Computational Investigations of the Reaction Mechanisms of Monofunctional DNA Glycosylases and the Relative Binding Affinities of Sirtuin 5 Inhibitors

Eli Naydenova

München 2021

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

Computational Investigations of the Reaction Mechanisms of Monofunctional DNA Glycosylases and the Relative Binding Affinities of Sirtuin 5 Inhibitors

Eli Naydenova

aus Sofia, Bulgarien

2021

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, __26.08.2021__ _____
(Eli Naydenova)

Dissertation eingereicht am: 26. August 2021

1. Gutachter: Prof. Dr. Christian Ochsenfeld

2. Gutachter: Prof. Dr. Regina de Vivie-Riedle

Mündliche Prüfung am: 29. Oktober 2021

Contents

Acknowledgements

Here, I would like acknowledge and thank those who supported and encouraged me during this interesting time.

I would like to thank **Prof. Dr. Christian Ochsenfeld** for giving me the opportunity to do my PhD thesis in his group and for the continuous support.

Further thanks go to **Prof. Dr. Regina de Vivie-Riedle** for being my second examiner.

I would like to thank **Dr. Johannes Dietschreit** for the helpful insights and for providing a different perspective to my projects. I thank **Beatriz von der Esch** for the interesting discussions and **Gökçen Savaşçi** for the IT introduction and support.

I also want to acknowledge **Dr. Sven Roßbach** and **Dr. Carina Glas** for our collaborations.

I thank all members of the Ochsenfeld group for the nice atmosphere and for all paper chats and Zoom coffee breaks over the years.

Finally, I want to thank my family for the understanding and support.

Abstract

Computational modeling offers detailed insight into the reactions and dynamics of biomolecules at an atomistic level. *In silico* methods complement experimental measurements and can contribute to protein engineering, drug discovery, and lead optimization. In this thesis, QM/MM calculations and MD simulations are employed for the exploration of the reaction mechanisms and the active site reorganization of the enzymes Uracil DNA glycosylase (UDG) and Thymine DNA glycosylase (TDG). UDG excises the RNA base uracil from single and double stranded DNA by hydrolysing the glycosidic bond between the sugar and the base. This reaction is shown in this work to proceed by the electrophilic migration of the sugar moiety in the direction of the water nucleophile. The addition of the water molecule and the simultaneous proton transfer to the side chain of either His148 or Asp145 occur without a barrier. Similarly to the excision of uracil, the N-glycosidic bond cleavage of the modified base 5-formylcytosine, catalyzed by TDG, was investigated in detail. The reaction mechanism features a re-organization of the substrate by a rotation of the phosphate group around the C4'-C5' bond, followed by the cleavage of the glycosidic bond. The latter features a dissociative oxocarbenium-like transition state and proceeds without a stable intermediate. The addition of the water nucleophile results in a delocalized positive charge, stabilized by an extensive H-bonding network, consisting of the abasic site and the residues Asn140 and Thr197.

A complete understanding of enzymatic activity often requires more than insight into the underlying reaction mechanism, as substrate specificity may depend on the binding affinity. The calculation of relative binding free energies allows for an *in silico* evaluation of the binding affinity of potential drug candidates. In this thesis, the alchemical non-equilibrium approach was employed for the calculation of the relative binding free energies of inhibitors of the NAD^+-dependent deacylase Sirtuin 5. The resulting values were compared with experimental measurements of the inhibitory activity for newly synthesized compounds.

Publications

This is a cumulative dissertation, comprising two articles in peer-reviewed journals (**I** and **II**) and one manuscript in preparation (**III**). In the following, all articles are listed together with the author's contribution to each of them.

I. **E. Naydenova**, S. Rossbach, and C. Ochsenfeld
"QM/MM Study of the Uracil DNA Glycosylase Reaction Mechanism: A Competition between Asp145 and His148"
J. Chem. Theory Comput., **15**, 8, 4344 (2019)
Contribution by the Author: *Most of the concept, all calculations and writing the manuscript*

II. **E. Naydenova**, J. C. B. Dietschreit, and C. Ochsenfeld
"Reaction Mechanism for the N-Glycosidic Bond Cleavage of 5-Formylcytosine by Thymine DNA Glycosylase"
J. Phys. Chem. B, **15**, 123, 4173 (2019)
Contribution by the Author: *Most of the concept, all calculations and writing the manuscript*

III. C. Glas, S. Lechner, **E. Naydenova**, N. Wössner, L. Yang, J. Dietschreit, H. Sun, M. Jung, C. Ochsenfeld, B. Küster, F. Bracher
"Development of hetero-triaryls as a new chemotype for subtype-selective and potent SIRT5 inhibition"
Contribution by the Author: *All MD simulations and calculations of the relative free energy differences*

1 Introduction

Enzymes are efficient biocatalysts, that facilitate complex chemical reactions under mild physiological conditions. Due to their high enantio-, chemo-, and stereoselectivity, catalytic efficiency and environmental sustainability, these biomolecules have found wide application in industrial synthesis [1, 2]. Understanding how enzymes achieve their high catalytic rates can provide key information for protein engineering and *de novo* enzyme design, which can in turn increase the thermodynamic stability or the substrate range of an enzyme. Knowledge of the underlying reaction mechanism can also be beneficial in rational drug design, as enzymes are often targets for potential drug candidates [3].

A wide range of mutagenesis, X-ray crystallography, spectroscopy, and kinetics experiments have been applied in the exploration of the structure and catalytic mechanisms of enzymes. Computational modeling complements these experimental procedures by providing atomistic insight into the reactions and dynamics of the studied biological systems. Quantum mechanics / molecular mechanics (QM/MM) energy and structure calculations as well as QM/MM molecular dynamics simulations are employed for the identification of transition states and reaction intermediates and can supply detailed information regarding the active site reorganization or the role of catalytic residues. Molecular dynamics and Monte Carlo methods allow for the exploration of the conformational landscape of a biomolecule and can be used for the study of enzyme-ligand interactions and the calculation of ligand binding affinities.

The goal of this thesis is to provide insights into the reactivity of several biological systems of interest, using QM/MM calculations and molecular dynamics simulations. This chapter will introduce the enzymes studied in this thesis - Uracil DNA glycosylase (UDG), Thymine DNA glycosylase (TDG), and the NAD^+-dependent deacylase Sirtuin 5, while the second chapter will provide the theoretical background for the methods employed in the investigation of the UDG and TDG reaction mechanisms and in the calculation of the relative binding affinities of the Sirtuin 5 inhibitors.

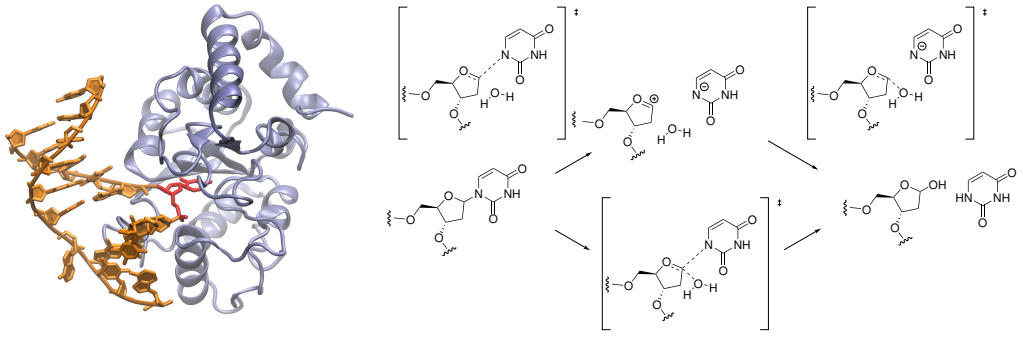
1.1 Uracil DNA Glycosylase (UDG) and Thymine DNA Glycosylase (TDG)

Base excision repair (BER) is one of the central DNA repair pathways that counter the cytotoxic and mutagenic effects of DNA damage [4]. BER corrects non-bulky DNA lesions, that arise due to the oxidation, deamination, and methylation of the nucleobases. If not repaired, these lesions can lead to point mutations and compromise the fidelity of DNA transcription and replication [4–6]. The first step of the BER pathway is the recognition and the subsequent excision of the modified

or mispaired base by a DNA glycosylase, followed by the coordinated action of at least three additional enzymes [7]. These enzymes facilitate the strand incision, end processing, gap filling, and DNA ligation to replace the excised base and restore the original DNA sequence.

DNA glycosylase enzymes can be broadly divided into two types. Monofunctional glycosylases catalyze the hydrolysis of the N-glycosidic bond, resulting in a free base and an abasic site, while bifunctional glycosylases use an amine nucleophile (usually a lysine side chain) to form a covalent Schiff base intermediate [6, 8]. In addition to excision of the damaged base, bifunctional enzymes also cleave the phosphordiester bond at the 3'-end of the lesion.

Uracil DNA glycosylase (UDG) is a monofunctional DNA glycosylase that removes the RNA base uracil, which occurs in single-stranded and double-stranded DNA due to the deamination of cytosine or the misincorporation of 2-deoxyuridine monophosphate [9, 10]. For this purpose, both the base and the ribose ring are extruded from the DNA helix and inserted into the UDG catalytic pocket (see Fig. 1a). Here, the base is stabilized by π-stacking and hydrogen bond interactions with the active site residues [11, 12]. The subsequent N-glycosidic bond cleavage proceeds with a water molecule acting as the nucleophile. There are two general mechanisms that have been proposed for this reaction (see Fig. 1b): the first is a stepwise mechanism with stable intermediate, comprised of an oxocarbenium-ion and an uracil anion, while the second mechanism is a concerted mechanism with a single dissociative oxocarbenium-like transition state [13, 14]. Publication **I** provides both a detailed investigation of the UDG reaction mechanism and insights into the role of the active site residues that coordinate the water nucleophile.



(a) **(b)**

Figure 1: (a) X-ray structure (PDB code: 1EMH) of double-stranded DNA (orange), containing the non-cleavable substrate analog presudouridine (red), in complex with UDG (blue). **(b)** Mechanisms, suggested for the N-glycosidic bond cleavage of uracil - hydrolysis with a stable oxocarbenium intermediate (top) or with a dissociative oxicarbenium-like transition state (bottom).

Thymine DNA glycosylase (TDG) belongs to the same superfamily as UDG. The enzyme was first discovered to excise the base thymine from T:G mismatches [15]. Unlike other DNA glycosylases, however, TDG is essential for embryonic development in mice [16]. Deleting the TDG gene or the catalytic inactivation of this enzyme results in lethality during early stage development, thereby implying a prominent role

for this enzyme in epigenetic regulation [17]. The presence of aberrant methylation patterns in TDG deficient cells indicates that this enzyme is involved in the process of active DNA demethylation [16, 17].

The conversion of cytosine (C) to 5-methylcytosine (mC) is an important epigenetic mechanism that is involved in transcription regulation, gene silencing, and X-chromosome inactivation [18]. Following the discovery of 5-methylcytosine in DNA in 1948, [19] the process of DNA methylation and the DNA methyltransferases, which catalyze this reaction, have been extensively studied. The mechanism of active demethlyation for mammals, however, remained elusive for a long time. The only biologically and biochemically verified pathway was established in 2011 and involves the iterative oxidation of mC to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxycytosine (caC), followed by the subsequent excision of the latter two bases by TDG (see Fig 2) [16, 17, 20, 21].

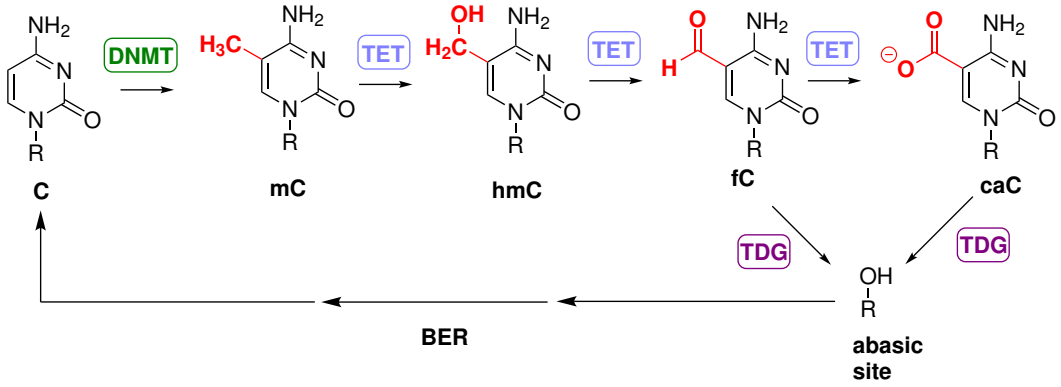


Figure 2: Active DNA demethylation pathway: mC is oxidized to hmC, fC, and caC by the TET enzymes and the latter two oxidation products are excised by TDG. The resulting abasic site is then converted back to cytosine by BER. Upon a methylation of C by a DNA methyltransferase (DNMT), the methylation site can be restored. This figure is adapted from Ref. [22]

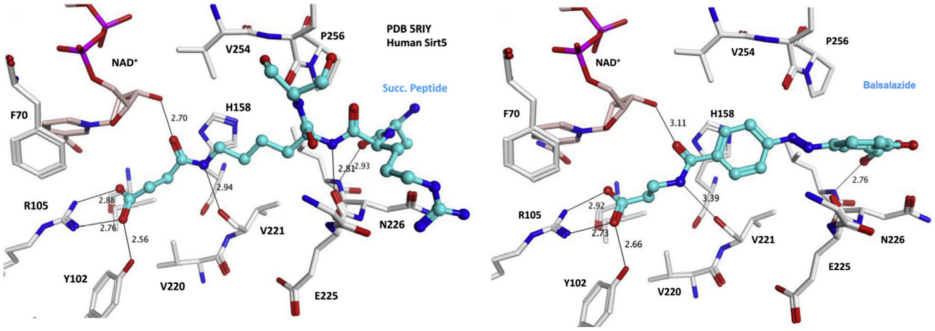
TDG employs different mechanisms for the N-glyocosidic bond cleavage of fC and caC. While the caC excision is acid-catalyzed, the TDG reaction rate for fC shows no pH dependence and there is no evidence of a protonation of this base prior to its excision [20, 23]. Similarly to UDG, a stepwise reaction mechanism with a stable oxocarbenium intermediate has been proposed for this reaction [23]. Within publication **II** the exact mechanism of the fC glycosidic bond hydrolysis is explored and the role of several active site residues is clarified.

1.2 Sirtuin 5 (SIRT5)

Post-transcriptional modifications (PTMs) expand the functional diversity of proteins [24]. They are involved in various biological processes, including gene expression, signal transduction, and protein-protein interactions [24–26]. PTMs refer to the reversible and irreversible covalent modifications of amino acids, that occur following the translation of RNA into a protein [26]. Such modifications include the deacetylation of lysine side chains, catalyzed by a total of four classes of lysine deacylases

(KDAC). While class I, II, and IV KDAC enzymes are Zn^{2+}-dependent, class III, called sirtuins, use NAD^+ as a cofactor [27]. In mammals, seven subtypes of sirtuins have been identified, three of which are located in the mitochondria including Sirtuin 5 (SIRT5) [28, 29].

Though SIRT5 exhibits only weak deacetylase activity, it has a high affinity for succinylated, malonylated, and glutarylated proteins and can effectively catalyze the removal of those groups [29–31]. The preference of SIRT5 for negatively charged substrates is presumed to stem from interactions between the carboxy-group of the substrate and the side chains of Arg105 and Tyr102 on account of a crystal structure, featuring a succinyllysine-containing peptide in the binding site (see Fig. 3a) [30].



In 2016, a high-throughput study identified the drug balsalazide as a potent SIRT5 inhibitor [32]. Docking studies indicate that, similarly to the succinylated substrate, the carboxy-group of the β-alanine side chain of the drug molecule has electrostatic and hydrogen bond interactions with Arg105 and Tyr102 (see Fig. 3b) [33]. Balsalazide has subtype selectivity for SIRT5, as it exhibits limited or no binding affinity for Sirtuin 1, 2, and 3 [33]. However, balsalazide's low solubility in water, minimal gut absorption, and the cleavage of the azo-bond by colonic bacterial azo reductases [34] rule out an application of this molecule as drug targeting SIRT5 [33]. Hence, the lead structure has to be optimized to develop promising drug candidates. Manuscript **III** presents the synthesis and characterization of Sirtuin 5 inhibitors with a structural similarity to balsalazide. An *in silico* approach is used to predict the relative binding affinities of selected compounds and the results are compared to the experiments.

2 Methods

This chapter outlines the methods used in the investigation of the reaction mechanisms of UDG and TDG and in the prediction of the binding affinities of the SIRT5 inhibitors. The first section is focused on the computational exploration of reaction mechanisms using QM/MM calculations, coordinate driving, and nudged elastic band methods, while the second section introduces relative binding free energy calculations, Jarzynski's identity [35, 36], and Bennett's acceptance ratio (BAR) [37].

2.1 Computational Investigation of Enzymatic Reaction Mechanisms

2.1.1 Molecular Mechanics (MM)

Molecular mechanics describes the interactions between atoms with the use of a simple potential energy function called force field [38–40]. In most force fields the potential $U(\mathbf{R})$ of N interacting particles is represented as the sum of bonded interactions (bond stretching, angle bending, torsion) and non-bonded interactions (van-der-Waals and electrostatics) and has the following general form:

$$U(\mathbf{R}) = \sum_{bonds} \frac{b_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{a_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{c_i}{2}(1 + cos(nw_i + \delta)) + \\ + \sum_{atom\ pairs} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \right) \tag{2.1}$$

Here $\mathbf{R} = (\mathbf{R}_1 ... \mathbf{R}_N)$ denotes the coordinates of the particles. The first two terms in eq. 2.1 describe the energy penalty associated with a deviation from an equilibrium bond length $l_{i,0}$ or angle $\theta_{i,0}$ with the use of simple harmonic functions and a respective force constant (a_i or b_i). The third term represents the change in the energy of a molecule upon a rotation around a chemical bond as a periodic function with rotational barrier heights c_i, periodicity n, and phase shift δ.

The non-bonded interactions described in the last term, are calculated for atoms that are either in different molecules or separated by at least three bonds [39]. The van-der-Waals repulsive and attractive interactions are modeled using a Lennard-Jones potential with σ_{ij} as the collision diameter, ε_{ij} as the well depth and r_{ij} as the distance between the two atoms. Calculations of the electrostatic interactions employ the Coulomb potential with q_i and q_j as the partial charges on atoms i and j, ε_0 as the dielectric constant in vacuum and ε_r as the relative dielectric constant.

Force fields contain a large set of parameters, that are obtained from extensive fitting procedures to both experimental data and quantum mechanical (QM) calculations. The assigned parameters may vary between different force fields, based on the parameterization protocol, the level of the QM calculations, or the referenced experimental data. Different force fields have been developed for proteins, nucleic acids, lipids, carbohydrates, and small molecules. In this thesis, the AMBER ff14SB force field [41] is used for the description of the proteins and AMBER ffOL15 [42] for DNA. As the parameterization of molecules not included in a force field can be time-consuming, general parameters for most organic molecules have been developed, which are compatible with the AMBER force fields [43]. These parameters were used here for the description of fC and the SIRT5 inhibitors.

2.1.2 Quantum Mechanics / Molecular Mechanics (QM/MM)

Force fields are suited for a wide range of applications, including Monte Carlo, molecular dynamics, and ligand docking simulations and the calculation of relative binding free energies. However, molecular mechanics cannot properly describe electronic reorganization due to charge transfer, electronic excitation, or chemical reactions [44]. In addition, the harmonic terms that describe bond deformations are not suitable for the modeling of bond breaking. Therefore, for the study of reaction mechanisms in biological systems often a combined QM/MM approach is employed [45]. The underlying principle of the QM/MM calculations is the partitioning of the system into an inner region (e.g., the substrate and the catalytically relevant active site residues of an enzyme), which is described using quantum mechanics and outer region (the rest of the protein environment and solvent), which is described using molecular mechanics [44–47]. The QM treatment allows modeling of the electronic rearrangements involved in a chemical reaction, while the MM treatment allows for the efficient inclusion of the wider environment and its effects on the reaction energetics.

The partitioning of the system into a QM and MM region is not a trivial problem, as the choice of a QM region can impact both the energetics and the studied properties of a system [48, 49]. In addition, if the QM/MM boundary crosses covalent bonds, separating the two regions by simply cutting through these bonds would result in unpaired electrons, which would in turn lead to unrealistic results. A straight-forward solution to this problem is to cap the unsaturated bonds with a link atom (usually a hydrogen), located along the bond vector between the QM and MM atoms [50, 51]. Introducing an extra atom would generally lead to additional degrees of freedom. However, the added constraint on the link atom coordinates in regards to the position of the QM and MM boundary atoms provides a solution for this problem. As the close proximity of the link and the MM boundary atoms (noted respectively as L and M1 in Fig. 4) could result in an overpolarization of the QM region, the M1 charge should be either removed or redistributed on the neighboring MM atoms (M2 in Fig. 4) [46]. The redistribution of the charges is referred as a charge-shift scheme [52]. Additional charges are added close to the M2 atoms to compensate for the change in the M1-M2 dipole moment.

According to the additive approach for QM/MM calculations, employed in this thesis, the QM/MM energy expression $E_{QM/MM}$ is the sum of the energy of the QM region

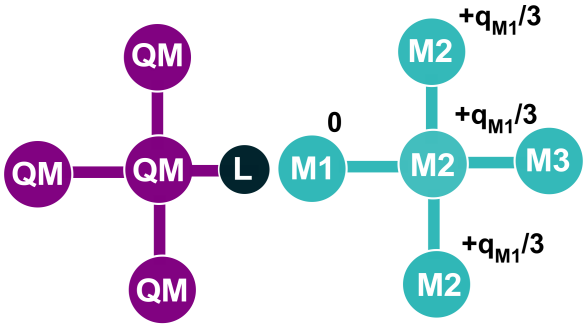


Figure 4: Representation of the atoms, close to the QM/MM boundary. The QM atoms are shown in magenta, the link atom (L) in black and the MM atoms in cyan. The charge q_{M1} of M1 is evenly redistributed on the M2 atoms, that are covalently bound to M1. This figure is adapted from Ref. [46]

(including the link atoms) E_{QM}, the energy of the MM region E_{MM} and a coupling term E_{QM-MM}, that describes the interaction between the QM and MM atoms (see eq. 2.2).

$$E_{QM/MM} = E_{QM} + E_{MM} + E_{QM-MM} \tag{2.2}$$

The coupling term E_{QM-MM} entails the bonded interactions E^{b}_{QM-MM} (bond stretching, angle bending and torsion), the van-der-Waals E^{vdW}_{QM-MM} and the electrostatic interactions E^{el}_{QM-MM} between the QM and MM atoms.

$$E_{QM-MM} = E^{b}_{QM-MM} + E^{vdW}_{QM-MM} + E^{el}_{QM-MM} \tag{2.3}$$

The bonded and the van-der-Waals interactions are calculated at the MM level. The electrostatic interactions between the QM and MM atoms can be accordingly modeled using the Coulomb potential. This approach (referred as mechanical embedding), however, requires accurate atom-centered point charges for the QM atoms, which can be problematic considering that the charge distribution of the QM region changes as the system undergoes a chemical reaction [53]. Further on, mechanical embedding does not allow for the polarization of the QM atoms by the surrounding environment [46, 47]. In order to improve upon the description of the electrostatic interaction, the MM point charges can be included in the QM Hamiltonian as one-electron operators (electrostatic embedding) [54, 55]. On account of the additional terms in the QM Hamiltonian, electrostatic embedding requires slightly more computational effort.

2.1.3 Classical Molecular Dynamics (MD) Simulations

Molecular dynamics simulations are an important tool that allows for the investigation of the time evolution of a system of interacting particles. The underlying idea is that the dynamics of the particles is governed by a general equation of motion. In classical MD simulations, this equation of motion is Newton's second law:

$$-\nabla U(\mathbf{R}) = \mathbf{M}\frac{d^2\mathbf{R}(t)}{dt^2} \tag{2.4}$$

with $\mathbf{F} = -\nabla U(\mathbf{R})$ as the force and $\mathbf{R}$ and $\mathbf{M}$ as the respective coordinates and mass. $U(\mathbf{R})$ can be assessed using the force fields, described in 2.1.1.

Multiple algorithms for the numerical integration of eq. 2.4 have been developed. In the leap-frog algorithm [56] the position $\mathbf{R}$ at the time $(t + \Delta t)$ is calculated using the position at time t and the velocity $\mathbf{V}$ at the time $t + \frac{1}{2}\Delta t$.

$$\begin{aligned} \mathbf{V}(t + \frac{1}{2}\Delta t) &= \mathbf{V}(t - \frac{1}{2}\Delta t) + \frac{\mathbf{F}(t)}{\mathbf{M}}\Delta t \\ \mathbf{R}(t + \Delta t) &= \mathbf{R}(t) + \mathbf{V}(t + \frac{1}{2}\Delta t)\Delta t \end{aligned} \tag{2.5}$$

The simulation time step Δt has to be chosen in such way as to capture the fastest motions in the system, which for biomolecules are usually the bond vibrations of the hydrogen atoms. This would restrict the time step for the simulation to less than 1 fs. To increase the time step, bonds involving hydrogens can be constrained using the SHAKE [57] or LINCS [58] algorithms, that allow for a simulation time step of 2 fs.

Periodic boundary condition

If we confine the particles of our studied system to a finite volume, those close to the boundary would differ in their behavior as compared to those located further away. To eliminate the artifacts from the system boundaries, periodic boundary conditions are introduced. For this purpose, a central box for the system is defined and this box is then multiplied in all directions. Atoms outside of the central box are images of the atoms of the central box, hence an atom can leave the box on one side only to re-enter on the opposite side. In order to avoid double counting, the minimal image convention is used. This restricts the calculation of the pairwise interaction between atoms i and j only to the one between the original atom i and atom j (original or image), that is closest to atom i [38, 39].

Cutoff and Particle Mesh Ewald

The calculation of the non-bonded term at each integration step is the computationally most expensive part, as it scales at $O(N^2)$. Considering that biomolecular systems in explicit solvent consist of tens of thousands of atoms, these calculations can become prohibitively expensive. In order to reduce the computational effort, a cutoff for the non-bonded interactions can be introduced, so that interactions between atoms beyond a certain cutoff distance are set to zero. As an abrupt truncation of the interactions at a certain distance would introduce instabilities in the MD simulation, a switching function can be applied that ensures the continuity of the potential [59].

While the use of a cutoff for the Lennard-Jones potential is not problematic due to its fast decay (r^{-6}), the truncation of the Coulomb term introduces significant artifacts in the simulations [60]. Rather then using a cutoff, the electrostatic interactions in MD simulations can be calculated using the particle mesh Ewald method (PME) [61]. PME is based on the Ewald summation. Here, every point charge is surrounded by

a Gaussian charge distribution of equal magnitude and opposite sign, that acts as an ionic atmosphere screening neighboring charges [62]. To ensure charge conservation, a charge distribution with the same shape but with an opposite sign is also added. This effectively divides the Coulomb potential into a short-range part, that is evaluated in real space and a long-range part, evaluated in Fourier space, allowing for an efficient calculation. Further improvement of the scaling behavior to $O(N \log(N))$ is achieved in PME by using the fast Fourier transform [61].

2.1.4 Calculation of Minimal Energy Paths (MEPs)

One of the important problems both in the computational study of enzymatic reaction mechanisms and in theoretical chemistry in general is the identification of the minimal energy path (MEP) for a studied chemical reaction. The MEP is the steepest descent path that connects the first order saddle point of a potential energy surface (PES) (the transition state) with the two local minima that correspond to the reactant and the product of a reaction. A straightforward approach for the exploration of reaction paths is the coordinate driving (CD) or adiabatic mapping method [63–66]. Here, we start our investigation in a local minimum of the PES and define a reaction coordinate (e.g., a distance between two atoms or a combination of distances) that is varied stepwise with the use of harmonic restraints. At each step, a minimization is carried out over the remaining degrees of freedom. While the CD method is quite intuitive and straightforward in its implementation, the resulting reaction path is largely dependent on the choice of a reaction coordinate [67]. Further on, if significant atomic rearrangements, connected to the reaction, are not explicitly included in the reaction coordinate, the CD approach could lead to discontinuities in the reaction profile [68].

An alternative to CD is the nudged elastic band (NEB) method [69–72]. NEB belongs to the class of a chain-of-states methods in which two local minima of the PES are interconnected by a chain of images (or states), that are simultaneously optimized. A spring force interaction is added between adjacent images to ensure equal spacing along the reaction path and to prevent the images from "sliding down" from high-energy regions.

The tangent of an image i ($\boldsymbol{\tau}_i$) can be calculated from the coordinates of the adjacent image with the highest energy [70]:

$$\boldsymbol{\tau}_i = \begin{cases} \boldsymbol{\tau}_i^+ & \text{if } E_{i+1} > E_i > E_{i-1}. \\ \boldsymbol{\tau}_i^- & \text{if } E_{i+1} < E_i < E_{i-1}. \end{cases} \tag{2.6}$$

with $\boldsymbol{\tau}_i^+ = |\mathbf{R}_{i+1} - \mathbf{R}_i|$ and $\boldsymbol{\tau}_i^- = |\mathbf{R}_{i-1} - \mathbf{R}_i|$

Here E_i refers to the energy of image i with coordinates $\mathbf{R}_i$. If both neighboring images are either higher in energy ($E_{i+1} > E_i < E_{i-1}$) or lower in energy ($E_{i+1} < E_i > E_{i-1}$) than image i, then the tangent is calculated as an energy weighted

average of the vectors of the adjacent images:

$$\boldsymbol{\tau}_i = \begin{cases} \boldsymbol{\tau}_i^+ \Delta E_i^{max} + \tau_i^- \Delta E_i^{min} & \text{if } E_{i+1} > E_{i-1}. \\ \boldsymbol{\tau}_i^+ \Delta E_i^{min} + \tau_i^- \Delta E_i^{max} & \text{if } E_{i+1} < E_{i-1}. \end{cases} \tag{2.7}$$

with

$$\begin{aligned} \Delta E_i^{max} &= \max(|E_{i+1} - E_i|, |E_{i-1} - E_i|) \\ \Delta E_i^{min} &= \min(|E_{i+1} - E_i|, |E_{i-1} - E_i|) \end{aligned} \tag{2.8}$$

The tangent is then normalized:

$$\hat{\boldsymbol{\tau}}_i = \frac{\boldsymbol{\tau}_i}{|\boldsymbol{\tau}_i|} \tag{2.9}$$

and both the force, acting on an image due to the energy landscape (referred as true force), and the spring force are decomposed into components parallel and perpendicular to $\hat{\boldsymbol{\tau}}_i$ (see Fig. 5). Only the parallel component of the spring force ($F_{i\parallel}^S$) and the perpendicular component of the true force ($F_{i\perp}$) are included in the expression for the total force acting on the image i (F_i^{NEB}):

$$F_i^{NEB} = F_{i\parallel}^S + F_{i\perp} \tag{2.10}$$

$F_{i\perp}$ and $F_{i\parallel}^S$ are calculated according to the following equations:

$$F_{i\perp} = -\nabla E(\mathbf{R}_i) + \nabla E(\mathbf{R}_i) \cdot \hat{\boldsymbol{\tau}}_i \hat{\boldsymbol{\tau}}_i \tag{2.11}$$

$$F_{i\parallel}^S = k[(\mathbf{R}_{i+1} - \mathbf{R}_i) - (\mathbf{R}_i - \mathbf{R}_{i-1})] \cdot \hat{\boldsymbol{\tau}}_i \tag{2.12}$$

with k as the spring constant.

This force projection is referred as "nudging" and prevents the spring force from interfering with convergence of the states to the MEP and the true force from affecting the spacing of the images. This effectively decouples the relaxation of the images from their distribution along the path.

The climbing image NEB method includes a slight modification of the original NEB implementation. Following several iterations, the image with the highest energy (j) is determined and the force acting on this image is then calculated according to the following equation [71]:

$$F_j^{CI-NEB} = -\nabla E(\mathbf{R}_j) + 2\nabla E(\mathbf{R}_j) \cdot \hat{\boldsymbol{\tau}}_j \cdot \hat{\boldsymbol{\tau}}_j \tag{2.13}$$

The image j does not feel the spring forces of the neighboring images and can therefore climb to the saddle point.

NEB has several advantages compared to the CD method. The calculation is not biased by the choice of a reaction coordinate and the simultaneous optimization of the images allows for a computationally efficient parallel computation of the images. At the same time, the NEB setup requires prior knowledge of the educt and product of the studied reaction, as an initial and final point are needed for the generation of the chain of images.

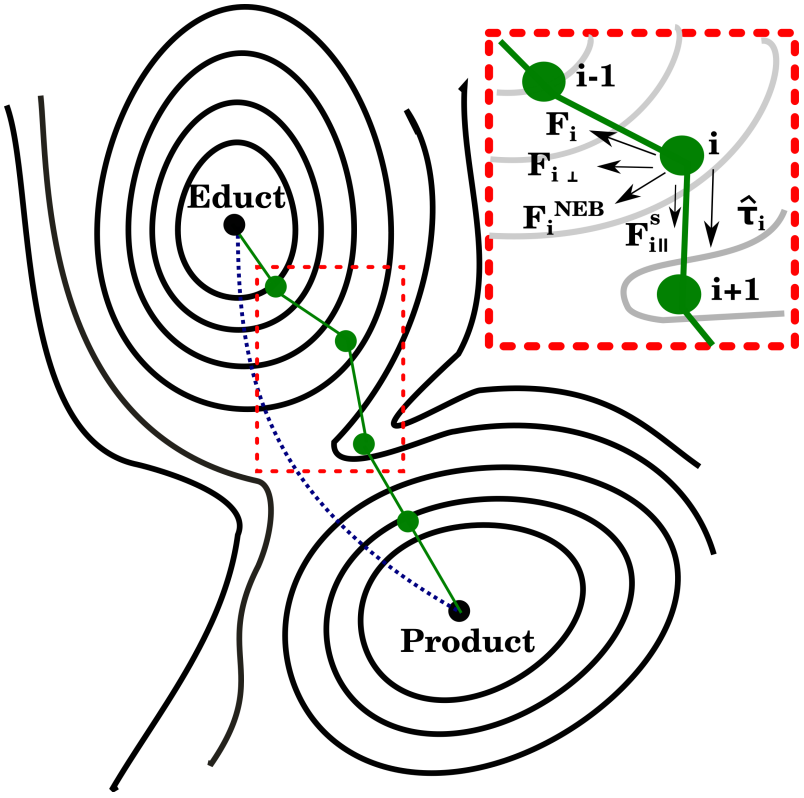


Figure 5: Representation of the NEB method. This figure shows a PES with the corresponding MEP indicated in blue. The chain of images at the start of the NEB are shown in green. The force acting on image i ($\boldsymbol{F}_i^{NEB}$) consists of the spring force component parallel to $\hat{\boldsymbol{\tau}}_i$ ($\boldsymbol{F}_{i\parallel}^{S}$) and the true force component perpendicular to $\hat{\boldsymbol{\tau}}_i$ ($\boldsymbol{F}_{i\perp}$). This figure is adapted from Ref. [72]

2.2 Relative Binding Free Energy (RBFE) Calculations

2.2.1 Free Energy Differences

According to the second law of thermodynamics, a closed system in thermodynamic equilibrium assumes a state with a minimum free energy. The difference in the free energies between state A and state B (ΔA_{AB}) determines the probability of finding a system in one state or in the other and is therefore of high interest.

The Helmholtz free energy A is related to the canonical partition function $Q(N, V, T)$, that describes a system with N particles, volume V, and temperature T:

$$A = -\frac{1}{\beta} \ln Q(N, V, T) = -\frac{1}{\beta} \ln \left(\frac{1}{h^{3N} N!} \int d\mathbf{R}_N d\mathbf{p}_N \quad e^{-\beta H(\mathbf{R}_N, \mathbf{p}_N)} \right) \tag{2.14}$$

with $\beta = 1/k_B T$ and k_B as the Boltzmann constant. Here, $H(\mathbf{R}_N, \mathbf{p}_N)$ is the Hamiltonian of the system of N particles with coordinates $\mathbf{R}_N$ and conjugate momenta $\mathbf{p}_N$ and h is Planck's constant. Assessing A from MD simulations is quite difficult, as it would require adequate sampling of the higher energy phase space regions [73],

which can be achieved only for small systems. The free energy difference ΔA_{AB}:

$$\Delta A_{AB} = A_B - A_A = -\frac{1}{\beta} \ln \frac{Q_B(N,V,T)}{Q_A(N,V,T)} \tag{2.15}$$

is more accessible, as contributions to the free energy from the high energy microstates would be discarded for both states with this error cancelation resulting in faster convergence for ΔA_{AB} [74].

Free energy differences can be calculated by equilibrium or non-equilibrium methods. The latter are based on Jarzynski's identity and Crooks Fluctuation Theorem, that are summarized in the following sections.

2.2.2 Jarzinski's Identity and Estimator

Consider a system in contact with a heat reservoir that depends on an external parameter λ and a process, which causes this parameter to evolve from an initial value ($\lambda = 0$) to a final value ($\lambda = 1$). $\lambda = 0$ corresponds to state A and $\lambda = 1$ corresponds to the state B of the system. According to the second law of thermodynamics, if we change λ infinitely slowly, then the system would remain in quasi-static equilibrium and the total work (W_∞) is then equal to the Helmholtz free energy difference ΔA_{AB} of the states A and B [35]:

$$W_\infty = \Delta A_{AB} = A_B - A_A \tag{2.16}$$

If λ is varied at a finite rate, then the work (W) would depend on the initial conditions of the system and an ensemble of measurements (where each measurement is started after allowing the system and reservoir to equilibrate at a certain temperature) would yield a distribution of W values. Due to work dissipation the average of this distribution $\langle W \rangle$ exceeds the Helmholtz free energy difference ΔA_{AB} :

$$\langle W \rangle \geq \Delta A_{AB} \tag{2.17}$$

For a trajectory that describes the time evolution of the system from state A ($\lambda = 0$) to state B ($\lambda = 1$), the work can be calculated by the following integral:

$$W = \int_{\lambda=0}^{\lambda=1} \frac{\partial H}{\partial \lambda} \, d\lambda \tag{2.18}$$

with H is the Hamiltonian of the system.

In 1997 Jarzynski derived the following identity, that relates equilibrium free-energy difference to an ensemble of non-equilibrium (finite time) work values [35, 36]:

$$e^{-\beta \Delta A_{AB}} = \langle e^{-\beta W} \rangle \tag{2.19}$$

with:

$$\langle e^{-\beta W} \rangle = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} e^{-\beta W_n} \tag{2.20}$$

N denotes the number of trajectories and W_n the work value for the nth transition. On the left side of eq. 2.19 is an exponential of the free energy difference of the

states A and B, and on the right side is an exponentially weighted average over an infinite number of non-equilibrium work trajectories. This equality is also valid for an NPT ensemble [75]

$$e^{-\beta \Delta G_{AB}} = \langle e^{-\beta W} \rangle \tag{2.21}$$

with ΔG_{AB} as the Gibbs Free Energy.

Based on eq. 2.21 the free energy difference for a finite number N transitions can be estimated, using the Jarzynski estimator:

$$\Delta \bar{G}_{AB} = -\frac{1}{\beta} \ln \left(\frac{1}{N} \sum_{n=1}^{N} e^{-\beta W_n} \right) \tag{2.22}$$

$\Delta \bar{G}_{AB}$ is the estimated value for the free energy difference. As exponential averaging depends strongly on the behavior at the tails of the distribution, which are generally not as well sampled as the rest of the distribution [76], this approach tends to behave poorly for a small sample size [77]. Therefore, the limited number of the transitions introduces a significant bias to the Jarzynski estimator [78].

2.2.3 Crooks Fluctuation Theorem (CFT) and Bennett's Acceptance Ratio (BAR)

While the Jarzynski identity employs only transitions from state A ($\lambda = 0$) to state B ($\lambda = 1$), Crooks Fluctuation Theory takes into account work values from both the forward (A→B) and the reverse (B→A) process. If all transitions are started from an equilibrium ensemble, the following relation can be derived [79–81]:

$$\frac{P_f(W)}{P_r(-W)} = e^{-\beta(W - \Delta G)} \tag{2.23}$$

$P_f(W)$ and $P_r(-W)$ are the probability distributions of the work values respectively for the forward and for the reverse paths. Starting from eq. 2.23, Shirts et. al derived a maximum likelihood estimator for the free energy difference, given a set of work values [76]:

$$\sum_{i=1}^{N_f} \frac{1}{1 + \frac{N_f}{N_r} e^{\beta(W_i - \Delta \bar{G})}} = \sum_{j=1}^{N_r} \frac{1}{1 + \frac{N_r}{N_f} e^{\beta(W_j - \Delta \bar{G})}} \tag{2.24}$$

N_f and N_r are the number of forward and reverse transitions. This estimator is known as Bennett's acceptance ratio (BAR), as it was first introduced by Bennett for the calculation of the free energy difference between two states, sampled at equilibrium [37]. ΔG can be calculated by iteratively solving equation 2.24. The BAR estimator results in the free energy difference, which maximizes the chance that the specified forward and reverse work distributions would be observed.

2.2.4 Absolute and Relative Binding Free Energy Calculations

In silico prediction of protein-ligand binding free energy is of high-interest for drug discovery, as it provides an estimate of the binding affinity of potential drug candidates [82, 83]. If such calculations are sufficiently fast and accurate, they can

reduce the time, or costs for drug lead optimization [84]. Binding affinities can be estimated from simulations using absolute binding free energy (ABFE) or relative binding free energy (RBFE) calculations [85]. The former involves sampling of the binding/unbinding of a ligand either by a physical or a non-physical pathway, which can be quite challenging especially for charged molecules [86, 87]. The removal of a ligand from the binding site constitutes as a large perturbation of the system, resulting in a smaller phase space overlap and a slower convergence for such simulations [74]. RBFE calculations, in turn, allow for the estimation of the relative binding free energy differences between two structurally similar molecules. Here, we take advantage of the fact, that the free energy of a system is a state variable and calculations of the free energy differences are therefore independent of the path taken in between [74]. This can be illustrated, using the thermodynamical cycle shown in Fig.6.

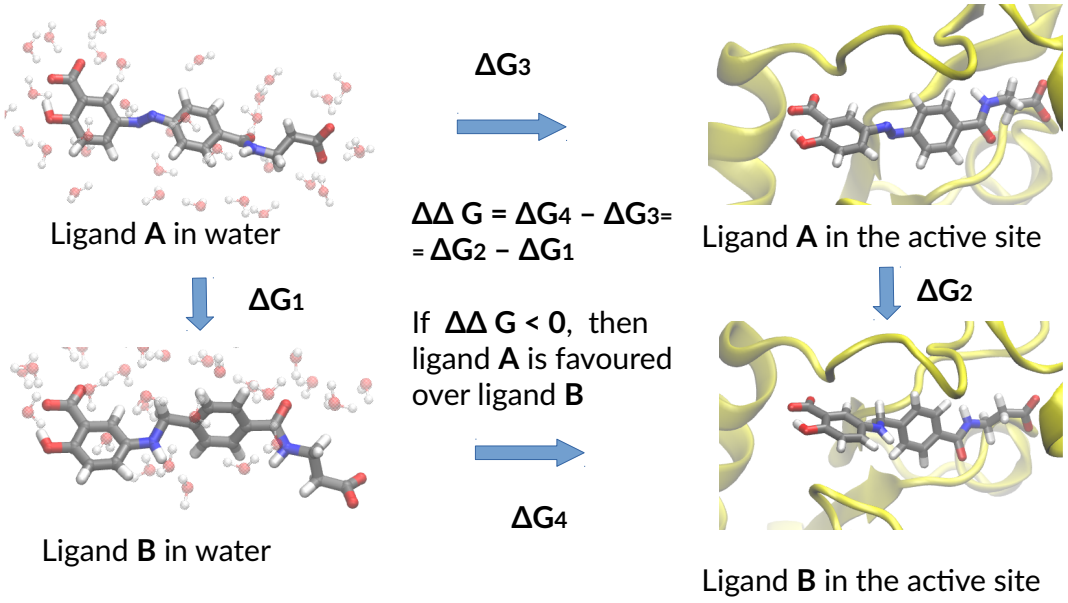


Figure 6: Schematic representation of a thermodynamic cycle. The horizontal arrows describe the binding of the ligands A an B to the protein. The vertical arrows represent the alchemical pathways for the transformation of ligand A to ligand B in solvent and in the active site. The relative binding affinity ($\Delta\Delta G$) calculated from the difference of ΔG_4 and ΔG_3 corresponds to the difference of ΔG_2 and ΔG_1.

The thermodynamic cycle here consists of the free energy difference upon the binding of a ligand A to the protein (ΔG_3) and upon the binding of a different ligand B to the same protein (ΔG_4) and the free energy difference for the alchemical (non-physical) transformation of ligand A to ligand B in solvent (ΔG_1) and in the protein active site (ΔG_2). As this is a closed cycle, the sum of all branches is zero. Therefore, the the relative binding affinity (double free energy difference $\Delta\Delta G$) of ligand B to ligand A can be calculated from the following equation:

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \tag{2.25}$$

If $\Delta\Delta G$ is negative, then ligand B binds more tightly to the protein than ligand A. As mentioned above, assessing ΔG_1 or ΔG_2 from MD simulations is computationally demanding due to the significant change in the environment. The free energy differences from the alchemical pathways are more accessible. The first calculations of relative free energy differences were performed in 1985 to compute the hydration free energies of ethane and methanol [88] and since then, different methods have emerged. In manuscript **III**, the alchemical non-equilibrium approach [74, 89, 90] was applied for the calculation of the relative binding free energies of selected SIRT5 inhibitors. The applied protocol includes equilibrium sampling of the respective ligands in water and in the active site, followed by short non-equilibrium transition simulations, that transform one ligand into the other, as represented in Fig. 7.

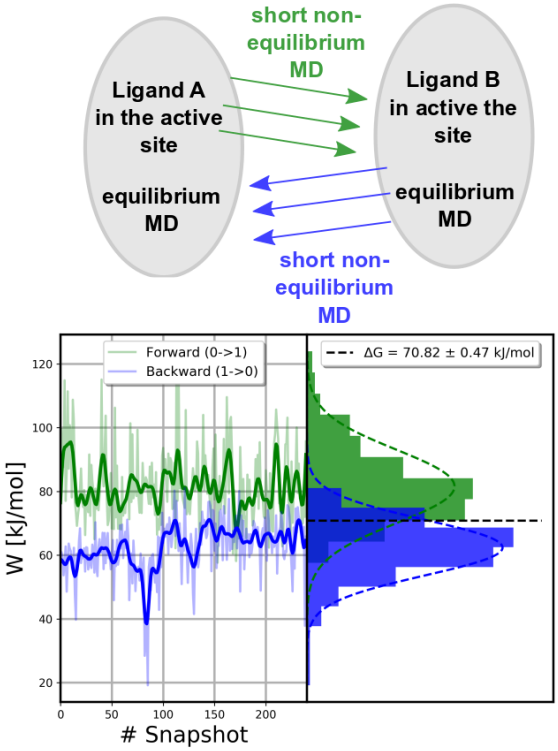


Figure 7: Schematic representation of the protocol for the relative binding free energy calculations. Using equilibrium MD simulations, equilibrium ensembles are generated for ligand A and ligand B in water and in the active site. Non-equilibrium MD simulations are started from different snapshots of the equilibrium ensembles in the forward (green) and in the reverse (blue) direction. The respective work values are represented both as a function of the starting structure (left) and as histograms (right). The free energy difference, calculated by using BAR, is shown as a black dotted line.

The transitions simulations were performed both in the forward (A→B) and in reverse (B→A) directions and the respective work values were calculated using eq. 2.18. The free energy difference is then estimated with BAR.

2.2.5 Soft-core Potentials

The alchemical transitions between the ligands A and B exploit nonphysical pathways that often include the annihilation and creation of particles. The insertion or deletion of an atom can result in divergence in $\partial H/\partial\lambda$ for λ close to zero or one, as (according to the Lennard-Jones and Coulomb potentials used in the MM force fields for the non-bonded interactions) the potential energy of two particles goes to infinity when the interatomic distance is very small. This would in turn result in instabilities during the numerical integration and in an unreliable free energy estimate [91]. To avoid singularities and numerical instabilities, the expression for the non-bonded interactions from eq. 2.1 is modified, so that the Coulomb and Lennard Jones potentials have finite values even for small interatomic distances [91]:

$$U_{nb}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r(\alpha_Q(1-\lambda)+r_{ij}^p)^{1/p}} + 4\lambda\varepsilon_{ij}\left(\frac{1}{(\alpha_{LJ}(1-\lambda)+(r_{ij}/\sigma_{ij})^s)^{12/s}} - \frac{1}{(\alpha_{LJ}(1-\lambda)+(r_{ij}/\sigma_{ij})^s)^{6/s}}\right) \tag{2.26}$$

Equation 2.26 presents the modified "soft-core" potential for the non-bonded interactions U_{nb} with α_{LJ} and α_Q as positive constants. In the GROMACS implementation of the softcore potential $s = p = 6$ [92]. By adding a λ-dependent “shifting” term to the interatomic distance of any Lennard-Jones and Coulombic interaction that is being annihilated or created in the transformation, the infinitely repulsive body or “hard core” is replaced by a potential energy barrier of finite magnitude or “soft core”, thereby preventing singularities in the potential.

3 Publications

3.1 Publication I: QM/MM Study of the Uracil DNA Glycosylase Reaction Mechanism: A Competition between Asp145 and His148

E. Naydenova, S. Rossbach, and C. Ochsenfeld

"QM/MM Study of the Uracil DNA Glycosylase Reaction Mechanism: A Competition between Asp145 and His148"

J. Chem. Theory Comput., **15**, 8, 4344 (2019)

Uracil DNA glycosylase catalyzes the N-glycosidic bond cleavage of uracil, thereby initiating the base excision repair mechanism for this DNA lesion. Here we employ hybrid quantum mechanics/molecular mechanics calculations to investigate the exact mechanism of the nucleophile attack and the role of the conserved His148 residue. Our calculations suggest that the C1'-N1 bond dissociation proceeds by a migration of the electrophilic sugar in the direction of the water nucleophile, resulting in a planar, oxocarbenium-like transition state. The subsequent nucleophile addition and proton transfer to a nearby base occur without a barrier. We assign the role of a proton acceptor to His148 and elucidate why mutations of this residue curtail the enzymatic activity but do not fully suppress it.

JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2019, 15, 4344−4350

pubs.acs.org/JCTC

QM/MM Study of the Uracil DNA Glycosylase Reaction Mechanism: A Competition between Asp145 and His148

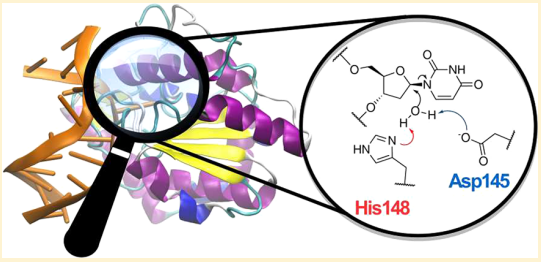
Eli Naydenova,[†,‡] Sven Roßbach,[†,‡] and Christian Ochsenfeld*[,†,‡]

[†]Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany

[‡]Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany

Supporting Information

ABSTRACT: Uracil DNA glycosylase catalyzes the *N*-glycosidic bond cleavage of uracil, thereby initiating the base excision repair mechanism for this DNA lesion. Here we employ hybrid quantum mechanics/molecular mechanics calculations to investigate the exact mechanism of the nucleophile attack and the role of the conserved His148 residue. Our calculations suggest that the C1′−N1 bond dissociation proceeds by a migration of the electrophilic sugar in the direction of the water nucleophile, resulting in a planar, oxocarbenium-like transition state. The subsequent nucleophile addition and proton transfer to a nearby base occur without a barrier. We assign the role of a proton acceptor to His148 and elucidate why mutations of this residue curtail the enzymatic activity but do not fully suppress it.



1. INTRODUCTION

DNA nucleobases undergo continuous modifications due to spontaneous decay or exposure to endogenous and exogenous genotoxic species.[1] To conserve the integrity of the genome, mutagenic alterations are counteracted by highly efficient DNA repair mechanisms such as the base excision repair (BER) pathway.[2] BER is predominantly employed for the repair of single DNA lesions and is initiated by the excision of the aberrant base. For the DNA damage uracil, an RNA nucleobase, which arises in DNA as a result of the deamination of cytosine or alternatively from the misincorporation of 2′-deoxyuridine monophosphate,[3] this excision is catalyzed by the enzyme uracil DNA glycosylase (UDG).[2] UDG employs a water nucleophile to cleave the *N*-glycosidic bond between the deoxyribose and uracil, resulting in an abasic site and an uracilate anion.[4]

The exact mechanism of the glycosidic bond cleavage has been the subject of several experimental[4−14] and computational studies.[15−17] X-ray structures of UDG bound to DNA revealed that the damaged nucleotide is rotated out of the DNA helix into the active site of the enzyme.[5,8] Bond dissociation, facilitated by the reorganization of the substrate in the UDG active site[8] and by a hydrogen bond between His268 and O2 of uracil[9] (see Figure 1), is presumed to proceed by a highly dissociative transition state or an oxocarbenium-like intermediate on the basis of kinetic isotope effect (KIE) measurements.[10]

Although the UDG-catalyzed glycosidic bond cleavage has been extensively studied, the exact mechanism of the nucleophile attack and the role of His148 in the reaction mechanism remain unknown. Initial quantum mechanics/molecular mechanics (QM/MM) investigations, performed with a QM region of 32 atoms, revealed a stepwise mechanism with a stable oxocarbenium-ion−uracil-anion intermediate.[15] Upon the addition of the nucleophile, a proton was simultaneously transferred to the negatively charged Asp145 (see Figure 1). His148, presumed to be doubly protonated and positively charged in this study, was surmised to have an anticatalytic influence on the reaction, as it was found to destabilize the transition state.[15,16] However, experimental findings contradict the latter with evidence of an increase in barrier height upon the selective mutation of this residue.[14] Further on, His148 was suggested to be neutral, with a single proton at Nε based on nuclear magnetic resonance (NMR) and binding studies.[14]

Because of the proximity of His148 to the nucleophile and its neutral protonation state, there exists a distinct possibility for this residue to either orient the water molecule for the C1′ attack or to act as the general base for its activation. A recent combined QM and semiempirical study investigated this possibility.[17] Contrary to the previous calculations, the reaction was found to proceed without a stable intermediate, and a proton transfer to His148 was preferred over a proton transfer to Asp145. However, for computational reasons, the model employed in ref 17 did not include the whole protein and heavily restrained the movements of the active-site residues as only the QM region (114 atoms, MPWB1K/6-31G*) was allowed to relax, whereas the atoms in the semiempirical region were kept frozen.[17] Therefore, a reinvestigation of the mechanism seems useful.

Received: December 28, 2018
Published: July 18, 2019

DOI: 10.1021/acs.jctc.8b01305
J. Chem. Theory Comput. 2019, 15, 4344−4350

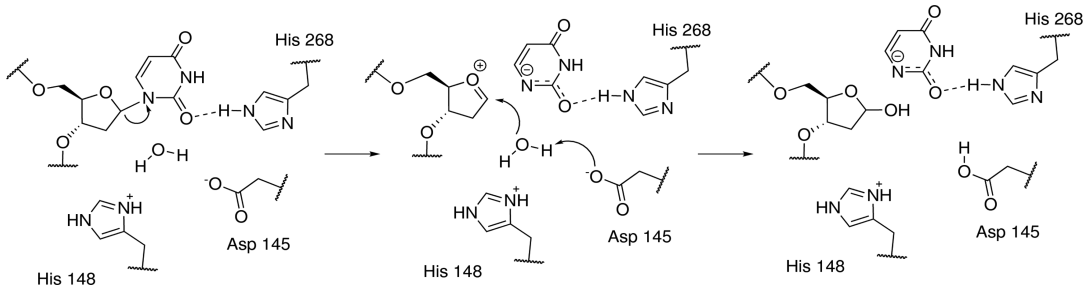


Figure 1. Reaction mechanism of the uracil *N*-glycosidic bond hydrolysis suggested in a QM/MM study by Dinner et al.[15]

In our present work, we systematically study the UDG-catalyzed glycosidic bond cleavage with a particular emphasis on the catalytic function of His148. Our results indicate a role of a general base for this residue that is consistent with experimental evidence. Employing a QM size of 696 atoms allows us to assess the stability of the products resulting from the nucleophile attack and provides a reliable description of the reaction energetics.

2. COMPUTATIONAL METHODS

2.1. System Setup. The present calculations are based on the X-ray structure of the UDG enzyme bound to DNA containing the substrate analogue 2′-deoxypseudouridine (ΨU) (PDB code 1EMH).[8] The C1 and N5 atoms of ΨU were interchanged in silico to obtain 2′-deoxyuridine (dU). The structure was protonated, neutralized with Na^+ ions, and solvated in a box of TIP3P water with a buffer of 10 Å around the solute using the LEAP module of AmberTools16.[18] Protonation states for the UDG titratable residues were assigned on the basis of PROPKA 3.0[19] pK_a calculations (see the Supporting Information (SI), Section 1), with the exception of His148, for which a neutral ionization state with a proton at Nε was assumed based on NMR and binding experiments.[14] Unrestrained force-field molecular dynamics (MD) simulations were unable to reproduce the reactive conformation of dU (see SI, Section 2); therefore, we used the crystal structure with proper solvation as a starting structure for our QM/MM simulations. To allow for the relaxation of the solvent and the ions, we performed 10 000 steps of conjugated gradient minimization with a frozen solute and an equilibration over 400 ps with both DNA and protein atoms held fixed. We employed the AMBER OL15 force field[20] for the DNA, the AMBER ff14SB force field[21] for the protein, and the NAMD 2.10 engine[22] for the minimization and equilibration.

2.2. QM/MM Calculations. The initially chosen QM region consists of the dU nucleotide, three water molecules, Wat600, Wat655, and Wat727 (numbers corresponding to the crystal structure), and the backbone and side chains of Gln144, Asp145, Pro146, Tyr147, His148, Phe158, Asn204, and His268 (170 atoms; see Figure S4). Whereas Asp145, Pro146, and His148 are situated within H-bonding distance of the nucleophile, the remainder of the QM residues are located near the nucleobase and assumed to either enhance the stability of the transition state or facilitate the departure of the leaving group.[11] To calculate the minimal energy paths (MEPs), we perfomed adiabatic mapping calculations and used the resulting educt and products as a starting point for the climbing-image nudged elastic band (NEB) calculations.[23] The structures of the climbing images were then further refined using the dimer method.[24]

All residues within 8 Å of the QM region (2096 atoms) were optimized, whereas the remaining MM residues were fixed. We employed the DL-POLY[25] implementation within ChemShell[26] for the MM component and the Q-Chem[27] program package for the QM component of the hybrid calculations. The QM region was treated at the B3LYP-D3[28,29] level of theory with the 6-31G** basis set.[30] Partial charges from the MM description were included in the QM Hamiltonian, and link atoms with a charge-shifting scheme[31] were applied at the QM/MM boundary.

2.3. QM Size Convergence. For the investigation of the QM size convergence of the activation and reaction energies, we employed the geometries of the stationary points resulting from the QM/MM reaction path calculations with a QM region of 170 atoms. We performed single-point calculations on the educt, transition states, and products. The number of QM atoms at each step of this study was gradually increased by including residues based on their distance to the initial QM region. A detailed description of the QM regions at each step of this convergence study is provided in the SI (Section 3.3). The largest QM sphere comprises all residues within 6 Å of the initial QM region (1498 atoms). A reasonable size convergence was considered to be achieved if the values for the reaction and activation energies agreed with each other within a tolerance of 2 kcal/mol.

An increase in the QM size typically has an important influence on the relative energies and system properties.[32−34] Whereas single-point calculations allow for an assessment of the convergence behavior, a reoptimization of the stationary points could still result in significant changes in the relative energies.[35,36] Therefore, we performed geometry reoptimizations of the reaction educt and products with the converged QM size of 696 atoms. In these calculations, only the residues of the initial QM region (170 atoms) were allowed to relax to prevent structural changes of the environment that are unrelated to the studied reaction (e.g., a rotation of a water molecule at the border of the QM region).

3. RESULTS AND DISCUSSION

3.1. Reactant Structure and Tautomeric Form. The starting point for our calculations is an X-ray structure of an UDG−DNA complex, containing the reactant mimic pseudouridine (ΨU). In the enzyme's active site, the sugar moiety of the substrate analogue is flattened from a 2′-endo to a mild 3′-exo pucker, and the nucleobase is rotated around the glycosidic bond.[8] After modifying ΨU to attain the original substrate and subsequent structural optimization, the reactant adopts a flattened 3′-endo conformation, thereby retaining the near-planarity of the deoxyribose and the overall orientation of the base (Figure 2). This enzyme-induced structure resembles the one of the substrate analog and is expected to facilitate bond cleavage and to favor an oxocarbenium-like transition state.[10]

DOI: 10.1021/acs.jctc.8b01305
J. Chem. Theory Comput. 2019, 15, 4344−4350

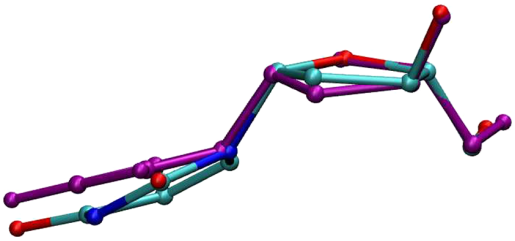


Figure 2. Superposition of the ΨU (in purple) and the minimized dU structures in the UDG active site. Hydrogen atoms are not shown for clarity.

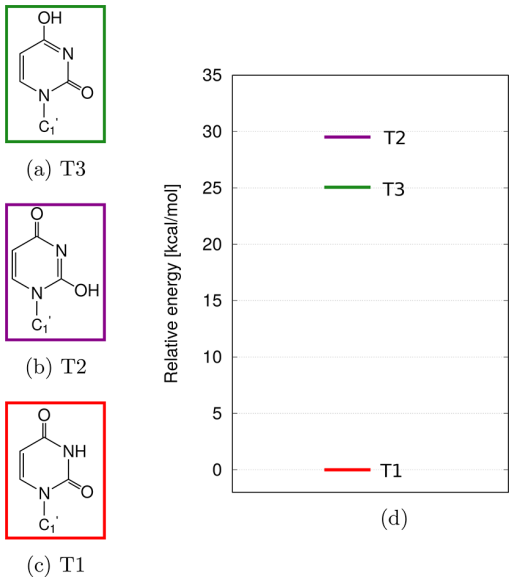


Figure 3. (a,b) Lactim and (c) lactam uracil tautomers. (d) Relative energies of the lactim tautomers compared with the lactam tautomer. These energies result from QM/MM optimizations of the tautomers in the UDG active site at the B3LYP-D3/6-31G** level with a QM region of 170 atoms. All residues within 8 Å of the QM region (2096 atoms) were relaxed, whereas the rest of the MM atoms were kept frozen.

Previous computational studies on the UDG reaction mechanism featured uracil in the lactam (keto) form (see Figure 3c) because this tautomer is the most predominant one in solution.[37] In principle, the lactam form of the base exists in an equilibrium with the lactim (enol) form, and in the enzyme active site it is possible for the latter to be favored over the former. To investigate this possibility, we performed QM/MM optimizations at the B3LYP-D3/6-31G** level of theory for the lactam (see Figure 3c) and lactim (see Figure 3a,b) tautomers in the UDG active site. The energies of T2 and T3 in the UDG−DNA complex relative to the one of the T1 tautomer are presented in Figure 3d.

Our calculations show that in the UDG active site the T2 and T3 tautomers are destabilized by >25 kcal/mol in comparison with the T1 tautomer on account of a reorganization of the H-bonding network and the loss of important contacts with the active-site residues upon a tautomeric shift (see SI, Section 3.2). This indicates that the lactam tautomer is the preferred substrate for the studied reaction.

3.2. Reaction Mechanism. In our investigations of the *N*-glycosidic bond cleavage mechanism, we examined the possibility of a nucleophile activation both by a proton transfer to Asp145 and by a proton transfer to His148. The calculated MEPs reveal that both reactions proceed according to the same mechanism (see Figure 4a). Over the course of the C1′−N1 bond dissociation, the sugar moiety migrates in the direction of the nucleophile, whereas the uracil base and the water molecule remain close to their initial positions (see Figure 5). A mechanism featuring an electrophilic migration is consistent with the results of X-ray studies of a transition-state mimic in the UDG active site.[13] The electrophilic sugar is attracted by the negative charge on Asp145, only to be captured by a water nucleophile, which is held fixed by its interaction with His148 and Asp145. Upon the addition of the water molecule, a proton is simultaneously transferred to a nearby base (see Figure 4b,c). The absence of a barrier for this step indicates that the attack occurs spontaneously on account of the reactivity of the electrophilic sugar (see Figure 4a). The reaction proceeds through an oxocarbenium-like transition state and without a stable intermediate.

Whereas the climbing images of the NEB calculations for a proton transfer to Asp145 and for a proton transfer to His148 differ slightly in their geometries, further refinement of the transition states using the dimer method[24] indicates the presence of only one transition-state structure with the geometry shown in Figure 5 (structure B) and an energy of 10.8 kcal/mol. This transition state is dissociative with a distance between the anomeric carbon of the sugar moiety and N1 of uracil of ∼2.2 Å and a distance between the nucleophile and uracil N1 of ∼2.5 Å (see Figure 5). The planar sp^2 geometry of the 2′-deoxyribose sugar moiety implies an extensive oxocarbenium character for the transition states, which is in accordance with the results of KIE studies.[10]

Whereas the barrier height for the bond hydrolysis is independent of the base employed in the reaction, the product of a proton transfer to Asp145 exhibits a higher stability than the one of a proton transfer to His148 (Figure 4). The difference in the reaction energy amounts to 3.5 kcal/mol and indicates that the addition of the nucleophile is more likely to proceed with Asp145 as the general base. Single-point calculations at the B3LYP-D3/cc-pVTZ, PBE0-D3/6-31G**, and PBE0-D3/cc-pVTZ levels result in changes in the reaction energies; however, the overall trends remain the same (see Table 1). For a QM size of 170 atoms, a proton transfer to Asp145 is slightly preferred over a proton transfer to His148.

3.3. QM Size Convergence. Increasing the size of the QM region typically has an important impact on the reaction energetics,[32−34,36] so that a systematic QM convergence study is crucial for reliable calculations. To assess the converged size of the QM region for the UDG reaction mechanism, we performed single-point calculations on the reaction educt, transition state, and products resulting from the initial QM/MM reaction path calculations (QM region of 170 atoms). We increased the QM size by including residues based on their distance to the initial QM region. Our results, presented in Figure 6, show the size convergence of the reaction energetics and reveal a QM region of 696 atoms to be a good compromise between convergence and size. It includes all enzyme, DNA, and solvent residues within 2.5 Å of the initial QM region.

Whereas single-point calculations are typically sufficient to assess the size of the converged QM region, it is worthwhile to study the influence of structure reoptimizations on the reaction energetics (see Section 2.3). A comparison of the reaction energies for the glycosidic bond cleavage with a proton transfer to Asp145 and a proton transfer to His148, resulting from

DOI: 10.1021/acs.jctc.8b01305
J. Chem. Theory Comput. 2019, 15, 4344−4350

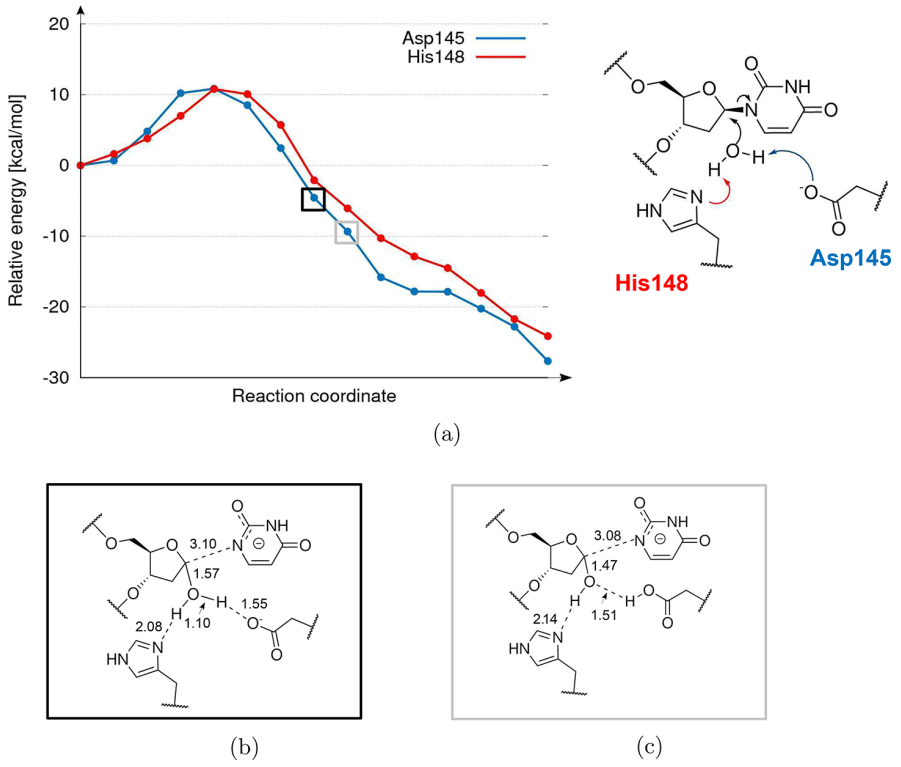


Figure 4. (a) Minimal energy paths for the glycosidic bond cleavage with a proton transfer to Asp145 (blue) and a proton transfer to His148 (red) based on NEB calculations at the B3LYP-D3/6-31G** level for a QM region of 170 atoms. Important distances (in angstroms) for the structures immediately (b) before and (c) after the proton transfer in the case of the Asp145 protonation pathway are noted in the black and gray squares.

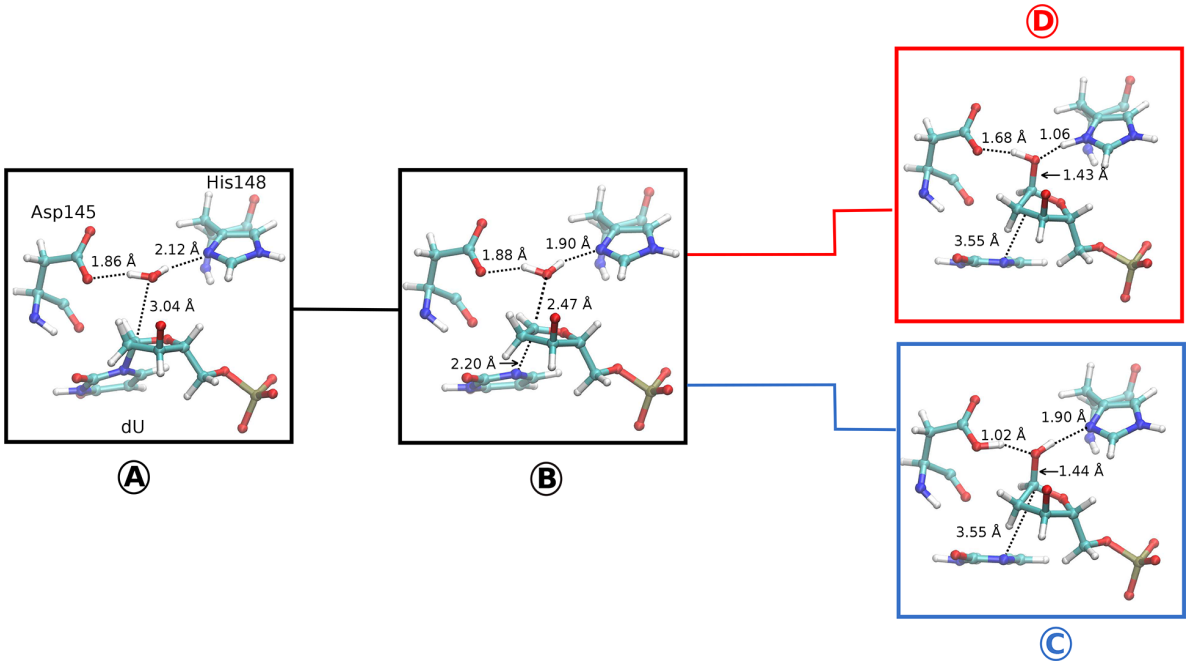


Figure 5. Structures of the (A) reaction educt, (B) transition state, and the reaction products (C) for a proton transfer to Asp145 and (D) for a proton transfer to His148.

calculations with a QM region of 170 and a QM region with 696 atoms both prior to and after the reoptimizations, is presented in Table 2.

