation step as ligand-independent, therefore only the barrier between INT1 and INT2 was calculated. The calculated barriers were corrected by the same value determined for the QM correction of the rate-determining step of **HT1146**. The constructed PMFs are shown in **Figure A12** and the corresponding barriers are presented in **Table 10**. Calculated barriers of the less active compounds **HT1071** and **HT2210** were found to be higher than the more potent ones', indicating that the activity drop for these compounds emerges from the lower  $k_{\text{inact}}$  values, which is more pronounced for **HT2210**.

**Table 10.** Experimental inhibition constants for the noncovalent binding ( $K_i$ ), corresponding binding free energies ( $\Delta G_{\text{exp}}$ ), experimental inhibition rate constants for the covalent reaction ( $k_{\text{inact}}$ ), corresponding free energy barriers ( $\Delta G_{\text{exp}}^\ddagger$ ) and experimental  $\log(k_{\text{inact}}/K_i)_{\text{exp}}$  values. Calculated (FF14SB) binding free energies ( $\Delta G_{\text{calc}}$ ), calculated (DFTB3/FF14SB) free energy barriers ( $\Delta G_{\text{calc}}^\ddagger$ ) and calculated  $\log(k_{\text{inact}}/K_i)_{\text{calc}}$  values for the selected immunoproteasome inhibitors. The applied temperature for both the experimental measurements and the calculations is 37 °C. (Energies are in  $\text{kcal}\cdot\text{mol}^{-1}$ . Experimental data were taken from ref. [97].)

Compound	$K_i$ [ $\mu\text{M}$ ]	$\Delta G_{\text{exp}}$	$k_{\text{inact}}$ $\times 10^3$	$\Delta G_{\text{exp}}^\ddagger$	$\log(k_{\text{inact}}/K_i)_{\text{exp}}^a$	$\Delta G_{\text{calc}}$	$\Delta G_{\text{calc}}^\ddagger$	$\log(k_{\text{inact}}/K_i)_{\text{calc}}^a$
<b>HT1146</b>	7.90	-7.2	0.93	22.4	2.1	-7.3	19.8	4.0
<b>HT2004</b>	1.4	-8.3	1.54	22.1	3.0	-9.3	19.0	6.0
<b>HT1042</b>	0.42	-9.0	0.38	23.0	3.0	-8.2	21.7	3.2
<b>HT1213</b>	1.1	-8.5	0.26	23.2	2.4	-8.3	21.6	3.4
<b>HT1071</b>					1.3	-6.8	22.5	1.7
<b>HT2210</b>					1.9	-7.8	24.1	1.3

<sup>a</sup> $k_{\text{inact}}$  is in  $\text{s}^{-1}$ , and  $K_i$  is in M units

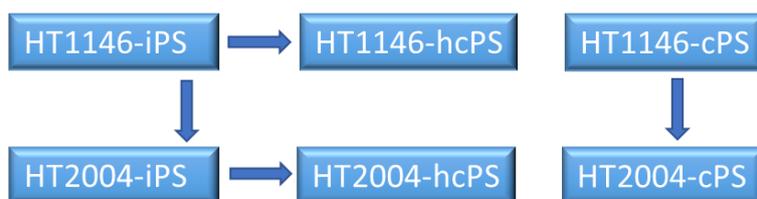
#### 5.4.4 Selectivity studies between iPS and cPS

**HT1146** exhibits slight selectivity between iPS and cPS in favor of iPS with  $k_{\text{inact}}/K_i=118$  and  $31.3 \text{ M}^{-1}\text{s}^{-1}$ , respectively, while **HT2004** is prominently more active against iPS, as shown by  $k_{\text{inact}}/K_i=1093$  and  $0.23 \text{ M}^{-1}\text{s}^{-1}$  for iPS and cPS, respectively. We performed further MD based free energy calculations to evaluate and clarify the significant selectivity difference between the two compounds. The differences in the  $k_{\text{inact}}/K_i$  ratios were described with free energy differences in the following way:

$$\log\left(\frac{k_{\text{inact}}[A]}{K_i[A]}\right) - \log\left(\frac{k_{\text{inact}}[B]}{K_i[B]}\right) = \frac{\log(e)}{RT}(-\Delta G^\ddagger[B] + \Delta G^\ddagger[A] - \Delta G[B] + \Delta G[A]) = \frac{\log(e)}{RT}(-\Delta\Delta G^\ddagger - \Delta\Delta G) \quad (20)$$

where  $\Delta G^\ddagger[A]$  and  $\Delta G[A]$  are the activation free energy and the noncovalent binding free energy, respectively, for compound A, while B denotes the other molecule. Differences in the transition barriers can be evaluated based on the constructed PMFs derived from QM/MM umbrella sampling calculations in iPS and in cPS, both with **HT1146** and **HT2004**. The binding free energy differences are calculated directly from TI simulations, applying the proper thermodynamic transformations (**Figure 47**).

The free energy barriers of the rate-determining step in iPS were calculated previously. To obtain the covalent barriers for cPS, we performed QM/MM MD simulations for the rate determining step, assuming the same reaction mechanism in cPS as found in iPS. Similarly, the previously computed binding free energy difference between **HT1146** and **HT2004** in iPS was complemented with the binding free energy difference in cPS, as obtained by thermodynamic integration. Experimentally derived and calculated free energies describing the binding differences between **HT1146** and **HT2004** are shown in **Table 11**.



**Figure 47.** Applied transformations during proteasome selectivity studies

**Table 11.** Comparison of experimentally derived and calculated (FF14SB and DFTB3/FF14SB) free energy differences between **HT1146** and **HT2004** binding (top two rows) and between binding to iPS and cPS (bottom two rows). (Experimental data were taken from ref. [97].)

Transformation	Experimental <sup>a</sup>	Calculated <sup>b</sup>
	$-\Delta\Delta G^\ddagger - \Delta\Delta G$ [kcal·mol <sup>-1</sup> ]	$-\Delta\Delta G^\ddagger - \Delta\Delta G$ [kcal·mol <sup>-1</sup> ]
HT1146 > HT2004 iPS	-1.4	-2.0 [0.0 <sup>c</sup> -2.0]
HT1146 > HT2004 cPS	3.0	4.5 [5.0 <sup>d</sup> -0.5]
HT1146 iPS > cPS	0.8	0.4 [2.0 <sup>e</sup> -1.6]
HT2004 iPS > cPS	5.1	6.7 [7.0 <sup>f</sup> -0.3]

<sup>a</sup>The sum of  $-\Delta G^\ddagger - \Delta G$  is obtained from the experimental  $\log(k_{\text{inact}}/K_i)$  difference using **Eq. 20**. <sup>b</sup>The differences in the barriers ( $\Delta\Delta G^\ddagger$ ) for the rate-determining step and binding free energy differences ( $\Delta\Delta G$ ) are separately calculated. <sup>c</sup>Obtained as the difference between the barriers of **HT2004** (11 kcal·mol<sup>-1</sup>) and **HT1146** (11 kcal·mol<sup>-1</sup>) in iPS. <sup>d</sup>Obtained as the difference between the barrier of **HT2004** (18 kcal·mol<sup>-1</sup>) and **HT1146** (13 kcal·mol<sup>-1</sup>) in cPS. <sup>e</sup>Obtained as the difference between the barrier of **HT1146** in cPS (13 kcal·mol<sup>-1</sup>) and iPS (11 kcal·mol<sup>-1</sup>). <sup>f</sup>Obtained as the difference between the barrier of **HT2004** in cPS (18 kcal·mol<sup>-1</sup>) and iPS (11 kcal·mol<sup>-1</sup>).

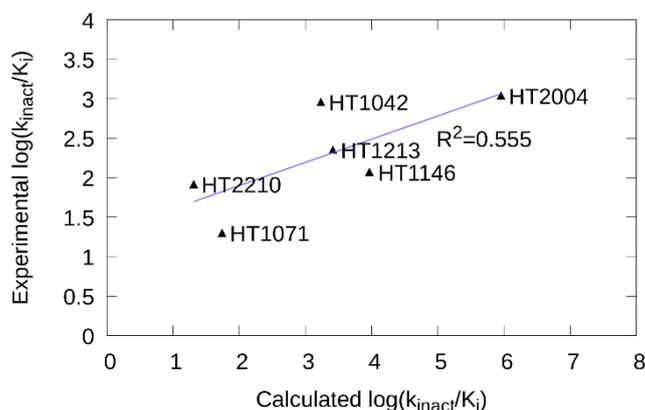
Although **Eq. 20** refers to different compounds in the same protein; an analogous equation can be written for a compound that binds to two similar proteins. In this latter case, letter A and B refers to iPS and cPS, respectively, both binding the same ligand. To compute the binding free energy difference for the same ligand but in different enzyme, we followed our previously developed protocol, described at the kinase selectivity studies. The structure and sequence analysis of the two active sites highlighted the Gln-Ser substitution at residue 53 as the most prominent difference, between the immuno- and constitutive proteasome. Although these pockets possess identical amino acid sequences, the different conformations of Met48 observed in  $\beta 5i$  and  $\beta 5c$  render the S1 pocket more spacious in  $\beta 5i$  than in  $\beta 5c$  [208,209]. The Gln53Ser mutation is proposed to affect the Met48 conformation. In addition, the interaction between Gln53 and Met48 in iPS is beneficial with an extended Met48 conformation that opens some space for the ligand between Ala49 and Met45.

Therefore, we performed sidechain mutational TIs with and without bound ligand (for more information please refer to **Chapter 4 – Materials and methods**) applying the aforementioned mutation to build the “hybrid” cPS. The results for both the covalent and non-covalent binding differences are shown in **Table 11**.

The barrier differences were derived from the already available PMFs of the covalent reactions of iPS and cPS with **HT1146** and **HT2004**.

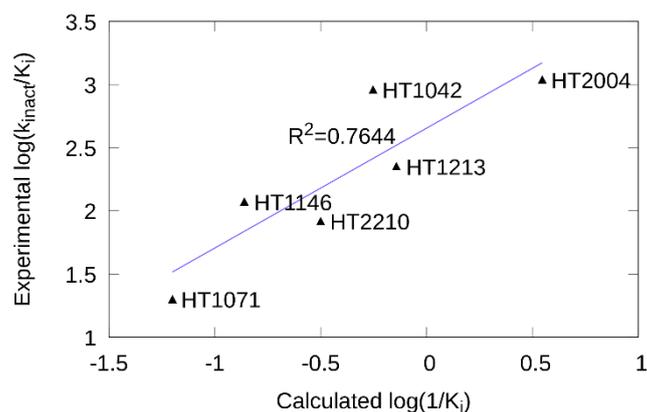
#### 5.4.5 Evaluation of the overall proteasome inhibition

The experimental  $K_i$  and  $k_{\text{inact}}$  values when converted to binding free energies ( $\Delta G$ ) and reaction barrier free energies ( $\Delta G^\ddagger$ ) cover narrow ranges of  $2 \text{ kcal}\cdot\text{mol}^{-1}$  and  $1 \text{ kcal}\cdot\text{mol}^{-1}$ , respectively. The range of the experimental free energies corresponding to the experimental  $k_{\text{inact}}/K_i$  ratios, namely the  $-\Delta G - \Delta G^\ddagger$  expression also around  $1 \text{ kcal}\cdot\text{mol}^{-1}$  (top 4 entries in **Table 11**). Therefore, the less active compounds **HT1071** and **HT2210** were added to the compound selection to broaden the range of experimental range. In this way, we evaluated the complete proteasome inhibition process including both the non-covalent and covalent steps for six compounds. The correlation between the calculated and experimental  $\log(k_{\text{inact}}/K_i)$  ratios are shown in **Figure 48**, and the data describing the complete proteasome inhibition are collected in **Table 10**. A reasonable correlation with  $R^2=0.555$  ( $p(R)=0.089$ )<sup>3</sup> for the six investigated oxathiazolones was observed. Interestingly, higher correlation  $R^2=0.76$ , ( $p(R)=0.022$ )<sup>3</sup> was found between the experimental  $\log(k_{\text{inact}}/K_i)$  and calculated  $\log(1/K_i)$  (**Figure 49**). This suggests that the differences in the iPS inhibitory activity of these compounds is primarily affected by the molecular recognition, the non-covalent binding. The examined compounds have identical warhead that have virtually no electronic crosstalk with the non-warhead region. Therefore, similar covalent reactivity is expected for these molecules that suggests the more prominent influence of the non-covalent recognition. Indeed, the low calculated affinity for noncovalent complex formation by **HT1071** explains the low inhibitory activity, which is actually the lowest among the compounds investigated. However, the modest activity of **HT2210** is a consequence of both the lower binding free energy and the low reactivity of this compound in the active site of iPS. This observation highlights the potential effect of the binding mode on the covalent reactivity.



**Figure 48.** Comparison of the experimental and calculated  $\log(k_{\text{inact}}/K_i)$  ( $k_{\text{inact}}$  in  $\text{s}^{-1}$  and  $K_i$  in M units).

<sup>3</sup> The p (probability) value of the correlation t-test.

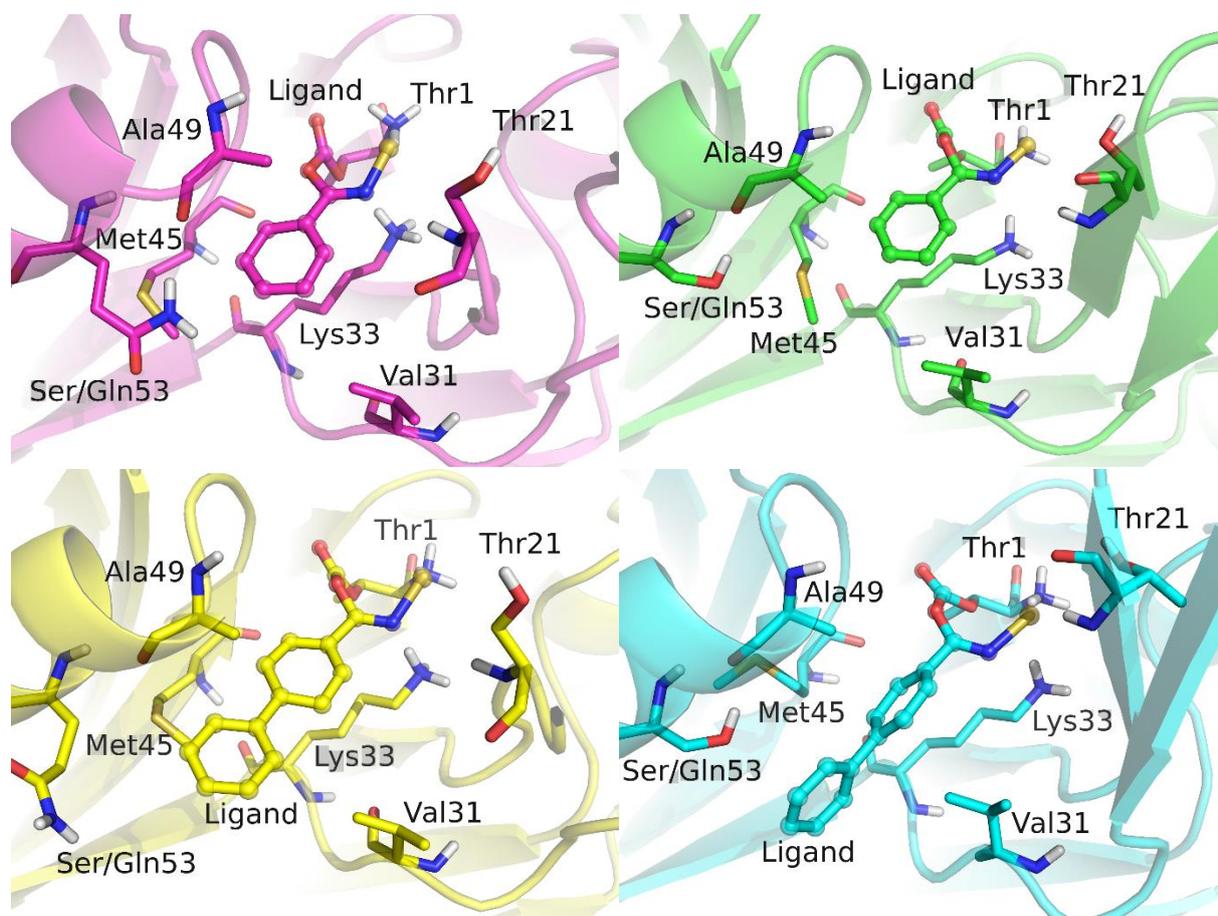


**Figure 49.** Comparison of the experimental  $\log(k_{inact}/K_i)$  and calculated  $\log(1/K_i)$  ( $K_i$  in M units).

