

Dissertation zur Erlangung des Doktorgrades  
der Fakultät für  
Chemie und Pharmazie  
der Ludwig-Maximilians-Universität München

**Quantum-chemical investigations of reaction  
mechanisms of enzymes and bioactive  
molecules**

**Iris Diana Blank**

aus

**Villingen-Schwenningen, Deutschland**

**2016**





Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. C. Ochsenfeld betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 11.10.2016

---

(Iris Blank)

Dissertation eingereicht am: 24.10.2016

1. Gutachter: Prof. Dr. Christian Ochsenfeld

2. Gutachter: Prof. Dr. Stephan Sieber

Mündliche Prüfung am: 1.12.2016



### Danksagung

Bei Prof. Dr. Christian Ochsenfeld bedanke ich mich für die spannenden Themenstellungen, die gute Betreuung und die angenehme Zusammenarbeit.

Prof. Dr. Stephan Sieber danke ich für die Zeit und Mühe der Erstellung des Zweitgutachtens und für die gute Zusammenarbeit in den Kooperationsprojekten.

Ich danke meinen Kollegen, die zu einer ausgesprochen netten Arbeitsatmosphäre beigetragen haben. Besonderer Dank gilt Sven Roßbach für viele wertvolle Diskussionen und die produktive wissenschaftliche Zusammenarbeit. Außerdem danke ich Sven für alles, was mir neben der Arbeit noch wichtig ist.

Meinen Eltern danke ich dafür, dass sie mein Studium und diese Arbeit ermöglicht haben.



# Zusammenfassung

In der vorliegenden Arbeit werden Reaktionsmechanismen von Enzymen und bioaktiven Substanzen mit Hilfe von Kraftfeld-Moleküldynamik (FF-MD) Simulationen und quantenmechanisch/molekülmechanischen (QM/MM) Rechnungen untersucht. Die Aufklärung von enzymatischen Reaktionsmechanismen liefert Informationen auf atomarer Ebene über die zugrunde liegenden Prozesse und kann einen wertvollen Beitrag zur Medikamentenentwicklung leisten, wie zum Beispiel zur Entwicklung von Antibiotika oder Anti-Tumor-Wirkstoffen.

Der erste Teil dieser Arbeit behandelt die Aufklärung des Reparaturprozesses oxidativer DNA Schäden. DNA wird ständig durch verschiedene Einflüsse beschädigt, beispielsweise durch mutagene Chemikalien, Strahlung oder reaktive Sauerstoffspezies (ROS). Normalerweise besitzen Zellen eine schnelle und effiziente Reparaturmaschinerie um die Anhäufung von DNA Schäden zu verhindern, was sonst zu Mutationen und zu Krankheiten wie Krebs, Alzheimer, Parkinson oder Diabetes II führen kann. Das Enzym Formamidopyrimidin-DNA-Glykosylase (Fpg) repariert die hochmutagenen oxidativen Purinschäden FapydG (2,6-diamino-4-hydroxy-5-Formamidopyrimidin), FapydA (4,6-diamino-5-Formamido-pyrimidin) und 8OG (7,8-dihydro-8-Oxoguanin). Im Rahmen dieser Arbeit wurde gezeigt, dass der bisher angenommene Basen-protonierte Mechanismus energetisch nicht möglich ist. Stattdessen wurde ein Ribose-protonierter Mechanismus gefunden, der in vollständiger Übereinstimmung mit experimentellen Beobachtungen steht. Der Basen-unabhängige Reaktionsmechanismus für *anti*-gebundene Schäden kann nun die Basenexzision einer Vielfalt von chemisch unterschiedlichen DNA-Schäden durch ein und dasselbe Enzym erklären. Darüber hinaus erklärt er auch warum unpolare Purin-Analoga durch Fpg geschnitten werden. Außerdem impliziert dieser Basen-unabhängige Mechanismus, dass die Diskriminierung der DNA-Schäden im Erkennungsschritt stattfinden muss und nicht erst im aktiven Zentrum von Fpg.

Zusätzlich zu DNA Reparaturprozessen wurde im zweiten Teil dieser Arbeit ein Analogon des Naturproduktes Duocarmycin untersucht. Duocarmycine sind wirksame Anti-Tumor-Wirkstoffe, deren DNA-bindende Wirkung bekannt ist. Interessanterweise binden Analoga, denen der DNA-bindende Teil fehlt, spezifisch an die Isoform 1A1 der Aldehyddehydrogenase (ALDH1A1). Durch strukturelle, biochemische, kinetische sowie theoretische Untersuchungen konnten die Details der Wechselwirkung des zytotoxischen Duocarmycin Analogons Seco Drug 2 und der hydrophoben Tasche von ALDH1A1 aufgeklärt werden. Die Wechselwirkung basiert auf Van-der-Waals und  $\pi$ - $\pi$ -Interaktionen gestapelter aromatischer Ringe. Außerdem wurde herausgefunden, dass Seco Drug 2 ALDH1A1 inhibiert, indem es kovalent an

ein dem katalytischen Cystein benachbartes Cystein bindet. Diese Bindung ist spezifisch für ALDH1A1 und wird Funktionsstudien von ALDH1A1 im gesamten Proteom ermöglichen.

Im dritten Teil dieser Arbeit wurden die Reaktivität und Selektivität der Reaktion aromatischer und aliphatischer Amine mit  $\beta$ -Lactonen und  $\beta$ -Lactamen untersucht um eine Reaktivitätsskala zu erstellen. Es ist bekannt, dass Anilin als aromatisches Amin  $\beta$ -Butyrolacton (BBL) nukleophil an der C3-Position angreift, wohingegen Benzylamin als aliphatisches Amin, BBL an der C1-Position attackiert. Diese ambidente Reaktivität von  $\beta$ -Lactonen (C1- im Vergleich zu C3-Angriff) konnte nun erklärt werden. Die entscheidende Rolle spielt dabei ein  $\text{H}_3\text{O}^+$  Molekül, das den Übergangszustand stabilisiert und so zur Verringerung der Reaktionsbarriere führt. Dadurch wird die Reaktion für Anilin ermöglicht und für Benzylamin wird sie barriereelos. Des Weiteren konnten die Barrieren der Reaktionen von monozyklische  $\beta$ -Lactonen und  $\beta$ -Lactamen sowie bityklischen  $\beta$ -Lactamen (Antibiotika) mit Aminogruppen durch theoretische Berechnungen und gemessene Kinetiken quantifiziert werden. Mit diesen Ergebnissen konnte eine Reaktivitätsskala aufgestellt werden.

Der letzte Teil dieser Arbeit konzentriert sich auf den Mechanismus der Pyridoxalkinase, einem Enzym, das im Vitamin B<sub>6</sub> Stoffwechsel involviert ist. Das aktive Zentrum der Pyridoxalkinase stellt eine Alternative zu der von Antibiotika überwiegend anvisierten Zellwand dar. Damit könnte eine neue, strukturell verschiedene Antibiotika-Klasse entwickelt werden, die vor allem wertvoll für die Behandlung multi-resistenter Bakterienstämme wäre. Die Funktion der Aminosäuren und deren Wechselwirkungen innerhalb des aktiven Zentrums von Pyridoxalkinase wurden untersucht und die Aminosäuren anhand ihrer Funktion klassifiziert. Zusätzlich wurde der vorgeschlagene Mechanismus der Phosphorylierung untersucht. Es konnte gezeigt werden, dass schon der erste Schritt des vorgeschlagenen Mechanismus, die Alkoholat-Bildung des Pyridoxals, unter physiologischen Bedingungen zu sofortiger Reprotonierung durch Wasser führt. Stattdessen wird die Protonierung des  $\gamma$ -Phosphats von ATP als erster Schritt vorgeschlagen, was auch experimentelle Beobachtung erklären kann. Durch den anschließenden Austausch der Koordinationssphäre von  $\text{Mg}^{2+}$  und nukleophilem Angriff der 5'-Alkoholgruppe am  $\gamma$ -Phosphat kommt es zur Bildung von Pyridoxalphosphat. Obwohl der Mechanismus noch nicht vollständig aufgeklärt ist, bildet er eine realistische Alternative zum bisher angenommenen Mechanismus der Pyridoxalkinase.

# Abstract

In this thesis, reaction mechanisms of enzymes and bioactive compounds are investigated using force-field molecular dynamics (FF-MD) simulations and quantum mechanics/molecular mechanics (QM/MM) calculations. The elucidation of enzymatic reaction mechanisms gives information on the underlying processes on atomistic level and can provide valuable input for drug development, like antibiotics or antitumor agents.

The first part of this thesis is the elucidation of the repair processes of oxidative DNA damages. DNA is constantly damaged by various influences, like mutagenic chemicals, radiation or reactive oxygen species (ROS). Usually, the cell has a fast and efficient repair machinery to prevent the accumulation of DNA damages, which would result in mutations and can lead to diseases like cancer, Alzheimer, Parkinson and Diabetes II. The enzyme Formamidopyrimidine-DNA glycosylase (Fpg) repairs the highly mutagenic oxidative purine lesions FapydG (2,6-diamino-4-hydroxy-5-formamido-pyrimidine), FapydA (4,6-diamino-5-formamidopyrimidine) and 8OG (7,8-dihydro-8-oxoguanine). We were able to show that the widely assumed base-protonated mechanism is energetically not feasible. Instead, we found a ribose-protonated mechanism, which is in full agreement with experimental observations. The base-independent reaction mechanism for *anti*-bound lesions can now explain the excision of a diversity of chemically different DNA damages by the same enzyme and the experimental finding that also nonpolar purine analogues can be excised. Additionally, this base-independent mechanism implicates that the discrimination of the DNA lesions occurs at the recognition step and not within the active site of Fpg.

In addition to the investigation of DNA repair processes, an analogue of the natural product duocarmycin, was studied in the second part of this thesis. Duocarmycins are potent antitumor agents that are known to bind DNA. Interestingly, analogues missing their DNA binding moiety, specifically target Aldehyde Dehydrogenase 1A1 (ALDH1A1), an isoform of aldehyde dehydrogenase. Structural, biochemical, kinetic and theoretical investigations revealed the details of the interaction pattern of the cytotoxic duocarmycin analogue seco drug 2 in the hydrophobic pocket of ALDH1A1 due to van der Waals interaction and  $\pi$ -stacking. Furthermore, it was found that seco drug 2 inhibits ALDH1A1 through covalently binding to the cysteine residue adjacent to the catalytic cysteine. This unique binding is not shared within other isoforms of ALDH and will enable functional studies of ALDH1A1 within whole proteomes.

In the third part of this thesis, the reactivity and selectivity of aromatic and aliphatic

amines reacting with  $\beta$ -lactones and  $\beta$ -lactams has been investigated, to establish a reactivity scale. It is known that aniline (as aromatic amine) attacks  $\beta$ -butyrolactone (BBL) at the C3 position, whereas benzylamine (as aliphatic amine) attacks at the C1 position. This ambident reactivity of  $\beta$ -lactones (C1 versus C3 attack) is now revealed and can be attributed to  $\text{H}_3\text{O}^+$  stabilization of the transition state, leading to a surmountable reaction barrier in case of aniline and an even lower barrier in case of benzylamine. Additionally, the observation that bicyclic  $\beta$ -lactams, a group of antibiotics, and monocyclic  $\beta$ -lactones can react with amino groups, whereas monocyclic  $\beta$ -lactams are only binding weakly to amines, was confirmed theoretically and experimentally due to the reaction barriers and kinetics. With these results, a reactivity scale was established.

The last part of this thesis focuses on the mechanism of pyridoxal kinase, an enzyme involved in vitamin B<sub>6</sub> metabolism. The active site of pyridoxal kinase could be a new target of antibiotics, instead of the mainly targeted cell wall biosynthesis. This could lead to a new structurally different class of antibiotics especially valuable in the treatment of multi-resistant bacterial strains. First, the importance of amino acids and their interaction pattern within the active site of pyridoxal kinase was investigated to classify the amino acids according to their function. Furthermore, the suggested reaction mechanism of the phosphorylation was studied and possible parts of a new reaction mechanism have been identified including different effects on the phosphorylation barrier. We showed that the suggested initial alcoholate formation of pyridoxal bound to Cys110 (CPL) leads to immediate re-protonation by water under physiological conditions, preventing the next reaction step. Instead, we suggest protonation of the  $\gamma$ -phosphate of ATP as initial step followed by changes in the coordination sphere of  $\text{Mg}^{2+}$  and nucleophilic attack of the 5'-alcohol of pyridoxal at the  $\gamma$ -phosphate of ATP. Although the mechanism is not yet fully elucidated, it provides a viable alternative to the previously assumed mechanism of pyridoxal kinase.



# Contents

<b>List of publications .....</b>	<b>0</b>
<b>I Introduction .....</b>	<b>2</b>
DNA repair of oxidative lesions .....	3
Aldehyde dehydrogenase 1A1 inhibition by duocarmycins .....	5
Selectivity and reactivity of $\beta$ -lactones and $\beta$ -lactames.....	5
Pyridoxal kinase.....	6
<b>II Theory .....</b>	<b>9</b>
Molecular Mechanics (MM) and Force Field Molecular Dynamics (FF-MD).....	9
Quantum Mechanics (QM).....	10
Quantum Mechanics/Molecular Mechanics (QM/MM).....	11
Calculating reactions.....	12
<b>III Conclusion and outlook.....</b>	<b>14</b>
<b>IV Bibliography.....</b>	<b>17</b>
<b>V Publications .....</b>	<b>24</b>
V.1 Paper [1]: "A Base-Independent Repair Mechanism for DNA Glycosylase - No Discrimination Within the Active Site", I. D. Blank, K. Sadeghian, C. Ochsenfeld, <i>Scientific Reports</i> 5, 10369 (2015). ....	24
V.2 Paper [2]: "DNA Glycosylase Repairs FapydA Base-Independently - Schiff Base in Equilibrium with Closed-Ribose Structure", T. Moeller, I. D. Blank, C. Ochsenfeld, (under revision).....	52
V.3 Paper [3]: "Ribose-protonated DNA base-excision repair: a combined theoretical and experimental study", K. Sadeghian, D. Flaig, I. D. Blank, S. Schneider, R. Strasser, D. Stathis, M. Winnacker, T. Carell and C. Ochsenfeld, <i>Angew. Chem. Int. Ed.</i> 53, 10044 (2014). ....	80
V.4 Paper [4]: "Structural, Biochemical, and Computational Studies Reveal the Mechanism of Selective Aldehyde Dehydrogenase 1A1 Inhibition by Cytotoxic Duocarmycin Analogues", M. F. Koch, S. Harteis, I. D. Blank, G. Pestel, L. F. Tietze, C. Ochsenfeld, S. Schneider, S. A. Sieber, <i>Angew. Chem. Int. Ed.</i> 54, 13550 (2015).....	
V.5 Paper [5]: "Kinetic and theoretical studies of $\beta$ -lactone reactivity - a quantitative scale for biological application", E. N. Wiedemann, F. A. Mandl, I. D. Blank, C. Ochsenfeld, A. R. Ofial, S. A. Sieber, <i>ChemPlusChem</i> , 80, 1673 (2015)...	186

V.6	Paper [6]: "Theoretical and experimental studies on <i>Staphylococcus aureus</i> pyridoxal kinase reveal a facile interaction network crucial for phosphorylation", I. D. Blank, V. Kirsch, S. A. Sieber, C. Ochsenfeld, (to be submitted).....	276
-----	--	-----

# List of publications

This work is a cumulative dissertation. The results of this work were so far published in four peer-reviewed journal papers, another project is under revision and one paper will be submitted soon. These six papers are considered the main part of this thesis. The following list specifies the contributions of the author in each of the publications in italics. All six publications and their corresponding supplementary information are collected in Chapter V of this work and are referred to in this work as paper [1]-[6].

**[1]** I. D. Blank, K. Sadeghian, C. Ochsenfeld, "A Base-Independent Repair Mechanism for DNA Glycosylase - No Discrimination Within the Active Site", *Scientific Reports* **5**, 10369 (2015).

Contribution by I. D. Blank: *All MD simulations and QM/MM calculations on the mechanism, the QM size convergence and most of the writing.*

**[2]** T. Moeller, I. D. Blank, C. Ochsenfeld, „DNA Glycosylase Repairs FapydA Base-Independently - Schiff Base in Equilibrium with Closed-Ribose Structure“, (under revision).

Contribution by I. D. Blank: *Supervision and partial writing.*

**[3]** K. Sadeghian, D. Flaig, I. D. Blank, S. Schneider, R. Strasser, D. Stathis, M. Winnacker, T. Carell and C. Ochsenfeld, "Ribose-protonated DNA base-excision repair: a combined theoretical and experimental study", *Angew. Chem. Int. Ed.* **53**, 10044 (2014).

Contribution by I. D. Blank: *Investigation of protonation states, mechanistic alternatives and review of the manuscript.*

**[4]** M. F. Koch, S. Harteis, I. D. Blank, G. Pestel, L. F. Tietze, C. Ochsenfeld, S. Schneider, S. A. Sieber, "Structural, Biochemical, and Computational Studies Reveal the Mechanism of Selective Aldehyde Dehydrogenase 1A1 Inhibition by Cytotoxic Duocarmycin Analogues", *Angew. Chem. Int. Ed.* **54**, 13550 (2015).

Contribution by I. D. Blank: *All MD simulations, analysis and partial writing.*

**[5]** E. N. Wiedemann, F. A. Mandl, I. D. Blank, C. Ochsenfeld, A. R. Ofial, S. A. Sieber, "Kinetic and theoretical studies of  $\beta$ -lactone reactivity - a quantitative scale for biological application", *ChemPlusChem*, **80**, 1673 (2015).

Contribution by I. D. Blank: *All calculations on the mechanism and writing of the corresponding parts.*

**[6]** I. D. Blank, V. Kirsch, S. A. Sieber, C. Ochsenfeld, "Theoretical and experimental studies on *Staphylococcus aureus* pyridoxal kinase reveal a facile interaction network crucial for phosphorylation", (to be submitted)

Contribution by I. D. Blank: *All MD simulations and QM/MM calculations and most of the writing.*

# Chapter I

## Introduction

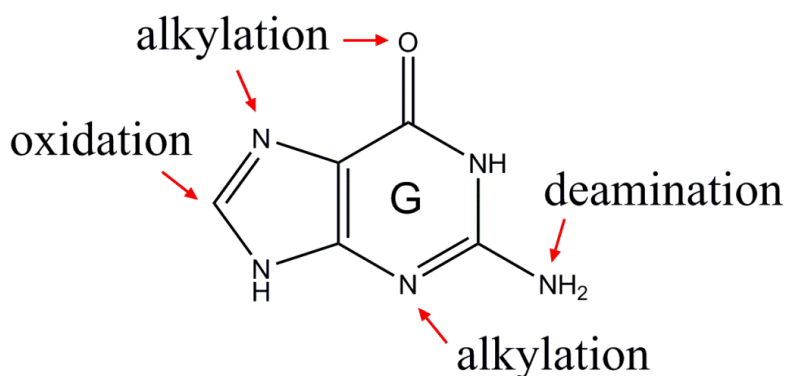
The investigation of enzymatic reactions is of utmost interest for drug development, because the understanding of the function of enzymes is often indispensable to develop specific inhibitors like antibiotics and drugs against a variety of diseases. The drug design process can be optimized to end up with very selective drugs that are specific for a single enzyme in the genome, ideally targeting only one out of many different isoforms. On the one hand, this allows for a better possibility to perform fundamental research on only one isoform, on the other hand, unintended side effects of drugs can be minimized.

Over the last decades, the number of high resolution protein structures deposited in the protein data bank (PDB) has grown enormously. This valuable structural information about enzymes, often in complexes with substrates, inhibitors or cofactors, gives hints concerning their functionality. Nevertheless, reaction mechanisms based on a series of X-ray or NMR structures of apo proteins and complexes suffer from missing information about the energetics of the reaction. Additionally, to crystallize reactive conformations, the reaction has to be stopped in the reactive conformation, which inevitably requires mutations, modifications of the substrate or cross-linking techniques that result in an altered system. The alteration is supposed to be negligible but a small structural difference can also influence the overall picture substantially. Theoretical approaches employ molecular dynamics (MD) simulations<sup>1,2</sup> starting from crystal structures to minimize these artifacts as well as possible artifacts introduced by the crystallization conditions, low temperature or neglected hydrogen atoms.

Starting from the resulting structures, reaction mechanisms can be calculated with QM/MM (quantum mechanics/molecular mechanics)<sup>3,4</sup> methods to study proposed and alternative mechanisms on atomistic level. In this way, it is possible to connect the static crystallographic geometries with the actual energetics and to gain an overall picture of the reaction. In this work, reaction mechanisms of different enzymes and bioactive compounds have been elucidated, as well as the interaction patterns and selectivity of drugs binding to a specific enzyme.

## DNA repair of oxidative lesions

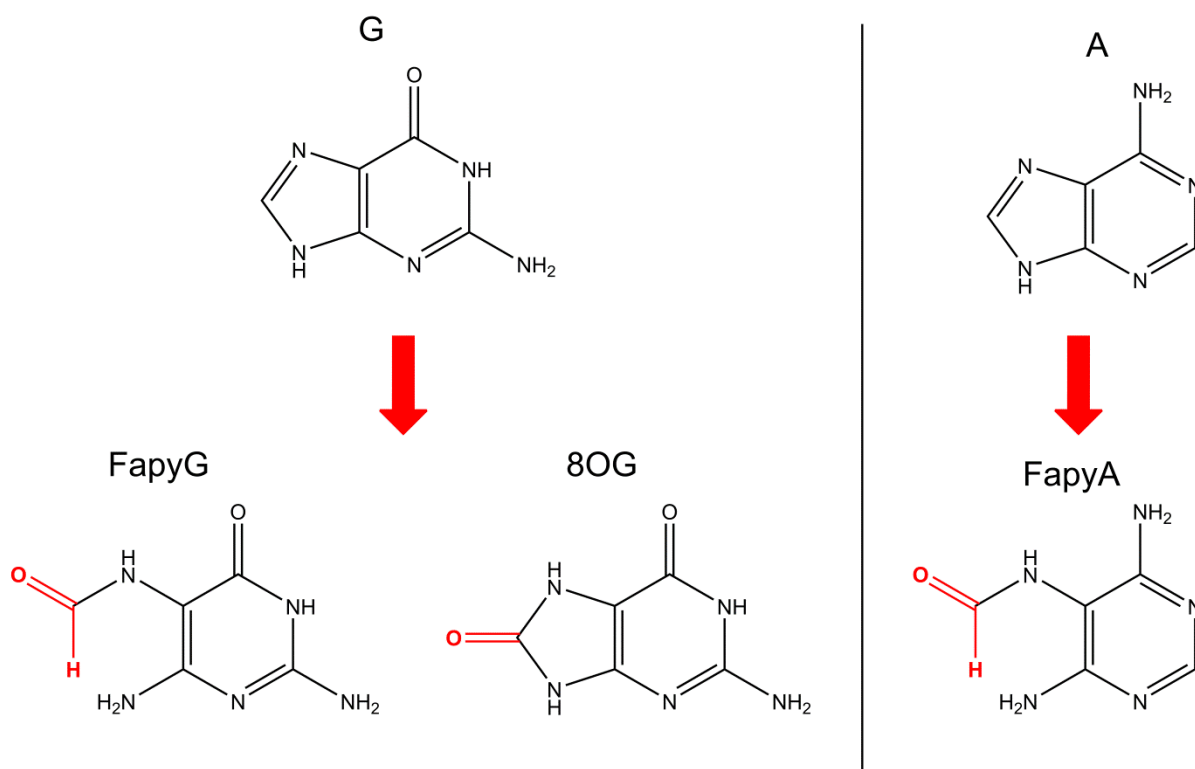
In the first part of this work, the reaction mechanism of a DNA repair enzyme that repairs oxidative damages has been investigated. DNA is permanently exposed to adverse conditions, like radiation, mutagenic chemicals or reactive oxygen species (ROS), which can lead to hydrolysis, alkylation or oxidation of DNA bases (Figure 1)<sup>5</sup>.



**Figure 1:** Different possible modifications of DNA bases, illustrated at the example of guanine.

Usually, the repair machinery of the cell ensures that DNA is repaired fast and efficiently preventing damage to the genome. Unrepaired DNA damages would lead to mutations. The accumulation of oxidative DNA damages has been associated with diseases like cancer<sup>6,7</sup>, Alzheimer<sup>8</sup>, Parkinson<sup>9</sup>, Diabetes II<sup>10</sup> and Multiple Sclerosis<sup>11</sup>. Despite the scope of an efficient repair process, the mechanistic details of the repair of oxidative DNA damages and their recognition remained scarcely explored. We investigated the repair mechanism of the chemically different oxidative purine damages, FapydG (2,6-diamino-4-hydroxy-5-formamido-pyrimidine), FapydA (4,6-diamino-5-formamidopyrimidine) and 8OG (7,8-dihydro-8-oxoguanine) (Figure 2). These most abundant oxidative DNA lesions are characterized by a high miscoding potential<sup>12</sup> leading to a high mutation frequency. Although 8OG is the best known lesion, the levels of the formamidopyrimidines FapydG and FapydA often exceed the levels of 8OG<sup>13,13</sup>, their repair efficiency is higher<sup>14</sup> and it is FapydG that has the highest mutation frequency among the oxidative damages<sup>14</sup>. Therefore, the repair mechanism of FapydG is investigated within this work.

The mentioned oxidative DNA damages are all repaired by the same enzyme, Formamidopyrimidine DNA glycosylase (Fpg)<sup>15</sup>, which is interesting to note, since the damages not only are chemically different, but also derive from different DNA bases. Additionally, the repair efficiency varies among the damages and also is dependent on the sequence context<sup>14</sup>. Hence, the molecular repair mechanism is of great interest to also gain better insights into the discrimination behavior of Fpg.



**Figure 2:** Oxidative DNA base damages deriving from guanine and adenine, investigated within this work.

The assumed overall repair process starts with Fpg sliding along a DNA strand and after a damaged base is recognized, the base is flipped into the active site of Fpg. As Fpg is known to be part of the base excision repair (BER) system<sup>15</sup>, it first excises the damaged nucleobase and in a subsequent step the ribose of the damaged nucleotide. This bifunctionality (glycosylase and AP lyase activity) has been found by enzyme activity assays<sup>15</sup> and was later validated by the X-ray structure of a Schiff base intermediate<sup>16</sup>, which is the intermediate structure between base and ribose excision. Since the damaged base is cleaved first, the widespread assumption was that the damaged base is protonated<sup>17</sup> in the initial step of the excision reaction to become a better leaving group. After base excision, the ribose ring is cleaved by hydrolysis at 5' and 3' position. The resulting gap in the backbone of the DNA strand, is then refilled by other enzymes<sup>18,19</sup>.

Based on the X-ray structure of Fpg containing FapydG [PDB-code: 1XC8]<sup>20</sup> or 8OG [PDB-code: 1R2Y]<sup>21</sup>, we calculated the reaction mechanism employing the QM/MM approach and found an alternative reaction mechanism with a ribose ring opening as the initial step instead of the previously assumed base protonation. These results are shown in section V.1 – V.3 and papers [1-3]. While the excision of the *syn*-bound 8OG involves protonation at O8 of the damaged base in the second step and subsequent deprotonation in the third step of the reaction (section V.3, paper [3]), we find a base-independent excision mechanism for *anti*-bound lesions like FapydG (section V.1, paper [1]) and FapydA (section V.2, paper [2]). This new insight can now explain the experimental findings that Fpg repairs a considerable amount of chemically different DNA damages. Even non-natural compounds like nonpolar

purine analogues<sup>22</sup> are excised by Fpg very likely in the same way. The excision mechanisms of a nonpolar purine analogue and *anti*-bound 8OG are currently under investigation in our group. The importance of the newly found mechanism and its implications to discrimination is discussed in paper [1 and 2].

## Aldehyde dehydrogenase 1A1 inhibition by duocarmycins

Duocarmycin natural products<sup>23</sup> bind to the narrow AT-rich minor groove of DNA and because of their remarkable cytotoxicity<sup>24,25</sup>, they are expected to be potential anticancer drugs. It was found, that analogues of duocarmycin lacking the indole moiety but still containing the spirocyclopropylcyclohexadienone moiety can also act as electrophiles within cells and specifically alkylate one isoform of aldehyde dehydrogenase, aldehyde dehydrogenase 1A1 (ALDH1A1)<sup>26,27</sup>.

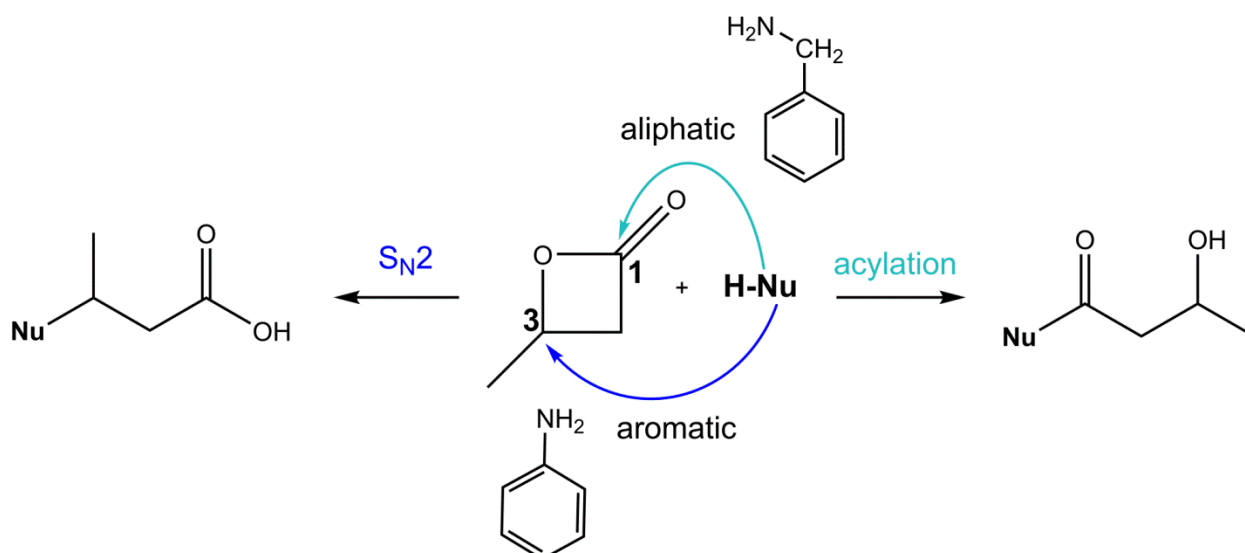
In this part of the work, the selectivity of the duocarmycin analogue, seco drug 2, towards ALDH1A1 is investigated. It turned out, that seco drug 2 covalently binds to a specific cysteine residue in the hydrophobic pocket, leading to inhibition of ALDH1A1. Investigation of the interaction between seco drug 2 and ALDH1A1 further revealed a distinct interaction pattern based on van der Waals interaction and  $\pi$ -stacking. A detailed discussion of the residues involved in stabilization of seco drug 2 is given in section V.4, paper [4]. These findings allow for a selective study of ALDH1A1 within the whole proteome, with the aim of better understanding metabolic pathways.

## Selectivity and reactivity of $\beta$ -lactones and $\beta$ -lactams

The third part of this thesis deals with the investigation of bioactive molecules and their reactivity. Amines are possible targets for antibiotics, that have lactones and lactams as building blocks<sup>28,29</sup>. Despite their related structure, lactones and lactams show different reactivity<sup>30</sup> and a reactivity scale would be very helpful to predict the reactivity of new antibiotics and increase the development efficiency of bioinspired covalent inhibitors.

The emphasis of this study is on the unexplained ambident reactivity of  $\beta$ -lactones towards primary amines (C1 versus C3 attack)<sup>31,32</sup>. It was unknown, why aliphatic and aromatic amines behave differently when nucleophilically attacking  $\beta$ -lactones. Aniline and benzylamine have been chosen as similar aromatic and aliphatic amines for the investigation of the reaction mechanisms with  $\beta$ -butyrolactone (BBL) as electrophile (Figure 3). The results have been validated by kinetics and product studies. The detailed investigation of these reaction mechanisms is part of section V.5 and paper [5].





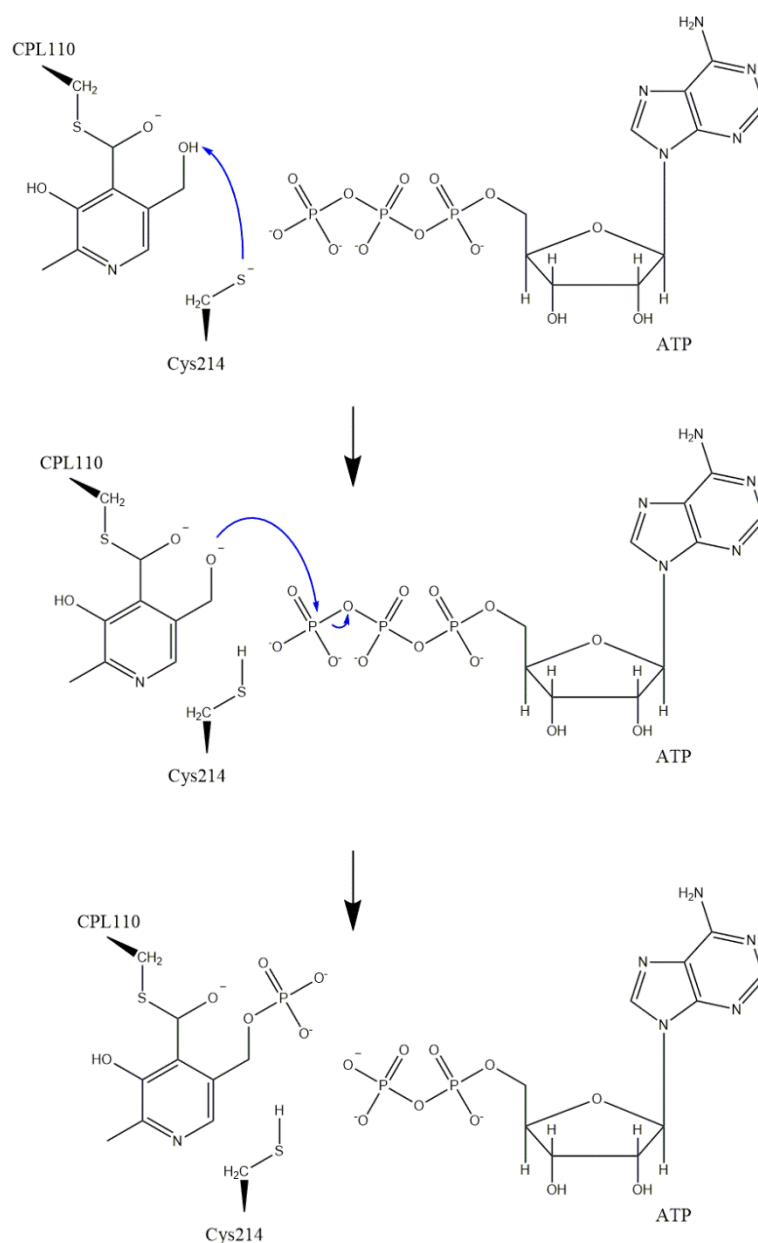
**Figure 3:** Reaction of  $\beta$ -butyrolactone (BBL) with aliphatic and aromatic amines.

## Pyridoxal kinase

Unfortunately, some bacteria, like Methicillin-Resistant *Staphylococcus aureus* (MRSA), have evolved resistance against several antibiotics (multidrug resistance), which renders their treatment very difficult. The resistance originates from the *mecA* gene, encoding<sup>33</sup> a modified penicillin-binding protein (PBP) with lower binding affinity for  $\beta$ -lactams in general. These PBPs are essential transpeptidases for cell wall synthesis. Most antibiotics are targeting the cell wall synthesis, when dealing with gram-positive bacteria like *Staphylococcus aureus*.

Facing multiresistant strains, other points of attack have to be considered to specifically intervene the bacterial metabolism and there is a high demand of completely new designed antibiotics that are targeting essential pathways for bacterial cells like vitamin biosynthesis. The aim is that the targeted strains cannot evolve resistance easily by adapting to a structurally similar drug because this would only delay the issue and soon lead to additional resistances.

In this study, the mechanistic details of pyridoxal kinase (PLK) are investigated. This enzyme synthesizes also pyridoxal phosphate (PLP) that is required by PLP-dependent enzymes, in many areas of the amino acid metabolism. Once the reaction mechanism is revealed, specific drugs can be developed to inhibit the active site of this enzyme. Mutational studies revealed that Cys110 is essential for PLP synthesis and that Cys214Asp mutants show higher turnover rates than the wild type<sup>34</sup>. By solving X-ray structures of PLK with and without bound substrates pyridoxal and ATP, it was found that Cys110 forms a hemithioacetal intermediate with pyridoxal, referred to as CPL in the following. Based on these findings, a mechanism has been postulated<sup>34</sup>. It was suggested, that after hemithioacetal formation and lid closure, Cys214 deprotonates the 5'-pyridoxal alcohol, which subsequently attacks the  $\gamma$ -phosphate of ATP. After this phosphorylation, PLP is released by the enzyme (Figure 4).



**Figure 4:** Postulated mechanism for phosphorylation of PL by PLK.

In section V.6 and paper [6], we investigate the interaction pattern between crucial amino acids in the active site of PLK by MD simulations to assess their potential influence on the reaction mechanism and correlated these results with their evolutionary conservation to illuminate their function within the enzyme. We classified the investigated amino acids according to their function:

- enzyme stability
- substrate binding
- phosphorylation mechanism

Additionally, we calculated the suggested mechanism and investigated possible parts of a new reaction mechanism, including different effects influencing the reaction barrier of the phosphorylation.

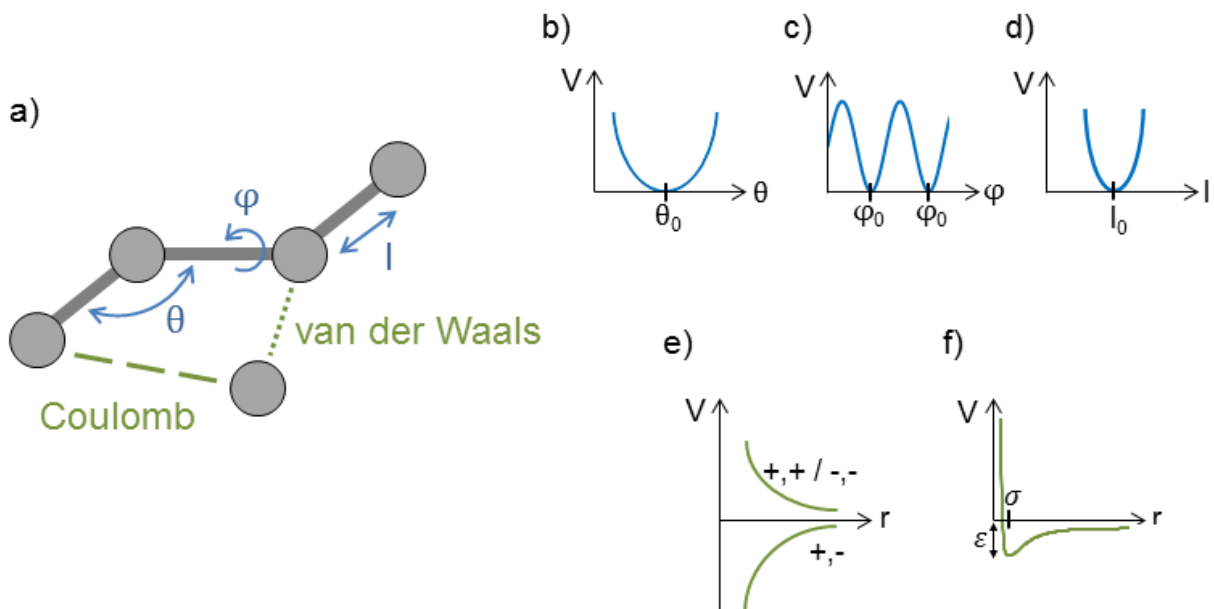
The following section contains a brief discussion of the methods applied in this work. It is not meant to give in detail information, available in many reviews ( see e.g. <sup>1,4,35</sup>), but to give an overview of the methods used in this thesis.

# Chapter II

## Theory

### Molecular Mechanics (MM) and Force Field Molecular Dynamics (FF-MD)

Although X-ray crystallography can provide three-dimensional pictures of large biomolecules with high resolution, these pictures are static and bring their own sources of artefacts due to mutations, chemical modifications of the substrate or cross-linking techniques necessary for crystallization. Therefore computational methods have become very important to further improve the obtained structures, to assess the stability of molecular systems or to understand dynamic processes. A common method to treat the dynamics of big biomolecules in solution within a reasonable time, are force fields.



**Figure 5:** a) Visualization of the force field terms (bonded interactions in blue, non-bonded interactions in green); potentials plotted for: b) angles, c) torsion angles, d) bond lengths, e) electrostatic interactions (Coulomb's law), f) van der Waals interactions (Lennard-Jones potential).

In force fields, the interaction of  $N$  atoms as a function of their position  $r_N$  is given by the potential  $V(r_1, \dots, r_N)$  of  $N$  interacting atoms as a function of the position  $r_i$ <sup>1,35</sup>.

$$U(r_1, \dots, r_N) = \sum_{bonds} \frac{a_i}{2} (l_i - l_{i0})^2 + \sum_{angles} \frac{b_i}{2} (\theta_i - \theta_{i0})^2 + \sum_{dihedrals} \frac{c_i}{2} (1 + \cos(n\varphi_i - \gamma_i)) + \sum_{i < j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i < j} k \frac{q_i q_j}{r_{ij}} \quad (1)$$

The potential consists of three terms describing bonded interactions and two terms describing non-bonded interactions (equation 1, Figure 5). The first two terms describe the deformation of the bond lengths  $l_i$  and angles  $\theta_i$  between bonds. Both are modeled in a harmonic form with the force constants  $a_i$  and  $b_i$  that ensure the correct chemical structure around the minima  $l_{i0}$  and  $\theta_{i0}$  and prevent bond breakage. The third term describes the dihedral angles, which are modeled using a sinusoidal function for a periodic description of staggered and eclipsed conformations. The fourth and fifth term represent the van der Waals interactions, modeled by the Lennard-Jones potential, and the electrostatic interactions by Coulomb's law. The time evolution of the system is described by the solution of Newton's equations of motion:

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \quad (2)$$

With  $r_i(t)$  as the position vector of the  $i^{\text{th}}$  atom and  $F_i$  as the force on the  $i^{\text{th}}$  atom with mass  $m_i$  at the time  $t$ . In an MD trajectory, the initial velocities are randomly generated according to the Boltzmann distribution at a given temperature, which are then propagated over time.

With molecular dynamics (MD) simulations interaction patterns and their statistics even in the millisecond time scale can be studied with a moderate usage of computational time. Interaction partners can be identified and different conformations e.g. substrate or inhibitor binding modes can be found.

Unfortunately, the classical force fields cannot describe bond breaking and formation and therefore a description of chemical reactions is impossible. In addition, electronic polarization is often ignored. To be able to study reactions, quantum mechanical methods are necessary.

## Quantum Mechanics (QM)

In quantum mechanical calculations, the electronic structure is described by wave functions with exclusive use of natural constants. The key component of quantum mechanics is the Schrödinger equation, which cannot be solved exactly for molecules containing more than one electron. Hence, approximations to the exact solution have been developed, leading to a hierarchy of wave function-based *ab-initio* methods<sup>36,37</sup>, with which one can achieve increasing accuracy but combined with arbitrary high computational costs.

The simplest *ab-initio* method in this hierarchy is the Hartree-Fock (HF) theory, in which each electron is subjected to the other electrons only by a mean field and

therefore neglecting the so-called electron correlation effect. Nevertheless, HF wave functions are the basis of post-HF methods, like Møller-Plesset perturbation theory<sup>36,38,39</sup> (e.g. MP2) or coupled cluster (CC)<sup>40-42</sup> theory, which show a considerable improvement in accuracy over HF calculations and allow to systematically approach the exact solution. Unfortunately, they are computational much more costly. The computational costs formally scale with the power of three of the system size for HF calculations, to the power of five for MP2 and even to the power of six and more for CC. This renders a description of systems containing a lot of atoms difficult or even impossible.

A cheaper possibility to calculate the electronic structure is based on the electron density. The basis of density functional theory (DFT) is the theorem of Hohenberg and Kohn<sup>43</sup>, which says that the electronic ground state energy can be described not only by a functional of the electronic wave function, but also as functional of the electron density. Unfortunately, the exact functional is unknown.

$$E[\rho] = E_{\text{NN}} + E_{\text{T}}[\rho] + E_{\text{V}}[\rho] + E_{\text{C}}[\rho] + E_{\text{X}}[\rho] + E_{\text{corr}}[\rho] \quad (3)$$

The energy of a system is a sum, containing terms for nuclear-nuclear repulsion ( $E_{\text{NN}}$ ), kinetic energy of the electrons ( $E_{\text{T}}$ ), nuclear-electron attraction ( $E_{\text{V}}$ ) and electron-electron Coulomb repulsion ( $E_{\text{C}}$ ) as well as the terms for the non-classical electron-electron exchange ( $E_{\text{X}}$ ) and correlation ( $E_{\text{corr}}$ ) energies. The exact values of  $E_{\text{X}}$ ,  $E_{\text{corr}}$  and  $E_{\text{T}}$  cannot be gained by HF theory. As the explicit functional of the electron density is unknown, approximations are used. Over the years, a very large number of different DFT functionals, varying in the approximation of  $E_{\text{XC}}$  ( $E_{\text{X}}+E_{\text{C}}$ ) has been developed, which are optimized for different purpose. In the following, the functionals used in this thesis are discussed shortly.

Generalized Gradient Approximation (GGA) functionals not only depend on the electron density, but also on the first derivation of the electron density. One of these type of functionals is BP86<sup>44,45</sup>, which is well suited for reaction energies and has relatively low computational cost. It is known, that the HOMO-LUMO gap decreases with increasing system size, which then leads to convergence problems<sup>46,47</sup>. This problem can be overcome by including HF exchange, like it is part of hybrid functionals, where the exchange-correlation term is composed of HF exchange and a pure DFT part. The most famous hybrid functional is B3LYP<sup>48,49</sup>, which contains 20% HF exchange. The highly parametrized M062X<sup>50</sup>, developed for ionic systems, contains 54% HF exchange. Including HF exchange in DFT functionals, results in an increase of computational costs but typically also increases the accuracy. Dispersion interactions are not or not well described by most of the DFT functionals. Therefore, an empiric dispersion correction developed by Grimme et al.<sup>51</sup> is used in the present work.

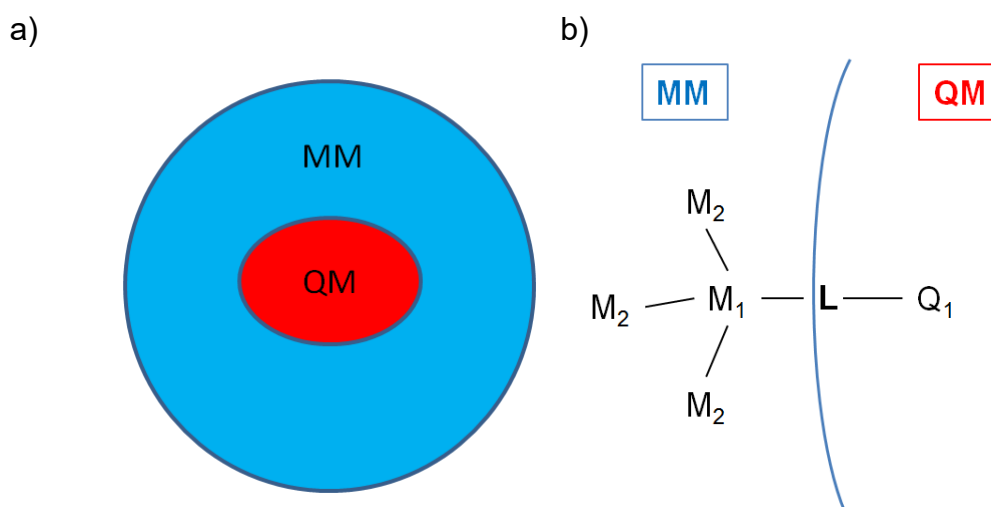
## Quantum Mechanics/Molecular Mechanics (QM/MM)

A purely quantum mechanical description is not yet feasible or extremely costly for the investigation of biosystems like whole proteins in solution. Therefore, a QM/MM

(Quantum Mechanics/Molecular Mechanics) approach<sup>3</sup> is employed and the system is divided into a region of interest that is treated at QM level to describe chemical reactions and the rest of the system that is described at MM level to take the interaction with the environment into account (see Figure 6). In this work, the additive QM/MM coupling scheme<sup>4,52</sup> is used. The total energy of the entire system is then given by

$$E_{QM/MM} = E_{MM} + E_{QM} + E_{QM-MM} \quad (4)$$

with the last term representing the explicit coupling between QM and MM region, which includes the bonded, van der Waals and electrostatic interactions.



**Figure 6:** a) schematic representation of the total system within the QM/MM approach. The system is divided into two subsystems, the QM and the MM region; b) the link atom approach for QM/MM border treatment.

We use electrostatic embedding<sup>53</sup>, which includes the MM point charges into the QM Hamiltonian as one-electron terms leading to polarization of the QM region through the surrounding MM point charges. The electrostatic interaction between QM and MM region is therefore treated at QM level. In many cases, the QM/MM border will cut covalent bonds, leading to dangling bonds in the QM region. In the QM calculation, these dangling bonds are saturated by an additional hydrogen atom as link atom (L), which is placed onto the bond between the QM atom  $Q_1$  and its neighboring MM atom  $M_1$  (see Figure 6). This leads to overpolarization of the link atom by the point charge of  $M_1$  that is therefore shifted and distributed to its neighboring MM atoms  $M_2$ . To compensate for the dipole moments created by the charge shift, additional dipoles with opposite sign are created close to the position of  $M_2$ <sup>54-56</sup>.

## Calculating reactions

During geometry optimization of a QM/MM described system, often including tens of thousands of atoms, the system is divided into a static part (often called frozen

region) and a part that is optimized (often called active region) surrounded by the static part. The active region should include not only the QM region, but also a surrounding MM sphere to avoid tension between the two subsystems. All other MM atoms are part of the frozen region. This region is supposed not to change during the reaction because of its distance to the active site, but the electrostatic interactions are still important.

To simulate reaction paths, non-equilibrium structures along the reaction coordinate are needed. In the adiabatic mapping approach, the bonds that are part of the reaction coordinate are successively increased or decreased by a constant amount. The energy profile is obtained by restrained optimizations along the chosen reaction coordinate with adiabatic changes of the environment. Often, the actual reaction coordinate cannot be properly described due to the complexity of the reaction, leading to jumps in the energy profile. In this case the reaction profile can be refined by the nudged elastic band (NEB) method<sup>57</sup>. It is especially helpful in cases with conformational changes like reorientation of an amino acid side chain. The NEB method is designed to find a minimal energy path between two geometries, by a set of images interpolating between the beginning and end of the reaction<sup>57</sup>. A spring force between neighboring images ensures that the images stay connected and evenly distributed along the reaction coordinate on the energy hypersurface, like an elastic band. During optimization of the band, the tangent to the path is evaluated at each image. The true force and the spring force are split into components parallel and perpendicular to the path. For the force projection (nudging) only the perpendicular component of the true force and the parallel component of the spring force is considered. The parallel component of the spring forces is responsible for the equal spacing between the images. Otherwise, following the true force along the path would result in an accumulation of images towards minima and high energy regions, like the transition state, would not be properly described. The transition state structure can be obtained by a climbing image<sup>58</sup> that is not affected by spring forces and follows the gradient along the band to a saddle point.

Another aspect to further improve the results obtained by QM/MM methods is to increase the QM region to get a better description, allowing polarization and to minimize the effects of the QM/MM border treatment. It has been shown that often big QM sizes are necessary to obtain a reliable description of the investigated reaction mechanism<sup>59,60,61,62</sup>.

To achieve size convergence, the QM region is systematically increased until there is no significant change in relative energies anymore. In paper **[1]**, we showed that it is sufficient to perform single point calculations to determine the necessary QM size. To get reliable QM size converged energies, optimization of the system is indispensable including the QM region, which was obtained through screening by single point calculations.



## Chapter III

### Conclusion and outlook

We introduced new base excision repair mechanisms for the enzyme Fpg that involve the oxidative DNA damages FapydG, FapydA, and 8OG. The repair mechanisms of these damages has in common, that the first reaction step is a ribose ring opening, initiated by proton abstraction by O4' from E2 while P1 nucleophilically attacks C1'. This is in contrast to the previously assumed direct glycosidic bond breakage. In difference to the *syn*-bound 8OG lesion, the *anti*-bound lesions FapydG and FapydA are excised by a base-independent reaction mechanism. This can explain the experimental findings that Fpg can also excise other chemically different DNA bases like 5-hydroxyuracil<sup>63</sup>, thymine glycol<sup>64</sup> and even nonpolar purine analogues<sup>22</sup>. The repair process of a nonpolar purine analogue and 8OG in *anti*-conformation is currently under investigation in our group and it also seems to follow the base-independent mechanism. As a result, it can be concluded that discrimination of the damaged bases has to occur before flipping the base into the active site, which implicates that discrimination is part of the recognition step. This new mechanism might also apply to related DNA repair processes, which could be investigated in future studies.

Our QM size convergence on FapydG shows that single point calculations are sufficient to determine the necessary QM region, although only the optimization of the system including the necessary QM region leads to a QM size converged energy profile.

Furthermore, the selective interaction of duocarmycin with ALDH1A1 has been elucidated. Structural, biochemical, kinetic and theoretical investigations revealed the details of the interaction pattern of duocarmycin with the hydrophobic pocket of ALDH1A1. It was found that the inhibitor duocarmycin is covalently bound to a cysteine adjacent to the catalytic cysteine. This is a unique binding, which is not shared within other isoforms of ALDH. Our findings are supposed to have a huge impact on drug design that targets DNA or a specific isoform of aldehyde dehydrogenase. This provides for the first time the opportunity to study the function of ALDH1A1 within whole proteomes, resulting in a better understanding of metabolic pathways. This approach shows that analyzing the interaction pattern of amino acids with a ligand and comparing the interaction pattern of wild type and mutants can explain the selectivity and interaction behavior of a drug. Hence, this approach can

be a valuable tool for drug design.

In the next part of this thesis, the different reactivity and selectivity in the reaction of aromatic and aliphatic amines with  $\beta$ -lactones has been investigated. QM/MM calculations of the reaction of aniline and benzylamine with BBL indicate a crucial role of  $\text{H}_3\text{O}^+$ , which has been confirmed experimentally. While the barrier of the C1 attack of benzylamine is lowered by the presence of  $\text{H}_3\text{O}^+$ , the reaction is also possible without. In the case of aniline, C3 attack is only possible in the presence of  $\text{H}_3\text{O}^+$ , which significantly stabilizes the transition state. Additionally, the reaction of an amino group with monocyclic  $\beta$ -lactams, bicyclic  $\beta$ -lactams (like penicillin G) and monocyclic  $\beta$ -lactones have been quantified by calculating the reaction barriers and measuring the kinetics. With this approach, the scope of application can be further expanded to other bioactive compounds.

In another project part of this thesis the phosphorylation reaction of pyridoxal kinase has been investigated. The interactions in the active site have been studied by MD simulations and correlated to experimental kinetic measurements of mutants. We were able to classify the amino acids into three categories, namely enzyme stability, substrate binding and orientation and participation in the phosphorylation mechanism. Furthermore, the reaction mechanism of the phosphorylation was investigated employing a QM/MM approach and different effects reducing the reaction barrier of the phosphorylation step have been suggested. Instead of the initial alcoholate formation of CPL<sup>34</sup>, which seems impossible under physiological conditions due to simultaneous re-protonation by water, we suggest protonation of the  $\gamma$ -phosphate of ATP by Cys214 as initial step. This can explain the higher activity of the C214D mutant compared to the wild type PLK<sup>34</sup>. In the second step a substitution in the solvation sphere of  $\text{Mg}^{2+}$  can occur, where the  $\gamma$ -phosphate is replaced by a water molecule. Lys179 can stabilize the  $\beta$ -phosphate of ATP by formation of a strong hydrogen bond, further lowering the phosphorylation barrier by 3 kcal/mol. The resulting barrier of 42 kcal/mol indicates, that not all parts of the mechanism are revealed yet, probably due to the uncertainty of the loop. Nevertheless, the reaction mechanism is a viable alternative to the previously proposed one<sup>34</sup>, because it is analogous to the reaction mechanism found for another kinase<sup>65</sup>. Comparison between theoretical and experimental observations will help to understand the orientation and interaction pattern of the unresolved but crucial loop residues in the X-ray structure of PLK in complex with ATP and pyridoxal. With detailed information about the loop conformation in the closed state, we could further improve the full reaction mechanism of pyridoxal kinase, which will help to develop a new class of antibiotics consisting of selective, active site targeting inhibitors for pyridoxal kinase to treat methicillin resistant strains of *Staphylococcus aureus*.

This work shows that theoretical investigations can provide new insights into biochemical problems. With quantum-chemical calculations the feasibility of suggested reaction mechanisms can be evaluated based on their energetics and new mechanisms can be found. The combination of experimental and theoretical

results leads to a deeper understanding of the underlying processes. Nevertheless, a special challenge of calculations on biomolecular systems is their big size. With the advent of GPU accelerated computing, even more challenging calculations on larger systems can be performed and the scope of application will be further extended.

# Chapter IV

## Bibliography

- 1 Durrant, J. D., McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biology* **9**, 1-9 (2011).
- 2 Hospital, A., Goñi, J. R., Orozco, M., Gelpí, J. L. Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry : AABC* **8**, 37-47 (2015).
- 3 Warshel, A. Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). *Angewandte Chemie International Edition* **53**, 10020-10031 (2014).
- 4 Senn, H. M., Thiel, W. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition* **48**, 1198-1229 (2009).
- 5 Bauer, N. C., Corbett, A. H., Doetsch, P. W. The current state of eukaryotic DNA base damage and repair. *Nucleic Acids Research* **43**, 10083-10101 (2015).
- 6 Cooke, M. S., Evans, M. D., Dizdaroglu, M., Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *The FASEB Journal* **17**, 1195-1214 (2003).
- 7 Evans, M. D., Dizdaroglu, M., Cooke, M. S. Oxidative DNA damage and disease: induction, repair and significance. *Mutation Research/Reviews in Mutation Research* **567**, 1-61 (2004).
- 8 Bradley-Whitman, M. A., Timmons, M. D., Beckett, T. L., Murphy, M. P., Lynn, B. C., Lovell, M. A. Nucleic Acid Oxidation: An early feature of Alzheimer's disease. *Journal of Neurochemistry* **128**, 294-304 (2014).
- 9 Alam, Z. I., Jenner, A., Daniel, S. E., Lees, A. J., Cairns, N., Marsden, C. D., Jenner, P., Halliwell, B. Oxidative DNA Damage in the Parkinsonian Brain: An Apparent Selective Increase in 8-Hydroxyguanine Levels in Substantia Nigra. *Journal of Neurochemistry* **69**, 1196-1203 (1997).

- 10 Rehman, A., Nourooz-Zadeh, J., Möller, W., Tritschler, H., Pereira, P., Halliwell, B. Increased oxidative damage to all DNA bases in patients with type II diabetes mellitus. *FEBS Letters* **448**, 120-122 (1999).
- 11 Vladimirova, O., O'Connor, J., Cahill, A., Alder, H., Butunoi, C., Kalman, B. Oxidative damage to DNA in plaques of MS brains. *Multiple Sclerosis* **4**, 413-418 (1998).
- 12 Tudek, B. Imidazole ring-opened DNA purines and their biological significance. *Journal of Biochemistry and Molecular Biology* **36**, 12-19 (2003).
- 13 Hu, J., de Souza-Pinto, N. C., Haraguchi, K., Hogue, B. A., Jaruga, P., Greenberg, M. M., Dizdaroglu, M., Bohr, V. A. Repair of Formamidopyrimidines in DNA Involves Different Glycosylases: Role of the OGG1, NTH1, and NEIL1 enzymes. *Journal of Biological Chemistry* **280**, 40544-40551 (2005).
- 14 Greenberg, M. M. The Formamidopyrimidines: Purine Lesions Formed in Competition With 8-Oxopurines From Oxidative Stress. *Accounts of Chemical Research* **45**, 588-597 (2012).
- 15 O'Connor, T. R., Laval, J. Physical association of the 2,6-diamino-4-hydroxy-5N-formamidopyrimidine-DNA glycosylase of Escherichia coli and an activity nicking DNA at apurinic/apyrimidinic sites. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 5222-5226 (1989).
- 16 Fromme, J. C., Verdine, G. L. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nature Structural & Molecular Biology* **9**, 544-552 (2002).
- 17 Zharkov, D. O., Rieger, R. A., Iden, C. R., Grollman, A. P. NH<sub>2</sub>-terminal Proline Acts as a Nucleophile in the Glycosylase/AP-Lyase Reaction Catalyzed by Escherichia coli Formamidopyrimidine-DNA Glycosylase (Fpg) Protein. *Journal of Biological Chemistry* **272**, 5335-5341 (1997).
- 18 McCullough, A. K., Dodson, M. L., Lloyd, R. S. Initiation of Base Excision Repair: Glycosylase Mechanisms and Structures. *Annual review of biochemistry* **68**, 255-285 (1999).
- 19 Stivers, J. T., Jiang, Y. L. A Mechanistic Perspective on the Chemistry of DNA Repair Glycosylases. *Chemical Reviews* **103**, 2729-2760 (2003).
- 20 Coste, F., Ober, M., Carell, T., Boiteux, S., Zelwer, C., Castaing, B. Structural Basis for the Recognition of the FapydG Lesion (2,6-Diamino-4-hydroxy-5-formamidopyrimidine) by Formamidopyrimidine-DNA Glycosylase. *Journal of*

- Biological Chemistry* **279**, 44074-44083 (2004).
- 21 Fromme, J. C., Verdine, G. L. DNA Lesion Recognition by the Bacterial Repair Enzyme MutM. *Journal of Biological Chemistry* **278**, 51543-51548 (2003).
- 22 McKibbin, P. L., Kobori, A., Taniguchi, Y., Kool, E. T., David, S. S. Surprising Repair Activities of Nonpolar Analogs of 8-oxoG Expose Features of Recognition and Catalysis by Base Excision Repair Glycosylases. *Journal of the American Chemical Society* **134**, 1653-1661 (2012).
- 23 T. Yasuzawa, Y. S., M. Ichimura, I. Takahashi, H. Sano. Structure of duocarmycin SA, a potent antitumor antibiotic. *Journal of Antibiotics* **44**, 445–447 (1991).
- 24 Boger, D. L. Duocarmycins: a new class of sequence selective DNA minor groove alkylating agents. *Chemtracts: Organic Chemistry* **4** 329–349 (1991).
- 25 Boger, D. L., Johnson, D. S. CC-1065 and the Duocarmycins: Understanding their Biological Function through Mechanistic Studies. *Angewandte Chemie International Edition in English* **35**, 1438-1474 (1996).
- 26 Wirth, T., Schmuck, K., Tietze, L. F., Sieber, S. A. Duocarmycin Analogues Target Aldehyde Dehydrogenase 1 in Lung Cancer Cells. *Angewandte Chemie International Edition* **51**, 2874-2877 (2012).
- 27 Wirth, T., Pestel, G. F., Ganai, V., Kirmeier, T., Schuberth, I., Rein, T., Tietze, L. F., Sieber, S. A. The Two Faces of Potent Antitumor Duocarmycin-Based Drugs: A Structural Dissection Reveals Disparate Motifs for DNA versus Aldehyde Dehydrogenase 1 Affinity. *Angewandte Chemie International Edition* **52**, 6921-6925 (2013).
- 28 Waxman, D. J., Strominger, J. L. Penicillin-Binding Proteins and the Mechanism of Action of Beta-Lactam Antibiotics<sup>1</sup>. *Annual review of biochemistry* **52**, 825-869 (1983).
- 29 Staub, I., Sieber, S. A.  $\beta$ -Lactams as Selective Chemical Probes for the in Vivo Labeling of Bacterial Enzymes Involved in Cell Wall Biosynthesis, Antibiotic Resistance, and Virulence. *Journal of the American Chemical Society* **130**, 13400-13409 (2008).
- 30 Bottcher, T., Sieber, S. A. [small beta]-Lactams and [small beta]-lactones as activity-based probes in chemical biology. *MedChemComm* **3**, 408-417 (2012).
- 31 Manso, J. A., Pérez-Prior, M. T., García-Santos, M. d. P., Calle, E., Casado, J. A Kinetic Approach to the Alkylating Potential of Carcinogenic Lactones.

- Chemical Research in Toxicology* **18**, 1161-1166 (2005).
- 32 Griesbeck, A., Seebach, D. Herstellung enantiomerenreiner Derivate von 3-Amino- und 3-Mercaptobuttersäure durch SN2-Ringöffnung des  $\beta$ -Lactons und eines 1,3-Dioxanons aus der 3-Hydroxybuttersäure. *Helvetica Chimica Acta* **70**, 1326-1332 (1987).
  - 33 Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H., Spratt, B. G. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proceedings of the National Academy of Sciences* **99**, 7687-7692 (2002).
  - 34 Nodwell, M. B., Koch, M. F., Alte, F., Schneider, S., Sieber, S. A. A Subfamily of Bacterial Ribokinases Utilizes a Hemithioacetal for Pyridoxal Phosphate Salvage. *Journal of the American Chemical Society* **136**, 4992-4999 (2014).
  - 35 Petrenko, R., Meller, J. in *eLS* (John Wiley & Sons, Ltd, 2001).
  - 36 Szabo, A., Ostlund, N. S. *Modern Quantum Chemistry*. (1996).
  - 37 Helgaker, T., P.Jørgensen, Olsen, J. *Molecular Electronic-Structure Theory*. (Wiley-VCH, 2000).
  - 38 Møller, C., Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **46**, 618-622 (1934).
  - 39 Cremer, D. Møller–Plesset perturbation theory: from small molecule methods to methods for thousands of atoms. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 509-530 (2011).
  - 40 Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *The Journal of Chemical Physics* **45**, 4256-4266 (1966).
  - 41 Čížek, J. in *Advances in Chemical Physics* 35-89 (John Wiley & Sons, Inc., 1968).
  - 42 Noga, J., Bartlett, R. J. The full CCSDT model for molecular electronic structure. *The Journal of Chemical Physics* **86**, 7041-7050 (1987).
  - 43 Hohenberg, P., Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **136**, B864-B871 (1964).
  - 44 Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review. A* **38**, 3098-3100 (1988).
  - 45 Perdew, J. P. Density-functional approximation for the correlation energy of the

- inhomogeneous electron gas. *Physical review. B, Condensed matter* **33**, 8822-8824 (1986).
- 46 Rudberg, E., Rubensson, E. H., Sałek, P. Kohn–Sham Density Functional Theory Electronic Structure Calculations with Linearly Scaling Computational Time and Memory Usage. *Journal of Chemical Theory and Computation* **7**, 340-350 (2011).
- 47 Greg, L., Daniel, J. C., Nicholas, D. M. H., Peter, D. H., Mike, C. P. Electrostatic considerations affecting the calculated HOMO–LUMO gap in protein molecules. *Journal of Physics: Condensed Matter* **25**, 152101 (2013).
- 48 Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **98**, 5648-5652 (1993).
- 49 Stephens, P. J., Devlin, F. J., Chabalowski, C. F., Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **98**, 11623-11627 (1994).
- 50 Zhao, Y., Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **120**, 215-241 (2008).
- 51 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* **132**, 154104 (2010).
- 52 Lin, H., Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts* **117**, 185-199 (2006).
- 53 Bakowies, D., Thiel, W. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *The Journal of Physical Chemistry* **100**, 10580-10594 (1996).
- 54 Sherwood, P., de Vries, A. H., Guest, M. F., Schreckenbach, G., Catlow, C. R. A., French, S. A., Sokol, A. A., Bromley, S. T., Thiel, W., Turner, A. J., Billeter, S., Terstegen, F., Thiel, S., Kendrick, J., Rogers, S. C., Casci, J., Watson, M., King, F., Karlsen, E., Sjøvoll, M., Fahmi, A., Schäfer, A., Lennartz, C. QUASI: A general purpose implementation of the QM/MM approach and its application to



- problems in catalysis. *Journal of Molecular Structure: THEOCHEM* **632**, 1-28 (2003).
- 55 Sherwood, P., H. de Vries, A., J. Collins, S., P. Greatbanks, S., A. Burton, N., A. Vincent, M., H. Hillier, I. Computer simulation of zeolite structure and reactivity using embedded cluster methods. *Faraday Discussions* **106**, 79-92 (1997).
  - 56 de Vries, A. H., Sherwood, P., Collins, S. J., Rigby, A. M., Rigutto, M., Kramer, G. J. Zeolite Structure and Reactivity by Combined Quantum-Chemical-Classical Calculations. *The Journal of Physical Chemistry B* **103**, 6133-6141 (1999).
  - 57 Henkelman, G., Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *The Journal of Chemical Physics* **113**, 9978-9985 (2000).
  - 58 Henkelman, G., Uberuaga, B. P., Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics* **113**, 9901-9904 (2000).
  - 59 Sumowski, C. V., Schmitt, B. B. T., Schweizer, S., Ochsenfeld, C. Quantum-Chemical and Combined Quantum-Chemical/Molecular-Mechanical Studies on the Stabilization of a Twin Arginine Pair in Adenovirus Ad11. *Angewandte Chemie International Edition* **49**, 9951-9955 (2010).
  - 60 Sumowski, C. V., Ochsenfeld, C. A Convergence Study of QM/MM Isomerization Energies with the Selected Size of the QM Region for Peptidic Systems. *The Journal of Physical Chemistry A* **113**, 11734-11741 (2009).
  - 61 Flaig, D., Beer, M., Ochsenfeld, C. Convergence of Electronic Structure with the Size of the QM Region: Example of QM/MM NMR Shieldings. *Journal of Chemical Theory and Computation* **8**, 2260-2271 (2012).
  - 62 Sumner, S., Söderhjelm, P., Ryde, U. Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins. *Journal of Chemical Theory and Computation* **9**, 4205-4214 (2013).
  - 63 Hatahet, Z., Kow, Y. W., Purmal, A. A., Cunningham, R. P., Wallace, S. S. New substrates for old enzymes. 5-Hydroxy-2'-deoxycytidine and 5-hydroxy-2'-deoxyuridine are substrates for Escherichia coli endonuclease III and formamidopyrimidine DNA N-glycosylase, while 5-hydroxy-2'-deoxyuridine is a substrate for uracil DNA N-glycosylase. *Journal of Biological Chemistry* **269**,

- 18814-18820 (1994).
- 64 Jurado, J., Sapparbaev, M., Matray, T. J., Greenberg, M. M., Laval, J. The Ring Fragmentation Product of Thymidine C5-Hydrate When Present in DNA Is Repaired by the Escherichia coli Fpg and Nth Proteins. *Biochemistry* **37**, 7757-7763 (1998).
- 65 Shi, T., Lu, Y., Liu, X., Chen, Y., Jiang, H., Zhang, J. Mechanism for the Autophosphorylation of CheA Histidine Kinase: QM/MM Calculations. *The Journal of Physical Chemistry B* **115**, 11895-11901 (2011).

# Chapter V

## Publications

- V.1 Paper [1]: "A Base-Independent Repair Mechanism for DNA Glycosylase - No Discrimination Within the Active Site", I. D. Blank, K. Sadeghian, C. Ochsenfeld, *Scientific Reports* 5, 10369 (2015).



# SCIENTIFIC REPORTS

OPEN

## A Base-Independent Repair Mechanism for DNA Glycosylase—No Discrimination Within the Active Site

Received: 09 December 2014

Accepted: 09 April 2015

Published: 27 May 2015

Iris D. Blank<sup>1,2</sup>, Keyarash Sadeghian<sup>1,2</sup> & Christian Ochsenfeld<sup>1,2</sup>

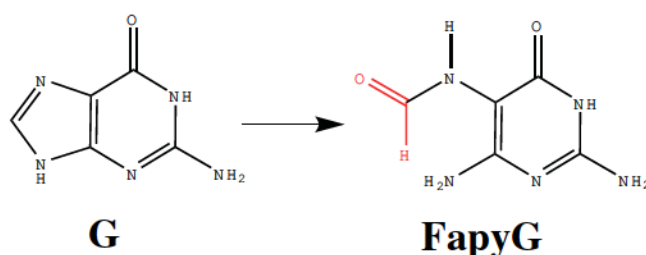
The ubiquitous occurrence of DNA damages renders its repair machinery a crucial requirement for the genomic stability and the survival of living organisms. Deficiencies in DNA repair can lead to carcinogenesis, Alzheimer, or Diabetes II, where increased amounts of oxidized DNA bases have been found in patients. Despite the highest mutation frequency among oxidized DNA bases, the base-excision repair process of oxidized and ring-opened guanine, FapydG (2,6-diamino-4-hydroxy-5-formamidopyrimidine), remained unclear since it is difficult to study experimentally. We use newly-developed linear-scaling quantum-chemical methods (QM) allowing us to include up to 700 QM-atoms and achieving size convergence. Instead of the widely assumed base-protonated pathway we find a ribose-protonated repair mechanism which explains experimental observations and shows strong evidence for a base-independent repair process. Our results also imply that discrimination must occur during recognition, prior to the binding within the active site.

DNA is constantly damaged by radiation, mutagenic chemicals, and reactive oxygen species, which leads to alkylation, hydrolysis, or oxidation of DNA bases. Therefore, the ability of cells to cope with DNA damages is crucial for their survival. Without effective repair machinery these damages can accumulate and soon affect the genomic stability, since they show a higher tendency for mismatches during replication. An increased level of oxidized DNA bases has been found in patients suffering from diseases such as Alzheimer<sup>1</sup>, Parkinson<sup>2</sup>, Multiple Sclerosis<sup>3</sup>, and Diabetes II<sup>4</sup>. Despite this importance, the reaction mechanism and, in particular, the discrimination within the repair process remained unclear, which is the focus of our present work.

We unravel the molecular repair mechanism for the case of the oxidative damage, FapydG (2,6-diamino-4-hydroxy-5-formamido-pyrimidine), (Fig. 1) that has the highest mutation frequency<sup>5</sup> of oxidative damages. The mechanism has also important implications for discrimination between damaged and undamaged bases. FapydG is repaired by the base excision repair enzyme Fpg (Formamidopyrimidine-DNA glycosylase, also known as MutM)<sup>6</sup>. It is assumed that Fpg slides along the DNA until it recognizes a damage, which is then flipped into the active site. For Fpg a Schiff base intermediate between the excision of the base and the ribose has been found [PDB-code: 1L1Z<sup>7</sup>], revealing bifunctionality (glycosylase and AP lyase activity)<sup>6</sup>. This means, that Fpg excises the base and the ribose successively. Since the base is excised first, base-protonation has been discussed as initial step, however, no strong evidence for such a process was provided so far. Once the damaged nucleotide has been excised, the resulting gap is going to be filled with the correct nucleotide by additional enzymes<sup>8,9</sup>.

In order to provide reliable insights into the repair of FapydG, computational studies are expected to be helpful, since direct experimental insights are highly difficult to obtain. There is only one crystal

<sup>1</sup>Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany. <sup>2</sup>Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany. Correspondence and requests for materials should be addressed to C.O. (email: christian.ochsenfeld@uni-muenchen.de)



**Figure 1.** Oxidation of guanine to FapyG (2,6-diamino-4-hydroxy-5-formamido-pyrimidine). We distinguish FapydG as the nucleotide and FapyG as the damaged base.

structure of the trapped educt state available [PDB-code: 1XC8]<sup>10</sup>. So far the proposed mechanism is only based on assumptions and no alternatives to a base-activated process have been explored. In this work, we employ theoretical methods, including quantum-chemical methods within a QM/MM approach, starting from the X-ray structure<sup>10</sup> in order to illuminate the overall cleavage reaction of FapyG. Since earlier studies have shown that often large QM spheres are necessary for a reliable theoretical description<sup>11–15</sup>, we converged the QM size using up to 700 QM atoms with linear-scaling SCF methods<sup>16–19</sup>.

## Methods

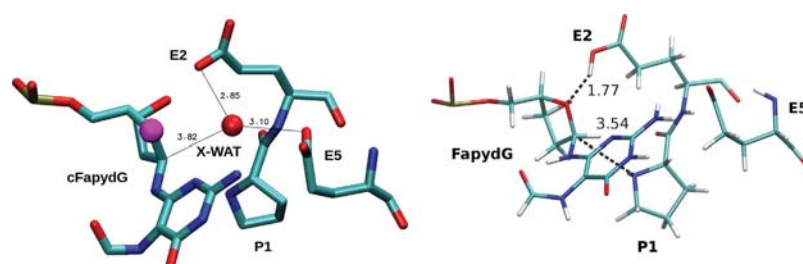
The crystal structure of wild-type Fpg in complex with a double strand DNA-fragment containing carbocyclic FapydG (cFapydG) from *Lactococcus lactis* was used as the starting structure [PDB code: 1XC8]. XLEAP (AmberTool)<sup>20</sup> has been used to add hydrogen atoms to the X-ray structure, to neutralize the system with sodium ions and to solvate it in a box of explicit TIP3P water<sup>21</sup> with a buffer of 10 Å around the solute. The parameters for the neutral proline residue and FapydG were taken from Perlow-Poehnelt *et al.*<sup>22</sup> and Song *et al.*<sup>23</sup>, respectively. We used ANTECHAMBER<sup>24,25</sup> to parametrize cFapydG. For force field molecular dynamics (FF-MD) simulations we used the NAMD engine<sup>26</sup> with Amber10 force field parameters<sup>20</sup>. Periodic boundary conditions and particle mesh Ewald summation (PME) with a cutoff value of 10 Å were employed (see SI-6.1). For QM/MM structure optimizations the DL-POLY implementation within ChemShell<sup>27</sup> (AMBER-FF) was combined with density functional theory (DFT) at the BP86-D3/6-31G\*\*<sup>28–32</sup> level of theory (unless specified otherwise) employing the Q-Chem program package<sup>33</sup> for the QM part. BP86-D3 was chosen for optimization due to its particular low weighted total mean absolute derivation for reaction energies (3.5 kcal/mol)<sup>34</sup> and its relatively low computational cost. The repair mechanism was calculated using both the adiabatic mapping approach and the nudged elastic band method of the DL-FIND<sup>35</sup> module implemented in ChemShell<sup>27</sup> (for system sizes see SI-6.2). The QM region was successively increased up to 700 atoms (see SI-5.1 for QM and relaxed regions). In SI-1 the influence of basis set and DFT-functional variation is shown.

## Results and Discussion

Starting point for our quantum-chemical study of the repair mechanism is the X-ray structure of Fpg in complex with cFapydG. It is important to note that the X-ray structures available for Fpg in complex with DNA not only differ in the damaged base and in the modifications necessary for trapping the educt state in the experiment, but also in the presence of a water molecule in the active site<sup>10,36,37</sup>. Therefore these influences together with the possible protonation states (not available from X-ray) are first systematically discussed in the following in order to obtain a realistic starting point for simulating the complex repair process: we start with the crystal structure of Fpg containing cFapydG in complex with DNA [PDB-code: 1XC8]<sup>10</sup>, discuss the influence of the carbon analogue, the proper protonation state, and then turn towards elucidating the reaction mechanism.

**Influence of the carbon analogue.** The only X-ray structure of FapydG<sup>10</sup> [PDB-Code: 1XC8] shows the educt state, where cFapydG is turned out of the DNA and placed into the active site of Fpg (Fig. 2). To allow crystallization of this reactive state, O<sub>4'</sub> of the ribose in FapydG has been substituted by a carbon atom, which is a strong hint, that the interaction between O<sub>4'</sub> and the active site is crucial for the reaction *in vivo*. Within the active site of this structure, a water molecule (X-WAT) has been observed next to the modified 4'-position. For the oxidative guanine damage 8OG (7,8-dihydro-8-oxoguanine), there are X-ray structures available with and without a carbon analogue<sup>36,37</sup> (see SI-2). These structures also differ in the presence of the water molecule. To analyze this difference, we investigate the behavior of the water molecule. We performed FF-MD simulations for the systems containing FapydG with/without X-WAT and cFapydG with/without X-WAT (SI-3.1). In both systems, FapydG and cFapydG with X-WAT, respectively, the presence of X-WAT destabilizes the active site (RMSD plots see SI-3.2) and it is very likely that X-WAT moves out into the solvent. In the combination FapydG with X-WAT, interaction between O<sub>4'</sub> and the protonated E2 of Fpg cannot be observed. In contrast, the system without X-WAT shows multiple events of E2-O<sub>4'</sub> interaction (SI-3.3). This interaction is crucial for our proposed base-independent





**Figure 2.** Left: Active site of the X-ray structure of Fpg in complex with cFapydG [PDB code: 1XC8] showing distances for water stabilization. The atomic position of the O → C substitution, which enabled this structure, is highlighted in magenta. Right: Protonation state of the active site of Fpg without X-WAT. E2 is in a protonated form, whereas E5 is not protonated; P1 is neutral.

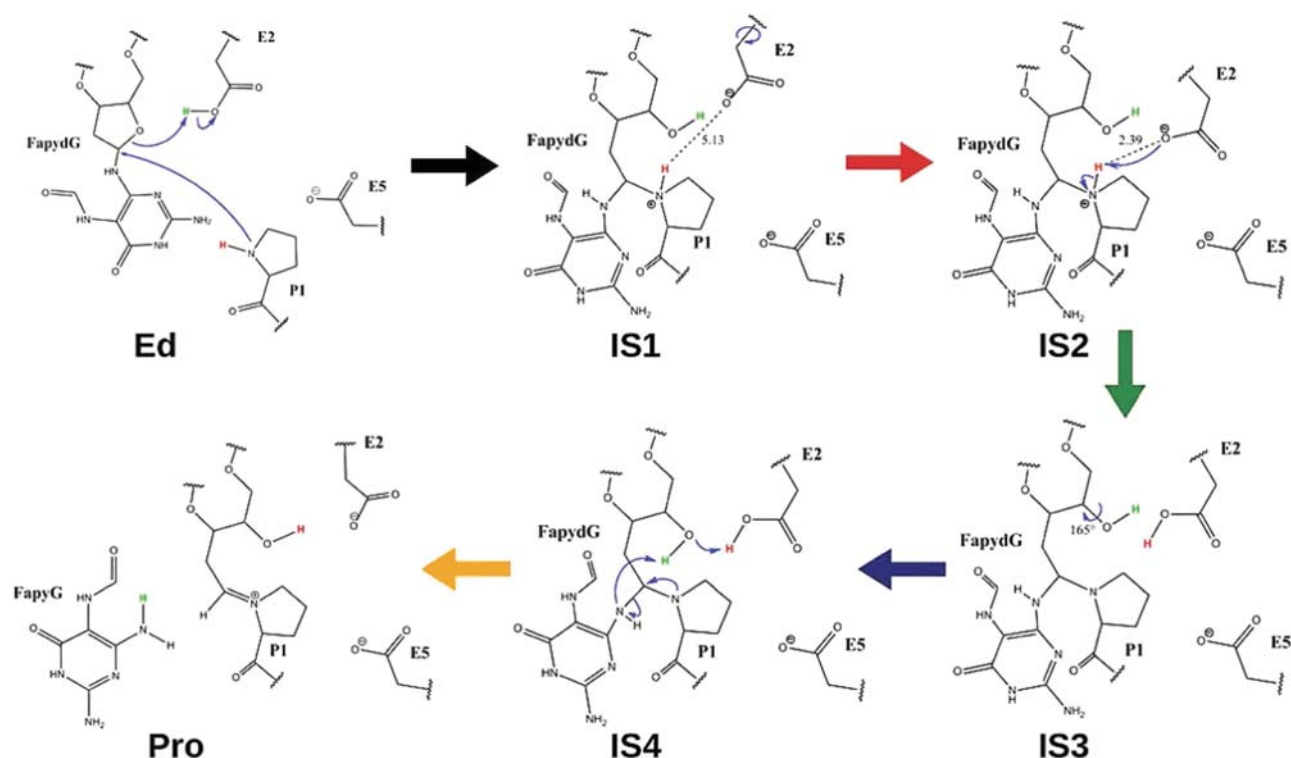
mechanism (see Fig. 3). If it is missing, the mechanism leads to a dead end (see Fig. 4) and can explain, why the carbon analogue allows the crystallization of this reactive state.

Overall, we conclude that due to the substitution of O<sub>4'</sub> to a C-atom in the X-ray structure<sup>10</sup>, cFapydG is less polar and H-bonds are formed with X-WAT instead of cFapydG. We suggest that the water molecule in the active site is an artifact of the carbocyclic compound cFapydG or c8OG in the X-ray structures [PDB-code: 1XC8<sup>10</sup> and 4CIS<sup>37</sup>], respectively, and is not part of the active site *in vivo*. Therefore, we will not consider the water molecule in the calculations any further.

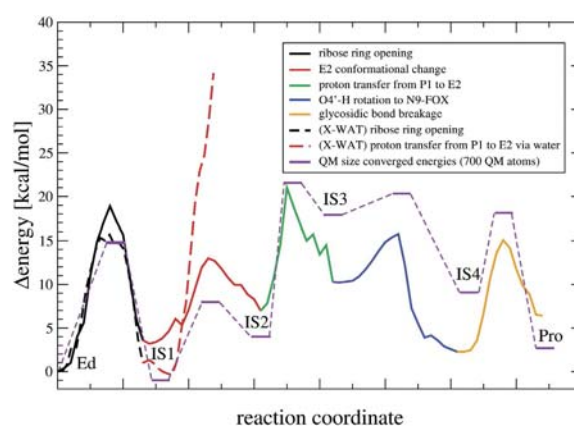
**Protonation state.** The correct protonation state of the active site is clearly decisive for the reaction mechanism. While X-ray data does not provide this information, in principle P1, E2, and E5 (see Fig. 2) can be protonated: However, protonation of the N-terminal P1 can be excluded since no nucleophilic attack at C<sub>1'</sub> could occur and, consequently, no Schiff base intermediate would be reached. For the two other possibilities, our QM/MM calculations indicate that E2 protonation is favored by 32 kcal/mol over E5 protonation. This is in line with PROPKA<sup>38–41</sup> predictions that estimate the pK<sub>a</sub> of E2 as 7.6 and of E5 as 5.5. This is also in line with the fact that E2 is located closer to the ribose ring than E5, so that most likely the protonated E2 is the proton donor for the first reaction step. The active site for our calculations is shown in Fig. 2.

**Repair mechanism.** For Fpg in general, a direct glycosidic bond cleavage mechanism has been proposed for 8OG for many years<sup>42</sup>. Here, the damaged base would be cleaved under nucleophilic attack of P1 while the ribose ring remains intact. Such a direct base excision requires, that the damaged base becomes a better leaving group by protonation. However, for FapydG this seems not possible, since according to our calculations neither energetically favored protonation sites of FapydG exist, nor are there any suitable proton donors in the cavity (as described further below). Furthermore, our QM/MM calculations show, that independent of the protonation state of the active site, the reaction barriers for glycosidic bond cleavage are higher than 30 kcal/mol. In this way, such a mechanism is most unlikely under physiological conditions - independent of the presence of X-WAT (see SI-4.1).

In addition to the direct base-excision pathway, another mechanism has been proposed in the literature for the repair by Fpg, which has received only little attention and for which no evidence has been provided<sup>8</sup>. Here, first the ribose is protonated before excision of the damaged base occurs. This is in line with a recent ribose-protonated mechanism we found for 8OG repair<sup>37</sup>, which, however, is not base-independent. In the first reaction step E2 is deprotonated by O<sub>4'</sub> while P1 nucleophilic attacks C<sub>1'</sub> during ribose ring opening leading to IS1 (intermediate state 1; Fig. 3). For this step we calculate a barrier of 14 kcal/mol (see Fig. 4; all energetics listed here are for the converged QM region with 700 atoms; see Fig. 5, SI-5.2 and Section “Details for QM size convergence”). The second reaction step is a reorientation of the E2 side chain (IS2), which allows deprotonation of P1. (As discussed earlier, we have shown X-WAT not to be present in the active site. In case of presence of X-WAT, the first step of the mechanism does not change significantly, while in the second step its presence prevents reorientation of E2, rendering the deprotonation of P1 highly unfavorable. The transfer of the acidic proton of P1 to other residues is due to distance and energetics not accessible under enzymatic conditions. Even the transfer via X-WAT to another residue is energetically unlikely.) In the third step P1 is deprotonated by E2 with a barrier of 17 kcal/mol (IS3). After this proton transfer, the fourth step is the reorientation of the alcohol group at C<sub>4'</sub> towards the damaged base (IS4) to avoid clashes with the protonated E2 residue. This step was calculated with a barrier of only 3 kcal/mol. The obtained stable intermediate is 8 kcal/mol higher in energy than the initial educt state (Ed). The last of the 5 steps is the base-excision, in which N<sub>9</sub> is protonated by the alcohol group at C<sub>4'</sub> which in turn abstracts a proton of E2 (Pro). The glycosidic bond breaks during Schiff base formation between C<sub>1'</sub> and P1. This crucial reaction step can now occur with a barrier of only 9 kcal/mol. The final product of the cleavage reaction are the free base FapyG and a stable Schiff base (Imine) between the DNA backbone and the N-terminal proline (P1) of Fpg. This



**Figure 3.** FapydG repair mechanism by Fpg. The color code of the arrows corresponds to the barriers in Fig. 4.

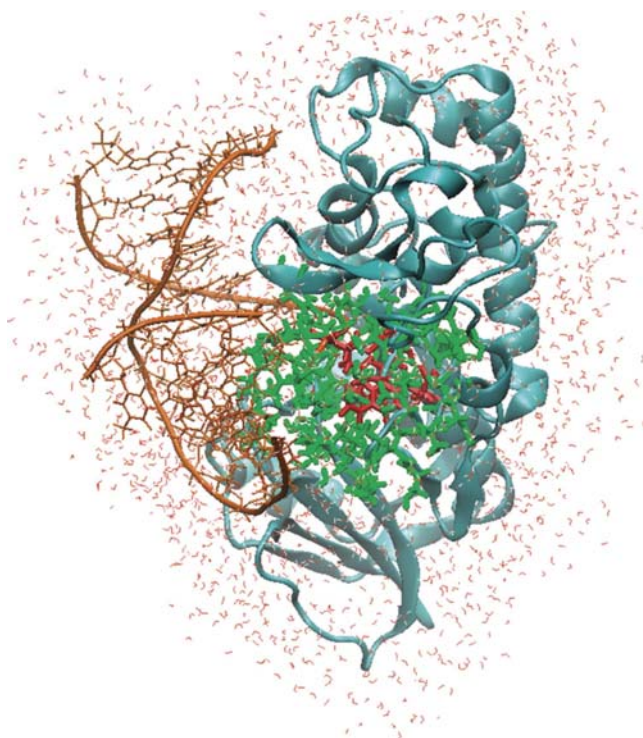


**Figure 4.** Reaction profile of the repair mechanism of FapydG with the color code of Fig.3. In dashed lines the reaction profile including X-WAT is shown. The system consists of 54412 atoms in total. 10 Å around N<sub>9</sub> of FapydG are optimized, including 87 QM atoms. The QM size converged energies for the barriers and intermediates using 700 QM atoms are shown in purple.

product structure is only 2 kcal/mol higher in energy as compared to the initial educt. The obtained Schiff base is also in agreement with the X-ray structure 1L1Z (see SI-4.2). The full repair mechanism is illustrated in Fig.3.

Overall, our repair mechanism is base-independent and can now explain the experimental observations, that a considerable number of different chemically modified DNA bases (pyrimidine<sup>43–45</sup> and purine bases<sup>10,46</sup>) - even nonpolar analogues<sup>47</sup> - can be excised by Fpg. Despite the structural differences, all these substrates have a N-glycosidic bond. This nitrogen is the only atom of the DNA base that is crucial in the mechanism, since it needs to be protonated to become a neutral leaving group, and is therefore an unspecific target for protonation. This implies, that discrimination of the DNA bases must occur in an earlier step of the DNA-enzyme interaction (recognition).





**Figure 5.** DNA repair enzyme Fpg in complex with damaged DNA. The active site is shown in red, the QM region including 700 atoms is shown in green.

size (QM region)	max. change in the reaction profile relative to next smaller QM size	max. change in the reaction profile due to optimization	average relaxation energy (absolute energy)
218	16.2	7.5	86.9
515	5.5	3.6	24.8
622	4.4	0.9	4.9
700	1.4	1.3	1.1

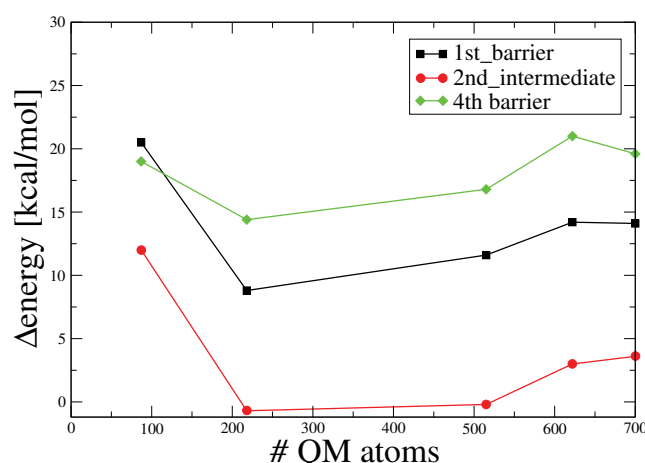
**Table 1.** Influence of increasing QM region and geometry optimization on the active site. Energy values in kcal/mol.

**Details of QM size convergence.** QM/MM approaches have been widely employed for describing, e.g., complex reactions in enzyme cavities (see, e.g., Ref. 48 for a recent review). Only with advent of linear-scaling QM/MM approaches (e.g., Ref. 19 for a recent review), the full convergence of results with the QM sphere has become possible, where it has been recognized that fairly large QM spheres are necessary for a reliable description of molecular processes<sup>11–15</sup>. For the present system, we have performed QM/MM convergence studies with up to 700 QM atoms (see Table 1 and Fig. 6): These indicate that although the reaction profile seems almost converged for 515 QM atoms, relaxation energies upon geometry optimization are only converged for larger spheres with about 700 QM atoms (for details see also SI-5.3). A similar QM size convergence has been found for calculating interaction energies<sup>37</sup>.

## Conclusion

We have presented a new base excision repair mechanism of the oxidative DNA damage FapydG that is base-independent and implies that no discrimination between damaged and undamaged bases occurs within the active site. Instead of the previously assumed direct glycosidic bond cleavage, our calculations strongly suggest a protonation of O<sub>4'</sub> with ribose ring opening as the first reaction step.

Here, it is important to note that a water molecule within the active site of the X-ray structure is most likely an artifact of employing a carbocyclic analogue to capture the educt state. The observed ribose



**Figure 6.** QM size convergence shown for selected points within the repair mechanism.

ring opening as the initial step and the formation of a Schiff base intermediate are also in line with the repair mechanism of 8OG<sup>37</sup>. In difference to 8OG, the opened imidazole ring of FapydG leads to an *anti*-conformation within the active site, a base-unspecific protonation and therefore to a base-independent mechanism. In this way, other oxidative DNA damages, like FapydA<sup>46</sup>, 5-hydroxyuracil<sup>43</sup> and thymine glycol<sup>44</sup>, can also be excised by Fpg. Even nonpolar analogues of 8OG are excised by Fpg, which has been reported by David and coworkers<sup>47</sup> and can now be rationalized by our new base-independent mechanism. Overall, we conclude as a consequence of the base-independent mechanism in the enzymatic cavity that discrimination is only part of the base-flip and recognition procedure. We are convinced that our new mechanism will help to elucidate similar DNA repair processes also in other organisms.

### Associated content

The figures were created using VMD<sup>49</sup>. For further details see Supplementary Information.