Contrary to previous assumptions based on our calculations with a QM region of 170 atoms, the results in Table 2 reveal that for the converged QM size, a proton transfer to His148 is slightly favored over a proton transfer to Asp145, as the product exhibits

DOI: 10.1021/acs.jctc.8b01305
J. Chem. Theory Comput. 2019, 15, 4344−4350

Table 1. Comparison of the Activation (E_A) and Reaction Energies (E_R) of *N*-Glycosidic Bond Cleavage Calculated at Different Levels of Theory with B3LYP-D3/6-31G Structures for the Stationary Points**

		B3LYP-D3		PBE0-D3	
		6-31G**	cc-pVTZ	6-31G**	cc-pVTZ
	E_A (kcal/mol)	10.8	12.1	13.9	15.5
Asp145	E_R (kcal/mol)	−27.6	−21.5	−28.6	−23.2
His148	E_R (kcal/mol)	−24.4	−19.5	−24.6	−20.3

Table 2. Reaction Energies (E_R) for a Proton Transfer to Asp145 and a Proton Transfer to His148, Resulting from Reaction Path Calculations with a QM Region of 170 Atoms, from Single-Point Calculations with a QM Region of 696 Atoms, and from the Reoptimized Structures Calculated at the B3LYP-D3/6-31G Level**

		initial QM, 170 atoms	converged QM, 696 atoms, single point	converged QM, 696 atoms, reoptimized
Asp145	E_R (kcal/mol)	−27.6	−24.4	−24.0
His148	E_R (kcal/mol)	−24.1	−24.7	−25.5

higher stability. Single-point calculations with the PBE0-D3 functional and with a QM region of 1200 and 1400 atoms confirm these results (see Table 3).

This indicates a role of a general base for His148 that is consistent with experimental evidence of an oscillating protonation state for this residue over the course of the reaction.[14] At the same time, our data show that the energetic differences between the two calculated reaction paths are very small, so that a reaction pathway through the protonation of Asp145 is still accessible, which offers an explanation to why mutations of His148 diminish the enzymatic activity but do not fully suppress it.[6] The selective mutation of Asp145 has a more profound influence on the reaction rate[6] because besides coordinating the water nucleophile, this residue provides stabilization to the dissociative transition state and attracts the electrophilic sugar for the subsequent nucleophile addition.[13] The existence of a small, residual enzymatic activity, however, upon the D145N mutation indicates that in this case, a *N*-glycosidic bond cleavage with a proton transfer to His148 can still take place.

4. CONCLUSIONS

In our QM/MM study of the UDG reaction mechanism, we explored the stability of the uracil tautomers in the enzyme's active site and identified the lactam tautomer of the base, which is predominant under physiological conditions in solution, as the preferred substrate for the studied reaction. Further investigations of the reaction mechanism indicate that whereas a proton transfer to Asp145 is favored when a small QM size of 170 atoms is employed, the nucleophile attack for the converged QM size of 696 atoms proceeds with His148 as the general base. This result is consistent with experimental studies that evidence a decrease in the enzymatic activity upon the mutation of His148 and a change in the protonation state of this residue over the course of the reaction. At the same time, an alternative reaction mechanism through a proton transfer to Asp145 explains why mutations of His148 do not result in a complete loss of enzymatic activity.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b01305.

Detailed information on the results of the PROPKA calculations, the force-field MD simulations, the influence

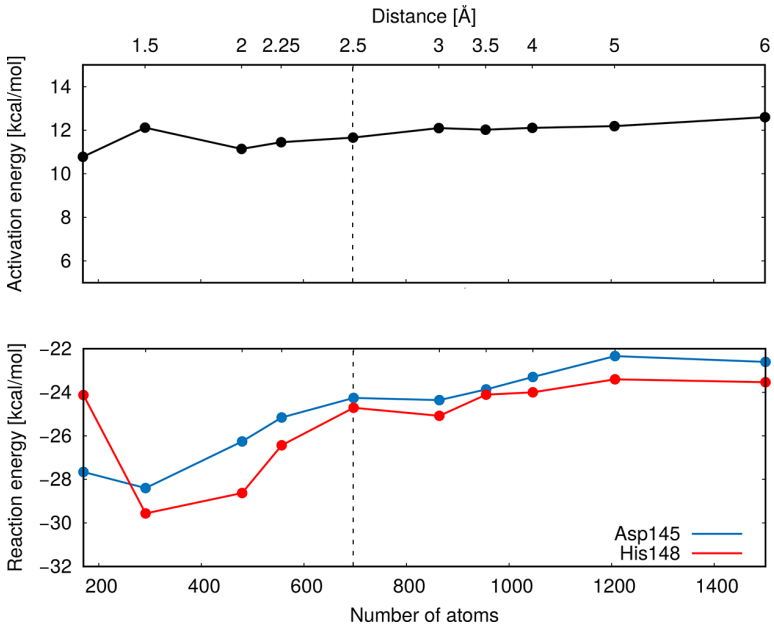


Figure 6. QM size convergence of the activation energy and the reaction energies for a proton transfer to Asp145 (blue) and a proton transfer to His148 (red). The dotted line indicates the QM region, for which a convergence of both the activation and the reaction energies within 2 kcal/mol is reached.

DOI: 10.1021/acs.jctc.8b01305
J. Chem. Theory Comput. 2019, 15, 4344−4350

Table 3. Reaction Energies (E_R) for a Proton Transfer to Asp145 and a Proton Transfer to His148, Resulting from Single-Point Calculations on the Reoptimized Geometries of the Stationary Points

		PBE0-D3/6-31G**	B3LYP-D3/6-31G**		
		696 atoms	696 atoms	1200 atoms	1400 atoms
Asp145	E_R (kcal/mol)	−24.6	−24.0	−21.2	−21.3
His148	E_R (kcal/mol)	−25.8	−25.5	−23.9	−24.1

of the tautomeric form, and the setup of QM/MM size convergence (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: christian.ochsenfeld@cup.uni-muenchen.de.

ORCID

Christian Ochsenfeld: 0000-0002-4189-6558

Funding

We acknowledge financial support by the DFG funding initiatives SFB749 (C7) and the Excellence Cluster EXC114 (CIPSM).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Johannes Dietschreit (LMU Munich) for the helpful discussions.

REFERENCES

(1) Lindahl, T.; Wood, R. D. Quality control by DNA repair. *Science* **1999**, *286*, 1897−905.

(2) Lindahl, T. An N-Glycosidase from Escherichia coli That Releases Free Uracil from DNA Containing Deaminated Cytosine Residues. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 3649−3653.

(3) Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **1993**, *362*, 709−715.

(4) Drohat, A. C.; Xiao, G.; Tordova, M.; Jagadeesh, J.; Pankiewicz, K. W.; Watanabe, K. A.; Gilliland, G. L.; Stivers, J. T. Heteronuclear NMR and crystallographic studies of wild-type and H187Q Escherichia coli uracil DNA glycosylase: Electrophilic catalysis of uracil expulsion by a neutral histidine 187. *Biochemistry* **1999**, *38*, 11876−11886.

(5) Slupphaug, G.; Mol, C. D.; Kavli, B.; Arvai, A. S.; Krokan, H. E.; Tainer, J. A. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **1996**, *384*, 87−92.

(6) Mol, C. D.; Arvai, A. S.; Slupphaug, G.; Kavli, B.; Alseth, I.; Krokan, H. E.; Tainer, J. A. Crystal Structure and Mutational Analysis of Human UraciI-DNA Glycosylase: Structural Basis for Specificity and Catalysis. *Cell* **1995**, *80*, 869−878.

(7) Parikh, S. S.; Mol, C. D.; Slupphaug, G.; Bharati, S.; Krokan, H. E.; Tainer, J. A. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **1998**, *17*, 5214−5226.

(8) Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A. Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5083−5088.

(9) Drohat, A. C.; Stivers, J. T. NMR evidence for an unusually low N1 pKa for uracil bound to uracil DNA glycosylase: Implications for catalysis. *J. Am. Chem. Soc.* **2000**, *122*, 1840−1841.

(10) Werner, R. M.; Stivers, J. T. Kinetic Isotope Effect Studies of the Reaction Catalyzed by Uracil DNA Glycosylase: Evidence for an Oxocarbenium Ion - Uracil Anion Intermediate. *Biochemistry* **2000**, *39*, 14054−14064.

(11) Handa, P.; Acharya, N.; Varshney, U. Effects of mutations at tyrosine 66 and asparagine 123 in the active site pocket of Escherichia coli uracil DNA glycosylase on uracil excision from synthetic DNA oligomers: evidence for the occurrence of long-range interactions between the enzyme and subs. *Nucleic Acids Res.* **2002**, *30*, 3086−3095.

(12) Jiang, Y. L.; Drohat, A. C.; Ichikawa, Y.; Stivers, J. T. Probing the limits of electrostatic catalysis by uracil DNA glycosylase using transition state mimicry and mutagenesis. *J. Biol. Chem.* **2002**, *277*, 15385−15392.

(13) Bianchet, M. A.; Seiple, L. A.; Jiang, Y. L.; Ichikawa, Y.; Amzel, L. M.; Stivers, J. T. Electrostatic guidance of glycosyl cation migration along the reaction coordinate of uracil DNA glycosylase. *Biochemistry* **2003**, *42*, 12455−12460.

(14) Parker, J. B.; Stivers, J. T. Uracil DNA glycosylase: Revisiting substrate-assisted catalysis by DNA phosphate anions. *Biochemistry* **2008**, *47*, 8614−8622.

(15) Dinner, A. R.; Blackburn, G. M.; Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **2001**, *413*, 752−755.

(16) Ma, A.; Hu, J.; Karplus, M.; Dinner, A. R. Implications of alternative substrate binding modes for catalysis by uracil-DNA glycosylase: An apparent discrepancy resolved. *Biochemistry* **2006**, *45*, 13687−13696.

(17) Przybylski, J. L.; Wetmore, S. D. A QM/QM investigation of the hUNG2 reaction surface: The untold tale of a catalytic residue. *Biochemistry* **2011**, *50*, 4218−4227.

(18) Case, D.; Cerutti, D. S.; Cheatham, T.; Darden, T.; Duke, R.; Giese, T. J.; Gohlke, H.; Götz, A.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T.-S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Kollman, P. A. *Amber 2016*; University of California: San Francisco, CA, 2017.

(19) Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.

(20) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurečka, P. Refinement of the SugarPhosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11*, 5723−5736.

(21) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

(22) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(23) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. Climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901−9904.

(24) Heyden, A.; Bell, A. T.; Keil, F. J. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *J. Chem. Phys.* **2005**, *123*, 224101.

(25) Smith, W.; Forester, T. R. DL_POLY_2.0: a general-purpose parallel molecular dynamics simulation package. *J. Mol. Graphics* **1996**, *14*, 136−141.

(26) 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.;

DOI: 10.1021/acs.jctc.8b01305
J. Chem. Theory Comput. 2019, 15, 4344−4350

Schäfer, A.; Lennartz, C. QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1−28.
(27) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W.; Harbach, P. H.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M.; Head-Gordon, M. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184−215.
(28) 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. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(29) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(30) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654−3665.
(31) de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, M.; Kramer, G. J. Zeolite Structure and Reactivity by Combined Quantum-ChemicalClassical Calculations. *J. Phys. Chem. B* **1999**, *103*, 6133−6141.
(32) 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* **2009**, *113*, 11734−11741.
(33) Roßbach, S.; Ochsenfeld, C. Influence of Coupling and Embedding Schemes on QM Size Convergence in QM/MM Approaches for the Example of a Proton Transfer in DNA. *J. Chem. Theory Comput.* **2017**, *13*, 1102−1107.
(34) Hu, L.; Söderhjelm, P.; Ryde, U. On the convergence of QM/MM energies. *J. Chem. Theory Comput.* **2011**, *7*, 761−777.
(35) Liao, R.-Z.; Thiel, W. Comparison of QM-Only and QM/MM Models for the Mechanism of Tungsten-Dependent Acetylene Hydratase. *J. Chem. Theory Comput.* **2012**, *8*, 3793−3803.
(36) Liao, R. Z.; Thiel, W. Convergence in the QM-only and QM/MM modeling of enzymatic reactions: A case study for acetylene hydratase. *J. Comput. Chem.* **2013**, *34*, 2389−2397.
(37) Tsuchiya, Y.; Tamura, T.; Fujii, M.; Ito, M. Keto-enol tautomer of uracil and thymine. *J. Phys. Chem.* **1988**, *92*, 1760−1765.

DOI: 10.1021/acs.jctc.8b01305
J. Chem. Theory Comput. 2019, 15, 4344−4350

Supporting Information

QM/MM study of the Uracil DNA Glycosylase Reaction Mechanism: a Competition between Asp145 and His148

Eli Naydenova,[†,‡] Sven Roßbach,[†,‡] and Christian Ochsenfeld[*,†,‡]

†*Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany.*

‡*Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany.*

E-mail: christian.ochsenfeld@cup.uni-muenchen.de

Contents

1 Protonation states

Table S1: A list of the pKa values for the UDG residues, calculated with PROPKA 3.0[1] and their respective protonation states within the system setup. The abbreviation HIE refers to a histidine with a proton at Nε.

Residue	**pKa**	**protonation state**	**Residue**	**pKa**	**protonation state**
GLU 83	4.60	deprotonated	HIS186	7.00	HIE
GLU 87	4.39	deprotonated	HIS189	6.16	HIE
LYS 90	10.45	protonated	ASP191	2.77	deprotonated
LYS 91	10.41	protonated	LYS197	10.47	protonated
HIS 92	5.79	HIE	HIS212	6.22	HIE
GLU 96	3.47	deprotonated	ARG210	14.06	protonated
LYS 99	11.41	protonated	HIS217	4.94	HIE
GLU111	3.55	deprotonated	LYS218	11.61	protonated
ARG113	10.70	protonated	GLU219	4.39	deprotonated
LYS104	10.32	protonated	ARG220	12.18	protonated
TYR101	13.50	protonated	GLU223	5.11	deprotonated
GLU112	2.35	deprotonated	ASP227	3.84	deprotonated
LYS114	11.13	protonated	TYR248	9.68	protonated
HIS115	6.75	HIE	LYS251	10.46	protonated
TYR116	10.02	protonated	LYS252	9.42	protonated
TYR119	10.92	protonated	ASP257	2.56	deprotonated
HIS123	6.14	HIE	ARG258	12.15	protonated
CYS132	11.41	protonated	LYS259	10.47	protonated
ASP133	3.19	deprotonated	ARG260	12.80	protonated
ASP136	4.22	deprotonated	HIS261	5.35	HIE
LYS135	10.15	protonated	HIS262	5.43	HIE
LYS138	9.21p	rotonated	HIS268	5.34	HIE
ASP145	6.50	deprotonated	ARG276	12.37	protonated
TYR147	15.83	protonated	CYS281	11.99	protonated
HIS148	7.22	HIE	ARG282	13.49	protonated
HIS154	4.64	HIE	HIS283	4.10	HIE
CYS157	13.32	protonated	LYS286	10.47	protonated
ARG162	12.59	protonated	GLU289	4.61	deprotonated
GLU171	4.24	deprotonated	LYS293	10.20	protonated
TYR174	13.27	protonated	LYS296	9.49	protonated
LYS175	10.18	protonated	LYS297	10.57	protonated
GLU176	5.34	deprotonated	ASP300	3.75	deprotonated
ASP180	3.12	deprotonated	LYS302	10.50	protonated
GLU182	4.70	deprotonated	GLU303	4.96	deprotonated
ASP183	3.80	deprotonated			

2 Force Field MD simulations

2.1 MD procedure

For the molecular dynamics simulations the amber force field parameters (ff14SB for the enzyme and OL15 for the DNA)[2] and the NAMD[3] program package were employed. Atom types for 2-deoxyuridine were assigned manually by comparison with uridine and 2-deoxythymidine. Atomic charges were determined by a RESP fit[4] with Q-Chem[5] on the B3LYP/6-31G* level. We employed periodic boundary conditions, particle mesh Ewald summation,[6] the RATTLE algorithm[7] and a step size of 2 fs. Two 10000 step minimizations were performed, first with a frozen solute and then with harmonic restraints of 1 kcal·mol^{-1}·Å^{-2} on the solute. The system was heated from 0 to 300 K in 1 K steps over 30 ps, using the same restraints, and an equilibration over 400 ps in a NVT ensemble was performed. Afterwards the system was equilibrated further for 200 ps in the NPT ensemble with gradually decreasing restraints. The timeframe for the production step was 100 ns. Several trajectories starting from the equilibrated system were computed. Temperature control was achieved by Langevin dynamics and pressure control by the Langevin piston Nosé-Hoover method.[8,9]

2.2 Results

The conformation of the substrate analogue 2-deoxypseudouridine bound in the UDG active site differs significantly from the one of 2-deoxyuridine in unbound DNA. The sugar moiety is flattened to a mild 3-exo pucker with a dihedral d1 between C1, C2, C3, and C4 of -7.5° and the base is rotated around the glycosidic bond (see Fig. S1).[10]

However, over the course of our MD simulations of the UDG-DNA complex, the planarity of the 2-deoxyribose sugar is lost (see Fig. S2) and uracil rotates around the glycosidic bond to retain a conformation, similar to the one in unbound DNA. This rotation is accompanied by the loss of a hydrogen bond between O2 of uracil and Hε of His268 (see Fig. S3). This enzyme-DNA interaction plays a significant role in the stabilization of the transition state

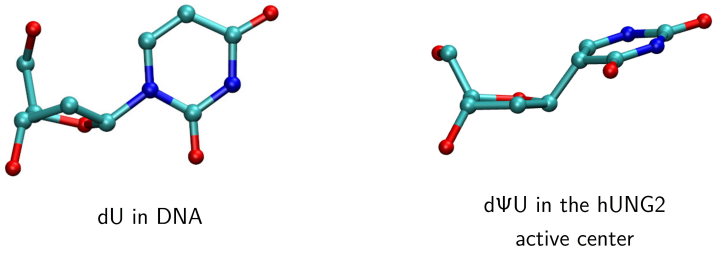


Figure S1: Conformation of the substrate analog (PDB Code: 1EMH) compared dU in unbound DNA. The depicted dU conformation depicted is analogous to the one of dT in a G/T mismatch (PDB Code: 113D).

and has been estimated to lower the barrier height by about 4.8 kcal/mol.[11] Furthermore, mutations of His268 abolish any enzymatic activity.[12]

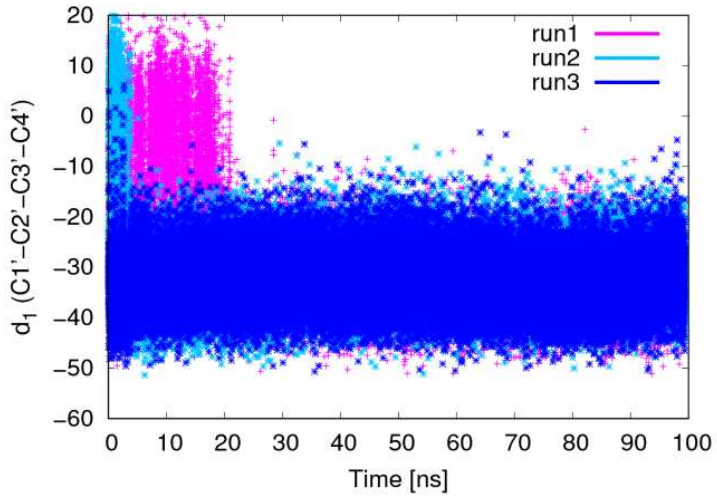


Figure S2: Values for the dihedral d1 between C1, C2, C3, and C4 of the 2'-deoxyribose sugar of dU over the course of the MD simulations. The planarity of the sugar moiety is lost after less than 20 ns of simulation time.

Therefore, the molecular dynamics simulations result in the loss of both the reactive substrate conformation and a UDG-DNA interaction, considered to be crucial for the reaction. It is possible that the extensive parametrization employed in force field molecular dynamics does not allow for the description of the unusual conformation of the nucleobase in the enzymes active site.

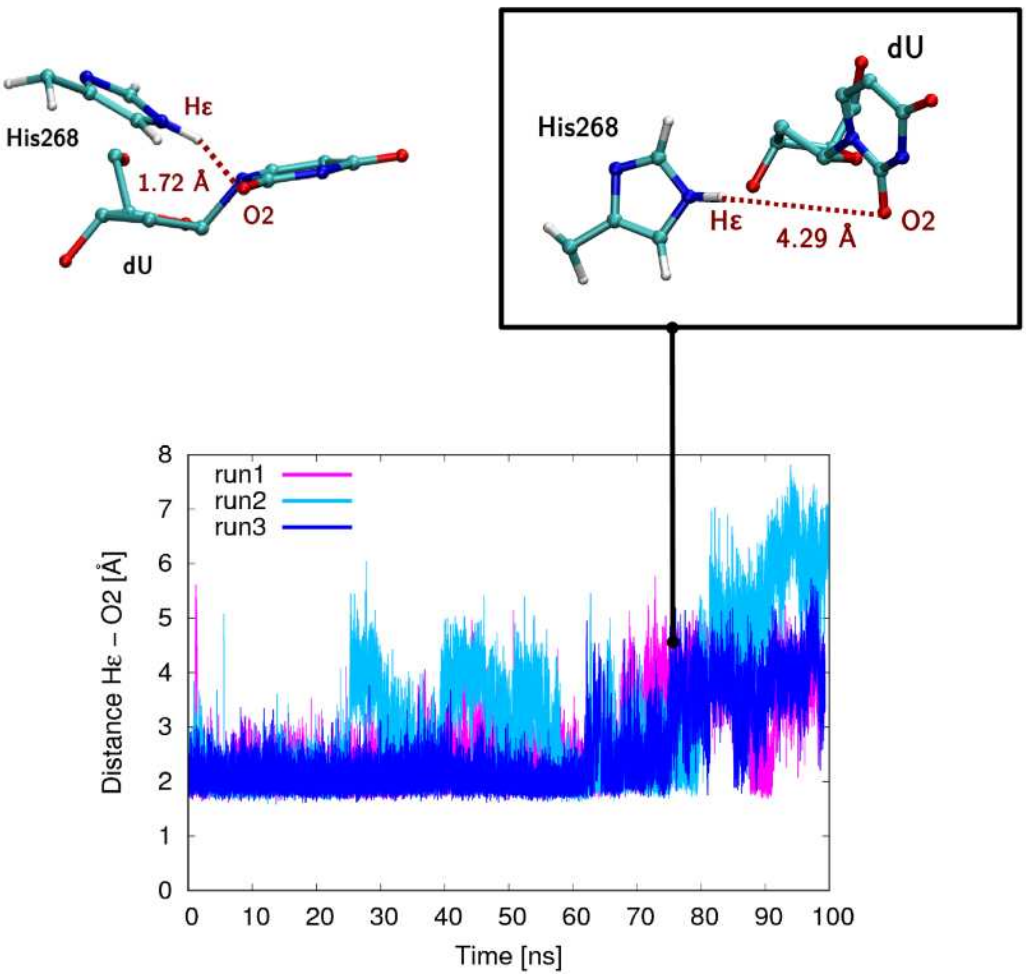


Figure S3: Conformation of dU at the start of the MD preparation and near the end of the MD production. The rotation of the base is accompanied by an increase in the distance between O2 of uracil and H of His268 over the course of the production runs. For clarity, the hydrogen atoms of 2-deoxyuridine are not shown.

3 QM/MM calculations

3.1 Initial QM region

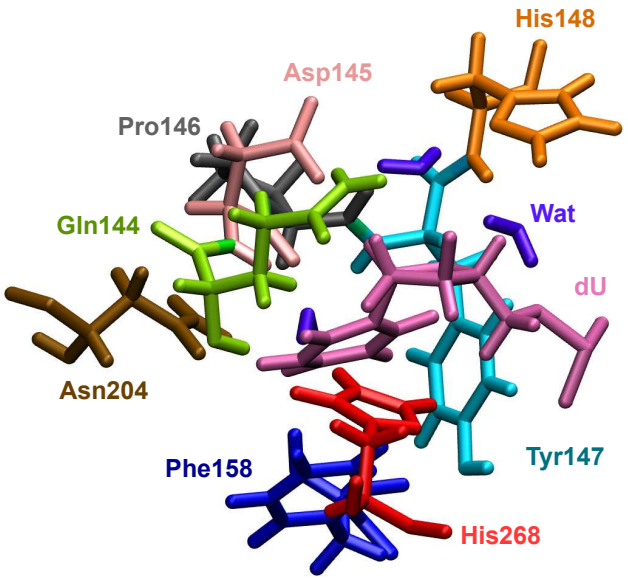


Figure S4: The QM region for the initial QM/MM calculations includes the dU nucleotide, three water molecules and the backbone and side chains of Gln144, Asp145, Pro146, Tyr147, His148, Phe158, Asn204, and His268 (170 atoms).

3.2 Influence of the tautomeric form

A comparison of the structures of the T1 and T2 tautomers indicates that the presence of a proton at O2 of uracil results in a weaker hydrogen bond interaction between O2 and Hε of His268 (see Fig. S5). The short H-bond formed between the carbonyl group of Asn204 and the H atom at N3 of the base in T1 is replaced by a H-bond between the Asn204 carbonyl group and the H atom at O2 in T2, which in term leads to an increase in the distance between the amino hydrogen of Asn204 and O4. All of these factors contribute to the destabilization of the T2 tautomer as compared to T1.

Compared to the T1 tautomer the distance between the O4 atom of uracil and both the amino hydrogen of Asn204 and the backbone of Phe158 in the T3 structure increases. A

hydrogen bond is formed between the proton at O4 and the carbonyl group of Asn204 to replace the H-bonding contact between N3H and the Asn204 carbonyl group in T1, that is lost upon the tautomeric shift to T3. While this tautomer is more stable than T2, it is still destabilized compared to T1.

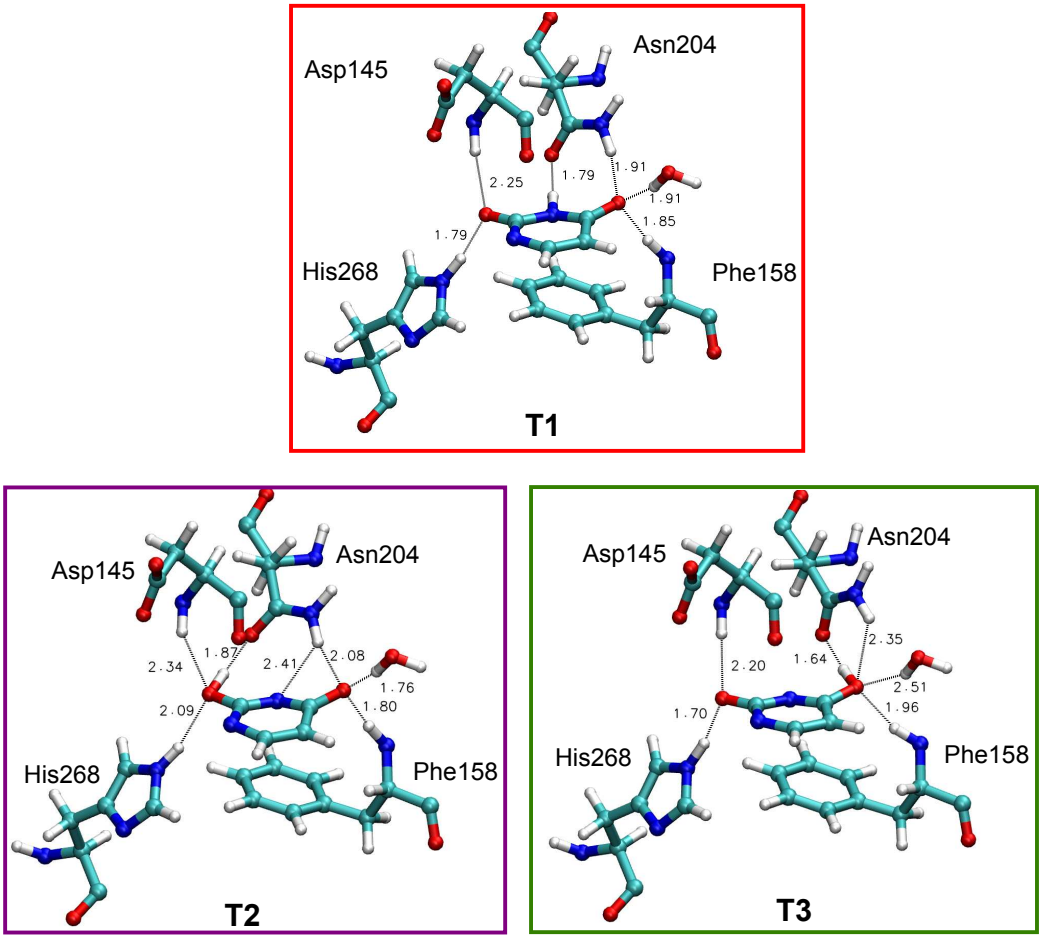


Figure S5: Representation of the T1, T2 and T3 uracil tautomers in the UDG active site. These structures are the result of QM/MM optimizations at the B3LYP-D3/6-31G** level with a QM region of 170 atoms. Important H-bonding contacts with the respective distances (in Å) are indicated with the dotted lines.

3.3 Details of the QM/MM size convergence

The following table includes a list of all DNA and enzyme residues added to the QM region at each step of the convergence study, with numbers corresponding to the X-ray structure (PDB code: 1EMH). The size of the QM region was increased by including residues with at least one atom located within the selected distance from the initial QM region.

Table S2: List of all residues, added to the QM region at each step of the QM/MM convergence study

Distance from the initial QM region	**Residues added to the QM region**	**Number of atoms**	**Charge of the QM region**
Initial QM	DU5, GLN144, ASP145, PRO146, TYR147, HIE148, PHE158, ASN204, HIE268 + 2 water molecules	170	-2
1.5 Å	DT4, GLY143, GLY149, CYS157, SER159, LEU203, ALA205, ALA267, PRO269	291	-3
2 Å	DG3, DA6, DT7, SER169, HIE212, ASN215, SER270 + SER273 + 10 water molecules	479	-6
2.25 Å	TYR119, VAL206, THR208, VAL209, GLY246 + 1 water molecule	556	-6
2.5 Å	GLN152, ALA153, PRO167, LEU170, ILE173, ALA214, TRP245, TYR248 + 2 water molecules	696	-6
3.5 Å	PRO120, ILE141, LEU142, HIE154, LEU156, VAL160, PRO168, LEU192, LEU207, ARG210, ALA211, GLN213, THR226, VAL274 + 7 water molecules	954	-5
4 Å	GLY155, SER216, SER247, ALA249, THR266, PHE278 + 6 water molecules	1012	-4
5 Å	PRO150, GLY190, TRP195, LEU202, HIE217, LYS218, LYS252, PRO271 + 7 water molecules	1205	-3
6 Å	VAL125, TRP128, ASN151, PRO166, GLU171, ASN172, TYR174, TRP222, LEU244, GLN250, LEU272, TYR275, ARG276 + 17 water molecules	1498	-3

References

(1) Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

(2) Case, D. et al. Amber 2016, University of California, San Francisco. 2017.

(3) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(4) Bayly, C. I.; Bayly, C. I.; Cieplak, P.; Cieplak, P.; Cornell, W. D.; Cornell, W. D.; Kollman, P. a.; Kollman, P. a. *J. Phys. Chem.* **1993**,

(5) Shao, Y. et al. *Mol. Phys.* **2015**, *113*, 184–215.

(6) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**,

(7) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24–34.

(8) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

(9) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(10) Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5083–5088.

(11) Drohat, A. C.; Stivers, J. T. *J. Am. Chem. Soc.* **2000**, *122*, 1840–1841.

(12) Mol, C. D.; Arvai, A. S.; Slupphaug, G.; Kavli, B.; Alseth, I.; Krokan, H. E.; Tainer, J. A. *Cell* **1995**, *80*, 869–878.

3.2 Publication II: Reaction Mechanism for the N-Glycosidic Bond Cleavage of 5-Formylcytosine by Thymine DNA Glycosylase

E. Naydenova, J. C. B. Dietschreit, and C. Ochsenfeld

"Reaction Mechanism for the N-Glycosidic Bond Cleavage of 5-Formylcytosine by Thymine DNA Glycosylase"

J. Phys. Chem. B, **15**, 123, 4173 (2019)

Thymine DNA glycosylase (TDG) initiates the base excision repair mechanism for the deamination and oxidation products of cytosine and 5-methylcytosine. This enzyme has a key role in epigenetic regulation, and its catalytic inactivation results in, e.g., mice embryo lethality. Here, we employ molecular dynamics simulations and quantum mechanics/molecular mechanics calculations to investigate the reaction mechanism of the TDG-catalyzed N-glycosidic bond hydrolysis of the modified base 5-formylcytosine. Our results reveal a reaction pathway, which in its first step features a reorganization of the substrate that lowers the barrier height for the subsequent C1'-N1 bond dissociation. The suggested mechanism is consistent with the experimental data, as it is not acid-catalyzed and proceeds through an oxocarbenium-like transition state. It also provides insights into the catalytic roles of the Thr197 and Asn140 residues.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2019, 123, 4173−4179

pubs.acs.org/JPCB

Reaction Mechanism for the *N*-Glycosidic Bond Cleavage of 5-Formylcytosine by Thymine DNA Glycosylase

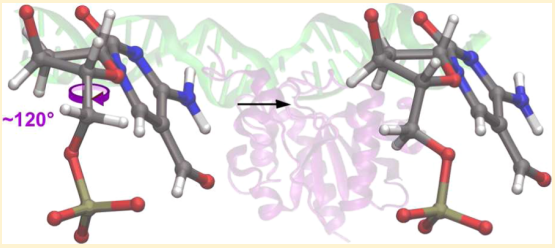
Eli Naydenova,[†,‡] Johannes C. B. Dietschreit,[†,‡] and Christian Ochsenfeld*[,†,‡]

[†]Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany

[‡]Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany

Supporting Information

ABSTRACT: Thymine DNA glycosylase (TDG) initiates the base excision repair mechanism for the deamination and oxidation products of cytosine and 5-methylcytosine. This enzyme has a key role in epigenetic regulation, and its catalytic inactivation results in, e.g., mice embryo lethality. Here, we employ molecular dynamics simulations and quantum mechanics/molecular mechanics calculations to investigate the reaction mechanism of the TDG-catalyzed *N*-glycosidic bond hydrolysis of the modified base 5-formylcytosine. Our results reveal a reaction pathway, which in its first step features a reorganization of the substrate that lowers the barrier height for the subsequent C1′−N1 bond dissociation. The suggested mechanism is consistent with the experimental data, as it is not acid-catalyzed and proceeds through an oxocarbenium-like transition state. It also provides insights into the catalytic roles of the Thr197 and Asn140 residues.



INTRODUCTION

Methylation of the DNA base cytosine to 5-methylcytosine (mC) is a key epigenetic modification known to impact gene expression and embryonic development.[1,2] Abnormal methylation patterns are a common feature in cancer and are associated with tumor suppressor gene silencing[3] and chromosome instability.[4] Although the reaction mechanisms of the DNA methyltransferases, which catalyze the covalent addition of a methyl group to cytosine, have been well characterized,[5] the understanding of the process of active DNA demethylation remains limited. The only established demethylation pathway in mammals proceeds through the stepwise oxidation of mC to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxycytosine (caC) mediated by the ten-eleven translocation enzymes.[6] The modified bases fC and caC are subsequently excised by the enzyme thymine DNA glycosylase (TDG),[7,8] thereby initiating the base excision repair pathway.

TDG is a monofunctional DNA glycosylase employing a water nucleophile to cleave the *N*-glycosidic bond between the base and the sugar moiety, resulting in an abasic site and a free base.[9] Although this enzyme was first identified to excise the canonical base thymine from T:G mispairs,[10] it has a wide variety of substrates, including deamination and oxidation products of both cytosine and mC.[8,11] One essential aspect of the TDG reaction mechanism, which is independent of the substrate involved, is the severe loss of the glycosylase activity upon N140A mutation.[12] The side chain of this residue coordinates the nucleophile, and the sharp decrease in the reaction rate is presumed to stem from the loss of this interaction.[13] Mutation of Thr197 also diminishes the catalytic activity, as its side chain positions Asn140 and its backbone contacts the putative nucleophile (see Figure 1a).[13]

For the substrate fC, the TDG reaction rate shows no dependence on the pH.[14] Considering the lack of any evidence for the involvement of a general acid in the reaction mechanism,[15] an activation of the modified base through its protonation is not feasible.[14,15] The *N*-glycosidic bond cleavage is therefore presumed to proceed through a highly dissociative transition state or an oxocarbenium intermediate (see Figure 1b),[14] with the departing base mostly stabilized through hydrogen bonds to backbone atoms of the enzyme.[15]

Although the analysis of the X-ray structure[15,16] and mutational studies[14,15] have provided significant insight into the underlying mechanism of the fC excision, an atomistic scale investigation of this reaction has not yet been performed. Here, we present a reaction mechanism for the TDG-catalyzed fC hydrolysis based on classical molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) simulations, which features a substrate reorganization that significantly lowers the barrier height for the subsequent glycosidic bond dissociation. Consistent with the experimental data, this mechanism is not acid catalyzed and provides further

Received: December 5, 2018
Revised: April 16, 2019
Published: May 1, 2019

DOI: 10.1021/acs.jpcb.8b11706
J. Phys. Chem. B 2019, 123, 4173−4179

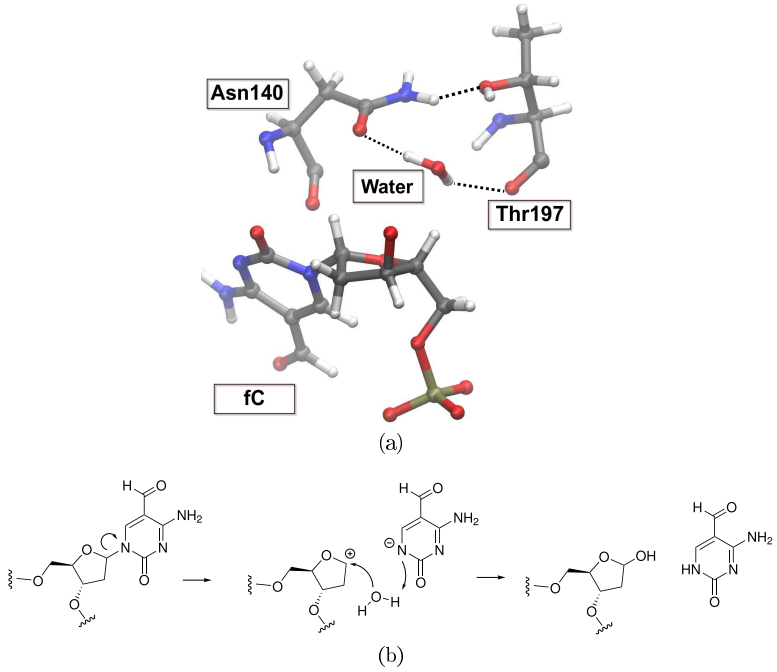


Figure 1. (a) Nucleophile-binding mechanism in TDG: the side chain of Asn140 and the backbone of Thr197 contact the water molecule. This representation stems from an X-ray structure of a TDG−DNA complex (PDB code: 5t2w)[15] with hydrogens added for clarity. (b) Suggested mechanism for the *N*-glycosidic bond cleavage of fC.[14] The reaction is expected to proceed through an oxocarbenium ion intermediate or a highly dissociative, oxocarbenium-like transition state.

insight into the catalytic role of the Asn140 and Thr197 residues.

COMPUTATIONAL DETAILS

Molecular Dynamics. Starting point of our simulations is the crystal structure of a TDG−DNA complex, featuring 2′-fluoroarabino-5-formylcytidine (PDB code: 5t2w).[15] The fluorine atom of the inhibitor was substituted by a hydrogen and the structure was protonated, solvated in a box of TIP3P[17] water with a buffer of 10 Å around the solute, and neutralized with Na^+ ions using the LEAP module of AmberTools16.[18] For the parametrization of fC, we employed antechamber[18] and the GAFF force field,[19] with charges calculated by a RESP fit at the HF/6-31G* level. Force-field MD simulations were performed with the NAMD 2.10 engine[20] and Amber force field parameters (ff14SB for the enzyme and OL15 for the DNA)[18] (for details on the MD simulation see the Supporting Information (SI), Section 1.1).

Considering the proximity of the residue His151 to fC and its unknown protonation state, we performed MD simulations featuring both a positively charged His151 and a neutral His151, with a hydrogen only on Nε (3 trajectories per protonation state, each 100 ns long). We disregarded the possibility for a neutral His151 with a proton at Nδ on the account of the observed short distance between Nε and the backbone oxygen of Pro125 in every crystal structure of a TDG−DNA complex, independent of the substrate or crystallization conditions.[15,21,22] This short distance is indicative of a hydrogen bond, which would be lost if there is no proton at Nε.

On the account of pK_a values, calculated with Karlsberg+[23−25] and an analysis of the H-bond occupancy, a neutral protonation state for His151 was adopted for the following QM/MM calculations (see the SI, Section 1.2). Though the selected His151 protonation state differs from the one employed by Kanaan et al.[26] in their investigation of the TDG-catalyzed excision of thymine, it is consistent with experimental studies, which show no pH dependency of the enzymatic activity and no evidence of the involvement of a general acid in the TDG-catalyzed fC excision.[14]

QM/MM Calculations. In the crystal structures of the TDG−DNA complex, the base is rotated approximately 40° around the glycosidic bond as compared to the free substrate (Figure S2). This conformation is found for different TDG substrates and crystallization conditions.[15,21,22] We, therefore, grouped the MD structures according to the torsion angle χ (O4′−C1′−N1−C2) and selected 20 frames for the QM/MM calculations from a group with χ values close to the ones found in the crystal structure. The minimal energy paths for these 20 frames were calculated using adiabatic mapping. For the frame with the lowest barrier height, the reaction path was refined with the climbing image nudged elastic band method (CI-NEB).[27]

The QM region includes the fC nucleotide, the water nucleophile, and the residues Ile139, Asn140, Ala145, His151, Tyr152, Asn191, and Thr197 (144 QM atoms). Asn140 and Thr197 are located close to the water molecule. The backbone amide group of Tyr152 contacts the formyl oxygen, whereas the backbones of Ile139 and Asn140 provide hydrogen bonds to the O2 atom of fC.[15]

Mutational studies reveal that the H151A mutation results in a 2-fold decrease in the TDG enzymatic activity.[14] A role of a general base for His151 has been discounted by experiments,[14] so this residue could help position the flipped base for its interaction with the rest of the active site residues. Experiments show that a mutation of Ala145 leads to a 3-fold decrease in

DOI: 10.1021/acs.jpcb.8b11706
J. Phys. Chem. B 2019, 123, 4173−4179

the enzymatic activity.[14] The side chain of this residue forms a nonpolar contact with the fC formyl carbon and is assumed to position the fC formyl oxygen for its interaction with Tyr152.[15] We included Ala145, His151, and Asn191 as part of the QM region, based on their proximity and interactions with the base. Although their influence on the enzymatic activity according to the mutational studies is minor,[14] their inclusion in the QM region allows for a comprehensive description of the reaction energetics.

Calculations of the QM part were performed at the B3LYP-D3[28] level of theory with the def2-svp basis set,[29] using the FermiONs++ program package.[30−32] For the MM part, we employed the DL-POLY interface[33] implemented in ChemShell[34] and Amber parameters.[18] All residues within 8 Å of the QM region were optimized to ensure the relaxation of the system, whereas the rest were kept frozen.

RESULTS AND DISCUSSION

Substrate Reorganization. To explore the underlying mechanism of TDG-catalyzed fC excision, we investigated the dissociation of the glycosidic bond between the C1′ and N1 atoms of fC both through adiabatic mapping and NEB calculations (see Figure 2a, blue curve). Our calculations indicate that over the course of the studied reaction, the base remains at its initial position, whereas the electrophilic sugar migrates in the direction of the nucleophile (see Figure 2b). The bond dissociation proceeds through a nearly planar oxocarbenium-like transition state, and the addition of the nucleophile is accompanied by a proton transfer to the nearby Asn140 residue. Over the course of our investigations, we employed different reaction coordinates for the adiabatic mapping procedure (e.g., r(C1′−N1), r(C1′−O(nuc)), r(C1′−N1)−r(C1′−O(nuc)), etc.) However, none of the calculated reaction paths was able to capture a stable intermediate. The NEB calculations also did not provide any evidence for the existence of such species.

Although a mechanism featuring an electrophile migration of the sugar moiety is reminiscent of the enzyme UDG,[35] the TDG-catalyzed glycosidic bond cleavage of fC has a high-energy barrier of 30.3 kcal/mol, and its product is unstable. At the same time, an alternative mechanism featuring an activation of the base through a N3 protonation has been refuted by experiments,[14,15] as the reaction is not acid-catalyzed, and a direct proton transfer from the nucleophile to the base is not feasible due to a distance of more than 4 Å in between. We, therefore, investigated the possibility of a preactivation of the substrate prior to the C1′−N1 bond dissociation.

A comparison of the crystal structures of the TDG reaction inhibitor, 2′-fluoroarabino-5-formylcytidine, and the reaction product, the abasic site (PDB codes: 5t2w and 4z7b), pinpoints a significant change in the conformation of the sugar. The dihedral γ between C3′, C4′, C5′, and O5′ differs in those two structures by approximately 120° (see Figure 3). The conformation of the educt, 5-formylcytidine, is similar to that of the inhibitor. We, therefore, calculated a reaction path for the rotation around γ, which resulted in an only slightly destabilized intermediate (reaction energy of +1.5 kcal/mol, D in Figure 2a). Following this substrate reorganization, the barrier height of a subsequent glycosidic bond cleavage is lowered by around 7 kcal/mol and the product of the reaction is more stable (see Figure 2a, red curve, activation energy of +23.3 kcal/mol). Similar to the reaction without the

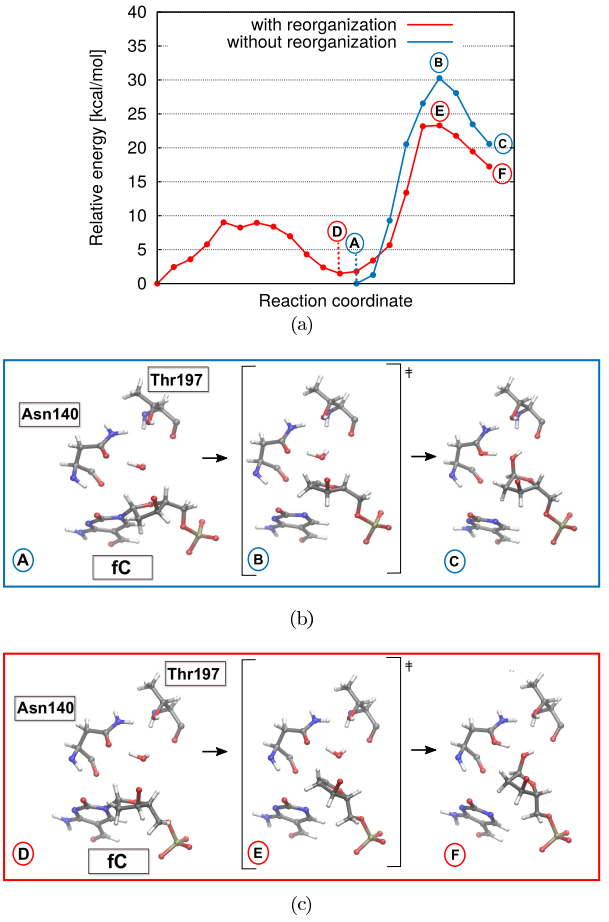


Figure 2. (a) Energy profiles for the glycosidic bond cleavage with (red curve) and without the substrate reorganization (blue curve). (b) Structures of the educt, transition state, and product for a reaction, which does not feature the substrate rearrangement. (c) Structures of the reaction educt, intermediate, and product for a reaction featuring the substrate rearrangement.

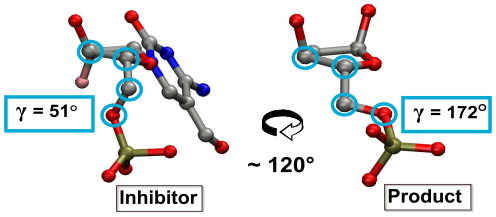


Figure 3. Conformation of the TDG inhibitor and reaction product (PDB codes 5t2w[15] and 4z7b[16]). The dihedral γ between the atoms O5′, C5′, C4′, and C3′ (shown in the blue circles) differs by about 120°. The conformation of the sugar moiety of the inhibitor 2′-fluoroarabino-5-formylcytidine is similar to that of fC and to that of pyrimidine DNA bases in B-DNA.

rearrangement, the C1′−N1 dissociation for the activated substrate proceeds through a planar, oxocarbenium-like transition state with the addition of the nucleophile resulting in a protonated Asn140 residue (see Figure 2c).

The results presented in Figure 2a suggest that TDG-catalyzed glycosidic bond hydrolysis proceeds via a stepwise mechanism, with the substrate rearrangement as the first step and the C1′−N1 bond dissociation coupled with the

DOI: 10.1021/acs.jpcb.8b11706
J. Phys. Chem. B 2019, 123, 4173−4179

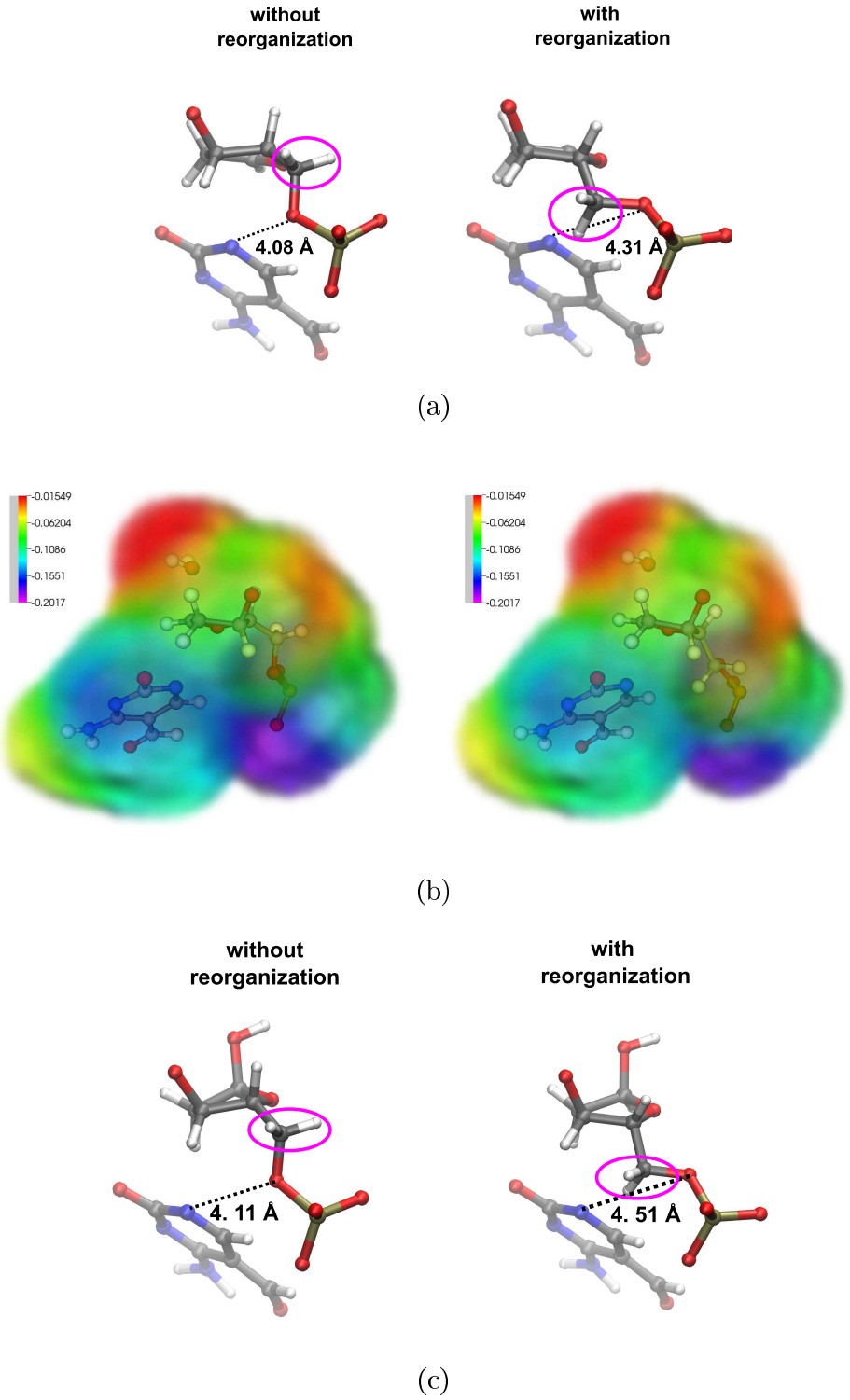


Figure 4. (a) Structures of the transition states of the *N*-glycosidic bond cleavage for a reaction without a substrate reorganization (left) and for a reaction featuring a substrate reorganization (right). (b) Electrostatic potential maps of both transition states, calculated with B3LYP and a def2-svp basis set. The rotation around the dihedral γ prior to the bond dissociation results in an increase in the distance between O5′ of the sugar and N1 of the base. In case of a reorganized substrate, the CH_2 group, highlighted in a pink circle, shields the negative charge of the phosphate group, thereby reducing its electrostatic repulsion with the negatively charged base. (c) Structures of the product of the *N*-glycosidic bond cleavage for a reaction without a substrate reorganization (left) and for a reaction featuring a substrate reorganization (right).

nucleophile addition as the second step. The decrease in barrier height following the rotation around γ raises the question of how the substrate reorganization facilitates the C1′−N1 bond dissociation. Over the course of the cleavage of

DOI: 10.1021/acs.jpcb.8b11706
J. Phys. Chem. B 2019, 123, 4173−4179

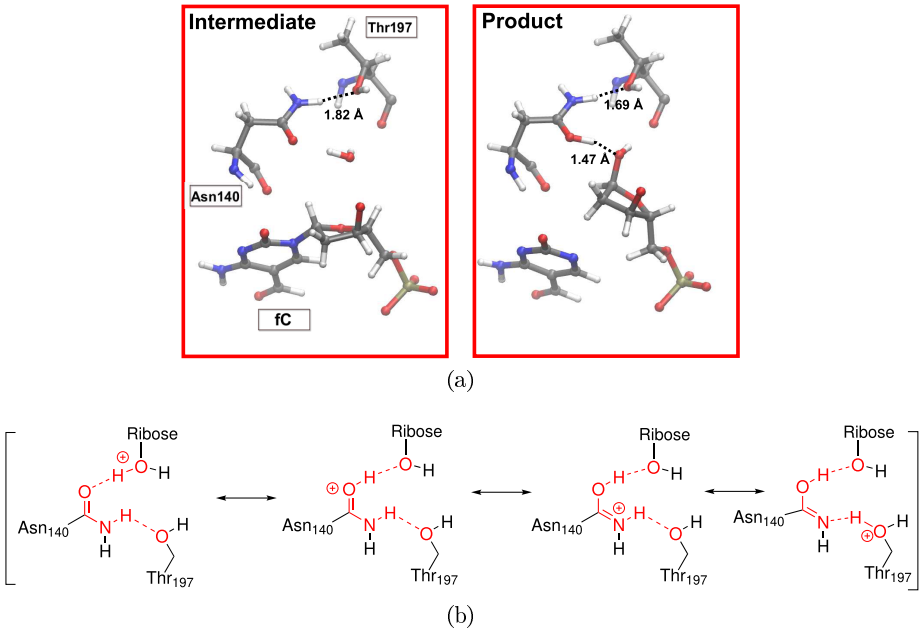


Figure 5. (a) Structures of the reaction intermediate (D in Figure 2) and product (F in Figure 2) for a reaction mechanism with substrate reorganization. Upon the addition of the water nucleophile, the hydrogen bond between the side chains of Asn140 and Thr197 is tightened. (b) Resonance structures for the positive charge delocalized over an extensive H-bonding network consisting of Asn140, Thr197, and the abasic site.

the glycosidic bond, a negative charge develops on fC, which results in an electrostatic repulsion between the base and O5′ of the sugar moiety. Through the rearrangement of the substrate prior to the excision, the distance between O5′ and the base in the TS and product increases (see Figure 4a,c), which can lead to a decrease in the barrier height. Further on, following the rotation, the negative charge of the phosphate group is also slightly shielded by the C5′ atom and its hydrogens (see Figure 4a,b). The substrate reorganization is accompanied by an increase in the distance between the O5′ atom of fC and the hydrogen of the hydroxyl group of Ser273 (see Figure S4).

An important characteristic for the TDG-catalyzed reactions is the dependence of the enzymatic activity on the N1 acidity of the leaving group, namely, the base.[11,15] The significant N1 acidity of fC is attributed to the resonance stabilization of the fC anion via charge delocalization to the formyl group.[14] Considering that the reaction results in an fC anion, a dependence of the enzymatic activity on the N1 acidity is consistent with the suggested mechanism. Compared to fC, the N1 acidity of cytosine and the respective stability of the cytosine anion are significantly lower,[14] which could explain why this base is not excised by TDG.

Role of Asn140 and Thr197. In the crystal structure of the TDG−DNA complex, there is a short distance between the Oδ of Thr197 and the amide group of Asn140, which suggests that a hydrogen bond is formed. On the account of this interaction and mutational studies, which evidence a decrease in the enzymatic activity upon the T197A mutation, this residue is presumed to position the Asn140 side chain, which, in turn, coordinates the water nucleophile. In our MD simulations, this hydrogen bond is retained with an overall occupancy of over 80%, which confirms the suspected role of this residue. Our QM/MM calculations, however, indicate a further involvement of Thr197 in the TDG reaction mechanism. As the addition of the nucleophile proceeds, the distance between the Thr197 hydroxyl oxygen and the Asn140 amide hydrogen decreases from to 1.82 to 1.69 Å (Figure 5a).

In solution, a proton transfer to the carbonyl group of an asparagine is unlikely to take place. However, in the TDG active site, the positive charge that develops upon the addition of the nucleophile is stabilized by hydrogen bond interactions between Asn140, Thr197, and the abasic site (see Figure 5a). The short distances between proton acceptors and donors for these interactions indicate that the reaction results in a proton delocalized over an extensive H-bonding network consisting of Asn140, Thr197, and the abasic site (see Figure 5b). The reaction energy for this product is relatively high (17.2 kcal/mol, see Figure 2a); however, further relaxation of the system is expected upon inclusion of dynamic effects and upon the release of the base into the solvent.

In some of the MD frames, a second water molecule can be found between the backbone of Thr197 and the nucleophile (Figure S5). We performed adiabatic mapping calculations for the glycosidic bond cleavage on one of these MD frames (Figure S6a), with structures of the reaction educt and product presented in Figure S6b. The presence of a second water molecule in the TDG active site alters the H-bonding network. The distance between the carbonyl group of Asn140 and the proton of the water nucleophile is longer, and the hydroxyl group of Thr197 is oriented toward the second water molecule (Figure S6b). Still, the reaction proceeds through a migration of the electrophile in the direction of the water nucleophile and results in a proton delocalized over an extensive H-bonding network consisting of Asn140, Thr197, the second water molecule, and the abasic site.

Continuation of the Mechanism. The enzyme UDG belongs to the same enzyme superfamily as TDG and employs a similar mechanism featuring an electrophilic migration of a sugar moiety for the glycosidic hydrolysis of uracil.[35] The reaction results in a negatively charged uracil and a proton transfer to a nearby Asp145 residue. Heteronuclear NMR

DOI: 10.1021/acs.jpcb.8b11706
J. Phys. Chem. B 2019, 123, 4173−4179

experiments have shown that following the glycosidic bond cleavage, uracil remains as an anion in the UDG active site until its release into solution.[36,37] Considering the similarities of the reaction mechanisms of TDG and UDG, it is possible that the same applies for fC in the TDG active site.

For TDG, the C1′−N1 bond dissociation of fC results in a negatively charged base and a positive charge delocalized on Asn140, Thr197, and the abasic site. The fC anion is stabilized by interactions with the enzyme backbone and the side chain of Asn191 (see Figure S7) and by the delocalization of the negative charge over the π-system.

A direct proton transfer of the delocalized proton to the base is not feasible, as the distance in-between amounts to more than 5 Å. However, it is possible that the fC anion retains its negative charge until its release into solution or that the proton is transferred during the release of the product. A proton transfer to the base might also involve a bulk water molecule.

The escape of the excised base is assumed to proceed along a solvent-filled channel that runs along the target DNA strand (see Figure 6).[16,22] An investigation of the product release along this channel would involve either steered MD simulations[38] or ideally extensive path integral MD simulations, which are, however, beyond the scope of this study.

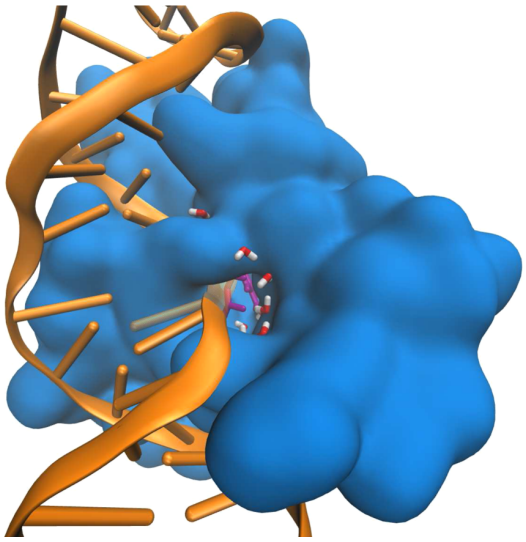


Figure 6. Representation of the solvent-filled channel in the TDG active site: fC is shown in magenta, TDG in blue, and the DNA in orange. The channel runs from the TDG active site to the enzyme surface along the target DNA strand.[16]

CONCLUSIONS

In this work, we employed force-field MD simulations and QM/MM calculations to explore the reaction mechanism of the TDG-catalyzed fC excision. Our results indicate that prior to the glycosidic bond cleavage, the reaction substrate is rearranged by the rotation of the phosphate group around the C4′−C5′ bond. This conformational change activates the substrate for the next step of the reaction pathway, i.e., the C1′−N1 bond dissociation, by diminishing the electrostatic repulsion between the anionic base and the phosphate backbone and reduces the barrier height by around 7 kcal/mol. The *N*-glycosidic bond cleavage proceeds through the migration of the electrophilic sugar in the direction of the bound water molecule, with the addition of the nucleophile, resulting in a positive charge delocalized over an extensive H-bonding network consisting of Asn140, Thr197, and the oxygen of the abasic site. Our calculations indicate an involvement of Thr197 in the reaction mechanism that goes beyond its previously suggested role and provide an explanation for the experimentally observed decrease in the enzymatic activity upon the T197A and N140A mutations.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b11706.

MD simulations; protonation state of His151; and stabilization of the anionic base (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: christian.ochsenfeld@cup.uni-muenchen.de. Phone: +49 89 2180 77921.

ORCID

Johannes C. B. Dietschreit: 0000-0002-5840-0002

Christian Ochsenfeld: 0000-0002-4189-6558

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge financial support by the DFG funding initiatives SFB749 (C7) and the Excellence Cluster EXC114 (CIPSM).

REFERENCES

(1) Cortázar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; MacDougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A. L.; Siegrist, F.; et al. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **2011**, *470*, 419−423.

(2) Cortellino, S.; Xu, J.; Sannai, M.; Moore, R.; Caretti, E.; Cigliano, A.; Le Coz, M.; Devarajan, K.; Wessels, A.; Soprano, D.; et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **2011**, *146*, 67−79.

(3) Kazanets, A.; Shorstova, T.; Hilmi, K.; Marques, M.; Witcher, M. Epigenetic silencing of tumor suppressor genes: Paradigms, puzzles, and potential. *Biochim. Biophys. Acta* **2016**, *1865*, 275−288.

(4) Jones, P. A.; Gonzalgo, M. L. Altered DNA methylation and genome instability: A new pathway to cancer? *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2103−2105.

(5) Cheng, X.; Blumenthal, R. M. Mammalian DNA methyltransferases: A structural perspective. *Structure* **2008**, *16*, 341−350.

(6) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011**, *333*, 1300−1303.

(7) He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **2011**, *333*, 1303−1307.

(8) Maiti, A.; Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **2011**, *286*, 35334−35338.

(9) Stivers, J. T.; Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **2003**, *103*, 2729−2759.

DOI: 10.1021/acs.jpcb.8b11706
J. Phys. Chem. B 2019, 123, 4173−4179

(10) Neddermann, P.; Jiricny, J. The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J. Biol. Chem.* **1993**, *268*, 21218−21224.
(11) Bennett, M.; Rodgers, M.; Hebert, A.; Ruslander, L.; Eisele, L.; Drohat, A. Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* **2006**, *128*, 12510−12519.
(12) Hardeland, U.; Bentele, M.; Jiricny, J.; Schar, P. Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J. Biol. Chem.* **2000**, *275*, 33449−33456.
(13) Maiti, A.; Noon, M. S.; MacKerell, A. D.; Pozharski, E.; Drohat, A. C. Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 8091−8096.
(14) Maiti, A.; Michelson, A. Z.; Armwood, C. J.; Lee, J. K.; Drohat, A. C. Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J. Am. Chem. Soc.* **2013**, *135*, 15813−15822.
(15) Pidugu, L. S.; Flowers, J. W.; Coey, C. T.; Pozharski, E.; Greenberg, M. M.; Drohat, A. C. Structural basis for excision of 5-formylcytosine by thymine DNA glycosylase. *Biochemistry* **2016**, *55*, 6205−6208.
(16) Malik, S. S.; Coey, C. T.; Varney, K. M.; Pozharski, E.; Drohat, A. C. Thymine DNA glycosylase exhibits negligible affinity for nucleobases that it removes from DNA. *Nucleic Acids Res.* **2015**, *43*, 9541−9552.
(17) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(18) Case, D.; Cerutti, D. S.; Cheatham, T.; Darden, T.; Duke, R.; Giese, T. J.; Gohlke, H.; Götz, A.; Greene, D.; Homeyer, N. et al. *Amber 2016*; University of California: San Francisco, 2017.
(19) 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.* **2004**, *25*, 1157−1174.
(20) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(21) Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G.; Luo, C.; Jiang, H.; He, C. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **2012**, *8*, 328−330.
(22) Coey, C. T.; Malik, S. S.; Pidugu, L. S.; Varney, K. M.; Pozharski, E.; Drohat, A. C. Structural basis of damage recognition by thymine DNA glycosylase: Key roles for N-terminal residues. *Nucleic Acids Res.* **2016**, *44*, 10248−10258.
(23) Kieseritzky, G.; Knapp, E.-W. Optimizing pKa computation in proteins with pH adapted conformations. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 1335−1348.
(24) Rabenstein, B.; Knapp, E. W. Calculated pH-dependent population and protonation of carbon-monoxy-myoglobin conformers. *Biophys. J.* **2001**, *80*, 1141−1150.
(25) Meyer, T.; Kieseritzky, G.; Knapp, E.-W. Electrostatic pKa computations in proteins: Role of internal cavities. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3320−3332.
(26) Kanaan, N.; Crehuet, R.; Imhof, P. Mechanism of the glycosidic bond cleavage of mismatched thymine in human thymine DNA glycosylase revealed by classical molecular dynamics and quantum mechanical/molecular mechanical calculations. *J. Phys. Chem. B* **2015**, *119*, 12365−12380.
(27) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. Climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901−9904.
(28) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, No. 154104.
(29) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(30) Kussmann, J.; Ochsenfeld, C. Pre-selective screening for matrix elements in linear-scaling exact exchange calculations. *J. Chem. Phys.* **2013**, *138*, No. 134114.
(31) Kussmann, J.; Ochsenfeld, C. Preselective screening for linear-scaling exact exchange-gradient calculations for graphics processing units and general strong-scaling massively parallel calculations. *J. Chem. Theory Comput.* **2015**, *11*, 918−922.
(32) Kussmann, J.; Ochsenfeld, C. Hybrid CPU/GPU integral engine for strong-scaling ab initio methods. *J. Chem. Theory Comput.* **2017**, *13*, 3153−3159.
(33) Smith, W.; Yong, C. W.; Rodger, P. M. DL_POLY: Application to molecular simulation. *Mol. Simul.* **2002**, *28*, 385−471.
(34) 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.; et al. QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1−28.
(35) Bianchet, M. A.; Seiple, L. A.; Jiang, Y. L.; Ichikawa, Y.; Amzel, L. M.; Stivers, J. T. Electrostatic guidance of glycosyl cation migration along the reaction coordinate of uracil DNA glycosylase. *Biochemistry* **2003**, *42*, 12455−12460.
(36) Stivers, J. T.; Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **2003**, *103*, 2729−2760.
(37) Drohat, A. C.; Stivers, J. T. *Escherichia coli* uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2. *Biochemistry* **2000**, *39*, 11865−11875.
(38) Da, L. T.; Shi, Y.; Ning, G.; Yu, J. Dynamics of the excised base release in thymine DNA glycosylase during DNA repair process. *Nucleic Acids Res.* **2018**, *46*, 568−581.

DOI: 10.1021/acs.jpcb.8b11706
J. Phys. Chem. B 2019, 123, 4173−4179

Supporting Information for

Reaction Mechanism for the N-Glycosidic Bond Cleavage of 5-Formylcytosine by Thymine DNA Glycosylase

Eli Naydenova,[†,‡] Johannes C.B. Dietschreit,[†,‡] and Christian Ochsenfeld[*,†,‡]

†*Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany.*

‡*Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany.*

E-mail: christian.ochsenfeld@cup.uni-muenchen.de

Phone: +49 89 2180 77921

Contents

1 Force Field MD simulations

1.1 MD procedure

For the molecular dynamics simulations we employed Amber 16 force field parameters[1] (ff14SB for the enzyme and OL15 for the DNA) and the NAMD 2.10 program package.[2] The preparation of the system included two 10000 step conjugate gradient minimizations, the first with a frozen solute and the second with harmonic restraints of 1 kcal·mol^{-1}·Å^{-2}. Using the same restraints, the system was heated from 0 to 300 K in 1 K steps over 30 ps, and an equilibration over 200 ps in an NVT ensemble was performed. The system was further equilibrated for 400 ps in the NPT ensemble with gradually decreasing restraints. In this step, we used Langevin dynamics for temperature control and Langevin piston Nosé-Hoover method for pressure control.[3,4] Starting from the equilibrated structure, we computed several 100 ns trajectories. Our simulations employed the RATTLE algorithm[5] and have a step size of 2 fs.