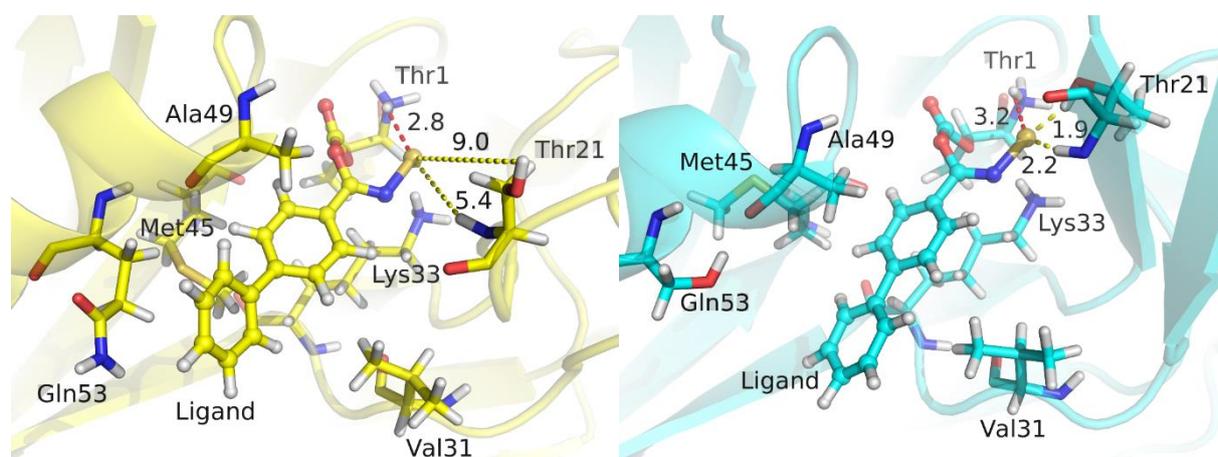
The binding mode and its consequences on the covalent step was examined in connection with the iPS and cPS selectivity studies. The structure of the non-covalent complex of the oxathiazolones positions the carbonyl group of the ring in the proximity of the  $O\gamma H$  group of Thr1 and orients the non-warhead region of the molecules towards the S1 binding pocket. The S1 pocket was found to be larger in the immunoproteasome [209] than in the constitutive counterpart. This led us to investigate how the abovementioned structural difference together with our computational results can explain the variation of the affinity of **HT1146** and **HT2004** towards iPS and cPS. Note that only the experimental  $k_{inact}/K_i$  ratio is available against cPS. An analysis of the computed energies showed that the selectivity of **HT2004** is primarily governed by the different reaction barriers in iPS and cPS with the latter possessing  $7 \text{ kcal}\cdot\text{mol}^{-1}$  larger barrier, while only a  $-0.3 \text{ kcal}\cdot\text{mol}^{-1}$  alteration was found in the ligand's binding free energy difference towards iPS and cPS. Visual inspection of the transition state structure of **HT2004** revealed a binding pose of the ligand with the proximal phenyl ring penetrating the extended space of the S1 pocket in iPS (**Figure 50**). This cavity between Met45 and Ala49 is formed due to the specific conformation of Met45 in iPS and allows the warhead of **HT2004** to adopt a position similar to what was observed for the oxathiazolone ring of **HT1146**. In contrast, the structure of the cPS active site shows a hydrogen bond between Ser53 (Gln53 in iPS) and the backbone carbonyl of Ala49 and does not allow Met45 to adopt the conformation with the extended space. Consequently, the proximal phenyl ring of **HT2004** in the cPS active site has a position shifted with respect to that in the iPS complex and resembles that seen in the **HT1146**-cPS complex. In addition, a slight rotation around the proximal phenyl ring is observed for **HT2004** in order to avoid steric clash between the terminal benzene ring and Val31. This rotation places the warhead of **HT2004** in cPS into a position different from the oxathiazolone orientation seen in the **HT1146**-iPS, **HT1146**-cPS and **HT2004**-iPS complexes. This shift hinders the proton transfer of the rate-determining step. The negatively charged sulfur is placed further from the attacking hydrogen atom; moreover, the interactions of the sulfur atom with Thr21 amide and alcoholic hydrogens prevent the negatively charged atom to approach the proton donating nitrogen (**Figure 51**).

According to the previously discussed observations the altered binding mode affects primarily the covalent reaction barrier through warhead positioning, while the non-covalent affinity is less affected. To conclude, small variations in the binding mode of compounds with similar intrinsic reactivities may result in highly different affinities towards the immunoproteasome. The observed changes in the warhead position relative to the targeted residue and in the protein-ligand interactions can substantially affect reactivity. This structural

characteristic has significant implications for the affinity and selectivity of covalent iPS inhibitor design.



**Figure 50.** Representative structures from umbrella sampling simulations around the transition states of the rate-determining step: **HT1146-iPS** (top left); **HT1146-cPS** (top right); **HT2004-iPS** (bottom left); **HT2004-cPS** (bottom right). The position of the oxathiazolone ring of **HT2004** in cPS is shifted owing to the lack of space between Met45 and Ala49 and to avoid steric clash with Val31.



**Figure 51.** **HT2004-iPS** (left) and **HT2004-cPS** (right) complex structures between the nucleophilic attack and the proton transfer of the rate-determining step. Distances between the ligand's sulfur atom and Thr21 H $\gamma$  and H are shown with yellow dashed lines. The distance between the transferred proton and the sulfur atom is shown with a red dashed line.

## CHAPTER 6 – CONCLUSIONS

### 6.1 Conclusions and thesis points

The main objective of this work was to develop a robust and efficient computational method to characterize the covalent inhibition of protein targets. Covalent inhibition is composed of a molecular recognition (described by  $K_i$ ) and a chemical reaction step (described by  $k_{\text{inact}}$ ), both requiring different computational techniques to evaluate. Therefore, the developed protocol contains thermodynamic integration estimating the non-covalent binding affinities, while QM/MM MD umbrella sampling simulations are used to construct the potential of mean force curve of the covalent reaction to evaluate reaction barrier heights determining the covalent reactivity.

First, we explored the relation between calculated intrinsic reactivities of small molecule electrophiles and experimental reactivity descriptors such as GSH half-life. These QM based studies gave us a perspective on the possible reaction mechanisms later adapted into the enzyme environment of covalently targetable proteins and on the scope and limitations of computational chemistry based reactivity predictions in covalent inhibitor design.

In the subsequent work we were able to expand the covalent reaction modeling into enzyme environment, namely evaluating the covalent inhibition of MurA by three sets of molecules containing various warheads and separate them into active and inactive compounds based on calculated reaction barriers. The good reproduction of the experimental results justified the assumption that the inhibitory activity is dominated by the chemical reaction and the non-covalent recognition can be omitted for the fragment sized compounds studied. Moreover, we proposed a mechanism for Cys115 activation in MurA using molecular dynamics and bioinformatics tools. According to this mechanism a proton transfer from Cys115 to the conserved His394 residue generates the activated thiolate.

The next step ahead was to evaluate both steps of the covalent inhibition. We selected five relevant protein targets with available experimental  $K_i$  and  $k_{\text{inact}}$  values. We performed MD simulations for assessing the non-covalent binding free energies and the reaction barriers of the specific protein-ligand complexes. Thereafter, we converted these calculated energies into  $K_i$  and  $k_{\text{inact}}$ , and we found reasonable correlation between calculated and experimental data.

Finally, we applied the computational scheme to evaluate a more complex inhibition mechanism, namely the covalent inhibition of immunoproteasome by oxathiazolone derivatives. The evaluation of the covalent binding required the determination of the reaction mechanism and its rate-determining step; therefore, we applied QM/MM steered MD and umbrella sampling simulations to clarify the mechanism. Once, the rate-determining step was known, we applied the previously developed protocol for six selected iPS inhibitors. Again, we observed good correlation with the experimental values.

Though, the primary aim of the developed protocol is to aid covalent drug design through compound ranking by activity prediction, the applicability is not limited exclusively to compound prioritization. We successfully used the developed computational scheme for selectivity studies evaluating the binding free energy difference of the complex formation of a ligand with related proteins, applying a “hybrid” protein TI approach. This technique was applied for selectivity assessments of compounds inhibiting related kinases, ITK, BTK and BMX; EGFR mutants; and also iPS and cPS.

My principal results are summarized in the following thesis points.

1. I have proposed a mechanism for the reaction catalyzed by MurA. This includes the loop closure upon UNAG binding, followed by a proton transfer between Cys115 and His394 that activates the cysteine residue for the nucleophilic attack towards the natural substrate PEP or the cysteine targeting inhibitors. I found that PEP is first protonated by His 394 and Cys115 forms a covalent adduct with the protonated PEP [170].

2. I have shown that the active-inactive separation of ligands is possible based on the calculated barriers of the inhibition reaction between MurA and nine selected compounds containing oxirane,  $\alpha$ -haloketone,  $\alpha$ -haloacetamide, acrylamide and vinyl-sulfone warheads [170].

3. I have developed a complex molecular dynamics based protocol characterizing the complete covalent protein inhibition including both the non-covalent and covalent steps. The protocol was validated successfully on a set of relevant protein targets, namely on KRAS, EGFR, ITK, BTK and BMX. Selectivity assessments were also carried out explaining binding affinity differences of specific compounds towards the related kinases and EGFR mutants [139,39].

4. I have clarified the mechanism of the covalent reaction between the Thr1 residue of immunoproteasome and an oxathiazolone derivative using QM/MM molecular dynamics umbrella sampling calculations. I have determined that the carbonate pathway is favored over the carbonyl route. I showed that the nucleophilic attack of the activated Thr1 O $\gamma$  and the proton transfer between the terminal amino group of Thr1 and the negatively charged sulfur of the ligand are asynchronous events composing the rate determining step [201].

5. I have shown that the differences in the iPS inhibitory activity of six examined oxathiazolone compounds are primarily affected by the molecular recognition, and less by the chemical reaction. I interpreted the immunoproteasome versus constitutive proteasome selectivity of two compounds. I found that the Gln53Ser mutation causes different Met48 conformation in cPS compared to iPS and this has a ligand size dependent effect on the binding and reactions of inhibitors [201].

## 6.2 Összefoglalás és tézispontok

Munkám fő célkitűzése egy robusztus és hatékony számítástechnikai kémiai módszer kifejlesztése volt, mellyel lehetővé válik a kovalens fehérje inhibíció teljes energetikai vizsgálata, beleértve a  $K_i$  konstanssal jellemezhető nem-kovalens molekuláris felismerési és a  $k_{inact}$  sebességi állandóval jellemezhető kovalens reakció lépését is. A használt protokoll az eltérő számítási szinteket megkövetelő lépéseknek megfelelően két módszer összekapcsolásából áll. A nem-kovalens lépés jellemzésére molekulamechanika alapú termodinamikai integrálást használ. A kovalens reakció gátmagasságának jellemzésére umbrella samplinget alkalmazó QM/MM molekuladinamikai szimulációkból számítja a reakció szabadentalpia görbét.

Elsőként kvantummechanikai számolásokkal vizsgáltuk az összefüggést az intrinzik (intrinsic) ligandum reaktivitás és a kísérletileg meghatározott reaktivitást jellemző leírók (pl.: GSH fél-élet idő) között. Az így elvégzett QM számolások rávilágítottak a kísérleti eredmények reprodukálásának lehetőségére tisztán *in silico* módszerekkel, illetve betekintést engedtek a későbbiekben az enzim környezetben is modellezett reakciók mechanizmusába.

Következő lépésként kiterjesztettük a kovalens reakciók modellezését fehérje környezetre is. Felállítottunk egy új lehetséges katalitikus és gátlási mechanizmust a MurA enzimre, melyben a Cys115 tioljának aktiválása a His394 általi deprotonálódással valósul meg, így elősegítve a kovalens reakciót a Cys115 és a természetes szubsztrát foszfoenol-piruvát (PEP) vagy ciszteint támadó inhibitorok között. Ezután az enzim kovalens gátlását is megvizsgáltuk kilenc potenciális inhibitor által, melyek között oxiránok,  $\alpha$ -haloketonok,  $\alpha$ -haloacetamidok, akrilamidok és vinil-szulfonok is megtalálhatók. A reakciók számolt gátja alapján sikeresen elkülönítettük az aktív vegyületeket az inaktívaktól, feltételezve, hogy a molekuláris felismerési lépés csak elhanyagolható befolyással bír a teljes kötődésre.

A következő lépésben figyelembe véve mindkét kötődési lépést, kiválasztottunk öt releváns fehérjét a megfelelő kísérleti  $K_i$  és  $k_{inact}$  adatokkal, majd MD szimulációk futtatásával meghatároztuk a kötődési szabadentalpiáját és a kovalens reakció gátját a megfelelő fehérje-ligandum pároknak. A kapott eredményeket  $K_i$  és  $k_{inact}$  adatokká alakítva összehasonlítottuk a számolt értékeket és a kísérleti adatokat. Az észlelt egyezés igazolta a módszer hatékonyságát kovalens inhibitorok jellemzésére.

Végezetül a kidolgozott sémát alkalmaztuk összetett mechanizmussal kötődő inhibitorok jellemzésére is, nevezetesen az immunoproteaszóma oxatiazolon származékokkal való gátlására. Ehhez szükség volt a reakció mechanizmusának tisztázására és a sebesség meghatározó lépés azonosítására. QM/MM MD futtatások után sikeresen azonosítottuk a karbonát utat, mint a karbontioát útnál valószínűbb mechanizmust. Ezen kívül az első aszinkron lépést — melyben a Thr1 alkoholátja nukleofil támadást intéz a ligandum ellen és megtörténik a proton átmenet a terminális aminocsoport és a ligandum negatív kénatomja között — sebességmeghatározó lépésként állapítottuk meg. Ezután az előzetesen leírtakat követve meghatároztuk hat iPS inhibitor  $K_i$  és  $k_{inact}$  értékeit, majd összevetettük a kísérleti eredményekkel. Ebben az esetben is jó korrelációt találtunk a számolt és mért értékek között.

Bár az elsődleges célja a kidolgozott módszert a vizsgált vegyületek kötődésének energetikai leírása és az ez alapján történő vegyület rangsorolás, munkám során bemutattam a protokoll szelektivitási vizsgálatokban és mechanizmuskutatóban való alkalmazhatóságát is. A módszer hatékony támogatást nyújthat kovalens inhibitorok azonosítását célzó gyógyszerkutatói munkában, főként szerkezet optimalizációs feladatokban.

Legfontosabb eredményeim tézispontok szerinti összefoglalása a következő.

1. Mechanizmust javasoltam a MurA enzim által katalizált reakcióra, mely tartalmaz egy a kofaktor UDP-N-acetylglucosamine (UNAG) általi hurok záródást, majd ezt követően egy Cys115 His394 közötti protonátmenetet, ezzel aktiválva a Cys115 oldallánc tiolját a természetes szubsztráttal (foszfoenol-piruvát, PEP) vagy ciszteint célzó kovalens inhibitorokkal szembeni nukleofil támadáshoz. A PEP kötődési mechanizmusára azt találtam, hogy először protonálódik a His394 által, majd adduktot képez a Cys115-tel [170].

2. Megmutattam, hogy a MurA Cys115 oldallánca és kilenc oxirán,  $\alpha$ -haloketon,  $\alpha$ -haloacetamid, akrilamid és vinil-szulfon származék között végbemenő kémiai reakciók számított gátmagasságai alapján a ligandumok aktív-inaktív besorolása elvégezhető [170].

3. Kifejlesztettem egy komplex molekuladinamika alapú szabadentalpiát számító protokollt a teljes kovalens inhibíció jellemzésére, beleértve a nem-kovalens és kovalens lépést is. A módszert sikeresen validáltam több releváns fehérjén, név szerint KRAS-on, EGFR-en, ITK-n, BTK-n és BMX-en. A módszer segítségével szelektivitási becsléseket is sikeresen tettem, néhány kiválasztott vegyület rokon kinázokkal, illetve EGFR mutánsokkal szembeni kötődési szabadentalpia különbségeinek számításával [139,39].