### References

- Bradley-Whitman, M. A. *et al.* Nucleic acid oxidation: an early feature of Alzheimer's disease. *J. Neurochem.* **128**, 294–304 (2014).
- Alam, Z. I. *et al.* Oxidative DNA Damage in the Parkinsonian Brain: An Apparent Selective Increase in 8-Hydroxyguanine Levels in Substantia Nigra. *J. Neurochem.* **69**, 1196–1203 (1997).
- Vladimirova, O. *et al.* Oxidative damage to DNA in plaques of MS brains. *Mult. Scler.* **4**, 413–418 (1998).
- Rehman, A. *et al.* Increased oxidative damage to all DNA bases in patients with type II diabetes mellitus. *FEBS Lett.* **448**, 120–122 (1999).
- Greenberg, M. M. The formamidopyrimidines: purine lesions formed in competition with 8-oxopurines from oxidative stress. *Acc. Chem. Res.* **45**, 588–97 (2012).
- O'Connor, T. R. & Laval, J. Physical association of the 2,6-diamino-4-hydroxy-5N-formamidopyrimidine-DNA glycosylase of *Escherichia coli* and an activity nicking DNA at apurinic/apyrimidinic sites. *Proc. Natl. Acad. Sci. USA* **86**, 5222–6 (1989).
- Fromme, J. C. & Verdine, G. L. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat. Struct. Biol.* **9**, 544–52 (2002).
- McCullough, A. K., Dodson, M. L. & Lloyd, R. S. Initiation of Base Excision Repair: Glycosylase Mechanisms and Structures. *Annu. Rev. Biochem.* **68**, 255–285 (1999).
- Stivers, J. T. & Jiang, Y. L. A Mechanistic Perspective on the Chemistry of DNA Repair Glycosylases. *Chem. Rev.* **103**, 2729–2760 (2003).
- Coste, F. *et al.* Structural basis for the recognition of the FapydG lesion (2,6-diamino-4-hydroxy-5-formamidopyrimidine) by formamidopyrimidine-DNA glycosylase. *J. Biol. Chem.* **279**, 44074–44083 (2004).
- Sumowski, C. V. & Ochsenfeld, C. A Convergence Study of QM/MM Isomerization Energies with the Selected Size of the QM Region for Peptidic Systems. *J. Phys. Chem. A* **113**, 11734–11741 (2009).
- Senn, H. M. & Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **48**, 1198–1229 (2009).
- Sumowski, C. V., Schmitt, B. T., Schweizer, S. & Ochsenfeld, C. Quantum-Chemical and Combined Quantum-Chemical/Molecular-Mechanical Studies on the Stabilization of a Twin Arginine Pair in Adenovirus Ad11. *Angew. Chem., Int. Ed.* **49**, 9951–9955 (2010).
- Flaig, D., Beer, M. & Ochsenfeld, C. Convergence of Electronic Structure with the Size of the QM Region: Example of QM/MM NMR Shieldings. *J. Chem. Theory Comput.* **8**, 2260–2271 (2012).
- Sumner, S., Söderhjelm, P. & Ryde, U. Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins. *J. Chem. Theory Comput.* **9**, 4205–4214 (2013).
- White, C. A., Johnson, B. G., Gill, P. M. W. & Head-Gordon, M. The continuous fast multipole method. *Chem. Phys. Lett.* **230**, 8–16 (1994).
- Ochsenfeld, C., White, C. A. & Head-Gordon, M. Linear and sublinear scaling formation of Hartree-Fock-type exchange matrices. *J. Chem. Phys.* **109**, 1663–1669 (1998).
- Ochsenfeld, C. Linear scaling exchange gradients for Hartree-Fock and hybrid density functional theory. *Chem. Phys. Lett.* **327**, 216–223 (2000).
- Kussmann, J., Beer, M. & Ochsenfeld, C. Linear-scaling self-consistent field methods for large molecules. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **3**, 614–636 (2013).

20. Case, D. A. *et al.* AMBER 11. University of California, San Francisco (2010).
21. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
22. Perlow-Poehnelt, R. A., Zharkov, D. O., Grollman, A. P. & Broyde, S. Substrate discrimination by formamidopyrimidine-DNA glycosylase: distinguishing interactions within the active site. *Biochemistry* **43**, 16092–105 (2004).
23. Song, K., Hornak, V., Santos, C. D., Grollman, A. P. & Simmerling, C. Molecular Mechanics Parameters for the FapydG DNA Lesion. *J. Comput. Chem.* **29**, 17–23 (2008).
24. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–74 (2004).
25. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **25**, 247–60 (2006).
26. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–802 (2005).
27. Sherwood, P. *et al.* QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct.: THEOCHEM* **632**, 1–28 (2003).
28. Vosko, S. H., Wilk, L. & Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **58**, 1200–1211 (1980).
29. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
30. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
31. Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **33**, 8822–8824 (1986).
32. 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.* **132**, 154104 (2010).
33. Shao, Y. *et al.* Advances in methods and algorithms in a modern quantum chemistry program package. *Physical Chemistry Chemical Physics* **8**, 3172–91 (2006).
34. Goerigk, L. & Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **13**, 6670–88 (2011).
35. Kästner, J. *et al.* DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **113**, 11856–11865 (2009).
36. Fromme, J. C. & Verdine, G. L. DNA lesion recognition by the bacterial repair enzyme MutM. *J. Biol. Chem.* **278**, 51543–8 (2003).
37. Sadeghian, K. *et al.* Ribose-Protonated DNA Base Excision Repair: A Combined Theoretical and Experimental Study. *Angew. Chem., Int. Ed.* **53**, 10044–10048 (2014).
38. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61**, 704–21 (2005).
39. Bas, D. C., Rogers, D. M. & Jensen, J. H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* **73**, 765–83 (2008).
40. Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
41. Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
42. Zharkov, D. O., Rieger, R. A., Iden, C. R. & Grollman, A. P. NH<sub>2</sub>-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by Escherichia coli formamidopyrimidine-DNA glycosylase (Fpg) protein. *J. Biol. Chem.* **272**, 5335–41 (1997).
43. Hatahet, Z., Kow, Y. W., Purmal, A. A., Cunningham, R. P. & Wallace, S. S. New substrates for old enzymes. 5-Hydroxy-2'-deoxycytidine and 5-hydroxy-2'-deoxyuridine are substrates for Escherichia coli endonuclease III and formamidopyrimidine DNA N-glycosylase, while 5-hydroxy-2'-deoxyuridine is a substrate for uracil DNA N-glycosylase. *Journal of Biological Chemistry* **269**, 18814–18820 (1994).
44. Jurado, J., Saparbaev, M., Matray, T. J., Greenberg, M. M. & Laval, J. The Ring Fragmentation Product of Thymidine C5-Hydrate When Present in DNA Is Repaired by the Escherichia coli Fpg and Nth Proteins. *Biochemistry* **37**, 7757–7763 (1998).
45. D'Ham, C., Romieu, A., Jaquinod, M., Gasparutto, D. & Cadet, J. Excision of 5,6-Dihydroxy-5,6-dihydrothymine, 5,6-Dihydrothymine, and 5-Hydroxycytosine from Defined Sequence Oligonucleotides by Escherichia coli Endonuclease III and Fpg Proteins: Kinetic and Mechanistic Aspects. *Biochemistry* **38**, 3335–3344 (1999).
46. Boiteux, S., Gajewski, E., Laval, J. & Dizdaroğlu, M. Substrate specificity of the Escherichia coli Fpg protein formamidopyrimidine-DNA glycosylase: excision of purine lesions in DNA produced by ionizing radiation or photosensitization. *Biochemistry* **31**, 106–110 (1992).
47. McKibbin, P. L., Kabori, A., Taniguchi, Y., Kool, E. T. & David, S. S. Surprising Repair Activities of Nonpolar Analogs of 8-oxoG Expose Features of Recognition and Catalysis by Base Excision Repair Glycosylases. *J. Am. Chem. Soc.* **134**, 1653–1661 (2012).
48. Warshel, A. Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). *Angew. Chem., Int. Ed.* **53**, 10020–10031 (2014).
49. Humphrey, W., Dalke, A. & Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).

## Acknowledgements

We acknowledge financial support by the DFG funding initiative SFB749 (TP C7) and the Excellence Cluster EXC114 (CIPSM). In addition, some computational time was provided by the LRZ Munich.

## Author Contributions

I.D.B., K.S. and C.O. wrote the main manuscript text and I.D.B. prepared the figures. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Blank, I. D. *et al.* A Base-Independent Repair Mechanism for DNA Glycosylase—No Discrimination Within the Active Site. *Sci. Rep.* **5**, 10369; doi: 10.1038/srep10369 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

## Supplementary Information

### **A Base-Independent Repair Mechanism for DNA Glycosylase — No Discrimination Within the Active Site**

Iris D. Blank<sup>a,b)</sup>, Keyarash Sadeghian<sup>a,b)</sup>, and Christian Ochsenfeld<sup>a,b)</sup>

<sup>a)</sup> Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU),  
Butenandtstr. 7, D-81377 Munich, Germany

<sup>b)</sup> Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry,  
University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany

## SI-1 Influence of basis set and DFT functional

Fig. SI-1 and Fig. SI-2 show that the deviation between the tested different basis sets and functionals is within 3kcal/mol.

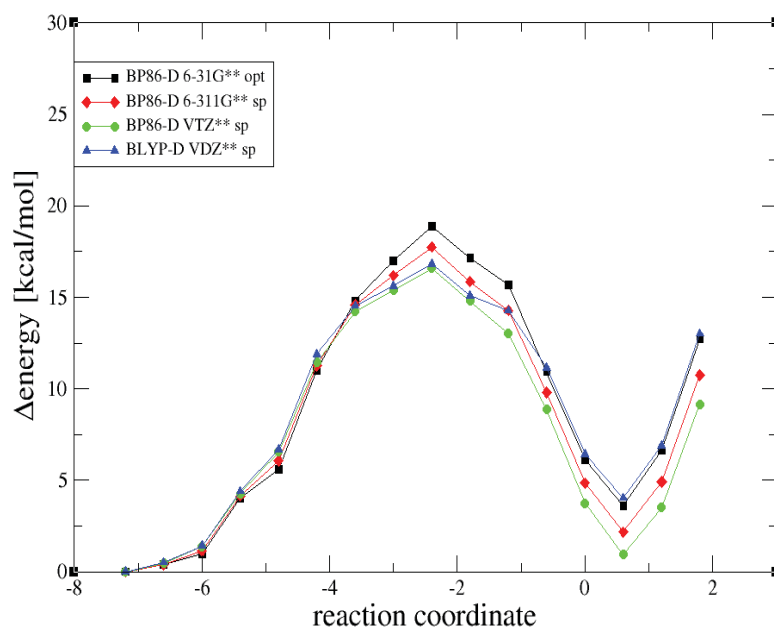


Fig. SI-1: Influence of the basis set and DFT functional on the first reaction barrier, the ribose protonation. All values are relative energies compared to their educt state (sp=single point; opt=optimization). The system consists of 54412 atoms in total and 87 atoms in the QM region.

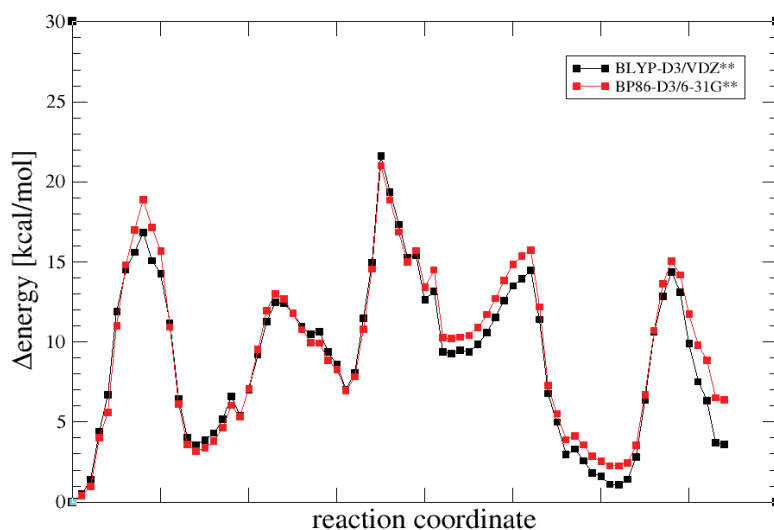


Fig. SI-2: Comparison of BP86-D3/6-31G\*\* with BLYP-D3/VDZ\*\* for the whole repair mechanism of FapydG by Fpg. The system consists of 54412 atoms in total and 87 atoms in the QM region.

## SI-2 Discussion of different X-ray-structures of Fpg

As only one structure exists for FapydG, X-ray structures including 8OG need to be used for comparison. Two different structures for the educt state have been obtained via X-ray crystallography. One structure was obtained by a E2Q mutation [PDB-code: 1R2Y [1]] the other structure by replacing 8OG with the carbocyclic analogue c8OG [PDB-code: 4CIS [2]]. The three structures have in common, that they trap the educt state, but it is worth mentioning, that only the crystal structures, where the damaged nucleotide was substituted by a carbocyclic compound[2, 3], contain a water molecule in the active site (X-WAT).

The only intermediate structure published so far is a Schiff base intermediate, where 8OG has been cleaved and the damaged base could not be resolved in the binding pocket[4] [PDB-code: 1L1Z]. Therefore no information can be gained about the presence of the water molecule in this structure.

The X-ray structure of the DNA-enzyme complex containing cFapydG [PDB-code: 1XC8] discussed in more detail:

In contrast to FapydG, cFapydG cannot be cleaved by the enzyme. cFapydG blocks the reaction by lacking the O<sub>4'</sub> atom, which was replaced by a carbon atom. In addition, no hydrogen bonds can be formed anymore to the ribose. Interactions between Fpg and O<sub>4'</sub> have been disabled by the modification, hence, the interaction pattern in the active site is expected to have significantly changed and to be disturbed. This indicates, that the interaction between the active site (especially E2) and O<sub>4'</sub> is crucial for the reaction.

The presence of a water molecule in the active site in some of the structures raises the question whether it is part of the excision mechanism *in vivo* or only an artifact of the crystallization conditions.

## SI-3 Details of the FF-MD simulations

### SI-3.1 Behavior of different systems in FF-MD

We employed FF-MD, which is a standard method to investigate the overall dynamics of systems by gaining a huge number of snapshots. This method can be applied because no chemical reaction takes place. To gain deeper insights into the behavior of the active site, multiple FF-MD simulations for all four systems have been performed: FapydG with (I) and without X-WAT (II), cFapydG with (III) and without X-WAT (IV). For proper statistics and to obtain a statistically significant analysis, at least 5 simulations have been performed for each system. The average runtime over all systems was 110 ns. The simulation time for each system is listed in Tab. SI-1. For system III more simulations have been performed to have better insight into the behavior of X-WAT. A long simulation time was used to investigate if other water molecules again enter the active site of Fpg.

*Tab. SI-1: FF-MD systems*

system	Total time	Number of simulations
I (FapydG including X-WAT)	100 ns	5
II (FapydG without X-WAT)	70 ns	5
III (cFapydG including X-WAT)	150 ns	12
IV (cFapydG without X-WAT)	120 ns	5

In the FF-MD simulations of FapydG including X-WAT (system I), X-WAT inhibits the protonated E2 to form a hydrogen bond with O<sub>4'</sub> which is needed to allow the starting conformation for the reaction to occur (see Section “repair mechanism”). X-WAT changes the hydrogen network within the active site substantially: It can form H-bonds to E5 (acceptor), amino-group of FapyG (donor), E2 (donor), and O<sub>4'</sub> (acceptor). Hence, X-WAT prevents the interaction between O<sub>4'</sub> and E2 due to steric hindrance. E2 interacts with O<sub>3'</sub> instead. X-WAT can also leave the active site

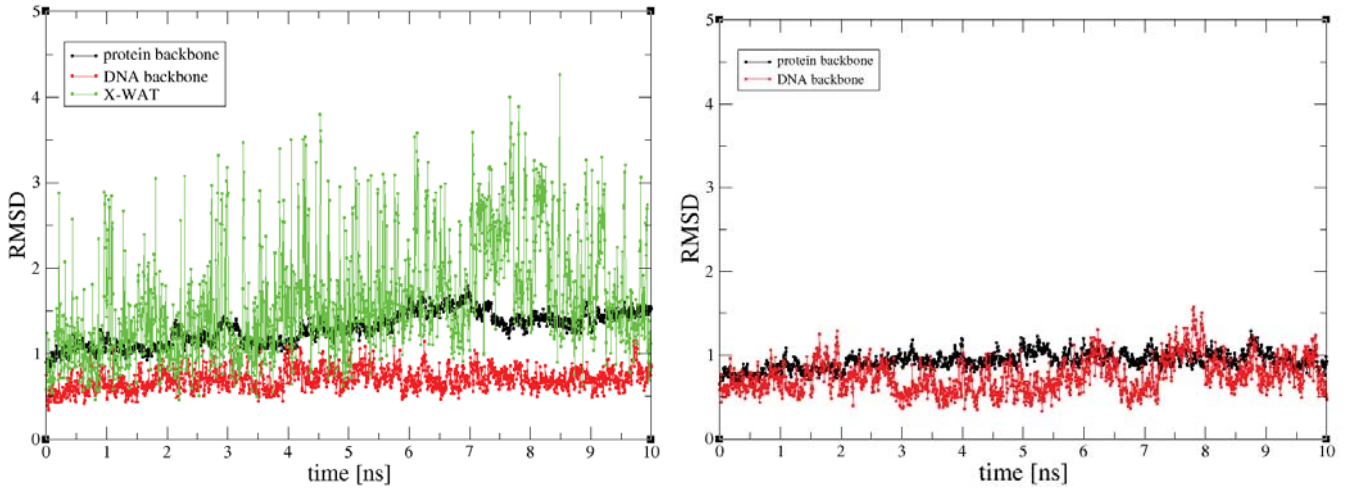


and then stay in the solvent within the simulation time. In this case no other water molecule enters the active site.

Also during FF-MD simulations of FapydG without X-WAT (system II), no entrance of a water molecule into the active site has been observed, and there are multiple events of interaction between protonated E2 with O<sub>4'</sub> (see SI-2.3). We also performed FF-MD calculations on the DNA-enzyme complex containing cFapydG instead of FapydG. In that case, X-WAT leaves the active site within the first ns in most cases (system III). In system IV no X-WAT is present in the beginning of the simulations. For this system we have observed that a water molecule is able to enter the active site with low probability after about 30 ns.

### SI-3.2 Structural analysis of FF-MD

As a stability check of the system during the FF-MD simulations, RMSD of the protein backbone, DNA and X-WAT are plotted in Fig. SI-3. As reference for the structure analysis we use the minimized and equilibrated X-ray structure. The selection of DNA consists of the backbone atoms of FapydG with pairing nucleotide as well as the previous and following nucleotide pair, where hydrogens were excluded.



SI-Fig.3: RMSD plots for systems I (left) and II (right) of the protein backbone, DNA, and X-WAT. The system is more stable without X-WAT, due to the direct interactions between  $O_{4'}$  and E2.

### SI-3.3 Interaction between ribose and E2

Tab. SI-2 shows that the crucial interaction between  $O_{4'}$  and E2 is much more probable without X-WAT. For definition of the systems see Tab. SI-1. Hydrogen bonding is defined as  $\leq 3 \text{ \AA}$  between donor and acceptor with an angle cutoff of  $45^\circ$ .

Tab. SI-2: Interactions between ribose and E2 in system I + II.

Interaction between	donor/acceptor	occupancy system I	occupancy system II
$O_{4'}$ and E2	donor GLH2-side	0.48%	36.97%
	acceptor FapyG278-side		
$O_{3'}$ and E2	donor GLH2-side	73.81%	37.52%
	acceptor FapyG278-side		

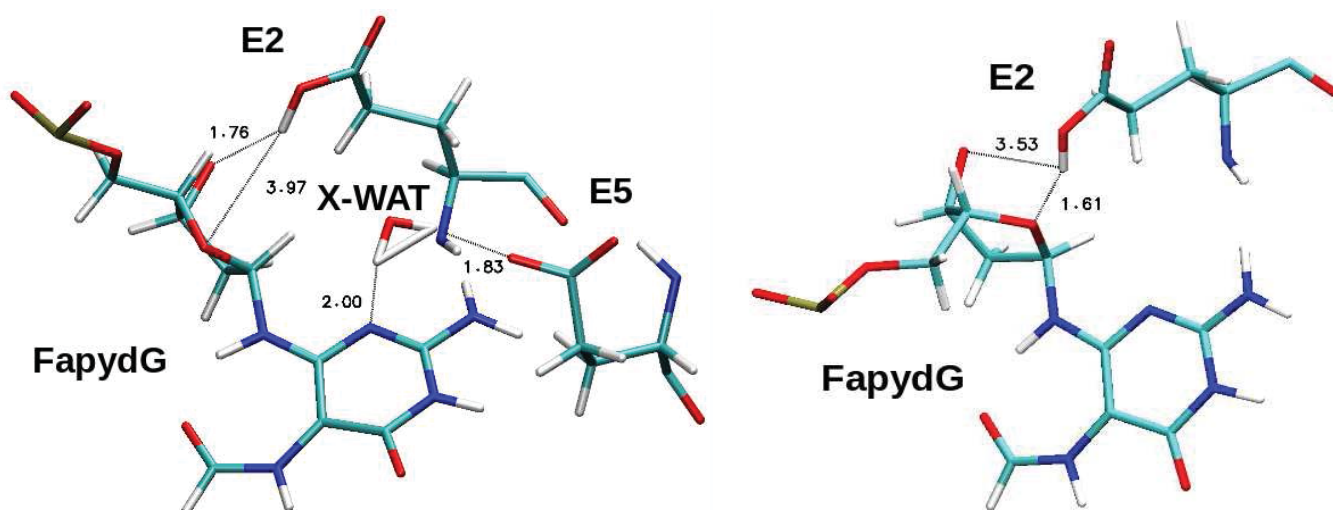


Fig. SI-4: Left: system I, where a snapshot out of a MD simulation is shown as example of the disturbed hydrogen network. The distance of E2 to  $O_{4'}$  is much bigger than to  $O_{3'}$  due to steric hindrance of X-WAT, which in this snapshot forms hydrogen bonds to  $N_3$  of FapyG and to E5, both as hydrogen donor. Right: system II, where a snapshot out of a MD simulation is shown as example of the hydrogen network. E2 is much closer to  $O_{4'}$  than to  $O_{3'}$ , so the system gets into a suitable starting conformation for the repair reaction.

## SI-4 Details of the reaction mechanism

### SI-4.1 Glycosidic bond breakage for all possible protonation states

For all eight possible protonation states (of P1, E2, E5) calculations of the direct glycosidic bond breakage have been performed. For each protonation state these calculations have been performed for the system containing X-WAT and lacking X-WAT. The nomenclature of the protonation states is listed in Tab. SI-3. The reaction profiles are depicted in Fig. SI-5.

Tab. SI-3: Overview of all eight possible protonation states in the active site of Fpg.

	P1	E2	E5
I	–	–	–
II	–	–	+
III	+	–	+
IV	+	–	–
V	–	+	–
VI	+	+	–
VII	–	+	+
VIII	+	+	+

(+) protonated; (–) unprotonated

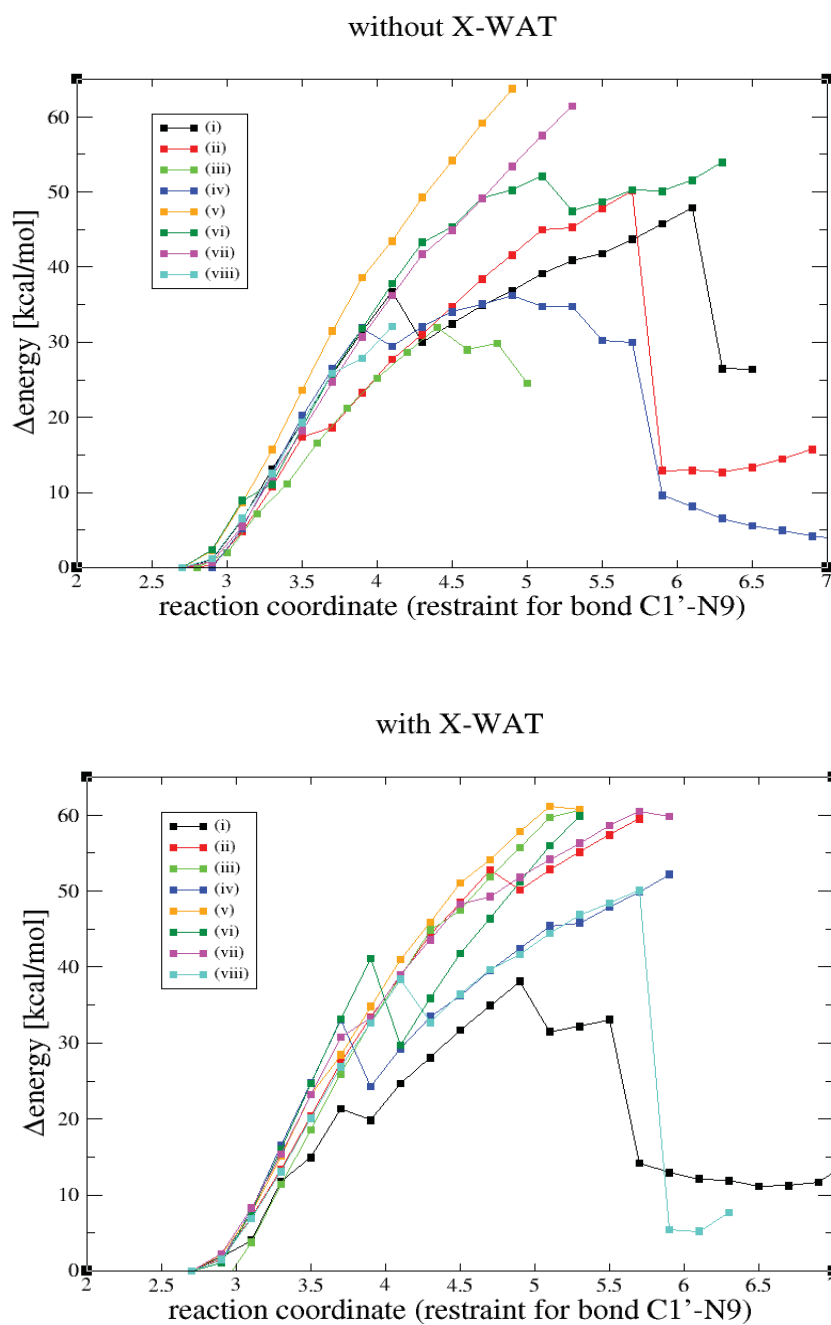


Fig. SI-5: Simulation of direct glycosidic bond breakage for all eight possible protonation states - without X-WAT and with X-WAT, respectively. No reaction barrier is lower than 30 kcal/mol. The system for these calculations consists of the whole equilibrated structure (54412 atoms), with 10 Å around N<sub>9</sub> of FapydG as relaxed region and about 90 QM atoms, depending on the protonation state and presence of X-WAT.

## SI-4.2 Structure of Schiff base intermediate

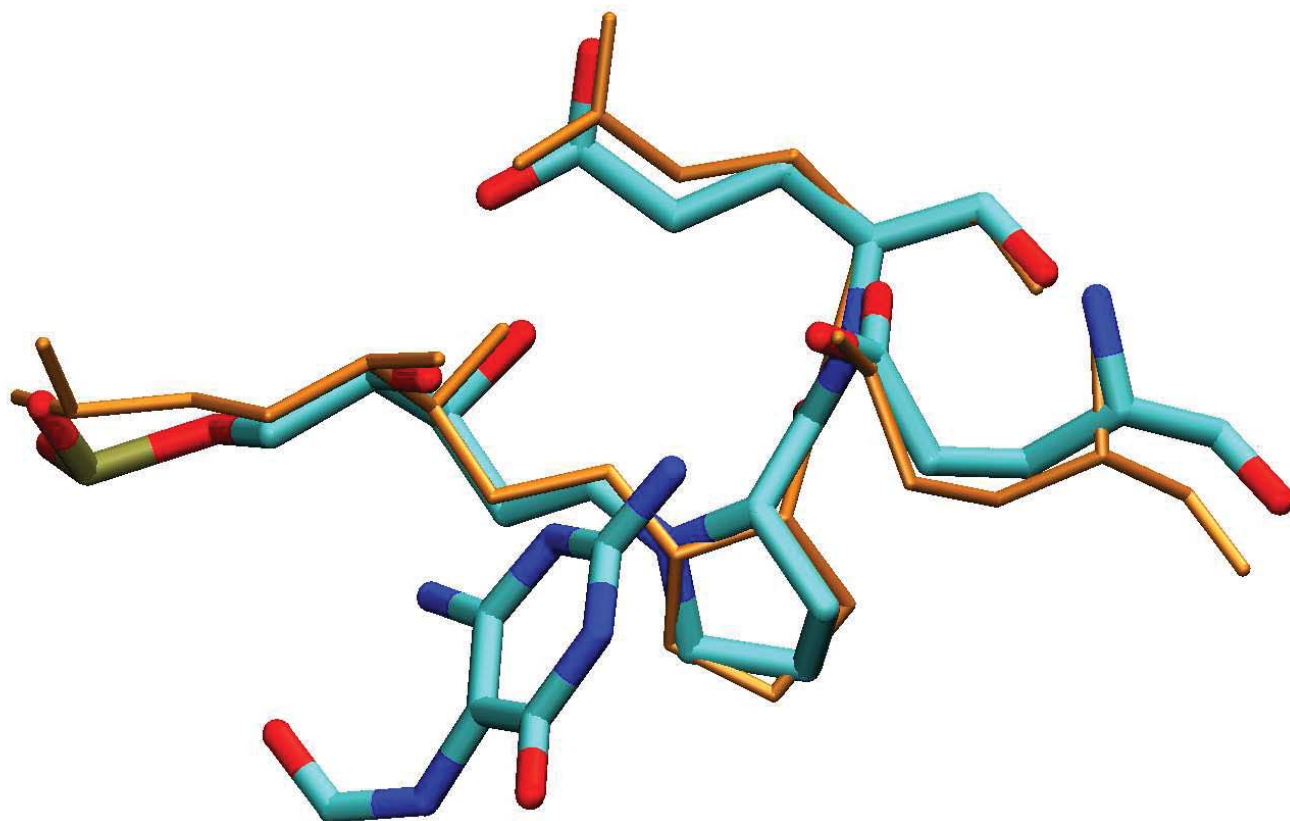


Fig. SI-6: Alignment of the obtained QM/MM calculated Schiff base and the corresponding X-ray structure [PDB-code: 1L1Z]. The calculated structure is shown in atomic colors, the structure obtained by X-ray crystallography is shown in orange. In the X-ray structure the cleaved base is not resolved.

## SI-5 Details on the QM size convergence (QM/MM)

### SI-5.1 Selected residues for each QM sphere within the QM size convergence study

Tab. SI-4: selection of residues included in QM/MM calculations for QM size convergence

QM atoms	Residues of QM atoms	Number of residues (QM)	Atoms in relaxed region	Residues of relaxed region	Number of residues in relaxed region	Charge (QM)
87	1 2 5 278	4	754	1-6, 57, 72-78, 109, 111, 161, 170-173, 216-223, 232, 235, 238, 260, 277-279, waters: 312 314 315 336 347 353 359 360 362 372 375 395 412 419 440 461 479 488 498 874 1295 1333 1337 1339 1344 1346 1498 1506 1522 1865 1888	67	-2
218	1 2 5 75 172 219 222 277 278 waters: 336 461 498 1337 1506	14	754	1-6, 57, 72-78, 109, 111, 161, 170-173, 216-223, 232, 235, 238, 260, 277-279, waters: 312 314 315 336 347 353 359 360 362 372 375 395 412 419 440 461 479 488 498 874 1295 1333 1337 1339 1344 1346 1498 1506 1522 1865 1888	67	-3
515	1 2 5 57 74 75 76 109 170 171 172 173 217 218 219 220 221 222 238 260 277 278 279 waters: 314 315 336 359 375 440 461 498 1337 1339 1344 1498 1506 1865	37	218	1 2 5 75 172 219 222 277 278 waters: 336 461 498 1337 1506	14	0
622	1 2 5 6 57 73 74 75 76 77 109 170 171 172 173 217 218 219 220 221 222 223 235 238 260 277 278 279 waters: 312 314 315 336 359 362 372 375 395 419 440 461 479 488 498 874 1333 1337 1339 1344 1498 1506 1522 1865 1888	53	218	1 2 5 75 172 219 222 277 278 waters: 336 461 498 1337 1506	14	0
700	1 2 4 5 6 57 72 73 74 75 76 77 78 109 170 171 172 173 217 218 219 220 221 222 223 232 235 238 260 277 278 279 waters: 312 314 315 336 359 360 362 372 375 395 412 419 440 461 479 488 498 874 1333 1337 1339 1344 1498 1506 1522 1865 1888	59	218	1 2 5 75 172 219 222 277 278 waters: 336 461 498 1337 1506	14	1

## SI-5.2 QM size convergence study of the reaction mechanism

Tab. SI-4: QM size convergence study of the reaction mechanism. Values in kcal/mol relative to the respective educt.

QM 87	QM 218	QM 515	QM 622	QM 700	comment
0	0	0	0	0	educt
17.7	12.0	10.3	10.8	10.5	
19.1	9.4	11.9	12.9	13.0	
20.5	8.8	11.6	14.2	14.1	1 <sup>st</sup> barrier
19.9	7.1	9.2	10.7	11.0	
7.6	-7.7	-4.0	-1.4 <sup>a)</sup>	-1.3 <sup>a)</sup>	1 <sup>st</sup> intermediate
16.7	1.8	0.9	7.5 <sup>a)</sup>	7.6 <sup>a)</sup>	2 <sup>nd</sup> barrier
16.5	6.0	2.4	4.1	4.3	
14.7	2.2	1.3	4.7	4.5	
12.0	-0.7	-0.2	3.0	3.6	2 <sup>nd</sup> intermediate
18.7	11.1	10.4	14.8	14.4	
24.9	23.0	17.4	20.1	19.9	
22.8	23.4	18.6	21.6	21.1	3 <sup>rd</sup> barrier
20.9	21.0	16.1	19.3	18.9	
15.1	9.8	13.2	17.2	16.9	3 <sup>rd</sup> intermediate
15.6	11.5	13.8	18.1	16.8	
16.1	15.7	14.8	17.7	17.2	
19.0	14.4	16.8	21.0	19.6	4 <sup>th</sup> barrier
7.8	9.7	9.3	12.4	11.0	
7.3	5.7	5.9	8.9	8.5	4 <sup>th</sup> intermediate
19.0	9.9	14.5	18.8	17.5	5 <sup>th</sup> barrier
19.3	17.7	15.2	17.9	16.8	
19.0	6.1	10.5	13.2	12.5	
11.8	-4.5	-1.5	1.9	2.5	product

<sup>a)</sup> Since the SCF cycle did not converge for BP86-D3/6-31G\*\* calculations due to the small HOMO-LUMO gap, we estimated this value via B3LYP-D3/6-31G\*\* calculations (for details see SI-5.3).



### SI-5.3 Discussion of estimated energies in the QM size convergence (QM/MM)

When we converged the relative energies for the larger systems of the barriers and intermediate structures, we encountered severe convergence problems for the 1<sup>st</sup> intermediate and one structure of the 2<sup>nd</sup> barrier calculated with BP86-D3. The problem of vanishing HOMO-LUMO gaps in DFT calculations for large systems is well known[5, 6]. To estimate the energy of these two structures, we choose B3LYP-D3, since the calculated HOMO-LUMO gap is larger for hybrid functionals, due to the exact HF exchange, allowing convergence of the SCF calculation.

We calculated single point (sp) energies with B3LYP-D3/6-31G\*\* on the structures, where the active site (218 QM atoms) has been optimized with BP86-D3/6-31G\*\*, and increased the QM region stepwise. Tab. SI-6 contains sp energies for all barriers and intermediates, calculated with B3LYP-D3/6-31G\*\*. Compared to the energies calculated with BP86, the B3LYP values are higher in energy, but the convergence behavior is almost the same. In Fig. SI-7 two examples are shown. Since the convergence behavior and the average increase of  $\Delta E$  from QM218 to QM700 show the same behavior for BP86-D3 and B3LYP-D3 (examples shown in Fig. SI-7), we used the increase of the B3LYP-D3 energies to estimate the missing energies of the BP86-D3 calculations. The average ratio of B3LYP-D3(700QM-218QM) to BP86-D3(700QM-218QM) is 1.2. We estimate the missing BP86-D3 energies by dividing the increase of B3LYP-D3 by this factor to get estimated the increase of BP86-D3 and add it to the QM218 energy of BP86-D3.

Based on this estimation, we conclude, that the exact energies of the two structures would not change the overall picture of our reaction profile. It should be noted, that although in general sp calculations are enough to estimate the necessary size of the QM region, that has to be included to obtain size converged energies, geometry optimizations are necessary to calculate the influence of this QM region on the energy.

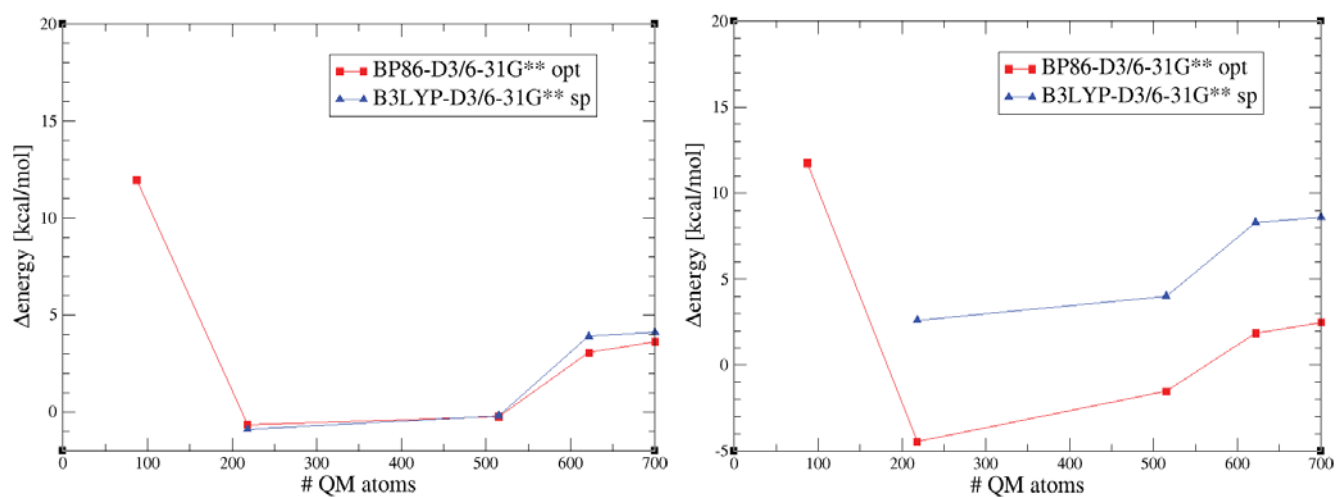


Fig. SI-7: Relative energies are plotted against the QM size within the QM/MM calculations for two different reaction points: 2<sup>nd</sup> intermediate (left), product (right) calculated with BP86-D3/6-31G\*\* (opt) and B3LYP-D3/6-31G\*\* (sp energy).

Tab. SI-6: QM size convergence study of the reaction mechanism, where the crucial points of the reaction have been calculated with increasing QM region in sp calculations (see text) with B3LYP-D3/6-31G\*\*. All values are in kcal/mol relative to the respective educt.

218	515	622	700	comment
0	0	0	0	educt
15.7	13.7	14.9	14.8	
13.4	17.3	19.5	19.4	
13.7	17.6	20.6	21.2	1 <sup>st</sup> barrier
12.7	15.3	18.6	18.8	
-3.1	0.8	4.5	4.6	1 <sup>st</sup> intermediate
4.8	7.5	11.6	11.7	
8.8	7.5	10.6	10.8	
5.6	7.2	11.4	11.5	2 <sup>nd</sup> barrier
2.6	4.0	8.3	8.6	2 <sup>nd</sup> intermediate
15.1	14.7	20.1	20.2	
26.9	20.9	24.7	24.6	
25.9	20.9	24.9	24.0	3 <sup>rd</sup> barrier
23.3	18.5	22.5	21.7	
11.8	15.7	20.4	19.6	3 <sup>rd</sup> intermediate
13.5	16.4	21.0	20.2	
17.6	18.9	22.4	21.6	
16.5	19.5	23.9	23.1	4 <sup>th</sup> barrier
10.6	9.5	12.9	12.3	
7.3	9.2	12.9	12.2	4 <sup>th</sup> intermediate
18.0	19.5	24.1	23.5	5 <sup>th</sup> barrier
25.8	21.5	24.0	23.3	
13.8	15.8	19.7	19.3	
-0.9	-0.2	3.9	4.1	product

## SI-6 Detailed Methods Section

### SI-6.1 Details for Molecular Dynamics Simulations

The following steps have been performed: The systems were energy minimized (NVT ensemble) in 3 steps, relaxing different degrees of freedom, using of the conjugate gradient algorithm: (1) only hydrogen atoms (2000 steps), (2) only solvent (3000 steps), (3) all atoms (5000 steps). The system was heated up to 300 K within 10 ps using Langevin dynamics, where a positional constraint of 1 kcal/mol/Å<sup>2</sup> was applied on non-water atoms. In the subsequent equilibration step we switched to the NPT ensemble employing the Langevin piston Nosé-Hoover method [7, 8]. At this stage the restraints on non-water atoms are reduced step by step down to zero (0.2 kcal/mol/Å<sup>2</sup> increments for every 20 ps). Equilibration was then performed for 300 ps, in which the coordinates were saved every 0.2 ps. Production runs were performed for at least 10 ns with timesteps of 2 fs using the SHAKE algorithm [9]. Coordinates were saved every 6 ps.

### SI-6.2 Details for QM/MM calculations

The system for these calculations consists of the whole equilibrated structure (54412 atoms), with 15 Å around N<sub>9</sub> of FapydG as relaxed region and 87 QM atoms. For the QM size convergence study the total system was reduced to protein, DNA, ions, water within 15Å of N<sub>9</sub> of FapydG, and 3 Å of water around both protein and DNA (in total 12031 atoms in the system).

## References

1. Fromme, J. C. & Verdine, G. L. DNA lesion recognition by the bacterial repair enzyme MutM. *J. Biol. Chem.* **278**, 51543–8 (2003).
2. Sadeghian, K. *et al.* Ribose-protonated DNA base-excision repair: a combined theoretical and experimental study. *Angew. Chem., Int. Ed.* doi: 10.1002/anie.201403334R1 (2014).
3. Coste, F. *et al.* Structural basis for the recognition of the FapydG lesion (2,6-diamino-4-hydroxy-5-formamidopyrimidine) by formamidopyrimidine-DNA glycosylase. *J. Biol. Chem.* **279**, 44074–44083 (2004).
4. Fromme, J. C. & Verdine, G. L. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat. Struct. Biol.* **9**, 544–52 (2002).
5. Rudberg, E., Rubensson, E. H. & Salek, P. Kohn–Sham Density Functional Theory Electronic Structure Calculations with Linearly Scaling Computational Time and Memory Usage. *J. Chem. Theory Comput.* **7**, 340–350 (2011).
6. Lever, G., Cole, D. J., Hine, N. D. M., Haynes, P. D. & Payne, M. C. Electrostatic considerations affecting the calculated HOMO-LUMO gap in protein molecules. *J. Phys.: Condens. Matter* **25**, 152101 (2013).
7. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
8. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (3 1985).
9. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327–341 (1977).



**V.2 Paper [2]: "DNA Glycosylase Repairs FapydA Base-Independently - Schiff Base in Equilibrium with Closed-Ribose Structure", T. Moeller, I. D. Blank, C. Ochsenfeld, (under revision).**





**DNA Glycosylase Repairs FapydA Base-Independently**  
— **Schiff Base in Equilibrium with Closed-Ribose Structure**

Tobias Moeller<sup>a,b</sup>, Iris D. Blank<sup>a,b</sup>, and Christian Ochsenfeld<sup>a,b</sup>\*

<sup>a)</sup> Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU Munich), Butenandtstr. 7, D-81377 Munich, Germany

<sup>b)</sup> Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU Munich), Butenandtstr. 5-13, D-81377 Munich, Germany

\*Correspondence to christian.ochsenfeld@uni-muenchen.de

**ABSTRACT:**

DNA is exposed to a plethora of damaging agents. This necessitates coping mechanisms of the cell in order to ensure cell survival. Oxidized DNA bases can be mutagenic and are associated with various diseases, such as, e.g., cancer, Huntington's disease, and Type II diabetes. Among oxidized DNA bases, FapydA (4,6-diamino-5-formamidopyrimidine) was shown to possess miscoding potential as well as to terminate DNA synthesis depending on DNA sequence context and specific DNA polymerase. However, the base excision mechanism for FapydA remained unclear. Our quantum-chemical calculations show that Fpg excises FapydA via a ribose-protonated mechanism. This is in line with a ribose-protonated mechanism elucidated recently by us for the excision of oxidative DNA damages by Fpg. FapydA is excised base-independently which explains experimental observations and consolidates this base-independent repair process. Furthermore, we find that the Schiff base found after base excision is in equilibrium with a closed-ribose structure which was not observed previously.

**KEYWORDS:**

QM/MM, DNA repair, MutM, Fpg, FapydG, FapydA, reaction mechanism, BER

## INTRODUCTION

DNA is permanently exposed to a plethora of damaging agents of endogenous and exogenous nature, such as reactive oxygen species (ROS) or radiation, respectively. ROS are a byproduct of metabolism and can lead to oxidative DNA damage. Among the most prominent oxidative lesions are 7,8-dihydro-8-oxoguanine [1] (8OG) and the formamidopyrimidines [2] FapydG (2,6-diamino-4-hydroxy-5-formamidopyrimidine) as well as FapydA (4,6-diamino-5-formamidopyrimidine) (see Fig. 1). These lesions are mutagenic and are associated with a multitude of diseases such as cancer, Huntington’s disease, and Type II diabetes [3, 4]. Therefore, it is vital to the cell to be able to cope with DNA damage. One way this is achieved is base excision repair [5] (BER). For the three lesions mentioned above, BER is performed by the bifunctional enzyme Fpg (Formamidopyrimidine-DNA-glycosylase, also known as MutM) [6–9]. While 8OG was found to be recognised by Fpg in a *syn*- or *anti*-conformation [10], both FapydA and FapydG were found to only be bound by Fpg in an *anti*-conformation [10]. The repair mechanism of Fpg has been unravelled only recently for 8OG [11] and FapydG [12] with the mechanism for FapydA remaining unclear. However, FapydA has been shown to possess miscoding potential [13, 14] as well as terminate DNA synthesis depending on DNA sequence context and polymerase [15]. Additionally, it has been shown that FapydA is excised more easily by Fpg than 8OG independently of the base pairing [2, 16]. Furthermore, it was shown that FapydA and FapydG are both excised at about the same speed by Fpg. Using quantum-chemical methods we elucidate here the repair mechanism of Fpg for FapydA as well as introduce an until now unknown equilibrium structure to the previously found Schiff base [17]. So far, it has been assumed that Fpg produces a Schiff base intermediate as the product of base excision which is also the starting point for subsequent ribose excision. This is based on an X-ray structure of a DNA-Fpg complex that was trapped using borohydride [PDB code: 1L1Z] [17]. Our work extends this notion through computational studies using the QM/MM approach (see, e.g., ref. [18] for a review), by introducing a closed-ribose structure that provides new impulses for the investigation of DNA repair.

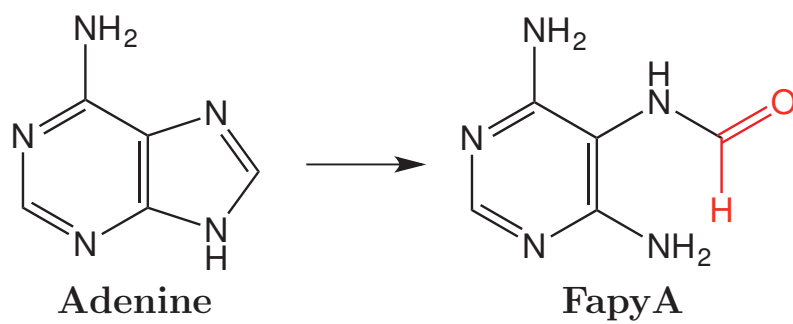


Figure 1: Comparison of Adenine and its oxidized form FapyA. FapyA in this case denotes the damaged base as opposed to the nucleotide FapydA.

## METHODS

The starting structure used in this work is based on the crystal structure of wild-type Fpg from *Lactococcus lactis* in complex with a DNA fragment which contains carbocyclic FapydG (cFapydG) [PDB code: 1XC8] [9]. We replaced cFapydG by FapydA using VMD [19] since no crystal structure of Fpg in complex with FapydA is available to our knowledge. The water in the reactive centre of the crystal structure was removed based on previous findings that showed it to be an artefact of crystallisation [12]. Hydrogen atoms were added to the X-ray structure, the system was neutralized with sodium ions and solvated in a box of explicit TIP3P water [20] with a buffer of 10 Å around the solute using XLEAP (AmberTool) [21]. The parameters for the neutral proline residue were taken from Perlow-Poehnelt et al. [10]. ANTECHAMBER [22, 23] was used to parametrize FapydA. Force field molecular dynamics (FF-MD) simulations were performed using the NAMD engine [24] with Amber10 force field parameters [21] under periodic boundary conditions and particle mesh Ewald summation (PME) with a cutoff value of 10 Å (see SI-4.1). Structure optimizations were performed at the QM/MM level employing the DL-POLY implementation within ChemShell [25] (AMBER-FF) utilising the Q-Chem program package [26] for the QM part. Density functional theory (DFT) was used at the BP86-D3/6-31G\*\* [27–31] level for reasons of comparability with computational studies on the mechanism of Fpg for excision of FapydG [12] as well as its weighted total mean absolute deviation for reaction energies of 3.5 kcal/mol [32] while maintaining sensible computational costs. Additionally, size convergence studies and an investigation of the influence of the DFT functional and basis set were done in previous work [12] which validates the method used here. The adiabatic mapping approach as well as the nudged elastic band (NEB) method of the DL-FIND [33] module implemented in ChemShell [25] were used for calculation of the repair mechanism.

## RESULTS AND DISCUSSION

### Preparation of the System

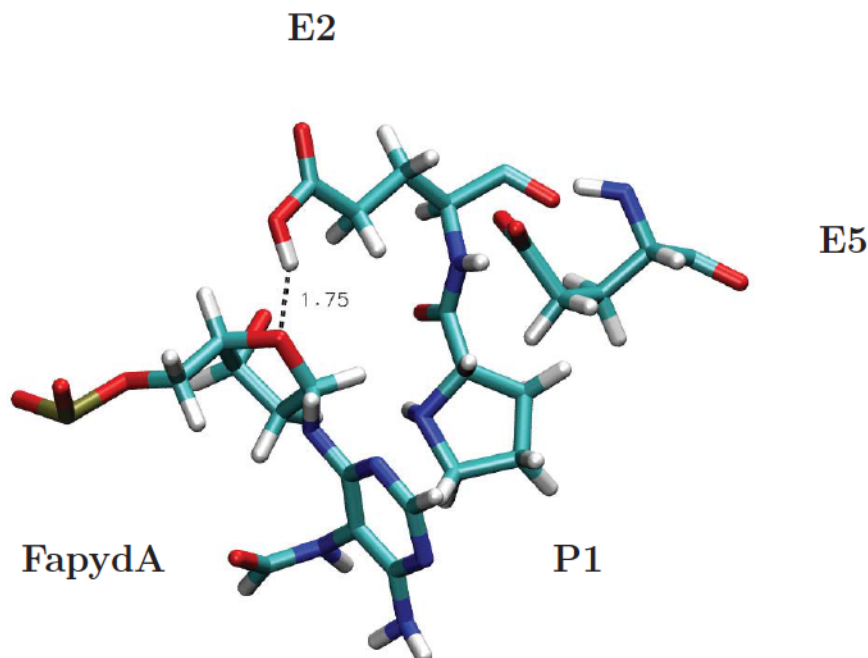


Figure 2: Active site of Fpg with P1 being neutral, E2 being protonated and, E5 not protonated. FapydA is bound in *anti*-conformation.

Since no crystal structure of Fpg in complex with FapydA exists to our knowledge, the starting structure was created by using the structure of Fpg in complex with a cFapydG-containing DNA fragment [PDB code: 1XC8] [9] and replacing cFapydG with FapydA. The crystal structure originally contains a water molecule in the active site which was shown to be an artefact of replacing O<sub>4'</sub> of the ribose with a carbon atom [12]. Hence, it was removed in this work as well. The protonation state, however, although discussed in previous work for the case of FapydG [12], was investigated in this work as well using QM/MM calculations and found to be the same for FapydA as reported before for FapydG. In principle, in the active site the N-terminal proline (P1) and two glutamates, E2, and E5, could be protonated. However, if P1 is protonated, it would not be able to attack

$C_1'$  nucleophilically due to its own positive charge and therefore no Schiff base intermediate would be formed. For a protonation state with neutral P1 and either E2 or E5 protonated we found that E2 being protonated is energetically favourable by 24 kcal/mol. This is in agreement with PROPKA [34–37] which gives a pKa of 7.70 for E2 and 4.88 for E5. FF-MD simulations were performed to equilibrate the system subsequently (SI-4.1). For the following calculations, a frame with a representative conformation was picked. The active site of the system before the first step of the reaction is shown in Fig. 2. A representation of the active and QM regions are shown in SI-4.2.

### Repair mechanism

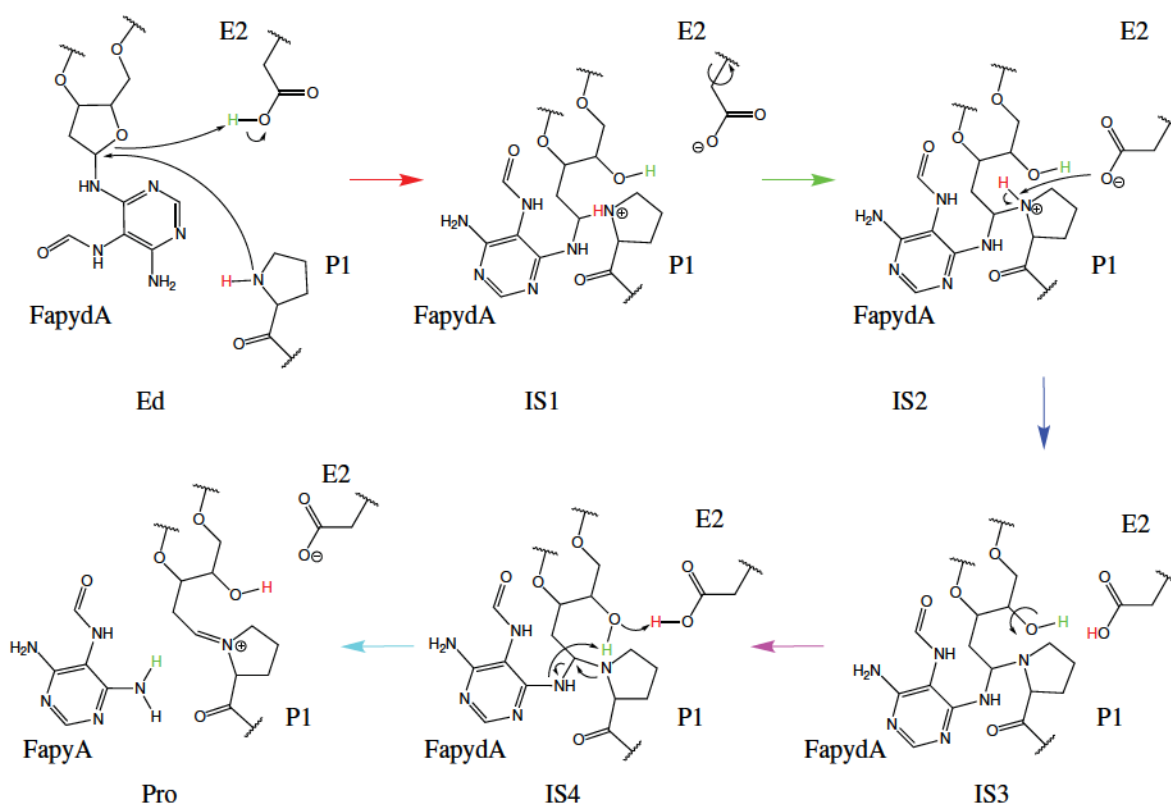


Figure 3: Overview of the calculated repair mechanism of Fpg for FapydA.

Two repair mechanisms have been proposed this far for base excision by Fpg. One suggested a direct glycosidic bond breakage for 8OG [38] (base-protonated mechanism). The other suggested protonation of ribose with subsequent leaving of the base [39] (ribose-protonated mechanism).

While the former was favoured in the literature for many years, our quantum chemical calculations revealed the latter to be energetically favourable in Fpg (also called MutM) for 8OG [11] and FapydG [12] as well as for 8OG in the human enzyme hOGG1 [40] which fulfils the same function as Fpg [41]. All three cases have in common that the initial step is the ribose-opening via a nucleophilic attack at  $C_{1'}$  by proline (Fpg) or lysine (hOGG1). In the case of FapydG, it was found that the mechanism is completely base-independent [12]. In our present work, we find a FapydG-analogous, base-independent, ribose-protonated mechanism. In a first step, the nitrogen of the N-terminal proline P1 attacks  $C_{1'}$  nucleophilically which leads to the ribose ring being opened. At the same time E2 gets deprotonated by  $O_{4'}$  leading to the intermediate state 1 (IS1, see Fig. 3). This step has a barrier of 16 kcal/mol which is slightly lower than for equivalent step of the base excision of FapydG (19 kcal/mol) [12]. After the first step, a rotation of the backbone of E2 takes place leading to the intermediate state 2 (IS2). This step prepares the system spatially for the next step and has a barrier of 17 kcal/mol. Its barrier is higher than the one found for FapydG of 10 kcal/mol. With E2 in place, the next step is the deprotonation of P1 by E2 with a barrier of 9 kcal/mol (IS3). The re-protonation of E2 allows it to protonate  $O_{4'}$  later in the reaction. This step has a significantly lower barrier than the equivalent step for FapydG which was found to be 15 kcal/mol [12]. In the second to last step, the alcohol group of  $C_{4'}$  rotates in order to avoid a clash between the protons of  $O_{4'}$  and E2 (IS4). This step has a barrier of 16 kcal/mol which is significantly higher than the one found for FapydG of 6 kcal/mol [12]. However, simultaneously to the reorientation of  $O_{4'}$ , the backbone of E2 undergoes another reorientation, similar to the one described before, only in reverse. The last step is the proton transfer from  $O_{4'}$  to N9 which subsequently leads to the leaving of the base. While  $O_{4'}$  is transferring its proton to N9 of the damaged base, E2 transfers its proton to  $O_{4'}$  in a concerted manner (see SI-1). This step leads to the formation of a Schiff base between P1 and  $C_{1'}$  and has a barrier of 14 kcal/mol which is slightly higher than the one found for FapydG (13 kcal/mol) [12]. The products of this reaction differ energetically from the educt by -17 kcal/mol. This is a significant difference to the energetic product-educt difference found for FapydG which is +6 kcal/mol. An overview of the individual



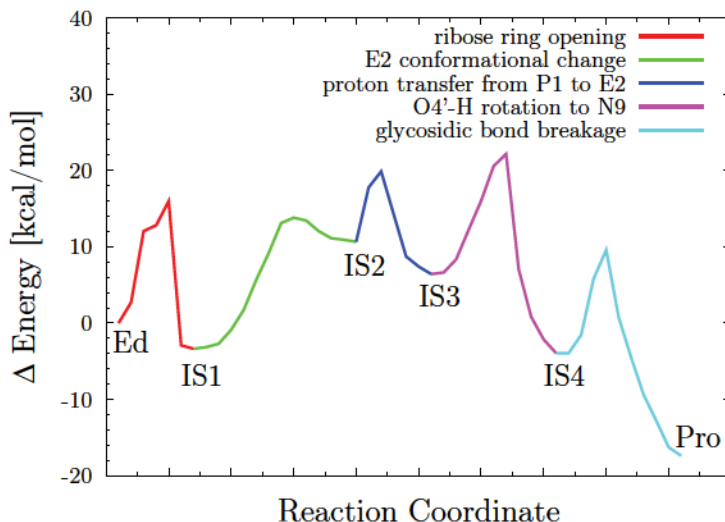


Figure 4: Energetic profile of the repair mechanism of FapydA by Fpg. The first step has been obtained from a calculation using the adiabatic mapping approach while the rest of the energies were refined by the NEB method.

barriers for FapydA and a comparison to FapydG are given in Figs. 4 and 5, respectively.

Although there are differences in the barriers when comparing FapydA and FapydG, the barriers for FapydA are always surmountable and the Schiff base product of base excision is in proper alignment to the reference X-ray structure [PDB code: 1L1Z] [17] (see SI-2). Moreover, while the previous work on FapydG [12] utilised single calculations for each reaction step, we performed NEB calculations that go from IS3 to the product with a given intermediate of IS4. This leads to an alternative path on the potential energy surface. Furthermore, in this work we see that the backbone of E2 twists while the alcohol group of C<sub>4'</sub> is turning in the step from IS3 to IS4. This is not nearly as pronounced in the reaction path reported before [12]. On the one hand, it leads to a higher barrier for this step for FapydA but on the other hand, it most likely is one of the reasons that the product is significantly lower in energy compared to FapydG. Moreover, it is reasonable to assume that small structural differences between the respective starting structures for FapydG and FapydA lead to differences in energy barriers. Another aspect to consider for the repair mechanism is the orientation of the base within the active site. FapydA behaves similar

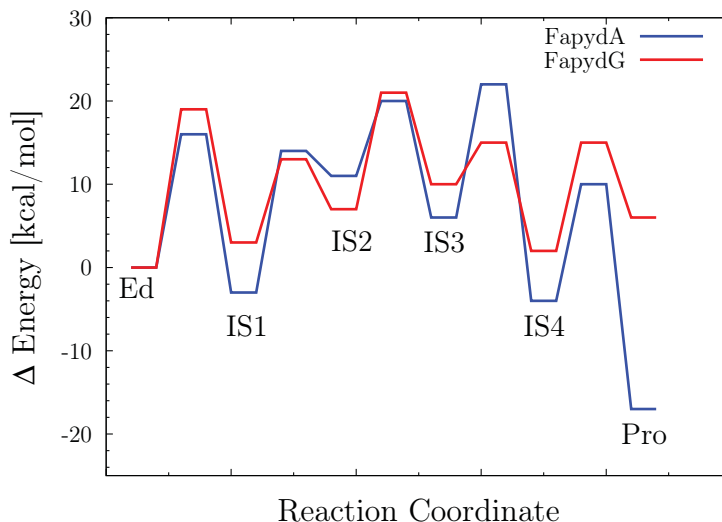


Figure 5: Comparison of energetics for the repair mechanism of Fpg for FapydA and FapydG [12].

to FapydG being repaired base-independently in the *anti*-conformation. This is in line with the observation that Fpg recognises FapydA and FapydG only in *anti*-conformation [10]. At the same time, our findings together with previous findings for FapydG [12] rationalize these observations and can explain why nonpolar analogues of 8OG can be repaired by Fpg [42]. Overall, we conclude that Fpg excises FapydA in the same way as FapydG. The base-protonated pathway has been investigated as well but was found to be energetically unfavourable (see SI-3).

### Equilibrium of Schiff base with closed ribose structure

For the last step of the repair mechanism, we observed different products depending on the setup of calculations: not only the expected Schiff base, but rather a structural isomer with a closed ribose ring as shown in Fig. 6 was obtained. While these calculations shared the same starting structure, only different restrained bonds in the adiabatic mapping approach were employed. Further calculations show that the barrier between the two final states is only about 2 kcal/mol with the resulting closed ribose structure being 1 kcal/mol lower in energy than the Schiff base structure. This implies an equilibrium between the two structures that has so far not been observed. This observation promises to be an interesting aspect to consider for future research on the

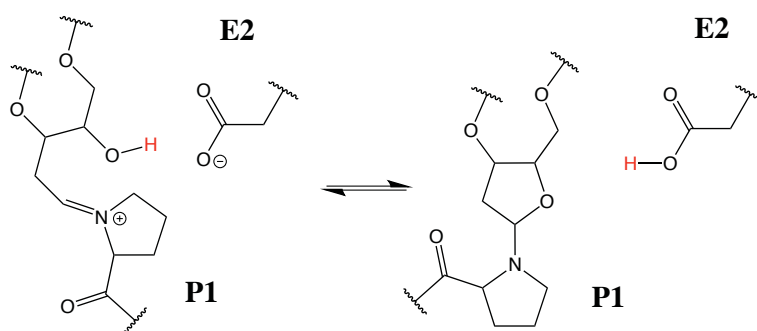


Figure 6: Equilibrium between Schiff base product and a closed-ribose isomer.

exact mechanism and energetics of the lyase reaction that follows base excision. It is possible that the crystal structure showing the Schiff base did end in this particular structure due to the use of borohydride [17] for blocking the subsequent lyase reaction and thusly the closed ribose structure was not discovered before.

## CONCLUSION

We have shown that Fpg utilises a base-independent base excision mechanism to excise FapydA, as it does for FapydG. The base-independence supports the notion that discrimination between damaged and undamaged base and the recognition of damaged bases is likely to happen before the base enters the active site of Fpg. Additionally, this work presents further evidence for the ribose-protonated repair mechanism where the ribose is opened in the first step of the reaction as opposed to the nucleobase-protonated mechanism [38, 43] suggested before. Since it was shown that Fpg excises FapydG [12] as well as FapydA in the same way, it is reasonable to assume that other oxidative DNA damages such as 5-hydroxyuracil [44] and thymine glycol [45] are excised via the same repair mechanism. In previous work, we suggested that the base-independent repair mechanism found for FapydG [12] rationalizes the observation that Fpg can excise nonpolar analogues of 8OG [42]. This idea obtains new support by our findings in the present work. In addition, it was shown that FapydA is only recognized by Fpg in the *anti*-conformation [10] which is rationalized by our work on FapydA (present work) and FapydG [12] since both can indeed be repaired in this conformation. Furthermore, our work shows that the observed Schiff base product of base excision by Fpg is only one possible structural isomer with a closed-ribose structure being the other. As a consequence, computational work has focused so far only on the Schiff base as a starting point based on X-ray structures [46, 47], whereas the equilibrium with the new closed-ribose structure offers new possibilities for future studies to tackle the mechanism of the  $\beta$ - and  $\delta$ -lyase reaction that follows base excision.

## **ASSOCIATED CONTENT**

The figures were created using VMD [19]. For further details see Supplementary Information.

## **AUTHOR INFORMATION**

### **Corresponding Author**

christian.ochsenfeld@uni-muenchen.de

### **Notes**

The authors declare no competing financial interest.

## **ACKNOWLEDGEMENTS**

Financial support is acknowledged by the DFG funding initiative SFB749 (TP C7) and the Excellence Cluster EXC114 (CIPSM).

## References

1. David, S. S., O'Shea, V. L. & Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **447**, 941–950 (2007).
2. Greenberg, M. M. The formamidopyrimidines: Purine lesions formed in competition with 8-oxopurines from oxidative stress. *Acc. Chem. Res.* **45**, 588–597 (2012).
3. Dizdaroglu, M., Kirkali, G. & Jaruga, P. Formamidopyrimidines in DNA: Mechanisms of formation, repair, and biological effects. *Free Radic. Biol. Med.* **45**, 1610–1621 (2008).
4. Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–1214 (2003).
5. Wallace, S. S. Base excision repair: A critical player in many games. *DNA Repair (Amst.)*. **19**, 14–26 (2014).
6. O'Connor, T. R. & Laval, J. Physical association of the 2,6-diamino-4-hydroxy-5N-formamidopyrimidine-DNA glycosylase of *Escherichia coli* and an activity nicking DNA at apurinic/apyrimidinic sites. *Proc. Natl. Acad. Sci. USA* **86**, 5222–5226 (1989).
7. Tchou, J. *et al.* 8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 4690–4694 (1991).
8. Boiteux, S., Gajewski, E., Laval, J. & Dizdaroglu, M. Substrate specificity of the *Escherichia coli* Fpg protein (formamidopyrimidine-DNA glycosylase): excision of purine lesions in DNA produced by ionizing radiation or photosensitization. *Biochemistry* **31**, 106–110 (1992).
9. Coste, F. *et al.* Structural basis for the recognition of the FapydG lesion (2,6-diamino-4-hydroxy-5-formamidopyrimidine) by formamidopyrimidine-DNA glycosylase. *J. Biol. Chem.* **279**, 44074–44083 (2004).
10. Perlow-Poehnelt, R. A., Zharkov, D. O., Grollman, A. P. & Broyde, S. Substrate discrimination by formamidopyrimidine-DNA glycosylase: Distinguishing interactions within the active site. *Biochemistry* **43**, 16092–16105 (2004).
11. Sadeghian, K. *et al.* Ribose-Protonated DNA Base Excision Repair: A Combined Theoretical and Experimental Study. *Angew. Chem. Int. Ed.* **126**, 10208–10212 (2014).
12. Blank, I. D., Sadeghian, K. & Ochsenfeld, C. A Base-Independent Repair Mechanism for DNA Glycosylase - No Discrimination Within the Active Site. *Sci. Rep.* **5**, 10369 (2015).
13. Delaney, M. O., Wiederholt, C. J. & Greenberg, M. M. FapydA induces nucleotide misincorporation translesionally by a DNA polymerase. *Angew. Chem. Int. Ed.* **41**, 771–773 (2002).

14. Tudek, B. Imidazole ring-opened DNA purines and their biological significance. *J. Biochem. Mol. Biol.* **36**, 12–19 (2003).
15. Grziewicz, M. A. *et al.* Fapyadenine is a moderately efficient chain terminator for prokaryotic DNA polymerases. *Free Radic. Biol. Med.* **28**, 75–83 (2000).
16. Krishnamurthy, N., Haraguchi, K., Greenberg, M. M. & David, S. S. Efficient removal of formamidopyrimidines by 8-oxoguanine glycosylases. *Biochemistry* **47**, 1043–1050 (2008).
17. Fromme, J. C. & Verdine, G. L. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat. Struct. Biol.* **9**, 544–552 (2002).
18. Senn, H. M. & Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **48**, 1198–1229 (2009).
19. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
20. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
21. Case, D. A. *et al.* *AMBER 11*. Univ. California, San Fr. (2010).
22. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
23. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260 (2006).
24. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
25. Sherwood, P. *et al.* QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct.: THEOCHEM* **632**, 1–28 (2003).
26. Shao, Y. *et al.* Advances in methods and algorithms in a modern quantum chemistry program package. *Physical Chemistry Chemical Physics* **8**, 3172–3191 (2006).
27. Vosko, S. H., Wilk, L. & Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **58**, 1200–1211 (1980).
28. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
29. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).

30. Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **33**, 8822–8824 (1986).
31. 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.* **132**, 154104 (2010).
32. Goerigk, L. & Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **13**, 6670–6688 (2011).
33. Kästner, J. *et al.* DL-FIND: an open-source geometry optimizer for atomistic simulations. *J. Phys. Chem. A* **113**, 11856–11865 (2009).
34. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
35. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
36. Bas, D. C., Rogers, D. M. & Jensen, J. H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins: Struct., Funct., Genet.* **73**, 765–783 (2008).
37. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct., Funct., Genet.* **61**, 704–721 (2005).
38. Zharkov, D. O., Rieger, R. A., Iden, C. R. & Grollman, A. P. NH<sub>2</sub>-terminal Proline Acts as a Nucleophile in the Glycosylase/AP-Lyase Reaction Catalyzed by Escherichia coli Formamidopyrimidine-DNA Glycosylase (Fpg) Protein. *J. Biol. Chem.* **272**, 5335–5341 (1997).
39. McCullough, A. K., Dodson, M. L. & Lloyd, R. S. Initiation of base excision repair: glycosylase mechanisms and structures. *Annu. Rev. Biochem.* **68**, 255–285 (1999).
40. Sadeghian, K. & Ochsenfeld, C. Unraveling the Base Excision Repair Mechanism of Human DNA Glycosylase. *J. Am. Chem. Soc.* **137**, 9824–9831 (2015).
41. Barnes, D. E. & Lindahl, T. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu. Rev. Genet.* **38**, 445–476 (2004).
42. McKibbin, P. L., Kobori, A., Taniguchi, Y., Kool, E. T. & David, S. S. Surprising repair activities of nonpolar analogs of 8-oxoG expose features of recognition and catalysis by base excision repair glycosylases. *J. Am. Chem. Soc.* **134**, 1653–1661 (2012).



43. Berti, P. J. & McCann, J. A. Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chem. Rev.* **106**, 506–555 (2006).
44. Hatahet, Z, Kow, Y. W., Purmal, A. A., Cunningham, R. P. & Wallace, S. S. New substrates for old enzymes. 5-Hydroxy-2'-deoxycytidine and 5-hydroxy-2'-deoxyuridine are substrates for Escherichia coli endonuclease III and formamidopyrimidine DNA N-glycosylase, while 5-hydroxy-2'-deoxyuridine is a substrate for uracil DNA N-glycos. *J. Biol. Chem.* **269**, 18814–18820 (1994).
45. Jurado, J, Saparbaev, M, Matray, T. J., Greenberg, M. M. & Laval, J. The ring fragmentation product of thymidine C5-hydrate when present in DNA is repaired by the Escherichia coli Fpg and Nth proteins. *Biochemistry* **37**, 7757–7763 (1998).
46. Sowlati-Hashjin, S. & Wetmore, S. D. Computational Investigation of Glycosylase and  $\beta$ -Lyase Activity Facilitated by Proline: Applications to FPG and Comparisons to hOgg1. *J. Phys. Chem. B* **118**, 14566–14577 (2014).
47. Sowlati-Hashjin, S. & Wetmore, S. D. Quantum mechanical study of the  $\beta$ - and  $\delta$ -lyase reactions during the base excision repair process: application to FPG. *Phys. Chem. Chem. Phys.* **17**, 24696–24706 (2015).

Author contribution statement:

T.M., I.D.B., and C.O. wrote the main manuscript text and T.M. prepared the figures. All authors reviewed the manuscript.

## Supplementary Information

### DNA Glycosylase Repairs FapydA Base-Independently — Schiff Base in Equilibrium with Closed-Ribose Structure

\*Correspondence to christian.ochsenfeld@uni-muenchen.de

Tobias Moeller<sup>a,b)</sup>, Iris D. Blank<sup>a,b)</sup>, and Christian Ochsenfeld<sup>a,b)</sup>

<sup>a)</sup> Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU Munich), Butenandtstr. 7, D-81377 Munich, Germany

<sup>b)</sup> Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU Munich), Butenandtstr. 5-13, D-81377 Munich, Germany

## SI-1 Atom distances during last step of repair mechanism

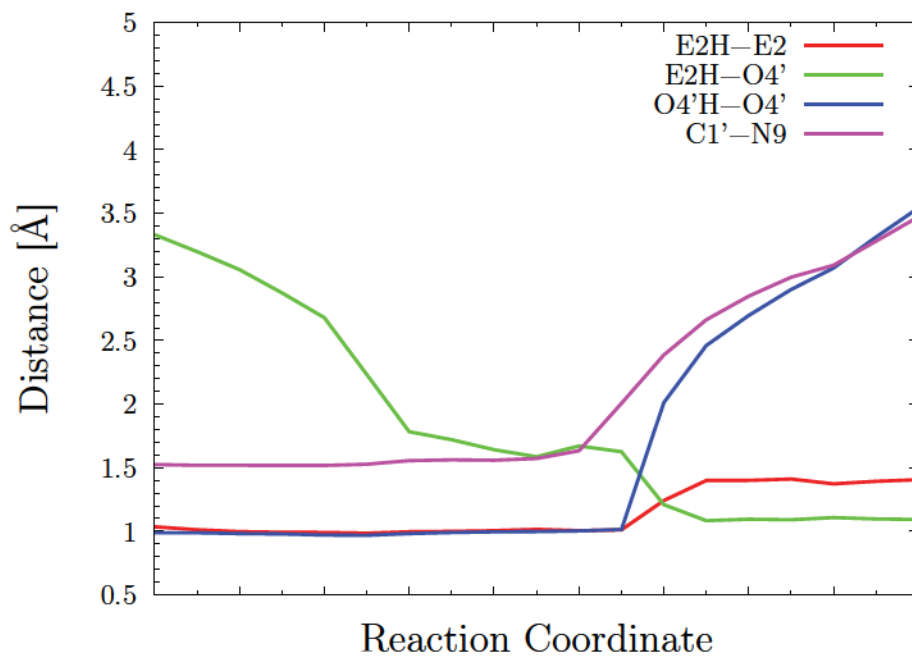


Figure 1: Atom distances during the last step of the reaction, the proton transfer from  $O_{4'}$  to N9 and from E2 to  $O_{4'}$ .

## SI-2 Structure of Schiff base product

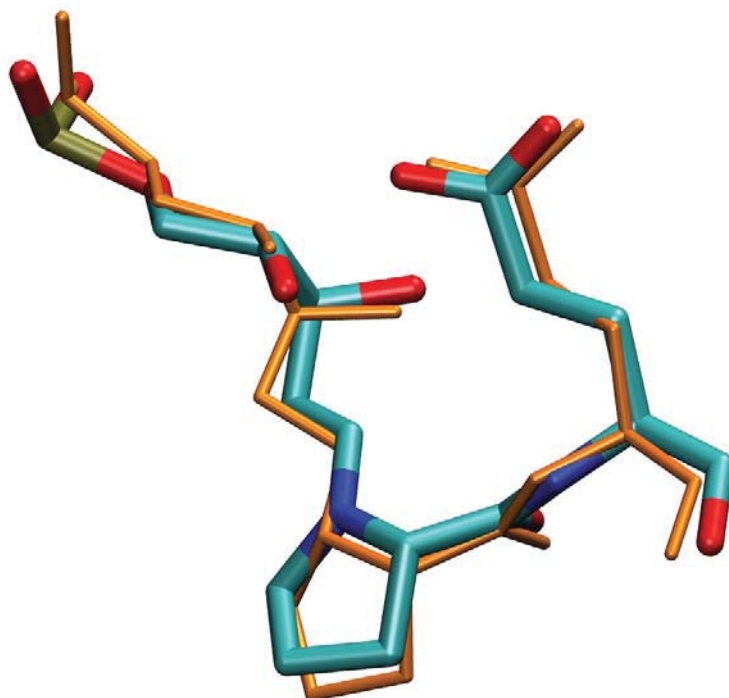


Figure 2: Alignment of the Schiff base product obtained from calculations (shown in atomic colors) and the corresponding X-ray structure [PDB code:1L1Z] [1] (shown in orange).

### SI-3 Investigation of base-protonated mechanism

While it was found for 8OG [2] as well as for FapydG [3] that the base-protonated pathway [4] is energetically much less favourable than the ribose-protonated [5] pathway, this matter has not been studied for FapydA. The base-protonated pathway suggests that in the first step P1 nucleophilically attacks C<sub>1'</sub>, which leads to the base leaving with the ribose ring remaining closed. This would, however, necessitate protonation of FapydA in order to make it a favourable leaving group since otherwise N9 would be negatively charged while C<sub>1'</sub> would be positively charged. This oxocarbenium ion has already been found to be energetically unfavourable [2]. Therefore, we investigated protonation of the base as a first step. The calculations gave barriers that were significantly higher (over 60 kcal/mol) than the one found for a ribose-protonated mechanism. The energetics for the glycosidic bond breakage of the reaction is shown in Fig. 3. The simulated reaction step was a nucleophilic attack of P1 on C<sub>1'</sub> with subsequent protonation of N9 by P1 and leaving of the base.

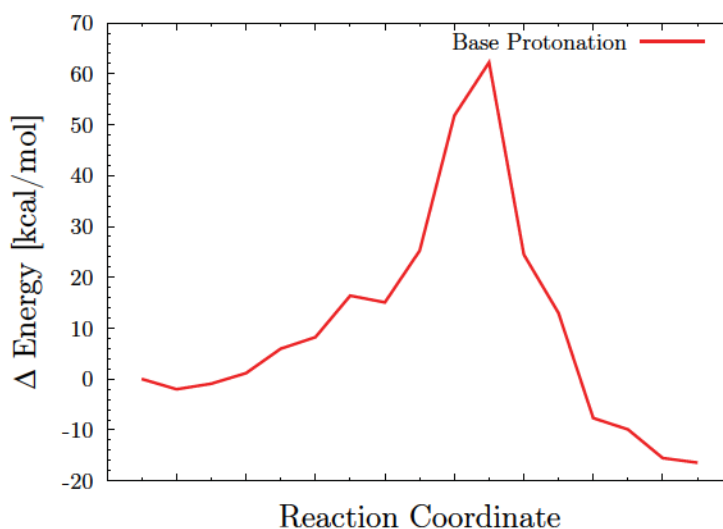


Figure 3: Energetics of a base-protonated pathway. The simulated reaction step was a nucleophilic attack of P1 on C<sub>1'</sub> with subsequent protonation of N9 by P1 and leaving of the base.

## SI-4 Detailed Methods Section

### SI-4.1 Details for Molecular Dynamics Simulations

To prepare the system, it was energy minimized in two steps in the NVT ensemble, relaxing different degrees of freedom, using the conjugate gradient algorithm: (1) only solvent with rest of the system frozen (10000 steps), (2) all atoms with the solvent being free and the rest being constrained (20000 steps). Afterwards the system was heated up to 300 K within 30 ps with the solute constrained. After heating up the system it was equilibrated in two steps. In a first step, the system is propagated for 200 ps with the solute being constrained. In the second step, the system is propagated for 200 ps in the NPT ensemble using the Langevin piston Nosé-Hoover method [6, 7] and the SHAKE algorithm [8]. The constraints on the solute are reduced step by step down to zero. Subsequently, a production run was performed for 60 ns. The RMSD for the protein backbone is shown in Fig. 4.

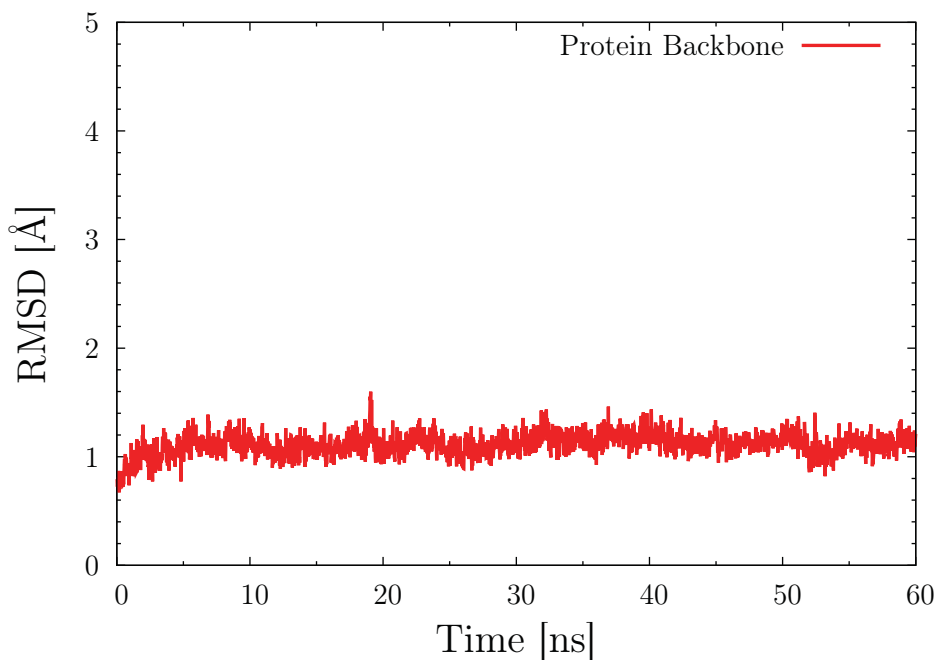


Figure 4: RMSD of the 60 ns production of the protein backbone.

## SI-4.2 Details for QM/MM calculations

The whole equilibrated structure was used (52214 atoms) with 10 Å around the selection of FapydA, E2, E5, and P1 as the relaxed region (2267 atoms) and 86 atoms in the QM region (FapydA, E2, E5, P1).

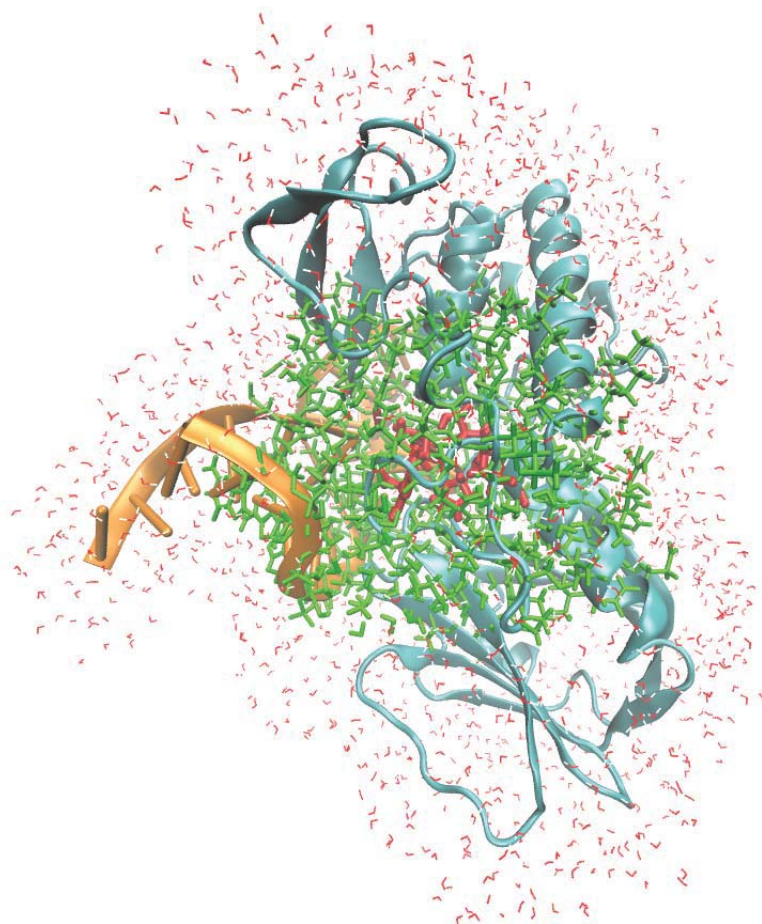


Figure 5: DNA duplex containing FapydA in complex with DNA repair enzyme Fpg. The QM region consisting of 86 atoms is shown in red, the active region consisting of 2267 atoms in green. For better visibility only water within 5Å of protein and DNA are shown.



## References

1. Fromme, J. C. & Verdine, G. L. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat. Struct. Biol.* **9**, 544–552 (2002).
2. Sadeghian, K. *et al.* Ribose-Protonated DNA Base Excision Repair: A Combined Theoretical and Experimental Study. *Angew. Chem. Int. Ed.* **126**, 10208–10212 (2014).
3. Blank, I. D., Sadeghian, K. & Ochsenfeld, C. A Base-Independent Repair Mechanism for DNA Glycosylase - No Discrimination Within the Active Site. *Sci. Rep.* **5**, 10369 (2015).
4. Zharkov, D. O., Rieger, R. A., Iden, C. R. & Grollman, A. P. NH<sub>2</sub>-terminal Proline Acts as a Nucleophile in the Glycosylase/AP-Lyase Reaction Catalyzed by Escherichia coli Formamidopyrimidine-DNA Glycosylase (Fpg) Protein. *J. Biol. Chem.* **272**, 5335–5341 (1997).
5. McCullough, A. K., Dodson, M. L. & Lloyd, R. S. Initiation of base excision repair: glycosylase mechanisms and structures. *Annu. Rev. Biochem.* **68**, 255–285 (1999).
6. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
7. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
8. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).



- V.3 Paper [3]: "Ribose-protonated DNA base-excision repair: a combined theoretical and experimental study", K. Sadeghian, D. Flaig, I. D. Blank, S. Schneider, R. Strasser, D. Stathis, M. Winnacker, T. Carell and C. Ochsenfeld, *Angew. Chem. Int. Ed.* 53, 10044 (2014).**



# Ribose-Protonated DNA Base Excision Repair: A Combined Theoretical and Experimental Study\*\*

Keyarash Sadeghian, Denis Flaig, Iris D. Blank, Sabine Schneider, Ralf Strasser, Dimitrios Stathis, Malte Winnacker, Thomas Carell,\* and Christian Ochsenfeld\*

**Abstract:** Living organisms protect the genome against external influences by recognizing and repairing damaged DNA. A common source of gene mutation is the oxidized guanine, which undergoes base excision repair through cleavage of the glycosidic bond between the ribose and the nucleobase of the lesion. We unravel the repair mechanism utilized by bacterial glycosylase, MutM, using quantum-chemical calculations involving more than 1000 atoms of the catalytic site. In contrast to the base-protonated pathway currently favored in the literature, we show that the initial protonation of the lesion's ribose paves the way for an almost barrier-free glycosidic cleavage. The combination of theoretical and experimental data provides further insight into the selectivity and discrimination of MutM's binding site toward various substrates.

While DNA repair is of central importance for the stability of the genome of organisms,<sup>[1,2]</sup> the underlying processes such as lesion recognition and repair are still poorly understood. Repair enzymes have to recognize single lesions within a vast majority of undamaged bases. The most prominent endogenous DNA damage is the oxidized guanine, 7,8-dihydro-8-oxoguanine (8OG). 8OG has the ability to form a Hoogsteen base pair with adenine, which is the basis for the G:C→A:T transversion mutation (Figure 1).<sup>[3]</sup>

Despite the existence of crystal structures showing 8OG-containing DNA in complex with the catalytically incompetent repair enzymes, one of the major unsolved questions is how 8OG is specifically recognized and excised in the

genome.<sup>[4–12]</sup> It is still not clear how the enzymes are able to exquisitely select 8OG and avoid harmful excision of canonical guanine (G). In bacteria, 8OG is repaired by MutM, also known as formamidopyrimidine glycosylase (Fpg).<sup>[13]</sup> MutM first recognizes 8OG and then guides it into the active site (lesion recognition complex, LRC).<sup>[14]</sup> The glycosidic bond is subsequently cleaved, creating an abasic intermediate. A crystal structure trapped after sodium borohydride treatment reveals that the enzyme, at some point, forms a covalent bond (Schiff base) between the anomeric center of 8OG and a proline residue in the active site.<sup>[15]</sup> However, the chemistry leading to this intermediate and contributing to the ability of the enzyme to discriminate G from 8OG is still unknown.

The currently favored hypothesis is that MutM protonates the 8OG base, thereby converting it into a good leaving group, which facilitates the nucleophilic attack of the anomeric C1' carbon by the proline nucleophile.<sup>[16]</sup> The heterocycle of 8OG is more electron-rich than that of G, hence making the former a better leaving group. Using newly developed linear-scaling quantum-chemical methods we report a thorough computational investigation of the repair mechanism of MutM. We address MutM's exquisite 8OG selectivity as well as the specific catalytic role of central amino acids in the active site. The results of our theoretical study are in full agreement with the presented experimental repair studies with different G derivatives and a new crystal structure showing 8OG embedded in a DNA duplex in complex with a wild-type and hence catalytically competent MutM (for details see the Supporting Information, SI). The theoretical treatment of the complex DNA-repair mechanism is a severe challenge for quantum chemical calculations, forcing us to utilize newly developed linear-scaling quantum-chemical methods.<sup>[17,18]</sup> These new methods enable us to calculate interaction energies with more than 1000 atoms of the active site and to study the full repair pathway with more than 500 atoms in the QM region of the QM/MM (quantum mechanics/molecular mechanics) approach. Such large QM regions have recently been shown to be crucial for reliable QM/MM results in other molecular systems.<sup>[19–23]</sup> The linear-scaling QM/MM calculations make it possible for the first time to clarify the repair mechanism and provide insight into the interaction energies between MutM and various substrates. We have used the dispersion-corrected BLYP-D3 density functional,<sup>[24–26]</sup> the split-valence polarized (SVP) basis set,<sup>[27]</sup> and the AMBER force field<sup>[28]</sup> to study the reaction pathways. The dispersion-corrected Hartree-Fock approximation was used for the QM part of the QM/MM interaction energy calculations (for full details see SI-2).

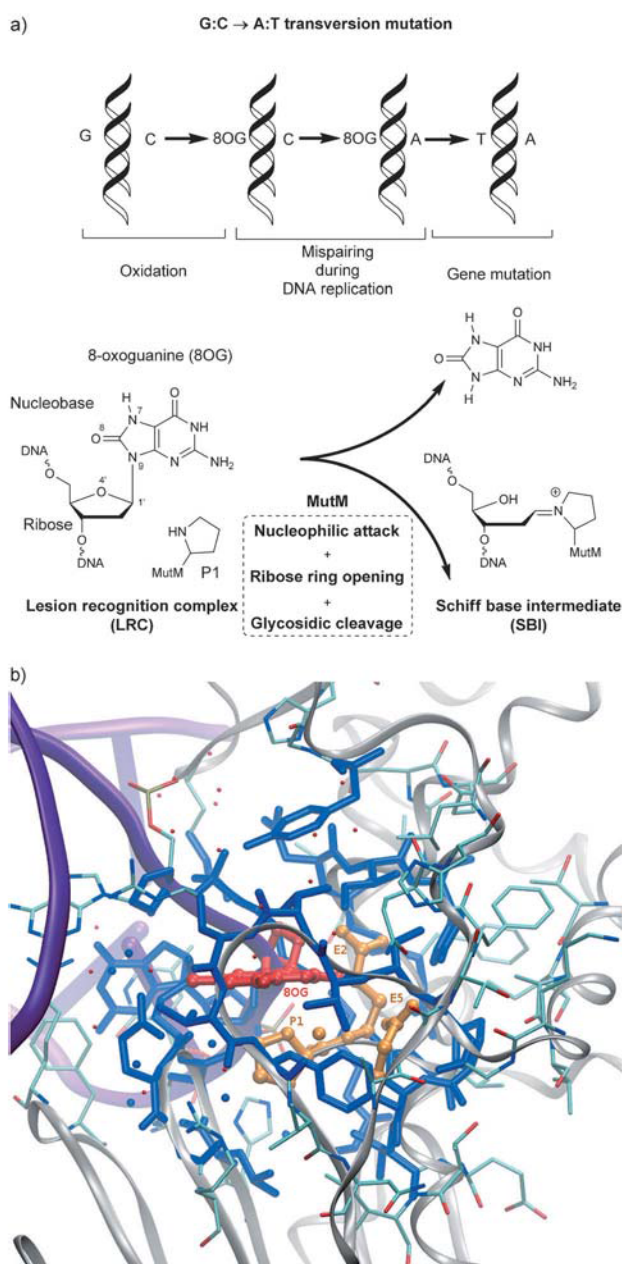
[\*] Dr. K. Sadeghian, Dr. D. Flaig, I. D. Blank, Prof. Dr. C. Ochsenfeld  
Chair of Theoretical Chemistry, Department of Chemistry  
University of Munich (LMU)  
Butenandtstrasse 7, 81377 Munich (Germany)  
Dr. K. Sadeghian, Dr. D. Flaig, I. D. Blank, Dr. S. Schneider,  
Dr. R. Strasser, Dr. D. Stathis, Dr. M. Winnacker, Prof. Dr. T. Carell,  
Prof. Dr. C. Ochsenfeld  
Center for Integrated Protein Science (CIPSM) at the Department of  
Chemistry, University of Munich (LMU)  
Butenandtstrasse 5–13, 81377 Munich (Germany)  
E-mail: thomas.carell@uni-muenchen.de  
christian.ochsenfeld@uni-muenchen.de

Dr. S. Schneider  
Department of Chemistry, Chair of Biochemistry  
TU München, Lichtenbergstrasse 4, 85747 Garching (Germany)

[\*\*] We acknowledge financial support by the DFG funding initiatives SFB749 (TP A4 and C7) and the Excellence Cluster EXC114 (CIPSM).

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201403334>.