1.2 Parameters for 2'-deoxyformylcytidine

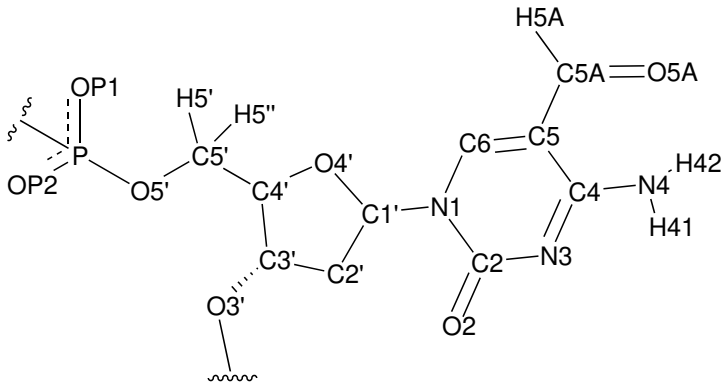


Figure S1: Atom names of the fC atoms

Table S1: Parameters for fC

Atom name	Charge	Atom name	Charge
P	0.9111983625	C5A	0.6632867980
OP1	-0.8119154556	O5A	-0.6262119667
OP2	-0.7599721674	H5A	0.0173058691
O5'	-0.2790439503	C4	1.1411368214
C5'	-0.2886251205	N4	-1.1592934909
H5'	0.1828576770	H41	0.4689297168
H5"	0.1190050671	H42	0.5172281692
C4'	0.2612349840	N3	-0.9756087825
H4'	0.0603807190	C2	0.9885195877
O4'	-0.4683633094	O2	-0.6907166092
C1'	0.2571176230	C3'	0.3052979526
H1'	0.1710959212	H3'	0.0649964109
N1	-0.3667171927	C2'	-0.6101344556
C6	0.3998355381	H2'	0.2480675126
H6	0.0641811214	H2"	0.1785463519
C5	-0.7203394180	O3'	-0.2632802849

2 Protonation state of His151

The histidine side chain can assume three different protonation states - positively charged (with a proton both on Nδ and Nε) or neutral with either a proton at Nε or a proton at Nδ. In order to determine the protonation state of the His151 residue, we performed MD simulations featuring a positively charged His151 (HIP) and featuring a neutral His151 with a hydrogen only on Nε (HIE)(3 trajectories per protonation state, each 100 ns long). We discounted the possibility for a neutral His151 with a proton at Nδ on the account of a short distance between Nε and the backbone oxygen of Pro125 in every crystal structure of a TDG-DNA complex, independent of substrate or crystallization conditions. This short distance is indicative of a hydrogen bond interaction, which would be lost if there is no proton at Nε.

2.1 pKa calculations

We performed pKa calculations with the cavity algorithm[6] of the Karlsberg+ program[7,8] on ten equally distanced in time MD frames from each MD trajectory. The mean pKa values per trajectory and their standard deviations (SD) are listed in Table S2.

Table S2: calculated pKa values

	HIE		**HIP**	
	mean pKa	**SD**	**mean pKa**	**SD**
traj1	5.7	1.81	9.56	1.85
traj2	6.57	1.67	10.22	0.98
traj3	4.88	1.9	9.99	0.42

For all trajectories, featuring a positively charged His151, the calculated pKa would suggest that His151 is doubly protonated, while for the trajectories featuring a neutral His151, a neutral protonation state is more probable. The high variation of the pKa values can partly be attributed to the proximity of the His150 residue, as different orientations of

its side chain affect significantly the pKa of His151.

2.2 Hydrogen bond occupancies

As the pKa calculations do not offer a definitive answer regarding the protonation state of His151, further investigations seemed necessary. Considering that a different protonation state can lead to a reorganization of the hydrogen bonding network, we decided to take a closer look at the occupancies of the hydrogen bonds for the HIP and HIE MD simulations. This comparison can elucidate which protonation state represents the conformation seen in the crystal structure.

A close contact between Nε of His151 and the backbone oxygen of Pro125, and between the amide group of Asn140 and the side chain alcohol of Thr197 is present in every crystal structure of the TDG-DNA complex, independent of the substrate. For fC, there is also a short distance between the backbone nitrogen of Tyr152 and the oxygen of the formyl group. The occupancies of these hydrogen bonds for both protonation states were calculated with the program cpptraj[9] and are presented in Table S3.

Table S3: Occupancies of the hydrogen bonds in %

		Neutral His151			**Protonated His151**		
Acceptor	**Donor**	traj1	traj2	traj3	traj1	traj2	traj3
Pro125 - O backbone	His151 - Nε	95.35	95.86	93.43	36.71	61.12	27.94
Thr197 - Oγ	Asn 140 - Nδ	99.74	99.78	80.47	68.88	99.75	99.53
fC - formyl O	Tyr152 - N backbone	99.89	99.87	99.54	93.79	94.44	44.49

For a neutral His151 residue, in all three trajectories the occupancy of these hydrogen bonds is more than 80%. This is not the case for the protonated form of this residue, where the occupancy of some of the hydrogen bonds is significantly lower. Most prominent is the loss of the interaction between the Pro125 backbone oxygen and Nε of His151. This analysis indicates that MD simulations featuring a neutral His151 tend to represent best the reactive conformation seen in the crystal structure.

3 Choice of a structure for the QM/MM simulations

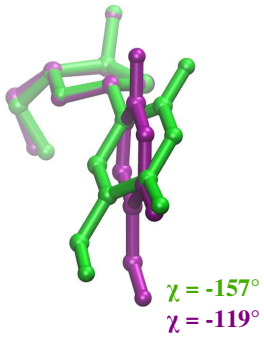


Figure S2: Superposition of the structures of fC in the TDG-DNA complex (green, PDB code: 5t2w) and in free DNA (violet, PDB code: 4qc7). χ is the dihedral angle between the O4', C1', N1, and C2 atoms. In the TDG-DNA complex the base is rotated about 40° around the glycosidic bond.

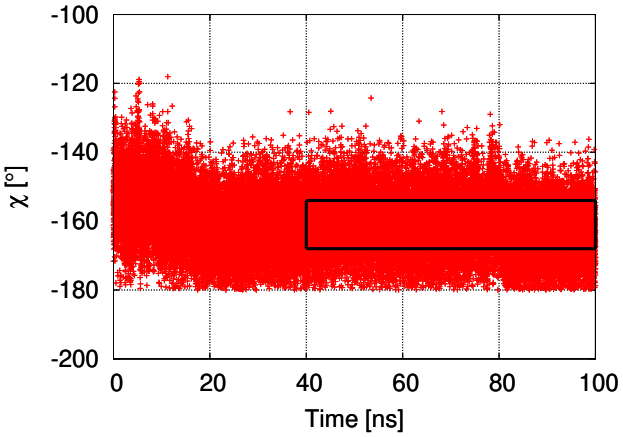


Figure S3: Values for the torsion angle χ over the course of an MD simulation. Structures of the TDG-DNA complex were grouped according to their χ value. Starting structures for the QM/MM calculations were selected from a group with χ between -153° and -170° and after at least 40 ns of simulation time (black box).

4 Substrate reorganization

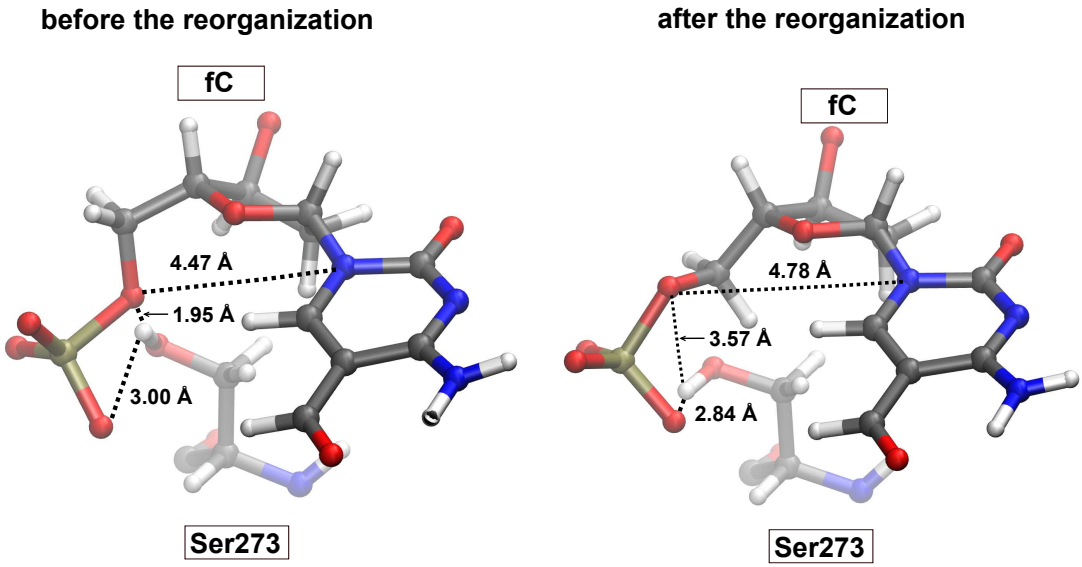


Figure S4: A comparison of the TDG active site prior and after the re-organization. Upon the rotation around γ, the distance between the C1' and the N1 atom of fC increases and the hydrogen bond interaction between Ser273 and the O5' of fC is lost.

5 Second water molecule in the TDG active site

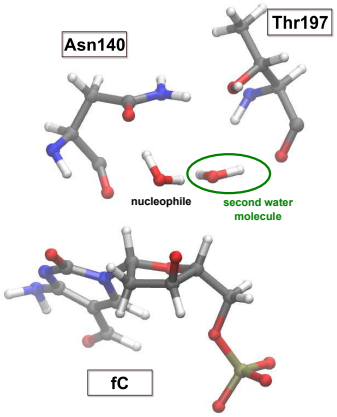
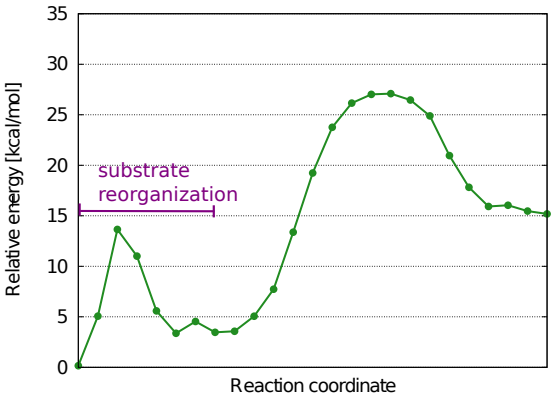
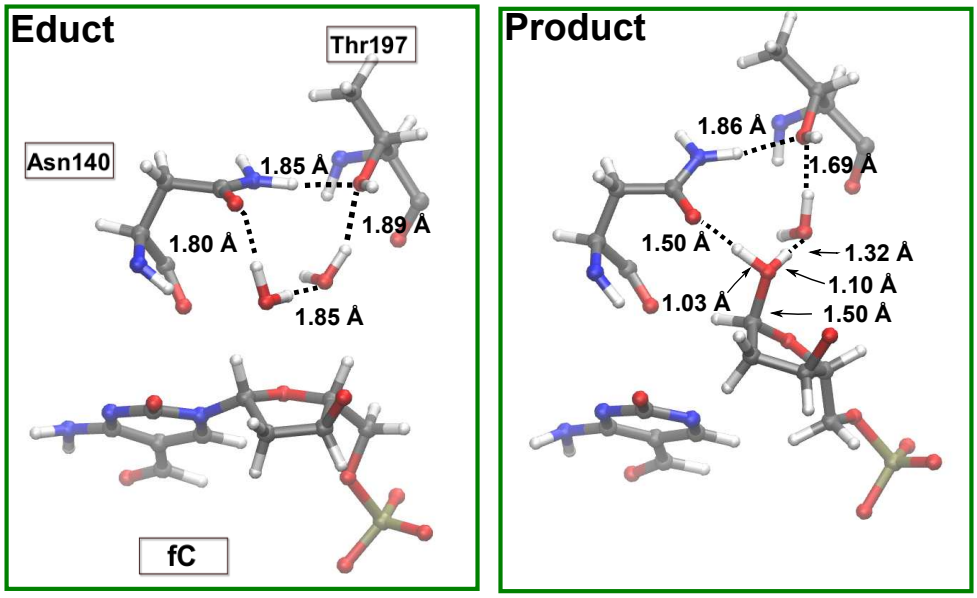


Figure S5: MD frame, featuring a second water molecule in the TDG active site



(a)



(b)

Figure S6: (a) Adiabatic mapping path for the substrate reorganization with a subsequent glycosidic bond cleavage for a reaction with a second water molecule in the TDG active site (reaction coordinate r(C1'-N1) - r (Onuc - C1')). (b) Structures of the reaction educt and product. The glycosidic bond hydrolysis results in a negatively charged fC anion and a positive charge delocalized on the abasic site, Asn140, Thr197, and the second water molecule.

6 Stabilization of the anionic base

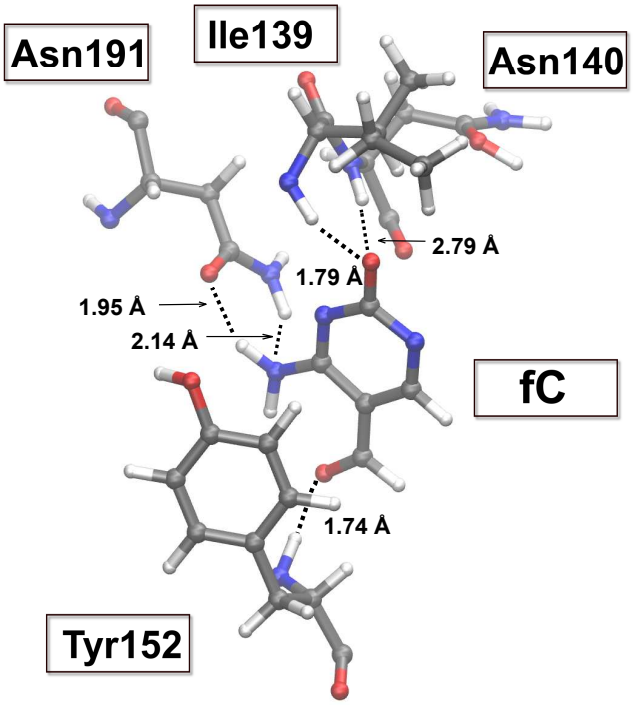


Figure S7: Structure of the product of the glycosidic bond cleavage. The negative charge on fC is stabilized by hydrogen bonds between O2 and the backbone hydrogens of Ile139 and Asn140, and between the formyl oxygen and the backbone hydrogen of Tyr152. Over the course of the glycosidic bond cleavage, the distance between the amino hydrogen of Asn191 and the N4 atom of fC decreases from 2.50 Å to 2.14 Å, while the distance between the carbonyl oxygen of Asn191 and the amino hydrogen remains the same. A stronger hydrogen bond between Asn191 and the N4 atom offers some stabilization for the product. However, mutations of Asn191 do not compromise the enzymatic activity,[10] so that the effect of this interaction is either quite small or its loss is compensated by the absence of the carbonyl group in the vicinity of the negatively charged product.

References

(1) Case, D.; Cerutti, D. S.; Cheatham, T.; Darden, T.; Duke, R.; Giese, T. J.; Gohlke, H.; Götz, A.; Greene, D.; Homeyer, N. et al. Amber 2016, University of California, San Francisco. 2017.

(2) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(3) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

(4) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(5) Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.

(6) Meyer, T.; Kieseritzky, G.; Knapp, E.-W. Electrostatic pKa computations in proteins: Role of internal cavities. *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 3320–3332.

(7) Kieseritzky, G.; Knapp, E.-W. Optimizing pKA computation in proteins with pH adapted conformations. *Proteins Struct. Funct. Bioinforma.* **2007**, *71*, 1335–1348.

(8) Rabenstein, B.; Knapp, E. W. Calculated pH-dependent population and protonation of carbon-monoxy-myoglobin conformers. *Biophys. J.* **2001**, *80*, 1141–1150.

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

(10) Maiti, A.; Michelson, A. Z.; Armwood, C. J.; Lee, J. K.; Drohat, A. C. Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J. Am. Chem. Soc.* **2013**, *135*, 15813–15822.

3.3 Manuscript III: Development of hetero-triaryls as a new chemotype for subtype-selective and potent SIRT5 inhibition

C. Glas, E. Naydenova, S. Lechner, N. Wössner, L. Yang, J. Dietschreit, H. Sun, M. Jung, C. Ochsenfeld, B. Küster, F. Bracher

"Development of hetero-triaryls as a new chemotype for subtype-selective and potent SIRT5 inhibition"

in preparation

Based on the recently identified Sirt5 inhibitor balsalazide, a launched drug with negligible bioavailability after oral administration, a as lead structure and early insights into its structure-activity relationships, we developed heteroaryl-triaryls as a novel chemotype of drug-like, potent and subtype-selective Sirt5 inhibitors. The unfavourable azo group of the lead structure was modified in a systematic and comprehensive manner, ending up with a few open-chain and, most importantly, five-membered heteroaromatic substitutes (isoxazole **CG_209**, triazole **CG_220**, pyrazole **CG_232**) with very encouraging in vitro activities (IC_{50} on Sirt5 in the low micromolar range, <10 μM). These advanced inhibitors were free of cytotoxicity and showed favourable pharmacokinetic properties, as confirmed by permeability into mitochondria using live cell imaging experiments. Furthermore, results from calculations of the relative free binding affinities of the analogues compared to balsalazide as reference compound agreed well with the trends for inhibitory activities obtained in the in vitro experiments. So, this method can be used to predict the affinity of closely related future potential Sirt5 inhibitors. Finally, Sirt5 was confirmed as target of balsalazide and one of its improved analogues in chemical proteomics experiments. Only two additional enzymes, glutaryl-CoA dehydrogenase (GCD) and nucleoside diphosphate kinase (NDPK), were found as off-targets in this investigation, once again confirming the high selectivity of the novel balsalazide-derived Sirt5 inhibitors. In summary, a combination of targeted chemical synthesis synthetic, biological work and computational studies led to a new generation of tailored Sirt5 inhibitors which represent valuable chemical tools for the investigation of the physiological role of Sirt5, but could also serve as advanced lead structures for drug candidates for systemic use.

Development of hetero-triaryls as a new chemotype for subtype-selective and potent Sirt5 inhibition

Carina Glas[1], Eli Naydenova[2], Severin Lechner[3], Nathalie Wössner[4], Liu Yang[5], Johannes Dietschreit[3], Hongyan Sun[5], Manfred Jung[4], Bernhard Küster[3], Christian Ochsenfeld[2], Franz Bracher[1*]

1: Department of Pharmacy – Center for Drug Research, Ludwig-Maximilians University, Butenandtstraße 5-13, 81377 Munich, Germany

2: Chair of Theoretical Chemistry – Department of Chemistry, Ludwig-Maximilians University, Butenandtstraße 7, 81377 Munich, Germany

3: Chair of Proteomics and Bioanalytics, Technical University of Munich, Emil-Erlenmeyer-Forum 5, 85354 Freising, Germany

4: Institute of Pharmaceutical Sciences, Albert-Ludwigs-University Freiburg, Albertstraße 25, 79104 Freiburg im Breisgau, Germany

5: Department of Chemistry and COSDAF (Centre of Super-Diamond and Advanced Films), City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong, China

*Correspondence to F. Bracher (franz.bracher@cup.lmu.de)

Keywords: balsalazide, sirt5, sirtuins, structure-activity relationships, inhibitor, histone deacylases, proteomics

Abstract

Based on the recently identified Sirt5 inhibitor balsalazide, a launched drug with negligible bioavailability after oral administration, a as lead structure and early insights into its structure-activity relationships, we developed heteroaryl-triaryls as a novel chemotype of drug-like, potent and subtype-selective Sirt5 inhibitors. The unfavourable azo group of the lead structure was modified in a systematic and comprehensive manner, ending up with a few open-chain and, most importantly, five-membered heteroaromatic substitutes (isoxazole **CG_209**, triazole **CG_220**, pyrazole **CG_232**) with very encouraging *in vitro* activities (IC_{50} on Sirt5 in the low micromolar range, <10 µM). These advanced inhibitors were free of cytotoxicity and showed favourable pharmacokinetic properties, as confirmed by permeability into mitochondria using live cell imaging experiments.

Furthermore, results from calculations of the relative free binding affinities of the analogues compared to balsalazide as reference compound agreed well with the trends for inhibitory activities obtained in the *in vitro* experiments. So, this method can be used to predict the affinity of closely related future potential Sirt5 inhibitors.

Finally, Sirt5 was confirmed as target of balsalazide and one of its improved analogues in chemical proteomics experiments. Only two additional enzymes, glutaryl-CoA dehydrogenase (GCD) and nucleoside diphosphate kinase (NDPK), were found as off-targets in this investigation, once again confirming the high selectivity of the novel balsalazide-derived Sirt5 inhibitors.

In summary, a combination of targeted chemical synthesis synthetic, biological work and computational studies led to a new generation of tailored Sirt5 inhibitors which represent valuable chemical tools for the investigation of the physiological role of Sirt5, but could also serve as advanced lead structures for drug candidates for systemic use.

INTRODUCTION

Lysine deacylases (KDACs) are associated with the post-translational modification of bioreversible ε-*N*-acylations of protein lysine residues involved in various cellular processes. KDACs thereby gain increasing interest as enzymatic targets in drug discovery. They are divided into Zn^{2+}-dependent classes I, II, and IV, as well as NAD^{+}-dependent class III, also named sirtuins.[1] On the basis of our previous work on the development[2-3] and biological characterisation[4-7] of sirtuin inhibitors, we have put our focus on the identification of small molecule inhibitors of the isoform Sirt5.[8] Sirt5 is one of seven human sirtuin isoforms. These share a highly conserved catalytic core of about 275 amino acids and possess variable C- and N-terminal domains, but exhibit different cellular localisation and function.[9-10] Sirt5 is primarily localised in the mitochondria and its major enzymatic activity *in vivo* and *in vitro* lies in the removal of acyl residues that are derived from dicarboxylic acids such as succinyl or malonyl.[11-13] Concerning its physiological role, Sirt5 is known to have regulatory effects on several biochemical pathways associated with cell metabolism.[10] Targets of Sirt5 include enzymes and proteins involved in glycolysis (e.g., through demalonylation of glyceraldehyde 3-phosphate dehydrogenase (GAPDH)),[14] the tricarboxylic acid cycle (e.g., by desuccinylation of the pyruvate dehydrogenase complex (PDC) and the succinate dehydrogenase (SDH))[11], and reducing reactive oxygen species (through deacylation of proteins such as SOD1,[15] IDH2, and G6PD)[16-17]. Sirt5 further regulates fatty acid metabolism,[11, 14, 18] ammonia detoxification,[19] and ketone body formation.[20-21] Dysregulation of Sirt5 is associated with a number of diseases such as metabolic disorders,[22-23] cancer,[24-25] and neurological dysfunctions.[26-27]

Potent and in particular simultaneously highly selective small molecule inhibitors of Sirt5 are still needed as chemical tools for further investigation of the physiological role of Sirt5 and the clinical significance of its modulation.

In our previous publication on Sirt5 inhibitors[8] we performed first investigations of the structure-activity relationships of balsalazide as a potent and subtype-selective Sirt5 inhibitor. Balsalazide, an approved drug, was initially identified as Sirt5 inhibitor by Guetschow *et al.* in 2016[28] in a high-throughput screening. With its terminal negatively charged carboxylate group balsalazide makes use of an interaction with the unique side chains of Arg105 and Tyr102[8] in the substrate binding site of Sirt5 that provide substrate specificity.[13] We designed and synthesised a first set of 13 analogues, where single functional groups of the lead structure were either deleted or minimally altered, except the core azo group. We could reveal that no structural feature of the salicylic acid motif or the *β*-alanine-derived side chain of balsalazide should be lightheadedly changed in order to retain Sirt5 inhibitory activity.[8] However, balsalazide, a prodrug of 5-aminosalicylic acid for the local treatment of inflammatory bowel disease, can be used as a chemical tool, but is not a promising candidate for application as a drug targeting Sirt5, due to its poor solubility, its minimal absorption after oral administration, and its liability for enzymatic degradation by gut bacteria under reductive cleavage of the azo moiety.[29] Therefore, we extended our SAR study with the intention to systematically replace functional groups, and in this set of analogues, with special focus on the critical azo group.

Activity-guided stepwise modification of the core azo group should lead to an increase in inhibitory activity against Sirt5 while maintaining subtype selectivity. Most importantly, this optimization should lead to a drug candidate with favorable physicochemical properties for future systemic use.

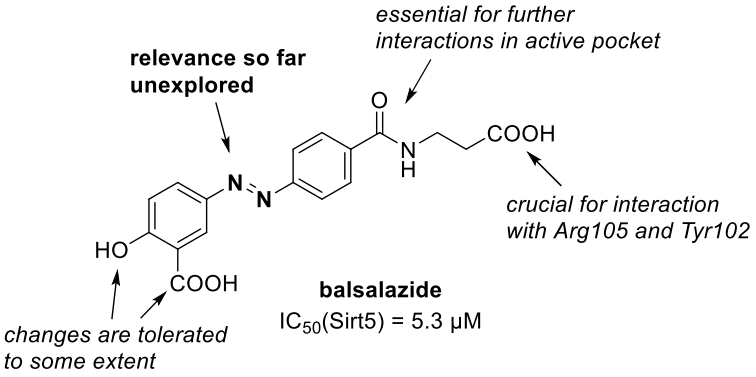


Figure 1: Structure of lead compound balsalazide with IC_{50} determined in our previous investigation.[8]

RESULTS AND DISCUSSION

1. Chemistry

Before replacing the core azo group with a variety of different spacers, we exchanged putatively essential functional groups of balsalazide with bioisosteres (Figure 3a and Figure 3, top left), generating tetrazole **CG_238**, primary salicylamide **CG_254**, and fluoro compound **CG_267**. In addition, one analogue **CG_190** was synthesised,

where an ethylenediamine linker was added to the salicylate subunit for immobilisation on beads in a chemical proteomics experiment (see Section 3 “Chemical Proteomics”). The syntheses of these analogues were in general performed following our previously published synthesis of balsalazide[8] with an azo coupling of an *in-situ* prepared arene diazonium salt and an electron-rich arene as the central step (**CG_238**, **CG_254**, and **CG_190**) or a Mills reaction of an aniline and a nitrosoarene building block (**CG_267**). A representative synthesis for this set of analogues is depicted in Figure 2a for tetrazole **CG_238**.

Next, the azo spacer was replaced by open-chained substitutes, where the number of heteroatoms was reduced step by step: imine **CG_71**, amines **CG_79** and **CG_112**, ether **CG_129**, stilbene **CG_111**, ethylene analogue **CG_142**, and alkyne **CG_219** (Figure 3, bottom). An exemplary synthesis for these inhibitors is shown for alkyne **CG_219** in Figure 2b. On the contrary, several analogues were designed and synthesised containing more heteroatoms as hydrogen bond acceptors and donors resulting in isomeric amides **CG_128** and **CG_133**, and the isomeric sulfonamides **CG_137** and **CG_140** (Figure 3, bottom). Here, the preparation of sulfonamide **CG_137** was chosen as synthetic example (Figure 2c) because this inhibitor showed very promising results in a first testing campaign for inhibitory activity against Sirt5 (see Figure 4). Based on these results structurally related analogues of **CG_137** were designed, where the sulfonamide group was *N*-methylated (**CG_168**, Figure 3, bottom) or the upper non-salicylate phenyl ring was replaced by aromatic 5- and 6-membered heterocycles in pyridine **CG_163**, and isomeric thiophenes **CG_169**, **CG_176**, **CG_177** (Figure 3, top right). Rigidisation of the lead structure was accomplished with *N*-aryl-1,2,3,4-tetrahydroisoquinoline **CG_224**[30] and quinoline **CG_268** (Figure 3, right). The synthesis of the latter was accomplished *via* Suzuki cross-coupling of an *in-situ* prepared salicylate-derived pinacolborane, as shown in Figure 2d. Lastly, the core azo group was replaced by 5-membered heteroaromatics: 1,2,4-oxadiazole **CG_264**, isoxazole **CG_209**, triazole **CG_220**, and pyrazole **CG_232**. These hetero-triaryls can be prepared by means of Huisgen 1,3-dipolar cycloadditions, as exemplified for triazole **CG_220** in Figure 2e. All synthetic procedures are described in detail in the Supporting Information (Part A: Syntheses).

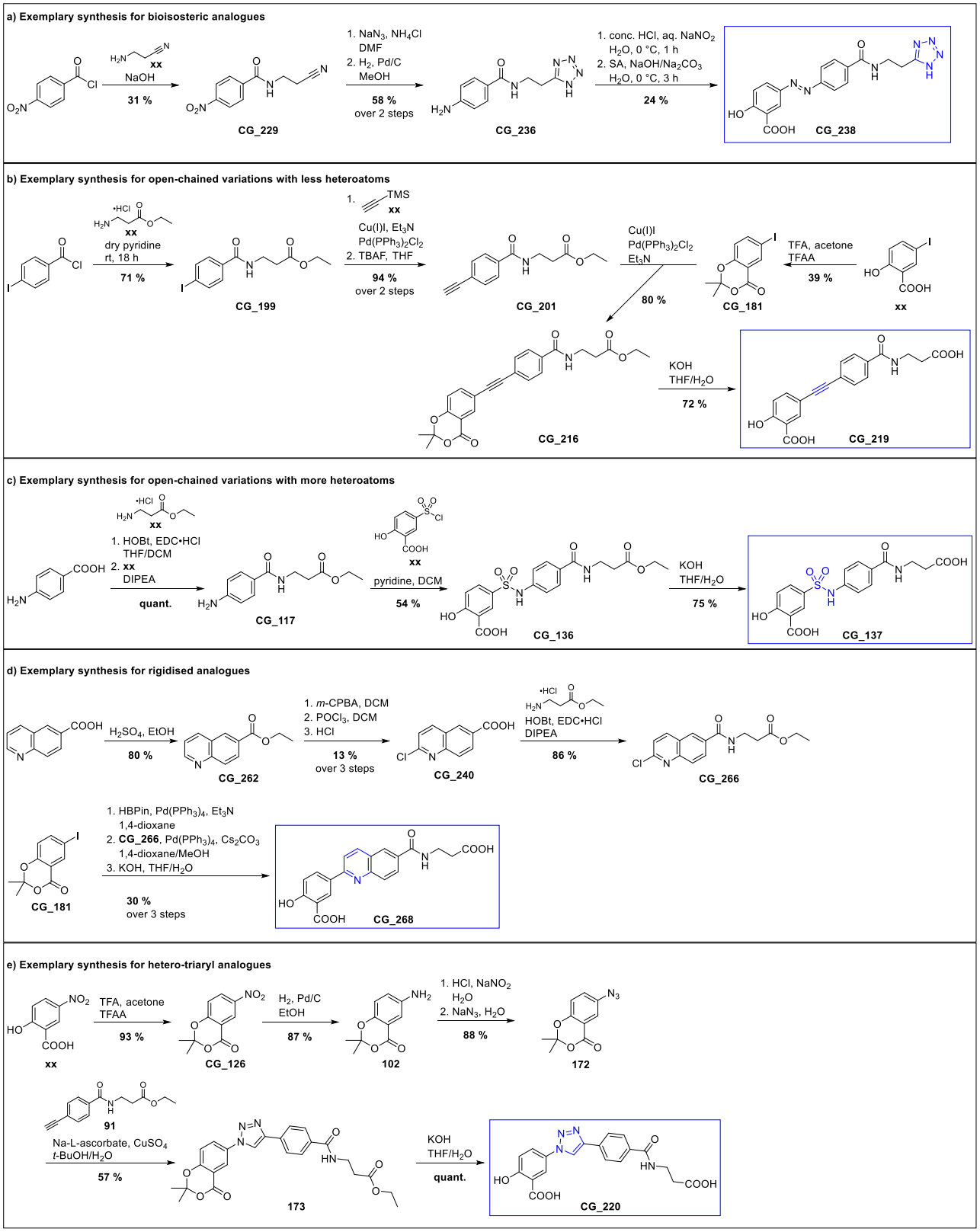


Figure 2: Exemplary syntheses for ***a****) a bioisosteric analogue with tetrazole* ***CG_238****,* ***b****) open-chained variations with less heteroatoms with alkyne* ***CG_219****,* ***c****) open-chained variation with more heteroatoms with sulfonamide* ***CG_137****,* ***d****) rigidised analogues with quinoline* ***CG_268****, and* ***e****) hetero-triaryls with triazole* ***CG_220****.*

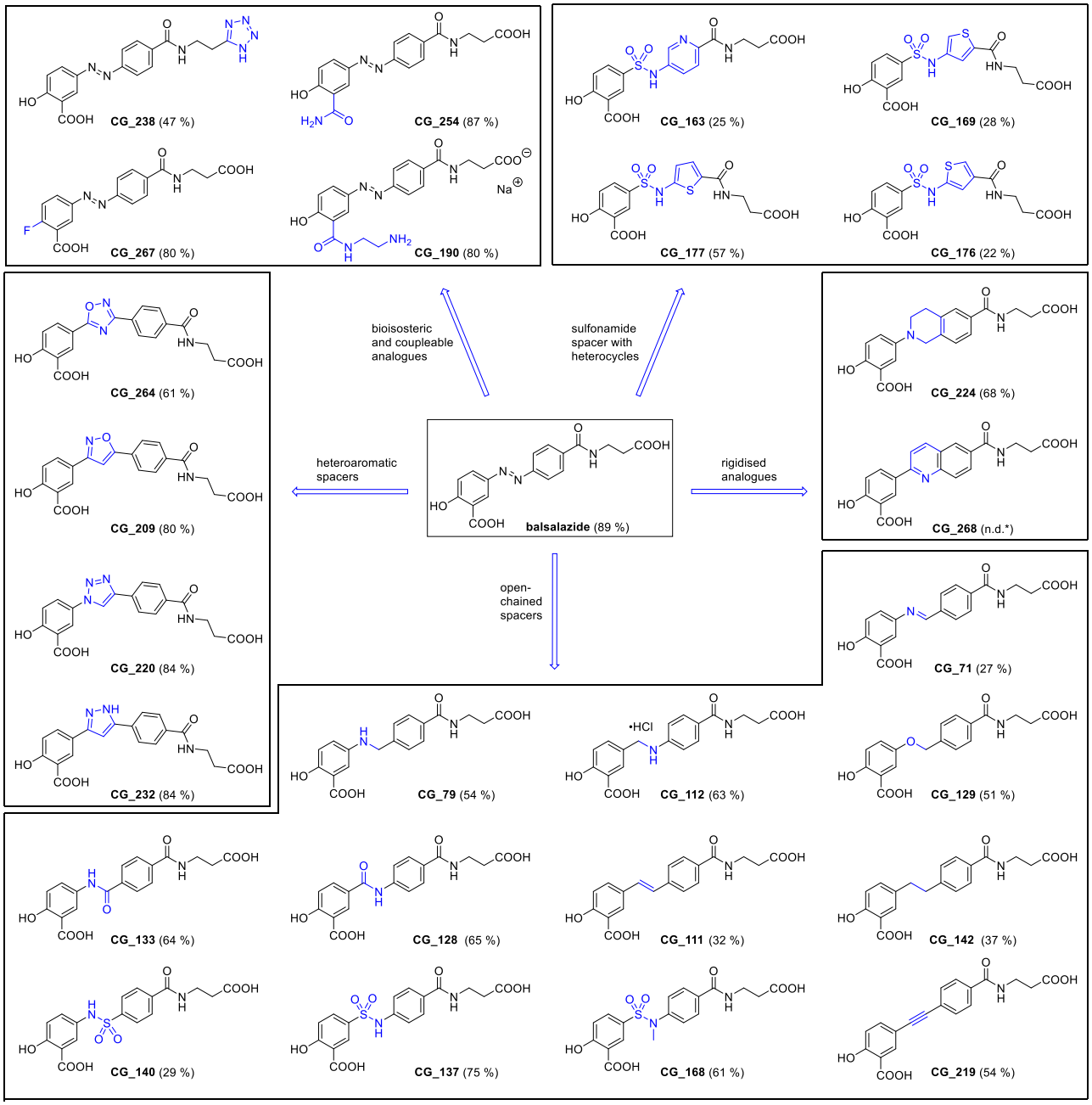


Figure 3: All 26 synthesised analogues of balsalazide including Sirt5 inhibition in % at 50 µM final assay concentration provided for each structure in parentheses.

2. Sirt5-inhibitory activity

For determination of the Sirt5-inhibitory activity of the synthesised balsalazide analogues we used the enzyme-based FLUOR DE LYS® Sirt5 fluorometric drug discovery assay kit according to manufacturer's instructions. All compounds, including balsalazide as reference, the known Sirt5 inhibitors nicotinamide and suramin, and protected precursors **CG-94/CG-103**, were tested at a final concentration of 50 µM in the initial screening (Figure 4).

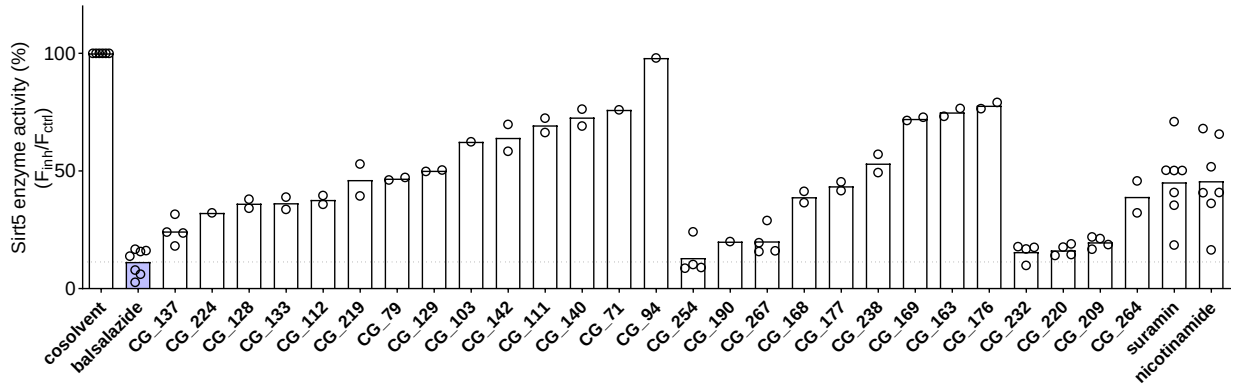


*Figure 4: Residual Sirt5 enzyme activity after 1 h incubation with the inhibitors in % (F_{inh}/F_{ctrl}). Balsalazide was included as reference and the known non-selective Sirt5 inhibitors nicotinamide and suramin as positive controls. Fluorescence measured upon incubation with the cosolvent DMSO alone (no inhibitor) was set to 100% enzyme activity. All inhibitors were tested at 50 µM final concentration. F = fluorescent units; λ_{ex} = 360 nm, λ_{em} = 460 nm; number of biological replicates stated with circles (◯) for each bar. Quinoline derivative **CG_268** could not be included due to auto-fluorescence. Screening was performed in three consecutive campaigns (**CG_137** – **CG_94**, then **CG_254** – **CG_176**, then **CG_232** – **CG_264**).*

In comparison to the lead structure balsalazide (89 % Sirt5 inhibition at 50 µM), the results shown in Figure 4 reveal that introduction of potential bioisosteric groups (Figure 3, top left) resulted in similar inhibitory activity against Sirt5 for primary amide **CG_254** and fluoro analogue **CG_267**, but a significant decrease in inhibitory activity for tetrazole **CG_238**. This leads to the conclusion that especially the distal carboxylate group of the *β*-alanine derived side chain cannot be replaced. Fortunately, ethylenediamine derivative **CG_190**, obtained by amidation of the salicylate subunit (and used for chemical proteomics experiments later), showed similar potency as balsalazide.

Among the analogues with focus on the core azo group, the set of open-chained variations showed only one analogue to be within the range of potency of balsalazide: sulfonamide **CG_137** (75% Sirt5 inhibition at 50 µM). Compared to these promising results for **CG_137**, sulfonamide **CG_140** (sulfonamide group inverted) showed a major drop in activity (29% inhibition at 50 µM). The pyridine and thiophene analogues derived from sulfonamide **CG_137** (Figure 3, top right) showed a significant loss in potency.

By rigidising the core of balsalazide, we obtained two analogues (**CG_224** and **CG_268**), of which only tetrahydroisoquinoline **CG_224** could be tested in this assay setup due to auto-fluorescence of quinolone derivative **CG_268**. It should be noted that despite of the good inhibition results for **CG_224** with 68% at 50 µM, this compound is not suitable for further experiments regarding the physiological role of Sirt5 because it was found to be unstable upon exposure to air.

Replacing the core azo group with 5-membered heterocycles finally gave way to comparable if not better inhibitors of Sirt5 than balsalazide. Only 1,2,4-oxadiazole **CG_264** showed a decrease in inhibitory activity. Pyrazole **CG_232**, triazole **CG_220** and isoxazole **CG_209** inhibited Sirt5 by 80 to 84% at 50 µM.

After this first screening for further identification of SAR, the compounds, which showed comparable inhibitory activity as the lead structure balsalazide (sulfonamide **CG_137**, isoxazole **CG_209**, triazole **CG_220**, tetrahydroisoquinoline **CG_224**, pyrazole **CG_232**, primary amide **CG_254**, and fluoro compound **CG_267**), as well as isomeric sulfonamide **CG_140** for comparison with **CG_137,** were subjected to testing for subtype selectivity. The compounds were tested for inhibition of Sirt1, 2, and 3 at 50 µM final assay concentration using an acetylated synthetic substrate[31] to identify their subtype selectivity over other class I sirtuins (Table 1). In addition, for the most promising inhibitors regarding preliminary Sirt5 inhibition results at 50 µM and subtype selectivity, further IC_{50} values for Sirt5 inhibition were determined (Table 1).

*Table 1: Summarised enzyme inhibition of selected compounds against Sirt1, 2, 3 and 5 at 50 µM, as well as IC_{50} values against Sirt5 for the most promising inhibitors. *n.i. = less than 10 % inhibition, [a] = assay system developed by Heltweg et al.[32], [b] = FLUOR DE LYS® SIRT5 fluorometric drug discovery assay kit, # = results from our previous publication[8].*

Compound name	IC_{50} [µM] or % inhibition at 50 µM			
	Sirt1[a]	Sirt2[a]	Sirt3[a]	Sirt5[b]
Balsalazide	n.i.*#	11 %#	n.i.#	5.3 µM#/13.8 µM
CG_137	n.i.	14 %	n.i.	12.5 µM

CG_140	n.i.	17 %	n.i.	29 %
CG_209	n.i.	n.i.	n.i.	11.5 µM
CG_220	n.i.	15 %	n.i.	7.2 µM
CG_224	14 %	16 %	n.i.	68 %
CG_232	n.i.	n.i.	n.i.	8.5 µM
CG_254	27 %	41 %	29 %	87 %
CG_267	n.i.	13 %	n.i.	23.6 µM

Overall, the results obtained for the four investigated sirtuins revealed high subtype selectivity for all tested inhibitors, except for primary amide **CG_254**. With significant inhibition values of 27% (Sirt1), 41% (Sirt2), and 29% (Sirt3) at 50 µM, this analogue was not considered for further IC_{50} value determination against Sirt5. In conclusion, these selected inhibitors showed, compared to the lead structure balsalazide, comparable inhibitory activity against Sirt5 in a first testing campaign and retained subtype selectivity over Sirt1, 2, and 3 (⩽14% Sirt1 inhibition, ⩽17% Sirt2 inhibition, <10% Sirt3 inhibition).

Determination of the IC_{50} values on Sirt5 (Table 1, Figure 5) showed that in particular the newly developed hetero-triaryl analogues of balsalazide (isoxazole **CG_209**, triazole **CG_220**, and pyrazole **CG_232**) experience high potency against Sirt5 in the low micromolar range (⩽11. 5 µM). In the scope of the present investigation, the IC_{50} value of balsalzide was again determined in the low micromolar range with 13.8 µM (previous publication: 5.3 µM[8]). Concerning both potency on Sirt5 and subtype selectivity, these novel hetero-triaryl analogues compare favourably with the lead structure balsalazide.

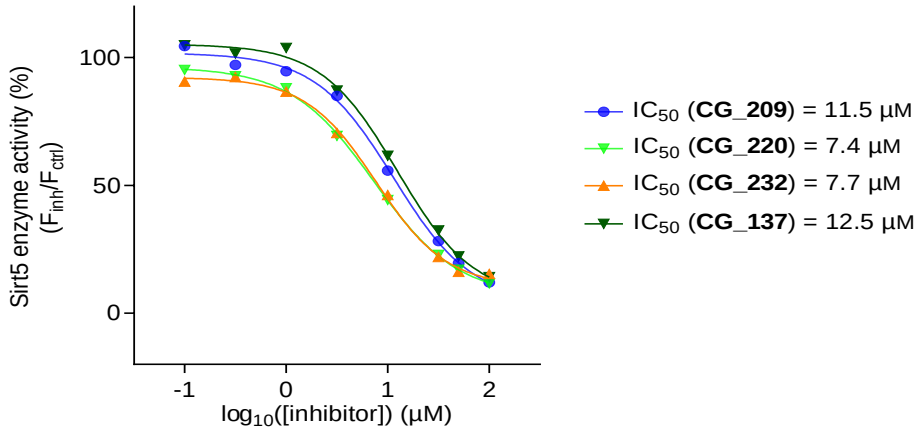


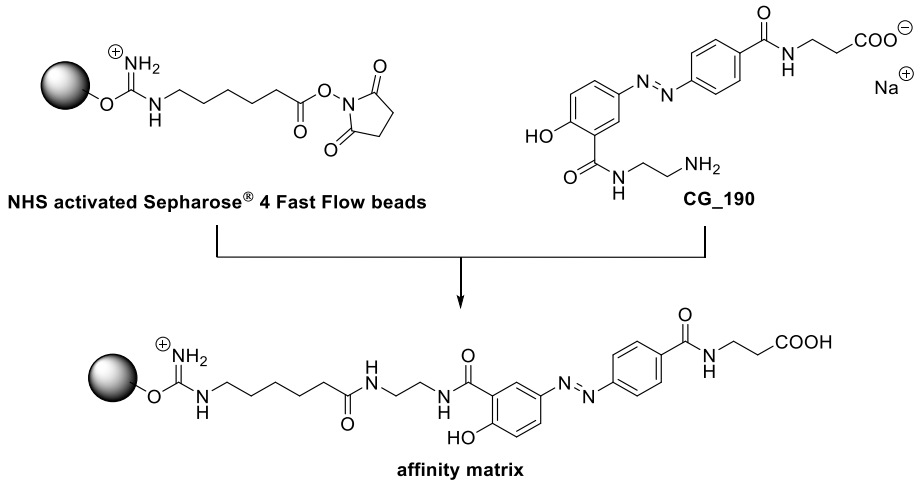
***Figure 5**: Inhibition of Sirt5 desuccinylase activity by **CG_137**, **CG_209**, **CG_220**, and **CG_232** determined using the FLUOR DE LYS® SIRT5 fluorometric drug discovery assay kit resulting in residual Sirt5 enzyme activity after 1 h incubation with the inhibitors in % (F_{inh}/F_{ctrl}). Depicted are dose-response curves and corresponding IC_{50} values for each inhibitor. F = fluorescence units; λ_{ex} = 360 nm, λ_{em} = 460 nm.*

In addition to these results, neither balsalazide nor the equipotent analogues showed any cytotoxic or antimicrobial effects (tested against diverse bacteria and fungi) in our routine testing (see Supporting Information), further underlining their high selectivity.

3. Chemical Proteomics

Chemical proteomics experiments were used to verify Sirt5 as a target of balsalazide and new inhibitors derived therefrom, and to identify possible off-targets in a full cell lysate of MCF7 cells using an immobilised derivative of balsalazide in a pull-down assay. MCF7 cells were chosen, since sirtuins, and in particular Sirt5, are overexpressed here.[33] In addition, competition experiments with free balsalazide and the selected potent drug-like inhibitors **CG_220** and **CG_232** should give further insight into the affinity of the different compounds to Sirt5.

For this purpose, the synthesised coupleable aminoethyl amide **CG_190** (Figure 3, top left) was coupled, under formation of an amide bond, to *N*-Hydroxysuccinimidyl-Sepharose® (NHS) 4 Fast Flow beads for immobilisation as shown in Scheme 1. The ethylenediamine linker was attached at a position, which, based on evidence from previous docking experiments,[8] supposedly plays no or only a minor role in the binding of balsalazide to the protein: the carboxylate group of the salicylic acid unit.



*Scheme 1: Coupling of **CG_190** to N-Hydroxysuccinimidyl-Sepharose® (NHS) 4 Fast Flow beads.*

Based on LC-MS/MS analysis, only three proteins showed dose-dependent reduction in intensity when incubated with the competing free compounds and can thus be considered to be bound by the molecules (see Figure S1 and Figure S2). These proteins were identified as the target enzyme Sirt5 and two off-targets: glutaryl-CoA-dehydrogenase (GCD; gene: GCDH) and nucleoside diphosphate kinase (NDPK; gene: NME4).

With free balsalazide the competition assay resulted in an EC_{50} value of 8.8 µM regarding Sirt5 (see Figure S1). This is in surprisingly good accordance with the IC_{50} value derived from the enzyme inhibition assay, since the binding assays performed herein often result in 10 – 100-fold lower affinity compared to enzyme-based assays assays. In addition to Sirt5, the proteins NDPK and GCD were also competed and resulted in EC_{50} values of 45.8 µM and 17.3 µM, respectively. This leads to the conclusion that balsalazide also shows binding to these proteins to some extent, but it does not tell if these enzymes are inhibited by balsalazide. A possible explanation for the observation regarding GCD is that balsalazide also might fit into the active site of GCD. Therefore, a competition assay with glutaryl-CoA was included, which is the physiological substrate of GCD. (see Figure S1 and Figure S2). In close analogy to the binding mode of succinylated or malonylated peptidic substrates and balsalazide for Sirt5, the terminal carboxylate of balsalazide could mimic the free carboxylic acid group of the glutaryl residue.

The resulting binding curves for glutaryl-CoA (see Figure S1) strongly undermine the hypothesis that balsalazide binds into the same binding pocket of GCD because immobilised **CG_190** was displaced with increasing concentrations of glutaryl-CoA. Glutaryl-CoA does not target Sirt5 and NDPK compared to balsalazide.

Additional competition assays with triazole **CG_220** and pyrazole **CG_232** showed that the same proteins were displaced from the affinity matrix as it was the case with balsalazide: Sirt5, NDPK and GCD (Figure S2), undermining molecular modes of action identical to balsalazide for the two advanced analogues.

The inhibitors triazole **CG_220** and pyrazole **CG_232** were slightly more potent in the previously performed enzyme-based assay than balsalazide. Here within these chemical proteomics experiments, the competition with compound **CG_232** resulted in EC_{50} values of only 25.7 µM for Sirt5 and 50.2 µm for GCD. For NDPK an EC_{50} value of 14.5 µM was obtained (balsalazide: EC_{50} = 45.8 µM). For **CG_220** this experiment resulted in an even better EC_{50} value of 1.2 µm for Sirt5 and thereby suggesting that this molecule exhibits the strongest binding affinity for Sirt5 within this set of compounds. Regarding NDPK, triazole **CG_220** was also more affine with an EC_{50} value of 7.8 µM. On the contrary, for the off-target GCD an EC_{50} value of only 59.5 µM was obtained.

Overall, these chemical proteomics experiments verified Sirt5 as a target protein for balsalazide as well as for the most potent inhibitors **CG_220** and **CG_232**. In addition, the two off-targets NDPK and GCD were identified. There are small molecules known that interact with NDPK such as ellagic acid[34] (inhibitor) and NMac1[35] (activator). The relevance and possibilities of these off-target findings must be further investigated.

4. Live cell imaging

In the previous experiments, we were able to see that newly developed inhibitors derived from balsalazide are able to interact with the target enzyme Sirt5 under conditions where the enzyme is removed from its natural surroundings in the cell.

We were then interested to see, if balsalazide and the most promising inhibitor **CG_220** are able to inhibit Sirt5 activity while it is still located in the mitochondria; thus, to see if this chemotype of Sirt5 inhibitors are mitochondria-membrane permeable. Therefore, live cell imaging experiments were conducted. A previous study[36] led to a novel class of succinylated peptide precursors that can be transformed into self-assembling building blocks through Sirt5 catalysis, leading to the formation of supramolecular nanofibers *in vitro* and in living cells. This Sirt5-mediated self-assembly of fluorescent peptide precursors can be used for imaging cellular Sirt5 in an activity-based fashion. Since a cell viability assay revealed that both inhibitors show no cytotoxicity in our routine testing (see Supporting Information) live cell experiments could be performed. The experiment with live mammalian cells included preincubation with different concentrations of our inhibitors for a defined time and subsequently with peptide 2[36] (exemplary results shown for inhibitor **CG_220** in Figure 6, experimental details and results for balsalazide can be found in the Supporting Information). Significant decrease of intracellular fluorescence was observed, suggesting that both inhibitors are mitochondria-membrane permeable and Sirt5 was inhibited and therefore the formation of nanofibers was suppressed. This was in particular well observable for the hetero-triaryl **CG_220**.

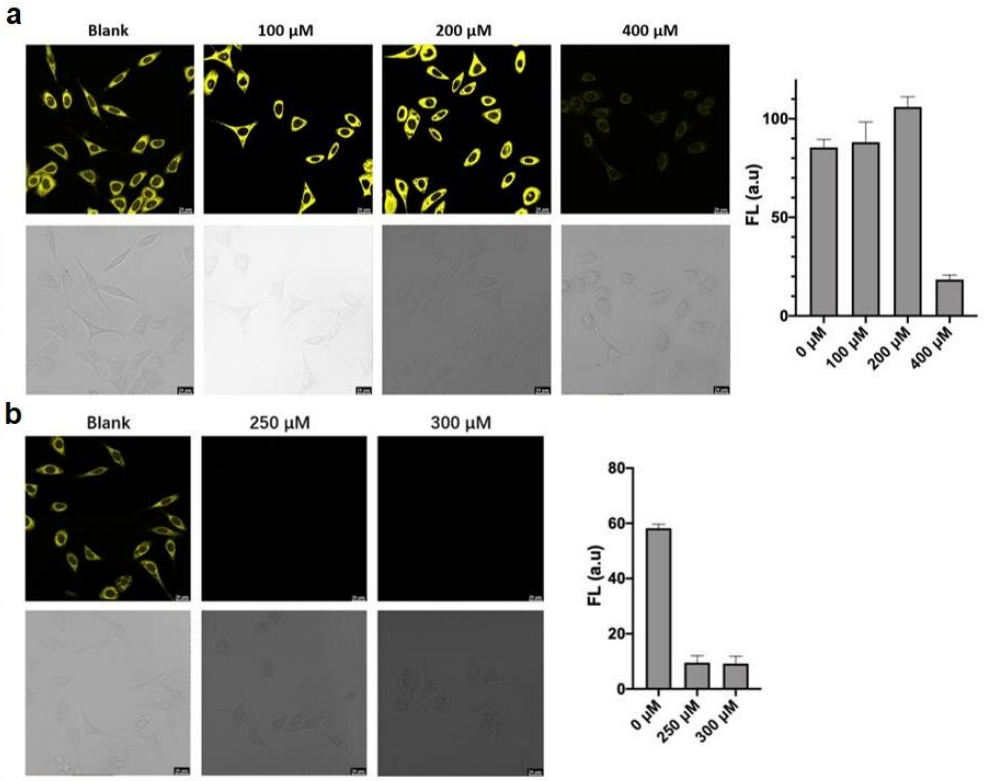


Figure 6: *Inhibition of Sirt5 ability to form nanofibers using peptide 2[36] in HeLa cells in the presence of various concentrations of* ***CG_220*** *after a) 45 min preincubation and b) 90 min preincubation in live cell imaging experiment.*

5. Relative binding free energy calculations

Our previous computational investigation of balsalazide derivatives in the Sirt5 binding site has indicated that the inhibitory strength of a compound is unrelated to its inhibition mode (covalent *vs.* non-covalent) and is likely dominated by protein-ligand interactions.[8] The binding affinity of an inhibitor can be predicted *in silico* in order to identify promising compounds prior to the synthesis or experimental measurements. For this purpose, we performed relative free binding energy calculations using thermodynamic integration and non-equilibrium MD simulations (for details on the calculations see Supporting Information). Balsalazide was selected as the lead structure and the relative binding free energies (ΔΔG) for twelve of the synthesised analogues respective to the lead structure were computed (see Figure 7). A negative ΔΔG value indicates a higher binding affinity for the respective inhibitor, compared to balsalazide. These calculations were performed only on the open-chained spacer variants except **CG_112** (see Figure 3) and additionally on **CG_267**, as introduction of rigid structures, ring breaking or a change in the charge state are known to produce significant errors in the estimation of the relative binding free energies, and were therefore avoided.[37-38]

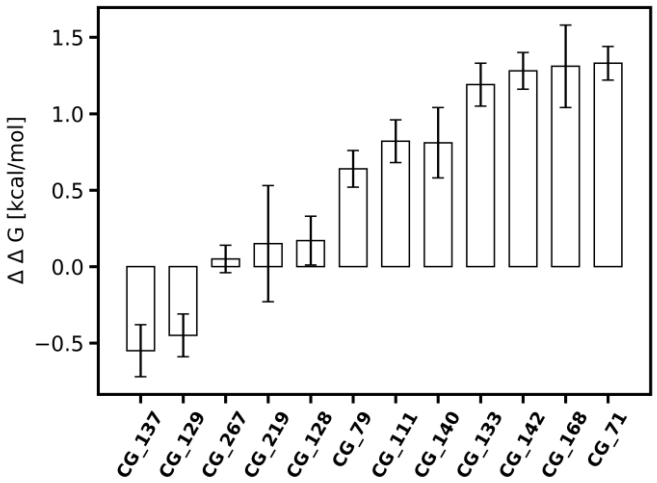


Figure 7: ΔΔG values for selected inhibitors respective to balsalazide in kcal/mol. Negative values signify a higher binding affinity to Sirt5 compared to balsalazide.

For ten of the twelve studied variants, the calculations showed a qualitative agreement with the trends in the *in-vitro* experiments for enzyme inhibition. Both the predicted binding affinity and the measured inhibitory activity of these compounds were lower compared to balsalazide (see Figure 4 and Figure 7). The simulations, however, overestimate the binding free energy of both ether **CG_129** and sulfonamide **CG_137**, as both variants were predicted to bind more tightly to Sirt5 than balsalazide.

A detailed analysis of the MD simulations identified several hydrogen bond interactions between **CG_137** and the backbone of Glu225 and Gly224 that are not observed in the simulations of balsalazide in the Sirt5 binding site (see Figure S3, Supporting Information). It is possible that our MD simulations oversample this conformation, resulting in an overestimation of the binding affinity of **CG_137** compared to balsalazide. For **CG_129** the discrepancy between the experimental results and the calculations is likely due to force field inaccuracies.

Overall, the relative binding free energy calculations presented here, show good qualitative agreement with the experiments for the majority of the studied closely related variants and can be employed for the prediction of the binding affinity of other potential Sirt5 inhibitors.

CONCLUSION

In conclusion, our previously published extensive structure-activity relationship studies revealed that balsalazide is a potent low micromolar inhibitor of Sirt5 with proven subtype selectivity over Sirt1, 2, and 3. However, balsalazide can be used as a chemical tool for the investigation of the physiological roles of Sirt5 but it is not a promising candidate for application as a drug targeting Sirt5, due to its minimal absorption from the gut, and liability for enzymatic degradation by gut bacteria under reductive cleavage of the azo moiety.[29] Therefore, we explored the relevance of balsalazide in additional biological experiments and further varied functional groups on our lead structure regarding bioisosterism with special focus on the azo spacer unit. We synthesised 26 novel analogues of balsalazide to not only increase the inhibitory activity while maintaining subtype selectivity, but also to change the structure into a drug candidate with appropriate physicochemical properties and high physiological stability for systemic use. Evaluation of these compounds revealed the most promising analogues sulfonamide **CG_137**, isoxazole **CG_209**, triazole **CG_220**, pyrazole **CG_232**, and fluoride **CG_267** as subtype-selective and in addition highly active Sirt5 inhibitors. All of the analogues with a modified spacer exhibited very good IC_{50} values in the low micromolar range ($\leqslant$12.5 μM). With triazole **CG_220** (IC_{50} = 7.2 μM) and pyrazole **CG_232** (IC_{50} = 8.5 μM) two inhibitors were developed that represent two very promising inhibitors of Sirt5 that can be used for future investigations of the physiological role of Sirt5 and its clinical relevance. In addition, because both inhibitors carry spacers (triazole and pyrazole) that are easily accessible using 1,3-dipolar cycloadditions, further optimisations can easily be performed to possibly increase affinity, and hence potency, further.

In further live cell imaging experiments, using Sirt5's ability to form supramolecular fluorescent nanofibers *via* catalysis, we were able to see a significant decrease of intracellular fluorescence, suggesting that both inhibitors are mitochondria-membrane permeable and Sirt5 was inhibited in its physiological environment. This was in particular well observable for the triazole **CG_220**.

Furthermore, computations of the relative free binding affinities of the analogues compared to balsalazide as reference compound agreed qualitatively with the trends for inhibitory activities obtained in the *in vitro* experiments. This shows the usefulness of such calculations in predicting the affinity of potential Sirt5 inhibitors. However, it has its limitations regarding structural changes of inhibitors such as rigidisation and ring breaking, so that the relative

binding affinities of the compounds with heteroaromatic spacers (that emerged as the best inhibitors designed in this study) could not be calculated at this stage.

A chemical proteomics experiment using an immobilised balsalazide derivative identified as targets of balsalazide and its analogues the desired sirtuin Sirt5. In addition, two more enzymes, namely glutaryl-CoA dehydrogenase (GCD) and nucleoside diphosphate kinase (NDPK), were found, where further studies have to give further insights into the relevance and possibilities of these findings. Competition assays with balsalazide, analogues triazole **CG_220** and pyrazole **CG_232**, and glutaryl-CoA revealed that triazole **CG_220** shows the strongest binding affinity to Sirt5 in this setup. Furthermore, balsalazide binds to GCD in the same manner as its natural substrate glutaryl-CoA. If balsalazide is thereby inhibiting GCD, the effects of the inhibition of both enzymes, GCD and Sirt5, could work synergistically. But this hypothesis has to be investigated further.

In summary, the synthetic and biological work carried out in the course of this extensive study and additional *in vitro* and computational studies revealed that changes on the problematic core azo group of balsalazide could not only increase the drug-likeness of the inhibitor but also potency, since these led with **CG_220** and **CG_232** to the development of at least two even more potent Sirt5 inhibitors with IC_{50} values in the low micromolar range (<10 µM) and proven subtype selectivity over Sirt1, 2, and 3. With these analogues subtype selectivity was maintained and inhibitory activity as well as physicochemical properties were improved. Hence, these novel inhibitors represent valuable chemical tools for the investigation of the physiological role of Sirt5, but could also serve as advanced lead structures for drug candidates for systemic use.

Abbreviated Methods:

Full and detailed experimental synthetic and biological protocols as well as description of computational methods can be found in the Supporting Information.

General procedures and representative compound synthesis, CG_220 and CG_232.

General Procedure A – Hydrogenation of nitroarenes *(The herein used general procedure follows Hofmann et al. [39])*

The appropriate nitro compound (1.00 equivalent) was dissolved to a concentration of 0.3 M in dry MeOH and Pd/C (10 wt % on activated carbon) was added under N_2 atmosphere. Hydrogenation was performed under 1 bar H_2 pressure at rt for 3 h. Subsequently, the reaction mixture was filtered through a pad of celite and the solvent was removed *in vacuo*. If necessary, the crude product was purified by flash column chromatography (FCC) using the indicated eluent.

General Procedure B – Azo coupling reaction *(The herein used general procedure follows a patent from Biorex Laboratories Ltd. from 1983 [40])*

The appropriate aniline (1.00 equivalent) was suspended in H_2O to a concentration of 0.3 M, before conc. HCl (3.40 equivalents) was added. The mixture was cooled to 0 °C and a cold 2.0 M solution of $NaNO_2$ (1.03 equivalents) in water was added dropwise. The mixture was stirred for 1 h and then added dropwise to a solution of the corresponding phenol (1.03 equivalents), dissolved in water to a concentration of 0.6 M containing NaOH (2.10 equivalents) and Na_2CO_3 (1.60 equivalents) at 0 °C. The reaction mixture was adjusted to pH 8 with 2 M aq. NaOH and stirred for 2 – 3 h (TLC monitoring), then poured into 0.5 M aq. HCl at 0 °C. The precipitate obtained was collected by filtration, washed with water and suction dried to give the corresponding azobenzene in sufficient purity, unless stated otherwise.

General Procedure C – Protection of salicylates as acetonides *(The herein used general procedure follows a patent from Applied Research Systems ARS Holding N.V. from 2005 [41])*

A mixture of the appropriate salicylic acid derivative (1.0 equivalents) in TFA (16 equivalents), TFAA (3.0 equivalents) and acetone (5.0 equivalents) was heated to 90 °C for 16 h. After cooling to room temperature, the reaction mixture was concentrated under reduced pressure. The residue was dissolved in EtOAc and washed with sat. aq. $NaHCO_3$ (2 x) and brine (2 x). The organic layer was dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC using the indicated eluent.

General Procedure D – Alkaline deprotection of esters and acetonides *(The herein used general procedure is based on Mukkamala et al. [42])*

The appropriate protected starting material (1.0 equivalents) was dissolved in THF to a concentration of 0.15 M followed by the addition of a solution of KOH (5.0 equivalents) in water (same volume as THF). The reaction mixture was stirred at the stated temperature for 1 – 3 h (TLC monitoring) and then acidified to pH 1 with 2 M aq. HCl at room temperature. It was proceeded with the indicated work-up procedure.

General Procedure E – Amide formation (I) with HATU and DIPEA *(The herein used general procedure follows a patent from Chiesi Farmaceutici S.P.A. from 2016 [43])*

The appropriate carboxylic acid (1.5 equivalents) was dissolved in DMF to a concentration of 0.3 M. Then DIPEA (3.0 equivalents) was added, followed by HATU (1.5 equivalents). The reaction mixture was stirred at room

temperature for 40 min. The appropriate amine (1.0 equivalent) was added, and the reaction mixture stirred at room temperature for 18 h. The reaction mixture was diluted with EtOAc and washed sequentially with water (2 x), 10 % aq. K_2CO_3 (1 x), brine (1 x) and 5 % aq. LiCl (1 x). The organic phase was dried using a phase separation paper and the filtrate was concentrated *in vacuo.* The crude product was then purified by FCC using the indicated eluent.

General Procedure F – Amide formation (II) with HOBt, EDC•HCl and DIPEA *(The herein used general procedure follows to a patent from Novo Nordisk A/S from 2004 [44])*

The appropriate carboxylic acid (1.0 equivalent) and 1-hydroxybenzotriazole (1.1 equivalents) were dissolved in THF to a concentration of 0.3 M (with respect to the carboxylic acid). EDC•HCl (1.1 equivalents) was added. If a semi-crystalline solid appeared, DCM (half the volume of THF) was added and a clear solution was obtained. The mixture was stirred for 10 min, then the appropriate amine (1.0 equivalent) and DIPEA (3.0 equivalents) were added. The mixture was stirred for 2 h at room temperature and then heated to reflux for 10 min. The reaction mixture was cooled to room temperature, then partitioned between water and DCM. The aqueous phase was further extracted with DCM (3 x) and the combined organic layers were dried using a phase separation paper and the solvent removed *in vacuo*. The crude product was purified by FCC using the indicated eluent.

General Procedure G – Amide formation (III) with HOBt, EDC•HCl and Et_3N *(The herein used general procedure follows Kloss et al. [45])*

The appropriate carboxylic acid/benzoic acid (1.0 equivalent), the corresponding amine (1.2 equivalents), EDC•HCl (1.2 equivalents) and HOBt (1.2 equivalents) were dissolved in dry DCM to a concentration of 0.3 – 0.4 M (with respect to the carboxylic acid) under N_2 atmosphere. Subsequently, triethylamine (1.8 equivalents) was added, and the mixture was stirred at room temperature for 24 h. The solution was diluted with ethyl acetate and washed with sat. aq. NH_4Cl (2 x) and water (1 x). The organic phase was dried using a phase separation paper and the solvent was evaporated. The crude product was then purified by FCC using the indicated eluent.

General Procedure H – Sulfonamide formation

The appropriate amine (1.0 equivalent) was dissolved in DCM to a concentration of 0.1 M. Then, pyridine (2.0 equivalents) and 5-chlorosulfonyl-2-hydroxybenzoic acid (**105**, 1.0 equivalents) were added at 0 °C under N_2 atmosphere. The mixture was stirred at 0 °C for 2 h and further on at room temperature until completion of the reaction (TLC monitoring). The solvent was removed *in vacuo* and the residue dissolved in $CHCl_3$:isopropanol (3:1) and washed with 1 M aq. HCl (3 x). The solvent was then evaporated *in vacuo* and the crude product purified by FCC using the indicated eluent.

*2,2-Dimethyl-6-nitro-4H-benzo[d][1,3]dioxin-4-one (**101**)*

Prepared according to **General Procedure C** from 5-nitrosalicylic acid (**100**, 1.00 g, 5.46 mmol). The crude product was purified by FCC (EtOAc/hexanes 10:90) and acetonide **101** (1.13 g, 5.08 mmol, 93 %) was obtained as a pale yellow solid. R_f: 0.16 (EtOAc/hexanes 10:90). m.p.: 94 °C. ^{1}H-NMR (400 MHz, $CDCl_3$): δ (ppm) = 8.88 (d, J = 2.8 Hz, 1H, 5-H), 8.43 (dd, J = 9.0, 2.8 Hz, 1H, 7-H), 7.13 (d, J = 9.1 Hz, 1H, 8-H), 1.79 (s, 6H, $C(CH_3)_2$). ^{13}C-NMR (101 MHz, $CDCl_3$): δ (ppm) = 160.4 (C-8a), 159.1 (C-4), 143.0 (C-6), 131.4 (C-7), 126.2 (C-5), 118.6 (C-8), 113.6 (C-4a), 107.9 (C-2), 26.1 ($C(CH_3)_2$). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2990, 1742, 1592, 1531, 1477, 1336, 1280, 1194, 1050, 924, 746. HR-MS (EI): m/z = $[M]^{\bullet+}$ calcd for $C_{10}H_9NO_5^{\bullet+}$: 223.0475; found: 223.0473. Purity (HPLC): 210 nm: 88 %; 254 nm: 92 % (method 3a).

*6-Amino-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one (**102**)*[46]

To a solution of nitro compound **101** (1.00 g, 4.48 mmol) in EtOH (7.5 mL) was added 10 % Pd/C (477 mg, 0.448 mmol) under N_2 atmosphere. Hydrogenation was performed under 1 bar H_2 pressure at room temperature for 12 h. Subsequently, the reaction mixture was filtered through a pad of celite, washed with EtOH (2 x 10 mL) and THF (1 x 10 mL) and concentrated *in vacuo* to give amine **102** (749 mg, 3.88 mmol, 87 %) as a yellow solid. R_f: 0.45 (EtOAc/hexanes 50:50). m.p.: 161 °C. ^{1}H-NMR (400 MHz, $CDCl_3$): δ (ppm) = 7.31 (d, J = 2.8 Hz, 1H, 5-H), 6.98 (dd, J = 8.7, 2.8 Hz, 1H, 7-H), 6.80 (d, J = 8.6 Hz, 1H, 8-H), 4.07 (s, 3H, NH_2), 1.70 (s, 6H, $C(CH_3)_2$). ^{13}C-NMR (101 MHz, $CDCl_3$): δ (ppm) = 161.6 (C-4), 149.5 (C-8a), 140.4 (C-6), 124.7 (C-7), 118.2 (C-8), 115.1 (C-5), 114.2 (C-4a), 106.5 (C-2), 25.8 ($C(CH_3)_2$).IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3469, 3371, 1711, 1494, 1324, 1275, 1198, 1048, 836. HR-MS (EI): m/z = $[M]^{\bullet+}$ calcd for $C_{10}H_{11}NO_3^{\bullet+}$: 193.0733; found: 193.0732. Purity (HPLC): 210 nm: >95 %; 254 nm: >95 % (method 3a).

*6-Azido-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one (**172**)*

Conc. HCl (0.62 mL) and a solution of $NaNO_2$ (85.7 mg, 1.24 mmol) in water (0.16 mL) was slowly added to amine **102** (200 mg, 1.04 mmol) in water (0.60 mL) at 0 °C. After 10 min, a solution of NaN_3 (1.24 mg, 1.24 mmol) in water (0.16 mL) was added dropwise and the reaction mixture stirred at 0 °C for 40 min. The mixture was then extracted with diethyl ether (3 x 25 mL). The combined organic extracts were dried using a phase separation paper and concentrated *in vacuo* to give product **172** (199 mg, 0.906 mmol, 88 %) as a yellow solid. R_f: 0.42 (EtOAc/hexanes 10:90). m.p.: 85 °C. ^{1}H-NMR (400 MHz, $CDCl_3$): δ (ppm) = 7.64 (d, J = 2.8 Hz, 1H, 5-H), 7.18 (dd, J = 8.7, 2.8 Hz, 1H, 7-H), 6.97 (d, J = 8.8 Hz, 1H, 8-H), 1.73 (s, 6H, $C(CH_3)_2$). ^{13}C-NMR (101 MHz, $CDCl_3$): δ (ppm) =160.4 (C-4), 153.1 (C-8a), 135.1 (C-6), 127.3 (C-7), 119.0 (C-5), 118.9 (C-8), 114.5 (C-4a), 106.8 (C-2),

25.8 ($C(CH_3)_2$). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2132, 1727, 1484, 1439, 1380, 1328, 1294, 1242, 1198, 1138, 1050, 979, 934, 836, 780, 749. HR-MS (EI): *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{10}H_9N_3O_3^{\bullet+}$: 219.0638; found: 219.0638. Purity (HPLC): 210 nm: >95 %; 254 nm: >95 % (method 2a).