4. Tisztáztam az immunoproteaszóma Thr1 oldallánca és oxatiazolon származékok között végbemenő kémiai reakció mechanizmusát, QM/MM molekuladinamikák használatával és az umbrella sampling mintavételezési technika segítségével. A mechanizmusok vizsgálata során a karbonát utat a karbontioát útnál energetikailag kedvezőbbnek találtam. Ezen kívül megmutattam, hogy az aktivált Thr1 O $\gamma$  általi nukleofil támadás, valamint a terminális aminocsoport és az oxatiazolon gyűrűből származó negatív töltésű kénatom közötti proton transzfer aszinkron kötődési események, és együttesen ezek alkotják a teljes reakciósor sebesség meghatározó lépését [201].

5. Bemutattam hat oxatiazolon származék esetében, hogy az immunoproteaszómával szemben mért gátlási aktivitásukban tapasztalt különbségeket elsődlegesen a molekuláris felismerési lépés során fellépő eltérések okozzák, a kémiai reakciótól való függésük kisebb mértékű. Értelmeztem két oxatiazolon származék immuno- és konstitutív proteaszómával szembeni eltérő szelektivitását, mely során egy Gln53Ser mutáció által okozott Met48 konformációváltást találtam, mint a szelektivitási különbség fő oka. A cPS-ban és az iPS-ban eltérő konformáció a ligandum méretétől függő hatást gyakorol az inhibitorok kötődésére és kémiai reakciójára [201].

### 6.3 List of the publications discussed in the thesis

1. L. M. Mihalovits, G. G. Ferenczy, G. M. Keserű; Catalytic Mechanism and Covalent Inhibition of UDP-N-Acetylglucosamine Enolpyruvyl Transferase (MurA): Implications to the Design of Novel Antibacterials *Journal of Chemical Information and Modeling*, 2019, **59**, 5161–5173. DOI: 10.1021/acs.jcim.9b00691. IF(2019): 4.549 FI: 4 **Ref.** [170]
2. L. M. Mihalovits, G. G. Ferenczy, G. M. Keserű; Affinity and Selectivity Assessment of Covalent Inhibitors by Free Energy Calculations *Journal of Chemical Information and Modeling*, 2020, **60**, 6579–6594. DOI: 10.1021/acs.jcim.0c00834. IF(2020): 4.956 FI: 3 **Ref.** [139]
3. L. M. Mihalovits, G. G. Ferenczy, G. M. Keserű; Mechanistic and thermodynamic characterization of oxathiazolones as potent and selective covalent immunoproteasome inhibitors *Computational and Structural Biotechnology Journal*, 2021, **19**, 4486-4496. DOI: 10.1016/j.csbj.2021.08.008. IF: 7.271 FI: 0 **Ref.** [201]
4. L. M. Mihalovits, G. G. Ferenczy, G. M. Keserű; The role of quantum chemistry in covalent inhibitor design *International Journal of Quantum Chemistry*, 2021 (early view). DOI: 10.1002/qua.26768. IF(2020): 2.444 FI: 0 **Ref.** [39]

## REFERENCES

- [1] Talele T, Khedkar S, Rigby A. Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Curr Top Med Chem* 2010;10:127–41. <https://doi.org/10.2174/156802610790232251>.
- [2] Sabe VT, Ntombela T, Jhamba LA, Maguire GEM, Govender T, Naicker T, et al. Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *Eur J Med Chem* 2021;224:113705. <https://doi.org/10.1016/j.ejmech.2021.113705>.
- [3] Johnson HWB, Lowe E, Anderl JL, Fan A, Muchamuel T, Bowers S, et al. Required Immunoproteasome Subunit Inhibition Profile for Anti-Inflammatory Efficacy and Clinical Candidate KZR-616 ((2 S,3 R)- N-((S)-3-(Cyclopent-1-en-1-yl)-1-((R)-2-methyloxiran-2-yl)-1-oxopropan-2-yl)-3-hydroxy-3-(4-methoxyphenyl)-2-((S)-2-(2-morpholin. *J Med Chem* 2018;61:11127–43. <https://doi.org/10.1021/acs.jmedchem.8b01201>.
- [4] Huber EM, Groll M. Inhibitors for the immuno- and constitutive proteasome: Current and future trends in drug development. *Angew Chemie - Int Ed* 2012;51:8708–20. <https://doi.org/10.1002/anie.201201616>.
- [5] Oldfield V, Keating GM, Perry CM. Rasagiline. *Drugs* 2007;67:1725–47. <https://doi.org/10.2165/00003495-200767120-00006>.
- [6] Savi P, Nurden P, Nurden AT, Levy-Toledano S, Herbert J-M. Clopidogrel: a review of its mechanism of action. *Platelets* 1998;9:251–5. <https://doi.org/10.1080/09537109876799>.
- [7] Adhikari AA, Seegar TCM, Ficarro SB, McCurry MD, Ramachandran D, Yao L, et al. Development of a covalent inhibitor of gut bacterial bile salt hydrolases. *Nat Chem Biol* 2020;16:318–26. <https://doi.org/10.1038/s41589-020-0467-3>.
- [8] Olesen SH, Ingles DJ, Yang Y, Schönbrunn E. Differential antibacterial properties of the MurA inhibitors terreic acid and fosfomycin. *J Basic Microbiol* 2014;54:322–6. <https://doi.org/10.1002/jobm.201200617>.
- [9] Planken S, Behenna DC, Nair SK, Johnson TO, Nagata A, Almaden C, et al. Discovery of N-((3R,4R)-4-Fluoro-1-(6-((3-methoxy-1-methyl-1H-pyrazol-4-yl)amino)-9-methyl-9H-purin-2-yl)pyrrolidine-3-yl)acrylamide (PF-06747775) through Structure-Based Drug Design: A High Affinity Irreversible Inhibitor Targeting Oncogenic EGFR Mutants. *J Med Chem* 2017;60:3002–19. <https://doi.org/10.1021/acs.jmedchem.6b01894>.
- [10] Lau SCM, Batra U, Mok TSK, Loong HH. Dacomitinib in the Management of Advanced Non-Small-Cell Lung Cancer. *Drugs* 2019;79:823–31. <https://doi.org/10.1007/s40265-019-01115-y>.
- [11] Sutanto F, Konstantinidou M, Dömling A. Covalent inhibitors: A rational approach to drug discovery. *RSC Med Chem* 2020;11:876–84. <https://doi.org/10.1039/d0md00154f>.
- [12] Ábrányi-Balogh P, Petri L, Imre T, Szijj P, Scarpino A, Hrast M, et al. A road map for prioritizing warheads for cysteine targeting covalent inhibitors. *Eur J Med Chem* 2018;160. <https://doi.org/10.1016/j.ejmech.2018.10.010>.
- [13] Vane J., Botting R. The mechanism of action of aspirin. *Thromb Res* 2003;110:255–8. [https://doi.org/10.1016/S0049-3848\(03\)00379-7](https://doi.org/10.1016/S0049-3848(03)00379-7).
- [14] Shannon DA, Weerapana E. Covalent protein modification: the current landscape of residue-specific electrophiles. *Curr Opin Chem Biol* 2015;24:18–26. <https://doi.org/10.1016/j.cbpa.2014.10.021>.
- [15] Baillie TA. Targeted Covalent Inhibitors for Drug Design. *Angew Chemie Int Ed*

- 2016;55:13408–21. <https://doi.org/10.1002/anie.201601091>.
- [16] Miseta A, Csutora P. Relationship Between the Occurrence of Cysteine in Proteins and the Complexity of Organisms. *Mol Biol Evol* 2000;17:1232–9. <https://doi.org/10.1093/oxfordjournals.molbev.a026406>.
- [17] Hallenbeck K, Turner D, Renslo A, Arkin M. Targeting Non-Catalytic Cysteine Residues Through Structure-Guided Drug Discovery. *Curr Top Med Chem* 2016;17:4–15. <https://doi.org/10.2174/1568026616666160719163839>.
- [18] Zhao Z, Bourne PE. Progress with covalent small-molecule kinase inhibitors. *Drug Discov Today* 2018;23:727–35. <https://doi.org/10.1016/j.drudis.2018.01.035>.
- [19] Uetrecht J, Naisbitt DJ. Idiosyncratic Adverse Drug Reactions: Current Concepts. *Pharmacol Rev* 2013;65:779–808. <https://doi.org/10.1124/pr.113.007450>.
- [20] Johnson DS, Weerapana E, Cravatt BF. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med Chem* 2010;2:949–64. <https://doi.org/10.4155/fmc.10.21>.
- [21] Flanagan ME, Abramite JA, Anderson DP, Aulabaugh A, Dahal UP, Gilbert AM, et al. Chemical and computational methods for the characterization of covalent reactive groups for the prospective design of irreversible inhibitors. *J Med Chem* 2014;57:10072–9. <https://doi.org/10.1021/jm501412a>.
- [22] Krippendorff B-F, Neuhaus R, Lienau P, Reichel A, Huisinga W. Mechanism-Based Inhibition: Deriving  $k_{inact}$  and  $k_{inact}$  Directly from Time-Dependent  $IC_{50}$  Values. *J Biomol Screen* 2009;14:913–23. <https://doi.org/10.1177/1087057109336751>.
- [23] Miyahisa I, Sameshima T, Hixon MS. Rapid Determination of the Specificity Constant of Irreversible Inhibitors ( $k_{inact} / K_I$ ) by Means of an Endpoint Competition Assay. *Angew Chemie Int Ed* 2015;54:14099–102. <https://doi.org/10.1002/anie.201505800>.
- [24] Bilotta M, Tradigo G, Veltri P. Bioinformatics Data Models, Representation and Storage. *Encycl. Bioinforma. Comput. Biol.*, Elsevier; 2019, p. 110–6. <https://doi.org/10.1016/B978-0-12-809633-8.20410-X>.
- [25] Batool M, Ahmad B, Choi S. A Structure-Based Drug Discovery Paradigm. *Int J Mol Sci* 2019;20:2783. <https://doi.org/10.3390/ijms20112783>.
- [26] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9. <https://doi.org/10.1038/s41586-021-03819-2>.
- [27] Zoete V, Grosdidier A, Michielin O. Docking, virtual high throughput screening and in silico fragment-based drug design. *J Cell Mol Med* 2009;13:238–48. <https://doi.org/10.1111/j.1582-4934.2008.00665.x>.
- [28] Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J Med Chem* 2006;49:534–53. <https://doi.org/10.1021/jm050540c>.
- [29] Scarpino A, Petri L, Knez D, Imre T, Ábrányi-Balogh P, Ferenczy GG, et al. WIDOCK: a reactive docking protocol for virtual screening of covalent inhibitors. *J Comput Aided Mol Des* 2021;35:223–44. <https://doi.org/10.1007/s10822-020-00371-5>.
- [30] Zhu K, Borrelli KW, Greenwood JR, Day T, Abel R, Farid RS, et al. Docking Covalent Inhibitors: A Parameter Free Approach To Pose Prediction and Scoring. *J Chem Inf Model* 2014;54:1932–40. <https://doi.org/10.1021/ci500118s>.
- [31] De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem* 2016;59:4035–61. <https://doi.org/10.1021/acs.jmedchem.5b01684>.
- [32] Acharya C, Coop A, E. Polli J, D. MacKerell A. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. *Curr Comput Aided-Drug Des* 2011;7:10–22.

- <https://doi.org/10.2174/157340911793743547>.
- [33] Neves BJ, Braga RC, Melo-Filho CC, Moreira-Filho JT, Muratov EN, Andrade CH. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front Pharmacol* 2018;9. <https://doi.org/10.3389/fphar.2018.01275>.
- [34] Bajusz D, Rácz A, Héberger K. Chemical Data Formats, Fingerprints, and Other Molecular Descriptions for Database Analysis and Searching. *Compr. Med. Chem. III*, vol. 3–8, 2017, p. 329–78. <https://doi.org/10.1016/B978-0-12-409547-2.12345-5>.
- [35] Vázquez J, López M, Gibert E, Herrero E, Luque FJ. Merging Ligand-Based and Structure-Based Methods in Drug Discovery: An Overview of Combined Virtual Screening Approaches. *Molecules* 2020;25:4723. <https://doi.org/10.3390/molecules25204723>.
- [36] Lonsdale R, Burgess J, Colclough N, Davies NL, Lenz EM, Orton AL, et al. Expanding the Armory: Predicting and Tuning Covalent Warhead Reactivity. *J Chem Inf Model* 2017;57:3124–37. <https://doi.org/10.1021/acs.jcim.7b00553>.
- [37] van Duin ACT, Dasgupta S, Lorant F, Goddard WA. ReaxFF: A Reactive Force Field for Hydrocarbons. *J Phys Chem A* 2001;105:9396–409. <https://doi.org/10.1021/jp004368u>.
- [38] Cee VJ, Volak LP, Chen Y, Bartberger MD, Tegley C, Arvedson T, et al. Systematic study of the glutathione (GSH) reactivity of N-arylacrylamides: 1. Effects of aryl substitution. *J Med Chem* 2015;58:9171–8. <https://doi.org/10.1021/acs.jmedchem.5b01018>.
- [39] Mihalovits LM, Ferenczy GG, Keserű GM. The role of quantum chemistry in covalent inhibitor design. *Int J Quantum Chem* 2021:qua.26768. <https://doi.org/10.1002/qua.26768>.
- [40] Barreteau H, Kovač A, Boniface A, Sova M, Gobec S, Blanot D. Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev* 2008;32:168–207. <https://doi.org/10.1111/j.1574-6976.2008.00104.x>.
- [41] Rogers HJ, Perkins HR, Ward JB. *Microbial Cell Walls and Membranes*. Dordrecht: Springer Netherlands; 1980. <https://doi.org/10.1007/978-94-011-6014-8>.
- [42] El Zoeiby A, Sanschagrin F, Levesque RC. Structure and function of the Mur enzymes: Development of novel inhibitors. *Mol Microbiol* 2003;47:1–12. <https://doi.org/10.1046/j.1365-2958.2003.03289.x>.
- [43] Skarzynski T, Kim DH, Lees WJ, Walsh CT, Duncan K. Stereochemical course of enzymatic enolpyruvyl transfer and catalytic conformation of the active site revealed by the crystal structure of the fluorinated analogue of the reaction tetrahedral intermediate bound to the active site of the C115A mutant of Mu. *Biochemistry* 1998;37:2572–7. <https://doi.org/10.1021/bi9722608>.
- [44] Eschenburg S, Kabsch W, Healy ML, Schonbrunn E. A new view of the mechanisms of UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate-3-phosphate synthase (AroA) derived from X-ray structures of their tetrahedral reaction intermediate states. *J Biol Chem* 2003;278:49215–22. <https://doi.org/10.1074/jbc.M309741200>.
- [45] Skarzynski T, Mistry A, Wonacott A, Hutchinson SE, Kelly VA, Duncan K. Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomicin. *Structure* 1996;4:1465–74. [https://doi.org/10.1016/S0969-2126\(96\)00153-0](https://doi.org/10.1016/S0969-2126(96)00153-0).
- [46] Schönbrunn E, Sack S, Eschenburg S, Perrakis A, Krekel F, Amrhein N, et al. Crystal structure of UDP-N-acetylglucosamine enolpyruvyltransferase, the target of the antibiotic fosfomicin. *Structure* 1996;4:1065–75. <https://doi.org/10.1016/S0969->