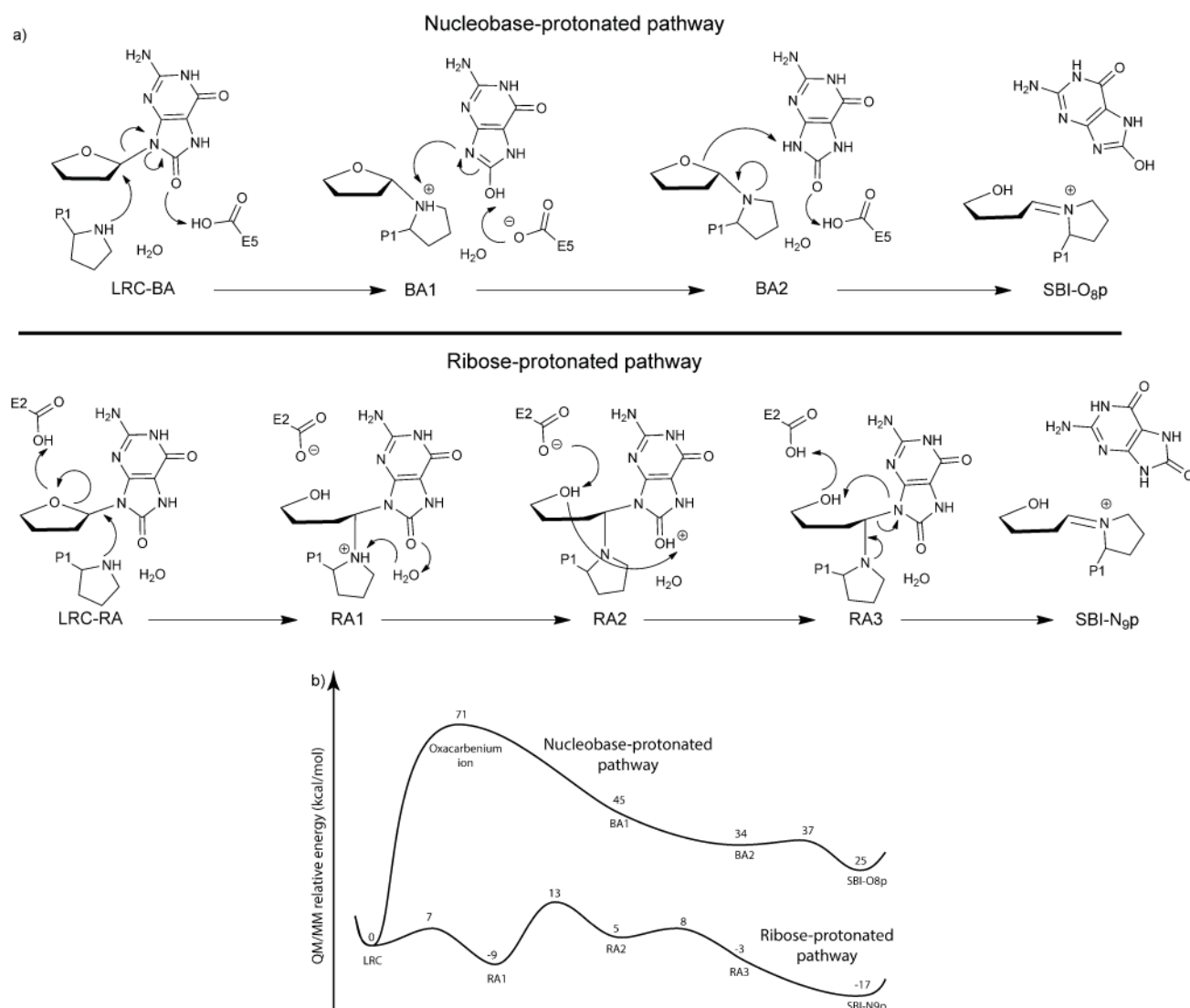
**Figure 1.** a) Top: 8OG can form a Hoogsteen base pair with A during DNA replication leading to G:C→A:T transversion mutations. Bottom: schematic representation of the glycosylase activity of MutM connecting the LRC and SBI states. b) A close-up view of the MutM-DNA complex showing the QM regions used in our QM/MM calculations. The QM region used to investigate the repair pathway includes not only 8OG (red) and the P1, E2, and E5 (orange) of MutM, but also the catalytic water (W) and all residues (blue) directly interacting with these fragments (→ QM region=509 atoms). For the calculations of the QM/MM interaction energies, we have extended the QM region to 1115 atoms, which are shown in green. Hydrogens are omitted for clarity. DNA and protein backbone are shown in pink and silver ribbon representations, respectively.

The currently preferred nucleobase-protonated pathway involves the initial cleavage of the glycosidic bond after protonation/H-bonding of the 8OG base (Figure 2a). This protonation increases the leaving group activity of the

nucleobase. The subsequently formed oxycarbenium ion is thought to be stabilized by the E2 residue. The essential proline nucleophile, P1, then reacts with the oxycarbenium ion as depicted in Figure 2a to form an imine that was trapped using sodium borohydride. We denote the intermediates of this mechanism as BA1 and BA2. The final product is the Schiff base intermediate (SBI), which includes the O8-protonated tautomer (SBI-O8p). Although this mechanism is in principle chemically feasible, our QM/MM data, shown in Figure 2b, mark this mechanism as surprisingly unlikely: The highest energy point found along this pathway is 71 kcalmol<sup>-1</sup> above the LRC. We also confirmed the BLYP-D3/MM energetics by comparison to single-point RIMP2/MM and B2PLYP/MM results (see SI-3). In addition, we find that the critical intermediates BA1 and BA2, featuring a bond between P1 and the ribose, lie energetically so high (45 and 34 kcalmol<sup>-1</sup> above the LRC, respectively) that both of them are thermodynamically not accessible. The 8OG base may assist the proton transfer required to open up the ribose ring by temporarily accepting the proton of P1 (→ barrierless formation of the BA2 state). However, also the SBI-O8p state is energetically so unfavorable (+25 kcalmol<sup>-1</sup>) that we have to exclude a glycosylase-driven repair reaction that is initiated by conversion of 8OG into a better leaving unit. For a brief comparison between the currently favored 8OG repair mechanism and the repair of uracil by the DNA uracil glycosylase<sup>[29]</sup> see SI-3.

In light of the high energies observed for the intermediates of the base-protonated pathway, we investigated other reaction pathways. We considered a mechanism, in part proposed earlier,<sup>[30]</sup> in which the nucleobase does not function as an electron-accepting unit (after protonation) but as an electron-donating substructure (Figure 2a). This ribose-protonated pathway involves the opening of the ribose as the first step. The latter is facilitated by protonated E2, a known critical residue in the active site. The ribose ring-opening is accompanied by the nucleophilic attack of P1 leading to the first intermediate, RA1 (Figure 2a). P1 may undergo deprotonation, forming the second intermediate RA2, followed by the reprotonation of E2 (→RA3). The final step in this pathway is the cleavage of the glycosidic bond to finally give the iminium ion (SBI-N9).

In contrast to the first mechanism, in the ribose-centered scenario the intended glycosidic cleavage takes place at the very last stage. The QM/MM data show clearly that this pathway is energetically more favorable throughout, with an indeed almost barrierless glycosidic bond cleavage reaction. The highest energy point on this path is only 13 kcalmol<sup>-1</sup> above the LRC and involves a proton transfer from P1 to the O8 position of 8OG. Also the final SBI state (SBI-N9p) is 17 kcalmol<sup>-1</sup> below the LRC, implying that the overall reaction along the ribose-protonated pathway is thermodynamically significantly more favored than the nucleobase-protonated one. The mechanism also explains the eminent function of E2 in the active site. Site-directed mutagenesis showed that in the E2Q mutant of MutM the repair function is fully abolished.<sup>[14]</sup> In the ribose-protonated mechanism the protonated form of E2 is essential for enabling ribose ring opening, hence allowing the nucleophilic attack of P1.



**Figure 2.** Overview of the nucleobase- and ribose-protonated pathways. a) Schematic representation of the studied mechanisms. b) Our QM/MM data obtained for both pathways. The QM/MM (BLYP-D3/AMBER) structure optimizations performed to obtain these QM/MM relative energies (with respect to the LRC) involve 509 atoms in the QM region. On the base-protonated pathway, we locate a potential oxacarbenium ion, which also has been discussed as a reaction intermediate in the context of the base excision repair by uracil DNA glycosylase.<sup>[29]</sup> For 8OG repair, however, we find this ion to be a very high-energy transition state. Furthermore, we observe that a single water molecule, directly interacting with P1, can play a crucial catalytic role in the ribose-protonated pathway. It stabilizes the RA1 intermediate (Figure SI-16) and mediates the proton transfer to make the RA2 formation more feasible. The presence of such a water molecule replacing the R79 residue is motivated in Section SI-1. The influence of zero-point energy correction is found to be minor (see SI-4).

The rate-determining step of the ribose-protonated mechanism is the proton transfer step connecting the RA1 ( $-9 \text{ kcal mol}^{-1}$ ) and RA2 ( $5 \text{ kcal mol}^{-1}$ ) structures. As evident from Figure 2a, the energetically highest point on this path is  $13 \text{ kcal mol}^{-1}$  above the LRC (i.e.  $22 \text{ kcal mol}^{-1}$  above RA1) and connects the RA1 and RA2 intermediates. At the first glance, these values may seem high compared with metabolic enzymes; however, as pointed out by O'Brien,<sup>[31]</sup> DNA repair operates under different kinetics: it is a rare event compared with the synthesis of relatively large amounts of substrates processed by metabolic pathways.

The ribose-protonated pathway presented here in full detail revives an idea proposed earlier,<sup>[30]</sup> which seems to receive no attention in recent literature. A possible reason for

neglecting this idea is that a base-invariant mechanism implies that any nucleobase (even G) that can find its way into the active site can, in principle, be cleaved from DNA. For this to happen, G must first successfully pass through both the base-encounter and the base-extrusion steps before reaching the base-stabilization stage (LRC in the case of 8OG). However, experimental and theoretical studies<sup>[7,8]</sup> suggest that MutM discriminates G already at the early encounter and extrusion stages. Our results, presented further below, strongly support the idea that MutM can also differentiate 8OG from G, even at the final LRC stage.

We have chosen a combination of experimental and computational approaches to probe the extent of MutM's active site discrimination towards various substrates. On the

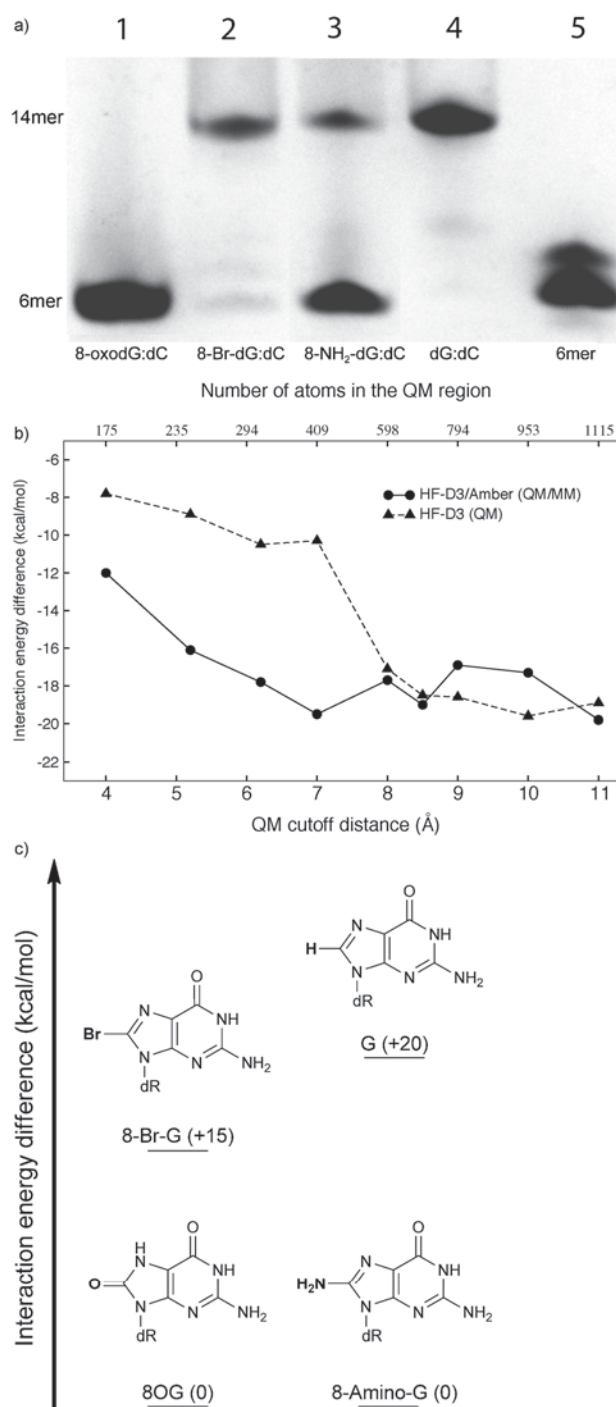


experimental side, we synthesized a series of G derivatives with different substituents at C8. The chemical groups introduced at this position are expected to influence the electronic density in the vicinity of the glycosidic bond. The structures of these G derivatives and the repair data obtained with short single-stranded oligonucleotides, in which the 8OG analogues were incorporated, are depicted in Figure 3c. Our data show fast repair of the 8OG-containing oligonucleotide (Figure 3a, lane 1). It is clearly visible that the 8-Br-G and canonical G are not repaired (lanes 2 and 4). Interestingly, we observe a fast repair of the  $\text{NH}_2$ -substituted 8OG analogue (lane 3). The electron-donating amino group increases the electron density of the G derivative thereby converting the base into a better electron donor as needed for the ribose-protonated mechanism. Apparently the 8-Br-G, which is even less electron-rich than G, is not repaired because it is an inadequate electron donor to support the ribose-protonated process. Besides these electronic properties, the binding of the compounds into the active site also needs to be considered.

To address the issue of MutM discrimination computationally, we performed QM/MM calculations, which allow us to compare the interaction of MutM with the experimentally investigated substrates. Our HF-D3/MM calculations show that G is discriminated at the LRC stage by 20 kcal mol<sup>-1</sup>. The mean absolute error of the HF-D3 method has been benchmarked to below 1 kcal mol<sup>-1</sup>.<sup>[32]</sup> To check the convergence of the obtained value with respect to QM size (Figure 3b), we included up to 1115 atoms into the QM region (see SI-2). To converge the G discrimination within 2 kcal mol<sup>-1</sup> with respect to the QM size, one must include at least 600 atoms into the QM region. QM-only calculations exhibit a much slower convergence as depicted in Figure 3b, so that even more atoms need to be included in the QM sphere for converged results. Our calculations indicate that the S219 residue, in H-bonding interaction with the substrate, is mainly responsible for the discrimination.

Furthermore, we have repeated these QM/MM calculations by replacing 8OG with the 8-NH<sub>2</sub>-G and 8-Br-G substrates, thereby assuming the *syn* binding mode as already observed for 8OG (see SI-5 for further details). The QM/MM calculations yield very similar binding affinities for 8-NH<sub>2</sub>-G and 8OG (Figure 3). The 8-Br-G substrate on the other hand is 15 kcal mol<sup>-1</sup> less stable than 8OG and therefore less likely to be bound to MutM in the same orientation as 8OG. Overall, MutM seems to be selective when it comes to stabilizing substrates in its binding site. To gain deeper insight into the exact arrangements of the important residues E2 and E5 relative to the lesion, we solved a crystal structure of MutM in complex with carbocyclic 8OG-containing DNA (PDB: 4CIS; see SI-7 for further details).

In conclusion, our linear-scaling QM/MM calculations with large QM regions (more than 500 atoms for the pathway and more than 1000 atoms in computing interaction energies) unravel the base excision repair mechanism for the glycosylase activity of the bacterial DNA repair enzyme, MutM. We show that through the opening of the ribose ring of 8OG, the final cleavage of the glycosidic bond becomes an almost barrier-free process. Our proposed mechanism is in full agreement with the experimental observations, in which the



**Figure 3.** Repair activity (experimental data) and binding discrimination (theoretical data) of MutM toward different substrates. a) Activity assay of MutM with C8-modified nucleobases: 1) 8OG:C, 2) 8-Br-G:C, 3) 8-NH<sub>2</sub>-G:C, 4) G:C, and 5) single-stranded 6-mer DNA (see SI-6). b) Convergence of the interaction energy difference (kcal mol<sup>-1</sup>) between 8OG versus G with the protein environment (negative values indicate the preference of binding of MutM toward 8OG). The fluctuations (in the order of 2 kcal mol<sup>-1</sup>) observed in the QM/MM values for  $R > 8$  Å are due to the inclusion of charged fragments in the QM region and their interaction with the remaining point-charge field (see SI-5 for the list of fragments). Such a long-range effect on the calculated QM/MM energies emphasizes the choice of sufficiently large QM regions in QM/MM calculations. c) QM/MM MutM–DNA interaction energy difference for different G derivatives, with 8OG–MutM taken as reference.



modifications of the ribose ring or mutations of the amino acids directly interacting with the ribose prohibit the glycosylase activity of MutM. We have computed the extent of discrimination by MutM towards different substrates by calculating the total interaction energies. We have shown that the catalytic site of MutM imposes considerable discrimination towards G, hence providing further support for the proposed ribose-protonated mechanism. This pathway is in agreement with the experimental activity data demonstrating MutM's selectivity in binding and repair affinity.

Received: March 14, 2014

Revised: May 12, 2014

Published online: July 25, 2014

**Keywords:** DNA repair · molecular modeling · oxidative damage · quantum chemistry · ribose

- [1] E. C. Friedberg, *Nature* **2003**, *421*, 436.
- [2] T. Lindahl, R. D. Wood, *Science* **1999**, *286*, 1897.
- [3] T. Lindahl, *Nature* **1993**, *362*, 709.
- [4] J. C. Fromme, G. L. Verdine, *Nat. Struct. Biol.* **2002**, *9*, 544.
- [5] P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine, X. S. Xie, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5752.
- [6] A. Banerjee, W. L. Santos, G. L. Verdine, *Science* **2006**, *311*, 1153.
- [7] Y. Qi, M. C. Spong, K. Nam, A. Banerjee, S. Jiralerspong, M. Karplus, G. L. Verdine, *Nature* **2009**, *462*, 762.
- [8] Y. Qi, M. C. Spong, K. Nam, M. Karplus, G. L. Verdine, *J. Biol. Chem.* **2010**, *285*, 1468.
- [9] K. Nam, G. L. Verdine, M. Karplus, *J. Am. Chem. Soc.* **2009**, *131*, 18208.
- [10] A. Banerjee, W. Yang, M. Karplus, G. L. Verdine, *Nature* **2005**, *434*, 612.
- [11] S. D. Bruner, D. P. Norman, G. L. Verdine, *Nature* **2000**, *403*, 859.
- [12] J. C. Fromme, S. D. Bruner, W. Yang, M. Karplus, G. L. Verdine, *Nat. Struct. Biol.* **2003**, *10*, 204.
- [13] S. S. David, V. L. O'Shea, S. Kundu, *Nature* **2007**, *447*, 941.
- [14] J. C. Fromme, G. L. Verdine, *J. Biol. Chem.* **2003**, *278*, 51543.
- [15] R. Gilboa, D. O. Zharkov, G. Golan, A. S. Fernandes, S. E. Gerchman, E. Matz, J. H. Kycia, A. P. Grollman, G. Shoham, *J. Biol. Chem.* **2002**, *277*, 19811.
- [16] P. J. Berti, J. A. B. McCann, *Chem. Rev.* **2006**, *106*, 506.
- [17] J. Kussmann, M. Beer, C. Ochsenfeld, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 614–636.
- [18] J. Kussmann, C. Ochsenfeld, *J. Chem. Phys.* **2013**, *138*, 134114.
- [19] C. V. Sumowski, C. Ochsenfeld, *J. Phys. Chem. A* **2009**, *113*, 11734.
- [20] L. Hu, P. Söderhjelm, U. Ryde, *J. Chem. Theory Comput.* **2011**, *7*, 761.
- [21] R.-Z. Liao, W. Thiel, *J. Chem. Theory Comput.* **2012**, *8*, 3793.
- [22] C. V. Sumowski, B. B. T. Schmitt, S. Schweizer, C. Ochsenfeld, *Angew. Chem.* **2010**, *122*, 10147; *Angew. Chem. Int. Ed.* **2010**, *49*, 9951.
- [23] D. Flaig, M. Beer, C. Ochsenfeld, *J. Chem. Theory Comput.* **2012**, *8*, 2260.
- [24] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- [25] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
- [26] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
- [27] A. Schäfer, H. Horn, R. Ahlrichs, *J. Chem. Phys.* **1992**, *97*, 2571.
- [28] D. Case, et al. "AMBER 10", Technical Report, University of California, San Francisco, **2008**.
- [29] A. R. Dinner, G. M. Blackburn, M. Karplus, *Nature* **2001**, *413*, 752.
- [30] A. K. McCullough, M. L. Dodson, R. S. Lloyd, *Annu. Rev. Biochem.* **1999**, *68*, 255.
- [31] P. J. O'Brien, *Chem. Rev.* **2006**, *106*, 720–752.
- [32] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.





Supporting Information

© Wiley-VCH 2014

69451 Weinheim, Germany

**Ribose-Protonated DNA Base Excision Repair: A Combined Theoretical and Experimental Study\*\***

*Keyarash Sadeghian, Denis Flaig, Iris D. Blank, Sabine Schneider, Ralf Strasser, Dimitrios Stathis, Malte Winnacker, Thomas Carell,\* and Christian Ochsenfeld\**

anie\_201403334\_sm\_miscellaneous\_information.pdf

## Contents

SI-1	Force-field simulation . . . . .	4
SI-1.1	Flexibility of MutM's active site . . . . .	4
SI-1.2	Force-field preparation of the R79in conformation . . . . .	7
SI-1.3	Force-field preparation of the R79out conformation . . . . .	9
SI-1.4	Preparation of R79outW structure . . . . .	9
SI-1.5	Structural analysis of the force-field trajectories . . . . .	10
SI-2	QM/MM setup . . . . .	14
SI-2.1	Choice of the QM-region . . . . .	15
SI-2.2	QM/MM structure optimization . . . . .	16
SI-2.3	QM/MM isomerization energies . . . . .	18
SI-2.4	Estimation of pKa values . . . . .	19
SI-3	Nucleobase-protonated pathway . . . . .	21
SI-3.1	Comparison with the mechanism of uracil glycosylase . . . . .	25
SI-4	Ribose-protonated pathway . . . . .	27
SI-4.1	QM/MM convergence with respect to QM-size . . . . .	33
SI-4.2	Zero-point energy corrections . . . . .	35
SI-5	Binding pocket discrimination . . . . .	37
SI-5.1	Total interaction energies . . . . .	37
SI-5.2	Investigation of individual residue contributions . . . . .	40
SI-5.3	Ribose-protonated pathway for other substrates . . . . .	44
SI-6	MutM activity assay . . . . .	48
SI-6.1	Cloning, protein expression and purification . . . . .	48
SI-6.2	Activity assays with MutM . . . . .	49

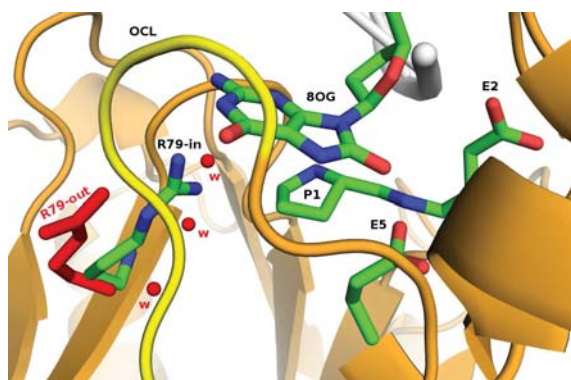
SI-7 Crystal structure of the carbocyclic 8OG -MutM complex . . . . .	51
SI-7.1 Co-crystallization, Data Collection, and Structure Determination	53

## SI-1 Force-field simulation

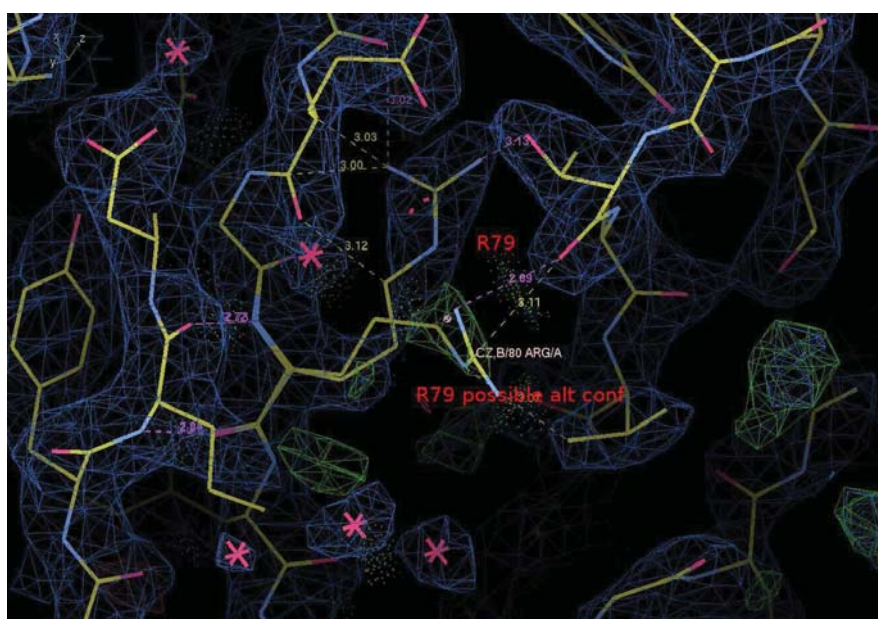
### SI-1.1 Flexibility of MutM's active site

In this work we have used the crystal structures with PDB-codes 1R2Y and 1R2Z [1] as model systems for our calculations. An overlay of these two structures is shown in Fig. SI-1(a). In the 1R2Y structure, with an 8OG substrate, R79 side chain is found to be deep inside the binding site. In the 1R2Z crystal structure, with dihydro-uracil (DHU) as substrate, however, R79 is pointing away from the binding site and water molecules are found there instead. The electron density map of the 1R2Y structure (Fig. SI-1(b)) also shows that there are two alternative positions for the R79 residue.

Further evidence for the presence of water molecules in the binding site as well as different orientation for the amino acid at the 79th position is provided by the crystal structure MutM from *E. coli*, trapped at the Schiff base intermediate (SBI), as depicted in Fig. SI-2(a). Here, the side chain of R78 of *E. coli* (corresponding to the R79 of MutM from *B. st.*) is pointing away from the lesion and is solvent-exposed. Instead of R78's side chain, three water molecules are found to occupy the very same portion of space in the active site.



(a)



(b)

Figure SI-1: (a) A close-up view at the binding site of *Bacillus stearothermophilus*-MutM trapped at the lesion recognition complex (LRC, PDB-code: 1R2Y) stage [1] (enzyme shown in orange and DNA in white with back-mutation of E2Q). Hydrogens are omitted for more clarity. Position of R79 and 3 water molecules (w) in the binding site of the 1R2Z structure are shown red (the DHU-substrate is not explicitly shown). The oxo-capping-loop (OCL), consisting of the residues 221-225, is highlighted in yellow. The R79-in and R79-out labels denote the two orientations of R79 found in the 1R2Y and 1R2Z structures, respectively. In the R79-out conformation, water molecules are found in the binding site. (b) The electron density map of the 1R2Y structure showing an alternative possible conformation for the R79 residue.



Also the crystal structure of MutM from *Lactococcus lactis* (*L. lactis*) bound to carbocyclic-FapydG (cFapydG) [2] provides strong evidence for the presence of water molecules in the active site (see Fig. SI-2(b)). Here, the K82 residue of *L. lactis* MutM, which corresponds to the R79 of *B. st.* MutM, is not resolved in the structure, hence indicating a rather dynamic orientation for this very residue.

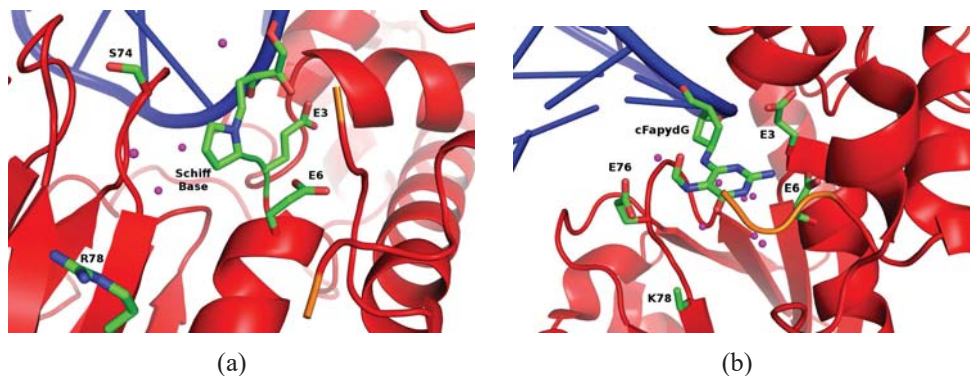


Figure SI-2: Close-up views of crystal structures of MutM-DNA complex showing the extent of water molecules (purple) penetrating the binding site. DNA is shown in blue and enzyme in red (a) Crystal structure of MutM from *E. coli* in complex with 8OG, where the repair process was trapped at the Schiff-base intermediate [3]. The side chain of the R78 residue (corresponding to the R79 of MutM from *B. st.*) is pointing away from the binding site. The binding site on the other hand is occupied by several water molecules. (b) The crystal structure of MutM from *Lactococcus lactis* in complex with cFapydG reported by Coste et. al. [2].

The open question is whether the dynamic 8OG capping loop (OCL, see Fig. SI-1(a)) is responsible for the movement of R79 and water penetration in the binding site. Since the answer to this question requires a separate study, we chose to investigate the repair mechanism with both R79 orientations found in the published crystal structures:

- R79 side chain found within the active site ( $\rightarrow$ **R79in**, Fig. SI-3(a))
- R79 found away from the active site ( $\rightarrow$ **R79out**, Fig. SI-3(b)).



- R79 found away from the active site and its position replaced by a few water molecules ( $\rightarrow$ **R79outW**, Fig. SI-3(c)).

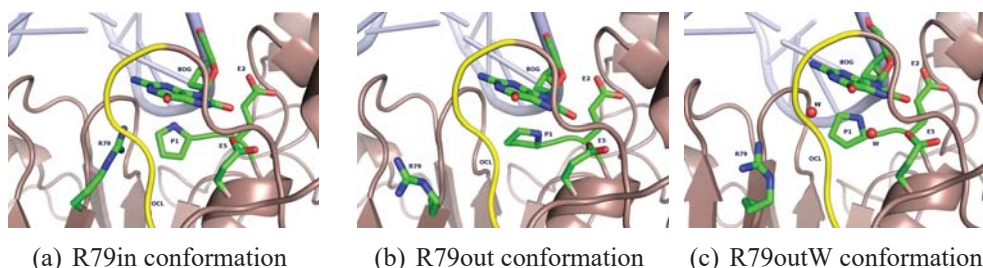


Figure SI-3: Close-up views of the model systems used for force-field simulations and subsequent QM/MM calculations. These model systems are based on the crystal structures 1R2Y and 1R2Z [1] where the R79 residue takes two different orientations, (a) "in", and (b) "out", towards the 8OG-substrate in the active site. We also consider the possibility of water molecules replacing R79 (c). The oxo-capping loop (OCL), consisting of residues 221-225 (see text, or Ref. [1]), is shown in yellow, enzyme in brown, and DNA in light blue.

## SI-1.2 Force-field preparation of the R79in conformation

The crystal structure with the PDB code 1R2Y was used as the starting point. The E2-residue was restored by manual modification of the PDB-file. The orientation of all Asn, Gln, and His residues were checked using the MolProbity server [4]. The LEAP-module of AMBER Tools [5] was used to neutralize the system by adding sodium ions. The neutralized system was then placed in a water box, thereby making sure that every solute molecule (enzyme, DNA, and counter ions) is covered by a solvent layer of at least 12 Å thickness. The parameters for the neutral proline residue and 8OG were taken from Refs. [6] and [7], respectively. The standard AMBER-10 force-field parameters and AMBER Tools [5] were used to parametrize the amino- and nucleic acids of the system. The TIP3 water model [8]

was used to describe the explicit solvation. The NAMD force-field engine [9] was used throughout this study.

The solvated and neutralised R79in structures, with either E2 or E5 protonated, were minimised in three steps. These steps involved sequential relaxation of (i) hydrogen atoms, (ii) water molecules, and finally (iii) all atoms. The all atom minimised structure were then heated up to 300 K in an NVT-ensemble, where a positional restraint of 1 kcal/mol/Å<sup>2</sup> on all atoms of the enzyme and DNA was applied. A time step of 2 fs was used throughout the simulation. Each system is then equilibrated for 300 ps in an NPT ensemble (1 atm pressure) where the restraints on the protein and DNA atoms are gradually switched off (reduction of 0.2 kcal/mol/Å<sup>2</sup> for every 20 ps). Once the restraints are set to zero, the equilibration stage (NPT-ensemble) is continued for further 60ps. The final structure, reached at the end of this equilibration stage was used for the QM/MM investigations (see further below). In order to analyse the structural stability of the solvated DNA-enzyme complexes with respect to both the protonation states of E2 and E5 residues, force field MD runs (6ns) were carried out for each system in an NVT-ensemble (at 300 K temperature).

All residues, except E2 and E5, were treated in their standard protonation state. For E2 and E5, we have considered the possibility of protonating either of these residues, where the other is treated in its carboxylate state. We have performed force-field MD simulations to check the structural stability of the binding site for either of the E2 and E5 protonation states. Since subsequent QM/MM calculations (Sec. SI-2.3) have shown that the E2-protonated state is energetically more stable, we will mainly concentrate on the results obtained from the E2-protonated setup. The E5-protonated conformations will only be discussed in terms of structural sta-

bility at the force-field level (see the analysis of the corresponding force-field MD trajectories). Unless stated otherwise, the QM/MM results presented in Sec. SI-3 and SI-4 are based on the setup where the E2-residue is protonated.

### **SI-1.3 Force-field preparation of the R79out conformation**

In order to model the R79out conformation, we superimposed the two 1R2Y and 1R2Z crystal structures. Orientation of the R79 side chain in the 8OG-MutM structure (1R2Y) was then modelled to have the best possible match with that observed in the DHU-MutM structure (1R2Z). Before neutralising and solvating the so obtained structure, a 2000 step force field structure minimisation was performed where only the R79 residue was relaxed. Having removed the electrostatic and steric clashes we proceeded as in the R79in case (stepwise minimizations, heating, equilibration, MD-run)

### **SI-1.4 Preparation of R79outW structure**

For the R79outW case, we took the equilibrated R79out structure from the MD-simulation, placed a single water molecule near the P1 residue, more or less in the initial position of the R79 side chain in the R79in structure. After a 1000 step minimisation, which only involved the relaxation of the newly added water molecule. We then performed a MD run at the force-field level for 3 ns. We found that a single water can easily leave the binding site already at the equilibration stage (data not shown). In other words, the space previously occupied by the R79 side chain, is too large for a single water molecule. As mentioned already in Sec. SI-1.1, the binding site can accept more than one water molecule. We

therefore decided to repeat the same procedure of adding water, but this time added two water molecules. The two water molecules were first allowed to relax, followed by an MD-run for 3 ns for the whole system. In the structure obtained at the end of the MD-run both water molecules remain in the binding site.

In addition, we included three water molecules in the binding site by following the same strategy as before. We observed that the binding site can also accept three water molecules. However, we found that the third water molecule is so far away from the P1 and 8OG moieties (more than 6 Å), that a major catalytic contribution is not anticipated. We therefore did not continue our investigations at the QM/MM level for this particular case.

## **SI-1.5 Structural analysis of the force-field trajectories**

In this section, we present the analysis of our force-field molecular dynamics simulations. Overall, we can show that the neighbourhood of the 8OG lesion in the binding site is stable, irrespective of the orientation of the R79. The binding site also remains stable even if we switch the protonation states of the E2 and E5 residues. A similar stability of the binding site is also observed if two or three water molecules replace the side-chain of R79 in the binding site.

For the purpose of analysis, and in analogy to the study by Song and co-workers[10], three sets of atoms, denoted as "Protein", "DNA", and "OCL", are defined (see below) and their root mean square deviation (RMSD) in their positions (in Å) are plotted. In each case, differing in the R79 position as well as the E2/E5 protonation states, the reference structure for the RMSD calculations is the corresponding equilibrated structure. Since we are investigating the repair mechanism here, a

fourth group of atoms was also defined, which we denote as N-term. The four sets of atoms are defined as follows:

- "Protein": includes the  $\alpha$ -carbon ("CA") atoms of the protein backbone.
- "DNA": includes the backbone heavy atom of 8OG, opposite cytosine and the two flanking base pairs.
- "OCL": includes the backbone heavy atoms of the oxo-capping-loop (OCL) consisting of the residues 222-231.
- "N-term": includes the backbone heavy atoms of the residues P1, E2, and E5 which are important for the catalytic activity of MutM.

In Fig. SI-4 the RMSD values of the Protein, DNA, OCL, and N-term (see above for definition) groups of atoms are plotted (E2 protonated. The E5-protonated plots can be found in Fig. SI-6). In both R79in and R79out cases, the RMSD values for both DNA and protein oscillate around or even below 1 Å. For the OCL and N-terminal groups, the RMSD values do not go beyond the 1 Å border. Overall the structural fluctuations are within a range which can be considered as stable. This analysis justifies the usage of the corresponding equilibrated structures as representative starting points for QM/MM calculations.

For the R79out cases with either 2 or 3 additional water molecules replacing the R79 side chain in the binding site, we have also performed an analysis for the obtained trajectories<sup>1</sup>. As shown in Fig. SI-5, the DNA and protein groups are stable (RMSD values fluctuate around 1 Å). The OCL and N-term sets of atoms show

---

<sup>1</sup>E2 is treated here in its protonated state. QM/MM calculations presented and discussed in Sec. SI-2.3, have shown that this protonation state is energetically more stable than the E5-protonated case. We therefore did not carry out any force-field MD-simulations for the E5-protonated case.

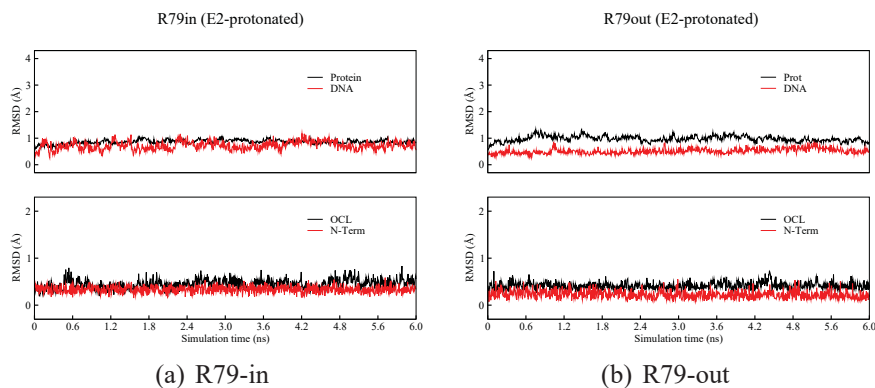


Figure SI-4: RMSD plots for the (a) R79-in and (b) R79-out structures. In both cases, the E2 residue is protonated and E5 is treated as carboxylate anion.

fluctuations below 1 Å, meaning that the overall stability of the binding site is not disturbed by replacing the R79 side chain by 2 or 3 water molecules. This analysis again justifies the choice of using the equilibrated structures for the QM/MM investigations.

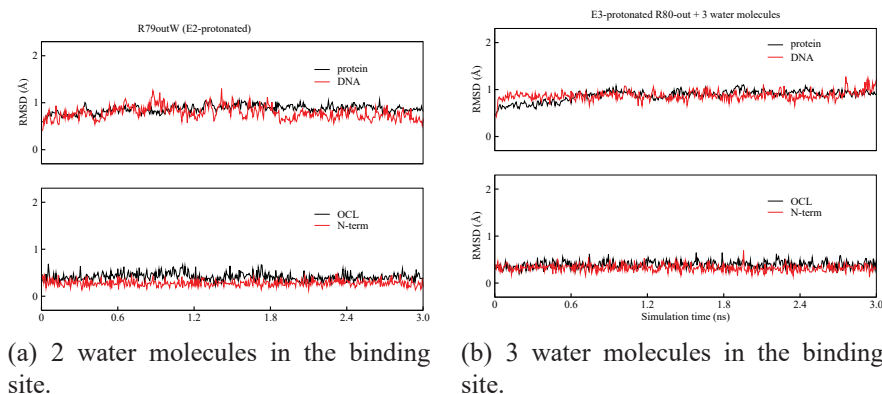


Figure SI-5: RMSD plots for the force-field MD simulations, where (a) two, or (b) three water molecules replace the side-chain of R79 in the binding site.

We also checked the trajectories with E5 protonated. The analysis of the corresponding trajectories (for both R79in and R79out), depicted in Fig. SI-6, shows

slightly higher RMSD values (but all below 2 Å ) for the protein and DNA groups than those obtained in the corresponding E2-protonated cases. The binding site, however, remains as stable as in the previously discussed cases.

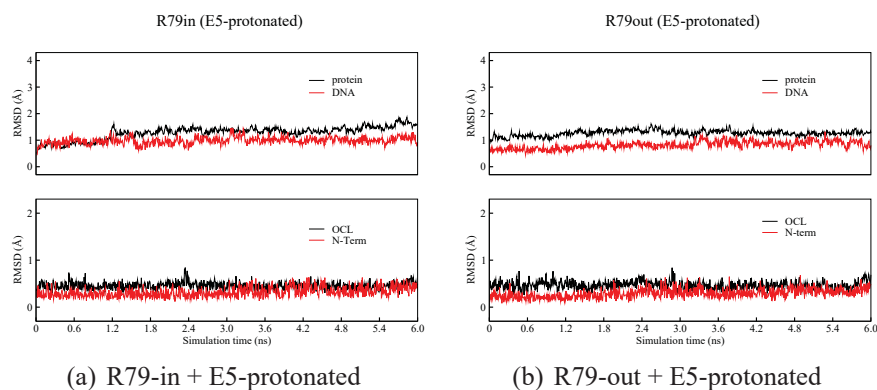


Figure SI-6: RMSD plots for the (a) R79-in and (b) R79-out structures. In both cases, the E5-residue is protonated whereas E2 is kept negatively charged.

## SI-2 QM/MM setup

For each of the R79in, R79out and R79outW setups (already described in Sect. SI-1.1, we have performed QM/MM calculations along the base-protonated and ribose-protonated pathways. The equilibrated structure for each of the setups was minimized at the force-field level and then taken as the starting point for the QM/MM calculations. In order to reduce the computation time of the QM/MM calculations, we decided to remove the outermost solvating water molecules. The fragments which were kept include: (i) the full protein, (ii), the full DNA, (iii) all counter-ions, (iv) all water molecules where at least one atom is within 5.0 Å of either protein, DNA or counter-ions, and (vi) all water molecules where at least one atom is within 12 Å of the 8OG-lesion. These reductions lead to structures with almost 10000 atoms.

In the present work, we employ the QM/MM-interface ChemShell [11] in connection with the quantum chemistry program packages Q-CHEM [12], TURBO-MOLE [13], and the newly developed QM-package FermiONs++ [14]. The DL-POLY [15] force-field engine integrated in ChemShell was used for the MM part of the calculations. The same set of AMBER force-field parameters used for the classical MD simulations was also used for the MM part of the QM/MM calculations. We have used the additive QM/MM scheme with electrostatic embedding. The charge-shift scheme [16] and link atoms (hydrogen) were employed to deal with the QM-MM boundary region. During the QM/MM structure optimizations, carried out using the DL-FIND [17] optimiser, also embedded in ChemShell, we relaxed all fragments (amino acids, nucleic acids, water counter ions) which lie within 10 Å of the N9 atom of the 8OG-residue. Everything beyond this selection



is kept frozen during QM/MM structure optimizations.

### SI-2.1 Choice of the QM-region

As for the QM-region, we have chosen three sets of QM-region, with increasing size which are then used for QM/MM structure optimization. These three regions, denoted here as A, B, and C, are shown in Fig. SI-7. A list of residues included in each QM-region together with the number of atoms can be found in Fig. SI-1.

Table SI-1: List of residues includes in each QM-region. The capping hydrogens are also included in the number of QM-atoms.

QM-region	number of atoms	Residues
A	78	8OG, P1, E2, E5 + 2 water molecules in the R79outW case
B	212	QM-A + Met76, E77, R79, I174, Y175, S219, T220, V221, R222, T223, Y224 and a water molecule (solvation)
C	509	QM-B + L3, P4, V5, K59, L74, R75, G78, G170, F171, G172, N173, G216, G217, Y240, R262, DA291, DG293 + 7 water molecules

Finally, we have used a distance criterion to check the convergence of the QM/MM relative energies with respect to the size of the QM-region. For this purpose, we chose the ribose-protonated pathway in the R79outW setup. We chose the N9 atom of the 8OG residue as the centre of selection. All residues within radius  $R$  ( $R=4, 5, 6, 7$  and  $8 \text{ \AA}$ ) were selected and defined as QM-regions. For all structures defined as minima and maxima along the path obtained from the QM/MM optimization with the QM-C region, single point energy calculations were performed at the BLYP-D3/AMBER level of theory using the 6-31G\*\* basis set.

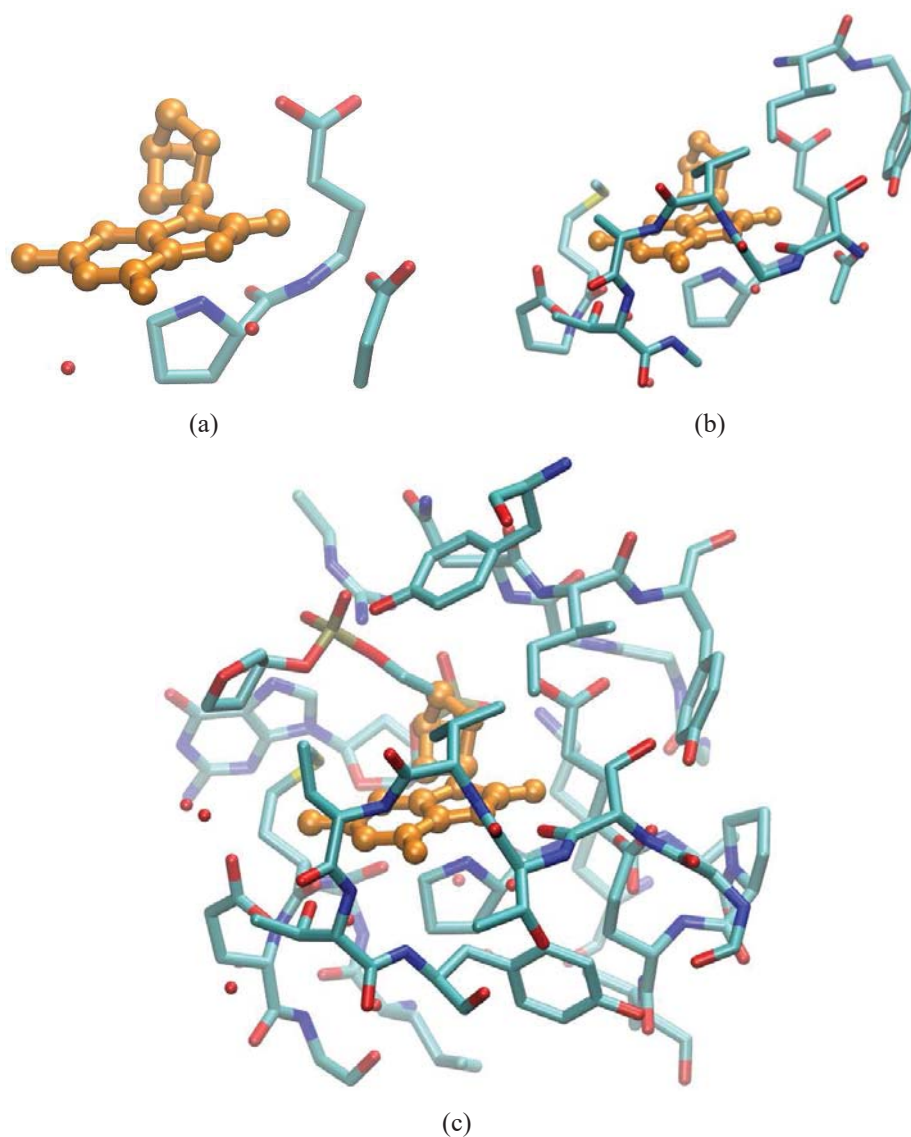


Figure SI-7: The three QM-regions (a) QM-A (78 atoms), (b) QM-B (212 atoms), and (c) QM-C (509 atoms), all shown in licorice representation. The 8OG residue is shown in balls/sticks and orange. Hydrogen atoms are not shown for more clarity. Water molecules are shown as red spheres.

## SI-2.2 QM/MM structure optimization

The SVP basis set [18] was used throughout all QM/MM structure optimizations. We chose density functional theory (DFT) to describe the QM-regions, A, B, and

C. The functionals BP-86 [19–23] and BLYP [19, 21, 24] were chosen in combination with the semi-empirical dispersion correction, as proposed by Grimme (DFT-D3) [25, 26]. For the QM part of the QM/MM structure optimizations with the smaller QM-regions A and B, we used the TURBOMOLE program package. For optimizations with the largest QM-region, i.e., QM-C, however, we switched to the development version of the FermiONs++ program package [14].

For the QM-A region, QM/MM structure optimizations were carried out at the BP86-D3/Amber, BLYP-D3/Amber and B3LYP-D3/Amber levels of theory. For the QM-B region, we carried out QM/MM structure optimizations at the BP86-D3/Amber and BLYP-D3/Amber levels of theory. For the QM-C region, only the BLYP-D3/Amber level was chosen. As later shown in Sect. SI-3 and SI-4, for the QM-A and QM-B regions, the results obtained for with different functional are qualitatively similar. For the largest QM-C region, therefore, we decided to use the BLYP-D3 DFT-functional.

For a more efficient convergence of the QM/MM structure optimizations, the optimized structures obtained along the reaction paths using the QM-A region were used for further optimization (keeping the restraint) with the QM-B region. A similar approach was employed for the largest QM-C region. Here, however, we decided to relax the atoms of the QM-A region only. Furthermore, for the QM-C region, not all points along the path were minimized. Instead we only minimized the structures corresponding to the minima and maxima. The results presented further below show that the minima and maxima obtained with the QM-B region are also reproduced with the QM-C region.

### SI-2.3 QM/MM isomerization energies

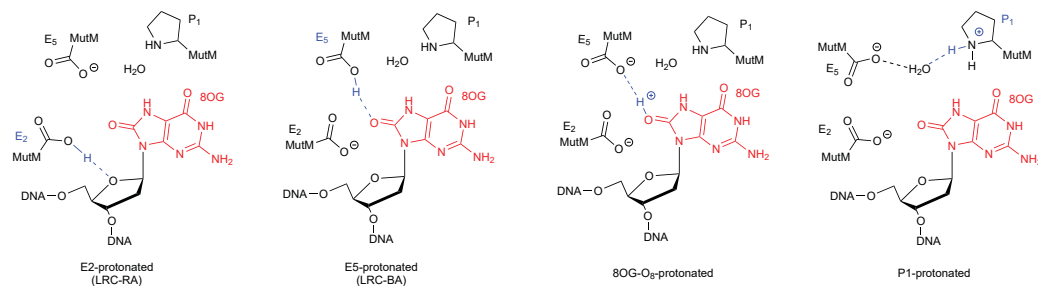


Figure SI-8: Different isomers within the binding site of the MutM. These correspond to different protonation states, where the proton shown in blue can reside on E2 (LRC-RA), E5 (LRC-BA), P1, and O8 of 8OG. The E2-protonated isomer is chosen as the reference structure for the calculations of the relative QM/MM energies in Tab. SI-2.

Before proceeding to the QM/MM results for the base- and ribose-protonated repair pathways, we show some QM/MM results for different isomerization states. Here we consider four possible protonation states, where the P1, E2, E5, or the 8OG-base at the O8-position, is protonated. These four protonated states are shown schematically shown in Fig. SI-8. The corresponding QM/MM isomerization energies are listed in Tab. SI-2. Here, the E2-protonated state (LRC-RCA) is taken as reference.

Table SI-2: QM/MM energies of isomerization (QM-B, BLYP-D3/Amber, in kcal/mol) relative to the E2-protonated isomer (see Fig. SI-8). The LRC-BA and LRC-RA notations are used to represent the starting "Lesion Recognition Complex" structures from which the base- (BA) and ribose-protonated (RA) pathways are initiated.

force-field setup	Protonation site			
	E2 (LRC-RA)	E5 (LRC-BA)	P1	8OG-O8
R79in	0	29.2	18.5	37.4
R79out	0	28.4	26.2	27.7
R79outW	0	18.3	8.5	18.1

In all cases, the LRC-RA state is the lowest energy state, irrespective of the R79 orientation and presence of water in the active site. All other isomers are higher energy, although the energetic difference seems to depend on the R79 orientation. **Interestingly, the O8-protonated state is energetically far more unstable than the LRC-RA state. This observation already provides the first evidence against a base-protonated mechanism.**

For the investigation of the ribose-protonated pathway, we have considered the LRC-RA state as the starting structure. Here, E2 is already protonated and can easily open the ribose ring as the first step of the reaction. For the base-protonated pathway, however, we start from the LRC-BA state, where a hydrogen bond between E5 and O8 is already available. In the context of a base-protonated mechanism, the LRC-BA state has the advantage of having E2 in the carboxylate form which can subsequently stabilize the oxacarbenium ion formed after the glycosidic cleavage. The overall energetics of intermediates encountered in both pathways is independent of the starting structure.

## SI-2.4 Estimation of pKa values

We have not estimated the pKa values for the E2 and E5 residues in this work. However, we would like to refer to the work of Simmerling *et. al.* [10] who have already performed such calculations.

(a) For E2: Simmerling and coworkers have obtained pKa values of 5.7 and 6.8 (depending on the method used for pKa prediction). Since these pH values are close to the neutral pH, Simmerling *et. al.* have also performed simulations with protonated E2. Based on their simulations, both protonation states for E2 are

probable.

(b) For E5: Simmerling and coworkers do not report any data on the pKa value of E5. In their simulation they have chosen the standard protonation state (i.e. negatively charged carboxylate side chain) for this residue.

One should point out that the conventional methods used to estimate the pKa values in the above mentioned study are all based on empirical parameters and static structures. The results obtained should therefore be considered with some caution. In contrast to the previous studies, here, we concentrate on the repair mechanism, where we have compared the base- and ribose-protonated mechanisms. Since a proton is required for both protonation processes, we must have a proton donating residue. For this purpose we have considered the possibility of either having the E2 or the E5 residue acting as proton donor. Our QM/MM isomerization energies (see Sect. SI-2.3) favor an E2 protonated state over an E5-protonated one. Our finding is also in agreement with the study by Karplus *et al* [27] who have shown that a protonated E2 residue stabilizes the 8OG-lesion compared with the deprotonated (negatively charged carboxylate) case.

### SI-3 Nucleobase-protonated pathway

In this section, we present our QM/MM results for the base-protonated mechanism. We first present the results of the R79outW setup, followed by the results obtained for the R79in and R79out setups (see Sect. SI-1.1 for definitions).

The reaction coordinates, RC-BA and RC-SBI-O8p, as depicted in Fig. SI-9, were chosen to follow the reaction path. QM/MM minimum energy paths of the nucleobase-protonated mechanism, obtained along the RC-BA1 and RC-SBI-O8p reaction coordinates are shown in Fig. SI-10. Here, we present the data obtained using different QM-regions and DFT-functionals. Based on these data we can rule out the nucleobase-protonated pathway. Irrespective of the chosen QM-region and DFT-functional, the first intermediate BA1 is thermodynamically not accessible. The barriers to reach the latter is around 70 kcal/mol, rendering this reaction pathway as impossible.

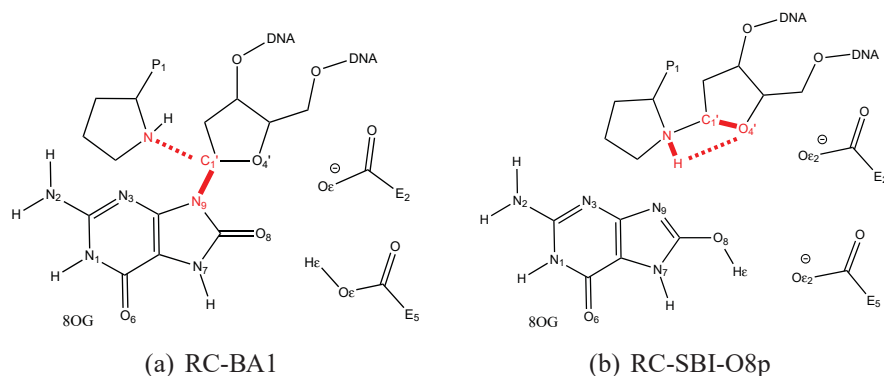


Figure SI-9: The reaction coordinates used to investigate the base-protonated pathway. (a) RC-BA: A combination of the nucleophilic attack of P1 and glycosidic cleavage (b) RC-SBI-O8p: Proton transfer from P1 to O4' accompanied by the opening of the ribose ring. Bonds being broken during the reaction are shown in bold red, those being formed with red dotted lines.

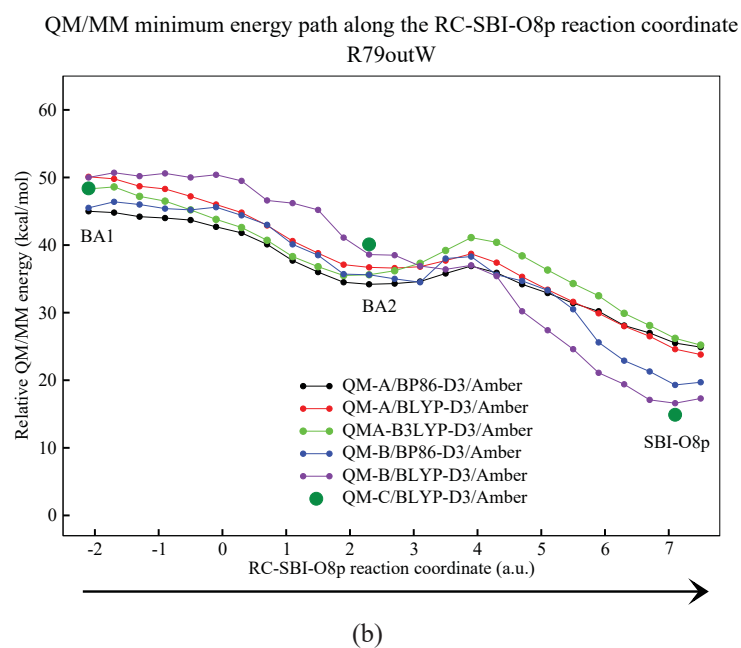
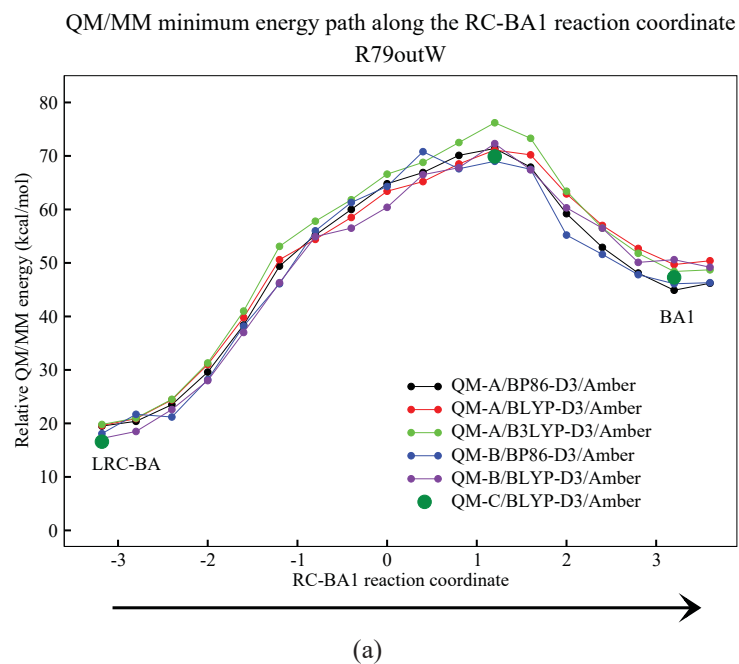


Figure SI-10: QM/MM minimum energy path along the nucleobase-protonated pathway (R79out-E2pW2 setup). All QM/MM energies are in kcal/mol and relative to the corresponding LRC-RA structure.



As evident from the QM/MM data on this pathway (see Fig. SI-10), for the QM-region A, the variation in the energetics obtained from the different DFT functionals is at most 5 kcal/mol. Upon increasing the QM-region, the energy of the highest point on the rate-determining step remains at about 70 kcal/mol. The overall unfavorable trend of the nucleobase-protonated pathway does not change by increasing the QM-size and/or changing the functional.

In addition to the calculations mentioned above, we have also estimated the energy difference between the LRC and the highest point on the path towards BA1 intermediate at other levels of theory. For this purpose we have performed single point calculations for the two structures at the RIMP2/MM, SOS-RIMP2/MM, SCS-RIMP2/MM, and B2PLYP/MM levels of theory. Since calculations with smaller QM-regions give the same qualitative picture as the one from larger QM-regions, we have decided to perform the additional QM/MM calculation with the QM-A region. The SVP basis set was chosen for this purpose. We have performed these calculations using a development version of the Fermions++ program package [14]. The results, presented in Tab. SI-3 confirm those obtained by the DFT approaches.

Table SI-3: QM/MM estimate for the energy difference between the LRC structure and the highest point on the rate determining (LRC-BA  $\rightarrow$  BA1) minimum energy path (QM-region A).

Method	QM/MM energy difference (kcal/mol)
BLYP-D3	+70
RIMP2/MM	+61.5
SOS-RIMP2/MM	+61.9
SCS-RIMP2/MM	+61.7
B2PLYP/MM	+56.2

As described in the next section, the ribose-protonated pathway leads to the

formation of the SBI state with a N9-protonated tautomer of 8OG ( $\rightarrow$ SBI-N9p). In order to estimate the intrinsic energy difference between the O8- and N9-protonated tautomer (both shown schematically in Fig. SI-11), we have performed gas phase QM calculations at the BP86-D3/SVP level of theory. Our calculations show that the N9-protonated tautomer is intrinsically 14.5 kcal/mol lower in energy than the O8-protonated case (gas-phase zero point energy effects are below 0.5 kcal/mol). Further tautomerization of the 8OG-base may therefore lead to a thermodynamically more favourable state ( $\rightarrow$ N9 protonation). However our calculations have shown that the whole nucleobase-protonated mechanism can be kinetically ruled out.

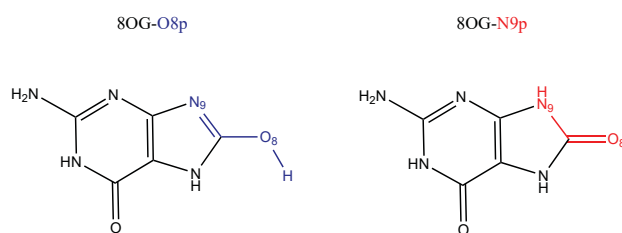


Figure SI-11: Both O8- and N9-protonated tautomers of 8OG.

As for the R79in and R79out setups, we have followed the reaction path along the RC-BA1 coordinate. The results are shown in fig. SI-12. We find the BA1 state to be energetically stable and therefore did not continue with the base-protonated pathway for this setup.

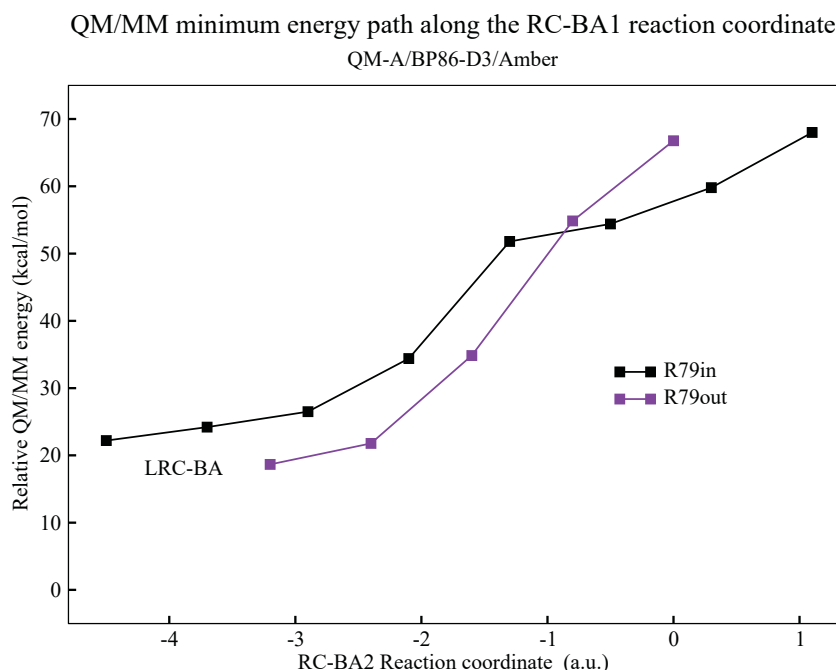


Figure SI-12: QM/MM minimum energy path along the RC-BA1 reaction coordinate obtained at the QM-A/BP86-D3/Amber for both R79in-E5p and R79out-E5p setups. All QM/MM energies (in kcal/mol) are relative to the corresponding LRC-RA structure. No covalent bond between P1 and ribose-C1' could be identified at the end of these paths. Therefore, we did not use the BA1 label in this figure

### SI-3.1 Comparison with the mechanism of uracil glycosylase

Our high energy structure (+70 kcal/mol) along the base-protonated mechanism contains an almost planar ribose ring and resembles the oxacarbenium ion which

has been discussed by Dinner *et. al.* [28]. Interestingly, Dinner and co-workers suggest the formation of such oxacarbenium ion as the first intermediate (minimum on the potential energy surface) upon base cleavage. In our case, for the 8OG excision repair by MutM, however, we locate such a cation near the transition state of the reaction path. A comprehensive comparison of repair mechanisms of the two repairing enzymes is beyond the scope of this work and subject of further study.

## SI-4 Ribose-protonated pathway

In the following, we discuss the QM/MM results of the ribose-protonated pathway in more detail. In analogy to the nucleobase-protonated pathway, we begin with the QM/MM results of the R79outW setup and then present/discuss the results obtained for the R79in and R79out setups (see Sect. SI-1.1 for the definition).

The reaction coordinates used to connect the initial LRC-RA and the final SBI-N9p states of the ribose-protonated pathway are shown in Fig. SI-13.

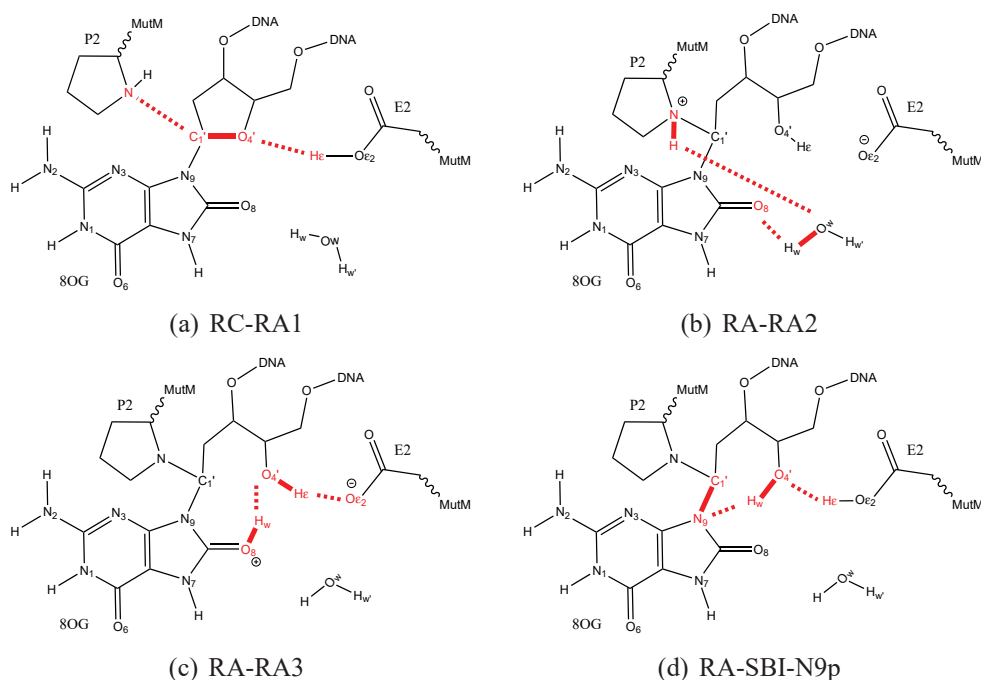


Figure SI-13: Reaction coordinates for the ribose-protonated pathway with the involvement of a catalytic water molecule in the active site. Bonds being broken during the reaction are shown in bold red, those being formed with red dotted lines.

The QM/MM minimum energies of the ribose-protonated pathway obtained with the QM-A region (different functionals) are shown in Fig. SI-14. The QM/MM energies obtained for the QM-B and QM-C regions are shown in Fig. SI-15.

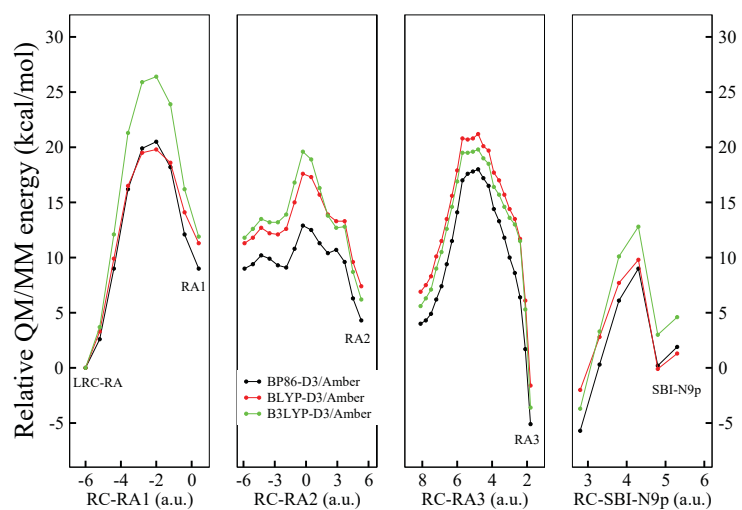


Figure SI-14: The ribose-protonated pathway for the R79out-E2pW setup. Energy differences (in kcal/mol) with respect to the LRC-RA structure are shown for the QM/MM structure optimizations with the QM-region A.

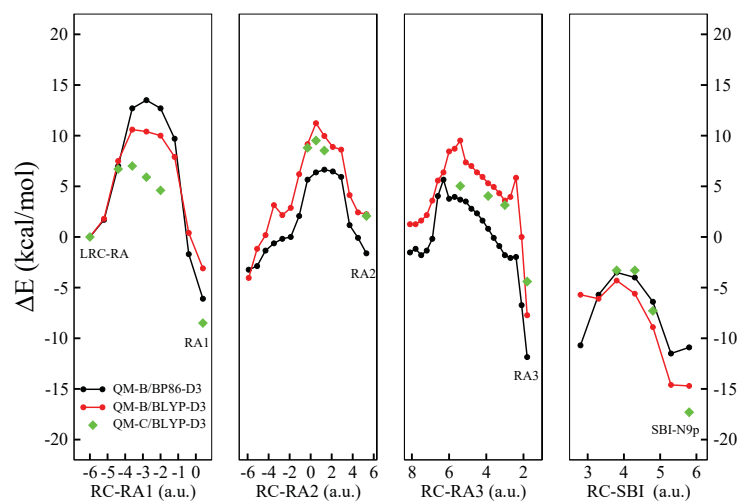
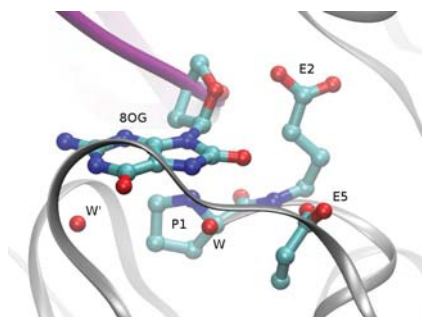


Figure SI-15: QM/MM minimum energy paths computed for the ribose-protonated mechanism with QM-regions B and C. All energies, in kcal/mol, are plotted relative to the QM/MM energy of the LRC-RA state.

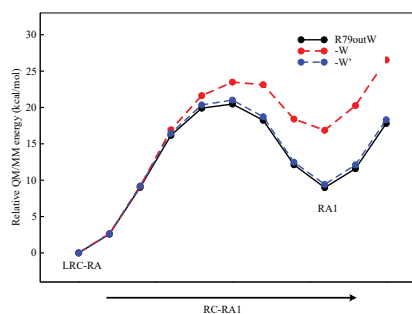
Based on the QM/MM data presented in Fig. SI-14 and Fig. SI-15 for the ribose-protonated pathway one can compare the energetics obtained with the different DFT functionals and/or QM-regions. In general, the variations are larger than those observed for the nucleobase-protonated pathway (see Fig. SI-10). Here, for the QM-region A, the B3LYP-D3/MM barrier ( $\approx 26$  kcal/mol) for the LRC-RA  $\rightarrow$  RA1 step is about 8 kcal/mol higher than observed for the BP86-D3/MM and BLYP-D3/MM cases. The 26 kcal/mol barrier estimate is still much lower than the 70 kcal/mol barrier involved in the nucleobase-protonated pathway. Upon increasing the QM-region to B and C, the barrier for the LRC-RA  $\rightarrow$  RA1 step is reduced even further (roughly 11 and 7 kcal/mol, respectively). As for all other steps of the ribose-protonated pathway, the relative energies obtained with QM-C are close (within 2 kcal/mol) to those of the QM-B region.

The presence of the catalytic water molecules near P1, replacing R79, seems to make a huge difference (compare with the QM/MM results obtained with the R79in and R79out setups further below, where such a catalytic water molecule in the binding site is absent). This water molecule can already influence the first step of the ribose-protonated pathway, namely the LRC-RA  $\rightarrow$  RA1 step. During our force-field MD-simulation, we have found that two water molecules (W and W', shown here in Fig. SI-16(a)) can reside in the binding site. We have also performed QM/MM calculations to estimate and compare the effect of these two water molecules. For this purpose we took the structures obtained along the RC-RA1 reaction coordinate (QM-A, BP86-D3/Amber), manually removed one of the water molecules, and then performed single point QM/MM calculations at the same level of theory. The results obtained are shown in Fig. SI-16(b). As evident from these results, removing the water molecule, W, leads to RA1-destabilization

of about 10 kcal/mol. Removing the other water molecule, W', on the other hand does not have a significant influence on the energetics.



(a) The two water molecules replacing R79 in the active site.



(b) Single point QM/MM energies along the RC-RA1 coordinate, where either W or W' is removed.

Figure SI-16: The influence of the water molecules, W and W', shown in (a) can be estimated from the QM/MM energies shown in (b).



The results of the R79in setup are shown in Figs. SI-17 and SI-18. The results of the R79out setup are shown in Figs. SI-19.

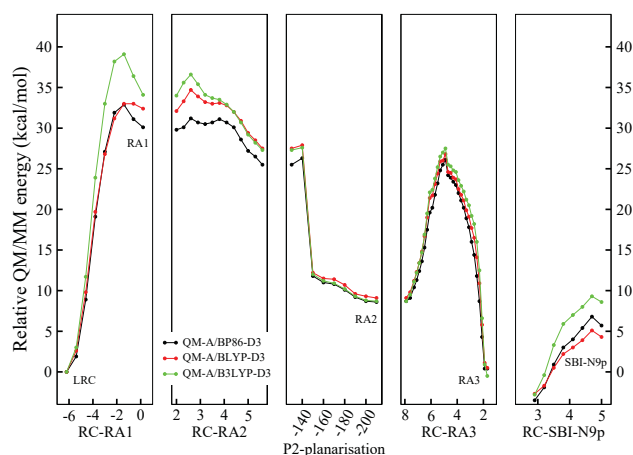


Figure SI-17: The ribose-protonated pathway for the R79in conformation. The RC-RA1-4 reaction coordinates are all in a.u. A dihedral angle defined by the  $C_{\alpha}(P2)$ ,  $N(P2)$ ,  $C_{\delta}(P2)$ , and  $C1'(8OG)$  atoms was used to account for a sudden planarization of the P1-ring upon deprotonation leading to the RA2 intermediate.

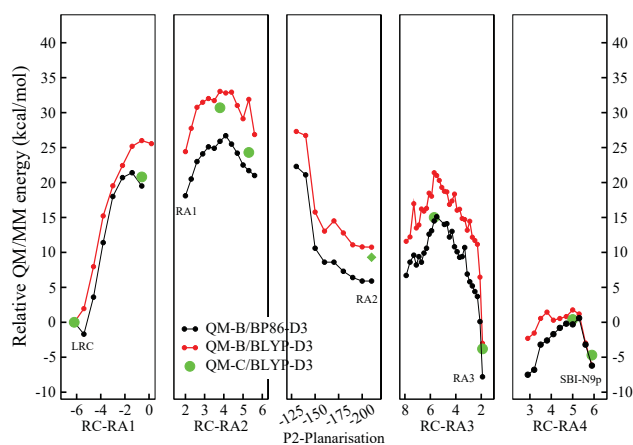


Figure SI-18: QM/MM energies for the ribose-protonated pathway in R79in conformation. Energy differences (in kcal/mol) with respect to LRC-RA. QM-regions B and C are chosen.

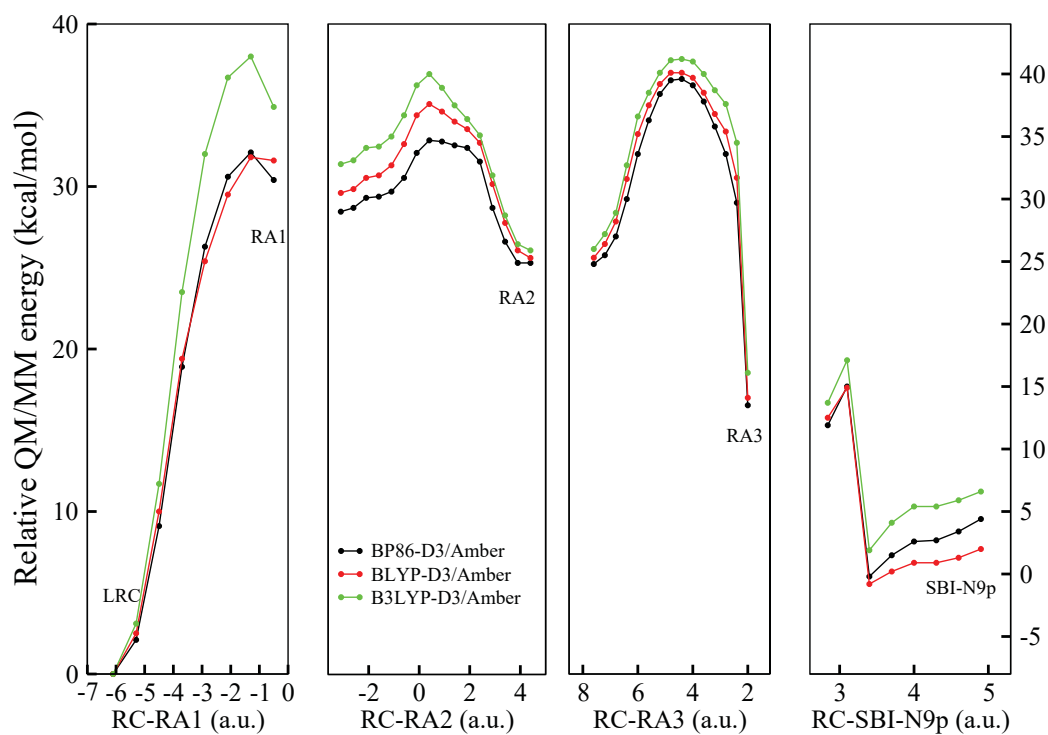


Figure SI-19: The ribose-protonated pathway for the R79out conformation. Energy differences (in kcal/mol) with respect to the LRC-RA state. The values of the reaction coordinates RC-RA1-4 are all in a.u.

### SI-4.1 QM/MM convergence with respect to QM-size

For the R79outW setup, we have also investigated the convergence of the QM/MM relative energies for the ribose-protonated pathway. For this purpose we selected all residues (amino acids, nucleic acids, water, ...) where at least one atom is within the radius  $R$  of the N9 atom of the 8OG base ( $R = 4, 5, \dots, 9$  Å). The residues included in the different QM-spheres are defined by the selection radius as given in Tab. SI-4. The corresponding QM/MM relative energies are shown in Tab. SI-5 which were obtained by single point calculations at the BLYP-D3/AMBER level of theory (6-31G\*\* basis set). The barrier estimate for the second step of the reaction, represented here by the energy of the RA2-max structure, changed from 27 kcal/mol (5 Å radius) to 14 kcal/mol (8-9 Å). The energy of the final product state, SBI-N9p, reaches its converged value, if we chose a radius of 8 Å (493 atoms).

Table SI-4: List of selected residues for different radii.

# atoms	Rad. (Å)	Charge	List of residues
93	4	-1	P1 E2 E5 I174 G93 WAT353 WAT390 WAT438
143	5	-1	+ M76 T94
278	6	0	+ R75 G172 N173 V221 R263 8OG292 WAT342
374	7	+1	+ K59 Y175 S219 Y224 Y241
493	8	0	+ L3 P4 V6 E77 T220 R222 T223 WAT329
606	9	0	+ L74 G78 G170 F171 V176 G218 L238 WAT323 WAT346 WAT361 WAT648 WAT307 WAT350

Table SI-5: QM/MM relative energies (kcal/mol) obtained along the ribose-protonated pathway. QM/MM single point energies were performed, at the BLYP-D3/AMBER level in dependence of the size of the QM-sphere. All residues within radius R are included into the QM-region of calculation.

Radius (Å)	4	5	6	7	8	9
# atoms	93	143	278	374	493	606
RA1-max	10.6	8.4	8.7	9.6	6.8	6.8
RA1	4.6	3.6	-2.1	-5.5	-6.9	-6.9
RA2-max	26.5	27.5	5.8	15.3	13.9	13.9
RA2	11.4	13.0	3.6	9.2	8.3	8.3
RA3-max	20.4	21.8	16.2	14.0	12.5	12.5
RA3	1.4	0.8	-1.4	-4.8	-3.3	-3.3
SBI-N9p-max	-2.6	-6.0	-9.2	-1.2	-4.5	-4.5
SBI-N9p	-8.0	-12.2	-17.4	-16.7	-16.0	-16.0

## SI-4.2 Zero-point energy corrections

In order to estimate the influence of zero point energy (ZPE) corrections, we have performed partial force-constant calculations (see for example Ref. [29]) at selected structures along both base-protonated and ribose-protonated reaction pathways. For this purpose we chose the minima and maxima obtained at the BLYP-D3/Amber level of theory with the QM-region, B. For the partial force constant calculations, we chose the QM-A region as the active-region for which the numerical Hessian components (two-point formula) were calculated. The remaining atoms of the QM-B region which do not belong to QM-A are assumed to have an infinite mass and therefore do not contribute to the ZPE-correction. This partitioning is justified, since we are interested in the contribution of those vibrational modes which are directly involved in the reaction.

The calculated ZPE corrections are listed in Table. SI-6, where the total ZPE of the LRC state is taken as reference. Negative values in this table correspond to a stabilization with respect to the LRC state, positive values mean destabilization.

As evident from the data presented in Tab. SI-6, the ZPE corrections do not influence the overall preference of the ribose-protonated pathway over the base-protonated pathway. For the ribose-protonated pathway, the ZPE corrections are below 3 kcal/mol. For the base-protonated pathway, the BA1 intermediate is stabilized by about 11 kcal/mol. Interestingly for the transition state (BA1-max), which resembles the oxacarbenium ion, however, the ZPE correction is about 1 kcal/mol. Overall, the ZPE corrections do not change the general picture and the base-protonated mechanism remains highly unlikely.

Table SI-6: Relative QM/MM ZPE corrections,  $\Delta$ -ZPE (in kcal/mol), for the reaction intermediates as well as the maximum energy structure connecting them. The total QM/MM-ZPE of the LRC structure is taken here as reference. For the maximum energy structure leading to the next intermediate (indicated as TS) the value of the single imaginary frequency found is also given.

State	$\Delta$ -ZPE	imaginary frequency ( $\text{cm}^{-1}$ )
Ribose-protonated pathway		
LRC	0.0	—
RA1 (TS)	-2.0	<i>i</i> 63.1
RA1	1.6	—
RA2 (TS)	-2.4	<i>i</i> 897.8
RA2	-1.0	—
RA3 (TS)	-0.2	<i>i</i> 43.3
RA3	-0.3	—
SBI	-1.1	—
Base-protonated pathway		
BA1 (TS)	-1.2	<i>i</i> 97.3
BA1	-10.7	—

## SI-5 Binding pocket discrimination

### SI-5.1 Total interaction energies

In the present work, we have also looked at the interaction energies between the 8OG base and the MutM enzyme. We then compare these interaction energies, obtained from the QM/MM calculations, with those of systems where G, 8-AminoG, and 8-Br-G, each replace 8OG in the binding site. For all the three alternative substrates we assume the very same binding mode (*syn*) as observed for 8OG. We took the LRC-RA structure (R79outW setup, QM/MM optimized with the QM-B region at the BLYP-D3/AMBER level), modified the 8OG fragment to the desired substrate and then re-optimized the structure at the same QM/MM level. For a direct comparison we made sure that the same set of atoms of the environment (QM or MM) are relaxed during the QM/MM structure optimization.

According to the super-molecular approach, the interaction ( $I$ ) between the DNA ( $N$ ) and the protein environment ( $P$ ) can be determined as difference of absolute energies of the total energy of the DNA-protein-complex and the absolute fragment energies, both in case of the damaged DNA (8OG) and also in case of the alternative DNA substrate ( $S$ ):

$$\begin{aligned} I_{N_{8OG}P} &= E_{N_{8OG}P} - E_{N_{8OG}} - E_P \\ - \left( I_{N_S P} &= E_{N_S P} - E_{N_S} - E_P \right) \\ \hline \Delta I_{NP} &= E_{N_{8OG}P} + E_{N_S} - E_{N_{8OG}} - E_{N_S P} \end{aligned} \quad (1)$$

The interaction energy difference  $\Delta I_{NP} = W_{N_{8OG}P} - W_{N_S P}$  states if or to what extent the enzyme binds preferably to 8OG vs. the alternative substrate ( $S = G, 8\text{-Amino-}$

Table SI-7: Composition of the QM region within the total DNA-protein system (QM cutoff distance  $r$ , number of QM atoms, charge of the QM region, residues within the QM region). Residues belonging to DNA, including the substrate S (8OG, G, 8-Amino-G, and 8-Br-G), are highlighted in red.

$r$ [Å]	Number	Charge	QM residues (DNA residues in red)
4.0	175	-2	P1, E2, E5, I174, S219, T220, V221, Y224, <b>S</b> , water
5.2	235	-1	+ Y175, R222, T223, water
6.2	294	-1	+ P4, M76, G218, Q235
7.0	409	-2	+ L3, G172, G232, L238, Y241, <b>DG 293</b> , water
8.0	598	-2	+ V6, R75, E77, N173, A213, G217, V225, F234, R263, <b>DA 291</b> , water
9.0	794	-1	+ ILE 8, K59, R79, G172, F171, V176, E178, I210, A231, T233, water
10.0	953	-1	+ E7, T8, L74, G78, T209, LYS 215, N226, H236, water
11.0	1115	-3	+ H73, L163, D177, V214, E230, H237, G242, H266, water

G, 8-Br-G). According to the definition above, a negative sign of  $\Delta I_{NP}$  means a preferred binding of the 8OG lesion.

To determine the errors for limited QM sizes, the QM size is increased upon convergence, starting from the central nucleotide 8OG or S, respectively (the QM cut-off distance refers to the position of the O8-C8-bond). The selection is based on pdb residues, i.e. as soon as the distance to any atom of a nucleotide or amino acid residue falls below the cut-off distance, the whole residue is treated within the QM region. Tab. SI-7 shows the resulting composition of the QM region in detail.

For the QM-only approach the environment beyond the cut-out at each QM



size is completely neglected, while the QM/MM approach includes the further environment as MM subsystem within an electrostatic QM/MM embedding (point charges at the QM/MM border are shifted according to Refs. [11, 16]). The QM description is based on the Hartree-Fock theory using an SVP basis set [18]. A counterpoise correction [30] was performed for the basis set superposition error. Dispersion correction (D3) according to Grimme et al. [31] was employed [30].

At this point we would also like to comment on Fig. 3b in the main part of the manuscript. The fluctuations in the QM/MM energies from 6.2 to 8 Angstrom result from the fact that we include charged residues (both nucleic and amino acids). The DNA-enzyme interface comprises a multitude of oppositely charged fragments (negative phosphates from the DNA and positively charged amino acids, or salt bridges from the repair enzyme). According to table SI-6, upon increasing the radius from 7 to 8 Å, we include many charged residues, namely R75, E77, R263 and DA291 (O5'-adjacent base to 8OG). This change in the electrostatics is expected to be dampened/compensated by the MM-charges in the QM/MM calculations. However, in QM-only calculations this is not the case as there are no MM-charges. One should also mention that G and 8OG have dipoles in opposite direction (see for example Banerjee et. al. Nature, 2005, 434, 612-8 for further discussion) and therefore quite sensitive towards changes in electrostatic fluctuations which are not shielded in the QM-calculations.

## SI-5.2 Investigation of individual residue contributions

The QM/MM data of the previous section indicates a total interaction energy preference of about 20 kcal/mol for MutM towards 8OG compared with G. The convergence behavior of the total energy difference with QM size (see Fig. 3b in main text) in combination with the data presented in Tab. SI-7 allows to find a set of amino acids which could be influential for the discrimination. However, assigning individual residue contributions is not possible from this data. We have therefore performed a deeper analysis on the role of individual residues by performing additional QM/MM calculations based on the Fig. SI-20.

	$I_{N,P}$	$=$	$E_{tot}$	$-E_N$	$-E_P$	total
$-$	$I_{N\setminus L,P}$	$=$	$E_{tot\setminus L}$	$-E_{N\setminus L}$	$-E_P$	without nucleotide( $L$ )
	$I_{L,P}$	$=$	$E_{tot}$	$-E_N$	$-E_{tot\setminus L}$	$+E_{N\setminus L}$
$-$	$I_{L,P\setminus R_i}$	$=$	$E_{tot\setminus R_i}$	$-E_N$	$-E_{tot\setminus (L,R_i)}$	$+E_{N\setminus L}$
	$I_{L,R_i}$	$=$	$E_{tot}$	$-E_{tot\setminus L}$	$-E_{tot\setminus R_i}$	$+E_{tot\setminus (L,R_i)}$
						$R_i - L - \text{contribution}$

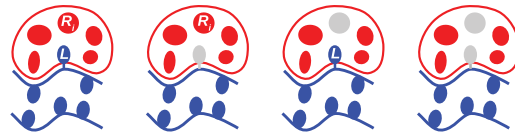


Figure SI-20: Scheme for deriving the interaction contribution  $I_{L,R_i}$  of a DNA nucleotide  $L$  with the protein residue  $R_i$ ;  $N$ : DNA (blue);  $P$ : Protein (red); the notation ' $\setminus$ ' means 'in absence of'.

According to the illustrated approach, by selectively removing a pair of interacting species, we can calculate the individual discrimination contributions of those residues of MutM which are in vicinity (8 Å) of the lesion within the binding pocket of the repair enzyme. Fig. SI-21 shows the details on the groups which have been removed as well as the positions of saturating with hydrogen atoms.

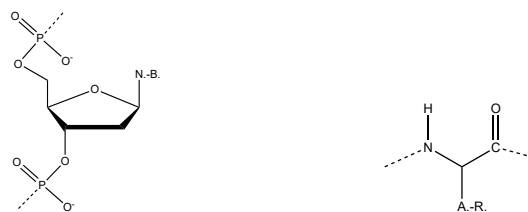


Figure SI-21: Structures of removed groups; Left: Nucleoside diphosphate with damaged or intact nucleobase (N.-B.); Right: total amino acid residue (with amino acid rest A.-R.); the dashed bonds are cut and saturated with hydrogen at the neighboring groups.

In Tab. SI-8 we present the QM/MM interaction contributions of the affected nucleotide diphosphate <sup>2</sup> with the closest 27 protein residues, for both cases of damaged and intact base. A negative difference of the interaction contribution indicates a preferred stabilization of the damage by the corresponding protein residue  $R_i$ . As the nucleoside diphosphate has a charge of -2, the Table shows a strong repulsion for negatively charged residues (E) and a strong attraction for positively charged residues (lysine, arginine) for both the 8-oxoG and G case. However, in the difference contribution, the strong electrostatic pure monopole-monopole contribution cancels out. Based on our analysis, presented in Tab. SI-8, the amino acids P1, E5, Y175, S219 and T220 differentiate most in binding the damaged vs. the intact DNA. The geometrical arrangement of the four residues that strongest prefer the damage is shown in Fig. SI-22. The strong interacting difference can be mainly attributed to different dipole-dipole interactions affected by the carbonyl oxygen O8 and H7 atom of the damage. In particular a strong hydrogen bond to the protein backbone in S219 can be influential in stabilizing

<sup>2</sup>The phosphate group of the nucleic acid adjacent to the lesion is also included in the QM-calculation. We used this approach to avoid any artifacts in the calculation which may arise by introducing a QM-MM boundary by breaking the highly polar O-P bond

the lesion inside the active site (18 kcal/mol, see Tab. SI-8).

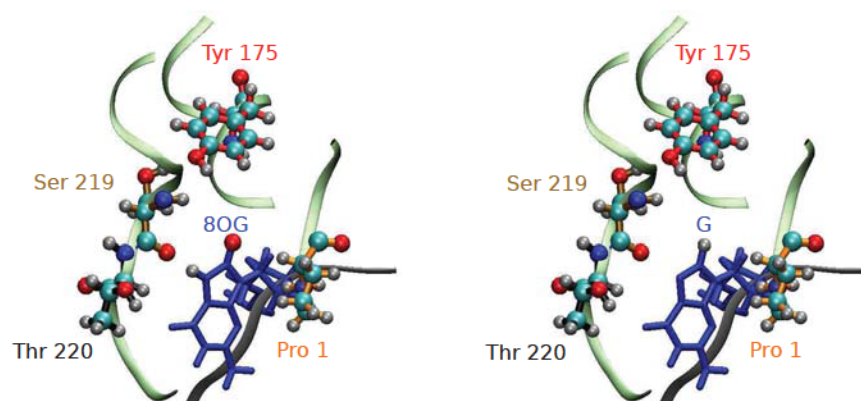


Figure SI-22: Structure cutout of residues P1, Y175, S219 and T220 and the nucleotide of the damage 8-oxoG (left) and the intact base G (right), respectively.

Table SI-8: Interaction contributions  $I_{L,R_i}$  in kcal/mol of the damaged nucleoside diphosphate (8-oxoG) or the intact nucleoside diphosphate (G), respectively, with the nearest 27 protein residues  $R_i$  and the difference of interaction contributions (calculated at HF-D3/SVP/Amber level with a counterpoise correction for the removed molecular groups). The differences arising from individual contributions add up to -26.5 kcal/mol (i.e. in favor of 8OG). Due to the many body effects, this sum does not reproduce the total interaction energy difference of -20 kcal/mol presented in the main text (see Fig. 3b).