*Ethyl 3-(4-(1-(2,2-dimethyl-4-oxo-4H-benzo[d][1,3]dioxin-6-yl)-1H-1,2,3-triazol-4-yl)benzamido)propanoate (**173**)*

An alkyne **91** stock solution (0.33 M in *tert*-BuOH, 600 µL, 0.200 mmol) and an azide **172** stock solution (0.13 M in *tert*-BuOH/DMSO 21:1, 628 µL, 0.0800 mmol) were mixed in a 15 mL round bottom flask. Then, a $CuSO_4$ stock solution (26.7 mM in water, 600 µL, 16.0 µmol) and a sodium ascorbate stock solution (0.13 M in water, 600 µL, 0.0800 mmol) were added. The reaction mixture was stirred at room temperature for 4 days. EtOAc (20 mL) was added and the organic phase was washed with 1 M aq. HCl (10 mL), sat. aq. $NaHCO_3$ (10 mL) and brine (10 mL). The organic phase was dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 60:40) to give triazole **173** (21.0 mg, 0.0452 mmol, 57 %) as a white solid. R_f: 0.29 (EtOAc/hexanes 60:40). m.p.: 215 °C. ^{1}H-NMR (400 MHz, $CDCl_3$): δ (ppm) = 8.29 (s, 1H, 5''-H), 8.26 (d, *J* = 2.6 Hz, 1H, 5'''-H), 8.16 (dd, *J* = 8.9, 2.7 Hz, 1H, 7'''-H), 7.98 (d, *J* = 8.2 Hz, 2H, 3'-H and 5'-H), 7.88 (d, *J* = 8.1 Hz, 2H, 2'-H and 6'-H), 7.20 (d, *J* = 8.9 Hz, 1H, 8'''-H), 6.93 (t, *J* = 6.1 Hz, 1H, CONH), 4.19 (q, *J* = 7.1 Hz, 2H, CH_2CH_3), 3.76 (q, *J* = 5.9 Hz, 2H, 3-H), 2.67 (t, *J* = 5.8 Hz, 2H, 2-H), 1.80 (s, 6H, $C(CH_3)_2$), 1.29 (t, *J* = 7.2 Hz, 3H, CH_2CH_3). ^{13}C-NMR (101 MHz, $CDCl_3$): δ (ppm) = 173.2 (C-1), 166.8 (CONH), 160.1 (C4'''), 156.2 (C-8a'''), 147.9 (C-4''), 134.4 (C-1'), 133.1 (C-4'), 132.1 (C-6'''), 129.1 (C-7'''), 127.9 (C-2' and C-6'), 126.1 (C-3' and C-5'), 120.9 (C-5'''), 119.4 (C-8'''), 118.4 (C-5''), 114.3 (C-4a'''), 107.5 (C-2'''), 61.0 (CH_2CH_3), 35.5 (C-3), 34.1 (C-2), 26.0 ($C(CH_3)_2$), 14.4 (CH_2CH_3). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3414, 1739, 1725, 1647, 1533, 1506, 1492, 1302, 1184, 1149, 1057, 1030, 930, 826, 768. HR-MS (ESI): *m*/*z* = $[M+H]^+$ calcd for $C_{24}H_{25}N_4O_6^+$: 465.1769; found: 465.1765. Purity (HPLC): 210 nm: >95 %; 254 nm: >95 % (method 2a).

*5-(4-(4-((2-Carboxyethyl)carbamoyl)phenyl)-1H-1,2,3-triazol-1-yl)-2-hydroxybenzoic acid (**CG_220**)*

Prepared according to **General Procedure D** from triazole **173** (30.0 mg, 0.0646 mmol) at room temperature. The solvents were evaporated *in vacuo*. To the crude product was added MeOH (10 mL) and the suspension filtered. The filtrate was concentrated *in vacuo* to give product **CG_220** (25.6 mg, 0.0646 mmol, quant.) as a white solid. R_f: 0.10 (EtOAc/hexanes+AcOH 80:20+1). m.p.: 295 °C. ^{1}H-NMR (400 MHz, $(CD_3)_2SO$): δ (ppm) = 12.22 (s, 2H, 1-COOH and 2'''-COOH), 9.41 (s, 1H, 5'-H), 8.62 (t, *J* = 5.5 Hz, 1H, CONH), 8.30 (d, *J* = 2.8 Hz, 1H, 6-H), 8.08 (dd, *J* = 9.0, 2.8 Hz, 1H, 4-H), 8.03 (d, *J* = 8.5 Hz, 2H, 2''-H and 6''-H), 7.96 (d, *J* = 8.5 Hz, 2H, 3''-H and 5''-H), 7.22 (d, *J* = 8.9 Hz, 1H, 3-H), 3.48 (td, *J* = 7.1, 5.3 Hz, 2H, 1'''-H), 2.54 (t, *J* = 7.1 Hz, 2H, 2'''-H). ^{13}C-NMR (101 MHz, $(CD_3)_2SO$): δ (ppm) = 172.9 (C-3'''), 170.8 (1-COOH), 165.7 (CONH), 160.9 (C-2), 146.5 (C-4'), 133.8 (C-4''), 132.9 (C-1''), 128.5 (C-5), 127.9 (C-3'' and C-5''), 127.3 (C-4), 125.0 (C-2'' and C-6''), 121.6 (C-6), 120.5 (C-5'), 118.7 (C-3), 114.1 (C-1), 35.6 (C-1'''), 33.8 (C-2'''). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3310, 3076, 2921, 2852, 1677, 1633, 1621, 1536, 1447, 1294, 1219, 1199, 1186, 1040, 850, 829, 767, 694. HR-MS (ESI): *m*/*z* = $[M-H]^-$ calcd for $C_{19}H_{15}N_4O_6^-$: 395.0997; found: 395.0994. Purity (HPLC): 210 nm: >95 %; 254 nm: >95 % (method 2a).

*2,2-Dimethyl-4-oxo-4H-benzo[d][1,3]dioxine-6-carbaldehyde (**50**)*[41]

Prepared according to **General Procedure C** from 5-formylsalicylic acid (**48**, 1.00 g, 6.02 mmol). The crude product was purified by FCC (EtOAc/hexanes 12:88) to give acetonide **50** (618 mg, 3.00 mmol, 50 %) as a pale yellow solid. R_f: 0.43 (EtOAc/hexanes 30:70). m.p.: 117 °C. ^{1}H-NMR (500 MHz, $CDCl_3$): δ (ppm) = 9.97 (d, *J* = 0.8 Hz, 1H, CHO), 8.47 (d, *J* = 2.1 Hz, 1H, 5-H), 8.12 (dd, *J* = 8.6, 2.1 Hz, 1H, 7-H), 7.12 (dd, *J* = 8.6, 0.8 Hz, 1H, 8-H), 1.78 (s, 6H, $C(CH_3)_2$). ^{13}C-NMR (126 MHz, $CDCl_3$): δ (ppm) = 189.7 (CHO), 160.6 (C-4), 160.0 (C-8a), 135.9 (C-7), 133.6 (C-5), 131.6 (C-6), 118.6 (C-8), 113.7 (C-4a), 107.4 (C-2), 26.1 ($C(CH_3)_2$). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 1737, 1686, 1611, 1580, 1265, 1196, 1108, 841, 779, 754, 581. HR-MS (EI): *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{11}H_{10}O_4^{\bullet+}$: 206.0574; found: 206.0577. Purity (HPLC): 210 nm: >95 %; 254 nm: >95 % (method 3a).

*(E)-N'-((2,2-Dimethyl-4-oxo-4H-benzo[d][1,3]dioxin-6-yl)methylene)-4-methylbenzenesulfonohydrazide (**179**)*

To a stirred solution of *p*-toluenesulfonyl hydrazide (**178**, 93.1 mg, 0.500 mmol) in MeOH (1.2 mL), aldehyde **50** (103 mg, 0.500 mmol) was added portionwise at room temperature. After 3 h, the reaction mixture was cooled to 0 °C and the product was collected by filtration, washed with a small quantity of cold MeOH (5.0 mL) and dried *in vacuo* to give the product **179** (153.3 mg, 0.409 mmol, 82 %) as a white solid. R_f: 0.37 (EtOAc/hexanes 35:65). m.p.: 199 °C. ^{1}H-NMR (400 MHz, $CDCl_3$): δ (ppm) = 8.02 (d, *J* = 2.2 Hz, 1H, 5'-H), 7.92 (dd, *J* = 8.7, 2.2 Hz, 1H, 7'-H), 7.89 – 7.85 (m, 2H, 2-H and 6-H), 7.75 (s, 1H, 6'-CH), 7.33 (d, *J* = 8.1 Hz, 2H, 3-H and 5-H), 6.97 (d, *J* = 8.6 Hz, 1H, 8'-H), 2.41 (s, 3H, 4-CH_3), 1.73 (s, 6H, $C(CH_3)_2$). ^{13}C-NMR (101 MHz, $CDCl_3$): δ (ppm) = 160.6 (C-4'), 157.6 (C-8a'), 145.7 (6'-CH), 144.6 (C-4), 135.3 (C-1), 134.3 (C-7'), 129.9 (C-3 and C-5), 129.5 (C-5'), 128.4 (C-6'), 128.1 (C-2 and C-6), 118.2 (C-8'), 113.6 (C-4a'), 107.1 (C-2'), 26.0 ($C(CH_3)_2$), 21.8 (4-CH_3). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3247, 1716, 1623, 1498, 1362, 1286, 1217, 1163, 1067, 1051, 966, 932, 881, 809, 667. HR-MS (ESI): *m*/*z* = $[M+H]^+$ calcd for $C_{18}H_{19}N_2O_5S^+$: 375.1009; found: 375.1007. Purity (HPLC): 210 nm: >95 %; 254 nm: >95 % (method 2a).

*Ethyl 3-(4-(3-(2,2-dimethyl-4-oxo-4H-benzo[d][1,3]dioxin-6-yl)-1H-pyrazol-5-yl)benzamido)propanoate (**177**)*

To compound **179** (53.4 mg, 0.143 mmol; 0.500 eq) in MeCN (1.2 mL), 5 M aq. NaOH (28.5 µL) was added and the mixture was stirred for 20 min. Then the mixture was added to the dipolarophile arylacetylene **91** (70.0 mg, 0.285 mmol; 1.0 eq), and the reaction mixture was stirred at 50 °C for 6 h. Then again, a mixture of **179** (0.500 eq) in MeCN and 5 M aq. NaOH (stirred for 20 min) was added and the mixture heated to 50 °C for further 16 h. After this time, again a mixture of **179** (0.500 eq) in MeCN and 5 M aq. NaOH (stirred for 20 min) was added and the mixture stirred further at 50 °C for 16 h, now containing 1.50 eq of compound **179** in total. After cooling to room temperature, EtOAc (15 mL) was added and the organic phase was washed with 1 M aq. HCl (10 mL), sat. aq. $NaHCO_3$ (10 mL) and brine (10 mL). The organic phase was dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 60:40) to give pyrazole **177** (37.0 mg, 0.0798 mmol, 28 %) as a pale yellow solid. R_f: 0.13 (EtOAc/hexanes 60:40). m.p.: 197 °C. ^{1}H-NMR (400 MHz, $(CD_3)_2SO$): due to tautomerisation NMR measured at 100 °C. δ (ppm) = 13.26 (s, 1H, pyrazole-H), 8.32 (d, J = 2.2 Hz, 1H, 5'''-H), 8.23 (t, J = 5.1 Hz, 1H, CONH), 8.13 (d, J = 8.6 Hz, 1H, 7'''-H), 7.90 (s, 4H, 2'-, 3'-, 5'- and 6'-H), 7.21 (s, 1H, 4''-H), 7.19 (d, J = 8.6 Hz, 1H, 8'''-H), 4.12 (q, J = 7.1 Hz, 2H, CH_2CH_3), 3.56 (td, J = 7.0, 5.6 Hz, 2H, 3-H), 2.61 (t, J = 7.0 Hz, 2H, 2-H), 1.75 (s, 6H, $C(CH_3)_2$), 1.21 (t, J = 7.1 Hz, 3H, CH_2CH_3). ^{13}C-NMR (126 MHz, $(CD_3)_2SO$): due to tautomerisation NMR measured at 100 °C. δ (ppm) = 170.7 (C-1), 165.6 (CONH), 159.4 (C-4'''), 154.6 (C-8a'''), 133.2 (C-7'''), 127.2 (C-3' and C-5'), 124.9 (C-5'''), 124.5 (C-2' and C-6'), 117.3 (C-8'''), 113.5 (C-6'''), 113.1 (C-4a'''), 106.1 ($C(CH_3)_2$), 99.9 (C-4''), 59.3 (CH_2CH_3), 35.2 (C-3), 33.6 (C-2), 25.0 ($C(CH_3)_2$), 13.5 (CH_2CH_3), at 100 °C C-3'' and C-5'' not visible. IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3315, 3238, 2927, 1742, 1702, 1628, 1542, 1494, 1277, 1259, 1183, 1139, 930, 796, 769. HR-MS (ESI): m/z = $[M+H]^+$ calcd for $C_{25}H_{26}N_3O_6^+$: 464.1816; found: 464.1814. Purity (HPLC): 210 nm: >95 %; 254 nm: >95 % (method 2a).

*5-(5-(4-((2-Carboxyethyl)carbamoyl)phenyl)-1H-pyrazol-3-yl)-2-hydroxybenzoic acid (**CG_232**)*

Prepared according to **General Procedure D** from compound **177** (30.0 mg, 0.0646 mmol) at room temperature. The solvents were evaporated *in vacuo*. To the crude product water (10 mL) was added, the remaining solid collected by filtration and dried *in vacuo* to give pyrazole **CG_232** (21.3 mg, 0.0539 mmol, quant.) as a beige solid. R_f: 0.10 (EtOAc/hexanes+AcOH 80:20+1). m.p.: 288 °C. ^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ (ppm) = 12.42 (s, 1H, 1-COOH or 2'''-COOH), 11.34 (s, 1H, pyrazol-NH), 8.57 (t, J = 5.5 Hz, 1H, CONH), 8.27 (d, J = 2.3 Hz, 1H, 6-H), 7.98 (dd, J = 8.6, 2.3 Hz, 1H, 4-H), 7.95 – 7.88 (m, 4H, 2''-, 3''-, 5''- and 6''-H), 7.26 (s, 1H, 4'-H), 7.07 (d, J = 8.6 Hz, 1H, 3-H), 3.48 (td, J = 7.1, 5.4 Hz, 2H, 1'''-H), 2.53 (t, J = 7.2 Hz, 2H, 2'''-H). ^{13}C-NMR (126 MHz, $(CD_3)_2SO$): δ (ppm) = 172.9 (C-3'''), 171.7 (1-COOH), 165.8 (CONH), 160.7 (C-2), 133.3 (C-1'' and C-4''), 132.5 (C-4), 127.8 (C-3'' and C-5''), 126.7 (C-6), 124.8 (C-2'' and C-6''), 122.7 (C-5), 117.8 (C-3), 113.3 (C-1), 99.8 (C-4'), 35.6 (C-1'''), 33.8 (C-2'''), C-3' and C-5' not visible. IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3391, 3141, 2923, 2503, 1722, 1666, 1598, 1568, 1546, 1489, 1301, 1181, 865, 858, 801, 770. HR-MS (ESI): m/z = $[M-H]^-$ calcd for $C_{20}H_{16}N_3O_6^-$: 394.1045; found: 394.1047. Purity (HPLC): 210 nm: >95 %; 254 nm: >95 % (method 2a).

FLUOR DE LYS® SIRT5 fluorometric drug discovery assay kit. For screening the inhibitory activity of synthesized inhibitors against recombinant human Sirt5 a commercially available Sirt5 fluorometric drug discovery kit (BML-AK513) from ENZO LIFE SCIENCES (ELS) AG (Lausen, Switzerland) was used. Fluorescence was measured on a FLUOstar Omega microplate reader (BMG Labtech, Ortenberg, Germany) (excitation wavelength: 360 nm, emission wavelength: 460 nm). Graphs were generated using GraphPad Prism 9.0 software (La Jolla, CA).

Acknowledgements

This research was supported by Deutsche Forschungsgemeinschaft with funds from SFB 1309 – 325871075 (Chemical Biology of Epigenetic Modifications) (F.B., C.O., B.K.). M.J. and N.W. thank the DFG for funding (Ju295/14-1, Si868/15-1 and **235777276**/GRK1976).

We thank Moritz Kornmayer for support in benchwork, Anna Niedrig for determination of HPLC purities and Martina Stadler for MTT and agar diffusion tests.

Author Contributions

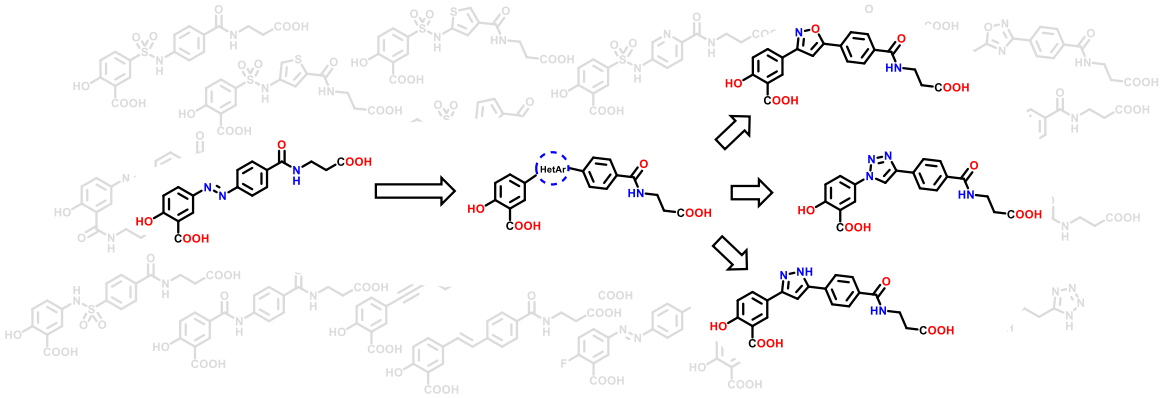
C.G. performed synthesis, enzyme assays, coordinated data assembly and wrote the manuscript. E.N and J.D. performed computational chemistry. S.L. performed chemical proteomics experiments. N.W. performed enzyme assays. L.Y. performed live cell imaging. H.S. supervised live cell imaging. M.J. supervised enzyme assays. B.K. supervised chemical proteomics experiments. C.O. supervised computational chemistry studies. F.B. designed the project, supervised synthesis, and wrote the manuscript.

Additional Information

Supplementary Information accompanies this paper.

Competing Interests: The authors declare no competing interests.

Graphical Abstract



References

[1] Schiedel, M.; Conway, S. J., *Curr. Opin. Chem. Biol.* **2018**, *45*, 166-178, doi: 10.1016/j.cbpa.2018.06.015.

[2] Huber, K.; Schemies, J.; Uciechowska, U.; Wagner, J. M.; Rumpf, T.; Lewrick, F.; Süss, R.; Sippl, W.; Jung, M.; Bracher, F., *J. Med. Chem.* **2010**, *53* (3), 1383-1386, doi: 10.1021/jm901055u.

[3] Ong, D. N.; Dittrich, S.; Swyter, S.; Jung, M.; Bracher, F., *Tetrahedron* **2017**, *73* (38), 5668-5679, doi: 10.1016/j.tet.2017.08.005.

[4] Lancelot, J.; Caby, S.; Dubois-Abdesselem, F.; Vanderstraete, M.; Trolet, J.; Oliveira, G.; Bracher, F.; Jung, M.; Pierce, R. J., *PLoS Neglected Trop. Dis.* **2013**, *7* (9), e2428, doi: 10.1371/journal.pntd.0002428.

[5] Veiga-Santos, P.; Reignault, L. C.; Huber, K.; Bracher, F.; De Souza, W.; De Carvalho, T. M. U., *Parasitology* **2014**, *141* (6), 814-825, doi: 10.1017/S0031182013001704.

[6] Verçoza, B. R. F.; Godinho, J. L. P.; de Macedo-Silva, S. T.; Huber, K.; Bracher, F.; de Souza, W.; Rodrigues, J. C. F., *Apoptosis* **2017**, *22* (9), 1169-1188, doi: 10.1007/s10495-017-1397-8.

[7] Gadelha, A. P. R.; Bravim, B.; Vidal, J.; Reignault, L. C.; Cosme, B.; Huber, K.; Bracher, F.; de Souza, W., *Int. J. Med. Microbiol.* **2019**, *309* (2), 130-142, doi: 10.1016/j.ijmm.2019.01.002.

[8] Glas, C.; Dietschreit, J. C. B.; Wössner, N.; Urban, L.; Ghazy, E.; Sippl, W.; Jung, M.; Ochsenfeld, C.; Bracher, F., *Eur. J. Med. Chem.* **2020**, *206*, 112676, doi: 10.1016/j.ejmech.2020.112676.

[9] Sanders, B. D.; Jackson, B.; Marmorstein, R., *Biochim. Biophys. Acta, Proteins Proteomics* **2010**, *1804* (8), 1604-1616, doi: 10.1016/j.bbapap.2009.09.009.

[10] Schiedel, M.; Robaa, D.; Rumpf, T.; Sippl, W.; Jung, M., *Med. Res. Rev.* **2018**, *38* (1), 147-200, doi: 10.1002/med.21436.

[11] Park, J.; Chen, Y.; Tishkoff, Daniel X.; Peng, C.; Tan, M.; Dai, L.; Xie, Z.; Zhang, Y.; Zwaans, Bernadette M. M.; Skinner, Mary E.; Lombard, David B.; Zhao, Y., *Mol Cell* **2013**, *50* (6), 919-930, doi: 10.1016/j.molcel.2013.06.001.

[12] Peng, C.; Lu, Z.; Xie, Z.; Cheng, Z.; Chen, Y.; Tan, M.; Luo, H.; Zhang, Y.; He, W.; Yang, K.; Zwaans, B. M. M.; Tishkoff, D.; Ho, L.; Lombard, D.; He, T.-C.; Dai, J.; Verdin, E.; Ye, Y.; Zhao, Y., *Mol. Cell. Proteomics* **2011**, *10* (12), M111.012658, doi: 10.1074/mcp.M111.012658.

[13] Du, J.; Zhou, Y.; Su, X.; Yu, J. J.; Khan, S.; Jiang, H.; Kim, J.; Woo, J.; Kim, J. H.; Choi, B. H.; He, B.; Chen, W.; Zhang, S.; Cerione, R. A.; Auwerx, J.; Hao, Q.; Lin, H., *Science* **2011**, *334* (6057), 806, doi: 10.1126/science.1207861.

[14] Nishida, Y.; Rardin, Matthew J.; Carrico, C.; He, W.; Sahu, Alexandria K.; Gut, P.; Najjar, R.; Fitch, M.; Hellerstein, M.; Gibson, Bradford W.; Verdin, E., *Mol Cell* **2015**, *59* (2), 321-332, doi: 10.1016/j.molcel.2015.05.022.

[15] Lin, Z.-F.; Xu, H.-B.; Wang, J.-Y.; Lin, Q.; Ruan, Z.; Liu, F.-B.; Jin, W.; Huang, H.-H.; Chen, X., *Biochem. Biophys. Res. Commun.* **2013**, *441* (1), 191-195, doi: 10.1016/j.bbrc.2013.10.033.

[16] Zhou, L.; Wang, F.; Sun, R.; Chen, X.; Zhang, M.; Xu, Q.; Wang, Y.; Wang, S.; Xiong, Y.; Guan, K.-L.; Yang, P.; Yu, H.; Ye, D., *EMBO Rep.* **2016**, *17* (6), 811-822, doi: 10.15252/embr.201541643.

[17] Xiangyun, Y.; Xiaomin, N.; linping, G.; Yunhua, X.; Ziming, L.; Yongfeng, Y.; Zhiwei, C.; Shun, L., *Oncotarget* **2016**, *8* (4), doi: 10.18632/oncotarget.14346.

[18] Zhang, Y.; Bharathi, S. S.; Rardin, M. J.; Uppala, R.; Verdin, E.; Gibson, B. W.; Goetzman, E. S., *PLoS One* **2015**, *10* (3), e0122297, doi: 10.1371/journal.pone.0122297.

[19] Nakagawa, T.; Lomb, D. J.; Haigis, M. C.; Guarente, L., *Cell* **2009**, *137* (3), 560-570, doi: 10.1016/j.cell.2009.02.026.

[20] Gertz, M.; Steegborn, C., *Cell. Mol. Life Sci.* **2016**, *73* (15), 2871-2896, doi: 10.1007/s00018-016-2180-7.

[21] Rardin, Matthew J.; He, W.; Nishida, Y.; Newman, John C.; Carrico, C.; Danielson, Steven R.; Guo, A.; Gut, P.; Sahu, Alexandria K.; Li, B.; Uppala, R.; Fitch, M.; Riiff, T.; Zhu, L.; Zhou, J.; Mulhern, D.; Stevens, Robert D.; Ilkayeva, Olga R.; Newgard, Christopher B.; Jacobson, Matthew P.; Hellerstein, M.; Goetzman, Eric S.; Gibson, Bradford W.; Verdin, E., *Cell Metab.* **2013**, *18* (6), 920-933, doi: 10.1016/j.cmet.2013.11.013.

[22] Kumar, S.; Lombard, D. B., *Antioxid. Redox Signaling* **2015**, *22* (12), 1060-1077, doi: 10.1089/ars.2014.6213.

[23] Parihar, P.; Solanki, I.; Mansuri, M. L.; Parihar, M. S., *Exp. Gerontol.* **2015**, *61*, 130-141, doi: 10.1016/j.exger.2014.12.004.

[24] Lu, W.; Zuo, Y.; Feng, Y.; Zhang, M., *Tumor Biol.* **2014**, *35* (11), 10699-10705, doi: 10.1007/s13277-014-2372-4.

[25] Lai, C.-C.; Lin, P.-M.; Lin, S.-F.; Hsu, C.-H.; Lin, H.-C.; Hu, M.-L.; Hsu, C.-M.; Yang, M.-Y., *Tumor Biol.* **2013**, *34* (3), 1847-1854, doi: 10.1007/s13277-013-0726-y.

[26] Lutz, M. I.; Milenkovic, I.; Regelsberger, G.; Kovacs, G. G., *NeuroMol. Med.* **2014**, *16* (2), 405-414, doi: 10.1007/s12017-014-8288-8.

[27] Liu, L.; Peritore, C.; Ginsberg, J.; Shih, J.; Arun, S.; Donmez, G., *Behav. Brain Res.* **2015**, *281*, 215-221, doi: 10.1016/j.bbr.2014.12.035.

[28] Guetschow, E. D.; Kumar, S.; Lombard, D. B.; Kennedy, R. T., *Anal. Bioanal. Chem.* **2016**, *408* (3), 721-31, doi: 10.1007/s00216-015-9206-0.
[29] Muijsers, R. B. R.; Goa, K. L., *Drugs* **2002**, *62* (11), 1689-1705, doi: 10.2165/00003495-200262110-00010.
[30] Glas, C.; Wirawan, R.; Bracher, F., *Synthesis* **2021**, (EFirst), A-L, doi: 10.1055/s-0040-1706002.
[31] Maurer, B.; Rumpf, T.; Scharfe, M.; Stolfa, D. A.; Schmitt, M. L.; He, W.; Verdin, E.; Sippl, W.; Jung, M., *ACS Med. Chem. Lett.* **2012**, *3* (12), 1050-3, doi: 10.1021/ml3002709.
[32] Heltweg, B.; Trapp, J.; Jung, M., *Methods* **2005**, *36* (4), 332-337, doi: 10.1016/j.ymeth.2005.03.003.
[33] Igci, M.; Kalender, M. E.; Borazan, E.; Bozgeyik, I.; Bayraktar, R.; Bozgeyik, E.; Camci, C.; Arslan, A., *Gene* **2016**, *586* (1), 123-128, doi: 10.1016/j.gene.2016.04.023.
[34] Buxton, I. L., *Proceedings of the Western Pharmacology Society* **2008**, *51*, 30-4, doi:
[35] Lee, J.-J.; Kim, H. S.; Lee, J.-S.; Park, J.; Shin, S. C.; Song, S.; Lee, E.; Choi, J.-E.; Suh, J.-W.; Lee, H.; Kim, E. E.; Seo, E. K.; Shin, D. H.; Lee, H.-Y.; Lee, H.-Y.; Lee, K.-J., *Scientific Reports* **2018**, *8* (1), 10909, doi: 10.1038/s41598-018-29101-6.
[36] Yang, L.; Peltier, R.; Zhang, M.; Song, D.; Huang, H.; Chen, G.; Chen, Y.; Zhou, F.; Hao, Q.; Bian, L.; He, M.-l.; Wang, Z.; Hu, Y.; Sun, H., *J. Am. Chem. Soc.* **2020**, *142* (42), 18150-18159, doi: 10.1021/jacs.0c08463.
[37] Gkeka, P.; Eleftheratos, S.; Kolocouris, A.; Cournia, Z., *J. Chem. Theory Comput.* **2013**, *9* (2), 1272-1281, doi: 10.1021/ct300899n.
[38] Cournia, Z.; Allen, B.; Sherman, W., *J. Chem. Inf. Model.* **2017**, *57* (12), 2911-2937, doi: 10.1021/acs.jcim.7b00564.
[39] Hofmann, D.; Gans, E.; Krüll, J.; Heinrich, M. R., *Chem. Eur. J.* **2017**, *23* (17), 4042-4045, doi: 10.1002/chem.201605359.
[40] Chan, R. P. K., US4412992A, **1983**.
[41] Swinnen, D.; Jorand-Lebrun, C.; Gerber, P.; Gonzalez, J.; Bombrun, A., WO2005097773A1, **2005**.
[42] Mukkamala, R.; Hossain, A.; Singh Aidhen, I., *Nat. Prod. Res.* **2017**, *31* (9), 1085-1090, doi: 10.1080/14786419.2016.1274891.
[43] Carzaniga, L.; Rancati, F.; Rizzi, A.; Linney, I.; Schmidt, W.; Barnes, M.; Knight, C., WO2016193241A1, **2016**.
[44] Kodra, J. T.; Behrens, C.; Madsen, P.; Joergensen, A. S.; Christensen, I. T., WO2004056763A2, **2004**.
[45] Kloss, F.; Lincke, T.; Hertweck, C., *Eur. J. Org. Chem.* **2011**, *2011* (8), 1429-1431, doi: 10.1002/ejoc.201001695.
[46] Swinnen, D.; Gerber, P.; Gonzalez, J.; Bombrun, A.; Jorand-Lebrun, C., WO2005012280A1, **2005**.

Development of hetero-triaryls as a new chemotype for subtype-selective and potent Sirt5 inhibition

Carina Glas[1], Eli Naydenova[2], Severin Lechner[3], Nathalie Wössner[4], Liu Yang[5], Johannes Dietschreit[3], Hongyan Sun[5], Manfred Jung[3], Bernhard Küster[3], Christian Ochsenfeld[2], Franz Bracher[1,*]

1: Department of Pharmacy – Center for Drug Research, Ludwig-Maximilians University, Butenandtstraße 5-13, 81377 Munich, Germany

2: Chair of Theoretical Chemistry – Department of Chemistry, Ludwig-Maximilians University, Butenandtstraße 7, 81377 Munich, Germany

3: Chair of Proteomics and Bioanalytics, Technical University of Munich, Emil-Erlenmeyer-Forum 5, 85354 Freising, Germany

4: Institute of Pharmaceutical Sciences, Albert-Ludwigs-University Freiburg, Albertstraße 25, 79104 Freiburg im Breisgau, Germany

5: Department of Chemistry and COSDAF (Centre of Super-Diamond and Advanced Films), City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong, China

ORCID-IDs: **C.G.** 0000-0002-5014-4921, **N.W.** 0000-0002-4402-2442, **J.D.** 0000-0002-5840-0002, **M.J.** 0000-0002-6361-7716, **C.O.** 0000-0002-4189-6558, **B.K.** 0000-0002-9094-1677, **F.B.** 0000-0003-0009-8629.

*Correspondence to F. Bracher (franz.bracher@cup.uni-muenchen.de)

Table of Contents

Part A: Chemical Synthesis

Conventions

Safety Hazards: no unusual safety hazards were encountered.

Solvents and reagents: All solvents were purchased from commercial sources and used without further purification. If needed, solvents were dried according to standard methods via distillation over drying agents and stored under nitrogen atmosphere with activated molecular sieves. Tetrahydrofuran (THF) was distilled over benzophenone and sodium. Standard vacuum line techniques were used and glassware was flame dried prior to use. Organic solvents were dried during workup using anhydrous Na_2SO_4, $MgSO_4$ or by filtration through a hydrophobic filter paper (MN 617 WA, Ø125 mmm, Macherey-Nagel).

Reactions, purification and chromatography: Reactions were monitored via thin layer silica gel chromatography (TLC) using polyester sheets POLYGRAM SIL G/UV_{254} coated with 0.2 mm silica gel (Macherey-Nagel). Plates were visualised using UV light (254 nm or 365 nm) or staining with $KMnO_4$, CAM (ceric ammonium molybdate), $FeCl_3$, bromocresol green, DNPH or ninhydrin reagents. Products were purified by flash column chromatography (normal-phase silica gel chromatography) using SiO_2 60 (0.040 - 0.063 mm, 230 - 400 mesh ASTM) from Merck. HPLC analytical measurements for determination of the purities of the products were carried out detecting at 210 nm and 254 nm using the following methods:

Method 1: Agilent Poroshell 120, EC-C18 2.7 µm (3.0 x 100 mm), injection volume 10 µL, temperature 50 °C

a) ACN/water/THF 40:59.9:0.1; flow 1.0 mL/min
b) ACN/phosphate buffer pH 3.0 70:30; flow 0.8 mL/min
c) ACN/phosphate buffer pH 3.0 30:70; flow 1.0 mL/min
d) ACN/phosphate buffer pH 3.0 50:50; flow 1.0 mL/min

Method 2: Zorbax Eclipse Plus, C18 5.0 µm (4.6 x 150 mm), injection volume 2 µL, temperature 50 °C

a) MeOH/phosphate buffer pH 5.0 gradient: 9/95 to 90/10; flow 1.0 mL/min

Method 3: Synergi 4u Hydro-RP 80A, 5.0 µm (4.6 x 150 mm), injection volume 20 µL, max. temperature 60 °C

a) (ACN+0.05 % TFA)/(water+0.05 % TFA) gradient: 10/90 to 100/0; flow 2.0 mL/min

Method 4: Thermo Scientific Hypersil GOLD, 3.0 µm (2.1 x 50 mm), injection volume 5 – 10 µL, temperature 31 °C

a) ACN/water+0.1 % TFA gradient: 10/90 to 99/1; flow 0.4 mL/min

Compound characterization: NMR spectra (^{1}H-NMR, ^{13}C-NMR, DEPT, COSY, HSQC, HMBC) were recorded using Avance III HD 400 MHz Bruker BioSpin and Avance III HD 500 MHz Bruker BioSpin (^{1}H-NMR: 400 MHz and 500 MHz, ^{13}C-NMR: 101 MHz and 126 MHz) using the deuterated solvent stated. Chemical shifts (δ) are quoted in parts per million (ppm) and referenced to the residual solvent peak. Multiplicities are denoted as s-singlet, d-doublet, t-triplet, q-quartet, and quin-quintet and derivatives thereof (br denotes a broad resonance peak). Coupling constants *J* are given in Hz and rounded to the nearest 0.1 Hz. E/Z ratios were determined using ^{1}H-NMR spectra. Infrared spectra were recorded from 4000 to 650 cm^{-1} on a PERKIN ELMER Spectrum BX-59343 FT-IR instrument. A Smiths Detection DuraSamp IR II Diamond ATR sensor was used for detection. The absorption bands are reported in wavenumbers [cm^{-1}]. High resolution mass spectra (HR-MS) were recorded using a Jeol MStation 700 or JMS GCmate II Jeol instrument for electron impact ionisation (EI). Thermo Finnigan LTQ was used for electrospray ionisation (ESI). Melting points were measured with a Büchi Schmelzpunktapparatur B-540 and are reported in °C.

Synthetic procedures

For the detailed synthesis of balsalazide see our latest publication in the European Journal of Medicinal Chemistry [1] *and for the preparation of rigidised analogue* ***CG_224*** *see our publication in Synthesis*[2].

Standard synthetic protocols:

General Procedure A – Hydrogenation *(The herein used general procedure follows work from Hofmann et al. [3])*

The appropriate nitro compound (1.00 equivalent) was dissolved to a concentration of 0.3 M in dry MeOH and Pd/C (10 wt % on activated carbon) was added under N_2 atmosphere. Hydrogenation was performed under 1 bar H_2 pressure at rt for 3 h. Subsequently, the reaction mixture was filtered through a pad of celite and the solvent was removed *in vacuo*. If necessary, the crude product was purified by flash column chromatography (FCC) using the indicated eluent.

General Procedure B – Azo coupling reaction *(The herein used general procedure follows a patent from Biorex Laboratories Ltd. from 1983 [4])*

The appropriate aniline (1.00 equivalents) was suspended in H_2O to a concentration of 0.3 M, before conc. HCl (3.40 equivalents) was added. The mixture was cooled to 0 °C and a cold 2.0 M solution of $NaNO_2$ (1.03 equivalents) in water was added dropwise. The mixture was stirred for 1 h and then added dropwise to a solution of the corresponding phenol (1.03

equivalents) dissolved in water to a concentration of 0.6 M containing sodium hydroxide (2.10 equivalents) and sodium carbonate (1.60 equivalents) at 0 °C. The reaction mixture was adjusted to pH 8 with 2 M aq. NaOH and stirred for 2 – 3 h (TLC monitoring), then poured into 0.5 M aq. HCl at 0 °C. The precipitate obtained was collected by filtration, washed with water and suction dried to give the corresponding azobenzene in sufficient purity if not stated otherwise.

General Procedure C – Protection to acetonides of salicylates *(The herein used general procedure follows a patent from APPLIED RESEARCH SYSTEMS ARS HOLDING N.V. from 2005 [5])*

A mixture of the appropriate salicylic acid derivative (1.0 equivalents) in TFA (16 equivalents), TFAA (3.0 equivalents) and acetone (5.0 equivalents) was heated to 90 °C for 16 h. After cooling to room temperature, the reaction mixture was concentrated under reduced pressure. The residue was dissolved in EtOAc and washed with sat. aq. $NaHCO_3$ (2 x) and brine (2 x). The organic layer was dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC using the indicated eluent.

General Procedure D – Alkaline deprotection of esters and acetonides *(The herein used general procedure is based on work from Mukkamala et al. [6])*

The appropriate protected starting material (1.0 equivalents) was dissolved in THF to a concentration of 0.15 M followed by the addition of a solution of KOH (5.0 equivalents) in water (same volume as THF). The reaction mixture was stirred at the stated temperature for 1 – 3 h (TLC monitoring) and then acidified to pH 1 with 2 M aq. HCl at room temperature. It was proceeded with the indicated work-up procedure.

General Procedure E – Amide formation (I) with HATU and DIPEA *(The herein used general procedure follows a patent from CHIESI FARMACEUTICI S.P.A. from 2016 [7])*

The appropriate carboxylic acid (1.5 equivalents) was dissolved in DMF to a concentration of 0.3 M. Then DIPEA (3.0 equivalents) was added, followed by HATU (1.5 equivalents). The reaction mixture was stirred at room temperature for 40 min. The appropriate amine (1.0 equivalents) was added, and the reaction mixture stirred at room temperature for 18 h. The reaction mixture was diluted with EtOAc and washed sequentially with water (2 x), 10 % aq. K_2CO_3 (1 x), brine (1 x) and 5 % aq. LiCl (1 x). The organic phase was dried using a phase separation paper and the filtrate was concentrated *in vacuo.* The crude product was then purified by FCC using the indicated eluent.

General Procedure F – Amide formation (II) with HOBt, EDC•HCl and DIPEA *(The herein used general procedure follows to a patent from NOVO NORDISK A/S from 2004 [8])*

The appropriate carboxylic acid (1.0 equivalents) and 1-hydroxybenzotriazole (1.1 equivalents) were dissolved in THF to a concentration of 0.3 M (with respect to the acid). EDC•HCl (1.1

equivalents) was added. If a semi-crystalline solid appeared, DCM (half the volume of THF) was added and a clear solution was obtained. The mixture was stirred for 10 min, then the appropriate amine (1.0 equivalents) and DIPEA (3.0 equivalents) were added. The mixture was stirred for 2 h at room temperature and then heated to reflux for 10 min. The reaction mixture was cooled to room temperature again, then partitioned between water and DCM. The aqueous phase was further extracted with DCM (3 x) and the combined organic layers were dried using a phase separation paper and the solvent removed *in vacuo*. The crude product was purified by FCC using the indicated eluent.

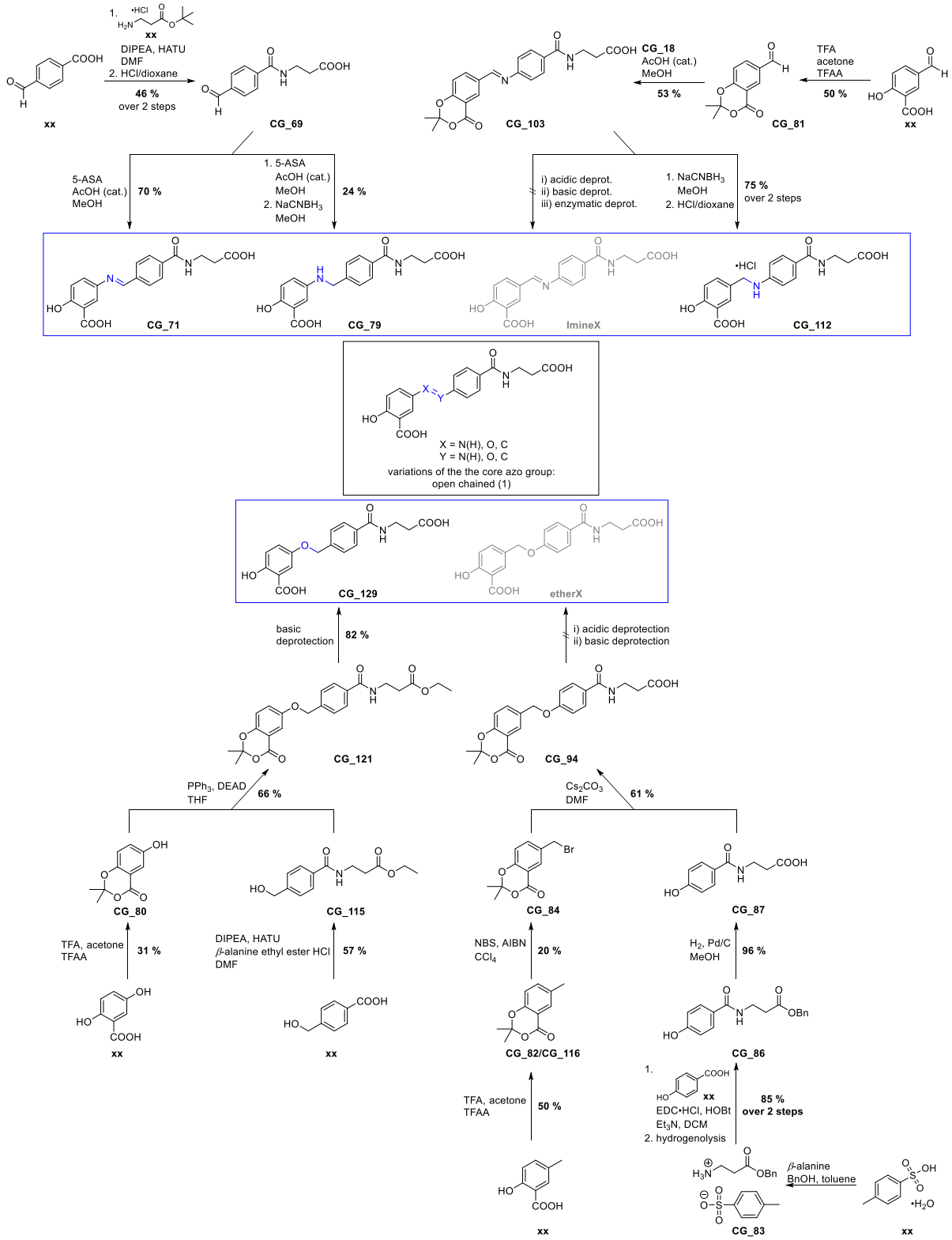
General Procedure G – Amide formation (III) with HOBt, EDC•HCl and Et_3N *(The herein used general procedure follows KLOSS et al. [9])*

The appropriate carboxylic acid/benzoic acid (1.0 equivalents), the corresponding amine (1.2 equivalents), EDC•HCl (1.2 equivalents) and HOBt (1.2 equivalents) were dissolved in dry DCM to a concentration of 0.3 – 0.4 M (with respect to the acid) under N_2 atmosphere. Subsequently, triethylamine (1.8 equivalents) was added, and the mixture was stirred at room temperature for 24 h. The solution was diluted with ethyl acetate and washed with sat. aq. NH_4Cl (2 x) and water (1 x). The organic phase was dried using a phase separation paper and the solvent was evaporated. The crude product was then purified by FCC using the indicated eluent.

General Procedure H – Sulfonamide formation

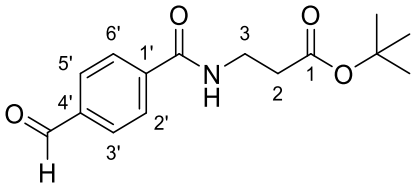
The appropriate amine (1.0 equivalents) was dissolved in DCM to a concentration of 0.1 M. Then, pyridine (2.0 equivalents) and 5-chlorosulfonyl-2-hydroxybenzoic acid (**105**, 1.0 equivalents) were added at 0 °C under N_2 atmosphere. The mixture was stirred at 0 °C for 2 h and afterwards at room temperature until completion of the reaction (TLC monitoring). The solvent was removed *in vacuo* and the residue dissolved in $CHCl_3$:isopropanol (3:1) and washed with 1 M aq. HCl (3 x). The solvent was then evaporated *in vacuo* and the crude product purified by FCC using the indicated eluent.

Synthesis of open-chained analogues imine CG_71, amines CG_79 and CG_112, and ether CG_129



Scheme S1: Preparation of open-chained analogues imine **CG_71**, amines **CG_79** and **CG_112**, and ether **CG_129**.

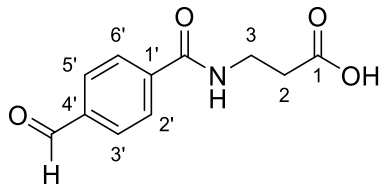
tert-Butyl 3-(4-formylbenzamido)propanoate (**46**) [7]



Prepared according to **General Procedure E** from 4-carboxybenzaldehyde (**44**, 1.50 g, 9.99 mmol) and *tert*-butyl 3-aminopropanoate hydrochloride (**45**, 1.21 g, 6.66 mmol). The crude product was purified by FCC (EtOAc/hexanes 35:65) to give aldehyde **46** (1.76 g, 6.35 mmol, 95 %) as a white solid.

R_f: 0.26 (EtOAc/hexanes 40:60). **m.p.:** 103 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 10.08 (s, 1H, CHO), 7.97 – 7.93 (m, 2H, 3'-H and 5'-H), 7.93 – 7.88 (m, 2H, 2'-H and 6'-H), 7.01 (t, J = 6.2 Hz, 1H, CONH), 3.74 – 3.68 (m, 2H, 3-H), 2.59 – 2.55 (m, 2H, 2-H), 1.47 (s, 9H, $C(CH_3)_3$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 191.7 (CHO), 172.5 (C-1), 166.2 (CONH), 139.8 (C-1'), 138.4 (C-4'), 130.0 (C-3' and C-5'), 127.8 (C-2' and C-6'), 81.6 (**C**$(CH_3)_3$), 35.8 (C-3), 35.0 (C-2), 28.3 (C(**C**$H_3)_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3299, 2724, 1708, 1634, 1552, 1330, 1146, 846, 826, 754, 653. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{15}H_{18}NO_4^-$: 276.1241; found: 276.1243. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

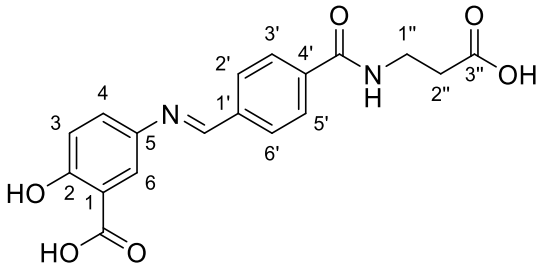
3-(4-Formylbenzamido)propanoic acid (**47**) [7]



A solution of 4 m HCl in 1,4-dioxane (6.5 mL) was added to aldehyde **46** (900 mg, 3.25 mmol) and the reaction mixture was stirred at room temperature for 3 h. The resulting precipitate was collected by filtration, washed with ether (20 mL), and dried *in vacuo* to afford product **47** (517 mg, 2.34 mmol, 72 %) as a white solid.

R_f: 0.33 (EtOAc/hexanes+AcOH 70:30+1). **m.p.:** 162 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.25 (s, 1H, COOH), 10.07 (s, 1H, CHO), 8.75 (t, *J* = 5.5 Hz, 1H, CONH), 8.03 – 8.00 (m, 2H, 2'-H and 6'-H), 8.00 – 7.97 (m, 2H, 3'-H and 5'-H), 3.48 (td, *J* = 7.1, 5.4 Hz, 2H, 3-H), 2.53 (t, *J* = 7.1 Hz, 2H, 2-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 192.9 (CHO), 172.8 (C-1), 165.4 (CONH), 139.4 (C-1'), 137.8 (C-4'), 129.4 (C-3' and C-5'), 127.9 (C-2' and C-6'), 35.7 (C-3), 33.6 (C-2). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3297, 1698, 1633, 1546, 1436, 1334, 1226. 928, 854, 821. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{11}H_{10}NO_4^-$: 220.0615; found: 220.0616. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

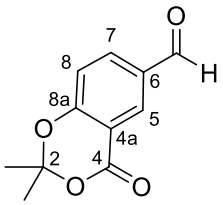
(*E*)-5-((4-((2-Carboxyethyl)carbamoyl)benzylidene)amino)-2-hydroxybenzoic acid (**CG_71**)



Aldehyde **47** (235 mg, 1.06 mmol) and 5-aminosalicylic acid (5-ASA; 54.2 mg, 0.354 mmol) were added to MeOH (6.0 mL). AcOH (4.0 µL) was then added to the reaction mixture, which turned into a clear orange solution after 5 min and was stirred at room temperature for 24 h. Precipitates were collected by centrifugation, washed with MeOH (3 x 8.0 mL), and dried *in vacuo* to provide imine **CG_71** (88.3 mg, 0.248 mmol, 70 %) as an orange solid.

R_f: 0.33 (EtOAc/hexanes+AcOH 70:30+1). **m.p.:** 240 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.26 (s, 2H, 1-COOH and 3''-COOH), 10.07 (s, 1H, OH), 8.76 (s, 1H, Ar-C**H**), 8.65 (t, *J* = 5.5 Hz, 1H, CONH), 8.03 – 7.98 (m, 2H, 2'-H and 6'-H), 7.98 – 7.91 (m, 2H, 3'-H and 5'-H), 7.77 (d, *J* = 2.7 Hz, 1H, 6-H), 7.59 (dd, *J* = 8.8, 2.7 Hz, 1H, 4-H), 7.03 (d, *J* = 8.8 Hz, 1H, 3-H), 3.48 (td, *J* = 7.1, 5.5 Hz, 2H, 1''-H), 2.54 (t, *J* = 7.1 Hz, 2H, 2''-H). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (C-3''), 171.6 (1-COOH), 165.7 (CONH), 160.0 (C-2), 158.3 (Ar-**C**H), 142.2 (C-5), 138.4 (C-1'), 136.4 (C-4'), 129.0 (C-4), 128.3 (C-2' and C-6'), 127.6 (C-3' and C-5'), 122.4 (C-6), 118.0 (C-3), 113.4 (C-1), 35.7 (C-1''), 33.7 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 1709, 1553, 1489, 1293, 1193, 1088, 1013, 838, 569. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{18}H_{15}N_2O_6^-$: 355.0936; found: 355.0938. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

2,2-Dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxine-6-carbaldehyde (**50**) [5]

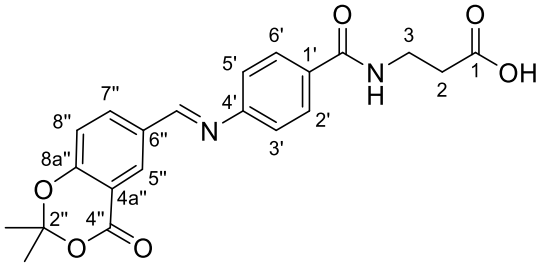


Prepared according to **General Procedure C** from 5-formylsalicylic acid (**48**, 1.00 g, 6.02 mmol). The crude product was purified by FCC (EtOAc/hexanes 12:88) to give acetonide **50** (618 mg, 3.00 mmol, 50 %) as a pale yellow solid.

R_f: 0.43 (EtOAc/hexanes 30:70). **m.p.:** 117 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 9.97 (d, *J* = 0.8 Hz, 1H, CHO), 8.47 (d, *J* = 2.1 Hz, 1H, 5-H), 8.12 (dd, *J* = 8.6, 2.1 Hz, 1H, 7-H), 7.12 (dd, *J* = 8.6, 0.8 Hz, 1H, 8-H), 1.78 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 189.7 (CHO), 160.6 (C-4), 160.0 (C-8a), 135.9 (C-7), 133.6 (C-5), 131.6 (C-6), 118.6 (C-8),

113.7 (C-4a), 107.4 (C-2), 26.1 ($C(CH_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 1737, 1686, 1611, 1580, 1265, 1196, 1108, 841, 779, 754, 581. **HR-MS (EI):** *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{11}H_{10}O_4^{\bullet+}$: 206.0574; found: 206.0577. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

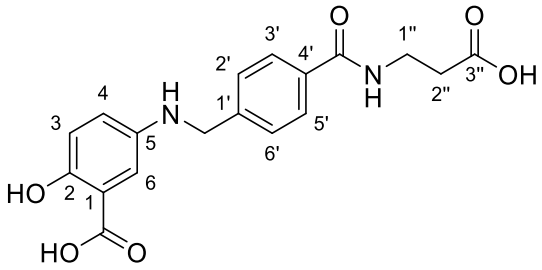
(*E*)-3-(4-(((2,2-Dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)methylene)amino)benzamido)-propanoic acid (**CG_103**)



Aldehyde **50** (250 mg, 1.21 mmol) was dissolved in MeOH (6.0 mL). Amine **3** (278 mg, 1.33 mmol) and AcOH (4.0 µL) were added and the reaction mixture was stirred at room temperature for 16 h. Precipitates were then collected by centrifugation, washed with MeOH (3 x 6.0 mL), and dried *in vacuo* to provide imine **CG_103** (255 mg, 0.634 mmol, 53 %) as a white solid.

R_f: 0.46 (EtOAc/hexanes+AcOH 30:70+1). **m.p.:** 203 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.23 (s, 1H, COOH), 8.71 (s, 1H, Ar-C**H**), 8.53 (t, *J* = 5.5 Hz, 1H, CONH), 8.44 (d, *J* = 2.1 Hz, 1H, 5''-H), 8.27 (dd, *J* = 8.5, 2.1 Hz, 1H, 7''-H), 7.92 – 7.88 (m, 2H, 2'-H and 6'-H), 7.36 – 7.31 (m, 2H, 3'-H and 5'-H), 7.29 (d, *J* = 8.6 Hz, 1H, 8''-H), 3.47 (td, J = 7.1, 5.3 Hz, 2H, 3-H), 2.54 – 2.51 (m, 2H, 2-H), 1.74 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (COOH), 165.7 (CONH), 160.1 (Ar-**C**H), 159.8 (C-4''), 157.7 (C-8a''), 153.5 (C-4'), 136.4 (C-7''), 131.8 (C-1'), 130.9 (C-6''), 130.3 (C-5''), 128.4 (C-2' and C-6'), 120.8 (C-3' and C-5'), 118.2 (C-8''), 113.2 (C-4a''), 107.0 (C-2), 35.6 (C-3), 33.8 (C-2), 25.3 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3357, 2359, 1728, 1707, 1648, 1591, 1498, 1296, 1231, 1174, 947, 854, 771, 672. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{21}H_{19}N_2O_6^-$: 395.1249; found: 395.1249. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

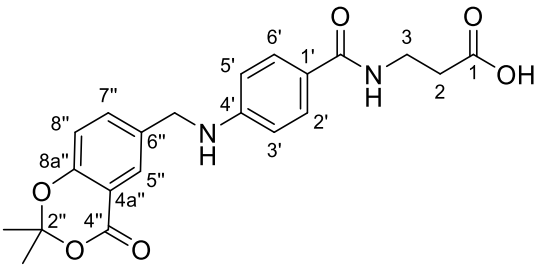
5-((4-((2-Carboxyethyl)carbamoyl)benzyl)amino)-2-hydroxybenzoic acid (**CG_79**)



Aldehyde **47** (190 mg, 0.859 mmol) and 5-aminosalicylic acid (5-ASA, 65.8 mg, 0.429 mmol) were added to MeOH (4.6 mL). AcOH (3.0 µL) was then added to the reaction mixture, which turned into a clear orange solution after 5 min and was stirred at room temperature for 24 h. Precipitates were collected by centrifugation and washed with MeOH (3 x 8.0 mL). The crude precipitates were then resuspended in MeOH (5.0 mL) and $NaBH_3CN$ (85.2 mg, 1.290 mmol) was added in portions at room temperature to the stirred mixture. After stirring for additional 4 h, water (5.0 mL) was added. The solvents were removed *in vacuo* and the crude product purified by FCC (100 % MeOH) to give amine **CG_79** (37.3 mg, 0.104 mmol, 24 %) as a yellow solid.

R_f: 0.22 (EtOAc/hexanes+AcOH 80:20+1), streaks. **m.p.:** 330 °C. **^{1}H-NMR (400 MHz, D_2O):** δ (ppm) = 7.71 – 7.65 (m, 2H, 3'-H and 5'-H), 7.45 (d, J = 8.1 Hz, 2H, 2'-H and 6'-H), 7.35 (d, J = 2.9 Hz, 1H, 6-H), 7.00 (dd, J = 8.8, 2.9 Hz, 1H, 4-H), 6.82 (d, J = 8.7 Hz, 1H, 3-H), 4.36 (s, 2H, Ar-C**H_2**), 3.58 (t, J = 7.0 Hz, 2H, 1''-H), 2.51 (t, J = 7.0 Hz, 2H, 2''-H). **^{13}C-NMR (101 MHz, D_2O):** δ (ppm) = 180.2 (C-3''), 175.0 (1-COOH), 170.4 (CONH), 154.2 (C-2), 141.7 (C-1'), 132.8 (C-4'), 128.4 (C-2' and C-6'), 127.3 (C-3' and C-5'), 123.0 (C-4), 118.4 (C-1), 117.8 (C-6), 117.0 (C-3), 49.5 (Ar-**C**H_2), 37.1 (C-1''), 36.5 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3335, 2000, 1659, 1631, 1566, 1494, 1484, 1413, 1300, 1234, 1018, 822. **HR-MS (ESI):** m/z = $[M-H]^-$ calcd for $C_{18}H_{17}N_2O_6^-$: 357.1092; found: 357.1094. **Purity (HPLC):** 210 nm: 95 %; 254 nm: 87 % (method 3a).

3-(4-(((2,2-Dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)methyl)amino)benzamido)propanoic acid (**51**)

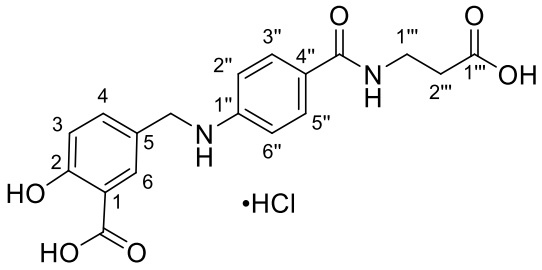


Imine **CG_103** (100 mg, 0.252 mmol) was dissolved in MeOH (2.5 mL) under N_2 atmosphere and $NaBH_3CN$ (50.1 mg, 0.757 mmol) was added in portions. After stirring for 4 h, the reaction was quenched with sat. aq. $NaHCO_3$ (10 mL). The mixture was concentrated *in vacuo* and directly purified by FCC (EtOAc/hexanes+AcOH 80:20+1) to give amine **51** (99.5 mg, 0.250 mmol, 99 %) as a white solid.

R_f: 0.31 (EtOAc/hexanes+AcOH 80:20+1). **m.p.:** 172 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 8.08 (t, J = 5.6 Hz, 1H, CONH), 7.84 (d, J = 2.2 Hz, 1H, 5''-H), 7.66 (dd, J = 8.5, 2.3 Hz, 1H, 7''-H), 7.59 – 7.55 (m, 2H, 2'-H and 6'-H), 7.09 (d, J = 8.4 Hz, 1H, 8''-H), 6.82 (t, J =

6.2 Hz, 1H, NH), 6.61 – 6.55 (m, 2H, 3'-H and 5'-H), 4.34 (d, *J* = 6.1 Hz, 2H, Ar-C**H₂**), 3.38 (td, *J* = 7.1, 5.4 Hz, 2H, 3-H), 2.42 (t, *J* = 7.2 Hz, 2H, 2-H), 1.67 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 173.3 (C-1), 166.1 (CONH), 160.3 (C-4''), 154.3 (C-8a''), 150.7 (C-4'), 136.0 (C-7''), 134.7 (C-6''), 128.6 (C-2' and C-6'), 127.4 (C-5''), 121.6 (C-1'), 117.3 (C-8''), 112.8 (C-4a''), 111.1 (C-3' and C-5'), 106.3 (C-2''), 44.9 (Ar-**C**H_2), 35.5 (C-3), 34.5 (C-2), 25.2 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3340, 2828, 1736, 1694, 1620, 1599, 1509, 1501, 1301, 1261, 1214, 1185, 915, 835, 766. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{21}H_{21}N_2O_6^-$: 397.1405; found: 397.1406. **Purity (HPLC):** 210 nm: 90 %; 254 nm: 93 % (method 3a).

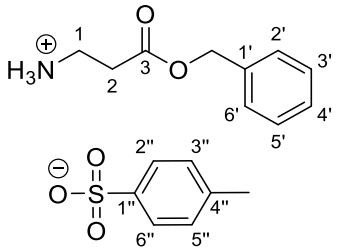
5-(((4-((2-Carboxyethyl)carbamoyl)phenyl)amino)methyl)-2-hydroxybenzoic acid hydrochloride (**CG_112**)



Amine **51** (35.0 mg, 0.0880 mmol) was dissolved in 1,4-dioxane (0.60 mL) followed by the dropwise addition of conc. HCl (45 µL). The reaction mixture was heated to 60 °C for 2 h, cooled to room temperature and the precipitate collected by filtration and dried *in vacuo* to give HCl-salt **CG_112** (26.2 mg, 0.0664 mmol, 76 %) as a white solid.

R_f: not determinable. **m.p.:** 170 °C. **^{1}H-NMR (500 MHz, CD_3OD):** *δ* (ppm) = 7.94 (d, *J* = 2.4 Hz, 1H, 6-H), 7.90 – 7.87 (m, 2H, 3''-H and 5''-H), 7.50 (dd, *J* = 8.6, 2.4 Hz, 1H, 4-H), 7.33 – 7.30 (m, 2H, 2''-H and 6''-H), 6.96 (d, *J* = 8.5 Hz, 1H, 3-H), 4.55 (s, 2H, Ar-C**H₂**), 3.63 (t, *J* = 6.8 Hz, 2H, 1'''-H), 2.66 (t, *J* = 6.8 Hz, 2H, 2'''-H). **^{13}C-NMR (126 MHz, CD_3OD):** *δ* (ppm) = 173.9 (C-1'''), 172.9 (1-COOH), 168.9 (CONH), 163.9 (C-2), 141.7 (C-1''), 138.0 (C-4), 133.9 (C-4''), 133.3 (C-6), 130.3 (C-3'' and C-5''), 124.1 (C-5), 122.1 (C-2'' and C-6''), 119.0 (C-3), 114.2 (C-1), 54.2 (Ar-**C**H_2), 37.1 (C-1'''), 34.6 (C-2'''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3321, 2544, 1666, 1641, 1596, 1278, 1185, 1112, 862, 848, 754. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{18}H_{17}N_2O_6^-$: 357.1092; found: 357.1092. **Purity (HPLC):** 210 nm: 91 %; 254 nm: 91 % (method 2a).

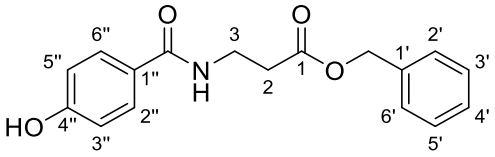
3-(Benzyloxy)-3-oxopropan-1-aminium 4-methylbenzenesulfonate (**53**) [9]



β-Alanine (445 mg, 5.00 mmol) and *p*-toluenesulfonic acid hydrate (**52**, 1.14 g, 6.00 mmol) were suspended in benzyl alcohol (1.55 mL, 15.0 mmol) and toluene (15 mL). The mixture was heated to reflux with a Dean-Stark equipment for 5 h and then cooled to room temperature. The resulting precipitate was collected by filtration and washed with diethyl ether (3 x 20 mL). It was then resuspended in EtOAc (300 mL) and heated to reflux. The hot suspension was repeatedly filtered, and the precipitate collected and dried *in vacuo* to give aminoester salt **53** (1.45 g, 3.33 mmol, 83 %) as a white crystalline solid.

R_f: not determinable. **m.p.:** 140 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 7.75 (s, 3H, NH_3^+), 7.50 – 7.46 (m, 2H, 2''-H and 6''-H), 7.41 – 7.32 (m, 5H, **Ar**-CH_2), 7.13 – 7.09 (m, 2H, 3''-H and 5''-H), 5.14 (s, 2H, Ar-C**H_2**), 3.06 (t, *J* = 6.9 Hz, 2H, 1-H), 2.72 (t, *J* = 6.9 Hz, 2H, 2-H), 2.29 (s, 3H, CH_3). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 170.2 (C-3), 145.6 (C-1''), 137.7 (C-4''), 135.8 (C-1'), 128.5 (C-3' and C-5'), 128.2 (C-4'), 128.09 (C-3'' and C-5''), 128.07 (C-2' and C-6'), 125.5 (C-2'' and C-6''), 66.0 (Ar-**C**H_2), 34.7 (C-1), 31.4 (C-2), 20.8 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3051, 1732, 1518, 1400, 1228, 1201, 1154, 1120, 1031, 1007, 815, 742, 865, 568. **HR-MS (ESI):** *m/z* = $[M+H]^+$ calcd for $C_{10}H_{14}NO_2^+$: 180.1019; found: 180.1018. **Purity (HPLC):** 254 nm: >95 % (method 4a).

Benzyl 3-(4-hydroxybenzamido)propanoate (**55**) [9]

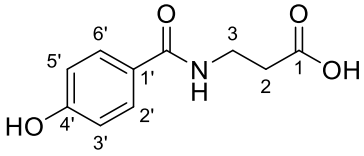


Prepared according to **General Procedure G** from 4-hydroxybenzoic acid (**54**, 276 mg, 2.00 mmol) and aminoester salt **53** (843 mg, 2.40 mmol). The crude product was purified by FCC (EtOAc/hexanes 50:50) to give product **55** (532 mg, 1.78 mmol, 89 %) as a white solid.

R_f: 0.24 (EtOAc/hexanes 50:50). **m.p.:** 112 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** *δ* (ppm) = 7.61 – 7.54 (m, 2H, 2''-H and 6''-H), 7.38 – 7.29 (m, 5H, **Ar**-CH_2), 6.86 – 6.82 (m, 2H, 3''-H and 5''-H), 6.79 (d, *J* = 6.0 Hz, 1H, CONH), 5.15 (s, 2H, Ar-C**H_2**), 3.73 (qd, *J* = 6.6, 6.1, 3.1 Hz, 2H, 3-H), 2.70 (t, J = 5.8 Hz, 2H, 2-H). **^{13}C-NMR (101 MHz, $CDCl_3$):** *δ* (ppm) = 172.9 (C-1), 167.8 (CONH), 159.9 (C-4''), 135.6 (C-1'), 129.1 (C-2'' and C-6''), 128.8 (C-3' and C-5'), 128.6 (C-

4'), 128.5 (C-2' and C-6'), 125.7 (C-1''), 115.7 (C-3'' and C-5''), 66.9 (Ar-**C**H_2), 35.6 (C-3), 34.2 (C-2). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3303, 1708, 1635, 1510, 1273, 1205, 1174, 960, 846, 757, 697. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{17}H_{16}NO_4^-$: 298.1085; found: 298.1087. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

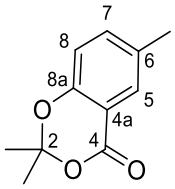
3-(4-Hydroxybenzamido)propanoic acid (**56**) [9]



Phenol **55** (500 mg, 1.67 mmol) was dissolved in dry MeOH (10 mL) and 10 % Pd/C (10 wt% on activated carbon, 50.0 mg) was added under N_2 atmosphere. Hydrogenation was performed under 1 bar H_2 pressure at room temperature for 1.5 h. Subsequently, the reaction mixture was filtered through a pad of celite and the solvent was removed *in vacuo* to give carboxylic acid **56** (336 mg, 1.61 mmol, 96 %) as a white solid.

R_f: 0.27 (EtOAc/hexanes+AcOH 70:30+1).
m.p.: 180 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 8.25 (t, *J* = 5.5 Hz, 1H, CONH), 7.73 – 7.64 (m, 2H, 2'-H and 6'-H), 6.83 – 6.74 (m, 2H, 3'-H and 5'-H), 3.41 (td, *J* = 7.2, 5.4 Hz, 2H, 3-H), 2.52 – 2.44 (m, 2H, 2-H). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 173.0 (C-1), 165.9 (CONH), 160.1 (C-4'), 129.0 (C-2' and C-6'), 125.1 (C-1'), 114.7 (C-3' and C-5'), 35.5 (C-3), 34.0 (C-2). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3439, 3247, 1703, 1637, 1503, 1331, 1205, 923, 851, 726. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{10}H_{10}NO_4^-$: 208.0615; found: 208.0616. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

2,2,6-Trimethyl-4*H*-benzo[*d*][1,3]dioxin-4-one (**58**)



Method A:[6]
Under N_2 atmosphere, first acetone (4.83 mL, 65.7 mmol) and then dropwise $SOCl_2$ (3.36 mL, 46.0 mmol) were added to a solution of 5-methylsalicylic acid (**57**, 5.00 g, 32.9 mmol) and DMAP (2.01 g, 16.4 mmol) in 1,2-dimethoxyethane (25 mL) at 0 °C and stirred for 1 h at this temperature. The reaction mixture was then allowed to reach room temperature and stirred for further 19 h. Sat. aq. $NaHCO_3$ (25 mL) was added to the mixture and the aq. phase was extracted with diethyl ether (3 x 35 mL). The combined organic layers were washed with brine

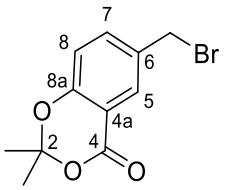
(2 x 20 mL), dried using a phase separation paper and the solution concentrated under reduced pressure. The crude product was purified by FCC (EtOAc/hexanes 5:95) to give acetonide **58** (1.61 g, 8.39 mmol, 26 %) as an off-white solid.

