

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

**Quantum-Chemical Investigations of
Enzymatic Reaction Mechanisms
and
Benchmarking of Range-Separated
Density-Functional Theory within the
Random Phase Approximation**

Andrea Kreppel

aus

Passau

2020

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, 10.01.2021

(Andrea Kreppel)

Disseration eingereicht am: 29.09.2020

1. Gutachter: Prof. Dr. Christian Ochsenfeld

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

Mündliche Prüfung am: 09.11.2020

Acknowledgments

Here, I would like to appreciate the professional and personal support that I have received during my Ph.D. studies.

First and foremost, I would like to express my gratitude to **Prof. Dr Christian Ochsenfeld**, who gave me the opportunity to do my Ph.D. thesis in his group. I thank him for the continuous support as well as the constructive discussions and advice while I could freely choose the research topics.

Furthermore, I would like to thank **Prof. Dr. Regina de Vivie-Riedle** for the preparation of the second opinion and the other members of my Ph.D. defense committee, **Prof. Dr. Thomas Carell**, **Prof. Dr. Franz Bracher**, **Prof. Dr. Hendrik Zipse**, and **Dr. Dorian Didier**.

Particularly, I would like to thank **Dr. Iris Blank**, who passed on her expertise to me. I would also like to thank **Henryk Laqua** and **Daniel Graf** for the constructive cooperation, as well as **Eli Naydenova**, **Beatriz von der Esch**, and **Johannes Dietschreit** for the useful discussions.

Of course, I would also like to thank **Eli Naydenova**, **Laurens Peters**, **Daniel Graf**, **Andreas Baumann**, and **Monika Kreppel** for the support in writing this thesis.

Last but not least, I would like to thank my **family** and **friends**, especially **Andreas Baumann**, for their understanding and support throughout the duration studies and preparation of my Ph.D. thesis.

Abstract

Enzymes are efficient biocatalysts, which are not only essential for living organisms but are also employed for industrial syntheses of chemical compounds. Insights into the exact enzymatic reaction mechanism is valuable for many fields, e.g., controlling the reaction in industrial settings, or development of enzyme inhibitors. Furthermore, it also contributes to a comprehensive understanding of the entire biochemical process around the enzyme.

The first part of this thesis deals with the elucidation of the catalytic mechanism of two enzymes. In a first project, the glycosylase activity of the bacterial DNA-glycosylase Fpg is investigated. An alternative, base-independent excision mechanism is found for the substrate 8OG. It is a variation of the previously found ribose-protonated excision mechanism that does not involve the excised base. With this 'base-independent' ribose-protonated mechanism, Fpg can excise 8OG in both *syn*- and *anti*-bound conformations. This is in contrast to the previously found 'base-specific' mechanism, which only proceeds with *syn*-bound 8OG. The energy profile of the rate-determining step of the base-independent excision mechanism, obtained with QM/MM calculations with a significant number of atoms in the QM sphere, results in a reaction barrier of similar height for *syn*- and *anti*-8OG, which is in good agreement with experimental measurements. From this it can be concluded that Fpg has no preference between *syn*- and *anti*-8OG in the base-independent excision mechanism.

In a second project, the decarboxylation mechanism of 5caU by the enzyme IDCase is examined. For this purpose, QM/MM energy profiles of possible decarboxylation pathways are compared. The comparison reveals that the reaction catalyzed by IDCase most likely follows a direct decarboxylation mechanism. Detailed investigations on this mechanism confirm that the direct decarboxylation of 5caU by IDCase is a one-step mechanism with simultaneous proton transfer and the C-C bond opening.

The description of an energy profile of a reaction mechanism requires an accurate method that performs well for thermodynamic and kinetic properties, as well as for non-covalent interactions. Semi-local and hybrid density functional theory rather gives varying results, especially for the description of reaction barrier heights. Range-separated hybrid DFT, based on a short-range PBE exchange-correlation functional, in combination with long-range random phase approximation correlation (RSHPBE+lrRPA) gives promising results in some small-scale studies and shows a relatively fast basis set size convergence. In the second part of this thesis RSHPBE+lrRPA is benchmarked on a large-scale to be able to assess its potential. Therefore RSHPBE+lrRPA is applied on the GMTKN55 data set. The results of the benchmark reveal that indeed a moderately sized triple- ζ basis set is mostly sufficient for RSHPBE+lrRPA. Furthermore, RSHPBE+lrRPA shows a stable performance over the complete test set with less fluctuations in the accuracy between the subsets than standard RPA. According to its results on the GMTKN55 data set, RSHPBE+lrRPA is comparable to a double-hybrid functional without empirical dispersion correction.

List of Publications

This work is a cumulative dissertation comprising two articles published in peer-review journals (**I**, **III**) and one just accepted article (**II**). In the following, all articles are listed together with the author's contributions.

I A. Kreppel, I. D. Blank, C. Ochsenfeld,
"Base-Independent DNA Base-Excision Repair of 8-Oxoguanine"
J. Am. Chem. Soc., **2018**, *140*, 4522–4526.
Contribution by the Author: *Most of the concept and all calculations,*
writing the manuscript

II A. Kreppel, C. Ochsenfeld,
"The Enzymatic Decarboxylation Mechanism of 5-carboxy Uracil - a Comprehensive
Quantum Chemical Study"
J. Chem. Theory Comput., accepted.
Contribution by the Author: *Most of the concept and all calculations,*
writing the manuscript

III A. Kreppel, D. Graf, H. Laqua, C. Ochsenfeld,
"Range-Separated Density-Functional Theory in Combination with the Random
Phase Approximation: An Accuracy Benchmark"
J. Chem. Theory Comput., **2020**, *16*, 2985–2994.
Contribution by the Author: *Joint concept with D. Graf and H. Laqua,*
all calculations, most of writing the manuscript

Contents

Acknowledgements	v
Abstract	vii
List of Publications	ix
Part I. Quantum-Chemical Investigations of Enzyme Reaction Mechanisms	1
1 Introduction	3
1.1 Base-Excision-Repair	3
1.2 Active DNA Demethylation	6
2 Theoretical Background	9
2.1 Molecular Force Fields	9
2.2 Density-Functional Theory	10
2.3 Quantum Mechanics/Molecular Mechanics	12
2.4 Molecular Dynamics Simulations	13
2.5 Structure Optimization	15
2.6 Constrained Structure Optimization	17
2.7 Nudged Elastic Band Method	18
3 Publications	21
Publication I: Base-Independent DNA Base-Excision Repair of 8-Oxoguanine	21
Publication II: The Enzymatic Decarboxylation Mechanism of 5-carboxy Uracil - a Comprehensive Quantum Chemical Study	39
4 Summary to Part I	113
Part II. Accuracy Benchmark of Range-Separated Density- Functional Theory including the Random Phase Approx- imation	115
5 Introduction	117
6 Theoretical Background	119
6.1 The Random-Phase-Approximation	119
6.1.1 Adiabatic Connection	119
6.1.2 Correlation Energy in Terms of Density Fluctuations	120
6.1.3 Fluctuation-Dissipation Theorem	121
6.1.4 Random-Phase-Approximation	123
6.2 Efficient Calculation of the RPA Correlation Energy	124
6.2.1 The Resolution-of-the-Identity	124

6.2.2	RPA Correlation Energy in the Atomic Orbital Formalism	126
6.3	Range-Separated DFT	127
6.3.1	Long-Range Corrected Density-Functionals	127
6.3.2	Range-Separated DFT with the Random-Phase-Approximation . . .	128
	Publication III: Range-Separated Density-Functional Theory in Combination with the Random Phase Approximation: An Accuracy Benchmark .	131
7	Conclusion	151

Part I.

Quantum-Chemical
Investigations
of
Enzyme
Reaction Mechanisms

1 Introduction

Enzymes are excellent catalysts, as they tolerate a wide range of complex molecules as substrates with high enantio- and regioselectivity [1, 2]. At the same time, this biocatalytic process is often more environmentally friendly than conventional chemical catalysis due to milder reaction conditions, lower physiological toxicity, and fewer by-products. Because of these remarkable properties, enzymatic catalysis is increasingly used in industrial synthesis [3–6].

Knowledge of the the catalytic mechanism or the catalytic residues of the enzyme can be used as a guide for protein engineering. By specific mutation of amino acids, for example, the substrate specificity can be increased or the substrate range can be expanded.

An important field of research involving enzymes is the development of specific and potent inhibitors that potentially can serve as drugs. In this context, high-resolution 3D structures of the enzyme are the basis of the design of non-covalently binding inhibitors in a computer-aided process [7–9]. Precise knowledge of the catalytic mechanism here is also advantageous, as it allows for the design of mechanism-based covalent inhibitors [10–12]. These examples show that knowledge of the exact reaction pathway catalyzed by enzymes are beneficial in many cases.

Theoretical approaches play, beside structural and mutational studies, or kinetic isotope measurements, an important role in the elucidation of enzymatic reactions, which, above all, also contributes to the complete understanding of biological processes.

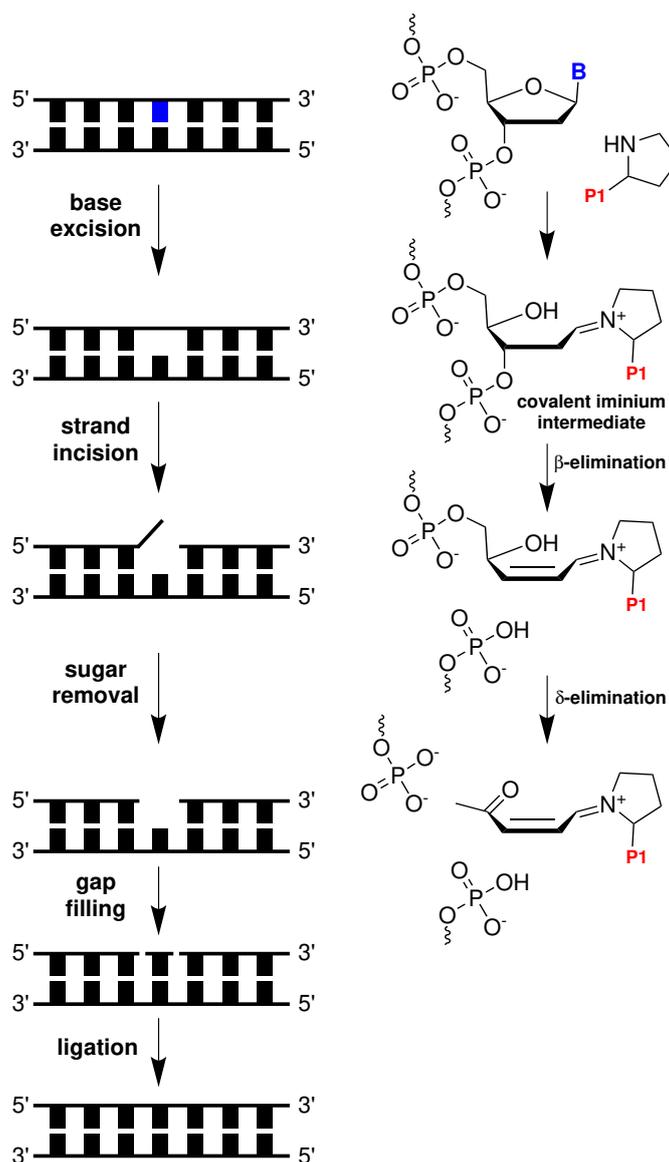
In the following, a brief overview of the two biological processes is given, to which this work contributes by the theoretical investigation of the enzyme reaction mechanisms.

1.1 Base-Excision-Repair

DNA is constantly exposed to exogenous and endogenous damaging sources, such as radiation, side products of cellular metabolism, or replication errors [13]. Such DNA damages include chemical modification of a base, chemical cross-linking of two bases, the breakage of one or both of the DNA strands, or the miss-incorporation of an incorrect base. Since these DNA lesions may be miscoding, several DNA repair pathways exist for maintaining genomic integrity [14].

Small, non-bulky DNA lesions derived from deamination, alkylation, or oxidation can be repaired in the base excision repair (BER) pathway [15–17]. This pathway is highly conserved among bacteria and humans [17] and generally proceeds via five key steps (see scheme 1, left) although several variations of the pathway exist [16]. In a first step, the lesion is recognized and excised by a DNA glycosylase, creating an apurinic/apyrimidinic (AP) site. Subsequently, the DNA backbone at the AP site is incised in an AP lyase reaction. Third, the remaining sugar fragment of the AP site is removed. The resulting gap is then filled by a DNA polymerase and, finally, sealed by a DNA ligase.

There are several different DNA glycosylases with different substrate specificity. All DNA glycosylases have in common that they flip the damaged base out of the DNA helix into its active site [18]. Generally, there are two different types of DNA glycosylases [19, 20]. Monofunctional glycosylases only catalyze the first step in BER by hydrolysis of the glycosidic bond. Bifunctional glycosylases, do not only excise the damaged base but also have lyase activity (second step of BER). All bifunctional enzymes catalyze the incision of the DNA backbone at the 3'-side of the AP site (β -elimination). Some bifunctional



Scheme 1: Five key steps in the base excision repair schematically sketched for the bifunctional DNA glycosylase Fpg. Overview of the corresponding chemical reactions catalyzed by Fpg are given on the right.

glycosylases, such as the formamidopyrimidine DNA glycosylase (Fpg) in prokaryotes [21, 22], can also completely remove the ribose moiety in a subsequent δ -elimination reaction (see scheme 1, right). It has been suggested that the β - and δ -elimination catalyzed by Fpg proceed in an E1cB like mechanism [23].

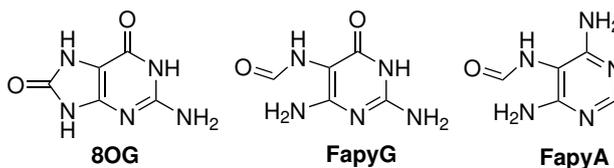


Figure 1: Oxidized purin bases, which are substrates for Fpg.

Fpg efficiently excises the oxidative purine lesions 7,8-dihydro-8-oxoguanine (8OG), 2,6-diamino-4-hydroxy-5-formamido-pyrimidine (FapyG) [24, 25] (see fig. 1) but also shows catalytic activity for several chemically different DNA lesions [23, 26] and even nonpolar purine analogues [27]. It recognizes these DNA lesion by sliding along the DNA strand and flipping the damaged bases into the active site [28, 29]. The excision of the base results, like for most bifunctional glycosylases, in an iminium intermediate which covalently links the enzyme with the DNA (Schiff base intermediate) [30–32]. The cleavage of the glycosidic bond by Fpg proceeds with initial ribose ring opening (ribose-protonated mechanism) [33, 34] rather than via initial base protonation as previously assumed [35].

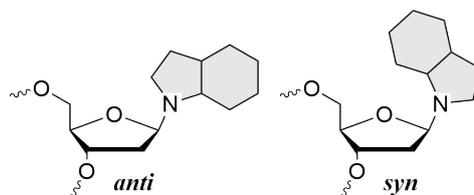
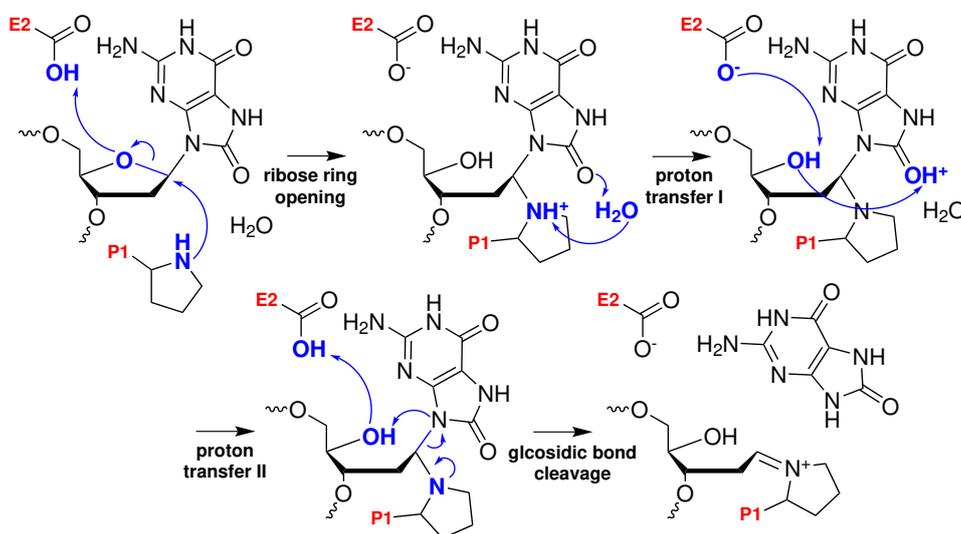


Figure 2: Purine nucleotide in *anti*- and *syn*-conformation. The purin base is sketched in grey.

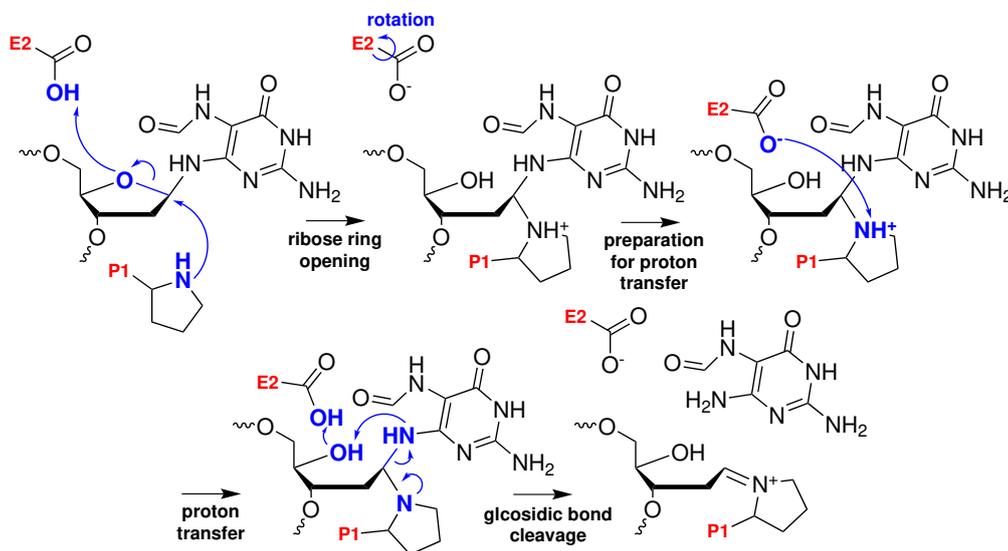
Fpg can accommodate purine nucleotides in *syn*- and *anti*-conformation (see fig. 2). For the *syn*-bound 8OG a ribose-protonated mechanism with base specific protonation has been discovered (scheme 2) [33]. The ribose ring opening in the first reaction step is induced by the nucleophilic attack of proline P1 at C1' of the ribose and facilitated by protonation of O4' by glutamate E2. The proton of P1 is then (indirectly) transferred in two successive steps via O8 of the base onto E2. In a last reaction step, the glycosidic bond is cleaved accompanied with an indirect proton transfer from E2 onto N9 of the base.



Scheme 2: Base-specific ribose-protonated mechanism of the excision of *syn*-bound 8OG by Fpg [33].

A completely base-independent variation of the ribose-protonated mechanism has been found by the investigation of the excision of the *anti*-bound lesion FapyG (scheme 3) [34].

Here, the ribose ring opening in the first reaction step follows the same mechanism as the base-specific pathway. The only difference between the two variations of the ribose-protonated mechanisms is how the proton is transferred from P1 to E2. After a rearrangement of the E2 side chain in the base-independent mechanism, the proton can be transferred directly. Thus, the damaged base is not involved in the proton transfer. The last reaction step, in which the glycosidic bond is cleaved, is again the same as in the base-specific mechanism.



Scheme 3: Excision of FapyG by Fpg via a base-independent ribose-protonated mechanism [34].

So far, crystal structures of Fpg in complex with 8OG exist only with the nucleotide in *syn*-conformation (PDB code: 1R2Y [36], 4CIS [33]), but there are indications that *anti*-8OG can also bind to Fpg [37–41]. However, it remained unclear, whether the glycosidic bond of the *anti*-bound 8OG actually can be cleaved by Fpg.

For this reason, the base-independent ribose-protonated excision mechanism was investigated for *anti*-8OG in **Publication I**. Furthermore, this base-independent mechanism is explored for *syn*-bound 8OG as an alternative excision pathway to the previously found base-specific mechanism [33].

1.2 Active DNA Demethylation

Epigenetic modifications are chemical alterations to DNA or histones that do not alter the primary DNA sequence [42]. These modifications mediate tissue specific expression profiles and are inherited by the daughter cell [43, 44]. In higher eukaryotes, 5-methyl cytosine (5mC, see fig. 3) is an important epigenetic modification [42–45].

Although the methylation pattern of somatic DNA is stably maintained, the methylation of DNA is dynamically regulated involving active DNA demethylation at specific stages of development [42–46]. Active DNA demethylation involves a chemical transformation of the methyl group. One possible transformation is the successive oxidation by the ten-eleven translocation (TET) enzymes.

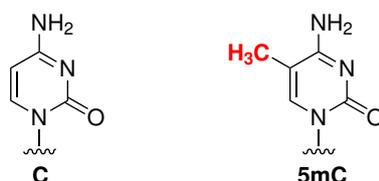
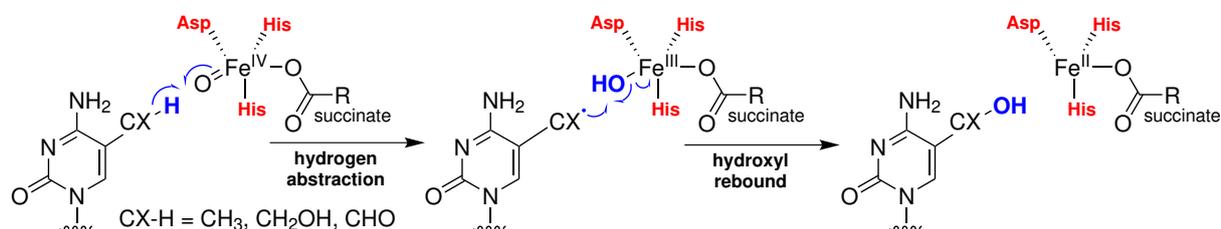
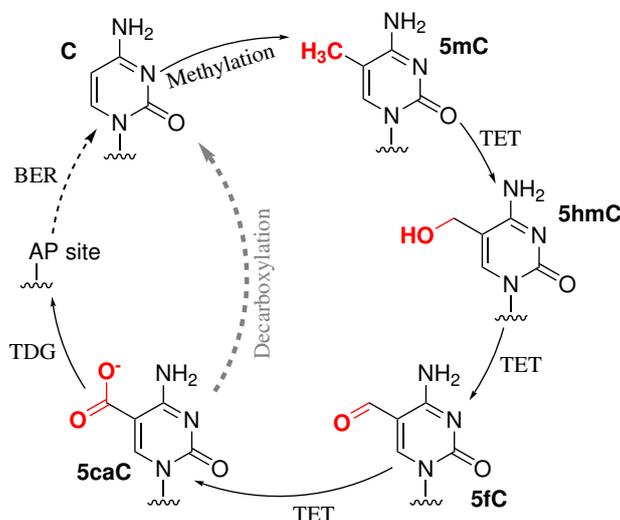


Figure 3: Canonical cytosine base (C) and 5-methyl cytosine (5mC).



Scheme 4: Oxidative reaction catalyzed by the TET proteins. The substrates 5mC, 5hmC, 5fC are oxidized to 5hmC, 5fC, and 5caC, respectively.

TET proteins are Fe(II)/ α -ketoglutarate dependent dioxygenases which iteratively convert 5mC into 5-hydroxymethyl cytosine (5hmC), 5-formyl cytosine (5fC), and 5-carboxy cytosine (5caC). The oxidative reactions catalyzed by the TET proteins involve a hydrogen abstraction from the substrate by the activated Fe(IV)-oxo complex and a subsequent hydroxyl rebound (scheme 4) [47–49]. The oxidized base 5caC can then be recognized and excised by the thymine-DNA glycosylase (TDG) and subsequently be replaced by a canonical cytosine base in the BER pathway [50, 51] (see scheme 5).

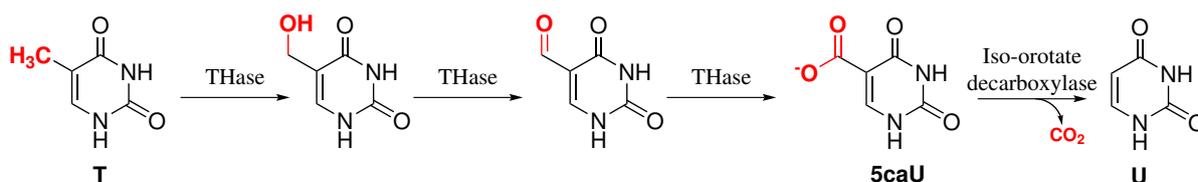


Scheme 5: TET mediated demethylation of 5caC. 5mC is successively oxidized to 5caC via 5hmC and 5fC. 5caC can then be replaced by an unmodified cytosine in the BER pathway (see section 1.1) or be actively decarboxylated by a yet undetected decarboxylase in an alternative pathway. The wavy line connected to N1 of the base indicates the connection to the desoxyribose.

However, there is evidence for an alternative active demethylation pathway downstream

of the TET mediated oxidation [52–55]. Studies have suggested that this pathway involves the enzymatic decarboxylation of 5caC to C by a yet undetected putative decarboxylase (scheme 5, dashed grey line). Although, the decarboxylation of 5caC was in the focus of several studies [53–55], the exact mechanism of this reaction remains unclear.

Due to the similarity of the TET mediated demethylation to the thymine(T)-to-uracil(U) conversion in the pyrimidine salvage pathway of some fungi, the latter might provide insights for the decarboxylation of 5caC in active demethylation. In this pathway, the methylgroup of the substrate T that is removed is successively oxidized in a similar way than in active DNA demetylation by a dioxygenase, the THase (scheme 6). The oxidized intermediate, 5-carboxy uracil (5caU) is then decarboxylated by iso-orotate decarboxylase (IDCase) resulting in U.



Scheme 6: Thymine-to-uracil conversion in the pyrimidine salvage pathway.

Although mechanisms have been proposed for the decarboxylation of 5caU by IDCase [56, 57], the catalytic mechanism of IDCase is not fully understood. For the related substrates, 5caC and orotidine, also several decarboxylation mechanisms have been proposed [54, 56–67]. **Publication II** compares the different decarboxylation mechanisms transferred to the decarboxylation of 5caU in the enzymatic environment of IDCase. Furthermore, it examines in detail the reaction coordinate of the mechanism that appears to be most likely the catalytic mechanism of IDCase.

2 Theoretical Background

Notation

In this chapter, the indices over atoms are denoted with capital letters I, J , whereas indices i run over occupied molecular orbitals. In the same way, it is differentiated between the coordinates of atoms, which are addressed with a capital \mathbf{R} , and the coordinates of electrons (lower case \mathbf{r}).

2.1 Molecular Force Fields

Molecular force fields (FFs) are used for the description of a molecular system. Within a FF, the atoms are considered as point masses. The force that is acting on an atom is defined by an effective potential described by the laws of classical mechanics [68]. This effective potential is a conservative field that only depends on the coordinates of the point masses.

For a conventional, non-polarizable FFs, the effective potential consists of the sum of the contributions of all interactions between the nuclei:

$$E_{\text{pot}} = \sum_{\text{bonds}} E_{\text{bonds}}^r + \sum_{\text{angles}} E_{\text{angles}}^\theta + \sum_{\text{torsion}} E_{\text{dihedrals}}^\gamma + \sum_{I < J} E_{\text{nonbonded}}^{IJ}. \quad (1)$$

For the covalent bond stretching interaction E_{bonds}^r between two atoms with distance R , and the bond angle bending E_{angles}^θ between three atoms defining the angle θ , often a harmonic approximation is chosen as in the AMBER force field [69]:

$$E_{\text{bonds}}^R = K_R(R - R_{eq})^2, \quad (2)$$

$$E_{\text{angles}}^\theta = K_\theta(\theta - \theta_{eq})^2, \quad (3)$$

where K_R, K_θ are force constants and r_{eq}, θ_{eq} equilibrium bond length and angle. The contribution of dihedral ϕ between four atoms $E_{\text{dihedrals}}^\phi$ can be described by one or a sum of several sinusoidal terms:

$$\frac{1}{2}V_n(1 + \cos(n\phi - \gamma_n)), \quad (4)$$

with periodicity n , associated amplitude (dihedral force constant) V_n , and phase shift γ_n . The nonbonded potential $E_{\text{nonbonded}}^{IJ}$ between atoms I and J with charges q_I and q_J separated with distance R_{IJ}

$$E_{\text{nonbonded}}^{IJ} = E_{\text{Coulomb}}^{IJ} + E_{\text{van der Waals}}^{IJ} \quad (5)$$

arises from Coulomb interactions, which are calculated according to the classical Coulomb's law

$$E_{\text{Coulomb}}^{IJ} = \frac{q_I q_J}{\epsilon R_{IJ}}, \quad (6)$$

with the reciprocal Coulomb's constant ϵ , and van der Waals interactions taken into account by the Lennard-Jones 12-6 potential [70, 71]:

$$E_{\text{van der Waals}}^{IJ} = \frac{A_{IJ}}{R_{IJ}^{12}} - \frac{B_{IJ}}{R_{IJ}^6}, \quad (7)$$

with the Lennard-Jones parameters $A_{I,J}$ and $B_{I,J}$. The parameters needed for the energy terms in eq. (2)-(5) are fitted to experimental data or data obtained by quantum chemical calculations and are normally fitted for a specific type of molecules such as amino acids, nucleic acids, or small organic molecules.

Consequently, the description of unusual residues, such as ligands in enzymes, is often difficult as it is not easy to estimate how good they are described by the existing parameters of the force field, or otherwise they would have to be parameterized in an often elaborate procedure which usually requires quantum chemical calculations. Some force fields provide tools which assist the parametrization procedure like, e.g., the MCPB.py tool [72] of the AMBER program suite for the parametrization of metal centers in proteins.

As the potential in eq. (1) contains a sum over all bonds in the system, the bonding situation between all atoms has to be predefined. For that reason bond breaking events and the formation of new bonds as they occur during chemical reactions cannot be described. Furthermore, the chemical environment is only included by the definition of "atom types" [73]. This leads to the disadvantage that these types have to be defined at the beginning of the simulation, so that reactions with changing electronic structure cannot be described.

A method that is able to describe chemical reactions has to allow for a flexible description of the electronic structure of the system. A popular choice here is the density-functional theory.

2.2 Density-Functional Theory

Density-functional theory (DFT) aims to describe a molecular system in terms of the electron density distribution $\rho(\mathbf{r})$. The motivation of DFT was the development of an "approximate practical method [...] of applying quantum mechanics" as Dirac claimed in 1929 [74]. At this time, quantum chemical calculation mostly employed wave functions that depend on $3N$ coordinates for a system with N electrons, which makes the calculations extremely complex and computationally costly. For that reason it seemed to be important to reduce the number of variables in the calculations, which shows why it was desirable to be able to express the energy of the system completely in terms of only one variable, $\rho(\mathbf{r})$, depending on three coordinates.

DFT was initiated by Thomas [75] and Fermi [76] in 1927/28 by the investigation of the basic nature of $\rho(\mathbf{r})$. They defined the first density-functional for an atom by writing the energy per volume in terms of $\rho(\mathbf{r})$ for which they assumed $\rho(\mathbf{r})$ to be slowly varying with respect to the position \mathbf{r} . This assumption is now known as the 'local density approximation' (LDA).

Proof for the possible exactness of the density-functional formalism was given in 1964 by Hohenberg and Kohn with the Hohenberg-Kohn theorems [77]. The first theorem guarantees that the ground-state density $\rho_0(\mathbf{r})$ explicitly determines the external potential V_{ext} of the nuclei to within a constant. With this one-to-one relationship all terms of the Hamiltonian operator are known in terms of $\rho_0(\mathbf{r})$. This means that all ground-state observables can be expressed as a density functional including the total energy of the system. Hohenberg and Kohn showed that the density dependent energy functional satisfies the variational principle:

$$E_0 = \min_{\rho(\mathbf{r})} E[\rho, V_{\text{ext}}], \quad (8)$$

with

$$E[\rho, V_{\text{ext}}] = \int d\mathbf{r} V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) + F[\rho]. \quad (9)$$

Formally eq. (8) gives the exact ground-state energy, however, the exact "universal" functional $F[\rho(\mathbf{r})]$ is so far not known.

Modern density-functionals are based on the Kohn-Sham formalism [78]. In this formulation, the physical many-particle system is mapped onto a system with non-interacting electrons that gives the same electron density $\rho(\mathbf{r})$ as the physical system [78]. To obtain the same distribution of the electrons, the external potential V_{ext} has to be replaced by a different external potential V_s in the non-interacting system.