- 2126(96)00113-X.
- [47] Schönbrunn E, Eschenburg S, Krekel F, Luger K, Amrhein N. Role of the loop containing residue 115 in the induced-fit mechanism of the bacterial cell wall biosynthetic enzyme MurA. *Biochemistry* 2000;39:2164–73. <https://doi.org/10.1021/bi991091j>.
- [48] Gautam A, Rishi P, Tewari R. UDP-N-acetylglucosamine enolpyruvyl transferase as a potential target for antibacterial chemotherapy: Recent developments. *Appl Microbiol Biotechnol* 2011;92:211–25. <https://doi.org/10.1007/s00253-011-3512-z>.
- [49] Eschenburg S, Schönbrunn E. Comparative X-ray analysis of the un-liganded fosfomycin-target MurA. *Proteins Struct Funct Genet* 2000;40:290–8. [https://doi.org/10.1002/\(SICI\)1097-0134\(20000801\)40:2<290::AID-PROT90>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-0134(20000801)40:2<290::AID-PROT90>3.0.CO;2-0).
- [50] Yoon HJ, Sang JL, Mikami B, Park HJ, Yoo J, Se WS. Crystal structure of UDP-N-acetylglucosamine enolpyruvyl transferase from *Haemophilus influenzae* in complex with UDP-N-acetylglucosamine and fosfomycin. *Proteins Struct Funct Genet* 2008;71:1032–7. <https://doi.org/10.1002/prot.21959>.
- [51] Zhu JY, Yang Y, Han H, Betzi S, Olesen SH, Marsilio F, et al. Functional consequence of covalent reaction of phosphoenolpyruvate with UDP-N-acetylglucosamine 1-carboxyvinyltransferase (MurA). *J Biol Chem* 2012;287:12657–67. <https://doi.org/10.1074/jbc.M112.342725>.
- [52] Marquardt JL, Brown ED, Walsh CT, Lane WS, Haley TM, Ichikawa Y, et al. Kinetics, Stoichiometry, and Identification of the Reactive Thiolate in the Inactivation of UDP-GlcNAc Enolpyruvyl Transferase by the Antibiotic Fosfomycin. *Biochemistry* 1994;33:10646–51. <https://doi.org/10.1021/bi00201a011>.
- [53] Kahan FM, Kahan JS, Cassidy PJ, Kropp H. The mechanism of action of fosfomycin (phosphonomycin). *Ann N Y Acad Sci* 1974;235:364–86. <https://doi.org/10.1111/j.1749-6632.1974.tb43277.x>.
- [54] Gardiner BJ, Stewardson AJ, Abbott IJ, Peleg AY. Nitrofurantoin and fosfomycin for resistant urinary tract infections: old drugs for emerging problems. *Aust Prescr* 2019;42:14. <https://doi.org/10.18773/austprescr.2019.002>.
- [55] Bachelier A, Mayer R, Klein CD. Sesquiterpene lactones are potent and irreversible inhibitors of the antibacterial target enzyme MurA. *Bioorg Med Chem Lett* 2006;16:5605–9. <https://doi.org/10.1016/j.bmcl.2006.08.021>.
- [56] Jin BS, Han SG, Lee WK, Ryoo SW, Lee SJ, Suh SW, et al. Inhibitory mechanism of novel inhibitors of UDP-N-acetylglucosamine enolpyruvyl transferase from *Haemophilus influenzae*. *J Microbiol Biotechnol* 2009;19:1582–9. <https://doi.org/10.4014/jmb.0905.05036>.
- [57] Shigetomi K, Shoji K, Mitsuhashi S, Ubukata M. The antibacterial properties of 6-tuliposide B. Synthesis of 6-tuliposide B analogues and structure–activity relationship. *Phytochemistry* 2010;71:312–24. <https://doi.org/10.1016/j.phytochem.2009.10.008>.
- [58] Schönbrunn E, Han H, Yang Y, Olesen SH, Becker A, Betzi S. The fungal product terreic acid is a covalent inhibitor of the bacterial cell wall biosynthetic enzyme UDP-N-acetylglucosamine 1-carboxyvinyltransferase (MurA). *Biochemistry* 2010;49:4276–82. <https://doi.org/10.1021/bi100365b>.
- [59] Ostrem JML, Shokat KM. Direct small-molecule inhibitors of KRAS: From structural insights to mechanism-based design. *Nat Rev Drug Discov* 2016;15:771–85. <https://doi.org/10.1038/nrd.2016.139>.
- [60] Singh D, Kumar Attri B, Kaur Gill R, Bariwal J. Review on EGFR Inhibitors: Critical Updates. *Mini-Reviews Med Chem* 2016;16:1134–66.

- <https://doi.org/10.2174/1389557516666160321114917>.
- [61] Milburn M, Tong L, DeVos A, Brunger A, Yamaizumi Z, Nishimura S, et al. Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* (80- ) 1990;247:939–45. <https://doi.org/10.1126/science.2406906>.
- [62] Prior IA, Lewis PD, Mattos C. A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Res* 2012;72:2457–67. <https://doi.org/10.1158/0008-5472.CAN-11-2612>.
- [63] Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* 2013;503:548–51. <https://doi.org/10.1038/nature12796>.
- [64] Xiong Y, Lu J, Hunter J, Li L, Scott D, Choi HG, et al. Covalent guanosine mimetic inhibitors of G12C KRAS. *ACS Med Chem Lett* 2017;8:61–6. <https://doi.org/10.1021/acsmedchemlett.6b00373>.
- [65] Canon J, Rex K, Saiki AY, Mohr C, Cooke K, Bagal D, et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* 2019;575:217–23. <https://doi.org/10.1038/s41586-019-1694-1>.
- [66] Briere DM, Li S, Calinisan A, Sudhakar N, Aranda R, Hargis L, et al. The KRAS G12C Inhibitor MRTX849 Reconditions the Tumor Immune Microenvironment and Sensitizes Tumors to Checkpoint Inhibitor Therapy. *Mol Cancer Ther* 2021;20:975–85. <https://doi.org/10.1158/1535-7163.MCT-20-0462>.
- [67] Roskoski R. The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacol Res* 2014;79:34–74. <https://doi.org/10.1016/j.phrs.2013.11.002>.
- [68] Arteaga CL. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. *J Clin Oncol* 2001;19:32S-40S.
- [69] Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR, et al. Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene* 2006;366:2–16. <https://doi.org/10.1016/j.gene.2005.10.018>.
- [70] Sharma S V., Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* 2007;7:169–81. <https://doi.org/10.1038/nrc2088>.
- [71] Barker AJ, Gibson KH, Grundy W, Godfrey AA, Barlow JJ, Healy MP, et al. Studies leading to the identification of ZD1839 (iressa<sup>TM</sup>): an orally active, selective epidermal growth factor receptor tyrosine kinase inhibitor targeted to the treatment of cancer. *Bioorg Med Chem Lett* 2001;11:1911–4. [https://doi.org/10.1016/S0960-894X\(01\)00344-4](https://doi.org/10.1016/S0960-894X(01)00344-4).
- [72] Moyer JD, Barbacci EG, Iwata KK, Arnold L, Boman B, Cunningham A, et al. Induction of apoptosis and cell cycle arrest by CP-358,774, an inhibitor of epidermal growth factor receptor tyrosine kinase. *Cancer Res* 1997;57:4838–48.
- [73] Carmi C, Lodola A, Rivara S, Vacondio F, Cavazzoni A, R. Alfieri R, et al. Epidermal Growth Factor Receptor Irreversible Inhibitors: Chemical Exploration of the Cysteine-Trap Portion. *Mini-Reviews Med Chem* 2011;11:1019–30. <https://doi.org/10.2174/138955711797247725>.
- [74] Ferrarotto R, Gold KA. Afatinib in the treatment of head and neck squamous cell carcinoma. *Expert Opin Investig Drugs* 2014;23:135–43. <https://doi.org/10.1517/13543784.2014.858696>.
- [75] Cross DAE, Ashton SE, Ghiorghiu S, Eberlein C, Nebhan CA, Spitzler PJ, et al. AZD9291, an Irreversible EGFR TKI, Overcomes T790M-Mediated Resistance to EGFR Inhibitors in Lung Cancer. *Cancer Discov* 2014;4:1046–61. <https://doi.org/10.1158/2159-8290.CD-14-0337>.
- [76] Zhou W, Ercan D, Chen L, Yun C-H, Li D, Capelletti M, et al. Novel mutant-selective

- EGFR kinase inhibitors against EGFR T790M. *Nature* 2009;462:1070–4. <https://doi.org/10.1038/nature08622>.
- [77] Mano H. Tec family of protein-tyrosine kinases: an overview of their structure and function. *Cytokine Growth Factor Rev* 1999;10:267–80. [https://doi.org/10.1016/S1359-6101\(99\)00019-2](https://doi.org/10.1016/S1359-6101(99)00019-2).
- [78] Mamontov P, Eberwine RA, Perrigoue J, Das A, Friedman JR, Mora JR. A negative role for the interleukin-2-inducible T-cell kinase (ITK) in human Foxp3 + T REG differentiation. *PLoS One* 2019;14:e0215963. <https://doi.org/10.1371/journal.pone.0215963>.
- [79] Rip J, Van Der Ploeg EK, Hendriks RW, Corneth OBJ. The role of Bruton's tyrosine kinase in immune cell signaling and systemic autoimmunity. *Crit Rev Immunol* 2018;38:17–62. <https://doi.org/10.1615/CritRevImmunol.2018025184>.
- [80] Qiu L, Wang F, Liu S, Chen XL. Current understanding of tyrosine kinase BMX in inflammation and its inhibitors. *Burn Trauma* 2014;2:121–4. <https://doi.org/10.4103/2321-3868.135483>.
- [81] Yang W-C, Collette Y, Nunés JA, Olive D. Tec Kinases, from Structure to Function. *Sci Signal* 2000;2000:tw5–tw5. <https://doi.org/10.1126/stke.2000.29.tw5>.
- [82] Zapf CW, Gerstenberger BS, Xing L, Limburg DC, Anderson DR, Caspers N, et al. Covalent inhibitors of interleukin-2 inducible T cell kinase (ItK) with nanomolar potency in a whole-blood assay. *J Med Chem* 2012;55:10047–63. <https://doi.org/10.1021/jm301190s>.
- [83] Thorarensen A, Dowty ME, Banker ME, Juba B, Jussif J, Lin T, et al. Design of a Janus Kinase 3 (JAK3) Specific Inhibitor 1-((2 S ,5 R )-5-((7 H -Pyrrolo[2,3- d ]pyrimidin-4-yl)amino)-2-methylpiperidin-1-yl)prop-2-en-1-one (PF-06651600) Allowing for the Interrogation of JAK3 Signaling in Humans. *J Med Chem* 2017;60:1971–93. <https://doi.org/10.1021/acs.jmedchem.6b01694>.
- [84] Abdeldayem A, Raouf YS, Constantinescu SN, Moriggl R, Gunning PT. Advances in covalent kinase inhibitors. *Chem Soc Rev* 2020;49:2617–87. <https://doi.org/10.1039/C9CS00720B>.
- [85] Baumeister W, Walz J, Zühl F, Seemüller E. The Proteasome: Paradigm of a Self-Compartmentalizing Protease. *Cell* 1998;92:367–80. [https://doi.org/10.1016/S0092-8674\(00\)80929-0](https://doi.org/10.1016/S0092-8674(00)80929-0).
- [86] Borissenko L, Groll M. 20S Proteasome and Its Inhibitors: Crystallographic Knowledge for Drug Development. *Chem Rev* 2007;107:687–717. <https://doi.org/10.1021/cr0502504>.
- [87] Bedford L, Lowe J, Dick LR, Mayer RJ, Brownell JE. Ubiquitin-like protein conjugation and the ubiquitinating-proteasome system as drug targets. *Nat Rev Drug Discov* 2011;10:29–46. <https://doi.org/10.1038/nrd3321>.
- [88] Nalepa G, Rolfe M, Harper JW. Drug discovery in the ubiquitin–proteasome system. *Nat Rev Drug Discov* 2006;5:596–613. <https://doi.org/10.1038/nrd2056>.
- [89] Groettrup M, Kraft R, Kostka S, Standera S, Stohwasser R, Kloetzel P-M. A third interferon- $\gamma$ -induced subunit exchange in the 20S proteasome. *Eur J Immunol* 1996;26:863–9. <https://doi.org/10.1002/eji.1830260421>.
- [90] Griffin TA, Nandi D, Cruz M, Fehling HJ, Van Kaer L, Monaco JJ, et al. Immunoproteasome assembly: Cooperative incorporation of interferon  $\gamma$  (IFN- $\gamma$ )-inducible subunits. *J Exp Med* 1998;187:97–104. <https://doi.org/10.1084/jem.187.1.97>.
- [91] Kloetzel P-M, Ossendorp F. Proteasome and peptidase function in MHC-class-I-mediated antigen presentation. *Curr Opin Immunol* 2004;16:76–81. <https://doi.org/10.1016/j.coi.2003.11.004>.
- [92] Sosič I, Gobec M, Brus B, Knez D, Živec M, Konc J, et al. Nonpeptidic Selective

- Inhibitors of the Chymotrypsin-Like ( $\beta 5$  i) Subunit of the Immunoproteasome. *Angew Chemie - Int Ed* 2016;55:5745–8. <https://doi.org/10.1002/anie.201600190>.
- [93] Kalim KW, Basler M, Kirk CJ, Groettrup M. Immunoproteasome Subunit LMP7 Deficiency and Inhibition Suppresses Th1 and Th17 but Enhances Regulatory T Cell Differentiation. *J Immunol* 2012;189:4182–93. <https://doi.org/10.4049/jimmunol.1201183>.
- [94] Ichikawa HT, Conley T, Muchamuel T, Jiang J, Lee S, Owen T, et al. Beneficial effect of novel proteasome inhibitors in murine lupus via dual inhibition of type I interferon and autoantibody-secreting cells. *Arthritis Rheum* 2012;64:493–503. <https://doi.org/10.1002/art.33333>.
- [95] Muchamuel T, Basler M, Aujay MA, Suzuki E, Kalim KW, Lauer C, et al. A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis. *Nat Med* 2009;15:781–7. <https://doi.org/10.1038/nm.1978>.
- [96] Lin G, Li D, De Carvalho LPS, Deng H, Tao H, Vogt G, et al. Inhibitors selective for mycobacterial versus human proteasomes. *Nature* 2009;461:621–6. <https://doi.org/10.1038/nature08357>.
- [97] Fan H, Angelo NG, Warren JD, Nathan CF, Lin G. Oxathiazolones selectively inhibit the human immunoproteasome over the constitutive proteasome. *ACS Med Chem Lett* 2014;5:405–10. <https://doi.org/10.1021/ml400531d>.
- [98] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* 2015;11:3696–713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- [99] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 2009:NA-NA. <https://doi.org/10.1002/jcc.21367>.
- [100] Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J Chem Theory Comput* 2016;12:281–96. <https://doi.org/10.1021/acs.jctc.5b00864>.
- [101] Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* 2004;25:1656–76. <https://doi.org/10.1002/jcc.20090>.
- [102] David Sherrill C, Schaefer HF. The Configuration Interaction Method: Advances in Highly Correlated Approaches, 1999, p. 143–269. [https://doi.org/10.1016/S0065-3276\(08\)60532-8](https://doi.org/10.1016/S0065-3276(08)60532-8).
- [103] Wilson S. The Perturbation Theory of Electron Correlation. III. Relativistic Many-Body Perturbation Theory, 1992, p. 253–71. [https://doi.org/10.1007/978-1-4615-7419-4\\_10](https://doi.org/10.1007/978-1-4615-7419-4_10).
- [104] Bartlett RJ, Musiał M. Coupled-cluster theory in quantum chemistry. *Rev Mod Phys* 2007;79:291–352. <https://doi.org/10.1103/RevModPhys.79.291>.
- [105] Hohenberg P, Kohn W. Inhomogeneous Electron Gas. *Phys Rev* 1964;136:B864–71. <https://doi.org/10.1103/PhysRev.136.B864>.
- [106] Kohn W, Sham LJ. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys Rev* 1965;140:A1133–8. <https://doi.org/10.1103/PhysRev.140.A1133>.
- [107] Vosko SH, Wilk L, Nusair M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can J Phys* 1980;58:1200–11. <https://doi.org/10.1139/p80-159>.
- [108] Burke K, Perdew JP, Wang Y. Derivation of a Generalized Gradient Approximation: The PW91 Density Functional. *Electron. Density Funct. Theory*, Boston, MA: Springer

- US; 1998, p. 81–111. [https://doi.org/10.1007/978-1-4899-0316-7\\_7](https://doi.org/10.1007/978-1-4899-0316-7_7).
- [109] Perdew JP. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys Rev B* 1986;33:8822–4. <https://doi.org/10.1103/PhysRevB.33.8822>.
- [110] Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A* 1988;38:3098–100. <https://doi.org/10.1103/PhysRevA.38.3098>.
- [111] Gill PMW, Johnson BG, Pople JA, Frisch MJ. The performance of the Becke—Lee—Yang—Parr (B—LYP) density functional theory with various basis sets. *Chem Phys Lett* 1992;197:499–505. [https://doi.org/10.1016/0009-2614\(92\)85807-M](https://doi.org/10.1016/0009-2614(92)85807-M).
- [112] Mardirossian N, Head-Gordon M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol Phys* 2017;115:2315–72. <https://doi.org/10.1080/00268976.2017.1333644>.
- [113] Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 2010;132:154104. <https://doi.org/10.1063/1.3382344>.
- [114] Grimme S. Density functional theory with London dispersion corrections. *WIREs Comput Mol Sci* 2011;1:211–28. <https://doi.org/10.1002/wcms.30>.
- [115] John. A. P, David L. B. Approximate molecular orbital theory. *J Chem Educ* 1971;48:A116. <https://doi.org/10.1021/ed048pA116.1>.
- [116] Dewar MJS, Thiel W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J Am Chem Soc* 1977;99:4899–907. <https://doi.org/10.1021/ja00457a004>.
- [117] Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J Am Chem Soc* 1985;107:3902–9. <https://doi.org/10.1021/ja00299a024>.
- [118] Stewart JJP. Optimization of parameters for semiempirical methods. III Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi. *J Comput Chem* 1991;12:320–41. <https://doi.org/10.1002/jcc.540120306>.
- [119] Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, et al. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys Rev B* 1998;58:7260–8. <https://doi.org/10.1103/PhysRevB.58.7260>.
- [120] Porezag D, Frauenheim T, Köhler T, Seifert G, Kaschner R. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys Rev B* 1995;51:12947–57. <https://doi.org/10.1103/PhysRevB.51.12947>.
- [121] Gaus M, Cui Q, Elstner M. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J Chem Theory Comput* 2011;7:931–48. <https://doi.org/10.1021/ct100684s>.
- [122] Lence E, Van Der Kamp MW, González-Bello C, Mulholland AJ. QM/MM simulations identify the determinants of catalytic activity differences between type II dehydroquinase enzymes. *Org Biomol Chem* 2018;16:4443–55. <https://doi.org/10.1039/c8ob00066b>.
- [123] Senn HM, Thiel W. QM/MM Methods for Biomolecular Systems. *Angew Chemie Int Ed* 2009;48:1198–229. <https://doi.org/10.1002/anie.200802019>.
- [124] Cao L, Ryde U. On the Difference Between Additive and Subtractive QM/MM Calculations. *Front Chem* 2018;6. <https://doi.org/10.3389/fchem.2018.00089>.
- [125] Field MJ, Bash PA, Karplus M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J Comput Chem* 1990;11:700–33. <https://doi.org/10.1002/jcc.540110605>.