Method B:

Prepared according to **General Procedure C** from 5-methylsalicylic acid (**57**, 6.70 g, 44.0 mmol). The crude product was purified by FCC (EtOAc/hexanes 7:93) to give acetonide **58** (4.26 g, 22.2 mmol, 50 %) as an off-white solid.

R_f: 0.33 (EtOAc/hexanes 7:93). **m.p.:** 87 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 7.75 (dp, *J* = 2.0, 0.8 Hz, 1H, 5-H), 7.37 – 7.32 (m, 1H, 7-H), 6.86 (d, *J* = 8.3 Hz, 1H, 8-H), 2.34 (d, *J* = 0.8 Hz, 3H, 6-C**H$_3$**), 1.72 (s, 6H, C(C**H$_3$**)$_2$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 161.5 (C-4), 154.1 (C-8a), 137.4 (C-7), 132.4 (C-6), 129.6 (C-5), 117.1 (C-8), 113.4 (C-4a), 106.4 (C-2), 25.9 (C(**C**H$_3$)$_2$), 20.7 (6-**C**H$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 1729, 1618, 1489, 1380, 1292, 1200, 1046, 932, 832, 779. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{11}H_{13}O_3^+$: 193.0859; found: 193.0859. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

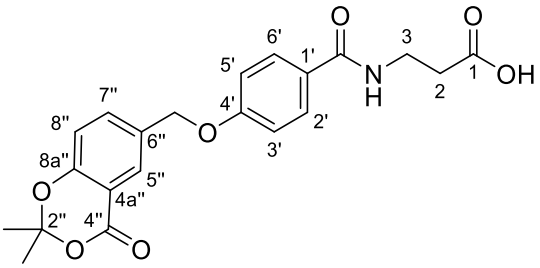
6-(Bromomethyl)-2,2-dimethyl-4*H*-benzo[*d*][1,3]dioxin-4-one (**59**) [6]



Acetonide **58** (1.00 g, 5.20 mmol) was dissolved in CCl_4 (13 mL) in a pressure tube. NBS (1.02 g, 5.72 mmol) and AIBN (87.0 mg, 0.520 mmol) were added and the reaction mixture heated to 80 °C for 16 h. Then, it was cooled to room temperature, filtered and the solvent of the filtrate removed *in vacuo*. The resulting residue was dissolved in DCM (15 mL), washed with brine (2 x 15 mL) and the solvent removed *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 5:95) to give benzyl bromide **59** (645 mg, 2.38 mmol, 46 %) as a white solid.

R_f: 0.20 (EtOAc/hexanes 5:95). **m.p.:** 149 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 7.99 – 7.96 (m, 1H, 5-H), 7.59 (dd, *J* = 8.5, 2.3 Hz, 1H, 7-H), 6.96 (d, *J* = 8.5 Hz, 1H, 8-H), 4.48 (s, 2H, CH_2), 1.74 (s, 6H, C(CH_3)$_2$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 160.7 (C-4), 156.1 (C-8a), 137.3 (C-7), 132.6 (C-6), 130.0 (C-5), 118.1 (C-8), 113.7 (C-4a), 106.8 (C-2), 32.2 (CH_2), 26.0 (C(CH_3)$_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 1736, 1615, 1492, 1302, 1201, 1053, 936, 847, 779, 705, 607. **HR-MS (EI):** *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{11}H_{11}BrO_3^{\bullet+}$: 269.9886; found: 269.9881. **Purity (NMR):** >95 %.

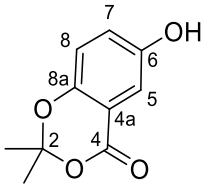
3-(4-((2,2-Dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)methoxy)benzamido)propanoic acid (**CG_94**)



To phenol **56** (100 mg, 0.478 mmol), bromide **59** (130 mg, 0.478 mmol) and Cs_2CO_3 (389 mg, 1.20 mmol) was added dry DMF (5.0 mL) and the mixture heated to 60 °C for 20 min. Water (50 mL) was added and stirring continued for 30 min. The aq. phase was acidified with 2 M aq. HCl and extracted with EtOAc (3 x 50 mL). The combined organic layers were washed with 1 M aq. HCl (20 mL), dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC (EtOAc/toluene+AcOH 50:50+1) to give ether **CG_94** (117 mg; 0.292 mmol, 61 %) as a colourless to white oil.

R_f: 0.29 (EtOAc/toluene+AcOH 50:50+1). **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 9.97 (s, 1H, COOH), 8.28 (t, *J* = 5.5 Hz, 1H, CONH), 7.89 (d, *J* = 2.2 Hz, 1H, 5''-H), 7.70 – 7.63 (m, 3H, 2'-H and 6'-H and 7''-H), 7.08 (d, J = 8.5 Hz, 1H, 8''-H), 6.79 – 6.74 (m, 2H, 3'-H and 5'-H), 5.10 (s, 2H, Ar-C**H_2**), 3.47 (td, *J* = 6.9, 5.4 Hz, 2H, 3-H), 2.62 (t, *J* = 6.9 Hz, 2H, 2-H), 1.68 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 171.3 (C-1), 166.0 (CONH), 160.1 (C-4'), 160.0 (C-4''), 155.2 (C-8a''), 137.0 (C-7''), 131.0 (C-6''), 129.0 (C-2' and C-6'), 128.9 (C-5''), 125.0 (C-1'), 117.4 (C-8''), 114.7 (C-3' and C-5'), 112.8 (C-4a''), 106.5 (C-2''), 64.6 (Ar-**C**H_2), 35.4 (C-3), 33.9 (C-2), 25.2 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2956, 2924, 2855, 2360, 1732, 1458, 1259, 1172, 1029, 799. **HR-MS (ESI):** *m/z* = $[M-H]^-$ calcd for $C_{21}H_{20}NO_7^-$: 398.1245; found: 398.1246. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

6-Hydroxy-2,2-dimethyl-4*H*-benzo[*d*][1,3]dioxin-4-one (**62**) [10]

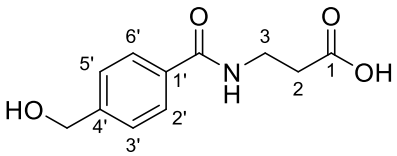


TFAA (2.71 mL, 19.5 mmol) and acetone (2.39 mL, 32.4 mmol) were added to a stirred mixture of 2,5-dihydroxybenzoic acid (**61**, 1.00 g, 6.49 mmol) and TFA (8.03 mL, 107 mmol) at 0 °C. The mixture was slowly warmed to room temperature and after 14 h the solvent was concentrated to 1/3 of its volume. EtOAc (5.0 mL) and sat. aq. $NaHCO_3$ (50 mL) were added and the mixture was stirred for 2 h. The layers were separated, and the aq. phase was

extracted with EtOAc (3 x 50 mL). The combined extracts were dried using a phase separation paper, concentrated *in vacuo* and purified by FCC (EtOAc/hexanes 25:75) to give acetonide **62** (388 mg, 2.00 mmol, 31 %) as an off-white solid.

R_f: 0.32 (EtOAc/hexanes 25:75). **m.p.:** 145 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** *δ* (ppm) = 7.48 (d, *J* = 3.0 Hz, 1H, 5-H), 7.11 (dd, *J* = 8.8, 3.0 Hz, 1H, 7-H), 6.87 (d, *J* = 8.8 Hz, 1H, 8-H), 5.94 (s, 1H, OH), 1.72 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, $CDCl_3$):** *δ* (ppm) = 162.0 (C-4), 151.4 (C-6), 150.1 (C-8a), 124.9 (C-7), 118.6 (C-8), 114.6 (C-5), 113.9 (C-4a), 106.8 (C-2), 25.8 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** *ṽ* (cm^{-1}) = 3313, 1710, 1623, 1500, 1468, 1381, 1328, 1283, 1197, 1060, 985, 887, 780, 640. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{10}H_9O_4^-$: 193.0506; found: 193.0507. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

3-(4-(Hydroxymethyl)benzamido)propanoic acid (**74**)

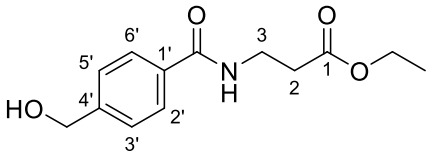


4-(Hydroxymethyl)benzoic acid (**73**, 500 mg, 3.29 mmol) was dissolved in dry DCM (20 mL). At 0 °C oxalyl dichloride (417 µL, 4.93 mmol) was added dropwise, followed by the addition of a catalytic amount of DMF (2 drops). The resulting mixture was stirred at room temperature for 2 h. After this time, the solvent was removed *in vacuo* to afford crude acid chloride.
To a solution of K_2CO_3 (1.09 g, 6.57 mmol) in a mixture of EtOAc/water (2:1, 20 mL) was added *β*-alanine (351 mg, 3.94 mmol). The resulting solution was cooled to 0 °C, followed by dropwise addition of the crude substituted benzoyl chloride dissolved in DCM (3.0 mL). The reaction mixture was warmed to room temperature and stirred for 16 h. Phases were separated, and the aq. phase further extracted with EtOAc (20 mL). The combined organic layers were discarded, and the aq. phase was adjusted to pH 1 – 2 by the addition of 2 M aq. HCl and extracted with EtOAc (3 x 20 mL). The combined organic layers were dried using a phase separation paper and the solvent removed *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes+AcOH 70:30+1) to give product **74** (374 mg, 1.68 mmol, 51 %) as a white solid.

R_f: 0.17 (EtOAc/hexanes+AcOH 70:30+1). **m.p.:** 102 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 12.13 (s, 1H, COOH), 8.70 (t, *J* = 6.0 Hz, 1H, OH), 8.47 (t, *J* = 5.5 Hz, 1H, CONH), 7.81 – 7.77 (m, 2H, 2'-H and 6'-H), 7.40 – 7.36 (m, 2H, 3'-H and 5'-H), 4.54 (s, 2H, Ar-C$\mathbf{H_2}$), 3.45 (td, *J* = 7.1, 5.4 Hz, 2H, 3-H), 2.52 (d, *J* = 1.8 Hz, 1H, 2-H). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 172.9 (COOH), 166.1 (CONH), 145.8 (C-4'), 132.7 (C-1'), 127.0 (C-2' and C-6'), 125.9 (C-3' and C-6'), 62.4 (Ar-$\mathbf{C}H_2$), 35.6 (C-3), 33.9 (C-2). **IR (ATR):** *ṽ* (cm^{-1}) = 3308, 2930,

2602, 1711, 1641, 1548, 1269, 1202, 1005, 850, 673. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{11}H_{12}NO_4^-$: 222.0772; found: 222.0774. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

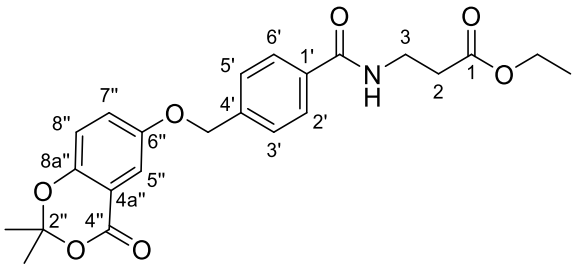
Ethyl 3-(4-(hydroxymethyl)benzamido)propanoate (**77**)



Prepared according to **General Procedure E** from 4-(hydroxymethyl)benzoic acid (**73**, 500 mg, 3.29 mmol) and *β*-alanine ethyl ester hydrochloride (**76**, 337 mg, 2.19 mmol). The crude product was purified by FCC (DCM/EtOAc 50:50) to give amide **77** (315 mg, 1.25 mmol, 57 %) as a colourless oil.

R_f: 0.23 (DCM/EtOAc 50:50). **^{1}H-NMR (400 MHz, $CDCl_3$):** *δ* (ppm) = 7.68 (d, *J* = 7.9 Hz, 2H, 2'-H and 6'-H), 7.35 (d, *J* = 7.8 Hz, 2H, 3'-H and 5'-H), 6.96 (t, *J* = 6.1 Hz, 1H, CONH), 4.71 (s, 2H, Ar-C**H_2**), 4.15 (q, *J* = 7.1 Hz, 2H, C**H_2**CH_3), 3.69 (q, *J* = 5.8 Hz, 2H, 3-H), 2.62 (t, *J* = 5.9 Hz, 2H, 2-H), 1.26 (t, *J* = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (101 MHz, $CDCl_3$):** *δ* (ppm) = 173.1 (C-1), 167.4 (CONH), 144.9 (C-4'), 133.4 (C-1'), 127.2 (C-2' and C-6'), 126.8 (C-3' and C-5'), 64.6 (Ar-**C**H_2), 61.0 (**C**H_2CH_3), 35.5 (C-3), 34.0 (C-2), 14.3 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3321, 1937, 1727, 1714, 1633, 1537, 1307, 1270, 1181, 1017, 853, 752. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{13}H_{18}NO_4^+$: 252.1230; found: 252.1231. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

Ethyl 3-(4-(((2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)oxy)methyl)benzamido)-propanoate (**78**)

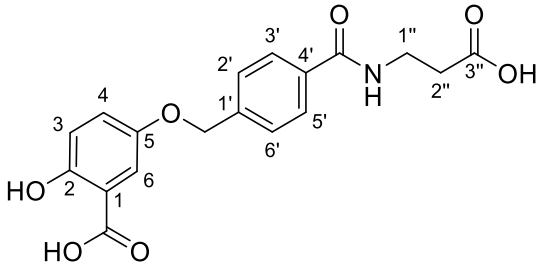


To a solution of primary alcohol **77** (50.0 mg, 0.199 mmol) and phenol **62** (38.6 mg, 0.199 mmol) in dry THF (1.0 mL), triphenylphosphine (52.7 mg, 0.199 mmol) was added at room temperature. The reaction mixture was cooled to 0 °C and DEAD (40 % in toluene, 90.6 µL, 0.199 mmol) was added dropwise and the reaction mixture was allowed to reach room temperature and was stirred for 16 h. The solvent was evaporated under reduced

pressure and the crude product purified by FCC (EtOAc/hexanes 50:50) to give ether **78** (56.4 mg, 0.132 mmol, 66 %) as a white solid.

R_f: 0.40 (EtOAc/hexanes 50:50). **m.p.:** 112 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 7.81 – 7.77 (m, 2H, 2'-H and 6'-H), 7.51 – 7.46 (m, 3H, 3'-H, 5'-H and 5''-H), 7.20 (dd, *J* = 9.0, 3.1 Hz, 1H, 7''-H), 6.91 (d, *J* = 9.0 Hz, 1H, 8''-H), 6.87 (d, *J* = 5.9 Hz, 1H, CONH), 5.10 (s, 2H, 4'-C**H_2**), 4.17 (q, *J* = 7.1 Hz, 2H, C**H_2**CH_3), 3.73 (q, *J* = 6.1 Hz, 2H, 3-H), 2.68 – 2.62 (m, 2H, 2-H), 1.72 (s, 6H, $C(CH_3)_2$), 1.27 (t, *J* = 7.1 Hz, 3H, CH_2C**H_3**). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 173.1 (C-1), 167.0 (CONH), 161.3 (C-4''), 153.8 (C-6''), 150.7 (C-8a''), 140.1 (C-4'), 134.3 (C-1'), 127.5 (C-3' and C-5'), 127.4 (C-2' and C-6'), 125.5 (C-7''), 118.7 (C-8''), 113.9 (C-4a''), 112.5 C-5''), 106.6 (C-2''), 70.1 (4'-**C**H_2), 61.0 (**C**H_2CH_3), 35.4 (C-3), 34.1 (C-2), 25.8 ($C($**C**$H_3)_2$), 14.3 (CH_2**C**H_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3340, 2958, 1726, 1629, 1494, 1291, 1193, 1178, 1135, 1058, 1015, 846. **HR-MS (ESI):** *m/z* = $[M+H]^+$ calcd for $C_{23}H_{26}NO_7^+$: 428.1704; found: 428.1705. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

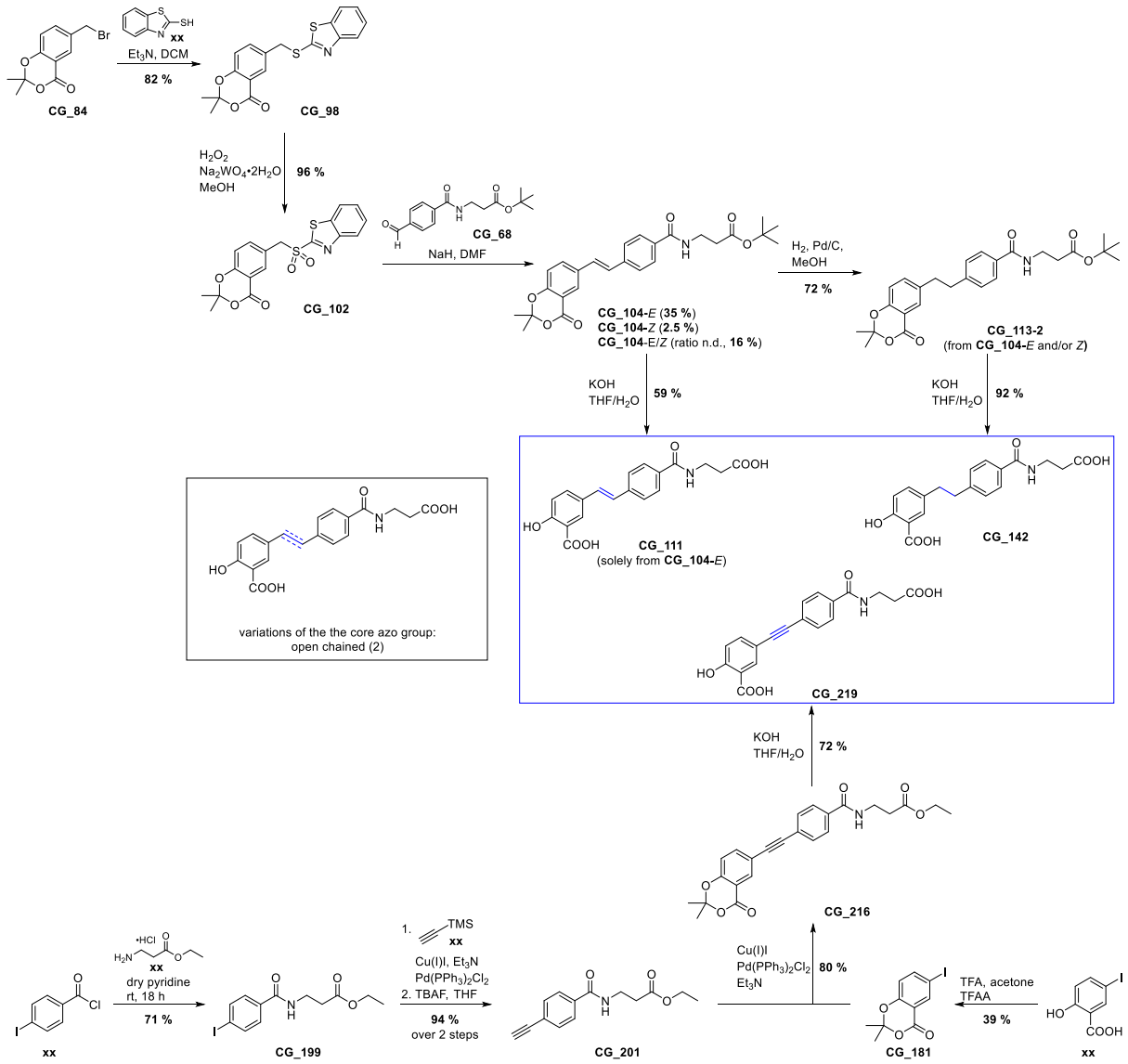
5-((4-((2-Carboxyethyl)carbamoyl)benzyl)oxy)-2-hydroxybenzoic acid (**CG_129**)



Prepared according to **General Procedure D** from compound **78** (40.0 mg, 0.0936 mmol) and heating the reaction mixture to reflux. The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo* to give pure ether **CG_129** (27.4 mg, 0.0763 mmol, 82 %) as a white solid.

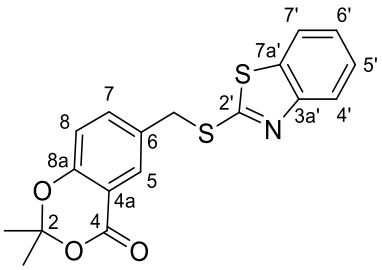
R_f: 0.34 (EtOAc/hexanes+AcOH 60:40+1). **m.p.:** 209 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.59 – 8.47 (m, 1H, CONH), 7.84 – 7.79 (m, 2H, 3'-H and 5'-H), 7.55 – 7.50 (m, 2H, 2'-H and 6'-H), 7.42 (d, *J* = 3.1 Hz, 1H, 6-H), 7.18 (dd, *J* = 9.0, 3.2 Hz, 1H, 4-H), 6.86 (d, *J* = 9.1 Hz, 1H, 3-H), 5.11 (s, 2H, 1'-C**H_2**), 3.67 – 3.60 (m, 2H, 1''-H), 2.64 (t, *J* = 6.9 Hz, 2H, 2''-H). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 175.4 (C-3''), 173.1 (1-COOH), 170.0 (CONH), 157.8 (C-2), 152.2 (C-5), 142.6 (C-1'), 135.1 (C-4'), 128.5 (C-3' and C-5' or C-2' and C-6'), 128.4 (C-3' and C-5' or C-2' and C-6'), 125.4 (C-4), 119.2 (C-3), 115.5 (C-6), 113.7 (C-1), 71.1 (1'-**C**H_2), 37.1 (C-1''), 34.6 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3311, 3921, 1692, 1676, 1624, 1486, 1441, 1292, 1209, 1182, 1084, 1034, 828. **HR-MS (ESI):** *m/z* = $[M-H]^-$ calcd for $C_{18}H_{16}NO_7^-$: 358.0932; found: 358.0933. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

Synthesis of stilbene CG_111, alkane CG_142, and alkyne CG_219



Scheme S2: Preparation of stilbene **CG_111**, alkane **CG_142**, and alkyne **CG_219**.

6-((Benzo[*d*]thiazol-2-ylthio)methyl)-2,2-dimethyl-4*H*-benzo[*d*][1,3]dioxin-4-one (**83**) [6]

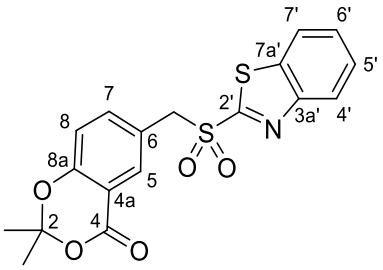


To a solution of bromide **59** (500 mg, 1.84 mmol) and 2-mercaptobenzothiazole (**82**; 370 mg, 2.21 mmol) in DCM (5.0 mL), triethylamine (514 µL, 3.69 mmol) was added at 0 °C. Then the mixture was stirred at room temperature for 2.5 h. Water (50 mL) was added and the aq. phase

extracted with DCM (3 x 30 mL). The combined organic layers were washed with water (50 mL) and brine (50 mL) and dried using a phase separation paper. The filtrate was concentrated under reduced pressure. The resulting crude product was purified by FCC (EtOAc/hexanes 10:90) to give thioether **83** (540 mg, 1.51 mmol, 82 %) as a white solid.

R_f: 0.20 (EtOAc/hexanes 10:90). **m.p.:** 102 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.05 (d, *J* = 2.3 Hz, 1H, 5-H), 7.92 (dd, *J* = 8.1, 1.0 Hz, 1H, 4'-H), 7.76 (ddd, *J* = 8.1, 1.3, 0.7 Hz, 1H, 7'-H), 7.69 (dd, *J* = 8.5, 2.3 Hz, 1H, 7-H), 7.44 (ddd, *J* = 8.3, 7.2, 1.2 Hz, 1H, 5'-H), 7.31 (ddd, *J* = 8.4, 7.3, 1.2 Hz, 1H, 6'-H), 6.92 (d, *J* = 8.5 Hz, 1H, 8-H), 4.60 (s, 2H, Ar-C**H_2**), 1.72 (s, 7H, $C(CH_3)_2$). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 165.9 (C-2'), 160.9 (C-4), 155.6 (C-8a), 152.9 (C-3a'), 137.4 (C-7), 135.4 (C-7a'), 131.3 (C-6), 130.3 (C-5), 126.4 (C-5'), 124.7 (C-6'), 121.7 (C-4'), 121.2 (C-7'), 117.8 (C-8), 113.6 (C-4a), 106.7 (C-2), 36.7 (Ar-**C**H_2), 26.0 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 1733, 1618, 1494, 1457, 1427, 1298, 1202, 1050, 993, 843, 754, 724. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{18}H_{16}NO_3S_2^+$: 358.0566; found: 358.0568. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

6-((Benzo[*d*]thiazol-2-ylsulfonyl)methyl)-2,2-dimethyl-4*H*-benzo[*d*][1,3]dioxin-4-one (**84**) [6]

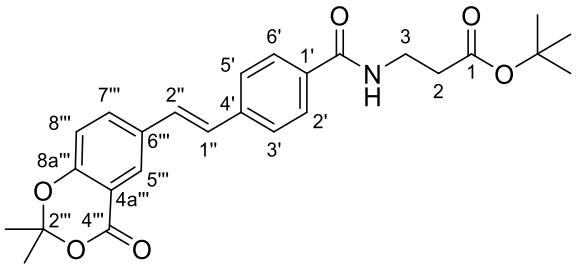


To a solution of thioether **83** (357 mg, 1.00 mmol) in MeOH (5.0 mL) was added Na_2WO_4 dihydrate (330 mg, 1.00 mmol) and 30 % H_2O_2 (1.70 mL, 15.0 mmol). The mixture was stirred at room temperature for 12 h. The solvent was then evaporated under reduced pressure and the residue resuspended in water (50 mL) and extracted with DCM (3 x 40 mL). The combined organic layers were dried using a phase separation paper and concentrated under reduced pressure to give product **84** (375 mg, 0.963 mmol. 96 %) as a white solid.

R_f: 0.39 (EtOAc/hexanes 35:65). **m.p.:** 187 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.26 (ddd, *J* = 8.3, 1.3, 0.7 Hz, 1H, 4'-H), 7.96 (ddd, *J* = 8.2, 1.3, 0.7 Hz, 1H, 7'-H), 7.84 (d, *J* = 2.2 Hz, 1H, 5-H), 7.66 (ddd, *J* = 8.4, 7.2, 1.3 Hz, 1H, 5'-H), 7.62 – 7.56 (m, 2H, 7-H and 6'-H), 6.94 (d, *J* = 8.5 Hz, 1H, 8-H), 4.74 (s, 2H, Ar-C**H_2**), 1.69 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 164.9 (C-2'), 160.2 (C-4), 156.8 (C-8a), 152.7 (C-3a'), 139.0 (C-7), 137.0 (C-7a'), 132.6 (C-5), 128.4 (C-6'), 128.0 (C-5'), 125.8 (C-4'), 122.4 (C-7'), 121.0 (C-6), 118.3 (C-8), 113.9 (C-4a), 106.9 (C-2), 60.1 (Ar-**C**H_2), 25.9 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 1729, 1620, 1494, 1302, 1202, 1147, 1056, 767, 619. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for

$C_{18}H_{16}NO_5S_2^+$: 390.0464; found: 390.0467. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

tert-Butyl (*E*/*Z*)-3-(4-(2-(2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)vinyl)benzamido)-propanoate (**85**-*E*/*Z*)



To a solution of sulfone **84** (850 mg, 2.18 mmol) in dry DMF (21 mL), sodium hydride (131 mg, 3.27 mmol) was added at 0 ºC under N_2 atmosphere. The solution turned reddish indicating the formation of carbanion. After 5 min of stirring, aldehyde **46** (605 mg, 2.18 mmol) was added. The temperature of the reaction was gradually raised from 0 °C to room temperature and stirring continued at this temperature for 16 h. The reaction mixture was quenched with sat. aq. NH_4Cl (15 mL). The aqueous phase was extracted with EtOAc (3 x 15 mL). The combined organic layers were washed with 10 % aq. NaOH (3 x 15 mL) to remove by-product hydroxybenzothiazole, followed by water (20 mL) and brine (20 mL). The solvent was then removed *in vacuo* and the crude product dissolved in MeOH (10 mL). To remove unreacted aldehyde, sat. $NaHSO_3$ (50 mL) was added to the solution, shaken for approx. 30 s, diluted with water (50 mL) and extracted with a solvent mixture of 10 % EtOAc in hexanes (3 x 50 mL). The obtained crude product was purified by FCC (EtOAc/hexanes 40:60) to give the *E*-isomer **85**-*E* (349 mg, 0.774 mmol, 35 %) as a white solid, the *Z*-isomer **85**-*Z* (24.9 mg, 0.0552 mmol, 3 %) as a pale yellow oil and a mixture of *E*- and Z-isomer **85**-*E*/*Z* as a pale yellow oil (154 mg, 0.341 mmol, 16 %).

E-Isomer:

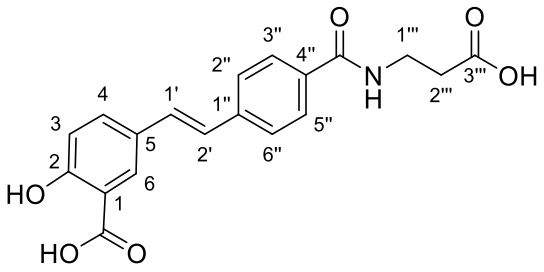
R_f: 0.27 (EtOAc/hexanes 40:60). **m.p.:** 166 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 8.12 (d, J = 2.2 Hz, 1H, 5'''-H), 7.79 – 7.75 (m, 2H, 2'-H and 6'-H), 7.70 (dd, J = 8.5, 2.3 Hz, 1H, 7'''-H), 7.56 – 7.52 (m, 2H, 3'-H and 5'-H), 7.13 (d, J = 16.4 Hz, 1H, 2''-H), 7.08 (d, J = 16.4 Hz, 1H, 1''-H), 6.98 (d, J = 8.5 Hz, 1H, 8'''-H), 6.89 (t, J = 6.0 Hz, 1H, CONH), 3.70 (q, J = 6.0 Hz, 2H, 3-H), 2.60 – 2.53 (m, 3H, 2-H), 1.76 (s, 6H, $C(CH_3)_2$), 1.47 (s, 9H, $C(CH_3)_3$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 172.5 (C-1), 166.8 (CONH), 161.2 (C-4'''), 155.7 (C-8a'''), 140.1 (C-4'), 134.6 (C-7'''), 133.6 (C-1'), 132.1 (C-6'''), 128.6 (C-2''), 128.2 (C-1''), 127.6 (C-2' and C-6'), 127.5 (C-5'''), 126.7 (C-3' and C-5'), 117.8 (C-8'''), 113.9 (C-4a'''), 106.8 (C-2'''), 81.4 ($\mathbf{C}(CH_3)_3$), 35.6 (C-3), 35.2 (C-2), 28.3 ($C(\mathbf{C}H_3)_2$), 26.0 ($C(\mathbf{C}H_3)_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3384,

2931, 1727, 1712, 1657, 1602, 1537, 1503, 1316, 1269, 1206, 1156, 1138, 1062, 945, 838. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{26}H_{30}NO_6^+$: 452.2068; found: 452.2069. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

Z-Isomer:

R_f: 0.34 (EtOAc/hexanes 40:60). **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 7.84 (d, *J* = 2.4 Hz, 1H, 5'''-H), 7.66 – 7.62 (m, 2H, 2'-H and 6'-H), 7.35 (dd, *J* = 8.6, 2.2 Hz, 1H, 7'''-H), 7.30 – 7.26 (m, 2H, 3' and 5'-H), 6.84 (t, *J* = 6.1 Hz, 1H, CONH), 6.76 (d, *J* = 8.5 Hz, 1H, 8'''-H), 6.64 (d, *J* = 12.3 Hz, 1H, 1''-H or 2''-H), 6.60 (d, *J* = 12.3 Hz, 1H, 1''-H or 2''-H), 3.68 (q, *J* = 6.0 Hz, 2H, 3-H), 2.59 – 2.51 (m, 2H, 2-H), 1.73 (s, 6H, $C(CH_3)_2$), 1.46 (s, 9H, $C(CH_3)_3$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 172.5 (C-1), 166.9 (CONH), 161.0 (C-4'''), 155.3 (C-8a'''), 140.2 (C-4'), 136.8 (C-7'''), 133.4 (C-1'), 131.6 (C-6'''), 130.3 (C-5'''), 130.0 (C-1''), 129.8 (C-2''), 129.1 (C-3' and C-5'), 127.2 (C-2' and C-6'), 117.2 (C-8'''), 113.6 (C-4a'''), 106.6 (C-2'''), 81.4 ($\mathbf{C}(CH_3)_3$), 35.6 (C-3), 35.2 (C-2), 28.3 ($C(\mathbf{C}H_3)_2$), 26.0 ($C(\mathbf{C}H_3)_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3336, 2979, 1726, 1642, 1614, 1493, 1276, 1203, 1154, 1049, 843, 779. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{26}H_{30}NO_6^+$: 452.2068; found: 452.2069. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

(*E*)-5-(4-((2-Carboxyethyl)carbamoyl)styryl)-2-hydroxybenzoic acid (**CG_111**)

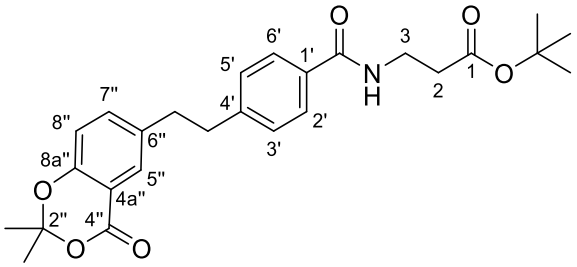


Prepared according to **General Procedure D** from compound **85**-*E* (80.0 mg, 0.177 mmol) and heating the reaction mixture to 60 °C. The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo*. The crude product was resuspended in 1 M aq. NaOH and 2 M aq. HCl was added until no more product precipitated. The crude product was collected by filtration and purified by FCC (MeOH/DCM+AcOH 5:95+1) to give stilbene **CG_111** (37.0 mg, 0.104 mmol, 59 %) as a white solid.

R_f: 0.19 (DCM/hexanes+AcOH 50:50+1). **m.p.:** 245 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 8.51 (s, 1H, CONH), 7.94 (d, *J* = 2.4 Hz, 1H, 6-H), 7.81 (d, *J* = 8.0 Hz, 2H, 3''-H and 5''-H), 7.59 (d, *J* = 8.0 Hz, 2H, 2''-H and 6''-H), 7.45 (dd, *J* = 8.4, 2.4 Hz, 1H, 4-H), 7.24 (d, *J* = 16.4 Hz, 1H, 1'-H), 6.98 (d, *J* = 16.3 Hz, 1H, 2'-H), 6.65 (d, *J* = 8.3 Hz, 1H, 3-H), 3.52 – 3.38

(m, 1H, 1'''-H), 2.35 – 2.23 (m, 3H, 2'''-H). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 177.0 (C-3'''), 171.2 (1-COOH), 165.6 (CONH), 161.6 (C-2), 140.5 (C-1''), 132.4 (C-4''), 130.8 (C-1'), 130.0 (C-4), 128.9 (C-6), 127.6 (C-3'' and C-5''), 125.5 (C-2'' and C-6''), 123.0 (C-2'), 120.2 (C-1), 116.6 (C-3), 36.6 (C-1'''), 36.5 (C-2'''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3296, 3021, 1678, 1634, 1440, 1330, 1291, 1203, 1085, 954, 838, 681. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{19}H_{16}NO_6^-$: 354.0983; found: 354.0984. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 1f).

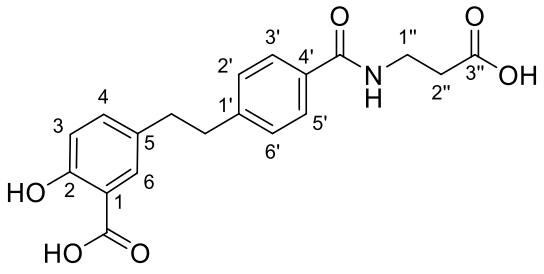
tert-Butyl 3-(4-(2-(2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)ethyl)benzamido)-propanoate (**86**)



Prepared according to **General Procedure A** from stilbene **85**-*E*/*Z* (100 mg, 0.221 mmol). Product **86** (72.2 mg, 0.159 mmol, 72 %) was obtained as a white solid.

R_f: 0.32 (DCM/hexanes+AcOH 50:50+1). **m.p.:** 102 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 7.78 (d, *J* = 2.3 Hz, 1H, 5''-H), 7.69 – 7.66 (m, 2H, 2'-H and 6'-H), 7.28 – 7.25 (m, 1H, 7''-H), 7.21 – 7.17 (m, 2H, 3'-H and 5'-H), 6.85 (d, *J* = 8.4 Hz, 1H, 8''-H), 6.84 – 6.81 (m, 1H, CONH), 3.68 (q, *J* = 6.0 Hz, 2H, 3-H), 2.98 – 2.93 (m, 2H, 6''-C**H_2**), 2.93 – 2.89 (m, 2H, 4'-C**H_2**), 2.58 – 2.53 (m, 2H, 2-H), 1.72 (s, 6H, $C(CH_3)_2$), 1.46 (s, 9H, $C(CH_3)_3$). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 172.5 (C-1), 167.2 (CONH), 161.5 (C-4''), 154.5 (C-8a''), 144.9 (C-4'), 136.9 (C-7''), 135.8 (C-6''), 132.6 (C-1'), 129.1 (C-5''), 128.8 (C-3' and C-5'), 127.2 (C-2' and C-6'), 117.2 (C-8''), 113.5 (C-4a''), 106.5 (C-2''), 81.4 ($\mathbf{C}(CH_3)_3$), 37.6 (4'-$\mathbf{C}H_2$), 36.7 (6''-$\mathbf{C}H_2$), 35.6 (C-3), 35.2 (C-2), 28.3 ($C(\mathbf{C}H_3)_3$), 25.9 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3396, 2929, 1712, 1657, 1494, 1308, 1255, 1210, 1160, 1058, 944, 848, 782. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{26}H_{32}NO_6^+$: 454.2224; found: 454.2224. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

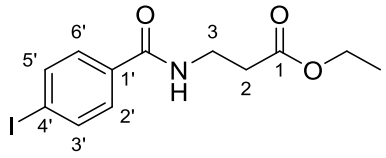
5-(4-((2-Carboxyethyl)carbamoyl)phenethyl)-2-hydroxybenzoic acid (**CG_142**)



Prepared according to **General Procedure D** from compound **86** (60.0 mg, 0.132 mmol) and heating the reaction mixture to reflux. The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo* to give the pure product **CG_142** (43.3 mg, 0.121 mmol, 92 %) as a white solid.

R_f: 0.34 (EtOAc/DCM+AcOH 50:50+1). **m.p.:** 200 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.21 (s, 1H, 1-COOH or 2''-COOH), 8.43 (t, *J* = 5.5 Hz, 1H, CONH), 7.76 – 7.70 (m, 2H, 3'-H and 5'-H), 7.62 (d, *J* = 2.3 Hz, 1H, 6-H), 7.35 (dd, *J* = 8.5, 2.4 Hz, 1H, 4-H), 7.31 – 7.26 (m, 2H, 2'-H and 6'-H), 6.85 (d, *J* = 8.4 Hz, 1H, 3-H), 3.44 (td, *J* = 7.1, 5.4 Hz, 2H, 1''-H), 2.92 – 2.86 (m, 2H, 1'-C**H$_2$**), 2.88 – 2.79 (m, 2H, 5-C**H$_2$**), 2.54 – 2.51 (m, 2H, 2''-H, collapses with DMSO). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (C-3''), 171.9 (1-COOH), 166.1 (CONH), 159.4 (C-2), 144.7 (C-1'), 135.9 (C-4), 132.0 (C-5 or C-4'), 131.8 (C-5 or C-4'), 129.6 (C-6), 128.3 (C-2' and C-6'), 127.1 (C-3' and C-5'), 116.9 (C-3), 112.6 (C-1), 36.9 (1'-**C**H_2), 35.6 (C-1''), 35.5 (5-**C**H_2), 33.8 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2924, 2855, 1722, 1666, 1630, 1442, 1331, 1291, 1196, 836, 674. **HR-MS (ESI):** *m/z* = $[M-H]^-$ calcd for $C_{19}H_{18}NO_6^-$: 356.1140; found: 356.1142. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

Ethyl 3-(4-iodobenzamido)propanoate (**88**) [11]

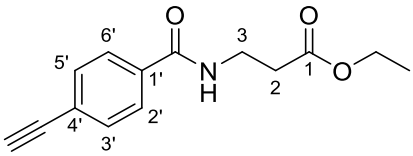


A solution of *β*-alanine ethyl ester hydrochloride (**76**, 1.56 g, 10.1 mmol) in dry pyridine (10 mL) was added to a suspension of 4-iodo-benzoylchloride (**87**, 2.00 g, 7.51 mmol) in dry pyridine (7.0 mL). The reaction mixture was stirred at room temperature for 18 h. The solvent was evaporated *in vacuo* and the residue was taken up in EtOAc (20 mL). The suspension was washed with 1 M aq. HCl (2 x 10 mL), then with 1 M aq. NaOH (2 x 10 mL) and dried using a phase separation paper. The solvent was evaporated *in vacuo* and product **88** (1.84 g, 5.31 mmol, 71 %) was obtained as an off-white solid.

R_f: 0.35 (EtOAc/hexanes 40:60). **m.p.:** 95 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 7.80 – 7.76 (m, 2H, 3'-H and 5'-H), 7.51 – 7.47 (m, 2H, 2'-H and 6'-H), 6.86 (t, *J* = 5.7 Hz, 1H, CONH), 4.17 (q, *J* = 7.1 Hz, 2H, C**H$_2$**CH_3), 3.71 (q, *J* = 6.1 Hz, 2H, 3-H), 2.63 (t, *J* = 5.9 Hz, 2H, 2-H), 1.27 (t, *J* = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 173.2 (C-1), 166.6 (CONH), 137.9 (C-3' and C-5'), 133.9 (C-1'), 128.7 (C-2' and C-6'), 98.6 (C-4'), 61.1 (**C**H_2CH_3), 35.5 (C-3), 33.9 (C-2), 14.3 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3298, 2935, 1723, 1634, 1585, 1538,

1323, 1180, 1151, 1077, 1006, 853, 838. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{12}H_{15}INO_3^+$: 348.0091; found: 348.0090. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

Ethyl 3-(4-ethynylbenzamido)propanoate (**91**)

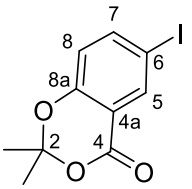


Iodo compound **88** (868 mg, 2.50 mmol) was added to Et_3N (25 mL) under N_2 atmosphere. The mixture was degassed with N_2 for 10 min. Copper(I) iodide (47.6 mg, 0.250 mmol) and $Pd(PPh_3)_2Cl_2$ (175 mg, 0.250 mmol) were added. Then, trimethylsilylacetylene (**89**, 707 µL, 5.00 mmol) was added dropwise over 3 min. The reaction mixture was stirred at 60 °C for 16 h. Then, the mixture was diluted with EtOAc (20 mL) and filtered through a pad of celite. The filtrate was washed with 1 M aq. HCl (10 mL), sat. aq. $NaHCO_3$ (10 mL) and brine (10 mL). The organic phase was dried using a phase separation paper and concentrated *in vacuo*. The crude TMS-protected product (**90**) was used without further purification in the next step (identification *via* ^{1}H-NMR and HRMS: $[M+H]^+$ calcd for $C_{17}H_{24}NO_3Si^+$: 318.1520; found: 318.1522).

To a solution of crude alkyne **90** (2.50 mmol) in THF (25 mL), TBAF (1.0 m in THF; 2.75 mL, 2.75 mmol) was added. The solution was stirred at room temperature for 15 min. After removal of the solvent *in vacuo*, the crude product was purified by FCC (EtOAc/hexanes 40:60) to give alkyne **91** (577 mg, 2.35 mmol, 94 % over two steps) as a light brown solid.

R_f: 0.31 (EtOAc/hexanes 40:60). **m.p.:** 55 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** *δ* (ppm) = 7.74 – 7.70 (m, 2H, 2'-H and 6'-H), 7.56 – 7.53 (m, 2H, 3'-H and 5'-H), 6.88 (t, *J* = 5.4 Hz, 1H, CONH), 4.18 (q, *J* = 7.2 Hz, 2H, C**H$_2$**CH$_3$), 3.72 (q, *J* = 6.0 Hz, 2H, 3-H), 3.19 (s, 1H, C≡C**H**), 2.64 (t, *J* = 5.9 Hz, 2H, 2-H), 1.28 (t, *J* = 7.2 Hz, 3H, CH_3). **^{13}C-NMR (101 MHz, $CDCl_3$):** *δ* (ppm) = 173.1 (C-1), 166.6 (CONH), 134.5 (C-1'), 132.4 (C-3' and C-5'), 127.1 (C-2' and C-6'), 125.5 (C-4'), 82.9 (**C**≡CH), 79.6 (C≡**C**H), 61.0 (**C**H$_2$CH$_3$), 35.5 (C-3), 34.0 (C-2), 14.3 (CH_3). **IR (ATR):** *ṽ* (cm^{-1}) = 3283, 3250, 2981, 2929, 1723, 1635, 1607, 1543, 1497, 1320, 1291, 1181, 1151, 1025, 852, 765, 664. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{14}H_{14}NO_3^-$: 244.0979; found: 244.0978. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

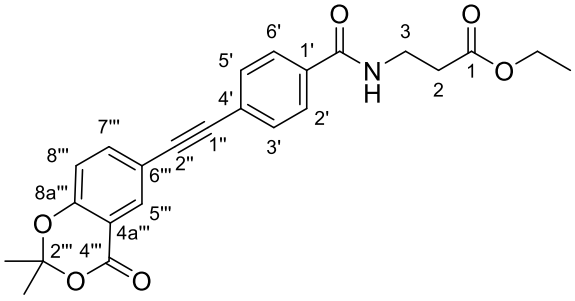
6-Iodo-2,2-dimethyl-4*H*-benzo[*d*][1,3]dioxin-4-one (**93**) [12]



5-Iodosalicylic acid (**92**, 2.00 g, 7.58 mmol) in TFA (15.9 mL, 212 mmol) was cooled to 0 °C under N_2 atmosphere. TFAA (10.0 mL, 72.0 mmol) and acetone (4.46 mL, 60.6 mmol) were added, and the solution was allowed to warm to room temperature and stirred for 48 h. The solution was concentrated *in vacuo* and the residue purified by FCC (EtOAc/hexanes 5:95) to give product **93** (895 mg, 2.94 mmol, 39 %) as a pale yellow solid.

R_f: 0.28 (EtOAc/hexanes 5:95). **m.p.:** 52 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 8.26 (d, *J* = 2.2 Hz, 1H, 5-H), 7.81 (dd, *J* = 8.6, 2.2 Hz, 1H, 7-H), 6.75 (d, *J* = 8.7 Hz, 1H, 8-H), 1.73 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 159.8 (C-4), 155.8 (C-8a), 145.0 (C-7), 138.3 (C-5), 119.6 (C-8), 115.7 (C-4a), 106.9 (C-2), 84.6 (C-6), 25.9 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3346, 2984, 2938, 1733, 1688, 1634, 1535, 1275, 1246, 1169, 1143, 856, 834, 662. **HR-MS (EI):** *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{10}H_9IO_3^{\bullet+}$: 303.9591; found: 303.9591. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

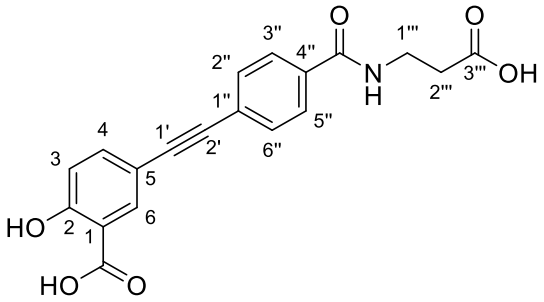
Ethyl 3-(4-((2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)ethynyl)benzamido)propanoate (**94**)



Iodo compound **93** (76.0 mg, 0.250 mmol) was added to triethylamine (2.5 mL) under N_2 atmosphere. The mixture was degassed with N_2 for 10 min. Copper(I) iodide (4.76 mg, 0.0250 mmol), $Pd(PPh_3)_2Cl_2$ (17.5 mg, 0.0250 mmol), and alkyne **91** (92.0 µL, 0.375 mmol) were added sequentially and the reaction mixture was stirred at 60 °C for 1 h. Then, the mixture was diluted with EtOAc (20 mL) and filtered through a pad of celite. The filtrate was washed with 1 M aq. HCl (10 mL), sat. aq. $NaHCO_3$ (10 mL) and brine (10 mL). The organic phase was dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 60:40) to give alkyne **94** (84.5 mg, 0.200 mmol, 80 %) as an off-white solid.

R_f: 0.24 (EtOAc/hexanes 40:60). **m.p.:** 132 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.15 (d, *J* = 2.1 Hz, 1H, 5'''-H), 7.77 – 7.74 (m, 2H, 2'-H and 6'-H), 7.69 (dd, *J* = 8.5, 2.1 Hz, 1H, 7'''-H), 7.58 – 7.55 (m, 2H, 3'-H and 5'-H), 6.97 (d, *J* = 8.5 Hz, 1H, 8'''-H), 6.90 (t, *J* = 6.0 Hz, 1H, CONH), 4.18 (q, *J* = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 3.74 (q, *J* = 6.0 Hz, 2H, 3-H), 2.65 (t, *J* = 5.9 Hz, 2H, 2-H), 1.75 (s, 6H, C(CH$_3$)$_2$), 1.28 (t, *J* = 7.1 Hz, 4H, CH$_2$C**H$_3$**). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 173.2 (C-1), 166.6 (CONH), 160.4 (C-4'''), 156.1 (C-8a'''), 139.4 (C-7'''), 134.1 (C-1'), 133.2 (C-5'''), 131.9 (C-3' and C-5'), 127.2 (C-2' and C-6'), 126.3 (C-4'), 117.8 (C-8'''), 117.7 (C-6'''), 113.9 (C-4a'''), 107.0 (C-2'''), 89.9 (C-2''), 89.0 (C-1''), 61.0 (**C**H$_2$CH$_3$), 35.5 (C-3), 34.0 (C-2), 26.0 (C(**C**H$_3$)$_2$), 14.3 (CH$_2$**C**H$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3390, 1725, 1641, 1537, 1505, 1290, 1180, 1159, 1142, 1058, 1031, 941, 849, 762, 688. **HR-MS (ESI):** *m*/*z* = [M+H]$^+$ calcd for $C_{24}H_{24}NO_6^+$: 422.1598; found: 422.1594. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

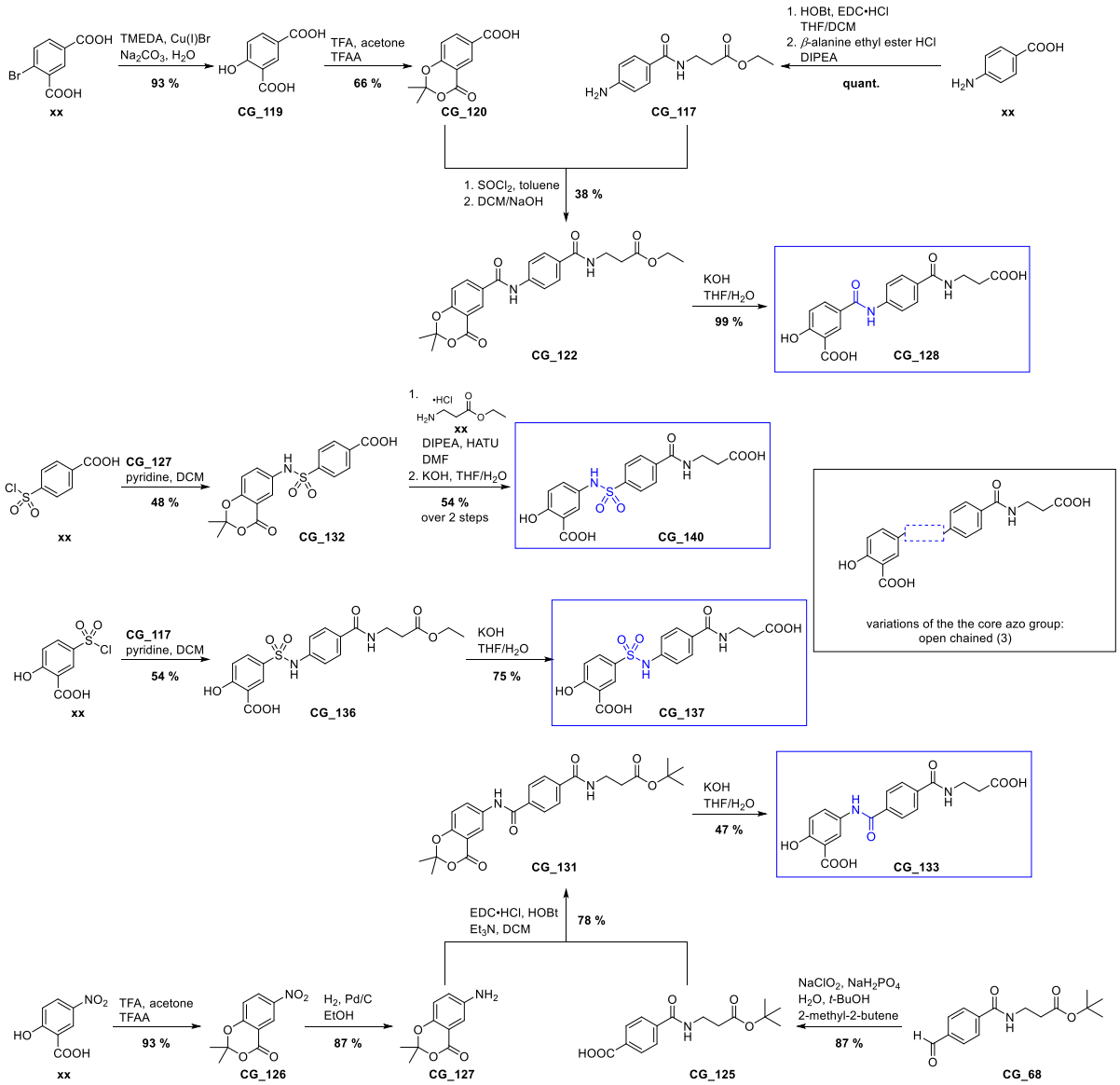
5-((4-((2-Carboxyethyl)carbamoyl)phenyl)ethynyl)-2-hydroxybenzoic acid (**CG_219**)



Prepared according to **General Procedure D** at room temperature from alkyne **94** (50.0 mg, 0.119 mmol). The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo* to give the pure product **CG_219** (30.3 mg, 0.0858 mmol, 72 %) as a white solid.

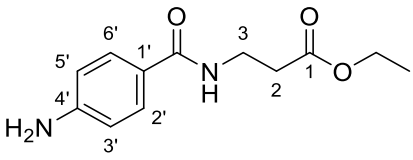
R_f: 0.45 (EtOAc/hexanes+AcOH 80:20+1). **m.p.:** 216 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.27 (s, 2H, 1-COOH and 2'''-COOH), 8.62 (t, *J* = 5.5 Hz, 1H, CONH), 7.96 (d, *J* = 2.3 Hz, 1H, 6-H), 7.89 – 7.84 (m, 2H, 3''-H and 5''-H), 7.68 (dd, *J* = 8.6, 2.2 Hz, 1H, 4-H), 7.65 – 7.60 (m, 2H, 2''-H and 6''-H), 7.01 (d, *J* = 8.6 Hz, 1H, 3-H), 3.46 (td, *J* = 7.1, 5.4 Hz, 2H, 1'''-H), 2.55 – 2.51 (m, 2H, 2'''-H). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (C-3'''), 170.9 (1-COOH), 165.5 (CONH), 161.5 (C-2), 138.0 (C-4), 133.9 (C-4''), 133.7 (C-6), 131.1 (C-2'' and C-6''), 127.5 (C-3'' and C-5''), 125.1 (C-1''), 118.0 (C-3), 114.1 (C-5), 112.5 (C-1), 90.6 (C-1'), 87.5 (C-2'), 35.6 (C-1'''), 33.7 (C-2'''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3292, 2921, 2207, 1694, 1667, 1628, 1538, 1505, 1421, 1325, 1294, 1200, 1162, 1086, 848, 833, 765, 672. **HR-MS (ESI):** *m*/*z* = [M-H]$^-$ calcd for $C_{19}H_{14}NO_6^-$: 352.0827; found: 352.0825. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

Synthesis of amides CG_128 and CG_133, and sulfonamides CG_137 and CG_140



Scheme S3: Preparation of amides **CG_128** and **CG_133**, and sulfonamides **CG_137** and **CG_140**.

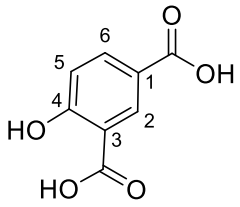
Ethyl 3-(4-aminobenzamido)propanoate (**95**) [8]



Prepared according to **General Procedure F** from 4-aminobenzoic acid (**14**, 500 mg, 3.65 mmol) and *β*-alanine ethyl ester hydrochloride (**76**, 560 mg, 3.65 mmol). The crude product was purified by FCC (EtOAc/hexanes 70:30) to give amide **95** (857 mg, 3.63 mmol, 99 %) as a beige solid.

R_f: 0.25 (EtOAc/hexanes 70:30). **m.p.:** 114 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 8.05 (t, *J* = 5.5 Hz, 1H, CONH), 7.57 – 7.51 (m, 2H, 2'-H and 6'-H), 6.55 – 6.51 (m, 2H, 3'-H and 5'-H), 5.63 (s, 2H, NH_2), 4.05 (q, *J* = 7.1 Hz, 2H, C**H_2**CH_3), 3.42 (td, *J* = 7.1, 5.5 Hz, 2H, 3-H), 2.55 – 2.51 (m, 3H, 2-H), 1.17 (t, *J* = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 171.4 (C-1), 166.3 (CONH), 151.4 (C-4'), 128.7 (C-2' and C-6'), 121.2 (C-1'), 112.6 (C-3' and C-5'), 59.9 (**C**H_2CH_3), 35.3 (C-3), 34.1 (C-2), 14.1 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3454, 3345, 3240, 1720, 1594, 1512, 1308, 1258, 1181, 1034, 842, 690. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{12}H_{17}N_2O_3^+$: 237.1234; found: 237.1236. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

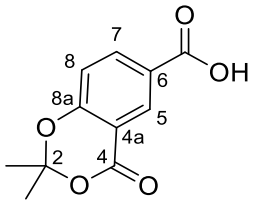
4-Hydroxyisophthalic acid (**97**) [13]



4-Bromoisophthalic acid (**96**, 1.00 g, 4.10 mmol) was dissolved in water (4.0 mL) and Na_2CO_3 (1.20 g, 11.0 mmol) was added. The resulting solution was stirred for 1.5 h at 85 °C.
Meanwhile TMEDA (31.0 mg, 270 µmol) and copper(I) bromide (18.0 mg, 126 µmol) were dissolved in water (0.50 mL) and stirred for 1 h.
The two solutions were then mixed and stirred for 18 h at 85 °C, cooled to room temperature and acidified with 1 M aq. HCl. The resulting precipitate was collected by filtration and dried *in vacuo* to give 4-hydroxyisophthalic acid (**97**, 695 mg, 3.81 mmol, 93 %) as a white solid.

R_f: 0.31 (EtOAc/hexanes+AcOH 30:70+1). **m.p.:** 303 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 12.95 (s, 1H, 1-COOH or 3-COOH), 11.86 (s, 1H, 1-COOH or 3-COOH), 10.23 (s, 1H, OH), 8.38 (d, *J* = 2.3 Hz, 1H, 2-H), 8.03 (dd, *J* = 8.7, 2.3 Hz, 1H, 6-H), 7.04 (d, *J* = 8.7 Hz, 1H, 5-H). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 171.2 (3-COOH), 166.3 (1-COOH), 164.4 (C-4), 136.2 (C-6), 132.3 (C-2), 121.7 (C-1), 117.5 (C-5), 113.1 (C-3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2993, 1666, 1584, 1444, 1414, 1290, 1198, 897, 847, 770, 691. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_8H_5O_5^-$: 181.0142; found: 181.0141. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

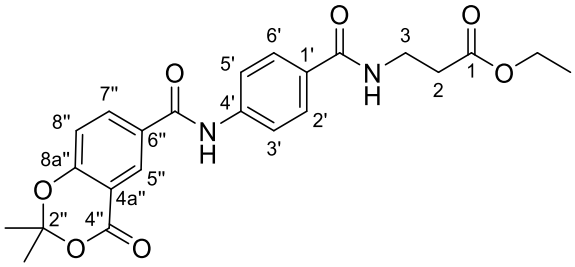
2,2-Dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxine-6-carboxylic acid (**98**) [5]



A suspension of 4-hydroxyisophthalic acid (**97**, 500 mg, 2.75 mmol) in acetone (1.01 mL, 13.7 mmol), TFA (3.09 mL, 41.2 mmol) and TFAA (1.03 mL, 7.41 mmol) was heated to 100 °C for 24 h. The reaction mixture was then cooled to room temperature and concentrated under reduced pressure. The residue was taken up with 1 M aq. HCl (15 mL) and extracted with EtOAc (3 x 20 mL). The combined organic layers were dried using a phase separation paper and the solvents were removed under reduced pressure. The crude product was purified by FCC (EtOAc/hexanes 20:80 → 40:60) to give acetonide **98** (404 mg, 1.82 mmol, 66 %) as a yellow solid.

R_f: 0.35 (EtOAc/hexanes 25:75). **m.p.:** 228 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 13.23 (s, 1H, COOH), 8.38 (d, *J* = 2.1 Hz, 1H, 5-H), 8.19 (dd, *J* = 8.6, 2.2 Hz, 1H, 7-H), 7.23 (d, *J* = 8.6 Hz, 1H, 8-H), 1.72 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 165.8 (COOH), 159.6 (C-4), 158.6 (C-8a), 137.6 (C-7), 130.6 (C-5), 125.6 (C-6), 117.9 (C-8), 112.8 (C-4a), 107.1 (C-2), 25.3 ($C(\mathbf{C}H_3)_2$).**IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2538, 1743, 1680, 1613, 1422, 1381, 1276, 1198, 1132, 927, 769. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{11}H_9O_5^-$: 221.0455; found: 221.0454. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

Ethyl 3-(4-(2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxine-6-carboxamido)benzamido)-propanoate (**99**)

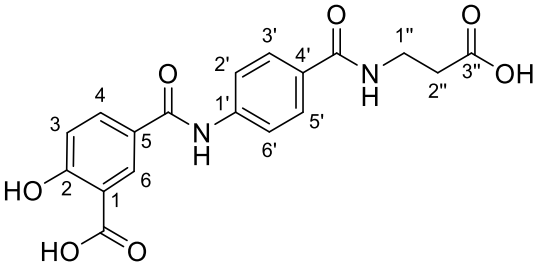


To a solution of benzoic acid **98** (100 mg, 0.450 mmol) in toluene (2.0 mL), $SOCl_2$ (164 µL, 2.25 mmol) was added. The reaction mixture was heated to reflux for 3 h. The solvent and excess of reagent were then removed under reduced pressure and the resulting crude acid chloride was dissolved in DCM (2.5 mL) and added to a flask containing amine **95** (106 mg, 0.45 mmol) in DCM (14 mL) and 10 % aq. NaOH (4.5 mL). The mixture was stirred for 1.5 h. Then the layers were separated, and the aqueous layer extracted with DCM (3 x 40 mL). The combined organic layers were washed with water (15 mL) and dried using a phase separation

paper. The solvent was removed *in vacuo* and the crude product purified by FCC (EtOAc/hexanes 40:60 → 80:20) to give amide **99** (75.2 mg, 0.171 mmol, 38 %) as a yellow solid.

R_f: 0.31 (EtOAc/hexanes 60:40). **m.p.:** 227 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 10.60 (s, 1H, 6''-CONH), 8.57 (d, *J* = 2.3 Hz, 1H, 5''-H), 8.48 (t, *J* = 5.5 Hz, 1H, 1'-CONH), 8.29 (dd, *J* = 8.7, 2.3 Hz, 1H, 7''-H), 7.89 – 7.81 (m, 4H, 2'-, 3'-, 5'- and 6'-H), 7.30 (d, *J* = 8.7 Hz, 1H, 8''-H), 4.07 (q, *J* = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 3.49 (q, *J* = 6.8 Hz, 2H, 3'-H), 2.58 (t, *J* = 7.0 Hz, 2H, 2-H), 1.74 (s, 6H, C(CH$_3$)$_2$), 1.18 (t, *J* = 7.1 Hz, 3H, CH$_2$C**H$_3$**). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 171.3 (C-1), 165.8 (1'-CONH), 163.9 (6''-CONH), 159.8 (C-4''), 157.9 (C-8a''), 141.6 (C-1'), 136.6 (C-7''), 129.4 (C-6''), 129.1 (C-4'), 129.0 (C-5''), 127.9 (C-2' and C-6'), 119.5 (C-3' and C-5'), 117.7 (C-8''), 112.6 (C-4a''), 107.0 (C-2''), 59.9 (**C**H$_2$CH$_3$), 35.5 (C-3), 33.9 (C-2), 25.3 (C(**C**H$_3$)$_2$), 14.1 (CH$_2$**C**H$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3403, 3332, 2924, 1736, 1721, 1640, 1536, 1504, 1258, 1182, 844, 762. **HR-MS (ESI):** *m*/*z* = [M+H]$^+$ calcd for $C_{23}H_{25}N_2O_7^+$: 441.1656; found: 441.1513. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

5-((4-((2-Carboxyethyl)carbamoyl)phenyl)carbamoyl)-2-hydroxybenzoic acid (**CG_128**)

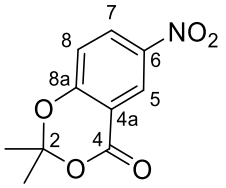


Prepared according to **General Procedure D** from compound **99** (50.0 mg, 0.114 mmol) and heating the reaction mixture to reflux. The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo*. The crude product was recrystallised from MeOH to give amide **CG_128** (42.0 mg, 0.113 mmol, 99 %) as a white solid.

R_f: 0.22 (EtOAc/hexanes+AcOH 80:20+1). **m.p.:** 280 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.22 (s, 2H, 1-COOH and 2''-COOH), 10.43 (s, 1H, 5-CONH), 8.50 (d, *J* = 2.4 Hz, 1H, 6-H), 8.44 (t, *J* = 5.5 Hz, 1H, 4'-CONH), 8.14 (dd, *J* = 8.7, 2.4 Hz, 1H, 4-H), 7.88 – 7.80 (m, 4H, 2'-, 3'-, 5'- and 6'-H), 7.10 (d, *J* = 8.7 Hz, 1H, 3-H), 3.45 (td, *J* = 7.1, 5.4 Hz, 2H, 1''-H), 2.54 – 2.51 (m, 2H, 2''-H collapses with DMSO). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (C-3''), 171.4 (1-COOH), 165.7 (4'-CONH), 164.4 (C-2), 163.7 (5-CONH), 141.8 (C-4'), 134.9 (C-4), 130.6 (C-6), 129.2 (C-1'), 127.8 (C-3' and C-5'), 125.4 (C-5), 119.4 (C-2' and C-6'), 117.3 (C-3), 113.0 (C-1), 35.6 (C-1''), 33.9 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3302, 1694, 1679,

1650, 1519, 1440, 1331, 1223, 1200, 841, 758, 658. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{18}H_{15}N_2O_7^-$: 371.0885; found: 371.0886. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 1f).

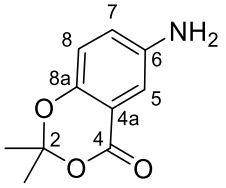
2,2-Dimethyl-6-nitro-4*H*-benzo[*d*][1,3]dioxin-4-one (**101**)



Prepared according to **General Procedure C** from 5-nitrosalicylic acid (**100**, 1.00 g, 5.46 mmol). The crude product was purified by FCC (EtOAc/hexanes 10:90) and acetonide **101** (1.13 g, 5.08 mmol, 93 %) was obtained as a pale yellow solid.

R_f: 0.16 (EtOAc/hexanes 10:90). **m.p.:** 94 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 8.88 (d, *J* = 2.8 Hz, 1H, 5-H), 8.43 (dd, *J* = 9.0, 2.8 Hz, 1H, 7-H), 7.13 (d, *J* = 9.1 Hz, 1H, 8-H), 1.79 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 160.4 (C-8a), 159.1 (C-4), 143.0 (C-6), 131.4 (C-7), 126.2 (C-5), 118.6 (C-8), 113.6 (C-4a), 107.9 (C-2), 26.1 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2990, 1742, 1592, 1531, 1477, 1336, 1280, 1194, 1050, 924, 746. **HR-MS (EI):** *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{10}H_9NO_5^{\bullet+}$: 223.0475; found: 223.0473. **Purity (HPLC):** 210 nm: 88 %; 254 nm: 92 % (method 3a).

6-Amino-2,2-dimethyl-4*H*-benzo[*d*][1,3]dioxin-4-one (**102**) [14]

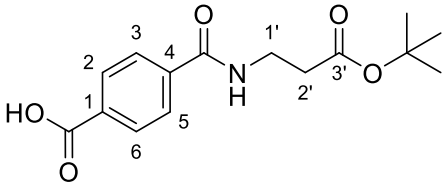


To a solution of nitro compound **101** (1.00 g, 4.48 mmol) in EtOH (7.5 mL) was added 10 % Pd/C (477 mg, 0.448 mmol) under N_2 atmosphere. Hydrogenation was performed under 1 bar H_2 pressure at room temperature for 12 h. Subsequently, the reaction mixture was filtered through a pad of celite, washed with EtOH (2 x 10 mL) and THF (1 x 10 mL) and concentrated *in vacuo* to give amine **102** (749 mg, 3.88 mmol, 87 %) as a yellow solid.

R_f: 0.45 (EtOAc/hexanes 50:50). **m.p.:** 161 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 7.31 (d, *J* = 2.8 Hz, 1H, 5-H), 6.98 (dd, *J* = 8.7, 2.8 Hz, 1H, 7-H), 6.80 (d, *J* = 8.6 Hz, 1H, 8-H), 4.07 (s, 3H, NH_2), 1.70 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 161.6 (C-4), 149.5 (C-8a), 140.4 (C-6), 124.7 (C-7), 118.2 (C-8), 115.1 (C-5), 114.2 (C-4a), 106.5 (C-2), 25.8 ($C(\mathbf{C}H_3)_2$).**IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3469, 3371, 1711, 1494, 1324, 1275, 1198, 1048, 836. **HR-MS**

(EI): *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{10}H_{11}NO_3^{\bullet+}$: 193.0733; found: 193.0732. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

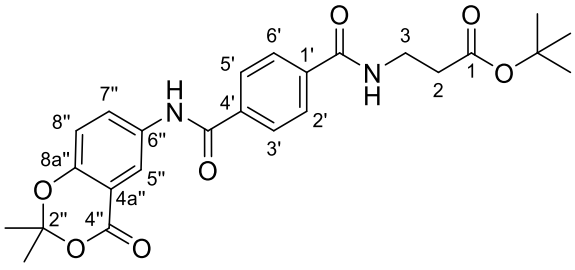
4-((3-(*tert*-Butoxy)-3-oxopropyl)carbamoyl)benzoic acid (**103**)



To a solution of aldehyde **46** (200 mg, 0.721 mmol) in *tert*-butylalcohol (0.93 mL), 2-methyl-2-butene (0.50 mL) was added. To this, a solution containing $NaClO_2$ (80 %, 106 mg, 0.938 mmol) and NaH_2PO_4 monohydrate (129 mg, 0.938 mmol) in water (0.74 mL) was added dropwise. The reaction mixture was stirred at room temperature for 4 h and then concentrated *in vacuo*, resuspended in water (10 mL), and extracted with hexanes (2 x 10 mL). The aqueous phase was acidified to pH 1 – 2 with 2 M aq. HCl, saturated with sodium chloride, and extracted with ether (3 x 15 mL). The combined organic extracts were dried using a phase separation paper and concentrated *in vacuo* to give benzoic acid **103** (182 mg, 0.622 mmol, 86 %) as a white solid.