Additionally, Kohn and Sham introduced the following decomposition for the "universal" functional in eq. (9):

$$F[\rho] = T_s[\{\phi_i\}] + E_J[\rho] + E_{\text{xc}}[\rho]. \quad (10)$$

The advantage of this formulation is that not only the energy contribution arising from the classical Coulomb potential for electrons

$$E_J[\rho] = \frac{1}{2} \iint d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \quad (11)$$

but also the kinetic energy of the non-interacting system T_s can be calculated exactly. $T_s[\{\phi_i\}]$ is evaluated in terms of the Kohn-Sham orbitals ϕ_i

$$T_s = -\frac{1}{2} \sum_i^{N_{el}} \int d\mathbf{r} \phi_i^*(\mathbf{r}) \nabla^2 \phi_i(\mathbf{r}) \quad (12)$$

which are obtained by self-consistently solving the Kohn-Sham equations

$$\left[-\frac{1}{2}\nabla^2 + V_s(\mathbf{r})\right]\phi_i(\mathbf{r}) = \varepsilon_i\phi(\mathbf{r}), \quad (13)$$

with the effective Kohn-Sham potential

$$V_s(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + \frac{\partial E_J[\rho]}{\partial \rho(\mathbf{r})} + \frac{\partial E_{\text{xc}}[\rho]}{\partial \rho(\mathbf{r})}. \quad (14)$$

The only approximation that has to be made is the one for the exchange-correlation (xc) energy E_{xc} , which accounts for electron exchange and correlation. It also provides a correction of T_s to obtain the kinetic energy of the electrons of the interacting system. Within the LDA an approximation for E_{xc} is given by

$$E_{\text{xc}}^{\text{LDA}} = \int d\mathbf{r} \rho(\mathbf{r})\varepsilon_{\text{xc}}^{\text{LDA}}[\rho(\mathbf{r})], \quad (15)$$

where $\varepsilon_{\text{xc}}[\rho(\mathbf{r})]$ is the exchange and correlation energy per particle of a homogeneous electron gas. An approximation that goes beyond the LDA is the generalized gradients approximation (GGA) [79]. Within the GGA, the exchange and correlation energy is not only dependent on $\rho(\mathbf{r})$ itself but also on its gradient $\nabla\rho(\mathbf{r})$:

$$E_{\text{xc}}^{\text{GGA}} = \int d\mathbf{r} \rho(\mathbf{r})\varepsilon_{\text{xc}}^{\text{GGA}}[\rho(\mathbf{r}), \nabla\rho(\mathbf{r})]. \quad (16)$$

This group includes for example the functional of Perdew, Burke, and Ernzerhof (PBE) [80] or a combination of the Becke 1988 exchange functional and the Perdew 86 correlation functional (BP86) [81, 82].

A further improvement over GGAs are functionals that include a fraction of exact exchange energy E_x^{exx} and a compatible description of the correlation energy [79]. Such hybrid density functionals mix some exact exchange into E_{xc} :

$$E_{\text{xc}}^{\text{hybrid}} = E_{\text{xc}}^{\text{GGA}} + a(E_x^{\text{exx}} - E_x^{\text{GGA}}). \quad (17)$$

The B3LYP functional [83] is probably the most prominent representative of this group. B3LYP is a combination of Becke’s three-parameter exchange [84] and the nonlocal correlation functional B_c^{LYP} of Lee, Yang, and Parr [85]:

$$E_{\text{xc}}^{\text{B3LYP}} = (1 - a_0)E_x^{\text{LDA}} + a_0E_x^{\text{exx}} + a_x\Delta E_x^{\text{B88}} + a_cE_c^{\text{LYP}} + (1 - a_c)E_c^{\text{VWN}}, \quad (18)$$

where ΔE_x^{B88} is Becke’s gradient correction to the exchange functional [81] and E_c^{VWN} the VWN local correlation expression [86]. The three semiempirical coefficients are set to $a_0 = 0.20$, $a_x = 0.72$, and $a_c = 0.81$ [83].

Although DFT certainly has met the aim to be a practical applicable QM method, which is confirmed by the frequent occurrence in publications [87], it is still not affordable for extended molecular systems like solvated enzymes. To be still able to describe enzymatic reactions at QM level, often a combined quantum mechanics/molecular mechanics approach is applied.

2.3 Quantum Mechanics/Molecular Mechanics

The idea of combining QM and classical energy calculations as a compromise between accuracy and computational cost for the calculation of large molecular systems was introduced by Warshel and Levitt [88]. Within this quantum mechanics/molecular mechanics (QM/MM) approach only the residues in the region of interest, for example those that are involved in the chemical reaction, are considered quantum mechanically (QM region), while the remaining residues are described by a molecular mechanics (MM) force field (MM region) (fig. 4). The handling of electrostatic interactions between MM and QM

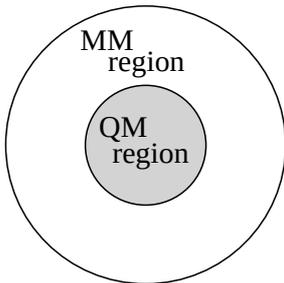


Figure 4: Partition of the complete system into a region which is considered at quantum mechanical level of theory (QM region) and a region described by a molecular mechanics force field (MM region).

region is important as the non-covalent interactions are significant over long ranges. A possible embedding scheme for this purpose is the electrostatic embedding [89, 90]. Here,

the electrostatic interactions are taken into account at QM level by including the partial charges of the MM atoms as point charges in the QM calculation.

Within the QM/MM separation, bonds that span over the QM-MM boarder require special treatment. A common approach is the "link atom" method, where the QM region is commonly saturated with hydrogen atoms which are not present in the MM calculation [91, 92]. These hydrogen link atoms (HL) are placed onto the connection between the QM (Q1) and MM boundary atoms (M1) (see fig. 5). The close position of HL to

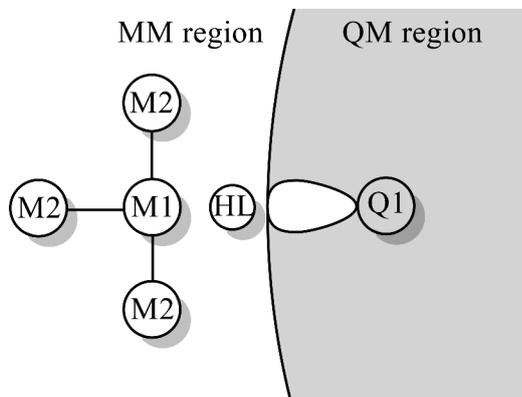


Figure 5: Schematic illustration of a link atom (HL) for the treatment of bonds crossing the QM/MM boundary. Representation following Ref. [92].

M1 leads in case of electrostatic embedding to an overpolarization of the Q1-HL bond in the QM calculation. This can be avoided by applying a charge shift correction, where the partial charge of M1 is redistributed onto/towards [91, 93] the second layer of MM atoms (M2). This causes a changed dipole moment which can be corrected by the addition of a pair of point charges with equal magnitude but opposite sign to each M2 atom along the M1-M2 bond direction [91].

The total energy of the complete system is calculated in an "additive scheme", like it is implemented in ChemShell [91], as a sum of the energy of the QM region E_{QM} and the MM region E_{MM} :

$$E_{\text{QM/MM}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{coupling}} \quad (19)$$

In E_{MM} all angle and dihedral contributions are contained that comprise at least one atom of the MM region but can comprise one or more atoms of the QM region [91]. For that reason the hybrid energy has be corrected by a coupling term E_{coupling} to avoid double counting of energy contributions.

The analytic derivative of the QM/MM energy with respect to the position of the atoms, which is needed in structure optimizations or dynamics simulations, has to be corrected if link atoms are present [91]. The artificial contribution of the link atoms to the derivative can be eliminated completely since the coordinates of link atom positions are determined by the coordinates of the respecting QM and MM boarder atoms.

2.4 Molecular Dynamics Simulations

In classical molecular dynamics (MD) simulations, the motion of atoms or molecules are described in terms of classical mechanics. The relationship between position \mathbf{R}_I of a

particle I with mass m_I and the force \mathbf{F}_I that is acting on it over time is defined by Newton's equations of motion [94]:

$$\frac{d^2\mathbf{R}_I(t)}{dt^2} = \frac{\mathbf{F}_I}{m_I}, \quad (20)$$

$$\mathbf{F}_I = -\frac{\partial E(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N)}{\partial \mathbf{R}_I}, \quad (21)$$

where the potential E is typically described by a molecular force field (see section 2.1). The position of a particle at an instant of time $t_0 + t$ can be obtained by integration over eq. (20), for which different numerical integrators exist. One example is the Velocity-Verlet algorithm [95, 96], which is illustrated in algorithm 1. In this algorithm, the position of N particles $\underline{\mathbf{R}} = (\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N)$ is propagated in discrete time steps Δt , where at each time step, at first the positions of the atoms are updated using the positions, velocities $\underline{\mathbf{v}}$, and accelerations $\underline{\mathbf{a}}$ of the atoms of the previous time step. Next, the accelerations of the current time step are calculated using the forces on the atoms caused by the potential (eq. (21)). Finally, the velocities are updated using the accelerations of the current and previous time step. A possible approach for the initial velocities $\underline{\mathbf{v}}_{\text{ini}}$ is to set them to zero, or use a Maxwellian distribution for it [68].

Algorithm 1 Verlocity Verlet algorithm.

Input: initial geometry configuration $\underline{\mathbf{R}}_{\text{ini}}$ at t_0 ,
array of masses \mathbf{m} of all atoms,
time step Δt ,
model for description of the energy E of the system

Output: geometry configuration at time t_{max}

Variables: velocities $\underline{\mathbf{v}}$ of all atoms,
acceleration \mathbf{a} acting on an atom,
forces $\mathbf{F}(\underline{\mathbf{R}})$ acting on an atom

```

1:  $\underline{\mathbf{R}}(t_0) := \underline{\mathbf{R}}_{\text{ini}}$ 
2:  $\underline{\mathbf{v}}(t_0) := \underline{\mathbf{v}}_{\text{ini}}$ 
3:  $\underline{\mathbf{a}}(t_0) := \frac{\mathbf{F}(\underline{\mathbf{R}}_{\text{ini}})}{\mathbf{m}}$ 
4: for  $t = t_0 + \Delta t$  to  $t = t_{\text{max}}$  step  $\Delta t$  do
5:   //update positions
6:    $\underline{\mathbf{R}}(t) = \underline{\mathbf{R}}(t - \Delta t) + \underline{\mathbf{v}}(t - \Delta t)\Delta t + \frac{1}{2}\underline{\mathbf{a}}(t - \Delta t)\Delta t^2$ 
7:   //calculate current accelerations
8:   for each atom  $I$ 
9:      $\mathbf{a}_I(t) = \frac{\mathbf{F}_I(\underline{\mathbf{R}}(t))}{m_I}$ 
10:  end for
11:  //update velocity
12:   $\underline{\mathbf{v}}(t) = \underline{\mathbf{v}}(t - \Delta t) + \frac{1}{2}(\underline{\mathbf{a}}(t - \Delta t) + \underline{\mathbf{a}}(t))\Delta t$ 
13: end for

```

To picture all motions of the system by the numeric integration adequately, the used time step has to be much smaller than the period of the highest vibrational frequency of

the system [68]. Due to the fast hydrogen bond stretching vibrations of over 3000 cm^{-1} , the time step is limited to about 1 fs. Constraining the length of all bonds to hydrogens to the equilibrium value, which can be done by algorithms like SHAKE [97], allows for a larger time step of up to 5 fs. Even if the calculations at single time step are not computationally demanding, the simulations today are restricted to the nano second scale due to the huge number of steps that have to be evaluated.

Since classical MD simulations rely on molecular force fields (section 2.1), such simulations can not describe a change in the electronic structure of the system. The description of a changing electronic environment would require an *ab initio* molecular dynamics simulation, in which the forces acting on the atoms are obtained by "on-the-fly" calculations during the propagation [98]. The high computational demand of *ab initio* MD simulations, however, restricts the simulation time rather on the pico second scale.

Solving Newton's equations of motion (eq. (20), eq. (21)) for a system with constant box size, a microcanonical ensemble is obtained [68]. Such an ensemble is also called *NVE* ensemble, as in these simulations the number of particles (N), the volume of the system (V), and the energy (E) stays constant during the simulation, while the temperature and pressure are variable. For physiological processes, constant temperature or constant-pressure simulations might be more realistic. A canonical *NVT* ensemble with constant temperature (T) or a *NPT* ensemble with constant pressure P can be obtained by a thermostat or barostat, which e.g. rescales the velocities [99–101] or couples the system with an external heat bath [102].

2.5 Structure Optimization

In general, structure optimization is known as searching procedure for an optimal spacial arrangement of a set of atoms. The overall optimal arrangement for the system is the ground state geometry at 0 K and corresponds to the global minimum of the potential energy surface (PES) with respect to the (relative) positions of the atoms. The PES (see fig. 6) of symmetric H_2O for example has two equivalent minima corresponding to the structures with 104.5° (GS₁) and 255.5° (GS₂) HOH-angles. For accessing the thermochemical properties of the system, all optimal arrangements, which correspond to local minima of the PES, are of interest.

At any local minimum of the PES, where the system with N atoms has the configuration $\underline{\mathbf{R}} = (\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N)$, the gradient of energy E , which is an array of the negative forces acting on each atom (eq. (21)), vanishes:

$$\mathbf{g}(\underline{\mathbf{R}}) = \nabla E(\underline{\mathbf{R}}) = -(\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_N), \quad (22)$$

$$\mathbf{g}(\underline{\mathbf{R}}) = 0 \quad (23)$$

and the Hessian $\mathbf{H}(\underline{\mathbf{R}})$, with

$$H_{IJ}(\underline{\mathbf{R}}) = \frac{\partial^2 E(\underline{\mathbf{R}})}{\partial \mathbf{R}_I \partial \mathbf{R}_J}, \quad (24)$$

has only positive eigenvalues.

Various algorithms have been developed for finding local minima of PES, which follow the general iterative procedure shown in algorithm 2. Starting with an initial geometry configuration $\underline{\mathbf{R}}_{\text{ini}}$, the geometry configuration is iteratively altered by $\Delta \underline{\mathbf{R}}$ until the forces acting on each atom are below a certain threshold. The forces are calculated according

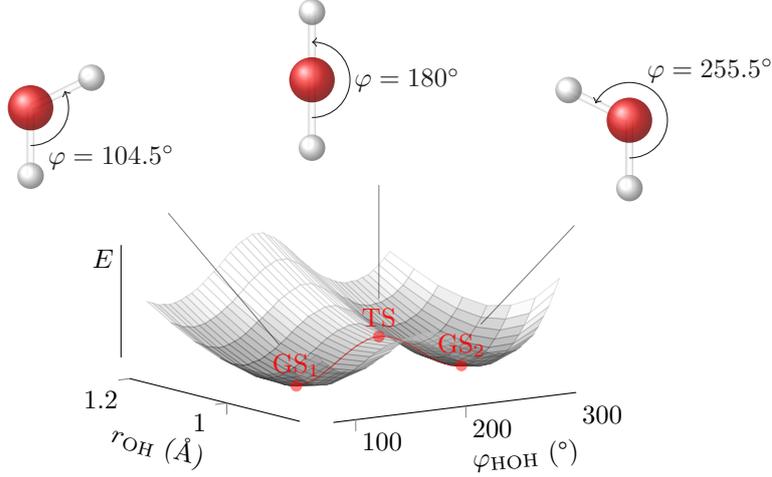


Figure 6: PES of a C_{2V} symmetric water molecule with respect to the OH-bond length r_{OH} and HOH-angle φ_{HOH} .

to any model for the description of molecular systems, for example using molecular force fields (section 2.1), quantum-mechanics (e.g., DFT section 2.2), or a QM/MM approach (section 2.3). The algorithms mainly differ in the calculation of the increments of the

Algorithm 2 General procedure for finding a local minimum on the PES.

Input: initial geometry configuration $\underline{\mathbf{R}}_{ini}$,

model for description of the energy E of the system

Output: geometry configuration for which the PES has a local minimum

Variables: current geometry configuration $\underline{\mathbf{R}}$, array of forces $\underline{\mathbf{F}}(\underline{\mathbf{R}})$ acting on all atoms

```

1:  $\underline{\mathbf{R}} := \underline{\mathbf{R}}_{ini}$ 
2: while true do
3:    $\underline{\mathbf{F}}(\underline{\mathbf{R}}) := \frac{-dE(\underline{\mathbf{R}})}{d\underline{\mathbf{R}}}$ 
4:   if all  $|\mathbf{F}_I| < \text{threshold}$  then
5:     return  $\underline{\mathbf{R}}$ 
6:   else
7:      $\underline{\mathbf{R}} := \underline{\mathbf{R}} + \Delta\underline{\mathbf{R}}$ 
8:   end if
9: end while
    
```

atom configuration $\Delta\underline{\mathbf{R}}$.

Newton and quasi-Newton methods are based on a local quadratic approximation of the PES at the position $\underline{\mathbf{R}}_n$ of the current step n [103, 104]:

$$E(\underline{\mathbf{R}}) = E(\underline{\mathbf{R}}_n) + \mathbf{g}_n^T (\underline{\mathbf{R}} - \underline{\mathbf{R}}_n) + \frac{1}{2} (\underline{\mathbf{R}} - \underline{\mathbf{R}}_n)^T \mathbf{H}_n (\underline{\mathbf{R}} - \underline{\mathbf{R}}_n), \quad (25)$$

with the gradient \mathbf{g}_n and Hessian \mathbf{H}_n of the current configuration $\underline{\mathbf{R}}_n$. Differentiation of eq. (25) with respect to the atom positions $\underline{\mathbf{R}}$ yields an approximation for the gradient:

$$\mathbf{g}(\underline{\mathbf{R}}) = \mathbf{g}_n + \mathbf{H}_n (\underline{\mathbf{R}} - \underline{\mathbf{R}}_n). \quad (26)$$

Considering that the gradient $\mathbf{g}(\underline{\mathbf{R}})$ in eq. (26) becomes zero at a stationary point, the displacement from the minimum $\Delta\underline{\mathbf{R}}(\underline{\mathbf{R}}) = \underline{\mathbf{R}} - \underline{\mathbf{R}}_n$ at point $\underline{\mathbf{R}}_n$ is given by:

$$\Delta\underline{\mathbf{R}}(\underline{\mathbf{R}}) = -\mathbf{H}_{n-1}^{-1}\mathbf{g}_n. \quad (27)$$

In the Newton method [105], the Hessian is evaluated for the calculation of each replacement step eq. (27). Since the explicit calculation of the Hessian is computationally demanding, other methods, the quasi-Newton methods, use an approximate Hessian which is obtained from the Hessian of the previous step \mathbf{H}_{n-1} by an update scheme:

$$\mathbf{H}_n = \mathbf{H}_{n-1} + \Delta\mathbf{H}. \quad (28)$$

This update scheme has to satisfy the Newton condition

$$\Delta\mathbf{g} = \mathbf{H}_n\Delta\underline{\mathbf{R}}, \quad (29)$$

with $\Delta\mathbf{g} = \mathbf{g}(\underline{\mathbf{R}}_n) - \mathbf{g}(\underline{\mathbf{R}}_{n-1})$ and $\Delta\underline{\mathbf{R}} = \underline{\mathbf{R}}_n - \underline{\mathbf{R}}_{n-1}$. A widely used update of the Hessian in eq. (28) is the Broyden-Fletcher-Goldfarb-Shannon [106–109] (BFGS) update:

$$\Delta\mathbf{H}^{\text{BFGS}} = \frac{\Delta\mathbf{g}\Delta\mathbf{g}^T}{\Delta\mathbf{g}\Delta\underline{\mathbf{R}}} - \frac{\mathbf{H}_{n-1}\Delta\underline{\mathbf{R}}\Delta\underline{\mathbf{R}}^T\mathbf{H}_{n-1}}{\Delta\underline{\mathbf{R}}^T\mathbf{H}_{n-1}\Delta\underline{\mathbf{R}}}. \quad (30)$$

As an initial Hessian, for example, a scaled identity matrix, a two point finite difference of the gradient, or even an explicitly calculated Hessian can be used.

Besides the local minima of a PES also the minimum energy paths connecting the minima are of interest from a chemical point of view. These paths represent the path of the conversion from one stable state of the system to another stable state, which represents a chemical reaction or a conformational rearrangement. The highest point along such a path is the transition state (TS) of the conversion. Such a TS is a maximum in one dimension of the PES (direction of the reaction coordinate) and a minimum in all other directions. In case of the water molecule (fig. 6), the minimum energy path between GS_1 and GS_2 describes the opening (or closing) of the HOH-angle (red line) and passes through the linear TS. Scanning such an reaction path on a PES can be done by series of constraint optimizations.

2.6 Constrained Structure Optimization

The energy profile of a chemical reaction can be obtained by subsequent optimizations along the reaction coordinate (RC). The RC is defined by a bond, angle, dihedral, or a linear combination of several bonds/angles/dihedrals. By applying harmonic restraints

$$V_{\text{res}} = \frac{1}{2}k(x - x_0)^2 \quad (31)$$

to each component x of the RC with the equilibrium value x_0 , the RC it can be constraint to a specific value [104, 110]. These restraints are added as a penalty function to the energy expression. This approach is known as "coordinate driving" [104] or sometime as "adiabatic mapping approach".

In large molecular systems the PES has a complex structure with a huge number of local minima. In order to find the desired reaction path on the PES it is often advantageous to smoothen the PES by removing the degrees of freedom of the system that are not of interest in the reaction. Therefore, the positions of atoms that are not within a certain distance of the reaction center can be kept fix. For the studied reactions in this thesis, the atoms of all residues within 10 Å of the QM region were kept frozen.

2.7 Nudged Elastic Band Method

Coordinate driving (see section 2.6) gives a good estimate for a low energy path between two minima on the PES but it does not assure that the found path is a MEP. The nudged elastic band (NEB) method, developed by Milles and Jónsson [111–113], guarantees a MEP by optimizing the whole path connecting the minima.

Within the NEB method, the energy profile of the MEP is characterized by equally distributed optimized geometries, so called images, along the path. The equal distribution is maintained by spring forces that connect the images. As initial guess for the path, the linear synchronous transit path (linear interpolation between reactant and product structure) can be used (see fig. 7, grey path) [104]. If intermediate structures are known they can be added to the guess [110, 114].

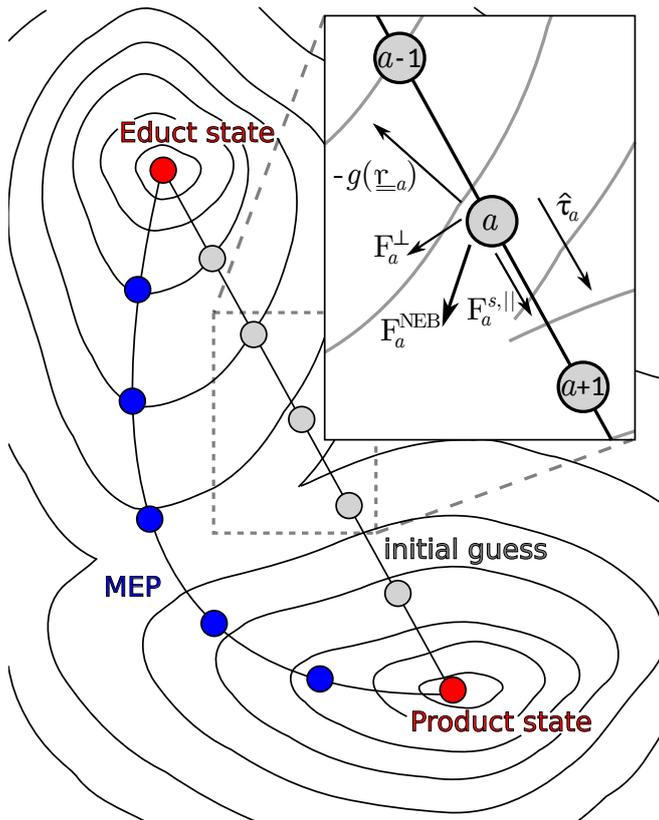


Figure 7: Illustration of the NEB method. Representation following Ref. [115]

For minimization of the path, each image is relaxed only perpendicular to the path. Therefore, an effective force $\mathbf{F}_a^{\text{NEB}}$ [115]

$$\mathbf{F}_a^{\text{NEB}} = \mathbf{F}_a^\perp + \mathbf{F}_a^{\text{s},\parallel}. \quad (32)$$

is used for geometry extrapolation of image a (see fig. 7). It is composed of the force \mathbf{F}_a^\perp arising from the potential, where the components parallel to the path have been removed according to

$$\mathbf{F}_a^\perp = -g(\mathbf{r}_a) + g(\mathbf{r}_a) \cdot \hat{\tau}_a \hat{\tau}_a, \quad (33)$$

with the gradient $g(\mathbf{r}_a)$ of the geometrical configuration in image a according to eq. (23) and the tangent $\hat{\tau}_a$ along the path. The second component of $\mathbf{F}_a^{\text{NEB}}$ is the force $\mathbf{F}_a^{\text{s},\parallel}$,

which is acting parallel along the path and is arising from the spring forces, assuring that the images stay equally distributed along the path.

For finding the saddle point along the MEP, a climbing image NEB can be calculated [116]. The "climbing image" gives the highest point along the path as it is relaxed only perpendicular to the path and it moves up along the path during the optimizations while it does not sense any spring forces to the neighboring images.

In **Publications I** and **II** the methods as described within this section were employed for the calculation of the energy profiles of possible enzymatic reaction pathways, which clarify the exact course of the catalytic reaction.

3 Publications

Publication I: Base-Independent DNA Base-Excision Repair of 8-Oxoguanine

A. Kreppel, I. D. Blank, C. Ochsenfeld,
"Base-Independent DNA Base-Excision Repair of 8-Oxoguanine"
J. Am. Chem. Soc., **2018**, *140*, 4522–4526

Abstract: Living organisms protect their genome from gene mutation by excising damaged DNA bases. Here, 8-oxoguanine (8OG) is one of the most abundant DNA lesions. In bacteria the base excision is catalyzed by the enzyme formamidopyrimidine-DNA-glycosylase (Fpg), for which two different orientations of 8OG binding into the active site of Fpg have been proposed: *syn*- and *anti*-conformation. Here, we present a new ribose-protonated repair mechanism for 8OG that is base-independent and can excise 8OG in both conformations. Using high-level QM/MM calculations with up to 588/573 atoms in the QM sphere, the activation barrier is computed in excellent agreement with the experimentally measured value. Since the excised base itself is not directly involved in the mechanism, this implies that lesion discrimination does not occur within the active site of the enzyme.

Reprinted with permission from:

A. Kreppel, I. D. Blank, C. Ochsenfeld,
"Base-Independent DNA Base-Excision Repair of 8-Oxoguanine"
J. Am. Chem. Soc., **2018**, *140*, 4522–4526

Copyright 2018 American Chemical Society.

<https://pubs.acs.org/doi/10.1021/jacs.7b11254>

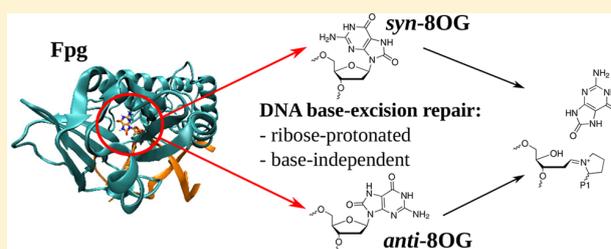
Base-Independent DNA Base-Excision Repair of 8-Oxoguanine

Andrea Kreppel, Iris D. Blank, and Christian Ochsenfeld*[†]

Chair of Theoretical Chemistry, and Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstraße 5-13, Munich, D-81377, Germany

Supporting Information

ABSTRACT: Living organisms protect their genome from gene mutation by excising damaged DNA bases. Here, 8-oxoguanine (8OG) is one of the most abundant DNA lesions. In bacteria the base excision is catalyzed by the enzyme formamidopyrimidine-DNA-glycosylase (Fpg), for which two different orientations of 8OG binding into the active site of Fpg have been proposed: *syn*- and *anti*-conformation. Here, we present a new ribose-protonated repair mechanism for 8OG that is base-independent and can excise 8OG in both conformations. Using high-level QM/MM calculations with up to 588/573 atoms in the QM sphere, the activation barrier is computed in excellent agreement with the experimentally measured value. Since the excised base itself is not directly involved in the mechanism, this implies that lesion discrimination does not occur within the active site of the enzyme.



INTRODUCTION

DNA is constantly exposed to exogenous and endogenous damaging sources, such as radiation, side products of cellular metabolism, or replication errors. Because of its low redox potential, the purine base guanine is especially susceptible to oxidative damage.^{1–3} One of the most abundant oxidative DNA lesions is 8-oxoguanine (8OG).¹ By forming a Hoogsteen base pair with adenine, the 8OG base defect can introduce GC to TA transversion, which explains its mutagenic potential.^{4,5} Elevated levels of 8OG are linked to various neurodegenerative diseases such as multiple sclerosis,⁶ Parkinson's disease,⁷ or Alzheimer's disease.⁸ This underlines the importance of understanding the biological processes of 8OG removal.

To prevent the cell from the negative effects of the oxidative DNA lesions, various mechanisms for recognition and repair exist. One repair pathway is the base excision repair (BER) that is conserved among bacteria and humans.⁹ In this pathway, the damaged base is first recognized and excised by a DNA-glycosylase creating an apurinic site (AP-site) with a covalent Schiff base intermediate that was resolved in a crystal structure (PDB code: 1L1Z).¹⁰ Consecutively the ribose moiety is excised in β - and δ -elimination reactions. In the Gram-positive bacterium *Lactococcus lactis* (*L. lactis*), these reactions are catalyzed by the bifunctional enzyme formamidopyrimidine DNA glycosylase (Fpg).¹¹

There are two possible binding modes for damaged purine bases in the active site of Fpg. While the crystal structure of *L. lactis* accommodates the damaged nucleotide 2,6-diamino-4-hydroxy-5-formamido-pyrimidine (FapyG) in *anti*-conformation,¹² the crystal structure of the homologue enzyme in *Geobacillus stearothermophilus* (*B. st.*) contains a *syn*-8OG nucleotide in its active site.¹³ These two conformations only

differ in the rotation of the base around the glycosidic bond (Figure 1).

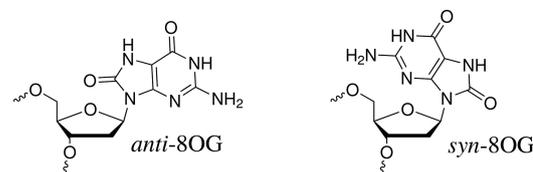


Figure 1. 8OG nucleotide in *anti*- and *syn*-conformation.

So far only crystal structures of Fpg with 8OG bound in *syn*-conformation have been obtained (PDB code: 1R2Y,¹³ 4CIS¹⁷). However, it is assumed that the substrate 8OG can bind in both orientations into the active site of Fpg,^{14–16} where the dihedral of the glycosidic bond of the nucleotide in *anti*-conformation is in the range of -140 to -70 degrees and in *syn*-conformation in the range of 55 to 160 degrees for the following reasons. It has been proposed that in these crystal structures the nucleotide binds artificially in *syn*-conformation as the preparation for crystallization blocks only the reaction of *syn*-8OG¹⁵ or the *syn*-binding mode is stabilized artificially by the mutations introduced for crystallization.¹⁸ Additionally, it has been found that both conformations of 8OG bind stably to Fpg's active site in MD simulations.^{14,15} In the simulation of both orientations similar interactions between base and active site residues have been established. The energetic difference between *syn*- and *anti*-bound 8OG in the active site of *B. st.* Fpg

Received: October 22, 2017

Published: March 26, 2018

is only about 3 kcal/mol with a rotational barrier of 8 and 12 kcal/mol, respectively.¹⁶

Although the binding of 8OG in both conformations appears to be reasonable, it is not clear whether the base can also be excised in both conformations. We recently found an excision mechanism for *syn*-bound 8OG with “base specific protonation” by performing systematic QM/MM calculations,¹⁷ but for *anti*-bound 8OG no excision mechanism is known so far. For another *anti*-bound guanine lesion, FapyG, we recently revealed an excision mechanism where the cleaved base itself is not involved.¹⁹ In the following we will refer to the latter mechanism as “base-independent excision mechanism”. Both the base-specific and the base-independent mechanism are alternative mechanisms of the ribose-protonated excision pathway of Fpg which we have shown to be energetically more favorable^{17,19,20} than the mechanism with direct cleavage of the glycosidic bond by protonation of the base that had been mainly proposed in the literature²¹ prior to our work.

In the present work, we introduce a base-independent excision mechanism for 8OG that is similar to the excision mechanism we found for FapyG. We further compare this base-independent excision mechanism for the excision of *syn*-bound 8OG with the excision of *anti*-bound 8OG. Quantum-chemical methods within a QM/MM approach are employed throughout for calculating the energy profile of the cleavage reaction of 8OG in both binding orientations. Additionally, we will show that this base-independent excision mechanism is an alternative reaction pathway for the excision of *syn*-bound 8OG to the base-specific excision mechanism we have found earlier.

MATERIALS AND METHODS

To achieve best possible comparability between the two conformations of 8OG in Fpg, both systems were prepared in exactly the same way. Each system was set up based upon the crystal structure (PDB code: 1XC8) of *L. lactis*, where Fpg is bound to double stranded DNA that contains carbocyclic FapydG in *anti*-orientation in the active pocket. The damaged base was altered into 8OG with the molefacture tool of VMD.²² At this step the only difference in preparation between the two systems was introduced: in the case of *syn*-8OG the base was rotated around the glycosidic bond to obtain the *syn*-binding mode. The dihedral of the glycosidic bond in the X-ray-based setup of *anti*-8OG was -64° and in the setup of *syn*-8OG the dihedral was set to 140° . In each system the crystal water was removed including the water molecule in the active site that had been found in our earlier work to be artificially due to the O4' to C mutation at the ribose introduced for crystallization.¹⁹ Also the protonation state of the active site residues was adopted from Blank et al.¹⁹ The N-terminal proline was set to be neutral, E2 to be protonated, and E5 to be deprotonated. Additionally, the four cysteines of the zinc finger were set to be deprotonated. All remaining residues were protonated according to standard protonation states, the systems were neutralized with sodium ions and dissolved in a TIP3P box with 10 Å of solvent around the solute. These preparation steps were done using the AmberTool TLEAP.²³

For generation of equilibrated starting structures for the combined quantum mechanical and classical (QM/MM) calculations, force field molecular dynamics (FF-MD) simulations with the AMBER11 force field²³ and the simulation engine NAMD²⁴ were performed (for details on the MD simulation see Supporting Information, (SI) section 1). The force field parameters for the N-terminal proline were taken from Perlow-Poehnelt et al.¹⁴ and for the zinc finger the zinc AMBER force field (ZAFF)²⁵ was used. The damaged base 8OG was parametrized using antechamber (SI, section 1.2).²³

Structure optimizations on the QM/MM level were performed using the DL-Find optimizer implemented in ChemShell.^{26,27} The region described at the QM level with density functional theory

(DFT) was chosen to include the residues P1, E2, and E5 and the damaged nucleotide 8OG (80 atoms) (Figure 2) for reasons of

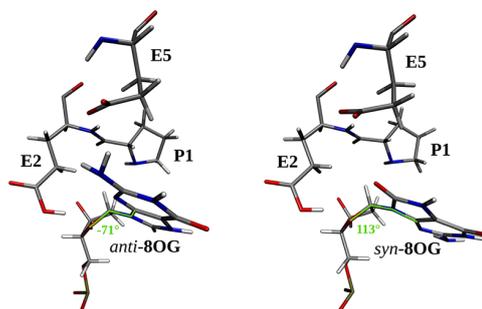


Figure 2. Residues in the QM region of the QM/MM optimized *anti*-8OG setup (left) and *syn*-8OG setup (right).

comparability to results of Sadhegian et al.¹⁷ and Blank et al.¹⁹ The influence of the size of the QM region was studied by performing QM/MM calculations with larger QM spheres (up to 588 atoms) on the first reaction step (for details see SI, section 4). In the structure optimizations the atoms in the radius of 10 Å around the QM residues were relaxed, the remaining atoms were kept fixed. The DFT calculations were performed with the QM program package Fermions++^{28,29} using BP86-D3 and B3LYP-D3 for the first reaction step and B3LYP-D3 for the remaining reaction. For a reasonable balance of accuracy and computational cost the split-valence basis set def2-SVP was selected. In our related study¹⁹ we show that the difference of QM/MM energy profiles calculated using different double- and triple- ζ basis sets are within 3 kcal/mol. For calculating the energy profile of the reaction, the adiabatic mapping approach (SI, section 2) and the nudged elastic band (NEB) method³⁰ implemented in DL-Find was used.