- [126] Théry V, Rinaldi D, Rivail J -L., Maignet B, Ferenczy GG. Quantum mechanical computations on very large molecular systems: The local self-consistent field method. *J Comput Chem* 1994;15. <https://doi.org/10.1002/jcc.540150303>.
- [127] Warshel A, Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* 1976;103:227–49. [https://doi.org/10.1016/0022-2836\(76\)90311-9](https://doi.org/10.1016/0022-2836(76)90311-9).
- [128] Pu J, Gao J, Truhlar DG. Generalized Hybrid Orbital (GHO) Method for Combining Ab Initio Hartree–Fock Wave Functions with Molecular Mechanics. *J Phys Chem A* 2004;108:632–50. <https://doi.org/10.1021/jp036755k>.
- [129] Antes I, Thiel W. Adjusted connection atoms for combined quantum mechanical and molecular mechanical methods. *J Phys Chem A* 1999;103:9290–5. <https://doi.org/10.1021/jp991771w>.
- [130] Zhang Y. Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theor Chem Acc* 2006;116:43–50. <https://doi.org/10.1007/s00214-005-0008-x>.
- [131] Abrams C, Bussi G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* 2013;16:163–99. <https://doi.org/10.3390/e16010163>.
- [132] Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, et al. Steered Molecular Dynamics 2011:39–65. [https://doi.org/10.1007/978-3-642-58360-5\\_2](https://doi.org/10.1007/978-3-642-58360-5_2).
- [133] Do PC, Lee EH, Le L. Steered Molecular Dynamics Simulation in Rational Drug Design. *J Chem Inf Model* 2018;58:1473–82. <https://doi.org/10.1021/acs.jcim.8b00261>.
- [134] Jarzynski C. Nonequilibrium Equality for Free Energy Differences. *Phys Rev Lett* 1997;78:2690–3. <https://doi.org/10.1103/PhysRevLett.78.2690>.
- [135] Xiong H, Crespo A, Marti M, Estrin D, Roitberg AE. Free Energy Calculations with Non-Equilibrium Methods: Applications of the Jarzynski Relationship. *Theor Chem Acc* 2006;116:338–46. <https://doi.org/10.1007/s00214-005-0072-2>.
- [136] Colizzi F, Perozzo R, Scapozza L, Recanatini M, Cavalli A. Single-Molecule Pulling Simulations Can Discern Active from Inactive Enzyme Inhibitors. *J Am Chem Soc* 2010;132:7361–71. <https://doi.org/10.1021/ja100259r>.
- [137] Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys* 1977;23:187–99. [https://doi.org/10.1016/0021-9991\(77\)90121-8](https://doi.org/10.1016/0021-9991(77)90121-8).
- [138] Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 1992;13:1011–21. <https://doi.org/10.1002/jcc.540130812>.
- [139] Mihalovits LM, Ferenczy GG, Keserü GM. Affinity and Selectivity Assessment of Covalent Inhibitors by Free Energy Calculations. *J Chem Inf Model* 2020;60:6579–94. <https://doi.org/10.1021/acs.jcim.0c00834>.
- [140] Jorgensen WL, Thomas LL. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J Chem Theory Comput* 2008;4:869–76. <https://doi.org/10.1021/ct800011m>.
- [141] Zwanzig RW. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J Chem Phys* 1954;22:1420–6. <https://doi.org/10.1063/1.1740409>.
- [142] Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 2013;27:221–34. <https://doi.org/10.1007/s10822-013-9644-8>.
- [143] Schrödinger Release 2020-1: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2016; Impact, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger,

- LLC, New York, NY, 2020. 2020.
- [144] Schrödinger Release 2021-3: LigPrep, Schrödinger, LLC, New York, NY, 2021. 2021.
- [145] Hunter JC, Gurbani D, Ficarro SB, Carrasco MA, Lim SM, Choi HG, et al. In situ selectivity profiling and crystal structure of SML-8-73-1, an active site inhibitor of oncogenic K-Ras G12C. *Proc Natl Acad Sci* 2014;111:8895–900. <https://doi.org/10.1073/pnas.1404639111>.
- [146] Sogabe S, Kawakita Y, Igaki S, Iwata H, Miki H, Cary DR, et al. Structure-Based Approach for the Discovery of Pyrrolo[3,2-d]pyrimidine-Based EGFR T790M/L858R Mutant Inhibitors. *ACS Med Chem Lett* 2013;4:201–5. <https://doi.org/10.1021/ml300327z>.
- [147] Engel J, Richters A, Getlik M, Tomassi S, Keul M, Termathe M, et al. Targeting Drug Resistance in EGFR with Covalent Inhibitors: A Structure-Based Design Approach. *J Med Chem* 2015;58:6844–63. <https://doi.org/10.1021/acs.jmedchem.5b01082>.
- [148] Thorarensen A, Dowty ME, Banker ME, Juba B, Jussif J, Lin T, et al. Design of a Janus Kinase 3 (JAK3) Specific Inhibitor 1-((2S,5R)-5-((7H-Pyrrolo[2,3-d]pyrimidin-4-yl)amino)-2-methylpiperidin-1-yl)prop-2-en-1-one (PF-06651600) Allowing for the Interrogation of JAK3 Signaling in Humans. *J Med Chem* 2017;60:1971–93. <https://doi.org/10.1021/acs.jmedchem.6b01694>.
- [149] Kutach AK, Villaseñor AG, Lam D, Belunis C, Janson C, Lok S, et al. Crystal Structures of IL-2-inducible T cell Kinase Complexed with Inhibitors: Insights into Rational Drug Design and Activity Regulation. *Chem Biol Drug Des* 2010;76:154–63. <https://doi.org/10.1111/j.1747-0285.2010.00993.x>.
- [150] Cui H, Baur R, Le Chapelain C, Dubiella C, Heinemeyer W, Huber EM, et al. Structural Elucidation of a Nonpeptidic Inhibitor Specific for the Human Immunoproteasome. *ChemBioChem* 2017;18:523–6. <https://doi.org/10.1002/cbic.201700021>.
- [151] Harshbarger W, Miller C, Diedrich C, Sacchettini J. Crystal Structure of the Human 20S Proteasome in Complex with Carfilzomib. *Structure* 2015;23:418–24. <https://doi.org/10.1016/j.str.2014.11.017>.
- [152] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem* 2004;47:1739–49. <https://doi.org/10.1021/jm0306430>.
- [153] Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J Med Chem* 2004;47:1750–9. <https://doi.org/10.1021/jm030644s>.
- [154] Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, et al. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J Med Chem* 2006;49:6177–96. <https://doi.org/10.1021/jm051256o>.
- [155] Schrödinger Release 2019-2: Prime, Schrödinger, LLC, New York, NY, 2019. n.d.
- [156] Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem* 2005;26:1752–80. <https://doi.org/10.1002/jcc.20292>.
- [157] Bowers KJ, Chow E, Huafeng X, Dror RO, Eastwood MP, Gregersen BA, et al. “Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters,” *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, Tampa, Florida, 2006, November 11-17, 2006.
- [158] Schrödinger Release 2019-2: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2019. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2019. n.d.
- [159] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML, Chandrasekhar,

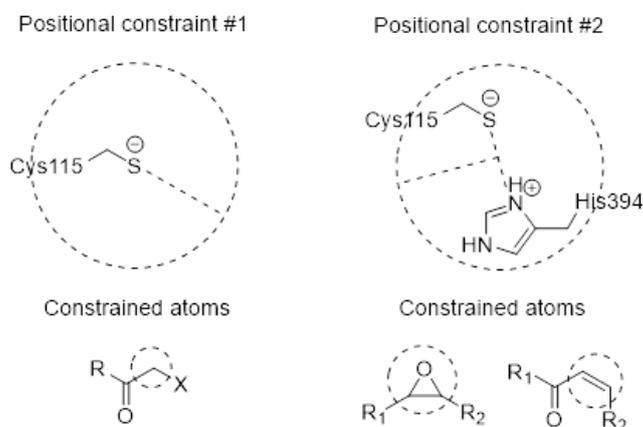
- Jayaraman Madura JD, et al. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–35. <https://doi.org/10.1063/1.445869>.
- [160] Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE, Cruzeiro IVWD, et al. AMBER 2018, University of California, San Francisco 2018.
- [161] Chen W, Deng Y, Russell E, Wu Y, Abel R, Wang L. Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges. *J Chem Theory Comput* 2018;14:6346–58. <https://doi.org/10.1021/acs.jctc.8b00825>.
- [162] Flanagan ME, Abramite JA, Anderson DP, Aulabaugh A, Dahal UP, Gilbert AM, et al. Chemical and computational methods for the characterization of covalent reactive groups for the prospective design of irreversible inhibitors. *J Med Chem* 2014;57:10072–9. <https://doi.org/10.1021/jm501412a>.
- [163] Marenich A V., Cramer CJ, Truhlar DG. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J Phys Chem B* 2009;113:6378–96. <https://doi.org/10.1021/jp810292n>.
- [164] Pulay P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem Phys Lett* 1980;73:393–8. [https://doi.org/10.1016/0009-2614\(80\)80396-4](https://doi.org/10.1016/0009-2614(80)80396-4).
- [165] Cortis CM, Friesner RA. Numerical solution of the Poisson-Boltzmann equation using tetrahedral finite-element meshes. *J Comput Chem* 1997;18:1591–608. [https://doi.org/10.1002/\(SICI\)1096-987X\(199710\)18:13<1591::AID-JCC3>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1096-987X(199710)18:13<1591::AID-JCC3>3.0.CO;2-M).
- [166] Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem* 2004;25:1157–74. <https://doi.org/10.1002/jcc.20035>.
- [167] Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 1977;23:327–41. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
- [168] Schrödinger Release 2021-1: Maestro, Schrödinger, LLC, New York, NY, 2021. n.d.
- [169] 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;55:589–99. <https://doi.org/10.1021/ci500720e>.
- [170] Mihalovits LM, Ferenczy GG, Keserü GM. Catalytic Mechanism and Covalent Inhibition of UDP-N-Acetylglucosamine Enolpyruvyl Transferase (MurA): Implications to the Design of Novel Antibacterials. *J Chem Inf Model* 2019;59:5161–73. <https://doi.org/10.1021/acs.jcim.9b00691>.
- [171] Promoting transparency and reproducibility in enhanced molecular simulations. *Nat Methods* 2019;16:670–3. <https://doi.org/10.1038/s41592-019-0506-8>.
- [172] Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. *Comput Phys Commun* 2014;185:604–13. <https://doi.org/10.1016/j.cpc.2013.09.018>.
- [173] Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* 2009;180:1961–72. <https://doi.org/10.1016/j.cpc.2009.05.011>.
- [174] Chai J-D, Head-Gordon M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys Chem Chem Phys* 2008;10:6615. <https://doi.org/10.1039/b810189b>.
- [175] Dunning TH. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J Chem Phys* 1989;90:1007–23. <https://doi.org/10.1063/1.456153>.

- [176] Woon DE, Dunning TH. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J Chem Phys* 1993;98:1358–71. <https://doi.org/10.1063/1.464303>.
- [177] Miertuš S, Scrocco E, Tomasi J. Electrostatic interaction of a solute with a continuum. A direct utilization of AB initio molecular potentials for the prevision of solvent effects. *Chem Phys* 1981;55:117–29. [https://doi.org/10.1016/0301-0104\(81\)85090-2](https://doi.org/10.1016/0301-0104(81)85090-2).
- [178] Steinbrecher T, Loeffler HH. Small molecule binding to T4-lysozyme L99A n.d. <http://ambermd.org/tutorials/advanced/tutorial9/> (accessed June 5, 2020).
- [179] Grossfield A. WHAM: an implementation of the weighted histogram analysis method n.d.
- [180] Jackson SG, Zhang F, Chindemi P, Junop MS, Berti PJ. Evidence of kinetic control of ligand binding and staged product release in MurA (Enolpyruvyl UDP-GlcNAc synthase)-catalyzed reactions. *Biochemistry* 2009;48:11715–23. <https://doi.org/10.1021/bi901524q>.
- [181] Eschenburg S, Priestman M, Schönbrunn E. Evidence that the fosfomycin target Cys115 in UDP-N-acetylglucosamine Enolpyruvyl Transferase (MurA) is essential for product release. *J Biol Chem* 2005;280:3757–63. <https://doi.org/10.1074/jbc.M411325200>.
- [182] Benesch RE, Benesch R. The Acid Strength of the -SH Group in Cysteine and Related Compounds. *J Am Chem Soc* 1955;77:5877–81. <https://doi.org/10.1021/ja01627a030>.
- [183] Baum EZ, Montenegro DA, Licata L, Turchi I, Webb GC, Foleno BD, et al. Identification and characterization of new inhibitors of the Escherichia coli MurA enzyme. *Antimicrob Agents Chemother* 2001;45:3182–8. <https://doi.org/10.1128/AAC.45.11.3182-3188.2001>.
- [184] Notredame C, Higgins DG, Heringa J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 2000;302:205–17. <https://doi.org/10.1006/jmbi.2000.4042>.
- [185] Krishnan S, Miller RM, Tian B, Mullins RD, Jacobson MP, Taunton J. Design of reversible, cysteine-targeted michael acceptors guided by kinetic and computational analysis. *J Am Chem Soc* 2014;136:12624–30. <https://doi.org/10.1021/ja505194w>.
- [186] Smith JM, Jami Alahmadi Y, Rowley CN. Range-separated DFT functionals are necessary to model thio-Michael additions. *J Chem Theory Comput* 2013;9:4860–5. <https://doi.org/10.1021/ct400773k>.
- [187] Krenske EH, Petter RC, Zhu Z, Houk KN. Transition states and energetics of nucleophilic additions of thiols to substituted  $\alpha,\beta$ -unsaturated ketones: Substituent effects involve enone stabilization, product branching, and solvation. *J Org Chem* 2011;76:5074–81. <https://doi.org/10.1021/jo200761w>.
- [188] Awoonor-Williams E, Isley WC, Dale SG, Johnson ER, Yu H, Becke AD, et al. Quantum Chemical Methods for Modeling Covalent Modification of Biological Thiols. *J Comput Chem* 2020;41:427–38. <https://doi.org/10.1002/jcc.26064>.
- [189] Silva JRA, Cianni L, Araujo D, Batista PHJ, De Vita D, Rosini F, et al. Assessment of the Cruzain Cysteine Protease Reversible and Irreversible Covalent Inhibition Mechanism. *J Chem Inf Model* 2020;60:1666–77. <https://doi.org/10.1021/acs.jcim.9b01138>.
- [190] Schwartz PA, Kuzmic P, Solowiej J, Bergqvist S, Bolanos B, Almaden C, et al. Covalent EGFR inhibitor analysis reveals importance of reversible interactions to potency and mechanisms of drug resistance. *Proc Natl Acad Sci U S A* 2014;111:173–8. <https://doi.org/10.1073/pnas.1313733111>.
- [191] Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc*