Protein residue $R_i$	8-oxoG	G	Difference
P1	-74.1	-68.2	-5.9
E2	92.6	91.1	1.6
L3	-2.8	-2.5	-0.3
P4	-4.8	-3.6	-1.2
E5	67.1	59.4	7.8
V6	-3.6	-3.0	-0.6
K59	-103.5	-101.2	-2.3
R75	-53.0	-52.6	-0.4
M76	-3.0	-3.1	0.1
R79	-46.1	-47.2	1.2
G172	-17.0	-16.4	-0.6
N173	-32.8	-32.6	-0.2
I174	-9.6	-8.0	-1.5
Y175	-11.2	-8.1	-3.1
A213	-0.5	-0.1	-0.4
G218	-3.7	-3.4	-0.3
S219	-12.3	5.9	-18.2
T220	-8.5	-4.8	-3.7
V221	-11.2	-12.1	0.9
R222	-63.3	-63.9	0.7
T223	-10.2	-10.0	-0.2
Y224	-5.2	-6.2	1.0
V225	3.5	3.9	-0.4
F234	-1.6	-2.2	0.6
Q235	1.6	2.1	-0.5
L238	-3.6	-3.5	-0.1
Y241	-24.8	-24.4	-0.4

### SI-5.3 Ribose-protonated pathway for other substrates

The results presented in Sect.SI-5.1 already favor the binding of 8OG and 8-Amino-G compared with the guanine (G) and 8-Br-G substrates. Our proposed ribose-protonated mechanism is *a priori* nucleobase-invariant. We therefore decided to check whether it is possible for MutM to carry out base-excision for G, 8-Amino-G, and the 8-Br-G substrate, in a similar fashion as it does for 8OG. For this purpose we employed the structures already used to calculate the interaction energy differences and performed QM/MM geometry optimizations along a ribose-protonated pathway. In analogy to the 8OG-case, the first step involves opening of the ribose ring followed by the nucleophilic attack of the N-terminal Proline on C1' of the substrate ( $\rightarrow$ RA1, see Fig. SI-23). From this stage point however, we decided to follow a somewhat different reaction coordinate before cleaving the glycosidic bond. Since the substituents, Br and H, at the 8-position of the 8-Br-G and G substrates, can not be formally protonated, we followed a reaction path which involves a proton transfer from P1, via the catalytic water molecule, to the N7 position of these alternative substrates. A similar approach/strategy was employed for the 8-Amino-G nucleobase. Once N7 is protonated ( $\rightarrow$ RA2', see Fig. SI-23), for all three substrates, we followed the reaction coordinate which corresponds to the cleavage of the glycosidic bond. For QM/MM geometry optimization, we have carried out restrained QM/MM geometry optimizations on minimum energy reaction paths at the BLYP-D3/Amber level using the SVP basis set. Here we employed the QM-regions A and B. In order to remain consistent with the previously discussed calculations, the region used for relaxation during structure optimization was chosen to match the region used for

the 8OG lesion. The QM/MM relative energies are shown in Fig. SI-24.

The QM/MM relative energies presented in Fig. SI-24 (BLYP-D3/Amber values, QM-B region) indicate that the barrier for the ring-opening for all three substrates is between 12-15 kcal/mol (for 8OG, we estimate a barrier of about 12 kcal/mol, see Fig. SI-15). For this step we do not observe any clear discrimination with respect to the different substrates. For the second step, i.e., protonation of N7, the highest point along the ribose-protonated excision pathway of 8-Amino-G is about 15 kcal/mol above its corresponding LRC state. For the G and 8-Br-G substrates on the other hand, the highest energy point is about 25 kcal/mol above their corresponding LRC structure (Fig. SI-24). For all substrates, once N7 is protonated, the glycosidic cleavage can occur without any notable barrier (below 2 kcal/mol in all cases).

Overall, one may assign the 10 kcal/mol higher barrier required to protonate the N7 position of the G and 8-Br-G nucleobases as an indication of the overall inability of MutM to excise these lesions. One should point out that our reaction profiles are all based on the assumption that the alternative substrates bind to MutM in the *syn* mode. Of course, we can not rule out the possibility of different binding modes for the investigated substrates. There are no x-ray structures available for the binding mode of the 8OG-analogues. In this work, therefore, we constrain ourselves to compare the ribose-protonated pathway for the mentioned substrates under the very same condition (including binding-mode preference). The trend observed in the ribose-protonated reaction profiles (Fig. SI-24) provide a further explanation for the experimental data which we present in this work. The QM/MM reaction profiles suggest that even if substrates such as G and 8-Br-G manage to overcome the "discrimination barrier" encountered during

the recognition and stabilization, still, there seems to be some indirect discrimination along the repair pathway towards the final glycosidic cleavage step. It is expected that such repair-discrimination, in combination with the much more dominant stabilization-discrimination (discussed in detail in the previous section and in the main text), allow MutM to avoid cleaving an undamaged G, which may have found its way into the catalytic site of the enzyme.

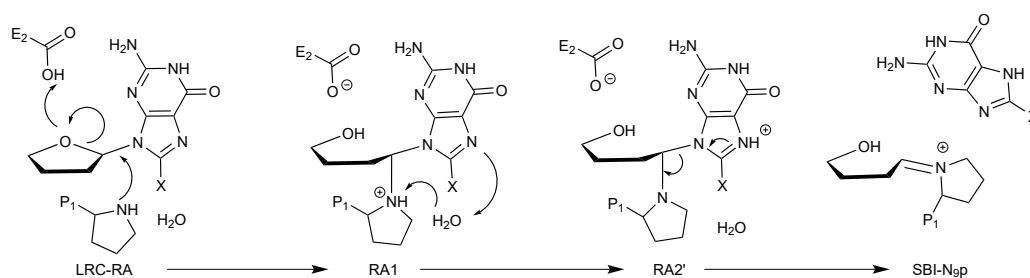


Figure SI-23: Scheme of a ribose-protonated pathway for the different 8OG-analogues: 8-Amino-G, G, and 8-Br-G.



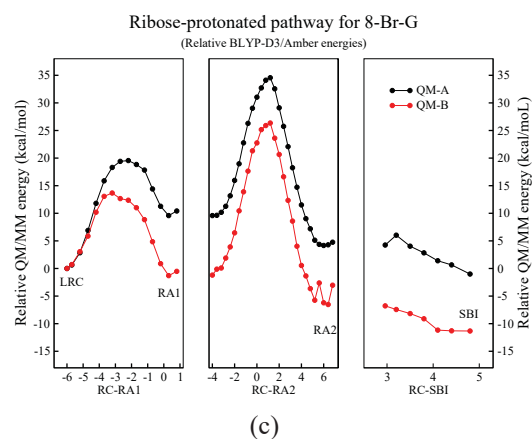
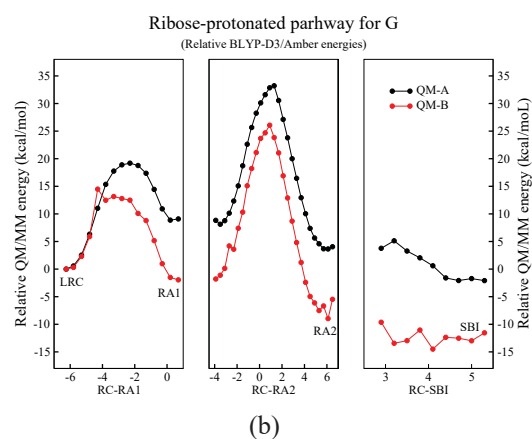
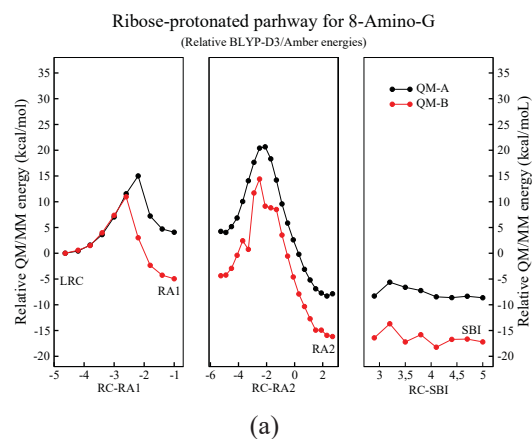


Figure SI-24: The relative QM/MM energy (BLYP-D3/Amber, SVP basis set) profiles along the ribose-protonated pathway for different substrates: (a) 8-Amino-G, (b) G, and (c) 8-Br-G. Here the results of both QM-regions A and B are presented.

## SI-6 MutM activity assay

### SI-6.1 Cloning, protein expression and purification

The *mutm* gene (*Lactococcus lactis* subsp. *cremoris* SK11 gi: YP\_808428.1) was amplified by polymerase chain reaction using the High Fidelity AccuPrimePfx polymerase (Invitrogen) and following primers: 5'-aatgccagagttaccagaagtgaaa-3' (forward) and 5'-tccttttggctgacagaatgggcaa-3' (reverse). The resulting product was cloned into the entry vector pENTRY-IBA10 via the StarGate® method (IBA, Göttingen). After verification of the correct sequence the entry vector was mixed with the acceptor vector pPSG-IBA3 to obtain the desired destination vector pPSG-IBA3-*mutm* (C-terminal Strep-tagII). The plasmid was transformed into BL21(DE3) cells (Invitrogen) and cells were grown in LB media at 37 °C until they reached an OD<sub>600</sub> of 0.8. After induction with IPTG (1 mM) and addition of ZnCl<sub>2</sub> (10 µM) the incubation was continued for 4 h at 30 °C. Cells were harvested by centrifugation and resuspended in Strep buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA) with Complete protease inhibitor mix (Roche, Mannheim). Cells were lysed by French press, and cell debris was removed by centrifugation. The lysate was loaded onto the Strep-Tactin column proceeding according to the manufacturers protocol (IBA). MutM containing fractions determined by SDS-PAGE analysis were pooled and loaded onto a Source S15 column (GE, Munich) equilibrated with Source S buffer (100 mM Tris-HCl pH 7.6, 50 mM NaCl, 1 mM EDTA, 5 mM EDTA, 5% Glycerin). The protein was eluted by linear gradient in Source S buffer containing 800 mM NaCl and the peak fractions containing MutM were pooled. Excess salt was removed by repeated concentration and dilution in Source S buffer. The protein was concentrated to 12 mg

ml<sup>-1</sup>, shock frozen in liquid nitrogen and stored at -80 °C. The correct mass of the protein was confirmed by MALDI-TOF analysis.

## **SI-6.2 Activity assays with MutM**

First a fluorescent labeled 14-mer counterstrand (Metabion, Germany) was annealed to the unlabeled 14-mer template DNA strand, either undamaged or containing the appropriate lesion in the middle of the strand (see Tab.SI-9. To perform the activity assay MutM (3.2  $\mu$ M) and modified double stranded DNA (40  $\mu$ M) were mixed in Source S buffer to a final volume of 10  $\mu$ L. After incubation for 30 min at 30° C, loading buffer (2  $\mu$ L) was added and the reaction was analyzed on a 20% denaturing gel and visualized by a LAS-3000 imaging system.

ODN 4 is an undamaged reference double strand and was also used as size standard of 14 bases. ODN 5 is a single stranded size standard of 6 bases. Activity of MutM would result in an excision of the base leading to a 6 mer. In the denaturing gels only the fluorescence labeled strand is detected. MutM is only active against double stranded DNA therefore only double strands were used. ODN 1 contains the 8-oxo-dG lesion in the middle of the duplex while ODN 2, and 3 contain an 8-Br-dG, or 8-NH<sub>2</sub>-dG at the same position.

Table SI-9: Table 1: Depiction of the used oligonucleotides for the activity assay.

ODN	Sequence			
1	5'-Fluo- 3'-	C T C T T T G A G A A A	8OG C	T T T C T C G -3' A A A G A G C -5'
2	5'-Fluo- 3'-	C T C T T T G A G A A A	8-Br-G C	T T T C T C G -3' A A A G A G C -5'
3	5'-Fluo- 3'-	C T C T T T G A G A A A	8-NH <sub>2</sub> -G C	T T T C T C G -3' A A A G A G C -5'
4	5'-Fluo- 3'-	C T C T T T G A G A A A	G C	T T T C T C G -3' A A A G A G C -5'
5	5'-Pho	C T C T T T	Fluo-3'	

## SI-7 Crystal structure of the carbocyclic 8OG -MutM complex

As already mentioned in the main text, in order to gain deeper insight into the exact arrangements of the important residues E2 and E5 relative to the lesion we solved a crystal structure of MutM in complex with 8OG containing DNA. The today available structures were obtained with a mutant version of MutM in which the critical E2 residue was replaced by a glutamine to generate a catalytically incompetent version [1]. In order to study the interaction of the lesion with the wild type enzyme we prepared a carbocyclic analogue of 8OG (c8OG), incorporated this building block as a phosphoramidite into a double stranded oligonucleotide and crystallized this duplex with wild type MutM.

MutM from *L. cremoris* and *B. stearothermophilus* share a sequence identity and similarity of 42%, and 61% (root mean square deviation superposition 1.25 Å), respectively, with the residues lining the active site and forming the 8OG capping loop (OCL) site conserved (Fig. SI-20). In the asymmetric unit of the crystal two protein molecules are bound to one DNA double strand, with one protein molecule interacting with the lesion. No significant difference between the two protein molecules can be observed and they superimpose with an root mean square deviation of 0.38 Å. The c8OG is flipped out from the DNA strand and 99° rotated, which is more than half way between *anti*- and *syn*-conformations, into the active site of MutM (Fig. SI-27 and Fig. SI-21). Although in the *B. staerothermophilus* E2Q complex 8OG is fully rotated into *syn* conformation, yet the active site residues are in structurally equivalent positions (Fig. 4a). However in the E2Q mutant N<sub>ε</sub>2 of Q2 forms a hydrogen bond with O8 of the lesion, which would not

be a favourable interaction with the E2 carboxy group in the wild type. The OCL is flexible and not defined in the electron density. It was previously observed that the OCL can be either flexible or closed [2]. Nevertheless, closure of the OCL is possible with the Watson-Crick face of the c8OG, forming H-bonds with the loop's main chain atoms (Fig. SI-21). It was postulated that the OCL couples lesion recognition to catalysis since mutation of the OCL completely abolishes enzyme activity [1].

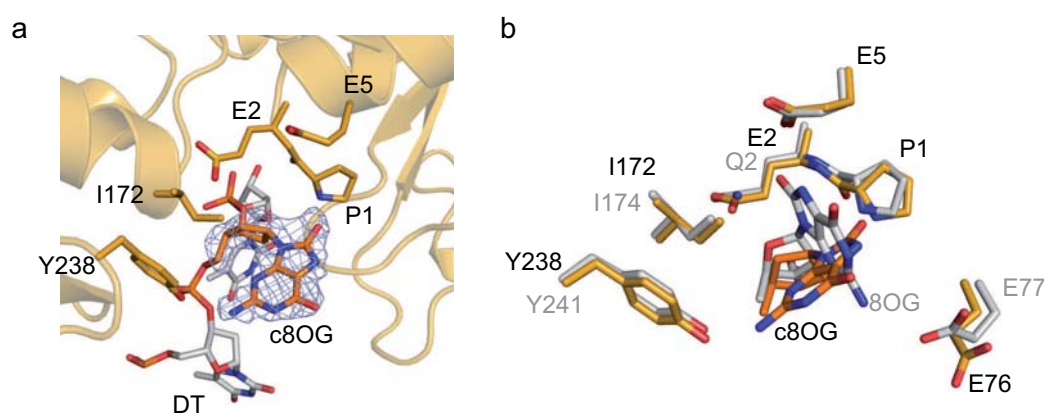


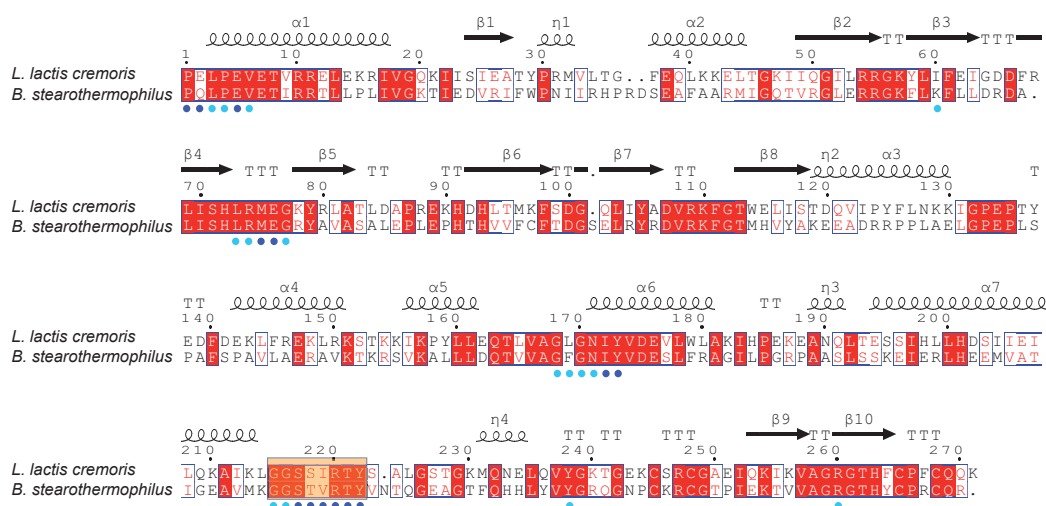
Figure SI-25: Structure of the *L. cremoris* MutM in binding c8OG-containing DNA. **a**, Simulated annealing Fo-DFc omit electron density of c8OG contoured at  $3\sigma$  in the active site of wild type MutM. **b**, Superposition of the *L. cremoris* MutM-c8OG and *B. stearothermophilus* MutM E2Q mutant 8OG complexes. In the *L. cremoris* MutM-c8OG complex (golden) c8OG (orange) is rotated by  $99^\circ$  taking up a position halfway between the *anti*- and *syn*- conformations. In the *B. stearothermophilus* MutM E2Q mutant and cross-linked 8OG complex (PDB code 1R2Y, grey), the 8OG is in the *syn*-conformation. Carbon atoms of the *B. stearothermophilus* MutM are coloured grey and *L. cremoris* MutM-c8OG in golden, with c8OG highlighted in orange.

### **SI-7.1 Co-crystallization, Data Collection, and Structure Determination**

The DNA was desalted with Sep-Pak<sup>®</sup> columns, lyophilized and then diluted and hybridized in a 1:1 molar ratio in Source S buffer. All complexes were crystallized by the 1:1 hanging drop vapour diffusion method. Prior to crystallization Protein (12.0 mg ml<sup>-1</sup>) and DNA were mixed in a 1:1.3 molar ratio respectively and incubated 30 min at 4°C. The samples were subsequently centrifuged to remove impurities. Crystals were obtained from solutions containing 100 mM Tris pH 7.5, 2 mM spermidine, 1 mM TCEP, 10  $\mu$ M ZnCl<sub>2</sub>, 90 mM LiSO<sub>4</sub> and 18% PEG. Crystals appeared after three weeks at 18 °C and were cryoprotected with ethylenglycol prior to flash freezing and storing in liquid nitrogen for data collection.

Data were collected at the synchrotron beamline PXI (Swiss Light Source (SLS), Villigen, Switzerland) and were processed with XDS [32]. Resolution cut-off were chosen using the correlation coefficient of random half-data sets (1/2 CC) of about 50% [33–35]. Crystals of MutM in complex with c8OG containing DNA belong to the orthorhombic space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> and were processed to about 2.3 Å spacing. The structure was solved by molecular replacement (PHASER [36]) using the coordinates of the protein in complex with DNA (PDB code 1XC8 [2]), sharing 97.5 % sequence identity as search model. In order to avoid model bias, the region around the lesion was deleted, the temperature factors were reset and simulated annealing (PHENIX [37]) was carried out prior to manual model building with COOT [38]. Refinement was carried out in REFMAC5 [39]. Electron density the lesion was clearly visible visible in the  $\sigma_A$  - weighted 2Fo-DFc and

Fo-DFc unbiased maps already after the first round of refinement. Restraints for c8OG were generated using PRODRG [40]. TLS-analysis was carried out with the TLSMD server[41] and TLS groups chosen accordingly. The model shows excellent geometry and no Ramachandran outliers [42]. For data collection, processing and refinement statistics see Tab. SI-10. Structural superpositions were done with SSM [43] using the protein main chain atoms as reference frame. Structural figures were prepared with PyMOL (Delano Scientific). The coordinates were deposited in the PDB at <http://www.ebi.ac.uk/pdbe/>. (PDB code: 4CIS)





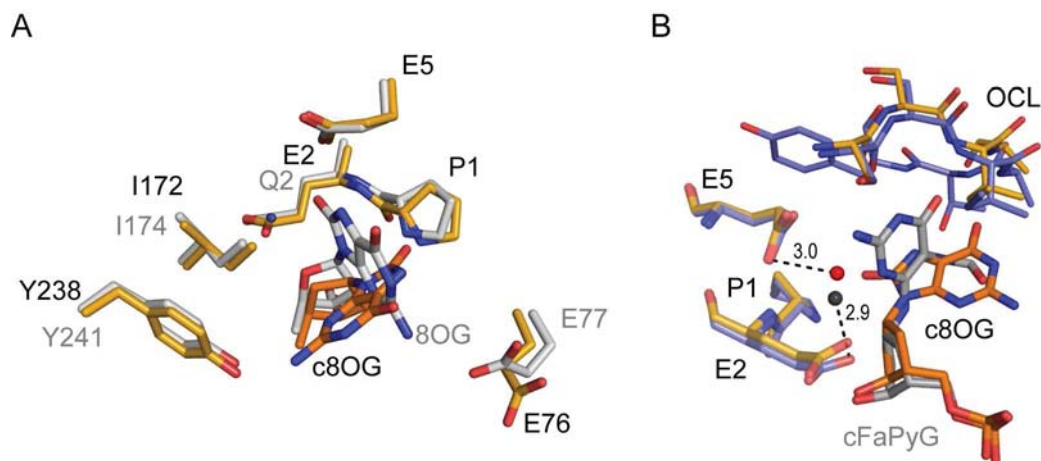


Figure SI-27: Superposition of the *L. cremoris* MutM-c8OG, cFaPyG and *B. stearothermophilus* MutM E2Q mutant 8OG complexes. A) In the *L. cremoris* MutM-c8OG complex (golden) c8OG (orange) is rotated by 99° taking up a position halfway between the *anti*- and *syn*- conformations. In the *B. stearothermophilus*, MutM E2Q mutant- cross-linked 8OG complex (PDB code 1R2Y, grey) the 8OG is in the *syn*-conformation. B) The OCL (residues 218-224) in the cFaPydG complex is closed and partially disordered in the c8OG complex. However c8OG could form hydrogen bonds with the main chain atoms of the closed OCL as in the cFaPyG complex. A water molecule in the active site of the c8OG (red sphere) and cFaPyG (black sphere) complex in H-bond distance to E5 and E2, respectively. Carbon atoms of the *B. stearothermophilus* MutM are coloured grey, *L. cremoris* MutM-c8OG in golden and cFaPyG blue, with c8OG highlighted in orange.

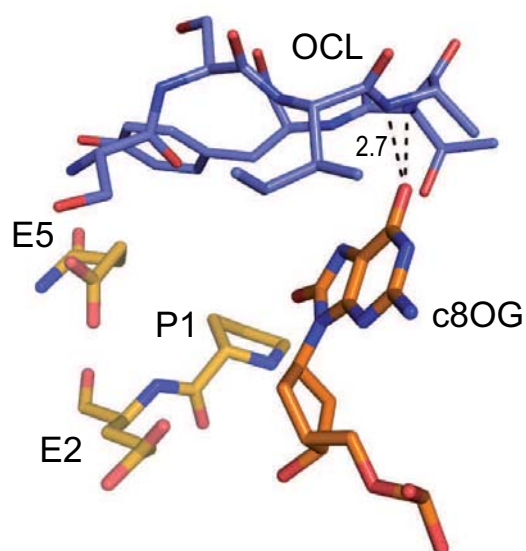


Figure SI-28: Interaction between c8OG and the modeled closed OCL loop. The loop (blue) was taken from the *L. cremoris* MutM in complex with cFapyG (PDB code 1XC8).

Table SI-10: Data processing and structure refinement statistics. Values in parentheses correspond to the highest resolution shell.

Wavelength (Å)	1.076
Resolution range (Å)	40.93 - 2.05 (2.12 - 2.05)
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell (Å)	43.06 112.69 132.88
$\alpha, \beta, \gamma(^{\circ})$	90 90 90
Total reflections	244,088 (18,628)
Unique reflections	41,439 (3,993)
Multiplicity	5.9 (4.7)
Completeness (%)	99.9 (99.3)
Mean I/sigma(I)	10.73 (1.50)
Wilson B-factor	28.86
R-merge	0.146 (1.185)
R-meas	0.16
CC1/2	0.996 (0.498)
CC*	0.999 (0.815)
R-work	0.214 (0.309)
R-free	0.239 (0.317)
Number of non-hydrogen atoms	5001
Macromolecules	4835
Ligands	31
Water	135
Protein residues	560
RMS(bonds)	0.017
RMS(angles)	1.8
Ramachandran favored (%)	98
Ramachandran outliers (%)	0
Average B-factor	30.1
Macromolecules	30.3
Ligands	27.6
Solvent	23.9

## References

- [1] Fromme, J. C.; Verdine, G. L. *The Journal of biological chemistry* **2003**, 278, 51543–8.
- [2] Coste, F.; Ober, M.; Carell, T.; Boiteux, S.; Zelwer, C.; Castaing, B. *The Journal of Biological Chemistry* **2004**, 279, 44074–83.
- [3] Gilboa, R.; Zharkov, D. O.; Golan, G.; Fernandes, A. S.; Gerchman, S. E.; Matz, E.; Kycia, J. H.; Grollman, A. P.; Shoham, G. *The Journal of Biological Chemistry* **2002**, 277, 19811–6.
- [4] Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; III, W. B. A.; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. *Nuc. Acids. Res.* **2007**, 35, 375-83.
- [5] Case, D. *et al.* “AMBER 10”, Technical Report, University of California, San Francisco, 2008.
- [6] Perlow-Poehnelt, R. a.; Zharkov, D. O.; Grollman, A. P.; Broyde, S. *Biochemistry* **2004**, 43, 16092–105.
- [7] Miller, J. H.; Fan-Chiang, C.-C. P.; Straatsma, T. P.; Kennedy, M. a. *Journal of the American Chemical Society* **2003**, 125, 6331–6.
- [8] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, 79, 926.
- [9] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.;

- Villa, E.; Chipot, C.; Skeel, R. D.; KalÃ, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.
- [10] Song, K.; Hornak, V.; de Los Santos, C.; Grollman, A. P.; Simmerling, C. *Biochemistry* **2006**, *45*, 10886–94.
- [11] Sherwood, P. *et al. Journal of Molecular Structure (Theochem)* **2003**, *1*, 632.
- [12] Shao, Y. *et al. Phys. Chem. Chem. Phys.* **2006**, *8*, 3172-3191.
- [13] “TURBOMOLE V6.5 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>.”, .
- [14] Kussmann, J.; Ochsenfeld, C. *The Journal of Chemical Physics* **2013**, *138*, 134114.
- [15] Smith, W.; Yong, C. W.; Rodger, P. M. *Molecular Simulation* **2002**, *28*, 385-471.
- [16] de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, M.; Kramer, G. J. *J. Phys. Chem. B* **1999**, *103*, 6133-6141.
- [17] Kastner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. *J Phys Chem A* **2009**, *113*, 11856-65.
- [18] Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571-2577.
- [19] Dirac, P. A. M. *Camb. Phil. Soc.* **1930**, *26*, 376-385.
- [20] Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200-1211.

- [21] Becke, A. D. *Phys. Rev. A* **1988**, 38, 3098–3100.
- [22] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785–789.
- [23] Perdew, J. P. *Phys. Rev. B* **1986**, 33, 8822–8824.
- [24] Slater, J. C. *Phys. Rev.* **1951**, 81, 385–390.
- [25] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *The Journal of Chemical Physics* **2010**, 132, 154104.
- [26] Grimme, S.; Ehrlich, S.; Goerigk, L. *Journal of Computational Chemistry* **2011**, 32, 1456–1465.
- [27] Nam, K.; Verdine, G. L.; Karplus, M. *Journal of the American Chemical Society* **2009**, 131, 18208–9.
- [28] Dinner, a. R.; Blackburn, G. M.; Karplus, M. *Nature* **2001**, 413, 752–5.
- [29] Abad, E.; Rommel, J. B.; Kaestner, J. *The Journal of biological chemistry* **2014**, M114.558494–.
- [30] Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, 19, 553-566.
- [31] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, 132, 154104.
- [32] Kabsch, W. *Acta Crystallogr. D, Biol. Crystallogr.* **2010**, 66, 133-44.
- [33] Diederichs, K.; Karplus, P. A. *Acta Crystallogr. D, Biol. Crystallogr.* **2013**, 69, 1215-22.
- [34] Evans, P. *Science* **2012**, 336, 986-7.

- [35] Karplus, P. A.; Diederichs, K. *Science* **2012**, *336*, 1030-3.
- [36] McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, *40*, 658-674.
- [37] Adams, P. D. *et al. Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213-21.
- [38] Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallographica Section D - Biological Crystallography* **2010**, *66*, 486.
- [39] Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr D, Biol Crystallogr* **2011**, *67*, 355-67.
- [40] Schuttelkopf, A. W.; van Aalten, D. M. *Acta Crystallogr. D, Biol. Crystallogr.* **2004**, *60*, 1355-63.
- [41] Painter, J.; Merritt, E. A. *J. Appl. Crystallogr.* **2006**, *39*, 109-111.
- [42] Chen, V. B.; Arendall, W. B., r.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 12-21.
- [43] Krissinel, E.; Henrick, K. *Acta Crystallogr D, Biol Crystallogr* **2004**, *60*, 2256-68.
- [44] Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. *Bioinformatics* **2007**, *23*, 2947-8.





- V.4 Paper [4]: "Structural, Biochemical, and Computational Studies Reveal the Mechanism of Selective Aldehyde Dehydrogenase 1A1 Inhibition by Cytotoxic Duocarmycin Analogues", M. F. Koch, S. Harteis, I. D. Blank, G. Pestel, L. F. Tietze, C. Ochsenfeld, S. Schneider, S. A. Sieber, *Angew. Chem. Int. Ed.* 54, 13550 (2015).**



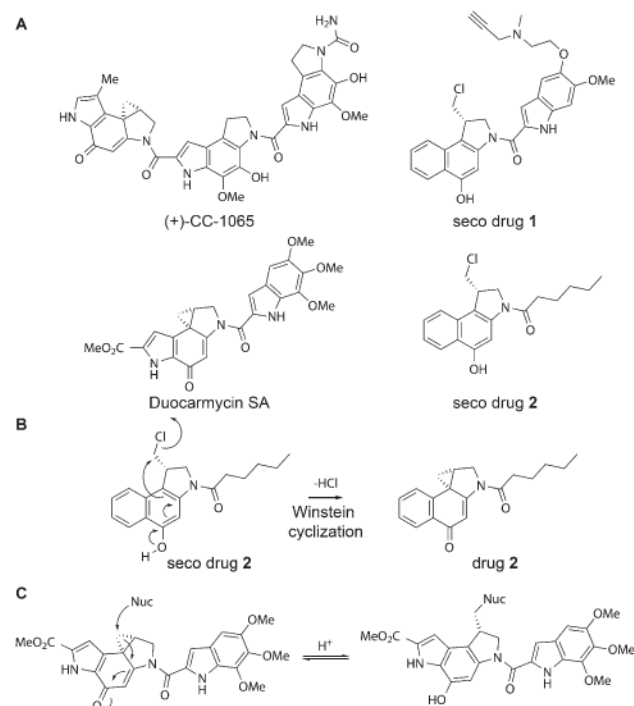


# Structural, Biochemical, and Computational Studies Reveal the Mechanism of Selective Aldehyde Dehydrogenase 1A1 Inhibition by Cytotoxic Duocarmycin Analogues

Maximilian F. Koch, Sabrina Harteis, Iris D. Blank, Galina Pestel, Lutz F. Tietze, Christian Ochsenfeld, Sabine Schneider,\* and Stephan A. Sieber\*

**Abstract:** Analogues of the natural product duocarmycin bearing an indole moiety were shown to bind aldehyde dehydrogenase 1A1 (ALDH1A1) in addition to DNA, while derivatives without the indole solely addressed the ALDH1A1 protein. The molecular mechanism of selective ALDH1A1 inhibition by duocarmycin analogues was unraveled through cocrystallization, mutational studies, and molecular dynamics simulations. The structure of the complex shows the compound embedded in a hydrophobic pocket, where it is stabilized by several crucial  $\pi$ -stacking and van der Waals interactions. This binding mode positions the cyclopropyl electrophile for nucleophilic attack by the noncatalytic residue Cys302, thereby resulting in covalent attachment, steric occlusion of the active site, and inhibition of catalysis. The selectivity of duocarmycin analogues for ALDH1A1 is unique, since only minor alterations in the sequence of closely related protein isoforms restrict compound accessibility.

Since their discovery in the 1970s, duocarmycin natural products such as (+)-CC-1065 and Duocarmycin SA have attracted much attention owing to their promising anticancer activity (Figure 1A).<sup>[1]</sup> The duocarmycin mode of action is based on its characteristic curved indole structure and a spirocyclopropylcyclohexadienone electrophile.<sup>[2]</sup> This alkylating moiety exists in conjugation with a vinylogous amide that tames its intrinsic reactivity. Duocarmycins are thus remarkably unreactive in solution.<sup>[3]</sup> Shape-selective recognition by the narrow AT-rich minor groove of DNA



**Figure 1.** A) Structures of (+)-CC-1065, Duocarmycin SA, and seco drugs 1 and 2. B) Conversion of seco drug 2 into the active molecule via Winstein cyclization. C) Nucleophilic attack at the cyclopropyl ring of Duocarmycin SA by DNA or protein (Nuc = Adenine N3 in DNA or Cys in ALDH1A1).

[\*] M. F. Koch,<sup>[†]</sup> S. Harteis,<sup>[†]</sup> Dr. S. Schneider, Prof. Dr. S. A. Sieber  
Center for Integrated Protein Science Munich CIPSM<sup>†</sup>Department of  
Chemistry, Technische Universität München  
Lichtenbergstrasse 4, 85747 Garching (Germany)  
E-mail: sabine.schneider@mytum.de  
stephan.sieber@tum.de

I. D. Blank, Prof. Dr. C. Ochsenfeld  
Center for Integrated Protein Science Munich CIPSM<sup>†</sup>Department of  
Chemistry, University of Munich  
Butenandtstrasse 7 (C), 81377 München (Germany)  
G. Pestel, Prof. Dr. L. F. Tietze  
Institute of Organic and Biomolecular Chemistry  
Georg-August-Universität Göttingen  
Tammannstrasse 2, 37033 Göttingen (Germany)

[†] These authors contributed equally to this work.

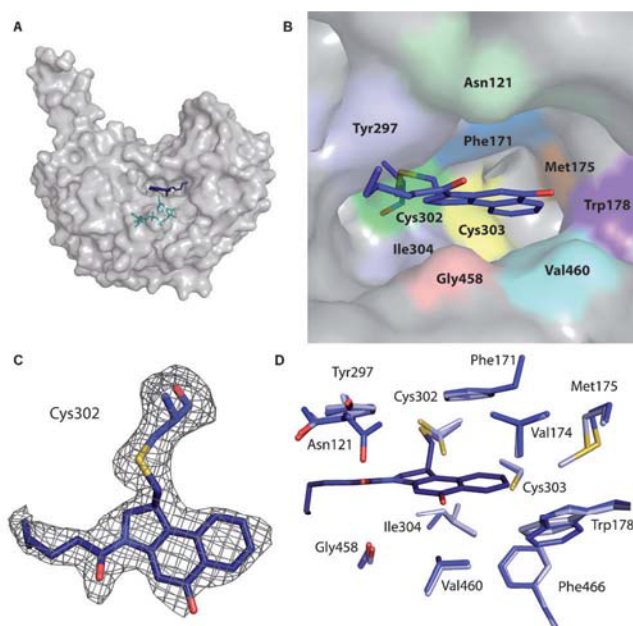
Supporting information (including experimental details and structure analysis) and ORCID(s) from the author(s) for this article are available on the WWW under <http://dx.doi.org/10.1002/anie.201505749>.

induces a conformational change that results in activation of the cyclopropyl moiety for nucleophilic attack by the 3' adenine N3 atom (Figure 1C).<sup>[2,4]</sup> Because of this unique target-based activation mechanism, duocarmycins were believed to solely address DNA; however, this notion was challenged when activity-based protein profiling (ABPP) with a duocarmycin-derived prodrug probe (seco drug 1, Figure 1A) in A459 lung cancer cells revealed aldehyde dehydrogenase 1A1 as an additional target.<sup>[5]</sup> Moreover, cytotoxic derivatives lacking the DNA-binding unit have been discovered<sup>[5,6]</sup> that do not exhibit pronounced in situ or in vitro DNA interaction. The affinity for ALDH1A1 was increased, for example, with seco drug 2 (Figure 1A), and imaging studies confirmed cytosolic localization without pronounced nuclear staining.<sup>[6b]</sup> Several independent RNA interference (RNAi) studies showed an important role for ALDH1A1 in cancer cell proliferation.<sup>[6b,7]</sup> However, the

exact function of ALDH1A1 in human biology is still not fully understood and tools to specifically manipulate its activity in the presence of related ALDH isoforms are needed.<sup>[8]</sup> Recent progress on the development of isoform-specific inhibitors in vitro has revealed some promising candidates. However, their selectivity in the context of the whole proteome has not yet been investigated.<sup>[9]</sup> Up to now, no structural data on the binding of seco drug 2 to ALDH1A1 were available and the proposed specific interaction between this protein and a molecule known to bind DNA was questioned.<sup>[10]</sup>

Since duocarmycin analogues are currently explored as drug–antibody conjugates (ADCs) a precise characterization of all possible target interactions would be desirable.<sup>[11]</sup> Herein, we reveal by high-resolution X-ray cocrystal structures of human and sheep ALDH1A1, a perfectly shaped pocket in the active site that precisely interacts with seco drug 2. Noncatalytic Cys302 covalently traps the compound and several aromatic amino acids engage in crucial interactions through van der Waals forces and  $\pi$ -stacking. The functional role of these residues was analyzed through mutational, kinetic, and computational studies.

The potency and DNA-binding ability of seco drug 2 were evaluated prior to crystallization through MTT cell toxicity and MS-based DNA interaction assays. Seco drug 2 effectively killed A549 cells with a half-maximal inhibitory concentration ( $IC_{50}$ ) of 28 nM (Figure S1 in the Supporting Information). In contrast to seco drug 1 and duocarmycin SA, both of which covalently modify AT-rich double-stranded DNA, seco drug 2 showed no such modification, thus suggesting a DNA-independent mode of action (Figure S2).<sup>[12]</sup> Since ALDH1A1 is a confirmed protein target with an important role in cell proliferation, we investigated this binding in greater detail. Recombinant ALDH1A1 from sheep<sup>[13]</sup> and human<sup>[9a]</sup> were cocrystallized with seco drug 2 (after Winstein cyclization, Figure 1B) in different crystal forms and the X-ray structures were determined to 1.8 and 2.1 Å resolution, respectively (Table S1 in the Supporting Information). The electron density maps of the compounds reveal binding in the preformed hydrophobic substrate-binding pocket next to the  $NAD^+$  cofactor without the introduction of conformational changes (Figure 2 and Figure S3). The apo and holo forms of the enzymes superimpose with an RMSD of 0.4 Å in both sheep and human structures (Figure S4). The flexible alkyl chain of seco drug 2 points towards the entrance of the active-site pocket and its definition by the electron density varies in the different structures (Figure S5). This is in agreement with inhibition results for indole-bearing duocarmycin analogues (seco drug 1) that need extra space in this region to accommodate their bulkier substituent.<sup>[5,6b]</sup> In fact, modeling shows that duocarmycin SA could fit into the channel without the induction of steric clashes (Figure S6). Surprisingly, in both crystal structures the electrophilic cyclopropyl moiety of seco drug 2 did not alkylate the nucleophilic active-site Cys303 (numbering according to the human structure<sup>[9a]</sup>) but solely the neighboring noncatalytic Cys302.<sup>[14]</sup> The covalent bond positions the adjacent tricyclic aromatic ring system in a tight-fitting hydrophobic binding pocket that stabilizes the ligand through interactions involving Phe171, Trp178 and Tyr297



**Figure 2.** X-ray crystal structure of seco drug 2 bound to human ALDH1A1 (PDB ID: 5AC2, this work). A) Overall structure of the complex, showing ALDH1A1 as a surface and the seco drug and  $NAD^+$  cofactor as blue and green stick models, respectively. B) Close-up view of the hydrophobic catalytic pocket, with seco drug 2 bound to Cys302. C) Simulated-annealing  $F_o - F_c$  difference omit electron density map of seco drug 2 and Cys302 contoured at  $2.5\sigma$ . D) Structural superposition of the active sites of apo ALDH1A1 (light blue, PDB ID: 4WJ9) and in complex with seco drug 2 (dark blue).

(Figure 2B). In line with the high sequence identity between the human and sheep enzymes (92 % sequence identity, 97 % sequence similarity; Figure S7) the two structures can be superimposed with an RMSD of 0.4 Å.

Based on this binding mode, we next analyzed the impact of exposed amino acids on interaction with the ligand. Owing to the high overall structural similarity, we performed all biochemical studies with the human enzyme and independently confirmed the results by MD calculations<sup>[15]</sup> with the sheep enzyme. We first evaluated the extent of covalent alkylation by using intact protein mass spectrometry (MS) and gel-based fluorescent labeling with an alkyne probe version of seco drug 2.<sup>[6b]</sup> Addition of seco drug 2 or the probe (10-fold excess) at pH 7.4 resulted in ALDH1A1 alkylation after more than 2 h incubation (Figure S8), thus demonstrating more-efficient binding compared to duocarmycin SA (Figure S6). MS/MS sequencing with the seco drug 2 probe confirmed the crystallographic results, with only Cys302 being modified (Figure S9). Even slightly higher pH values (8.0–8.5) were associated with the detection of additional modification sites, including Cys456 and Cys464, as we reported earlier.<sup>[5]</sup> This assignment can be explained by the elevated nucleophilicity of cysteine thiols at the given pH values, which may reflect nonphysiological conditions. Previous mutational studies with ALDH1A1 had highlighted only Cys303 as essential for catalysis and clearly showed that the neighboring Cys302 is dispensable.<sup>[14]</sup>

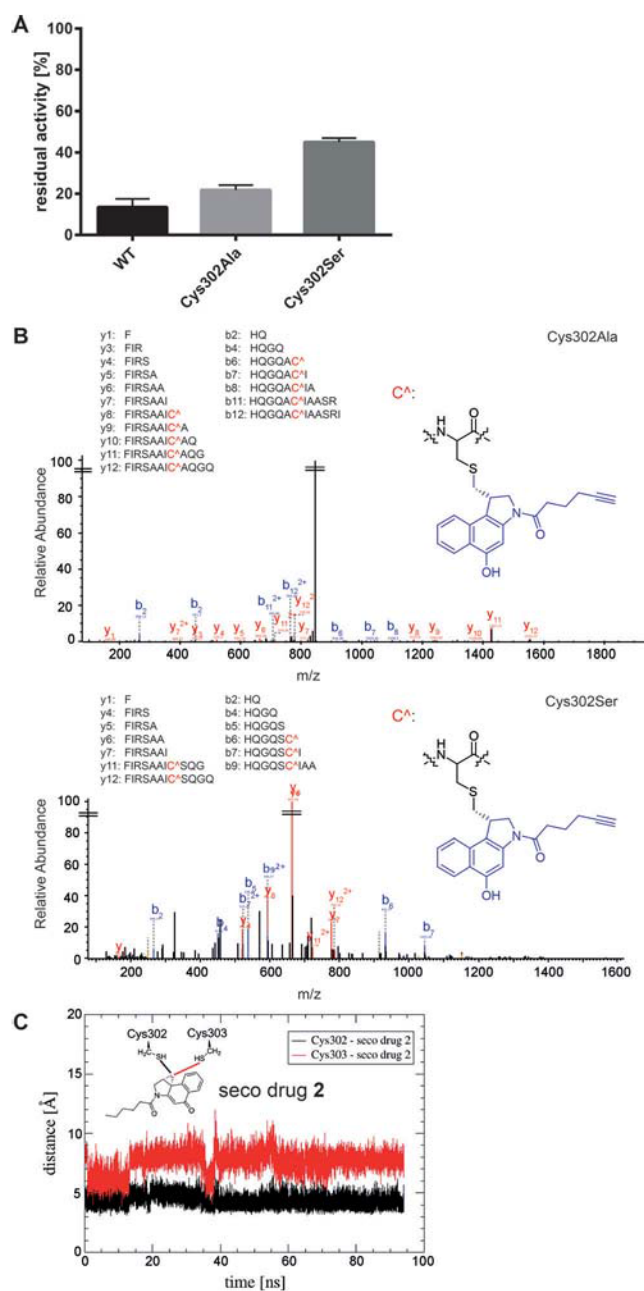


It is thus likely that the binding of seco drug **2** to Cys302 obstructs the active site and thereby prevents substrate access and turnover. Unexpectedly, ALDH1A1 Cys302Ala or Cys302Ser mutants were still inhibited by seco drug **2** (Figure 3A). MS/MS sequencing of both compound-treated mutant enzymes revealed exclusive alkylation of the neighboring active-site Cys303, thus suggesting that once the preferred nucleophile is eliminated, this proximal residue can substitute in the alkylation (Figure 3B).

To obtain a more quantitative measure of the binding kinetics, we determined  $K_i$  and  $k_{\text{inact}}$  values according to the method of Kitz and Wilson for covalent inhibitors.<sup>[16]</sup> The  $K_i$  values for the mutants were up to 10 times higher than those for the wild type, thus suggesting reduced affinity, especially in case of the Ser substitution (Figure S10). The binding preferences of seco drug **2** for either Cys302 or Cys303 were independently validated by molecular dynamics (MD) simulations, which confirmed Cys302 as the favored binding site owing to a stable interaction with the cyclopropyl moiety (Figure 3C). However, Cys303 is within reach of the cyclopropyl moiety and could thus easily bind in the mutants upon slight movement of Trp 178, as observed in a previous drug-protein cocrystal structure by Hurley et al. (PDB ID: 4X4L; Figure S11).<sup>[9a]</sup> In line with experimental data, the calculations further revealed that the Cys302Ala mutant provides better accessibility to Cys303 compared to the bulkier Cys302Ser mutant (Figure S12).

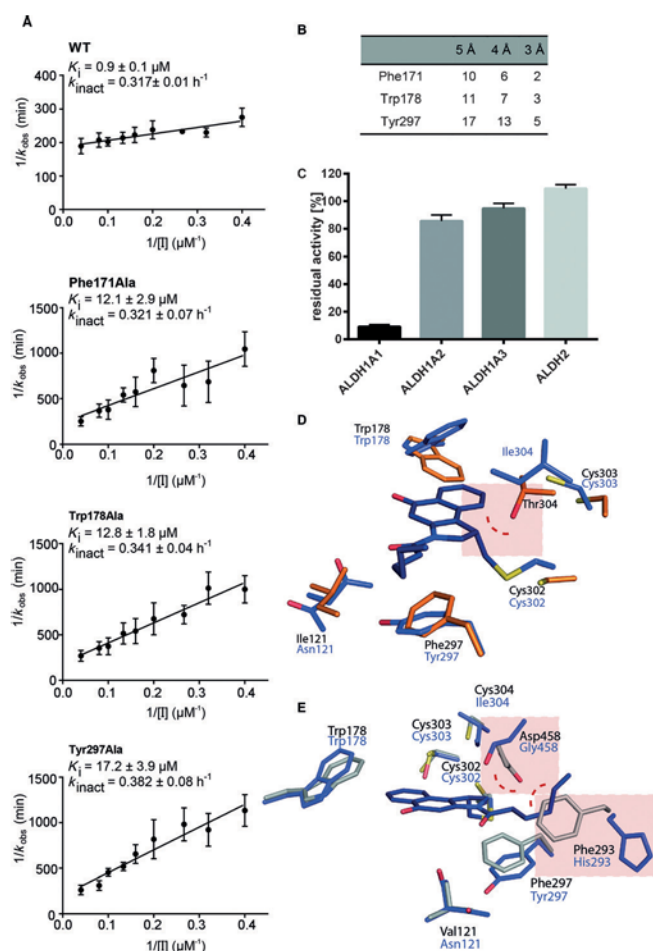
With a mechanistic appreciation of the alkylation reaction, we next focused on the hydrophobic pocket and its interactions with the ligand. The structures revealed several aromatic amino acid residues that stabilize the molecule by van der Waals forces and  $\pi$ -stacking (Figure 2B). To analyze their role in seco drug **2** binding, Phe 171, Trp 178, and Tyr 297 were individually exchanged for Ala and the corresponding mutant proteins were tested for compound inhibition. While the  $k_{\text{inact}}$  values were comparable to wild type and thus indicative of similar covalent enzyme inactivation rates, the  $K_i$  values were significantly elevated (up to 17-fold) in the mutant proteins, thus suggesting a reduction in stabilizing interactions (Figure 4A). MD calculations confirmed this hypothesis and revealed  $\pi$ -stacking and van der Waals interactions with Trp 178, as well as just van der Waals interactions with Phe 171 and Tyr 297 (Figure 4B and Figure S13).<sup>[17]</sup> Interestingly, Tyr 297 showed strong and stable van der Waals interactions with the whole aliphatic chain of seco drug **2** (Figure 4B and Figure S14) during all MD simulations. Compared to Trp 178 and Phe 171, there are more atoms of Tyr 297 within reach of seco drug **2**, thus highlighting this residue as the most important interaction partner (Figure 4B). Additional van der Waals interactions with Leu 174, Ile 304, and Val 460 were also observed (Figure S13).

The unique fit into this  $\pi$ -stacking network, as well as the shape, accessibility, and polarity of the corresponding binding pocket, provide an explanation for the previously observed proteome selectivity for ALDH1A1.<sup>[5,6b]</sup> Interestingly, even the closely related ALDH isoforms ALDH1A2, ALDH1A3, and ALDH2 (all about 70% sequence identity) were not or only slightly inhibited (ALDH1A2) by seco drug **2**. This is



**Figure 3.** A) Propionaldehyde-based activity assay with wild type (WT) enzyme and the binding-site mutants Cys302Ala and Cys302Ser. Residual activity was measured relative to the uninhibited enzyme activity (shown in %) after 5 h of incubation with a 50-fold excess of seco drug **2** (mean  $\pm$  SD,  $n = 3$ ). B) LC-MS/MS binding-site identification in human ALDH1A1 for Cys302Ser and Cys302Ala mutants with a probe version of seco drug **2**. Displayed are fragmented peptides (y and b ions) identified by MS/MS sequencing. The modified cysteine residues are indicated by C<sup>+</sup> (highlighted in red). C) Distance of the sulfur atoms of Cys302 and Cys303 to the cyclopropyl moiety of seco drug **2** from MD calculations. Cys302 is always closer to the cyclopropyl moiety than Cys303, which supports the selectivity of seco drug **2** for Cys302.

likely due to steric clashes, for example, as a result of the substitution of Ile 304 (ALDH1A1) by Thr (ALDH1A2/3) or of Gly 458/His 293 (ALDH1A1) by Asp 458/Phe 293 (ALDH2) (Figure 4C–E). Duocarmycin analogues are thus



**Figure 4.** A) Determination of  $K_i$  and  $k_{\text{inact}}$  values according to the method of Kitz and Wilson for WT ALDH1A1 and mutants where the interacting residues for seco drug 2 (Phe171Ala, Trp178Ala, and Tyr297Ala) were mutated to alanine. B) Average number of Phe 171, Trp 178, and Tyr 297 atoms interacting with seco drug 2 at the indicated distances (according to MD calculations). C) Propionaldehyde-based activity assay with related human isoforms ALDH1A2, ALDH1A3, and ALDH2. Residual activity was measured after 5 h of incubation with a 50-fold excess of seco drug 2. D) Active site of human ALDH1A1 with seco drug 2 (blue, PDB ID: 5AC2, this work) and the apo structure of ALDH1A2 (orange, PDB ID: 1B19). E) Active site of human ALDH1A1 with seco drug 2 (blue, PDB ID: 5AC2, this work) and the apo structure of ALDH2 (gray, PDB ID: 3N80).

erence for ALDH1A1 compared to other ALDH isoforms, their proteome-wide selectivity has not yet been evaluated.

In summary, this study elucidates the basis for the selective interaction between a duocarmycin analogue and ALDH1A1. A suite of structural, biochemical, kinetic, and computational methods revealed binding to a hydrophobic pocket, stabilization through van der Waals interactions and  $\pi$ -stacking, and covalent attachment to a cysteine residue directly adjacent to the catalytic residue. This mechanism is unique to ALDH1A1, and even closely related ALDH isoforms are not capable of binding to seco drug 2 owing to steric restrictions at their entrance channel and the catalytic site. The discovery of this unprecedented binding mode is thus not only of utmost importance for drug development, for example, to enable adjustment of the compound design to be exclusively DNA or ALDH1A1 specific (see the Supporting Information), but also provides the first tool for a focused study of ALDH1A1 function in whole proteomes. In addition, future studies will address the question of whether there are noncovalent binders of seco drug 2 that could contribute to the mode of action.

## Acknowledgements

S.A.S. acknowledges funding by the Deutsche Forschungsgemeinschaft. S.A.S. and C.O. also acknowledges SFB749 and the excellence cluster CIPSM. L.F.T. was supported by VW-Stiftung and State of Lower Saxony. S.Sch. acknowledges funding by SFB749, CIPSM and Fonds der chemischen Industrie. We would like to thank Prof. Michael Groll for access to the crystallization facility, Megan Wright for critical revision of the manuscript, Volker Kirsch for help with cytotoxicity studies, and Carola Draxler for help with MALDI measurements.

**Keywords:** ALDH1A1 · duocarmycin · enzymes · inhibitors · protein structure

**How to cite:** *Angew. Chem. Int. Ed.* **2015**, *54*, 13550–13554  
*Angew. Chem.* **2015**, *127*, 13754–13758

- a) D. G. Martin, C. G. Chidester, D. J. Duchamp, S. A. Mizsak, *J. Antibiot.* **1980**, *33*, 902–903; b) T. Yasuzawa, Y. Saitoh, M. Ichimura, I. Takahashi, H. Sano, *J. Antibiot.* **1991**, *44*, 445–447; c) D. L. Boger, D. S. Johnson, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 3642–3649.
- K. S. MacMillan, D. L. Boger, *J. Med. Chem.* **2009**, *52*, 5771–5780.
- L. F. Tietze, B. Krewer, H. Frauendorf, *Eur. J. Mass Spectrom.* **2009**, *15*, 661–672.
- D. L. Boger, R. M. Garbaccio, *Bioorg. Med. Chem.* **1997**, *5*, 263–276.
- T. Wirth, K. Schmuck, L. F. Tietze, S. A. Sieber, *Angew. Chem. Int. Ed.* **2012**, *51*, 2874–2877; *Angew. Chem.* **2012**, *124*, 2928–2931.
- a) L. F. Tietze, J. M. von Hof, M. Muller, B. Krewer, I. Schuberth, *Angew. Chem. Int. Ed.* **2010**, *49*, 7336–7339; *Angew. Chem.* **2010**, *122*, 7494–7497; b) T. Wirth, G. F. Pestel, V. Ganai, T. Kirmeier, I. Schuberth, T. Rein, L. F. Tietze, S. A. Sieber,

important tools for analyzing the role of ALDH1A1 in cancer cells and stem-cell development. In fact, such tools are urgently needed since commercially available ALDH1A1 inhibitors like diethylaminobenzaldehyde and disulfiram are known to be unselective and also bind ALDH2.<sup>[18]</sup> Recent efforts in broad inhibitor screening combined with rational design have provided attractive starting points for more isoform-specific inhibitors. Among these are reversible binders, which have been shown by structural studies to address the hydrophobic pocket through a different binding mode to that of seco drug 2.<sup>[9a,19]</sup> In contrast to seco drug 2, their inhibitory effect and specificity for ALDH1A1 is based on an extension into a cavity next to the entrance to the binding pocket (Figure S15). Although these molecules exhibit a pref-



- Angew. Chem. Int. Ed.* **2013**, *52*, 6921–6925; *Angew. Chem.* **2013**, *125*, 7059–7063.
- [7] L. Yue, Z. M. Huang, S. Fong, S. Leong, J. G. Jakowatz, A. Charruyer-Reinwald, M. Wei, R. Ghadially, *Melanoma Res.* **2015**, *25*, 138–148.
- [8] P. Marcato, C. A. Dean, C. A. Giacomantonio, P. W. Lee, *Cell Cycle* **2011**, *10*, 1378–1384.
- [9] a) C. A. Morgan, T. D. Hurley, *J. Med. Chem.* **2015**, *58*, 1964–1975; b) A. C. Kimble-Hill, B. Parajuli, C. H. Chen, D. Mochly-Rosen, T. D. Hurley, *J. Med. Chem.* **2014**, *57*, 714–722; c) M. Khanna, C. H. Chen, A. Kimble-Hill, B. Parajuli, S. Perez-Miller, S. Baskaran, J. Kim, K. Dria, V. Vasiliou, D. Mochly-Rosen, T. D. Hurley, *J. Biol. Chem.* **2011**, *286*, 43486–43494.
- [10] a) M. Tercel, S. P. McManaway, E. Leung, H. D. Liyanage, G. L. Lu, F. B. Pruijn, *Angew. Chem. Int. Ed.* **2013**, *52*, 5442–5446; *Angew. Chem.* **2013**, *125*, 5552–5556; b) L. F. Tietze, S. A. Sieber, *Angew. Chem. Int. Ed.* **2013**, *52*, 5447–5449; *Angew. Chem.* **2013**, *125*, 5557–5559.
- [11] R. C. Elgersma, R. G. Coumans, T. Huijbregts, W. M. Menge, J. A. Joosten, H. J. Spijker, F. M. de Groot, M. M. van der Lee, R. Ubink, D. J. van den Dobbelsteen, D. F. Egging, W. H. Dokter, G. F. Verheijden, J. M. Lemmens, C. M. Timmers, P. H. Beusker, *Mol. Pharm.* **2015**, *12*, 1813–1835.
- [12] L. Tietze, B. Krewer, J. M. Von Hof, H. Frauendorf, I. Schuberth, *Toxins* **2009**, *1*, 134.
- [13] S. A. Moore, H. M. Baker, T. J. Blythe, K. E. Kitson, T. M. Kitson, E. N. Baker, *Structure* **1998**, *6*, 1541–1551.
- [14] J. Farres, T. T. Wang, S. J. Cunningham, H. Weiner, *Biochemistry* **1995**, *34*, 2592–2598.
- [15] R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, D. E. Shaw, *Annu. Rev. Biophys.* **2012**, *41*, 429–452.
- [16] R. Kitz, I. B. Wilson, *J. Biol. Chem.* **1962**, *237*, 3245–3249.
- [17] a) C. R. Martinez, B. L. Iverson, *Chem. Sci.* **2012**, *3*, 2191–2201; b) R. S. Paton, J. M. Goodman, *J. Chem. Inf. Model.* **2009**, *49*, 944–955.
- [18] J. S. Moreb, D. Ucar, S. Han, J. K. Amory, A. S. Goldstein, B. Ostmark, L. J. Chang, *Chem.-Biol. Interact.* **2012**, *195*, 52–60.
- [19] C. A. Morgan, T. D. Hurley, *Chem.-Biol. Interact.* **2015**, *234*, 29–37.

Received: June 24, 2015

Revised: August 5, 2015

Published online: September 16, 2015







## Supporting Information

### **Structural, Biochemical, and Computational Studies Reveal the Mechanism of Selective Aldehyde Dehydrogenase 1A1 Inhibition by Cytotoxic Duocarmycin Analogues**

*Maximilian F. Koch, Sabrina Harteis, Iris D. Blank, Galina Pestel, Lutz F. Tietze, Christian Ochsenfeld, Sabine Schneider,\* and Stephan A. Sieber\**

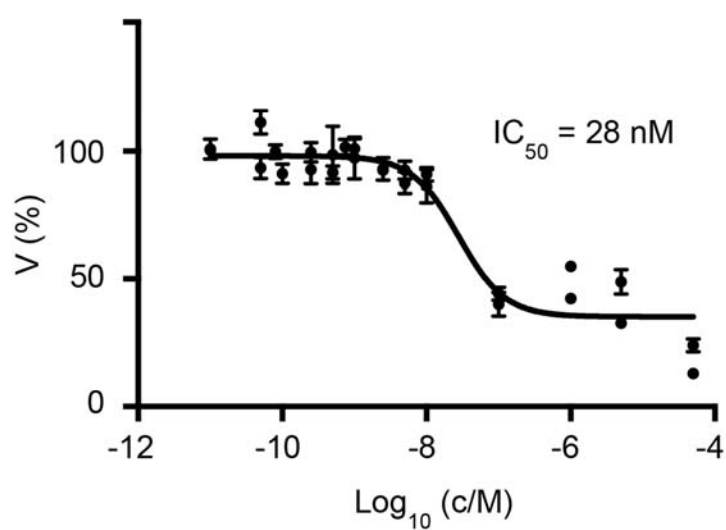
anie\_201505749\_sm\_miscellaneous\_information.pdf

# Supporting Information

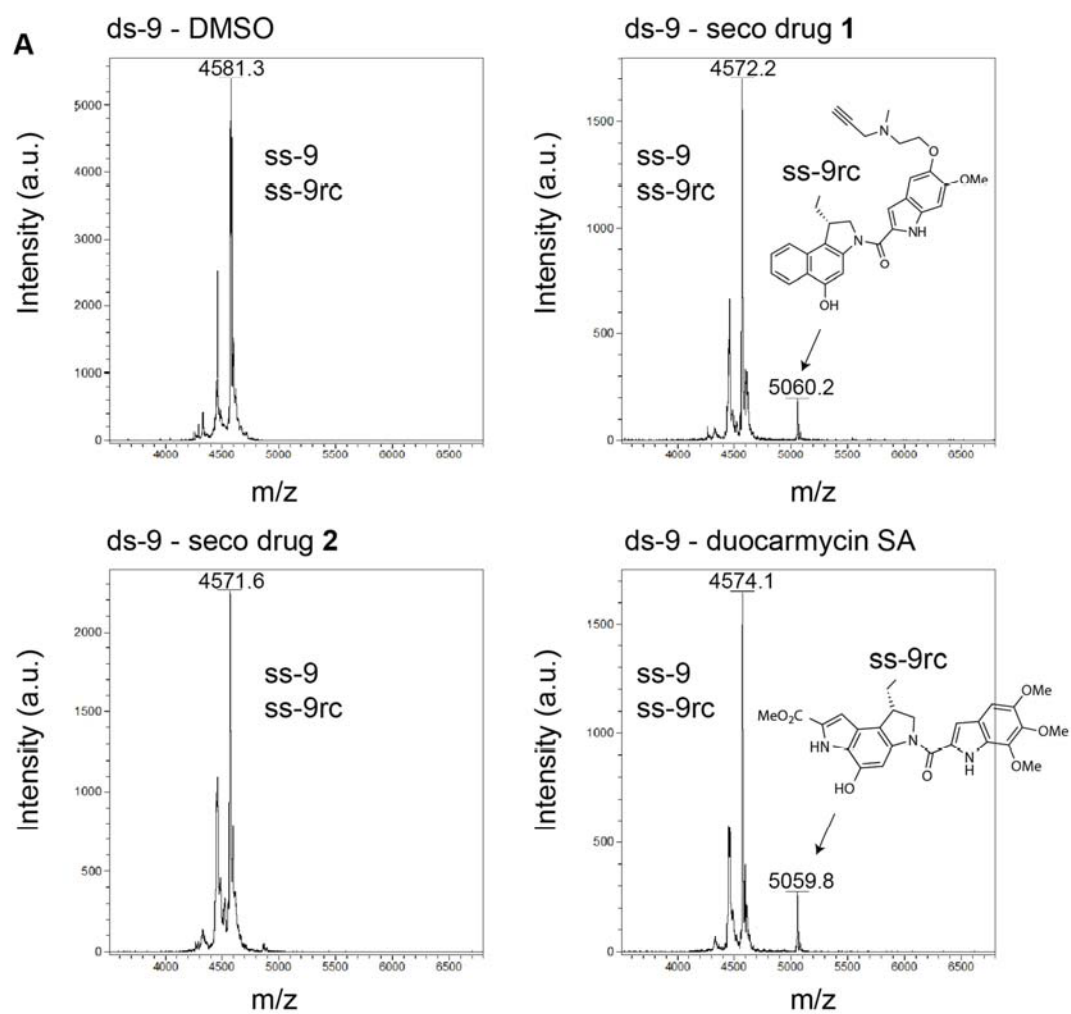
## Table of contents

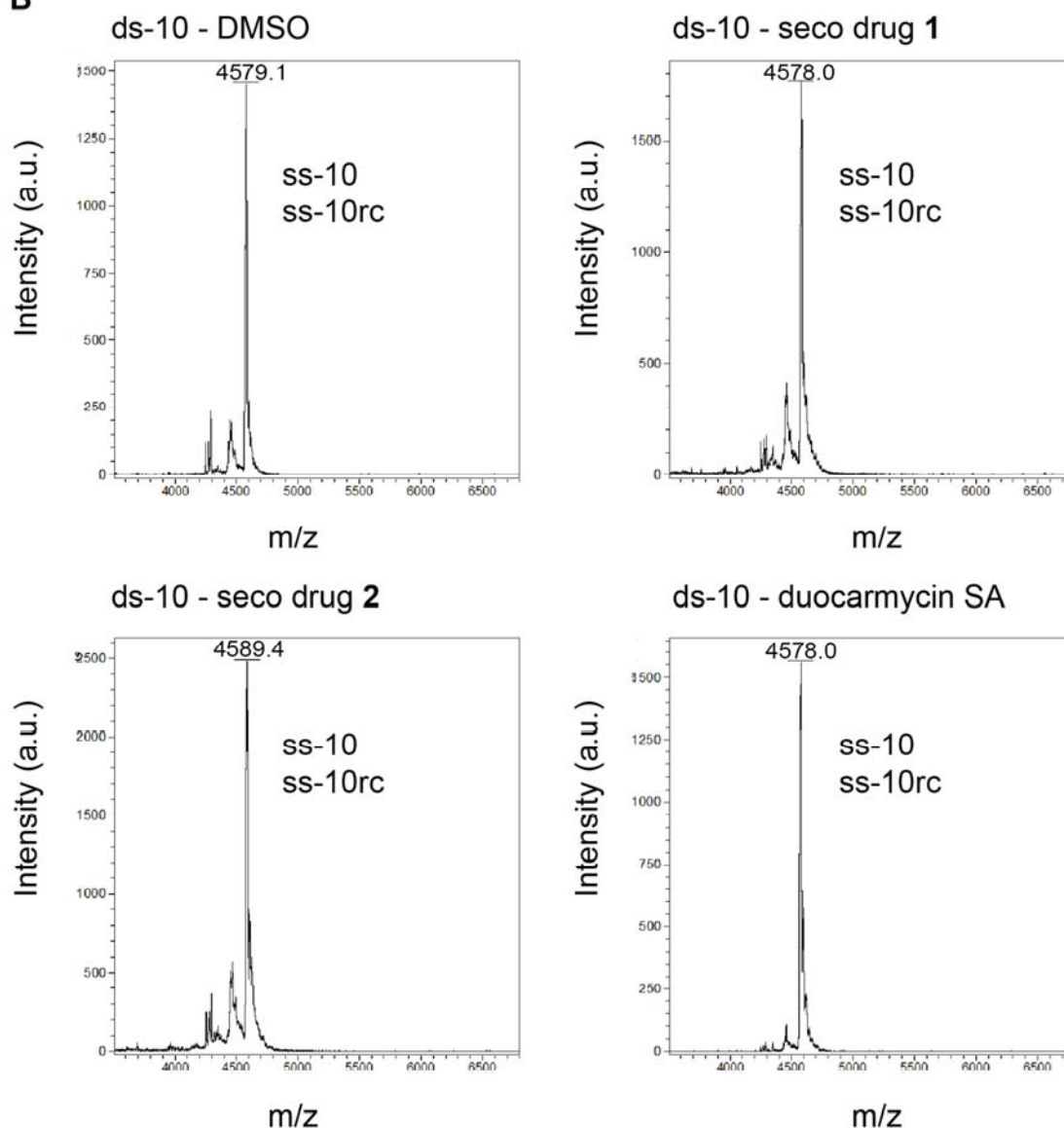
1. Figures .....	2
2. Tables .....	13
3. Experimental procedures.....	16
3.1 Cloning, expression and purification of ALDH1A1 .....	16
3.2 Crystallization, data collection and structure determination.....	16
3.3 Aldehyde dehydrogenase activity assay .....	18
3.4 Determination of $K_i$ and $k_{inact}$ .....	19
3.5 Intact protein mass spectrometry.....	19
3.6 Time dependent intact protein alkylation experiment .....	20
3.7 Gel-based fluorescent labeling experiment.....	20
3.8 Bottom-up proteomics for binding site identification .....	21
3.9 MTT cytotoxicity assay .....	22
3.10 DNA-alkylation experiment .....	24
3.11 Putative design criteria for selective ALDH1A1 inhibition and DNA-alkylation.....	24
4. Molecular dynamics simulations .....	25
5. References.....	26

## 1. Figures

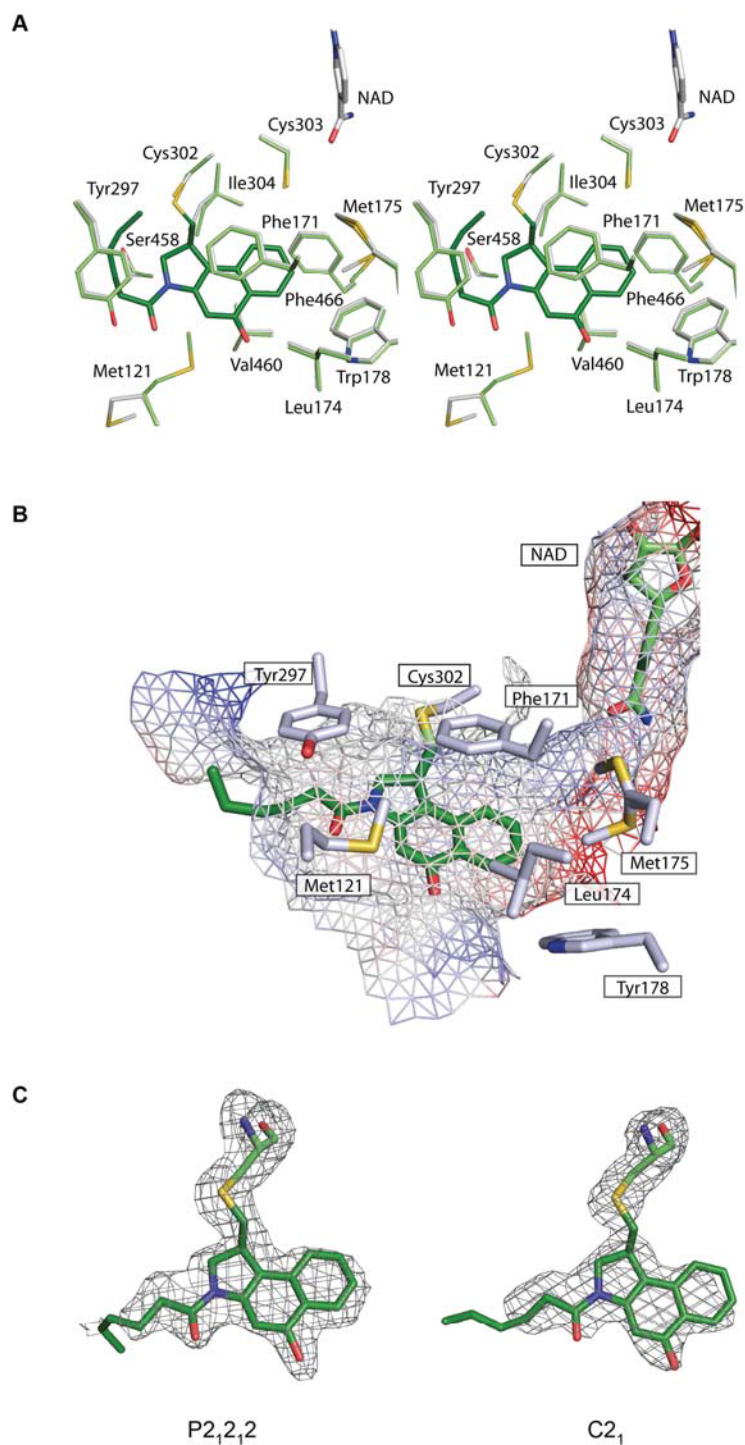


**Figure S1.** Fit of the viability  $V$  (MTT assay) for different concentrations ( $c$ ) of seco drug 2 ( $\text{IC}_{50} = 28 \text{ nM}$  with a 95% confidential interval of 12-61 nM).

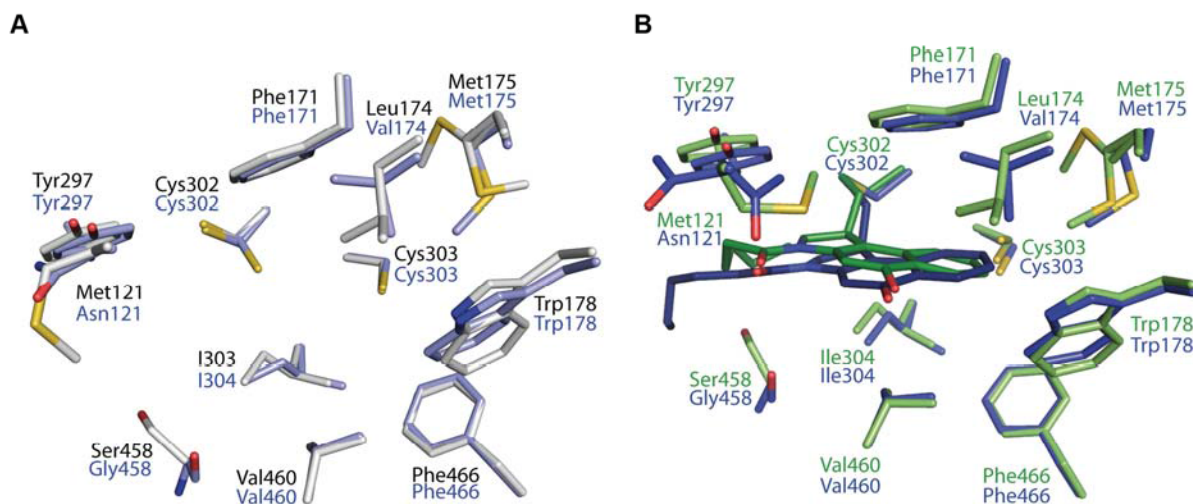


**B**

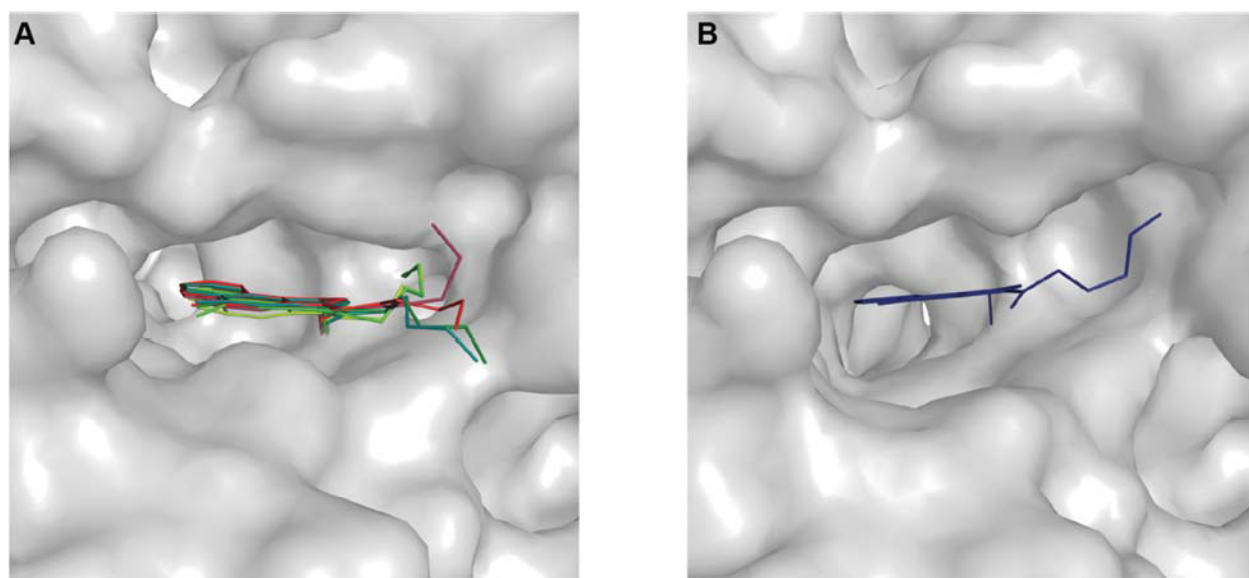
**Figure S2.** DNA-alkylation by seco drug 1, seco drug 2 and duocarmycin SA. The dsDNA was incubated with 25 fold excess duocarmycin SA, seco drug 1 or seco drug 2 for 24 h at 25 °C, followed by mass determination using matrix-assisted laser desorption ionization (MALDI).<sup>[1]</sup> A) DNA with AT-rich sequence (ds-9). DNA-alkylation can be observed for seco drug 1 and duocarmycin SA. For seco drug 2 no DNA-alkylation is detected. B) AT-deficient control dsDNA. No alkylation can be observed.



**Figure S3.** A) Stereo view of the binding pocket of sheep ALDH1A1, in complex with seco drug 2 (green) and the superimposed apo structure (white). B) Surface representation of the binding pocket of sheep ALDH1A1 with bound seco drug 2. The surface of the binding pocket is shown as mesh, with the charge distribution indicated by coloring (red = negative, blue = positive). C) Simulated annealing Fo-DFc omit electron density map of seco drug 2 bound to Cys301 of sheep ALDH1A1, contoured at 2.5 $\sigma$ , in the two different crystal forms (PDB codes 5AC1 (P2<sub>1</sub>2<sub>1</sub>2) and 5AC2 (C2<sub>1</sub>)).

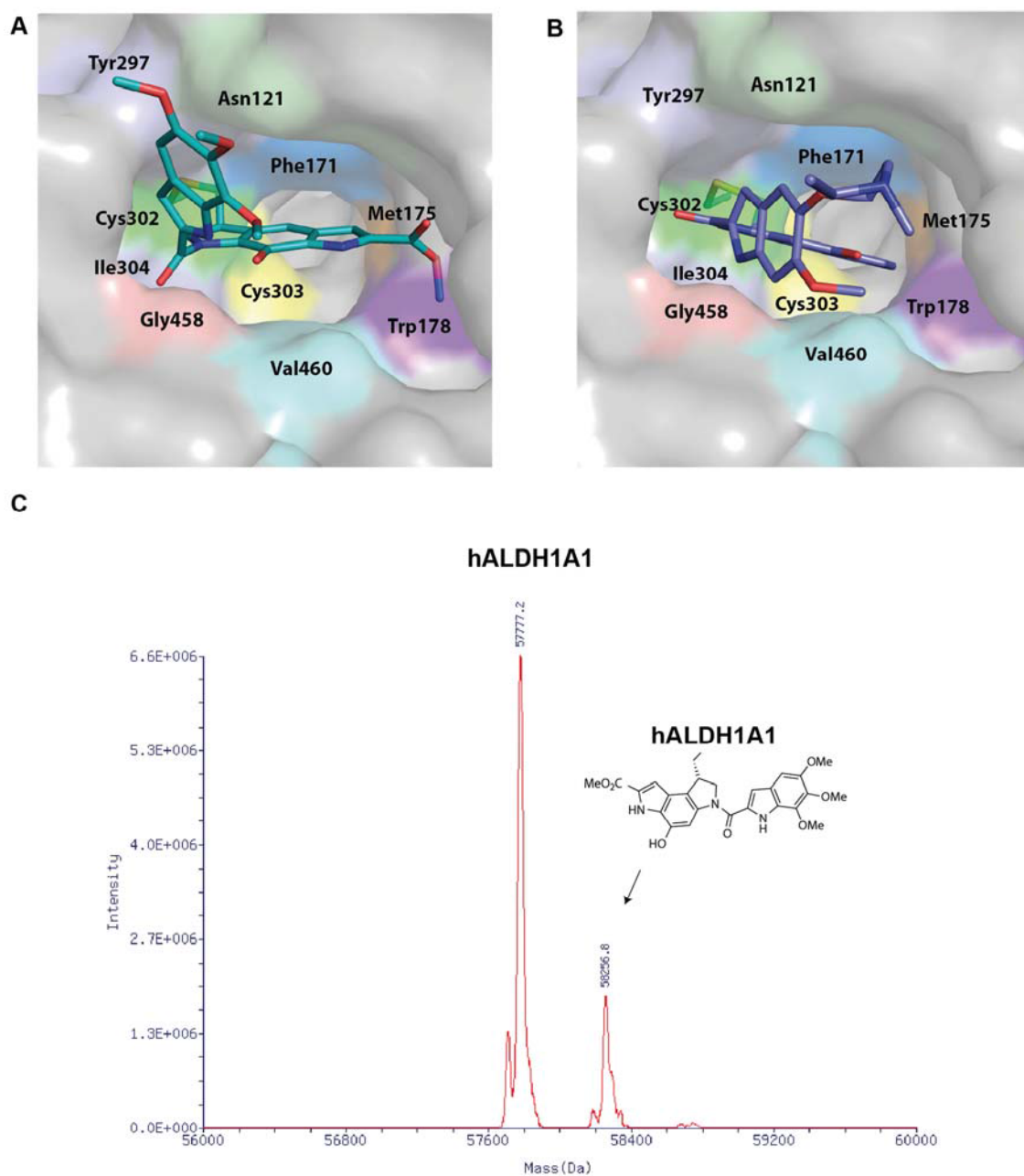


**Figure S4.** A) Structural superposition of the apo crystal structures of sheep (grey, this work, PDB code 5ABM) and human ALDH1A1 (blue, PDB code 4WJ9). B) Comparison of seco drug 2 bound to human ALDH1A1 (blue, PDB code 5AC2) and sheep ALDH1A1 (green, PDB code 5AC0/5AC1).



**Figure S5.** A) Comparison of seco drug 2 bound to the sheep ALDH1A1 monomers in the asymmetric unit of the crystal in the space group  $C2_1$  (4 molecules, green/cyan, PDB code 5AC0) and  $P2_12_12$  (2 molecules, red, PDB code 5AC1). B) Seco drug 2 bound to human ALDH1A1 ( $P4_1$ , one molecule, PDB code 5AC2).

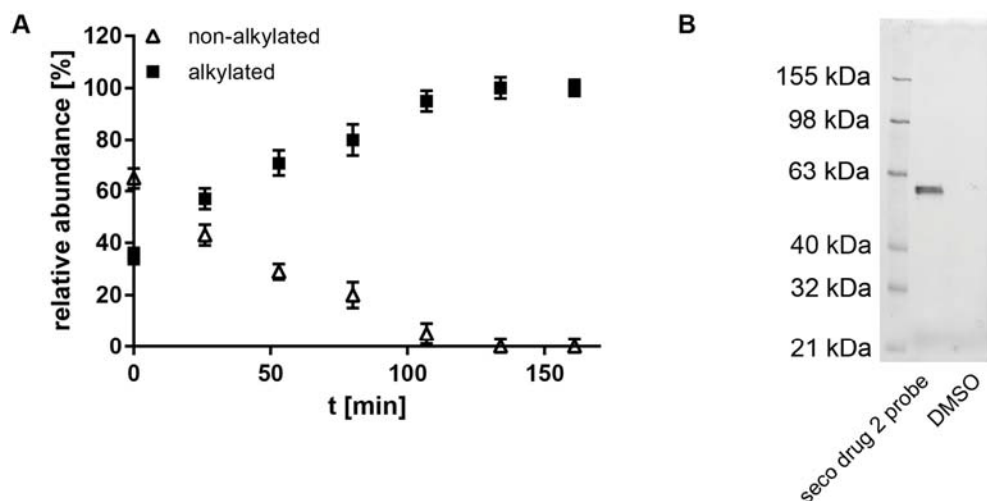




**Figure S6.** Model of duocarmycin SA bound to Cys302 (A) and seco drug 1 (B) bound to Cys 302 in the active site pocket of human ALDH1A1. C) Confirmation of alkylation and binding of duocarmycin SA to wild type hALDH1A1 by ESI-LC-MS. hALDH1A1 was incubated with 25 fold excess of duocarmycin SA for 3 h at 30 °C.

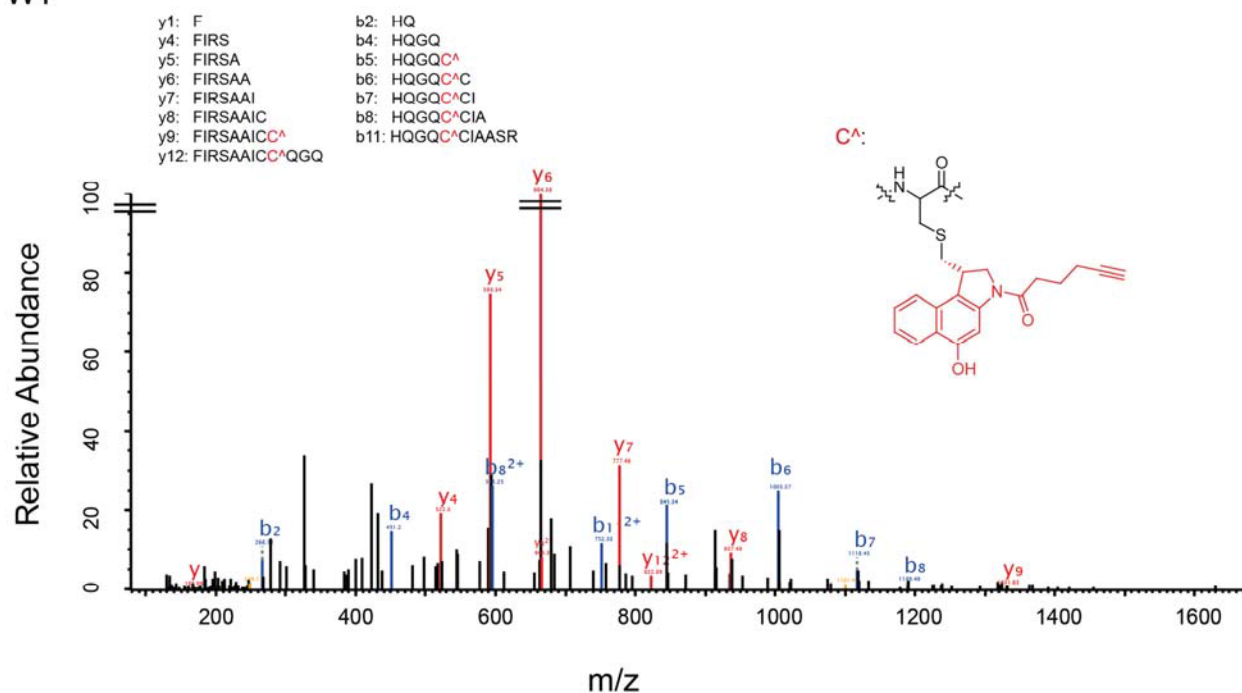




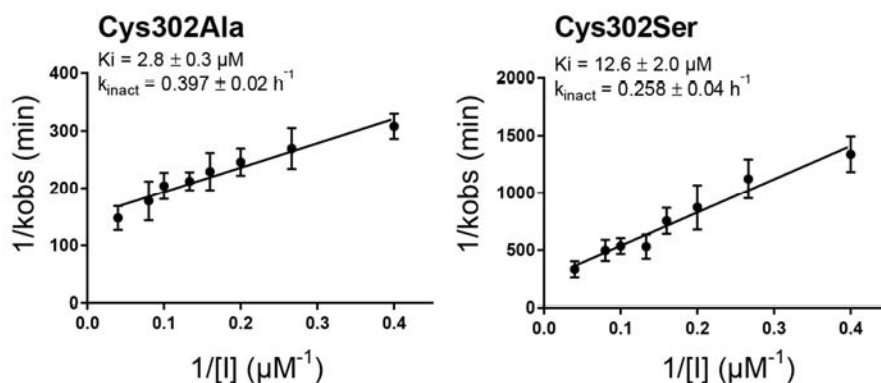


**Figure S8.** hALDH1A1 alkylation by seco drug 2. A) Time course of ALDH1A1 alkylation by relative quantification by intact protein MS after addition of 10 fold excess of seco drug 2 to human wild type ALDH1A1. B) In-gel fluorescent labeling of human wild type ALDH1A1 by click chemistry after addition of a 10 fold excess of seco drug 2 probe at pH 7.4 and incubation of 2.5 h.

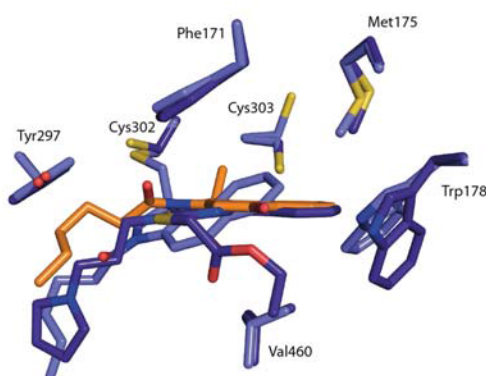
WT



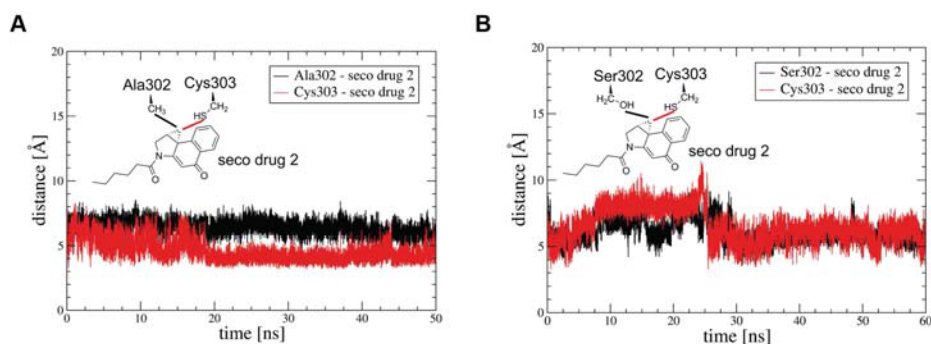
**Figure S9.** LC-MS/MS binding site identification in human ALDH1A1 with seco drug 2 probe. Displayed are fragmented peptides (y and b ions) identified by MS/MS sequencing. The modified cysteine-containing ions are indicated by: C<sup>A</sup> (highlighted in red).



**Figure S10.** Determination of  $K_i$  and  $k_{obs}$  according the method of Kitz and Wilson<sup>[3]</sup> for the ALDH1A1 binding site mutants Cys302Ala and Cys302Ser.

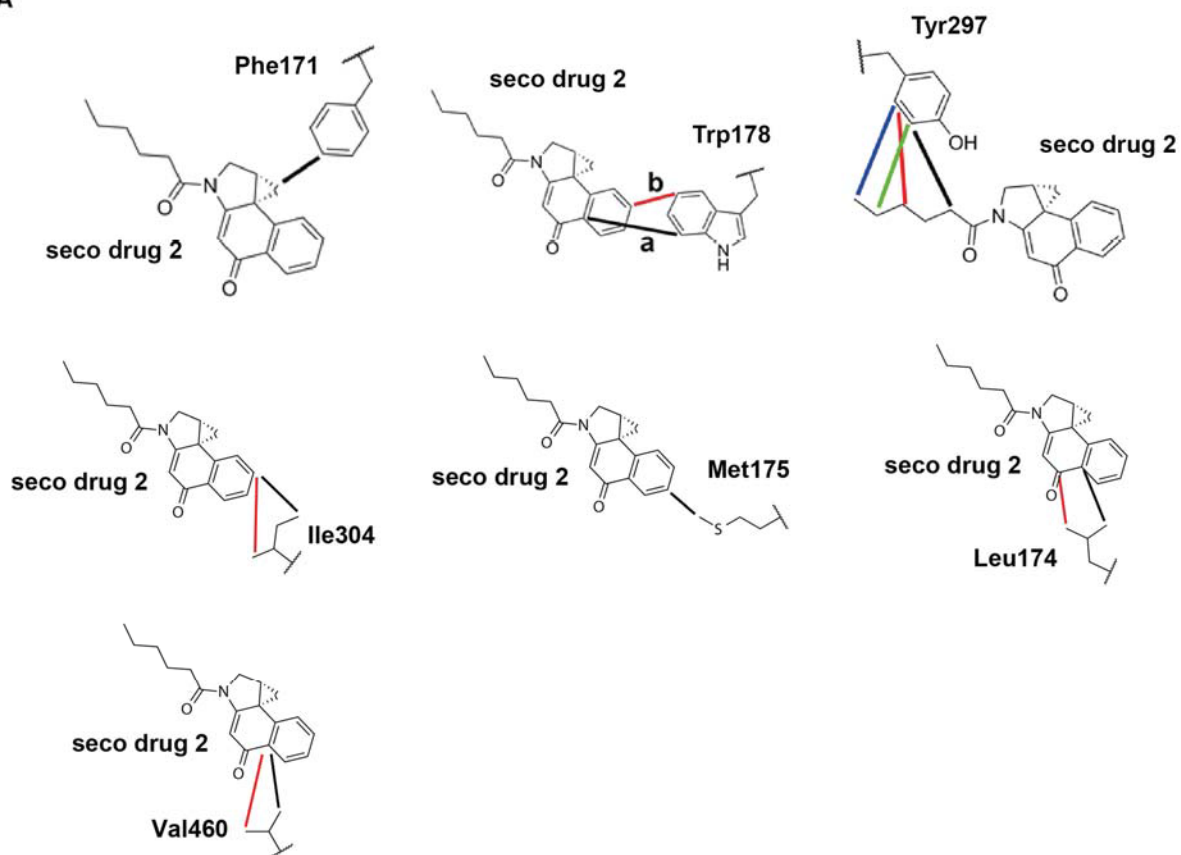


**Figure S11.** Model for seco drug 2 binding to Cys303. Shift of Trp178 in the hALDH1A1 structure in complex with the CM037 inhibitor (light blue, PDB code 4X4L).<sup>[4]</sup> This shift would allow seco drug 2 (dark blue) to move (orange) into proximity to Cys303.

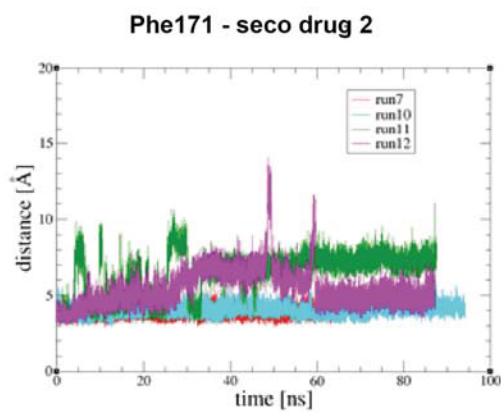


**Figure S12.** Structural analysis of the Cys302Ala and Cys302Ser interaction with seco drug 2 by FF-MD. A) Distance of the  $C_\beta$  atom of Cys302 and sulfur atom of Cys303 to the cyclopropyl moiety of seco drug 2 from MD simulations. Cys303 is closer to the cyclopropyl moiety than Ala302 and shows strong and permanent interaction. B) Distance of the oxygen side chain atom of Ser302 and sulfur atom of Cys303 to the cyclopropyl moiety of seco drug 2 from MD simulations. Cys303 shows multiple interactions with the cyclopropyl moiety.