RESULTS AND DISCUSSION

The base-specific excision mechanism that we have found earlier for *syn*-bound 8OG is initiated by a ribose-opening (step 1, Figure 3). This ribose-opening takes place via protonation of the ribose at O4' by a glutamate (E2) and nucleophilic attack of the N-terminal proline at C1' of the ribose. In the second part of the mechanism the proton of P1 is transferred to E2 in two consecutive steps, first via a water molecule onto position O8 of the base (step 2a') and then via the opened ribose ring onto E2 (step 2b'). The last reaction step (step 3) comprises the actual cleavage of the glycosidic bond by protonation of the base at N9 and formation of the Schiff base intermediate.

The base-independent excision pathway of 8OG examined in our present work, shares the first reaction step with the base-specific excision mechanism. Both excision pathways are ribose-protonated. The only difference is how the proton of P1 is transferred to E2 in the second part of the excision mechanism. In our new base-independent pathway for 8OG, the proton is transferred directly (step 2b) and not in two consecutive steps. Before this direct proton transfer, a reorientation of the glutamate side chain of E2 toward P1 has to occur (step 2a). After the proton transfer the E2 side chain returns in its initial position (step 2c). The actual cleavage of the glycosidic bond in the last reaction step is again the same mechanism as in the base-specific excision pathway.

Comparison of the Base-Independent Excision Pathway for *syn*-8OG and *anti*-8OG. In the following we will investigate whether both *syn*- and *anti*-bound 8OG are excised by Fpg with a base-independent mechanism under physiological conditions. Initial work^{17,19} indicated that the first step

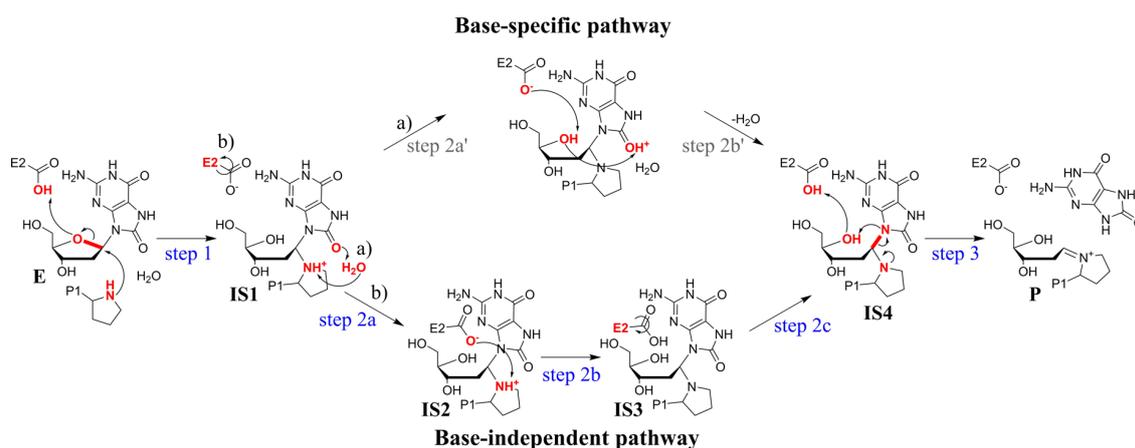


Figure 3. Two alternative paths of the ribose-protonated excision pathway of Fpg: (a) with base-specific protonation found by Sadeghian et al.¹⁷ and (b) a new base-independent mechanism where the excised base is not involved directly.

in base excision of 8OG has the highest energy barrier and is consequently rate determining. We have therefore paid particular attention to computing this step.

The reaction profile of this reaction step is calculated by adiabatic mapping for both binding modes of 8OG for several initial structures (see Figure 4). Different snapshots from the

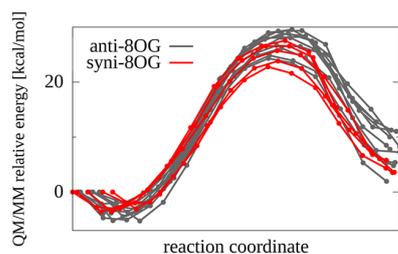


Figure 4. Energy profile for the first reaction step (step 1) of the base excision mechanism of 8OG by Fpg in which the base is bound into the active site in *anti*-conformation (gray) and in *syn*-conformation (red).

respective MD simulation were taken as initial structures. For *syn*-bound 8OG an energy barrier in the range of 25 to 31 kcal/mol with a Boltzmann average³¹ of 26 kcal/mol was calculated for the first reaction step. For the *anti*-conformation an energy barrier in the range of 28 to 33 kcal/mol with a Boltzmann average of 29 kcal/mol was found. The calculated energy barriers are in both cases higher than the experimentally determined activation enthalpy of 19.2 kcal/mol.³² This can be explained by methodological errors and simplifications in the employed simulation model. Using, for example, the DFT-functional BP86 for calculating the energy profile, lower energy barriers are obtained (SI, section 3). Furthermore, it is well-known that increasing the QM region can have a huge impact on the calculated QM/MM energy profile of enzyme reactions.^{33–37} In our case, the energy barriers are significantly reduced for larger QM spheres (Figure 5). This influence is considerably stronger for the *anti*-conformation than for the *syn*-conformation. Here, it is obvious that the geometries of the *syn*- and *anti*- systems differ in the orientation of the base in the active site, which explains that the influence of an increased QM region is different for both systems. With a size-converged QM-region (QM C, 588/573 atoms) the energy barrier has for

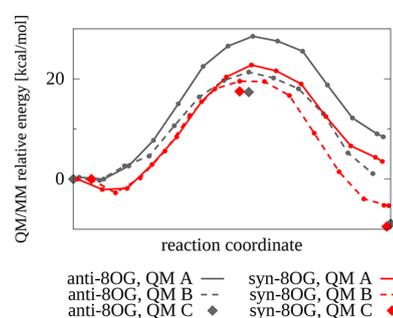


Figure 5. QM/MM energy profile for the first reaction step (step 1) of the base excision mechanism of *anti*-8OG (gray) and *syn*-8OG (red) by Fpg with QM-regions containing 80 atoms (QM A, solid line), 230 atoms (QM B, dashed line), and 588/573 atoms (QM C, squares). BP3LYP-D3/def2-svp//AMBER11-FF were used for calculation. For computational details see SI, section 4.

syn- and *anti*-8OG the same height of 17 kcal/mol which is in excellent agreement with the experimental value of 19.2 kcal/mol.³² As the ribose opening has for *anti*- and *syn*-bound 8OG the same energy barrier, it can be assumed that the first reaction step of the base excision takes place for both conformations under physiological conditions.

To confirm that in both cases the remaining reaction path can also be overcome under physiological conditions, the energy profile of the entire reaction path of the base excision for each conformation was calculated using adiabatic mapping and refined by NEB (Figure 6). The remaining reaction path of the base-independent excision by Fpg consists of four other reaction steps. The second reaction step (step 2a, Figure 3), following the ribose ring opening, is a conformational rearrangement of E2 that prepares for the proton transfer from P1 to E2 in step 2b. After the proton transfer another conformational reorganization occurs (step 2c) to arrange the active site for the actual cleavage of the glycosidic bond in step 3.

While the energy profile of the first and second reaction step is similar for both conformations, the energy profiles of the proton transfer in step 2b differs. The energy barrier for this reaction step is only slightly higher for *syn*-8OG. The energy profile of the subsequent rearrangement of the active site also differs between *syn*- and *anti*-8OG. While in the latter case this

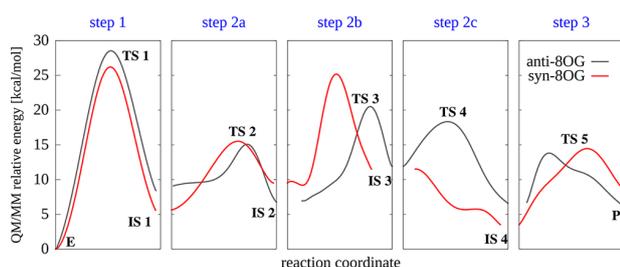


Figure 6. Comparison of the energy profile of the base excision for 8OG bound in *syn*-conformation (red) and *anti*-conformation (gray) to Fpg. For a more detailed view at the energy profile and the definition of the reaction coordinate see SI, section 2.

reaction step has an energy barrier with considerable height, in the case of *syn*-8OG this reaction step is almost barrier free. The actual bond cleavage, however, has again a similar energy profile. In both cases, the reaction is endothermic which can be justified with the back reaction being hindered as the base diffuses out of the active pocket.

Summing up, the energy profiles of the first step exhibit a higher energy barrier than those of the subsequent steps. This implies for both conformations that, if the reaction barrier of the first reaction step can be overcome under physiological conditions, the whole base excision mechanism can occur under these conditions. The maximum energy barriers along the excision pathway of 8OG in both conformations are almost identical for large QM spheres, indicating that excision from both conformations is accessible. The computed barrier of 17 kcal/mol is in excellent agreement with the experimental value of 19.2 kcal/mol.³²

Comparison of the Energetics of Both Excision Pathways for *syn*-8OG. To estimate which one of the two alternative excision pathways actually takes place for the *syn*-bound 8OG, the kinetics of both reaction pathways have to be compared. For this reason, the energy barriers of the base-independent excision pathway calculated in this work are compared to our previously calculated energy profile of the excision of *syn*-8OG according to the base-specific excision mechanism.¹⁷

The energy barrier of the ribose ring opening in reaction step 1 has for both pathways about the same height with 26 and 25 kcal/mol, respectively (Table 1). This was expected, as it is the same reaction mechanism for the first step in both cases (see Figure 3). Thus, the energetics of the subsequent proton transfer is critical for the difference of the kinetics. In the case of

Table 1. Comparison of the Energy Barriers in kcal/mol between the Base-Specific and the Base-Independent Excision Mechanism for *syn*-8OG for Comparable QM Sphere (80 Atoms) and Computational Method^a

reaction step	base-specific ^b	base-independent
step 1	26	25
step 2a'/step 2a	8	9
step 2b'/step 2b	14	15
step 2c		1
step 3	15	10

^aThe energy barriers were calculated at QM/MM level using B3LYP-D3/AMBER11-FF with def-SVP and def2-SVP basis set, respectively.

^bThe values for the base-specific excision pathway were taken from ref 17, Figure SI 14.

the base-specific excision mechanism, the proton transfer occurs in two steps. At first the base is protonated at position O8 via a water molecule (step 2a'). Subsequently, residue E2 is protonated by the base via the OH-group at position C4' of the opened ribose-ring (step 2b'). These two reaction steps have an energy barrier of 8 and 14 kcal/mol, respectively. For the direct proton transfer from P1 to E2 in the base-independent excision mechanism, the E2 side chain has to orientate toward P1. This reorientation has an energy barrier of 9 kcal/mol (step 2a) and the energy barrier for the actual proton transfer is 15 kcal/mol (step 2b). Afterward, the E2 side chain rotates back in its initial position almost barrier free (step 2c). The proton transfer has in both pathways similar energy barriers and thus it is expected that both pathways occur with similar reaction rates. The last reaction step is the cleavage of the glycosidic bond (step 3). Although in this reaction step both pathways have again the same mechanism, the energy barrier heights with 15 and 10 kcal/mol differ. This minor difference in the reaction profile may be attributed to the fact, that the previous calculations on the base-specific excision mechanism were performed on the homologous enzyme of *B. st.* (PDB code: 1R2Y¹³).

CONCLUSION

We have presented a new alternative excision pathway for 8OG with a base-independent mechanism. This mechanism is similar to the excision mechanism we found recently for FapyG, another *anti*-bound purine lesion. The comparison of the energetics of this new “base-independent” excision to our previously found “base-specific” base excision mechanism for *syn*-bound 8OG indicates that *syn*-bound 8OG can be excised by both alternative excision pathways of Fpg.

Furthermore, the excision of *syn*-bound and *anti*-bound 8OG according to the base-independent mechanism is compared. These investigations demonstrate the importance of using a QM sphere with sufficient size in QM/MM studies. Here, the rate determining energy barrier is reduced by up to 11 kcal/mol using a size-converged QM region (588/573 atoms) compared to using a QM sphere comprising 80 atoms. The QM/MM energy profiles of the size-converged QM region suggest that 8OG can be excised by Fpg in both conformations, *syn* and *anti*, due to same height of the maximum energy barriers along both paths and yield good agreement with experimental measurements.³² The occurrence of both excision paths apparently depends on whether both binding modes of 8OG are naturally observed in the active site of Fpg which is an interesting aspect for future investigations. Here, it would be desirable to gain insights into the dynamics of the functional Fpg-8OG complex in solution. Insights into the extrusion process of 8OG out of double stranded DNA into the active site of Fpg resulting in an *anti*-binding mode would also be interesting as it has been done computationally for the *syn*-binding mode.^{38,39}

While the fast sliding of Fpg along the double stranded DNA for base interrogation⁴⁰ and a recent UV resonance Raman study⁴¹ as well as computational studies^{38,39} provide evidence that there occurs substrate discrimination before any binding of the base into the active site, these studies were not able to rule out that there is an additional discrimination in the active site. The present work, together with our earlier work on the excision mechanism of Fpg and its human homologue,^{17,19,20} provides strong evidence for the ribose-protonated and base-independent excision of DNA bases by Fpg. Therefore, the results of our study lead to the conclusion that no discrimination between DNA lesions and canonical DNA

nucleotides occurs within the active site. This conclusion is supported by the fact that Fpg excises also nonpolar analogues of 8OG,² indicating that no hydrogen bonds in the active site can be responsible for substrate recognition.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b11254.

Details of the MD simulations; details of the calculation of the QM/MM energy profiles; BP86 energy profile of step 1; details of QM size convergence (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*christian.ochsenfeld@uni-muenchen.de

ORCID

Christian Ochsenfeld: 0000-0002-4189-6558

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful for financial support through the DFG funding initiative SFB749 (TP C07) and the Excellence Cluster EXC114 (CIPSM).

■ REFERENCES

- (1) Burrows, C. J.; Muller, J. G. *Chem. Rev.* **1998**, *98*, 1109–1152.
- (2) McKibbin, P. L.; Kobori, A.; Taniguchi, Y.; Kool, E. T.; David, S. *J. Am. Chem. Soc.* **2012**, *134*, 1653–1661.
- (3) Jovanovic, S. V.; Simic, M. G. *J. Phys. Chem.* **1986**, *90*, 974–978.
- (4) Le Page, F.; Guy, A.; Cadet, J.; Sarasin, A.; Gentil, A. *Nucleic Acids Res.* **1998**, *26*, 1276–1281.
- (5) Cheng, K. C.; Cahill, D. S.; Kasai, H.; Nishimura, S.; Loeb, L. A. *J. Biol. Chem.* **1992**, *267*, 166–172.
- (6) Vladimirova, O.; O'Connor, J.; Cahill, A.; Alder, H.; Butunoi, C.; Kalman, B. *Mult. Scler.* **1998**, *4*, 413–418.
- (7) Alam, Z. I.; Jenner, A.; Daniel, S. E.; Lees, A. J.; Cairns, N.; Marsden, C. D.; Jenner, P.; Halliwell, B. *J. Neurochem.* **1997**, *69*, 1196–1203.
- (8) Bradley-Whitman, M. A.; Timmons, M. D.; Beckett, T. L.; Murphy, M. P.; Lynn, B. C.; Lovell, M. A. *J. Neurochem.* **2014**, *128*, 294–304.
- (9) Wallace, S. S. *DNA Repair* **2014**, *19*, 14–26.
- (10) Fromme, J. C.; Verdine, G. L. *Nat. Struct. Biol.* **2002**, *9*, 544–552.
- (11) O'Connor, T. R.; Laval, J. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 5222–5226.
- (12) Coste, F.; Ober, M.; Carell, T.; Boiteux, S.; Zelwer, C.; Castaing, B. *J. Biol. Chem.* **2004**, *279*, 44074–44083.
- (13) Fromme, J. C.; Verdine, G. L. *J. Biol. Chem.* **2003**, *278*, 51543–51548.
- (14) Perlow-Poehnelt, R. A.; Zharkov, D. O.; Grollman, A. P.; Broyde, S. *Biochemistry* **2004**, *43*, 16092–16105.
- (15) Song, K.; Hornak, V.; de Los Santos, C.; Grollman, A. P.; Simmerling, C. *Biochemistry* **2006**, *45*, 10886–10894.
- (16) Song, K.; Kelso, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C. *J. Am. Chem. Soc.* **2007**, *129*, 14536–14537.
- (17) Sadeghian, K.; Flaig, D.; Blank, I. D.; Schneider, S.; Strasser, R.; Stathis, D.; Winnacker, M.; Carell, T.; Ochsenfeld, C. *Angew. Chem., Int. Ed.* **2014**, *53*, 10044–10048.
- (18) Boiteux, S.; Coste, F.; Castaing, B. *Free Radical Biol. Med.* **2017**, *107*, 179–201.
- (19) Blank, I. D.; Sadeghian, K.; Ochsenfeld, C. *Sci. Rep.* **2015**, *5*, 10369.
- (20) Sadeghian, K.; Ochsenfeld, C. *J. Am. Chem. Soc.* **2015**, *137*, 9824–9831.
- (21) Zharkov, D. O.; Rieger, R. A.; Iden, C. R.; Grollman, A. P. *J. Biol. Chem.* **1997**, *272*, 5335–5341.
- (22) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (23) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; De- Bolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- (24) 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.
- (25) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M. *J. Chem. Theory Comput.* **2010**, *6*, 2935–2947.
- (26) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. *J. Phys. Chem. A* **2009**, *113*, 11856–11865.
- (27) ChemShell, a Computational Chemistry Shell. www.chemshell.org (accessed 2017).
- (28) Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2013**, *138*, 134114.
- (29) Kussmann, J.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2015**, *11*, 918–922.
- (30) Jónsson, H.; Miles, G.; Jacobsen, K. W. *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific, 1998; pp 385–404.
- (31) Cooper, A. M.; Kästner, J. *ChemPhysChem* **2014**, *15*, 3264–3269.
- (32) Minetti, C. A. S. A.; Remeta, D. P.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 70–75.
- (33) Sumowski, C. V.; Ochsenfeld, C. *J. Phys. Chem. A* **2009**, *113*, 11734–11741.
- (34) Hu, L.; Söderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.* **2011**, *7*, 761–777.
- (35) Liao, R. Z.; Thiel, W. *J. Chem. Theory Comput.* **2012**, *8*, 3793–3803.
- (36) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. *J. Phys. Chem. B* **2016**, *120*, 11381–11394.
- (37) Roßbach, S.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2017**, *13*, 1699–1705.
- (38) Qi, Y.; Spong, M. C.; Nam, K.; Banerjee, A.; Jiralerspong, S.; Karplus, M.; Verdine, G. L. *Nature* **2009**, *462*, 762–766.
- (39) Li, H.; Endutkin, A. V.; Bergonzo, C.; Campbell, A. J.; de Los Santos, C.; Grollman, A.; Zharkov, D. O.; Simmerling, C. *Nucleic Acids Res.* **2016**, *44*, 683–94.
- (40) Blainey, P. C.; van Oijen, A. M.; Banerjee, A.; Verdine, G. L.; Xie, X. S. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5752–5757.
- (41) Jayanth, N.; Puranik, M. *J. Phys. Chem. B* **2017**, *121*, 5679–5687.

Base-independent DNA base-excision repair of 8-Oxoguanine

Supporting Information

Andrea Kreppel, Iris D. Blank, Christian Ochsenfeld

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

March 8, 2018

Contents

1	MD simulation	S2
1.1	MD simulation protocol	S2
1.2	FF parameters used for the non-standart residue 8OG	S4
1.3	Binding modes of the base during MD simulation	S6
2	QM/MM energy profile of the base excision of 8OG by Fpg	S8
3	QM/MM energy profile of step 1 using the BP86 DFT functional	S9
4	QM size convergence of the QM/MM energy profile of step 1	S10

1 MD simulation

1.1 MD simulation protocol

For the MD preparation and production runs the AMBER 11 FF [1] was used with the simulation engine NAMD v2.7 [2].

The systems were simulated in three different runs for 60 ns each. Bonds to hydrogens are constrained by the SETTLE algorithm [3] allowing for a timestep of 2 fs. The temperature was held constant at 300 K using Langevin dynamics and the pressure was held at 1.013 atm by the Nosé -Hoover Langevin piston pressure control [4, 5]. The electrostatic interactions are evaluated using the particle mesh Ewald summation method. The van der Waals interactions were cut off by means of a switching function in the interval from 10 to 13.5 Å. Full electrostatic interactions were evaluated every 4 fs. Coordinates were saved every 2 ps.

The systems were equilibrated before performing the production runs according to the following protocol: In the first preparation step the water molecules were relaxed with 10000 steps of conjugate gradient minimization while the protein, DNA and the Zn^{2+} -ion were kept frozen. Second, it was optimized again with 20000 steps of conjugate gradient minimization while the protein, DNA and Zn^{2+} were kept under restraints of $1 \text{ kcal } \text{Å}^{-2}$.

In the next step the system was heated to 300 K by increasing the temperature for 1 K every 100 time steps under the same restraints as in the second minimization. The heated system was equilibrated in two steps while bonds to hydrogens are constrained by the SETTLE algorithm [3]: 1.) for 200 ps the temperature was rescaled to 300 K every 1 ps; 2.) for 100 ps the temperature was controlled by Langevin dynamics and it was switched to a NPT ensemble with the Nosé -Hoover Langevin piston pressure control [4, 5]. Next the restraints were reduced stepwise (see Table S1) until the system is fully unrestrained after 400 ps of equilibration within the NPT ensemble.

Table S1: Restraints used for the protein and DNA during equilibration in NPT ensemble.

restraint [kcal mol ⁻¹ Å ⁻²]	simulation time
1.0	100 ps
0.8	30 ps
0.6	30 ps
0.5	30 ps
0.4	30 ps
0.3	20 ps
0.2	20 ps
0.1	20 ps
0.0	120 ps

1.2 FF parameters used for the non-standart residue 8OG

Table S2: Used GAFF atom types and partial charges for the residue 8OG.

Atom name	GAFF atom type	xyz - coordinates			partial charges
O1P	o	-0.9045	-2.944	5.824	6.397
P	p5	1.1484	-3.157	4.303	6.338
O2P	o	-0.8395	-4.466	3.833	5.68
O5'	os	-0.5862	-1.96	3.796	5.199
C5'	c3	0.1834	-0.685	4.387	5.307
C4'	c3	0.1041	0.187	3.931	4.143
O4'	os	-0.4336	1.544	4.407	4.269
C3'	c3	0.1221	0.345	2.421	4.045
O3'	os	-0.2088	0.789	2.096	2.701
C2'	c3	-0.1464	1.518	2.137	4.965
C1'	c3	0.2595	2.422	3.349	4.741
N9	n	-0.366	3.171	3.797	5.888
C4	cc	0.1646	4.184	4.725	5.759
N3	nc	-0.3	4.569	5.281	4.606
C2	cd	0.4092	5.549	6.17	4.721
N2	nh	-0.8458	6.001	6.814	3.623
N1	n	-0.4375	6.075	6.512	5.943
C6	c	0.6265	5.671	5.994	7.197
O6	o	-0.6715	6.163	6.451	8.243
C5	cd	-0.2174	4.709	4.967	7.025
N7	n	-0.1581	3.995	4.173	7.928
O	o	-0.4605	2.352	2.532	7.716
C	c	0.5815	3.073	3.392	7.255
H	hn	0.3375	6.827	7.211	5.976
H1	hn	0.4628	5.714	6.361	2.758
H2	hn	0.3165	4.311	3.908	8.867
H3	hn	0.4198	7	7.05	3.533
H4	h2	0.0607	3.156	3.157	3.947
H5	hc	0.0537	2.042	1.207	4.718
H6	hc	0.0757	1.186	2.098	6.009
H7	h1	0.0967	-0.577	1.907	4.347
H8	h1	0.0857	-0.211	4.329	3.2
H9	h1	0.0297	-0.165	4.106	6.245
H10	h1	0.0377	-0.746	5.488	5.303

Table S3: Bond Parameters added for the 8OG residue.

Bond	Force Constant	Ideal Length	Analogy With
OS-p5	230	1.61	OS-P
os-P	230	1.61	OS-P

Table S4: Bond Angle Parameters added for the 8OG residue.

Bond Angle	Force Constant	Ideal Angle	Analogy With
OS-p5-os	45	102.6	OS-P -OS
os-P-OS	45	102.6	OS-P -OS
OS-p5-o	100	108.23	O2-P -OS
os-P-O2	100	108.23	O2-P -OS
CT-OS-p5	100	120.5	CT-OS-P
c3-os-P	100	120.5	CT-OS-P

Table S5: Torsion Parameters added for the 8OG residue.

Torsion	Phase	Force Constant	Phase Angle	Periodicity	Analogy With
CT-OS-p5-os	1	0.25	0	-3	OS-P -OS-CT
CT-OS-p5-os	1	1.2	0	2	OS-P -OS-CT
c3-os-P-OS	1	0.25	0	-3	OS-P -OS-CT
c3-os-P-OS	1	1.2	0	2	OS-P -OS-CT
CT-OS-p5-o	1	0.25	0	-3	OS-P -OS-CT
CT-OS-p5-o	1	1.2	0	2	OS-P -OS-CT
c3-os-P-O2	1	0.25	0	-3	OS-P -OS-CT
c3-os-P-O2	1	1.2	0	2	OS-P -OS-CT

Table S6: Improper Torsion Parameters added for the 8OG residue.

Improper Torsion	Force Constant	Phase Angle	Periodicity
c-c3-n-cc	1.1	180	2
cd-n-cc-nc	1.1	180	2
n-nc-cd-nh	1.1	180	2
cd-hn-nh-hn	1.1	180	2
c-cd-n-hn	1.1	180	2
cd-n-c-o	10.5	180	2
c-cc-cd-n	1.1	180	2
n-n-c-o	10.5	180	2

1.3 Binding modes of the base during MD simulation

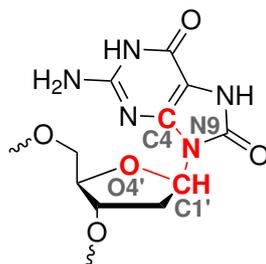


Figure S1: The dihedral of the glycosidic bond is defined as the dihedral between the four atoms: O4'-C1'-N9-C4 (marked in red).

To be able to evaluate if the binding mode of *syn*-8OG and *anti*-8OG within the active site of Fpg during the MD simulation stays the same, the dihedral of the glycosidic bond (O4'-C1'-N9-C4, see Figure S1) was analyzed.

Song *et al.* [6] calculated a free energy profile by umbrella sampling for the rotation around the 8OG glycosidic bond in the Fpg active site. They found a minimum in the energy profile at about -70° that corresponds to the 8OG *anti*-conformation and another minimum at about 55° for the *syn*-conformation. In the simulation of *anti*-bound 8OG the initial dihedral of the glycosidic bond was -64° while in the simulation of *syn*-bound 8OG the initial dihedral of the glycosidic bond was set to 140° .

The dihedral in the *anti*-8OG simulation stays between -180° and -50° in all three runs (Figure S2). Most time the dihedral stays around -150° to -100° . This indicates that 8OG remains within the region of the *anti*-conformation. In the *syn*-8OG simulation the dihedral stays between 50° and 150° in all three runs (Figure S2) which indicates the *syn*-conformation of 8OG. This is in line with the findings of Perlow-Poehnelt *et al.* [7]. They simulated 8OG in both conformations, *syn* and *anti*, and found that the dihedral of *anti*-8OG stays stably in between -140° and -90° and the dihedral of *syn*-8OG stays in between 100° and 160° . Our MD simulation demonstrates that 8OG binds stably in both conformations into the Fpg active site and does not switch between the two different binding modes within the simulated time.

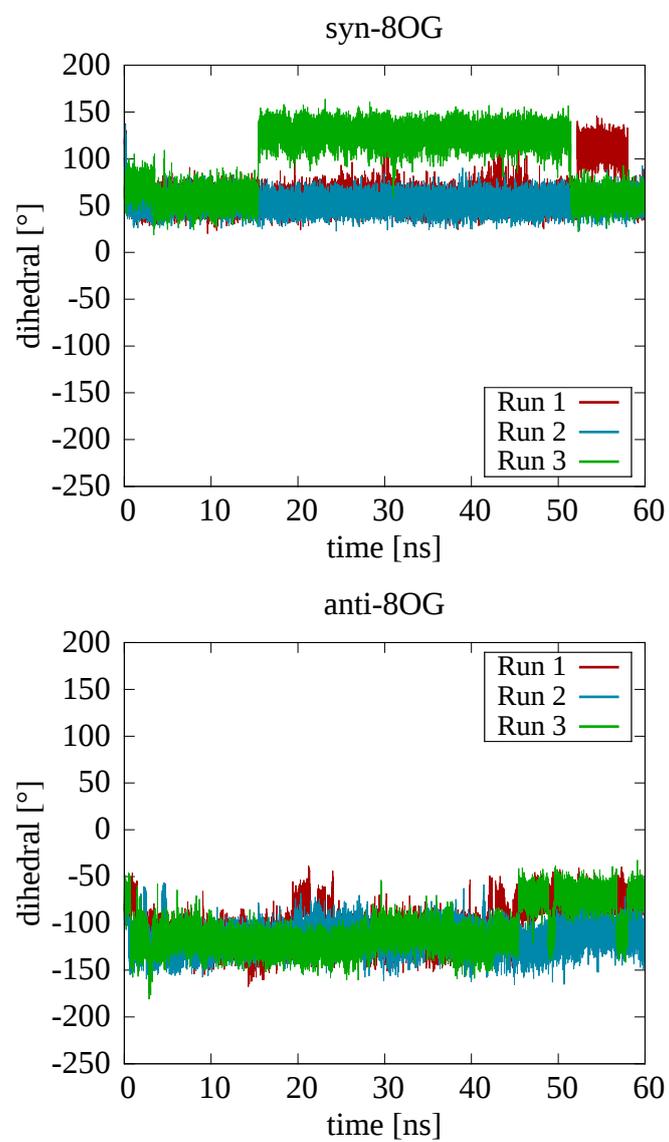


Figure S2: Dihedral of the glycosidic bond (O4'-C1'-N9-C4) during the MD production run 1, 2, and 3 of the simulation of *syn*-8OG (top) and *anti*-8OG (bottom).

2 QM/MM energy profile of the base excision of 8OG by Fpg

The energy profile of the base excision was calculated using an "adiabatic mapping approach". In this approach, constrained optimizations were performed along the reaction coordinate. This was done adding harmonic restraints of the form $V = K/2(x - x_0)^2$ to the QM/MM optimization with the optimizer tool DL-FIND where x_0 , that was defined as a linear combination of bond distance or an torsion angle, was varied gradually for 0.1 Å per restrained bond or 10° per restrained dihedral along the reaction coordinate. The force constant K was chosen to be 4 Hartree/Bohr² per restrained bond or 4 Hartree/rad² dihedral.

For the ribose-opening in reaction step 1, the reaction coordinate (rc1) was defined as distance $d(O4',C1')-d(E2:O,O4')-d(C1',P1:N)$ (Figure S3). The reaction coordinate (rc2) for the reorientation of the glutamate side chain in step 2a was defined as dihedral $E2:C1-E2:C2-E2:C3-E2:CA$. The proton transfer in step 2b was described by the distance $d(P1:H,P1:N)-d(P1:H,E2:O)$ as reaction coordinate (rc3). In the last adiabatic mapping path, the reaction coordinate (rc4) was defined as distance $d(N9,C1')-d(O4':H,N9)+d(O4',O4'H)-d(E2:H,O4')$. This path comprises two energy barriers where within the first one, the residues in the active site reorientate and within the second one, the actual cleavage of the glycosidic bond via protonation of the base occurs.