R_f: 0.29 (EtOAc/hexanes 50:50). **m.p.:** 164 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 13.18 (s, 1H, COOH), 8.69 (t, *J* = 5.6 Hz, 1H, CONH), 8.07 – 7.96 (m, 2H, 2-H and 6-H), 7.94 – 7.87 (m, 2H, 3-H and 5-H), 3.46 (td, J = 7.0, 5.4 Hz, 2H, 1'-H), 2.50 – 2.46 (m, 2H, 2'-H, collapses with DMSO), 1.39 (s, 9H, C(CH_3)$_3$). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 170.6 (C-3'), 166.8 (COOH), 165.5 (CONH), 138.2 (C-4), 133.0 (C-1), 129.3 (C-2 and C-6), 127.4 (C-3 and C-5), 79.9 (**C**(CH_3)$_3$), 35.7 (C-1'), 34.9 (C-2'), 27.7 (C(**C**H_3)$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3324, 2980, 1725, 1688, 1640, 1537, 1432, 1276, 1152, 1074, 875, 690. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{15}H_{18}NO_5^-$: 292.1190; found: 292.1190. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

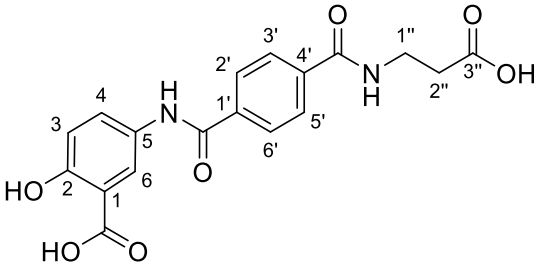
tert-Butyl 3-(4-((2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)carbamoyl)benzamido)-propanoate (**104**)



Prepared according to **General Procedure G** from benzoic acid **103** (100 mg, 0.341 mmol) and amine **102** (79.0 mg, 0.409 mmol). The crude product was purified by FCC (EtOAc/hexanes+AcOH 50:50+1) to give amide **104** (125 mg, 0.267 mmol, 78 %) as a white solid.

R_f: 0.29 (EtOAc/hexanes+AcOH 50:50+1). **m.p.:** 213 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 10.54 (s, 1H, 4'-CONH), 8.70 (t, *J* = 5.6 Hz, 1H, 1'-CONH), 8.39 (d, *J* = 2.6 Hz, 1H, 5''-H), 8.08 – 8.03 (m, 3H, 3'-H and 5'-H and 7''-H), 7.99 – 7.94 (m, 2H, 2'-H and 6'-H), 7.17 (d, *J* = 8.9 Hz, 1H, 8''-H), 3.49 (td, *J* = 7.0, 5.5 Hz, 2H, 3-H), 2.55 – 2.52 (m, 1H, 2-H), 1.71 (s, 6H, $C(CH_3)_2$), 1.40 (s, 9H, $C(CH_3)_3$). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 170.6 (C-1), 165.5 (1'-CONH), 164.9 (4'-CONH), 160.2 (C-4''), 151.5 (C-8a''), 137.1 (C-1'), 136.7 (C-4'), 134.2 (C-6''), 129.2 (C-7''), 127.7 (C-3' and C-5'), 127.2 (C-2' and C-6'), 120.0 (C-5''), 117.6 (C-8''), 112.9 (C-4a''), 106.5 (C-2''), 79.9 ($\mathbf{C}(CH_3)_3$), 35.7 (C-3), 34.9 (C-2), 27.7 ($C(\mathbf{C}H_3)_3$), 25.2 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3370, 2981, 1751, 1722, 1644, 1538, 1494, 1305, 1283, 1257, 1199, 1137, 832. **HR-MS (ESI):** *m/z* = $[M-H]^-$ calcd for $C_{25}H_{27}N_2O_7^-$: 467.1824; found: 467.1825. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

5-(4-((2-Carboxyethyl)carbamoyl)benzamido)-2-hydroxybenzoic acid (**CG_133**)

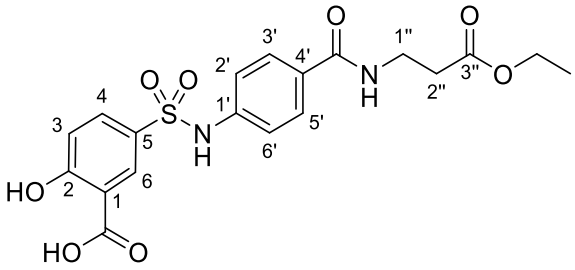


Amide **104** (80.0 mg, 0.171 mmol) was dissolved in THF (1.2 mL) followed by the addition of a solution of KOH (47.9 mg, 0.854 mmol) in water (1.2 mL). The reaction mixture was heated to reflux for 2.5 h, cooled to room temperature and the solvents removed *in vacuo*. To the residue was added water (3.0 mL) and 2 M aq. HCl until no further product precipitated. The solid was collected by filtration, washed with 1 M aq. HCl (5.0 mL) and water (5.0 mL), and dried to give product **CG_133** (30.1 mg, 0.0807 mmol, 47 %) as a white solid.

R_f: 0.14 (EtOAc/DCM+AcOH 50:50+1). **m.p.:** 286 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 14.02 (s, 1H, 1-COOH or 2''-COOH or 2-OH), 12.28 (s, 1H, 1-COOH or 2''-COOH or 2-OH), 11.08 (s, 1H, 1-COOH or 2''-COOH or 2-OH), 10.33 (s, 1H, 1'-CONH), 8.69 (t, *J* = 5.5 Hz, 1H, 4'-CONH), 8.28 (d, *J* = 2.7 Hz, 1H, 6-H), 8.05 – 8.01 (m, 2H, 2'-H and 6'-H), 7.97 – 7.94 (m, 2H, 3'-H and 5'-H), 7.89 (dd, *J* = 9.0, 2.7 Hz, 1H, 4-H), 6.98 (d, *J* = 8.9 Hz, 1H, 3-H), 3.49 (td, *J* = 7.1, 5.4 Hz, 2H, 1''-H), 2.54 (t, *J* = 7.1 Hz, 2H, 2''-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (C-3''), 171.7 (1-COOH), 165.5 (4'-CONH), 164.5 (1'-CONH), 157.5 (C-2),

136.9 (C-1' or C-4'), 136.8 (C-1' or C-4'), 130.7 (C-5), 128.7 (C-4), 127.6 (C-2' and C-6'), 127.2 (C-3' and C-5'), 121.9 (C-6), 117.1 (C-3), 112.4 (C-1), 35.7 (C-1''), 33.7 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3290, 1677, 1633, 1547, 1451, 1327, 1280, 1196, 860, 684. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{18}H_{15}N_2O_7^-$: 371.0885; found: 371.0886. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

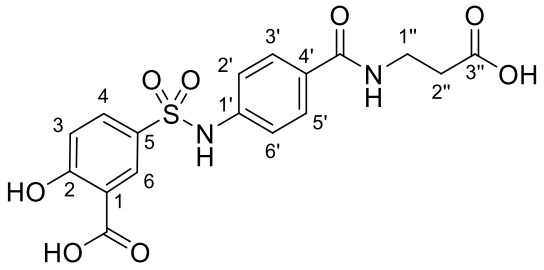
5-(*N*-(4-((3-Ethoxy-3-oxopropyl)carbamoyl)phenyl)sulfamoyl)-2-hydroxybenzoic acid (**106**)



Prepared according to **General Procedure H** from amine **95** (236 mg, 1.00 mmol). The crude product was purified by FCC (EtOAc/hexanes 70:30 → MeOH/EtOAc 5:95) to give sulfonamide **106** (237 mg, 0.544 mmol, 54 %) as an off-white solid.

R_f: 0.29 (EtOAc/DCM+AcOH 50:50+1). **m.p.:** 118 °C. **^{1}H-NMR (500 MHz, CD_3OD):** δ (ppm) = 8.33 (d, *J* = 2.5 Hz, 1H, 6-H), 7.73 (dd, *J* = 8.8, 2.5 Hz, 1H, 4-H), 7.67 – 7.63 (m, 2H, 3'-H and 5'-H), 7.19 – 7.16 (m, 2H, 2'-H and 6'-H), 6.88 (d, *J* = 8.8 Hz, 1H, 3-H), 4.12 (q, *J* = 7.2 Hz, 2H, C**H_2**CH_3), 3.58 (t, *J* = 6.9 Hz, 2H, 1''-H), 2.60 (t, *J* = 6.8 Hz, 2H, 2''-H), 1.22 (t, *J* = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (126 MHz, CD_3OD):** δ (ppm) = 173.6 (1-COOH), 173.5 (C-3''), 169.7 (CONH), 166.9 (C-2), 142.6 (C-1'), 133.3 (C-4), 131.8 (C-6), 130.8 (C-4'), 129.9 (C-5), 129.5 (C-3' and C-5'), 120.3 (C-2' and C-6'), 118.5 (C-3), 118.2 (C-1), 61.7 (**C**H_2CH_3), 37.0 (C-1''), 34.9 (C-2''), 14.5 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2926, 1726, 1625, 1610, 1568, 1335, 1312, 1198, 1157, 1107, 924, 842, 767, 656. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{19}H_{19}N_2O_8S^-$: 435.0868; found: 435.0865. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 1f).

5-(*N*-(4-((2-Carboxyethyl)carbamoyl)phenyl)sulfamoyl)-2-hydroxybenzoic acid (**CG_137**)

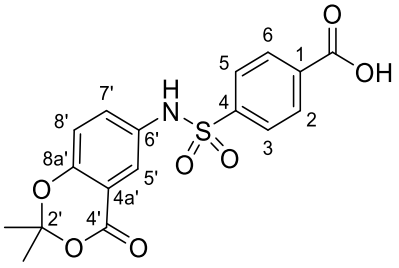


Prepared according to **General Procedure D** from sulfonamide **106** (100 mg, 0.229 mmol) and heating the reaction mixture to reflux. The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase

separation paper and concentrated *in vacuo*. The pure product **CG_137** (70.4 mg, 0.172 mmol, 75 %) was obtained as a pale beige solid.

R_f: 0.16 (EtOAc/hexanes+AcOH 80:20+1), streaks. **m.p.:** 228 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.18 (s, 1H, 1-COOH or 2''-COOH), 10.53 (s, 1H, OH), 8.37 (t, *J* = 5.5 Hz, 1H, CONH), 8.19 (d, *J* = 2.5 Hz, 1H, 6-H), 7.80 (dd, *J* = 8.8, 2.5 Hz, 1H, 4-H), 7.72 – 7.67 (m, 2H, 3'-H and 5'-H), 7.16 – 7.11 (m, 2H, 2'-H and 6'-H), 7.04 (d, *J* = 8.8 Hz, 1H, 3-H), 3.39 (td, *J* = 7.2, 5.5 Hz, 2H, 1''-H), 2.46 (t, J = 7.1 Hz, 2H, 2''-H). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (C-3''), 170.2 (1-COOH), 165.5 (CONH), 164.5 (C-2), 140.3 (C-1'), 133.0 (C-4), 129.8 (C-6), 129.6 (C-4'), 129.2 (C-5), 128.4 (C-3' and C-5'), 118.5 (C-2' and C-6'), 118.3 (C-3), 114.3 (C-1), 35.5 (C-1''), 33.8 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3401, 3146, 2924, 1693, 1634, 1609, 1504, 1334, 1303, 1208, 1159, 1109, 932, 725. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{17}H_{15}N_2O_8S^-$: 407.0555; found: 407.0553. **Purity (HPLC):** 210 nm: >95 %; 254 nm: 95 % (method 3a).

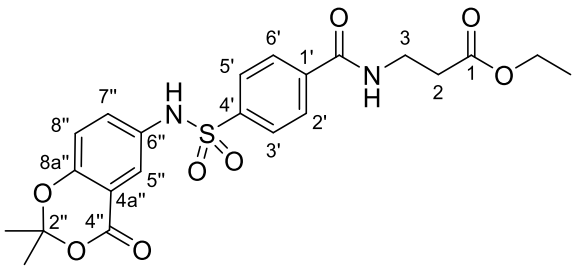
4-(*N*-(2,2-Dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)sulfamoyl)benzoic acid (**108**)



To a solution of amine **102** (300 mg, 1.55 mmol) in DCM (15 mL) was added pyridine (188 µL, 2.33 mmol) and 4-(chlorosulfonyl)benzoic acid (**107**, 343 mg, 1.55 mmol) at 0 °C under N_2 atmosphere. The mixture was stirred for 2 h at 0 °C and then allowed to warm to room temperature. The solvent was removed *in vacuo* and the crude product purified by recrystallisation from hot MeOH to give sulfonamide **108** (125 mg, 0.267 mmol, 78 %) as a white solid.

R_f: 0.67 (EtOAc/DCM+AcOH 50:50+1). **m.p.:** 260 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.14 – 8.09 (m, 2H, 2-H and 6-H), 7.84 – 7.80 (m, 2H, 3-H and 5-H), 7.57 (d, *J* = 2.7 Hz, 1H, 5'-H), 7.37 (dd, *J* = 8.8, 2.7 Hz, 1H, 7'-H), 6.95 (d, *J* = 8.8 Hz, 1H, 8'-H), 1.67 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 168.0 (COOH), 162.0 (C-4'), 154.9 (C-8a'), 144.5 (C-4), 136.1 (C-1), 133.6 (C-6'), 132.3 (C-7'), 131.4 (C-2 and C-6), 128.3 (C-3 and C-5), 123.5 (C-5'), 119.5 (C-8'), 115.0 (C-4a'), 108.1 (C-2'), 25.7 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3270, 1740, 1681, 1493, 1392, 1287, 1162, 1127, 899, 864, 726. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{17}H_{14}NO_7S^-$: 376.0496; found: 376.0498. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

Ethyl 3-(4-(*N*-(2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)sulfamoyl)benzamido)-propanoate (**109**)



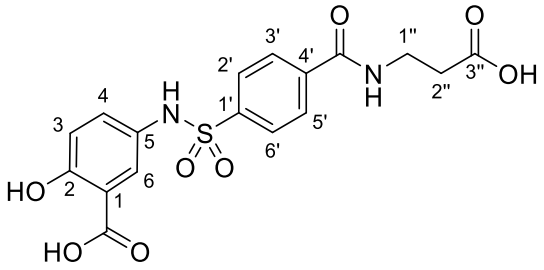
Prepared according to **General Procedure E** from benzoic acid **108** (200 mg, 0.530 mmol) and *β*-alanine ethyl ester hydrochloride (**76**; 81.4 mg, 0.530 mmol). The crude product was purified by FCC (EtOAc/hexanes 70:30) to give product **109** (136 mg, 0.284 mmol, 54 %) as a pale yellow crystalline solid.

R_f: 0.62 (EtOAc/hexanes 70:30).

m.p.: 57 °C.

^{1}H-NMR (400 MHz, $(CD_3)_2SO$): δ (ppm) = 10.47 (s, 1H, SO_2NH), 8.73 (t, J = 5.4 Hz, 1H, CONH), 7.95 – 7.90 (m, 2H, 2'-H and 6'-H), 7.81 – 7.76 (m, 2H, 3'-H and 5'-H), 7.52 (d, J = 2.7 Hz, 1H, 5''-H), 7.36 (dd, J = 8.8, 2.7 Hz, 1H, 7''-H), 7.04 (d, J = 8.9 Hz, 1H, 8''-H), 4.05 (q, J = 7.1 Hz, 2H, C**H**$_2$CH$_3$), 3.47 (t, J = 6.9 Hz, 2H, 3-H), 2.56 (t, J = 6.9 Hz, 2H, 2-H), 1.63 (s, 6H, C(CH$_3$)$_2$), 1.15 (t, J = 7.1 Hz, 3H, CH$_2$C**H**$_3$). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 171.2 (C-1), 164.9 (CONH), 159.8 (C-4''), 152.3 (C-8a''), 141.2 (C-4'), 138.2 (C-1'), 132.4 (C-6''), 130.1 (C-7''), 128.1 (C-2' and C-6'), 126.7 (C-3' and C-5'), 120.8 (C-5''), 118.5 (C-8''), 113.3 (C-4a''), 106.6 (C-2''), 60.0 (**C**H$_2$CH$_3$), 35.5 (C-3), 33.5 (C-2), 25.2 (C(**C**H$_3$)$_2$), 14.0 (CH$_2$**C**H$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2988, 1722, 1651, 1490, 1287, 1200, 1166. 1092, 985, 835, 733. **HR-MS (ESI):** m/z = [M-H]$^-$ calcd for $C_{22}H_{23}N_2O_8S^-$: 475.1181; found: 475.1179. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

5-((4-((2-Carboxyethyl)carbamoyl)phenyl)sulfonamido)-2-hydroxybenzoic acid (**CG_140**)

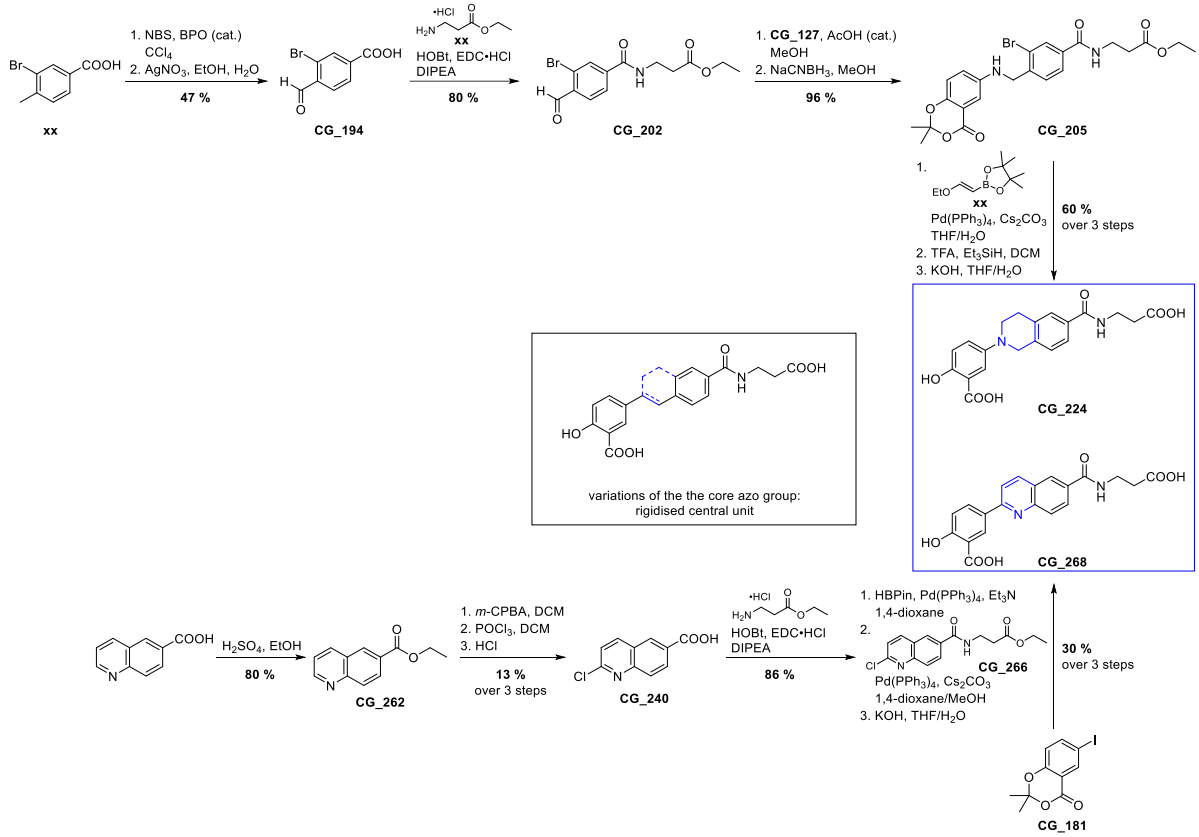


Prepared according to **General Procedure D** from compound **109** (85.0 mg, 0.178 mmol) and heating the reaction mixture to reflux. The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase

separation paper and concentrated *in vacuo* to give pure sulfonamide **CG_140** (70.4 mg, 0.172 mmol, 97 %) as a white solid.

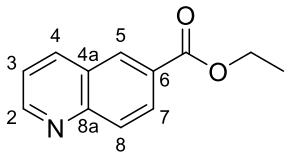
R_f: 0.14 (EtOAc/DCM+AcOH 50:50+1). **m.p.:** 223 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.24 (s, 1H, 1-COOH or 2''-COOH), 10.08 (s, 1H, SO_2NH), 8.71 (t, *J* = 5.5 Hz, 1H, CONH), 7.95 – 7.89 (m, 2H, 3'-H and 5'-H), 7.75 – 7.70 (m, 2H, 2'-H and 6'-H), 7.46 (d, *J* = 2.8 Hz, 1H, 6-H), 7.16 (dd, *J* = 8.9, 2.7 Hz, 1H, 4-H), 6.83 (d, J = 8.9 Hz, 1H, 3-H), 3.44 (td, *J* = 7.2, 5.4 Hz, 2H, 1''-H), 2.52 – 2.51 (m, 1H, 2''-H, collapses with DMSO). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.8 (C-3''), 171.1 (1-COOH), 165.0 (CONH), 158.7 (C-2), 141.3 (C-1'), 138.1 (C-4'), 130.0 (C-4), 128.2 (C-5), 128.0 (C-3' and C-5'), 126.8 (C-2' and C-6'), 123.8 (C-6), 117.8 (C-3), 113.6 (C-1), 35.7 (C-1''), 33.5 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3296, 3248, 1698, 1678, 1633, 1445, 1336, 1227, 1166, 1152, 836, 741, 667. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{17}H_{15}N_2O_8S^-$: 407.0555; found: 407.0554. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 3a).

Synthesis of rigidised analogues CG_224 [2] and CG_268



Scheme S4: Preparation of rigidized analogues 1,2,3,4-tetrahydroisoquinoline **CG_224** (for synthetic details see our latest publication [2]) and quinoline **CG_268**.

Ethyl quinoline-6-carboxylate (**122**) [15]

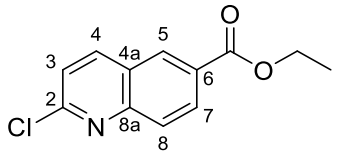


To a solution of 6-quinolinecarboxylic acid (**121**, 3.00 g, 17.3 mmol) in EtOH (120 mL), H_2SO_4 (6.0 mL) was added. The solution was heated to reflux for 16 h. After cooling to room temperature, water (250 mL) was added, and the solution was basified using solid K_2CO_3. The solution was then extracted with DCM (3 x 100 mL). The combined organic layers were dried using a phase separation paper and the solvent was removed *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 20:80) to give ethyl ester **122** (2.79 g, 13.9 mmol, 80 %) as a pale yellow to white solid.

R_f: 0.21 (EtOAc/hexanes 20:80). **m.p.:** 47 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 9.01 (dd, J = 4.3, 1.7 Hz, 1H, 2-H), 8.59 (d, J = 1.9 Hz, 1H, 5-H), 8.31 (dd, J = 8.8, 1.9 Hz, 1H, 7-H), 8.29 – 8.25 (m, 1H, 4-H), 8.17 – 8.12 (m, 1H, 8-H), 7.47 (dd, J = 8.3, 4.2 Hz, 1H, 3-H), 4.46 (q, J = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 1.45 (t, J = 7.1 Hz, 3H, CH$_3$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ

(ppm) = 166.3 (**C**OOEt), 152.6 (C-2), 150.2 (C-8a), 137.5 (C-4), 131.0 (C-5), 129.9 (C-8), 129.1 (C-7), 128.6 (C-6), 127.6 (C-4a), 122.0 (C-3), 61.5 (**C**H_2CH_3), 14.5 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3248, 2991, 1720, 1625, 1502, 1480, 1462, 1362, 1273, 1243, 1187, 1095, 1022, 908, 846, 797, 784, 743. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{12}H_{12}NO_2^+$: 202.0863; found: 202.0861. **Purity (HPLC):** 254 nm: >95 % (method 4a).

Ethyl 2-chloroquinoline-6-carboxylate (**124**) [16]

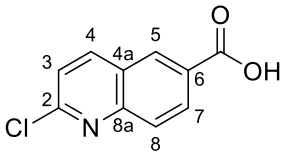


To a mixture of ethyl quinoline-6-carboxylate (**122**, 735 mg, 3.65 mmol) in dry DCM (10.5 mL), *m*-chloroperbenzoic acid (819 mg, 4.75 mmol) was added at 0 °C and the reaction mixture was stirred at room temperature for 16 h (formation of the product confirmed *via* LC-MS). The mixture was then washed with 10 % aq. Na_2SO_3 (15 mL), sat. aq. $NaHCO_3$ (15 mL) and brine (15 mL). The organic layer was dried using a phase separation paper and concentrated *in vacuo* to give crude *N*-oxide intermediate **123**.

To the residue, dry DCM (8.5 mL) and phosphoryl chloride (4.47 mL, 47.5 mmol) were added, and the mixture was stirred at 50 °C for 16 h. The reaction mixture was concentrated *in vacuo*, the residue was diluted with DCM (20 mL) and washed with sat. aq. $NaHCO_3$ (2 x 15 mL) and brine (15 mL). The organic phase was concentrated *in vacuo* and the crude product was purified by FCC (EtOAc/toluene 1:99) to give product **124** (179 mg, 0.760 mmol, 21 %) as a white solid.

R_f: 0.32 (EtOAc/toluene 1:99). **m.p.:** 97 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.58 (d, *J* = 1.9 Hz, 1H, 5-H), 8.34 (dd, *J* = 8.8, 1.9 Hz, 1H, 7-H), 8.22 – 8.20 (m, 1H, 4-H), 8.06 (dd, *J* = 8.8, 0.7 Hz, 1H, 8-H), 7.46 (d, *J* = 8.6 Hz, 1H, 3-H), 4.45 (q, *J* = 7.1 Hz, 2H, C**H_2**CH_3), 1.45 (t, *J* = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 165.9 (6-**C**OOEt), 153.1 (C-2), 149.8 (C-8a), 140.1 (C-4), 130.6 (C-5), 130.4 (C-7), 129.1 (C-6), 129.0 (C-8), 126.2 (C-4a), 123.4 (C-3), 61.7 (**C**H_2CH_3), 14.5 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3064, 2903, 1715, 1621, 1453, 1393, 1261, 1197, 1142, 1085, 1023, 917, 814, 781, 748, 664. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{12}H_{11}ClNO_2^+$: 236.0473; found: 236.0472. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

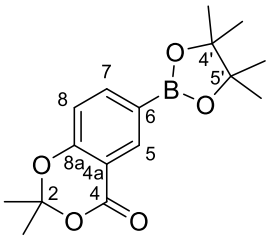
2-Chloroquinoline-6-carboxylic acid (**125**) [16]



A suspension of compound **124** (150 mg, 0.636 mmol) in HCl (32 %, 3.0 mL) was heated to 95 °C for 75 min, cooled to room temperature and concentrated to dryness. The residue was purified FCC (EtOAc/hexanes+AcOH 30:70+1 → 40:60+1) to give product **125** (82.0 mg, 0.395 mmol, 62 %) as a white solid.

R_f: 0.38 (EtOAc/hexanes+AcOH 30:70+1). **m.p.:** 244 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 13.37 (s, 1H, COOH), 8.73 (d, *J* = 1.9 Hz, 1H, 5-H), 8.65 (dd, *J* = 8.7, 0.7 Hz, 1H, 4-H), 8.26 (dd, *J* = 8.8, 2.0 Hz, 1H, 7-H), 8.04 – 8.01 (m, 1H, 8-H), 7.70 (d, *J* = 8.6 Hz, 1H, 3-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 166.7 (COOH), 152.0 (C-2), 148.9 (C-8a), 141.3 (C-4), 130.8 (C-5), 130.2 (C-7), 129.3 (C-6), 128.2 (C-8), 126.2 (C-4a), 123.3 (C-3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2922, 2528, 1677, 1621, 1580, 1450, 1311, 1280, 1196, 1137, 1088, 809, 852, 813, 785, 771, 750. **HR-MS (EI):** *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{10}H_6ClNO_2^{\bullet+}$: 207.0082; found: 207.0085. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

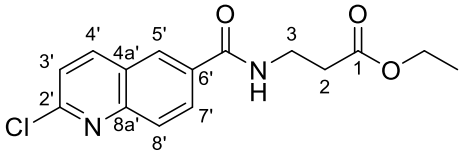
2,2-Dimethyl-6-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-4*H*-benzo[*d*][1,3]dioxin-4-one (**126**)



A solution of iodo compound **93** (152 mg, 0.500 mmol), KOAc (147 mg, 0.500 mmol), and B_2pin_2 (140 mg, 0.551 mmol) in 1,4-dioxane (5.0 mL) was degassed for 10 min and put under N_2 atmosphere. $PdCl_2$(dppf)•DCM (12.2 mg, 0.0150 mmol) was added, and the mixture stirred at 80 °C for 16 h. Another 0.2 eq of B_2pin_2 were added and stirring continued at 80 °C for further 16 h. The reaction mixture was cooled to room temperature, water (15 mL) was added, and the organic materials extracted with EtOAc (3 x 20 mL). The combined organic layers were dried using a phase separation paper and concentrated *in vacuo*. The crude product was put through a short silica gel column (EtOAc/hexanes 10:90) and concentrated *in vacuo*. The resulting material was then diluted with EtOAc (10 mL) and IBX (406 mg, 1.45 mmol) was added under N_2 atmosphere. The reaction mixture was heated to 80 °C for 2 h while stirring rapidly. The heterogeneous mixture was cooled to room temperature, filtered through a plug of silica and concentrated under reduced pressure. The crude product was purified by FCC (EtOAc/hexanes 5:95) to give pinacol ester **126** (33.9 mg, 0.111 mmol, 22 %) as a white solid.

R_f: 0.17 (EtOAc/hexanes 10:90). **m.p.:** 96 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.45 (d, J = 1.7 Hz, 1H, 5-H), 7.96 (dd, J = 8.2, 1.6 Hz, 1H, 7-H), 6.94 (d, J = 8.2 Hz, 1H, 8-H), 1.73 (s, 6H, 2-($C\mathbf{H_3}$)$_2$), 1.33 (s, 12H, 4'-($C\mathbf{H_3}$)$_2$ and 5'-($C\mathbf{H_3}$)$_2$). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 161.1 (C-4), 158.4 (C-8a), 142.7 (C-7), 137.1 (C-5), 123.4 (C-6), 116.7 (C-8), 113.2 (C-4a), 106.5 (C-2), 84.3 (C-4' and C-5'), 26.0 (2-($\mathbf{C}H_3$)$_2$), 25.0 (4'-($\mathbf{C}H_3$)$_2$ and 5'-($\mathbf{C}H_3$)$_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2987, 2362, 1738, 1613, 1379, 1353, 1323, 1276, 1202, 1141, 1128, 1076, 1047, 982, 930, 851, 841, 783, 680, 656. **HR-MS (ESI):** m/z = $[M+H]^+$ calcd for $C_{16}H_{22}BO_5^+$: 305.1555; found: 305.1556. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

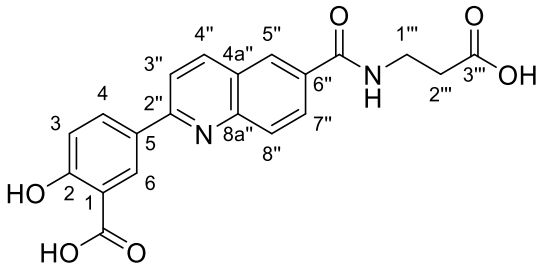
Ethyl 3-(2-chloroquinoline-6-carboxamido)propanoate (**128**)



Prepared according to **General Procedure F** from quinoline **125** (70.0 mg, 0.337 mmol) and β-alanine ethyl ester hydrochloride (**76**, 51.8 mg, 0.337 mmol). The crude product was purified by FCC (EtOAc/hexanes 50:50) to give amide **128** (89.2 mg, 0.291 mmol, 86 %) as a white solid.

R_f: 0.30 (EtOAc/hexanes 50:50). **m.p.:** 130 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 8.30 (t, J = 1.4 Hz, 1H, 5'-H), 8.19 (d, J = 8.6 Hz, 1H, 4'-H), 8.07 (s, 1H, 7'-H), 8.06 (s, 1H, 8'-H), 7.46 (d, J = 8.6 Hz, 1H, 3'-H), 7.07 (t, J = 5.1 Hz, 1H, CONH), 4.19 (q, J = 7.2 Hz, 2H, $C\mathbf{H_2}CH_3$), 3.79 (q, J = 5.9 Hz, 2H, 3-H), 2.69 (t, J = 5.8 Hz, 2H, 2-H), 1.29 (t, J = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 173.2 (C-1), 166.3 (CONH), 152.6 (C-2'), 149.2 (C-8a'), 139.8 (C-4'), 132.9 (C-6'), 129.3 (C-8'), 128.4 (C-7'), 127.5 (C-5'), 126.4 (C-4a'), 123.5 (C-3'), 61.1 ($\mathbf{C}H_2CH_3$), 35.7 (C-3), 34.0 (C-2), 14.3 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3358, 2983, 1719, 1638, 1620, 1538, 1504, 1441, 1318, 1279, 1216, 1179, 1147, 1094, 1027, 908, 820, 780, 748, 664. **HR-MS (ESI):** m/z = $[M+H]^+$ calcd for $C_{15}H_{16}ClN_2O_3^+$: 307.0844; found: 307.0847. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

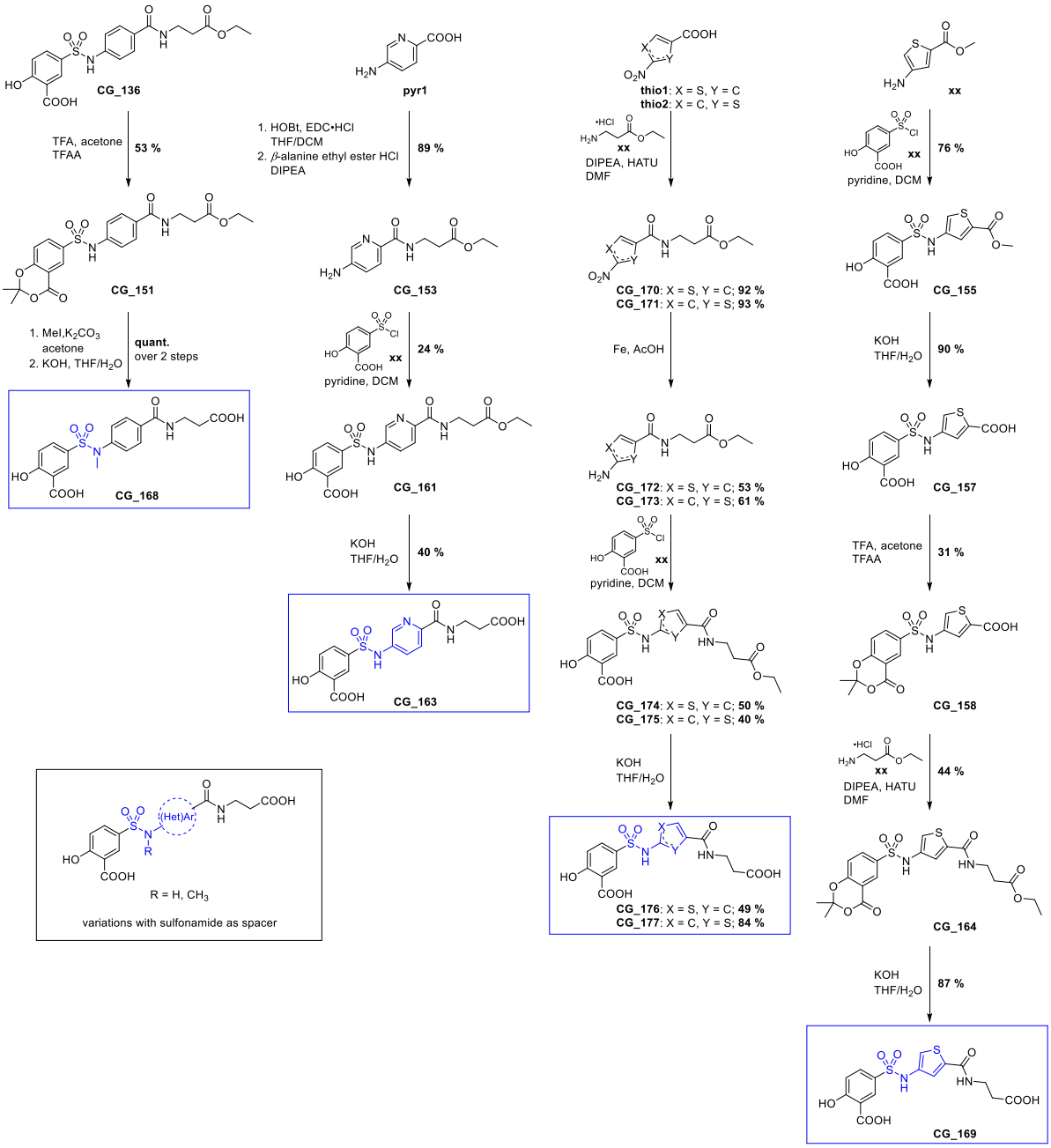
5-(6-((2-Carboxyethyl)carbamoyl)quinolin-2-yl)-2-hydroxybenzoic acid (**CG_268**)



Under N_2 atmosphere, iodo compound **93** (76.0 mg, 0.2500 mmol) and $Pd(PPh_3)_4$ (8.67 mg, 7.50 µmol) were dissolved in dry 1,4-dioxane (1.5 mL) and purged with N_2 for 10 min. The reaction mixture was heated to 80 °C and dry triethylamine (348 µL, 2.50 mmol) and HBPin (72.6 µL, 0.500 mmol) were added. After 3 h at 80 °C, the reaction mixture was allowed to cool to room temperature. Dry MeOH (1.5 mL), compound **128** (76.7 mg, 0.250 mmol) and Cs_2CO_3 (204 mg, 0.625 mmol) were added sequentially and the resulting mixture heated to 100 °C for 16 h. The reaction mixture was allowed to cool to room temperature and water (10 mL) was added. The organic material was extracted with EtOAc (3 x 15 mL). The aqueous phase was then additionally acidified with 1 M aq. HCl and again extracted with EtOAc (3 x 15 mL). The combined organic phases were dried using a phase separation paper and the solvent evaporated *in vacuo*. To the residue was added THF (1.7 mL) followed by the addition of a solution of KOH (70.1 mg, 1.25 mmol) in water (1.7 mL). The reaction mixture was stirred at 60 °C for 2 h. Water (5 mL) was added, and the mixture acidified to pH 1 with 2 M aq. HCl. The resulting precipitates were collected by filtration and dried *in vacuo*. Product **CG_268** (28.6 mg, 0.0752 mmol, 30 % over three steps) was obtained as a brown solid.

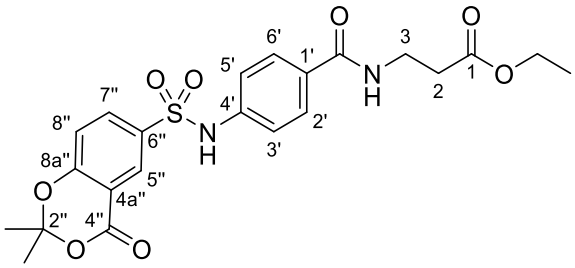
R_f: 0.15 (MeOH/DCM+AcOH 5:95+1). **m.p.:** 170 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 11.62 (s, 1H, 1-COOH or 2'''-COOH), 8.81 – 8.77 (m, 2H, 6-H and CONH), 8.56 (d, *J* = 8.8 Hz, 1H, 4''-H), 8.51 (d, *J* = 1.9 Hz, 1H, 5''-H), 8.47 (dd, *J* = 8.8, 2.4 Hz, 1H, 4-H), 8.22 (d, *J* = 8.9 Hz, 1H, 3''-H), 8.19 (dd, *J* = 9.0, 1.9 Hz, 1H, 7''-H), 8.13 (d, *J* = 8.8 Hz, 1H, 8''-H), 7.17 (d, *J* = 8.7 Hz, 1H, 3-H), 3.54 (q, *J* = 6.7 Hz, 2H, 1'''-H), 2.58 (t, *J* = 7.1 Hz, 2H, 2'''-H). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (C-3'''), 171.7 (1-COOH), 165.7 (CONH), 162.6 (C-2), 156.1 (C-2''), 148.1 (C-8a''), 138.7 (C-4''), 134.4 (C-4), 131.9 (C-6''), 129.6 (C-6), 129.1 (C-5), 128.5 (C-8''), 128.3 (C-7''), 127.7 (C-5''), 126.0 (C-4a''), 118.9 (C-3''), 117.8 (C-3), 113.6 (C-1), 35.8 (C-1'''), 33.8 (C-2'''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2919, 2505, 1713, 1634, 1594, 1538, 1493, 1432, 1393, 1353, 1297, 1207, 1153, 903, 826. **HR-MS (ESI):** *m/z* = $[M-H]^-$ calcd for $C_{20}H_{15}N_2O_6^-$: 379.0936; found: 379.0937. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

Synthesis of analogues CG_168, CG_163, CG_176, CG_177, and CG_169



Scheme S5: Preparation of compounds **CG_168**, **CG_163**, **CG_176**, **CG_177**, **CG_169**.

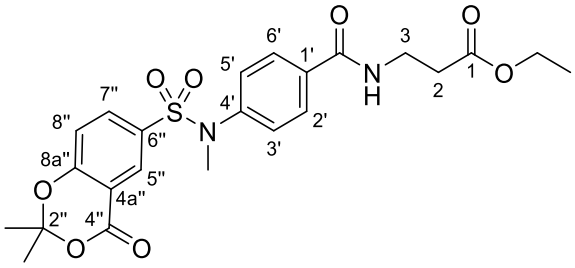
Ethyl 3-(4-((2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxine)-6-sulfonamido)benzamido)-propanoate (**129**)



Prepared according to **General Procedure C** from sulfonamide **106** (75.0 mg, 0.172 mmol). The crude product was purified by FCC (EtOAc/hexanes+AcOH 50:50+1) to give product **129** (43.3 mg, 0.0909 mmol, 53 %) as a yellow solid.

R_f: 0.37 (EtOAc/hexanes+AcOH 70:30+1). **m.p.:** 164 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 8.49 (d, J = 2.4 Hz, 1H, 5''-H), 7.93 (dd, J = 8.7, 2.4 Hz, 1H, 7''-H), 7.69 – 7.64 (m, 2H, 2'-H and 6'-H), 7.48 (s, 1H, SO_2NH), 7.20 – 7.15 (m, 2H, 3'-H and 5'-H), 7.01 (d, J = 8.7 Hz, 1H, 8''-H), 6.85 (t, J = 5.9 Hz, 1H, CONH), 4.17 (q, J = 7.2 Hz, 2H, C**H_2**CH_3), 3.70 (q, J = 6.0 Hz, 2H, 3-H), 2.63 (t, J = 5.8 Hz, 2H, 2-H), 1.74 (s, 6H, $C(CH_3)_2$), 1.27 (t, J = 7.1 Hz, 3H, CH_2C**H_3**). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 173.2 (C-1), 166.6 (CONH), 159.6 (C-4''), 159.4 (C-8a''), 139.5 (C-4'), 135.1 (C-7''), 133.7 (C-6''), 131.1 (C-1'), 129.8 (C-5''), 128.7 (C-2' and C-6'), 120.1 (C-3' and C-5'), 118.6 (C-8''), 113.8 (C-4a''), 107.7 (C-2''), 61.1 (**C**H_2CH_3), 35.5 (C-3), 34.0 (C-2), 26.1 (C(**C**H_3)$_2$), 14.3 (CH_2**C**H_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3330, 3256, 2923, 2853, 1732, 1635, 1610, 1542, 1391, 1335, 1288, 1185, 1164, 1123, 1075, 935, 828, 694. **HR-MS (ESI):** *m/z* = $[M-H]^-$ calcd for $C_{22}H_{23}N_2O_8S^-$: 475.1181; found: 475.1181. **Purity (HPLC):** 254 nm: >95 % (method 4a).

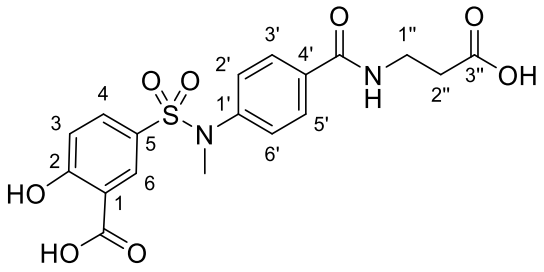
Ethyl 3-(4-((*N*,2,2-trimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxine)-6-sulfonamido)benzamido)-propanoate (**130**)



Methyl iodide (13.0 µL, 0.210 mmol) was added to a mixture of compound **129** (50.0 mg, 0.105 mmol) and K_2CO_3 (43.6 mg, 0.262 mmol) in acetone (7.3 mL). The mixture was stirred at room temperature for 16 h, filtered and the solvent was removed *in vacuo* to give *N*-methylated sulfonamide **130** (51.4 mg, 0.105 mmol, quant.) as a yellowish solid.

R_f: 0.39 (toluene/EtOAc+AcOH 65:35+1). **m.p.:** 55 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.03 (dd, *J* = 2.3, 0.5 Hz, 1H, 5''-H), 7.80 – 7.75 (m, 2H, 2'-H and 6'-H), 7.71 (dd, *J* = 8.7, 2.4 Hz, 1H, 7''-H), 7.30 – 7.26 (m, 2H, 3'-H and 5'-H), 7.18 (dd, *J* = 8.7, 0.5 Hz, 1H, 8''-H), 4.15 (q, *J* = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 3.63 (t, *J* = 6.8 Hz, 2H, 3-H), 3.24 (s, 3H, NCH$_3$), 2.65 (t, *J* = 6.8 Hz, 2H, 2-H), 1.76 (s, 6H, C(CH$_3$)$_2$), 1.24 (t, *J* = 7.2 Hz, 3H, CH$_2$C**H$_3$**). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 173.5 (C-1), 169.2 (CONH), 160.9 (C-4''), 160.7 (C-8a''), 145.6 (C-4'), 136.7 (C-7''), 134.3 (C-1'), 132.2 (C-6''), 130.6 (C-5''), 129.1 (C-2' and C-6'), 127.2 (C-3' and C-5'), 119.8 (C-8''), 114.7 (C-4a''), 108.9 (C-2''), 61.7 (**C**H$_2$CH$_3$), 38.4 (NCH$_3$), 37.1 (C-3), 34.9 (C-2), 25.9 (C(**C**H$_3$)$_2$), 14.5 (CH$_2$**C**H$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3410, 2985, 1728, 1643, 1607, 1477, 1353, 1291, 1255, 1173, 1158, 1131, 1058, 1024, 930, 876, 852, 771, 720. **HR-MS (ESI):** *m*/*z* = [M+H]$^+$ calcd for $C_{23}H_{27}N_2O_8S^+$: 491.1483; found: 491.1480. **Purity (HPLC):** 254 nm: >95 % (method 4a).

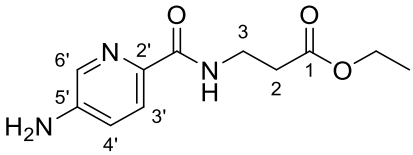
5-(*N*-(4-((2-Carboxyethyl)carbamoyl)phenyl)-N-methylsulfamoyl)-2-hydroxybenzoic acid (**CG_168**)



Prepared according to **General Procedure D** at room temperature from compound **130** (46.0 mg, 0.0938 mmol). The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo* to give the pure product **CG_168** (39.2 mg, 0.0928 mmol, quant.) as a yellow solid.

R_f: 0.30 (EtOAc/hexanes+AcOH 40:60+1), streaks. **m.p.:** 219 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.03 (d, *J* = 2.4 Hz, 1H, 6-H), 7.81 – 7.76 (m, 2H, 3'-H and 5'-H), 7.50 (dd, *J* = 8.8, 2.4 Hz, 1H, 4-H), 7.28 – 7.22 (m, 2H, 2'-H and 6'-H), 6.99 (d, *J* = 8.8 Hz, 1H, 3-H), 3.62 (t, *J* = 7.0 Hz, 2H, 1''-H), 3.20 (s, 3H, CH$_3$), 2.64 (t, *J* = 7.0 Hz, 2H, 2''-H). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 175.4 (C-3''), 172.6 (1-COOH), 169.3 (CONH), 166.9 (C-2), 145.9 (C-1'), 134.9 (C-4), 134.0 (C-4'), 132.2 (C-6), 129.0 (C-3' and C-5'), 127.4 (C-5), 127.1 (C-2' and C-6'), 119.0 (C-3), 115.5 (C-1), 38.3 (CH$_3$), 37.0 (C-1''), 34.6 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2923, 1731, 1652, 1608, 1540, 1500, 1476, 1429, 1351, 1293, 1256, 1174, 1062, 931, 875, 771, 700. **HR-MS (ESI):** *m*/*z* = [M-H]$^-$ calcd for $C_{18}H_{17}N_2O_8S^-$: 421.0711; found: 421.0710. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 1f).

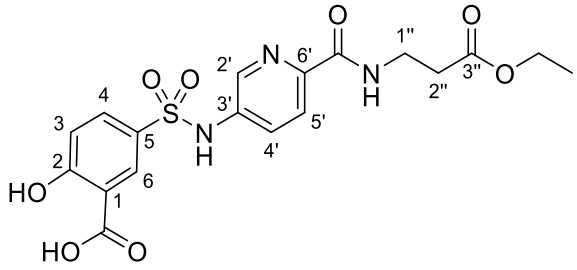
Ethyl 3-(5-aminopicolinamido)propanoate (**134**)



Prepared according to **General Procedure F** from 5-aminopyridine-2-carboxylic acid (**131**, 500 mg, 3.62 mmol) and *β*-alanine ethyl ester hydrochloride (**76**, 556 mg, 3.62 mmol). The crude product was purified by FCC (EtOAc/hexanes 70:30) to give amine **134** (763 mg, 3.22 mmol, 89 %) as a pale yellow oil.

R_f: 0.50 (EtOAc/hexanes 80:20). **^{1}H-NMR (400 MHz, CD_3OD):** *δ* (ppm) = 7.96 (dd, *J* = 2.7, 0.7 Hz, 1H, 6'-H), 7.77 (dd, *J* = 8.5, 0.7 Hz, 1H, 3'-H), 7.02 (dd, *J* = 8.5, 2.7 Hz, 1H, 4'-H), 4.15 (q, *J* = 7.2 Hz, 2H, C**H$_2$**CH$_3$), 3.64 (t, *J* = 6.6 Hz, 2H, 3-H), 2.62 (t, *J* = 6.6 Hz, 2H, 2-H), 1.24 (t, *J* = 7.2 Hz, 3H, CH$_3$). **^{13}C-NMR (101 MHz, CD_3OD):** *δ* (ppm) = 173.7 (C-1), 167.6 (CONH), 149.1 (C-5'), 138.9 (C-2'), 136.3 (C-6'), 124.2 (C-3'), 120.7 (C-4'), 61.7 (**C**H$_2$CH$_3$), 36.08 (C-3), 35.2 (C-2), 14.5 (CH$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3340, 3229, 2982, 1720, 1644, 1585, 1519, 1478, 1323, 1252, 1188, 1161, 1025, 1014, 849, 734. **HR-MS (ESI):** *m*/*z* = [M+H]$^+$ calcd for $C_{11}H_{16}N_3O_3^+$: 238.1186; found: 238.1186. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 1d).

5-(*N*-(6-((3-Ethoxy-3-oxopropyl)carbamoyl)pyridin-3-yl)sulfamoyl)-2-hydroxybenzoic acid (**137**)

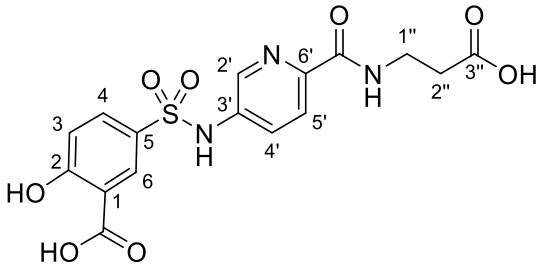


Prepared according to **General Procedure H** from compound **134** (100 mg, 0.421 mmol). The crude product was purified by FCC (MeOH/DCM+AcOH 3:97+1 → 10:90+1) to give product **137** (58.2 mg, 0.133 mmol, 32 %) as a white solid.

R_f: 0.33 (MeOH/DCM+AcOH 10:90+1). **m.p.:** 209 °C. **^{1}H-NMR (500 MHz, CD_3OD):** *δ* (ppm) = 8.37 (d, *J* = 2.5 Hz, 1H, 6-H), 8.32 (d, *J* = 2.6 Hz, 1H, 2'-H), 7.91 (d, *J* = 8.5 Hz, 1H, 5'-H), 7.72 (d, *J* = 8.6 Hz, 1H, 4-H), 7.66 (dd, *J* = 8.6, 2.6 Hz, 1H, 4'-H), 6.87 (d, *J* = 8.6 Hz, 1H, 3-H), 4.12 (q, *J* = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 3.63 (t, *J* = 6.7 Hz, 2H, 1''-H), 2.61 (t, *J* = 6.7 Hz, 2H, 2''-H), 1.21 (t, *J* = 7.1 Hz, 3H, CH$_3$). **^{13}C-NMR (126 MHz, CD_3OD):** *δ* (ppm) = 174.6 (1-COOH), 173.5 (C-3''), 167.2 (C-2), 166.4 (**C**ONH), 145.7 (C-6'), 141.1 (C-2'), 139.4 (C-3'), 132.7 (C-4), 131.9

(C-6), 129.3 (C-5), 128.1 (C-4'), 123.6 (C-5'), 120.0 (C-1), 118.5 (C-3), 61.8 (**C**H_2CH_3), 36.3 (C-1''), 34.9 (C-2''), 14.5 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2925, 1716, 1636, 1582, 1442, 1375, 1324, 1261, 1163, 1110, 1072, 1021, 917, 834, 661. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{18}H_{18}N_3O_8S^-$: 436.0820; found: 436.0819. **Purity (HPLC):** 210 nm: >95 %; 254 nm: 95 % (method 1f).

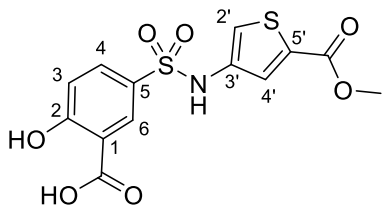
5-(*N*-(6-((2-Carboxyethyl)carbamoyl)pyridin-3-yl)sulfamoyl)-2-hydroxybenzoic acid (**CG_163**)



Prepared according to **General Procedure D** at room temperature from compound **137** (50.0 mg, 0.114 mmol). The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo* to give the pure product **CG_163** (18.6 mg, 0.0454 mmol, 40 %) as a white solid.

R_f: 0.328 (MeOH/DCM+AcOH 10:90+1). **m.p.:** 228 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.34 – 8.32 (m, 2H, 6-H and 2'-H), 7.96 (dd, *J* = 8.5, 0.7 Hz, 1H, 5'-H), 7.85 (dd, *J* = 8.8, 2.5 Hz, 1H, 4-H), 7.69 (dd, *J* = 8.6, 2.6 Hz, 1H, 4'-H), 7.01 (d, J = 8.8 Hz, 1H, 3-H), 3.62 (t, *J* = 6.7 Hz, 2H, 1''-H), 2.60 (t, *J* = 6.7 Hz, 2H, 2''-H). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 175.4 (C-3''), 172.4 (1-COOH), 166.8 (C-2), 166.2 (CONH), 146.4 (C-6'), 141.2 (C-2'), 138.8 (C-3'), 134.5 (C-4), 131.7 (C-6), 130.7 (C-5), 128.5 (C-4'), 123.7 (C-5'), 119.4 (C-3), 115.1 (C-1), 36.3 (C-1''), 34.6 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2923, 1694, 1637, 1587, 1538, 1455, 1327, 1215, 1161, 1110, 1075, 920, 873, 845, 706. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{16}H_{14}N_3O_8S^-$: 408.0507; found: 408.0505. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 1f).

2-Hydroxy-5-((5-(methoxycarbonyl)thiophene)-3-sulfonamido)benzoic acid (**141**)

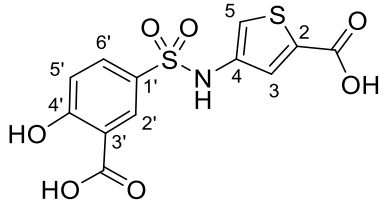


Prepared according to **General Procedure H** from methyl 4-aminothiophene-2-carboxylate (**140**, 490 mg, 3.12 mmol) and 5-chlorosulfonyl-2-hydroxybenzoic acid (**105**, 738 mg,

3.12 mmol). The crude product was purified by FCC (MeOH/DCM+AcOH 5:95+1) to give sulfonamide **141** (847 mg, 2.37 mmol, 76 %) as an off-white solid.

R_f: 0.33 (MeOH/DCM+AcOH 5:95+1). **m.p.:** 206 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.26 (d, *J* = 2.4 Hz, 1H, 6-H), 7.79 (dd, *J* = 8.8, 2.5 Hz, 1H, 4-H), 7.49 (d, *J* = 1.7 Hz, 1H, 4'-H), 7.18 (d, *J* = 1.7 Hz, 1H, 2'-H), 7.00 (d, *J* = 8.8 Hz, 1H, 3-H), 3.83 (s, 3H, C**H**$_3$). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 172.6 (1-COOH), 166.6 (C-2), 163.5 (5'-**C**OOMe), 137.4 (C-3'), 134.5 (C-4), 134.0 (C-5'), 131.6 (C-6), 130.9 (C-5), 129.2 (C-4'), 120.4 (C-2'), 119.1 (C-3), 114.8 (C-1), 52.8 (**C**H$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3229, 1702, 1660, 1608, 1445, 1364, 1293, 1156, 1073, 991, 761, 663. **HR-MS (ESI):** *m*/*z* = [M-H]$^-$ calcd for $C_{13}H_{10}NO_7S_2^-$: 355.9904; found: 355.9905. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 1e).

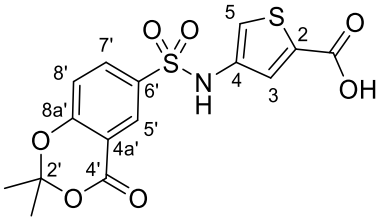
4-((3-Carboxy-4-hydroxyphenyl)sulfonamido)thiophene-2-carboxylic acid (**142**)



Prepared according to **General Procedure D** from compound **141** (500 mg, 1.40 mmol) and heating the reaction mixture to reflux. The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo* to give the pure product **142** (433 mg, 1.26 mmol, 90 %) as an off-white solid.

R_f: 0.40 (EtOAc/DCM+AcOH 30:70+1). **m.p.:** 259 °C. **^{1}H-NMR (500 MHz, CD_3OD):** δ (ppm) = 8.27 (d, *J* = 2.4 Hz, 1H, 2'-H), 7.80 (dd, *J* = 8.9, 2.4 Hz, 1H, 6'-H), 7.47 (d, *J* = 1.7 Hz, 1H, 3-H), 7.16 (d, *J* = 1.7 Hz, 1H, 5-H), 7.02 (d, *J* = 8.8 Hz, 1H, 5'-H). **^{13}C-NMR (126 MHz, CD_3OD):** δ (ppm) = 172.4 (3'-COOH), 166.5 (C-4'), 164.6 (2-COOH), 137.3 (C-4), 135.2 (C-2), 134.7 (C-6'), 131.6 (C-2'), 131.0 (C-1'), 129.2 (C-3), 120.3 (C-5), 119.2 (C-5'), 114.4 (C-3'). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3241, 1675, 1437, 1356, 1259, 1159, 969, 797, 733, 715, 659. **HR-MS (ESI):** *m*/*z* = [M-H]$^-$ calcd for $C_{12}H_8NO_7S_2^-$: 341.9748; found: 341.9747. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 1f).

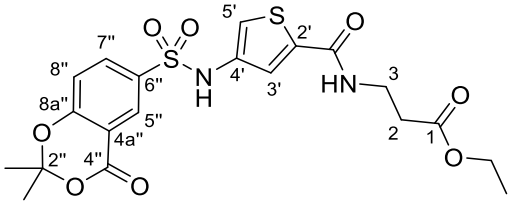
4-((2,2-Dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxine)-6-sulfonamido)thiophene-2-carboxylic acid (**143**)



Prepared according to **General Procedure C** from compound **142** (250 mg, 0.728 mmol). The crude product was purified by FCC (EtOAc/DCM+AcOH 30:70+1) to give acetonide **143** (87.3 mg, 0.228 mmol, 31 %) as a dark yellow oily solid.

R_f: 0.55 (EtOAc/DCM+AcOH 30:70+1). **m.p.:** 70 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.29 (d, J = 2.3 Hz, 1H, 5'-H), 7.96 (dd, J = 8.7, 2.4 Hz, 1H, 7'-H), 7.44 (d, J = 1.7 Hz, 1H, 3-H), 7.22 (d, J = 1.7 Hz, 1H, 5-H), 7.17 (d, J = 8.8 Hz, 1H, 8'-H), 1.73 (s, 7H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 164.5 (2-COOH), 161.0 (C-4'), 160.4 (C-8a'), 136.9 (C-4), 136.1 (C-7'), 135.4 (C-6'), 135.3 (C-2), 130.0 (C-5'), 129.2 (C-3), 120.8 (C-5), 119.7 (C-8'), 114.7 (C-4a'), 108.9 (C-2'), 25.8 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3109, 2923, 1683, 1609, 1476, 1435, 1293, 1162, 1132, 1077, 980, 782, 681. **HR-MS (ESI):** m/z = $[M-H]^-$ calcd for $C_{15}H_{12}NO_7S_2^-$: 382.0061; found: 382.0064. **Purity (HPLC):** 210 nm: 94 %; 254 nm: 89 % (method 1e).

Ethyl 3-(4-((2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxine)-6-sulfonamido)thiophene-2-carboxamido)propanoate (**144**)

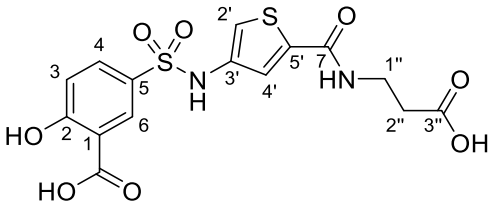


Prepared according to **General Procedure E** from compound **143** (51.0 mg, 0.133 mmol) and *β*-alanine ethyl ester hydrochloride (**76**, 40.8 mg, 0.266 mmol). The crude product was purified by FCC (EtOAc/hexanes+AcOH 50:50+1) to give product **144** (28.0 mg, 0.0580 mmol, 44 %) as a beige solid.

R_f: 0.45 (EtOAc/hexanes+AcOH 50:50+1). **m.p.:** 88 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.29 (d, J = 2.3 Hz, 1H, 5''-H), 7.96 (dd, J = 8.7, 2.4 Hz, 1H, 7''-H), 7.48 (d, J = 1.6 Hz, 1H, 3'-H), 7.17 (d, J = 8.7 Hz, 1H, 8''-H), 7.07 (d, J = 1.6 Hz, 1H, 5'-H), 4.13 (q, J = 7.2 Hz, 2H, $C\mathbf{H}_2CH_3$), 3.57 (t, J = 6.9 Hz, 2H, 3-H), 2.60 (t, J = 6.9 Hz, 2H, 2-H), 1.73 (s, 6H, $C(CH_3)_2$), 1.23 (t, J = 7.1 Hz, 3H, $CH_2C\mathbf{H}_3$). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 173.4 (C-1), 163.7 (CONH), 161.0 (C-4'), 160.4 (C-8a''), 139.8 (C-2'), 136.7 (C-4'), 136.1 (C-7''), 135.4 (C-6''),

130.0 (C-5''), 125.0 (C-3'), 119.7 (C-8''), 118.9 (C-5'), 114.7 (C-4a''), 108.8 (C-2''), 61.7 (**C**H_2CH_3), 36.9 (C-3), 34.9 (C-2), 25.8 (C(**C**$H_3)_2$), 14.5 (CH_2**C**H_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3243, 1742, 1726, 1611, 1559, 1531, 1477, 1367, 1293, 1163, 1120, 1077, 980, 846, 723. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{20}H_{21}N_2O_8S_2^-$: 481.0745; found: 481.0744. **Purity (HPLC):** 210 nm: 89 %; 254 nm: >95 % (method 1e).

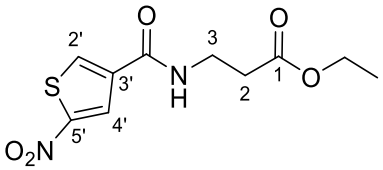
5-(*N*-(5-((2-Carboxyethyl)carbamoyl)thiophen-3-yl)sulfamoyl)-2-hydroxybenzoic acid (**CG_169**)



Prepared according to **General Procedure D** from compound **144** (60.0 mg, 0.124 mmol) at room temperature. The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo* to give the pure product **CG_169** (44.9 mg, 0.108 mmol, 87 %) as a beige solid.

R_f: 0.20 (MeOH/DCM 10:90). **m.p.:** 246 °C. **^{1}H-NMR (500 MHz, CD_3OD):** δ (ppm) = 8.25 (d, *J* = 2.5 Hz, 1H, 6-H), 7.80 (dd, *J* = 8.8, 2.4 Hz, 1H, 4-H), 7.47 (d, *J* = 1.6 Hz, 1H, 4'-H), 7.04 (d, *J* = 1.6 Hz, 1H, 2'-H), 7.02 (d, *J* = 8.9 Hz, 1H, 3-H), 3.56 (t, *J* = 7.0 Hz, 2H, 1''-H), 2.60 (t, *J* = 7.0 Hz, 2H, 2''-H). **^{13}C-NMR (126 MHz, CD_3OD):** δ (ppm) = 175.3 (C-3''), 172.3 (1-COOH), 166.5 (C-2), 163.8 (CONH), 139.6 (C-5'), 137.1 (C-3'), 134.8 (C-4), 131.6 (C-6), 131.1 (C-5), 125.2 (C-4'), 119.2 (C-3), 118.8 (C-2'), 114.1 (C-1), 36.9 (C-1''), 34.6 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3374, 3239, 2924, 2502, 2417, 1723, 1678, 1602, 1342, 1215, 1166, 1076, 805, 670. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{15}H_{13}N_2O_8S_2^-$: 413.0119; found: 413.0118. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

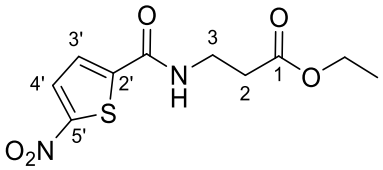
Ethyl 3-(5-nitrothiophene-3-carboxamido)propanoate (**147**)



Prepared according to **General Procedure E** from 5-nitrothiophene-3-carboxylic acid (**145**, 500 mg, 2.89 mmol) and *β*-alanine ethyl ester hydrochloride (**76**, 444 mg, 2.89 mmol). The crude product was purified by FCC (EtOAc/hexanes 40:60) to give **147** (725 mg, 2.66 mmol, 92 %) as a white solid.

R_f: 0.25 (EtOAc/hexanes 40:60). **m.p.:** 102 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.34 (d, J = 1.9 Hz, 1H, 4'-H), 8.32 (d, J = 1.9 Hz, 1H, 2'-H), 4.15 (q, J = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 3.61 (t, J = 6.8 Hz, 2H, 3-H), 2.64 (t, J = 6.8 Hz, 2H, 2-H), 1.24 (t, J = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 173.3 (C-1), 163.5 (CONH), 153.5 (C-5'), 137.4 (C-3'), 135.9 (C-2'), 128.1 (C-4'), 61.7 (**C**H$_2$CH$_3$), 36.9 (C-3), 34.8 (C-2), 14.5 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3373, 3283, 1716, 1624, 1556, 1512, 1335, 1294, 1186, 1014, 830, 815, 729. **HR-MS (ESI):** m/z = [M-H]$^-$ calcd for $C_{10}H_{11}N_2O_5S^-$: 271.0394; found: 271.0393. **Purity (HPLC):** 210 nm: >95 %; 254 nm: 95 % (method 1e).

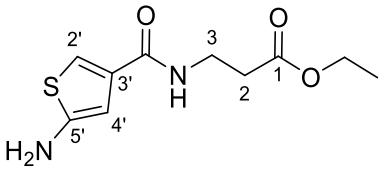
Ethyl 3-(5-nitrothiophene-2-carboxamido)propanoate (**148**)



Prepared according to **General Procedure E** from 5-nitrothiophene-2-carboxylic acid (**146**, 500 mg, 2.89 mmol) and β-alanine ethyl ester hydrochloride (**76**, 444 mg, 2.89 mmol). The crude product was purified by FCC (EtOAc/hexanes 40:60) to give **148** (729 mg, 2.68 mmol, 93 %) as a yellow solid.

R_f: 0.28 (EtOAc/hexanes 40:60). **m.p.:** 86 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 7.96 (d, J = 4.3 Hz, 1H, 4'-H), 7.62 (d, J = 4.4 Hz, 1H, 3'-H), 4.15 (q, J = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 3.63 (t, J = 6.8 Hz, 2H, 3-H), 2.65 (t, J = 6.8 Hz, 2H, 2-H), 1.24 (t, J = 7.2 Hz, 3H, CH_3). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 173.2 (C-1), 162.4 (CONH), 155.4 (C-5'), 146.4 (C-2'), 129.8 (C-4'), 128.1 (C-3'), 61.8 (**C**H$_2$CH$_3$), 37.2 (C-3), 34.7 (C-2), 14.5 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3334, 1714, 1633, 1566, 1513, 1337, 1321, 1292, 1214, 1192, 1029, 817, 754, 655. **HR-MS (ESI):** m/z = [M-H]$^-$ calcd for $C_{10}H_{11}N_2O_5S^-$: 271.0394; found: 271.0393. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 1e).

Ethyl 3-(5-aminothiophene-3-carboxamido)propanoate (**149**)

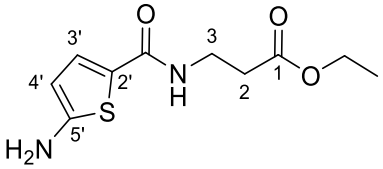


To a mixture of nitro compound **147** (100 mg. 0.358 mmol) and AcOH (1.0 mL), iron powder (78.9 mg, 1.41 mmol) was added, and the resulting reaction mixture was stirred at room temperature for 4 h. The reaction mixture was filtered through a pad of celite and the solvent removed *in vacuo*. To the residue was added EtOAc (40 mL) and sat. aq. $NaHCO_3$ (40 mL).

Phases were separated and the aq. phase extracted with EtOAc (3 x 40 mL). The combined organic extracts were washed with brine (20 mL), dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 55:45 → 60:40) to give the amine **149** (234 mg, 0.966 mmol, 53 %) as a dark brown oil.

R_f: 0.16 (EtOAc/hexanes 40:60). **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 7.07 (d, J = 1.7 Hz, 1H, 2'-H), 6.41 (d, J = 1.7 Hz, 1H, 4'-H), 4.14 (q, J = 7.1 Hz, 2H, C**H_2**CH_3), 3.55 (t, J = 6.9 Hz, 2H, 3-H), 2.60 (t, J = 6.9 Hz, 2H, 2-H), 1.24 (t, J = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 173.5 (C-1), 166.2 (CONH), 154.6 (C-5'), 136.8 (C-3'), 115.8 (C-2'), 106.4 (C-4'), 61.7 (**C**H_2CH_3), 36.6 (C-3), 35.0 (C-2), 14.5 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3314, 2981, 1718, 1627, 1522, 1473, 1374, 1294, 1249, 1183, 1026, 668. **HR-MS (ESI):** m/z = $[M+H]^+$ calcd for $C_{10}H_{15}N_2O_3S^+$: 243.0798; found: 243.0799. **Purity (HPLC):** 210 nm: 87 %; 254 nm: 87 % (method 2a).