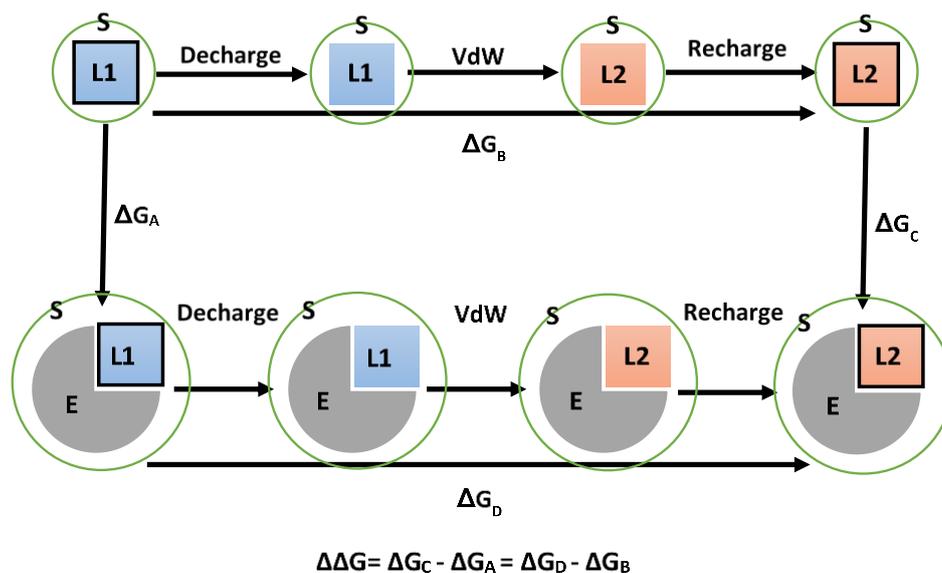
- 2015;137:2695–703. <https://doi.org/10.1021/ja512751q>.
- [192] He X, Liu S, Lee TS, Ji B, Man VH, York DM, et al. Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein-Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* 2020;5:4611–9. <https://doi.org/10.1021/acsomega.9b04233>.
- [193] Simonson T, Carlsson J, Case DA. Proton Binding to Proteins: pKa Calculations with Explicit and Implicit Solvent Models. *J Am Chem Soc* 2004;126:4167–80. <https://doi.org/10.1021/ja039788m>.
- [194] <http://biophysics.cs.vt.edu/index.php>; accessed 15/09/2020 n.d.
- [195] Gruden M, Andjeklović L, Jissy AK, Stepanović S, Zlatar M, Cui Q, et al. Benchmarking density functional tight binding models for barrier heights and reaction energetics of organic molecules. *J Comput Chem* 2017;38:2171–85. <https://doi.org/10.1002/jcc.24866>.
- [196] Wang X, Li R, Cui W, Li Q, Yao J. QM/MM free energy Simulations of an efficient Gluten Hydrolase (Kuma030) Implicate for a Reactant-State Based Protein-Design Strategy for General Acid/Base Catalysis. *Sci Rep* 2018;8:7042. <https://doi.org/10.1038/s41598-018-25471-z>.
- [197] Wang X, Bakanina Kissanga GM, Li E, Li Q, Yao J. The catalytic mechanism of: S - acyltransferases: Acylation is triggered on by a loose transition state and deacylation is turned off by a tight transition state. *Phys Chem Chem Phys* 2019;21:12163–72. <https://doi.org/10.1039/c9cp02248a>.
- [198] McGregor LM, Jenkins ML, Kerwin C, Burke JE, Shokat KM. Expanding the Scope of Electrophiles Capable of Targeting K-Ras Oncogenes. *Biochemistry* 2017;56:3178–83. <https://doi.org/10.1021/acs.biochem.7b00271>.
- [199] Copeland RA. Evaluation of enzyme inhibitors in drug discovery. A guide for medicinal chemists and pharmacologists. *Methods Biochem Anal* 2005;46:1–265.
- [200] Schwöbel JAH, Wondrousch D, Koleva YK, Madden JC, Cronin MTD, Schüürmann G. Prediction of michael-type acceptor reactivity toward glutathione. *Chem Res Toxicol* 2010;23:1576–85. <https://doi.org/10.1021/tx100172x>.
- [201] Mihalovits LM, Ferenczy GG, Keserű GM. Mechanistic and Thermodynamic Characterization of Oxathiazolones as Potent and Selective Covalent Immunoproteasome Inhibitors. *Comput Struct Biotechnol J* 2021. <https://doi.org/10.1016/j.csbj.2021.08.008>.
- [202] Lin G, Li D, De Carvalho LPS, Deng H, Tao H, Vogt G, et al. Inhibitors selective for mycobacterial versus human proteasomes. *Nature* 2009;461:621–6. <https://doi.org/10.1038/nature08357>.
- [203] Uranga J, Hasecke L, Proppe J, Fingerhut J, Mata RA. Theoretical Studies of the Acid–Base Equilibria in a Model Active Site of the Human 20S Proteasome. *J Chem Inf Model* 2021. <https://doi.org/10.1021/acs.jcim.0c01459>.
- [204] Serrano-Aparicio N, Świderek K, Moliner V. Theoretical study of the inhibition mechanism of human 20S proteasome by dihydroeponemycin. *Eur J Med Chem* 2019;164:399–407. <https://doi.org/10.1016/j.ejmech.2018.12.062>.
- [205] Serrano-Aparicio N, Moliner V, Świderek K. Nature of Irreversible Inhibition of Human 20S Proteasome by Salinosporamide A. The Critical Role of Lys–Asp Dyad Revealed from Electrostatic Effects Analysis. *ACS Catal* 2021;11:3575–89. <https://doi.org/10.1021/acscatal.0c05313>.
- [206] Wei D, Lei B, Tang M, Zhan CG. Fundamental reaction pathway and free energy profile for inhibition of proteasome by epoxomicin. *J Am Chem Soc* 2012;134:10436–50. <https://doi.org/10.1021/ja3006463>.
- [207] Wei D, Tang M, Zhan CG. Fundamental reaction pathway and free energy profile of

- proteasome inhibition by syringolin A (SylA). *Org Biomol Chem* 2015;13:6857–65. <https://doi.org/10.1039/c5ob00737b>.
- [208] Hsu H-C, Singh PK, Fan H, Wang R, Sukenick G, Nathan C, et al. Structural Basis for the Species-Selective Binding of N,C-Capped Dipeptides to the Mycobacterium tuberculosis Proteasome. *Biochemistry* 2017;56:324–33. <https://doi.org/10.1021/acs.biochem.6b01107>.
- [209] Huber EM, Basler M, Schwab R, Heinemeyer W, Kirk CJ, Groettrup M, et al. Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity. *Cell* 2012;148:727–38. <https://doi.org/10.1016/j.cell.2011.12.030>.

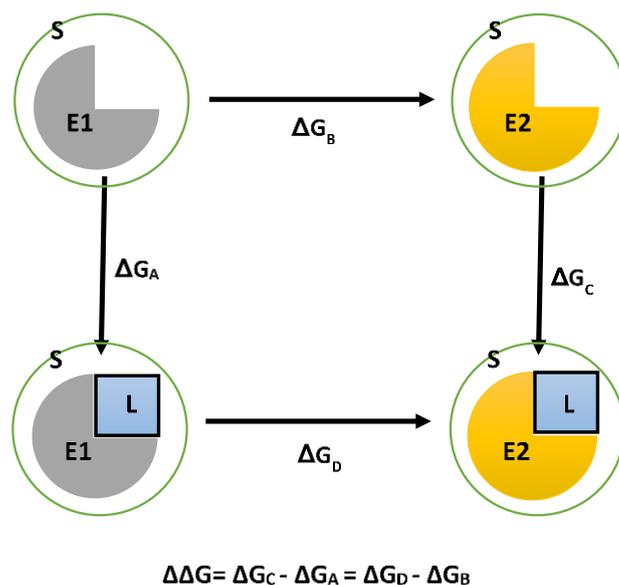
## APPENDIX A



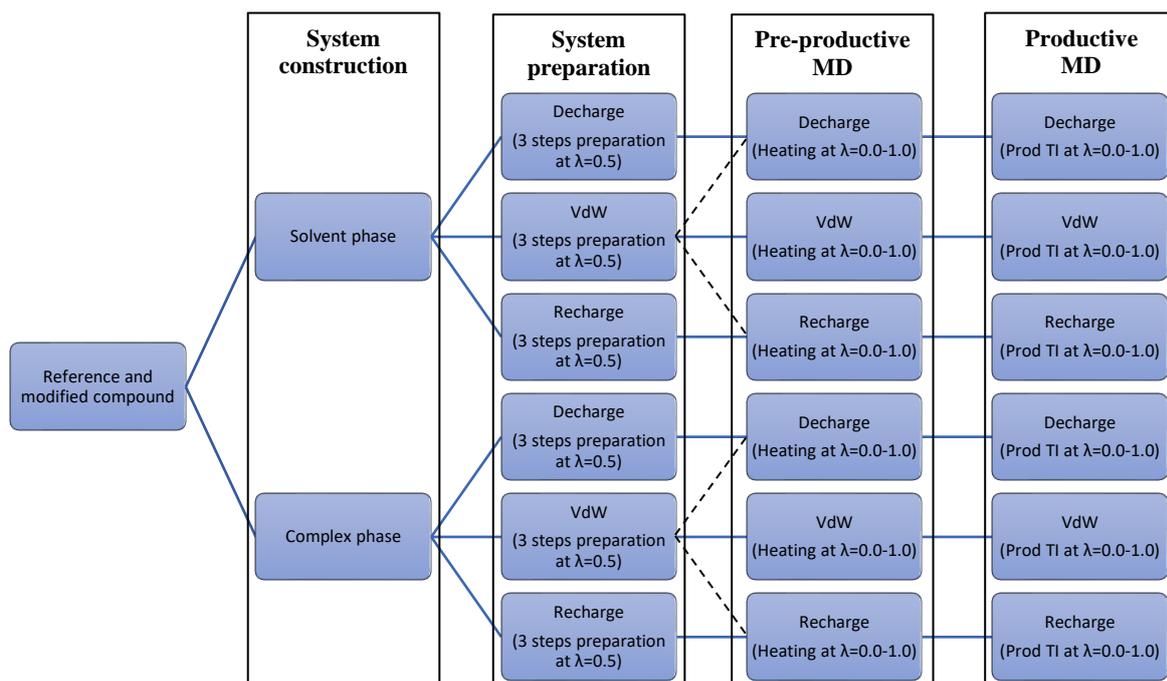
**Figure A1.** Positional constraints applied in MurA dockings. Constraint #2 positions atoms within 5Å from the midpoint of the S-N segment. Two out of the three atoms of the oxirane ring are required to satisfy the constraint.



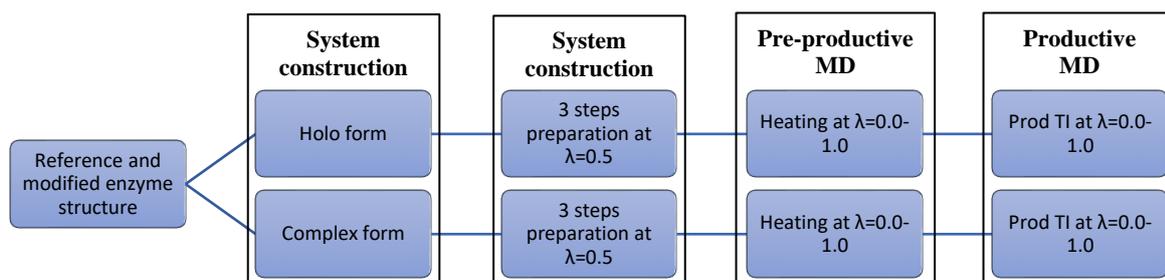
**Figure A2.** General scheme of a thermodynamic cycle calculating ligand binding free energy differences (L: ligand; E: enzyme; S: solvent; L1, L2: ligands, ligands with charged SC atoms are framed).



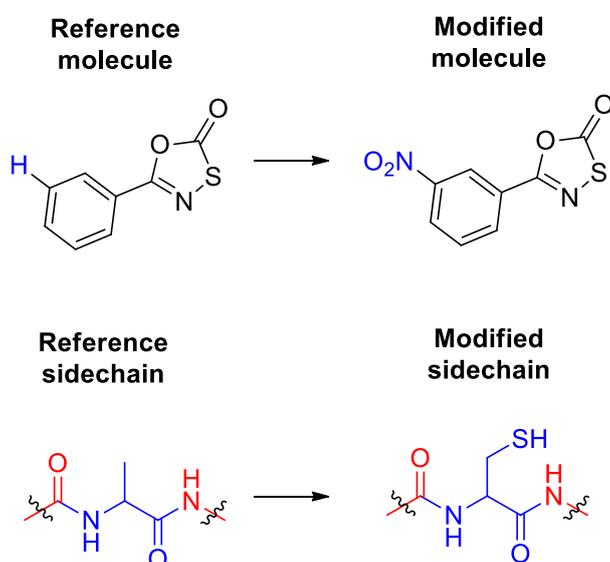
**Figure A3.** General scheme of a thermodynamic cycle calculating the binding free energy of the same ligand in related enzyme (L: ligand; E1, E2: enzyme; S: solvent.).



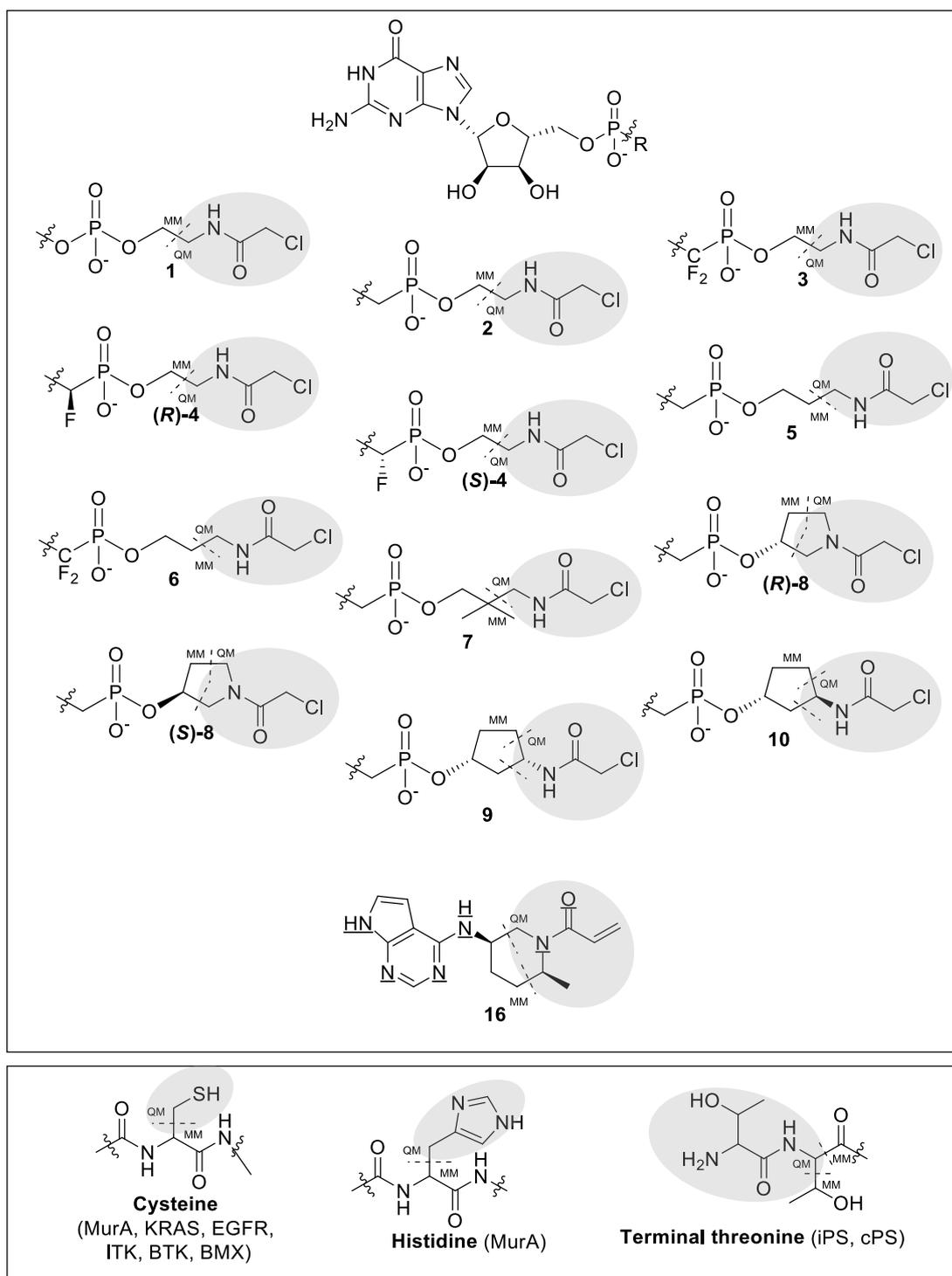
**Figure A4.** General workflow of ligand transformational thermodynamic integrations using the dual topology approach. During the TI simulation of KRAS and EGFR the sub-steps branch after the vdw system preparation only (dashed black lines).