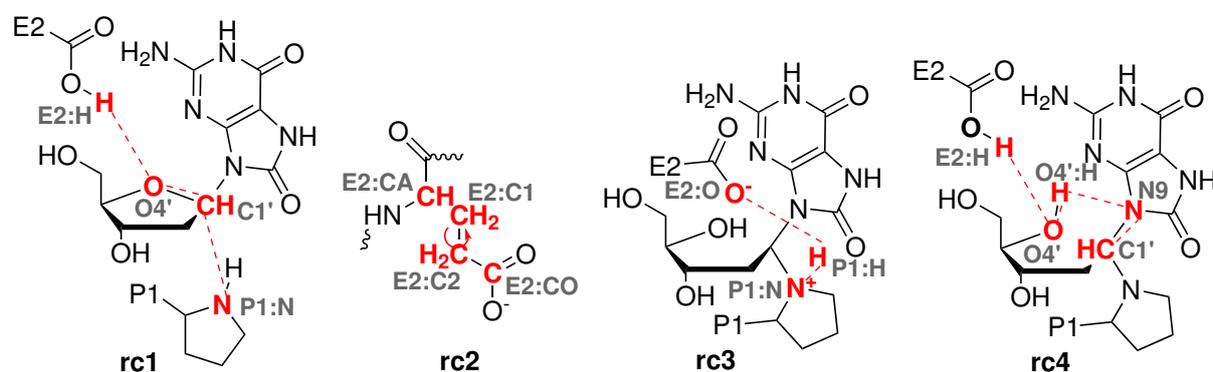


Figure S3: Bond distances or dihedral angle used for the description of the base excision mechanism plotted in red.

The detailed energy profile obtained by adiabatic mapping is plotted in Figure S4. Here the energy profile of step 3 of the *syn*-reaction path and step 2a, 2c and 3 of the *anti*-reaction path was refined by NEB using the adiabatic path as initial guess.

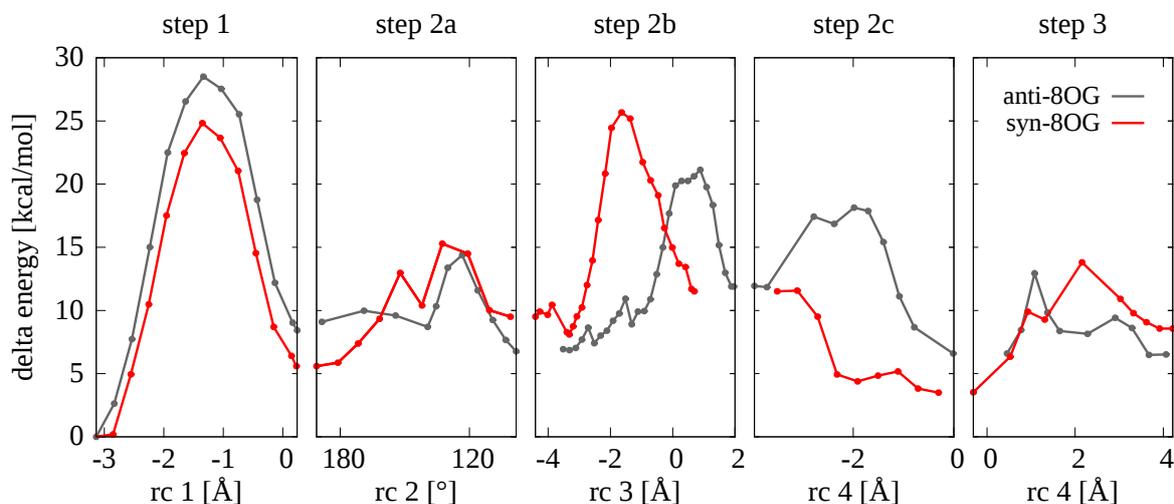


Figure S4: Comparison of the QM/MM energy profile of the base excision for 8OG bound in *syn*-conformation (grey) and *anti*-conformation (red) to Fpg using B3LYP-D3/def2-svp//AMBER11-FF for calculation.

3 QM/MM energy profile of step 1 using the BP86 DFT functional

The DFT-functional BP86 yields an energy barrier in the range of 20 to 23 kcal/mol with an Boltzmann average [8] of 21 kcal/mol for *syn*-bound 8OG for the first reaction step. For the *anti*-conformation an energy barrier in the range of 21 to 27 kcal/mol with an Boltzmann average of 22 kcal/mol is found (Figure S5). The BP86 energy barrier is in both cases lower than the B3LYP value, where an average energy barrier of 26 kcal/mol for *syn*-8OG and 29 kcal/mol for *anti*-8OG is obtained. This difference illustrates the impact of the chosen DFT functional. Nevertheless, comparing the energy barriers obtained using BP86 allows for the same conclusions. Here, the energy barriers of the two different conformations are even more similar than comparing the energy profiles calculated using B3LYP.

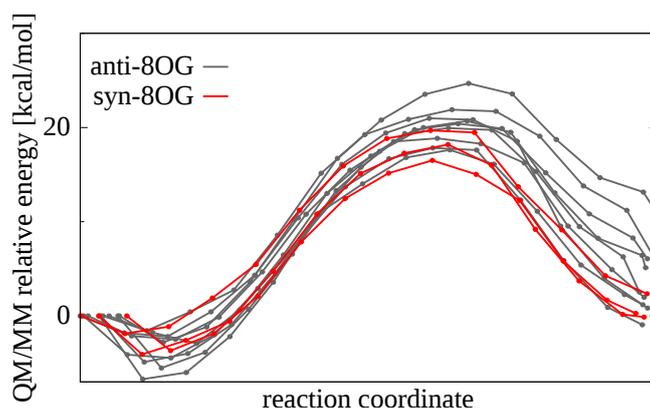


Figure S5: Energy profile for the first reaction step (step 1) of the base excision mechanism of *anti*-8OG (grey) and *syn*-8OG (red) by Fpg for a variety of different starting structures using BP86-D3/def2-svp//AMBER11-FF for calculation.

4 QM size convergence of the QM/MM energy profile of step 1

To estimate the error introduced by using a small QM region for calculating the QM/MM energy profile, the QM/MM energy profile of reaction step 1 was calculated using different QM regions (see Table S7). QM-region C is converged with respect to the QM size as it contains the corresponding residues to the QM size converged QM-region of our previous study on *B. st.* Fpg. For QM-region A and B the reaction profile was calculated using adiabatic mapping where all residues in a radius of 10 Å around the QM-region were relaxed. By reason of computational costs, not the whole reaction profile was calculated for QM-region C. Here, the educt, product and transition structure (geometry with highest energy along the path) of the respective path of QM-region B were optimized or optimized with constraints where only the atoms contained in QM-region A were allowed to relax.

Table S7: Residues included in each QM-region. The capping hydrogens are not included in the number of QM-atoms.

QM-region	#atoms (anti/syn)	Charge	Residues
QM A	80	-2	P1, E2, E5, 8OG
QM B	230	-2	P1, E2, E5, R74, M75, I172, I219, Y222, dT8, 8OG, and 6 water
QM C	588/573	0	P1, E2, L3, P4, E5, V6, K57, L73, R74, M75, G76, K78, G168, L169, G170, N171, I172, Y173, G215, G216, S217, S218, I219, R220, T221, Y222, R260, dT6, dT8, 8OG, and 17 or 12 water

References

- (1) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- (2) Phillips, J. C. et al. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (3) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (4) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (5) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (6) Song, K.; Kelso, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C. *J. Am. Chem. Soc* **2007**, *129*, 14536–14537.
- (7) Perlow-Poehnelt, R. A.; Zharkov, D. O.; Grollman, A. P.; Broyde, S. *Biochemistry* **2004**, *43*, 16092–16105.
- (8) Cooper, A. M.; Kästner, J. *ChemPhysChem* **2014**, *15*, 3264–3269.

Publication II: The Enzymatic Decarboxylation Mechanism of 5-carboxy Uracil - a Comprehensive Quantum Chemical Study

A. Kreppel, C. Ochsenfeld,

"The Enzymatic Decarboxylation Mechanism of 5-carboxy Uracil - a Comprehensive Quantum Chemical Study", *J. Chem. Theory Comput.*, accepted.

Abstract: Dynamic regulation of DNA methylation is an important process for the control of gene expression in mammals. It is believed that in the demethylation pathway of 5-methyl cytosine the intermediate 5-carboxy cytosine (5caC) can be actively decarboxylated alongside the substitution in the base-excision repair. For the active decarboxylation of 5caC a decarboxylase has not been identified so far. Due to the similar chemistry of the decarboxylation of 5-carboxy uracil (5caU) to uracil (U) in the pyrimidine salvage pathway catalyzed by the iso-orotate decarboxylase (IDCase) the study of this reaction might give valuable insights for the active 5caC decarboxylation process. In this work, we employ QM/MM calculations and find that the catalytic mechanism of IDCase proceeds via a direct decarboxylation mechanism. Detailed investigations on the reaction coordinate reveal that it is a one-step mechanism with concerted proton-transfer and C-C bond opening.

Reprinted with permission from:

A. Kreppel, C. Ochsenfeld,

"The Enzymatic Decarboxylation Mechanism of 5-carboxy Uracil - a Comprehensive Quantum Chemical Study"
J. Chem. Theory Comput., accepted.

Copyright 2020 American Chemical Society.

<https://pubs.acs.org/doi/pdf/10.1021/acs.jctc.0c00616>

The Enzymatic Decarboxylation Mechanism of 5-Carboxy Uracil: A Comprehensive Quantum Chemical Study

Andrea Kreppel and Christian Ochsenfeld*



Cite This: <https://dx.doi.org/10.1021/acs.jctc.0c00616>



Read Online

ACCESS |



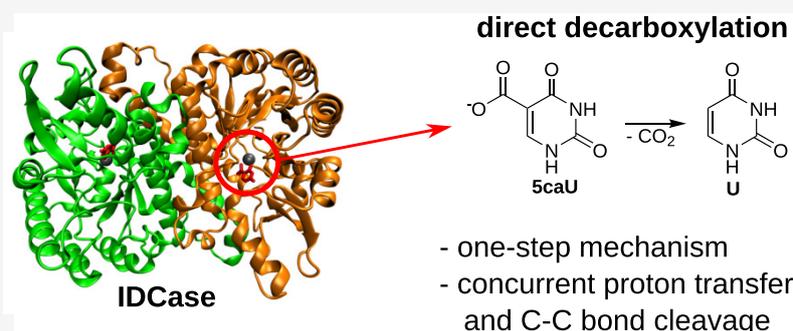
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: Dynamic regulation of DNA methylation is an important process for the control of gene expression in mammals. It is believed that in the demethylation pathway of 5-methyl cytosine, the intermediate 5-carboxy cytosine (5caC) can be actively decarboxylated alongside the substitution in the base excision repair. For the active decarboxylation of 5caC, a decarboxylase has not been identified so far. Due to the similar chemistry of the decarboxylation of 5-carboxy uracil (5caU) to uracil (U) in the pyrimidine salvage pathway catalyzed by the iso-ortate decarboxylase (IDCase), the study of this reaction might give valuable insights into the active 5caC decarboxylation process. In this work, we employ quantum chemical and molecular mechanic calculations and find that the catalytic mechanism of IDCcase proceeds via a direct decarboxylation mechanism. Detailed investigations on the reaction coordinate reveal that it is a one-step mechanism with concerted proton transfer and C–C bond opening.

1. INTRODUCTION

Epigenetic modifications regulate patterns of gene expression by the alteration of DNA accessibility and the chromatin structure.¹ One of the best-characterized epigenetic modifications is the methylation of the fifth position of cytosine (C) (see Figure 1) that is found in most plant, animal, and fungal models.² Although the DNA methylation pattern in mammalian somatic cells is stably maintained, there is dynamic regulation of DNA methylation involving DNA demethylation at specific stages of development.^{3–6} A proposed mechanism for an active DNA demethylation involves the iterative oxidation of 5mC to 5caC by the Ten-eleven Translocation (TET) dioxygenases^{7–11} and its subsequent excision by the DNA-glycosylase TDG in the base excision repair (BER) pathway.^{9,12}

However, there is evidence of an alternative active demethylation pathway downstream of the TET-mediated oxidation.^{13–16} Studies have suggested that this pathway involves the enzymatic decarboxylation of 5caC to C by a yet undetected putative decarboxylase. Although the demethylation of 5caC was in the focus of several studies,^{14–16} the exact mechanism of this reaction remains unclear.

Due to the similarity of the TET-mediated demethylation to the thymine(T)-to-uracil(U) conversion in the pyrimidine

salvage pathway, the latter might provide insights into the decarboxylation of 5caC in the active demethylation pathway. In the pyrimidine salvage pathway, the substrate T is also successively oxidized by a dioxygenase to 5-carboxyuracil (5caU) (Figure 2), which in turn is decarboxylated by the enzyme iso-ortate decarboxylase (IDCase) resulting in U.¹⁸

For the decarboxylation mechanism of 5caU and the related substrates orotidine and 5caC, various mechanisms have been proposed.^{15,19–30} Among these, Xu *et al.*¹⁹ suggested that the decarboxylation of 5caU takes place via a tetrahedral intermediate formed by nucleophile addition at the carboxyl carbon of 5caU (see Scheme 1A). Another decarboxylation mechanism, which was proposed for orotidine (Scheme 1B(a)) and 5caC (Scheme 1B(b)), involves the activation of the substrate by nucleophilic addition to an aromatic carbon in close proximity to the carboxyl group.^{15,28,30} For the catalytic

Received: June 16, 2020

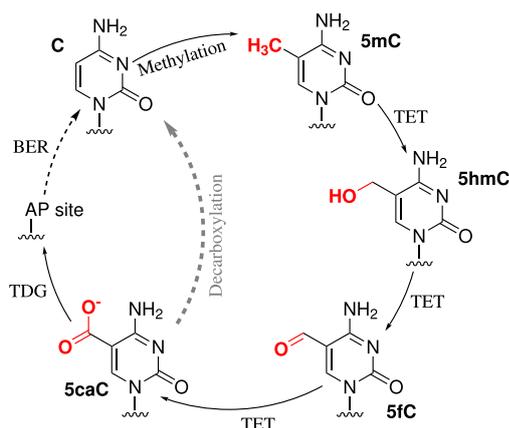


Figure 1. Active DNA demethylation. 5-Methylcytosine (**5mC**) is successively oxidized to 5-carboxycytosine (**5caC**) via 5-hydroxymethylcytosine (**5hmC**) and 5-formylcytosine (**5fC**). **5caC** can then be replaced by an unmodified **C** in the base excision repair (BER) pathway or actively decarboxylated by a yet undetected decarboxylase in an alternative pathway. The wavy line connected to N1 of the base indicates the connection to the desoxyribose.

activity of the enzyme orotidine-5'-monophosphate decarboxylase (ODCase) on the substrate **orotidine**, also a direct decarboxylation mechanism has been proposed^{20–22,26,31} (Scheme 1C(a)). A slightly different direct decarboxylation mechanism has been suggested for the deprotonated tautomeric form of **5caU** (**taut-5caU**)²⁹ (Scheme 1C(b)).

In this work, we compare these three decarboxylation mechanisms for the decarboxylation of **5caU** in the enzymatic environment of IDCase using combined quantum chemical and molecular mechanic (QM/MM) calculations. A detailed study of this enzyme and its reaction mechanism could provide also insights into the active decarboxylation of **5caC**.³²

2. COMPUTATIONAL DETAILS

The system was prepared based on chains A and C of the crystal structure (PDB code: 4HK6) of *Cordyceps militaris* IDCase in complex with the inhibitor 5-nitrouracil (**5niU**). The nitro group of **5niU** was altered manually into a carboxy group. In this work, we consider the diketo form of **5caU** as a substrate for IDCase since the diketo form is the most stable tautomer of uracil.^{33–42} Even though it has been questioned in literature whether the metal ion in the active center really is a zinc ion,²⁹ we still retain the Zn^{2+} ion in the active site like originally proposed in the crystal structure since the metal ion seems to have rather a structuring function and is not actively involved in the reaction. The work by Sheng *et al.*²⁹ supports this decision since they showed that the type of metal has no major effect on the energetics of the reaction.

The crystal water was removed except for the water molecule at the conserved water position in the active site (Wat3, see Figure 3). All residues were protonated according to standard protonation states at pH 7. The system was

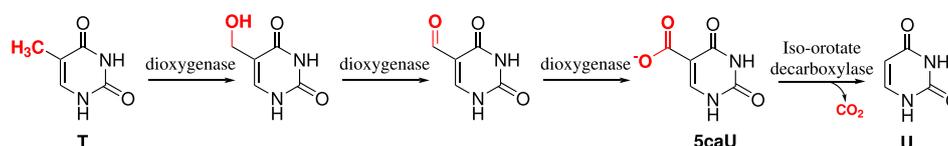
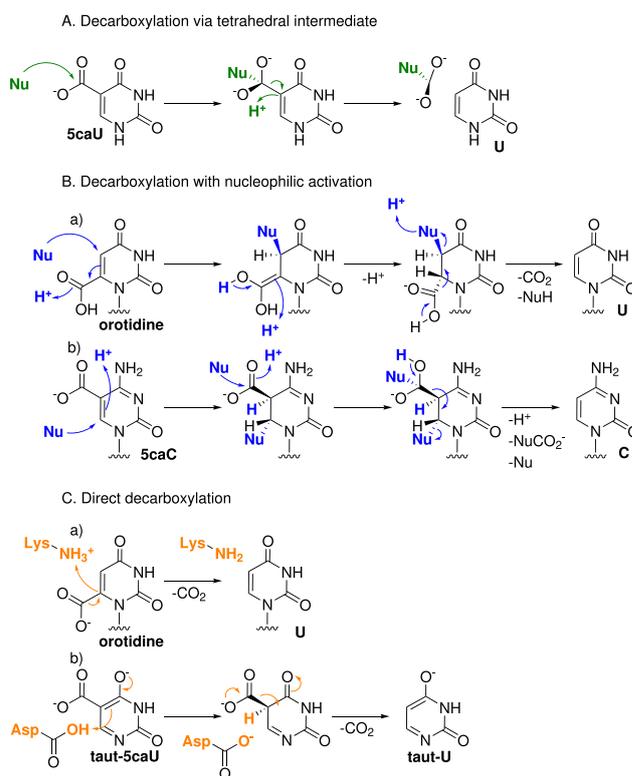


Figure 2. Thymine-to-uracil conversion in the pyrimidine salvage pathway as described in ref 17.

Scheme 1. Proposed Decarboxylation Mechanisms of **5caU** and the Related Substrates Orotidine and **5caC**^a



^a(A) Decarboxylation of **5caU** via a tetrahedral intermediate.¹⁹ (B) Decarboxylation of (a) **orotidine**²⁸ and (b) **5caC**^{15,30} initiated by the activation with nucleophile addition. (C) Direct decarboxylation of (a) **orotidine** catalyzed by ODCase^{20–22} and (b) **taut-5caU** by IDCase.²⁹ The wavy lines indicate the connection to ribose or desoxyribose.

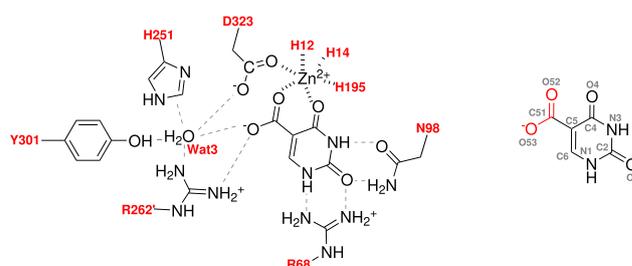


Figure 3. Active site of IDCase (left). Substrate **5caU** with atom names in gray (right). The carboxyl group that is removed by IDCase is marked in red.

neutralized with sodium ions and dissolved in an orthorhombic box with at least 10 Å of TIP3P⁴³ solvent around the solute, which resulted in a box with a volume of about 8.35×10^{-22} L.

To generate equilibrated starting structures for QM/MM calculations, the enzyme–substrate complexes were simulated

with classical molecular dynamics using the ff14SB force field contained in the AMBER16 program package.^{44,45} For the decarboxylation according to pathways B and C, the system was additionally relaxed by classical molecular dynamics simulation after the activation steps. The parameters for the Zn^{2+} metal center containing **ScaU** or the intermediate of pathway B were obtained using the AmberTools⁴⁵ MCPB.py and ANTECHAMBER and can be found in the Supporting Information. The simulations were performed with the simulation engine NAMD⁴⁶ (MD simulation protocol can be found in the Supporting Information, Section S1).

QM/MM structure optimizations employed the DL-Find optimizer implemented in ChemShell^{47–49} with a convergence criterion for the maximum energy change of 5.0000×10^{-4} Hartree, maximum gradient component of 2.0000×10^{-3} Hartree/Bohr, maximum step component of 8.0000×10^{-3} Bohr, root-mean-square (RMS) gradient of 1.3333×10^{-3} Hartree/Bohr, and RMS step of 5.3333×10^{-3} Bohr. The QM part of the calculations was performed with the program package FermiONS++^{50–52} using the B3LYP functional⁵³ with Grimme D3 dispersion correction^{54,55} and the def2-SVP⁵⁶ basis set, if not stated otherwise. All minima along the reaction pathways showed no imaginary modes in numerical frequency calculations for the QM atoms. The frequency calculations were performed with the same QM/MM setup as the structure optimizations. The highest point along each pathway was optimized as a transition state using the dimer method^{57,58} implemented in ChemShell.

In the QM calculations, the substrate **ScaU**, aspartate D323, Zn^{2+} , and the ligating histidines (H12, H14, and H195) are included and, for pathways A and B, also a water molecule is included. In the first step of pathway C (initial protonation of D323), also protonated H251 was included in the QM sphere as the proton-donating residue for the reaction.

In the structure optimizations at QM/MM level, the atoms in a radius of 10 Å around the QM residues were relaxed, and the remaining atoms were kept fixed.

3. RESULTS AND DISCUSSION

3.1. Energy Profiles of the Possible Decarboxylation Pathways of ScaU by IDCase. The enzyme IDCase is a homodimer where each monomer has an active site consisting of one Zn^{2+} ion that is coordinated by three histidines (H12, H14, and H195), one aspartate (D323), and the substrate 5-carboxyuracil (**ScaU**), which is illustrated in Figure 3. **ScaU** is further stabilized in the active site by the interactions of an asparagine of the second monomer (R262'). Close to **ScaU**, a conserved water molecule (Wat3) was found in crystal structures,¹⁹ which forms hydrogen bonds to D323, another histidine (H251), a tyrosine (Y301), and R262'. In the MD simulation, we found further water molecules in the active site. There are two water molecules forming hydrogen bonds to the carboxyl group of **ScaU** and D323 (Figure 4). These water positions can also be seen in the radial distribution function of the water molecules around the carboxyl oxygen of **ScaU**.

In the following, it is discussed how the three decarboxylation mechanisms displayed in Scheme 1 can be applied to the enzymatic decarboxylation of **ScaU** catalyzed by IDCase. The different mechanisms are compared based on the obtained QM/MM energy profiles.

3.1.1. Decarboxylation via a Tetrahedral Intermediate. For the decarboxylation mechanism of **ScaU** by IDCase via a tetrahedral intermediate, as proposed by Xu *et al.*¹⁹ (see

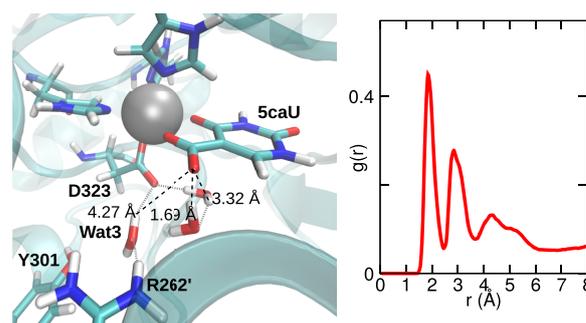


Figure 4. Common situation of water molecules in the active site of IDCase observed in MD simulation (left). Distances of the water molecules to the carboxyl oxygen of **ScaU** are given at the black dashed lines, and further hydrogen bonds are indicated with gray dashed lines. The position of these water molecules can also be seen in the radial distribution function of the water molecules around the carboxyl oxygen of **ScaU** (right).

Scheme 1A), a water molecule was considered to act as the nucleophile in this work (see Scheme 2). In the initial reaction step, the water molecule performs a nucleophilic attack at the carboxyl carbon C51 (for atom naming scheme, see Figure 3, right) while it simultaneously transfers a hydrogen to the non-ligating oxygen O53 of the carboxyl group. The QM/MM energy profile (see Figure 5, green) shows that this reaction step has a rather high energy barrier with about 47 kcal/mol and results in an unstable tetrahedral intermediate (**A1**) that is about 29 kcal/mol higher in energy than the reactant complex (**ScaU**). In the following reaction step, the C–C bond is cleaved and a proton is transferred to the C5 position of the substrate. The barrier for this step is about 23 kcal/mol, which means that this reaction step accommodates the highest point in energy along the reaction path with a relative energy of 52 kcal/mol to the reactant complex (**ScaU**). The final product complex (**U**) lies about 4 kcal/mol higher in energy than **ScaU**.

Although suggested by Xu *et al.*,¹⁹ we consider D323 or an OH^- unlikely to be the nucleophile in this reaction for the following reasons: Aspartate D323 seems unlikely to be the nucleophile as we could not obtain any stable tetrahedral intermediate with QM/MM calculations (see Supporting Information, Section S3.1). It also appears to be questionable that OH^- functions as a nucleophile in this reaction: We could not obtain a stable deprotonation product of H_2O with D323, H251, or **ScaU** as a base in QM/MM calculations (Supporting Information, Section S3.2). Furthermore, OH^- is not likely to exist for a long time in the active site as it will deprotonate any nearby residue (tyrosine Y301 or the positively charged arginine R262'), forming H_2O .

3.1.2. Decarboxylation with Nucleophilic Activation. In this pathway, an activation of **ScaU** precedes the decarboxylation in the same way as proposed for the decarboxylation of orotidine or **ScaC** (see Scheme 1B). **ScaU** is activated by the addition of a nucleophile to the C6 position. Since in the active site of IDCase no nucleophilic residue is located close to the C6 position of **ScaU**, a water molecule is considered to act as the nucleophile (Scheme 3(1)). During the nucleophilic attack, a hydrogen is transferred to D323, which is shifted in a second reaction step to the C5 position of **ScaU**. This activation has relatively low energy barriers with about 13 and 9 kcal/mol, respectively (see Figure 5, blue). The overall reaction energy of the activation step is slightly positive with the intermediate **B2** of about 5 kcal/mol higher in energy than the reactant complex

Scheme 2. Mechanism for the Decarboxylation of 5caU by IDCase via a Tetrahedral Intermediate

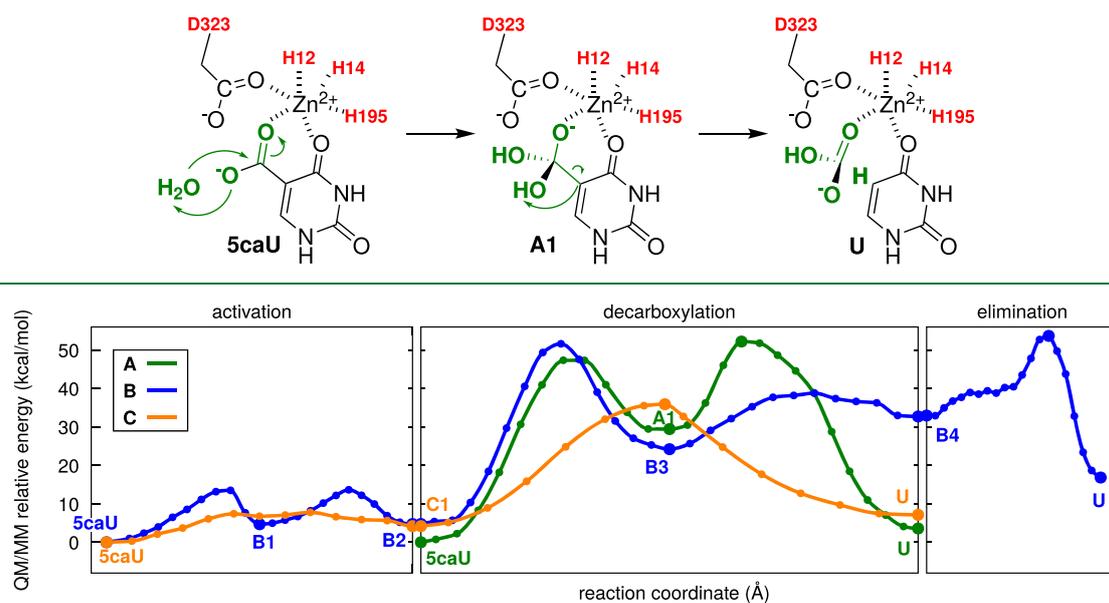


Figure 5. QM/MM energy profiles of the decarboxylation of 5caU by IDCase according to pathways A (green), B (blue), and C (orange) (for the corresponding mechanisms, see Schemes 2–4). For the highest point along each pathway, a transition state optimization was performed. The definition of the reaction coordinates, selection criteria for the starting geometries, and three-dimensional structures of reactant complexes, intermediates, and product complexes can be found in the Supporting Information, Section S2.

(5caU). The fact that all reactants are trapped by complexation with Zn^{2+} in perfect position for the backward reaction, which has even lower energy barriers than the forward reaction, indicates that this activation is clearly a reversible process.

On the basis of the decarboxylation mechanism proposed for 5caC (Scheme 1B(b)), the activated substrate 5caU (B2) is then attacked by another nucleophile in the C51 position (Scheme 3(2)). Here, again a water molecule acts as a nucleophile and transfers a hydrogen to O53 of the activated substrate in the same way as in the decarboxylation pathway A. For this reason, an MD simulation was performed at intermediate B2 to let water molecules re-enter the active site (for further information, see the Supporting Information, Section S4), which can then perform the nucleophilic attack. The QM/MM energy profile of this reaction step reveals the same high energy barrier as in pathway A (about 47 kcal/mol). At the resulting intermediate B3, which is about 19 kcal/mol higher in energy than the activated substrate B2, the actual C–C bond cleavage takes place with an energy barrier of about 14 kcal/mol, forming an unstable anionic intermediate (B4, about 28 kcal/mol above B2). During the release of H_2CO_3 , a proton is transferred to D323, resulting in a protonated D323 and HCO_3^{3-} .

In a third part of the reaction pathway (Scheme 3(3)), the final product U is obtained by elimination of the added nucleophile. The elimination is facilitated by protonation of the leaving nucleophile by the protonated D323 and has an energy barrier of about 21 kcal/mol. This reaction step also contains the highest point along the entire pathway, which results in an overall reaction barrier for pathway B of about 54 kcal/mol. The overall reaction energy of this pathway is positive with about +17 kcal/mol.

Based on the proposed mechanism for orotidine (Scheme 1B, mechanism a), it would also be thinkable that the decarboxylation and elimination take place in a single

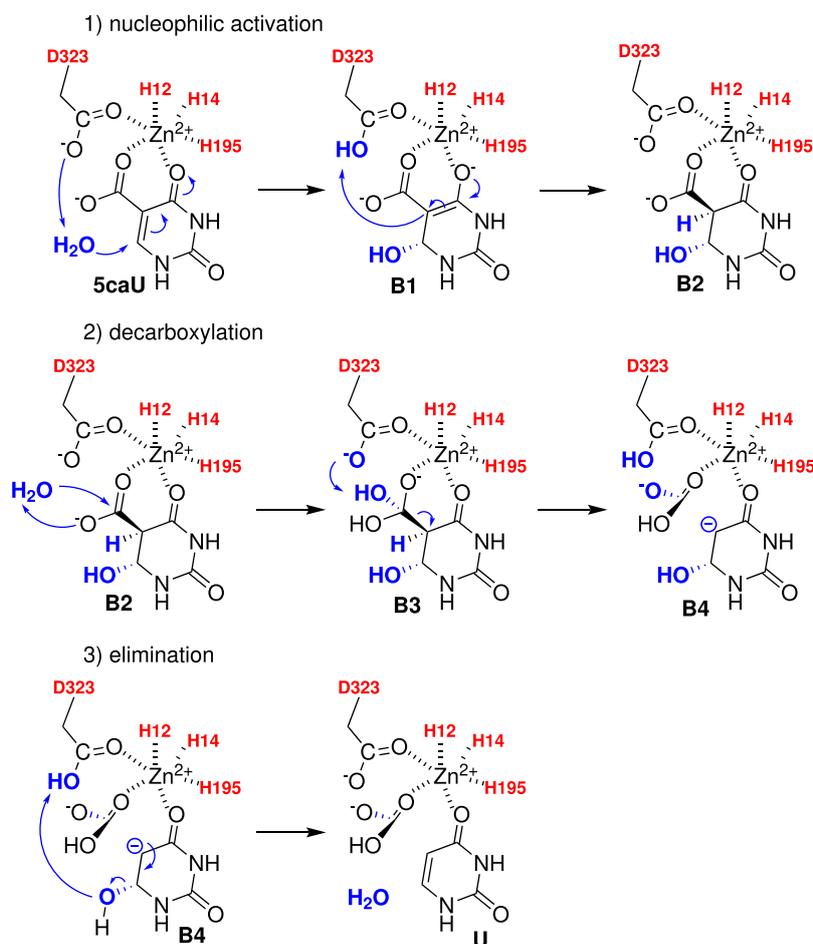
concerted reaction step. We also investigated this possibility with QM/MM calculations (see the Supporting Information, Section S5). As we have obtained for this alternative pathway a higher energy barrier (about 50 kcal/mol; 55 kcal/mol relative to 5caU), we focus on the above-described mechanism in this work.

3.1.3. Direct Decarboxylation. During the direct decarboxylation of orotidine (Scheme 1C), a proton is transferred to the substrate from a nearby residue. In the enzyme IDCase, the aspartate D323 in close proximity to the substrate is in ideal position for this proton transfer (Scheme 4) during which the C–C bond opens and CO_2 is released. The QM/MM profile (see Figure 5, decarboxylation step, orange) gives an energy barrier of about 31 kcal/mol and a reaction energy of about +3 kcal/mol for the direct decarboxylation of 5caU. Repeated scanning of the energy surface by iteratively calculating the forward and reverse reaction path gives almost identical energy profiles (Supporting Information, Section S6). Recalculation of the energy profile with an increased QM region (Supporting Information, Section S7) gives a similar energy barrier (about 28 kcal/mol) and only a slightly lower reaction energy of about –2 kcal/mol. These results support the initial calculation of the profile.