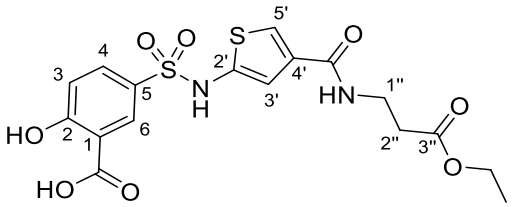
Ethyl 3-(5-aminothiophene-2-carboxamido)propanoate (**150**)



To a mixture of nitro compound **148** (500 mg. 1.84 mmol) and AcOH (4.80 mL), iron powder (410 mg, 7.35 mmol) was added, and the resulting reaction mixture was stirred at room temperature for 2 h. The reaction mixture was filtered through a pad of celite and the solvent removed *in vacuo*. To the residue was added EtOAc (40 mL) and sat. aq. $NaHCO_3$ (40 mL). Phases were separated and the aq. phase extracted with EtOAc (3 x 40 mL). The combined organic extracts were washed with brine (20 mL), dried using a phase separation paper and concentrated *in vacuo* to give amine **150** (272 mg, 1.12 mmol, 61 %) in sufficient purity as a pale brown oil.

R_f: 0.35 (EtOAc/hexanes 70:30). **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 7.25 (d, J = 4.0 Hz, 1H, 3'-H), 5.97 (d, J = 4.1 Hz, 1H, 4'-H), 4.14 (q, J = 7.1 Hz, 2H, C**H_2**CH_3), 3.54 (t, J = 6.9 Hz, 2H, 3-H), 2.59 (t, J = 6.9 Hz, 2H, 2-H), 1.24 (t, J = 7.1 Hz, 4H, CH_3). **^{13}C-NMR (101 MHz, CD_3OD):** δ (ppm) = 173.6 (C-1), 165.5 (CONH), 161.6 (C-5'), 131.2 (C-3'), 121.3 (C-2'), 106.8 (C-4'), 61.7 (**C**H_2CH_3), 36.7 (C-3), 35.3 (C-2), 14.5 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3315, 3210, 1715, 1607, 1516, 1460, 1299, 1184, 1144, 1025, 735. **HR-MS (ESI):** m/z = $[M+H]^+$ calcd for $C_{10}H_{15}N_2O_3S^+$: 243.0798; found: 243.0798. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

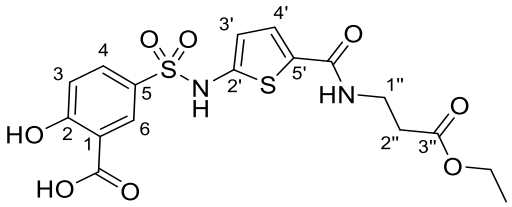
5-(*N*-(4-((3-Ethoxy-3-oxopropyl)carbamoyl)thiophen-2-yl)sulfamoyl)-2-hydroxybenzoic acid (**151**)



Prepared according to **General Procedure H** from amine **149** (100 mg, 0.413 mmol). The crude product was purified by FCC (MeOH/DCM+AcOH 3:97+1 → 7:93+1) to give sulfonamide **151** (91.1 mg, 0.206 mmol, 50 %) as a brown solid.

R_f: 0.44 (MeOH/DCM+AcOH 10:90+1). **m.p.:** 90 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.27 (s, 1H, 6-H), 7.73 (d, *J* = 8.6 Hz, 1H, 4-H), 7.62 (s, 1H, 5'-H), 6.98 (t, *J* = 4.6 Hz, 2H, 3-H and 3'-H), 4.12 (q, *J* = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 3.54 (t, *J* = 6.8 Hz, 2H, 1''-H), 2.59 (t, *J* = 6.7 Hz, 2H, 2''-H), 1.22 (t, *J* = 7.1 Hz, 3H, CH$_3$). **^{13}C-NMR (126 MHz, CD_3OD):** δ (ppm) = 173.4 (C-3''), 173.2 (1-COOH), 166.9 (C-2), 165.2 (CONH), 141.1 (C-2'), 136.3 (C-4'), 134.3 (C-4), 132.0 (C-6), 129.7 (C-5), 125.5 (C-5'), 120.3 (C-3'), 118.9 (C-3), 116.5 (C-1), 61.7 (**C**H$_2$CH$_3$), 36.7 (C-1''), 34.9 (C-2''), 14.5 (CH$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3102, 1704, 1631, 1476, 1362, 1327, 1294, 1162, 1109, 1073, 900, 836, 657. **HR-MS (ESI):** *m*/*z* = [M-H]$^-$ calcd for $C_{17}H_{17}N_2O_8S_2^-$: 441.0432; found: 441.0434. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

5-(*N*-(5-((3-Ethoxy-3-oxopropyl)carbamoyl)thiophen-2-yl)sulfamoyl)-2-hydroxybenzoic acid (**152**)

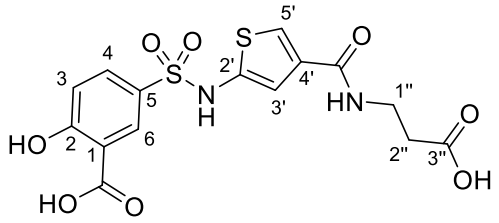


Prepared according to **General Procedure H** from compound **150** (100 mg, 0.413 mmol). The crude product was purified by FCC (MeOH/DCM+AcOH 3:97+1 → 7:93+1) to give sulfonamide **152** (73.3 mg, 0.166 mmol, 40 %) as a brown solid.

R_f: 0.20 (MeOH/DCM 5:95). **m.p.:** 93 °C. **^{1}H-NMR (400 MHz, CD_3OD):** δ (ppm) = 8.34 – 8.29 (m, 1H, 6-H), 7.77 (dd, *J* = 8.8, 1.8 Hz, 1H, 4-H), 7.34 (d, *J* = 4.0 Hz, 1H, 4'-H), 6.99 (d, *J* = 8.7 Hz, 1H, 3-H), 6.59 (d, *J* = 4.0 Hz, 1H, 3'-H), 4.12 (q, *J* = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 3.54 (t, *J* = 6.8 Hz, 2H, 1''-H), 2.58 (t, *J* = 6.8 Hz, 2H, 2''-H), 1.21 (t, *J* = 7.1 Hz, 3H, CH$_3$). **^{13}C-NMR (126 MHz, CD_3OD):** δ (ppm) = 173.4 (C-3''), 173.0 (1-COOH), 167.0 (C-2), 164.2 (CONH), 145.7 (C-2'), 134.2 (C-4), 133.4 (C-5'), 132.0 (C-6), 129.8 (C-5), 128.7 (C-4'), 119.5 (C-3'), 118.9 (C-3), 116.4 (C-1), 61.7 (**C**H$_2$CH$_3$), 36.8 (C-1''), 35.0 (C-2''), 14.5 (CH$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2982,

1711, 1609, 1528, 1457, 1318, 1202, 1162, 1110, 1074, 1018, 892, 832, 805. **HR-MS (ESI):** *m/z* = $[M-H]^-$ calcd for $C_{17}H_{17}N_2O_8S_2^-$: 441.0432; found: 441.0434. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

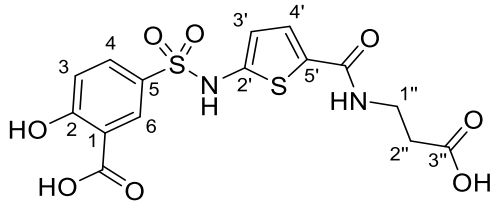
5-(*N*-(4-((2-Carboxyethyl)carbamoyl)thiophen-2-yl)sulfamoyl)-2-hydroxybenzoic acid (**CG_176**)



Prepared according to **General Procedure D** from compound **151** (70.0 mg, 0.158 mmol) at room temperature. The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo* to give the pure product **CG_176** (32.4 mg, 0.0782 mmol, 49 %) as a white solid.

R_f: 0.25 (MeOH/DCM+AcOH 10:90+1). **m.p.:** 254 °C. **^{1}H-NMR (500 MHz, CD_3OD):** δ (ppm) = 8.26 (d, *J* = 2.4 Hz, 1H, 6-H), 7.81 (dd, *J* = 8.8, 2.4 Hz, 1H, 4-H), 7.64 (d, *J* = 1.7 Hz, 1H, 5'-H), 7.04 (d, *J* = 8.9 Hz, 1H, 3-H), 6.99 (d, *J* = 1.7 Hz, 1H, 3'-H), 3.54 (t, *J* = 7.0 Hz, 2H, 1''-H), 2.58 (t, *J* = 6.9 Hz, 2H, 2''-H). **^{13}C-NMR (126 MHz, CD_3OD):** δ (ppm) = 175.3 (C-3''), 172.3 (1-COOH), 166.7 (C-2), 165.1 (CONH), 140.9 (C-2'), 136.3 (C-4'), 135.1 (C-4), 131.9 (C-6), 130.5 (C-5), 125.6 (C-5'), 120.5 (C-3'), 119.3 (C-3), 114.1 (C-1), 36.7 (C-1''), 34.6 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3384, 3234, 2877, 1723, 1678, 1602, 1565, 1535, 1330, 1201, 1167, 1109, 1076, 896, 738, 660. **HR-MS (ESI):** *m/z* = $[M-H]^-$ calcd for $C_{15}H_{13}N_2O_8S_2^-$: 413.0119; found: 413.0118. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

5-(*N*-(5-((2-Carboxyethyl)carbamoyl)thiophen-2-yl)sulfamoyl)-2-hydroxybenzoic acid (**CG_177**)

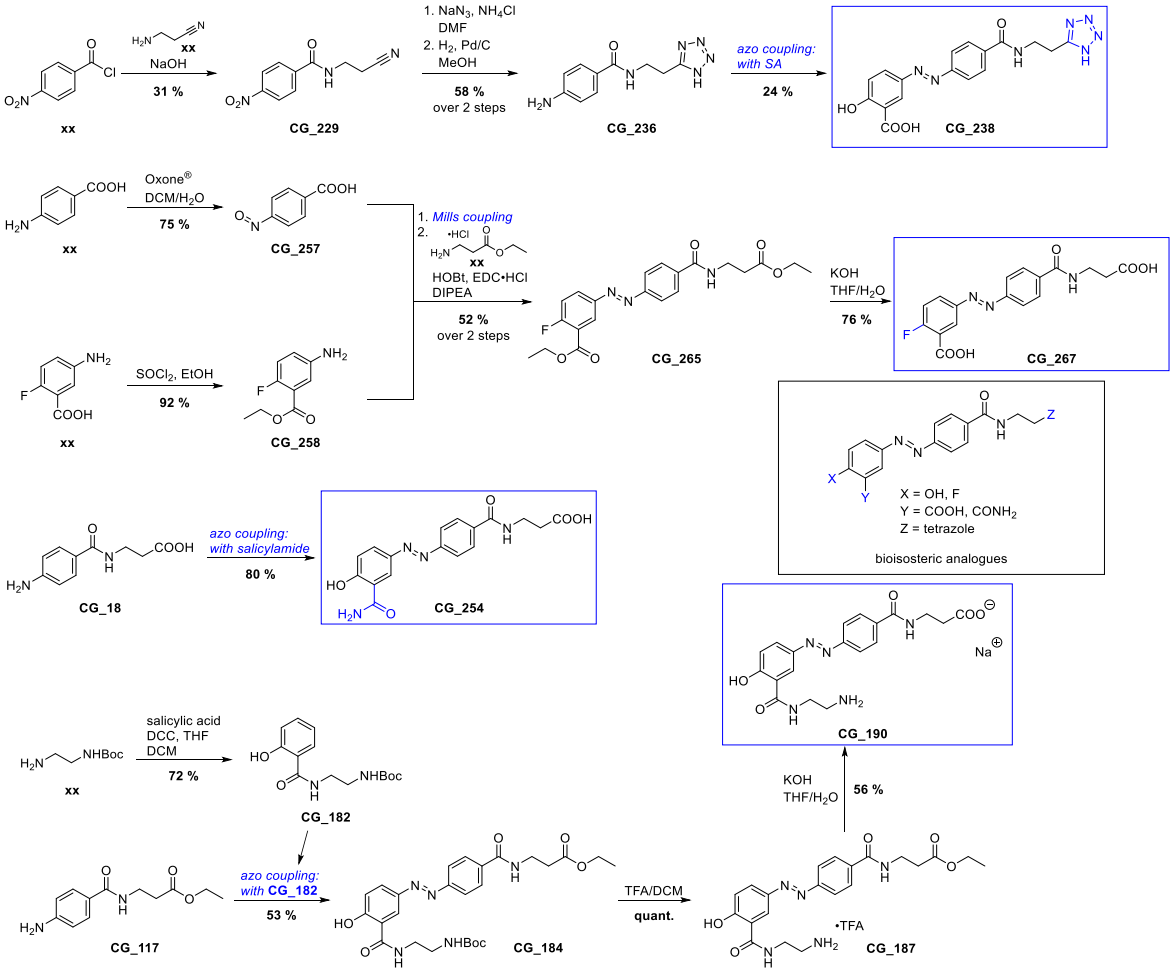


Prepared according to **General Procedure D** from compound **152** (60.0 mg, 0.136 mmol) at room temperature. The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and

concentrated *in vacuo* to give the pure product **CG_177** (47.3 mg, 0.114 mmol, 84 %) as a yellow solid.

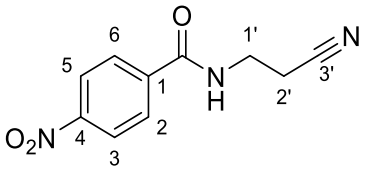
R_f: 0.24 (MeOH/DCM+AcOH 10:90+1), streaks. **m.p.:** 86 °C. **^{1}H-NMR (500 MHz, CD_3OD):** δ (ppm) = 8.30 (d, *J* = 2.4 Hz, 1H, 6-H), 7.82 (dd, *J* = 8.9, 2.4 Hz, 1H, 4-H), 7.36 (d, *J* = 4.1 Hz, 1H, 4'-H), 7.03 (d, *J* = 8.8 Hz, 1H, 3-H), 6.60 (d, *J* = 4.0 Hz, 1H, 3'-H), 3.54 (t, *J* = 7.0 Hz, 2H, 1''-H), 2.58 (t, *J* = 7.0 Hz, 2H, 2''-H). **^{13}C-NMR (126 MHz, CD_3OD):** δ (ppm) = 175.3 (C-3''), 172.4 (1-COOH), 166.8 (C-2), 164.2 (CONH), 145.6 (C-2'), 134.8 (C-4), 133.6 (C-5'), 131.9 (C-6), 130.3 (C-5), 128.7 (C-4'), 119.7 (C-3'), 119.2 (C-3), 114.7 (C-1), 36.9 (C-1''), 34.7 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3076, 2931, 1682, 1558, 1525, 1455, 1317, 1201, 1157, 1106, 1073, 894, 798, 654. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{15}H_{13}N_2O_8S_2^-$: 413.0119; found: 413.0118. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

Synthesis of bioisosteres CG_238, CG_267, CG_254, and immobilisable CG_190



Scheme S6: Preparation of bioisosteres **CG_238**, **CG_267**, **CG_254**, and immobilisable **CG_190**.

N-(2-Cyanoethyl)-4-nitrobenzamide (**158**) [17]

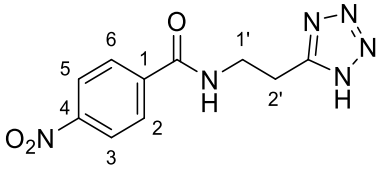


At 0 °C, 4-nitrobenzoyl chloride (**1**, 1.30 g, 6.99 mmol) and 10 % aq. NaOH (3.22 mL, 8.04 mmol) were gradually added alternately into 3-aminopropionitrile (**157**, 522 µL, 7.13 mmol). After stirring for 30 min, the precipitates were collected by filtration, washed with water (15 mL) and dried *in vacuo*. The crude product was purifed by FCC (EtOAc/hexanes+AcOH 50:50+1) and subsequent recrystillisation from hot acetone. Amide **158** (471 mg, 2.15 mmol, 31 %) was obtained as white crystals.

R_f: 0.20 (EtOAc/hexanes+AcOH 50:50+1). **m.p.:** 156 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 9.18 (t, J = 5.7 Hz, 1H, CONH), 8.36 – 8.33 (m, 2H, 3-H and 5-H), 8.10 – 8.06 (m, 2H,

2-H and 6-H), 3.53 (td, *J* = 6.4, 5.5 Hz, 2H, 1'-H), 2.80 (t, *J* = 6.5 Hz, 2H, 2'-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 164.9 (CONH), 149.2 (C-4), 139.5 (C-1), 128.7 (C-2 and C-6), 123.7 (C-3 and C-5), 119.2 (C-3'), 35.6 (C-1'), 17.4 (C-2'). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3306, 3078, 2252, 1650, 1598, 1548, 1521, 1422, 1346, 1320, 1297, 1081, 876, 846, 689, 676. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{10}H_8N_3O_3^-$: 218.0571; found: 218.0571. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

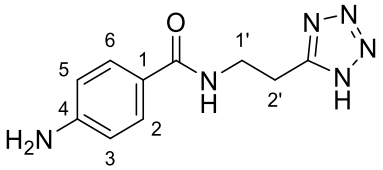
N-(2-(1*H*-Tetrazol-5-yl)ethyl)-4-nitrobenzamide (**159**)



Nitrile **158** (300 mg, 1.37 mmol), NaN_3 (93.4 mg, 1.44 mmol) and NH_4Cl (80.5 mg, 1.51 mmol) were suspended in dry DMF (1.2 mL) and the reaction mixture heated to 120 °C. After 16 h, the mixture was cooled to room temperature and the solid residue triturated with 2 M aq. HCl (5.0 mL), filtered, washed with water (25 mL) and dried *in vacuo*. Tetrazole **159** (274 mg, 1.05 mmol, 77 %) was obtained as an off-white solid.

R_f: 0.29 (MeOH/DCM+AcOH 5:95+0.5). **m.p.:** 225 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 16.15 (s, 1H, tetrazole-H), 8.97 (t, *J* = 5.7 Hz, 1H, CONH), 8.33 – 8.30 (m, 2H, 3-H and 5-H), 8.04 – 8.01 (m, 2H, 2-H and 6-H), 3.65 (td, *J* = 6.9, 5.6 Hz, 2H, 1'-H), 3.18 (t, *J* = 6.9 Hz, 2H, 2'-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 164.8 (CONH), 154.1 (tetrazole-C), 149.1 (C-4), 139.9 (C-1), 128.7 (C-2 and C-6), 123.5 (C-3 and C-5), 37.8 (C-1'), 23.2 (C-2'). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3302, 3026, 2889, 2755, 1643, 1596, 1554, 1512, 1353, 1345, 1324, 1311, 1265, 1108, 1056, 872, 846, 724, 716. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{10}H_9N_6O_3^-$: 261.0742; found: 261.0740. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

N-(2-(1*H*-Tetrazol-5-yl)ethyl)-4-aminobenzamide (**160**)

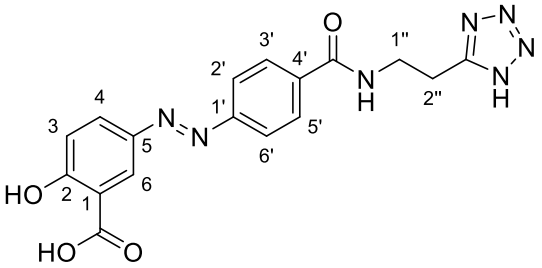


Prepared according to **General Procedure A** from nitro compound **159** (250 mg, 0.953 mmol). Aryl amine **160** (1.76 g, 8.44 mmol, 75 %) was obtained as a beige solid.

R_f: 0.15 (MeOH/DCM+AcOH 5:95+0.5). **m.p.:** 252 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 16.05 (s, 1H, tetrazole-H), 8.18 (t, *J* = 5.7 Hz, 1H, CONH), 7.54 – 7.51 (m, 2H, 2-H and 6-H), 6.54 – 6.51 (m, 2H, 3-H and 5-H), 5.61 (s, 2H, NH_2), 3.57 (td, *J* = 7.1, 5.6 Hz, 2H, 1'-H),

3.11 (t, *J* = 7.1 Hz, 2H, 2'-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 166.4 (CONH), 154.3 (tetrazole-C), 151.7 (C-4), 128.7 (C-2 and C-6), 120.9 (C-1), 112.5 (C-3 and C-5), 37.4 (C-1'), 23.7 (C-2'). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3487, 3386, 2606, 1619, 1595, 1548, 1508, 1425, 1309, 1293, 1272, 1190, 1032, 843, 768. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{10}H_{11}N_6O^-$: 231.1000; found: 231.0999. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

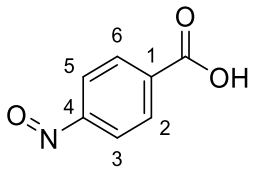
(*E*)-5-((4-((2-(1*H*-Tetrazol-5-yl)ethyl)carbamoyl)phenyl)diazenyl)-2-hydroxybenzoic acid (**CG_238**)



Prepared according to **General Procedure B** from aryl amine **160** (150 mg, 0.646 mmol) and salicylic acid (91.9 mg, 0.665 mmol). The crude product was purified by FCC (MeOH/DCM+AcOH 5:95+1) to give tetrazole **CG_238** (58.0 mg, 0.152 mmol, 24 %) as an orange solid.

R_f: 0.24 (MeOH/DCM+AcOH 7:93+1). **m.p.:** 264 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 16.16 (s, 1H, tetrazole-H), 11.96 (s, 1H, 1-COOH), 8.79 (t, *J* = 5.6 Hz, 1H, CONH), 8.33 (d, *J* = 2.6 Hz, 1H, 6-H), 7.97 (d, *J* = 8.2 Hz, 2H, 3'-H and 5'-H), 7.93 (dd, *J* = 8.8, 2.5 Hz, 1H, 4-H), 7.87 (d, *J* = 8.3 Hz, 2H, 2'-H and 6'-H), 6.91 (d, *J* = 8.8 Hz, 1H, 3-H), 3.66 (q, *J* = 6.5 Hz, 2H, 1''-H), 3.18 (t, *J* = 7.0 Hz, 2H, 2''-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 170.8 (1-COOH), 167.8 (C-2), 165.8 (CONH), 154.1 (tetrazole-C), 153.7 (C-1'), 143.1 (C-5), 135.2 (C-4'), 128.4 (C-3' and C-5'), 127.6 (C-4), 126.4 (C-6), 121.8 (C-2' and C-6'), 118.2 (C-3), 117.3 (C-1), 37.7 (C-1''), 23.4 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3318, 2924, 1628, 1540, 1485, 1386, 1296, 1248, 1178, 1158, 1071, 856, 835, 799, 770, 699. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{17}H_{16}N_7O_4^+$: 382.1258; found: 382.1253. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

4-Nitrosobenzoic acid (**161**) [18]

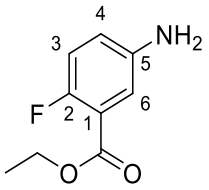


4-Aminobenzoic acid (**14**, 500 mg, 3.65 mmol) was dissolved in DCM (5.0 mL). A solution of Oxone® (1.79 g, 7.29 mmol) in water (23 mL) was added and the solution stirred under N_2

atmosphere for 1 h. The precipitate was collected by filtration, washed with water (20 mL), and dried *in vacuo* to give nitroso derivative **161** (411 mg, 2.72 mmol, 75 %) as a yellow solid.

R_f: 0.43 (EtOAc/hexanes+AcOH 30:70+1). **m.p.:** 220 °C (decomposition). **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 13.49 (s, 1H, COOH), 8.28 – 8.24 (m, 2H, 2-H and 6-H), 8.05 – 8.01 (m, 2H, 3-H and 5-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 166.2 (COOH), 165.0 (C-4), 136.6 (C-1), 131.0 (C-2 and C-6), 120.6 (C-3 and C-5). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3060, 2830, 2666, 2552, 1686, 1600, 1427, 1294, 1262, 1112, 928, 871, 845, 792, 762, 690. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_7H_4NO_3^-$: 150.0197; found: 150.0196. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

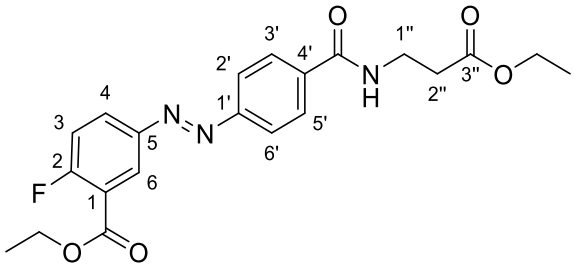
Ethyl 5-amino-2-fluorobenzoate (**163**) [19]



A solution of 5-amino-2-fluorobenzoic acid (**162**, 500 mg, 3.22 mmol) in EtOH (10 mL) was cooled to 0 °C and $SOCl_2$ (470 µL, 6.45 mmol) was added dropwise over 10 min. The resulting solution was then allowed to warm to room temperature over 1 h and then heated to reflux for 3 h. After this time, the mixture was concentrated *in vacuo* and water (25 mL) was added. The obtained purple acidic mixture was washed with an EtOAc/hexanes mixture (1:1; 2 x 15 mL). The organic phase was further extracted with 0.5 M aq. HCl (1 x 15 mL). The combined aqueous acidic layers were washed with an EtOAc/hexanes mixture (1:1; 15 mL), cooled to 0 °C, basified using solid $NaHCO_3$ and then extracted with EtOAc (3 x 15 mL). The combined organic phase was washed with brine (20 mL), dried using a phase separation paper and concentrated under reduced pressure to give product **163** (541 mg, 2.95 mmol, 92 %) as a purple oil, which solidifies below room temperature.

R_f: 0.54 (EtOAc/hexanes 50:50). **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 7.19 (dd, *J* = 5.8, 3.1 Hz, 1H, 6-H), 6.92 (dd, *J* = 10.3, 8.7 Hz, 1H, 3-H), 6.78 (ddd, *J* = 8.7, 3.9, 3.1 Hz, 1H, 4-H), 4.37 (q, *J* = 7.1 Hz, 2H, CH_2), 3.64 (s, 2H, NH_2), 1.38 (t, *J* = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 164.8 (d, **C**OOEt), 155.5 (d, C-2), 142.4 (d, C-5), 120.7 (d, C-4), 119.2 (d, C-1), 117.7 (d, C-3), 117.3 (C-6), 61.4 (CH_2), 14.4 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3441, 3358, 3211, 2981, 1698, 1633, 1589, 1500, 1446, 1367, 1308, 1264, 1243, 1213, 1138, 1088, 1023, 817, 769. **HR-MS (EI):** *m*/*z* = $[M]^{\bullet+}$ calcd for $C_9H_{10}FNO_2^{\bullet+}$: 183.0690; found: 183.0688. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

Ethyl (*E*)-5-((4-((3-ethoxy-3-oxopropyl)carbamoyl)phenyl)diazenyl)-2-fluorobenzoate (**165**)

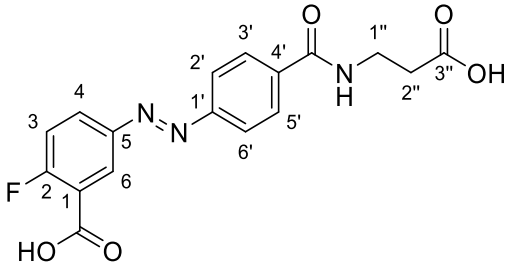


Nitroso derivative **161** (60.4 mg, 0.400 mmol) and amine **163** (61.1 mg, 0.333 mmol) were suspended in AcOH/DMSO (1:1; 2.5 mL). The reaction mixture was stirred at room temperature for 48 h. Water (10 mL) was added, the precipitate collected by filtration, washed with water (15 mL), and dried *in vacuo* to give crude azobenzene **164**.

Amide coupling was then performed according to **General Procedure F** from crude azo-derivative **164** (0.333 mmol) and *β*-alanine ethyl ester hydrochloride (**76**, 51.2 mg, 0.333 mmol). The crude product was purified by FCC (EtOAc/hexanes 40:60) to give azobenzene **165** (71.2 mg, 0.171 mmol, 52 % over two steps) as an orange solid.

R_f: 0.34 (EtOAc/hexanes 40:60). **m.p.:** 113 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** *δ* (ppm) = 8.54 (dd, *J* = 6.8, 2.6 Hz, 1H, 6-H), 8.11 (ddd, *J* = 8.8, 4.4, 2.6 Hz, 1H, 4-H), 7.98 – 7.95 (m, 2H, 2'-H and 6'-H), 7.94 – 7.91 (m, 2H, 3'-H and 5'-H), 7.29 (dd, *J* = 10.0, 8.8 Hz, 1H, 3-H), 6.97 (t, *J* = 5.8 Hz, 1H, CONH), 4.45 (q, *J* = 7.1 Hz, 2H, 1-COOC**H_2**CH_3), 4.19 (q, *J* = 7.1 Hz, 2H, 2''-COOC**H_2**CH_3), 3.76 (q, *J* = 6.0 Hz, 2H, 1''-H), 2.68 (t, *J* = 6.0 Hz, 2H, 2''-H), 1.44 (t, *J* = 7.1 Hz, 3H, 1-$COOCH_2C$**H_3**), 1.29 (t, *J* = 7.2 Hz, 3H, 2''-$COOCH_2C$**H_3**). **^{13}C-NMR (126 MHz, $CDCl_3$):** *δ* (ppm) = 173.2 (C-3''), 166.6 (CONH), 163.8 (d, 1-**C**$OOCH_2CH_3$), 163.6 (d, C-2), 154.0 (C-1'), 148.6 (d, C-5), 136.7 (C-4'), 128.2 (C-3' and C-5'), 128.0 (d, C-4), 127.9 (d, C-6), 123.2 (C-2' and C-6'), 120.0 (d, C-1), 118.1 (d, C-3), 61.9 (1-COO**C**H_2CH_3), 61.1 (2''-COO**C**H_2CH_3), 35.6 (C-1''), 34.0 (C-2''), 14.4 (1-$COOCH_2C$**H_3**), 14.3 (2''-$COOCH_2C$**H_3**). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3298, 2983, 1728, 1713, 1636, 1617, 1605, 1537, 1481, 1422, 1287, 1240, 1188, 1154, 1074, 1018, 858, 844, 781. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{21}H_{23}FN_3O_5^+$: 416.1616; found: 416.1615. **Purity (HPLC):** 210 nm: 88 %; 254 nm: 88 % (method 2a).

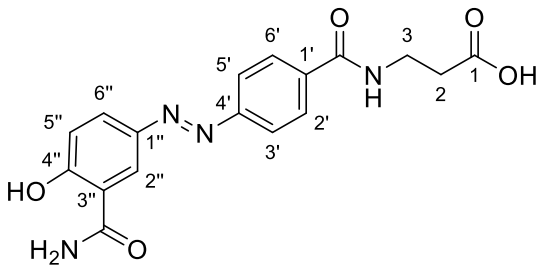
(*E*)-5-((4-((2-carboxyethyl)carbamoyl)phenyl)diazenyl)-2-fluorobenzoic acid (**CG_267**)



Prepared according to **General Procedure D** from azobenzene **165** (60.0 mg, 0.144 mmol) at room temperature. The resulting precipitates were collected by filtration and dried *in vacuo* to give product **CG_267** (39.6 mg, 0.110 mmol, 76 %) as an orange solid.

R_f: 0.15 (MeOH/DCM+AcOH 2:98+1). **m.p.:** 248 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 13.52 (s, 1H, 1-COOH or 2''-COOH), 12.36 (s, 1H, 1-COOH or 2''-COOH), 8.73 (t, *J* = 5.5 Hz, 1H, CONH), 8.38 (dd, *J* = 6.9, 2.7 Hz, 1H, 6-H), 8.21 (ddd, *J* = 8.8, 4.4, 2.7 Hz, 1H, 4-H), 8.07 – 8.03 (m, 2H, 3'-H and 5'-H), 8.01 – 7.97 (m, 2H, 2'-H and 6'-H), 7.59 (dd, *J* = 10.3, 8.8 Hz, 1H, 3-H), 3.49 (td, *J* = 7.1, 5.4 Hz, 2H, 1''-H), 2.55 (t, *J* = 7.1 Hz, 2H, 2''-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (C-3''), 165.4 (CONH), 164.3 (d, 1-COOH), 162.8 (d, C-2), 153.0 (C-1'), 148.0 (d, C-5), 136.9 (C-4'), 129.1 (d, C-4), 128.5 (C-3' and C-5'), 125.8 (d, C-6), 122.6 (C-2' and C-6'), 120.4 (d, C-1), 118.6 (d, C-3), 35.7 (C-1''), 33.7 (C-2''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3361, 2921, 2666, 1697, 1683, 1639, 1619, 1536, 1449, 1284, 1237, 1134, 1080, 920, 855, 834, 772, 696. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{17}H_{13}FN_3O_5^-$: 358.0845; found: 358.0846. **Purity (HPLC):** 210 nm: 90 %; 254 nm: 90 % (method 2a).

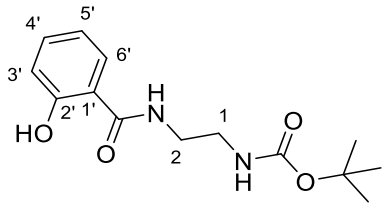
(*E*)-3-(4-((3-Carbamoyl-4-hydroxyphenyl)diazenyl)benzamido)propanoic acid (**CG_254**)



Prepared according to **General Procedure B** from amine **3** (100 mg, 0.480 mmol) and salicylamide (67.8 mg, 0.495 mmol). Azobenzene **CG_254** (136 mg, 0.382 mmol, 80 %) was obtained as an orange solid.

R_f: 0.28 (MeOH/DCM+AcOH 5:95+1). **m.p.:** 278 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 13.75 (s, 1H, OH), 12.28 (s, 1H, COOH), 8.76 (s, 1H, $CONH_2$), 8.69 (t, *J* = 5.5 Hz, 1H, CONH), 8.60 (d, *J* = 2.4 Hz, 1H, 2''-H), 8.14 – 8.10 (m, 1H, $CONH_2$), 8.05 – 8.01 (m, 2H, 2'-H and 6'-H), 7.99 (dd, *J* = 8.9, 2.4 Hz, 1H, 6''-H), 7.91 – 7.87 (m, 2H, 3'-H and 5'-H), 7.07 (d, *J* = 9.0 Hz, 1H, 5''-H), 3.49 (td, *J* = 7.1, 5.4 Hz, 2H, 3-H), 2.54 (t, *J* = 7.1 Hz, 2H, 2-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (COOH), 171.4 ($CONH_2$), 165.5 (CONH), 164.8 (C-4''), 153.4 (C-4'), 144.3 (C-1''), 135.9 (C-1'), 128.5 (C-2' and C-6'), 126.9 (C-2''), 125.9 (C-6''), 122.0 (C-3' and C-5'), 118.9 (C-5''), 114.9 (C-3''), 35.7 (C-3), 33.7 (C-2). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3439, 3346, 3815, 1691, 1657, 1630, 1527, 1488, 1436, 1375, 1311, 1252, 1218, 1115, 954, 892, 855, 769. **HR-MS (EI):** *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{17}H_{16}N_4O_5^{\bullet+}$: 356.1115; found: 356.1113. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

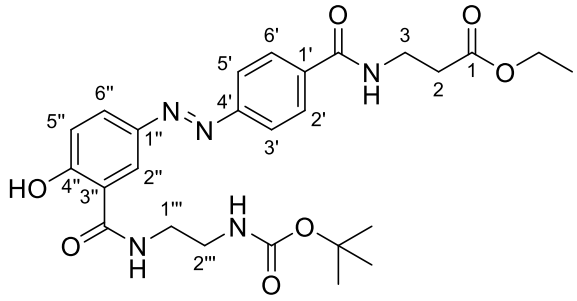
tert-Butyl (2-(2-hydroxybenzamido)ethyl)carbamate (**167**) [20]



To a solution of salicylic acid (150 mg, 1.09 mmol) and *tert*-butyl *N*-(2-aminoethyl)carbamate (**166**, 174 mg, 1.09 mmol) in dry THF (1.2 mL), a solution of 1,3-dicyclocarbodiimide (246 mg, 1.19 mmol) in DCM (0.40 mL) was added dropwise at 0 °C. The reaction was stirred for 18 h and then filtered. The filtrate was concentrated *in vacuo*. The residue was dissolved in EtOAc (20 mL), washed with 5 % aq. $NaHCO_3$ (10 mL), dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 30:70) to give **167** (220 mg, 0.786 mmol, 72 %) as a colourless oil.

R_f: 0.27 (EtOAc/hexanes 30:70). **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 7.73 (s, 1H, 1-NHCO), 7.47 (dd, *J* = 8.0, 1.5 Hz, 1H, 6'-H), 7.37 (ddd, *J* = 8.6, 7.2, 1.5 Hz, 1H, 4'-H), 6.96 (dd, *J* = 8.3, 1.2 Hz, 1H, 3'-H), 6.84 (ddd, *J* = 8.2, 7.2, 1.2 Hz, 1H, 5'-H), 5.00 (s, 1H, 2-NHCO), 3.57 – 3.50 (m, 2H, 2-H), 3.43 (dd, *J* = 6.7, 3.9 Hz, 2H, 1-H), 1.44 (s, 9H, C(**C**H_3)$_3$). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 170.7 (2-NHCO), 161.7 (C-2'), 158.2 (1-NHCO), 134.2 (C-4'), 126.1 (C-6'), 118.8 (C-5'), 118.5 (C-3'), 114.3 (C-1'), 80.7 (**C**$(CH_3)_3$), 42.6 (C-2), 39.7 (C-1), 28.4 (C(**C**H_3)$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3345, 2979, 2936, 1688, 1636, 1533, 1489, 1365, 1250, 1161, 1147, 856, 751. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{14}H_{19}N_2O_4^-$: 279.1350; found: 279.1350. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

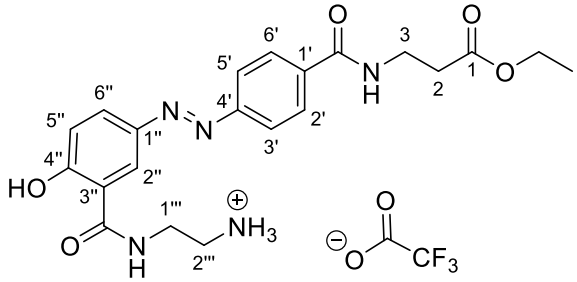
Ethyl (*E*)-3-(4-((3-((2-((*tert*-butoxycarbonyl)amino)ethyl)carbamoyl)-4-hydroxyphenyl)diazenyl)benzamido)propanoate (**168**)



Prepared according to **General Procedure B** from amine **95** (100 mg, 0.423 mmol) and compound **167** (122 mg, 0.436 mmol). Azobenzene **168** (119 mg, 0.226 mmol, 53 %) was obtained as a yellow solid.

R_f: 0.38 (EtOAc/hexanes 60:40). **m.p.:** 180 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 13.36 (s, 1H, OH), 9.12 (t, *J* = 5.8 Hz, 1H, 1'''-NHCO), 8.71 (t, *J* = 5.5 Hz, 1H, 3-NHCO), 8.56 (d, *J* = 2.4 Hz, 1H, 2''-H), 8.05 – 8.01 (m, 2H, 2'-H and 6'-H), 7.99 (dd, *J* = 8.9, 2.4 Hz, 1H, 6''-H), 7.90 (dd, *J* = 8.5, 1.7 Hz, 2H, 3'-H and 5'-H), 7.10 (d, *J* = 8.8 Hz, 1H, 5''-H), 6.97 (t, *J* = 5.9 Hz, 1H, 2'''-NHCO), 4.08 (q, *J* = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 3.52 (td, *J* = 6.9, 5.4 Hz, 2H, 3-H), 3.37 (q, *J* = 6.0 Hz, 2H, 1'''-H), 3.17 (q, *J* = 6.2 Hz, 2H, 2'''-H), 2.60 (t, *J* = 7.0 Hz, 2H, 2-H), 1.36 (s, 9H, C(C**H$_3$**)$_3$), 1.18 (t, *J* = 7.1 Hz, 3H, CH$_2$C**H$_3$**). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 171.3 (C-1), 168.6 (1'''-NHCO), 165.6 (3-NHCO), 163.5 (C-4''), 155.8 (2'''-NHCO), 153.4 (C-4'), 144.5 (C-1''), 135.9 (C-1'), 128.5 (C-2' and C-6'), 126.1 (C-6''), 125.8 (C-2''), 122.0 (C-3' and C-5'), 118.6 (C-5''), 115.6 (C-3''), 77.7 (**C**(CH$_3$)$_3$), 60.0 (**C**H$_2$CH$_3$), 39.5 (C-1''', collapses with DMSO), 39.2 (C-2''', collapses with DMSO), 35.6 (C-3), 33.7 (C-2), 28.2 (C(**C**H$_3$)$_3$), 14.1 (CH$_2$**C**H$_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3345, 2980, 2937, 1731, 1688, 1634, 1592, 1536, 1487, 1367, 1275, 1249, 1169, 1026, 989, 856, 833. **HR-MS (EI):** *m*/*z* = [M]$^{\bullet+}$ calcd for $C_{26}H_{33}N_5O_7^{\bullet+}$: 527.2374; found:527.2399. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

Ethyl (*E*)-3-(4-((3-((2-aminoethyl)carbamoyl)-4-hydroxyphenyl)diazenyl)benzamido)propanoate • TFA (**169**)

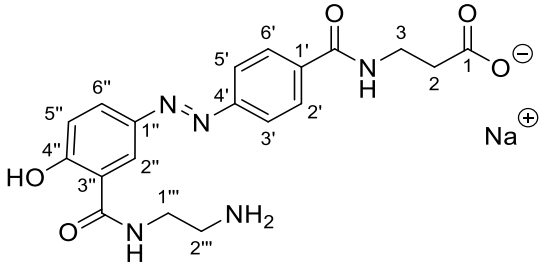


Azobenzene **168** (70.0 mg, 0.133 mmol) was dissolved in DCM (1.2 mL) and TFA (0.80 mL) was added dropwise. The reaction mixture was stirred for 2 h at room temperature. After this time, the product was concentrated *in vacuo* to give TFA-salt **169** (69.6 mg, 0.129 mmol, quant.) as an orange solid.

R_f: 0.17 (MeOH/DCM+AcOH 10:90+1). **m.p.:** 190 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 13.05 (s, 1H, OH), 9.17 (t, *J* = 5.7 Hz, 1H, 1'''-NHCO), 8.73 (t, *J* = 5.5 Hz, 1H, 3-NHCO), 8.55 (d, *J* = 2.4 Hz, 1H, 2''-H), 8.07 – 7.98 (m, 3H, 2'-H and 6'-H and 6''-H), 7.92 – 7.88 (m, 2H, 3'-H and 5'-H), 7.86 (s, 3H, 2'''-NH$_3^+$), 7.15 (d, *J* = 8.9 Hz, 1H, 5''-H), 4.08 (q, *J* = 7.1 Hz, 2H, C**H$_2$**CH$_3$), 3.60 (q, *J* = 6.0 Hz, 2H, 1'''-H), 3.52 (td, *J* = 6.9, 5.4 Hz, 2H, 3-H), 3.06 (s, 2H, 2'''-H), 2.60 (t, *J* = 7.0 Hz, 2H, 2-H), 1.18 (t, *J* = 7.1 Hz, 3H, CH$_3$). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 171.3 (C-1), 168.7 (1'''-NHCO), 165.5 (3-NHCO), 163.0 (C-4''), 153.4 (C-4'), 144.5 (C-1''), 136.0 (C-1'), 128.5 (C-2' and C-6'), 126.5 (C-6''), 126.0 (C-2''), 122.0 (C-5'), 118.6 (C-5''), 116.1 (C-3''), 60.0 (**C**H$_2$CH$_3$), 38.5 (C-2'''), 37.0 (C-1'''), 35.6 (C-3), 33.7 (C-

2), 14.1 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3282, 3081, 2921, 1723, 1690, 1631, 1550, 1480, 1302, 1258, 1200, 1161, 1128, 811, 974, 720. **HR-MS (ESI):** *m/z* = $[M+H]^+$ calcd for $C_{21}H_{26}N_5O_5^+$: 428.1928; found: 428.1924. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

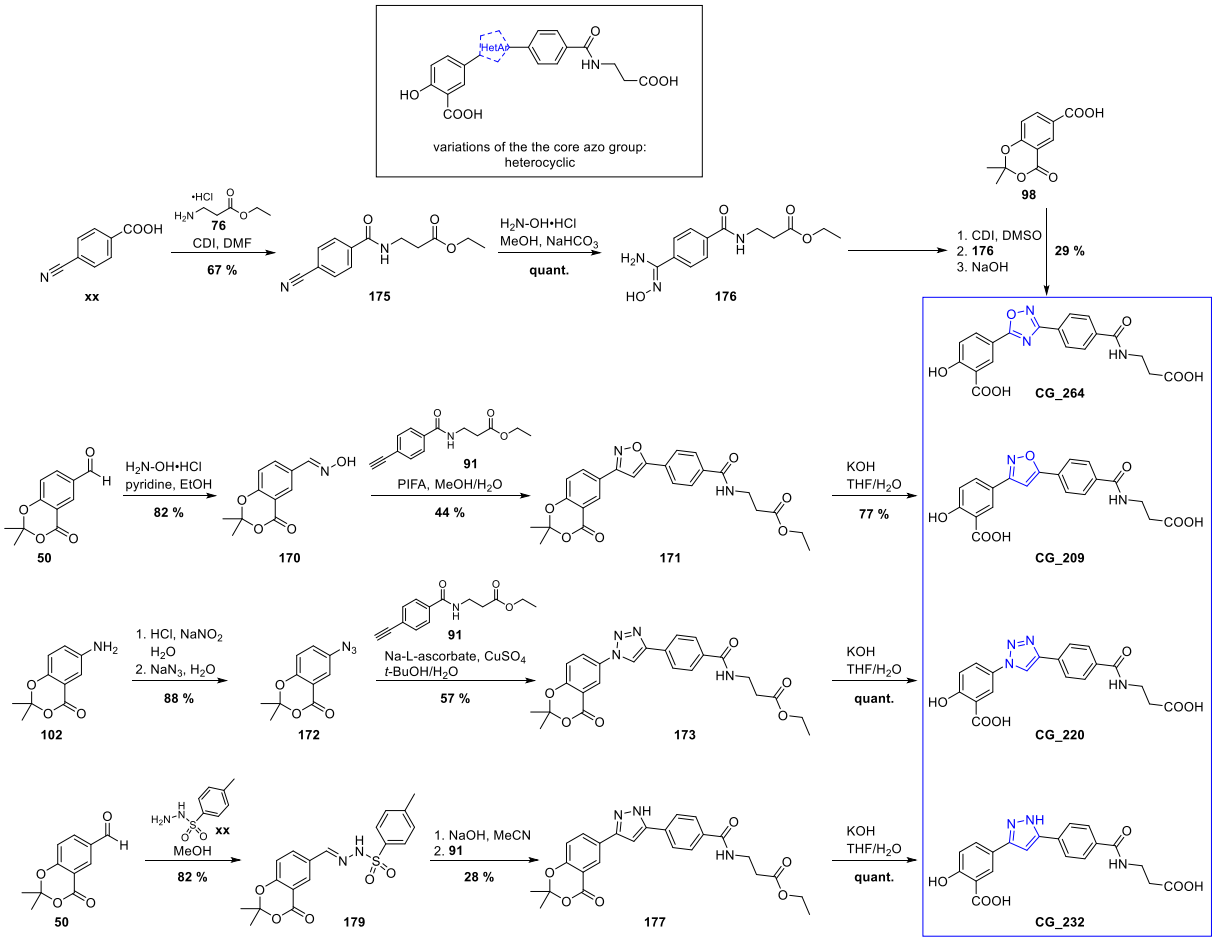
(*E*)-3-(4-((3-((2-Aminoethyl)carbamoyl)-4-hydroxyphenyl)diazenyl)benzamido)propanoic acid sodium salt (**CG_190**)



Prepared according to **General Procedure D** from compound **169** (40.0 mg, 0.0739 mmol) at room temperature. After 1.5 h, the pH of the reaction mixture was adjusted to 5 (calculated isoelectric point of the product) with 1 M aq. NaOH/1 M aq. HCl and concentrated *in vacuo*. The crude product was dissolved in MeOH/water (1:9) and an SPE purification was performed (two column volumes of MeOH/water (1:9) to wash out any remaining salts, followed by two column volumes MeOH to elute the product). Product **CG_190** (17.4 mg, 0.0413 mmol, 56 %) was obtained as an orange solid.

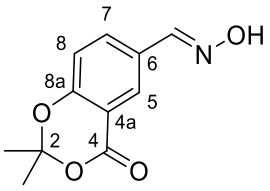
R_f: not determinable. **m.p.:** 234 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 11.67 (s, 1H, 1'''-NHCO), 8.57 (t, *J* = 5.4 Hz, 1H, 3-NHCO), 8.37 (d, *J* = 2.8 Hz, 1H, 2''-H), 7.97 – 7.89 (m, 2H, 2'-H and 6'-H), 7.74 – 7.70 (m, 2H, 3'-H and 5'-H), 7.68 (dd, *J* = 9.1, 2.9 Hz, 1H, 6''-H), 6.44 (d, *J* = 9.2 Hz, 1H, 5''-H), 3.56 – 3.50 (m, 1H, 1'''-H), 3.48 (q, *J* = 7.0 Hz, 1H, 3-H), 2.98 (s, 2H, 2'''-H), 2.56 – 2.52 (m, 2H, 2-H), NH_2 not clearly visible. **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 176.7 (C-4''), 173.0 (C-1), 169.0 (1'''-NHCO), 165.8 (3-NHCO), 154.7 (C-4'), 138.6 (C-1''), 133.1 (C-1'), 131.1 (C-2''), 128.2 (C-2' and C-6'), 124.7 (C-6''), 123.4 (C-5''), 120.8 (C-3' and C-5'), 117.9 (C-3''), 40.0 (C-2'''), 36.8 (C-1'''), 35.7 (C-3), 34.0 (C-2). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3632, 3295, 2359, 1628, 1529, 1491, 1409, 1360, 1300, 1258, 1181, 894, 857, 840, 767, 662. **HR-MS (ESI):** *m/z* = $[M-H]^-$ calcd for $C_{19}H_{20}N_5O_5^-$: 398.1470; found: 398.1468. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

Synthesis of hetero-triaryls CG_264, CG_209, CG_220, and CG_232



Scheme S7: Preparation of hetero-triaryls **CG_264**, **CG_209**, **CG_220** and **CG_232**.

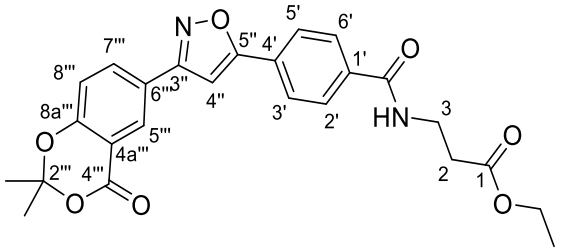
(*E*)-2,2-Dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxine-6-carbaldehyde oxime (**170**) [21]



Aldehyde **50** (200 mg, 0.970 mmol), hydroxylamine hydrochloride (135 mg, 1.94 mmol) and pyridine (204 µL, 2.52 mmol) were dissolved in EtOH (5.0 mL) and the reaction mixture was heated to 80 °C for 16 h. The reaction mixture was cooled to room temperature and the solvent evaporated *in vacuo*. Water (10 mL) and EtOAc (10 mL) were added, layers were separated, and the aqueous phase further extracted with EtOAc (2 x 15 mL). The combined organic layers were dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 30:70) to give aldoxime **170** (177 mg, 0.798 mmol, 82 %) as a white solid.

R_f: 0.44 (EtOAc/hexanes 30:70). **m.p.:** 160 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 8.12 (s, 1H, C**H**=N-OH), 8.10 (d, *J* = 2.2 Hz, 1H, 5-H), 7.88 (dd, *J* = 8.6, 2.2 Hz, 1H, 7-H), 7.00 (d, *J* = 8.6 Hz, 1H, 8-H), 1.75 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 160.7 (C-4), 157.3 (C-8a), 148.7 (CH=N-OH), 134.2 (C-7), 129.1 (C-5), 127.0 (C-6), 118.2 (C-8), 113.8 (C-4a), 107.0 (C-2), 26.0 ($C(\mathbf{C}H_3)_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3289, 2946, 1740, 1616, 1498, 1380, 1298, 1263, 1202, 1136, 1115, 1048, 967, 946, 848, 751, 667. **HR-MS (EI):** *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{11}H_{11}NO_4^{\bullet+}$: 221.0683; found: 221.0689. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

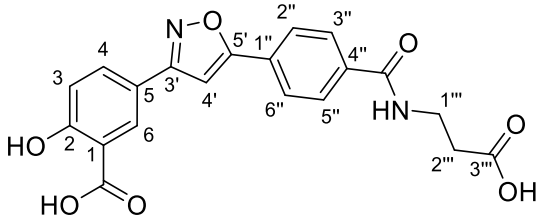
Ethyl 3-(4-(3-(2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)isoxazol-5-yl)benzamido)propanoate (**171**)



Aldoxime **170** (67.6 mg, 0.306 mmol) and alkyne **91** (50.0 mg, 0.204 mmol) were suspended in MeOH/water (5:1; 2.0 mL, 0.1 m with respect to alkyne). PIFA (131 mg, 0.306 mmol) was added as 0.5 eq per 2 h. After addition of 1.5 eq of PIFA in total, the reaction mixture was stirred at room temperature for 16 h. Then EtOAc (15 mL) was added, the mixture filtered through a pad of silica and solvents evaporated *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 50:50) to give isoxazole **171** (41.7 mg, 0.0898 mmol, 44 %) as a white solid.

R_f: 0.24 (EtOAc/hexanes 50:50). **m.p.:** 200 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.37 (d, *J* = 2.3 Hz, 1H, 5'''-H), 8.19 (dd, *J* = 8.6, 2.2 Hz, 1H, 7'''-H), 7.90 (s, 4H, 2'-, 3'-, 5'- and 6'-H), 7.10 (d, *J* = 8.6 Hz, 1H, 8'''-H), 6.97 (t, *J* = 6.1 Hz, 1H, CONH), 6.94 (s, 1H, 4''-H), 4.19 (q, *J* = 7.2 Hz, 2H, $C\mathbf{H_2}CH_3$), 3.76 (q, *J* = 6.0 Hz, 2H, 3-H), 2.67 (t, *J* = 5.9 Hz, 2H, 2-H), 1.79 (s, 6H, $C(CH_3)_2$), 1.29 (t, *J* = 7.2 Hz, 3H, $CH_2C\mathbf{H_3}$). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 173.2 (C-1), 169.9 (C-5''), 166.4 (CONH), 161.7 (C-3''), 160.7 (C-4'''), 157.5 (C-8a'''), 136.0 (C-1'), 134.7 (C-7'''), 129.9 (C-4'), 128.3 (C-5'''), 127.9 (C-2' and C-6'), 126.1 (C-3' and C-5'), 124.0 (C-6'''), 118.4 (C-8'''), 113.9 (C-4a'''), 107.1 (C-2'''), 98.5 (C-4''), 61.1 ($\mathbf{C}H_2CH_3$), 35.6 (C-3), 34.0 (C-2), 26.0 ($C(\mathbf{C}H_3)_2$), 14.3 ($CH_2\mathbf{C}H_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3408, 2981, 2905, 1735, 1724, 1648, 1624, 1539, 1500, 1378, 1292, 1251, 1181, 1142, 1056, 927, 850, 764. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{25}H_{25}N_2O_7^+$: 465.1656; found: 465.1654. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

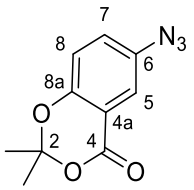
5-(5-(4-((2-Carboxyethyl)carbamoyl)phenyl)isoxazol-3-yl)-2-hydroxybenzoic acid (**CG_209**)



Prepared according to **General Procedure D** at room temperature from isoxazole **171** (35.0 mg, 0.0754 mmol). The organic material was extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried using a phase separation paper and concentrated *in vacuo*. The pure product **CG_209** (23.0 mg, 0.0580 mmol, 77 %) was obtained as a white solid.

R_f: 0.27 (EtOAc/hexanes+AcOH 80:20+1). **m.p.:** 272 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.24 (s, 1H, 1-COOH or 2'''-COOH), 8.69 (t, *J* = 5.5 Hz, 1H, CONH), 8.33 (d, *J* = 2.3 Hz, 1H, 6-H), 8.05 (dd, *J* = 8.6, 2.3 Hz, 1H, 4-H), 8.01 (s, 4H, 2''-, 3''-, 5''- and 6''-H), 7.76 (s, 1H, 4'-H), 7.13 (d, *J* = 8.7 Hz, 1H, 3-H), 3.49 (q, *J* = 6.7 Hz, 2H, 1'''-H), 2.54 (t, *J* = 7.2 Hz, 2H, 2'''-H). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (C-3'''), 171.3 (1-COOH), 168.9 (C-5'), 165.4 (CONH), 162.6 (C-2), 161.8 (C-3'), 135.7 (C-4''), 133.3 (C-4), 129.0 (C-1''), 128.6 (C-6), 128.1 (C-3'' and C-5''), 125.5 (C-2'' and C-6''), 119.4 (C-5), 118.2 (C-3), 114.0 (C-1), 99.5 (C-4'), 35.7 (C-1'''), 33.7 (C-2'''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3332, 2922, 2853, 1697, 1668, 1638, 1587, 1538, 1423, 1399, 1286, 1237, 1201, 855, 794, 769. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{20}H_{15}N_2O_7^-$: 395.0885; found: 395.0885. **Purity (HPLC):** 210 nm: 93 %; 254 nm: 93 % (method 2a).

6-Azido-2,2-dimethyl-4*H*-benzo[*d*][1,3]dioxin-4-one (**172**)

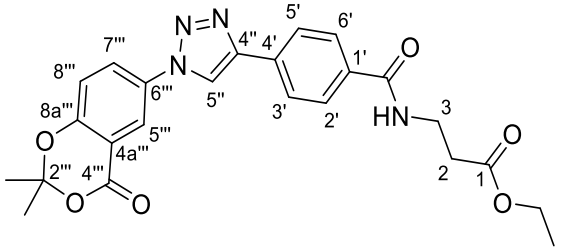


Conc. HCl (0.62 mL) and a solution of $NaNO_2$ (85.7 mg, 1.24 mmol) in water (0.16 mL) was slowly added to amine **102** (200 mg, 1.04 mmol) in water (0.60 mL) at 0 °C. After 10 min, a solution of NaN_3 (1.24 mg, 1.24 mmol) in water (0.16 mL) was added dropwise and the reaction mixture stirred at 0 °C for 40 min. The mixture was then extracted with diethyl ether (3 x 25 mL). The combined organic extracts were dried using a phase separation paper and concentrated *in vacuo* to give product **172** (199 mg, 0.906 mmol, 88 %) as a yellow solid.

R_f: 0.42 (EtOAc/hexanes 10:90). **m.p.:** 85 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 7.64 (d, *J* = 2.8 Hz, 1H, 5-H), 7.18 (dd, *J* = 8.7, 2.8 Hz, 1H, 7-H), 6.97 (d, *J* = 8.8 Hz, 1H, 8-H), 1.73 (s,

6H, C(CH_3)$_2$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) =160.4 (C-4), 153.1 (C-8a), 135.1 (C-6), 127.3 (C-7), 119.0 (C-5), 118.9 (C-8), 114.5 (C-4a), 106.8 (C-2), 25.8 (C(**C**H_3)$_2$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 2132, 1727, 1484, 1439, 1380, 1328, 1294, 1242, 1198, 1138, 1050, 979, 934, 836, 780, 749. **HR-MS (EI):** *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{10}H_9N_3O_3^{\bullet+}$: 219.0638; found: 219.0638. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

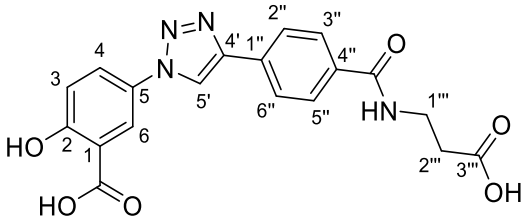
Ethyl 3-(4-(1-(2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)-1H-1,2,3-triazol-4-yl)benzamido)propanoate (**173**)



An alkyne **91** stock solution (0.33 m in *tert*-BuOH, 600 µL, 0.200 mmol) and an azide **172** stock solution (0.13 m in *tert*-BuOH/DMSO 21:1, 628 µL, 0.0800 mmol) were added in a 15 mL round bottom flask. Then, a $CuSO_4$ stock solution (26.7 mm in water, 600 µL, 16.0 µmol) and a sodium ascorbate stock solution (0.13 m in water, 600 µL, 0.0800 mmol) were added. The reaction mixture was stirred at room temperature for 4 days. EtOAc (20 mL) was added, and the organic phase was washed with 1 M aq. HCl (10 mL), sat. aq. $NaHCO_3$ (10 mL) and brine (10 mL). The organic phase was dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 60:40) to give triazole **173** (21.0 mg, 0.0452 mmol, 57 %) as a white solid.

R_f: 0.29 (EtOAc/hexanes 60:40). **m.p.:** 215 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 8.29 (s, 1H, 5''-H), 8.26 (d, *J* = 2.6 Hz, 1H, 5'''-H), 8.16 (dd, *J* = 8.9, 2.7 Hz, 1H, 7'''-H), 7.98 (d, *J* = 8.2 Hz, 2H, 3'-H and 5'-H), 7.88 (d, *J* = 8.1 Hz, 2H, 2'-H and 6'-H), 7.20 (d, *J* = 8.9 Hz, 1H, 8'''-H), 6.93 (t, *J* = 6.1 Hz, 1H, CONH), 4.19 (q, *J* = 7.1 Hz, 2H, C**H_2**CH_3), 3.76 (q, *J* = 5.9 Hz, 2H, 3-H), 2.67 (t, *J* = 5.8 Hz, 2H, 2-H), 1.80 (s, 6H, C(CH_3)$_2$), 1.29 (t, *J* = 7.2 Hz, 3H, CH_2C**H_3**). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 173.2 (C-1), 166.8 (CONH), 160.1 (C4'''), 156.2 (C-8a'''), 147.9 (C-4''), 134.4 (C-1'), 133.1 (C-4'), 132.1 (C-6'''), 129.1 (C-7'''), 127.9 (C-2' and C-6'), 126.1 (C-3' and C-5'), 120.9 (C-5'''), 119.4 (C-8'''), 118.4 (C-5''), 114.3 (C-4a'''), 107.5 (C-2'''), 61.0 (**C**$H_2$$CH_3$), 35.5 (C-3), 34.1 (C-2), 26.0 (C(**C**H_3)$_2$), 14.4 (CH_2**C**H_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3414, 1739, 1725, 1647, 1533, 1506, 1492, 1302, 1184, 1149, 1057, 1030, 930, 826, 768. **HR-MS (ESI):** *m*/*z* = $[M+H]^+$ calcd for $C_{24}H_{25}N_4O_6^+$: 465.1769; found: 465.1765. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

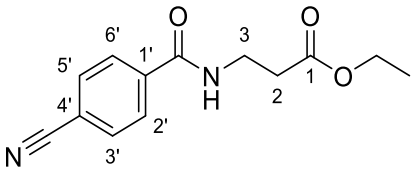
5-(4-(4-((2-Carboxyethyl)carbamoyl)phenyl)-1*H*-1,2,3-triazol-1-yl)-2-hydroxybenzoic acid (**CG_220**)



Prepared according to **General Procedure D** from triazole **173** (30.0 mg, 0.0646 mmol) at room temperature. The solvents were evaporated *in vacuo*. To the crude product was added MeOH (10 mL) and the suspension filtered. The filtrate was concentrated *in vacuo* to give product **CG_220** (25.6 mg, 0.0646 mmol, quant.) as a white solid.

R_f: 0.10 (EtOAc/hexanes+AcOH 80:20+1). **m.p.:** 295 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.22 (s, 2H, 1-COOH and 2'''-COOH), 9.41 (s, 1H, 5'-H), 8.62 (t, *J* = 5.5 Hz, 1H, CONH), 8.30 (d, *J* = 2.8 Hz, 1H, 6-H), 8.08 (dd, *J* = 9.0, 2.8 Hz, 1H, 4-H), 8.03 (d, *J* = 8.5 Hz, 2H, 2''-H and 6''-H), 7.96 (d, *J* = 8.5 Hz, 2H, 3''-H and 5''-H), 7.22 (d, *J* = 8.9 Hz, 1H, 3-H), 3.48 (td, *J* = 7.1, 5.3 Hz, 2H, 1'''-H), 2.54 (t, *J* = 7.1 Hz, 2H, 2'''-H). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 172.9 (C-3'''), 170.8 (1-COOH), 165.7 (CONH), 160.9 (C-2), 146.5 (C-4'), 133.8 (C-4''), 132.9 (C-1''), 128.5 (C-5), 127.9 (C-3'' and C-5''), 127.3 (C-4), 125.0 (C-2'' and C-6''), 121.6 (C-6), 120.5 (C-5'), 118.7 (C-3), 114.1 (C-1), 35.6 (C-1'''), 33.8 (C-2'''). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3310, 3076, 2921, 2852, 1677, 1633, 1621, 1536, 1447, 1294, 1219, 1199, 1186, 1040, 850, 829, 767, 694. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{19}H_{15}N_4O_6^-$: 395.0997; found: 395.0994. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

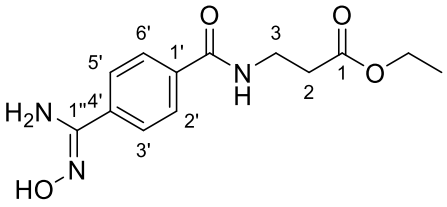
Ethyl 3-(4-cyanobenzamido)propanoate (**175**)



4-Cyanobenzoic acid (**174**, 368 mg, 2.50 mmol) was treated with solid CDI (422 mg, 2.60 mmol) in dry DMF (1.5 mL) at 0 °C. After the initial vigorous evolution of gas, the reaction mixture was allowed to warm to room temperature and stirred until all gas evolution ceased. Solid *β*-alanine ethyl ester hydrochloride (**76**, 388 mg, 2.52 mmol) was added, and the mixture was stirred until a clear solution resulted. Water (10 mL) was added, and the mixture extracted with EtOAc (3 x 15 mL). The combined organic phase was washed with 0.5 M aq. NaOH (15 mL) and brine (15 mL), dried using a phase separation paper and concentrated *in vacuo* to give product **175** (414 mg, 1.68 mmol, 67 %) as a white solid.

R_f: 0.25 (EtOAc/hexanes+AcOH 40:60+1). **m.p.:** 82 °C. **^{1}H-NMR (500 MHz, $CDCl_3$):** δ (ppm) = 7.88 – 7.85 (m, 2H, 2'-H and 6'-H), 7.74 – 7.72 (m, 2H, 3'-H and 5'-H), 7.00 (t, *J* = 5.3, 4.7 Hz, 1H, CONH), 4.18 (q, *J* = 7.1 Hz, 2H, C**H_2**CH_3), 3.73 (q, *J* = 6.0 Hz, 2H, 3-H), 2.65 (t, *J* = 5.9 Hz, 2H, 2-H), 1.28 (t, *J* = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (126 MHz, $CDCl_3$):** δ (ppm) = 173.2 (C-1), 165.6 (CONH), 138.4 (C-1'), 132.6 (C-3' and C-5'), 127.8 (C-2' and C-6'), 118.2 (CN), 115.3 (C-4'), 61.1 (**C**H_2CH_3), 35.6 (C-3), 33.8 (C-2), 14.3 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3344, 3298, 2980, 2230, 1727, 1637, 1543, 1324, 1278, 1192, 1150, 1090, 1026, 857. **HR-MS (EI):** *m*/*z* = $[M]^{\bullet+}$ calcd for $C_{13}H_{14}N_2O_3^{\bullet+}$: 246.0999; found: 246.0998. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

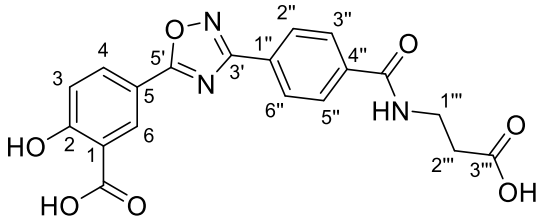
Ethyl (*Z*)-3-(4-(*N*'-hydroxycarbamimidoyl)benzamido)propanoate (**176**)



To a stirred solution of nitrile **175** (300 mg, 1.22 mmol) in MeOH (3.0 mL), $NaHCO_3$ (113 mg, 1.34 mmol) and hydroxylamine hydrochloride (84.7 mg, 1.22 mmol) were added. The resulting reaction mixture was heated to reflux for 4 h. Then, the reaction mixture was cooled to room temperature and filtered. The filtrate was concentrated under reduced pressure to give product **176** (339 mg, 1.21 mmol, quant.) as a white solid.

R_f: 0.30 (EtOAc/hexanes 90:10). **m.p.:** 159 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** δ (ppm) = 9.79 (s, 1H, OH), 8.62 (t, *J* = 5.6 Hz, 1H, CONH), 7.84 – 7.80 (m, 2H, 2'-H and 6'-H), 7.76 – 7.73 (m, 2H, 3'-H and 5'-H), 5.89 (s, 2H, NH_2), 4.06 (q, *J* = 7.1 Hz, 2H, C**H_2**CH_3), 3.49 (td, *J* = 6.9, 5.4 Hz, 2H, 3-H), 2.58 (t, *J* = 7.0 Hz, 2H, 2-H), 1.17 (t, *J* = 7.1 Hz, 3H, CH_3). **^{13}C-NMR (101 MHz, $(CD_3)_2SO$):** δ (ppm) = 171.3 (C-1), 165.8 (CONH), 150.2 (C-1''), 135.8 (C-4'), 134.4 (C-1'), 127.0 (C-2' and C-6'), 125.1 (C-3' and C-5'), 59.9 (**C**H_2CH_3), 35.5 (C-3), 33.8 (C-2), 14.1 (CH_3). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3442, 3299, 2983, 1653, 1623, 1539, 1472, 1372, 1323, 1300, 1267, 1205, 1021, 929, 856, 821, 776, 687. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{13}H_{16}N_3O_4^-$: 278.1146; found: 278.1147. **Purity (HPLC):** 210 nm: 92 %; 254 nm: 92 % (method 2a).

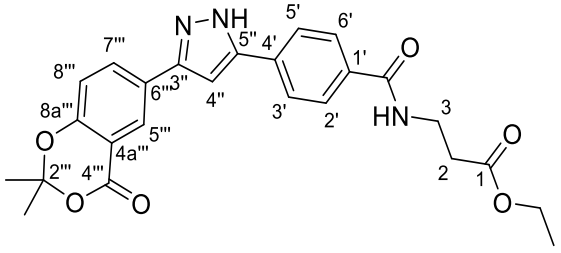
5-(3-(4-((2-Carboxyethyl)carbamoyl)phenyl)-1,2,4-oxadiazol-5-yl)-2-hydroxybenzoic acid (**CG_264**)



Under N_2 atmosphere, carboxylic acid **98** (100 mg, 0.450 mmol) was dissolved in dry DMSO (0.50 mL). Solid CDI (79.6 mg, 0.491 mmol) was added, and the reaction mixture stirred at room temperature for 30 min. After this time, amidoxime **176** (114 mg, 0.409 mmol) was added. The reaction mixture was stirred at room temperature for another 18 h. Then, solid NaOH (19.6 mg, 0.491 mmol) was added rapidly. The mixture was stirred at room temperature for 2 h and afterwards diluted with cold water. The resulting precipitate was collected by filtration, washed with cold water (15 mL), and dried *in vacuo*. The crude product was purified by FCC (MeOH/DCM+AcOH 4:96+1 → 7:93+1) to give 1,2,4-oxadiazole **CG_264** (46.6 mg, 0.117 mmol, 29 %) as a pink solid.