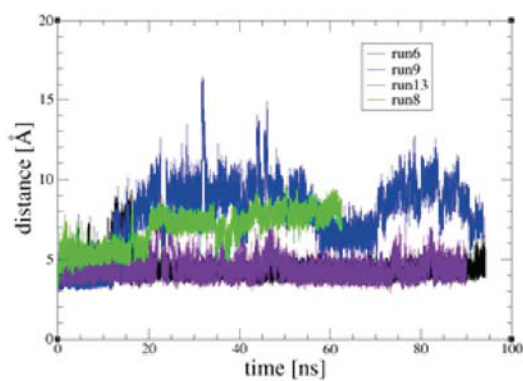
A



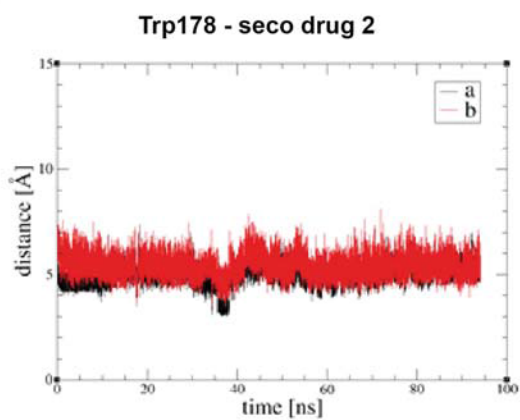
B



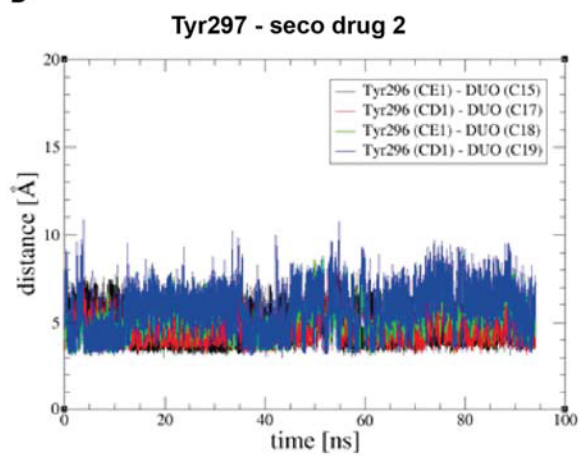
Phe171 - seco drug 2

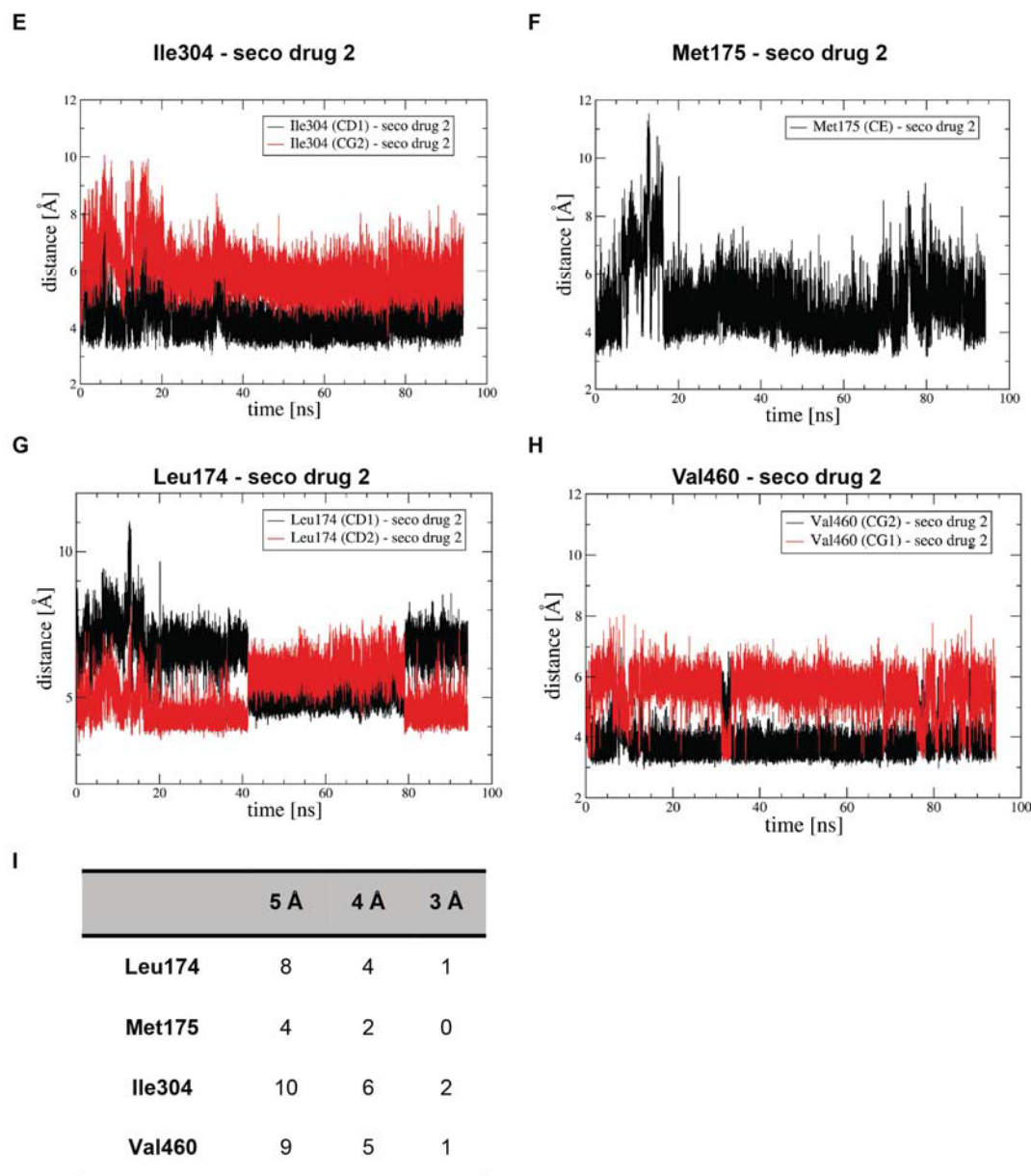


C



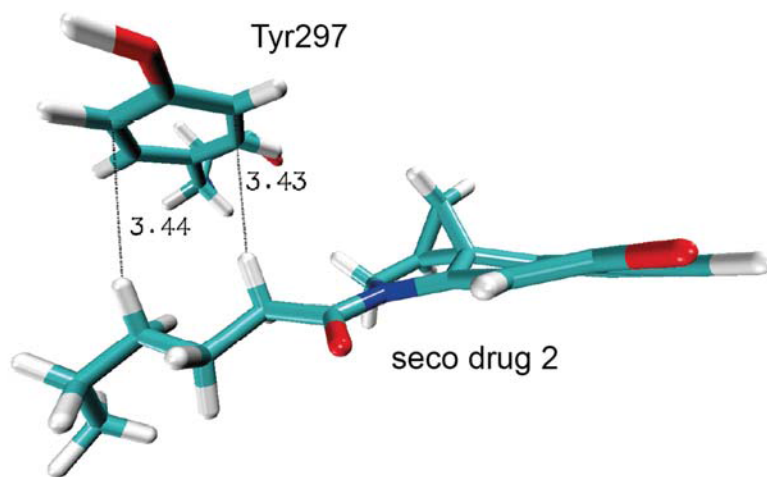
D



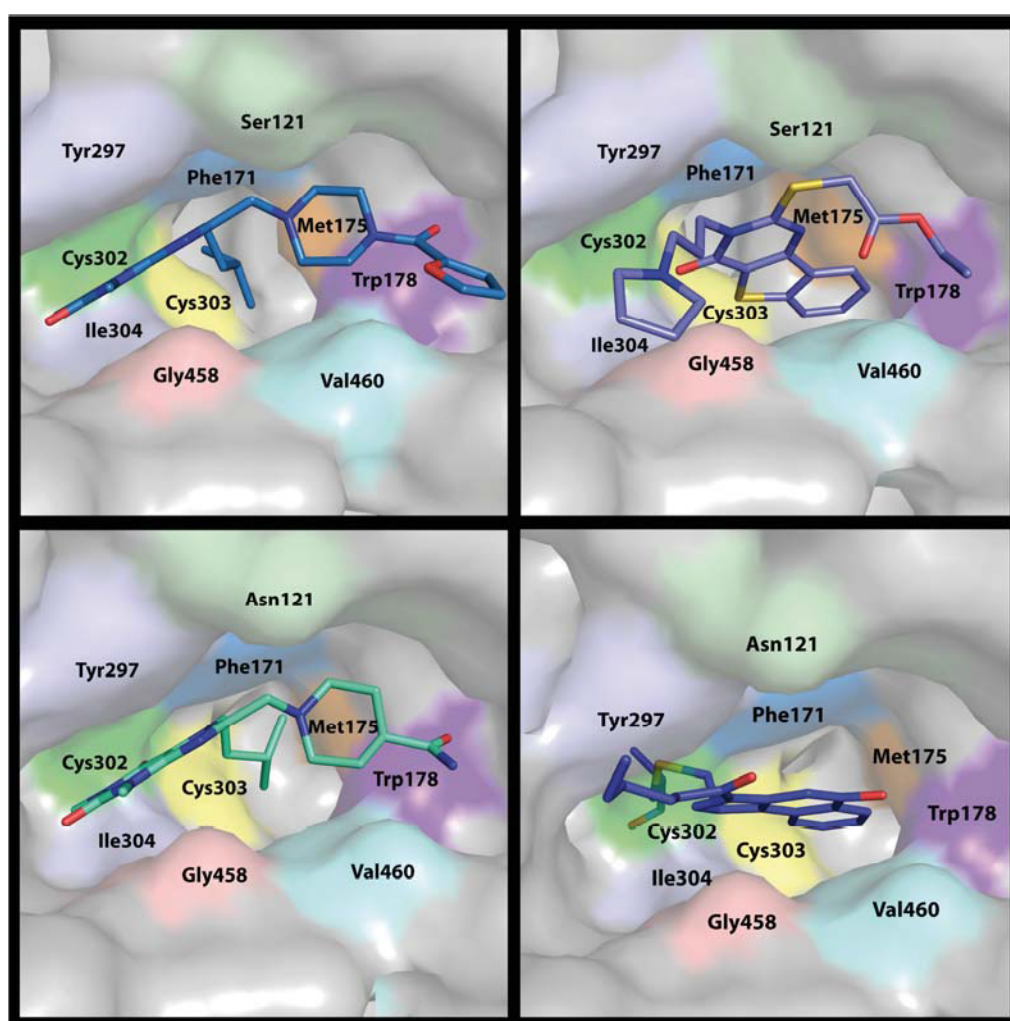


**Figure S13.** A) Definition of the plotted distance between the specified atoms for the interaction of Phe171, Trp178, Tyr297, Ile304, Met175, Leu174 and Val460 with seco drug 2. B) In 50% of MD simulations Phe171 shows very stable interaction with the cyclopropyl moiety of seco drug 2. C) Distances between Trp178 and seco drug 2 for a representative MD simulation. D) The interaction of Tyr297 with seco drug 2 out of one representative MD simulation. Tyr297 interacts with all atoms of the aliphatic chain of seco drug 2. This is shown for the distances between tyrosine atoms C<sub>ε</sub>1 and C<sub>δ</sub>1 and the carbon atoms of the aliphatic chain of seco drug 2. E) Distances between Ile304 and seco drug 2 for a representative MD simulation. F) Distances between Met175 and seco drug 2 for a representative MD simulation. G) Distances between Leu174 and seco drug 2 for a representative MD simulation. H) Distances between Val460 and seco drug 2 for a representative MD simulation. I) Average number of Leu174, Met175, Ile304 and Val460 atoms interacting with seco drug 2 at the indicated distances (MD calculations).





**Figure S14.** MD snapshot of the Interaction between Tyr297 and seco drug 2.



**Figure S15.** Comparison of hALDH1A1 inhibitors bound in the active site pocket. The inhibitors are bound at the entrance of the active side. Seco drug 2 is covalently bound to Cys302 located in the inner pocket. A) CM026 (PDB code 4WP7), B) CM037 (PDB code 4X4L), C) CM053 (PDB code 4WPN) and D) seco drug 2 (this work, PDB code 5AC2).

## 2. Tables

SI Table 1.

	oALDH1A1	oALDH1A1-seco drug 2	oALDH1A1-seco drug 2	hALDH1A1-seco drug 2
Wavelength (Å)	1	1	1	1
Resolution range (Å)	51.8 - 1.7 (1.76 - 1.7)	50.2 - 1.9 (1.96 - 1.9)	49.4 - 2.1 (2.16 - 2.1)	109.1 - 1.85 (1.91 - 1.85)
Space group	C 1 2 1	P 21 21 2	C 1 2 1	P 4 2 2
Unit cell	188.0 80.8 171.4 90 118 90	93.6 150.6 80.7 90 90 90	188.3 80.8 172.3 90 118 90	109.1 109.1 83.2 90 90 90
Total reflections	1,684,395 (158,838)	722,636 (61,494)	903,139 (67,598)	1,135,506 (102,214)
Unique reflections	249,604 (24,273)	90,805 (8,837)	134,836 (11,553)	43,648 (4,180)
Multiplicity	6.7 (6.5)	6.8 (6.0)	6.7 (5.8)	26.0 (24.5)
Completeness (%)	99.7 (97.3)	99.8 (98.1)	98.4 (84.9)	99.8 (98.1)
Mean I/sigma(I)	12.6 (2.1)	10.56 (1.43)	8.9 (1.7)	17.8 (1.2)
Wilson B-factor	19.5	21.8	28.9	27.4
R-merge	0.106 (0.861)	0.125 (1.02)	0.177 (0.921)	0.188 (2.676)
R-meas	0.115	0.136	0.192	0.192
CC1/2	0.998 (0.71)	0.998 (0.71)	0.993 (0.70)	0.999 (0.57)
CC*	0.999 (0.91)	0.999 (0.91)	0.998 (0.91)	1 (0.85)
R-work	0.180 (0.299)	0.153 (0.281)	0.229 (0.338)	0.193 (0.371)
R-free	0.203 (0.308)	0.197 (0.317)	0.265 (0.368)	0.228 (0.387)
Number of non-hydrogen atoms	16,987	8,593	16,425	4,094
macromolecules	15,464	7,760	15,427	3,849
ligands	180	134	268	69
Protein residues	1,984	994	1,988	494
RMS(bonds)	0.017	0.019	0.015	0.010
RMS(angles)	1.73	1.85	1.63	1.42
Ramachandran favored (%)	98	97	97	97
Ramachandran outliers (%)	0	0	0	0.2
Clashscore	2.83	2.5	2.8	1.4
Average B-factor	22.2	25.9	27.9	32.4
macromolecules	21.8	25.2	27.8	32.4
ligands	22.2	31.9	37.9	35.10
solvent	26.2	31.9	25.9	31.6

**SI Table 2.** Primers for site directed mutagenesis of hALDH1A1.

Name		Primer Sequence (5' to 3')
C301A	forw	gga tgc ggc tat aca agc ctg gcc ctg gtg gta g
C301A	rev	cta cca cca ggg cca ggc ttg tat agc cgc atc c
C301S	forw	gcg gct ata caa ctc tgg ccc tgg tgg
C301S	rev	cca cca ggg cca gag ttg tat agc cgc
F170A	forw	gca taa cca acg ggg cat tcc aag gaa tga ttg ggc cac ata
F170A	rev	tat gtg gcc aaa tca ttc ctt gga atg ccc cgt tgg tta tgc
W177A	forw	cag gcc cta tct tgc caa tga gca taa cca acg gga aat t
W177A	rev	aat ttc ccg ttg gtt atg ctc att gcg aag ata ggg cct g
Y296A	forw	cac tgg ccc tgg tgg gcg aat acc cca tgg tg
Y296A	rev	cac cat ggg gta ttc gcc cac cag ggc cag tg

**SI Table 3.** Peptides identified by mass spectrometry containing modified cysteine residues C<sup>+</sup> (highlighted in red). The table shows the fragment ion mass (MH<sup>+</sup>), the fragment ion charge (z), the mass error, the minimal reached p value (PEP) and the score value.

ALDH1A1	Peptide	MH <sup>+</sup>	z	Mass error [ppm]	PEP	Score
WT	HQGQC <sup>+</sup> CIAASRIF	1432.7	3	-0.31	1E-04	58
C301S	HQGQSC <sup>+</sup> IAASRIF	1416.7	3	0.52	9E-08	64
C301A	HQGQAC <sup>+</sup> IAASRIF	1400.7	3	1.3	4E-37	103

**SI Table 4.** Overview of performed MD-simulations.

system	total time [ns]	number of simulations
WT	720	13
C302A	550	8
C302S	550	8

**SI Table 5.** Base sequences of the four single stranded (ss) DNA oligonucleotides used in the present study.

Name	Base Sequence (5' to 3')
ss-9	cgg ctt ata tga ccg
ss-9rc	cgg tca tat aag ccg
ss-10	tgg ctg cga gca cct
ss-10rc	agg tgc tcg cag cca



### 3. Experimental Procedures

#### 3.1 Cloning, expression and purification of ALDH1A1

A codon optimized sequence (GeneArt) of the sheep ALDH1A1 (sALDH1A1) was cloned in frame with a sequence coding for an N-terminal Streptacin-affinity and *Tabacco Etch Virus* (TEV) protease cleavage site into pKM-RQ using the Gateway system (Gateway®, life technologies).

Point mutations were introduced in hALDH1A1 using Quick Change Mutagenesis (QuikChange II, Agilent Technologies) using the oligonucleotides listed in SI Table 2.

For expression, sALDH1A1-pDEST007 was transformed into *Escherichia coli* BL21 (DE3) and cells grown in LB medium supplemented with 100 µg/ml Ampicillin at 37 °C until an OD<sub>600</sub> of 0.6 was reached. Expression was induced by adding anhydrotetracyclin to a final concentration of 432 nM and the culture was incubated for 12 h at 18 °C. Cells were lysed by sonication and the protein purified by Streptactin-affinity (StrepTrap<sup>TM</sup> HP 5 mL, Qiagen). After removal of the Strep-tag using TEV protease, the enzyme was subject to size exclusion chromatography (HiLoad 16/60 Superdex 200 prep grade, GE healthcare), concentrated in 50 mM Tris-HCl pH 7.5, 10% glycerin and stored at -80 °C. hALDH1A1 was expressed and purified as previously described.<sup>[5]</sup>

#### 3.2 Crystallization, data collection and structure determination

Freshly purified sALDH1A1 was concentrated to 9-10 mg/mL in HEPES (10 mM, pH 7), 1 mM NAD and 5 mM DTT (protein crystallization buffer 1) and crystallized using 100 mM Bis-Tris, pH 6.0, 4.5–7% PEG5000 and 150-225 mM MgCl<sub>2</sub> at 4°C. Crystals typically appeared after 10-14 days.

For co-crystallization of sheep and human ALDH1A1 with seco drug **2**, 570.7  $\mu$ L seco drug **2** (154.5  $\mu$ M) in Tris-HCl 50 mM, pH 8.5 were pre-incubated for 2 h at 30°C and 300 rpm. Afterwards the pH was adjusted to 7.4, 864  $\mu$ L ALDH1A1 (28  $\mu$ M) were added and the reaction mixture of ALDH1A1 and activated seco drug **2** was incubated at 4 °C for 3 d. Binding of seco drug **2** was confirmed by intact protein mass spectrometric analysis using an LTQ-FT-ICR. Crystallization of sALDH1A1 in complex with seco drug **2** was accomplished as described above for the apo enzyme. hALDH1A1 in complex with seco drug **2** was concentrated to 4 mg/mL and crystallized in 100 mM sodium Bis-Tris, pH 6.5, 8–12% PEG3350, 200 mM NaCl, and 5–10 mM YbCl<sub>3</sub> at 25 °C overnight.

Diffraction data were collected at the synchrotron beam lines PXI (Swiss Light Source, Villigen, Switzerland) and ID29 (European Synchrotron Radiation Facility, Grenoble, France). Crystals of sheep and human ALDH1A1 in complex with or without seco drug **2** belonged to different space groups and the data were processed with XDS<sup>[6]</sup> to 1.7–2.1 Å spacing, with a resolution cut-off of CC 1/2 of 50%.<sup>[7]</sup> For data processing statistics see Supplementary Table 1. The structures were solved by molecular replacement or different fourier methods using the ALDH1A1 coordinates (PDB code: 1BXS, 4WP7) in PHASER<sup>[8]</sup> and REFMAC5<sup>[9]</sup>, respectively. In order to reduce model bias prior to molecular replacement/rigid body refinement the coordinates for NAD were removed and the temperature factors were reset, followed by simulated annealing in PHENIX.<sup>[10]</sup> Clear peaks for seco drug **2** and Cys302 were visible in the simulated-annealing omit Fo-DFc electron density map. Rounds of model building and refinement were carried out in COOT<sup>[11]</sup> and REFMAC5.<sup>[9]</sup> Restraints for seco drug **2** were created using JLigand<sup>[12]</sup> and the number of TLS groups was determined using the TLSMD server.<sup>[13]</sup> Ytterbium ions in the structure of

the hALDH1A1 were placed according to the anomalous difference fourier map calculated with FFT.<sup>[8b]</sup> Final coordinates were submitted to the PDB\_REDO server.<sup>[14]</sup> Diffraction data and refinement statistics are summarized in Supplementary Table 1. Structural super positions were done with SSM<sup>[15]</sup> and all structural figures were prepared with PyMol (Delano Scientific, San Carlos, CA). Atomic coordinates were submitted to the Protein Data Bank (<http://www.ebi.ac.uk/pdbe/>) with the PDB codes: sALDH1A1 (APO) = 5ABM, sALDH1A1 (seco drug **2**, P2<sub>1</sub>2<sub>1</sub>2) = 5AC0, sALDH1A1 (seco drug **2**, C2<sub>1</sub>) = 5AC1 and hALDH1A1 (seco drug **2**) = 5AC2.

### 3.3 Aldehyde dehydrogenase activity assay

The activity of sALDH1A1, hALDH1A, ALDH1A1-C302A, ALDH1A1-C302S, ALDH1A2 (Origene, ORIGTP323250), ALDH1A3 (life technologies, 11636-H07E-50) and ALDH2 (Abnova, P3479) were carried out as outlined below. To ensure the formation of active duocarmycin drug, initially 500  $\mu$ M seco drug **2** was pre-incubated in Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C. This was followed by adding 2.5  $\mu$ l activated seco drug **2** to 2.5  $\mu$ l ALDH1A1 (10  $\mu$ M), resulting in final concentration of 1  $\mu$ M enzyme and 50  $\mu$ M seco drug **2** in a total volume of 50  $\mu$ L at pH 7.4, and was further incubated for 5 h at 30 °C. Subsequently, 50  $\mu$ L substrate mixture containing 50 mM Tris-HCl, 100 mM KCl, 5 mM  $\beta$ -mercaptoethanol, 1 mM  $\beta$ -NAD<sup>+</sup> and 10 mM propanal were added to initiate the enzymatic reaction. The product formation of NADH was monitored by measuring the absorption increase at  $\lambda$  = 340 nm at 37 °C in 96-well-plates with an Infinite 200 PRO NanoQuant microplate reader.

### 3.4 Determination of $K_i$ and $k_{inact}$ of ALDH1A1

To evaluate the influence of residues lining the binding site of hALDH1A1 the kinetic parameters  $K_i$  and  $k_{inact}$  of wild type and mutant hALDH1A1 (Cys302Ala, Cys302Ser, Phe171Ala, Trp178Ala, Tyr297Ala) were determined by the method of Kitz and Wilson.<sup>[3, 16]</sup> Initially, in order to ensure the formation of active duocarmycin drug, seco drug **2** was pre-incubated in Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C at different concentrations. Afterwards, 2.5  $\mu$ L ALDH1A1 (8.2  $\mu$ M) was added to give an enzyme concentration of 410 nM in a total volume of 50  $\mu$ L (pH 7.4). The reaction mixture was incubated and samples were taken at different time points. Subsequently, 50  $\mu$ L substrate mixture containing 50 mM Tris-HCl, 100 mM KCl, 5 mM  $\beta$ -mercaptoethanol, 1 mM  $\beta$ -NAD<sup>+</sup> and 10 mM propanal were added to the samples in order to initiate the enzymatic reaction. The product formation of NADH was monitored by measuring the absorption increase at  $\lambda = 340$  nm at 37 °C in 96-well-plates with an Infinite 200 PRO NanoQuant microplate reader. All measurements were carried out in technical triplicates and biological duplicates. To determine the kinetic parameters the natural log of the remaining enzyme activity after a certain pre-incubation time was plotted against the pre-incubation time for several inhibitor concentrations [I]. The observed rate of inactivation  $k_{obs}$  is derived from the negative slopes of the linear fit. A further double reciprocal plot of  $k_{obs}$  against [I] leads to  $K_i$  and  $k_{inact}$  whereas the values are derived from the slope and the intercept of the linear fit.

### 3.5 Intact protein mass spectrometry

Intact protein measurements were performed on a Dionex Ultimate 3000 HPLC system coupled to Thermo LTQ-FT Ultra mass spectrometer. Samples were on-line desalted using a Massprep desalting cartridge (Waters) and further analyzed in

electrospray ionization (ESI) positive mode (spray voltage 4.0 kV, tube lens 110 V, capillary voltage 48 V, sheath gas 60 arb, aux gas 10 arb) with a resolution of 200,000 and a mass range of 600-2000 m/z. Collected data were deconvoluted using ProMass™ for Xcalibur 2.8 software (Thermo Scientific).

### **3.6 Time dependent intact protein alkylation experiment of human ALDH1A1**

Time dependent alkylation of seco drug **2** with hALDH1A1 was performed via intact protein mass spectrometry measurements as described above. Thereby 160 µM seco drug **2** were pre-incubated in Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C. Afterwards 13 µl of human ALDH1A1 (16 µM) and 13 µl activated seco drug **2** (160 µM) were added in a total volume of 130 µL PBS (final pH 7.4) and incubated at 30 °C. 2 µL of the reaction mixture were injected every 27 min into the LC-MS system and analyzed with ProMass™ for Xcalibur 2.8 software (Thermo Scientific).

Alkylation of duocarmycin SA with human ALDH1A1 was performed as described above for seco drug **2**. Therefore 400 µM duocarmycin SA were pre-incubated in Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C. Afterwards, 13 µl of human ALDH1A1 (16 µM) and 13 µl activated duocarmycin SA (400 µM) were added in a total volume of 130 µL PBS (final pH 7.4) and incubated for 3 h at 30 °C. 2 µL of the reaction mixture were injected into the LC-MS system and analyzed with ProMass™ for Xcalibur 2.8 software (Thermo Scientific).

### **3.7 Gel-based fluorescent labeling experiment**

160 µM seco drug **2** probe in Tris-HCl 50 mM, pH 8.5 were pre-incubated for 2 h at 30 °C. Subsequently, 13 µl of activated seco drug **2**, 10 µl of human ALDH1A1

20



(1.6  $\mu$ M) and probe (16  $\mu$ M) were mixed in a total volume of 130  $\mu$ L at pH 7.4 and incubated for 2.5 h at 30 °C. Afterwards, the reaction mixtures were clicked using rhodamine azide, Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) and applied on a 10% polyacrylamide gel as described by *Wirth et al.*<sup>[17]</sup>

### 3.8 Bottom-up proteomics for binding site identification

To discover the binding site of seco drug **2** in mutant enzymes ALDH1A1 (Cys302Ala and Cys302Ser), the recombinant enzymes were labeled, tryptic digested and the peptide fragments analyzed by ESI tandem mass spectrometry. 26  $\mu$ L of seco drug **2** (577  $\mu$ M) in Tris-HCl 50 mM, pH 8.5 were activated for 2 h at 30 °C and 300 rpm. 86  $\mu$ L ALDH1A1 (20  $\mu$ M) and 1.2 mL PBS were added to give a final concentration of 1.3  $\mu$ M ALDH1A1 and 11.4  $\mu$ M seco drug **2** (final pH 7.4). The reaction mixture was incubated at 30 °C and 300 rpm. Binding was controlled by intact protein measurement (>90%). In order to remove unreacted probe the mixture was washed twice with PBS with a centrifugal filter device (10 kDa molecular weight cut-off) and adjusted to a protein concentration of 1 mg/mL. For the reduction of disulfide bonds 1.3  $\mu$ L DTT (1 mM) were added and incubated for 15 min at 37 °C and 800 rpm. After reduction, free thiols were alkylated by the addition of 7.1  $\mu$ L iodoacetamide (30 mM) and incubated for 30 min at 22 °C and 800 rpm in the dark. Then, 2  $\mu$ L chymotrypsin (0.01 nmol) and 1.4  $\mu$ L CaCl<sub>2</sub> (10 mM) solution were incubated for 15 h at 37 °C and 800 rpm. Tryptic digest was stopped with 7  $\mu$ L formic acid (5%). Afterwards, the protein solutions were desalted with stage tips (Empore disk-C18, 47 mm, Agilent Technologies) and filtered via modified nylon 0.45  $\mu$ M low protein binding centrifugal filter (VWR). The digested peptides were analyzed on a UltiMate 3000 nano HPLC system (Dionex, Sunnyvale, California, USA) coupled to a Orbitrap Fusion™

Tribrid™ mass spectrometer (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA). Samples were loaded on a Acclaim C18 PepMap100 75 µm ID x 2 cm trap and transferred to a Acclaim C18 PepMap RSLC, 75 µm ID x 15 cm separation column (0.1% FA, 5% DMSO, gradient 10 min 3% ACN, 120 min from 3% to 25% ACN, 5 min to 40% ACN, 0.1 min to 90% ACN and 4.9 min hold at 90% ACN, 0.1 min to 3% ACN and 9.9 min 3% ACN). The mass spectrometer was operated in data dependent top speed mode selecting the most intense precursors with a minimal threshold of 5E3. Precursors were measured in the orbitrap at a resolution of 120,000 and an ion target of 4E5 (max inj. time of 50 ms) in a scan range from 300 to 1700 m/z. Monoisotopic precursor selection was enabled. Charge states from 1 to 7 were triggered. Dynamic exclusion duration was set to 60 s with a mass tolerance of 10 ppm. Precursors were isolated in the quadrupole (isol. window 1.6 m/z) and fragmentation was performed using higher-energy collisional dissociation (HCD). Resulting fragments were measured in the ion trap using a rapid scan rate (AGC target: 1E4 and max inj. time 40 ms). Assignment of the measured peptides to proteins was done with MaxQuant 1.4.0.8 software. Default settings were used except for the following: Precursor mass tolerance: 4.5 ppm; Fragment mass tolerance: 0.5 Da. Protein database: hALDH1A1 Uniprot, enzyme: chymotrypsin (specificity Tyr, Phe and Trp), variable modifications: oxidation (M) +15.995 Da, carbamidomethyl (C) +57.021 Da; SDB6 (C) +291.126 Da.

### 3.9 MTT cytotoxicity assay

Varying concentrations of seco drug **2** were pre-incubated for 2 h in 100 µL medium (without FBS) with a final DMSO concentration of 1%. A549 cells were grown in 96-well-plates at a concentration of 7000 cells per cavity. After removing the growth

22

medium the pre-incubated medium containing the reactive drug was added and the cells incubated for 24 h at 37 °C and 5% CO<sub>2</sub>. After 24 h of exposure, 20 µL (5 mg/ml) filtered 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) stock solution in PBS were added and the medium thoroughly mixed by gentle pipetting. The cells were incubated at 37 °C and 5% CO<sub>2</sub> for 1 h to allow the MTT to be metabolized and the reaction was controlled under the microscope. Subsequently, the medium and cell debris were removed and the produced formazan resuspended in 200 µL DMSO by placing the well-plate on a shaking table for 2 min and 650 rpm. The optical density was read out at  $\lambda = 570$  nm and  $\lambda = 630$  nm with an Infinite 200 PRO NanoQuant microplate reader and the background was subtracted at  $\lambda = 630$  nm. Cells incubated with 1% DMSO served as positive control.

For calculation of IC<sub>50</sub> values, residual viabilities for the respective compound concentration were fitted to

$$V = \frac{100}{1 + 10^{(\log(IC_{50}) - \log(c)) \cdot N}}$$

with

$V$ : viability [%]

$c$ : Inhibitor concentration [M]

$N$ : Hill slope

using Graphpad Prism 6.0.

Experiments were carried out in triplicates and two independent experiments. IC<sub>50</sub> values are given as mean values and 95% confidence interval.



### 3.10 DNA-alkylation experiment

For the DNA alkylation experiment single stranded DNA listed in SI Table 5 were dissolved in 150 mM NaCl and 10 mM Tris-HCl pH 7.5 to a final concentration of 1 mM. For ds DNA formation ss-9 and ss-9rc as well as ss-10 and ss-10rc were mixed in a 1:1 ratio and hybridized via a temperature gradient (95 °C to 4 °C) within 2 h. For the DNA-alkylation experiment seco drug **1**, seco drug **2** and duocarmycin SA were pre-incubated for activation in 5 µL Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C with a final concentration of 5 mM. Afterwards, 0.5 µl of the activated duocarmycins (5 mM) were mixed with 7.5 µL water and 2 µL hybridized DNA (50 µM) at a final pH of 7.4 and incubated for 24 h at 25 °C. The samples were analyzed *via* MALDI using a Bruker autoflex II unit with an MTP AnchorChip var/384 target. Prior to the measurements, the samples were desalted for 20 min using MF-Millipore membrane filters (0.025 µM). As MALDI matrix a mixture of 3-hydroxypicolinic acid (50 mg), ammonium hydrogencitrate (10 mg) in 500 µL ddH<sub>2</sub>O and 500 µL acetonitrile was applied.

### 3.11 Putative design criteria for selective ALDH1A1 inhibition or DNA-alkylation

By increasing the size of the aromatic ring system bearing the electrophile, the binding to ALDH1A1 is prevented and results in compounds solely targeting dsDNA. For instance attachment of an additional 5 or 6 membered ring to the alkylating benzene moiety would most likely result in occlusion from the ALDH1A1 active site due to steric hindrance with Trp178 and Met175. Based on the solution structure of duocarmycin with dsDNA<sup>[18]</sup> (PDB code 1DSA) such a modification is unlikely to disturb the interaction with DNA.

In contrast some small alterations of the hydrophobic ring system on seco drug 2, such as addition of methyl groups could be tolerated by the ALDH1A1 active site, and might fine tune the specific interaction.

#### 4. Molecular dynamics simulations

XLEAP (AmberTool)<sup>[19]</sup> has been used to add hydrogen atoms to the X-ray structure and to solvate the solutes (ALDH1A1 from sheep as monomer with ligands) in a box of explicit TIP3P water<sup>[20]</sup> with a buffer of 10 Å around the solute. ANTECHAMBER<sup>[21]</sup> was used to parameterize seco drug 2 and NAD. For force field minimizations the NAMD engine was used<sup>[22]</sup> with Amber10 force field parameters.<sup>[19]</sup>

Periodic boundary conditions and particle mesh Ewald summation (PME) with a cutoff value of 12 Å were employed. The system was energy minimized (NVT ensemble) using the conjugate gradient algorithm within 20000 steps. A positional constraint of 1 kcal/mol/Å<sup>2</sup> on non-water atoms was applied.

The system was heated up to 300 K within 30 ps. In the subsequent equilibration step, the system was equilibrated for 230 ps. In the second equilibration step we switched to the NPT ensemble employing the Langevin piston Nosé-Hoover method.<sup>[23]</sup> At this stage the constraints on non-water atoms are reduced step by step down to zero (0.1 kcal/mol/Å<sup>2</sup> increments for every 20 ps). Production runs were performed for 10-100 ns using the SHAKE algorithm<sup>[24]</sup> with timesteps of 2 fs.

For proper statistics and to obtain a statistically significant analysis, at least 8 simulations have been performed for each system. The average run time over all systems was 600 ns. The simulation time for each system is listed in SI Table 4.

## 5. References

- [1] L. Tietze, B. Krewer, J. M. Von Hof, H. Frauendorf, I. Schuberth, *Toxins* **2009**, *1*, 134.
- [2] a) M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, *Bioinformatics* **2007**, *23*, 2947-2948; b) X. Robert, P. Gouet, *Nucleic Acids Res.* **2014**, *42*, W320-W324.
- [3] R. Kitz, I. B. Wilson, *J. Biol. Chem.* **1962**, *237*, 3245-3249.
- [4] C. A. Morgan, T. D. Hurley, *J. Med. Chem.* **2015**, *58*, 1964-1975.
- [5] T. Wirth, G. F. Pestel, V. Ganai, T. Kirmeier, I. Schuberth, T. Rein, P. L. F. Tietze, P. S. A. Sieber, *Angew. Chem. Int. Ed.* **2013**, *52*, 6921-6925.
- [6] W. Kabsch, *Acta Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 133-144.
- [7] a) P. R. Evans, G. N. Murshudov, *Acta Crystallogr. D: Biol. Crystallogr.* **2013**, *69*, 1204-1214; b) P. Evans, *Science* **2012**, *336*, 986-987; c) K. Diederichs, P. A. Karplus, *Acta Crystallogr. D: Biol. Crystallogr.* **2013**, *69*, 1215-1222; d) P. A. Karplus, K. Diederichs, *Science (New York, N.Y.)* **2012**, *336*, 1030-1033.
- [8] a) A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658-674; b) M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr. D: Biol. Crystallogr.* **2011**, *67*, 235-242.
- [9] G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallographica Section D: Biological Crystallography* **2011**, *67*, 355-367.
- [10] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 213-221.
- [11] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 486-501.
- [12] A. A. Lebedev, P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin, G. N. Murshudov, *Acta Crystallogr. D: Biol. Crystallogr.* **2012**, *68*, 431-440.
- [13] a) J. Painter, E. A. Merritt, *J. Appl. Crystallogr.* **2006**, *39*, 109-111; b) M. D. Winn, G. N. Murshudov, M. Z. Papiz, in *Methods in Enzymology, Vol. Volume 374* (Eds.: Charles W. Carter, Jr., M. S. Robert), Academic Press, **2003**, pp. 300-321.
- [14] R. P. Joosten, F. Long, G. N. Murshudov, A. Perrakis, *IUCrJ* **2014**, *1*, 213-220.
- [15] E. Krissinel, K. Henrick, *Acta Crystallogr.* **2004**, *60*, 2256-2268.
- [16] F. Ghanbari, K. Rowland-Yeo, J. C. Bloomer, S. E. Clarke, M. S. Lennard, G. T. Tucker, A. Rostami-Hodjegan, *Curr. Drug Metab.* **2006**, *7*, 315-334.
- [17] T. Wirth, G. F. Pestel, V. Ganai, T. Kirmeier, I. Schuberth, T. Rein, L. F. Tietze, S. A. Sieber, *Angew. Chem. Int. Ed.* **2013**, *52*, 6921-6925.
- [18] C. H. Lin, D. J. Patel, *J. Mol. Biol.* **1995**, *248*, 162-179.
- [19] D. Case, T. A. Darden, T. E. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, M. Crowley, R. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V.

- Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. Kollman, *Amber 11*.
- [20] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *Chem. Phys.* **1983**, 79, 926-935.
- [21] a) J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, 25, 1157-1174; b) J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, 25, 247-260.
- [22] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, 26, 1781-1802.
- [23] a) S. Nosé, *J. Chem. Phys.* **1984**, 81, 511-519; b) W. G. Hoover, *Physical Review A* **1985**, 31, 1695-1697.
- [24] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, 23, 327-341.

- V.5 Paper [5]: "Kinetic and theoretical studies of  $\beta$ -lactone reactivity - a quantitative scale for biological application", E. N. Wiedemann, F. A. Mandl, I. D. Blank, C. Ochsenfeld, A. R. Ofial, S. A. Sieber, *ChemPlusChem*, 80, 1673 (2015).**





# Kinetic and Theoretical Studies of Beta-Lactone Reactivity—A Quantitative Scale for Biological Application

Elija N. Wiedemann,<sup>[a]</sup> Franziska A. Mandl,<sup>[b]</sup> Iris D. Blank,<sup>[a]</sup> Christian Ochsenfeld,<sup>\*,[a]</sup> Armin R. Ofial,<sup>\*,[a]</sup> and Stephan A. Sieber<sup>\*,[b]</sup>

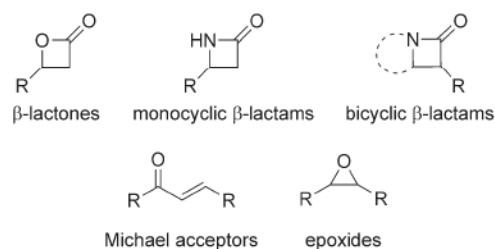
Natural products comprise a rich source for bioactive molecules with medicinal relevance. Many of these contain electrophilic scaffolds that bind conserved enzyme active sites covalently. Prominent examples include beta-lactams and beta-lactones, which specifically acylate serine residues in diverse peptidases. Although these scaffolds appear similar, their bioactivities and corresponding protein targets vary. To quantify and dissect these differences in bioactivities, the kinetics of the reactions of beta-butyrolactone with a set of reference amines in buffered aqueous solution at 37 °C have been analyzed. Different product ratios of C1 versus C3 attack on the beta-butyrolactone have been observed, depending on the aliphatic or ar-

omatic nature of the standard amine used. Quantum mechanics/molecular mechanics (QM/MM) calculations reveal that a H<sub>3</sub>O<sup>+</sup> molecule has a crucial role in stabilizing C3 attack by aniline, through coordination of the lactone ring oxygen. In agreement with their weak proteome reactivity, monocyclic beta-lactams did not react with the set of standard nucleophiles studied herein. Bicyclic beta-lactams, however, exhibited a lower activation barrier, and thus, reacted with standard nucleophiles. This study represents a starting point for semiquantitative reactivity scales for natural products, which, in analogy to chemical reactivity scales, will provide predictions for electrophilic modifications in biological systems.

## Introduction

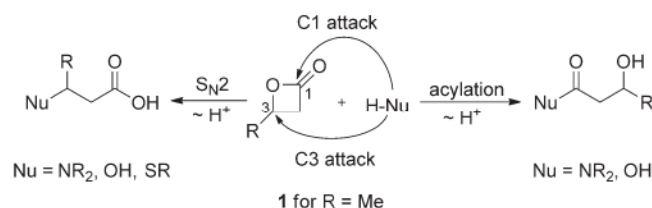
Natural products frequently exhibit diverse electrophilic moieties designed to trap nucleophilic amino acid residues and thereby inhibit cellular proteins.<sup>[1]</sup> Evolution has fine-tuned these electrophiles to address a variety of cellular nucleophiles, for example, in the active site of enzymes. These tailored scaffolds include beta-lactams, beta-lactones, epoxides, and Michael acceptors, which vary in their reactivity and selectivity towards amino-, thio-, and hydroxyl-substituted amino acids (Scheme 1).

The beta-lactams are among the most potent antibiotics and they exert their cellular activity by reacting with the active-site serine in penicillin-binding proteins (PBPs), thereby inhibiting cell wall biosynthesis, and thus, bacterial viability.<sup>[2]</sup> Importantly, monocyclic beta-lactams lack pronounced protein reactivity.<sup>[3]</sup> However, scaffolds in which the lactam is activated by, for example, sulfonylation (aztreonam) or as a bicycle (peni-



Scheme 1. Common electrophilic motifs in natural products.

cillins), exhibit desirable reactivity and bioactivity.<sup>[4]</sup> The beta-lactones, on the other hand, do not require activation to react with serine and cysteine residues, which suggests that monocyclic lactones and lactams show significantly different electrophilicities under physiological conditions.<sup>[5]</sup> Previous ring-opening studies with 3-alkylated beta-lactones revealed S-nucleophiles to require activation by deprotonation for an efficient C3 attack, a mixed S<sub>N</sub>2 reaction (attack at C3) and acylation (attack at C1) with amines, and pH-dependent C1 or C3 attack in hydrolysis reactions (Scheme 2).<sup>[6,7]</sup>



Scheme 2. Possible beta-lactone ring-opening pathways of different nucleophiles.

[a] E. N. Wiedemann,<sup>+</sup> I. D. Blank,<sup>+</sup> Prof. Dr. C. Ochsenfeld, Dr. A. R. Ofial  
Department Chemie  
Ludwig-Maximilians-Universität München  
Butenandtstrasse 5–13, 81377 München (Germany)  
E-mail: christian.ochsenfeld@cup.uni-muenchen.de  
ofial@lmu.de

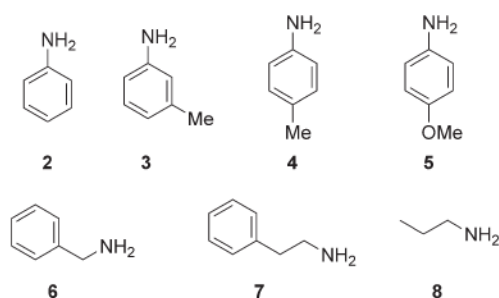
[b] F. A. Mandl,<sup>+</sup> Prof. Dr. S. A. Sieber  
Center for Integrated Protein Science CIPSM  
Institute of Advanced Studies IAS, Department Chemie  
Lehrstuhl für Organische Chemie II, Technische Universität München  
Lichtenbergstrasse 4, 85748 Garching (Germany)  
E-mail: stephan.sieber@tum.de

[<sup>+</sup>] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cplu.201500246>.

Owing to a lack of reliable reactivity and selectivity profiles under physiological conditions, the identification of nucleophiles that distinguish between C1 and C3 attack, as well as an explanation for the selectivity towards these two sites, has not been possible. To mimic biological systems accurately, one must consider aqueous media as well as a series of nucleophiles that imitate the reactivity of active sites. A step in this direction was the evaluation of the alkylating potential of lactones with respect to their reaction rates with 4-(*p*-nitrobenzyl)pyridine in mixtures of water/dioxane.<sup>[7]</sup> A crucial parameter for the determination of natural-product electrophilicities is the selection of suitable nucleophiles that can be described by suitable kinetic equations. Previously, the reactivities of structurally diverse carbocations and neutral electrophiles were characterized with a set of reference nucleophiles, and the corresponding second-order rate constants were evaluated according to the linear free energy relationship  $\log k(20^\circ\text{C}) = s_N(N + E)$ .<sup>[8]</sup> The electrophile-independent parameter  $N$  is a measure of the nucleophilicity of a dissolved reactant, whereas  $s_N$  provides information about its sensitivity towards changes in the electrophilicity of reaction partners. The electrophilicity,  $E$ , is a nucleophile-independent reactivity parameter. Because the same value of  $E$  is assigned to the reactivity of reference electrophiles (benzhydrylium ions and quinone methides) in different solvents, all solvent effects on reactivity are considered in the nucleophilicity parameters  $N$  and  $s_N$ .

The kinetics of electrophilic compounds under physiological conditions has previously been studied by different groups, for example, to derive the carcinogenic properties of lactones or lactams from their alkylation potential.<sup>[7,9]</sup> To determine the reactivities of lactones and lactams towards nucleophiles with known  $N$  and  $s_N$  parameters,<sup>[8d]</sup> we started by investigating the reactions of beta-butyrolactone (**1**) with a set of standard amino-substituted nucleophiles (Scheme 3) and determined its electrophilic reactivity, as well as the regioselectivity of nucleophilic attack with experimental and theoretical methods. In agreement with their minimal biological effects, monocyclic beta-lactams did not exhibit any reactivity with nucleophiles **6–8**, whereas reactions occurred when the beta-lactam was part of a strained bicyclic ring system. Because these reactions were performed under aqueous conditions, this study provides a basis for a rational approach to fine-tuning electrophilic scaffolds as selective inhibitors in biological applications.

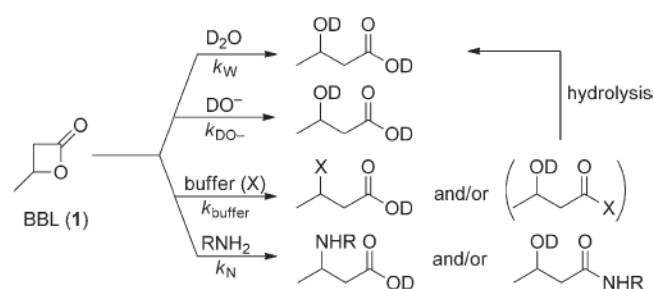


Scheme 3. Standard amino-substituted nucleophiles used herein: aniline (**2**), *m*-toluidine (**3**), *p*-toluidine (**4**), *p*-anisidine (**5**), benzylamine (**6**), 2-phenylethylamine (**7**), *n*-propylamine (**8**).

## Results and Discussion

### General considerations

To obtain a comprehensive picture of the reactivity of **1** under physiological conditions, we used a series of primary amines (Scheme 3) in aqueous solution ( $\text{D}_2\text{O}$ ) at  $37^\circ\text{C}$  and constant pH. In these reaction mixtures, four different nucleophiles are capable of reacting with **1**: water, hydroxide ions, the conjugate bases  $X$  of the buffer system, and the selected nucleophile (Scheme 4).



Scheme 4. Reactions of **1** with amines in buffered aqueous solutions.

The rate law for the decay of the concentration of **1** [Eq. (1)] has to consider the contributions of all four reaction channels.

$$-d[\mathbf{1}]/dt = (k_W + k_{\text{DO}^-}[\text{DO}^-] + k_{\text{buffer}}[\text{buffer}] + k_N[\text{amine}])[\mathbf{1}] \quad (1)$$

Previous kinetic studies on the pH dependency of the hydrolysis of **1** have shown that the uncatalyzed reaction of **1** with water follows a first-order rate law with a first-order rate constant,  $k_W$ . Attack by hydroxide ions ( $\text{OH}^-$ ) is only relevant at  $\text{pH} > 9$ , whereas acid catalysis accelerates hydrolysis only in the region of  $\text{pH} < 1$ .<sup>[10]</sup> Blackburn and Dodds have shown that, in buffered solutions, the reaction of pyridines and imidazoles with  $\beta$ -propiolactone could be described by Equation (2a).<sup>[11]</sup> Significant increases in the rate of the hydrolysis of **1** were only observed in the presence of highly concentrated buffers ( $> 1.5 \text{ mol L}^{-1}$ ).<sup>[10b]</sup> Based on an extensive study by Gresham et al. on the ring opening of  $\beta$ -propiolactone by acetate or anions of inorganic acids,<sup>[12]</sup> we expected that buffers not only participated in the stabilization of the pH, but that their conjugate bases were also potential nucleophiles capable of attacking **1** ( $k_{\text{buffer}}[\text{buffer}]$ ).<sup>[13]</sup> Because we found that the conjugate bases of the buffers indeed gave covalent adducts with **1** (see below), we extended Equation (2a) to Equation (2b).

$$k = k_W + k_{\text{DO}^-}[\text{DO}^-] + k_N[\text{amine}] \quad (2a)$$

$$k_2 = k_W + k_{\text{DO}^-}[\text{DO}^-] + k_{\text{buffer}}[\text{buffer}] + k_N[\text{amine}] \quad (2b)$$

$$k_1 = k_W + k_{\text{DO}^-}[\text{DO}^-] + k_{\text{buffer}}[\text{buffer}] \quad (3)$$

Owing to the high initial concentration of water, the constant value of  $[\text{DO}^-]$  in buffered solution and only partial con-



sumption of buffer and amine, the concentrations of all four relevant nucleophiles do not change significantly during the kinetic experiments. The consumption of **1** is, therefore, expected to follow an exponential function,  $[1]_t = [1]_0 e^{-kt}$ , in which  $k$  corresponds to  $k_1$  for the hydrolysis of **1** in buffered aqueous solutions [Eq. (3)], or to  $k_2$  for the reaction of **1** in buffered aqueous solutions of amines [Eq. (2b)].

### Reactivity of **1** in aqueous buffer solutions

In the first step, we determined the rate of the consumption of **1** in a 9:1 mixture of  $D_2O/[D_6]DMSO$  in the absence of standard amine nucleophiles by using time-resolved  $^1H$  NMR spectroscopy. The consumption of **1** through hydrolysis and nucleophilic attack of different buffers (pH range from 4.75 to 8.10) was investigated (Table 1). In each case, the consumption of **1** followed a monoexponential decay and produced a mixture of 3-HBA and the buffer adduct (see the Supporting Information). Least-squares fitting with the function  $[1]_t = [1]_0 e^{-k_1 t}$  yielded the first-order rate constants,  $k_1$ , as defined by Equation (3), which reflected the reactivity of **1** towards the solvent at constant pH for a 0.10 M buffer system.

The attack of **1** by buffer components was recorded by  $^1H$  NMR spectroscopy throughout the kinetic experiments (right column of Table 1) and independently confirmed by LC-HRMS analysis (see the Supporting Information). The presence of buffer changed the rate constant of the consumption of **1** only within a factor of  $< 1.25$  (Table 1, entries 2–5), compared with the non-buffered hydrolysis of **1** (Table 1, entry 1), which agrees with minimal formation of adducts of **1**. As illustrated in Table 1, entries 1–5, compound **1** reacted with all buffered aqueous solutions at comparable rates, so we decided to proceed with aqueous solutions of TEA for further kinetic investigations of reactions of **1** with primary amines. This is beneficial because the tertiary amine TEA provides a constant pH of 7.75, which is similar to physiological conditions in living cells (Table 1, entry 4).

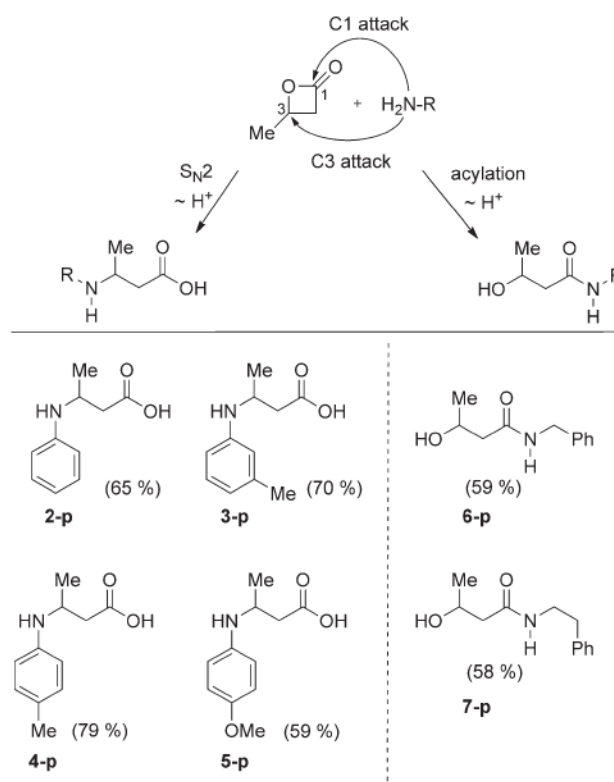
**Table 1.** Rate constants of the consumption of **1** in buffered and non-buffered mixtures of  $D_2O/[D_6]DMSO$  9/1 (v/v) at 37 °C ( $[1]_0 = 25$  mM).

Entry	Buffer	pH	$k_1$ [ $s^{-1}$ ]	$t_{1/2}$ [h]	3-HBA/buffer adduct
1	without buffer	–	$4.73 \times 10^{-5[a]}$	4.1	100:0
2	acetate	4.75	$4.14 \times 10^{-5}$	4.7	96:4
3	$Na_2HPO_4/NaH_2PO_4$	7.00	$5.85 \times 10^{-5}$	3.3	86:14
4	TEA	7.75	$5.47 \times 10^{-5}$	3.5	83:17 → 88:12 <sup>[b]</sup>
5	TRIS	8.10	$5.79 \times 10^{-5}$	3.3	79:21

[a] Comparison with data from Ref. [14]: with  $\Delta H^\ddagger = 87.6$  kJ mol $^{-1}$  and  $\Delta S^\ddagger = -77.9$  J mol $^{-1}$  K $^{-1}$ , the second-order rate constant for the hydrolysis of **1** is calculated to be  $k(37^\circ C) = 1.1 \times 10^{-6} M^{-1} s^{-1}$ . Multiplying  $k(37^\circ C)$  with  $[D_2O]$  for the mixture of  $D_2O/[D_6]DMSO$  9/1 in our experiment gives a first-order rate constant for the hydrolysis of **1** of  $k_w(37^\circ C) = 5.5 \times 10^{-5} s^{-1}$ , which is in sufficient agreement with the rate constant determined under our conditions. [b] The ratio of 3-HBA/buffer adduct increased during the course of the reaction (14 h) probably because of subsequent hydrolysis of the quaternary ammonium salt formed by C3 attack of TEA on **1** (for details, see the Supporting Information, Section 3.1.5). TEA = triethanolamine, TRIS = tris(hydroxymethyl)amino-methane, 3-HBA = 3-hydroxybutyric acid.

### Reactivity of **1** towards amines in aqueous solutions

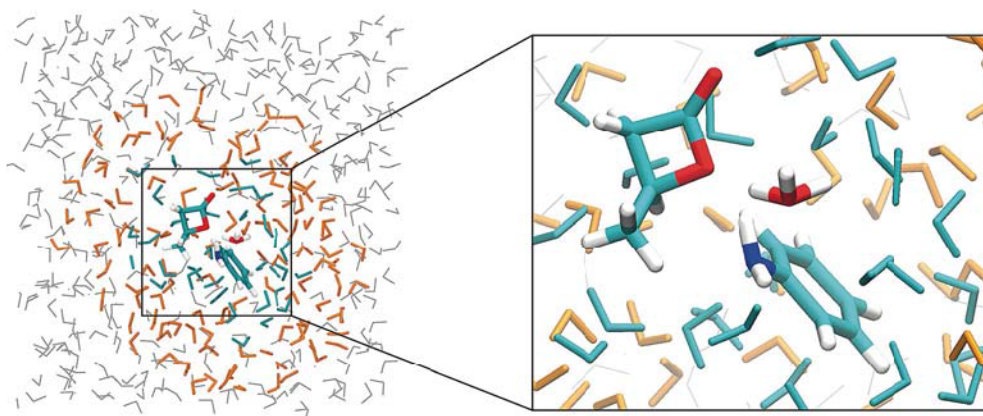
We commenced by studying the products of the reactions of **1** with a panel of nucleophilic amines (**2**, **3**, **4**, **5**, **6**, and **7**; Scheme 3) in aqueous solutions. Interestingly, we observed that the amine structure directly influenced the preferred site of attack at **1**. Aromatic amines opened the lactone ring through  $S_N2$  attack at the C3 position of **1** (O-alkyl fission), yielding the beta-amino acids, whereas primary aliphatic amines underwent acylation through reaction at the C1 site of **1** (O-acyl fission), forming 3-hydroxy amides (Scheme 5).



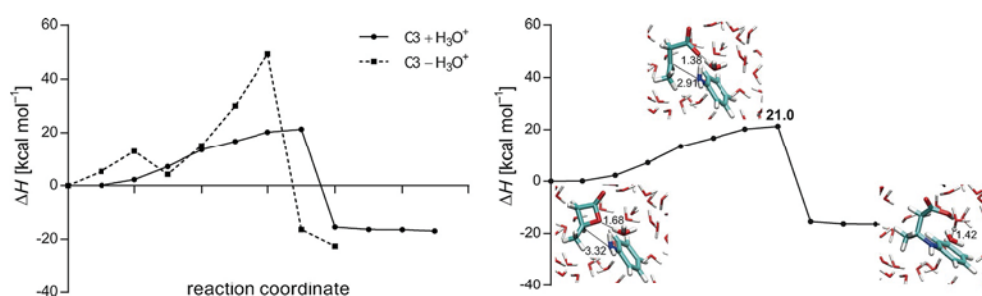
**Scheme 5.** Products (**2-p**–**7-p**) of the reactions of **1** with amines **2**–**7** in  $H_2O/CH_3CN$  9/1 (v/v) (yields refer to products isolated).

### QM/MM studies on the C1 versus C3 selectivity of amine attack on **1**

To explain the unexpected preference for either C1 or C3 attack of aliphatic and aromatic amines on **1**, we performed theoretical studies by using a combined quantum mechanics/molecular mechanics (QM/MM) approach with linear-scaling quantum-chemical methods (QM; see for example, Ref. [15] and references therein).<sup>[16]</sup> The QM/MM method has been widely used to describe complex reactions. By using this strategy, reaction mechanisms were calculated for both the adiabatic mapping approach and the nudged elastic band method (for computational details, see the Supporting Information). The composition of the calculated systems is illustrated for the reaction of **1** with **2** in Figure 1.



**Figure 1.** Illustration of the calculated systems obtained by using the QM/MM approach, as exemplified for the reaction of **1** with **2**. The reactants (depicted in atomic color) are part of the QM region (cyan; 156 atoms) in which chemical reactions can be described. The remaining part is described by MM, in which the region depicted in orange is optimized (including the QM region; overall 681 atoms) and the gray region is fixed (2235 atoms in total).



**Figure 2.** Calculated reaction profiles of C3 attack of **2** on **1** with (—) and without (----)  $\text{H}_3\text{O}^+$  (left) with structures out of the QM/MM simulation for the starting material, transition state, and product (right).

In this reaction,  $\text{H}_3\text{O}^+$  stabilizes the transition state of nucleophilic attack of **2** at C3 and thereby lowers the activation barrier by  $28 \text{ kcal mol}^{-1}$  to  $21 \text{ kcal mol}^{-1}$  (Figure 2). Specifically,  $\text{H}_3\text{O}^+$  approaches the lactone ring oxygen to a minimum distance of  $1.38 \text{ \AA}$ , forming a strong hydrogen bond (Figure 2, right), and thus, facilitating lactone ring opening. In a subsequent step, compound **2** forms a covalent bond with C3 and simultaneously protonates the lactone oxygen to form a stable product ( $-17 \text{ kcal mol}^{-1}$  relative to the initial state). In contrast, in the case of C1 attack,  $\text{H}_3\text{O}^+$  protonates the lactone ring oxygen instead of **2**, leaving **2** protonated when forming the product. This product, however, is not stable ( $+13 \text{ kcal mol}^{-1}$  relative to the initial state) and immediately regenerates the starting material (Figure S7 in the Supporting Information). Owing to the dependence of this reaction on  $\text{H}_2\text{O}/\text{H}_3\text{O}^+$ , the reaction did not occur in nonaqueous solvents, such as acetonitrile and DMSO (Figures S2 and S3 and Table S1 in the Supporting Information). The consumption of **1** at a rate constant of  $7.9 \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$  was observed, however, in  $\text{D}_2\text{O}/[\text{D}_6]\text{DMSO}$  9/1 (v/v) at  $37^\circ\text{C}$  (Table 2).

On the other hand, nucleophilic attack of the aliphatic amine **6** at C1 does not require  $\text{H}_3\text{O}^+$  stabilization. Calculations demonstrate that attack at C1 and protonation of the lactone

ring by the amine is favored (activation barrier of  $12 \text{ kcal mol}^{-1}$ ) over attack at C3 (activation barrier of  $37 \text{ kcal mol}^{-1}$ ; Table S1 in the Supporting Information). In the transition state, the C1 atom is in a tetrahedral configuration and the lactone ring remains intact. Subsequently, proton transfer from **6** to the ring oxygen induces lactone opening. If  $\text{H}_3\text{O}^+$  is considered in these simulations, the reaction also occurs via the same tetrahedral transition state, but  $\text{H}_3\text{O}^+$  protonates the ring oxygen instead. Finally, the positively charged nitrogen transfers a proton to a nearby water molecule (Figure 3).

The computational data is in agreement with experimental measurements of the rate constants for the reaction of **6** with **1** in the presence ( $5.8 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ , Table 2) and absence of water ( $2.5 \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$ ; Figure S4 in the Supporting Information). These values demonstrate that C1 attack indeed takes place, even in the absence of  $\text{H}_3\text{O}^+$  and  $\text{H}_2\text{O}$  (Figures S1 and S3 in the Supporting Information). Because almost all natural-

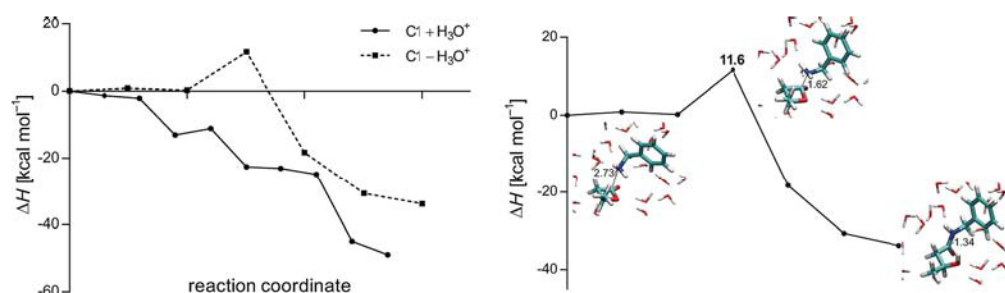
ly occurring beta-lactones exhibit bulky substituents at the C3 position, the biologically relevant ring-opening reaction in enzyme active sites is likely to be at the C1 position. This is supported by cocrystal structures of the thioesterase domain

**Table 2.** Second-order rate constants of the reactions of amines with **1** in  $0.10 \text{ M}$  TEA-buffered aqueous solutions ( $[\text{1}]_0 = 25 \text{ mM}$ , pH 7.75 in  $\text{D}_2\text{O}/[\text{D}_6]\text{DMSO}$  9/1 (v/v), at  $37^\circ\text{C}$ , determined by  $^1\text{H}$  NMR spectroscopy).

Nucleophile	$\text{p}K_{\text{a}}^{[\text{a}]}$	$N^{[\text{b}]}$	$s_N^{[\text{b}]}$	$k_{\text{N}} [\text{M}^{-1} \text{ s}^{-1}]$
<b>2</b>	9.4	12.99	0.73	$7.9 (\pm 0.5) \times 10^{-4}[\text{c}]$
<b>3</b>	9.13	n.d.	n.d.	$9.9 \times 10^{-4}$
<b>4</b>	8.92	13.00	0.79	$1.4 \times 10^{-3}$
<b>5</b>	8.64	14.28 <sup>[d,e]</sup>	0.68 <sup>[d,e]</sup>	$1.6 \times 10^{-3}$
<b>6</b>	4.64	13.44	0.55	$5.8 \times 10^{-3}[\text{f}]$
<b>7</b>	4.11	13.40 <sup>[d]</sup>	0.57 <sup>[d]</sup>	$1.1 \times 10^{-2}[\text{f}]$

[a] From Ref. [19]. [b] Reactivity parameters  $N$  and  $s_N$  of amines in water from Ref. [20]. [c] Average  $k_{\text{N}}$  of four kinetic runs at variable concentrations of **2**; the same value for  $k_{\text{N}}$  is obtained from the slope of the linear correlation of  $k_2$  versus  $[\text{2}]_0$  (see the Supporting Information, Section 4.1.5). [d] This study. [e] The previously published nucleophilicity parameters for **5** ( $N=16.53$ ,  $s_N=0.50$ ) from Ref. [20] were revised in this study (see the Supporting Information). [f] Rate constants of lower accuracy; an error of  $\pm 15\%$  is assumed.





**Figure 3.** Calculated reaction profiles of the C1 attack of **6** on **1** with (—) and without (----)  $\text{H}_3\text{O}^+$  (left) with structures from the QM/MM simulation for the starting material, transition state, and product (right). Accounting for  $\text{H}_3\text{O}^+$  in the calculated system leads to barrier-free attack at C1; however, minor product formation is also possible in the absence of  $\text{H}_3\text{O}^+$  (left).

of human fatty acid synthase inhibited by Orlistat,<sup>[17]</sup> as well as kinetic studies that show rapid hydrolysis of the ester intermediate of lactone-bound enzyme.<sup>[18]</sup>

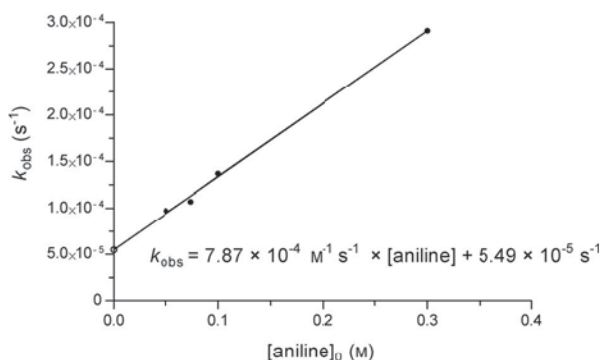
#### Kinetic studies of the reactions of **1** with amine nucleophiles in buffered media

Next, reactivity studies were performed by following the kinetics of the turnover of **1** with the same set of primary aliphatic amines and ring-substituted anilines in aqueous TEA-buffered solutions (Table 2). To fully solubilize these molecules, a small amount of DMSO (10 vol%) was necessary as a cosolvent. Again, the consumption of **1** followed a monoexponential function,  $[1]_t = [1]_0 e^{-k_2 t}$ , from which  $k_2$  was derived. By rearranging Equation (4),  $k_N$  can be calculated from Equation (5), in which  $k_1$  is the first-order rate constant of the background reactions of **1** in a mixture of TEA-buffered  $\text{D}_2\text{O}/[\text{D}_6]\text{DMSO}$  (9:1) (Table 1, entry 4) and  $[\text{amine}]_{\text{eff}}$  corresponds to the concentrations of free amines at pH 7.75.

$$k_2 = k_1 + k_N[\text{amine}]_{\text{eff}} \quad (4)$$

$$k_N = (k_2 - k_1)/[\text{amine}]_{\text{eff}} \quad (5)$$

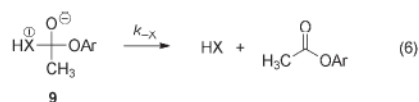
The linear dependence of the first-order rate constant  $k_2$  on the concentration of **2** (Figure 4) is in agreement with the rate



**Figure 4.** Linear dependence of the first-order rate constant  $k_2$  ( $=k_{\text{obs}}$ ) for the reaction of **1** with **2** on the concentration of **2** ( $k_{\text{obs}}$  for  $[2]_0 = 0$  corresponds to the hydrolysis of **1** in TEA-buffered solution; not considered for the depicted linear correlation).

laws shown in Equations (2a,b) and (4). The intercept on the ordinate matches, within experimental error, the first-order rate constant  $k_1$  [Eq. (3)] for the background reaction in TEA-buffered solution (Table 1, entry 4), which demonstrates that the background reaction is not significantly affected by the presence of the amine. The consistency of the determined  $k_N$  values is demonstrated by Hammett analysis of the data for anilines, which shows a good correlation of the rate constants  $k_N$  with the values of  $\sigma_p^-$  (and  $\sigma_m$ ) of the substituents at 2–5 ( $\log k_N = -1.21\sigma - 3.09$ ;  $R^2 = 0.9906$ ; see also Figure S41 in the Supporting Information).

Ritchie characterized the relative leaving-group abilities of nucleophiles in the tetrahedral intermediates **9** [Eq. (6)] by analyzing the rate constants for the reactions of several nucleophiles  $X$  (in water) with carbocations, N-acyl pyridinium ions, and aryl esters.<sup>[21]</sup>



The relative rates for the expulsion of  $X$  are remarkably different for water and **2** compared with those for primary aliphatic amines: relative  $\log k_{-X}$ :  $\text{H}_2\text{O}$  (5.20) = **2** (5.20)  $\gg$   $\text{BuNH}_2$  (1.21)  $>$   $\text{EtNH}_2$  (0.82).<sup>[21]</sup> These results from kinetic measurements may provide an alternative interpretation for the observed variation of the regioselectivity of the attack of **1** by different nucleophiles. For water and anilines, attack at C1 of **1** is highly reversible, whereas stable products are formed by the irreversible  $\text{S}_{\text{N}}2$  reaction at C3 of **1** (see Scheme 5). In accordance with this interpretation, the hydrolysis of (S)-**1** in buffered solution at pH 7 was reported to give selective inversion of the stereocenter to yield (R)-3-HBA.<sup>[10b]</sup> Although the nucleophilicity of **6** and **7** is in the same range as that of 2–5, products of C1 attack at **1** were obtained with the primary alkylamines **6** and **7**. This observation is in accord with Ritchie's  $\log k_X$  values for tetrahedral intermediates formed by nucleophilic attack on carboxylic esters, which indicate that the primary aliphatic amines are  $10^4$ -fold weaker nucleofuges than anilines or water.

Amines 2–7 employed in this study cover a relatively small range of nucleophilicity in water ( $12.99 < N < 14.28$ ), which corresponds to a reactivity difference of only a factor of 20 towards a certain electrophile (if differences in  $s_N$  are neglected).

Because anilines attack at C3, whereas aliphatic amines attack at C1, the rate constants for nucleophilic attack at 1 are additionally influenced to a different extent by anomeric effects.<sup>[22]</sup>

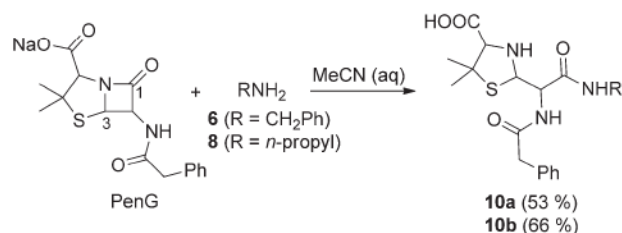
Mayr's reference electrophiles have been assigned the same value of  $E$  in different solvents.<sup>[8a-e]</sup> It can be expected, however, that the use of solvent-independent  $E$  parameters is not possible for all electrophiles. If products with highly localized charges are formed during nucleophilic attack, solvent-dependent  $E$  parameters are required.<sup>[23]</sup> In this study, the rate constant of the reaction of 1 with 6 decreases by a factor of 23 when changing from aqueous solution to DMSO, although an increase in the rate constants by about one order of magnitude ( $N/s_N$  for 6 in DMSO: 15.28/0.65) is expected when assuming the same electrophilic reactivity of 1 in both solvents.<sup>[8d]</sup> As a consequence, it is not possible to determine a reliable  $E$  parameter for C1 or C3 attack on 1. Hence, the discussion of  $\beta$ -lactam reactivities in the following paragraph has been based on relative rate constants towards amines.

### Electrophilic reactivity of beta-lactams

Monocyclic beta-lactams exhibit limited reactivity with enzyme active sites, whereas their bicyclic analogues are potent antibiotics with pronounced protein reactivity.<sup>[3]</sup> This physiological observation is in agreement with our experimental and theoretical studies, which reveal no turnover of 4-methyl-1-azetidin-2-one with nucleophiles owing to an insuperable activation barrier of 40–70 kcal mol<sup>-1</sup> (Table S2 and Figures S5 and S6 in the Supporting Information). In contrast, penicillin G, which is a bicyclic beta-lactam, reacts with 6, 7, and 8 at significantly higher rates, ranging from  $8 \times 10^{-2}$  (for 6) via  $9 \times 10^{-2}$  (for 7) to  $4 \times 10^{-1}$  M<sup>-1</sup>s<sup>-1</sup> (for 8), a trend that was further confirmed by QM/MM calculations (Figure 5). These rates exceed those observed for monocyclic lactones. Accordingly, the acylation products 10a,b of the reactions of penicillin G with 6 and 8 have been isolated and characterized by NMR spectroscopy and HRMS (Scheme 6).

### Conclusion

Electro- and nucleophilicity parameters have been obtained for a variety of compound classes to provide a prognostic value for chemical reactions.<sup>[8]</sup> Natural products exhibit many electro-



Scheme 6. Products of the reactions of sodium penicillin G with amines 6 and 8 in H<sub>2</sub>O/CH<sub>3</sub>CN 9/1 (v/v) (yields refer to products isolated).

philic motifs that have been customized by evolution to irreversibly bind the active sites of enzymes. The corresponding biological activities are as diverse as these scaffolds and range from anticancer to antibiotic properties. Recent drug-discovery efforts have focused on natural-product-inspired covalent inhibitors; however, a quantitative scale to describe and parameterize their reactivity under physiological conditions is currently lacking. We therefore used kinetic experiments to evaluate pharmacologically relevant electrophilic scaffolds. Importantly, the ambident reactivity of 1 (C1 vs. C3 attack) towards anilines and primary aliphatic amines was explained by QM/MM calculations by using linear-scaling QM methods. Strikingly, these calculations revealed that H<sub>3</sub>O<sup>+</sup> molecules were crucial for catalysis owing to their stabilizing effects on the transition state of aniline attack at the C3 position of the lactone. Although the energy path for this reaction was significantly lower than that of C1 attack, compound 6 preferred C1 attack (with or without H<sub>3</sub>O<sup>+</sup> catalysis) owing to an insuperable C3 energy barrier. All of these calculations were supported by experimental data, which confirmed the influence of water for the corresponding reaction rates. The value of this platform has been demonstrated in the direct comparison of the reactivity of 1, 4-methyl-1-azetidin-2-one, and penicillin G. The phenomenological observation that penicillin G and beta-lactones are reactive with protein active sites, whereas monocyclic beta-lactams are weak binders, could be proven experimentally and theoretically. With this proof of principle study, the scope of application can easily be expanded to other biologically relevant electrophilic scaffolds.

### Acknowledgements

We thank Prof. Dr. H. Mayr (LMU München) for generous support and Dr. D. Stephenson (LMU München) for help with the NMR spectroscopy measurements. We acknowledge financial support by the DFG funding initiative SFB749 (Projects A3, B1, C7), the Excellence Cluster EXC114 (CIPSM), and an ERC starting grant (250924-antibacterials).

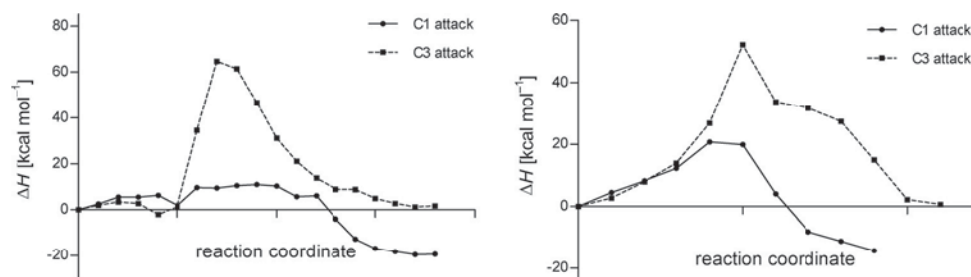


Figure 5. Calculated reaction profiles of 8 (left) and 6 (right) with penicillin G: —: attack at C1; ----: attack at C3. Left: The first barrier represents the reorientation of 8 during approach towards penicillin G. The second barrier represents nucleophilic attack at C1 and C3.



**Keywords:** computational chemistry · kinetics · lactones · reaction mechanisms · structure–activity relationships

- [1] a) M. Gersch, J. Kreuzer, S. A. Sieber, *Nat. Prod. Rep.* **2012**, *29*, 659–682; b) C. Drahl, B. F. Cravatt, E. J. Sorensen, *Angew. Chem. Int. Ed.* **2005**, *44*, 5788–5809; *Angew. Chem.* **2005**, *117*, 5936–5958; c) J. A. H. Schwöbel, Y. K. Koleva, S. J. Enoch, F. Bajot, M. Hewitt, J. C. Madden, D. W. Roberts, T. W. Schultz, M. T. D. Cronin, *Chem. Rev.* **2011**, *111*, 2562–2596.
- [2] D. J. Waxman, J. L. Strominger, *Annu. Rev. Biochem.* **1983**, *52*, 825–869.
- [3] I. Staub, S. A. Sieber, *J. Am. Chem. Soc.* **2008**, *130*, 13400–13409.
- [4] a) M. I. Page, A. P. Laws, *Tetrahedron* **2000**, *56*, 5631–5638; b) P. Proctor, N. P. Gensmantel, M. I. Page, *J. Chem. Soc. Perkin Trans. 2* **1982**, 1185–1192.
- [5] T. Böttcher, S. A. Sieber, *MedChemComm* **2012**, *3*, 408–417.
- [6] a) A. Noel, B. Delpech, D. Crich, *Org. Biomol. Chem.* **2012**, *10*, 6480–6483; b) A. Griesbeck, D. Seebach, *Helv. Chim. Acta* **1987**, *70*, 1326–1332.
- [7] J. A. Manso, M. T. Pérez-Prior, M. d. P. García-Santos, E. Calle, J. Casado, *Chem. Res. Toxicol.* **2005**, *18*, 1161–1166.
- [8] a) H. Mayr, M. Patz, *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 938–957; *Angew. Chem.* **1994**, *106*, 990–1010; b) H. Mayr, A. R. Ofial, *J. Phys. Org. Chem.* **2008**, *21*, 584–595; c) H. Mayr, T. Bug, M. F. Gotta, N. Hering, B. Irrgang, B. Janker, B. Kempf, R. Loos, A. R. Ofial, G. Remennikov, H. Schimmel, *J. Am. Chem. Soc.* **2001**, *123*, 9500–9512; d) a comprehensive listing of nucleophilicity parameters ( $N$ ,  $s_N$ ) and electrophilicity parameters ( $E$ ) is available at <http://www.cup.lmu.de/oc/mayr/DBintro.html>.
- [9] a) B. L. Van Duuren, B. M. Goldschmidt, *J. Med. Chem.* **1966**, *9*, 77–79; b) B. L. Van Duuren, *Ann. N. Y. Acad. Sci.* **1969**, *163*, 633–650.
- [10] a) F. A. Long, M. Purchase, *J. Am. Chem. Soc.* **1950**, *72*, 3267–3273; b) A. R. Olson, R. J. Miller, *J. Am. Chem. Soc.* **1938**, *60*, 2687–2692.
- [11] G. M. Blackburn, H. L. H. Dodds, *J. Chem. Soc. Perkin Trans. 2* **1974**, 377–382.
- [12] a) T. L. Gresham, J. E. Jansen, F. W. Shaver, J. T. Gregory, *J. Am. Chem. Soc.* **1948**, *70*, 999–1001; b) T. L. Gresham, J. E. Jansen, F. W. Shaver, *J. Am. Chem. Soc.* **1948**, *70*, 1003–1004.
- [13] For related kinetic studies of the reactions of  $\beta$ -propiolactone with anionic nucleophiles, see also: P. D. Bartlett, G. Small, *J. Am. Chem. Soc.* **1950**, *72*, 4867–4869.
- [14] M. T. Pérez-Prior, J. A. Manso, M. d. P. García-Santos, E. Calle, J. Casado, *J. Org. Chem.* **2005**, *70*, 420–426.
- [15] J. Kussmann, M. Beer, C. Ochsenfeld, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 614–636.
- [16] a) A. Warshel, *Angew. Chem. Int. Ed.* **2014**, *53*, 10020–10031; *Angew. Chem.* **2014**, *126*, 10182–10194; b) H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229; *Angew. Chem.* **2009**, *121*, 1220–1254.
- [17] C. W. Pemble, L. C. Johnson, S. J. Kridel, W. T. Lowther, *Nat. Struct. Mol. Biol.* **2007**, *14*, 704–709.
- [18] M. Gersch, K. Famulla, M. Dahmen, C. Göbl, I. Malik, K. Richter, V. S. Korotkov, P. Sass, H. Rübsamen-Schaeff, T. Madl, H. Brötz-Oesterhelt, S. A. Sieber, *Nat. Commun.* **2015**, *6*, 6320.
- [19] a) Y. Altun, *J. Solution Chem.* **2004**, *33*, 479–497; b) D. Xiong, Z. Li, H. Wang, J. Wang, *Green Chem.* **2013**, *15*, 1941–1948; c) P. J. Battye, E. M. Ihsan, R. B. Moodie, *J. Chem. Soc. Perkin Trans. 2* **1980**, 741–748; d) C. H. Arrowsmith, H. X. Guo, A. J. Kresge, *J. Am. Chem. Soc.* **1994**, *116*, 8890–8894; e) R. F. Jameson, G. Hunter, T. Kiss, *J. Chem. Soc. Perkin Trans. 2* **1980**, 1105–1110.
- [20] F. Brotzel, Y. C. Chu, H. Mayr, *J. Org. Chem.* **2007**, *72*, 3679–3688.
- [21] C. D. Ritchie, *J. Am. Chem. Soc.* **1975**, *97*, 1170–1179.
- [22] T. A. Nigst, H. Mayr, *Eur. J. Org. Chem.* **2013**, 2155–2163.
- [23] H. Mayr, *Tetrahedron* **2015**, *71*, 5095–5111.

Manuscript received: May 26, 2015

Final Article published: July 16, 2015



**Kinetic and theoretical studies of beta-lactone reactivity – a quantitative scale for  
biological application**

*Elija N. Wiedemann,<sup>a,†</sup> Franziska A. Mandl,<sup>b,†</sup> Iris D. Blank,<sup>a,†</sup> Christian Ochsenfeld,<sup>a,\*</sup>  
Armin R. Ofial,<sup>a,\*</sup> Stephan A. Sieber<sup>b,\*</sup>*

[a] *E. N. Wiedemann,<sup>†</sup> I. D. Blank,<sup>†</sup> Prof. Dr. C. Ochsenfeld, Dr. A. R. Ofial*

Department Chemie, Ludwig-Maximilians-Universität München

Butenandtstraße 5-13, 81377 München (Germany)

[b] *F. A. Mandl,<sup>†</sup> Prof. Dr. S. A. Sieber*

Center for Integrated Protein Science CIPSM

Institute of Advanced Studies IAS, Department Chemie

Lehrstuhl für Organische Chemie II, Technische Universität München

Lichtenbergstrasse 4, 85747 Garching (Germany)

[<sup>†</sup>] These authors contributed equally to this work.

\* To whom correspondence should be addressed: [stephan.sieber@tum.de](mailto:stephan.sieber@tum.de), [ofial@lmu.de](mailto:ofial@lmu.de),  
[christian.ochsenfeld@cup.uni-muenchen.de](mailto:christian.ochsenfeld@cup.uni-muenchen.de)

## Table of Content

1.	Supporting figures and tables .....	S4
2.	General methods and materials .....	S8
3.	Stability of $\beta$ -butyrolactone (BBL) in different buffer systems – Determination of first-order rate constants $k_1$ .....	S10
3.1.	Kinetic data .....	S10
3.2.	HRMS-studies .....	S18
4.	Kinetics of the reaction of BBL with amines in TEA-buffered D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v) – Determination of second-order rate constants $k_N$ .....	S22
4.1.	Kinetics of the reaction of BBL with aniline in TEA-buffered D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v) .....	S23
4.2.	Kinetics of the reaction of BBL with <i>m</i> -toluidine in TEA-buffered D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v) .....	S27
4.3.	Kinetics of the reaction of BBL with <i>p</i> -toluidine in TEA-buffered D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v) .....	S28
4.4.	Kinetics of the reaction of BBL with <i>p</i> -anisidine in TEA-buffered D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v) .....	S29
4.5.	Kinetics of the reaction of BBL with benzylamine in TEA-buffered D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v) .....	S30
4.6.	Kinetics of the reaction of BBL with phenylethylamine in TEA-buffered D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v) .....	S31
4.7.	Rate constants for the reactions of BBL with amines .....	S32
4.8.	Hammett plot for the reactions of BBL with anilines .....	S32
5.	Kinetics of the reaction of BBL with benzylamine in d <sub>6</sub> -DMSO .....	S33
6.	Determination of second-order rate constants of the reaction of BBL with carbanions in DMSO .....	S34



7.	Determination of second-order rate constants of the reaction of sodium penicillin G with various amines in D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v) .....	S42
7.1.	Consumption of sodium penicillin G in TEA-buffered solution – Determination of the first-order rate constant $k_1$ .....	S43
7.2.	Kinetics of the reaction of sodium penicillin G with benzylamine in TEA-buffered D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v).....	S46
7.3.	Kinetics of the reaction of sodium penicillin G with phenylethylamine in TEA-buffered D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v) .....	S47
7.4.	Kinetics of the reaction of sodium penicillin G with <i>n</i> -propylamine in TEA-buffered D <sub>2</sub> O/d <sub>6</sub> -DMSO 9/1 (v/v) .....	S48
7.5.	Rate constants for the reactions of sodium penicillin G with different aliphatic amines.....	S49
8.	Determination of <i>N</i> and <i>s<sub>N</sub></i> parameters of amines in H <sub>2</sub> O .....	S50
8.1.	Phenylethylamine.....	S51
8.2.	<i>p</i> -Anisidine.....	S57
9.	Product studies .....	S61
9.1.	Reactions of amines and anilines with BBL .....	S61
9.2.	Reactions of amines with sodium penicillin G .....	S64
10.	Syntheses of lactams .....	S66
11.	NMR Spectra.....	S68
12.	References .....	S78

## 1. Supporting figures and tables

**Table S1:** Calculated reaction barriers (enthalpies) of the reaction of  $\beta$ -butyrolactone (BBL) with aniline and benzylamine in water and in water including  $\text{H}_3\text{O}^+$ .

	aniline		benzylamine	
attack position lactone	C1	C3	C1	C3
in water including $\text{H}_3\text{O}^+$ (kcal/mol)	21*	21	0	78
in water (kcal/mol)	53	49	12	37

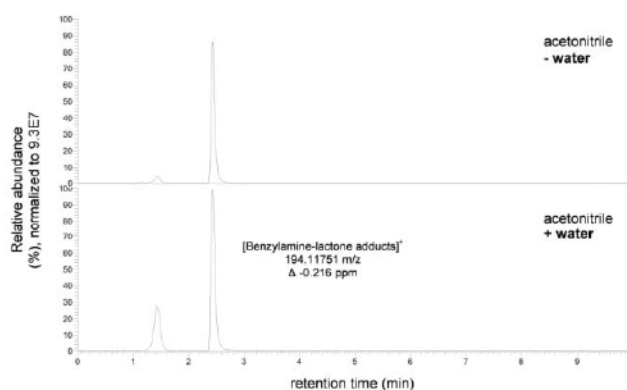
\* product not stable

**Table S2:** Calculated reaction barriers (enthalpies) of the reaction of 4-methyl-1-azetidin-2-one with aniline and benzylamine in water.

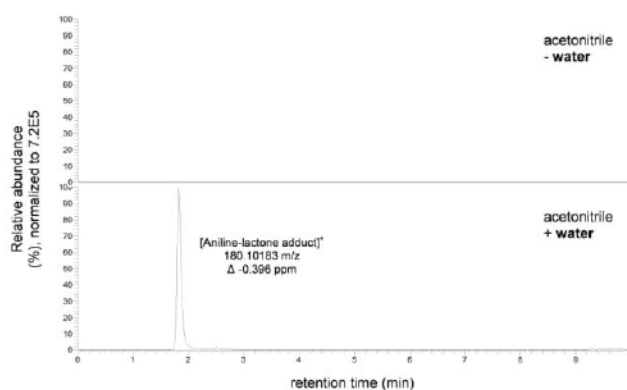
	aniline		benzylamine	
attack position lactam	C1	C3	C1	C3
in water (kcal/mol)	42	67	44	68

**Table S3:** Calculated reaction barriers (enthalpies) of the reaction of penicillin G with *n*-propylamine and benzylamine in water.

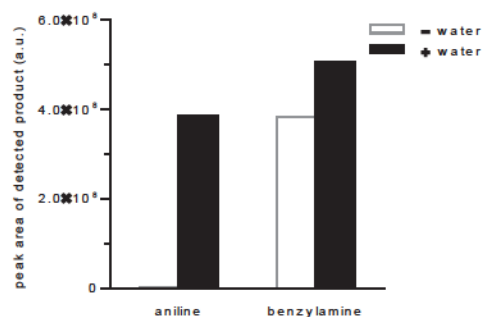
	<i>n</i> -propylamine		benzylamine	
attack position penicillin G	C1	C3	C1	C3
in water (kcal/mol)	11	65	21	52



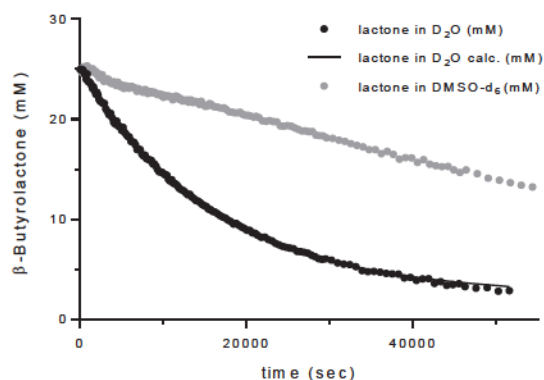
**Figure S1:** Mass range [194.11659 m/z – 194.11853 m/z] of full-scan [50-600 m/z] reversed-phase HPLC-ESI-HRMS (positive polarity) traces of the conversion of  $\beta$ -butyrolactone (BBL) with benzylamine after the addition of water compared to the reaction in acetonitrile.



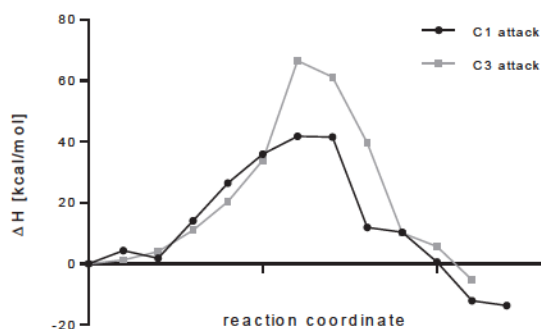
**Figure S2:** Mass range [180.10101 m/z – 180.10281] of full-scan [50-600 m/z] reversed-phase HPLC-ESI-HRMS (positive polarity) traces of the conversion of BBL with aniline after the addition of water compared to the reaction in acetonitrile.



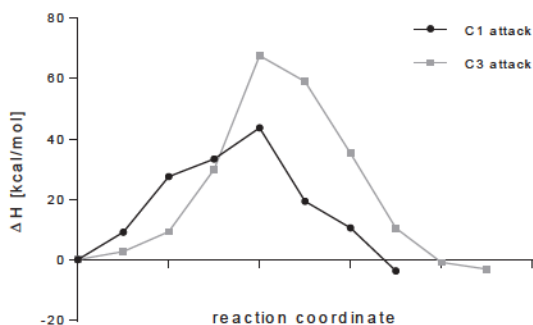
**Figure S3:** Relative comparison of corresponding peak areas (Figure S1, 2) of formed products of the reactions of BBL with aniline and benzylamine with (black) and without (white) water in acetonitrile measured by reversed-phase HPLC-ESI-HRMS (positive polarity).



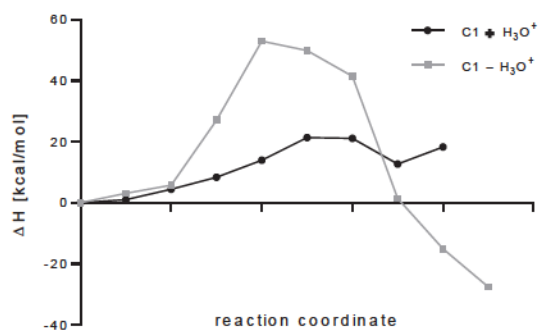
**Figure S4:** Decay of the concentration of BBL (determined by  $^1H$  NMR spectroscopy) while reacting with benzylamine in an aqueous TEA-buffered reaction mixture at 37 °C (black curve,  $[BBL]_0 = 25$  mM,  $[PhCH_2NH_2]_0 = 50$  mM,  $[PhCH_2NH_2]_{eff} = 1.2 \times 10^{-3}$  M,  $[TEA]_0 = 0.10$  M,  $pH = 7.75$ , solvent  $D_2O/DMSO-d_6$  9/1 (v/v),  $k_{obs} = 6.2 \times 10^{-5} s^{-1}$ ,  $k_N = 5.8 \times 10^{-3} M^{-1} s^{-1}$ ) and while reacting with benzylamine in  $d_6$ -DMSO at 37 °C versus time (grey curve,  $[BBL]_0 = 25$  mM,  $[PhCH_2NH_2]_0 = 50$  mM,  $k_N = 2.5 \times 10^{-4} M^{-1} s^{-1}$ ) is shown.



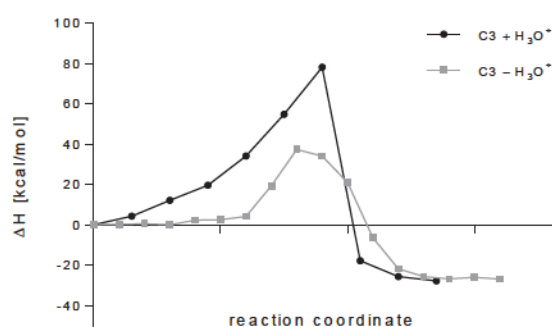
**Figure S5:** Calculated reaction profiles of aniline and 4-methyl-1-azetidin-2-one in water. The attack at C1 is shown in black and the attack at C3 in grey.



**Figure S6:** Calculated reaction profiles of benzylamine and 4-methyl-1-azetidin-2-one in water. The attack at C1 is shown in black and the attack at C3 in grey.



**Figure S7:** Calculated reaction profiles of the C1 attack of aniline on BBL with (black) and without (grey)  $H_3O^+$ .



**Figure S8:** Calculated reaction profiles of the C3 attack of benzylamine on BBL with (black) and without (grey)  $H_3O^+$ .

## 2. General methods and materials

All chemical reagents and solvents used for product studies were of reagent grade or of higher purity and used without further purification as obtained from the commercial sources Alfa-Aesar, AppliChem, Fluka/Sigma-Aldrich, Merck and Roth. Chemical reagents and solvents used for kinetic NMR-studies were purified by distillation. Concentrations under reduced pressure were performed by rotary evaporation at 40 °C. Yields refer to purified, dried and spectroscopically pure compounds.