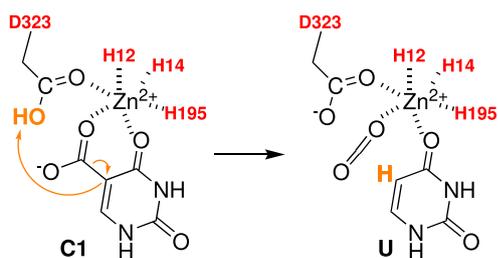
As 5niU is an inhibitor of IDCase, it should not be converted by the enzyme or only with a poor reaction rate. This means that the energy barrier of the reaction with 5niU (Scheme 5) should be much higher than that of the reaction of the actual substrate 5caU. Comparing the QM/MM energy profile of the decarboxylation step according to pathway C starting from the same initial structure for the substrate 5caU and the inhibitor 5niU (manually altered) shows that the energy barrier for 5niU is with about 70 kcal/mol considerably higher than that for 5caU with about 31 kcal/mol (Figure 6).

Since D323 is not protonated in pathways A and B, an estimate of the protonation costs must be added in pathway C

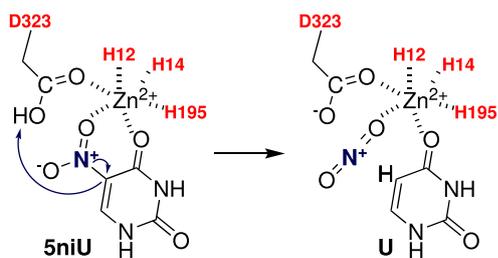
Scheme 3. Mechanism for the Decarboxylation of 5caU by IDCase with a Nucleophilic Activation (1) before the Decarboxylation Part (2) and Subsequent Nucleophile Elimination (3)



Scheme 4. Direct Decarboxylation Mechanism of 5caU Catalyzed by IDCase



Scheme 5. Hypothetical Conversion of the Inhibitor 5niU by IDCase according to Pathway C



in order to be comparable to pathways A and B. To this end, we consider the protonation of D323 by the neighboring H251

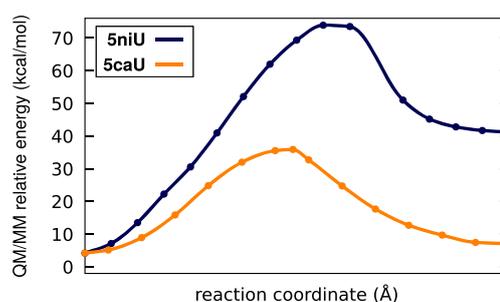
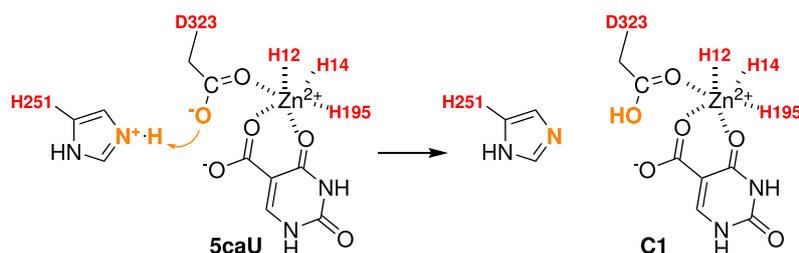


Figure 6. QM/MM energy profiles of the reaction of substrate 5caU (orange) and the inhibitor 5niU (dark blue) according to pathway C.

in protonated form (Scheme 6). The QM/MM energy profile (Figure 5, activation step, orange) for this proton transfer has a rather low energy barrier (about 8 kcal/mol) and indicates a protonation cost of about 4 kcal/mol. With this protonation cost, the overall barrier for decarboxylation according to pathway C including the protonation of D323 is about 35 kcal/mol.

3.2. Comparison of the Different Decarboxylation Pathways. Although the substrate is activated prior to the actual decarboxylation in pathway B, the subsequent nucleophilic attack in the decarboxylation part of the mechanism has a relatively high energy barrier with about 47 kcal/mol (Figure 5). The subsequent elimination of the

Scheme 6. Protonation of D323 by the Neighboring Histidine H251 in Preparation for the Decarboxylation Step in Pathway C



activator nucleophile is also not energetically favorable and accommodates the highest point in energy along the entire pathway B with about 54 kcal/mol relative to the reactant complex **5caU**. The highest point in energy along pathway A has a similar height with about 52 kcal/mol above the reactant complex **5caU**, whereas the overall reaction barrier of the direct decarboxylation including the preceding proton transfer (pathway C) is considerably lower with about 35 kcal/mol, which makes this pathway energetically more favorable. This result is almost independent on the size of the QM region or the QM method used in the calculation (see the Supporting Information, Sections S7 and S8).

For this reason, it is highly unlikely for the decarboxylation reaction to proceed according to pathways A or B. Thus, the activation process in pathway B seems to be only a reversible side reaction.

Since pathway C has a significantly lower energy barrier than pathways A and B, the direct decarboxylation pathway is most likely the catalytic mechanism of IDC_{ase}. The fact that the inhibitor **5niU** has a considerably higher energy barrier for the conversion according to pathway C than the substrate **5caU** clearly underpins this assumption. Furthermore, this finding is in line with the proposed mechanism of the tautomeric form of the substrate.²⁹

3.3. Function of the IDC_{ase} Active Site Residues in the Decarboxylation Reaction. In the direct decarboxylation mechanism of IDC_{ase}, D323 is essential for the proton transfer to C5 of **5caU**. This is in line with the mutation assay of Xu *et al.*,¹⁹ where D323 was found to be essential for the activity of IDC_{ase}. For the decarboxylation according to pathway C, D323 must be present in protonated form. In this work, we have shown that H251 can transfer a proton to D323. However, Xu *et al.*¹⁹ only observed a reduced activity of IDC_{ase} and not a complete depletion when mutating H251 to alanine. This leads to the assumption that there are alternative proton transfer pathways. It is possible that the residues Y301 and R262', for which reduced or no activity of IDC_{ase} was determined upon mutation, could transfer a proton to D323 via the water molecule Wat3. This remains to be investigated in future studies.

3.4. Reaction Coordinate of the Direct Decarboxylation Pathway of 5caU by IDC_{ase}. For the direct decarboxylation of the related substrate **orotidine** by the enzyme ODC_{ase} (see Scheme 1C), there is some disagreement in the literature on the exact reaction path. Some studies indicate a two-step decarboxylation pathway via a carbanionic intermediate,^{20,21,31,59,60} while others propose a one-step mechanism with concerted C–C bond breakage and proton transfer.^{22,61,62}

The obtained QM/MM energy profile of the direct decarboxylation of **5caU** in the enzymatic environment of IDC_{ase} (Figure 5, decarboxylation step, orange) clearly shows

a one-step process without intermediates (Figure 7a). Since investigations on a small active site model rather suggest a two-step process for the decarboxylation of **taut-5caU**,²⁹ we aim to rule out all ambiguities on the reaction coordinate by further investigations in the following. Therefore, we examined whether there are alternative paths that correspond to a two-step process (Figure 7b) by calculating the potential energy surface (PES) with respect to the C–C bond opening and the proton transfer. In this context, the proton transfer is defined as the difference between the distance of the transferred hydrogen D323:H to D323:O and the distance of D323:H to C5 of **5caU** (for atom names, see Figures 3 and 7).

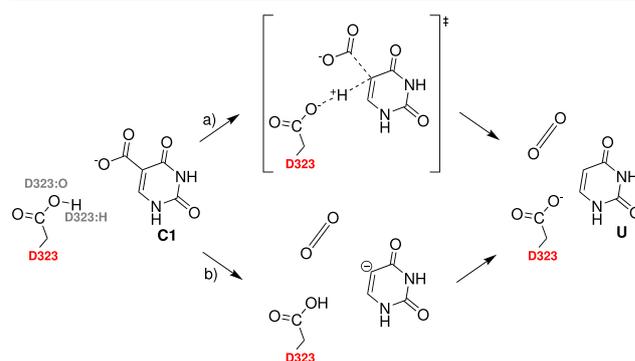


Figure 7. Direct decarboxylation of **5caU** (a) with concerted proton transfer in a one-step mechanism or (b) in a two-step mechanism via an anionic intermediate.

The PES in Figure 8 (top) clearly shows that there are no intermediates, which rules out a two-step mechanism. Furthermore, it can be seen that there exist no alternative reaction paths to the only minimal energy path (MEP) on the surface. The MEP (Figure 8, orange) connects the reactant complex (**5caU**) in the top left corner of the PES with the product complex (**U**) in the bottom right corner with an almost straight line along the diagonal, which means that the C–C bond opening occurs almost perfectly simultaneous to the proton transfer during the reaction.

This detail of the reaction coordinate, however, is sensitive to the used method in the QM calculations. Since we observed that the reaction path slightly varies between different QM methods (Supporting Information, Section S9), we decided to recalculate the PES using HF3c⁶³ (Figure 8, bottom). Here, the MEP (Figure 8, cyan) is not a straight line between **5caU** and **U**. At the transition state (TS) along the HF3c-MEP, the C–C bond with a length of 2.3 Å is slightly more opened than in the TS along the B3LYP-MEP (2.1 Å), whereas the proton transfer is further advanced in the B3LYP-TS, which can easily be seen by comparing the TS structures graphically (Figure 9). In contrast to the B3LYP-MEP, the HF3c-MEP rather suggests

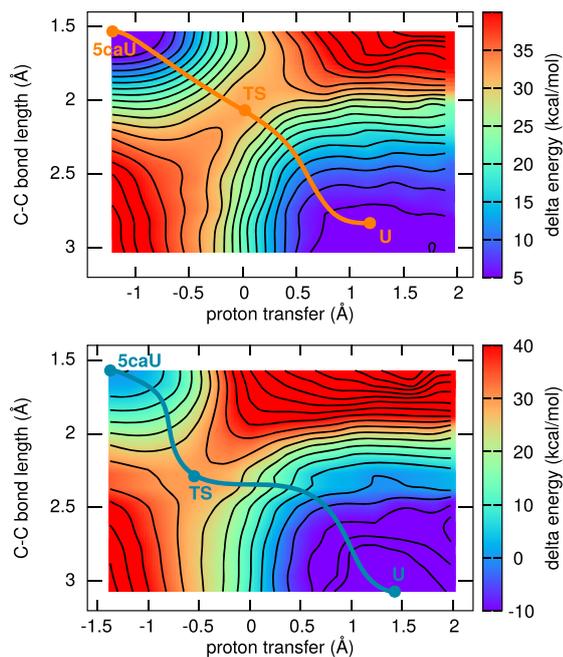


Figure 8. QM/MM potential energy surface of the direct decarboxylation of 5caU by IDCase with respect to the proton transfer and the C–C bond opening using B3LYP-D3 (top) and HF3c (bottom) for the QM calculations. The proton transfer is defined as the distance difference $d(\text{D323:O}, \text{D323:H}) - d(\text{D323:H}, \text{C5})$. The minimal energy paths (MEP) from the reactant complex (5caU) to the product complex (U) are plotted in orange and blue, where the highest point along each path was optimized as a transition state (TS) using the dimer method.^{57,58} The detailed data can be found in the Supporting Information (Section S10).

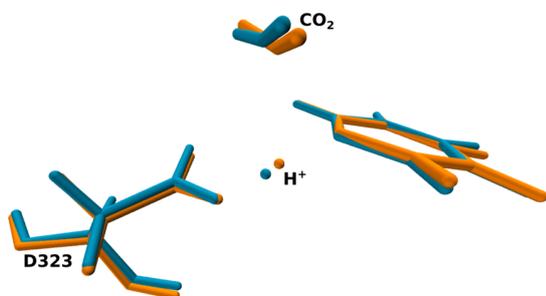


Figure 9. Graphical comparison of the optimized transition structures (TS) along the MEPs (Figure 8) obtained by B3LYP-D3 (orange) and HF3c (cyan).

a more sequential C–C bond opening and proton transfer. Although the path of the MEP on the PES obtained using HF3c to some degree differs from the MEP obtained by B3LYP-D3, both methods neither produce intermediates nor show alternative pathways. This confirms the observation that this reaction is a one-step mechanism.

Altogether, the results prove that the direct decarboxylation of 5caU by IDCase proceeds via a one-step mechanism to which no alternative pathways were found. Furthermore, the investigations indicate that the proton transfer takes place, at least to a certain extent, simultaneously with the proton transfer.

4. CONCLUSIONS

In this work, we investigated the reaction mechanism of the IDCase-catalyzed decarboxylation of 5caU using QM/MM calculations.

Our results show that the reaction proceeds via a direct decarboxylation mechanism, which is energetically more favorable than the previously proposed decarboxylation mechanism of 5caU via a tetrahedral intermediate, or the decarboxylation with a nucleophilic activation.

Detailed investigations on the reaction coordinate of the direct decarboxylation mechanism revealed that it is a one-step mechanism with concerted proton transfer and C–C bond opening.

For this mechanism, D323 must be present in protonated form since it transfers the proton to the substrate. In this work, we could show that D323 can be protonated by the neighboring histidine H251, but there might be alternative ways in which D323 can be protonated. One possibility would be a proton transfer from Y301 or R262' via a water molecule at the Wat3 position, which could be investigated in the future.

Our findings regarding the detailed catalytic mechanism of the decarboxylation of 5caU expand the knowledge of active decarboxylation processes, which can help to better understand the active decarboxylation in the demethylation pathway of SmC.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.0c00616>.

MD simulation protocol, additional information on the reaction profiles, and detailed values of the 2D energy surfaces (PDF)

Force field parameters for the metal center in the reactant complex (ZIP)

Force field parameters for the metal center for the intermediate B2 (ZIP)

Force field parameters for the metal center for the intermediate C1 (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Christian Ochsenfeld – Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), D-81377 Munich, Germany; Max Planck Institute for Solid State Research, D-70569 Stuttgart, Germany; orcid.org/0000-0002-4189-6558; Email: c.ochsenfeld@fkf.mpg.de

Author

Andrea Kreppel – Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), D-81377 Munich, Germany

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.0c00616>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG) via the SFB1309-32587107. The authors are grateful to Eli Naydenova and Beatriz von der

Esch (LMU Munich) for useful discussions. In addition, we thank Dr. J. Kussmann (LMU Munich) for providing a development version of the FermiONs++ program package. C.O. further acknowledges financial support as a Max-Planck Fellow at the MPI-FKF Stuttgart.

REFERENCES

- (1) Handy, D. E.; Castro, R.; Loscalzo, J. Epigenetic modifications: Basic mechanisms and role in cardiovascular disease. *Circulation* **2011**, *123*, 2145–2156.
- (2) Law, J. A.; Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **2010**, *11*, 204–220.
- (3) Oswald, J.; Engemann, S.; Lane, N.; Mayer, W.; Olek, A.; Fundele, R.; Dean, W.; Reik, W.; Walter, J. Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **2000**, *10*, 475–478.
- (4) Mayer, W.; Niveleau, A.; Walter, J.; Fundele, R.; Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **2000**, *403*, 501–502.
- (5) Hajkova, P.; Erhardt, S.; Lane, N.; Haaf, T.; El-Maarri, O.; Reik, W.; Walter, J.; Surani, M. A. Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **2002**, *117*, 15–23.
- (6) Sasaki, H.; Matsui, Y. Epigenetic events in mammalian germ-cell development: Reprogramming and beyond. *Nat. Rev. Genet.* **2008**, *9*, 129–140.
- (7) Kriaucionis, S.; Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science* **2009**, *324*, 929–930.
- (8) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **2009**, *324*, 930–935.
- (9) He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C.-X.; Zhang, K.; He, C.; Xu, G.-L. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **2011**, *333*, 1303–1307.
- (10) 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.
- (11) Pfaffeneder, T.; Hackner, B.; Truß, M.; Münzel, M.; Müller, M.; Deiml, C. A.; Hagemeyer, C.; Carell, T. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. - Int. Ed.* **2011**, *50*, 7008–7012.
- (12) Song, C. X.; Szulwach, K. E.; Dai, Q.; Fu, Y.; Mao, S. Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G. L.; Jin, P.; He, C. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **2013**, *153*, 678–691.
- (13) Guo, F.; Li, X.; Liang, D.; Li, T.; Zhu, P.; Guo, H.; Wu, X.; Wen, L.; Gu, T. P.; Hu, B.; Walsh, C. P.; Li, J.; Tang, F.; Xu, G. L. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **2014**, *15*, 447–459.
- (14) Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeyer, C.; Truss, M.; Carell, T. Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chem. - Int. Ed.* **2012**, *51*, 6516–6520.
- (15) Schiesser, S.; Pfaffeneder, T.; Sadeghian, K.; Hackner, B.; Steigenberger, B.; Schröder, A. S.; Steinbacher, J.; Kashiwazaki, G.; Höfner, G.; Wanner, K. T.; Ochsenfeld, C.; Carell, T. Deamination, Oxidation, and C-C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine. *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.
- (16) Iwan, K.; Rahimoff, R.; Kirchner, A.; Spada, F.; Schröder, A. S.; Kosmatchev, O.; Ferizaj, S.; Steinbacher, J.; Parsa, E.; Müller, M.; Carell, T. 5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo. *Nat. Chem. Biol.* **2018**, *14*, 72–78.
- (17) Wu, S. C.; Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607–620.
- (18) Smiley, J. A.; Kundracik, M.; Landfried, D. A.; Barnes, V. R., Sr.; Axhemi, A. A. Genes of the thymidine salvage pathway: Thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*. *Biochim. Biophys. Acta - Gen. Subj.* **2005**, *1723*, 256–264.
- (19) Xu, S.; Li, W.; Zhu, J.; Wang, R.; Li, Z.; Xu, G.-L.; Ding, J. Crystal structures of iso-orotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase. *Cell Res.* **2013**, *23*, 1296–1309.
- (20) Warshel, A.; Štrajbl, M.; Villà, J.; Florián, J. Remarkable rate enhancement of orotidine 5'-monophosphate decarboxylase is due to transition-state stabilization rather than to ground-state destabilization. *Biochemistry* **2000**, *39*, 14728–14738.
- (21) Raugei, S.; Cascella, M.; Carloni, P. A proficient enzyme: Insights on the mechanism of orotidine monophosphate decarboxylase from computer simulations. *J. Am. Chem. Soc.* **2004**, *126*, 15730–15737.
- (22) Wu, N.; Mo, Y.; Gao, J.; Pai, E. F. Electrostatic stress in catalysis: Structure and mechanism of the enzyme orotidine monophosphate decarboxylase. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2017–2022.
- (23) Beak, P.; Siegel, B. Mechanism of Decarboxylation of 1,3-Dimethylorotic Acid. A Model for Orotidine 5'-Phosphate Decarboxylase. *J. Am. Chem. Soc.* **1976**, *98*, 3601–3606.
- (24) Lee, J. K.; Houk, K. N. A proficient enzyme revisited: The predicted mechanism for orotidine monophosphate decarboxylase. *Science* **1997**, *276*, 942–945.
- (25) Appleby, T. C.; Kinsland, C.; Begley, T. P.; Ealick, S. E. The crystal structure and mechanism of orotidine 5'-monophosphate decarboxylase. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2005–2010.
- (26) Lee, T. S.; Chong, L. T.; Chodera, J. D.; Kollman, P. A. An alternative explanation for the catalytic proficiency of orotidine 5'-phosphate decarboxylase. *J. Am. Chem. Soc.* **2001**, *123*, 12837–12848.
- (27) Shostak, K.; Jones, M. E. Orotidylate Decarboxylase: Insights into the Catalytic Mechanism from Substrate Specificity Studies. *Biochemistry* **1992**, *31*, 12155–12161.
- (28) Silverman, R. B.; Groziak, M. P. Model Chemistry for a Covalent Mechanism of Action of Orotidine 5'-Phosphate Decarboxylase. *J. Am. Chem. Soc.* **1982**, *104*, 6434–6439.
- (29) Sheng, X.; Plasch, K.; Payer, S. E.; Ertl, C.; Hofer, G.; Keller, W.; Braeuer, S.; Goessler, W.; Glueck, S. M.; Himo, F.; Faber, K. Reaction mechanism and substrate specificity of Iso-orotate decarboxylase: A combined theoretical and experimental study. *Front. Chem.* **2018**, *6*, 608.
- (30) Schön, A.; Kaminska, E.; Schelter, F.; Ponkkonen, E.; Korytiaková, E.; Schiffrers, S.; Carell, T. Analysis of an active deformation mechanism of 5-formyl-deoxycytidine (fdC) in stem cells. *Angew. Chem., Int. Ed.* **2020**, *1–5594*.
- (31) Hu, H.; Boone, A.; Yang, W. Mechanism of OMP decarboxylation in orotidine 5'-monophosphate decarboxylase. *J. Am. Chem. Soc.* **2008**, *130*, 14493–14503.
- (32) Xu, G. L.; Wong, J. Oxidative DNA demethylation mediated by Tet enzymes. *Natl. Sci. Rev.* **2015**, *2*, 318–328.
- (33) Czermiński, R.; Lesyng, B.; Pohorille, A. Tautomerism of oxypyridines and oxypyrimidines: Theoretical study with complete optimization of geometry. *Int. J. Quantum Chem.* **1979**, *16*, 1141–1148.
- (34) Scanlan, M. J.; Hillier, I. H. Accurate Prediction of the Relative Energies of the Six Tautomers of Uracil. *Chem. Phys. Lett.* **1983**, *98*, 545–547.
- (35) Norinder, U. A theoretical reinvestigation of the nucleic bases adenine, guanine, cytosine, thymine and uracil using AM1. *J. Mol. Struct. THEOCHEM* **1987**, *151*, 259–269.
- (36) Tsuchiya, Y.; Tamura, T.; Fujii, M.; Ito, M. Keto-enol tautomer of uracil and thymine. *J. Phys. Chem.* **1988**, *92*, 1760–1765.

- (37) Wolken, J. K.; Tureček, F. Proton affinity of uracil. A computational study of protonation sites. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 1065–1071.
- (38) Kryachko, E. S.; Nguyen, M. T.; Zeegers-Huyskens, T. Theoretical study of tautomeric forms of uracil. I. Relative order of stabilities and their relation to proton affinities and deprotonation enthalpies. *J. Phys. Chem. A* **2001**, *105*, 1288–1295.
- (39) Zhang, R.; Ceulemans, A.; Nguyen, M. T. A theoretical study of uracil and its tautomers in their lowest-lying triplet state. *Mol. Phys.* **2005**, *103*, 983–994.
- (40) Sochacka, E.; Lodyga-Chruscinska, E.; Pawlak, J.; Cypriak, M.; Bartos, P.; Ebenryter-Olbinska, K.; Leszczynska, G.; Nawrot, B. C5-substituents of uridines and 2-thiouridines present at the wobble position of tRNA determine the formation of their keto-enol or zwitterionic forms - A factor important for accuracy of reading of guanosine at the 3'-end of the mRNA codons. *Nucleic Acids Res.* **2017**, *45*, 4825–4836.
- (41) Colasurdo, D. D.; Pila, M. N.; Iglesias, D. A.; Laurella, S. L.; Ruiz, D. L. Tautomerism of uracil and related compounds: A mass spectrometry study. *Eur. J. Mass Spectrom.* **2018**, *24*, 214–224.
- (42) Hartono, Y. D.; Ito, M.; Villa, A.; Nilsson, L. Computational Study of Uracil Tautomeric Forms in the Ribosome: The Case of Uracil and 5-Oxyacetic Acid Uracil in the First Anticodon Position of tRNA. *J. Phys. Chem. B* **2018**, *122*, 1152–1160.
- (43) 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.
- (44) 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.
- (45) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- (46) 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.
- (47) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell—a modular software package for QM/MM simulations. *WIREs Comput. Mol. Sci.* **2014**, *4*, 101–110.
- (48) ChemShell, a Computational Chemistry Shell. www.chemshell.org.
- (49) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct. THEOCHEM* **2003**, *632*, 1–28.
- (50) Kussmann, J.; Ochsenfeld, C. Pre-selective screening for matrix elements in linear-scaling exact exchange calculations. *J. Chem. Phys.* **2013**, *138*, 134114.
- (51) 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.
- (52) Kussmann, J.; Ochsenfeld, C. Hybrid CPU/GPU Integral Engine for Strong-Scaling Ab Initio Methods. *J. Chem. Theory Comput.* **2017**, *13*, 3153–3159.
- (53) 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.
- (54) 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.
- (55) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (56) 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.
- (57) Henkelman, G.; Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **1999**, *111*, 7010–7022.
- (58) 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.
- (59) Toth, K.; Amyes, T. L.; Wood, B. M.; Chan, K.; Gerlt, J. A.; Richard, J. P. Product deuterium isotope effect for orotidine 5'-monophosphate decarboxylase: Evidence for the existence of a short-lived carbanion intermediate. *J. Am. Chem. Soc.* **2007**, *129*, 12946–12947.
- (60) Amyes, T. L.; Wood, B. M.; Chan, K.; Gerlt, J. A.; Richard, J. P. Formation and stability of a vinyl carbanion at the active site of orotidine 5'-monophosphate decarboxylase: pKa of the C-6 proton of enzyme-bound UMP. *J. Am. Chem. Soc.* **2008**, *130*, 1574–1575.
- (61) Begley, T. P.; Appleby, T. C.; Ealick, S. E. The structural basis for the remarkable catalytic proficiency of orotidine 5'-monophosphate decarboxylase. *Curr. Opin. Struct. Biol.* **2000**, *10*, 711–718.
- (62) Begley, T. P.; Ealick, S. E. Enzymatic reactions involving novel mechanisms of carbanion stabilization. *Curr. Opin. Chem. Biol.* **2004**, *8*, 508–515.
- (63) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672–1685.

Supporting Information: Comprehensive Computational Study of the Enzymatic Decarboxylation Mechanism of 5-Carboxy Uracil

Andrea Kreppel[†] and Christian Ochsenfeld^{*,†,‡}

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

[‡]*Max Planck Institute for Solid State Research, Heisenbergstr. 1, D-70569 Stuttgart,
Germany*

E-mail: c.ochsenfeld@fkf.mpg.de

Contents

S1 MD simulation protocol	S3
S2 Details on the QM/MM energy profiles of the investigated decarboxylation pathways	S4
S2.1 Pathway A	S4
S2.2 Pathway B	S6
S2.3 Pathway C	S7
S3 Different nucleophiles in decarboxylation pathway A of IDCase	S8
S3.1 D323	S8
S3.2 OH ⁻	S9
S4 Water molecules in the active site for intermediate B2	S11
S5 Alternative decarboxylation and nucleophile elimination in pathway B	S11
S6 Influence of the initial structure on the energy profile of pathway C	S13
S7 QM/MM energy profiles of the decarboxylation step using an enlarged QM region	S14
S8 QM/MM energy profiles of the decarboxylation step using different QM methods	S15
S9 Dependence of the reaction coordinate for pathway C on the QM method	S16
S10 2D QM/MM energy profile of the direct decarboxylation	S18
References	S20

S1 MD simulation protocol

For the MD preparation and production, the ff14SB force field contained in the AMBER16 program package^{1,2} was used within the simulation engine NAMD v2.7.³

The reactant complex and intermediate C1 were simulated for 60 ns and the intermediate B2 in pathway B for 30 ns. Bonds to hydrogens were constrained by the SETTLE algorithm⁴ allowing for timesteps of 2 fs. The temperature was held constant at 300 K using Langevin dynamics and the pressure at 1.013 atm by the Nosé-Hoover Langevin piston pressure control.^{5,6} The electrostatic interactions were evaluated using the particle mesh Ewald summation method. The van der Waals interactions were cut off by means of a switching function in the interval from 10.0 to 13.5 Å. Full electrostatic interactions were evaluated every 4 fs. Coordinates were saved every 2 ps.

The systems were equilibrated before performing production runs according to the following protocol: In the first preparation step the water molecules were relaxed with 10000 steps of conjugate gradient minimization while the protein, DNA, and the Zn²⁺-ion were kept frozen. Second, the system was optimized again with 20000 steps of conjugate gradient minimization while the protein, DNA, and Zn²⁺ were kept under restraints of 1 kcal/Å².

Then, the system was heated to 300 K by increasing the temperature by 1 K every 100 time steps under the same restraints as in the second minimization. The heated system was equilibrated in two steps while bonds to hydrogens were constrained by the SETTLE algorithm:⁴ 1.) for 200 ps the temperature was rescaled to 300 K every 1 ps; 2.) for 100 ps the temperature was controlled by Langevin dynamics and it was switched to an NPT ensemble with the Nosé-Hoover Langevin piston pressure control.^{5,6} Next the restraints were reduced stepwise (see Table S1) until the system was fully unrestrained after 400 ps of equilibration within the NPT ensemble.

Table S1: Restraints used for protein and DNA during equilibration in an NPT ensemble.

restraint [kcal/(molÅ ²)]	simulation time
1.0	100 ps
0.8	30 ps
0.6	30 ps
0.5	30 ps
0.4	30 ps
0.3	20 ps
0.2	20 ps
0.1	20 ps
0.0	120 ps

S2 Details on the QM/MM energy profiles of the investigated decarboxylation pathways

The QM/MM energy profiles of the decarboxylation pathways of IDCase (fig. 5) were obtained by performing constrained optimizations along the reaction coordinate (RC) with the optimizer tool DL-FIND.⁷ The RC was defined as linear combination of atom1-atom2 distances ($d(\text{atom1}, \text{atom2})$) which describe the bonds that open or form in the reaction.

S2.1 Pathway A

In path A, the RC for the first reaction step (**5caU** \rightarrow **A1**, fig. S1) was defined as distance $d(\text{Wat:O}, \text{Wat:H}) - d(\text{Wat:H}, \text{O53}) - d(\text{Wat:O}, \text{C51})$. Therefore, a starting structure was chosen from the MD simulation based on a minimal distance $d(\text{Wat:O}, \text{C51})$. For RC of the second step (**A1** \rightarrow **U**, fig. S1) distance $d(\text{C51}, \text{C5}) + d(\text{O53}, \text{Wat:H}) - d(\text{Wat:H}, \text{C5})$ was chosen.

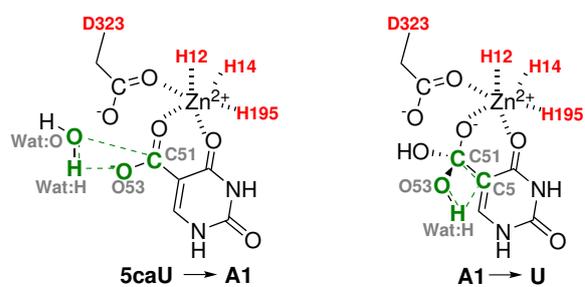


Figure S1: Illustration of the included distances in the reaction coordinates along the decarboxylation pathway A.

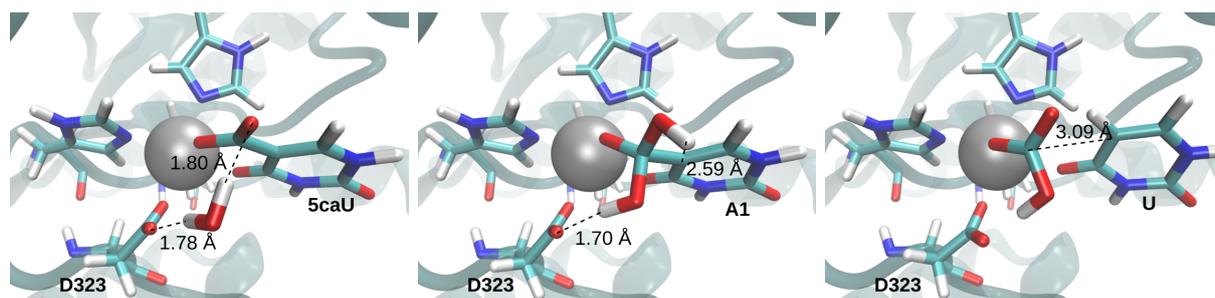


Figure S2: Reactant complex (left), intermediate A1 (middle), and product complex (right) along pathway A.

S2.2 Pathway B

For the first step in the activation part of the decarboxylation pathway B (**5caU** \rightarrow **B1**, fig. S3), the distance $d(\text{Wat:H,Wat:O})-d(\text{Wat:O,C6})-d(\text{Wat:H,D323:O})$ was used for the RC. A starting structure was selected from the MD simulation based on a minimal $d(\text{Wat:O,C6})$ distance. In the second step (**B1** \rightarrow **B2**, fig. S3) the distance $d(\text{D323:O,D323:H})-d(\text{D323:H,C5})$ was used as RC.

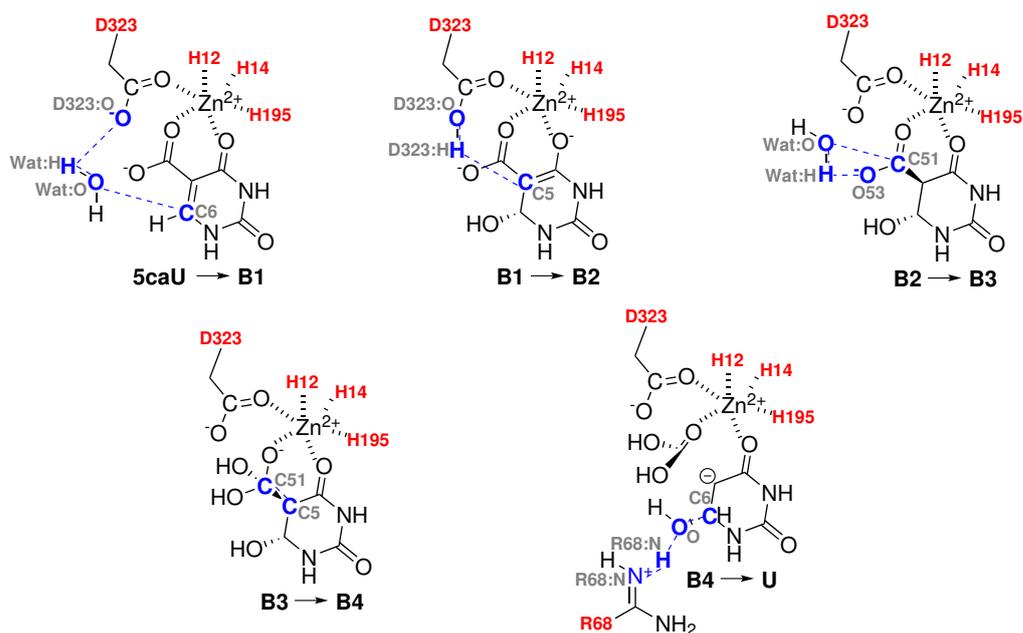


Figure S3: Illustration of the included distances in the reaction coordinates along the decarboxylation pathway B.