R_f: 0.31 (MeOH/DCM+AcOH 5:95+1). **m.p.:** 248 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.25 (s, 1H, 1-COOH or 2'''-COOH), 8.74 (t, *J* = 5.4 Hz, 1H, CONH), 8.56 (d, *J* = 2.3 Hz, 1H, 6-H), 8.20 (dd, *J* = 8.7, 2.3 Hz, 1H, 4-H), 8.17 (dd, *J* = 8.3, 1.7 Hz, 2H, 2''-H and 6''-H), 8.04 – 8.01 (m, 2H, 3''-H and 5''-H), 7.12 (d, *J* = 8.7 Hz, 1H, 3-H), 3.49 (td, *J* = 7.1, 5.4 Hz, 2H, 1'''-H), 2.55 (t, *J* = 7.1 Hz, 2H, 2'''-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** δ (ppm) = 175.2 (C-5'), 172.9 (C-3'''), 170.5 (1-COOH), 167.6 (C-3'), 166.0 (C-2), 165.5 (CONH), 136.9 (C-4''), 133.8 (C-4), 130.8 (C-6), 128.6 (C-3'' and C-5''), 127.1 (C-2'' and C-6''), 118.7 (C-3), 115.8 (C-5), 113.2 (C-1), 35.7 (C-1'''), 33.7 (C-2'''), C-1'' not visible. **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3307, 2925, 1688, 1632, 1548, 1471, 1411, 1354, 1302, 1247, 1206, 909, 863, 756, 720. **HR-MS (ESI):** *m*/*z* = $[M-H]^-$ calcd for $C_{19}H_{14}N_3O_7^-$: 396.0837; found: 396.0835. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

Ethyl 3-(4-(3-(2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)-1*H*-pyrazol-5-yl)benzamido)-propanoate (**177**)



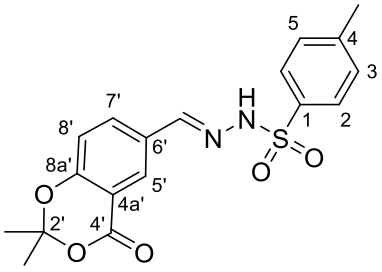
Method A:

To compound **179** (53.4 mg, 0.143 mmol; 0.5 eq) in MeCN (1.2 mL), 5 M aq. NaOH (28.5 µL) was added and the mixture was stirred for 20 min. Then the mixture was added to the dipolarophile **91** (70.0 mg, 0.285 mmol; 1.0 eq), and the reaction mixture was stirred at 50 °C for 6 h. Then again, a mixture of **179** (0.5 eq) in MeCN and 5 M aq. NaOH (stirred for 20 min) was added and the mixture heated to 50 °C for further 16 h. After this time, again a mixture of **179** (0.5 eq) in MeCN and 5 N aq. NaOH (stirred for 20 min) was added and the mixture stirred further at 50 °C for 16 h, now containing 1.5 eq of compound **179** in total. After cooling to room temperature, EtOAc (15 mL) was added and the organic phase was washed with 1 M aq. HCl (10 mL), sat. aq. $NaHCO_3$ (10 mL) and brine (10 mL). The organic phase was dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 60:40) to give pyrazole **177** (37.0 mg, 0.0798 mmol, 28 %) as a pale yellow solid.

Method B (not depicted in **Scheme S7**):
To a mixture of $PdCl_2$ (5.0 mol%, 3.87 mg, 0.0218 mmol), $Cr(CO)_6$ (96.1 mg, 0.437 mmol), iodo compound **93** (133 mg, 0.437 mmol), and alkyne **91** (150 mg, 0.612 mmol) in DMF (1.5 mL) was added hydrazine monohydrate (31.8 µL, 0.655 mmol). The resulting mixture was heated to 70 °C for 4 h, then cooled to room temperature and quenched with water (10 mL). The mixture was extracted with EtOAc (3 x 10 mL), the organic phase was dried using a phase separation paper and concentrated *in vacuo*. The crude product was purified by FCC (EtOAc/hexanes 65:35) to give pyrazole **177** (15.0 mg, 0.0324 mmol, 7 %) as a pale yellow solid.

R_f: 0.13 (EtOAc/hexanes 60:40). **m.p.:** 197 °C. **^{1}H-NMR (400 MHz, $(CD_3)_2SO$):** due to tautomerisation NMR measured at 100 °C. δ (ppm) = 13.26 (s, 1H, pyrazole-H), 8.32 (d, J = 2.2 Hz, 1H, 5'''-H), 8.23 (t, J = 5.1 Hz, 1H, CONH), 8.13 (d, J = 8.6 Hz, 1H, 7'''-H), 7.90 (s, 4H, 2'-, 3'-, 5'- and 6'-H), 7.21 (s, 1H, 4''-H), 7.19 (d, J = 8.6 Hz, 1H, 8'''-H), 4.12 (q, J = 7.1 Hz, 2H, C**H_2**CH_3), 3.56 (td, J = 7.0, 5.6 Hz, 2H, 3-H), 2.61 (t, J = 7.0 Hz, 2H, 2-H), 1.75 (s, 6H, $C(CH_3)_2$), 1.21 (t, J = 7.1 Hz, 3H, CH_2C**H_3**). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** due to tautomerisation NMR measured at 100 °C. δ (ppm) = 170.7 (C-1), 165.6 (CONH), 159.4 (C-4'''), 154.6 (C-8a'''), 133.2 (C-7'''), 127.2 (C-3' and C-5'), 124.9 (C-5'''), 124.5 (C-2' and C-6'), 117.3 (C-8'''), 113.5 (C-6'''), 113.1 (C-4a'''), 106.1 (**C**$(CH_3)_2$), 99.9 (C-4''), 59.3 (**C**H_2CH_3), 35.2 (C-3), 33.6 (C-2), 25.0 (C(**C**$H_3)_2$), 13.5 (CH_2**C**H_3), at 100 °C C-3'' and C-5'' not visible. **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3315, 3238, 2927, 1742, 1702, 1628, 1542, 1494, 1277, 1259, 1183, 1139, 930, 796, 769. **HR-MS (ESI):** m/z = $[M+H]^+$ calcd for $C_{25}H_{26}N_3O_6^+$: 464.1816; found: 464.1814. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

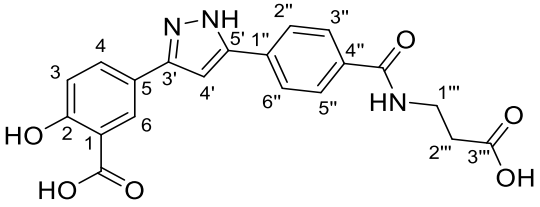
(*E*)-*N*'-((2,2-dimethyl-4-oxo-4*H*-benzo[*d*][1,3]dioxin-6-yl)methylene)-4-methylbenzenesulfonohydrazide (**179**)



To a stirred solution of *p*-toluenesulfonyl hydrazide (**178**, 93.1 mg, 0.500 mmol) in MeOH (1.2 mL), aldehyde **50** (103 mg, 0.500 mmol) was added portionwise at room temperature. After 3 h, the reaction mixture was cooled to 0 °C and the product was collected by filtration, washed with a small quantity of cold MeOH (5.0 mL) and dried *in vacuo* to give the product **179** (153.3 mg, 0.409 mmol, 82 %) as a white solid.

R_f: 0.37 (EtOAc/hexanes 35:65). **m.p.:** 199 °C. **^{1}H-NMR (400 MHz, $CDCl_3$):** δ (ppm) = 8.02 (d, J = 2.2 Hz, 1H, 5'-H), 7.92 (dd, J = 8.7, 2.2 Hz, 1H, 7'-H), 7.89 – 7.85 (m, 2H, 2-H and 6-H), 7.75 (s, 1H, 6'-CH), 7.33 (d, J = 8.1 Hz, 2H, 3-H and 5-H), 6.97 (d, J = 8.6 Hz, 1H, 8'-H), 2.41 (s, 3H, 4-CH_3), 1.73 (s, 6H, $C(CH_3)_2$). **^{13}C-NMR (101 MHz, $CDCl_3$):** δ (ppm) = 160.6 (C-4'), 157.6 (C-8a'), 145.7 (6'-**C**H), 144.6 (C-4), 135.3 (C-1), 134.3 (C-7'), 129.9 (C-3 and C-5), 129.5 (C-5'), 128.4 (C-6'), 128.1 (C-2 and C-6), 118.2 (C-8'), 113.6 (C-4a'), 107.1 (C-2'), 26.0 ($C(\mathbf{C}H_3)_2$), 21.8 (4-$\mathbf{C}H_3$). **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3247, 1716, 1623, 1498, 1362, 1286, 1217, 1163, 1067, 1051, 966, 932, 881, 809, 667. **HR-MS (ESI):** m/z = $[M+H]^+$ calcd for $C_{18}H_{19}N_2O_5S^+$: 375.1009; found: 375.1007. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

5-(5-(4-((2-Carboxyethyl)carbamoyl)phenyl)-1*H*-pyrazol-3-yl)-2-hydroxybenzoic acid (**CG_232**)



Prepared according to **General Procedure D** from compound **177** (30.0 mg, 0.0646 mmol) at room temperature. The solvents were evaporated *in vacuo*. To the crude product water (10 mL) was added, the remaining solid collected by filtration and dried *in vacuo* to give pyrazole **CG_232** (21.3 mg, 0.0539 mmol, quant.) as a beige solid.

R_f: 0.10 (EtOAc/hexanes+AcOH 80:20+1). **m.p.:** 288 °C. **^{1}H-NMR (500 MHz, $(CD_3)_2SO$):** δ (ppm) = 12.42 (s, 1H, 1-COOH or 2'''-COOH), 11.34 (s, 1H, pyrazol-NH), 8.57 (t, J = 5.5 Hz, 1H, CONH), 8.27 (d, J = 2.3 Hz, 1H, 6-H), 7.98 (dd, J = 8.6, 2.3 Hz, 1H, 4-H), 7.95 – 7.88 (m,

4H, 2''-, 3''-, 5''- and 6''-H), 7.26 (s, 1H, 4'-H), 7.07 (d, *J* = 8.6 Hz, 1H, 3-H), 3.48 (td, *J* = 7.1, 5.4 Hz, 2H, 1'''-H), 2.53 (t, *J* = 7.2 Hz, 2H, 2'''-H). **^{13}C-NMR (126 MHz, $(CD_3)_2SO$):** *δ* (ppm) = 172.9 (C-3'''), 171.7 (1-COOH), 165.8 (CONH), 160.7 (C-2), 133.3 (C-1'' and C-4''), 132.5 (C-4), 127.8 (C-3'' and C-5''), 126.7 (C-6), 124.8 (C-2'' and C-6''), 122.7 (C-5), 117.8 (C-3), 113.3 (C-1), 99.8 (C-4'), 35.6 (C-1'''), 33.8 (C-2'''), C-3' and C-5' not visible. **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3391, 3141, 2923, 2503, 1722, 1666, 1598, 1568, 1546, 1489, 1301, 1181, 865, 858, 801, 770. **HR-MS (ESI):** *m/z* = $[M-H]^-$ calcd for $C_{20}H_{16}N_3O_6^-$: 394.1045; found: 394.1047. **Purity (HPLC):** 210 nm: >95 %; 254 nm: >95 % (method 2a).

Part B: Biological Data

Known small molecule Sirt5 inhibitors

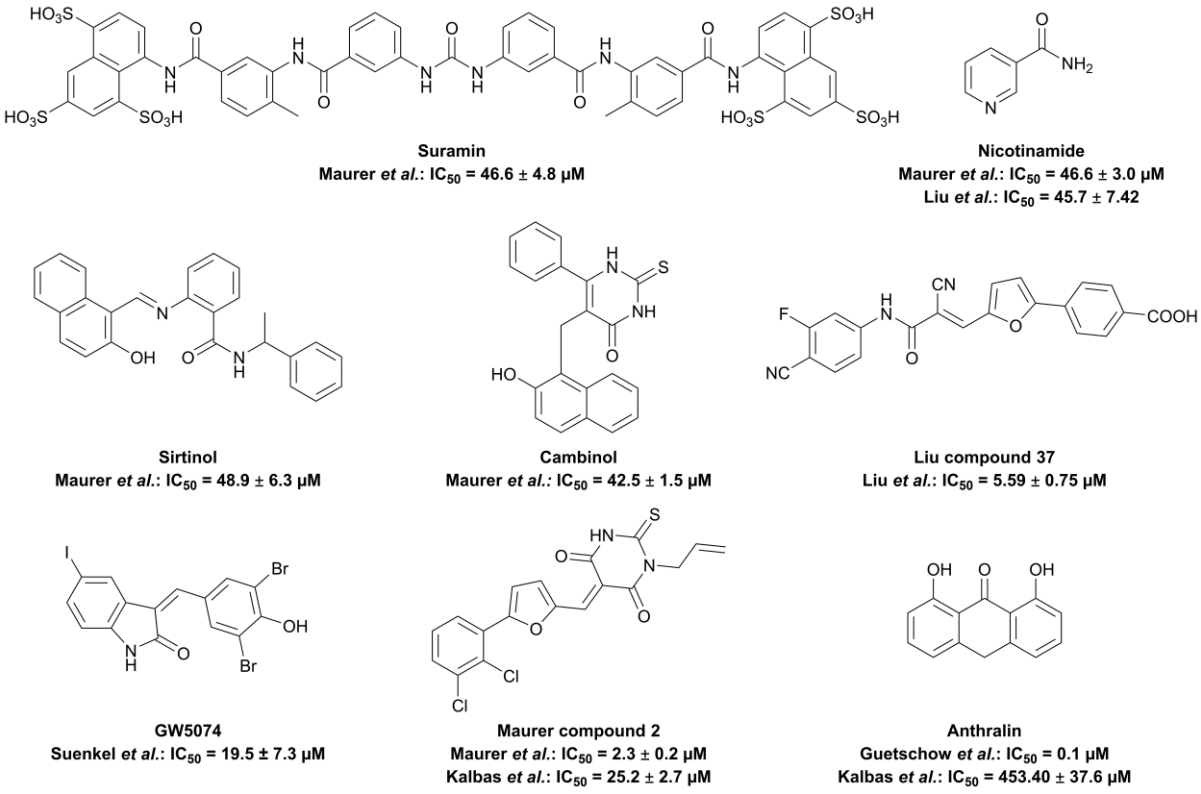


Figure S1: Structures and reported IC_{50} values of known small molecule Sirt5 inhibitors: suramin [22], nicotinamide [22-23], sirtinol [22], cambinol [22], Liu compound 37 [23], GW5074 [24], Maurer compound 2 [22, 25], anthralin [25-26].

MTT assay

For the MTT assay, [27] HL-60 cells were maintained under standard cell culture conditions in RPMI 1640 medium supplemented with 10 % fetal bovine serum (FCS) without antibiotics. Cells were grown and incubated at 37 °C in a 5 % CO_2 atmosphere.

HL-60 cells were seeded in 96-well plates at 9 x 10^4 cells/well and incubated for 24 h before treating with test compounds for 24 h with a final well volume of 100 µL, 1 % DMSO; three technical replicates; the co-solvent control was treated with 1 % DMSO only. Triton-X 100 was used as a positive control. 3-[4,5-dimethylthiazole-2-yl]-2,5-diphenyltetrazolium bromide (MTT; 5 mg in 1.0 mL PBS) was added to the cells for 2 h. Then 190 µL DMSO was added. After 1 h, absorbance of the MTT metabolite formazan was measured at 570 nm using a MRX Microplate Reader (Dynex Technologies, Chantilly, USA). Absorbance data was averaged over the technical replicates, then normalised to viable cell count from the co-solvent control cells (% control) as 100 %, where 0 % viability was assumed to correspond to absorbance zero.

All final compounds showed no cytotoxic effects (IC_{50} >50 µM).

Agar diffusion test

Solutions with 1 % (m/V) compound in DMSO were prepared. Of these solutions 3.0 µL were given on a test platelet (diameter 6 mm, Macherey-Nagel), equivalent to 30 µg substance. The same was done for the reference substances clotrimazole (antifungal) and tetracycline (antibacterial). Blind control was conducted with DMSO. The test platelets were then dried for 24 h at room temperature. Microorganisms were obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) in Braunschweig and cultivated according to the DSMZ recommendations in liquid culture. For the agar diffusion test, different agars were required. For Candida glabrata (DSM number: 11226), Hyphopichia burtonii (DSM number: 70663), Yarrowia lipolytica (DSM number: 1345), Escherichia coli (DSM number: 426) and Pseudomonas marginalis (DSM number: 7527) all-culture agar (AC-agar) of Sigma Aldrich was used. 35.2 g AC-agar and 20 g agar were suspended in 1.0 L water and autoclaved. For Staphylococcus equorum (DSM number: 20675) and Streptococcus entericus (DSM number: 14446) an agar is likewise prepared from 10.0 g casein peptone, 5.0 g yeast extract, 5.0 g glucose and 5.0 g sodium chloride in 1.0 L water. For Aspergillus niger (DSM number: 1988) 32 g potato dextrose agar and 20 g agar in 1.0 L water were used. After treatment in the autoclave 15 mL each of the warm, liquid agar was filled into Petri dishes under aseptic conditions and cooled to 8 °C for one hour. The germs were then brought onto the different agars using cotton swabs. The platelets containing the substances, the reference, and the blind control were put onto the agar. The agar plates were incubated for 36 h at 32 °C (bacteria) or 28 °C (yeasts). Then the diameters of growth inhibition were measured manually.

All final compounds showed no antimicrobial effects in this routine testing.

Enzyme based assays

FLUOR DE LYS® SIRT5 fluorometric drug discovery assay kit

For screening the inhibitory activity of synthesized inhibitors against recombinant human Sirt5 a commercially available Sirt5 fluorometric drug discovery kit (BML-AK513) from ENZO LIFE SCIENCES (ELS) AG (Lausen, Switzerland) was used. The assay was performed in Corning ½ volume 96-well black NBS microplates provided in the kit. In brief, for the desuccinylation reaction to take place, potential inhibitors (10 mM DMSO stock solutions) were incubated with Sirt5, FLUOR DE LYS®-succinyl, and NAD^+ in assay buffer for 60 min at 37 °C. After which time the deacetylation reaction was stopped by addition of FLUOR DE LYS®-developer and

nicotinamide. The mixture was incubated for 15 min. Fluorescence was measured on a FLUOstar Omega microplate reader (BMG Labtech, Ortenberg, Germany) (λ_{ex} = 355 nm, λ_{em} = 460 nm) and averaged over the technical replicates. A mixture with only DMSO was used as no inhibition control and set to 100 % enzyme activity. All values were normalized to this co-solvent control after blank subtraction. No compound showed autofluorescence under the given conditions.

IC_{50} values were determined with GraphPad Prism 8.0 software (La Jolla, CA). Dose response curves were fit using a nonlinear log(inhibitor) vs. response model with variable slope.

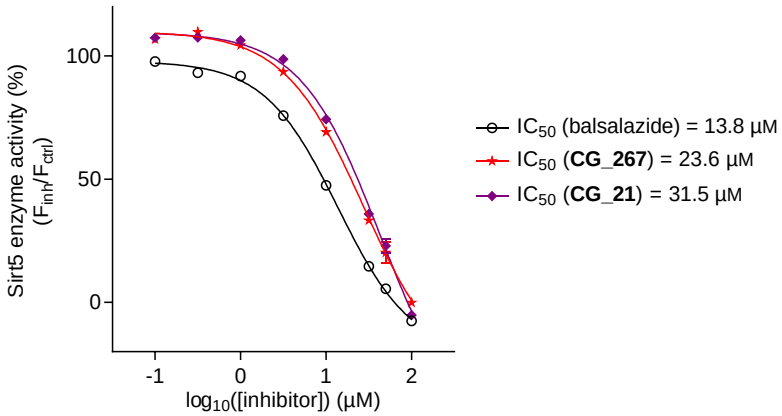


Figure S1: Inhibition of Sirt5 desuccinylase activity by balsalazide, **CG_267** and **CG_21** determined using the FLUOR DE LYS® SIRT5 fluorometric drug discovery assay kit resulting in residual Sirt5 enzyme activity after 1 h incubation with the inhibitors in % (F_{inh}/F_{ctrl}). F = fluorescence units; λ_{ex} = 360 nm, λ_{em} = 460 nm.

SIRT1, 2, and 3 fluorometric enzyme assay

Inhibition of recombinant **Sirt1/2/3** was determined using a homogeneous fluorescence deacetylase assay [28]. Sirt1/2/3 was mixed with the fluorescent substrate ZMAL (final assay concentration, 10.5 µM), NAD^+ (6 mM, final assay concentration 500 µM), the inhibitor in DMSO or DMSO as a control [5 % (v/v)], and filled up with assay buffer (25 mM Tris·HCl, 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$, pH 8.0) to a volume of 60 µL, incubated for 4 h (37 °C, 180 rpm), then treated with 60 µL of trypsin solution [50 mM Tris·HCl, 100 mM NaCl, 8 mM nicotinamide, 5.5 U/µL trypsin, 6.7 % DMSO (v/v)] and incubated again (37 °C, 180 rpm, 20 min). The fluorescence intensity was then measured with a microplate reader (BMG Polarstar, λ_{ex} = 390 nm, λ_{em} = 460 nm). To ensure initial state conditions, the conversion of ZMAL was adjusted to 10 – 30 % substrate conversion without inhibitor. A blank control without enzyme and a 100 % conversion control with the fluorescent metabolite AMC instead of ZMAL were performed. Inhibition rates were calculated in reference to the DMSO control after blank subtraction. All inhibition experiments were run at least twice in duplicates.

Chemical Proteomics

Cell culture and lysis:

MCF7 cells were maintained under standard cell culture conditions in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10 % fetal bovine serum (FBS). Cells were grown and incubated at 37 °C in a 5 % CO_2 atmosphere. Upon reaching 90 % confluency, cell medium was discarded, cells were washed twice with PBS and then lysed in lysis buffer containing 50 mM Tris/HCl, pH 7.5, 5 % glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 1 mM Na_3VO_4, 0.8 % Igepal, 0.375 mM NaF, 1 mM DTT including protease inhibitors (SigmaFast protease inhibitor tablet S8820) and phosphatase inhibitors (Phosphatase Inhibitor Cocktail 3, Sigma-Aldrich, Munich, Germany). After one freeze-thaw cycle (-80 °C) the lysate was ultracentrifuged for 30 min at 4 °C and 50000 g and supernatant was stored at -80 °C for further procedures.

Preparation of the affinity matrix:

NHS-Sepharose beads (Amersham Biosciences) were washed with dry DMSO (40 mL/mL beads) and reacted with **CG_190** (1 µmol/mL beads) for 20 h on an end-over-end shaker at room temperature in the dark in the presence of triethylamine (15 µL/mL beads) in DMSO (1 vol of DMSO for 1 vol of beads). Aminoethanol (50 µL/mL beads) was then added to react remaining NHS-groups, and the mixture was further shaken for an extra 20 h at room temperature in the dark. The beads were then washed with DMSO (30 mL/mL beads) and ethanol (3 x 10 mL/mL beads) and stored in ethanol (1 mL/mL beads) at 4 °C. Aliquots of the supernatants before and after coupling were controlled by LC−MS to conclude of the completion of the reactions.

Competition assay:

The cell lysates were diluted with equal volumes of 1 x compound pull down (CP) buffer (50 mM Tris/HCl, pH 7.5, 5 % glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 20 mM NaF, 1 mM sodium ortho-vanadate, 1 mM DTT), protease inhibitor (SigmaFast protease inhibitor tablet S8820) and phosphatase inhibitors (Phosphatase Inhibitor Cocktail 3, Sigma-Aldrich, Munich, Germany). If required, lysates were further diluted to a final protein concentration of 5 mg/mL using 1 x CP buffer supplemented with 0.4 % Igepal (CP-0.4).

For selectivity profiling experiments in 96-well plates, the diluted cell mix lysates (2.5 mg of total proteins/well) were incubated for 30 min at 30 °C in an end-over-end shaker with 0 µm (DMSO control), 0.1 µM, 0.3 µM, 1.0 µM, 3.0 µM, 10 µM, 30 µM, 100 µM, or 300 µM of the Sirt5 inhibitor dissolved in DMSO. 18 µL of the affinity matrix were incubated with the lysates at 30 °C for 30 min in a filter plate. The DMSO control lysate was recovered and incubated similarly with 18 µL affinity matrix as a pull down of pull-down experiment in order to calculate the depletion factor. The beads were then washed (3 times 1 mL of CP-0.4 buffer followed by 2 times 1 mL of CP-0.2) and the bound proteins subsequently denatured in a buffer consisting of 8 M urea, 40 mM Tris-HCl (pH 7.4) and 50 mM DTT at 40 °C and 700 rpm. After adding 4

µL of 550 mM chloracetamide for 30 min at 37 °C, the urea concentration was diluted with 250 µL Tris-HCl (pH 7.4). Denatured proteins were digested by adding 30 µL of 10 ng/µL trypsin and incubating over night at 700 rpm. The digestion was stopped with 10 µL of 10 % (v/v) formic acid (FA) and peptides were desalted on a Sep-Pak tC18 µElution Plate (Waters). Desalted peptides dried down on a Speedvac and stored at -20 °C for LC-MS/MS measurement.

LC-MS/MS Analysis:

Label-free nano-flow LC-MSMS measurement of peptides was performed on Dionex Ultimate3000 nano HPLC (Thermo Fisher Scientific) coupled to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). The peptides were delivered to a trap column (100 µm × 2 cm, packed in-house with Reprosil-Pur C18-AQ 5 µm resin, Dr. Maisch) at a flow rate of 5 µL/min in 100 % solvent A (0.1 % formic acid in HPLC grade water) and separated on an analytical column (75 µm × 40 cm, packed in-house with Reprosil-Gold C18, 3 µm resin, Dr. Maisch) at a flow rate of 300 nL/min using a 51 min gradient ranging from 4 % to 32 % solvent C in B (solvent B: 0.1 % FA and 5 % DMSO in HPLC grade water, solvent C: 0.1 % FA and 5 % DMSO in acetonitrile). The eluent was sprayed via stainless steel emitters (Thermo) at a spray voltage of 2.1 kV and a heated capillary temperature of 275 °C. The instrument was operated in data-dependent mode. Full scan MS spectra (*m*/*z* 360−1300) were acquired in the Orbitrap at 60 000 resolution using an automatic gain control (AGC) target value of 4e5 charges and a maximum injection time of 50 ms.

For LC-MS2 analysis of the full proteome, up to 12 peptide precursors were selected for fragmentation by higher energy collision-induced dissociation (HCD; 1.7 *m*/*z* isolation window, AGC value of 2e5, maximum injection time of 75 ms, Dynamic Exclusion of 30 s) using 30 % normalised collision energy (NCE) and analysed at a resolution of 15,000 in the Orbitrap.

Protein Identification and Quantification:

Protein and peptide identification and quantification was performed using MaxQuant[29] (Version 1.6.1.0). The tandem MS data was searched against human canonical sequences from the Swissprot database (20193 entries, downloaded 22.03.2016) using the Andromeda search engine. [30] *N*-terminal acetylation and oxidation of methionine were set as variable modification, while carbamidomethylation of cysteine was set as fixed modification. Trypsin/P was set as proteolytic enzyme. The minimum peptide length was set to seven and all data were adjusted to 1 % protein false discovery rate (FDR). Matching between runs was enabled with a match time window of 0.7 min and alignment window of 20 min.

LC-MS/MS data analysis

Subsequent data analysis was performed on identified and quantified protein groups. The relative binding of proteins to the affinity matrix in dependence of free competing compound was calculated as ratio of the raw intensities of the protein at given compound concentration relative to the raw intensity of the DMSO sample. EC_{50} values were derived from a four-parameter log-logistic regression using an internal pipeline that utilises the 'drc' package in R. [31]

To calculate the K_d^{app} value, the EC_{50} value was multiplied with a correction factor derived from the ratio of the amount of protein captured from two consecutive pulldowns of the same DMSO control lysate.[32] The targets were manually annotated based on the sigmoidal shape of the binding curve with a dose-dependent decrease of binding to the affinity matrix. Moreover, number and dose-dependent decrease of unique peptides and MSMS counts per condition were considered.

Results from the LC-MS/MS analysis

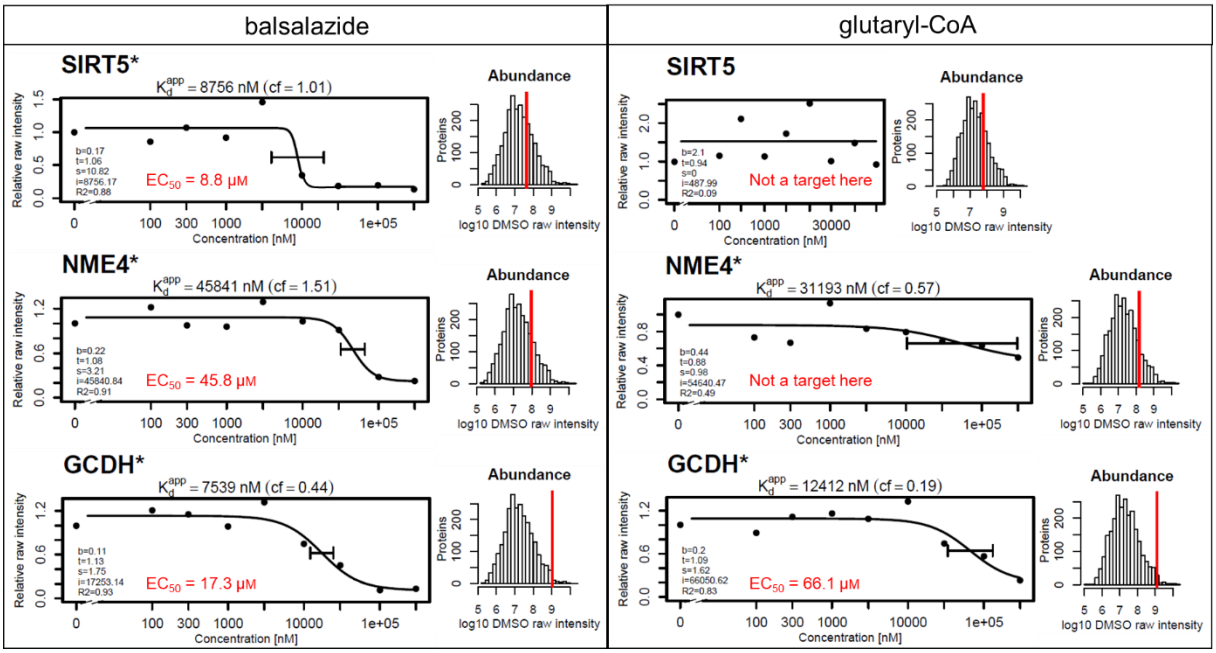


Figure S2: MS analysis results of the competition assay with immobilised CG_190 vs. balsalazide or glutaryl-CoA depicted as binding curves. Abundance = MS-based intensity of the target protein (red line) within the intensity distribution of all proteins in the DMSO control. * indicates that the protein is classified as a high confidence target.

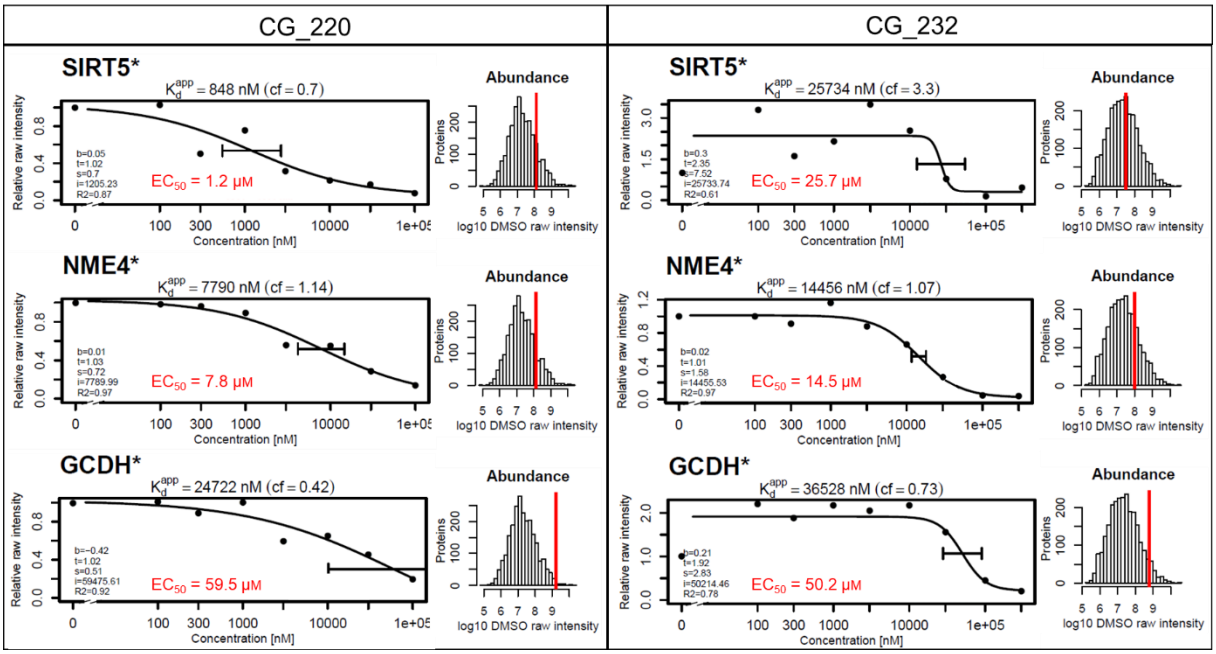


Figure S3: MS analysis results of the competition assay with immobilised CG_190 vs. CG_220 or CG_232 depicted as binding curves. Abundance: MS-based intensity of the target protein (red line) within the intensity distribution of all proteins in the DMSO control. * indicates that the protein is classified as a high confidence target.

Live cell imaging via desuccinylation-triggered peptide self-assembly

The cellular uptake and the intracellular behaviour of the studied peptides were characterised using confocal microscopy. Briefly, HeLa cells were seeded in 8-well ibidi culture plates at a density of 15,000 cells/well and allowed to attach to the plate for 12 h in growth medium at

37 °C with 5 % CO_2. Subsequently, the growth medium was replaced by fresh medium containing appropriate concentrations of peptides and suramin sodium salt. For the time lapse experiment, the cells were imaged immediately and every 3 h afterward. The cells were put back to incubate at 37 °C and 5 % CO_2 atmosphere between successive measurements. Fluorescent images were recorded on a Leica TCS SPE Confocal Laser Scanning Microscope using a 40×objective. For NBD dye, 458 nm was used as the excitation wavelength and 495−650 nm as the emission bandwidth. For Hoechst 33342, 405 nm was used as the excitation wavelength and 430−470 nm was used as the emission bandwidth.

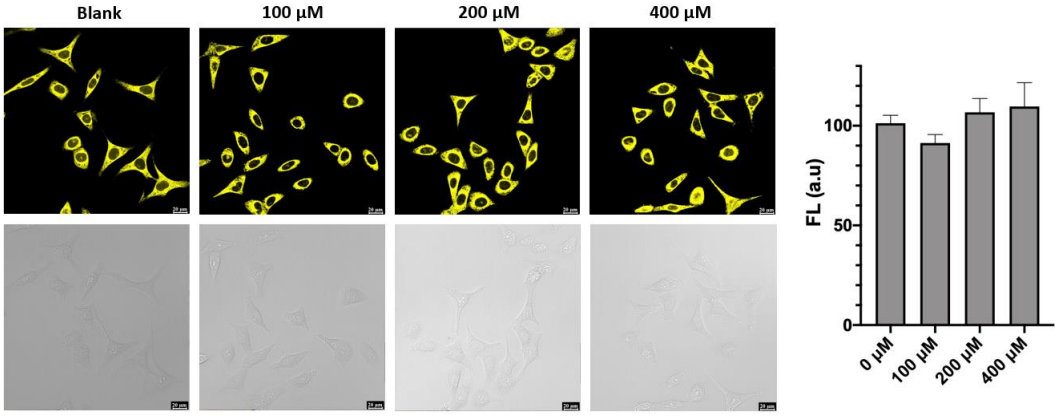


Figure S4: Inhibition of Sirt5 ability to form nanofibers using peptide 2[33] in HeLa cells in the presence of various concentrations of balsalazide after 45 min preincubation in live cell imaging experiment.

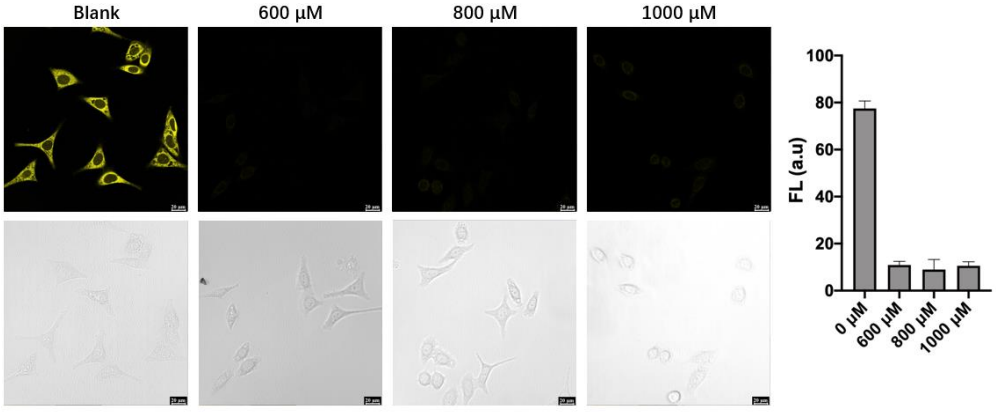


Figure S52: Inhibition of Sirt5 ability to form nanofibers using peptide 2[33] in HeLa cells in the presence of various concentrations of balsalazide after 90 min preincubation in live cell imaging experiment.

Part C: Relative binding free energy calculations

Computational details:

For the calculation of the relative binding free energies (ΔΔG), the non-equilibrium alchemical approach [34-35] in conjunction with molecular dynamics simulations was employed. Starting point of the system preparation was the structure of Sirt5 (PDB 3RIY) in a complex with NAD^+

and with balsalazide docked in the active site from a previous study. [1] The AMBERFF14SB [36] force field parameters for the protein, ZAFF [37] parameters for the zinc finger and NAD^+ parameters derived by Pavelites *et al*. [38] were employed. The Sirt5 inhibitors were parametrised using the Generalized Amber Force Field (GAFF) [39] in combination with AM1-BCC charges, [40] calculated with Antechamber. [41] In order to compute relative binding energies, pairs of inhibitors were formed, where most newly proposed structures were paired with balsalazide as many are variants of this lead structure. For each pair of inhibitors, hybrid structures and topologies were generated with the pmx tool [42-43] according to the single topology approach. The ligands and ligand–protein complexes were then placed in dodecahedral boxes of TIP3P [44] water with 1.5 nm distance to the box edge and the systems were neutralised with sodium ions. The hybrid systems were simulated in the end states A (representing one of the ligands, usually balsalazide) and state B (representing the second ligand) according to the following protocol.

The systems in state A and B were energy minimised using the steepest descent algorithm, followed by 250 ps of heating to a temperature of 298 K. Then 8 ns production runs were performed in an NPT ensemble with a time step of 2 fs. For temperature and pressure control during the molecular dynamics simulations, the velocity rescaling thermostat [45] with a time constant of 0.1 ps and the Parrinello–Rahman barostat [46] with a time constant of 5 ps and a target pressure of 1 bar, respectively, were employed. All bonds were constrained using the LINCS algorithm. [47] The long-range electrostatics were evaluated using the Particle Mesh Ewald (PME) [48-49] with a cutoff radius of 1.1 nm and a Fourier grid spacing of 0.12 nm. The van-der-Waals interactions were smoothly switched off between 1.0 and 1.1 nm. During the non-equilibrium transition simulations, the non-bonded interactions were treated with a soft-core potential. [50] All simulations were performed with GROMACS 2019. [51]

The first 2 ns of the production runs were discarded to account for equilibration, then 80 snapshots were extracted from each trajectory as starting structures for the alchemical transitions. For each starting structure, a non-equilibrium transition simulation from A to B (or *vice versa*) was performed over a time of 50 ps by gradually increasing the perturbation parameter λ from 0 to 1 (and *vice versa*). For the ligands **CG_111** and **CG_137**, a transition simulation length of 100 ps was used to improve the convergence. At each step of the non-equilibrium transition simulations, the derivative of the Hamiltonian with respect to λ was recorded. The whole procedure, starting from the energy minimisation up to the non-equilibrium transitions, was repeated 3 times per ligand pair, both for the ligand-water systems and for the ligand-protein complexes.

The free energy difference (ΔG) between the states A and B was estimated from the forward and reverse transitions with 240 work values in each direction (see Figure S6) using Bennet's

Acceptance Ratio (BAR) [52-53] based on the Crooks Fluctuation Theorem. [54-55] The error for the ΔG values was estimated by bootstrapping. The relative binding free energies (ΔΔ*G*) were obtained by subtracting the ΔG between inhibitors in water from the ΔG of the enzyme-inhibitor complexes making use of the thermodynamic cycle.

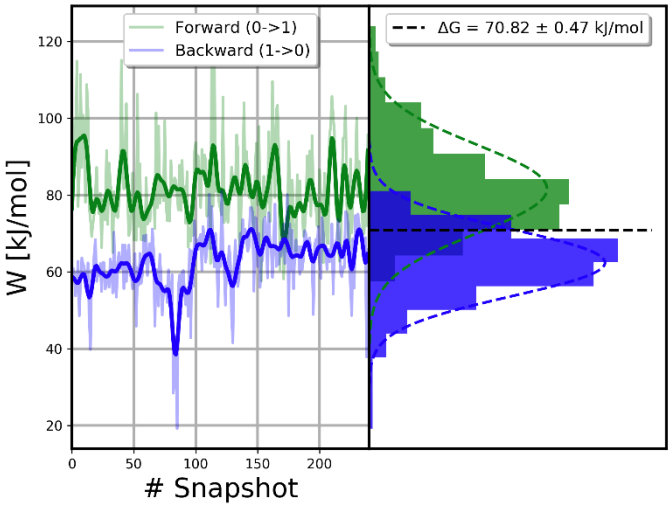


Figure S6: Summary of the results of the non-equilibrium transition simulations of balsalazide to CG_71 in the Sirt5 binding site. On the left side of the plot, the work values of the forward and reverse transitions for each starting structure are shown. On the right side of the plot, the distributions of the work values are depicted as histograms. A lack of an overlap between the two distributions would indicate convergence problems. This plot was generated, using the analyze_dhdl.py script of pmx. [42-43]. The estimated free energy difference (ΔG) is depicted as a black line.

<u>Representative snapshots from the MD simulations of **CG_137** and balsalazide in the Sirt5 binding site:</u>

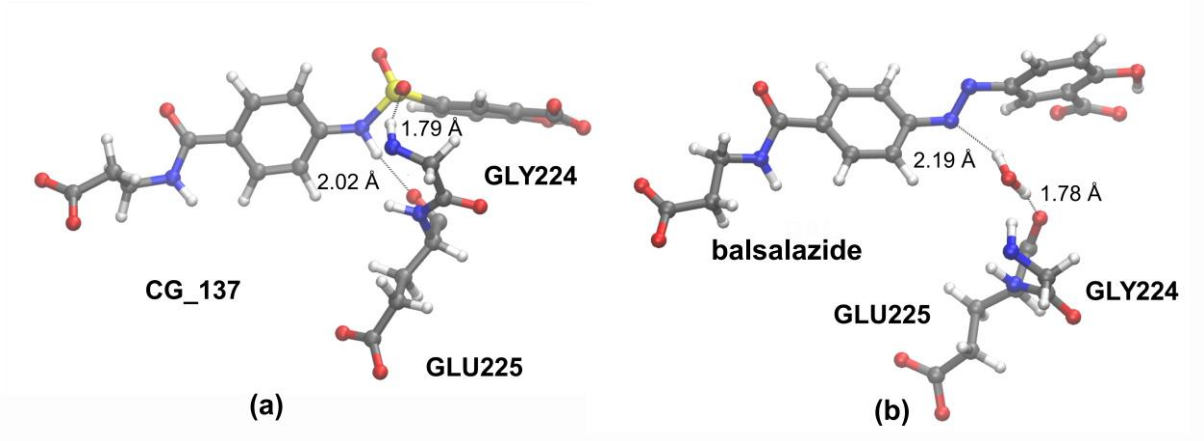


Figure S7: Representative snapshots of **(a) CG_137** and **(b)** balsalazide in the Sirt5 binding site. Over the course of the MD simulations of **CG_137** in the binding site, hydrogen bonds were formed between the sulfonamide spacer and the backbone atoms of Gly225 and Glu224. In the simulations of balsalazide, only water-mediated contacts between the inhibitor and Gly224 or Glu225 were observed. An oversampling of the configuration of **CG_137** in the Sirt5 binding site shown here could result in an overestimation of the binding affinity of **CG_137**.

Supplementary Information Bibliography

[1] Glas, C.; Dietschreit, J. C. B.; Wössner, N.; Urban, L.; Ghazy, E.; Sippl, W.; Jung, M.; Ochsenfeld, C.; Bracher, F., Identification of the subtype-selective Sirt5 inhibitor balsalazide through systematic SAR analysis and rationalization via theoretical investigations. *Eur. J. Med. Chem.* **2020,** *206*, 112676, doi: 10.1016/j.ejmech.2020.112676.
[2] Glas, C.; Wirawan, R.; Bracher, F., A Short Approach to N-Aryl-1,2,3,4-tetrahydroisoquinolines from N-(2-Bromobenzyl)anilines via a Reductive Amination/Palladium-Catalyzed Ethoxyvinylation/Reductive N-Alkylation Sequence. *Synthesis* **2021,** (EFirst), A-L, doi: 10.1055/s-0040-1706002.
[3] Hofmann, D.; Gans, E.; Krüll, J.; Heinrich, M. R., Sustainable Synthesis of Balsalazide and Sulfasalazine Based on Diazotization with Low Concentrations of Nitrogen Dioxide in Air. *Chem. Eur. J.* **2017,** *23* (17), 4042-4045, doi: 10.1002/chem.201605359.
[4] Chan, R. P. K. 2-Hydroxy-5-phenylazobenzoic acid derivatives and method of treating ulcerative colitis therewith. US4412992A, **1983**.
[5] Swinnen, D.; Jorand-Lebrun, C.; Gerber, P.; Gonzalez, J.; Bombrun, A. 1,1'-(1,2-Ethynediyl)bis-benzene Derivatives as PTP 1-B Inhibitors. WO2005097773A1, **2005**.
[6] Mukkamala, R.; Hossain, A.; Singh Aidhen, I., Valuable building block for the synthesis of lunularic acid, hydrangeic acid and their analogues. *Nat. Prod. Res.* **2017,** *31* (9), 1085-1090, doi: 10.1080/14786419.2016.1274891.
[7] Carzaniga, L.; Rancati, F.; Rizzi, A.; Linney, I.; Schmidt, W.; Barnes, M.; Knight, C. Compounds having muscarinic receptor antagonist and beta2 adrenergic receptor agonist activity. WO2016193241A1, **2016**.
[8] Kodra, J. T.; Behrens, C.; Madsen, P.; Joergensen, A. S.; Christensen, I. T. Novel Glucagon Antagonists. WO2004056763A2, **2004**.
[9] Kloss, F.; Lincke, T.; Hertweck, C., Highly Efficient Total Synthesis of the Clostridium-Derived anti-MRSA Antibiotic Closthioamide. *Eur. J. Org. Chem.* **2011,** *2011* (8), 1429-1431, doi: 10.1002/ejoc.201001695.
[10] Nilsson, P.; Katkevics, M.; Pelcman, B. Bis-Aryl Compounds For Use as Medicaments WO2009/127822A2, **2009**.
[11] Gerzon, K.; Humerickhouse, R. A.; Besch Jr., H. R.; Bidasee, K. R. Derivatives of Ryanodine and Dehydroryanodine US5679701A, **1997**.
[12] Huynh, T. H. V.; Mantel, M. L. H.; Mikkelsen, K.; Lindhardt, A. T.; Nielsen, N. C.; Otzen, D.; Skrydstrup, T., A Versatile Approach to β-Amyloid Fibril-Binding Compounds Exploiting the Shirakawa/Hayashi Protocol for trans-Alkene Synthesis. *Org. Lett.* **2009,** *11* (4), 999-1002, doi: 10.1021/ol8029593.
[13] Cho, Y. L.; Heo, H. J.; Oh, K. M.; Lee, H. S.; Park, C. S.; Chae, S. E.; Yun, J. Y.; Kwon, H. J.; Yang, Y. J.; Kang, D. H.; Kim, Y. Z.; Woo, S. H.; Park, T. K. Novel Cephalosporin Derivatives and Pharmaceutical Compositions Thereof. US2012264727A1, **2012**.
[14] Swinnen, D.; Gerber, P.; Gonzalez, J.; Bombrun, A.; Jorand-Lebrun, C. Alkynyl Aryl Carboxamides. WO2005012280A1, **2005**.
[15] Baccon-Sollier, P. L.; Malki, Y.; Maye, M.; Ali, L. M. A.; Lichon, L.; Cuq, P.; Vincent, L.-A.; Masurier, N., Imidazopyridine-fused [1,3]diazepinones: modulations of positions 2 to 4 and their impacts on the anti-melanoma activity. *J. Enzyme Inhib. Med. Chem.* **2020,** *35* (1), 935-949, doi: 10.1080/14756366.2020.1748024.
[16] Jones, K.; Rye, C.; Chessum, N.; Cheeseman, M.; Pasqua, A. E.; Pike, K. G.; Faulder, P. F. Fused 1,4-Dihydrodioxin Derivatives as Inhibitors of Heat Shock Transcription Factor 1. WO2015049535A1, **2015**.
[17] Ueda, T.; Okamoto, Y.; Tsuji, T.; Muraoka, M., Syntheses of 2-acylaminoacetamidine and 3-acylaminopropionamidine derivatives. *Chem. Pharm. Bull.* **1968,** *16* (12), 2355-2361, doi: 10.1248/cpb.16.2355.
[18] Tibiletti, F.; Simonetti, M.; Nicholas, K. M.; Palmisano, G.; Parravicini, M.; Imbesi, F.; Tollari, S.; Penoni, A., One-pot synthesis of meridianins and meridianin analogues via indolization of nitrosoarenes. *Tetrahedron* **2010,** *66* (6), 1280-1288, doi: 10.1016/j.tet.2009.12.020.
[19] Berger, M.; Hübner, J.; Ter Laak, A.; GORJANNACZ, M.; Fernandez-Montalvan, A. E.; Rodeschini, V.; Roche, D. Furane Derivatives as Inhibitors of ATAD2. WO2017093272A1, **2017**.
[20] Wacker, D. A.; Zhao, G.; Kwon, C.; Varnes, J. G.; Stein, P. D. Modulators of Serotonin Receptors. WO2005035533A1, **2005**.
[21] Paudyal, M. P.; Wu, L.; Zhang, Z.-Y.; Spilling, C. D.; Wong, C. F., A new class of salicylic acid derivatives for inhibiting YopH of Yersinia pestis. *Bioorg. Med. Chem.* **2014,** *22* (24), 6781-6788, doi: 10.1016/j.bmc.2014.10.042.
[22] Maurer, B.; Rumpf, T.; Scharfe, M.; Stolfa, D. A.; Schmitt, M. L.; He, W.; Verdin, E.; Sippl, W.; Jung, M., Inhibitors of the NAD(+)-Dependent Protein Desuccinylase and Demalonylase Sirt5. *ACS Med. Chem. Lett.* **2012,** *3* (12), 1050-3, doi: 10.1021/ml3002709.

[23] Liu, S.; Ji, S.; Yu, Z. J.; Wang, H. L.; Cheng, X.; Li, W. J.; Jing, L.; Yu, Y.; Chen, Q.; Yang, L. L.; Li, G. B.; Wu, Y., Structure-based discovery of new selective small-molecule sirtuin 5 inhibitors. *Chem. Biol. Drug. Des.* **2018,** *91* (1), 257-268, doi: 10.1111/cbdd.13077.
[24] Suenkel, B.; Fischer, F.; Steegborn, C., Inhibition of the human deacylase Sirtuin 5 by the indole GW5074. *Bioorg. Med. Chem. Lett.* **2013,** *23* (1), 143-6, doi: 10.1016/j.bmcl.2012.10.136.
[25] Kalbas, D.; Liebscher, S.; Nowak, T.; Meleshin, M.; Pannek, M.; Popp, C.; Alhalabi, Z.; Bordusa, F.; Sippl, W.; Steegborn, C.; Schutkowski, M., Potent and Selective Inhibitors of Human Sirtuin 5. *J. Med. Chem.* **2018,** *61* (6), 2460-2471, doi: 10.1021/acs.jmedchem.7b01648.
[26] Guetschow, E. D.; Kumar, S.; Lombard, D. B.; Kennedy, R. T., Identification of sirtuin 5 inhibitors by ultrafast microchip electrophoresis using nanoliter volume samples. *Anal. Bioanal. Chem.* **2016,** *408* (3), 721-31, doi: 10.1007/s00216-015-9206-0.
[27] Mosmann, T., Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods* **1983,** *65* (1), 55-63, doi: 10.1016/0022-1759(83)90303-4.
[28] Heltweg, B.; Trapp, J.; Jung, M., In vitro assays for the determination of histone deacetylase activity. *Methods* **2005,** *36* (4), 332-337, doi: 10.1016/j.ymeth.2005.03.003.
[29] Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008,** *26* (12), 1367-1372, doi: 10.1038/nbt.1511.
[30] Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M., Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011,** *10* (4), 1794-1805, doi: 10.1021/pr101065j.
[31] Ritz, C.; Baty, F.; Streibig, J. C.; Gerhard, D., Dose-Response Analysis Using R. *PLOS ONE* **2016,** *10* (12), e0146021, doi: 10.1371/journal.pone.0146021.
[32] Médard, G.; Pachl, F.; Ruprecht, B.; Klaeger, S.; Heinzlmeir, S.; Helm, D.; Qiao, H.; Ku, X.; Wilhelm, M.; Kuehne, T.; Wu, Z.; Dittmann, A.; Hopf, C.; Kramer, K.; Kuster, B., Optimized Chemical Proteomics Assay for Kinase Inhibitor Profiling. *J. Proteome Res.* **2015,** *14* (3), 1574-1586, doi: 10.1021/pr5012608.
[33] Yang, L.; Peltier, R.; Zhang, M.; Song, D.; Huang, H.; Chen, G.; Chen, Y.; Zhou, F.; Hao, Q.; Bian, L.; He, M.-l.; Wang, Z.; Hu, Y.; Sun, H., Desuccinylation-Triggered Peptide Self-Assembly: Live Cell Imaging of SIRT5 Activity and Mitochondrial Activity Modulation. *J. Am. Chem. Soc.* **2020,** *142* (42), 18150-18159, doi: 10.1021/jacs.0c08463.
[34] Aldeghi, M.; Gapsys, V.; de Groot, B. L., Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation. *ACS Cent. Sci.* **2018,** *4* (12), 1708-1718, doi: 10.1021/acscentsci.8b00717.
[35] Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; van Vlijmen, H.; Tresadern, G.; de Groot, B. L., Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chem. Sci.* **2020,** *11* (4), 1140-1152, doi: 10.1039/C9SC03754C.
[36] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-3713, doi: 10.1021/acs.jctc.5b00255.
[37] Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M., Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010,** *6* (9), 2935-2947, doi: 10.1021/ct1002626.
[38] Pavelites, J. J.; Gao, J.; Bash, P. A.; Mackerell Jr., A. D., A molecular mechanics force field for NAD+ NADH, and the pyrophosphate groups of nucleotides. *J. Comput. Chem.* **1997,** *18* (2), 221-239, doi: 10.1002/(sici)1096-987x(19970130)18:2<221::aid-jcc7>3.0.co;2-x.
[39] 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.* **2004,** *25* (9), 1157-1174, doi: 10.1002/jcc.20035.
[40] Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002,** *23* (16), 1623-1641, doi: 10.1002/jcc.10128.
[41] 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.* **2006,** *25* (2), 247-260, doi: 10.1016/j.jmgm.2005.12.005.
[42] Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L., pmx: Automated protein structure and topology generation for alchemical perturbations. *J. Comput. Chem.* **2015,** *36* (5), 348-354, doi: 10.1002/jcc.23804.
[43] Gapsys, V.; de Groot, B. L., pmx Webserver: A User Friendly Interface for Alchemistry. *J. Chem. Inf. Model.* **2017,** *57* (2), 109-114, doi: 10.1021/acs.jcim.6b00498.
[44] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935, doi: 10.1063/1.445869.
[45] Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126* (1), 014101, doi: 10.1063/1.2408420.

[46] Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190, doi: 10.1063/1.328693.
[47] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** *18* (12), 1463-1472, doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
[48] Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092, doi: 10.1063/1.464397.
[49] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103* (19), 8577-8593, doi: 10.1063/1.470117.
[50] Gapsys, V.; Seeliger, D.; de Groot, B. L., New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy Calculations. *J. Chem. Theory Comput.* **2012,** *8* (7), 2373-2382, doi: 10.1021/ct300220p.
[51] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25, doi: 10.1016/j.softx.2015.06.001.
[52] Bennett, C. H., Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976,** *22* (2), 245-268, doi: 10.1016/0021-9991(76)90078-4.
[53] Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S., Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett.* **2003,** *91* (14), 140601, doi: 10.1103/PhysRevLett.91.140601.
[54] Crooks, G. E., Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E* **2000,** *61* (3), 2361-2366, doi: 10.1103/PhysRevE.61.2361.
[55] Sivak, D. A.; Crooks, G. E., Near-Equilibrium Measurements of Nonequilibrium Free Energy. *J. Stat. Phys.* **2012,** *108* (15), 150601, doi: 10.1103/PhysRevLett.108.150601.

4 Summary

The first two projects of this thesis explore the reaction mechanisms of the monofunctional DNA glycosylases UDG and TDG. In publication **I**, QM/MM calculations are employed for the investigation of the N-glycosidic bond hydrolysis of uracil, catalyzed by the enzyme UDG. Over the course of the C1'- N1 bond dissociation, the sugar moiety migrates in the direction of the nucleophile, whereas the uracil base and the water molecule remain close to their initial positions. This results in a dissociative oxocarbenium-like transition state. Upon the addition of the water molecule, a proton is simultaneously transferred to a near-by base. The absence of a barrier for this step suggests that this reaction occurs spontaneously due to the reactivity of the electrophilic sugar. While both His148 and Asp145, which coordinate the water nucleophile, can act as a proton acceptor, our investigation suggest that His148 is slightly favored. This result is consistent with experimental studies that evidence a decrease in the enzymatic activity upon the mutation of His148 and a change in the protonation state of this residue over the course of the reaction [93, 94]. At the same time, an alternative reaction mechanism through a proton transfer to Asp145 explains why mutations of His148 do not result in a complete loss of enzymatic activity. In publication **II** we investigate the TDG catalyzed glycosidic bond cleavage of the modified base 5-formylcytosine (fC), which occurs naturally in DNA as an intermediate of the active cytosine demethylation pathway. Our calculations revealed that the reaction mechanism involves a rearrangement of the substrate by a rotation around the dihedral between O5', C5', C4', and C3'. This lowers the barrier height for the subsequent C1'-N1 bond dissociation as it reduces the repulsion between the phosphate group and the negatively charged reaction product. Similarly to the UDG catalyzed reaction, the bond cleavage proceeds by an oxocarbenium-like transition state with no evidence of a stable intermediate. The subsequent addition of the water nucleophile results in a positive charge that is delocalized over an extensive hydrogen bond network, consisting of Asn140, Thr197, and the abasic site. This indicates an involvement of Thr197 and Asn140 in the reaction mechanism that goes beyond the coordination of the water molecule, which would explain the significant decrease in enzymatic activity upon the mutation of these residues [95]. TDG facilitates the excision of another intermediate of the active demethylation pathway - 5-carboxycytosine (caC). This reaction is, however, pH-dependent and involves the residue Asn191, which is not essential for the fC removal [23]. While it is possible for the caC excision to have similarities to the fC reaction mechanism shown here, further studies are necessary to elucidate the exact reaction proceedings.

Manuscript **III** presents the synthesis and investigation of inhibitors of the NAD^+-dependent deacylase Sirtuin 5. Using non-equilibrium MD simulations and BAR, the relative binding affinities of selected Sirtuin 5 inhibitors as compared to balsalazide were computed. The results for most studied compounds agree qualitatively with the

potencies obtained in the *in vitro* experiments and showed that this approach can be employed for further lead optimization. Nevertheless, there are some limitations of the methodology, as relative free energy calculation of ring breaking and formation are known to suffer from systematic problems [96]. Therefore further studies of the binding affinity of SIRT5 inhibitors should either employ absolute binding free energy calculation or a different reference ligand.

Bibliography

[1] E. M. Abdelraheem, H. Busch, U. Hanefeld, and F. Tonin, React. Chem. Eng. **4**, 1878 (2019).

[2] A. S. Bommarius and M. F. Paye, Chem. Soc. Rev. **42**, 6534 (2013).

[3] G. A. Holdgate, T. D. Meek, and R. L. Grimley, Nat. Rev. Drug Discov. **17**, 115 (2018).

[4] A. C. Drohat and C. T. Coey, Chem. Rev. **116**, 12711 (2016).

[5] H. Lans, J. H. Hoeijmakers, W. Vermeulen, and J. A. Marteijn, Nat. Rev. Mol. Cell Biol. **20**, 766 (2019).

[6] J. T. Stivers and Y. L. Jiang, Chem. Rev. **103**, 2729 (2003).

[7] A. B. Robertson, A. Klungland, T. Rognes, and I. Leiros, Cell. Mol. Life Sci. **66**, 981 (2009).

[8] A. L. Jacobs and P. Schär, Chromosoma **121**, 1 (2012).

[9] T. Lindahl, Proc. Natl. Acad. Sci. USA **71**, 3649 (1974).

[10] T. Lindahl and R. D. Wood, Science **286**, 1897 (1999).

[11] S. S. Parikh, G. Walcher, G. D. Jones, G. Slupphaug, H. E. Krokan, G. M. Blackburn, and J. A. Tainer, Proc. Natl. Acad. Sci. USA **97**, 5083 (2000).

[12] G. Slupphaug, C. D. Mol, B. Kavli, A. S. Arvai, H. E. Krokan, and J. A. Tainer, Nature **384**, 87 (1996).

[13] A. R. Dinner, G. M. Blackburn, and M. Karplus, Nature **413**, 752 (2001).

[14] J. L. Przybylski and S. D. Wetmore, Biochemistry **50**, 4218 (2011).

[15] K. Wiebauer and J. Jiricny, Proc. Natl. Acad. Sci. USA **87**, 5842 (1990).

[16] D. Cortázar, C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, A. Wirz, D. Schuermann, A. L. Jacobs, F. Siegrist, R. Steinacher, J. Jiricny, A. Bird, and P. Schär, Nature **470**, 419 (2011).

[17] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, and A. Bellacosa, Cell **146**, 67 (2011).

[18] S. B. Baylin, Science **277**, 1948 (1997).

[19] R. D. Hotchkiss, J. Biol. Chem. **175**, 315 (1948).

[20] A. Maiti and A. C. Drohat, J. Biol. Chem. **286**, 35334 (2011).

[21] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang, Science **333**, 1300 (2011).

[22] L. S. Pidugu, J. W. Flowers, C. T. Coey, E. Pozharski, M. M. Greenberg, and A. C. Drohat, Biochemistry **55**, 6205 (2016).

[23] A. Maiti, A. Z. Michelson, C. J. Armwood, J. K. Lee, and A. C. Drohat, J. Am. Chem. Soc. **135**, 15813 (2013).

[24] Y. L. Deribe, T. Pawson, and I. Dikic, Nat. Struct. Mol. Biol. **17**, 666 (2010).

[25] A. J. Bannister and T. Kouzarides, Cell Res. **21**, 381 (2011).

[26] C. T. Walsh, S. Garneau-Tsodikova, and G. J. Gatto Jr., Angew. Chemie Int. Ed. **44**, 7342 (2005).

[27] X. J. Yang and E. Seto, Oncogene **26**, 5310 (2007).

[28] J.-Y. Huang, M. D. Hirschey, T. Shimazu, L. Ho, and E. Verdin, Biochim. Biophys. Acta - Proteins Proteomics **1804**, 1645 (2010).

[29] C. Peng, Z. Lu, Z. Xie, Z. Cheng, Y. Chen, M. Tan, H. Luo, Y. Zhang, W. He, K. Yang, B. M. Zwaans, D. Tishkoff, L. Ho, D. Lombard, T. C. He, J. Dai, E. Verdin, Y. Ye, and Y. Zhao, Mol. Cell. Proteomics **10**, 111 (2011).

[30] J. Du, Y. Zhou, X. Su, J. J. Yu, S. Khan, H. Jiang, J. Kim, J. Woo, J. H. Kim, B. H. Choi, B. He, W. Chen, S. Zhang, R. A. Cerione, J. Auwerx, Q. Hao, and H. Lin, Science **334**, 806 (2011).

[31] M. Tan, C. Peng, K. A. Anderson, P. Chhoy, Z. Xie, L. Dai, J. Park, Y. Chen, H. Huang, Y. Zhang, J. Ro, G. R. Wagner, M. F. Green, A. S. Madsen, J. Schmiesing, B. S. Peterson, G. Xu, O. R. Ilkayeva, M. J. Muehlbauer, T. Braulke, C. Mühlhausen, D. S. Backos, C. A. Olsen, P. J. McGuire, S. D. Pletcher, D. B. Lombard, M. D. Hirschey, and Y. Zhao, Cell Metab. **19**, 605 (2014).

[32] E. D. Guetschow, S. Kumar, D. B. Lombard, and R. T. Kennedy, Anal. Bioanal. Chem. **408**, 721 (2016).

[33] C. Glas, J. C. Dietschreit, N. Wössner, L. Urban, E. Ghazy, W. Sippl, M. Jung, C. Ochsenfeld, and F. Bracher, Eur. J. Med. Chem. **206**, 112676 (2020).

[34] R. B. R. Muijsers and K. L. Goa, Drugs **62**, 1689 (2002).

[35] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[36] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).

[37] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).

[38] S. A. Adcock and J. A. McCammon, Chem. Rev. **106**, 1589 (2006).

[39] M. J. Field, *A Practical Introduction to the Simulation of Molecular Systems*, Cambridge University Press, 2 edition, 2007.

[40] X. Cheng and I. Ivanov, Methods Mol. Biol. **929**, 243 (2012).

[41] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, J. Chem. Theory Comput. **11**, 3696 (2015).

[42] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo, and P. Jurečka, J. Chem. Theory Comput. **11**, 5723 (2015).

[43] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).

[44] R. Lonsdale, J. N. Harvey, and A. J. Mulholland, Chem. Soc. Rev. **41**, 3025 (2012).

[45] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).

[46] H. M. Senn and W. Thiel, Angew. Chemie Int. Ed. **48**, 1198 (2009).

[47] M. W. Van Der Kamp and A. J. Mulholland, Biochemistry **52**, 2708 (2013).

[48] C. V. Sumowski, B. B. T. Schmitt, S. Schweizer, and C. Ochsenfeld, Angew. Chemie Int. Ed. **49**, 9951 (2010).

[49] D. Flaig, M. Beer, and C. Ochsenfeld, J. Chem. Theory Comput. **8**, 2260 (2012).

[50] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).

[51] G. Groenhof, Methods Mol. Biol. **924**, 43 (2013).

[52] P. E. Sinclair, A. de Vries, P. Sherwood, C. Richard A. Catlow, and R. A. van Santen, J. Chem. Soc., Faraday Trans. **94**, 3401 (1998).

[53] H. Lin and D. G. Truhlar, Theor. Chem. Acc. **117**, 185 (2007).

[54] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).

[55] T. H. Rod and U. Ryde, J. Chem. Theory Comput. **1**, 1240 (2005).

[56] R. W. Hockney, S. P. Goel, and J. W. Eastwood, J. Comput. Phys. **14**, 148 (1974).

[57] W. F. Van Gunsteren and H. J. Berendsen, Mol. Phys. **34**, 1311 (1977).

[58] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, J. Comput. Chem. **18**, 1463 (1997).

[59] D. Van Der Spoel and P. J. Van Maaren, J. Chem. Theory Comput. **2**, 1 (2006).

[60] P. J. Steinbach and B. R. Brooks, J. Comput. Chem. **15**, 667 (1994).

[61] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).

[62] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, 2 edition, 1989.

[63] M. J. S. Dewar and S. Kirschner, J. Am. Chem. Soc. **93**, 4292 (1971).

[64] I. H. Williams and G. M. Maggiora, J. Mol. Struct. **89**, 365 (1982).

[65] K. P. Eurenius, D. C. Chatfield, B. R. Brooks, and M. Hodoscek, Int. J. Quantum Chem. **60**, 1189 (1996).

[66] H. B. Schlegel, *Optimization of Equilibrium Geometries and Transition Structures*, pages 249–286, 1987.

[67] H. L. Woodcock, M. Hodošček, and B. R. Brooks, J. Phys. Chem. A **111**, 5720 (2007).

[68] A. S. Edison, J. Am. Chem. Soc **124**, 7875 (2002).

[69] H. Jónsson, G. Mills, and K. W. Jacobsen, *Nudged elastic band method for finding minimum energy paths of transitions*, in Classical and Quantum Dynamics in Condensed Phase Simulations, edited by B. J. Berne, G. Ciccotti, and D. F. Coker, pages 385–404, World Scientific, 1998.

[70] G. Henkelman and H. Jónsson, J. Chem. Phys. **113**, 9978 (2000).

[71] G. Henkelman, B. P. Uberuaga, and H. Jónsson, J. Chem. Phys. **113**, 9901 (2000).

[72] D. Sheppard, R. Terrell, and G. Henkelman, J. Chem. Phys. **128** (2008).

[73] D. L. Beveridge and F. M. DiCapua, Annu. Rev. Biophys. Biophys. Chem. **18**, 431 (1989).

[74] V. Gapsys, S. Michielssens, J. Peters, B. de Groot, and H. Leonov, *Calculation of Binding Free Energies*, in Molecular Modeling of Proteins: Methods in Molecular Biology (Methods and Protocols), edited by A. Kukol, Humana Press, 2015.

[75] M. A. Cuendet, J. Chem. Phys. **125** (2006).

[76] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 1 (2003).

[77] D. M. Zuckerman and T. B. Woolf, Phys. Rev. Lett. **89**, 16 (2002).

[78] J. Gore, F. Ritort, and C. Bustamante, Proc. Natl. Acad. Sci. U. S. A. **100**, 12564 (2003).

[79] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).

[80] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[81] R. Chelli, S. Marsili, A. Barducci, and P. Procacci, J. Chem. Phys. **126** (2007).

[82] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, J. Med. Chem. **59**, 4035 (2016).

[83] L. Pérez-Benito, N. Casajuana-Martin, M. Jiménez-Rosés, H. Van Vlijmen, and G. Tresadern, J. Chem. Theory Comput. **15**, 1884 (2019).

[84] W. L. Jorgensen, Acc. Chem. Res. **42**, 724 (2009).

[85] Z. Cournia, B. Allen, and W. Sherman, J. Chem. Inf. Model. **57**, 2911 (2017).

[86] C. Öhlknecht, J. W. Perthold, B. Lier, and C. Oostenbrink, J. Chem. Theory Comput. **16**, 7721 (2020).

[87] J. Cruz, L. Wickstrom, D. Yang, E. Gallicchio, and N. Deng, J. Chem. Theory Comput. **16**, 2803 (2020).

[88] W. L. Jorgensen and C. Ravimohan, J. Chem. Phys. **83**, 3050 (1985).

[89] M. Aldeghi, V. Gapsys, and B. L. De Groot, ACS Cent. Sci. **4**, 1708 (2018).

[90] V. Gapsys, L. Pérez-Benito, M. Aldeghi, D. Seeliger, H. Van Vlijmen, G. Tresadern, and B. L. De Groot, Chem. Sci. **11**, 1140 (2020).

[91] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, Chem. Phys. Lett. **222**, 529 (1994).

[92] M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS User Manual*.

[93] C. D. Mol, A. S. Arvai, G. Slupphaug, B. Kavli, I. Alseth, H. E. Krokan, and J. A. Tainer, Cell **80**, 869 (1995).

[94] J. B. Parker and J. T. Stivers, Biochemistry **47**, 8614 (2008).

[95] A. Maiti, M. S. Noon, A. D. MacKerell, E. Pozharski, and A. C. Drohat, Proc. Natl. Acad. Sci. **109**, 8091 (2012).

[96] S. Liu, L. Wang, and D. L. Mobley, J. Chem. Inf. Model. **55**, 727 (2015).