**$^1\text{H}$  NMR and  $^{13}\text{C}$  NMR** spectra were recorded at  $(26 \pm 1)$  °C on Bruker Avance 360, Avance 500 and Avance III 500 spectrometers.  $^1\text{H}$  NMR kinetic measurements were performed on Varian VNMRS 400 or Varian 200 spectrometers at 37 °C. Chemical shifts ( $\delta$ ) were referenced to the residual proton or carbon signal of the deuterated solvent in ppm ( $d_6$ -DMSO:  $\delta_{\text{H}}$  2.50,  $\delta_{\text{C}}$  39.52;  $\text{CDCl}_3$ :  $\delta_{\text{H}}$  7.26,  $\delta_{\text{C}}$  77.16;  $\text{D}_2\text{O}$ :  $\delta_{\text{H}}$  4.79).<sup>[1]</sup> Coupling constants ( $J$ ) are reported in Hertz (Hz) and multiplicity is reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet or unresolved.

**Flash column chromatography** was performed on Merck silica gel (Geduran Si 60, 0.040–0.063 mm) using a forced flow eluant at 0.3 – 0.5 bar pressure. Analytical thin layer chromatography (TLC) was performed on aluminium-backed TLC silica gel 60 F254 plates by Merck, and compounds were visualized by UV detection (254 nm) or staining with aqueous  $\text{KMnO}_4/\text{Na}_2\text{CO}_3$  solution and subsequent heating.

**Reversed-phase HPLC analysis** was performed on a Waters 2695 separation module, equipped with a Waters PDA 2996 and a Waters XBridge C18 column (3.5  $\mu\text{m}$ ,  $4.6 \times 0.10$  M, flow = 1.2 mL/min, method: gradient 2% B  $\rightarrow$  100% B over 30 min). For preparative scale RP-HPLC separation a Waters 2545 quaternary gradient module in combination with a Waters PDA 2998 and a Waters XBridge C18 (5.0  $\mu\text{m}$ ,  $30 \times 150$  mm, flow = 50 mL/min, method: gradient 2% B  $\rightarrow$  100% B over 17 min) column was used. The mobile phase for elution consisted of a gradient mixture of 0.1% (v/v) TFA in water (buffer A, HPLC grade) and 0.1% (v/v) TFA in acetonitrile (buffer B, HPLC grade).

**Reversed-phase HPLC-ESI-HR-MS analysis** was performed on a Thermo Finnigan LTQ FT-ICR equipped with a Dionex Ultimate 3000 separation module eluting on a Waters XBridge C18 column (3.5  $\mu\text{m}$ ,  $4.6 \times 0.10$  M, flow = 1.1 mL/min). The column temperature was maintained at 30 °C. The mobile phase for elution consisted of a gradient mixture of

0.1% (v/v) formic acid in water (buffer A, HPLC-MS grade) and 0.1% (v/v) formic acid in acetonitrile:water 90/10 (buffer B, HPLC-MS grade).

### Computational Details

XLEAP (AmberTool)<sup>[2]</sup> has been used to solvate the solutes (lactone/lactam/penicillin G and the corresponding nucleophile) in a box of explicit TIP3P water<sup>[3]</sup> with a buffer of 10 Å around the solute.

ANTECHAMBER<sup>[4]</sup> was used to parameterize the molecules. For force field minimizations the NAMD engine was used.<sup>[5]</sup> Periodic boundary conditions and particle mesh Ewald summation (PME) with a cutoff value of 12 Å were employed. The system was energy minimized (NVT ensemble) using the conjugate gradient algorithm within 10000 steps. The system was heated up to 300 K within 30 ps. After this, the system was equilibrated for 230 ps. In the second equilibration step we switched to the NPT ensemble employing the Langevin piston Nosé-Hoover method.<sup>[6]</sup> Production runs were performed for 10 ns using the SHAKE algorithm.<sup>[7]</sup> In all steps a positional constraint of 1 kcal/mol/Å<sup>2</sup> on non-water atoms was applied.

For QM/MM structure optimizations the DL-POLY implementation within ChemShell<sup>[8]</sup> was combined with density functional theory (DFT) at the BP86-D3/6-31G\*\*<sup>[9]</sup> level of theory employing the Q-Chem program package<sup>[10]</sup> for the QM part. BP86-D3 was chosen for optimization due to its particular low “weighted total mean absolute deviation” (WTMAD) for reaction energies (3.5 kcal/mol)<sup>[11]</sup> and its relatively low computational cost. The reaction mechanisms were calculated using both the adiabatic mapping approach and the nudged elastic band method of the DL-FIND<sup>[12]</sup> module implemented in ChemShell.<sup>[8]</sup> The system for these calculations consists of the whole structure (total system size is up to 3144 atoms), with 8 Å around the solutes as relaxed region (up to 1035 atoms) and 3.5 Å around the solutes as QM region (up to 222 QM atoms).



### 3. Stability of $\beta$ -butyrolactone (BBL) in different buffer systems – Determination of first-order rate constants $k_1$

#### 3.1. Kinetic data

##### 3.1.1. General

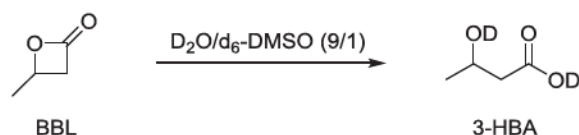
A 0.11 M buffer solution in D<sub>2</sub>O (540  $\mu$ L) was mixed with a solution of BBL (60  $\mu$ L of a 0.25 M solution in d<sub>6</sub>-DMSO) in an NMR tube at 37 °C to produce a 9/1 (v/v) D<sub>2</sub>O/d<sub>6</sub>-DMSO solvent mixture. The conversion of BBL was monitored by using time-resolved <sup>1</sup>H NMR spectroscopy. Concentrations of BBL, [BBL]<sub>t</sub>, were calculated from the ratio of the integral of the methyl resonances of BBL (d,  $\delta$  = 1.70 ppm) to the integral of the methyl resonances in the range  $\delta$  = 1.43–1.27 ppm.

The first-order rate constants  $k_{\text{obs}}$  ( $= k_1$  in equation (3) of the main text) were determined by least-squares fitting of the exponential function  $[\text{BBL}]_t = [\text{BBL}]_0 \cdot \exp(-k_{\text{obs}}t) + C$  to the time-dependent [BBL]<sub>t</sub>.

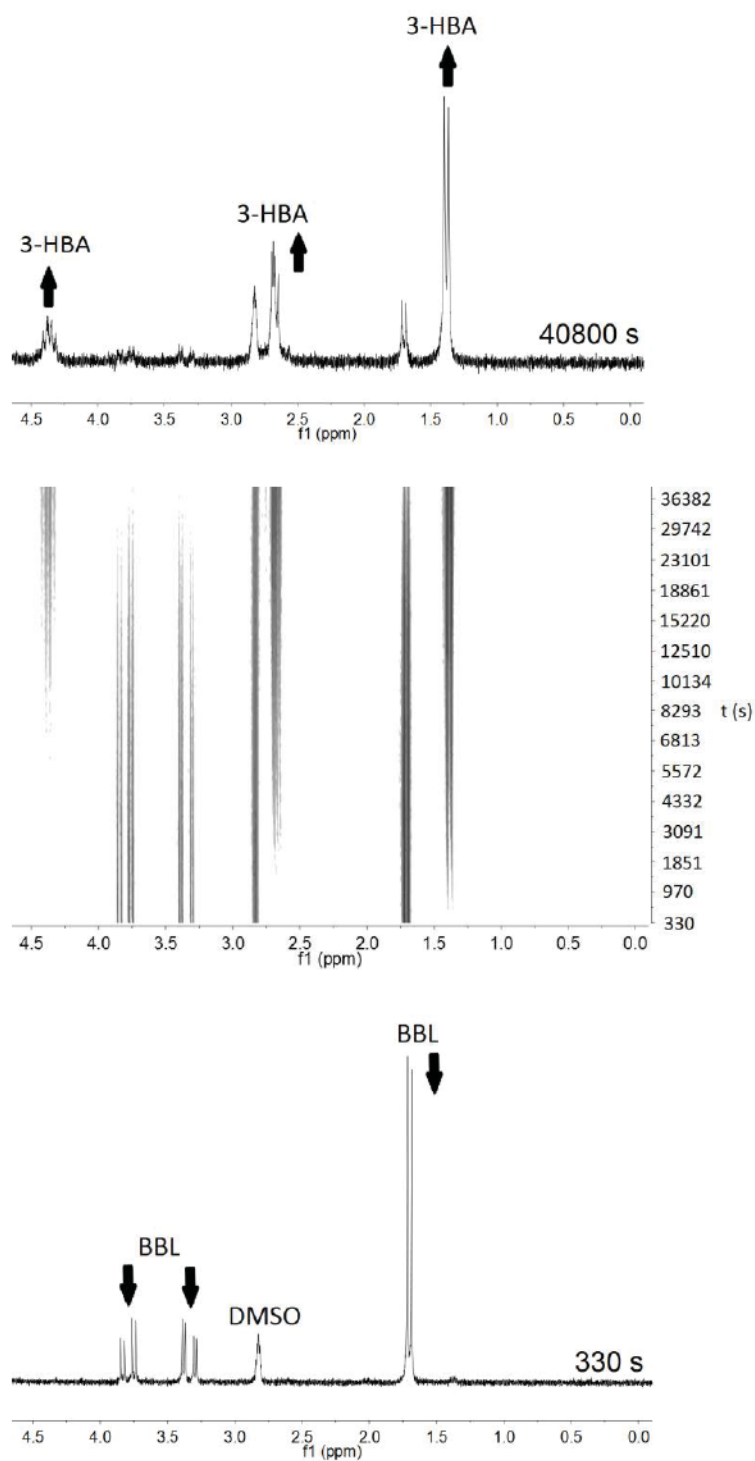
$$k_{\text{obs}} = k_1 = k_{\text{W}} + k_{\text{DO}^-}[\text{DO}^-] + k_{\text{buffer}}[\text{buffer}] \quad (3)$$

##### 3.1.2. Hydrolysis of BBL without buffer

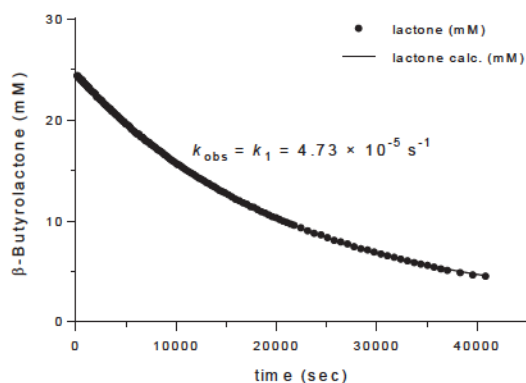
D<sub>2</sub>O (540  $\mu$ L) and BBL (60  $\mu$ L of a 0.25 M solution in d<sub>6</sub>-DMSO) were mixed in an NMR tube at 37 °C. The conversion of BBL was monitored by using time-resolved <sup>1</sup>H NMR spectroscopy. Concentrations of BBL, [BBL]<sub>t</sub>, were calculated from the ratio of the integral of the methyl resonances of BBL (d,  $\delta$  = 1.70 ppm) to the integral of the methyl resonances assigned to 3-hydroxybutanoic acid (3-HBA: d,  $\delta$  = 1.38 ppm).







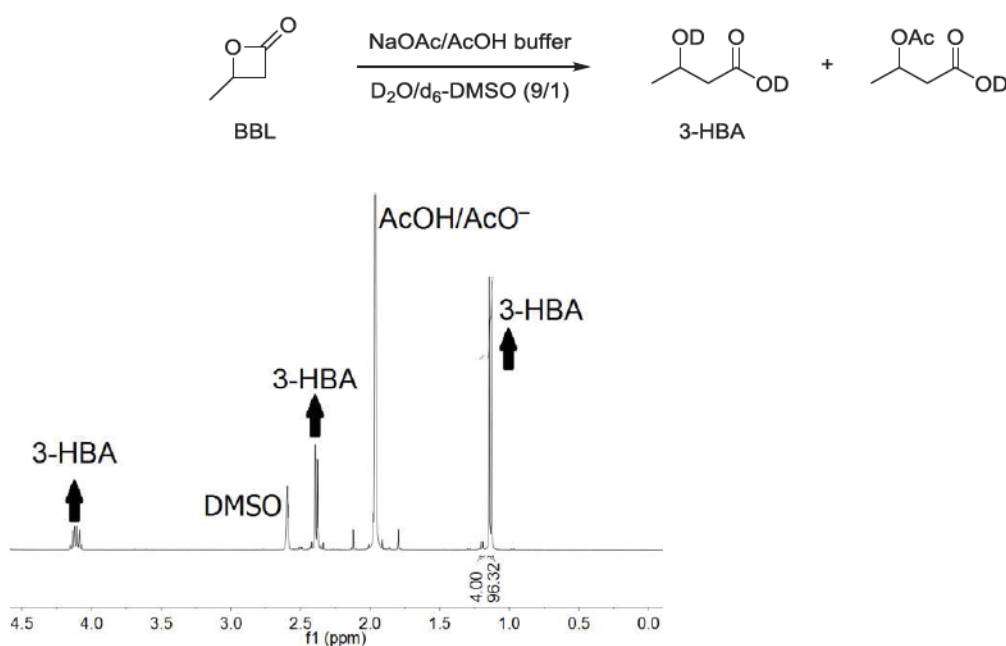
**Figure S9:** Exemplary time-resolved  $^1\text{H}$  NMR spectra of the hydrolysis of BBL in  $\text{D}_2\text{O}/d_6\text{-DMSO}$  9/1 (v/v) at 37  $^\circ\text{C}$  ( $[\text{BBL}]_0 = 25 \text{ mM}$ ). Start of the reaction (bottom), time-resolved topview of  $^1\text{H}$  NMR spectra (middle), showing the decrease of BBL resonances and increase of 3-HBA resonances, and final  $^1\text{H}$  NMR spectrum after 40800 sec (top).



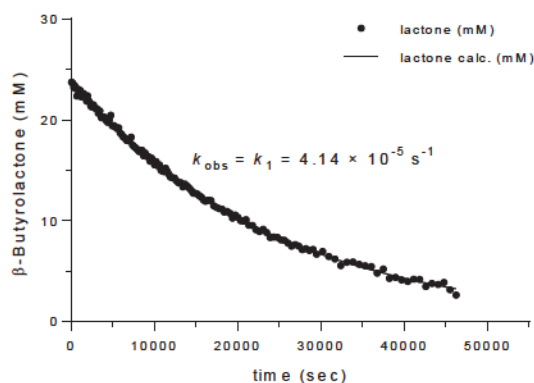
**Figure S10:** Decay of the concentration of BBL in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v) at 37 °C versus time ([BBL]<sub>0</sub> = 25 mM, pH = 7.00). The black line shows a mono-exponential fit of the data.

### 3.1.3. Hydrolysis of BBL in the presence of acetate buffer

Acetate buffer (540 μL of a 0.11 M solution of NaOAc/HOAc in D<sub>2</sub>O, pH 4.75) and BBL (60 μL of a 0.25 M solution in d<sub>6</sub>-DMSO) were mixed in an NMR tube at 37 °C. The conversion of BBL was monitored by using time-resolved <sup>1</sup>H NMR spectroscopy. Concentrations of BBL, [BBL]<sub>t</sub>, were calculated from the ratio of the integral of the methyl resonances of BBL (d, δ = 1.70 ppm) to the integral of the methyl resonances in the range δ = 1.43–1.27 ppm.



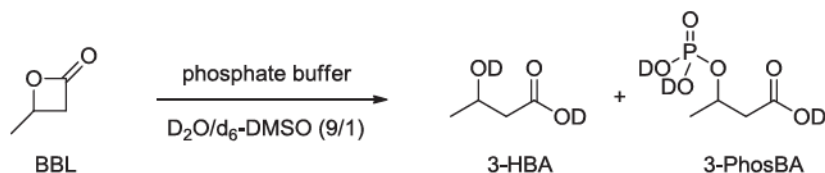
**Figure S11:** <sup>1</sup>H NMR spectrum (400 MHz) of BBL in an acetate-buffered reaction mixture after 60 h at 37 °C ([BBL]<sub>0</sub> = 25 mM, [AcOH]<sub>0</sub> = 0.10 M, pH = 4.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v)).

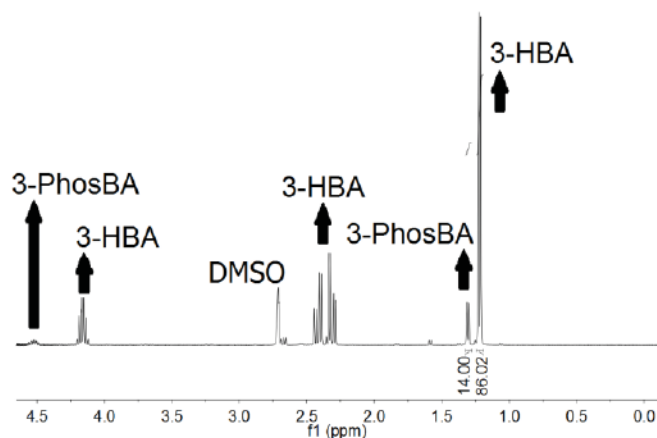


**Figure S12:** Decay of the concentration of BBL in an acetate-buffered reaction mixture at 37 °C ( $[BBL]_0 = 25$  mM,  $[AcOH]_0 = 0.10$  M, pH = 4.75 in  $D_2O/d_6$ -DMSO 9/1 (v/v)). The black line shows a mono-exponential fit of the data.

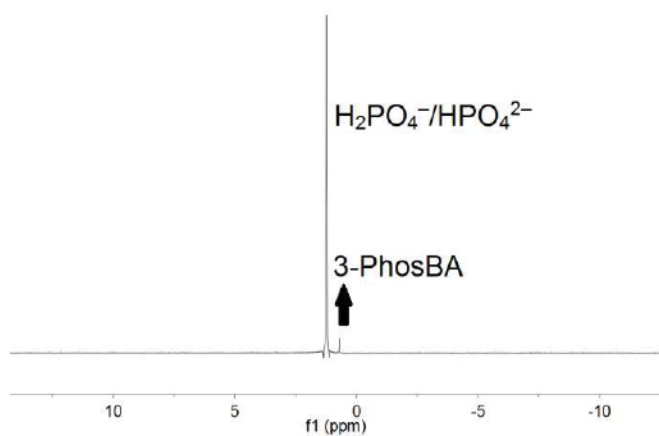
#### 3.1.4. Hydrolysis of BBL in the presence of phosphate buffer

Phosphate buffer (540  $\mu$ L of a 0.11 M solution of  $Na_2HPO_4/NaH_2PO_4$  in  $D_2O$ , pH 7.00) and BBL (60  $\mu$ L of a 0.25 M solution in  $d_6$ -DMSO) were mixed in an NMR tube at 37 °C. The conversion of BBL was monitored by using time-resolved  $^1H$  NMR spectroscopy. Concentrations of BBL,  $[BBL]_t$ , were calculated from the ratio of the integral of the methyl resonances of BBL (d,  $\delta = 1.70$  ppm) to the integral of the methyl resonances in the range  $\delta = 1.43$ – $1.27$  ppm that were assigned to the formation of 3-HBA and 3-(phosphonooxy)butanoic acid (3-PBA, Figure S13). C-4 attack of BBL by dihydrogenphosphate or hydrogenphosphate ions to form 3-PBA is further corroborated by  $^{31}P$  NMR spectroscopic analysis of the reaction mixture, in which the resonance at  $\delta_P = 0.69$  ppm was assigned to the formation of a phosphoric acid ester (Figure S14).

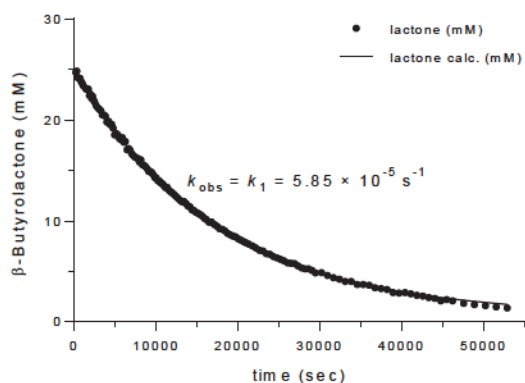




**Figure S13:**  $^1\text{H}$  NMR spectrum (400 MHz) of BBL in a phosphate-buffered reaction mixture after 49 h at 37 °C ( $[\text{BBL}]_0 = 25 \text{ mM}$ ,  $[\text{NaH}_2\text{PO}_4]_0 + [\text{Na}_2\text{HPO}_4]_0 = 0.10 \text{ M}$ ,  $\text{pH} = 7.00$  in  $\text{D}_2\text{O}/d_6\text{-DMSO}$  9/1 (v/v)).



**Figure S14:**  $^{31}\text{P}\{-^1\text{H}\}$  NMR spectrum (162 MHz) of BBL in a phosphate-buffered reaction mixture after 49 h at 37 °C ( $[\text{BBL}]_0 = 25 \text{ mM}$ ,  $[\text{NaH}_2\text{PO}_4]_0 + [\text{Na}_2\text{HPO}_4]_0 = 0.10 \text{ M}$ ,  $\text{pH} = 7.00$ , solvent  $\text{D}_2\text{O}/d_6\text{-DMSO}$  9/1 (v/v)). The resonance at  $\delta = 1.23 \text{ ppm}$  is assigned to the phosphate buffer, the resonance at  $\delta = 0.69 \text{ ppm}$  is assigned to 3-(phosphonoxy)butanoic acid (3-PhosBA).

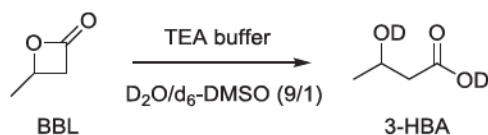


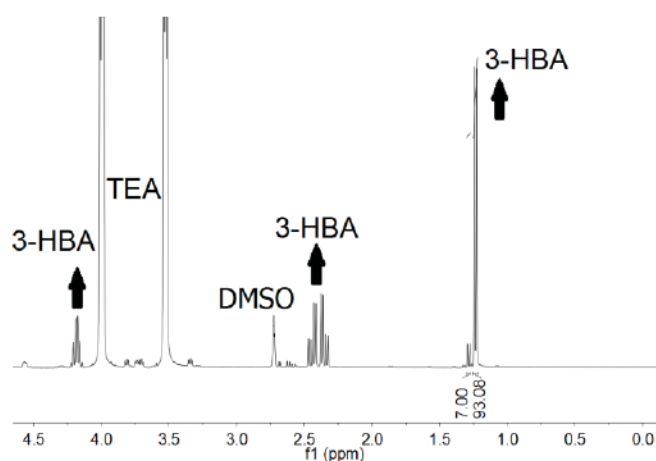
**Figure S15:** Decay of the concentration of BBL in a phosphate-buffered reaction mixture at 37 °C ( $[BBL]_0 = 25$  mM,  $[NaH_2PO_4]_0 + [Na_2HPO_4]_0 = 0.10$  M, pH = 7.00 in  $D_2O/d_6$ -DMSO 9/1 (v/v)). The black line shows a mono-exponential fit of the data.

### 3.1.5. Hydrolysis of BBL in the presence of triethanolamine buffer (TEA buffer)

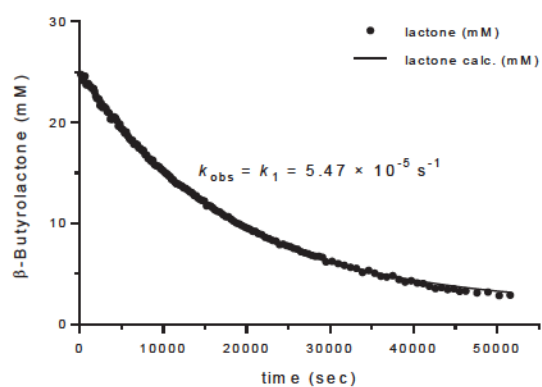
TEA buffer (540  $\mu$ L of a 0.11 M solution of TEA/TEA·HCl in  $D_2O$ , pH 7.75) and BBL (60  $\mu$ L of a 0.25 M solution in  $d_6$ -DMSO) were mixed in an NMR tube at 37 °C. The conversion of BBL was monitored by using time-resolved  $^1H$  NMR spectroscopy. Concentrations of BBL,  $[BBL]_t$ , were calculated from the ratio of the integral of the methyl resonances of BBL (d,  $\delta = 1.70$  ppm) to the integral of the methyl resonances in the range  $\delta = 1.43$ –1.27 ppm (Figure S16).

Beside 3-HBA, an adduct of BBL and TEA is formed. The 3-HBA/BBL-TEA-adduct ratio increases from 5.0 to 7.5 during the reaction (14 h, Figure S18), however reaches 13.3 after 13 days (Figure S16).

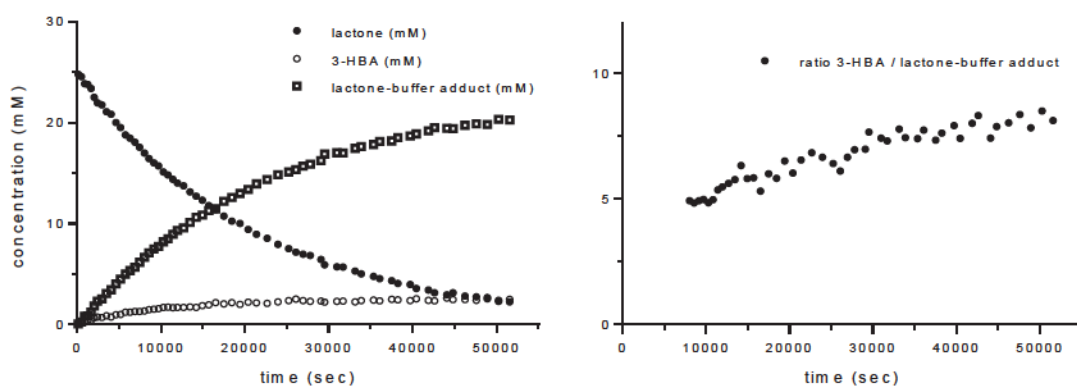




**Figure S16:**  $^1\text{H}$  NMR spectrum (400 MHz) of BBL in a TEA-buffered reaction mixture after 13 days at 37 °C ( $[\text{BBL}]_0 = 25 \text{ mM}$ ,  $[\text{TEA}]_0 = 0.10 \text{ M}$ ,  $\text{pH} = 7.75$  in  $\text{D}_2\text{O}/d_6\text{-DMSO}$  9/1 (v/v)).



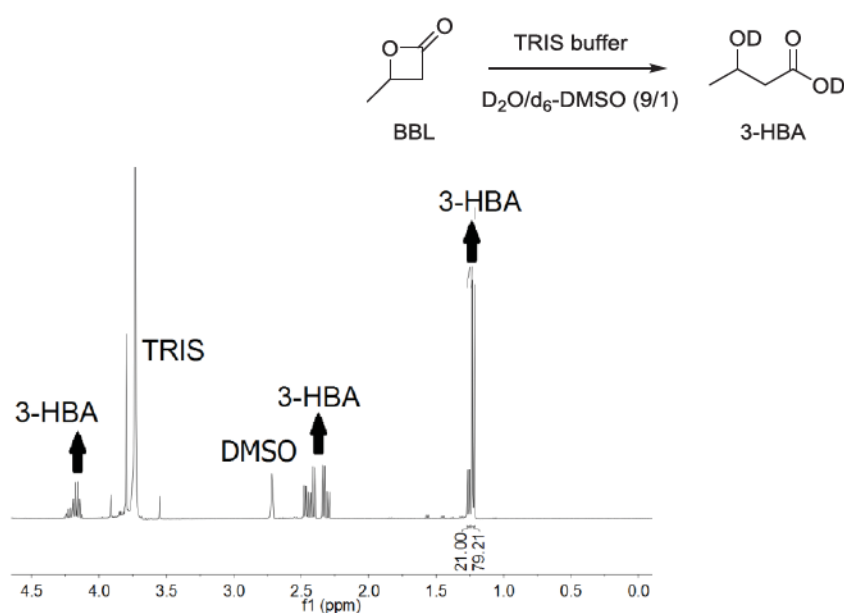
**Figure S17:** Decay of the concentration of BBL in a TEA-buffered reaction mixture at 37 °C ( $[\text{BBL}]_0 = 25 \text{ mM}$ ,  $[\text{TEA}]_0 = 0.10 \text{ M}$ ,  $\text{pH} = 7.75$  in  $\text{D}_2\text{O}/d_6\text{-DMSO}$  9/1 (v/v)). The black line shows a mono-exponential fit of the data.



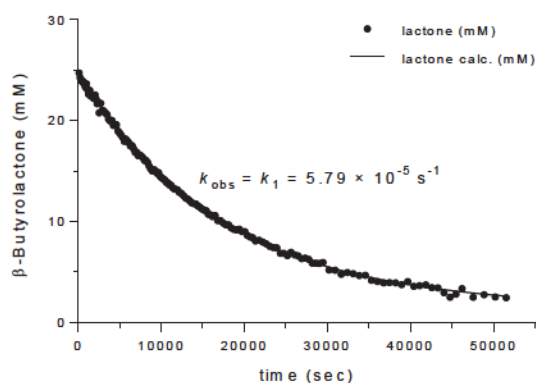
**Figure S18:** Concentration profiles in a TEA-buffered reaction mixture of BBL at 37 °C ( $[\text{BBL}]_0 = 25 \text{ mM}$ ,  $[\text{TEA}]_0 = 0.10 \text{ M}$ ,  $\text{pH} = 7.75$  in  $\text{D}_2\text{O}/d_6\text{-DMSO}$  9/1 (v/v) (left) and 3-HBA/BBL-buffer adduct ratio (right).

### 3.1.6. Hydrolysis of BBL in the presence of tris(hydroxymethyl)aminomethane buffer (TRIS buffer)

TRIS buffer (540  $\mu\text{L}$  of a 0.11 M solution of TRIS/TRIS $\cdot\text{HCl}$  in  $\text{D}_2\text{O}$ , pH 8.10) and BBL (60  $\mu\text{L}$  of a 0.25 M solution in  $\text{d}_6$ -DMSO) were mixed in an NMR tube at 37  $^\circ\text{C}$ . The conversion of BBL was monitored by using time-resolved  $^1\text{H}$  NMR spectroscopy. Concentrations of BBL,  $[\text{BBL}]_t$ , were calculated from the ratio of the integral of the methyl resonances of BBL ( $\delta$ , 1.70 ppm) to the integral of the methyl resonances in the range  $\delta = 1.43\text{--}1.27$  ppm (Figure S18).



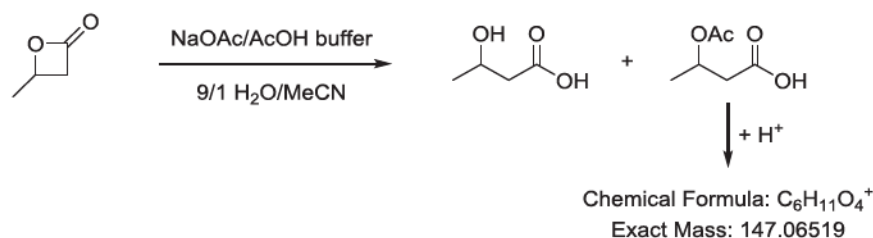
**Figure S19:**  $^1\text{H}$  NMR spectrum (400 MHz) of BBL in a TRIS-buffered reaction mixture after 9 d at 37  $^\circ\text{C}$  ( $[\text{BBL}]_0 = 25 \text{ mM}$ ,  $[\text{TRIS}]_0 = 0.10 \text{ M}$ ,  $\text{pH} = 8.10$  in  $\text{D}_2\text{O}/\text{d}_6\text{-DMSO}$  9/1 (v/v)).



**Figure S20:** Decay of the concentration of BBL in a TRIS-buffered reaction mixture at 37  $^\circ\text{C}$  ( $[\text{BBL}]_0 = 25 \text{ mM}$ ,  $[\text{TRIS}]_0 = 0.10 \text{ M}$ ,  $\text{pH} = 8.10$  in  $\text{D}_2\text{O}/\text{d}_6\text{-DMSO}$  9/1 (v/v)). The black line shows a mono-exponential fit of the data.

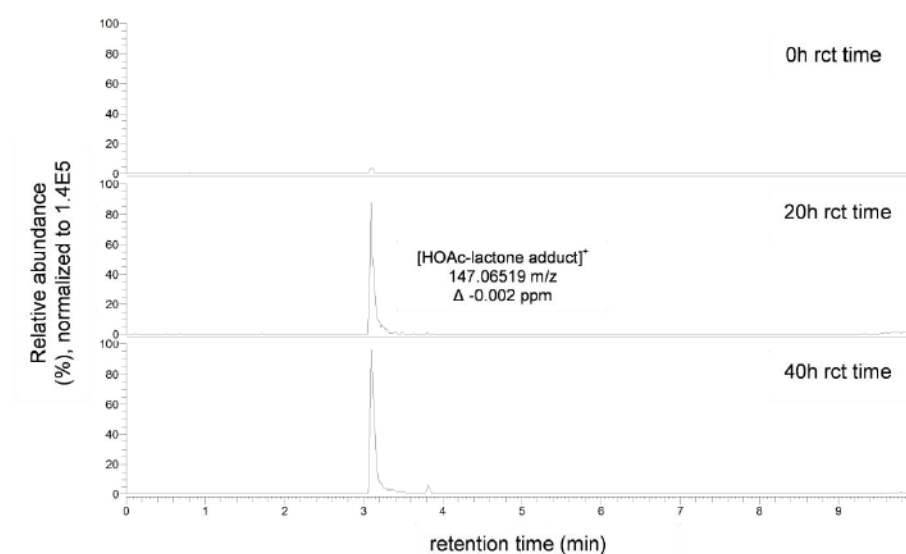
## 3.2. HRMS-studies

### 3.2.1. Reaction of BBL with acetate buffer



To a solution of sodium acetate (246 mg, 3.0 mmol) and acetic acid (173  $\mu$ L, 181 mg, 3.0 mmol) in a 9/1 (v/v)  $H_2O/CH_3CN$ -mixture (60 mL) with pH 4.75 was added BBL (122  $\mu$ L, 129 mg, 1.5 mmol). The reaction mixture was stirred at ambient temperature and samples thereof were analyzed by HPLC-ESI-HRMS after 0, 20, and 40 h.

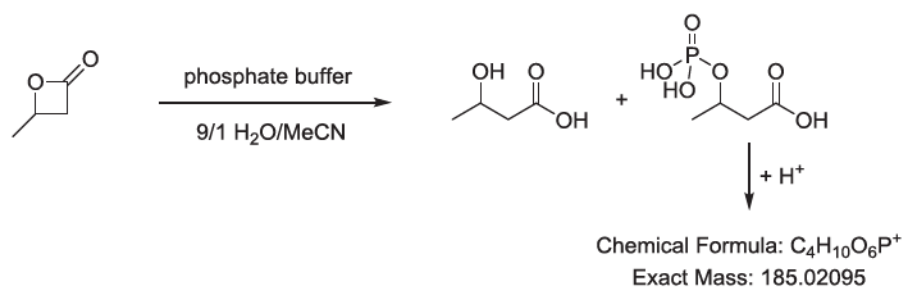
HRMS (ESI): calcd. for  $C_6H_{11}O_4$   $[M+H]^+$ : 147.06519; found: 147.06519.



**Figure S21:** Mass range  $[147.06444\text{ m/z} - 147.06592\text{ m/z}]$  of full-scan  $[50-600\text{ m/z}]$  reversed-phase HPLC-ESI-HRMS traces (positive polarity) of the reaction of BBL in an aqueous acetate-buffered solution after 0 h, 20 h, and 40 h.

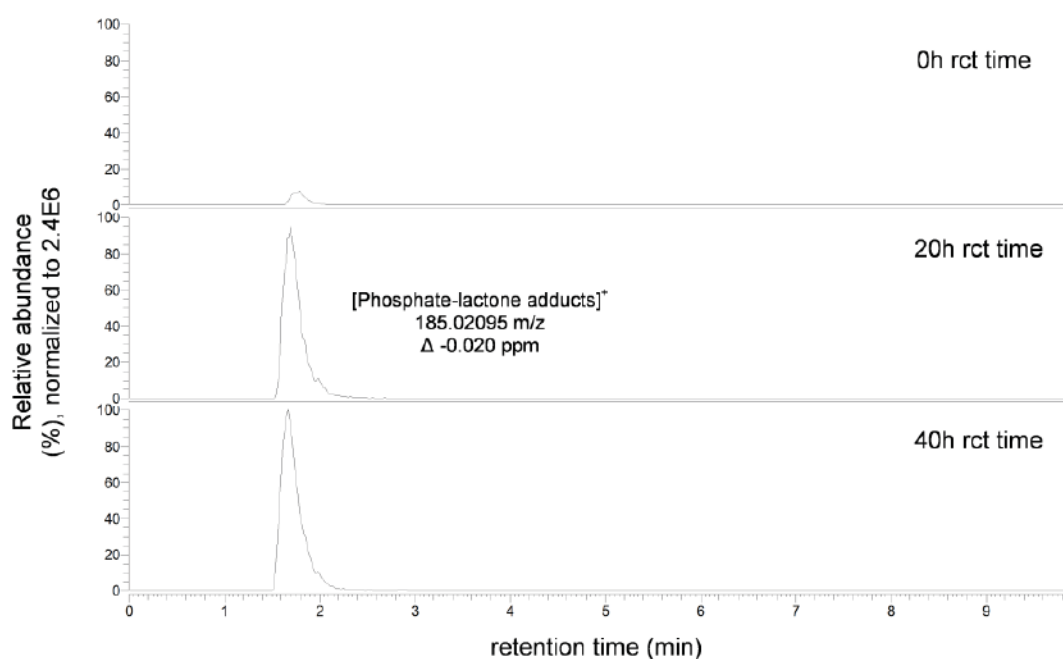


### 3.2.2. Reaction of BBL with phosphate buffer



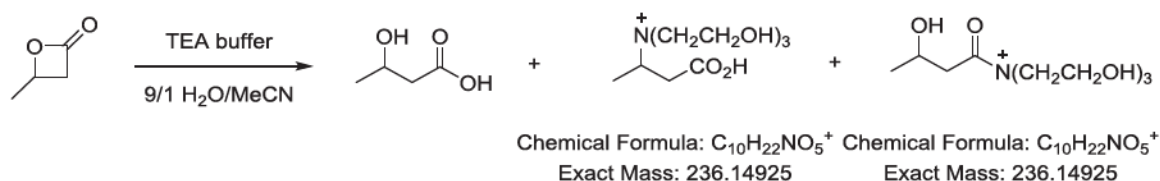
To a solution of sodium hydrogen phosphate (617 mg, 4.35 mmol) and sodium dihydrogen phosphate (228 mg, 1.65 mmol) in a 9/1 (v/v)  $H_2O/CH_3CN$ -mixture (60 mL) with pH 7.00 was added BBL (122  $\mu$ L, 129 mg, 1.5 mmol). The reaction mixture was stirred at ambient temperature and samples thereof were analyzed by HPLC-ESI-HRMS after 0, 20, and 40 h.

HRMS (ESI): calcd. for  $C_4H_{10}O_6P$   $[M+H]^+$ : 185.02095; found: 185.02095.



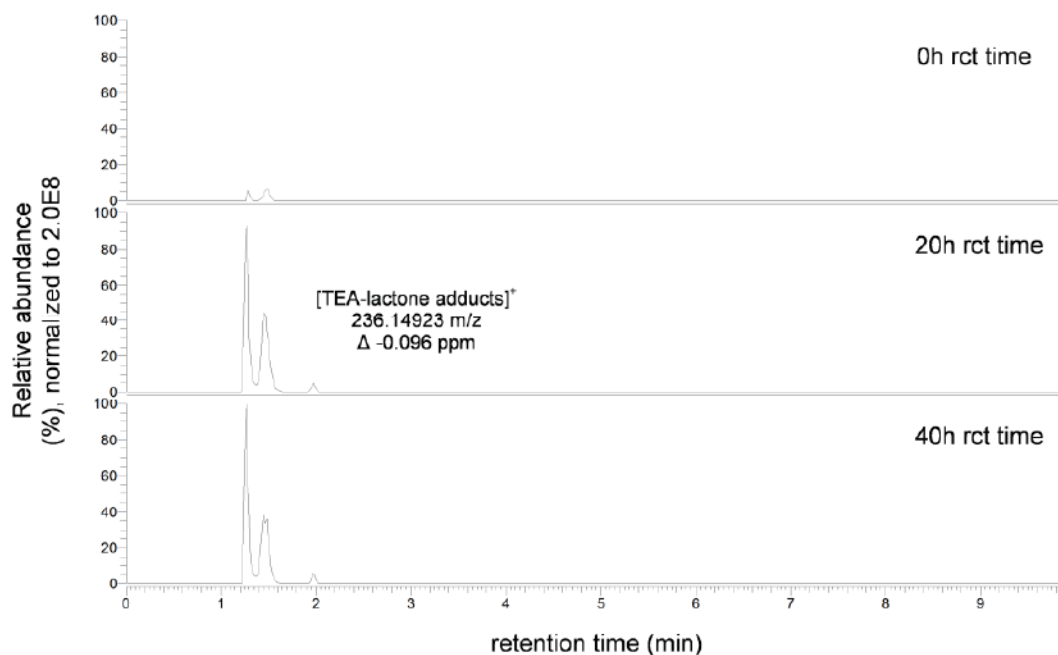
**Figure S22:** Mass range  $[185.02002\text{ m/z} - 185.02188\text{ m/z}]$  of full-scan  $[50-600\text{ m/z}]$  reversed-phase HPLC-ESI-HRMS (positive polarity) traces of the reaction of BBL in an aqueous phosphate-buffered solution after 0 h, 20 h, and 40 h.

### 3.2.3. Reaction of BBL with triethanolamine buffer (TEA buffer)



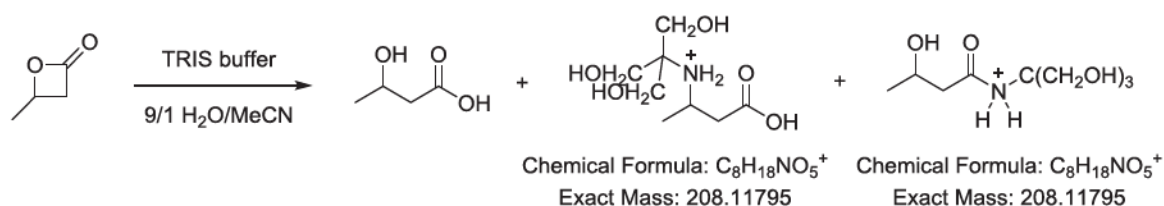
To a solution of TEA (946 mg, 6.0 mmol) in a 9/1 (v/v)  $H_2O/CH_3CN$ -mixture (60 mL) with pH 7.75 was added BBL (122  $\mu$ L, 129 mg, 1.5 mmol). The reaction mixture was stirred at ambient temperature and samples thereof were analyzed by HPLC-ESI-HRMS after 0, 20, and 40 h.

HRMS (ESI): calcd. for  $C_{10}H_{22}NO_5 [M]^+$ : 236.14925; found: 236.14923.



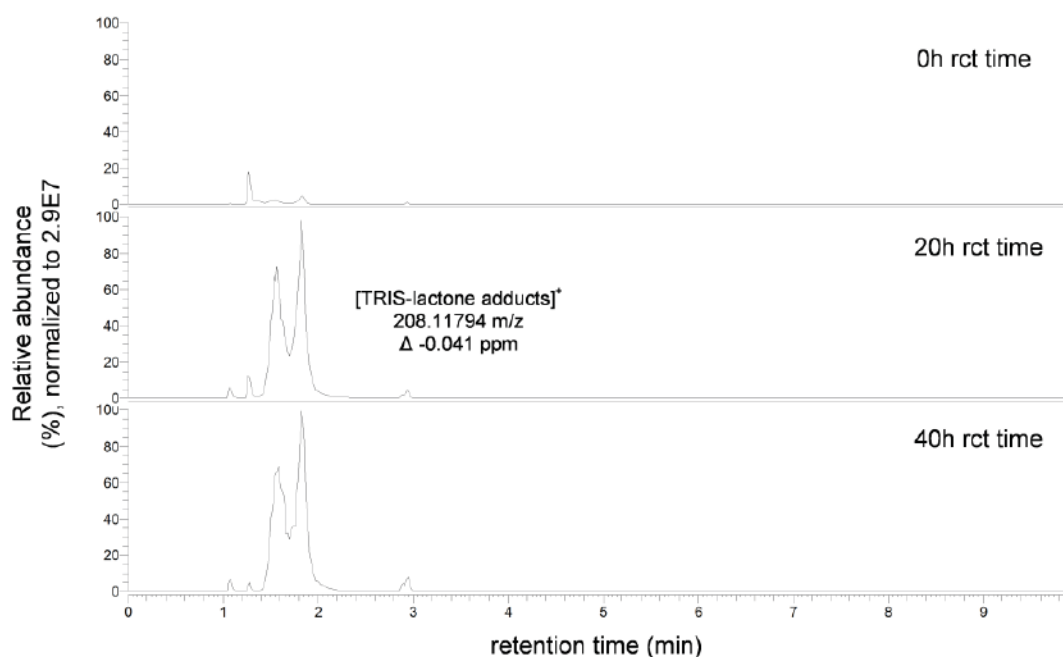
**Figure S23:** Mass range [236.13882 m/z – 236.14118 m/z] of full-scan [50–600 m/z] reversed-phase HPLC-ESI-HRMS (positive polarity) traces of the reaction of BBL in an aqueous TEA-buffered solution after 0 h, 20 h, and 40 h.

### 3.2.4. Reaction of BBL with tris(hydroxymethyl)aminomethane buffer (TRIS buffer)



To a solution of TRIS (796  $\mu$ L, 895 mg, 6.0 mmol) in a 9/1 (v/v)  $H_2O/CH_3CN$ -mixture (60 mL) with pH 8.10 was added BBL (122  $\mu$ L, 129 mg, 1.5 mmol). The reaction mixture was stirred at ambient temperature and samples thereof were analyzed by HPLC-ESI-HRMS after 0, 20, and 40 h.

HRMS (ESI): calcd. for  $C_8H_{18}NO_5$   $[M+H]^+$ : 208.11795; found: 208.11794.



**Figure S24:** Mass range  $[208.11691\text{ m/z} - 208.11899\text{ m/z}]$  of full-scan  $[50-600\text{ m/z}]$  reversed-phase HPLC-ESI-HRMS (positive polarity) traces of the reaction of BBL in an aqueous TRIS-buffered solution after 0 h, 20 h, and 40 h.

#### 4. Kinetics of the reaction of BBL with amines in TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v) – Determination of second-order rate constants $k_N$

##### General procedure

A TEA-buffered (pH 7.75) solution of an amine (0.056 M) in D<sub>2</sub>O (540  $\mu$ L) was mixed with a solution of BBL (60  $\mu$ L of a 0.25 M solution in d<sub>6</sub>-DMSO) in an NMR tube at 37 °C to produce a 9/1 (v/v) D<sub>2</sub>O/d<sub>6</sub>-DMSO solvent mixture. The conversion of BBL was monitored by using time-resolved <sup>1</sup>H NMR spectroscopy. Concentrations of BBL, [BBL]<sub>t</sub>, were calculated from the ratio of the integral of the methyl resonances of BBL (d,  $\delta$  = 1.70 ppm) to the integral of the methyl resonances in the range  $\delta$  = 1.43–1.27 ppm.

The first-order rate constants  $k_{\text{obs}}$  ( $= k_2$  in equation (4) of the main text) were determined by least-squares fitting of the exponential function  $[\text{BBL}]_t = [\text{BBL}]_0 \cdot \exp(-k_{\text{obs}}t) + C$  to the time-dependent [BBL]<sub>t</sub>.

$$k_{\text{obs}} = k_2 = k_1 + k_N[\text{amine}]_{\text{eff}} \quad (4)$$

According to equation (5) of the main text, i.e.

$$k_N = (k_2 - k_1)/[\text{amine}]_{\text{eff}} \quad (5),$$

$k_N$  is then calculated from the experimentally determined  $k_2$ , the first-order rate constant of the BBL decay in TEA-buffered solution ( $k_1 = 5.47 \times 10^{-5} \text{ s}^{-1}$ , see Section 3.1.5) and the effective concentration of free amine at pH 7.75, [amine]<sub>eff</sub>, which is calculated according to Henderson Hasselbalch equation with  $\text{p}K_{\text{aH}^+}$  values from ref <sup>[13]</sup>.

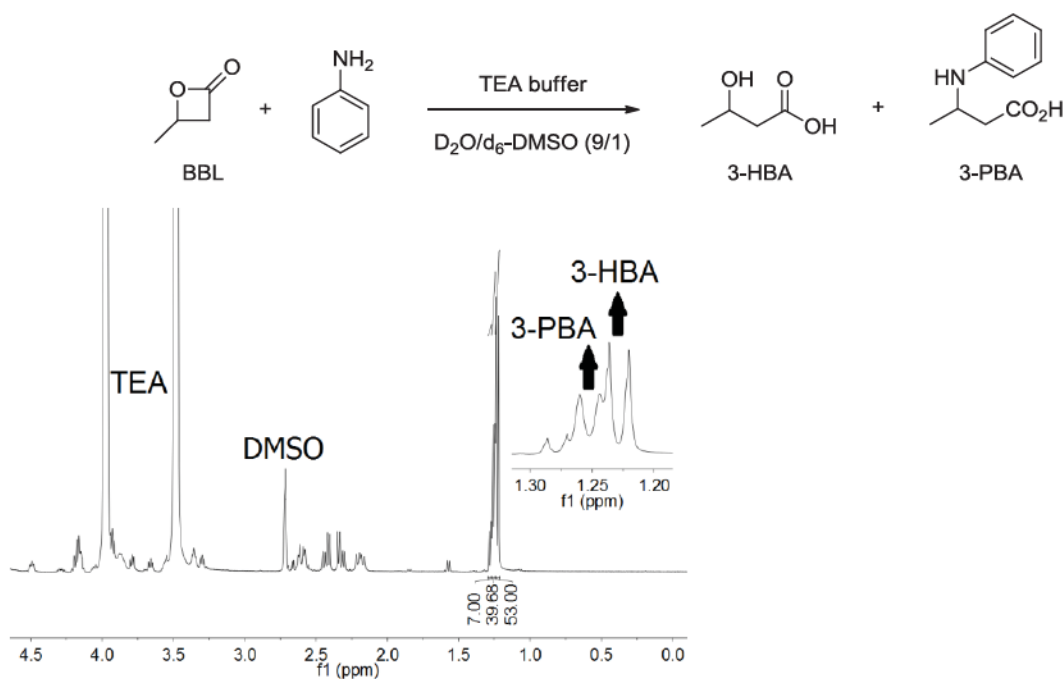
#### 4.1. Kinetics of the reaction of BBL with aniline in TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v)

##### 4.1.1. Reaction of BBL with [aniline]<sub>0</sub> = 50 mM

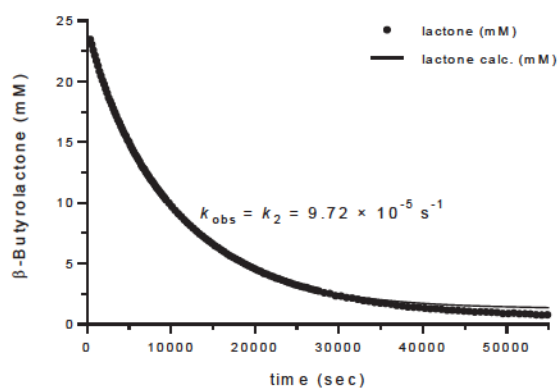
[BBL]<sub>0</sub> = 25 mM,

[PhNH<sub>2</sub>]<sub>0</sub> = 50 mM, [PhNH<sub>2</sub>]<sub>eff</sub> = 50 mM,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C.



**Figure S25:** <sup>1</sup>H NMR spectrum (400 MHz) of the crude reaction mixture after completion of the reaction of BBL with aniline in a TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v) solution.



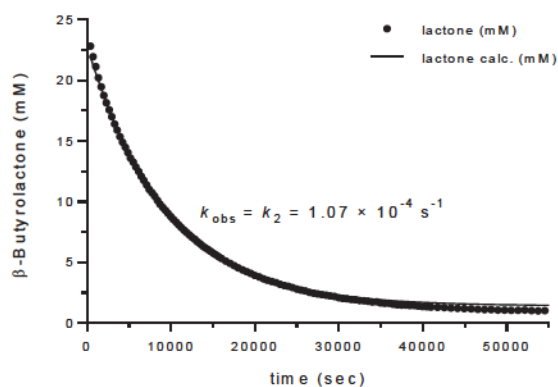
**Figure S26:** Decay of the concentration of BBL while reacting with aniline in a TEA-buffered reaction mixture at 37 °C ([aniline]<sub>0</sub> = 50 mM,  $k_{\text{N}} = 8.50 \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

#### 4.1.2. Reaction of BBL with [aniline]<sub>0</sub> = 74 mM

[BBL]<sub>0</sub> = 25 mM,

[PhNH<sub>2</sub>]<sub>0</sub> = 74 mM, [PhNH<sub>2</sub>]<sub>eff</sub> = 74 mM,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C.



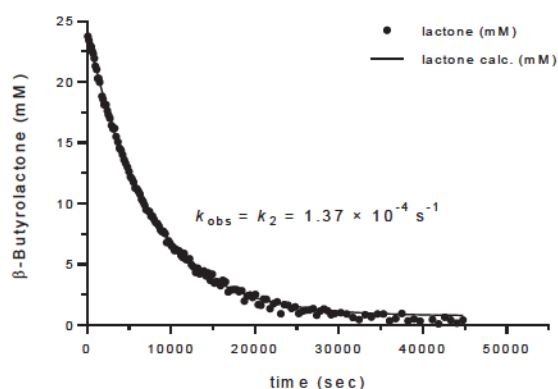
**Figure S27:** Decay of the concentration of BBL while reacting with aniline in a TEA-buffered reaction mixture at 37 °C ([aniline]<sub>0</sub> = 74 mM,  $k_{\text{N}} = 7.07 \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

#### 4.1.3. Reaction of BBL with [aniline]<sub>0</sub> = 0.10 M

[BBL]<sub>0</sub> = 25 mM,

[PhNH<sub>2</sub>]<sub>0</sub> = 0.10 M, [PhNH<sub>2</sub>]<sub>eff</sub> = 0.10 M,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C.



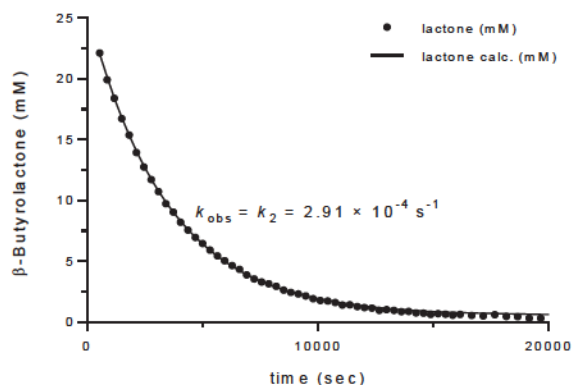
**Figure S28:** Decay of the concentration of BBL while reacting with aniline in a TEA-buffered reaction mixture at 37 °C ([aniline]<sub>0</sub> = 0.10 M,  $k_{\text{N}} = 8.23 \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

#### 4.1.4. Reaction of BBL with [aniline]<sub>0</sub> = 0.30 M

[BBL]<sub>0</sub> = 25 mM,

[PhNH<sub>2</sub>]<sub>0</sub> = 0.30 M, [PhNH<sub>2</sub>]<sub>eff</sub> = 0.30 M,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C.



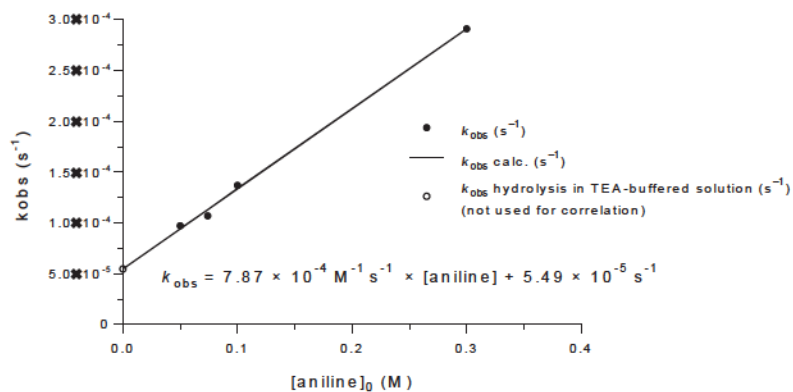
**Figure S29:** Decay of the concentration of BBL while reacting with aniline in a TEA-buffered reaction mixture at 37 °C ([aniline]<sub>0</sub> = 0.30 M,  $k_N = 7.88 \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

#### 4.1.5. Determination of the second-order rate constant $k_N$ for the reaction of BBL with aniline

**Table S4:** Rate constants for the reactions of BBL with aniline at 37 °C at pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v) at variable aniline concentrations.

[aniline] <sub>0</sub> (M)	$k_{\text{obs}}$ (s <sup>-1</sup> )	$k_N$ (M <sup>-1</sup> s <sup>-1</sup> )
$5.0 \times 10^{-2}$	$9.72 \times 10^{-5}$	$8.50 \times 10^{-4}$
$7.4 \times 10^{-2}$	$1.07 \times 10^{-4}$	$7.07 \times 10^{-4}$
$1.0 \times 10^{-1}$	$1.37 \times 10^{-4}$	$8.23 \times 10^{-4}$
$3.0 \times 10^{-1}$	$2.91 \times 10^{-4}$	$7.88 \times 10^{-4}$
averaged $k_N$		$7.92 \times 10^{-4}$
Standard dev		$\pm 5.4 \times 10^{-5}$

The averaged  $k_N$  is in excellent agreement with the  $k_N$  derived from the slope of the linear correlation of  $k_{\text{obs}}$  with [aniline]<sub>0</sub> shown in Figure S30.



**Figure S30:** Corelation of  $k_{obs}$  with  $[aniline]_0$ . For comparison also  $k_{obs}$  of hydrolysis in TEA-buffered solution is plotted.

The average value  $k_N(aniline) = 7.92 (\pm 0.54) \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$  is used in Table 3 of the main text.

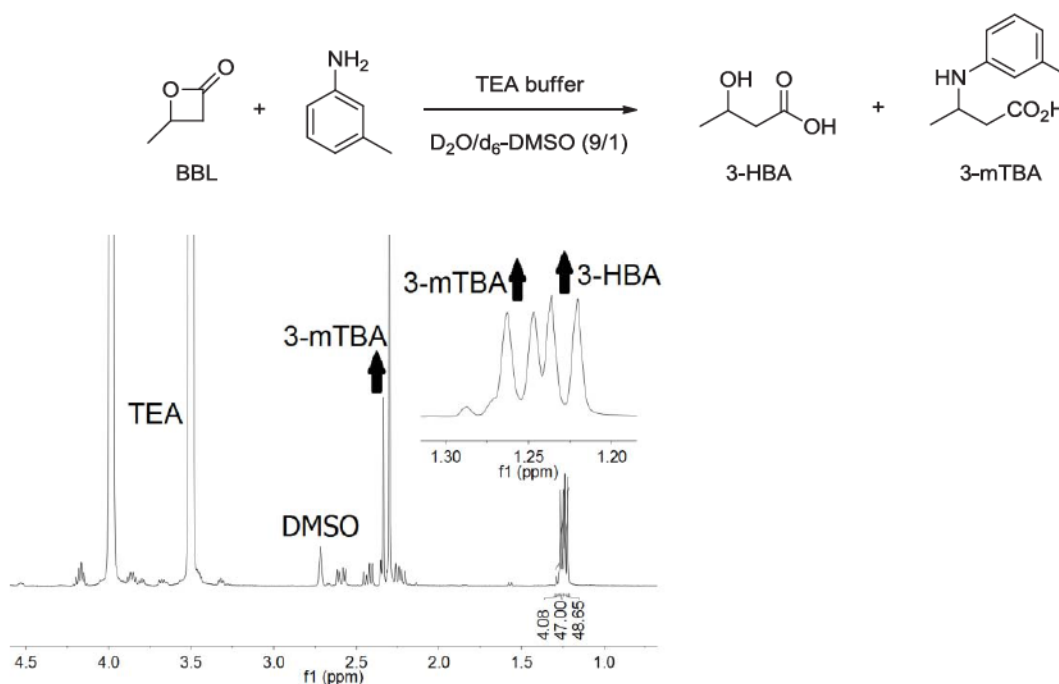


## 4.2. Kinetics of the reaction of BBL with *m*-toluidine in TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v)

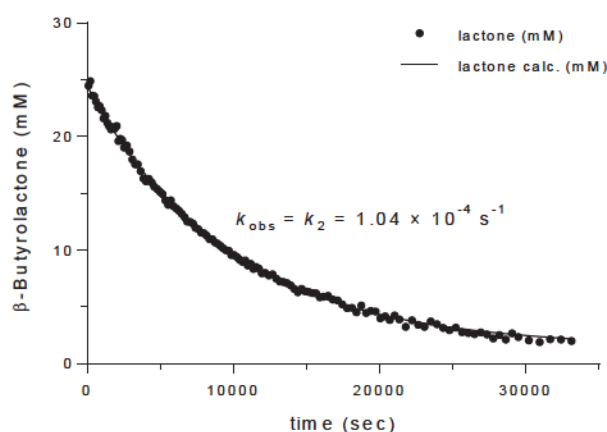
[BBL]<sub>0</sub> = 25 mM,

[*m*-tol-NH<sub>2</sub>]<sub>0</sub> = 50 mM, [*m*-tol-NH<sub>2</sub>]<sub>eff</sub> = 50 mM,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C.



**Figure S31:** <sup>1</sup>H NMR spectrum (400 MHz) of the crude reaction mixture after completion of the reaction of BBL with *m*-toluidine in a TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v) solution.



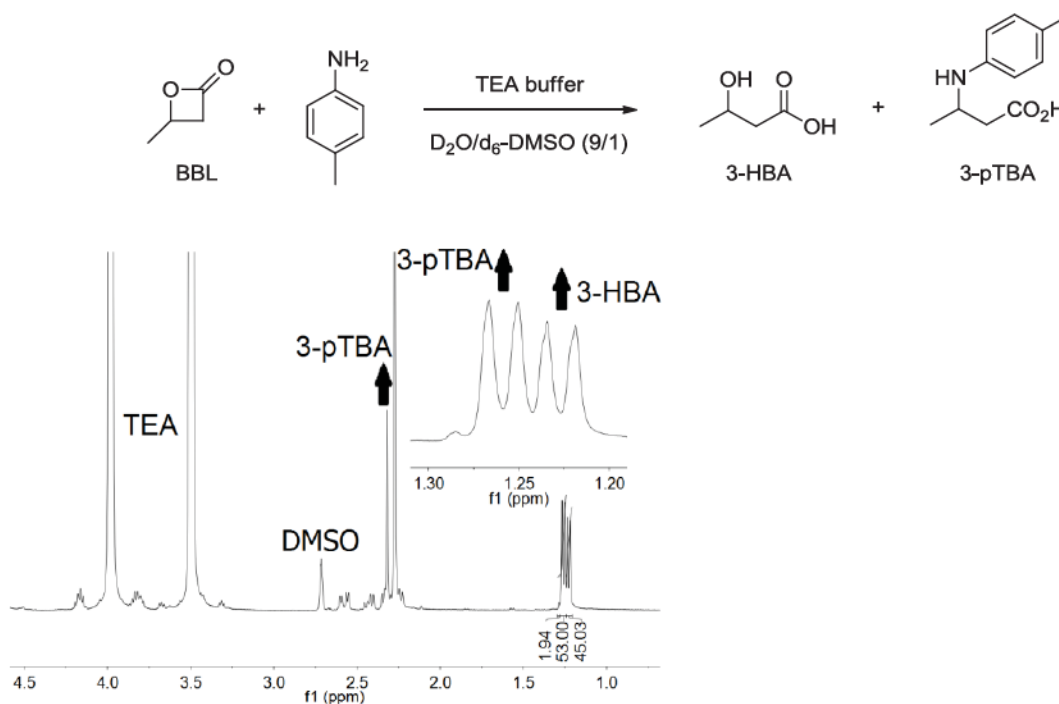
**Figure S32:** Decay of the concentration of BBL while reacting with *m*-toluidine in a TEA-buffered reaction mixture at 37 °C ( $k_N = 9.90 \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

### 4.3. Kinetics of the reaction of BBL with *p*-toluidine in TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v)

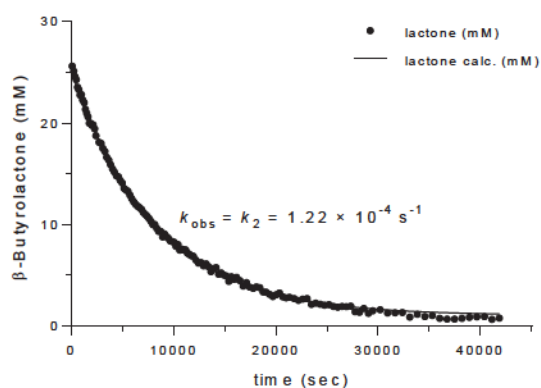
[BBL]<sub>0</sub> = 25 mM,

[*p*-tol-NH<sub>2</sub>]<sub>0</sub> = 50 mM, [*p*-tol-NH<sub>2</sub>]<sub>eff</sub> = 50 mM,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C.



**Figure S33:** <sup>1</sup>H NMR spectrum (400 MHz) of the crude reaction mixture after completion of the reaction of BBL with *p*-toluidine in a TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v) solution.



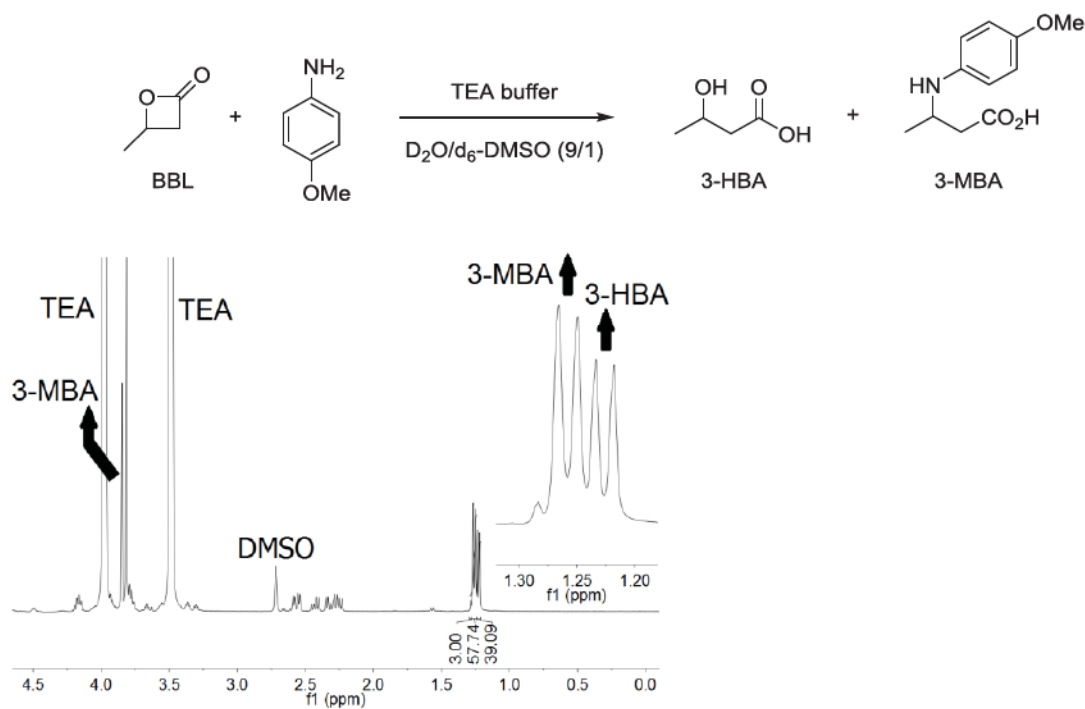
**Figure S34:** Decay of the concentration of BBL while reacting with *p*-toluidine in a TEA-buffered reaction mixture at 37 °C ( $k_{\text{N}} = 1.35 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

#### 4.4. Kinetics of the reaction of BBL with *p*-anisidine in TEA-buffered D<sub>2</sub>O/*d*<sub>6</sub>-DMSO 9/1 (v/v)

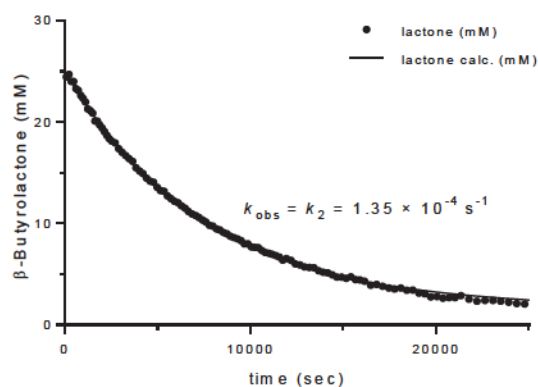
[BBL]<sub>0</sub> = 25 mM,

[*p*-ani-NH<sub>2</sub>]<sub>0</sub> = 50 mM, [*p*-ani-NH<sub>2</sub>]<sub>eff</sub> = 50 mM,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/*d*<sub>6</sub>-DMSO 9/1 (v/v), 37 °C.



**Figure S35:** <sup>1</sup>H NMR spectrum (400 MHz) of the crude reaction mixture after completion of the reaction of BBL with *p*-anisidine in a TEA-buffered D<sub>2</sub>O/*d*<sub>6</sub>-DMSO 9/1 (v/v) solution.



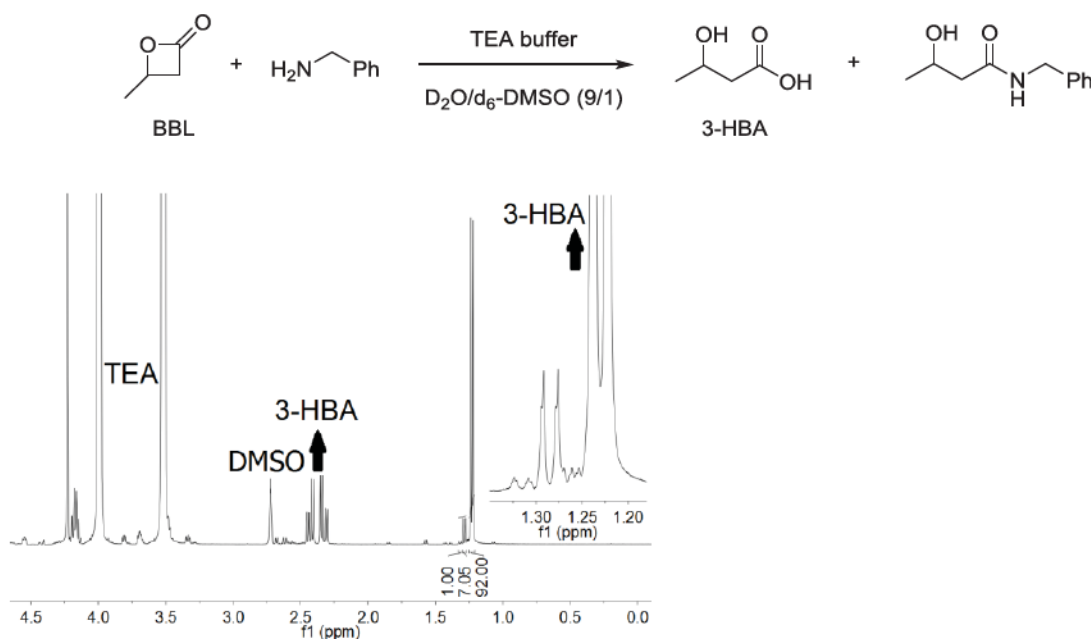
**Figure S36:** Decay of the concentration of BBL while reacting with *p*-anisidine in a TEA-buffered reaction mixture at 37 °C ( $k_{\text{N}} = 1.62 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

#### 4.5. Kinetics of the reaction of BBL with benzylamine in TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v)

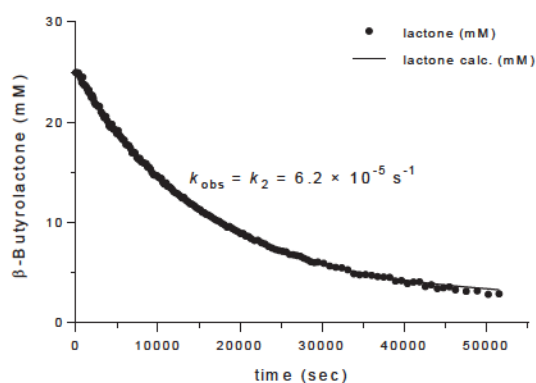
[BBL]<sub>0</sub> = 25 mM,

[PhCH<sub>2</sub>NH<sub>2</sub>]<sub>0</sub> = 50 mM, [PhCH<sub>2</sub>NH<sub>2</sub>]<sub>eff</sub> = 1.2 mM,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C.



**Figure S37:** <sup>1</sup>H NMR spectrum (400 MHz) of the crude reaction mixture after completion of the reaction of BBL with benzylamine in a TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v) solution.



**Figure S38:** Decay of the concentration of BBL while reacting with benzylamine in a TEA-buffered reaction mixture at 37 °C ( $k_{\text{N}} = 5.8 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

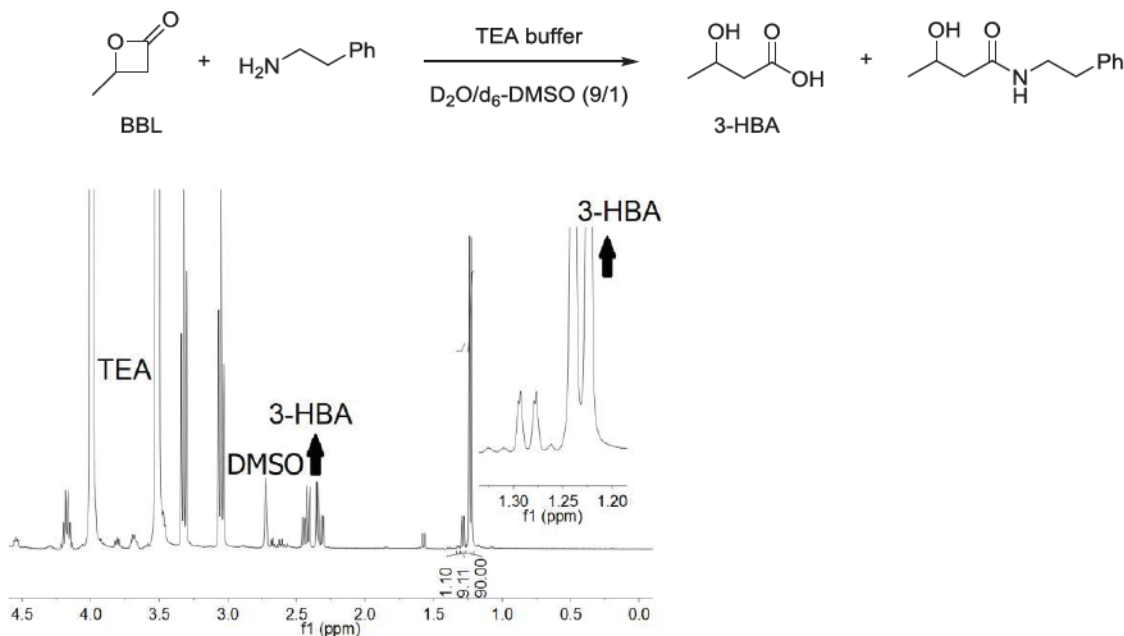
Due to the small concentration of unprotonated benzylamine under reaction conditions, we proofed the formation of *N*-benzyl-3-hydroxybutanamide by HPLC-HRMS (ESI) analysis of the reaction mixture. Calcd. for C<sub>11</sub>H<sub>16</sub>NO<sub>2</sub> [M+H]<sup>+</sup> : *m/z* = 194.11756; found: *m/z* = 194.11753.

#### 4.6. Kinetics of the reaction of BBL with phenylethylamine in TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v)

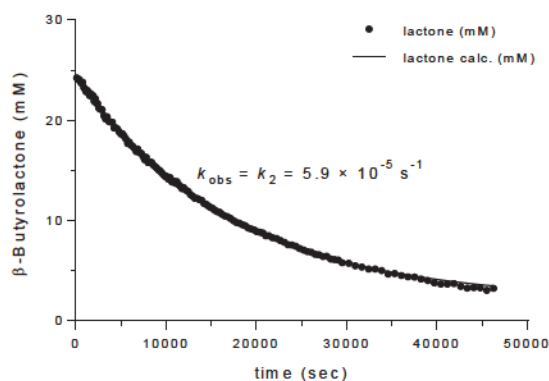
[BBL]<sub>0</sub> = 25 mM,

[PhCH<sub>2</sub>CH<sub>2</sub>NH<sub>2</sub>]<sub>0</sub> = 50 mM, [PhCH<sub>2</sub>CH<sub>2</sub>NH<sub>2</sub>]<sub>eff</sub> = 0.36 mM,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C.



**Figure S39:** <sup>1</sup>H NMR spectrum (400 MHz) of the crude reaction mixture after completion of the reaction of BBL with phenylethylamine in a TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v) solution.



**Figure S40:** Decay of the concentration of BBL while reacting with phenylethylamine in a TEA-buffered reaction mixture at 37 °C ( $k_N = 1.1 \times 10^{-2} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

Due to the small concentration of unprotonated phenylethylamine under reaction conditions, we proofed the formation of 3-hydroxy-*N*-phenethylbutanamide by HPLC-HRMS (ESI) analysis of the reaction mixture. Calcd. for C<sub>12</sub>H<sub>18</sub>NO<sub>2</sub> [M+H]<sup>+</sup>:  $m/z = 208.13321$ ; found:  $m/z = 208.13317$ .

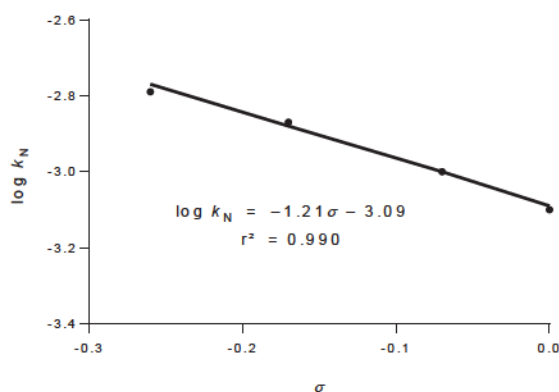
## 4.7. Rate constants for the reactions of BBL with amines

**Table S5:** Rate constants for the reactions of BBL with different amines at 37 °C and pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v).

	aniline	<i>m</i> -toluidine	<i>p</i> -toluidine	<i>p</i> -anisidine	benzyl- amine	phenylethyl- amine
$pK_{aH^+}$	4.60	4.87	5.08	5.36	9.36	9.89
$c_{\text{eff}}$ (M)	-	$5.0 \times 10^{-2}$	$5.0 \times 10^{-2}$	$5.0 \times 10^{-2}$	$1.2 \times 10^{-3}$	$3.6 \times 10^{-4}$
$k_{\text{obs}}$ (s <sup>-1</sup> )	-	$1.04 \times 10^{-4}$	$1.22 \times 10^{-4}$	$1.35 \times 10^{-4}$	$6.2 \times 10^{-5}$	$5.9 \times 10^{-5}$
$k_N$ (M <sup>-1</sup> s <sup>-1</sup> )	$7.92 \times 10^{-4}$ [a]	$9.90 \times 10^{-4}$	$1.35 \times 10^{-3}$	$1.62 \times 10^{-3}$	$5.8 \times 10^{-3}$	$1.1 \times 10^{-2}$
$\sigma_m$ or $\sigma_p$ <sup>[b]</sup>	0	-0.07	-0.17	-0.26	-	-

[a] Average value, see Section 4.1. [b] From: C. Hansch, A. Leo, R. W. Taft, *Chem. Rev.* **1991**, 91, 165–195.

## 4.8. Hammett plot for the reactions of BBL with anilines

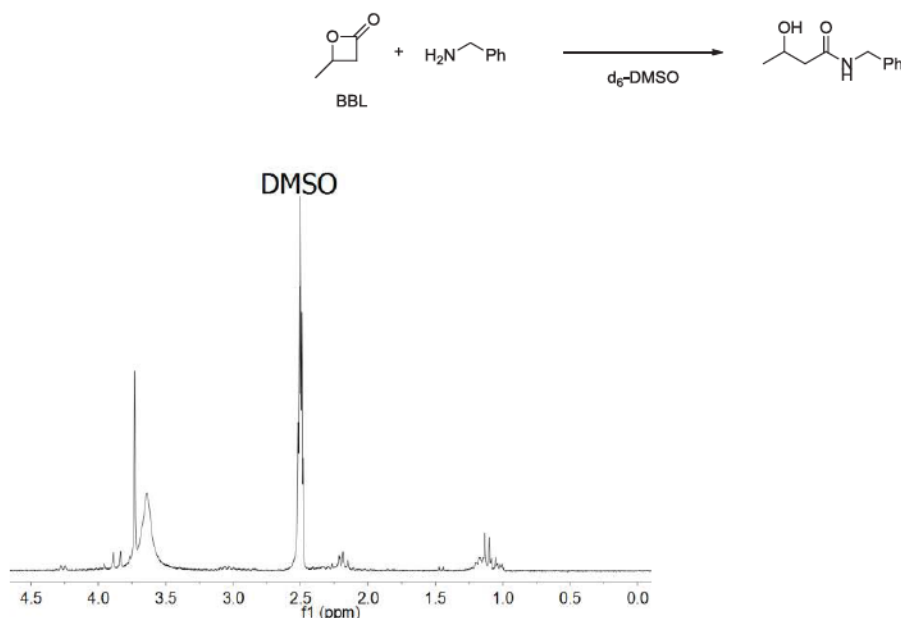


**Figure S41:** Hammett plot of the reaction of BBL with anilines at 37 °C and pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), with data from Table S5.

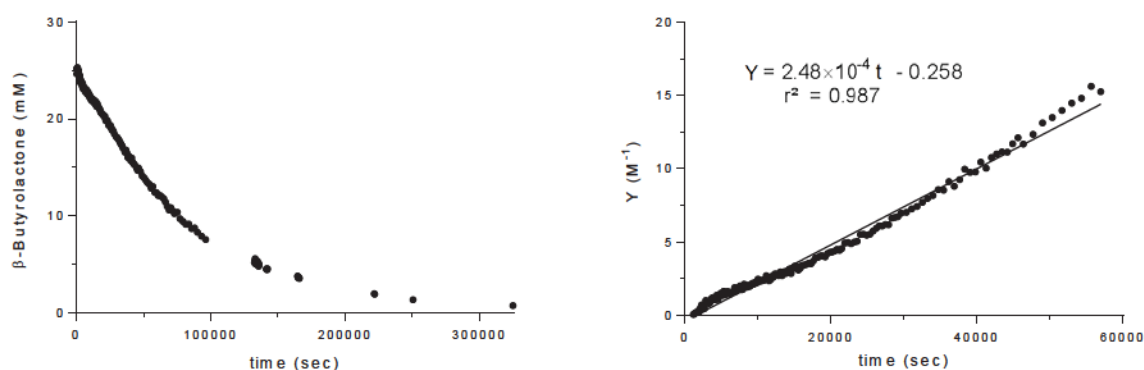
## 5. Kinetics of the reaction of BBL with benzylamine in d<sub>6</sub>-DMSO

[BBL]<sub>0</sub> = 25 mM,

[PhCH<sub>2</sub>NH<sub>2</sub>]<sub>0</sub> = 50 mM, in d<sub>6</sub>-DMSO, 37 °C.



**Figure S42:** <sup>1</sup>H NMR spectrum (200 MHz) of the crude reaction mixture after completion of the reaction of BBL with benzylamine d<sub>6</sub>-DMSO solution.



**Figure S43:** Decay of the concentration of BBL while reacting with benzylamine d<sub>6</sub>-DMSO at 37 °C (left). Determination of the second-order rate constant by plotting time versus  $Y = ([Nu]_0 - [E]_0)^{-1} \ln([E]_0([E]_t + [Nu]_0 - [E]_0) / [Nu]_0[E]_t)$  ( $k_N = 2.5 \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$ , evaluated from data for the first half-life; right).

## 6. Determination of second-order rate constants of the reaction of BBL with carbanions in DMSO

### General Procedure

Solutions of the potassium salts of the carbanions **9–12** were generated by deprotonation of the corresponding CH acid with 1.05 equivalents of KO<sup>t</sup>Bu in dry DMSO prior to the measurements. The kinetics of the reactions of BBL with the carbanions **9–12** were studied in DMSO at 20 °C by using stopped-flow techniques with UV-vis detection. First-order conditions were achieved by using a large excess of BBL, which led to a mono-exponential decrease of the absorptions of the carbanions (*A*) during their reactions with BBL, from which the first-order rate constants  $k_{\text{obs}}$  (s<sup>-1</sup>) were obtained by a least-squares fitting of the function  $A_t = A_0 \cdot \exp(-k_{\text{obs}}t) + C$ . The second-order rate constants  $k_2$  (M<sup>-1</sup> s<sup>-1</sup>) were then obtained from the slope of a plot of  $k_{\text{obs}}$  versus [BBL]<sub>0</sub>.

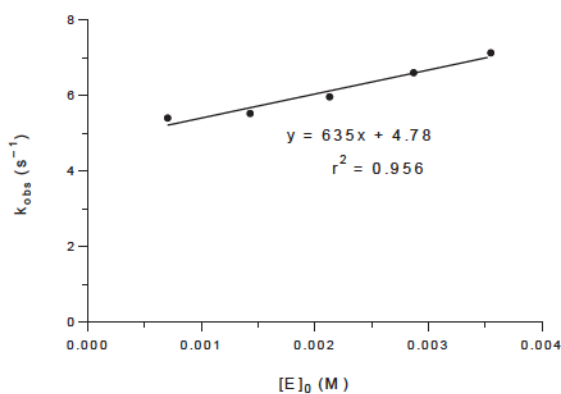


**Table S6:** Reaction of BBL with potassium cyano(4-nitrophenyl)methanide **9** (at 20 °C, stopped-flow, detection at 539 nm).

E = BBL; Nu = potassium cyano(4-nitrophenyl)methanide (**9**)

[Nu] <sub>0</sub> (M)	[E] <sub>0</sub> (M)	[E] <sub>0</sub> /[Nu] <sub>0</sub>	<i>k</i> <sub>obs</sub> (s <sup>-1</sup> )
3.56 × 10 <sup>-5</sup>	7.04 × 10 <sup>-4</sup>	20	5.41
3.56 × 10 <sup>-5</sup>	1.43 × 10 <sup>-3</sup>	40	5.53
3.56 × 10 <sup>-5</sup>	2.13 × 10 <sup>-3</sup>	60	5.97
3.56 × 10 <sup>-5</sup>	2.87 × 10 <sup>-3</sup>	81	6.61
3.56 × 10 <sup>-5</sup>	3.55 × 10 <sup>-3</sup>	100	7.14

$$k_2 (20\text{ °C}) = 635\text{ M}^{-1}\text{ s}^{-1}$$



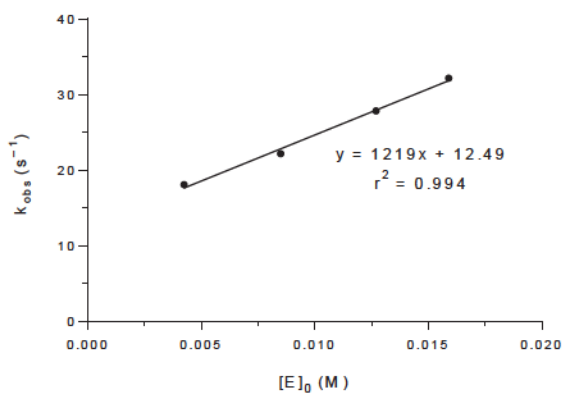
**Figure S44:** Dependence of the first-order rate constant *k*<sub>obs</sub> of the reaction of **9** with BBL at 20 °C on the concentration of BBL.

**Table S7:** Reaction of BBL with potassium cyano(4-nitrophenyl)methanide **9** (at 30 °C, stopped-flow, detection at 539 nm).

E = BBL; Nu = potassium cyano(4-nitrophenyl)methanide (**9**)

[Nu] <sub>0</sub> (M)	[E] <sub>0</sub> (M)	[E] <sub>0</sub> /[Nu] <sub>0</sub>	<i>k</i> <sub>obs</sub> (s <sup>-1</sup> )
6.20 × 10 <sup>-5</sup>	4.24 × 10 <sup>-3</sup>	68	18.1
6.20 × 10 <sup>-5</sup>	8.49 × 10 <sup>-2</sup>	137	22.2
6.20 × 10 <sup>-5</sup>	1.27 × 10 <sup>-2</sup>	205	27.9
6.20 × 10 <sup>-5</sup>	1.59 × 10 <sup>-2</sup>	257	32.2

$$k_2 (30\text{ °C}) = 1.22 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$$



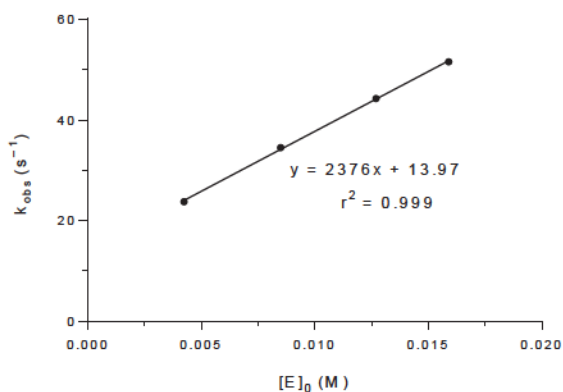
**Figure S45:** Dependence of the first-order rate constant *k*<sub>obs</sub> of the reaction of **9** with BBL at 30 °C on the concentration of BBL.

**Table S8:** Reaction of BBL with potassium cyano(4-nitrophenyl)methanide **9** (at 40 °C, stopped-flow, detection at 539 nm).

E = BBL; Nu = potassium cyano(4-nitrophenyl)methanide (**9**)

[Nu] <sub>0</sub> (M)	[E] <sub>0</sub> (M)	[E] <sub>0</sub> /[Nu] <sub>0</sub>	<i>k</i> <sub>obs</sub> (s <sup>-1</sup> )
6.20 × 10 <sup>-5</sup>	4.24 × 10 <sup>-3</sup>	68	23.8
6.20 × 10 <sup>-5</sup>	8.49 × 10 <sup>-2</sup>	137	34.5
6.20 × 10 <sup>-5</sup>	1.27 × 10 <sup>-2</sup>	205	44.3
6.20 × 10 <sup>-5</sup>	1.59 × 10 <sup>-2</sup>	257	51.6

$$k_2(40\text{ °C}) = 2.38 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$$



**Figure S46:** Dependence of the first-order rate constant *k*<sub>obs</sub> of the reaction of **9** with BBL at 40 °C on the concentration of BBL.

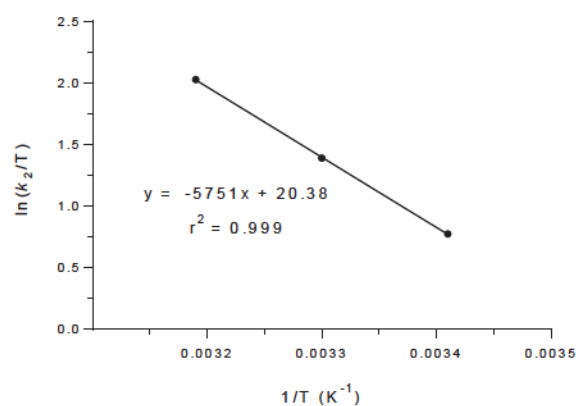
**Table S9:** Rate constants of the reaction of BBL with potassium cyano(4-nitrophenyl)methanide (**9**) at different temperatures.

T (°C)	$k_2$ (M <sup>-1</sup> s <sup>-1</sup> )	1/T (K <sup>-1</sup> )	ln( $k_2/T$ )
20	$6.35 \times 10^2$	$3.41 \times 10^{-3}$	$7.73 \times 10^{-1}$
30	$1.22 \times 10^3$	$3.30 \times 10^{-3}$	1.39
40	$2.38 \times 10^3$	$3.19 \times 10^{-3}$	2.03

$$\Delta H^\ddagger = 47.9 \text{ kJ mol}^{-1}$$

$$\Delta S^\ddagger = -27.9 \text{ J mol}^{-1} \text{ K}^{-1}$$

$$k_2 (37^\circ \text{C}) = 1.95 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$$



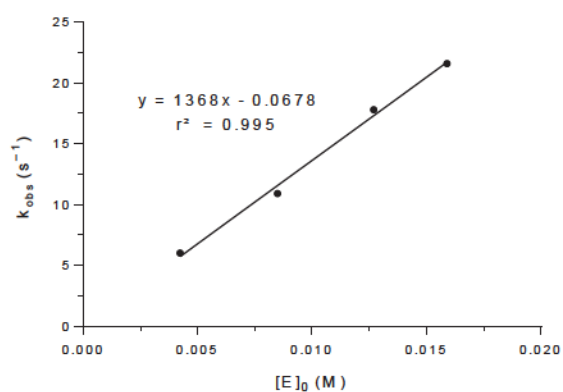
**Figure S47:** Eyring plot for the reactions of **9** with BBL at different temperatures.

**Table S10:** Reaction of BBL with potassium 2-ethoxy-1-(4-nitrophenyl)-2-oxoethan-1-ide **10** (at 20 °C, stopped-flow, detection at 550 nm).

E = BBL; Nu = potassium 2-ethoxy-1-(4-nitrophenyl)-2-oxoethan-1-ide (**10**)

[Nu] <sub>0</sub> (M)	[E] <sub>0</sub> (M)	[E] <sub>0</sub> /[Nu] <sub>0</sub>	<i>k</i> <sub>obs</sub> (s <sup>-1</sup> )
5.17 × 10 <sup>-5</sup>	4.24 × 10 <sup>-3</sup>	82	6.02
5.17 × 10 <sup>-5</sup>	8.49 × 10 <sup>-2</sup>	164	10.9
5.17 × 10 <sup>-5</sup>	1.27 × 10 <sup>-2</sup>	246	17.8
5.17 × 10 <sup>-5</sup>	1.59 × 10 <sup>-2</sup>	308	21.6

$$k_2 (20\text{ }^{\circ}\text{C}) = 1.37 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$$



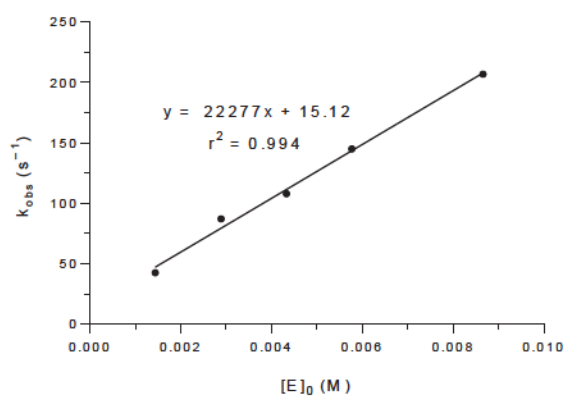
**Figure S48:** Dependence of the reaction of **10** with BBL at 20 °C on the concentration of BBL.

**Table S11:** Reaction of BBL with potassium 1-(4-cyanophenyl)-2-ethoxy-2-oxoethan-1-ide **11** (at 20 °C, stopped-flow, detection at 405 nm).

E = BBL; Nu = potassium 1-(4-cyanophenyl)-2-ethoxy-2-oxoethan-1-ide (**11**)

[Nu] <sub>0</sub> (M)	[E] <sub>0</sub> (M)	[E] <sub>0</sub> /[Nu] <sub>0</sub>	<i>k</i> <sub>obs</sub> (s <sup>-1</sup> )
1.04 × 10 <sup>-4</sup>	1.44 × 10 <sup>-3</sup>	14	42.6
1.04 × 10 <sup>-4</sup>	2.89 × 10 <sup>-3</sup>	28	87.2
1.04 × 10 <sup>-4</sup>	4.33 × 10 <sup>-3</sup>	42	108
1.04 × 10 <sup>-4</sup>	5.77 × 10 <sup>-3</sup>	56	145
1.04 × 10 <sup>-4</sup>	8.66 × 10 <sup>-3</sup>	84	207

$$k_2 (20\text{ °C}) = 2.23 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$$



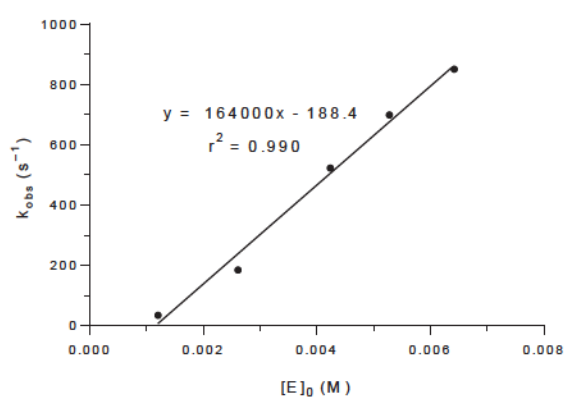
**Figure S49:** Dependence of the first-order rate constant *k*<sub>obs</sub> of the reaction of **11** with BBL at 20 °C on the concentration of BBL.