At the intermediate state **B2**, an MD simulation was performed in order to allow for a refilling of the water positions in the active site. For the decarboxylation steps a structure was selected from this simulation based on a minimal $d(\text{Wat:O,C51})$ distance. The first step of the decarboxylation part was described by the RC $d(\text{Wat:O,Wat:H})-d(\text{Wat:H,O53})-d(\text{Wat:O,C51})$ (**B2** \rightarrow **B3**, fig. S3) and the second step by $d(\text{C5,C51})$ (**B3** \rightarrow **B4**, fig. S3). The elimination step (**B4** \rightarrow **U**, fig. S3) proceeds via the RC $d(\text{R68:H,O})-d(\text{O,C6})-d(\text{R68:H,R68:N})$.

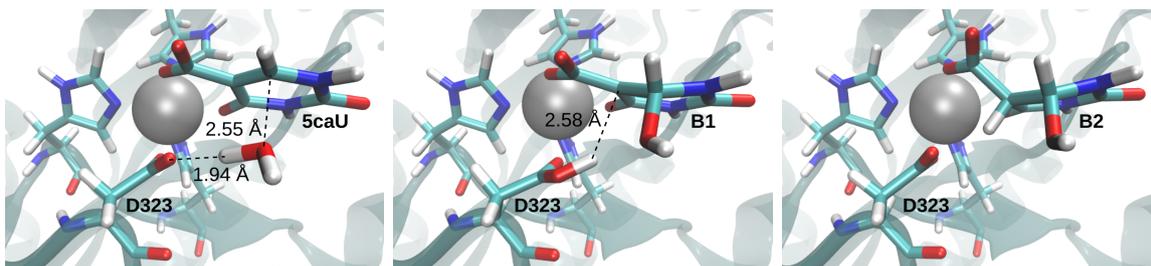


Figure S4: Reactant complex (left), intermediate B1 (middle), and intermediate B2 (right) along the activation part of pathway B.

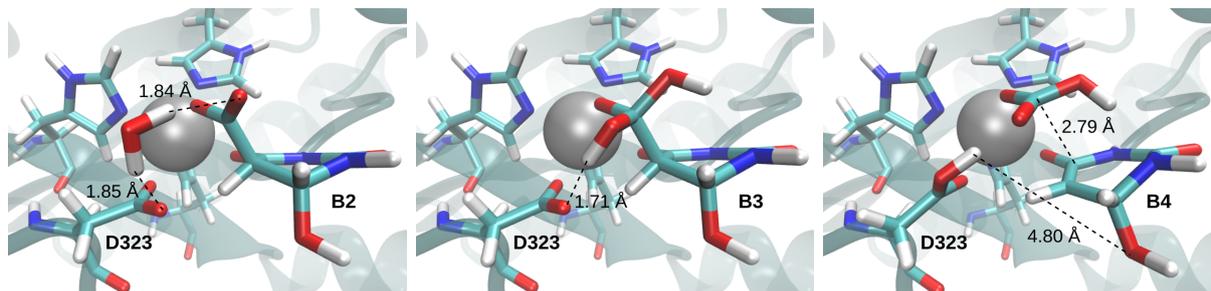


Figure S5: Intermediate B2 (left), intermediate B3 (middle), and intermediate B4 (right) along the decarboxylation part of pathway B.

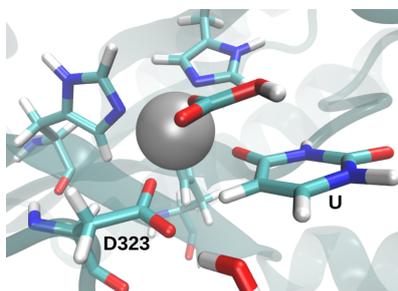


Figure S6: Final product complex after elimination of the nucleophile in pathway B.

S2.3 Pathway C

The preceding proton transfer in pathway C ($5caU \rightarrow C1$, fig. S7) is described by the RC $d(H251:N, H251:H) - d(H251:H, D323:O)$ and a starting structure was selected based on minimal distance $d(H251:H, D323:O)$. The intermediate C1 was relaxed by an MD simulation. From this MD simulation a structure based on a minimal distance $d(D323:H, C5)$ was selected for the successive decarboxylation ($C1 \rightarrow U$, fig. S7) which is described by the RC $d(D323:H, C5) - d(C5, C51) - d(D323:H, D323:O)$.

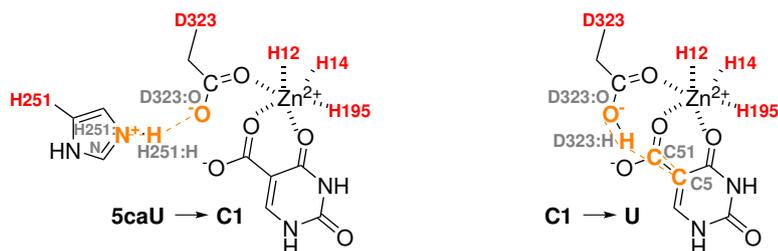


Figure S7: Illustration of the included distances in the reaction coordinates along the decarboxylation pathway C.

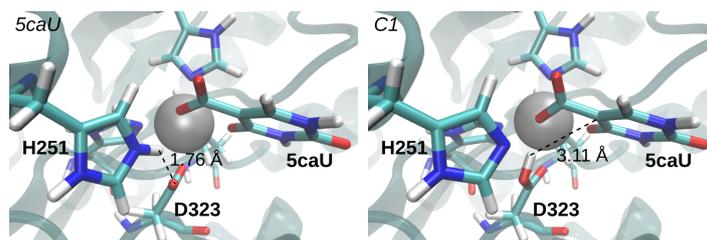


Figure S8: Reactant complex (left) and product complex (right) of the preceding protonation of D323 in pathway C.

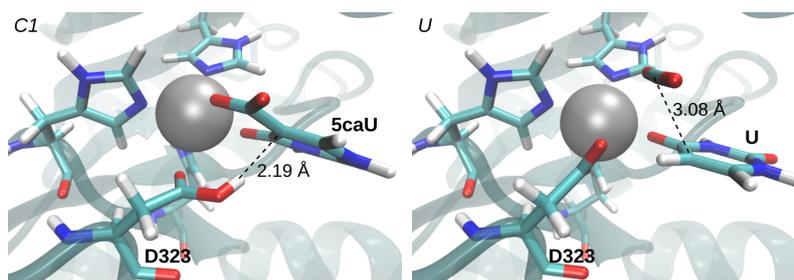


Figure S9: Reactant complex (left) and product complex (right) of the direct decarboxylation pathway C.

S3 Different nucleophiles in decarboxylation pathway A of IDCase

S3.1 D323

To investigate whether D323 could serve as a nucleophile for the nucleophilic attack at C51 of 5caU in the decarboxylation pathway A, the QM/MM energy profile was calculated for this reaction. The profile was obtained by constraint optimizations along the reaction coordinate

defined as the distance $d(\text{D323:O}, \text{5caU:C51})$ (for the atom names, see fig. S10, left). In the QM calculations the substrate 5caU, the Zn^{2+} ion, the ligating histidines (H12, H14, H195), and D323 were included. As the energy profile only rises along the reaction coordinate and does not show a local minimum for the intermediates (see fig. S10, right), it can be assumed that a stable tetrahedral intermediate cannot be obtained by the nucleophilic attack of D323.

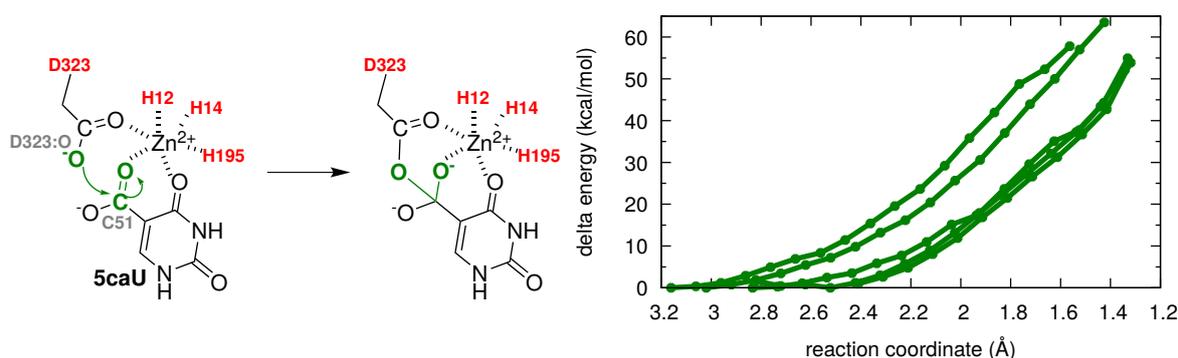


Figure S10: Nucleophilic attack of D323 at the carboxyl group of 5caU (left) and the QM/MM energy profile of this reaction obtained using B3LYP-D3⁸⁻¹² and a def2-SVP basis¹³ set for the QM calculations and the AMBER 16 FF^{1,2} for the description of the MM part.

S3.2 OH^-

For the investigation whether an OH^- ion can be formed in the active site of IDCase which could function as a nucleophile in the decarboxylation pathway A, the QM/MM energy profile of the deprotonation of H_2O by D323, histidine H251, or the substrate itself was calculated. In the QM calculations the substrate 5caU, the Zn^{2+} ion, the ligating histidines (H12, H14, H195), D323, and a water molecule were included. The reaction coordinates were defined as the linear combination of the distances: $d(\text{Wat:O}, \text{Wat:H}) - d(\text{Wat:H}, \text{X})$, where X is D323:O, O53 of 5caU, or H251:N, respectively (for the atom names, see fig. S11, left). In case that H251 would function as the proton acceptor in the deprotonation reaction, this residue was also included in the QM region. Here, again the QM/MM energy profiles steadily rise

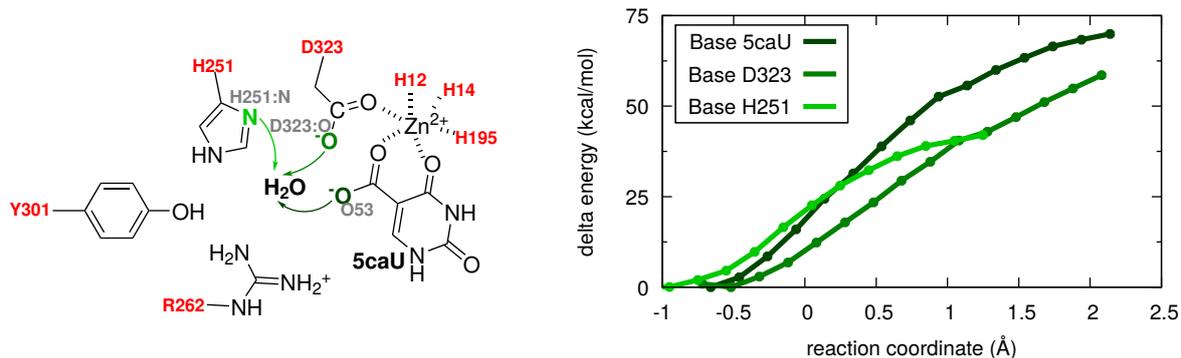


Figure S11: Formation of an OH⁻ in the active site of IDCase by deprotonation of a water molecule by H251 (light green), D323 (medium green), or 5caU (dark green) (left) and the QM/MM energy profile of this reaction obtained using B3LYP-D3 and a def2-SVP basis set for the QM calculations and the AMBER 16 FF for the description of the MM part (right).

and do not exhibit any local minima for the deprotonation product complexes (see fig. S11, right). This reveals that an OH⁻ ion in the active site is very unfavorable.

S4 Water molecules in the active site for intermediate B2

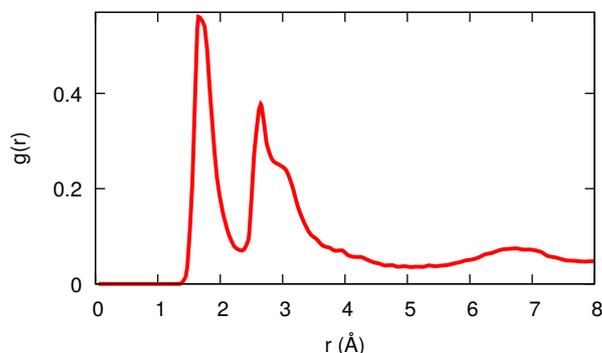


Figure S12: Radial distribution function of water molecules around the carboxyl oxygen of intermediate B2.

For the intermediate B2 an MD simulation was performed to allow for a refilling of the water positions in the active site as a water molecule is needed for the nucleophilic attack in the subsequent reaction step. The radial distribution function (fig. S12) of water molecules around the carboxyl oxygen of B2 shows that there are again two water positions at similar distances, of around 1.9 \AA and 3.0 \AA , as in the simulation of the reactant complex (fig. 4 in the manuscript). The Wat3 position, however, is shifted to a slightly larger distance. This can be explained by the fact that due to the change in the hybridization at C5, the carboxyl group of B2 moves slightly further away from the residues that form hydrogen bonds to Wat3 (Y301, R262 ').

S5 Alternative decarboxylation and nucleophile elimination in pathway B

On the basis of the proposed decarboxylation mechanism with a nucleophile addition for the substrate orotidine¹⁴ (scheme 1 in the manuscript, B. a), the activation step of the substrate 5caU (scheme 3 in the manuscript, B. 1) could also be followed by a concerted CO_2 loss and nucleophile elimination resulting in the final product uracil in just one reaction step

(fig. S13, left). The obtained QM/MM energy profile for this alternative pathway (fig. S13, right), however, makes this alternative pathway unlikely due to the relatively large energy barrier of about 50 kcal/mol.

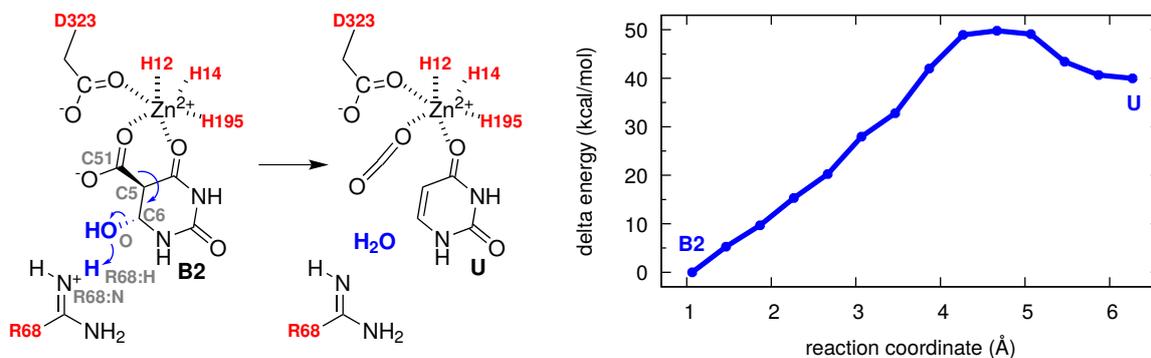


Figure S13: Alternative pathway of decarboxylation and nucleophile elimination in mechanism B (left) and the QM/MM energy profile of this reaction obtained using B3LYP-D3 and a def2-SVP basis set for the QM calculations and the AMBER 16 FF for the description of the MM part (right). Here, the reaction coordinate is defined as $d(C5,C51)+d(C6,O)-d(O,R68:H)+d(R68:H,R68:N)$.

S6 Influence of the initial structure on the energy profile of pathway C

To reduce the influence of the starting geometry taken from the classical MD simulation using the parameters obtained by MCPB.py for the metal center, the energy surface of the reaction was explored by iteratively calculating the forward and reverse reaction several times (fig. S14). As the energy profiles do not significantly vary between the iterations, it can be assumed that the starting geometries taken from the FF MD simulation are reasonable.

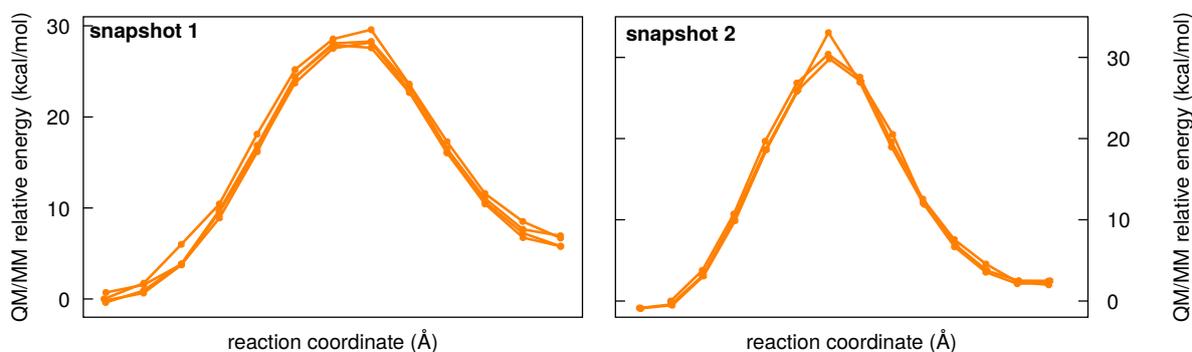


Figure S14: QM/MM energy profiles of the decarboxylation pathways C obtained using B3LYP-D3 and a def2-SVP basis set calculated several times in forward and reverse direction, initially starting from two different snapshots taken from the FF MD simulation.

S7 QM/MM energy profiles of the decarboxylation step using an enlarged QM region

The decarboxylation step of pathways A, B, and C were recalculated using an extended QM region with comparable size for all three pathways. The QM region includes in all three cases the residues H12, H14, R68, N98, H195, L218, F222, H251, V299, Y301, T322, H324, F326, D323, R262', Zn²⁺, **5caU**, and 4 water molecules. The obtained energy profiles with the enlarged QM region (fig. S15, dashed lines) resemble the respecting energy profiles with a minimal QM region (fig. S15, solid lines). The only difference is in pathway B, where the intermediate B3 is lower in energy than in the corresponding energy profile obtained with a minimal QM region. This can be explained by the formation of a hydrogen bond between R262' and the intermediate in the calculation with the enlarged QM region.

The calculations with the enlarged QM region confirm that decarboxylation along pathway C has the lowest energy barrier.

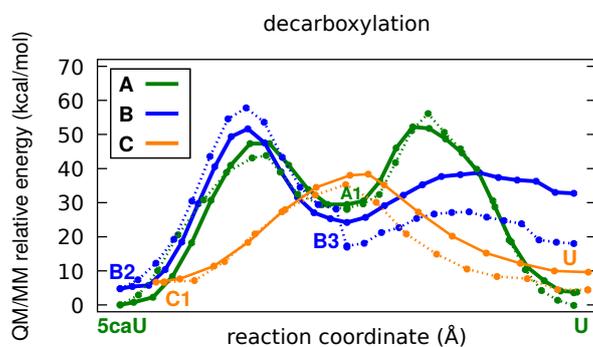


Figure S15: QM/MM energy profiles of the decarboxylation step according to pathways A (green), B (blue), and C (orange) with a minimal QM region (solid) and the enlarged QM region (dashed).

S8 QM/MM energy profiles of the decarboxylation step using different QM methods

The energy profiles of the decarboxylation step were recalculated using different density functional methods for QM region: B97-2,¹⁵ cam-B3LYP,¹⁶ and M06-2x¹⁷ all with Grimme's D3 dispersion^{11,12} correction and a def2-SVP¹³ basis set. In addition calculations using HF3c¹⁸ were performed. The obtained profiles (fig. S16) for all former methods indicate that the decarboxylation according to pathway C has the lowest energy barrier.

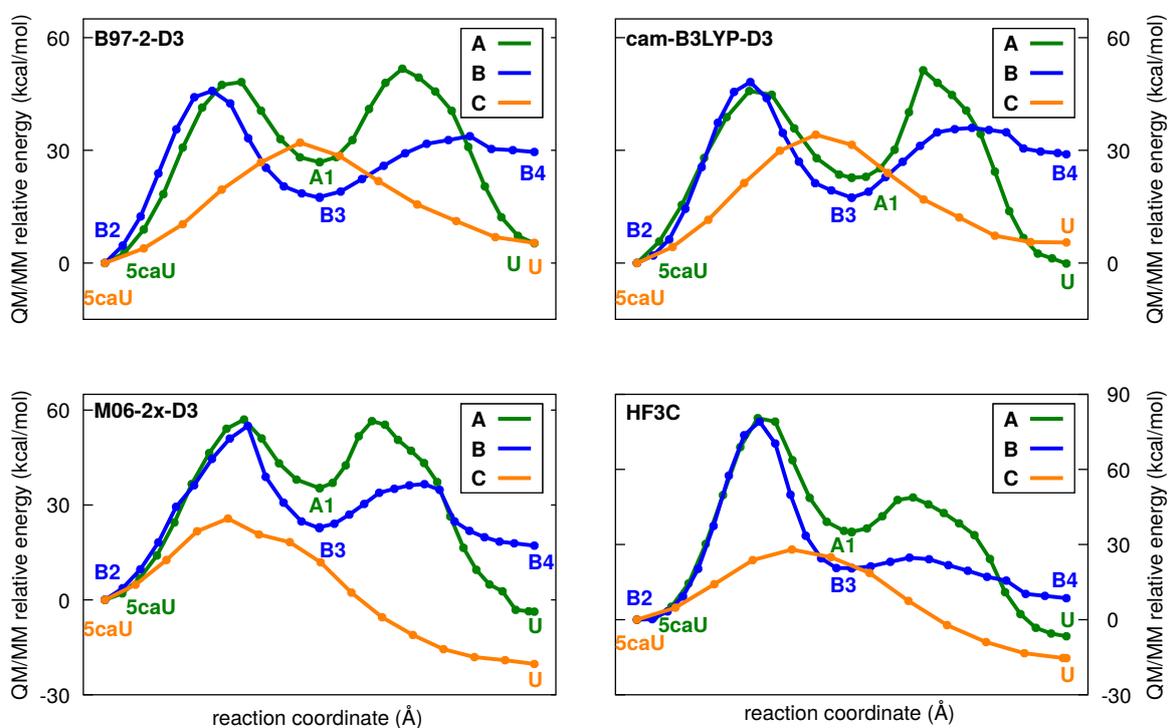


Figure S16: QM/MM energy profiles of the decarboxylation step according to pathways A (green), B (blue), and C (orange) obtained using B97-2-D3/def2-SVP (top left), cam-B3LYP-D3/def2-SVP (top right), M06-2x-D3/def2-SVP (bottom left), and HF3c (bottom right) for the QM region.

S9 Dependence of the reaction coordinate for pathway C on the QM method

The energy profiles for the direct decarboxylation mechanism obtained by using different QM methods (fig. S17) reveal slight differences in the paths. The energy profiles can be

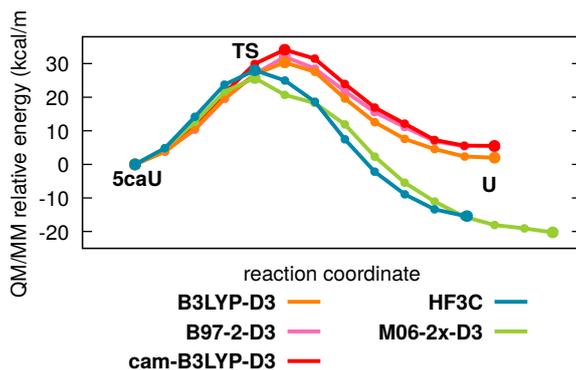


Figure S17: QM/MM energy profile of the direct decarboxylation of **5caU** by IDCCase (pathway C) using the density functionals B3LYP (orange), B97-2 (pink), cam-B3LYP (red), and M06-2x (green) and the method HF3c (cyan) for the QM calculations. For all density functional calculations Grimme D3 dispersion correction was used and the basis set def2-SVP.

classified into two groups. While the profiles obtained using B97-2-D3 and cam-B3LYP strongly resemble the B3LYP-D3 profile, the profiles obtained by HF3c and M06-2x-D3 have the maxima (transition state, **TS**) shifted to an earlier point in the reaction coordinate. Furthermore, the product complexes (**U**) are significant lower in energy for the second group of QM approaches.

This behavior can also be observed by comparing the structures of reactant complex (**5caU**), **TS**, and **U** obtained by the different methods as compared to the corresponding B3LYP-D3 structures. While the RMSD values for D323 and **5caU** of the B97-2-D3 and cam-B3LYP-D3 structures compared to the B3LYP-D3 structures are below 0.1 (see table S2), the HF3c and M06-2x-D3 structures show larger deviations.

A closer look at the **TS** structures (fig. S18) reveals that in the **TS** structures of B3LYP-D3, B97-2-D3, and cam-B3LYP-D3 the proton transfer is at a more advanced state indicating

Table S2: Differences for structures of the reactant state (**5caU**), transition state (**TS**), and product state (**U**) along the direct decarboxylation pathway (see Figure S17) as compared to the respective structure obtained at B3LYP-D3 level given as root-mean-square deviation (RMSD) of residue D323 and **5caU** in Å.

	B97-2-D3	cam-B3LYP	M06-2x-D3	HF3c
5caU	0.04	0.03	0.10	0.12
TS	0.08	0.04	0.21	0.18
U	0.03	0.04	0.18	0.25

a decarboxylation with concurrent proton transfer whereas in the HF3c and M06-2x-D3 TS structures the C-C bond is further opened rather suggesting a decarboxylation mechanism via a carbanion.

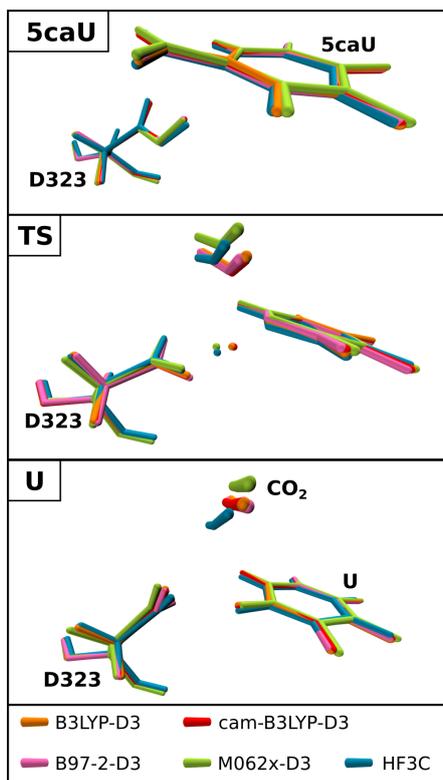


Figure S18: Graphical comparison of the reactants (**5caU**), transition state (**TS**), and product (**U**) complexes along the direct decarboxylation pathway obtained by different QM methods (see Figure S17).

S10 2D QM/MM energy profile of the direct decarboxylation

The 2D QM/MM energy profile of the direct decarboxylation was obtained by optimizations, where the C-C bond length as well as the proton transfer from D323 to C5 of 5caU are constraint to a specific value. The proton transfer is therefore defined as distance $d(\text{D323:H}, \text{D323:O}) - d(\text{D323:H}, \text{C5})$. The 2D energy profile was calculated using B3LYP (table S3) and HF3c (table S4).

Table S3: Relative QM/MM energies in kcal/mol of the direct decarboxylation with respect to the C-C bond opening and the proton transfer in Å. For the QM calculations the B3LYP density functional with Grimme's D3 dispersion correction and a def2-SVP basis set was used.

proton transfer (Å)	C-C bond length (Å)															
	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
-1.2	0.0	1.6	6.3	12.3	18.3	24.4	29.2	34.1	37.1	40.4	43.2	44.0	45.3	46.7	46.4	47.4
-1.0	0.7	2.3	6.7	12.4	18.1	23.8	28.2	32.3	35.3	38.2	40.2	40.9	41.6	42.7	42.3	43.1
-0.8	3.2	4.6	8.5	13.8	19.0	24.1	28.0	31.6	34.0	36.1	37.2	37.9	38.6	39.3	39.1	39.7
-0.6	7.9	8.6	12.1	17.0	21.5	25.8	29.1	31.8	33.8	35.1	35.4	35.3	36.0	35.8	35.6	36.0
-0.4	13.6	14.4	17.2	21.2	24.8	28.5	31.0	32.9	33.3	33.5	33.9	33.3	33.7	32.9	32.6	32.9
-0.2	20.5	20.5	22.8	25.6	28.6	30.6	32.1	32.9	32.4	31.6	30.2	30.3	30.8	29.4	28.9	28.8
0.0	26.6	25.9	27.5	29.5	31.8	32.1	32.0	30.2	28.5	27.3	25.4	24.8	23.4	21.9	21.9	22.2
0.2	30.7	29.4	30.3	31.5	31.7	31.2	29.4	27.0	23.9	20.4	19.9	19.1	17.1	15.6	15.6	15.8
0.4	34.7	32.9	32.5	33.4	31.8	29.7	25.4	21.5	18.4	15.5	13.5	12.0	10.5	10.3	10.2	10.1
0.6	38.1	35.9	35.4	34.6	33.0	28.2	23.2	18.7	15.4	12.3	9.9	8.3	6.7	6.4	6.7	5.5
0.8	42.4	40.0	38.3	37.4	34.4	28.0	22.5	17.6	14.0	11.0	8.4	6.7	5.1	4.7	4.4	3.5
1.0	44.9	43.3	42.2	39.8	31.8	25.2	20.9	17.8	14.1	10.9	8.1	6.5	4.8	3.3	3.8	3.1
1.2	46.1	46.0	44.5	37.2	31.5	25.5	21.1	15.9	14.9	11.3	8.6	6.9	4.1	3.4	4.2	3.5
1.4	49.7	48.8	41.9	37.4	32.1	25.6	21.8	16.1	13.1	10.4	7.5	6.7	4.5	3.9	4.9	3.4
1.6	53.5	44.2	41.4	37.9	32.8	26.3	22.0	16.9	13.3	10.8	7.9	7.4	5.3	4.8	4.2	4.1
1.8	36.3	44.4	41.9	37.6	32.6	22.9	22.8	17.6	13.1	11.7	8.7	6.5	5.8	3.5	4.8	5.0
2.0	29.5	44.6	42.9	38.2	33.2	22.3	20.6	17.7	13.6	9.4	8.9	7.0	6.9	3.9	5.5	2.5

Table S4: Relative QM/MM energy in kcal/mol of the direct decarboxylation with respect to the C-C bond opening and the proton transfer in Å. For the QM calculations the HF3c method was used.

proton transfer (Å)	C-C bond length (Å)															
	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0	3.1
-1.4	0.0	1.5	6.3	11.8	18.4	24.8	29.1	34.3	38.6	40.9	43.8	46.2	48.4	49.9	50.8	52.4
-1.2	0.0	1.7	6.1	11.5	17.8	23.0	27.7	32.4	36.0	37.9	40.1	40.9	43.0	43.4	44.4	45.7
-1.0	2.6	3.9	7.8	12.6	18.2	22.9	27.0	30.6	33.2	34.2	35.3	36.0	37.6	37.6	38.5	39.4
-0.8	6.8	7.3	11.6	15.5	20.5	24.3	27.1	29.4	31.2	31.1	31.3	31.4	31.6	32.3	32.6	33.6
-0.6	15.0	14.7	17.6	21.1	24.6	26.9	28.8	29.5	29.1	28.4	27.5	26.9	27.1	27.0	27.4	27.9
-0.4	26.9	25.7	27.3	29.0	30.7	31.0	30.4	28.9	27.9	26.3	24.5	23.3	23.3	22.6	22.9	23.0
-0.2	40.0	37.9	37.9	37.5	35.9	33.1	31.1	27.3	23.6	21.1	19.7	18.3	17.8	17.3	17.5	17.0
0.0	51.2	47.8	46.0	43.1	38.8	32.8	27.8	21.7	17.8	14.7	13.2	10.9	9.4	9.3	9.2	8.7
0.2	57.0	52.6	49.2	44.0	37.1	28.6	23.1	15.4	10.8	7.3	5.3	3.2	1.3	0.8	-0.3	0.5
0.4	61.8	56.6	50.5	43.4	35.3	24.8	15.5	10.1	4.4	1.4	-0.7	-4.9	-5.1	-5.7	-7.0	-6.1
0.6	64.6	58.7	52.9	44.3	32.9	20.4	11.8	4.7	0.4	-2.2	-6.9	-9.2	-9.6	-11.1	-11.0	-10.3
0.8	68.5	62.5	55.0	41.7	32.1	18.5	10.3	2.6	-3.5	-4.9	-9.1	-11.1	-11.7	-13.7	-13.7	-20.5
1.0	69.6	62.7	54.7	40.2	27.6	16.1	7.4	2.0	-4.6	-5.5	-9.5	-12.6	-21.2	-22.1	-22.8	-23.5
1.2	73.8	66.4	51.2	40.0	27.4	15.9	7.0	-0.2	-4.6	-4.9	-9.1	-12.6	-21.3	-22.7	-24.0	-24.9
1.4	77.1	69.2	51.3	40.5	27.8	16.3	7.5	-0.1	-4.5	-8.3	-12.0	-11.8	-20.1	-22.2	-23.8	-25.1
1.6	78.3	72.4	51.8	41.4	28.0	17.2	8.1	0.7	-3.5	-8.2	-11.6	-10.7	-17.8	-20.3	-22.3	-24.0
1.8	69.4	58.6	51.7	41.4	29.0	17.7	9.5	1.4	-2.8	-7.1	-10.9	-10.4	-14.3	-17.3	-19.8	-21.9
2.0	69.0	58.1	52.3	42.3	23.5	17.3	9.8	2.9	-2.6	-12.1	-9.4	-10.1	-9.8	-14.3	-16.6	-18.6