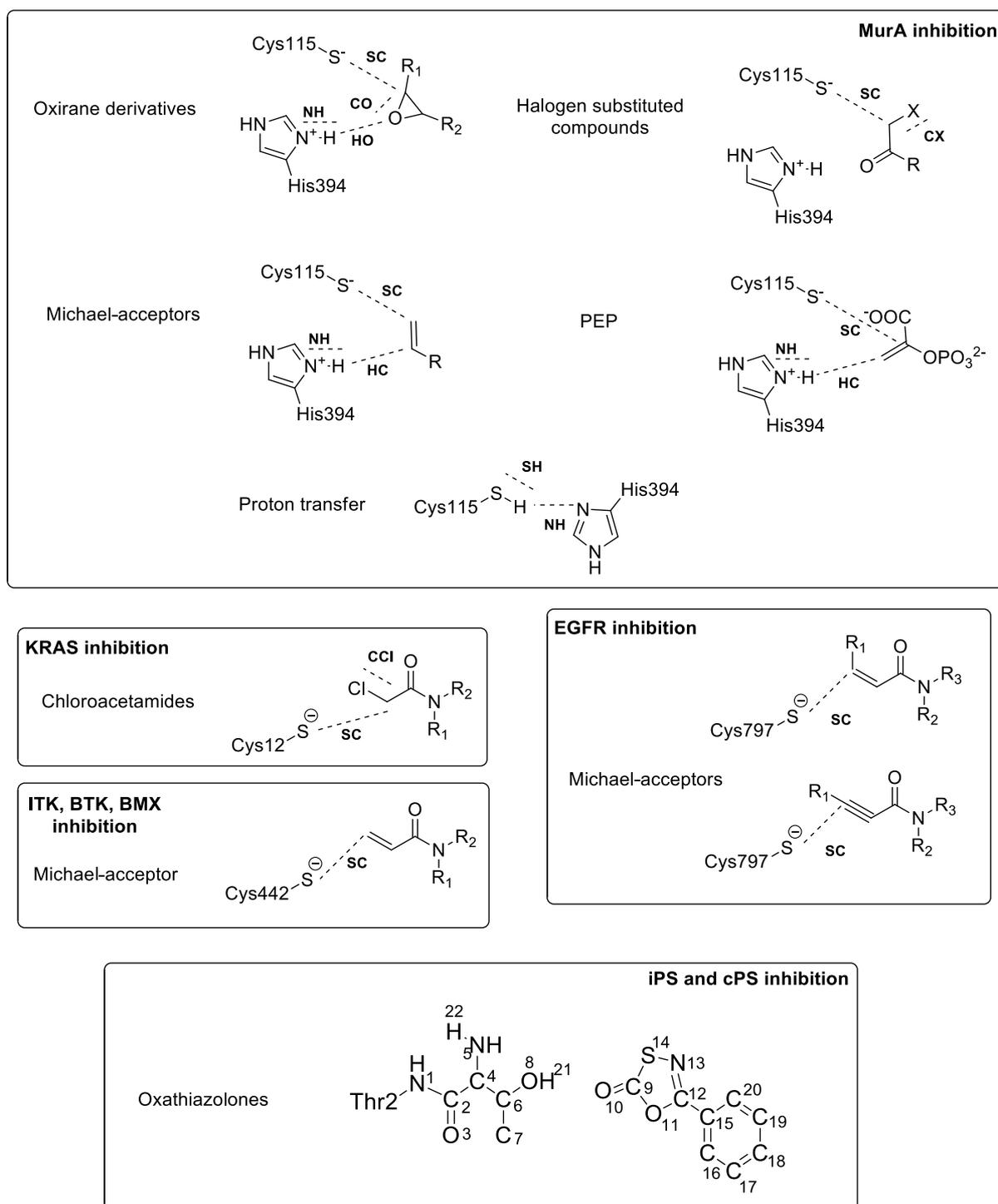
**Figure A5.** General workflow of sidechain mutational thermodynamic integrations using the single topology approach.



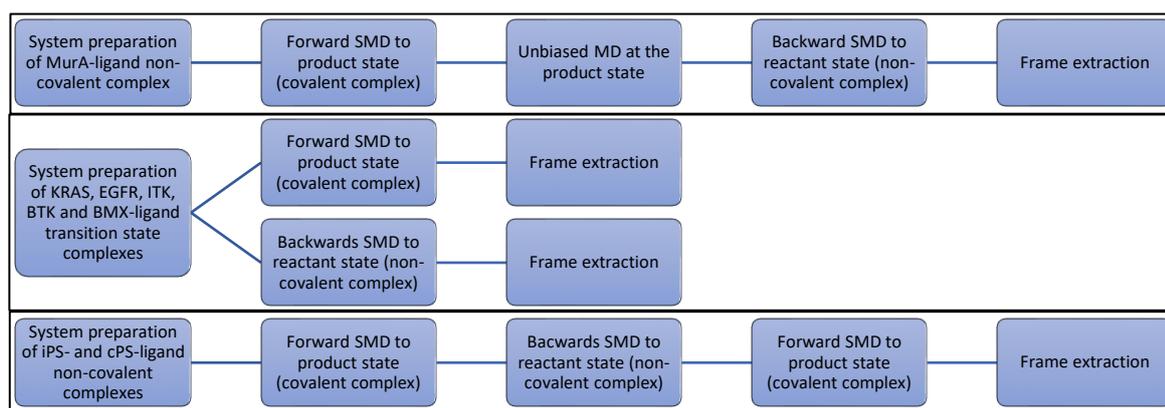
**Figure A6.** Representative examples for the scheme of the TI alchemical transformations. In ligand modifications the whole ligand is part of the TI region, while the atoms subject to modification (shown in blue) are part of the softcore region. The remaining ligand atoms are linearly scaled. In sidechain mutational TI the complete sidechain is part of the softcore and therefore the TI region. Atoms in red are not part any of the special regions.



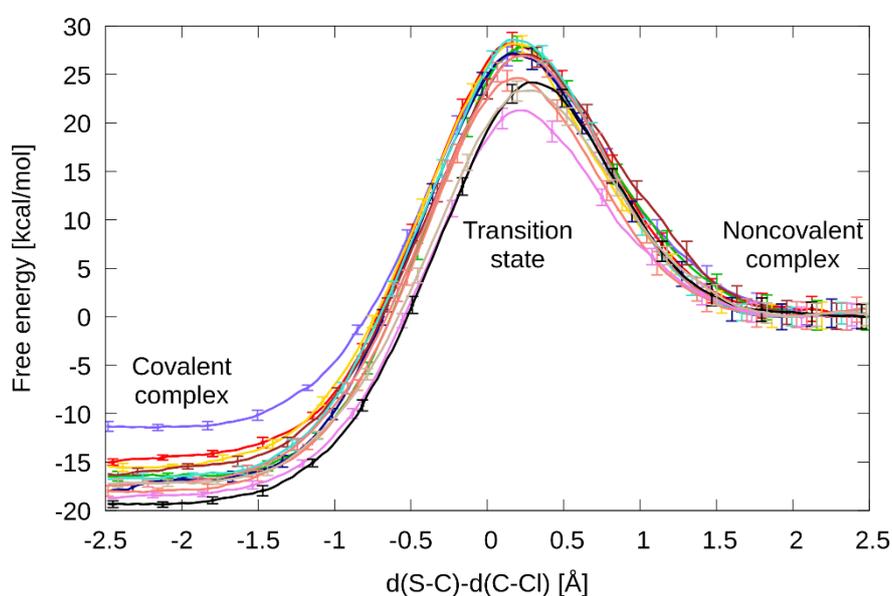
**Figure A7.** QM and MM regions during QM/MM simulations. The inhibitors against MurA, EGFR and iPS (cPS) were not separated, all their atoms were part of the QM region. In the bottom box the QM cuts are shown for the sidechain of the targeted enzymes.



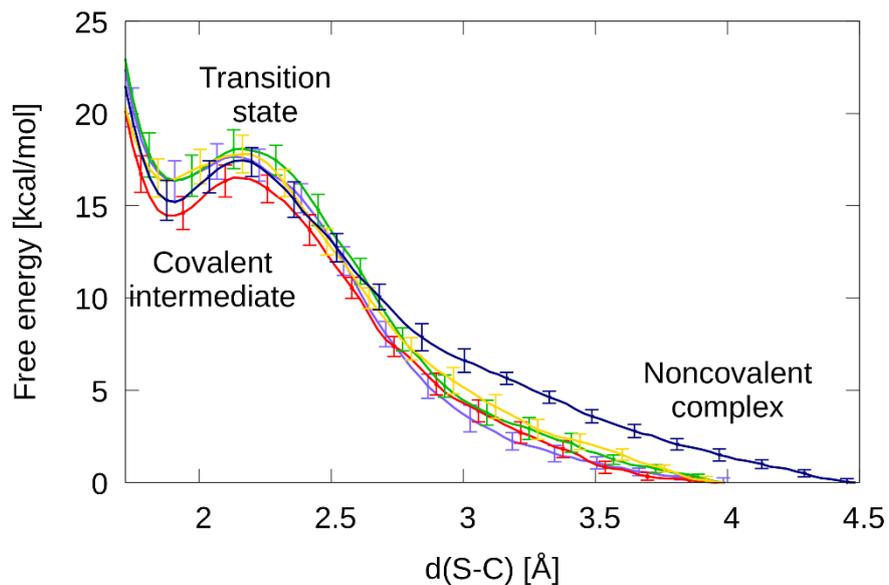
**Figure A8.** Applied distances as reaction coordinates during the QM/MM SMD and Umbrella Sampling simulations (the specific combinations are described in **Table A1**).



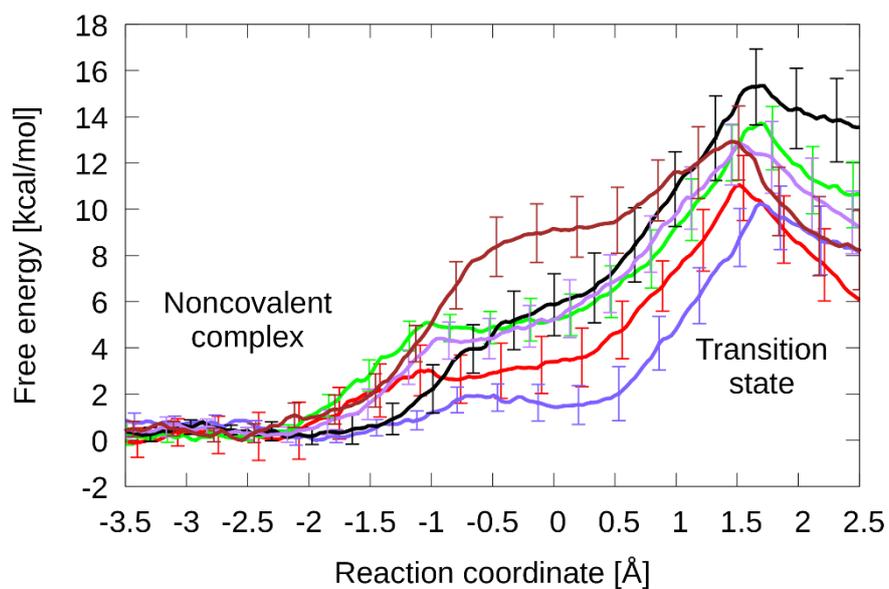
**Figure A9.** Workflow of the SMD simulations preparing the initial structures for subsequent umbrella sampling simulations. During the mechanistical studies of iPS, the same protocol was followed for every reaction step as described in the third section of the figure.



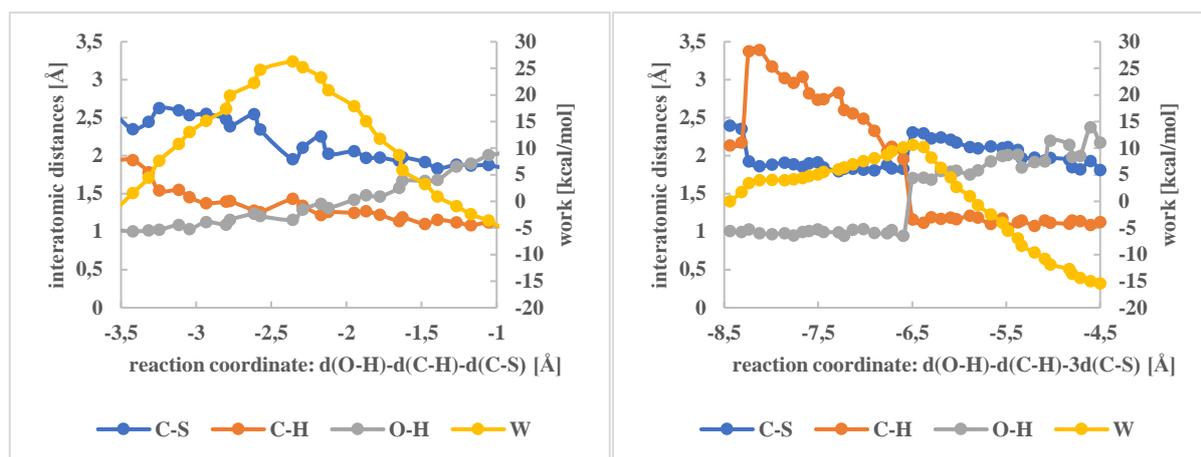
**Figure A10.** PMF curves of the KRAS inhibitors **1** (blue), **2** (red), **3** (green), (**R**)-**4** (gold), (**S**)-**4** (navy), **5** (brown), **6** (turquoise), **7** (coral), (**R**)-**8** (salmon), (**S**)-**8** (violet), **9** (black), **10** (bisque).



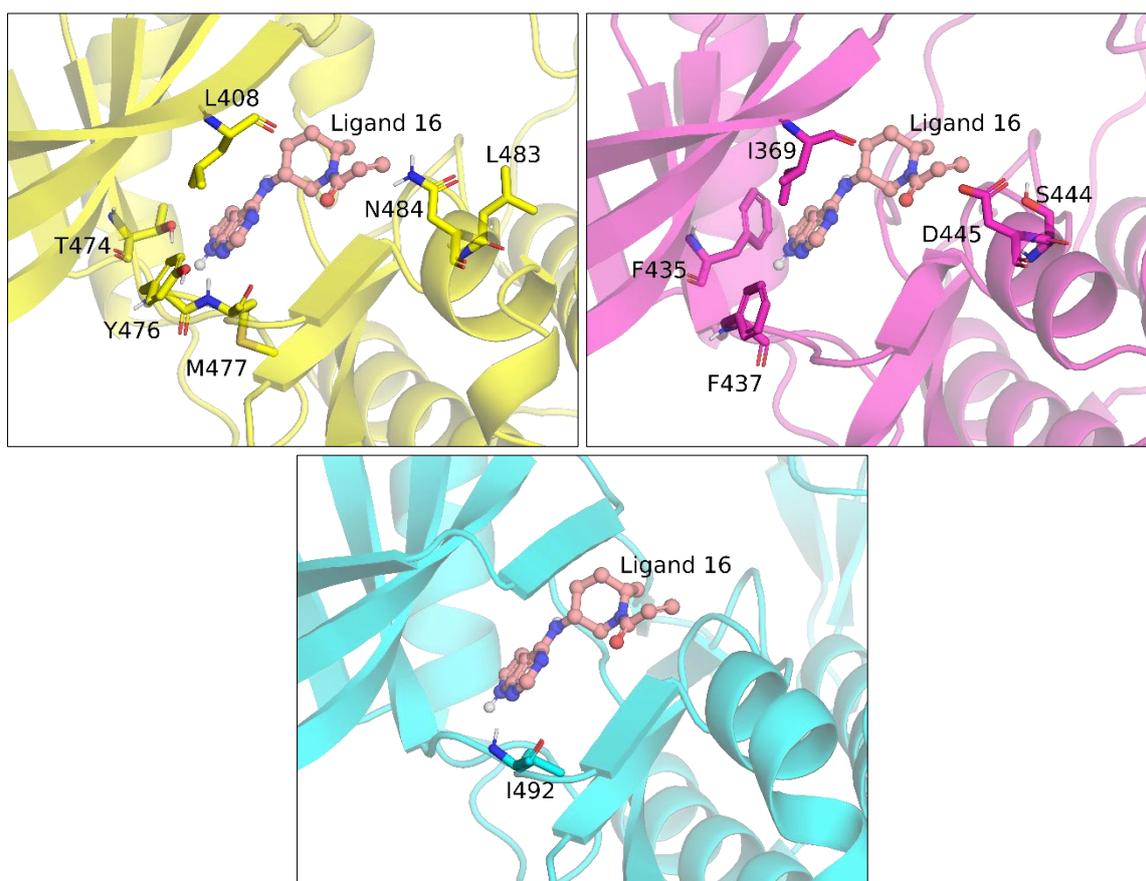
**Figure A11.** PMF curves of the EGFR inhibitors **11** (blue), **12** (red), **13** (green), **14** (gold), **15** (navy).