**Table S12:** Reaction of BBL with potassium 1-(4-bromophenyl)-2-ethoxy-2-oxoethan-1-ide **12** (at 20 °C, stopped-flow, detection at 352 nm).

E = BBL; Nu = potassium 1-(4-bromophenyl)-2-ethoxy-2-oxoethan-1-ide (**12**)

[Nu] <sub>0</sub> (M)	[E] <sub>0</sub> (M)	[E] <sub>0</sub> /[Nu] <sub>0</sub>	<i>k</i> <sub>obs</sub> (s <sup>-1</sup> )
2.43 × 10 <sup>-4</sup>	1.20 × 10 <sup>-3</sup>	5	35.5
2.43 × 10 <sup>-4</sup>	2.61 × 10 <sup>-3</sup>	11	185
2.43 × 10 <sup>-4</sup>	4.24 × 10 <sup>-3</sup>	17	523
2.43 × 10 <sup>-4</sup>	5.28 × 10 <sup>-3</sup>	21	699
2.43 × 10 <sup>-4</sup>	6.42 × 10 <sup>-3</sup>	26	850

$$k_2 (20\text{ °C}) = 1.64 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$$



**Figure S50:** Dependence of the first-order rate constant *k*<sub>obs</sub> of the reaction of **12** with BBL at 20 °C on the concentration of BBL.

## 7. Determination of second-order rate constants of the reaction of sodium penicillin G with various amines in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v)

### General procedure

A TEA-buffered (pH 7.75) solution of an amine (0.056 M) in D<sub>2</sub>O (540  $\mu$ L) was mixed with a solution of penicillin G (60  $\mu$ L of a 0.25 M solution in d<sub>6</sub>-DMSO) in an NMR tube at 37 °C to produce a 9/1 (v/v) D<sub>2</sub>O/DMSO solvent mixture. The conversion of penicillin G (PenG) was monitored by using time-resolved <sup>1</sup>H NMR spectroscopy. Concentrations of PenG, [PenG]<sub>t</sub>, were calculated from ratio of the integral of the aromatic protons (7.63 to 7.40 ppm, m, 5 H; 15 H respectively, if the nucleophile also gave signals in this range) to all protons of the bicyclic moiety (5.67 ppm, d, *J* = 3.9 Hz, 1 H and 5.59 ppm, d, *J* = 3.8 Hz, 1 H).

The first-order rate constants  $k_{\text{obs}}$  ( $= k_2$  in equation (4) of the main text) were determined by least-squares fitting of the exponential function  $[\text{PenG}]_t = [\text{PenG}]_0 \cdot \exp(-k_{\text{obs}}t) + C$  to the time-dependent [PenG]<sub>t</sub>.

$$k_{\text{obs}} = k_2 = k_1 + k_{\text{N}}[\text{amine}]_{\text{eff}} \quad (4)$$

According to equation (5) of the main text, i.e.

$$k_{\text{N}} = (k_2 - k_1)/[\text{amine}]_{\text{eff}} \quad (5),$$

$k_{\text{N}}$  is then calculated from the experimentally determined  $k_2$ , the first-order rate constant of the PenG decay in TEA-buffered solution ( $k_1 = 3.3 \times 10^{-5} \text{ s}^{-1}$ , see Section 7.1) and the effective concentration of free amine at pH 7.75.

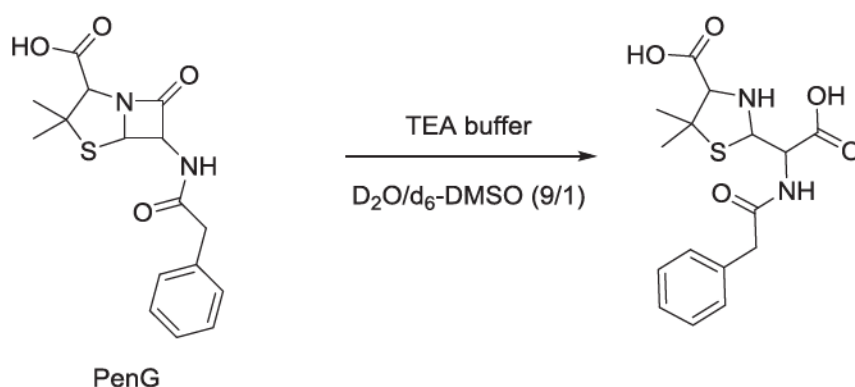


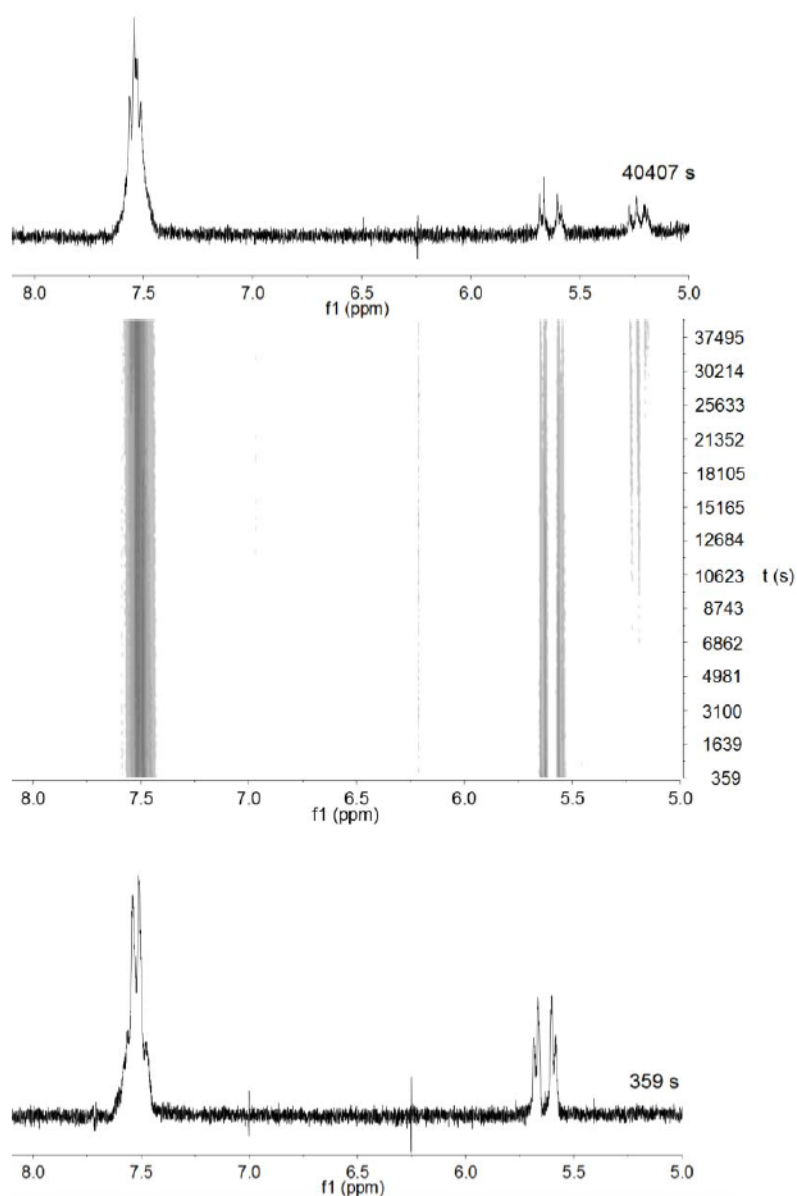
### 7.1. Consumption of sodium penicillin G in TEA-buffered solution – Determination of the first-order rate constant $k_1$

TEA buffer (540  $\mu\text{L}$  of a 0.11 M solution of TEA/TEA·HCl in  $\text{D}_2\text{O}$ , pH 7.75) and penicillin G (60  $\mu\text{L}$  of a 0.25 M solution in  $\text{d}_6$ -DMSO) were mixed in an NMR tube at 37  $^\circ\text{C}$ . The conversion of penicillin G was monitored by using time-resolved  $^1\text{H}$  NMR spectroscopy. Concentrations of penicillin G,  $[\text{PenG}]_t$ , were calculated from ratio of the integral of the aromatic protons (7.63 to 7.40 ppm, m, 5 H) to all protons of the bicyclic moiety (5.67 ppm, d,  $J = 3.9$  Hz, 1 H and 5.59 ppm, d,  $J = 3.8$  Hz, 1 H).

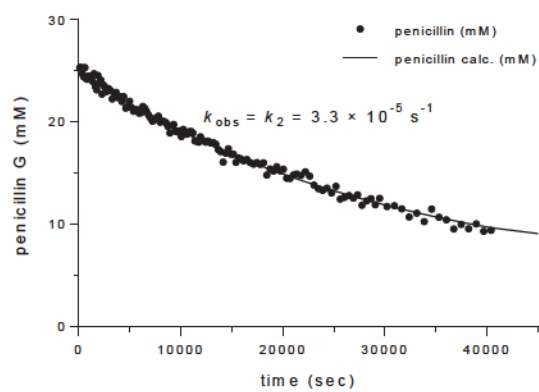
$[\text{PenG}]_0 = 25$  mM,

$[\text{TEA buffer}]_0 = 0.10$  M, pH 7.75 in  $\text{D}_2\text{O}/\text{d}_6\text{-DMSO}$  9/1 (v/v), 37  $^\circ\text{C}$





**Figure S51:** Exemplary time-resolved  $^1\text{H}$  NMR spectra of the hydrolysis of penicillin G in TEA-buffered  $\text{D}_2\text{O}/\text{d}_6\text{-DMSO}$  9/1 (v/v) at 37 °C ( $[\text{penicillin G}]_0 = 25 \text{ mM}$ ,  $\text{pH} = 7.75$ ). Start of the reaction (bottom), time-resolved topview of  $^1\text{H}$  NMR spectra showing the decrease of the followed penicillin G resonances (middle), and final  $^1\text{H}$  NMR spectrum after 40407 sec (top).



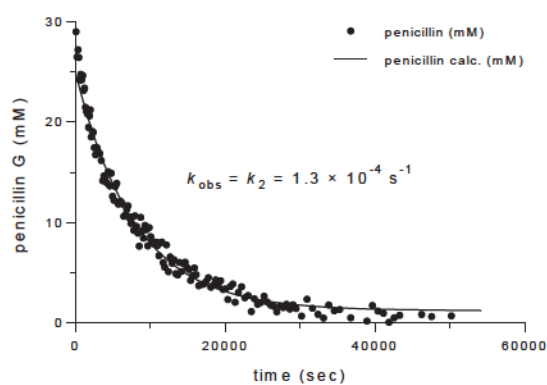
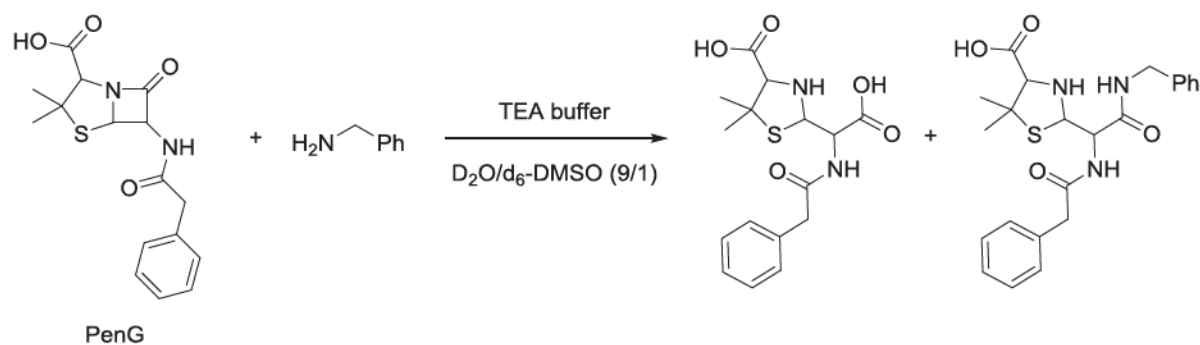
**Figure S52:** Decay of the concentration of penicillin G in a TEA-buffered reaction mixture at 37 °C. The black line shows a mono-exponential fit of the data.

## 7.2. Kinetics of the reaction of sodium penicillin G with benzylamine in TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v)

[PenG]<sub>0</sub> = 25 mM,

[PhCH<sub>2</sub>NH<sub>2</sub>]<sub>0</sub> = 50 mM, [PhCH<sub>2</sub>NH<sub>2</sub>]<sub>eff</sub> = 1.2 mM,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C



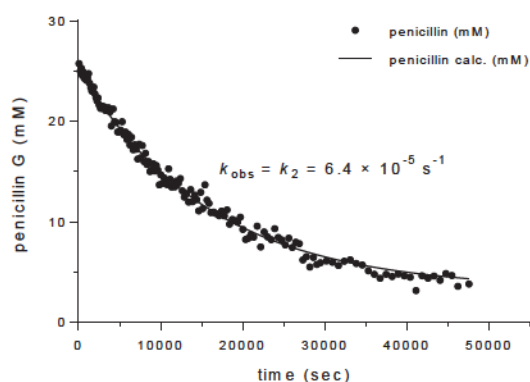
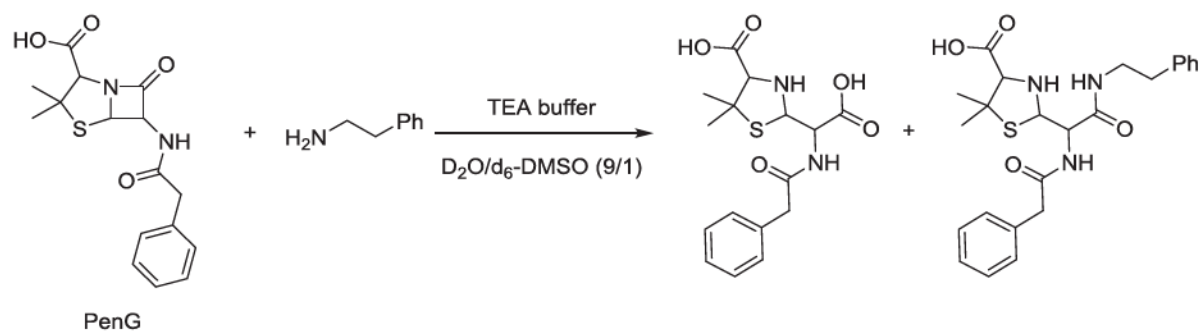
**Figure S53:** Decay of the concentration of penicillin G while reacting with benzylamine in a TEA-buffered reaction mixture at 37 °C ( $k_N = 8.1 \times 10^{-2} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

### 7.3. Kinetics of the reaction of sodium penicillin G with phenylethylamine in TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v)

[PenG]<sub>0</sub> = 25 mM,

[PhCH<sub>2</sub>CH<sub>2</sub>NH<sub>2</sub>]<sub>0</sub> = 50 mM, [PhCH<sub>2</sub>CH<sub>2</sub>NH<sub>2</sub>]<sub>eff</sub> = 0.36 mM,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C



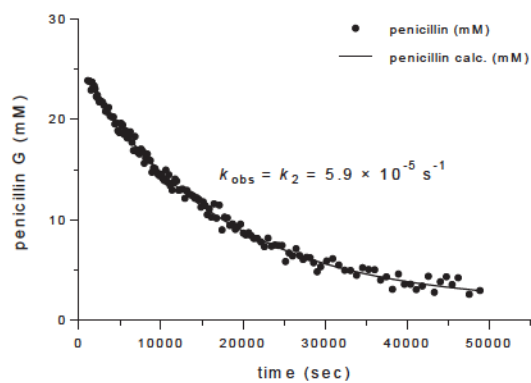
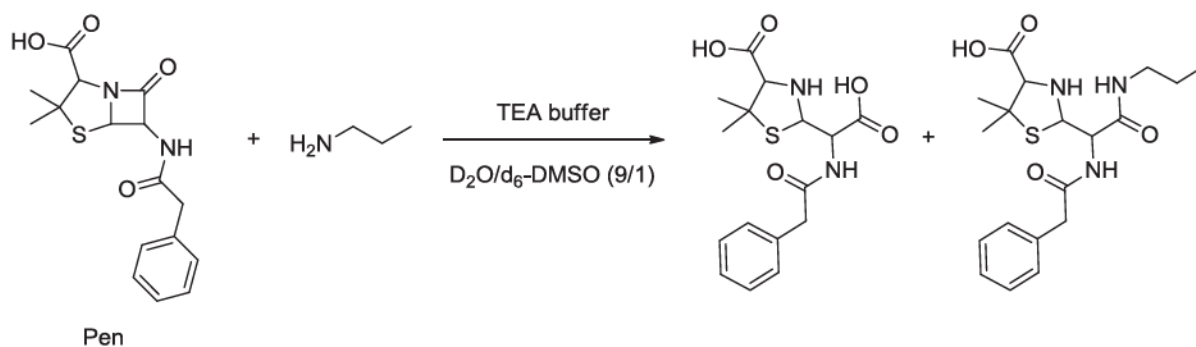
**Figure S14:** Decay of the concentration of penicillin G while reacting with phenylethylamine in a TEA-buffered reaction mixture at 37 °C ( $k_N = 8.6 \times 10^{-2} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

#### 7.4. Kinetics of the reaction of sodium penicillin G with *n*-propylamine in TEA-buffered D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v)

[PenG]<sub>0</sub> = 25 mM,

[*n*-PrNH<sub>2</sub>]<sub>0</sub> = 50 mM, [*n*-PrNH<sub>2</sub>]<sub>eff</sub> = 0.058 mM,

[TEA buffer]<sub>0</sub> = 0.10 M, pH 7.75 in D<sub>2</sub>O/d<sub>6</sub>-DMSO 9/1 (v/v), 37 °C



**Figure S55:** Decay of the concentration of penicillin G while reacting with *n*-propylamine in a TEA-buffered reaction mixture at 37 °C ( $k_N = 4.5 \times 10^{-1} \text{ M}^{-1} \text{ s}^{-1}$ ). The black line shows a mono-exponential fit of the data.

## 7.5. Rate constants for the reactions of sodium penicillin G with different aliphatic amines

**Table S13:** Rate constants for the reactions of penicillin G (PenG) with different aliphatic amines at 37 °C and pH 7.75.

	benzylamine	phenylethylamine	<i>n</i> -propylamine
$pK_{aH^+}$	9.36	9.89	10.68
$c_{\text{eff}}$ (M)	$1.2 \times 10^{-3}$	$3.6 \times 10^{-4}$	$5.8 \times 10^{-5}$
$k_{\text{obs}}$ (s <sup>-1</sup> )	$1.3 \times 10^{-4}$	$6.4 \times 10^{-5}$	$5.9 \times 10^{-5}$
$k_N$ (M <sup>-1</sup> s <sup>-1</sup> )	$8.1 \times 10^{-2}$	$8.6 \times 10^{-2}$	$4.5 \times 10^{-1}$

## 8. Determination of $N$ and $s_N$ parameters of amines in $H_2O$

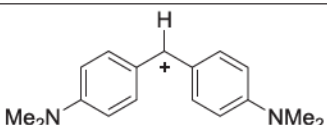
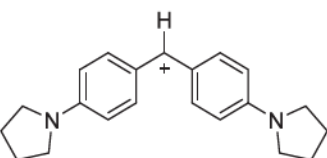
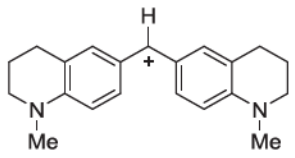
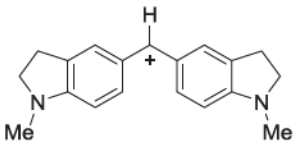
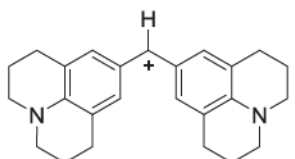
The nucleophilic reactivities of 2-phenylethylamine and *p*-anisidine in water were determined by using the benzhydrylium method as described previously.<sup>[15]</sup> Second-order rate constants for the attack of  $OH^-$  ions at the colored reference electrophiles ( $k_{2,OH^-}$ ) have previously been reported.<sup>[16]</sup> As outlined in more detail previously,<sup>[15]</sup> the second-order rate constants for the attack of the amine at the benzhydrylium ions ( $k_N$ ) in equation S1 can be obtained from the slope of a linear plot of  $k_{1\Psi}$  (equation S2) vs the amine concentration (equation S3).

$$k_{obs} = k_W + k_{2,OH^-}[OH^-] + k_N[amine] \quad (S1)$$

$$k_{1\Psi} = k_{obs} - k_{2,OH^-}[OH^-] = k_{obs} - k_{1\Psi, OH^-} \quad (S2)$$

$$k_{1\Psi} = k_W + k_N[amine] \quad (S3)$$

The following benzhydrylium tetrafluoroborates were used as reference electrophiles for the determination of the nucleophilicity parameters  $N$  and  $s_N$  of the amines according to equation (1) in the main text.

reference electrophile	abbreviation	$E$ parameter <sup>[a]</sup>
	(dma) <sub>2</sub> CH <sup>+</sup>	−7.02
	(pyr) <sub>2</sub> CH <sup>+</sup>	−7.69
	(thq) <sub>2</sub> CH <sup>+</sup>	−8.22
	(ind) <sub>2</sub> CH <sup>+</sup>	−8.76
	(jul) <sub>2</sub> CH <sup>+</sup>	−9.45
[a] from <sup>[17]</sup>		



## 8.1. Phenylethylamine

**Table S14:** Reaction of phenylethylamine with  $(\text{dma})_2\text{CH}^+\text{BF}_4^-$  (at 20 °C in water, cosolvent: 1 vol-%  $\text{CH}_3\text{CN}$ , stopped-flow, detection at 607 nm).

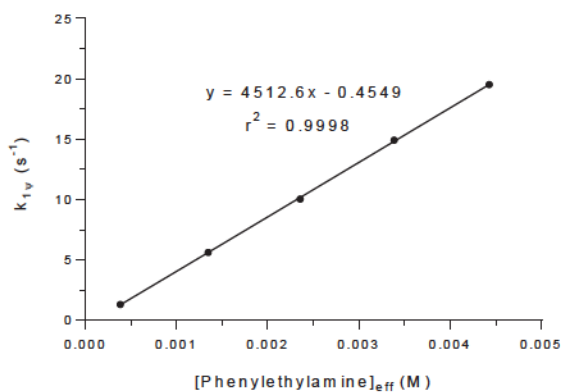
$\text{E} = (\text{dma})_2\text{CH}^+\text{BF}_4^-$ ; Nu = phenylethylamine

$[\text{E}]_0$ (M)	$[\text{Nu}]_0$ (M)	$[\text{Nu}]_{\text{eff}}$ (M)	$[\text{OH}^-]$ (M)	$[\text{Nu}]_{\text{eff}}/[\text{E}]_0$	$k_{\text{obs}}$ ( $\text{s}^{-1}$ )	$k_{1\psi, \text{OH}^-}$ ( $\text{s}^{-1}$ )	$k_{1\psi}$ ( $\text{s}^{-1}$ )
$7.11 \times 10^{-5}$	$5.57 \times 10^{-4}$	$3.84 \times 10^{-4}$	$1.73 \times 10^{-4}$	5	1.35	$2.26 \times 10^{-2}$	1.33
$7.11 \times 10^{-5}$	$1.67 \times 10^{-3}$	$1.35 \times 10^{-3}$	$3.23 \times 10^{-4}$	19	5.67	$4.24 \times 10^{-2}$	5.63
$7.11 \times 10^{-5}$	$2.79 \times 10^{-3}$	$2.36 \times 10^{-3}$	$4.28 \times 10^{-4}$	33	10.1	$5.60 \times 10^{-2}$	10.04
$7.11 \times 10^{-5}$	$3.90 \times 10^{-3}$	$3.39 \times 10^{-3}$	$5.13 \times 10^{-4}$	48	15.0	$6.72 \times 10^{-2}$	14.93
$7.11 \times 10^{-5}$	$5.01 \times 10^{-3}$	$4.43 \times 10^{-3}$	$5.86 \times 10^{-4}$	62	19.6	$7.68 \times 10^{-2}$	19.52

$$\text{p}K_{\text{B}} (\text{PhCH}_2\text{CH}_2\text{NH}_2) = 4.11^{[18]}$$

$$k_{2, \text{OH}^-} = 131 \text{ M}^{-1} \text{ s}^{-1}^{[16]}$$

$$k_{\text{N}} = 4.51 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$$



**Figure S56:** Determination of the second-order rate constant  $k_{\text{N}} = 4.51 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$  from the dependence of the first-order rate constant  $k_{1\psi}$  on the concentration of phenylethylamine.

**Table S15:** Reaction of phenylethylamine with  $(\text{pyr})_2\text{CH}^+\text{BF}_4^-$  (at 20 °C in water, cosolvent: 1 vol-%  $\text{CH}_3\text{CN}$ , stopped-flow, detection at 607 nm.)

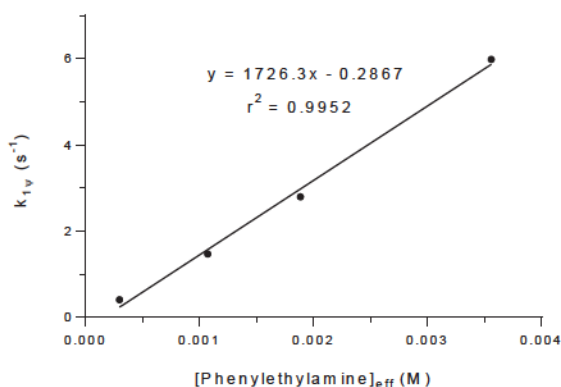
$\text{E} = (\text{pyr})_2\text{CH}^+\text{BF}_4^-$ ;  $\text{Nu} = \text{phenylethylamine}$

$[\text{E}]_0$ (M)	$[\text{Nu}]_0$ (M)	$[\text{Nu}]_{\text{eff}}$ (M)	$[\text{OH}^-]$ (M)	$[\text{Nu}]_{\text{eff}}/[\text{E}]_0$	$k_{\text{obs}}$ ( $\text{s}^{-1}$ )	$k_{1\psi, \text{OH}^-}$ ( $\text{s}^{-1}$ )	$k_{1\psi}$ ( $\text{s}^{-1}$ )
$1.37 \times 10^{-5}$	$4.54 \times 10^{-4}$	$3.01 \times 10^{-4}$	$1.53 \times 10^{-4}$	22	0.411	$7.41 \times 10^{-3}$	0.404
$1.37 \times 10^{-5}$	$1.36 \times 10^{-3}$	$1.07 \times 10^{-3}$	$2.88 \times 10^{-4}$	79	1.48	$1.40 \times 10^{-2}$	1.47
$1.37 \times 10^{-5}$	$2.27 \times 10^{-3}$	$1.89 \times 10^{-3}$	$3.83 \times 10^{-4}$	138	2.81	$1.86 \times 10^{-2}$	2.79
$1.37 \times 10^{-5}$	$4.09 \times 10^{-3}$	$3.56 \times 10^{-3}$	$5.26 \times 10^{-4}$	261	6.00	$2.55 \times 10^{-2}$	5.97

$$\text{p}K_{\text{B}} (\text{PhCH}_2\text{CH}_2\text{NH}_2) = 4.11$$

$$k_{2, \text{OH}^-} = 48.5 \text{ M}^{-1} \text{ s}^{-1} [16]$$

$$k_{\text{N}} = 1.73 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$$



**Figure S57:** Determination of the second-order rate constant  $k_{\text{N}} = 1.73 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$  from the dependence of the first-order rate constant  $k_{1\psi}$  on the concentration of phenylethylamine.

**Table S16:** Reaction of phenylethylamine with  $(\text{thq})_2\text{CH}^+\text{BF}_4^-$  (at 20 °C in water, cosolvent: 1 vol-%  $\text{CH}_3\text{CN}$ , stopped-flow, detection at 607 nm).

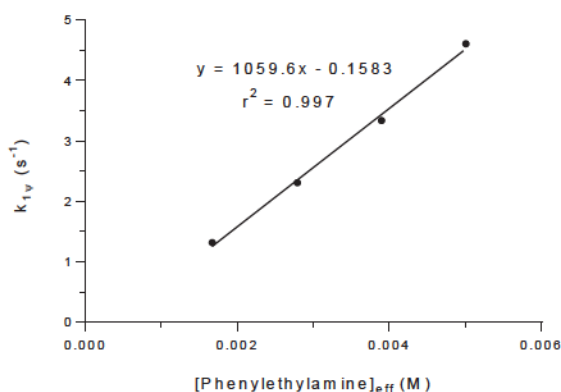
$\text{E} = (\text{thq})_2\text{CH}^+\text{BF}_4^-$ ;  $\text{Nu} = \text{phenylethylamine}$

$[\text{E}]_0 \text{ (M)}$	$[\text{Nu}]_0 \text{ (M)}$	$[\text{Nu}]_{\text{eff}} \text{ (M)}$	$[\text{OH}^-] \text{ (M)}$	$[\text{Nu}]_{\text{eff}}/[\text{E}]_0$	$k_{\text{obs}} \text{ (s}^{-1}\text{)}$	$k_{1\psi, \text{OH}^-} \text{ (s}^{-1}\text{)}$	$k_{1\psi} \text{ (s}^{-1}\text{)}$
$1.27 \times 10^{-5}$	$1.67 \times 10^{-3}$	$1.35 \times 10^{-3}$	$3.23 \times 10^{-4}$	106	1.33	$7.63 \times 10^{-3}$	1.32
$1.27 \times 10^{-5}$	$2.79 \times 10^{-3}$	$2.36 \times 10^{-3}$	$4.28 \times 10^{-4}$	185	2.32	$1.01 \times 10^{-2}$	2.31
$1.27 \times 10^{-5}$	$3.90 \times 10^{-3}$	$3.39 \times 10^{-3}$	$5.13 \times 10^{-4}$	266	3.35	$1.21 \times 10^{-2}$	3.34
$1.27 \times 10^{-5}$	$5.01 \times 10^{-3}$	$4.42 \times 10^{-3}$	$5.86 \times 10^{-4}$	347	4.62	$1.38 \times 10^{-2}$	4.61

$$\text{p}K_{\text{B}} (\text{PhCH}_2\text{CH}_2\text{NH}_2) = 4.11$$

$$k_{2, \text{OH}^-} = 23.6 \text{ M}^{-1} \text{ s}^{-1} [16]$$

$$k_{\text{N}} = 1.06 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$$



**Figure S58:** Determination of the second-order rate constant  $k_{\text{N}} = 1.06 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$  from the dependence of the first-order rate constant  $k_{1\psi}$  on the concentration of phenylethylamine.

**Table S17:** Reaction of phenylethylamine with (ind)<sub>2</sub>CH<sup>+</sup>BF<sub>4</sub><sup>−</sup> (at 20 °C in water, cosolvent: 1 vol-% CH<sub>3</sub>CN, stopped-flow, detection at 607 nm).

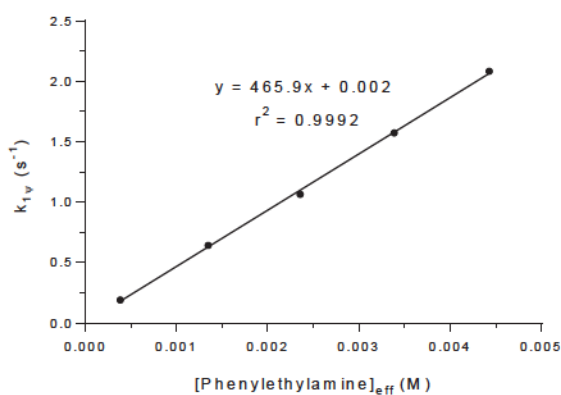
E = (ind)<sub>2</sub>CH<sup>+</sup>BF<sub>4</sub><sup>−</sup>; Nu = phenylethylamine

[E] <sub>0</sub> (M)	[Nu] <sub>0</sub> (M)	[Nu] <sub>eff</sub> (M)	[OH <sup>−</sup> ] (M)	[Nu] <sub>eff</sub> /[E] <sub>0</sub>	k <sub>obs</sub> (s <sup>−1</sup> )	k <sub>1ψ, OH<sup>−</sup></sub> (s <sup>−1</sup> )	k <sub>1ψ</sub> (s <sup>−1</sup> )
1.80 × 10 <sup>−5</sup>	5.57 × 10 <sup>−4</sup>	3.84 × 10 <sup>−4</sup>	1.73 × 10 <sup>−4</sup>	21	0.193	1.87 × 10 <sup>−3</sup>	0.191
1.80 × 10 <sup>−5</sup>	1.67 × 10 <sup>−3</sup>	1.35 × 10 <sup>−3</sup>	3.23 × 10 <sup>−4</sup>	75	0.646	3.49 × 10 <sup>−3</sup>	0.643
1.80 × 10 <sup>−5</sup>	2.79 × 10 <sup>−3</sup>	2.36 × 10 <sup>−3</sup>	4.28 × 10 <sup>−4</sup>	131	1.07	4.62 × 10 <sup>−3</sup>	1.07
1.80 × 10 <sup>−5</sup>	3.90 × 10 <sup>−3</sup>	3.39 × 10 <sup>−3</sup>	5.13 × 10 <sup>−4</sup>	188	1.58	5.54 × 10 <sup>−3</sup>	1.57
1.80 × 10 <sup>−5</sup>	5.01 × 10 <sup>−3</sup>	4.42 × 10 <sup>−3</sup>	5.86 × 10 <sup>−4</sup>	246	2.09	6.33 × 10 <sup>−3</sup>	2.08

$$\text{p}K_{\text{B}} (\text{PhCH}_2\text{CH}_2\text{NH}_2) = 4.11$$

$$k_{2,\text{OH}^-} = 10.8 \text{ M}^{-1} \text{ s}^{-1} [16]$$

$$k_{\text{N}} = 466 \text{ M}^{-1} \text{ s}^{-1}$$



**Figure S59:** Determination of the second-order rate constant  $k_{\text{N}} = 466 \text{ M}^{-1} \text{ s}^{-1}$  from the dependence of the first-order rate constant  $k_{1\psi}$  on the concentration of phenylethylamine.

**Table S18:** Reaction of phenylethylamine with  $(\text{jul})_2\text{CH}^+\text{BF}_4^-$  (at 20 °C in water, cosolvent: 1 vol-%  $\text{CH}_3\text{CN}$ , stopped-flow, detection at 630 nm).

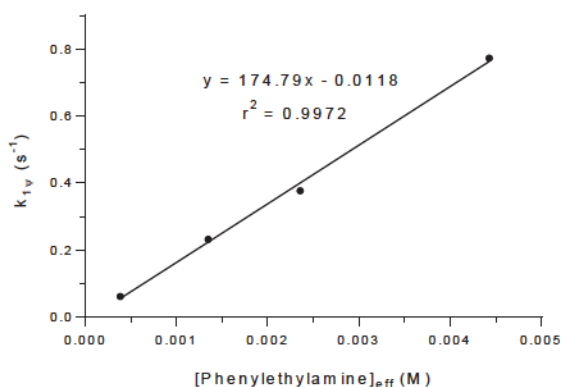
$\text{E} = (\text{jul})_2\text{CH}^+\text{BF}_4^-$ ;  $\text{Nu} = \text{phenylethylamine}$

$[\text{E}]_0$ (M)	$[\text{Nu}]_0$ (M)	$[\text{Nu}]_{\text{eff}}$ (M)	$[\text{OH}^-]$ (M)	$[\text{Nu}]_{\text{eff}}/[\text{E}]_0$	$k_{\text{obs}}$ ( $\text{s}^{-1}$ )	$k_{1\psi, \text{OH}^-}$ ( $\text{s}^{-1}$ )	$k_{1\psi}$ ( $\text{s}^{-1}$ )
$2.77 \times 10^{-5}$	$5.57 \times 10^{-4}$	$3.84 \times 10^{-4}$	$1.73 \times 10^{-4}$	14	0.0621	$5.94 \times 10^{-4}$	$6.15 \times 10^{-2}$
$2.77 \times 10^{-5}$	$1.67 \times 10^{-3}$	$1.35 \times 10^{-3}$	$3.23 \times 10^{-4}$	49	0.233	$1.11 \times 10^{-3}$	$2.32 \times 10^{-1}$
$2.77 \times 10^{-5}$	$2.79 \times 10^{-3}$	$2.36 \times 10^{-3}$	$4.28 \times 10^{-4}$	85	0.378	$1.47 \times 10^{-3}$	$3.77 \times 10^{-1}$
$2.77 \times 10^{-5}$	$5.01 \times 10^{-3}$	$4.42 \times 10^{-3}$	$5.86 \times 10^{-4}$	160	0.774	$2.02 \times 10^{-3}$	$7.72 \times 10^{-1}$

$$\text{p}K_{\text{B}} (\text{PhCH}_2\text{CH}_2\text{NH}_2) = 4.11$$

$$k_{2, \text{OH}^-} = 3.44 \text{ M}^{-1} \text{ s}^{-1} [16]$$

$$k_{\text{N}} = 175 \text{ M}^{-1} \text{ s}^{-1}$$

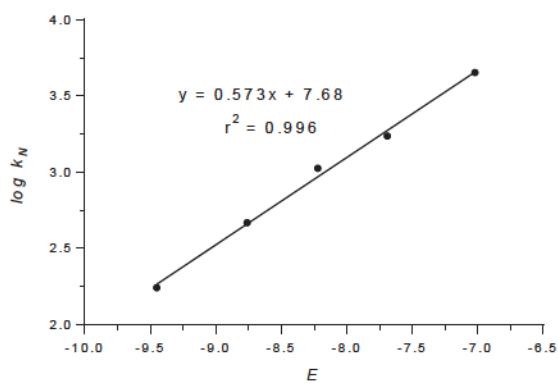


**Figure S60:** Determination of the second-order rate constant  $k_{\text{N}} = 175 \text{ M}^{-1} \text{ s}^{-1}$  from the dependence of the first-order rate constant  $k_{1\psi}$  on the concentration of phenylethylamine.

**Table S19:** Rate Constants for the reactions of phenylethylamine with different benzhydrylium ions as reference electrophiles (20 °C, in water).

reference electrophile	<i>E</i> parameter	<i>k<sub>N</sub></i> (20°C) (M <sup>-1</sup> s <sup>-1</sup> )
(dma) <sub>2</sub> CH <sup>+</sup>	-7.02	4.51 × 10 <sup>3</sup>
(pyr) <sub>2</sub> CH <sup>+</sup>	-7.69	1.73 × 10 <sup>3</sup>
(thq) <sub>2</sub> CH <sup>+</sup>	-8.22	1.06 × 10 <sup>3</sup>
(ind) <sub>2</sub> CH <sup>+</sup>	-8.76	4.66 × 10 <sup>2</sup>
(jul) <sub>2</sub> CH <sup>+</sup>	-9.45	1.75 × 10 <sup>2</sup>

Reactivity parameters for phenylethylamine (in water): *N* = 13.40; *s<sub>N</sub>* = 0.57



**Figure S61:** Plot of *log k<sub>N</sub>* vs. the electrophilicity parameters *E* for the reactions of benzhydrylium ions with phenylethylamine.

## 8.2. *p*-Anisidine

**Table S20:** Reaction of *p*-anisidine with  $(\text{dma})_2\text{CH}^+\text{BF}_4^-$  (at 20 °C in water, cosolvent: 1 vol-%  $\text{CH}_3\text{CN}$ , stopped-flow, detection at 607 nm).

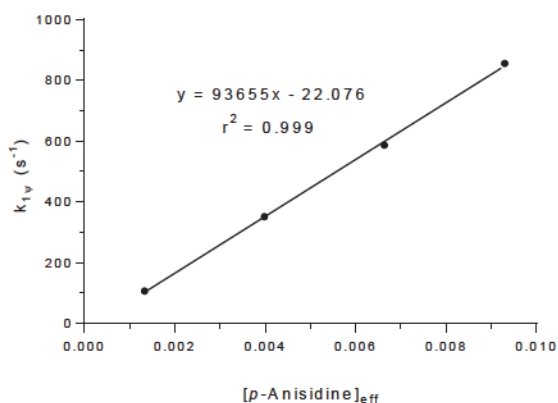
$\text{E} = (\text{dma})_2\text{CH}^+\text{BF}_4^-$ ;  $\text{Nu} = p\text{-anisidine}$

$[\text{E}]_0$ (M)	$[\text{Nu}]_0$ (M)	$[\text{Nu}]_{\text{eff}}$	$[\text{OH}^-]$ (M)	$[\text{Nu}]_{\text{eff}}/[\text{E}]_0$	$k_{\text{obs}}$ ( $\text{s}^{-1}$ )	$k_{1\psi, \text{OH}^-}$ ( $\text{s}^{-1}$ )	$k_{1\psi}$ ( $\text{s}^{-1}$ )
$1.90 \times 10^{-4}$	$1.33 \times 10^{-3}$	$1.33 \times 10^{-3}$	$1.63 \times 10^{-6}$	7	106	$2.13 \times 10^{-4}$	$1.06 \times 10^2$
$1.90 \times 10^{-4}$	$3.98 \times 10^{-3}$	$3.98 \times 10^{-3}$	$2.82 \times 10^{-6}$	21	351	$3.69 \times 10^{-4}$	$3.51 \times 10^2$
$1.90 \times 10^{-4}$	$6.64 \times 10^{-3}$	$6.64 \times 10^{-3}$	$3.64 \times 10^{-6}$	35	587	$4.77 \times 10^{-4}$	$5.87 \times 10^2$
$1.90 \times 10^{-4}$	$9.29 \times 10^{-3}$	$9.29 \times 10^{-3}$	$4.30 \times 10^{-6}$	49	856	$5.64 \times 10^{-4}$	$8.56 \times 10^2$

$$\text{p}K_{\text{B}} = 8.70^{[13\text{a}]}$$

$$k_{2, \text{OH}^-} = 131 \text{ M}^{-1} \text{ s}^{-1} [16]$$

$$k_{\text{N}} = 9.37 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$$



**Figure S62:** Determination of the second-order rate constant  $k_{\text{N}} = 9.37 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$  from the dependence of the first-order rate constant  $k_{1\psi}$  on the concentration of *p*-anisidine.

**Table S21:** Reaction of *p*-anisidine with  $(\text{pyr})_2\text{CH}^+\text{BF}_4^-$  (at 20 °C in water, cosolvent: 1 vol-%  $\text{CH}_3\text{CN}$ , stopped-flow, detection at 607 nm).

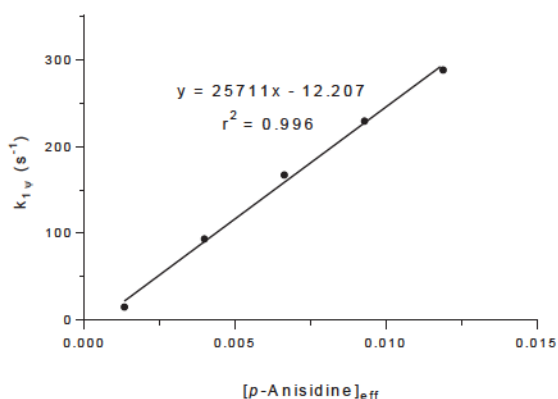
$\text{E} = (\text{pyr})_2\text{CH}^+\text{BF}_4^-$ ;  $\text{Nu} = p\text{-anisidine}$

$[\text{E}]_0 (\text{M})$	$[\text{Nu}]_0 (\text{M})$	$[\text{Nu}]_{\text{eff}}$	$[\text{OH}^-] (\text{M})$	$[\text{Nu}]_{\text{eff}}/[\text{E}]_0$	$k_{\text{obs}} (\text{s}^{-1})$	$k_{1\psi, \text{OH}^-} (\text{s}^{-1})$	$k_{1\psi} (\text{s}^{-1})$
$2.95 \times 10^{-5}$	$1.33 \times 10^{-3}$	$1.33 \times 10^{-3}$	$1.63 \times 10^{-6}$	45	14.6	$7.90 \times 10^{-5}$	$1.46 \times 10^1$
$2.95 \times 10^{-5}$	$3.98 \times 10^{-3}$	$3.98 \times 10^{-3}$	$2.82 \times 10^{-6}$	135	93.3	$1.37 \times 10^{-4}$	$9.33 \times 10^1$
$2.95 \times 10^{-5}$	$6.64 \times 10^{-3}$	$6.64 \times 10^{-3}$	$3.64 \times 10^{-6}$	225	167	$1.76 \times 10^{-4}$	$1.67 \times 10^2$
$2.95 \times 10^{-5}$	$9.29 \times 10^{-3}$	$9.29 \times 10^{-3}$	$4.30 \times 10^{-6}$	315	229	$2.09 \times 10^{-4}$	$2.29 \times 10^2$
$2.95 \times 10^{-5}$	$1.19 \times 10^{-2}$	$1.19 \times 10^{-2}$	$4.87 \times 10^{-6}$	405	288	$2.36 \times 10^{-4}$	$2.88 \times 10^2$

$$\text{p}K_{\text{B}} = 8.70$$

$$k_{2, \text{OH}^-} = 48.5 \text{ M}^{-1} \text{ s}^{-1[16]}$$

$$k_{\text{N}} = 2.57 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$$



**Figure S63:** Determination of the second-order rate constant  $k_{\text{N}} = 2.57 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$  from the dependence of the first-order rate constant  $k_{1\psi}$  on the concentration of *p*-anisidine.



**Table S22:** Reaction of *p*-anisidine with (ind)<sub>2</sub>CH<sup>+</sup>BF<sub>4</sub><sup>−</sup> (at 20 °C in water, cosolvent: 1 vol-% CH<sub>3</sub>CN, stopped-flow, detection at 607 nm).

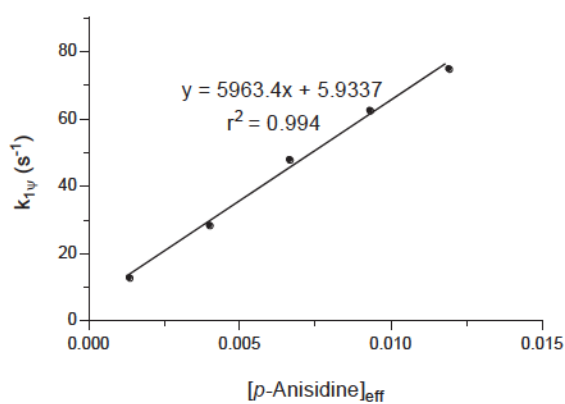
E = (ind)<sub>2</sub>CH<sup>+</sup>BF<sub>4</sub><sup>−</sup>; Nu = *p*-anisidine

[E] <sub>0</sub> (M)	[Nu] <sub>0</sub> (M)	[Nu] <sub>eff</sub>	[OH <sup>−</sup> ] (M)	[Nu] <sub>eff</sub> /[E] <sub>0</sub>	k <sub>obs</sub> (s <sup>−1</sup> )	k <sub>1ψ, OH<sup>−</sup></sub> (s <sup>−1</sup> )	k <sub>1ψ</sub> (s <sup>−1</sup> )
3.23 × 10 <sup>−5</sup>	1.33 × 10 <sup>−3</sup>	1.33 × 10 <sup>−3</sup>	1.63 × 10 <sup>−6</sup>	41	13.0	1.76 × 10 <sup>−5</sup>	13.0
3.23 × 10 <sup>−5</sup>	3.98 × 10 <sup>−3</sup>	3.98 × 10 <sup>−3</sup>	2.82 × 10 <sup>−6</sup>	123	28.8	3.04 × 10 <sup>−5</sup>	28.6
3.23 × 10 <sup>−5</sup>	6.64 × 10 <sup>−3</sup>	6.64 × 10 <sup>−3</sup>	3.64 × 10 <sup>−6</sup>	205	48.1	3.93 × 10 <sup>−5</sup>	48.1
3.23 × 10 <sup>−5</sup>	9.29 × 10 <sup>−3</sup>	9.29 × 10 <sup>−3</sup>	4.31 × 10 <sup>−6</sup>	288	62.7	4.65 × 10 <sup>−5</sup>	62.7
3.23 × 10 <sup>−5</sup>	1.19 × 10 <sup>−2</sup>	1.19 × 10 <sup>−2</sup>	4.87 × 10 <sup>−6</sup>	370	75.1	5.26 × 10 <sup>−5</sup>	75.1

$$\text{p}K_{\text{B}} = 8.70$$

$$k_{2,\text{OH}^-} = 10.8 \text{ M}^{-1} \text{ s}^{-1[16]}$$

$$k_{\text{N}} = 5.96 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$$

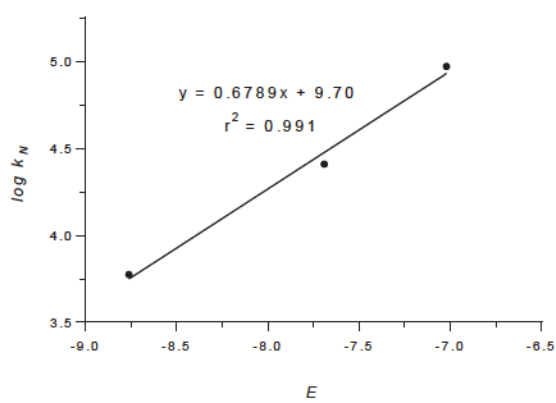


**Figure S64:** Determination of the second-order rate constant  $k_{\text{N}} = 5.96 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$  from the dependence of the first-order rate constant  $k_{1\psi}$  on the concentration of *p*-anisidine.

**Table S23:** Rate constants for the reactions of *p*-anisidine with different electrophiles (20 °C).

reference electrophile	<i>E</i> parameter	$k_N$ (20°C) (M <sup>-1</sup> s <sup>-1</sup> )
(dma) <sub>2</sub> CH <sup>+</sup>	-7.02	$9.37 \times 10^4$
(pyr) <sub>2</sub> CH <sup>+</sup>	-7.69	$2.57 \times 10^4$
(ind) <sub>2</sub> CH <sup>+</sup>	-8.76	$5.96 \times 10^3$

Reactivity parameters for *p*-anisidine (in water):  $N = 14.28$ ;  $s_N = 0.68$

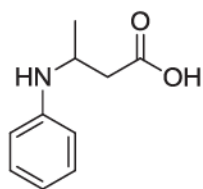


**Figure S65:** Plot of  $\log k_N$  vs. the electrophilicity parameters  $E$  for the reactions of benzhydrylium ions with *p*-anisidine.

## 9. Product studies

### 9.1. Reactions of amines and anilines with BBL

#### 3-(Phenylamino)butanoic acid [FM348\_6]



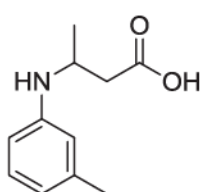
To a solution of aniline (93.1 mg, 1.00 mmol) in a 9/1-mixture of H<sub>2</sub>O and CH<sub>3</sub>CN (30 mL) was added BBL (81.5  $\mu$ L, 1.0 mmol). The reaction mixture was stirred until HPLC-ESI-HRMS analysis showed no further increase in product formation. The solvent was evaporated under reduced pressure. Purification by reversed-phase HPLC yielded 3-(phenylamino)butanoic acid (65%) as a yellowish oil. The obtained NMR data is in accordance with previously published results.<sup>[19]</sup>

**HRMS (ESI):** calcd. for C<sub>10</sub>H<sub>14</sub>NO<sub>2</sub> [M+H]<sup>+</sup>: 180.10191; found: 180.10179.

**<sup>1</sup>H NMR** (360 MHz, DMSO):  $\delta$  7.26–7.21 (m, 2 H), 6.89–6.83 (m, 3 H), 3.88–3.75 (m, 1 H), 2.58 (dd,  $J$  = 15.6, 5.4 Hz, 1 H), 2.35 (dd,  $J$  = 15.6, 7.7 Hz, 1 H), 1.19 (d,  $J$  = 6.4 Hz, 3 H).

**<sup>13</sup>C NMR** (90.6 MHz, DMSO):  $\delta$  172.3, 143.2, 129.4, 120.5, 116.4, 48.4, 39.7, 18.9.

#### 3-(*m*-Tolylamino)butanoic acid [FM349\_1-3]



To a solution of *m*-toluidine (107 mg, 1.00 mmol) in a 9/1-mixture of H<sub>2</sub>O and CH<sub>3</sub>CN (30 mL) was added BBL (81.5  $\mu$ L, 1.00 mmol). The reaction mixture was stirred until HPLC-ESI-HRMS analysis showed no further increase in product formation. The solvent was evaporated under reduced pressure. Purification by reversed-phase HPLC yielded 3-(*m*-tolylamino)butanoic acid (70%) as a yellowish solid.

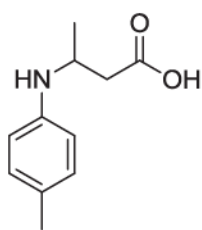
**T<sub>m</sub>:** 127 °C

**HRMS (ESI):** calcd. for C<sub>11</sub>H<sub>16</sub>NO<sub>2</sub> [M+H]<sup>+</sup>: 194.11756; found: 194.11745.

**<sup>1</sup>H NMR** (500 MHz, DMSO):  $\delta$  7.16 (t,  $J$  = 7.7 Hz, 1 H), 6.86–6.63 (m, 3 H), 3.86–3.76 (m, 1 H), 2.59 (dd,  $J$  = 15.7, 5.3 Hz, 1 H), 2.37 (dd,  $J$  = 15.7, 7.8 Hz, 1 H), 2.26 (s, 3 H), 1.19 (d,  $J$  = 6.4 Hz, 3 H).

**<sup>13</sup>C NMR** (90.6 MHz, DMSO):  $\delta$  172.1, 142.1, 138.7, 129.2, 122.2, 117.7, 114.4, 49.0, 39.4, 21.1, 18.6.

### 3-(*p*-Tolylamino)butanoic acid [FM350\_6-8]



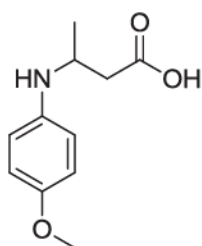
To a solution of *p*-toluidine (107 mg, 1.00 mmol) in a 9/1-mixture of H<sub>2</sub>O and CH<sub>3</sub>CN (30 mL) was added BBL (81.5 μL, 1.00 mmol). The reaction mixture was stirred until HPLC-ESI-HRMS analysis showed no further increase in product formation. The solvent was evaporated under reduced pressure. Purification by reversed-phase HPLC yielded 3-(*p*-tolylamino)butanoic acid (79%) as an orange oil.

**HRMS (ESI):** calcd. for C<sub>11</sub>H<sub>16</sub>NO<sub>2</sub> [M+H]<sup>+</sup>: 194.11756; found: 194.11746.

**<sup>1</sup>H NMR** (500 MHz, DMSO): δ 7.18 (d, *J* = 8.1 Hz, 2 H), 7.05 (d, *J* = 8.1 Hz, 2 H), 3.92–3.64 (m, 1 H), 2.63 (dd, *J* = 16.0, 5.1 Hz, 1 H), 2.43 (dd, *J* = 16.0, 8.0 Hz, 1 H), 2.26 (s, 3 H), 1.21 (d, *J* = 6.5 Hz, 3 H).

**<sup>13</sup>C NMR** (90.6 MHz, DMSO): δ 172.0, 136.6, 134.3, 130.3, 120.1, 51.7, 38.5, 20.6, 17.9.

### 3-((4-Methoxyphenyl)amino)butanoic acid [FM351\_4-7]



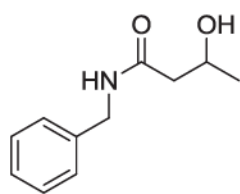
To a solution of 4-methoxyaniline (123 mg, 1.00 mmol) in a 9/1-mixture of H<sub>2</sub>O and CH<sub>3</sub>CN (30 mL) was added BBL (81.5 μL, 1.00 mmol). The reaction mixture was stirred until HPLC-ESI-HRMS analysis showed no further increase in product formation. The solvent was evaporated under reduced pressure. Purification by reversed-phase HPLC yielded 3-((4-methoxyphenyl)amino)butanoic acid (59%) as a colorless oil.

**HRMS (ESI):** calcd. for C<sub>11</sub>H<sub>16</sub>NO<sub>3</sub> [M+H]<sup>+</sup>: 210.11247; found: 210.11240.

**<sup>1</sup>H NMR** (360 MHz, DMSO): δ 7.26 (d, *J* = 8.8 Hz, 2 H), 7.01 (d, *J* = 8.9 Hz, 2 H), 3.84–3.77 (m, *J* = 13.1, 6.4 Hz, 1 H), 3.75 (s, 3 H), 2.66 (dd, *J* = 16.3, 4.9 Hz, 1H), 2.50–2.42 (m, 1 H), 1.23 (d, *J* = 6.5 Hz, 3 H).

**<sup>13</sup>C NMR** (90.6 MHz, DMSO): δ 171.5, 157.8, 129.9, 122.9, 115.1, 55.5, 53.1, 37.8, 17.2.

### ***N*-Benzyl-3-hydroxybutanamide [FM353\_28-30]**



To a solution of benzylamine (109  $\mu$ L, 1.00 mmol) in a 9/1-mixture of  $\text{H}_2\text{O}$  and  $\text{CH}_3\text{CN}$  (30 mL) was added BBL (81.5  $\mu$ L, 1.00 mmol). The reaction mixture was stirred until HPLC-ESI-HRMS analysis showed no further increase in product formation. The solvent was evaporated under reduced pressure. Purification by reversed-phase HPLC yielded *N*-benzyl-3-hydroxybutanamide (59%) as a white powder. The obtained NMR data is in accordance with previously published results.<sup>[20]</sup>

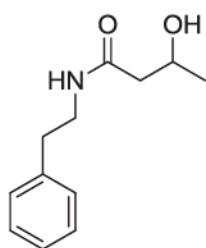
**T<sub>m</sub>:** 69 °C

**HRMS (ESI):** calcd. for  $\text{C}_{11}\text{H}_{16}\text{NO}_2$   $[\text{M}+\text{H}]^+$ : 194.11756; found: 194.11750.

**$^1\text{H}$  NMR** (500 MHz, DMSO):  $\delta$  8.32 (t,  $J$  = 5.6 Hz, 1 H), 7.35–7.21 (m, 5 H), 4.70 (d,  $J$  = 2.3 Hz, 1 H), 4.29 (d,  $J$  = 6.0 Hz, 2 H), 4.04 (d,  $J$  = 5.4 Hz, 1 H), 2.31 (dd,  $J$  = 13.8, 7.3 Hz, 1 H), 2.21 (dd,  $J$  = 13.8, 5.8 Hz, 1 H), 1.10 (d,  $J$  = 6.2 Hz, 3 H).

**$^{13}\text{C}$  NMR** (90.6 MHz, DMSO):  $\delta$  170.7, 139.6, 128.2, 127.2, 126.7, 63.8, 45.3, 41.9, 39.5, 23.4.

### **3-Hydroxy-*N*-phenylethylbutanamide [FM370\_13-15]**



To a solution of 2-phenylethylamine (127  $\mu$ L, 122 mg, 1.00 mmol) in a 9/1-mixture of  $\text{H}_2\text{O}$  and  $\text{CH}_3\text{CN}$  (30 mL) was added BBL (81.5  $\mu$ L, 1.00 mmol). The reaction mixture was stirred until HPLC-ESI-HRMS analysis showed no further increase in product formation. The solvent was evaporated under reduced pressure and purification by reversed-phase HPLC yielded 3-hydroxy-*N*-phenylethylbutanamide (58%) as a white powder. The obtained NMR data is in accordance with previously published results.<sup>[21]</sup>

**T<sub>m</sub>:** 51 °C

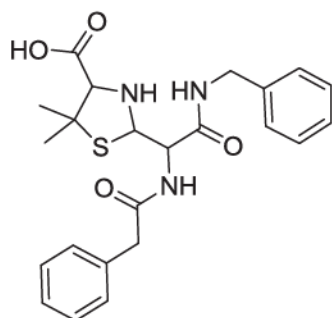
**HRMS (ESI):** calcd. for  $\text{C}_{12}\text{H}_{18}\text{NO}_2$   $[\text{M}+\text{H}]^+$ : 208.13321; found: 208.13304.

**$^1\text{H}$  NMR** (500 MHz, DMSO):  $\delta$  7.89 (t,  $J$  = 4.9 Hz, 1H), 7.43–7.08 (m, 5 H), 4.89 (s, 1 H), 4.11–3.84 (m, 1 H), 3.37–3.12 (m, 2 H), 2.71 (t,  $J$  = 7.4 Hz, 2 H), 2.21 (dd,  $J$  = 13.8, 7.1 Hz, 1 H), 2.10 (dd,  $J$  = 13.8, 5.9 Hz, 1 H), 1.05 (d,  $J$  = 6.2 Hz, 3 H).

**$^{13}\text{C}$  NMR** (90.6 MHz, DMSO):  $\delta$  170.6, 139.5, 128.6, 128.3, 126.0, 63.8, 45.3, 40.1, 35.2, 23.3.

## 9.2. Reactions of amines with sodium penicillin G

### 2-(2-(Benzylamino)-2-oxo-1-(2-phenylacetamido)ethyl)-5,5-dimethylthiazolidine-4-carboxylic acid



To a solution of benzylamine (31.5  $\mu$ L, 30.3 mg, 0.25 mmol) in a 9/1-mixture of  $\text{H}_2\text{O}$  and  $\text{CH}_3\text{CN}$  (5 mL) was added sodium benzylpenicillin (89 mg, 0.25 mmol). The reaction mixture was stirred until HPLC-ESI-HRMS analysis showed no further increase in product formation. The solvent was evaporated under reduced pressure and purification by reversed-phase HPLC yielded 2-(2-(benzylamino)-2-oxo-1-(2-phenylacetamido)ethyl)-

5,5-dimethylthiazolidine-4-carboxylic acid (53%) as a white powder (isolated as a mixture of diastereomers 45:55 dr). The obtained NMR data is in accordance with previously published results.<sup>[22]</sup>

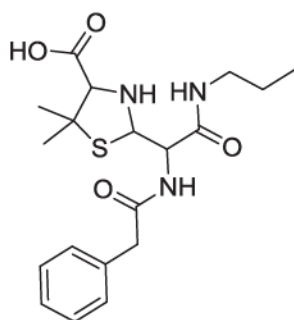
$T_m$ : 114  $^{\circ}\text{C}$ <sup>[23]</sup>

**HRMS (ESI)**: calcd. for  $\text{C}_{22}\text{H}_{28}\text{N}_3\text{O}_4\text{S}$   $[\text{M}+\text{H}]^+$ : 442.17950; found: 442.17977.

**$^1\text{H}$  NMR** (500 MHz, DMSO):  $\delta$  8.70 (t,  $J$  = 5.7 Hz, 1 H), 8.55 (t,  $J$  = 5.8 Hz, 1 H), 8.45 (d,  $J$  = 8.6 Hz, 1 H), 8.34 (d,  $J$  = 9.2 Hz, 1 H), 7.34–7.17 (m, 20 H), 4.89 (dd,  $J$  = 12.3, 7.0 Hz, 2 H), 4.81–4.75 (m, 1 H), 4.51 (t,  $J$  = 8.5 Hz, 1 H), 4.38–4.15 (m, 6 H), 3.64 – 3.56 (m, 2 H), 3.56–3.43 (m, 4 H), 1.56 (s, 3 H), 1.52 (s, 3 H), 1.19 (s, 3 H), 1.14 (s, 3 H).

**$^{13}\text{C}$  NMR** (126 MHz, DMSO):  $\delta$  170.5, 170.4, 170.3, 169.9, 169.6, 169.1, 139.1, 138.8, 136.6, 136.2, 129.14, 129.12, 128.24, 128.22, 128.2, 128.19, 128.1, 127.2, 127.1, 126.8, 126.7, 126.4, 126.2, 73.5, 72.2, 67.7, 66.9, 58.4, 57.2, 54.8, 42.2, 42.04, 42.01, 41.9, 28.4, 27.9, 27.7, 27.1.

**5,5-Dimethyl-2-(2-oxo-1-(2-phenylacetamido)-2-(propylamino)ethyl)thiazolidine-4-carboxylic acid**



To a solution of propylamine (20.5  $\mu$ L, 30.3 mg, 0.25 mmol) in a 9/1-mixture of  $\text{H}_2\text{O}$  and  $\text{CH}_3\text{CN}$  (5 mL) was added sodium benzylpenicillin (89 mg, 0.25 mmol). The reaction mixture was stirred until HPLC-ESI-HRMS analysis showed no further increase in product formation. The solvent was evaporated under reduced pressure and purification by reversed-phase HPLC yielded 5,5-dimethyl-2-(2-oxo-1-(2-phenylacetamido)-2-(propylamino)ethyl)-

thiazolidine-4-carboxylic acid (66%) as a white powder (isolated as a mixture of diastereomers 45:55 dr.)

**T<sub>m</sub>:** 121  $^{\circ}\text{C}$

**HRMS (ESI):** calcd. for  $\text{C}_{19}\text{H}_{28}\text{N}_3\text{O}_4\text{S}$   $[\text{M}+\text{H}]^+$ : 394.17950; found: 394.17956.

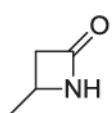
**$^1\text{H}$  NMR** (360 MHz, DMSO):  $\delta$  8.38 (d,  $J$  = 8.7 Hz, 1 H), 8.28 (d,  $J$  = 9.1 Hz, 1 H), 8.17 (t,  $J$  = 5.5 Hz, 1 H), 8.02 (t,  $J$  = 5.5 Hz, 1 H), 7.33–7.15 (m, 10 H), 4.86–4.83 (m, 2 H), 4.71–4.67 (m, 1 H), 4.45–4.40 (t,  $J$  = 8.5 Hz, 1 H), 3.62 (d,  $J$  = 5.0 Hz, 2 H), 3.51 (dd,  $J$  = 9.7, 2.8 Hz, 4 H), 3.04–2.96 (m, 4 H), 1.55 (s, 3 H), 1.51 (s, 3 H), 1.44–1.34 (m, 4 H), 1.19 (s, 3 H), 1.13 (s, 3 H), 0.83 (q,  $J$  = 7.7 Hz, 6 H).

**$^{13}\text{C}$  NMR** (90.6 MHz, DMSO):  $\delta$  170.4, 170.2, 170.1, 169.6, 169.1, 168.6, 136.4, 136.0, 129.1, 129.0, 128.1, 128.0, 126.3, 126.1, 73.2, 72.0, 67.5, 66.7, 57.9, 57.0, 56.8, 56.7, 54.8, 54.2, 42.0, 41.8, 28.3, 27.7, 27.5, 26.9, 22.1, 22.0, 11.3.



## 10. Syntheses of lactams

### 4-Methyl-1-azetidin-2-one



To a stirred solution of 3-aminobutanoic acid (7.50 g, 72.2 mmol) and 2-chloro-1-methylpyridinium iodide (20.4 g, 80 mmol) in dry CH<sub>3</sub>CN (500 mL) was added dropwise a solution of triethylamine (22.0 mL, 16.2 g, 160 mmol) in dry CH<sub>3</sub>CN (100 mL) under nitrogen over 10 min. The mixture was heated to reflux for 2 h. Then the solvent was evaporated under reduced pressure. Purification by flash column chromatography on silica (5.0 × 35 cm, EtOAc/EtOH = 10:1) yielded 4-methylazetidin-2-one (2.90 g, 34 mmol, 46%) as an yellowish oil. The obtained NMR data is in accordance with previously published results.<sup>[24]</sup>

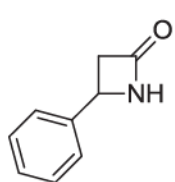
$R_f$  = 0.48 (EtOAc:EtOH 10:1)

**HRMS (ESI):** calcd. for C<sub>4</sub>H<sub>8</sub>NO [M+H]<sup>+</sup>: 86.06004; found: 86.06006.

**<sup>1</sup>H NMR** (360 MHz, CDCl<sub>3</sub>):  $\delta$  6.34 (br s, 1 H), 3.79–3.72 (m, 1 H), 3.07 (ddd,  $J$  = 14.8, 5.0, 2.1 Hz, 1 H), 2.52 (ddd,  $J$  = 14.8, 2.3, 1.4 Hz, 1 H), 1.34 (d,  $J$  = 6.1 Hz, 3 H).

**<sup>13</sup>C NMR** (90.6 MHz, CDCl<sub>3</sub>):  $\delta$  168.1, 45.1, 44.0, 21.3.

### 4-Phenyl-1-azetidin-2-one



*N*-Chlorosulfonyl isocyanate (16.0 mL, 25.9 g, 194 mmol, 1.2 eq.) was added dropwise to a stirred solution of styrene (18.5 mL, 16.8 g, 161 mmol, 1.0 eq.) in dry diethyl ether (75 mL) at room temperature over 10 min under nitrogen.

After stirring at ambient temperature for 2 h the solvent was removed under reduced pressure to give an oily residue, which was then redissolved in diethyl ether (100 mL) and added dropwise over 10 min to a vigorously stirred solution of water (150 mL), sodium carbonate (45.0 g, 106 mmol, 3.0 eq.), sodium sulfite (30.0 g, 126 mmol, 1.5 eq.) and ice (100 g). The solution was stirred for 1 h and filtered through celite under vacuum. The two layers were separated and the aqueous layer was extracted with diethyl ether (5 x 200 mL). The combined organic extracts were dried over sodium sulfate, filtered and the solvent evaporated under vacuum to yield the product (9.24 g, 62.8 mmol, 40%) as a white solid. The obtained NMR data is in accordance with previously published results.<sup>[25]</sup>

$T_m$ : 104 °C<sup>[25]</sup>



**HRMS (ESI):** calcd. for  $\text{C}_9\text{H}_{10}\text{NO}_5$   $[\text{M}+\text{H}]^+$ : 148.07569; found: 148.07572

**$^1\text{H}$  NMR** (360.13 MHz,  $\text{CDCl}_3$ ):  $\delta$  7.07 (m, 5 H), 6.54 (s, 1 H), 4.71 (dd,  $J = 5.3, 2.5$  Hz, 1 H), 3.43 (ddd,  $J = 14.9, 5.3, 2.5$  Hz, 1 H), 2.86 (ddd,  $J = 14.9, 2.5, 0.9$  Hz, 1 H).

**$^{13}\text{C}$  NMR** (90.6 MHz,  $\text{CDCl}_3$ ):  $\delta$  168.3, 140.3, 128.9, 128.3, 125.7, 50.5, 48.1.

## 11. NMR Spectra

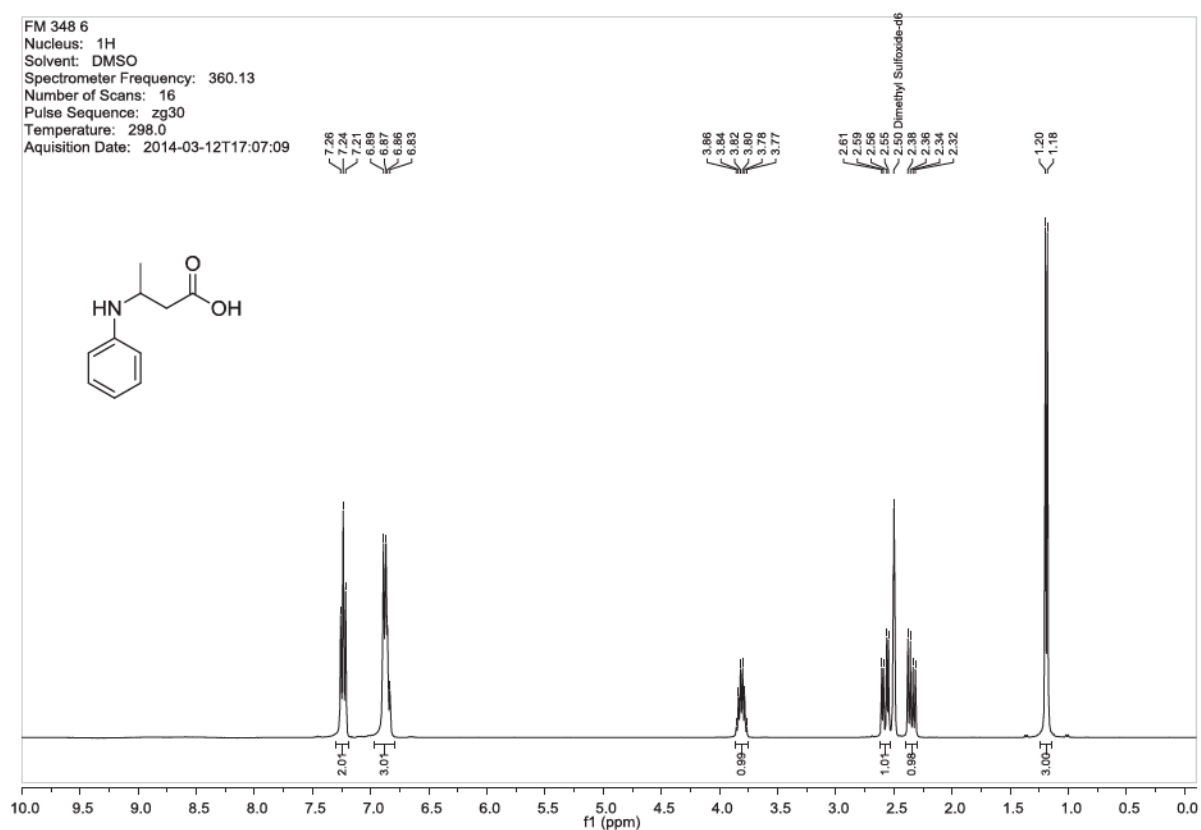
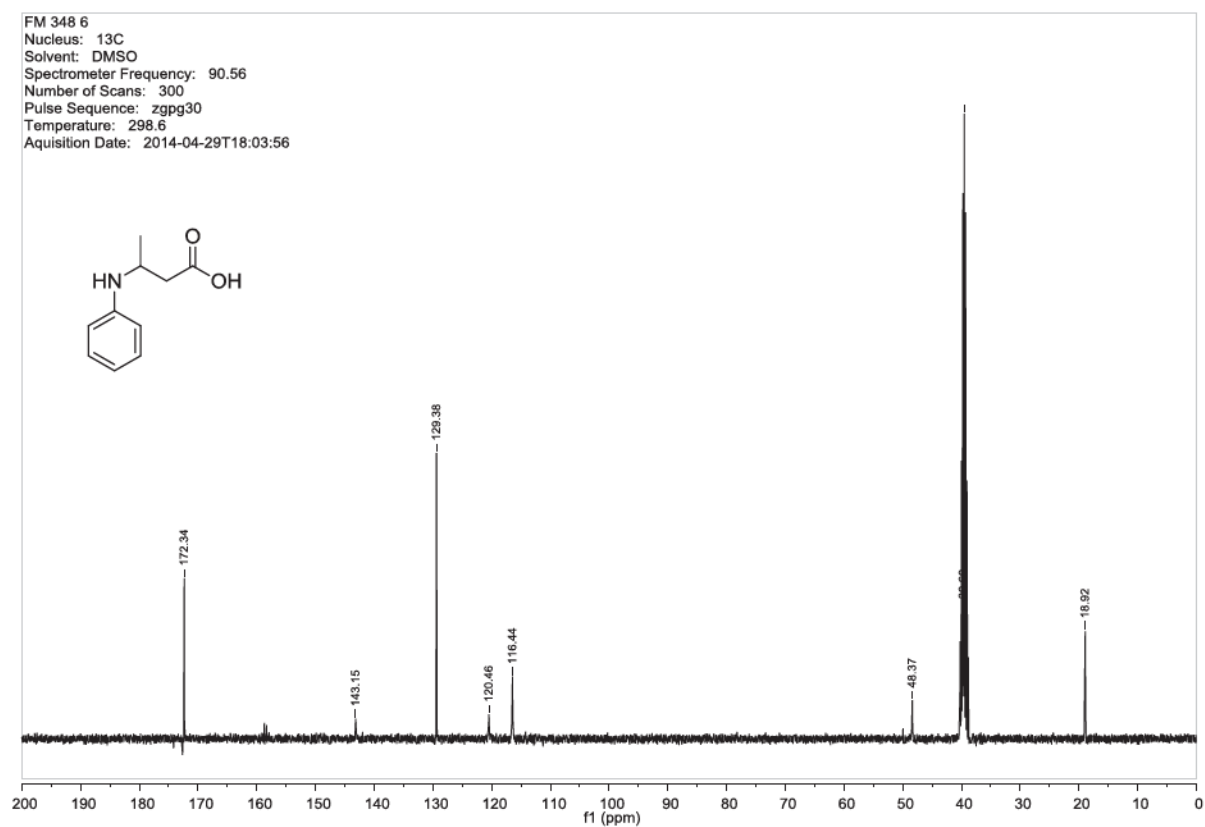


Figure S66:  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) of 3-(phenylamino)butanoic acid.



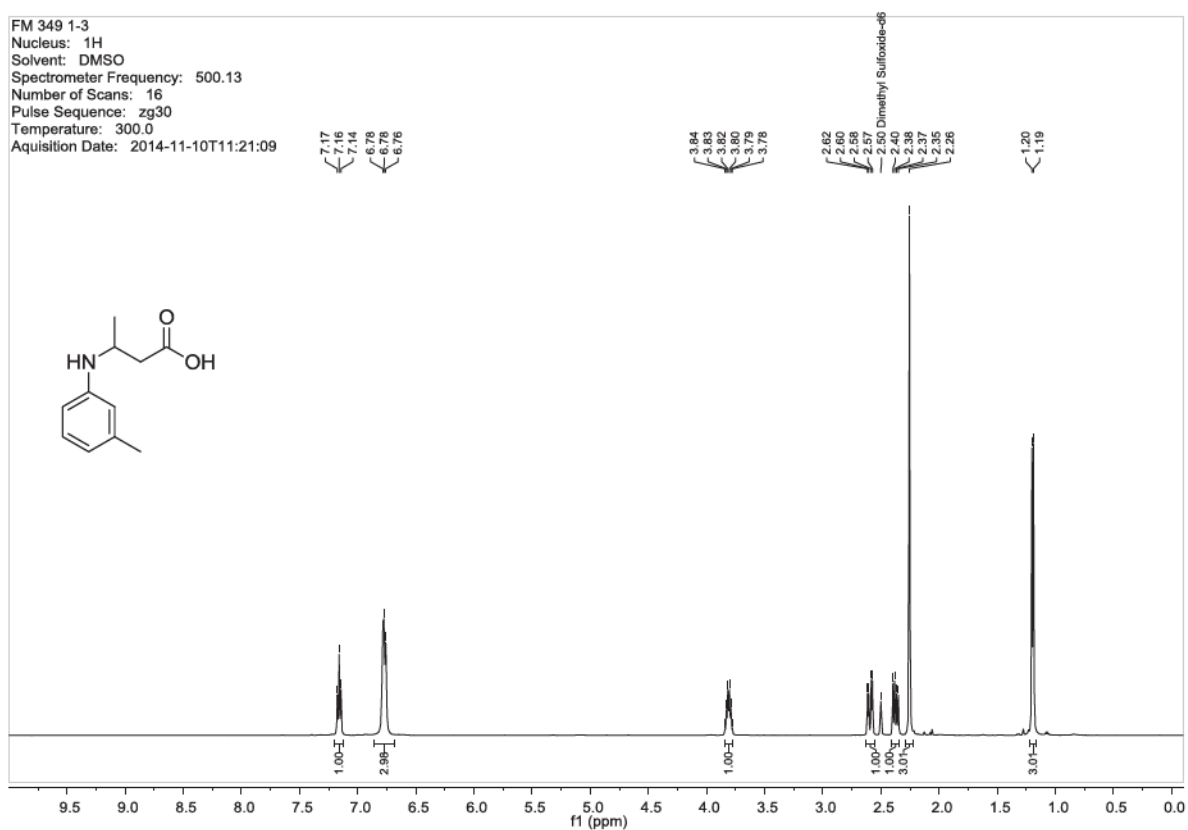
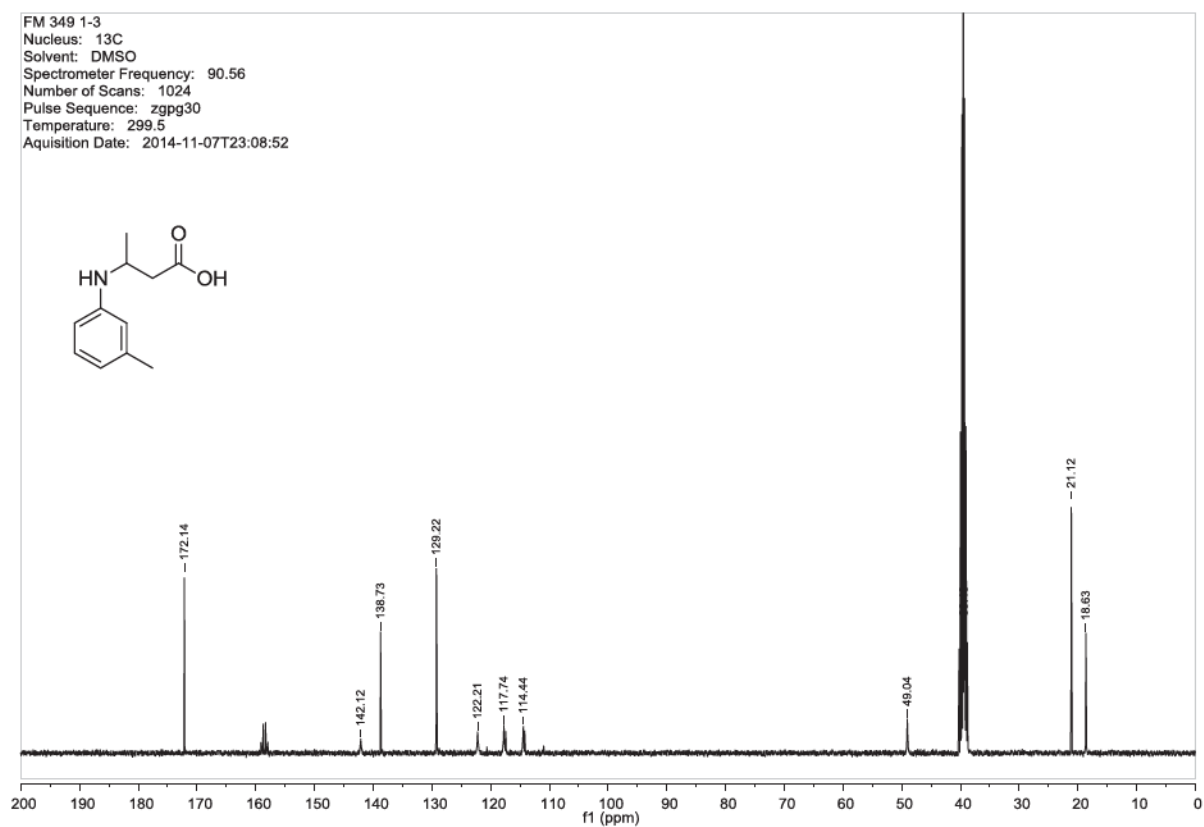


Figure S67:  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) of 3-(*m*-tolylamino)butanoic acid.



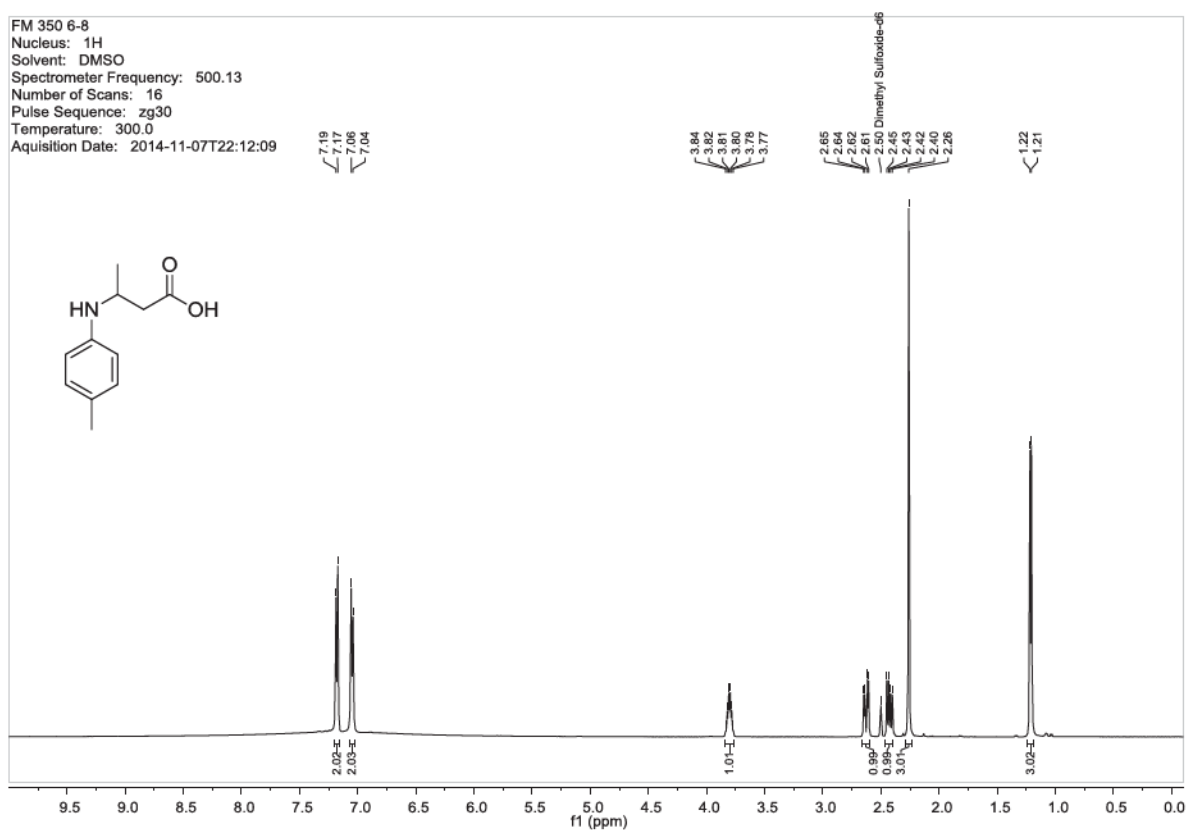
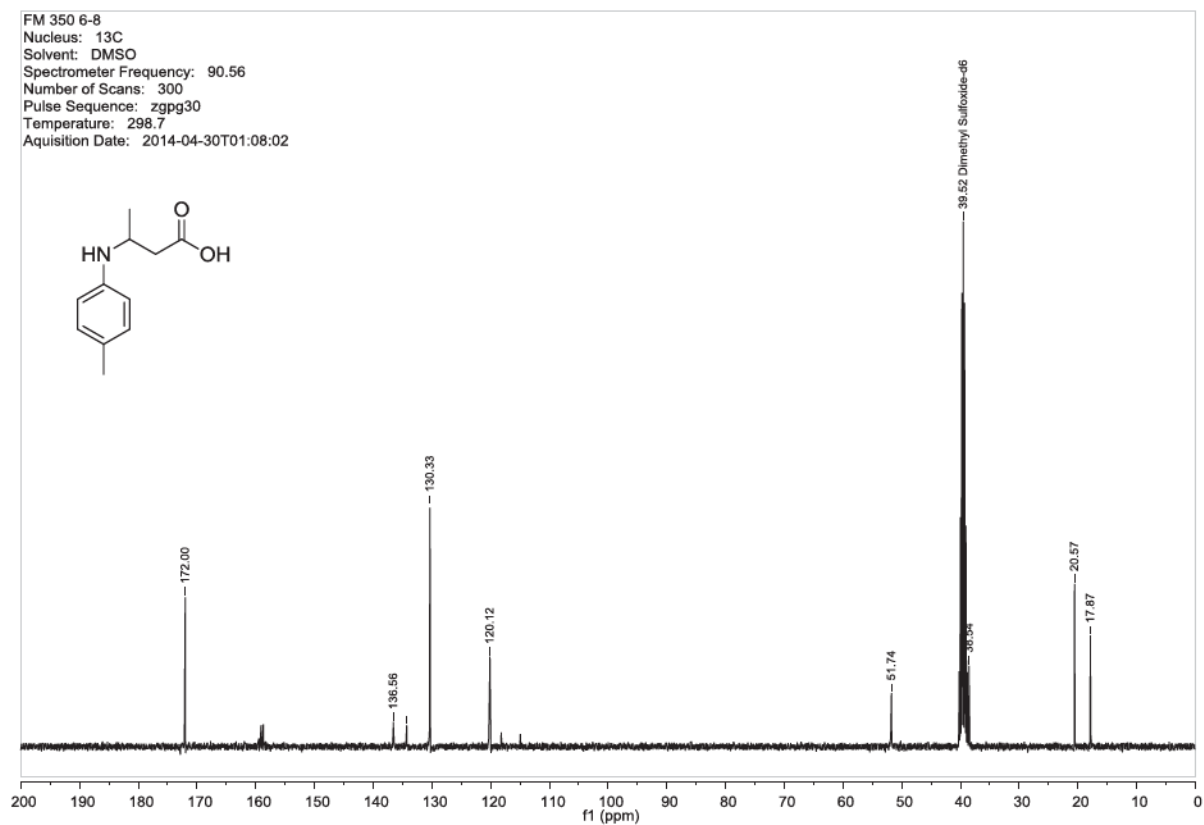
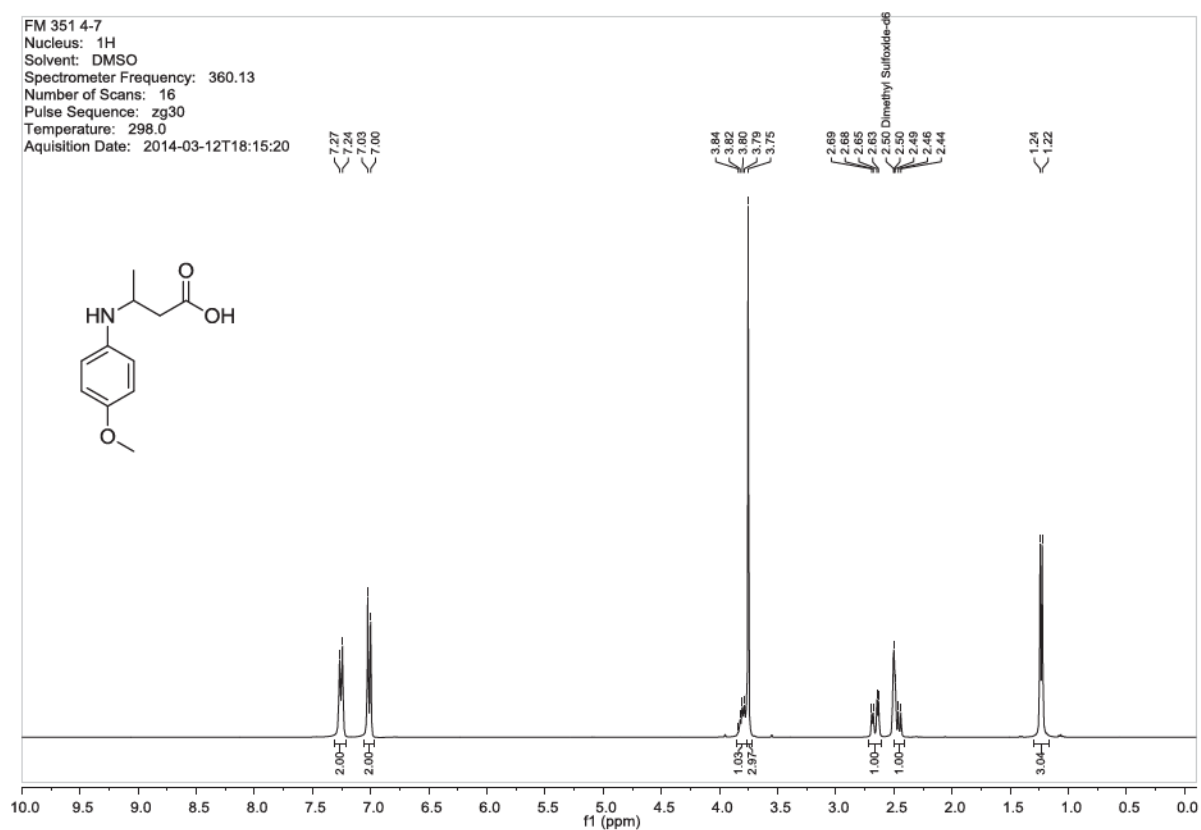
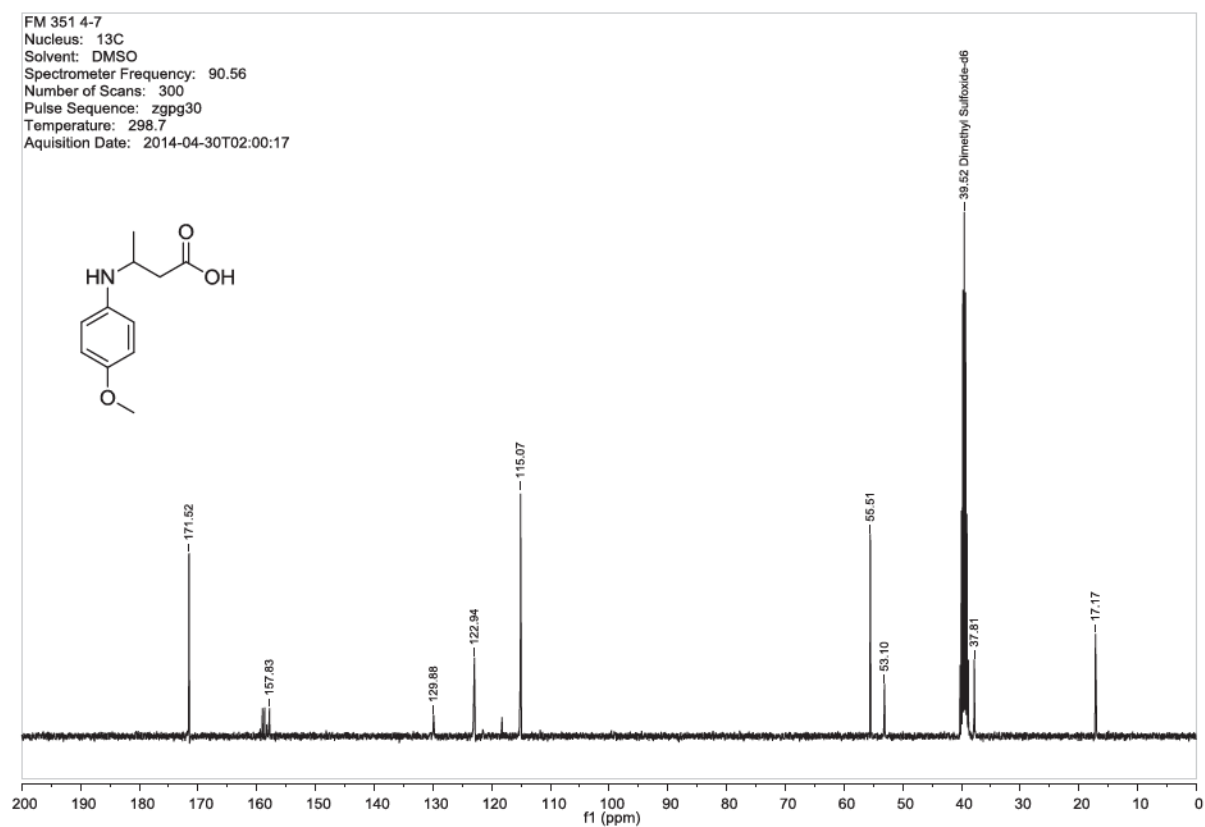


Figure S68:  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) of 3-(p-tolylamino)butanoic acid.