References

- (1) 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.
- (2) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- (3) 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.
- (4) Miyamoto, S.; Kollman, P. A. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (5) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (6) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (7) 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.; al, E. 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.
- (8) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

- (9) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (10) 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.
- (11) 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.
- (12) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (13) 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.
- (14) Silverman, R. B.; Groziak, M. P. Model Chemistry for a Covalent Mechanism of Action of Orotidine 5'-Phosphate Decarboxylase. *J. Am. Chem. Soc.* **1982**, *104*, 6434–6439.
- (15) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. Development and assessment of new exchange-correlation functionals. *J. Chem. Phys.* **1998**, *109*, 6264–6271.
- (16) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (17) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

- (18) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems.
J. Comput. Chem. **2013**, *34*, 1672–1685.

reactant.zip

AP1.lib

```

!!index array str
"AP1"
!entry.AP1.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"N" "N" 0 1 131072 1 7 -0.516300
"H" "H" 0 1 131072 2 1 0.307524
"CA" "CX" 0 1 131072 3 6 0.038100
"HA" "H1" 0 1 131072 4 1 0.046889
"CB" "2C" 0 1 131072 5 -1 0.061640
"HB2" "HC" 0 1 131072 6 1 0.015088
"HB3" "HC" 0 1 131072 7 1 0.015088
"CG" "CO" 0 1 131072 8 6 0.548084
"OD1" "Y4" 0 1 131072 9 39 -0.424539
"OD2" "O2" 0 1 131072 10 8 -0.543024
"C" "C" 0 1 131072 11 6 0.536600
"O" "O" 0 1 131072 12 8 -0.581900
!entry.AP1.unit.atomsptinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "2C" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CO" 0 -1 0.0
"OD1" "Y4" 0 -1 0.0
"OD2" "O2" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.AP1.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.AP1.unit.childsequence single int
2
!entry.AP1.unit.connect array int
1
11
!entry.AP1.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 11 1
5 6 1
5 7 1
5 8 1
8 9 1
8 10 1
11 12 1
!entry.AP1.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5

```

```

"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
!entry.AP1.unit.name single str
"AP1"
!entry.AP1.unit.positions table  dbl x  dbl y  dbl z
107.444000 27.817000 -45.958000
106.795000 27.273000 -46.508000
107.622000 27.466000 -44.544000
107.019000 26.589000 -44.309000
109.087000 27.151000 -44.232000
109.639000 27.054000 -45.167000
109.480000 27.995000 -43.665000
109.242000 25.887000 -43.423000
108.387000 25.605000 -42.550000
110.216000 25.158000 -43.684000
107.079000 28.515000 -43.570000
107.516000 28.596000 -42.412000
!entry.AP1.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x  int
c5x  int c6x
0 0 0 0 0 0
!entry.AP1.unit.residues table  str name  int seq  int childseq  int startatomx
str restype  int imagingx
"AP1" 1 13 1 "?" 0
!entry.AP1.unit.residuesPdbSequenceNumber array int
0
!entry.AP1.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.AP1.unit.velocities table  dbl x  dbl y  dbl z
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

```

HD1.lib

```

!!index array str
"HD1"
!entry.HD1.unit.atoms table  str name  str type  int typex  int resx  int flags
int seq  int elmnt  dbl chg
"N" "N" 0 1 131072 1 7 -0.415700
"H" "H" 0 1 131072 2 1 0.279411
"CA" "CX" 0 1 131072 3 6 0.018800
"HA" "H1" 0 1 131072 4 1 0.060456
"CB" "CT" 0 1 131072 5 6 -0.065319
"HB2" "HC" 0 1 131072 6 1 0.055017
"HB3" "HC" 0 1 131072 7 1 0.055017
"CG" "CC" 0 1 131072 8 6 0.034479
"ND1" "NA" 0 1 131072 9 7 -0.226513

```

```

"HD1" "H" 0 1 131072 10 1 0.327899
"CE1" "CR" 0 1 131072 11 6 -0.076611
"HE1" "H5" 0 1 131072 12 1 0.196130
"NE2" "Y1" 0 1 131072 13 39 0.006025
"CD2" "CV" 0 1 131072 14 6 -0.117407
"HD2" "H4" 0 1 131072 15 1 0.048910
"C" "C" 0 1 131072 16 6 0.597300
"O" "O" 0 1 131072 17 8 -0.567900
!entry.HD1.unit.atomsperinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "CT" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CC" 0 -1 0.0
"ND1" "NA" 0 -1 0.0
"HD1" "H" 0 -1 0.0
"CE1" "CR" 0 -1 0.0
"HE1" "H5" 0 -1 0.0
"NE2" "Y1" 0 -1 0.0
"CD2" "CV" 0 -1 0.0
"HD2" "H4" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.HD1.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD1.unit.childsequence single int
2
!entry.HD1.unit.connect array int
1
16
!entry.HD1.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 16 1
5 6 1
5 7 1
5 8 1
8 9 1
8 14 1
9 10 1
9 11 1
11 12 1
11 13 1
13 14 1
14 15 1
16 17 1
!entry.HD1.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7

```



```

!!index array str
"HD2"
!entry.HD2.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"N" "N" 0 1 131072 1 7 -0.415700
"H" "H" 0 1 131072 2 1 0.315732
"CA" "CX" 0 1 131072 3 6 0.018800
"HA" "H1" 0 1 131072 4 1 0.074548
"CB" "CT" 0 1 131072 5 6 0.006053
"HB2" "HC" 0 1 131072 6 1 0.064288
"HB3" "HC" 0 1 131072 7 1 0.064288
"CG" "CC" 0 1 131072 8 6 -0.108807
"ND1" "NA" 0 1 131072 9 7 -0.106727
"HD1" "H" 0 1 131072 10 1 0.288428
"CE1" "CR" 0 1 131072 11 6 -0.109303
"HE1" "H5" 0 1 131072 12 1 0.164624
"NE2" "Y2" 0 1 131072 13 39 -0.080737
"CD2" "CV" 0 1 131072 14 6 -0.072529
"HD2" "H4" 0 1 131072 15 1 0.157750
"C" "C" 0 1 131072 16 6 0.597300
"O" "O" 0 1 131072 17 8 -0.567900
!entry.HD2.unit.atomsptinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "CT" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CC" 0 -1 0.0
"ND1" "NA" 0 -1 0.0
"HD1" "H" 0 -1 0.0
"CE1" "CR" 0 -1 0.0
"HE1" "H5" 0 -1 0.0
"NE2" "Y2" 0 -1 0.0
"CD2" "CV" 0 -1 0.0
"HD2" "H4" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.HD2.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD2.unit.childsequence single int
2
!entry.HD2.unit.connect array int
1
16
!entry.HD2.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 16 1
5 6 1
5 7 1
5 8 1
8 9 1
8 14 1
9 10 1
9 11 1
11 12 1
11 13 1

```

```

13 14 1
14 15 1
16 17 1
!entry.HD2.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
!entry.HD2.unit.name single str
"HD2"
!entry.HD2.unit.positions table dbl x dbl y dbl z
102.372000 26.402000 -42.470000
102.935000 27.238000 -42.402000
101.866000 25.806000 -41.237000
101.273000 24.920000 -41.464000
103.029000 25.397000 -40.349000
103.603000 26.271000 -40.041000
102.648000 24.884000 -39.466000
104.000000 24.455000 -41.008000
103.643000 23.230000 -41.438000
102.713000 22.843000 -41.364000
104.717000 22.615000 -41.975000
104.640000 21.607000 -42.383000
105.764000 23.453000 -41.888000
105.358000 24.594000 -41.291000
106.069000 25.403000 -41.120000
100.953000 26.685000 -40.431000
100.944000 27.912000 -40.583000
!entry.HD2.unit.residueconnect table int c1x int c2x int c3x int c4x int
c5x int c6x
0 0 0 0 0
!entry.HD2.unit.residues table str name int seq int childseq int startatomx
str restype int imagingx
"HD2" 1 18 1 "?" 0
!entry.HD2.unit.residuesPdbSequenceNumber array int
0
!entry.HD2.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD2.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

```

```

0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

```

HD3.lib

```

!!index array str
"HD3"
!entry.HD3.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"N" "N" 0 1 131072 1 7 -0.415700
"H" "H" 0 1 131072 2 1 0.291533
"CA" "CX" 0 1 131072 3 6 0.018800
"HA" "H1" 0 1 131072 4 1 0.064505
"CB" "CT" 0 1 131072 5 6 0.004339
"HB2" "HC" 0 1 131072 6 1 0.042852
"HB3" "HC" 0 1 131072 7 1 0.042852
"CG" "CC" 0 1 131072 8 6 -0.005839
"ND1" "NA" 0 1 131072 9 7 -0.180240
"HD1" "H" 0 1 131072 10 1 0.339328
"CE1" "CR" 0 1 131072 11 6 -0.017975
"HE1" "H5" 0 1 131072 12 1 0.147713
"NE2" "Y3" 0 1 131072 13 39 -0.116107
"CD2" "CV" 0 1 131072 14 6 -0.158255
"HD2" "H4" 0 1 131072 15 1 0.140386
"C" "C" 0 1 131072 16 6 0.597300
"O" "O" 0 1 131072 17 8 -0.567900
!entry.HD3.unit.atomsperinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "CT" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CC" 0 -1 0.0
"ND1" "NA" 0 -1 0.0
"HD1" "H" 0 -1 0.0
"CE1" "CR" 0 -1 0.0
"HE1" "H5" 0 -1 0.0
"NE2" "Y3" 0 -1 0.0
"CD2" "CV" 0 -1 0.0
"HD2" "H4" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.HD3.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD3.unit.childsequence single int
2
!entry.HD3.unit.connect array int
1
16
!entry.HD3.unit.connectivity table int atom1x int atom2x int flags
1 2 1

```

```

1 3 1
3 4 1
3 5 1
3 16 1
5 6 1
5 7 1
5 8 1
8 9 1
8 14 1
9 10 1
9 11 1
11 12 1
11 13 1
13 14 1
14 15 1
16 17 1
!entry.HD3.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
!entry.HD3.unit.name single str
"HD3"
!entry.HD3.unit.positions table dbl x dbl y dbl z
107.955000 15.437000 -42.957000
107.529000 14.761000 -42.338000
108.478000 16.663000 -42.382000
109.147000 17.160000 -43.084000
107.311000 17.603000 -42.069000
106.629000 17.674000 -42.916000
106.775000 17.216000 -41.203000
107.723000 19.013000 -41.727000
108.267000 19.339000 -40.537000
108.449000 18.683000 -39.791000
108.523000 20.664000 -40.509000
108.976000 21.127000 -39.633000
108.128000 21.192000 -41.681000
107.629000 20.201000 -42.456000
107.252000 20.427000 -43.454000
109.285000 16.345000 -41.149000
108.916000 15.460000 -40.373000
!entry.HD3.unit.residueconnect table int c1x int c2x int c3x int c4x int
c5x int c6x
0 0 0 0 0
!entry.HD3.unit.residues table str name int seq int childseq int startatomx
str restype int imagingx
"HD3" 1 18 1 "?" 0
!entry.HD3.unit.residuesPdbSequenceNumber array int
0
!entry.HD3.unit.solventcap array dbl
-1.000000
0.0

```



```

2
!entry .5U1.unit.connect array int
0
0
!entry .5U1.unit.connectivity table int atom1x int atom2x int flags
1 2 1
2 3 1
2 8 1
3 4 1
3 13 1
4 5 1
4 6 1
6 7 1
6 14 1
7 8 1
7 12 1
8 9 1
9 10 1
9 11 1
!entry .5U1.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
!entry .5U1.unit.name single str
"5U1"
!entry .5U1.unit.positions table dbl x dbl y dbl z
108.429000 23.690000 -40.002000
109.256000 24.256000 -39.283000
109.033000 24.247000 -37.898000
109.930000 24.891000 -37.046000
109.751000 24.889000 -35.826000
111.040000 25.535000 -37.593000
111.258000 25.538000 -38.895000
110.390000 24.910000 -39.791000
110.750000 24.936000 -41.145000
111.764000 25.592000 -41.511000
110.183000 24.199000 -41.989000
112.081000 26.016000 -39.203000
108.234000 23.781000 -37.518000
111.709000 26.016000 -37.027000
!entry .5U1.unit.residueconnect table int c1x int c2x int c3x int c4x int
c5x int c6x
0 0 0 0 0
!entry .5U1.unit.residues table str name int seq int childseq int startatomx
str retype int imagingx
"5U1" 1 15 1 "?" 0
!entry .5U1.unit.residuesPdbSequenceNumber array int
0
!entry .5U1.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0

```

```

!entry.5U1.unit.velocities table  dbl x  dbl y  dbl z
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

```

ZN1.lib

```

!!index array str
"ZN1"
!entry.ZN1.unit.atoms table  str name  str type  int typex  int resx  int flags
int seq  int elmnt  dbl chg
"ZN" "M1" 0 1 131072 1 -1 0.319919
!entry.ZN1.unit.atomsptinfo table  str pname  str ptype  int ptypex  int
pelmnt  dbl pchg
"ZN" "M1" 0 -1 0.0
!entry.ZN1.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.ZN1.unit.childsequence single int
2
!entry.ZN1.unit.connect array int
0
0
!entry.ZN1.unit.hierarchy table  str abovetype  int abovex  str belowtype  int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
!entry.ZN1.unit.name single str
"ZN1"
!entry.ZN1.unit.positions table  dbl x  dbl y  dbl z
108.141000 23.400000 -42.046000
!entry.ZN1.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x  int
c5x  int c6x
0 0 0 0 0 0
!entry.ZN1.unit.residues table  str name  int seq  int childseq  int startatomx
str restype  int imagingx
"ZN1" 1 2 1 "?" 0
!entry.ZN1.unit.residuesPdbSequenceNumber array int
0
!entry.ZN1.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.ZN1.unit.velocities table  dbl x  dbl y  dbl z
0.0 0.0 0.0

```

reactant_complex.frmod

REMARK GOES HERE, THIS FILE IS GENERATED BY MCPB.PY					
MASS					
M1	65.4		Zn ion		
Y1	14.01	0.530	sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)		
Y2	14.01	0.530	sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)		
Y3	14.01	0.530	sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)		
Y4	16.00	0.434	carboxyl and phosphate group oxygen		
Y5	16.00	0.434	Oxygen with one connected atom		
Y6	16.00	0.434	Oxygen with one connected atom		
BOND					
M1-Y5	18.4	2.1689	Created by Seminario method using MCPB.py		
M1-Y6	50.8	2.0269	Created by Seminario method using MCPB.py		
Y1-M1	31.9	2.1549	Created by Seminario method using MCPB.py		
Y2-M1	13.6	2.2841	Created by Seminario method using MCPB.py		
Y3-M1	10.4	2.3244	Created by Seminario method using MCPB.py		
Y4-M1	53.3	1.9943	Created by Seminario method using MCPB.py		
CO-Y4	656.0	1.2500			
CR-Y1	488.0	1.335	JCC,7,(1986),230; HIS		
CR-Y2	488.0	1.335	JCC,7,(1986),230; HIS		
CR-Y3	488.0	1.335	JCC,7,(1986),230; HIS		
Y1-CV	410.0	1.394	JCC,7,(1986),230; HIS		
Y2-CV	410.0	1.394	JCC,7,(1986),230; HIS		
Y3-CV	410.0	1.394	JCC,7,(1986),230; HIS		
Y5-c	637.7	1.2183	SOURCE1_SOURCE5 27083 0.0110		
c -Y6	637.7	1.2183	SOURCE1_SOURCE5 27083 0.0110		
ANGL					
CO-Y4-M1	43.16	130.44	Created by Seminario method using MCPB.py		
CR-Y1-M1	37.69	118.91	Created by Seminario method using MCPB.py		
CR-Y2-M1	57.02	127.23	Created by Seminario method using MCPB.py		
CR-Y3-M1	41.60	123.63	Created by Seminario method using MCPB.py		
M1-Y1-CV	46.93	127.57	Created by Seminario method using MCPB.py		
M1-Y2-CV	57.04	126.81	Created by Seminario method using MCPB.py		
M1-Y3-CV	47.02	129.58	Created by Seminario method using MCPB.py		
M1-Y5-c	7.97	117.23	Created by Seminario method using MCPB.py		
M1-Y6-c	38.35	132.75	Created by Seminario method using MCPB.py		
Y1-M1-Y2	31.15	91.92	Created by Seminario method using MCPB.py		
Y1-M1-Y3	24.04	89.07	Created by Seminario method using MCPB.py		
Y1-M1-Y4	17.52	89.20	Created by Seminario method using MCPB.py		
Y1-M1-Y5	28.58	171.60	Created by Seminario method using MCPB.py		
Y1-M1-Y6	31.98	86.48	Created by Seminario method using MCPB.py		
Y2-M1-Y3	32.07	82.84	Created by Seminario method using MCPB.py		
Y2-M1-Y4	40.97	87.96	Created by Seminario method using MCPB.py		
Y2-M1-Y5	49.90	94.65	Created by Seminario method using MCPB.py		
Y2-M1-Y6	39.69	164.78	Created by Seminario method using MCPB.py		
Y3-M1-Y4	36.99	170.57	Created by Seminario method using MCPB.py		
Y3-M1-Y5	28.14	86.56	Created by Seminario method using MCPB.py		
Y3-M1-Y6	31.15	82.01	Created by Seminario method using MCPB.py		
Y4-M1-Y5	20.22	96.25	Created by Seminario method using MCPB.py		
Y4-M1-Y6	38.83	107.14	Created by Seminario method using MCPB.py		
Y6-M1-Y5	52.19	85.81	Created by Seminario method using MCPB.py		
2C-CO-Y4	70.0	117.00			
CC-CV-Y1	70.0	120.00	AA his		
CC-CV-Y2	70.0	120.00	AA his		
CC-CV-Y3	70.0	120.00	AA his		
CR-Y1-CV	70.0	117.00	AA his		
CR-Y2-CV	70.0	117.00	AA his		
CR-Y3-CV	70.0	117.00	AA his		
NA-CR-Y1	70.0	120.00	AA his		
NA-CR-Y2	70.0	120.00	AA his		
NA-CR-Y3	70.0	120.00	AA his		
O2-CO-Y4	80.0	126.00			
Y1-CR-H5	50.0	120.00	AA his		
Y1-CV-H4	50.0	120.00	AA his		

Y2-CR-H5	50.0	120.00	AA his				
Y2-CV-H4	50.0	120.00	AA his				
Y3-CR-H5	50.0	120.00	AA his				
Y3-CV-H4	50.0	120.00	AA his				
Y5-c -cd	69.14	123.93		SOURCE3	SOURCE5	3463	2.3073
Y5-c -n	74.22	123.05		SOURCE3	SOURCE5	8454	1.5552
Y6-c -o	77.94	130.25		SOURCE4	SOURCE5	1037	1.2396
cd-c -Y6	69.14	123.93		SOURCE3	SOURCE5	3463	2.3073
DIHE							
X -CR-Y1-X	2	10.0	180.0	2.0	JCC,7,(1986),	230	
X -CR-Y2-X	2	10.0	180.0	2.0	JCC,7,(1986),	230	
X -CR-Y3-X	2	10.0	180.0	2.0	JCC,7,(1986),	230	
CC-CV-Y1-CR	2	4.8	180.0	2.0	JCC,7,(1986),	230	
CC-CV-Y2-CR	2	4.8	180.0	2.0	JCC,7,(1986),	230	
CC-CV-Y3-CR	2	4.8	180.0	2.0	JCC,7,(1986),	230	
CR-Y1-CV-H4	2	4.8	180.0	2.0	JCC,7,(1986),	230	
CR-Y2-CV-H4	2	4.8	180.0	2.0	JCC,7,(1986),	230	
CR-Y3-CV-H4	2	4.8	180.0	2.0	JCC,7,(1986),	230	
2C-CO-Y4-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CC-CV-Y1-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CC-CV-Y2-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CC-CV-Y3-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CO-Y4-M1-Y5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CO-Y4-M1-Y6	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y1-M1-Y2	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y1-M1-Y3	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y1-M1-Y4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y1-M1-Y5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y1-M1-Y6	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y2-M1-Y3	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y2-M1-Y4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y2-M1-Y5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y2-M1-Y6	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y3-M1-Y4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y3-M1-Y5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CR-Y3-M1-Y6	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
CX-2C-CO-Y4	1	0.031	180.0	-4.0			
CX-2C-CO-Y4	1	0.0	0.0	-3.0			
CX-2C-CO-Y4	1	0.769	180.0	-2.0			
CX-2C-CO-Y4	1	0.0	0.0	1.0			
M1-Y1-CR-H5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
M1-Y1-CV-H4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
M1-Y2-CR-H5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
M1-Y2-CV-H4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
M1-Y3-CR-H5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
M1-Y3-CV-H4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
M1-Y5-c -cd	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
M1-Y5-c -n	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
M1-Y6-c -cd	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
M1-Y6-c -o	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
NA-CR-Y1-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
NA-CR-Y2-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
NA-CR-Y3-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
O2-CO-Y4-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
Y1-M1-Y2-CR	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
Y1-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
Y1-M1-Y3-CR	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
Y1-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
Y1-M1-Y4-CO	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
Y1-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
Y1-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
Y2-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
Y2-M1-Y3-CR	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	
Y2-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by	MCPB.py	

Y2-M1-Y4-CO	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y4-CO	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-CO-2C-HC	1	0.0	0.0	2.0	
Y4-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-c -n -hn	1	2.5	180.0	-2.0	JCC,7,(1986),230
Y5-c -n -hn	1	2.0	0.0	1.0	J.C. cistrans-NMA DE
Y6-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
c -Y6-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
IMPR					
X -X -c -Y6		10.5	180.	2.	JCC,7,(1986),230
X -X -c -Y5		10.5	180.	2.	JCC,7,(1986),230
X -O2-CO-Y4		10.5	180.	2.	
X -Y6-c -o		1.1	180.	2.	JCC,7,(1986),230
Y6-cd-c -o		1.1	180.0	2.0	General improper
torsional angle (1 general atom type)					
Y5-cd-c -n		10.5	180.0	2.0	General improper
torsional angle (2 general atom types)					
NONB					
M1	1.3950	0.0149170000			IOD set for Zn2+ ion from Li et al.
JCTC, 2013, 9, 2733					
Y1	1.8240	0.1700			OPLS
Y2	1.8240	0.1700			OPLS
Y3	1.8240	0.1700			OPLS
Y4	1.6612	0.2100			OPLS
Y5	1.6612	0.2100			OPLS
Y6	1.6612	0.2100			OPLS

B2.zip

AP2.lib

```
!!index array str
"AP2"
!entry.AP2.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"N" "N" 0 1 131072 1 7 -0.516300
"H" "H" 0 1 131072 2 1 0.222972
"CA" "CX" 0 1 131072 3 6 0.038100
"HA" "H1" 0 1 131072 4 1 0.483111
"CB" "2C" 0 1 131072 5 -1 -0.036112
"HB2" "HC" 0 1 131072 6 1 0.013225
"HB3" "HC" 0 1 131072 7 1 0.013225
"CG" "CO" 0 1 131072 8 6 0.520184
"OD1" "Y4" 0 1 131072 9 39 -0.348232
"OD2" "O2" 0 1 131072 10 8 -0.567590
"C" "C" 0 1 131072 11 6 0.536600
```

```

"O" "O" 0 1 131072 12 8 -0.581900
!entry.AP2.unit.atomsptinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "2C" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CO" 0 -1 0.0
"OD1" "Y4" 0 -1 0.0
"OD2" "O2" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.AP2.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.AP2.unit.childsequence single int
2
!entry.AP2.unit.connect array int
1
11
!entry.AP2.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 11 1
5 6 1
5 7 1
5 8 1
8 9 1
8 10 1
11 12 1
!entry.AP2.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
!entry.AP2.unit.name single str
"AP2"
!entry.AP2.unit.positions table dbl x dbl y dbl z
16.299000 -7.957000 -7.712000
16.914000 -7.154000 -7.784000
16.072000 -8.547000 -6.401000
16.638000 -7.886000 -5.717000
14.617000 -8.492000 -5.932000
14.191000 -7.539000 -6.288000
14.017000 -9.305000 -6.367000
14.405000 -8.496000 -4.418000
15.346000 -8.050000 -3.683000
13.288000 -8.872000 -4.013000

```

```

16.735000 -9.916000 -6.153000
16.283000 -10.669000 -5.304000
!entry.AP2.unit.residueconnect table int clx int c2x int c3x int c4x int
c5x int c6x
0 0 0 0 0 0
!entry.AP2.unit.residues table str name int seq int childseq int startatomx
str retype int imagingx
"AP2" 1 13 1 "?" 0
!entry.AP2.unit.residuesPdbSequenceNumber array int
0
!entry.AP2.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.AP2.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

```

HD4.lib

```

!!index array str
"HD4"
!entry.HD4.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"N" "N" 0 1 131072 1 7 -0.415700
"H" "H" 0 1 131072 2 1 0.276993
"CA" "CX" 0 1 131072 3 6 0.018800
"HA" "H1" 0 1 131072 4 1 0.091424
"CB" "CT" 0 1 131072 5 6 -0.115147
"HB2" "HC" 0 1 131072 6 1 0.073174
"HB3" "HC" 0 1 131072 7 1 0.073174
"CG" "CC" 0 1 131072 8 6 0.073046
"ND1" "NA" 0 1 131072 9 7 -0.293159
"HD1" "H" 0 1 131072 10 1 0.343402
"CE1" "CR" 0 1 131072 11 6 -0.066129
"HE1" "H5" 0 1 131072 12 1 0.153000
"NE2" "Y1" 0 1 131072 13 39 0.252200
"CD2" "CV" 0 1 131072 14 6 0.053650
"HD2" "H4" 0 1 131072 15 1 -0.466573
"C" "C" 0 1 131072 16 6 0.597300
"O" "O" 0 1 131072 17 8 -0.567900
!entry.HD4.unit.atomsPertInfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "CT" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CC" 0 -1 0.0
"ND1" "NA" 0 -1 0.0
"HD1" "H" 0 -1 0.0

```

```

"CE1" "CR" 0 -1 0.0
"HE1" "H5" 0 -1 0.0
"NE2" "Y1" 0 -1 0.0
"CD2" "CV" 0 -1 0.0
"HD2" "H4" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.HD4.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD4.unit.childsequence single int
2
!entry.HD4.unit.connect array int
1
16
!entry.HD4.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 16 1
5 6 1
5 7 1
5 8 1
8 9 1
8 14 1
9 10 1
9 11 1
11 12 1
11 13 1
13 14 1
14 15 1
16 17 1
!entry.HD4.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
!entry.HD4.unit.name single str
"HD4"
!entry.HD4.unit.positions table dbl x dbl y dbl z
18.763000 -2.872000 -7.414000
18.757000 -3.302000 -8.334000
18.021000 -3.534000 -6.333000
18.209000 -2.992000 -5.394000
16.510000 -3.501000 -6.585000
16.198000 -2.471000 -6.809000
16.289000 -4.131000 -7.461000

```



```

"O" "O" 0 1 131072 17 8 -0.567900
!entry.HD5.unit.atomsptinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "CT" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CC" 0 -1 0.0
"ND1" "NA" 0 -1 0.0
"HD1" "H" 0 -1 0.0
"CE1" "CR" 0 -1 0.0
"HE1" "H5" 0 -1 0.0
"NE2" "Y2" 0 -1 0.0
"CD2" "CV" 0 -1 0.0
"HD2" "H4" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.HD5.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD5.unit.childsequence single int
2
!entry.HD5.unit.connect array int
1
16
!entry.HD5.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 16 1
5 6 1
5 7 1
5 8 1
8 9 1
8 14 1
9 10 1
9 11 1
11 12 1
11 13 1
13 14 1
14 15 1
16 17 1
!entry.HD5.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14

```



```

"HA" "H1" 0 1 131072 4 1 0.045575
"CB" "CT" 0 1 131072 5 6 0.062437
"HB2" "HC" 0 1 131072 6 1 0.026213
"HB3" "HC" 0 1 131072 7 1 0.026213
"CG" "CC" 0 1 131072 8 6 0.029350
"ND1" "NA" 0 1 131072 9 7 -0.362165
"HD1" "H" 0 1 131072 10 1 0.402905
"CE1" "CR" 0 1 131072 11 6 0.133832
"HE1" "H5" 0 1 131072 12 1 0.085493
"NE2" "Y3" 0 1 131072 13 39 -0.047658
"CD2" "CV" 0 1 131072 14 6 -0.222451
"HD2" "H4" 0 1 131072 15 1 0.139429
"C" "C" 0 1 131072 16 6 0.597300
"O" "O" 0 1 131072 17 8 -0.567900
!entry.HD6.unit.atomsperinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "CT" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CC" 0 -1 0.0
"ND1" "NA" 0 -1 0.0
"HD1" "H" 0 -1 0.0
"CE1" "CR" 0 -1 0.0
"HE1" "H5" 0 -1 0.0
"NE2" "Y3" 0 -1 0.0
"CD2" "CV" 0 -1 0.0
"HD2" "H4" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.HD6.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD6.unit.childsequence single int
2
!entry.HD6.unit.connect array int
1
16
!entry.HD6.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 16 1
5 6 1
5 7 1
5 8 1
8 9 1
8 14 1
9 10 1
9 11 1
11 12 1
11 13 1
13 14 1
14 15 1
16 17 1
!entry.HD6.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1

```


0.0 0.0 0.0

5U2.lib

```
!!index array str
"5U2"
!entry.5U2.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"O4" "Y5" 0 1 131072 1 39 -0.396703
"C4" "c" 0 1 131072 2 6 0.385699
"N3" "n" 0 1 131072 3 7 -0.278772
"C2" "c" 0 1 131072 4 6 0.575752
"O2" "o" 0 1 131072 5 8 -0.546921
"N1" "n" 0 1 131072 6 7 -0.677770
"C6" "c3" 0 1 131072 7 6 0.490804
"C5" "c3" 0 1 131072 8 6 0.020273
"C" "c" 0 1 131072 9 6 0.476458
"O53" "o" 0 1 131072 10 8 -0.536126
"O52" "Y6" 0 1 131072 11 39 -0.497238
"H1" "h2" 0 1 131072 12 1 -0.013450
"H2" "hn" 0 1 131072 13 1 0.254951
"H3" "hn" 0 1 131072 14 1 0.360375
"O" "oh" 0 1 131072 15 8 -0.590050
"H11" "ho" 0 1 131072 16 1 0.385678
"H21" "hc" 0 1 131072 17 1 0.094243
!entry.5U2.unit.atomsptinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"O4" "Y5" 0 -1 0.0
"C4" "c" 0 -1 0.0
"N3" "n" 0 -1 0.0
"C2" "c" 0 -1 0.0
"O2" "o" 0 -1 0.0
"N1" "n" 0 -1 0.0
"C6" "c3" 0 -1 0.0
"C5" "c3" 0 -1 0.0
"C" "c" 0 -1 0.0
"O53" "o" 0 -1 0.0
"O52" "Y6" 0 -1 0.0
"H1" "h2" 0 -1 0.0
"H2" "hn" 0 -1 0.0
"H3" "hn" 0 -1 0.0
"O" "oh" 0 -1 0.0
"H11" "ho" 0 -1 0.0
"H21" "hc" 0 -1 0.0
!entry.5U2.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.5U2.unit.childsequence single int
2
!entry.5U2.unit.connect array int
0
0
!entry.5U2.unit.connectivity table int atom1x int atom2x int flags
1 2 1
2 3 1
2 8 1
3 4 1
3 13 1
4 5 1
4 6 1
6 7 1
6 14 1
```

```

7 8 1
7 12 1
7 15 1
8 9 1
8 17 1
9 10 1
9 11 1
15 16 1
!entry .5U2.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
!entry .5U2.unit.name single str
"5U2"
!entry .5U2.unit.positions table dbl x dbl y dbl z
15.210000 -8.181000 -0.553000
14.233000 -8.878000 -0.261000
14.290000 -9.615000 0.896000
13.305000 -10.466000 1.375000
13.480000 -11.109000 2.417000
12.164000 -10.502000 0.655000
11.986000 -10.048000 -0.728000
13.005000 -8.971000 -1.097000
12.330000 -7.589000 -1.439000
11.100000 -7.548000 -1.436000
13.113000 -6.642000 -1.763000
10.995000 -9.580000 -0.793000
15.155000 -9.556000 1.452000
11.437000 -11.098000 1.042000
12.118000 -11.116000 -1.617000
11.275000 -11.516000 -1.894000
13.352000 -9.232000 -2.119000
!entry .5U2.unit.residueconnect table int c1x int c2x int c3x int c4x int
c5x int c6x
0 0 0 0 0 0
!entry .5U2.unit.residues table str name int seq int childseq int startatomx
str restype int imagingx
"5U2" 1 18 1 "?" 0
!entry .5U2.unit.residuesPdbSequenceNumber array int
0
!entry .5U2.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry .5U2.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

```

```
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
```

ZN2.lib

```
!!index array str
"ZN2"
!entry.ZN2.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"ZN" "M1" 0 1 131072 1 -1 0.049551
!entry.ZN2.unit.atomsperinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"ZN" "M1" 0 -1 0.0
!entry.ZN2.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.ZN2.unit.childsequence single int
2
!entry.ZN2.unit.connect array int
0
0
!entry.ZN2.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
!entry.ZN2.unit.name single str
"ZN2"
!entry.ZN2.unit.positions table dbl x dbl y dbl z
15.166000 -6.666000 -2.115000
!entry.ZN2.unit.residueconnect table int c1x int c2x int c3x int c4x int
c5x int c6x
0 0 0 0 0 0
!entry.ZN2.unit.residues table str name int seq int childseq int startatomx
str restype int imagingx
"ZN2" 1 2 1 "?" 0
!entry.ZN2.unit.residuesPdbSequenceNumber array int
0
!entry.ZN2.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.ZN2.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0
```