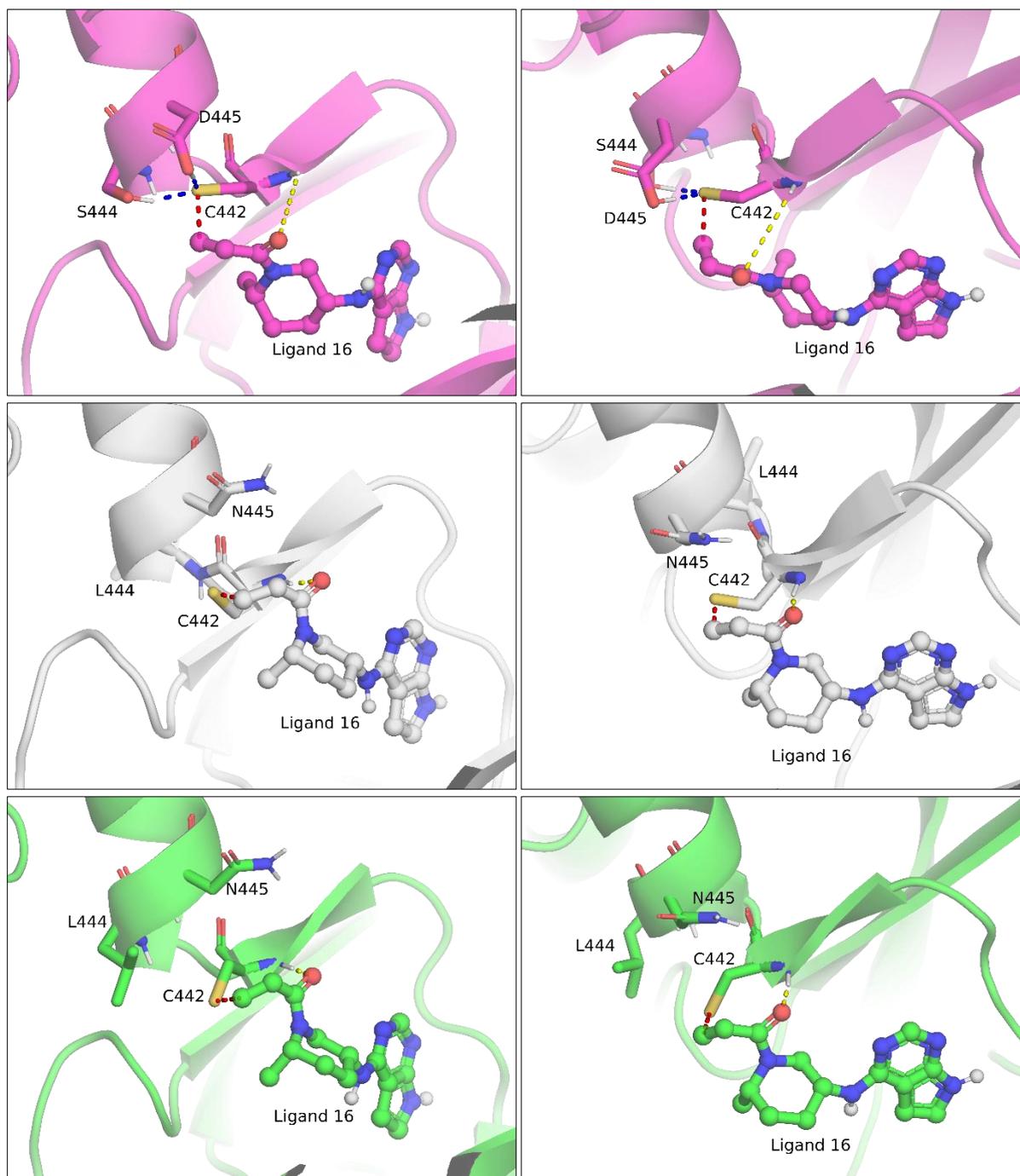
**Figure A12.** PMF curves of the rate-determining step of the iPS inhibitors **HT1146** (red), **HT2004** (blue), **HT1071** (green), **HT2210** (black), **HT1213** (purple), **HT1042** (brown).



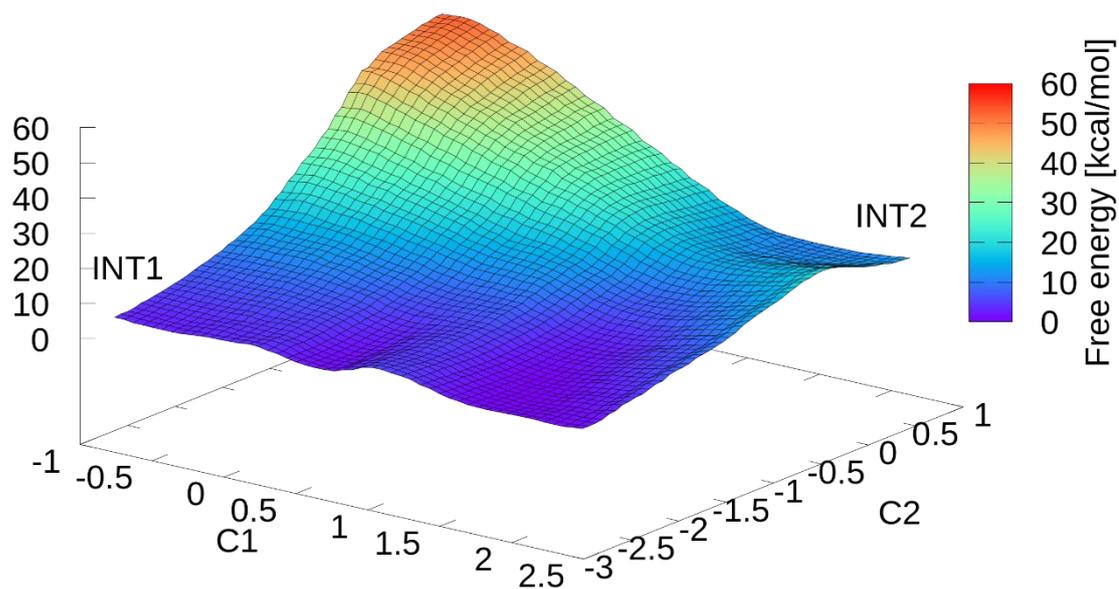
**Figure A13.** SMD curves of the addition of Cys797 of EGFR on **11**. The reaction coordinate is combined distances:  $RC = d(O-H) - d(C-H) - c \cdot d(C-S)$ . Calculations were performed with setting  $c = 0.3, 1.0, 1.5, 3.0$  and  $4.0$ . The C-S and C-H bonds are formed concerted when  $c = 1.0$  or  $1.5$  (left  $C = 1.0$ ) and they are formed consecutively when  $c = 3.0$ , or  $c = 4.0$  (right  $c = 3.0$ ). The maximum of the work is considerably lower along the consecutive path. The SMD simulations were performed analogously to the inhibitor-enzyme systems (see **Chapter 4 – Materials and methods**) except that the equilibration was performed in the reactant state.



**Figure A14.** Aligned structures of the kinases ITK (3MIY, pink), BTK (5P9F, gold) and BMX (3SXS, turquoise) complexed with **16**. Ligands are shown in ball and sticks, residues subject to mutations are represented as sticks.



**Figure A15.** Transition state structures in two different views (left and right) taken from the end of the trajectory of QM/MM MD US simulations with **16** at the 2.1 Å reaction coordinate (ITK, pink; hybrid BTK, gray; hybrid BMX, green). Ligands are shown in ball and sticks, residues are represented as sticks. The covalent bond between the warhead and Cys442 (ITK numbering) and the hydrogen bond between the warheads' carbonyl atom and the amide hydrogen of the Cys442 are indicated as red and yellow dashed lines, respectively. Additional intramolecular hydrogen bonds in ITK are marked with blue dashed lines.



**Figure A16.** Two-dimensional potential of mean force for the rate determining step of the carbonate route (**HT1146**-iPS complex). C1 and C2 correspond to the (C9-S14)-(O8-C9) and (N5-H22)-(S14-H22) distance combinations, respectively (see **Figure 43** and **Figure A8** for atom numbering).

## APPENDIX B

**Table A1.** Applied distances as reaction coordinates during the QM/MM SMD and Umbrella Sampling simulations (distances are clarified in **Figure A8**).

### MurA

Ligand	RC (first step)	RC (second step)
<i>M1-M3 (oxiranes)</i>	<b>d(S-C) - d(C-O)</b>	<b>d(H-O) - d(N-H)</b>
<i>M4-M6 (halo compounds)</i>	<b>d(S-C) - d(C-Br)</b>	-
<i>M7-M9 (Michael-acceptors)</i>	<b>d(S-C)</b>	<b>d(H-C) - d(N-H)</b>
<i>PEP</i>	<b>d(S-C)</b>	<b>d(H-C) - d(N-H)</b>
<i>Proton transfer</i>	<b>d(S-H) - d(N-H)</b>	-

### KRAS

Ligand	RC (first step)
<i>1-10 (chloroacetamides)</i>	<b>d(S-C) - d(C-Cl)</b>

### EGFR

Ligand	RC (first step)
<i>11-15,16-18 (Michael-acceptors)</i>	<b>d(S-C)</b>

### ITK, BTK, BMX

Ligand	RC (first step)
<i>16 (Michael-acceptor)</i>	<b>d(S-C)</b>

### iPS

Ligand	RC (first step)	RC (second step)	RC (third step)
<i>HT1146 (carbonate)</i>	<b>d(O8-H21) - d(H21-N5)</b>	<b>d(C9-S14) - d(O8-C9) + d(N5-H22) - d(H22-S14)</b>	<b>d(C9-O11) - d(N5-C5) then d(N5-H21) - d(H21-N13)</b>
<i>HT1146 (carbothioate)</i>	<b>d(O8-H21)-d(H21-N5)</b>	<b>d(C9-O11) - d(O8-C9) + d(N5-H22) - d(22-N13)</b>	<b>d(C9-S14) - d(N3-C9) + d(N5-H21) - d(H21-S14)</b>
<i>Remaining oxathiazolones</i>	-	<b>d(C9-S14) - d(O8-C9) + d(N5-H22) - d(H22-S14)</b>	-

### cPS

Ligand	RC (first step)	RC (second step)	RC (third step)
<i>HT1146, HT2004 (oxathiazolones)</i>	-	<b>d(C9-S14) - d(O8-C9) + d(N5-H22) - d(H22-S14)</b>	-

**Table A2.** Experimental and calculated free energy differences of the transformations during TI simulations.

**KRAS**

Reference compound	Perturbed compound	$\Delta\Delta G_{\text{exp}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc}}$ [kcal·mol <sup>-1</sup> ]
<b>2</b>	<b>1</b>	-3.4	-4.6
<b>2</b>	<b>(R)-4</b>	0.6	-0.9
<b>2</b>	<b>(S)-4</b>	0.6	2.2
<b>2</b>	<b>5</b>	-1.2	-1.4
<b>5</b>	<b>6</b>	0.0	-1.6
<b>6</b>	<b>3</b>	0.0	-0.5
<b>5</b>	<b>7</b>	1.5	3.7
<b>2</b>	<b>(R)-8</b>	-0.7	1.5
<b>(R)-8</b>	<b>(S)-8</b>	-0.7	-0.4
<b>5</b>	<b>10</b>	2.7	1.7
<b>10</b>	<b>9</b>	0.2	1.9

**EGFR**

Reference compound	Perturbed compound	$\Delta\Delta G_{\text{exp}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc}}$ [kcal·mol <sup>-1</sup> ]
<b>11</b>	<b>12</b>	1.6	2.8
<b>12</b>	<b>13</b>	0.4	0.0
<b>12</b>	<b>14</b>	2.3	4.3
<b>12</b>	<b>15</b>	1.6	1.5
<b>13</b>	<b>15<sup>+</sup></b>	1.6	-0.5

**iPS**

Reference compound	Perturbed compound	$\Delta\Delta G_{\text{exp}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc}}$ [kcal·mol <sup>-1</sup> ]
<b>HT1146</b>	<b>HT1042</b>	-1.8	-0.9
<b>HT1146</b>	<b>HT2004</b>	-1.0	-2.0
<b>HT1146</b>	<b>HT1071</b>	-	0.5
<b>HT1071</b>	<b>HT2210</b>	-	-1.0
<b>HT1146</b>	<b>HT1213</b>	-1.2	-1.0

**Table A3.** Calculated free energy differences of the transformations estimating the error of the TI simulations.

**KRAS**

Cycle	$\Delta\Delta G_{\text{calc1}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc2}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc3}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc4}}$ [kcal·mol <sup>-1</sup> ]	Error [kcal·mol <sup>-1</sup> ]
<b>2 &gt; 1 &gt; 3 &gt; 2</b>	-4.60	7.62	-1.60	-	1.42
<b>2 &gt; (R)-4 &gt; (S)-4 &gt; 2</b>	-0.91	0.35	-2.19	-	-2.75
<b>2 &gt; 5 &gt; 7 &gt; 2</b>	-1.41	0.62	-4.04	-	-1.75
<b>2 &gt; (R)-8 &gt; (S)-8 &gt; 2</b>	1.45	-0.35	-1.42	-	-0.32

RMSD 1.78  
kcal·mol<sup>-1</sup>

**EGFR**

Cycle	$\Delta\Delta G_{\text{calc1}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc2}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc3}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc4}}$ [kcal·mol <sup>-1</sup> ]	Error [kcal·mol <sup>-1</sup> ]
<b>13 &gt; 12 &gt; 15 &gt; 15</b>	-0.02	0.60	0.60	-	1.42

**iPS**

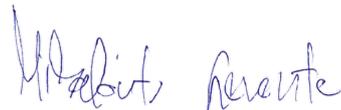
Cycle	$\Delta\Delta G_{\text{calc1}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc2}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc3}}$ [kcal·mol <sup>-1</sup> ]	$\Delta\Delta G_{\text{calc4}}$ [kcal·mol <sup>-1</sup> ]	Error [kcal·mol <sup>-1</sup> ]
<b>HT1146 &gt; HT2210 &gt; HT1071 &gt; HT1146</b>	-1.07	0.99	-0.48	-	-0.56
<b>HT1146 &gt; HT1071 &gt; HT1042 &gt; HT1146</b>	0.48	-1.05	0.86	-	0.30
<b>HT1146 &gt; HT2210 &gt; HT1071 &gt; HT1042 &gt; HT1146</b>	-1.07	0.99	-0.48	0.86	0.26

RMSD 0.40  
kcal·mol<sup>-1</sup>

## NYILATKOZAT

Alulírott **Mihalovits Levente Márk** kijelentem, hogy ezt a doktori értekezést magam készítettem és abban csak a megadott forrásokat használtam fel. Minden olyan részt, amelyet szó szerint, vagy azonos tartalomban, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Budapest, 2022.01.04.

  
.....