**Figure S69:**  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) of 3-((4-methoxyphenyl)amino)butanoic acid.



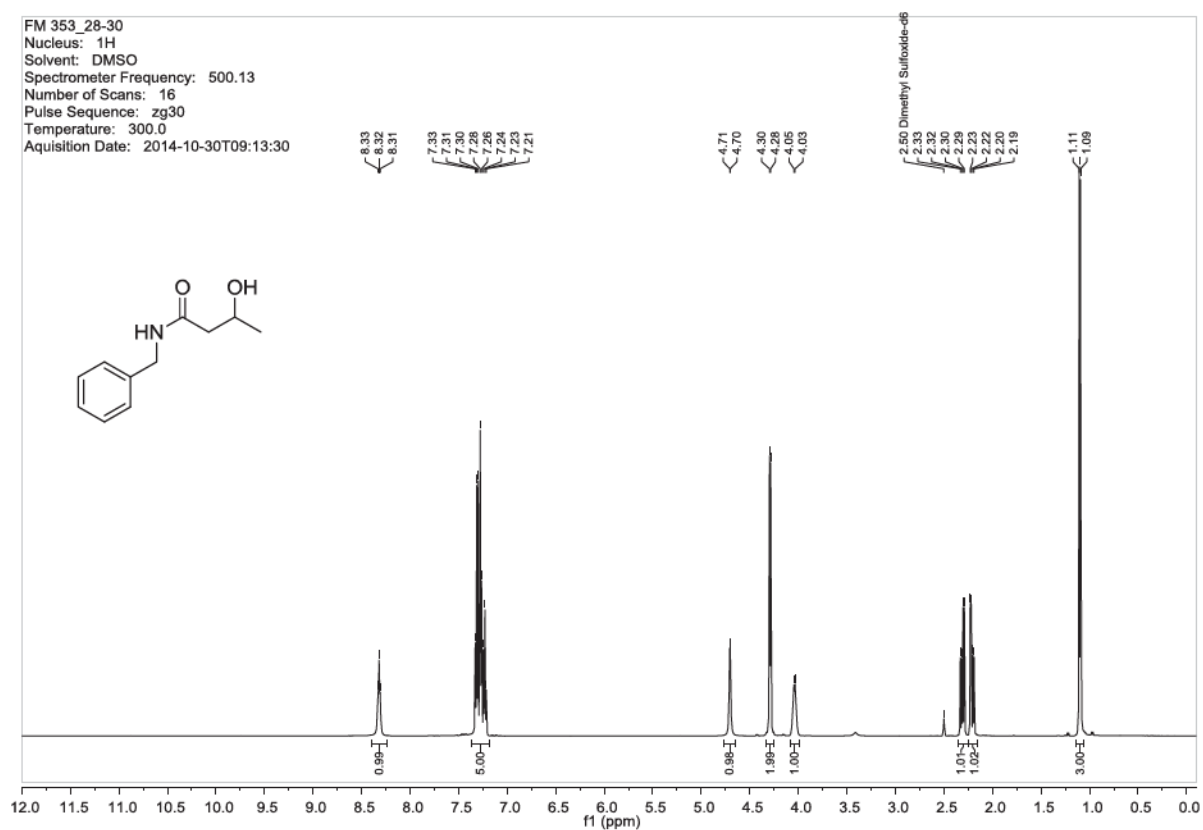
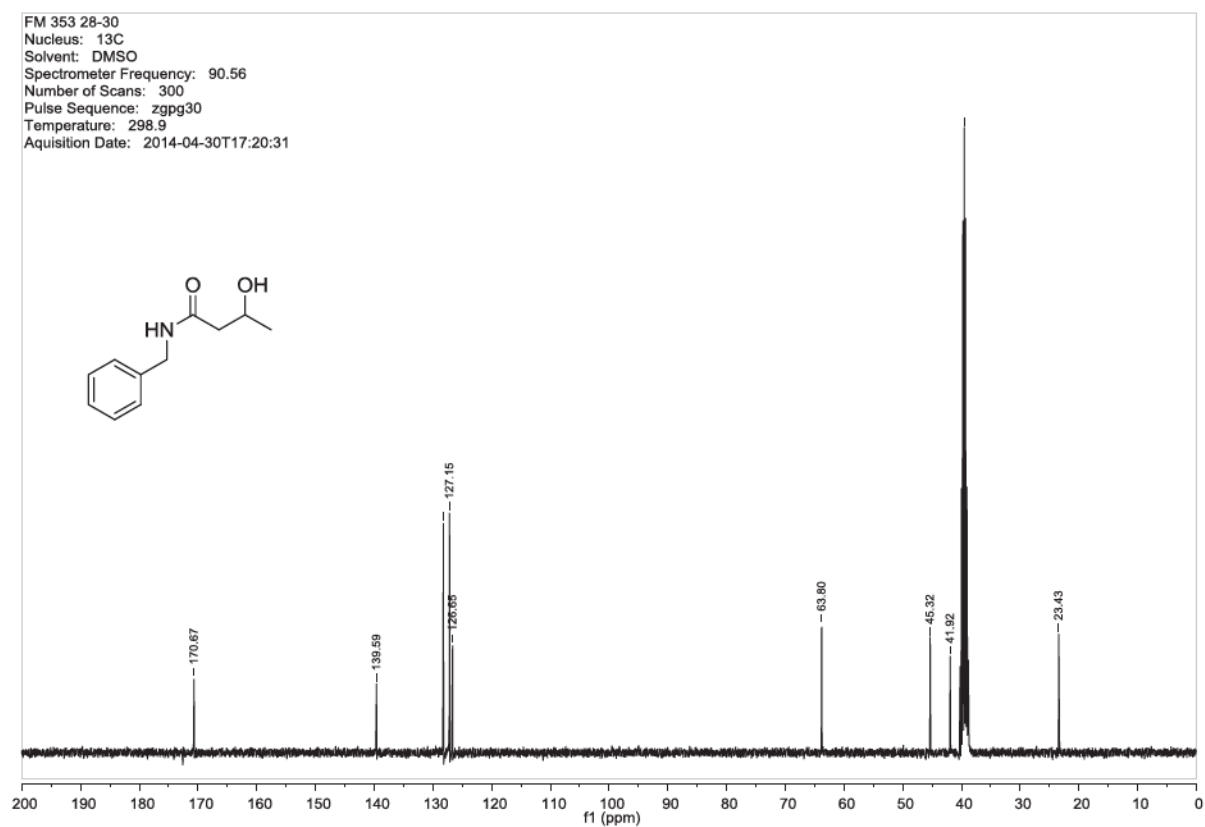
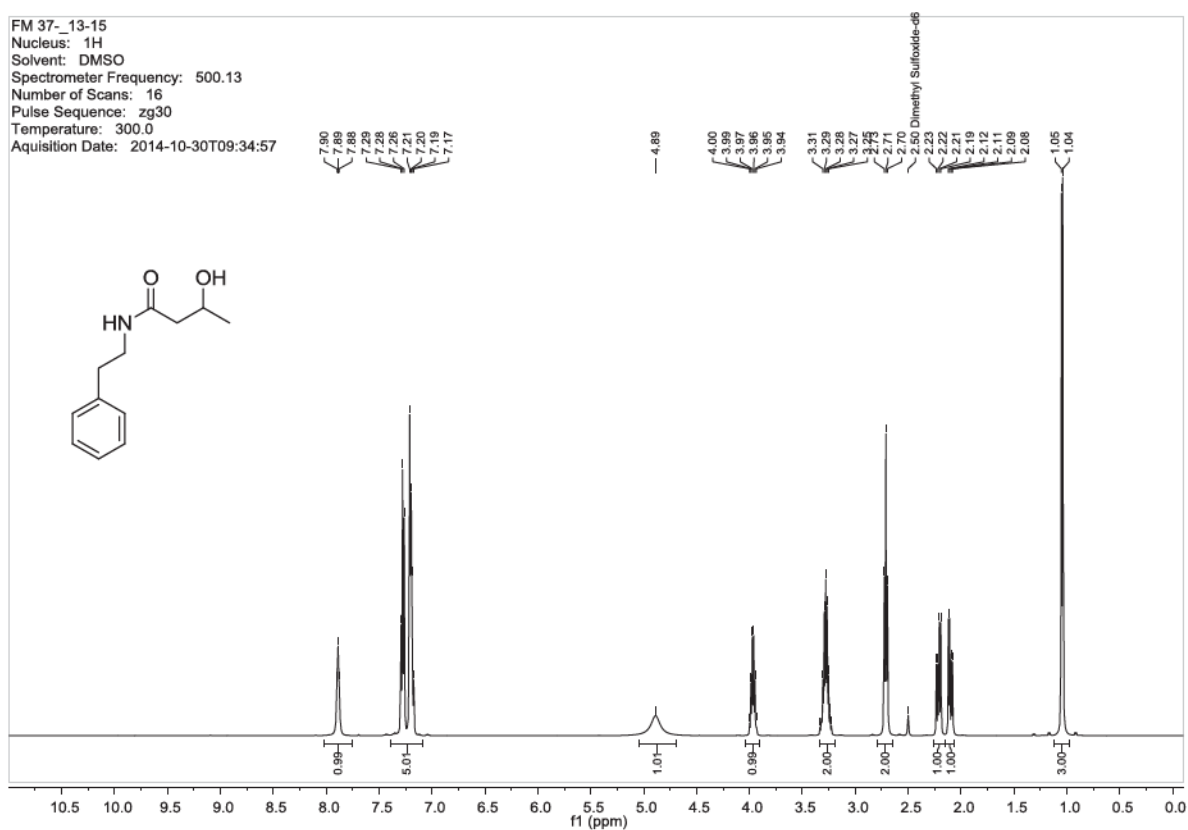
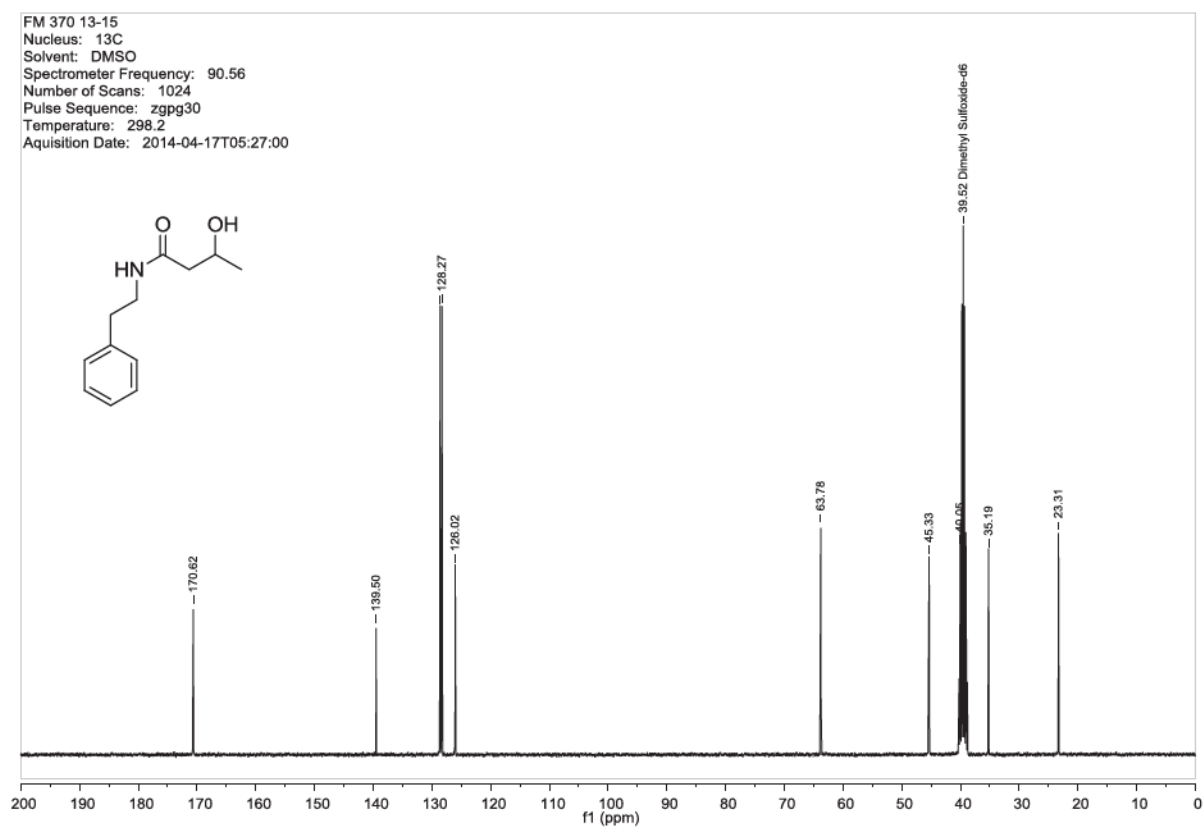


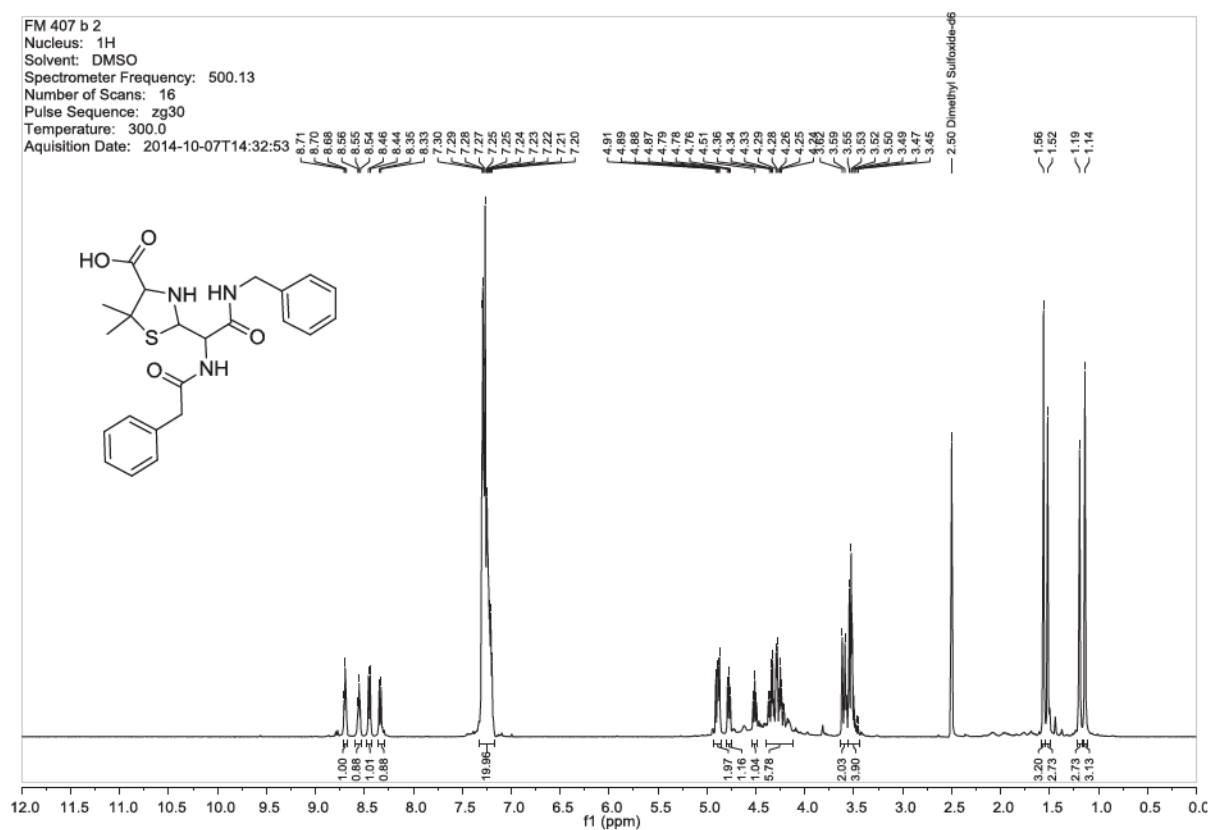
Figure S70:  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) of *N*-benzyl-3-hydroxybutanamide.



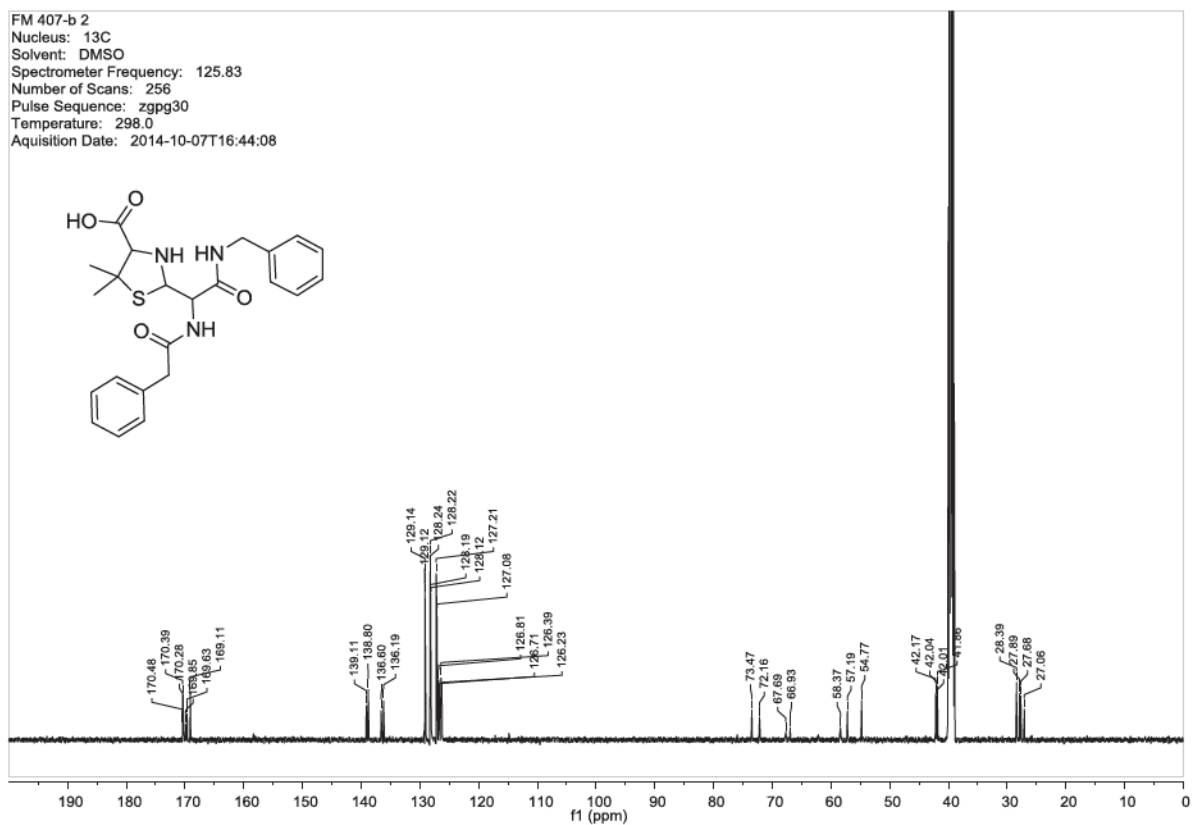


**Figure S71:**  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) of 3-hydroxy-N-phenethylbutanamide.

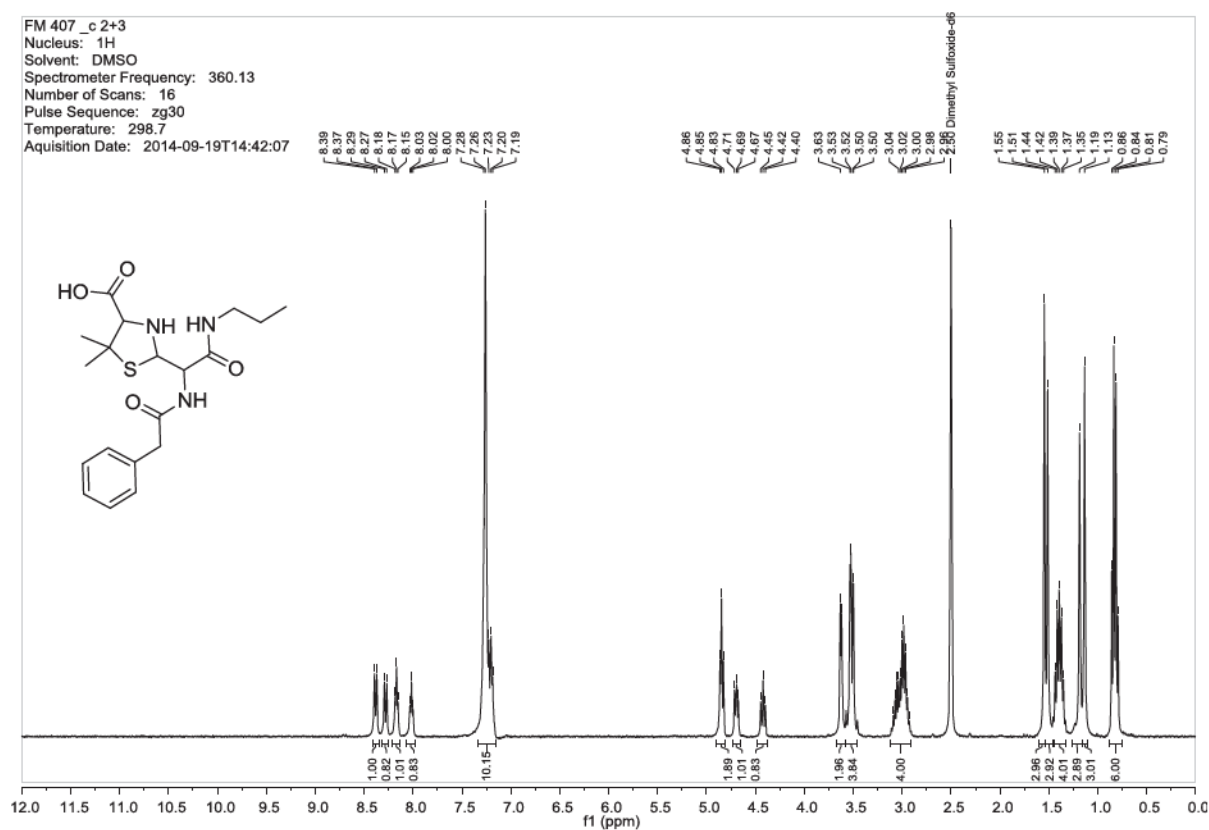




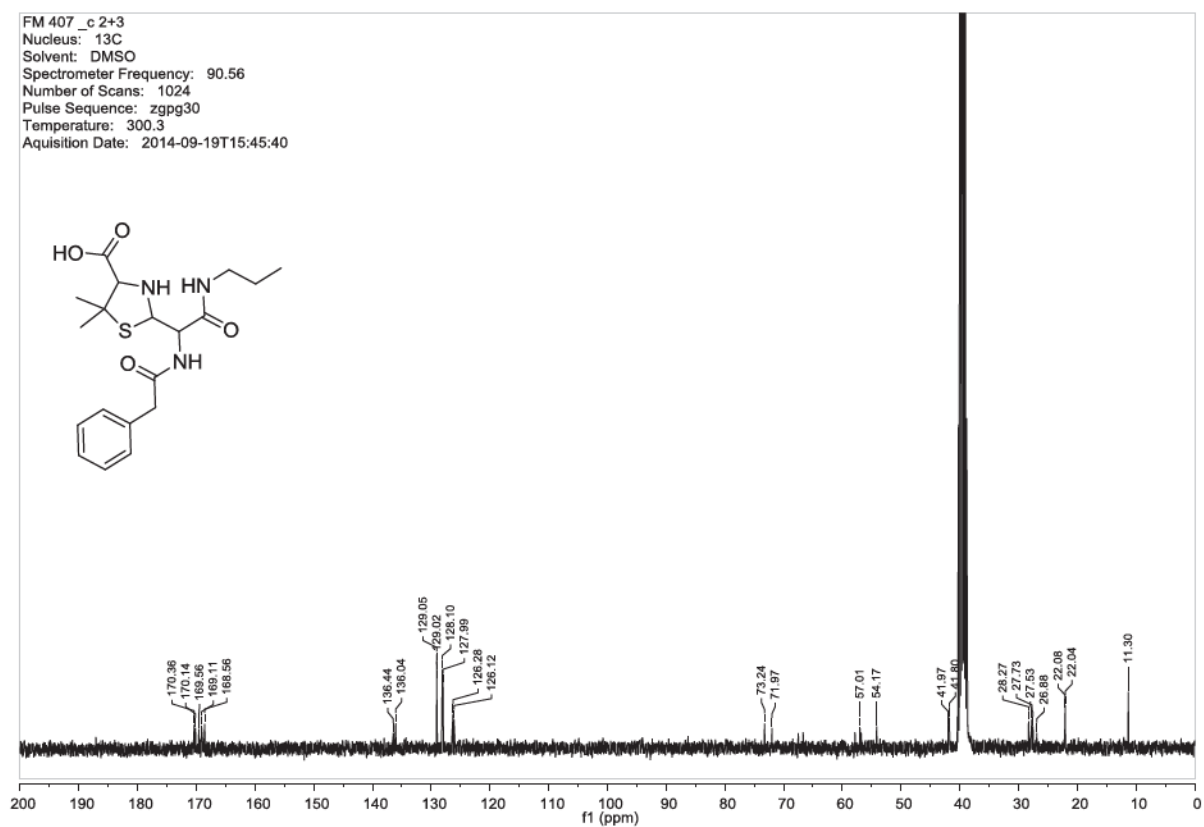
**Figure S22:**  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) of 2-(2-(benzylamino)-2-oxo-1-(2-phenylacetamido)ethyl)-5,5-dimethylthiazolidine-4-carboxylic acid.







**Figure S73:**  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) of 5,5-dimethyl-2-(2-oxo-1-(2-phenylacetamido)-2-(propylamino)ethyl)thiazolidine-4-carboxylic acid.



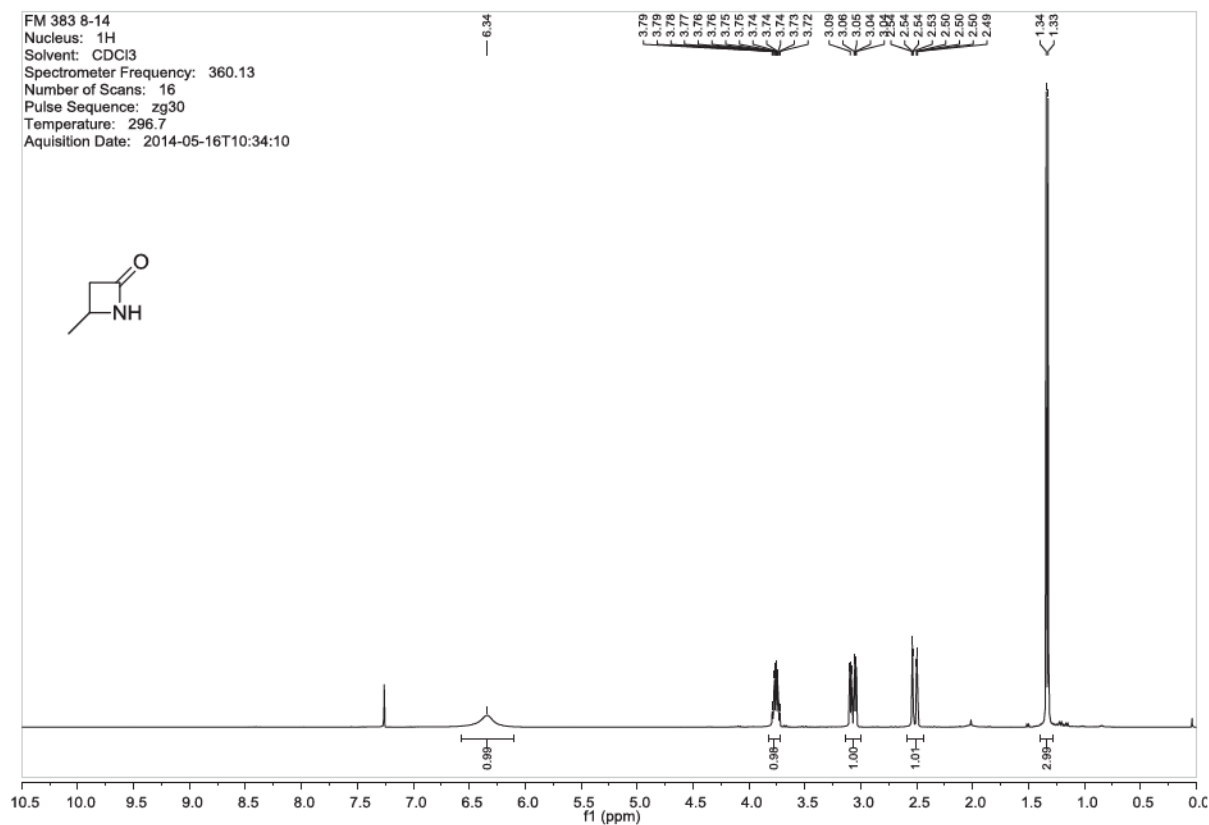
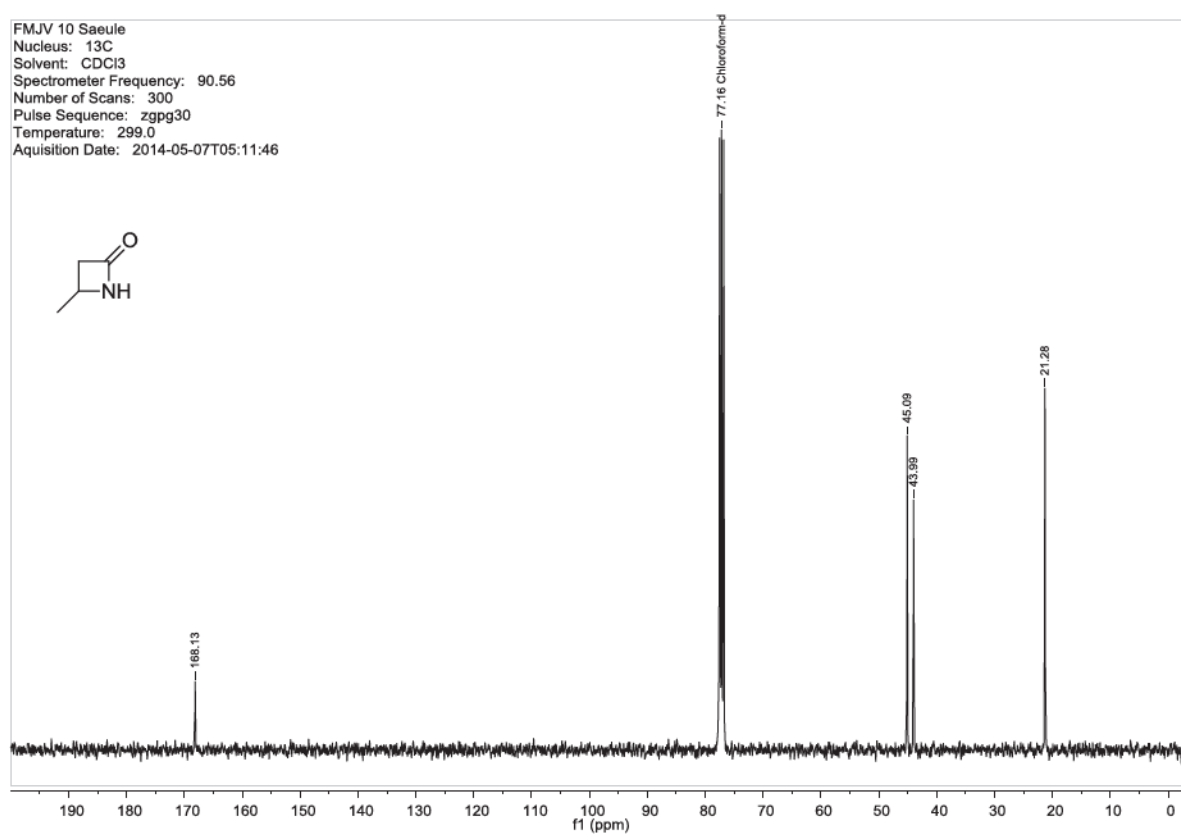


Figure S74:  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) of 4-methyl-1-azetidin-2-one.



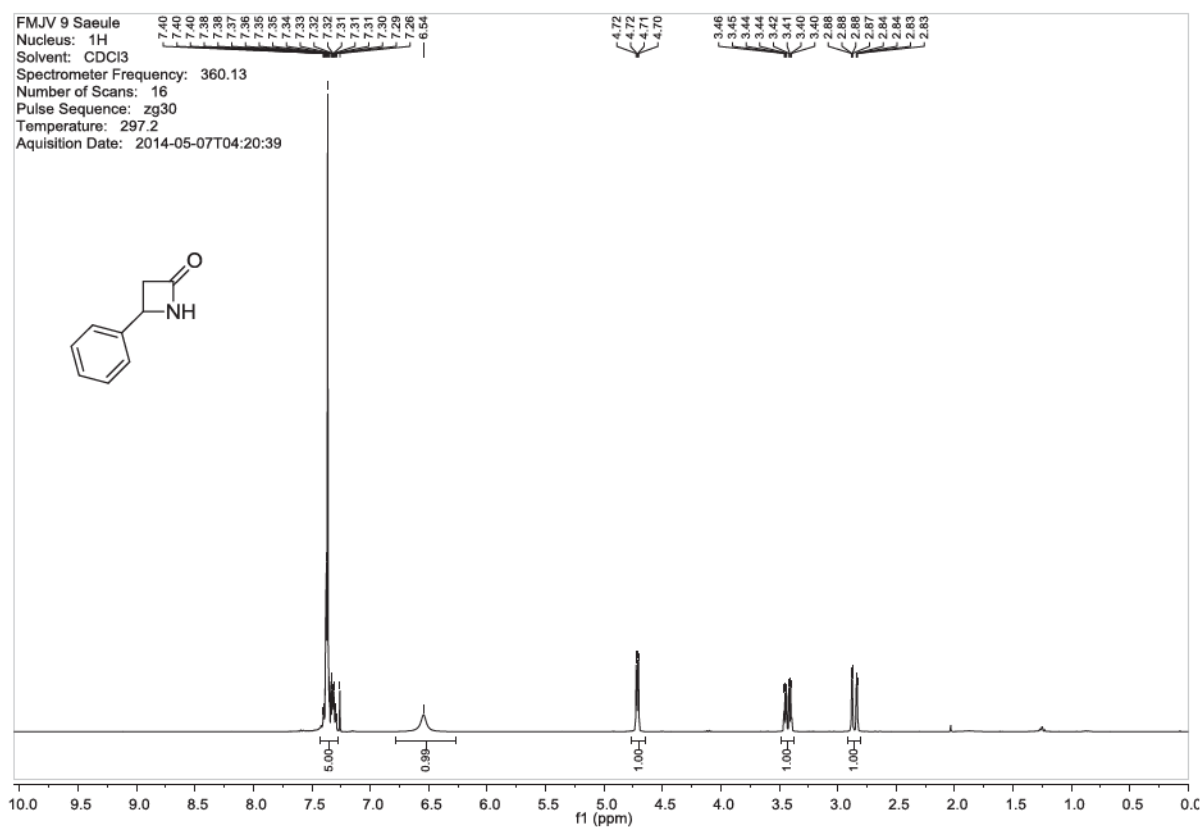
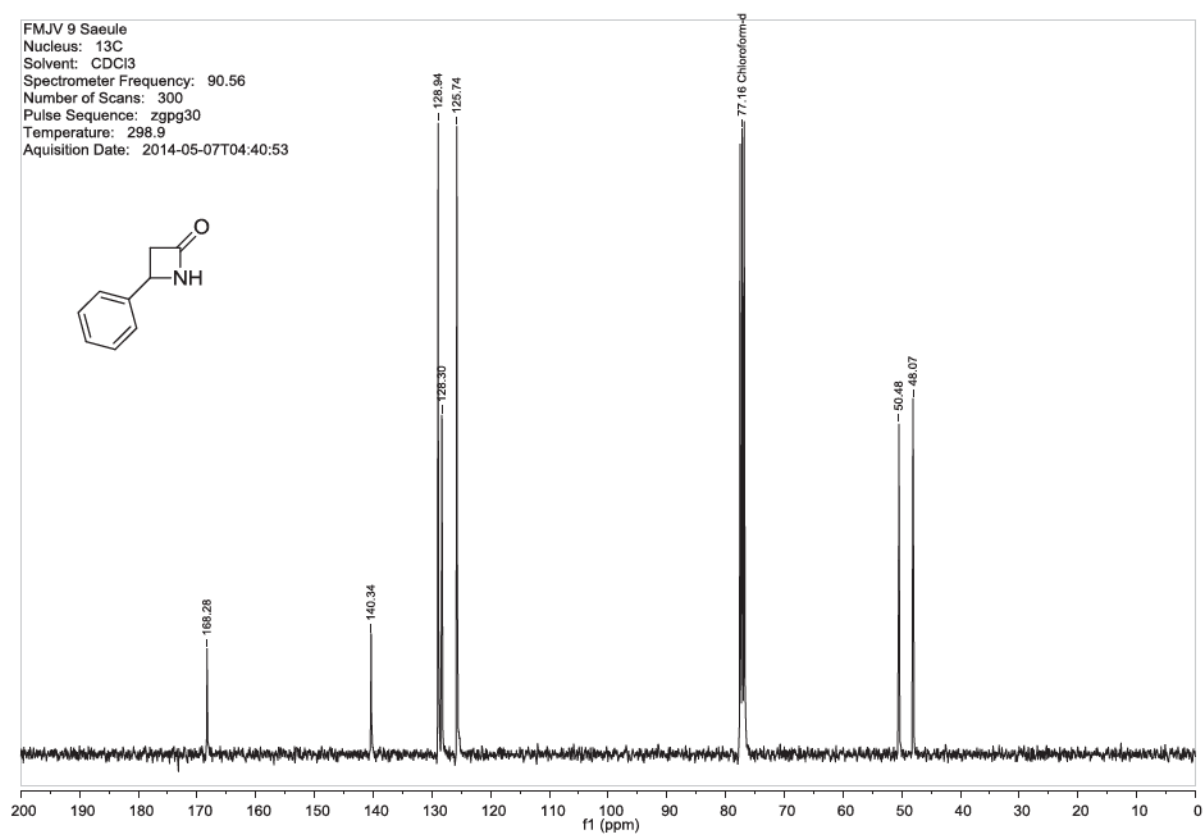


Figure S75:  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) of 4-phenyl-1-azetidin-2-one.



## 12. References

- [1] H. E. Gottlieb, V. Kotlyar, A. Nudelman, *J. Org. Chem.* **1997**, *62*, 7512-7515.
- [2] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R.E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, AMBER 11, University of California, San Francisco, **2010**.
- [3] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926-935.
- [4] a) J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157-1174; b) J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247-260.
- [5] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781-1802.
- [6] a) W. G. Hoover, *Phys. Rev. A* **1985**, *31*, 1695-1697; b) S. Nosé, *J. Chem. Phys.* **1984**, *81*, 511-519.
- [7] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327-341.
- [8] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer, C. Lennartz, *J. Mol. Struct. THEOCHEM* **2003**, *632*, 1-28.
- [9] a) S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200-1211; b) A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098-3100; c) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785-789; d) J. P. Perdew, *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822-8824; e) S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- [10] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio Jr, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Yeh Lin, T. Van Voorhis, S. Hung Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. Min Rhee, J. Ritchie, E. Rosta, C. David Sherrill, A. C. Simmonett, J. E. Subotnik, H. Lee Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172-3191.
- [11] L. Goerigk, S. Grimme, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670-6688.

- [12] J. Kästner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander, P. Sherwood, *J. Phys. Chem. A* **2009**, *113*, 11856-11865.
- [13] a) Y. Altun, *J. Solution Chem.* **2004**, *33*, 479-497; b) K. K. R. Poduval, M. Stobiecka, W. F. A. Dehaen, W. Dehaen, H. Radecka, J. Radecki, *Supramol. Chem.* **2010**, *22*, 413 – 419.
- [14] G. L. Flynn, *J. Pharm. Sci.* **1980**, *69*, 1109-1109.
- [15] F. Brotzel, Y. C. Chu, H. Mayr, *J. Org. Chem.* **2007**, *72*, 3679-3688.
- [16] S. Minegishi, H. Mayr, *J. Am. Chem. Soc.* **2003**, *125*, 286-295.
- [17] H. Mayr, T. Bug, M. F. Gotta, N. Hering, B. Irrgang, B. Janker, B. Kempf, R. Loos, A. R. Ofial, G. Remennikov, H. Schimmel, *J. Am. Chem. Soc.* **2001**, *123*, 9500-9512.
- [18] R. F. Jameson, G. Hunter, T. Kiss, *J. Chem. Soc. Perkin Trans. 2* **1980**, 1105-1110.
- [19] R. Gosmini, V. L. Nguyen, J. Toum, C. Simon, J.-M. G. Brusq, G. Krysa, O. Mirguet, A. M. Riou-Eymard, E. V. Boursier, L. Trottet, P. Bamborough, H. Clark, C.-w. Chung, L. Cutler, E. H. Demont, R. Kaur, A. J. Lewis, M. B. Schilling, P. E. Soden, S. Taylor, A. L. Walker, M. D. Walker, R. K. Prinjha, E. Nicodème, *J. Med. Chem.* **2014**, *57*, 8111-8131.
- [20] J.-i. Sakaki, S. Kobayashi, M. Sato, C. Kaneko, *Chem. Pharm. Bull.* **1989**, *37*, 2952-2960.
- [21] R. P. Maskey, I. Kock, M. Shaaban, I. Grün-Wollny, E. Helmke, F. Mayer, I. Wagner-Döbler, H. Laatsch, *Polym. Bull.* **2002**, *49*, 87-93.
- [22] D. S. Holmes, R. C. Bethell, N. Cammack, I. R. Clemens, J. Kitchin, P. McMeekin, C. L. Mo, D. C. Orr, B. Patel, *J. Med. Chem.* **1993**, *36*, 3129-3136.
- [23] T. Ogawa, K. Tomisawa, K. Sota, *Chem. Pharm. Bull.* **1988**, *36*, 1957-1962.
- [24] G. Bringmann, T. Geuder, *Synthesis* **1991**, *1991*, 829-831.
- [25] K. Hemming, M. N. Khan, V. V. R. Kondakal, A. Pitard, M. I. Qamar, C. R. Rice, *Org. Lett.* **2011**, *14*, 126-129.



**V.6 Paper [6]: "Theoretical and experimental studies on *Staphylococcus aureus* pyridoxal kinase reveal a facile interaction network crucial for phosphorylation", I. D. Blank, V. Kirsch, S. A. Sieber, C. Ochsenfeld, (to be submitted).**

## **Preface**

This paper draft does not yet include the experimental data, like kinetic studies on mutants, etc. Only the results of the theoretical investigation are shown.





**Theoretical studies on *Staphylococcus aureus* pyridoxal kinase  
reveal a facile interaction network crucial for phosphorylation**

I. D. Blank<sup>a</sup>, V. Kirsch<sup>b</sup>, S. A. Sieber<sup>b</sup>, C. Ochsenfeld<sup>a</sup>

- [a] I. D. Blank, Prof. Dr. C. Ochsenfeld  
Center for Integrated Protein Science Munich (CIPSM) at the Department of  
Chemistry, University of Munich (LMU Munich)  
Chair of Theoretical Chemistry, Department of Chemistry, University of Munich  
(LMU Munich)  
Butenandtstr. 7, D-81377 Munich, Germany
- [b] V. Kirsch, Prof. Dr. S. A. Sieber  
Center for Integrated Protein Science Munich (CIPSM)  
Institute of Advanced Studies IAS, Department of Chemistry  
Chair of Organic Chemistry II, Technische Universität München (TU Munich)  
Lichtenbergstrasse 4, D-85747 Garching, Germany

## Introduction

New strategies to target *S. aureus* are urgently needed as the spread of methicillin resistant strains (MRSA) limits current therapeutic treatments. *Staphylococcus aureus* represents a major pathogen causing severe infections of skin, lung and bones. Research into novel pathways includes metabolism and enzymes with essential functions for the bacterial cell. *S. aureus* requires pyridoxal phosphate (PLP) to maintain the activity of essential PLP-dependent enzymes. PLP is either supplied by de novo synthesis, pyridoxal (PL) uptake or salvage of PL liberated from enzyme turnover. Conversion of PL to PLP is catalyzed by specialized kinases (PLKs). PLKs utilize a conserved basic residue (cysteine 214) in their active site that deprotonates the 5'-hydroxy group of PL thereby facilitating a nucleophilic attack onto  $\gamma$ -phosphate of ATP, leading to PLP. Another cysteine, cysteine 110, is also supposed to have essential functions for the mechanism and is located in a flexible lid region (CC-PLK). This cysteine 110 was first identified as the binding site of rugulactone, an irreversible natural product inhibitor of CC-PLK<sup>1</sup>. Structural studies with apo and holo enzymes revealed a conformational change upon PL and ATP binding leading to an unexpected hemithioacetal intermediate formed between the 4'-aldehyde of PL and Cys110<sup>2</sup>. Indeed mutational studies with a Cys110Ala variant confirmed an essential role of this residue for PL catalysis. Interestingly, related substrates lacking the 4'-aldehyde such as pyridoxine (PN) as well as 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP), displayed a 4-10 fold reduced turnover compared to wild type PLK but could still be processed by the Cys110Ala mutant. Moreover, while a mutation of catalytic Cys214 to aspartate enhanced PL turnover by 2-fold, the same mutant was incapable of PN and HMP turnover. Based on these results a mechanism of PL phosphorylation has been postulated by which substrate binding reshapes the active site and induces lid closure. Thereby, Cys110 forms a hemithioacetal resulting in a net negative charge at the 4'-oxygen atom which was hypothesized to intramolecularly deprotonate the adjacent 5'-hydroxy group. In turn this nucleophile reacts with  $\gamma$ -phosphate of ATP resulting in PLP formation, lid opening and release. We here utilize in depth molecular dynamic simulations to dissect the interaction network crucial for PL binding and turnover. Based on these results, a new phosphorylation mechanism was unraveled using the QM/MM (Quantum mechanics/molecular mechanics) approach and revisits the originally proposed one. The results of this study show that hemithioacetal formation fixes the complex to facilitate proton

transfer from Cys214 to  $\gamma$ -phosphate of ATP. Subsequently, a facile coordination sphere composed of an  $\text{Mg}^{2+}$  ion and crucial residues in close proximity align the 5'OH group for phosphotransfer.

### Computational Details

The crystal structure of pyridoxal kinase from *Staphylococcus aureus* in complex with AMP-PCP and pyridoxal (PDB code: 4C5N)<sup>2</sup> was used as the starting structure.

We used the monomer in chain D, because in this chain Cys214 is resolved. The missing loop (residue 111 to 117) was modeled via homology modelling based on the crystal structure 4C5M (chain B). Water molecules within a distance of 5 Å of AMP-PCP and pyridoxal were kept. AMP-PCP was changed back to ATP. Based on other kinase structures, ATP normally coordinates to  $\text{Mg}^{2+}$ . As the coordination sphere of the water molecule (residue 2035) strongly indicates an  $\text{Mg}^{2+}$  ion, it was replaced by  $\text{Mg}^{2+}$  (see Figure 1).

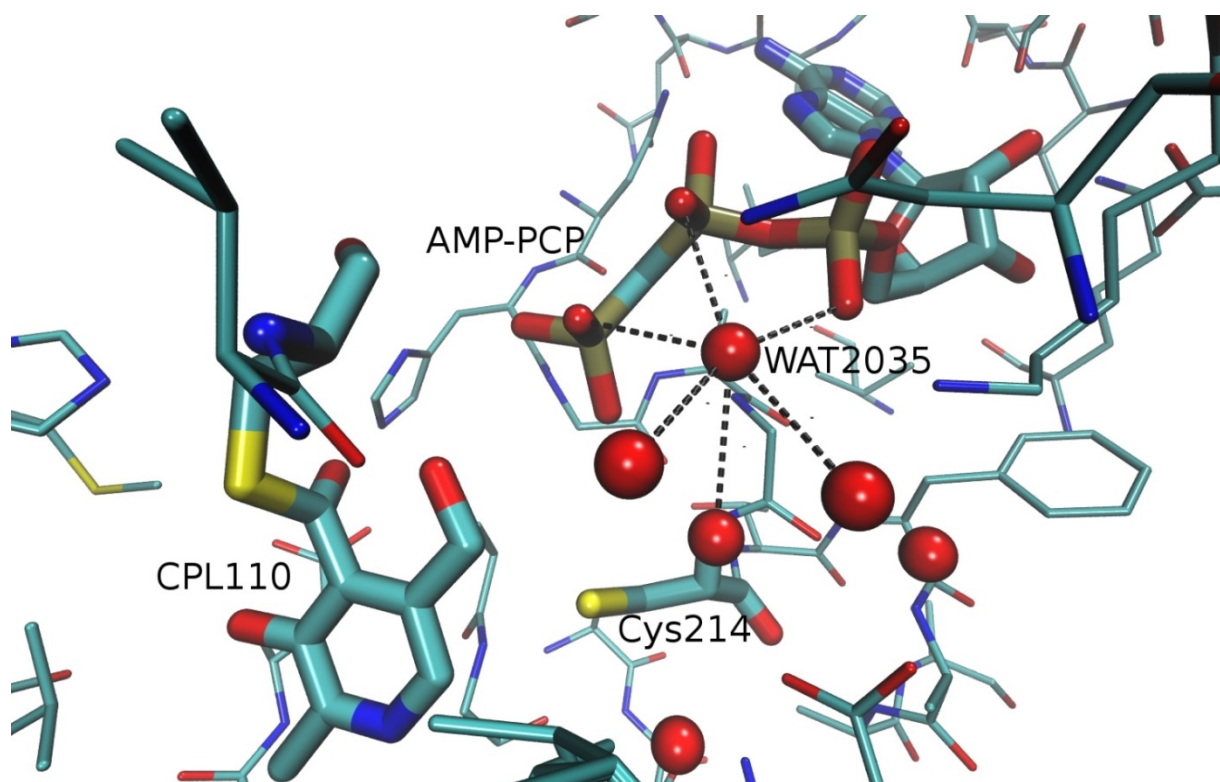


Figure 1: X-Ray structure (PDB code: 4C5N), where WAT2035 shows a typical coordination sphere for a  $\text{Mg}^{2+}$  ion and is therefore replaced by  $\text{Mg}^{2+}$  in our work.

The protonation state of the active site was determined based on PROPKA<sup>3-6</sup>. XLEAP (AmberTool)<sup>7</sup> has been used to add hydrogen atoms to the X-ray structure, to neutralize the system with sodium ions and to solvate it in a box of TIP3P water<sup>8</sup> with

a buffer of 10 Å around the solute. The parameters for ATP were taken from Meagher et al.<sup>9</sup>. GAFF parameters<sup>10</sup> have been assigned to Pyridoxal bound to Cys214 (CPL) with ANTECHAMBER<sup>11</sup>. For force field molecular dynamics (FF-MD) simulations we used the NAMD engine<sup>12</sup> with ff10 (ff99SB for proteins, ff99bsc0 and chi.OL3 for nucleic acids)<sup>7</sup>. Particle mesh Ewald summation (PME) with a cutoff value of 10 Å and periodic boundary conditions were employed (see SI, section MD preparation). For QM/MM structure optimizations the DL-POLY implementation within ChemShell<sup>13</sup> (AMBER-FF) was combined with density functional theory (DFT) at the M06-2X-D3/6-31G\*\*<sup>14</sup> level of theory, employing the Q-Chem program package<sup>15</sup> for the QM part. M06-2X was chosen for optimization due to its particular low weighted total mean absolute derivation for reaction energies<sup>16</sup>. Geometry optimizations have been performed with the DL-FIND<sup>17</sup> module implemented in ChemShell<sup>13</sup>. Reaction paths were calculated using the adiabatic mapping approach with up to 180 atoms in the QM region and 2054 atoms in the relaxed region (Figure 2).

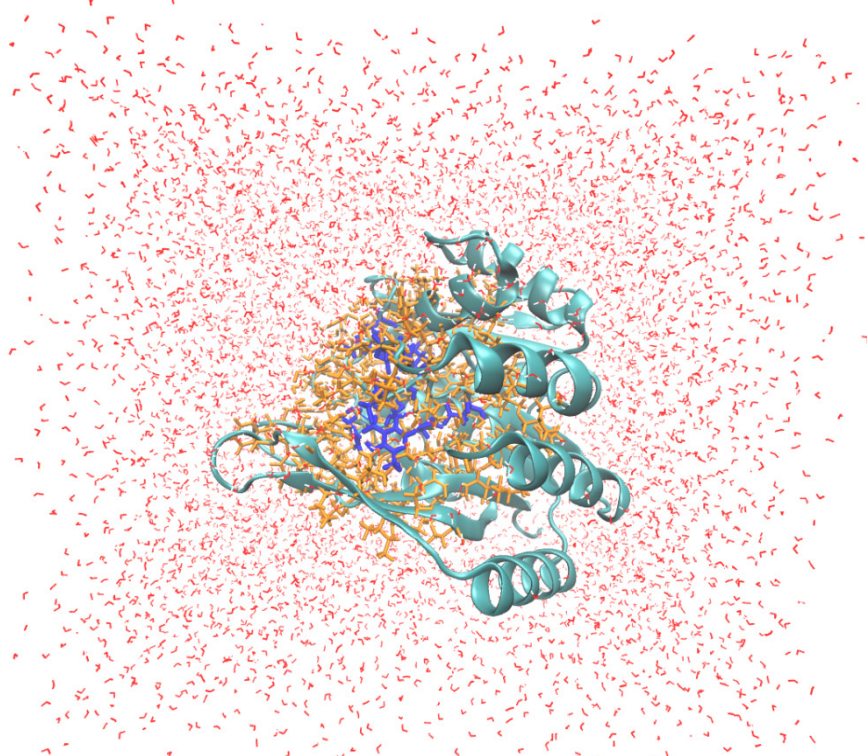


Figure 2: The simulated system is shown, where the QM region is depicted in blue and the relaxed region in orange.

## Results and Discussion

There is only one crystal structure of pyridoxal kinase in complex with AMP-PCP and pyridoxal (PDB code: 4C5N)<sup>2</sup>, where pyridoxal is covalently bound to Cys214 forming a hemithioacetal.

In this crystal structure, the ATP analogue AMP-PCP was used where the O( $\beta$ - $\gamma$ ) is substituted by a carbon atom which inhibits the phosphorylation. The non-hydrolysable AMP-PCP and ATP can potentially have different conformations within the binding site<sup>18</sup>. This could lead to a different coordination pattern, also influencing the number and position of Mg<sup>2+</sup> ions in the active site. Kinases are known to have one or two Mg<sup>2+</sup> ions in complex with ATP in the binding site. To investigate the number of Mg<sup>2+</sup> for pyridoxal kinase, we set up two systems with one and two Mg<sup>2+</sup> ions, respectively.

PROPKA<sup>3-6</sup> estimated a pK<sub>a</sub> value of 10.73 for Cys214, indicating that this residue is protonated under physiological pH which was confirmed by our quantum chemical calculations (see Section Phosphorylation Mechanism). This protonation state is in contradiction with the proposed mechanism<sup>2</sup>, which starts with the deprotonated Cys214 abstracting a hydrogen from the alcohol-group<sup>2</sup>. Afterwards the resulting alcoholate attacks the  $\gamma$ -phosphate of ATP. Our QM/MM calculations show, that the alcoholate is not stable due to immediate re-protonation by a water molecule nearby. We investigate the interaction pattern of surrounding amino acids to deduce their function and obtain indications about an alternative mechanism.

### Interaction pattern

We use FF-MD simulations to analyze the interaction pattern of several amino acids near the substrate in the active site, which are depicted in Figure 3. The resulting interaction pattern is shown in Figure 4. To reveal their function and their potential influence on the reaction mechanism, we correlated the results with the evolutionary conservation and find that Lys176, Lys77 and Lys111 seem to have more interaction partners than Lys179, which can explain their conservation. His210, Asp23 and Asp105 show strong interactions, bridging different secondary structure elements, which could indicate their importance in enzyme structure stability and therefore explain their evolutionary conservation. A detailed listing of the observed interactions and their concluded functions is given in Table 1.

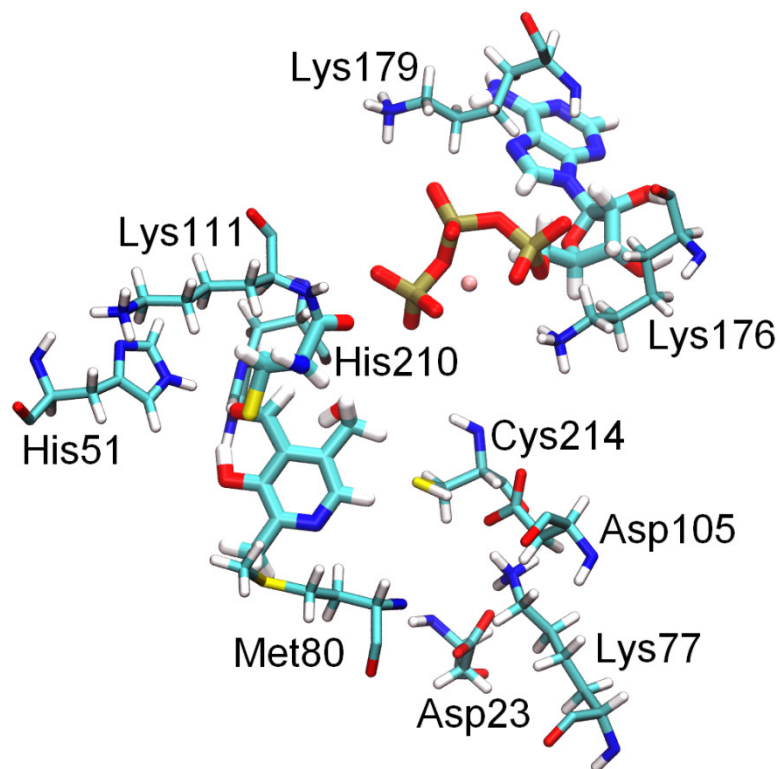


Figure 3: Active site residues of pyridoxal kinase.

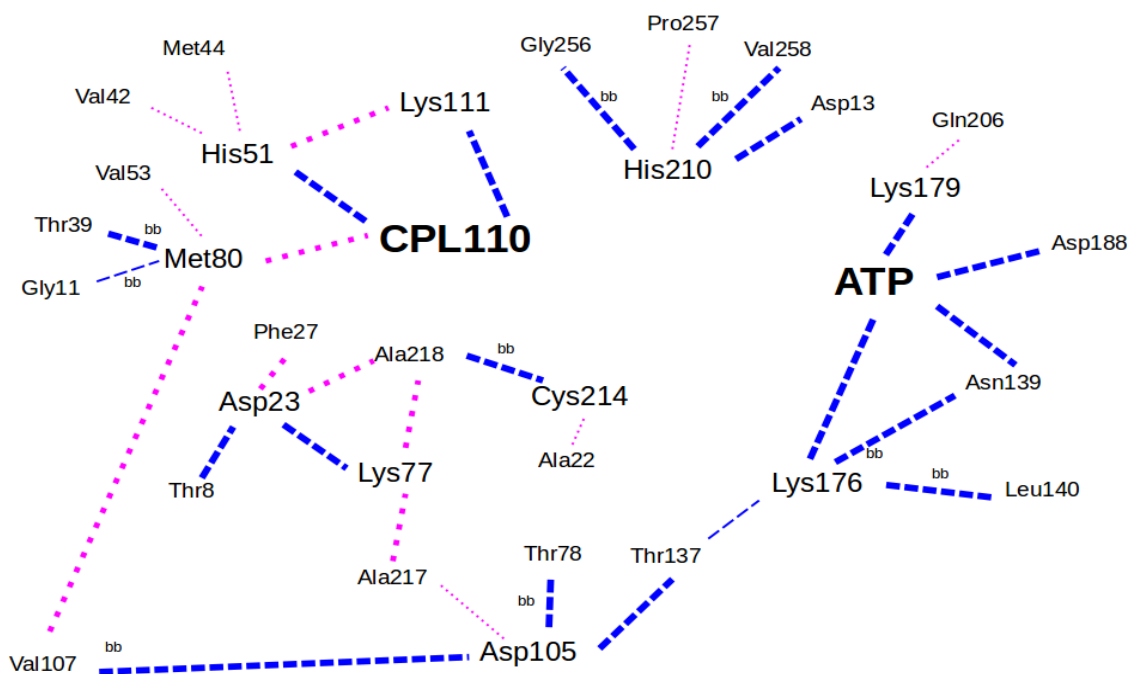


Figure 4: Interaction pattern between crucial residues. Non-polar interactions are colored in magenta, polar interactions in blue. Backbone interactions are indicated with bb.

Table 1: Interactions and function of different amino acids; based on the analysis of FF-MD simulations.

residue	interaction	function
Cys110	necessary for covalently binding of PL and correct orientation of PL within the active site	reaction
Lys111	ammonium group forms hydrogen bonds with oxygen of 5'-hydroxy group of CPL110 which will be phosphorylated; side chain forms van der Waals interactions with the imidazole ring system of His51	orientation of substrate
Lys176	ammonium group forms hydrogen bonds with $\alpha$ -phosphate of ATP and to the backbone carbonyl group of Thr137; backbone-backbone interaction of Lys176 to Leu140; Lys176 backbone carbonyl group interacts with Asn139 side chain	enzyme structure stability, ATP binding/orientation of substrate
Lys179	ammonium group forms hydrogen bonds with $\beta$ -phosphate of ATP; slight van der Waals interaction of side chain with Gln206	ATP binding/orientation of substrate, reaction
Lys77	ammonium group forms hydrogen bonds with Asp23 which connects the beta-sheet containing Lys77 with the nearby alpha-helix; Ala217 and Ala218 form van der Waals interactions with aliphatic part of Lys77 side chain	enzyme structure stability
His210	hydrogen bonds to Asp13; van der Waals interaction with Pro257; backbone-backbone interaction with Gly256 and Val258	enzyme structure stability
His51	van der Waals interactions with Lys111; slight van der Waals interactions with Val42 and Met44; forms hydrogen bond with oxygen in hemithioacetal of CPL110	orientation of substrate

Asp105	carboxylate group forms hydrogen bonds to backbone of Val107 and hydroxyl group of Thr137; aliphatic part of side chain forms van der Waals interaction with Ala217; backbone-backbone interaction with Thr78	enzyme structure stability
Cys214	slight van der Waals interactions with Ala22; backbone-backbone interaction with Ala218	reaction, enzyme structure stability
Met80	slight backbone-backbone interaction with Gly11; carbonyl oxygen forms hydrogen bond to hydroxyl group of Thr39; forms van der Waals interaction with Val53, Val107 and CPL110 (ring system and methyl group)	enzyme structure stability, orientation of substrate
Asp23	carboxylate group forms hydrogen bonds with hydroxyl group of Thr8 and ammonium group of Lys77; Phe27 and Ala218 form van der Waals interactions to aliphatic part of side chain of Asp23	enzyme structure stability
ATP (ribose/ phosphates)	H2' and H3' of ATP form hydrogen bonds to carboxylate group of Asp188; Asn139 forms hydrogen bonds to $\alpha$ -phosphate	reaction

This detailed investigation shows that the investigated amino acids can be grouped into three different sections: 1.) amino acids important for enzyme structure stabilization, 2.) amino acids important for binding/orientation of substrates and 3.) amino acids potentially important for the mechanism.

Amino acids important for enzyme structure stabilization are Asp23, Met80, Cys214, Asp105, Lys77, Lys176 and His210. For binding and orientation of the substrate PL, the amino acids Met80, His51, Lys179, Lys176 and Lys111 show important interactions to both substrates, either ATP or PL.

Finally, amino acids, which are potentially important for the mechanism, are Cys214, Cys110 and Lys179. Lys179 is potentially crucial for the mechanism because it brings the  $\gamma$ -phosphate in a favorable position for the phosphorylation reaction and could stabilize the transition state by interacting with the  $\beta$ -phosphate.



### Phosphorylation mechanism (QM/MM)

We investigated different possible mechanisms for the phosphorylation of pyridoxal which is covalently bound to Cys110, including the assumed mechanism. It has been suggested that as initial reaction the deprotonated Cys214 abstracts a proton of the 5' alcohol group<sup>2</sup>. Our calculations show, that the hydroxyl group will not reach the alcoholate form due to simultaneous re-protonation by a water molecule nearby. For a direct phosphorylation we calculated a barrier of 63 kcal/mol (see Figure 5). If Cys214 were deprotonated, the barrier would decrease by 11 kcal/mol to 52 kcal/mol, because of the concerted proton transfer from CPL to the negatively charged Cys214. This indicates that Cys214 is favorable in a deprotonated form during the phosphorylation reaction (see SI). As these reaction barriers are much too high for enzymatic reactions, direct phosphorylation seems not to be feasible. Although, unprotonated Cys214 lowers the barrier by 11 kcal/mol, according to PROPKA it is protonated in the beginning of the reaction. However, Cys214 can transfer its proton to  $\gamma$ -phosphate before the phosphorylation step. This is in line with experimental findings<sup>2</sup>, that the C214D mutant exhibits a considerably higher activity than the WT. Aspartate has a lower  $pK_a$  value compared to cysteine which is therefore able to protonate the  $\gamma$ -phosphate more easily.

In contrast to the proposed mechanism, we suggest protonation of the  $\gamma$ -phosphate as initial reaction step (Figure 6). This involves a reorientation of the thiol group of Cys214, away from Asp105, heading towards the  $\gamma$ -phosphate of ATP followed by a proton transfer of Cys214 to the  $\gamma$ -phosphate of ATP via water. This first reaction step has a barrier of 18 kcal/mol and is feasible within an enzymatic reaction (Figure 7).

If we now calculate the phosphorylation reaction as next step, the barrier for the phosphorylation decreases by 6 kcal/mol to 46 kcal/mol, which is still too high for an enzymatic reaction. The  $\gamma$ -phosphate is coordinated to  $Mg^{2+}$ , which prevents a movement towards CPL. Preparing for the phosphate transfer reaction, we substitute the  $\gamma$ -phosphate in the coordination sphere of  $Mg^{2+}$  by a water molecule. The coordination exchange does not result in a significantly decrease of the reaction barrier in the subsequent phosphorylation, but the stability of the product changes from 21 kcal/mol to -6 kcal/mol (reaction profile see SI) compared to the product when the  $\gamma$ -phosphate is still coordinated to  $Mg^{2+}$  after phosphorylation. Interchanging the first two reaction steps, does not lead to energetically favorable

geometries (see discussion in SI).

To make the reaction more favorable, we include Lys179 into the reaction coordinate, because we assessed Lys179 to be helpful by stabilizing the  $\beta$ -phosphate within the phosphorylation reaction. The reaction barrier for the phosphorylation decreases by 3 kcal/mol as Lys179 forms a strong hydrogen bond of its ammonium group with the  $\beta$ -phosphate of ATP. During this phosphorylation, the proton of the attacking alcohol group is transferred to Cys214 simultaneously. The first two steps are similar to the mechanism found for an auto-phosphorylation of a histidine kinase<sup>19</sup>. In contrast to this mechanism, we find that proton transfer from Lys179 to the  $\beta$ -phosphate is unfavorable. Instead, we observe a stabilizing effect. Unfortunately, there are no other amino acids nearby that can protonate the  $\beta$ -phosphate. This could be attributed to the loop conformation, not representing the actual reactive conformation of the closed loop *in vivo*, which also extensive MD simulations cannot correct. Together with a probably different coordination pattern of the analogue AMP-PCP, this could explain the big distance between CPL and  $\gamma$ -phosphate (4.36 Å) before the phosphate transfer reaction, which artificially increases the reaction barrier. We also investigated the influence of a second  $\text{Mg}^{2+}$  ion, which does not facilitate the phosphorylation reaction (see SI). So we conclude that only one  $\text{Mg}^{2+}$  is part of the active site.

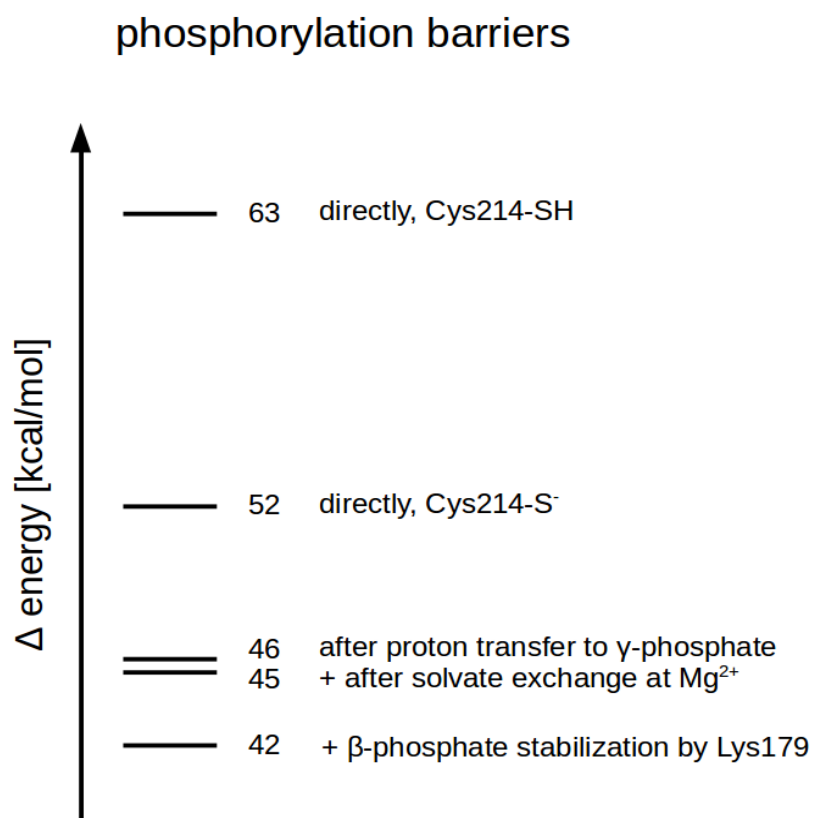


Figure 5: Phosphorylation barrier of CPL under different conditions, calculated at the M06-2X-D3/6-31G\*\* level of theory.

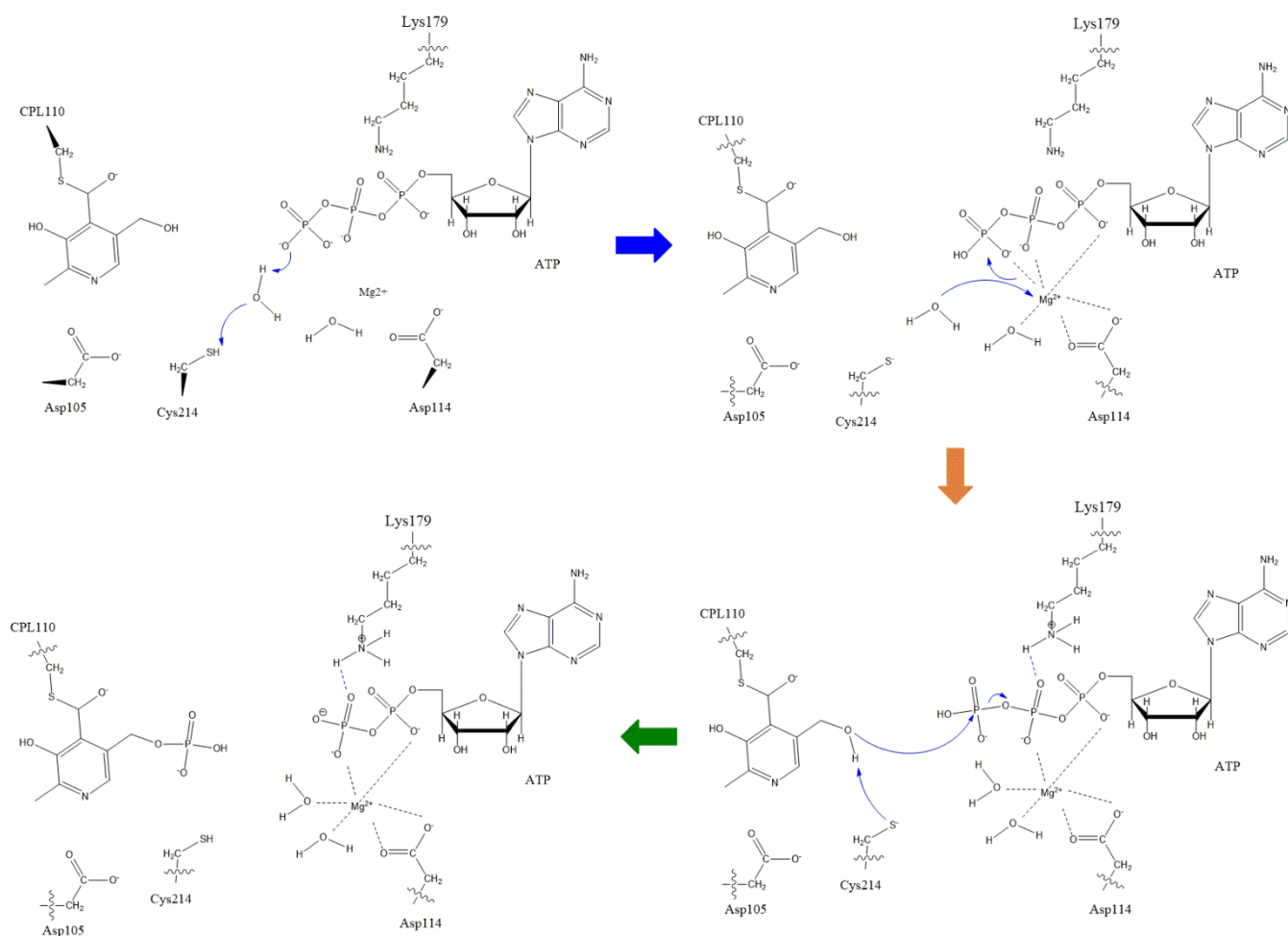


Figure 6: Phosphorylation mechanism of CPL by PLK. The color code of the arrows corresponds to the barriers in Figure 7.

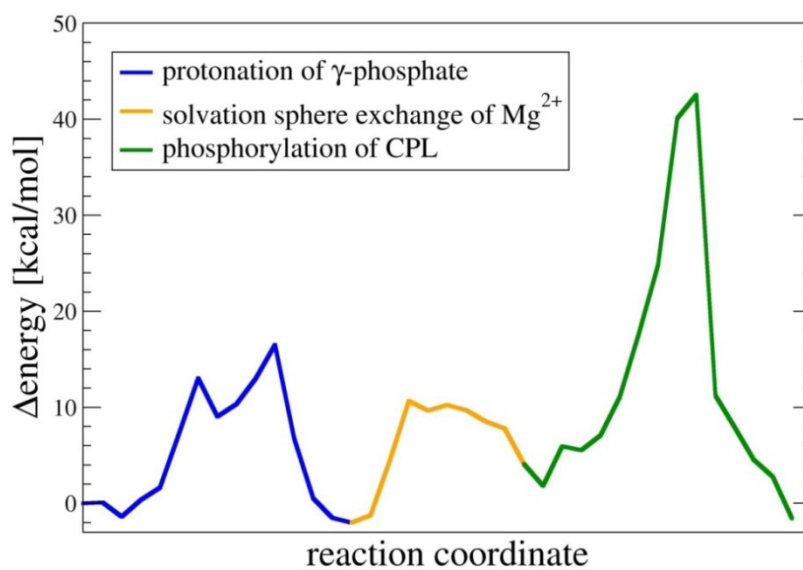


Figure 7: Reaction profile of the phosphorylation mechanism with the color code corresponding to Figure 6. Calculations have been performed at the M06-2X-D3/6-31G\*\* level of theory.

## Conclusion

We investigated the interaction pattern of amino acids within the active site by MD simulations and quantum-chemical calculations to illuminate their function within the enzyme and classified the investigated amino acids according to their function (enzyme stability, substrate binding and reaction mechanism). Additionally, the investigation of different influences on the phosphorylation reaction of PLK at different stages of the overall reaction mechanism lead us to the conclusion that only one  $Mg^{2+}$  ion is present in the active site, coordinating all phosphate groups of ATP.

We found that the suggested initial alcoholate formation of CPL<sup>2</sup> does not exist under physiological conditions due to simultaneous re-protonation by water. However, our calculations suggest that an initial proton transfer reaction from Cys214 to the  $\gamma$ -phosphate of ATP is feasible. This initial reaction could explain the higher activity of the C214D mutant compared to the wild type PLK<sup>2</sup>. In the second step a change in the solvation sphere of  $Mg^{2+}$  can take place, so that the  $\gamma$ -phosphate is not coordinated to  $Mg^{2+}$  anymore, leading to an energetically favorable phosphorylation product in the third step. The first and second reaction steps of our new mechanism are in line with the mechanism found for a histidine kinase<sup>19</sup>. In that case, a lysine residue transfers a proton via water to the  $\beta$ -phosphate. Our calculations indicate that Lys179 is not able to transfer a proton to the  $\beta$ -phosphate in the case of PLK, but stabilizes the  $\beta$ -phosphate through a strong hydrogen bond, leading to a decreased reaction barrier. The resulting barrier height of 42 kcal/mol shows, that not all details of the mechanism are revealed yet, probably due to the uncertainty of the loop. Nevertheless, the reaction mechanism is a viable alternative to the previously proposed one<sup>2</sup>, because it is analogous to the reaction mechanism found for another kinase<sup>19</sup>. Comparison between theoretical and experimental observations can contribute to the understanding of the orientation and interaction pattern of the unresolved but crucial loop residues in the X-ray structure of PLK in complex with ATP and PL. The correct loop conformation will probably further lower the final barrier and allow the full elucidation of the phosphorylation mechanism of PL by PLK.

## References

- 1 Nodwell, M. B., Menz, H., Kirsch, S. F., Sieber, S. A. Rugulactone and its Analogues Exert Antibacterial Effects through Multiple Mechanisms Including Inhibition of Thiamine Biosynthesis. *ChemBioChem* **13**, 1439-1446 (2012).
- 2 Nodwell, M. B., Koch, M. F., Alte, F., Schneider, S., Sieber, S. A. A Subfamily of Bacterial Ribokinases Utilizes a Hemithioacetal for Pyridoxal Phosphate Salvage. *Journal of the American Chemical Society* **136**, 4992-4999 (2014).
- 3 Bas, D. C., Rogers, D. M., Jensen, J. H. Very fast prediction and rationalization of pKa values for protein–ligand complexes. *Proteins: Structure, Function, and Bioinformatics* **73**, 765-783 (2008).
- 4 Li, H., Robertson, A. D., Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics* **61**, 704-721 (2005).
- 5 Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M., Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of Chemical Theory and Computation* **7**, 2284-2295 (2011).
- 6 Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation* **7**, 525-537 (2011).
- 7 Case D. A., D. T. A., Cheatham III T. E., Simmerling C. L., and Wang J. et al. AMBER 11. *University of California, San Francisco* (2010).
- 8 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983).
- 9 Meagher, K. L., Redman, L. T., Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *Journal of computational chemistry* **24**, 1016-1025 (2003).
- 10 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **25**, 1157-1174 (2004).
- 11 Wang, J., Wang, W., Kollman, P. A., Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics & modelling* **25**, 247-260 (2006).
- 12 Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., Schulten, K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781-1802 (2005).
- 13 Sherwood, P., de Vries, A. H., Guest, M. F., Schreckenbach, G., Catlow, C. R. A.,

- French, S. A., Sokol, A. A., Bromley, S. T., Thiel, W., Turner, A. J., Billeter, S., Terstegen, F., Thiel, S., Kendrick, J., Rogers, S. C., Casci, J., Watson, M., King, F., Karlsen, E., Sjøvoll, M., Fahmi, A., Schäfer, A., Lennartz, C. QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *Journal of Molecular Structure: THEOCHEM* **632**, 1-28 (2003).
- 14 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* **132**, 154104 (2010).
  - 15 Shao, Y., Molnar, L. F., Jung, Y., Kussmann, J., Ochsenfeld, C., Brown, S. T., Gilbert, A. T. B., Slipchenko, L. V., Levchenko, S. V., O'Neill, D. P., DiStasio Jr, R. A., Lochan, R. C., Wang, T., Beran, G. J. O., Besley, N. A., Herbert, J. M., Yeh Lin, C., Van Voorhis, T., Hung Chien, S., Sodt, A., Steele, R. P., Rassolov, V. A., Maslen, P. E., Korambath, P. P., Adamson, R. D., Austin, B., Baker, J., Byrd, E. F. C., Dachsel, H., Doerksen, R. J., Dreuw, A., Dunietz, B. D., Dutoi, A. D., Furlani, T. R., Gwaltney, S. R., Heyden, A., Hirata, S., Hsu, C.-P., Kedziora, G., Khalliulin, R. Z., Klunzinger, P., Lee, A. M., Lee, M. S., Liang, W., Lotan, I., Nair, N., Peters, B., Proynov, E. I., Pieniazek, P. A., Min Rhee, Y., Ritchie, J., Rosta, E., David Sherrill, C., Simmonett, A. C., Subotnik, J. E., Lee Woodcock Iii, H., Zhang, W., Bell, A. T., Chakraborty, A. K., Chipman, D. M., Keil, F. J., Warshel, A., Hehre, W. J., Schaefer III, H. F., Kong, J., Krylov, A. I., Gill, P. M. W., Head-Gordon, M. Advances in methods and algorithms in a modern quantum chemistry program package. *Physical Chemistry Chemical Physics* **8**, 3172-3191 (2006).
  - 16 Goerigk, L., Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Physical Chemistry Chemical Physics* **13**, 6670-6688 (2011).
  - 17 Kästner, J., Carr, J. M., Keal, T. W., Thiel, W., Wander, A., Sherwood, P. DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations†. *The Journal of Physical Chemistry A* **113**, 11856-11865 (2009).
  - 18 Picard, M., Toyoshima, C., Champeil, P. The Average Conformation at Micromolar [Ca<sup>2+</sup>] of Ca<sup>2+</sup>-ATPase with Bound Nucleotide Differs from That Adopted with the Transition State Analog ADP·AlFx or with AMPPCP under Crystallization Conditions at Millimolar [Ca<sup>2+</sup>]. *Journal of Biological Chemistry* **280**, 18745-18754 (2005).
  - 19 Shi, T., Lu, Y., Liu, X., Chen, Y., Jiang, H., Zhang, J. Mechanism for the Autophosphorylation of CheA Histidine Kinase: QM/MM Calculations. *The Journal of Physical Chemistry B* **115**, 11895-11901 (2011).





## Supporting Information

### **Theoretical studies on *Staphylococcus aureus* pyridoxal kinase reveal a facile interaction network crucial for phosphorylation**

I. D. Blank<sup>a</sup>, V. Kirsch<sup>b</sup>, S. A. Sieber<sup>b</sup>, C. Ochsenfeld<sup>a</sup>

- [a] I. D. Blank, Prof. Dr. C. Ochsenfeld  
Center for Integrated Protein Science Munich (CIPSM) at the Department of  
Chemistry, University of Munich (LMU Munich)  
Chair of Theoretical Chemistry, Department of Chemistry, University of Munich  
(LMU Munich)  
Butenandtstr. 7, D-81377 Munich, Germany
- [b] V. Kirsch, Prof. Dr. S. A. Sieber  
Center for Integrated Protein Science Munich (CIPSM)  
Institute of Advanced Studies IAS, Department of Chemistry  
Chair of Organic Chemistry II, Technische Universität München (TU Munich)  
Lichtenbergstrasse 4, D-85747 Garching, Germany

## **1. MD preparation**

The system was energy minimized (NVT ensemble) using the conjugate gradient algorithm (1 fs timestep) within 10 fs with non-water atoms kept fixed. Minimization was continued for 20 ps, where a positional constraint of 1 kcal/mol/Å<sup>2</sup> on non-water atoms was applied.

The system was heated up to 300 K within 30 ps. In the following step the system was equilibrated for 100 ps. In the second equilibration step we switched to the NPT ensemble employing the Nosé-Hoover Langevin piston method<sup>1,2</sup> and simulated for 50 fs until we decreased the constraints on non-water atoms step by step to zero (see Table 1). For proper statistics and to obtain a statistically significant analysis, production runs were performed for at least 150 ns in total (min. 5x30 ns) using the SHAKE<sup>3</sup> algorithm.

SI-Table 1: Reduction of constraints during equilibration step 2.

simulated time [ps]	constraint [kcal/mol/Å <sup>2</sup> ]
15	0.8
15	0.6
15	0.5
15	0.4
10	0.3
10	0.2
10	0.1
60	0.0

## **2. Analysis of the interaction pattern of crucial amino acids with their environment**

In this analysis, only direct interaction was taken into account, which means, that interactions mediated by water molecules are not considered. For the interaction partners see SI-Table 2.

SI-Table 2: Interaction partners (II) and their average number of atoms within 3 Å of the investigated residue (I).

I residue	II (interaction to)	average number of atoms of II within 3 Å of I
Lys111	His51	3
	CPL110	8
Lys176	Leu140	2
	Asn139	3
	ATP	6
	Thr137	2
Lys179	ATP	6
	Gln206	2
Lys77	Asp23	3
	Ala217	3
	Ala218	4
His210	Gly256	2
	Pro257	2
	Val258	3
	Asp13	2
His51	Lys111	2
	CPL110	2
	Met44	3
	Val42	2
Asp105	Val107	2
	Ala217	2
	Thr78	2
	Thr137	2

I residue	II (interaction to)	average number of atoms of II within 3 Å of I
Cys214	Ala22	2
	Ala218	3
Met80	CPL110	2
	Val107	2
	Gly11	2
	Thr39	2
	Val53	2
Asp23	Thr8	4
	Lys77	4
	Phe27	3
	Ala218	3

### **3. The one $\text{Mg}^{2+}$ system**

#### **Phosphorylation as initial reaction step**

Considering the phosphorylation proposed by Nodwell et al.<sup>4</sup> as first reaction step leads to unfavorable high barriers. In case that Cys214 is protonated (Cys-SH), the resulting barrier is 63 kcal/mol and the product is 55 kcal/mol more unstable than the educt. Cys214 is not involved within this phosphorylation reaction and the  $\gamma$ -phosphate of ATP is not protonated. At the end of the reaction, CPL still holds the proton of the now phosphorylated 5'-alcohol group. When Cys214 is negatively charged (Cys-S<sup>-</sup>), the resulting barrier is 52 kcal/mol and the formed product is 34 kcal/mol above the educt.

#### **Decreasing the phosphorylation barrier by protonation of the $\gamma$ -phosphate of ATP**

After the first reaction step, the proton transfer of Cys214 to the  $\gamma$ -phosphate of ATP, the barrier for a phosphorylation is 46 kcal/mol and the product is 21 kcal/mol above the educt. During the phosphate transfer, the proton of the 5'-alcohol group is abstracted by Cys214 to recover its thiol form. The phosphate, which was transferred to CPL is still in a protonated form.

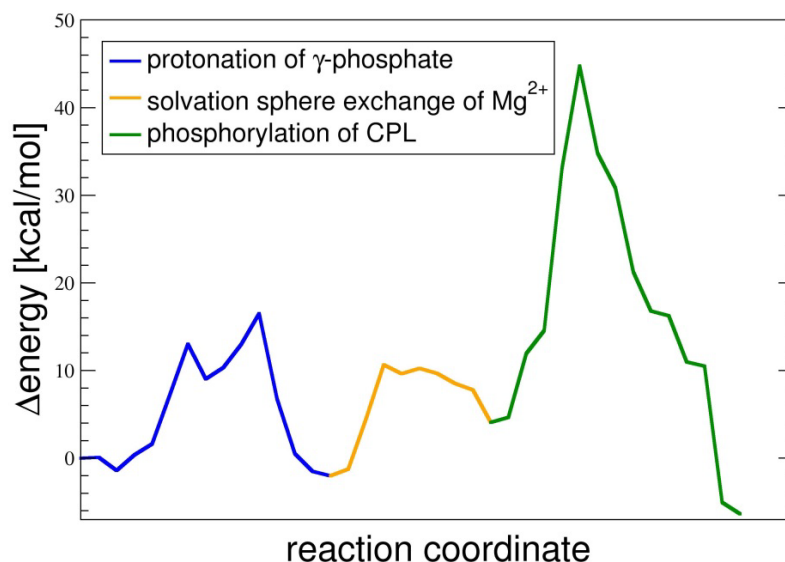
#### **Discussion about the interchangeability of step 1 ( $\gamma$ -phosphate protonation) and step 2 (solvation exchange of $\text{Mg}^{2+}$ )**

We see that first the  $\gamma$ -phosphate of ATP has to be protonated to minimize the electrostatic interaction between  $\text{Mg}^{2+}$  and the  $\gamma$ -phosphate before exchanging the coordination sphere of  $\text{Mg}^{2+}$  because a less negatively charged  $\gamma$ -phosphate can be removed in an easier manner.

In case, that we first exchange the coordination sphere and subsequently protonate the  $\gamma$ -phosphate, no further stable structure has been found so far.

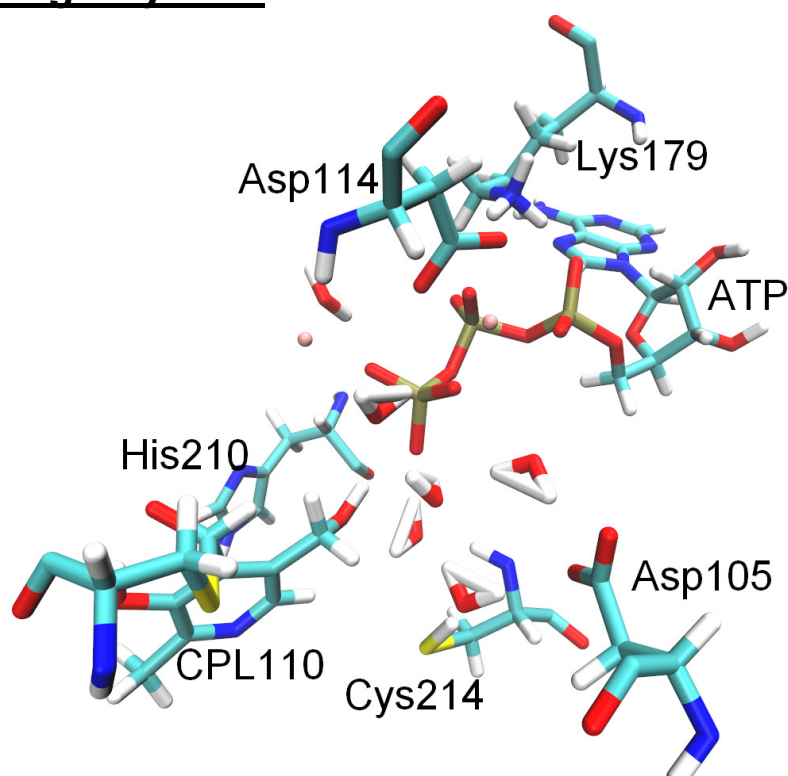
#### **Phosphorylation after $\gamma$ -phosphate protonation and solvation exchange at $\text{Mg}^{2+}$**

Phosphorylation after  $\gamma$ -phosphate protonation and solvation exchange at  $\text{Mg}^{2+}$ , still without the stabilization of Lys179, leads to a decrease of the product energy, see SI-Figure 1.

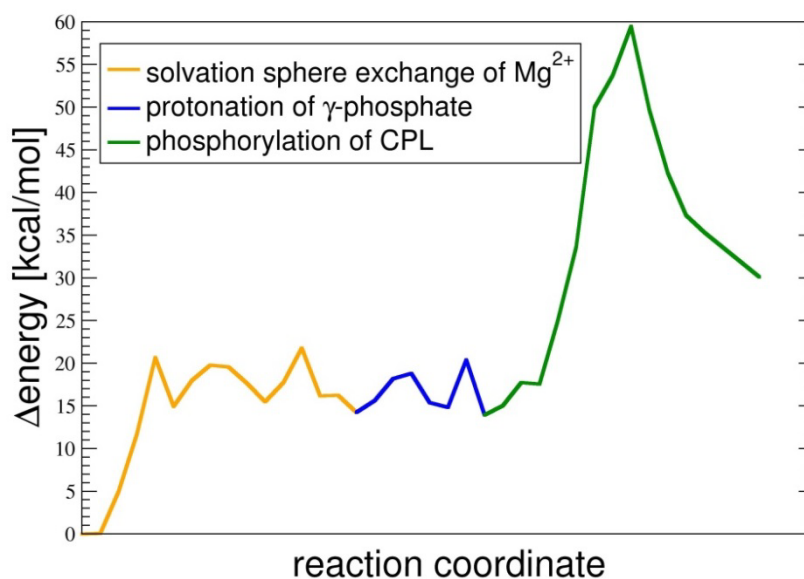


SI-Figure 1:  $\gamma$ -phosphate protonation, solvation exchange at  $\text{Mg}^{2+}$  and phosphorylation without the stabilization of Lys179. The previous solvation exchange leads to a decrease of the product energy compared to phosphorylation directly after protonation of  $\gamma$ -phosphate.

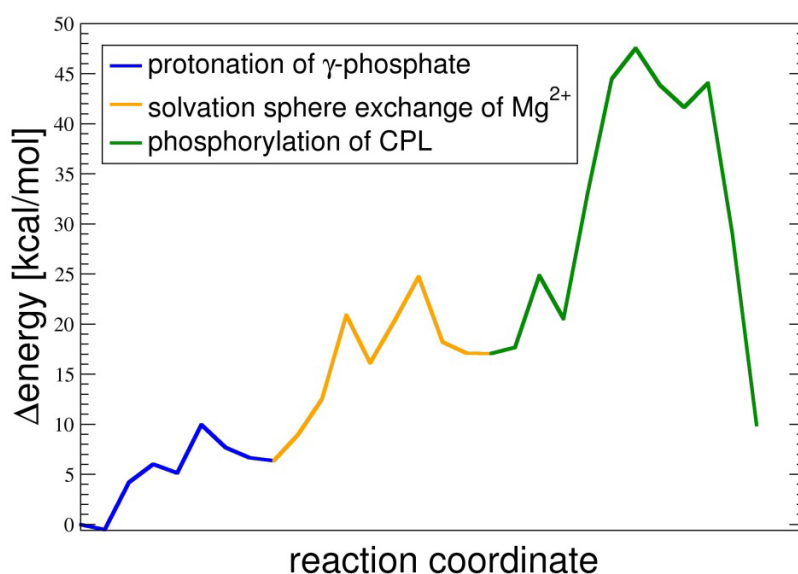
#### 4. The two $\text{Mg}^{2+}$ system



SI-Figure 2: The active site residues of PLK and the substrates for phosphorylation are depicted for the two  $\text{Mg}^{2+}$  system. The two  $\text{Mg}^{2+}$  ions are colored in pink.



SI-Figure 3: reaction profile for the two  $Mg^{2+}$  system, where the first part of the reaction is the exchange of the solvation sphere of  $Mg^{2+}$ , followed by protonation of the  $\gamma$ -phosphate of ATP and the phosphorylation step. Within the phosphorylation step, CPL is deprotonated by Asp105 via a water molecule. Lys179 does not stabilize the phosphorylation procedure.



SI-Figure 4: reaction profile for the two  $Mg^{2+}$  system, where the first part of the reaction is the protonation of the  $\gamma$ -phosphate of ATP followed by the exchange of the solvation sphere of  $Mg^{2+}$  and the phosphorylation step.

## **5. Conclusion to the amount of $\text{Mg}^{2+}$ ions in the active site:**

We investigated the influence of the amount of  $\text{Mg}^{2+}$  ions in the active site of PLK and conclude, that the third reaction step should be independent of the presence of a second  $\text{Mg}^{2+}$  ion, because the third barrier (phosphorylation), which is the highest barrier in the reaction, has the same height for the one  $\text{Mg}^{2+}$  and the two  $\text{Mg}^{2+}$  system. The reaction profile of the one  $\text{Mg}^{2+}$  system is energetically more favorable than that of the two  $\text{Mg}^{2+}$  system because of the relative energy of the intermediate states and the product. This indicates that the one  $\text{Mg}^{2+}$  ion mechanism is the favored one.

## **6. References**

- 1      Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **81**, 511–519 (1984).
- 2      Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
- 3      Ryckaert, J.-P., Ciccotti, G., Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).
- 4      Nodwell, M. B., Koch, M. F., Alte, F., Schneider, S., Sieber, S. A. A Subfamily of Bacterial Ribokinases Utilizes a Hemithioacetal for Pyridoxal Phosphate Salvage. *Journal of the American Chemical Society* **136**, 4992-4999 (2014).