B2.frmod

REMARK GOES HERE, THIS FILE IS GENERATED BY MCPB.PY					
MASS					
M1	65.4			Zn ion	
Y1	14.01	0.530		sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)	
Y2	14.01	0.530		sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)	
Y3	14.01	0.530		sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)	
Y4	16.00	0.434		carboxyl and phosphate group oxygen	
Y5	16.00	0.434		Oxygen with one connected atom	
Y6	16.00	0.434		Oxygen with one connected atom	
BOND					
M1-Y5	4.0	2.4197		Created by Seminario method using MCPB.py	
M1-Y6	58.6	2.0309		Created by Seminario method using MCPB.py	
Y1-M1	39.1	2.1505		Created by Seminario method using MCPB.py	
Y2-M1	50.3	2.1142		Created by Seminario method using MCPB.py	
Y3-M1	19.5	2.2407		Created by Seminario method using MCPB.py	
Y4-M1	22.4	2.0838		Created by Seminario method using MCPB.py	
CO-Y4	656.0	1.2500			
CR-Y1	488.0	1.335		JCC,7,(1986),230; HIS	
CR-Y2	488.0	1.335		JCC,7,(1986),230; HIS	
CR-Y3	488.0	1.335		JCC,7,(1986),230; HIS	
Y1-CV	410.0	1.394		JCC,7,(1986),230; HIS	
Y2-CV	410.0	1.394		JCC,7,(1986),230; HIS	
Y3-CV	410.0	1.394		JCC,7,(1986),230; HIS	
Y5-c	637.7	1.2183		SOURCE1_SOURCE5	27083 0.0110
c -Y6	637.7	1.2183		SOURCE1_SOURCE5	27083 0.0110
ANGL					
CO-Y4-M1	35.45	140.67		Created by Seminario method using MCPB.py	
CR-Y1-M1	48.85	117.67		Created by Seminario method using MCPB.py	
CR-Y2-M1	36.16	127.87		Created by Seminario method using MCPB.py	
CR-Y3-M1	50.69	119.06		Created by Seminario method using MCPB.py	
M1-Y1-CV	50.14	131.61		Created by Seminario method using MCPB.py	
M1-Y2-CV	39.58	124.69		Created by Seminario method using MCPB.py	
M1-Y3-CV	50.66	127.23		Created by Seminario method using MCPB.py	
M1-Y5-c	49.90	112.70		Created by Seminario method using MCPB.py	
M1-Y6-c	49.84	126.74		Created by Seminario method using MCPB.py	
Y1-M1-Y2	24.77	103.25		Created by Seminario method using MCPB.py	
Y1-M1-Y3	28.10	90.54		Created by Seminario method using MCPB.py	
Y1-M1-Y4	27.77	91.34		Created by Seminario method using MCPB.py	
Y1-M1-Y5	25.61	168.62		Created by Seminario method using MCPB.py	
Y1-M1-Y6	33.44	88.90		Created by Seminario method using MCPB.py	
Y2-M1-Y3	32.69	89.20		Created by Seminario method using MCPB.py	
Y2-M1-Y4	34.99	85.88		Created by Seminario method using MCPB.py	
Y2-M1-Y5	25.83	86.78		Created by Seminario method using MCPB.py	
Y2-M1-Y6	30.13	166.42		Created by Seminario method using MCPB.py	
Y3-M1-Y4	38.87	175.02		Created by Seminario method using MCPB.py	
Y3-M1-Y5	29.08	84.16		Created by Seminario method using MCPB.py	
Y3-M1-Y6	40.40	84.59		Created by Seminario method using MCPB.py	
Y4-M1-Y5	27.60	94.77		Created by Seminario method using MCPB.py	
Y4-M1-Y6	45.09	100.05		Created by Seminario method using MCPB.py	
Y6-M1-Y5	38.16	80.59		Created by Seminario method using MCPB.py	
2C-CO-Y4	70.0	117.00			
CC-CV-Y1	70.0	120.00		AA his	
CC-CV-Y2	70.0	120.00		AA his	
CC-CV-Y3	70.0	120.00		AA his	
CR-Y1-CV	70.0	117.00		AA his	
CR-Y2-CV	70.0	117.00		AA his	
CR-Y3-CV	70.0	117.00		AA his	
NA-CR-Y1	70.0	120.00		AA his	
NA-CR-Y2	70.0	120.00		AA his	
NA-CR-Y3	70.0	120.00		AA his	
O2-CO-Y4	80.0	126.00			
Y1-CR-H5	50.0	120.00		AA his	
Y1-CV-H4	50.0	120.00		AA his	

Y2-CR-H5	50.0	120.00	AA his					
Y2-CV-H4	50.0	120.00	AA his					
Y3-CR-H5	50.0	120.00	AA his					
Y3-CV-H4	50.0	120.00	AA his					
Y5-c -c3	67.40	123.20		SOURCE3	SOURCE5	10083	1.8011	
Y5-c -n	74.22	123.05		SOURCE3	SOURCE5	8454	1.5552	
Y6-c -o	77.94	130.25		SOURCE4	SOURCE5	1037	1.2396	
c3-c -Y6	67.40	123.20		SOURCE3	SOURCE5	10083	1.8011	
DIHE								
X -CR-Y1-X	2	10.0	180.0	2.0	JCC,7,(1986),	230		
X -CR-Y2-X	2	10.0	180.0	2.0	JCC,7,(1986),	230		
X -CR-Y3-X	2	10.0	180.0	2.0	JCC,7,(1986),	230		
X -CV-Y1-X	2	4.8	180.0	2.0	JCC,7,(1986),	230		
X -CV-Y2-X	2	4.8	180.0	2.0	JCC,7,(1986),	230		
X -CV-Y3-X	2	4.8	180.0	2.0	JCC,7,(1986),	230		
2C-CO-Y4-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CC-CV-Y1-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CC-CV-Y2-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CC-CV-Y3-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CO-Y4-M1-Y5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CO-Y4-M1-Y6	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y1-M1-Y2	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y1-M1-Y3	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y1-M1-Y4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y1-M1-Y5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y1-M1-Y6	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y2-M1-Y3	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y2-M1-Y4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y2-M1-Y5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y2-M1-Y6	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y3-M1-Y4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y3-M1-Y5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CR-Y3-M1-Y6	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
CX-2C-CO-Y4	1	0.031	180.0	-4.0				
CX-2C-CO-Y4	1	0.0	0.0	-3.0				
CX-2C-CO-Y4	1	0.769	180.0	-2.0				
CX-2C-CO-Y4	1	0.0	0.0	1.0				
M1-Y1-CR-H5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
M1-Y1-CV-H4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
M1-Y2-CR-H5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
M1-Y2-CV-H4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
M1-Y3-CR-H5	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
M1-Y3-CV-H4	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
M1-Y5-c -c3	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
M1-Y5-c -n	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
M1-Y6-c -c3	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
M1-Y6-c -o	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
NA-CR-Y1-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
NA-CR-Y2-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
NA-CR-Y3-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
O2-CO-Y4-M1	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y1-M1-Y2-CR	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y1-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y1-M1-Y3-CR	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y1-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y1-M1-Y4-CO	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y1-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y1-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y2-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y2-M1-Y3-CR	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y2-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y2-M1-Y4-CO	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y2-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y2-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		
Y3-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by	MCPB.py		

Y3-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y4-CO	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-CO-2C-HC	1	0.0	0.0	2.0	
Y4-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-c -c3-hc	1	0.8	0.0	-1.0	Junmei et al, 1999
Y5-c -c3-hc	1	0.0	0.0	-2.0	Junmei et al, 1999
Y5-c -c3-hc	1	0.08	180.0	3.0	Junmei et al, 1999
Y5-c -n -hn	1	2.5	180.0	-2.0	JCC,7,(1986),230
Y5-c -n -hn	1	2.0	0.0	1.0	J.C.cistrans-NMA DE
Y6-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
c -Y6-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
hc-c3-c -Y6	1	0.8	0.0	-1.0	Junmei et al, 1999
hc-c3-c -Y6	1	0.0	0.0	-2.0	Junmei et al, 1999
hc-c3-c -Y6	1	0.08	180.0	3.0	Junmei et al, 1999
IMPR					
X -X -c -Y6		10.5	180.	2.	JCC,7,(1986),230
X -X -c -Y5		10.5	180.	2.	JCC,7,(1986),230
X -O2-CO-Y4		10.5	180.	2.	
X -Y6-c -o		1.1	180.	2.	JCC,7,(1986),230
Y6-c3-c -o		1.1	180.0	2.0	General improper
torsional angle (1 general atom type)					
Y5-c3-c -n		10.5	180.0	2.0	General improper
torsional angle (2 general atom types)					
NONB					
M1	1.3950	0.0149170000			IOD set for Zn2+ ion from Li et al.
JCTC, 2013, 9, 2733					
Y1	1.8240	0.1700			OPLS
Y2	1.8240	0.1700			OPLS
Y3	1.8240	0.1700			OPLS
Y4	1.6612	0.2100			OPLS
Y5	1.6612	0.2100			OPLS
Y6	1.6612	0.2100			OPLS

C1.zip

AH1.lib

```
!!index array str
"AH1"
!entry.AH1.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"N" "N" 0 1 131072 1 7 -0.415700
"H" "H" 0 1 131072 2 1 0.256851
"CA" "CX" 0 1 131072 3 6 0.034100
"HA" "H1" 0 1 131072 4 1 0.144785
"CB" "2C" 0 1 131072 5 -1 -0.074705
"HB2" "HC" 0 1 131072 6 1 0.076313
"HB3" "HC" 0 1 131072 7 1 0.076313
"CG" "C" 0 1 131072 8 6 0.357534
"OD1" "Y4" 0 1 131072 9 39 -0.306676
```

```

"OD2" "OH" 0 1 131072 10 8 -0.331287
"HD2" "HO" 0 1 131072 11 1 0.396906
"C" "C" 0 1 131072 12 6 0.597300
"O" "O" 0 1 131072 13 8 -0.567900
!entry.AH1.unit.atomsperinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "2C" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "C" 0 -1 0.0
"OD1" "Y4" 0 -1 0.0
"OD2" "OH" 0 -1 0.0
"HD2" "HO" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.AH1.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.AH1.unit.childsequence single int
2
!entry.AH1.unit.connect array int
1
12
!entry.AH1.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 12 1
5 6 1
5 7 1
5 8 1
8 9 1
8 10 1
10 11 1
12 13 1
!entry.AH1.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
!entry.AH1.unit.name single str
"AH1"
!entry.AH1.unit.positions table dbl x dbl y dbl z
-15.647000 15.050000 -2.959000
-16.296000 14.506000 -3.509000
-15.469000 14.699000 -1.545000
-16.072000 13.822000 -1.310000

```

```

-14.004000 14.384000 -1.233000
-13.452000 14.287000 -2.168000
-13.611000 15.228000 -0.666000
-13.849000 13.120000 -0.424000
-14.704000 12.838000 0.449000
-12.875000 12.391000 -0.685000
-12.347000 12.806000 -1.426000
-16.012000 15.748000 -0.571000
-15.575000 15.829000 0.587000
!entry .AH1.unit.residueconnect table int c1x int c2x int c3x int c4x int
c5x int c6x
0 0 0 0 0 0
!entry .AH1.unit.residues table str name int seq int childseq int startatomx
str restype int imagingx
"AH1" 1 14 1 "?" 0
!entry .AH1.unit.residuesPdbSequenceNumber array int
0
!entry .AH1.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry .AH1.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

```

HD1.lib

```

!!index array str
"HD1"
!entry .HD1.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"N" "N" 0 1 131072 1 7 -0.415700
"H" "H" 0 1 131072 2 1 0.300950
"CA" "CX" 0 1 131072 3 6 0.018800
"HA" "H1" 0 1 131072 4 1 0.088638
"CB" "CT" 0 1 131072 5 6 -0.003876
"HB2" "HC" 0 1 131072 6 1 0.034179
"HB3" "HC" 0 1 131072 7 1 0.034179
"CG" "CC" 0 1 131072 8 6 -0.006438
"ND1" "NA" 0 1 131072 9 7 -0.142043
"HD1" "H" 0 1 131072 10 1 0.322852
"CE1" "CR" 0 1 131072 11 6 -0.168037
"HE1" "H5" 0 1 131072 12 1 0.214538
"NE2" "Y1" 0 1 131072 13 39 -0.022398
"CD2" "CV" 0 1 131072 14 6 -0.006285
"HD2" "H4" 0 1 131072 15 1 -0.086187
"C" "C" 0 1 131072 16 6 0.597300
"O" "O" 0 1 131072 17 8 -0.567900
!entry .HD1.unit.atomsptinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0

```

```

"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "CT" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CC" 0 -1 0.0
"ND1" "NA" 0 -1 0.0
"HD1" "H" 0 -1 0.0
"CE1" "CR" 0 -1 0.0
"HE1" "H5" 0 -1 0.0
"NE2" "Y1" 0 -1 0.0
"CD2" "CV" 0 -1 0.0
"HD2" "H4" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.HD1.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD1.unit.childsequence single int
2
!entry.HD1.unit.connect array int
1
16
!entry.HD1.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 16 1
5 6 1
5 7 1
5 8 1
8 9 1
8 14 1
9 10 1
9 11 1
11 12 1
11 13 1
13 14 1
14 15 1
16 17 1
!entry.HD1.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
!entry.HD1.unit.name single str
"HD1"

```



```

"ND1" "NA" 0 1 131072 9 7 -0.079637
"HD1" "H" 0 1 131072 10 1 0.279209
"CE1" "CR" 0 1 131072 11 6 -0.047942
"HE1" "H5" 0 1 131072 12 1 0.152869
"NE2" "Y2" 0 1 131072 13 39 -0.045535
"CD2" "CV" 0 1 131072 14 6 -0.118474
"HD2" "H4" 0 1 131072 15 1 0.184158
"C" "C" 0 1 131072 16 6 0.597300
"O" "O" 0 1 131072 17 8 -0.567900
!entry.HD2.unit.atomsperinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "CT" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CC" 0 -1 0.0
"ND1" "NA" 0 -1 0.0
"HD1" "H" 0 -1 0.0
"CE1" "CR" 0 -1 0.0
"HE1" "H5" 0 -1 0.0
"NE2" "Y2" 0 -1 0.0
"CD2" "CV" 0 -1 0.0
"HD2" "H4" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.HD2.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD2.unit.childsequence single int
2
!entry.HD2.unit.connect array int
1
16
!entry.HD2.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 16 1
5 6 1
5 7 1
5 8 1
8 9 1
8 14 1
9 10 1
9 11 1
11 12 1
11 13 1
13 14 1
14 15 1
16 17 1
!entry.HD2.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6

```



```

!!index array str
"HD3"
!entry.HD3.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"N" "N" 0 1 131072 1 7 -0.415700
"H" "H" 0 1 131072 2 1 0.306754
"CA" "CX" 0 1 131072 3 6 0.018800
"HA" "H1" 0 1 131072 4 1 0.061482
"CB" "CT" 0 1 131072 5 6 0.085771
"HB2" "HC" 0 1 131072 6 1 0.015053
"HB3" "HC" 0 1 131072 7 1 0.015053
"CG" "CC" 0 1 131072 8 6 0.026292
"ND1" "NA" 0 1 131072 9 7 -0.235137
"HD1" "H" 0 1 131072 10 1 0.386667
"CE1" "CR" 0 1 131072 11 6 -0.044730
"HE1" "H5" 0 1 131072 12 1 0.144740
"NE2" "Y3" 0 1 131072 13 39 0.090375
"CD2" "CV" 0 1 131072 14 6 -0.332675
"HD2" "H4" 0 1 131072 15 1 0.212094
"C" "C" 0 1 131072 16 6 0.597300
"O" "O" 0 1 131072 17 8 -0.567900
!entry.HD3.unit.atomsptinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"N" "N" 0 -1 0.0
"H" "H" 0 -1 0.0
"CA" "CX" 0 -1 0.0
"HA" "H1" 0 -1 0.0
"CB" "CT" 0 -1 0.0
"HB2" "HC" 0 -1 0.0
"HB3" "HC" 0 -1 0.0
"CG" "CC" 0 -1 0.0
"ND1" "NA" 0 -1 0.0
"HD1" "H" 0 -1 0.0
"CE1" "CR" 0 -1 0.0
"HE1" "H5" 0 -1 0.0
"NE2" "Y3" 0 -1 0.0
"CD2" "CV" 0 -1 0.0
"HD2" "H4" 0 -1 0.0
"C" "C" 0 -1 0.0
"O" "O" 0 -1 0.0
!entry.HD3.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD3.unit.childsequence single int
2
!entry.HD3.unit.connect array int
1
16
!entry.HD3.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 16 1
5 6 1
5 7 1
5 8 1
8 9 1
8 14 1
9 10 1
9 11 1
11 12 1
11 13 1

```

```

13 14 1
14 15 1
16 17 1
!entry.HD3.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
!entry.HD3.unit.name single str
"HD3"
!entry.HD3.unit.positions table dbl x dbl y dbl z
-15.136000 2.670000 0.042000
-15.562000 1.994000 0.661000
-14.613000 3.896000 0.617000
-13.944000 4.393000 -0.085000
-15.780000 4.836000 0.930000
-16.462000 4.907000 0.083000
-16.316000 4.449000 1.796000
-15.368000 6.246000 1.272000
-14.824000 6.572000 2.462000
-14.642000 5.916000 3.208000
-14.568000 7.897000 2.490000
-14.115000 8.360000 3.366000
-14.963000 8.425000 1.318000
-15.462000 7.434000 0.543000
-15.839000 7.660000 -0.455000
-13.806000 3.578000 1.850000
-14.175000 2.693000 2.626000
!entry.HD3.unit.residueconnect table int c1x int c2x int c3x int c4x int
c5x int c6x
0 0 0 0 0
!entry.HD3.unit.residues table str name int seq int childseq int startatomx
str restype int imagingx
"HD3" 1 18 1 "?" 0
!entry.HD3.unit.residuesPdbSequenceNumber array int
0
!entry.HD3.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.HD3.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

```

```

0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

```

5U1.lib

```

!!index array str
"5U1"
!entry .5U1.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"O4" "Y5" 0 1 131072 1 39 -0.392056
"C4" "c" 0 1 131072 2 6 0.299615
"N3" "n" 0 1 131072 3 7 -0.204974
"C2" "c" 0 1 131072 4 6 0.383000
"O2" "o" 0 1 131072 5 8 -0.463280
"N1" "n" 0 1 131072 6 7 -0.208096
"C6" "cc" 0 1 131072 7 6 -0.003759
"C5" "cd" 0 1 131072 8 6 0.059506
"C" "c" 0 1 131072 9 6 0.336093
"O53" "oh" 0 1 131072 10 8 -0.566925
"O52" "Y6" 0 1 131072 11 39 -0.372866
"H1" "h4" 0 1 131072 12 1 0.152480
"H2" "hn" 0 1 131072 13 1 0.286817
"H3" "hn" 0 1 131072 14 1 0.327955
!entry .5U1.unit.atomsptinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"O4" "Y5" 0 -1 0.0
"C4" "c" 0 -1 0.0
"N3" "n" 0 -1 0.0
"C2" "c" 0 -1 0.0
"O2" "o" 0 -1 0.0
"N1" "n" 0 -1 0.0
"C6" "cc" 0 -1 0.0
"C5" "cd" 0 -1 0.0
"C" "c" 0 -1 0.0
"O53" "oh" 0 -1 0.0
"O52" "Y6" 0 -1 0.0
"H1" "h4" 0 -1 0.0
"H2" "hn" 0 -1 0.0
"H3" "hn" 0 -1 0.0
!entry .5U1.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry .5U1.unit.childsequence single int
2
!entry .5U1.unit.connect array int
0
0
!entry .5U1.unit.connectivity table int atom1x int atom2x int flags
1 2 1
2 3 1
2 8 1
3 4 1
3 13 1
4 5 1
4 6 1

```

```

6 7 1
6 14 1
7 8 1
7 12 1
8 9 1
9 10 1
9 11 1
!entry .5U1.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
!entry .5U1.unit.name single str
"5U1"
!entry .5U1.unit.positions table dbl x dbl y dbl z
-14.662000 10.923000 2.997000
-13.835000 11.489000 3.716000
-14.058000 11.480000 5.101000
-13.161000 12.124000 5.953000
-13.340000 12.122000 7.173000
-12.051000 12.768000 5.406000
-11.833000 12.771000 4.104000
-12.701000 12.143000 3.208000
-12.341000 12.169000 1.854000
-11.327000 12.825000 1.488000
-12.908000 11.432000 1.010000
-11.010000 13.249000 3.796000
-14.857000 11.014000 5.481000
-11.382000 13.249000 5.972000
!entry .5U1.unit.residueconnect table int c1x int c2x int c3x int c4x int
c5x int c6x
0 0 0 0 0
!entry .5U1.unit.residues table str name int seq int childseq int startatomx
str restype int imagingx
"5U1" 1 15 1 "?" 0
!entry .5U1.unit.residuesPdbSequenceNumber array int
0
!entry .5U1.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry .5U1.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

```

```
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
```

ZN1.lib

```
!!index array str
"ZN1"
!entry.ZN1.unit.atoms table str name str type int typex int resx int flags
int seq int elmnt dbl chg
"ZN" "M1" 0 1 131072 1 -1 0.162282
!entry.ZN1.unit.atomsperinfo table str pname str ptype int ptypex int
pelmnt dbl pchg
"ZN" "M1" 0 -1 0.0
!entry.ZN1.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.ZN1.unit.childsequence single int
2
!entry.ZN1.unit.connect array int
0
0
!entry.ZN1.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
"U" 0 "R" 1
"R" 1 "A" 1
!entry.ZN1.unit.name single str
"ZN1"
!entry.ZN1.unit.positions table dbl x dbl y dbl z
-14.950000 10.633000 0.953000
!entry.ZN1.unit.residueconnect table int c1x int c2x int c3x int c4x int
c5x int c6x
0 0 0 0 0 0
!entry.ZN1.unit.residues table str name int seq int childseq int startatomx
str restype int imagingx
"ZN1" 1 2 1 "?" 0
!entry.ZN1.unit.residuesPdbSequenceNumber array int
0
!entry.ZN1.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.ZN1.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0
```

C1.frmod

```
REMARK GOES HERE, THIS FILE IS GENERATED BY MCPB.PY
MASS
M1 65.4 Zn ion
Y1 14.01 0.530 sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)
Y2 14.01 0.530 sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)
Y3 14.01 0.530 sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)
Y4 16.00 0.434 carbonyl group oxygen
Y5 16.00 0.434 Oxygen with one connected atom
Y6 16.00 0.434 Oxygen with one connected atom
BOND
```

M1-Y5	6.3	2.3679	Created by Seminario method using MCPB.py				
M1-Y6	83.0	1.9622	Created by Seminario method using MCPB.py				
Y1-M1	54.4	2.0972	Created by Seminario method using MCPB.py				
Y2-M1	69.6	2.0525	Created by Seminario method using MCPB.py				
Y3-M1	58.0	2.0844	Created by Seminario method using MCPB.py				
Y4-M1	0.0	2.9169	Created by Seminario method using MCPB.py				
C -Y4	570.0	1.229	JCC,7,(1986),230; AA,CYT,GUA,THY,URA				
CR-Y1	488.0	1.335	JCC,7,(1986),230; HIS				
CR-Y2	488.0	1.335	JCC,7,(1986),230; HIS				
CR-Y3	488.0	1.335	JCC,7,(1986),230; HIS				
Y1-CV	410.0	1.394	JCC,7,(1986),230; HIS				
Y2-CV	410.0	1.394	JCC,7,(1986),230; HIS				
Y3-CV	410.0	1.394	JCC,7,(1986),230; HIS				
Y5-c	637.7	1.2183	SOURCE1 SOURCE5	27083	0.0110		
c -Y6	637.7	1.2183	SOURCE1_SOURCE5	27083	0.0110		
ANGL							
C -Y4-M1	34.15	116.63	Created by Seminario method using MCPB.py				
CR-Y1-M1	53.39	121.42	Created by Seminario method using MCPB.py				
CR-Y2-M1	51.17	123.08	Created by Seminario method using MCPB.py				
CR-Y3-M1	45.66	125.77	Created by Seminario method using MCPB.py				
M1-Y1-CV	51.72	130.83	Created by Seminario method using MCPB.py				
M1-Y2-CV	50.22	129.59	Created by Seminario method using MCPB.py				
M1-Y3-CV	47.68	126.88	Created by Seminario method using MCPB.py				
M1-Y5-c	35.73	120.42	Created by Seminario method using MCPB.py				
M1-Y6-c	35.81	128.87	Created by Seminario method using MCPB.py				
Y1-M1-Y2	21.52	102.24	Created by Seminario method using MCPB.py				
Y1-M1-Y3	27.23	98.28	Created by Seminario method using MCPB.py				
Y1-M1-Y4	22.48	76.78	Created by Seminario method using MCPB.py				
Y1-M1-Y5	30.47	171.92	Created by Seminario method using MCPB.py				
Y1-M1-Y6	29.28	91.23	Created by Seminario method using MCPB.py				
Y2-M1-Y3	30.49	101.91	Created by Seminario method using MCPB.py				
Y2-M1-Y4	23.34	79.32	Created by Seminario method using MCPB.py				
Y2-M1-Y5	22.36	83.90	Created by Seminario method using MCPB.py				
Y2-M1-Y6	40.87	143.24	Created by Seminario method using MCPB.py				
Y3-M1-Y4	23.48	175.06	Created by Seminario method using MCPB.py				
Y3-M1-Y5	26.67	85.42	Created by Seminario method using MCPB.py				
Y3-M1-Y6	31.10	109.85	Created by Seminario method using MCPB.py				
Y4-M1-Y5	21.82	99.48	Created by Seminario method using MCPB.py				
Y4-M1-Y6	23.23	70.64	Created by Seminario method using MCPB.py				
Y6-M1-Y5	32.51	80.75	Created by Seminario method using MCPB.py				
2C-C -Y4	80.0	120.40					
CC-CV-Y1	70.0	120.00	AA his				
CC-CV-Y2	70.0	120.00	AA his				
CC-CV-Y3	70.0	120.00	AA his				
CR-Y1-CV	70.0	117.00	AA his				
CR-Y2-CV	70.0	117.00	AA his				
CR-Y3-CV	70.0	117.00	AA his				
NA-CR-Y1	70.0	120.00	AA his				
NA-CR-Y2	70.0	120.00	AA his				
NA-CR-Y3	70.0	120.00	AA his				
OH-C -Y4	80.0	120.00	(check with Junmei for: theta0:120.0?)				
Y1-CR-H5	50.0	120.00	AA his				
Y1-CV-H4	50.0	120.00	AA his				
Y2-CR-H5	50.0	120.00	AA his				
Y2-CV-H4	50.0	120.00	AA his				
Y3-CR-H5	50.0	120.00	AA his				
Y3-CV-H4	50.0	120.00	AA his				
Y5-c -cd	69.14	123.93	SOURCE3 SOURCE5	3463	2.3073		
Y5-c -n	74.22	123.05	SOURCE3_SOURCE5	8454	1.5552		
Y6-c -oh	75.92	122.10	SOURCE3_SOURCE5	2859	0.8497		
cd-c -Y6	69.14	123.93	SOURCE3_SOURCE5	3463	2.3073		
DIHE							
X -CR-Y1-X	2	10.0	180.0	2.0	JCC,7,(1986),230		

X -CR-Y2-X	2	10.0	180.0	2.0	JCC,7,(1986),230
X -CR-Y3-X	2	10.0	180.0	2.0	JCC,7,(1986),230
X -CV-Y1-X	2	4.8	180.0	2.0	JCC,7,(1986),230
X -CV-Y2-X	2	4.8	180.0	2.0	JCC,7,(1986),230
X -CV-Y3-X	2	4.8	180.0	2.0	JCC,7,(1986),230
2C-C -Y4-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
C -Y4-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
C -Y4-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CC-CV-Y1-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CC-CV-Y2-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CC-CV-Y3-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y1-M1-Y2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y1-M1-Y3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y1-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y1-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y1-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y2-M1-Y3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y2-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y2-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y2-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y3-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y3-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y3-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-2C-C -Y4	1	0.0	0.0	1.0	
HO-OH-C -Y4	1	2.3	180.0	-2.0	Junmei et al, 1999
HO-OH-C -Y4	1	1.9	0.0	1.0	Junmei et al, 1999
M1-Y1-CR-H5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y1-CV-H4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y2-CR-H5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y2-CV-H4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y3-CR-H5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y3-CV-H4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y5-c -cd	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y5-c -n	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y6-c -cd	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y6-c -oh	3	0.00	0.00	3.0	Treat as zero by MCPB.py
NA-CR-Y1-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
NA-CR-Y2-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
NA-CR-Y3-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
OH-C -Y4-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y2-CR	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y3-CR	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y4-C	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y3-CR	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y4-C	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y4-C	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-C -2C-HC	1	0.08	180.0	-3.0	Junmei et al, 1999 (HC-CT-C -O)
Y4-C -2C-HC	1	0.0	0.0	-2.0	
Y4-C -2C-HC	1	0.8	0.0	1.0	
Y4-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y6-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py

Y5-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-c -n -hn	1	2.5	180.0	-2.0	JCC,7,(1986),230
Y5-c -n -hn	1	2.0	0.0	1.0	J.C.cistrans-NMA DE
Y6-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-M1-Y3-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-M1-Y5-c	3	0.00	0.00	3.0	Treat as zero by MCPB.py
c -Y6-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
IMPR					
X -X -c -Y5		10.5	180.	2.	JCC,7,(1986),230
X -X -C -Y4		10.5	180.	2.	JCC,7,(1986),230
X -X -c -Y6		10.5	180.	2.	JCC,7,(1986),230
2C-OH-C -Y4		10.5	180.	2.	
Y5-cd-c -n		10.5	180.0	2.0	General improper
torsional angle (2 general atom types)					
NONB					
M1	1.3950	0.0149170000			IOD set for Zn2+ ion from Li et al.
JCTC, 2013, 9, 2733					
Y1	1.8240	0.1700			OPLS
Y2	1.8240	0.1700			OPLS
Y3	1.8240	0.1700			OPLS
Y4	1.6612	0.2100			OPLS
Y5	1.6612	0.2100			OPLS
Y6	1.6612	0.2100			OPLS

4 Summary to Part I

The first part of this thesis provides detailed mechanistic insights, which lead to a deeper understanding of the underlying biochemical processes.

In publication I, the glycosylase activity of the bacterial DNA glycosylase Fpg was subject of investigations. Within this study, we found an alternative pathway for the excision of the substrate 8OG, analogous to the excision mechanism found for FapyG [34]. This alternative pathway shares the initial ribose-opening with the previously found ribose-protonated excision mechanism of 8OG [33], but does not involve the base in the mechanism contrary to the previous 'base-specific' ribose-protonated mechanism. Thus, the alternative pathway is a base-independent variation of the ribose-protonated excision. It can excise 8OG in both conformations, *syn* and *anti* whereas the base-specific mechanism only excises *syn*-bound 8OG.

According to the obtained QM/MM energy profiles using large QM spheres with up to 588 atoms, we found that Fpg has no preference in the base-independent excision between both conformations. The heights of the maximum energy barriers along the excision pathways of *syn*- and *anti*-bound 8OG are about the same and are in good agreement with experimental measurements.

In publication II, different possible pathways for the decarboxylation of 5caU by IDCase were compared. The QM/MM profiles revealed that the catalytic mechanism of IDCase involves most likely a direct decarboxylation mechanism, which has turned out to be energetically more favorable than a decarboxylation via a tetrahedral intermediate or with nucleophilic activation. Detailed investigations on the direct decarboxylation of 5caU by IDCase revealed that it is a one-step mechanism with concerted C-C bond opening and proton transfer.

Part II.

Accuracy Benchmark
of
Range-Separated
Density-Functional Theory
with
Random Phase Approximation

5 Introduction

For the investigation of reaction mechanisms, accurate and reliable methods are necessary. In order to reliably describe the energetics and dynamics of an enzymatic reaction, said method must perform well for thermodynamic and kinetic properties and should give a good description of non-covalent interactions. Semi-local and hybrid density functionals, which are widely used in this context, produce mixed results [117, 118]. Especially, they often seem to give a poor description of reaction barrier heights.

In a small-scale study by Toulouse and coworkers [119], range-separated hybrid DFT, based on a short-range PBE exchange-correlation functional, in combination with long-range random phase approximation correlation (RSHPBE+lrRPA) seemed to perform well for energy barrier heights. Additionally, promising results of RSHPBE+lrRPA were obtained for atomization energies [119] and non-covalent interactions [120–125]. As the basis set dependence of RSHPBE+lrRPA is not as pronounced as that of standard RPA [119, 122, 125–127], this method could provide reasonable accuracies for larger molecular systems.

Before employing a new method, however, it is essential to benchmark it extensively in order to assess how well it is suitable for an application. Furthermore, a comprehensive benchmark can provide information on how to improve the method.

In publication III, a comprehensive picture of the performance of RSHPBE+lrRPA over a broad range of molecular chemistry is obtained. RSHPBE+lrRPA is applied to the GMTKN55 data set [118], a benchmark database for general main group thermochemistry, kinetics and non-covalent interactions, comprising 1505 relative energies grouped in 55 subsets.

6 Theoretical Background

6.1 The Random-Phase-Approximation

The random phase approximation (RPA) is a non-perturbative electronic structure method for the calculation of the ground-state correlation energy of many-electron systems. In the following, the derivation of the RPA correlation energy in terms of the adiabatic connection (AC) fluctuation-dissipation framework is briefly summarized.

6.1.1 Adiabatic Connection

In the Kohn-Sham approach (see section 2.2), a physical many-electron system is linked to a system with non-interacting electrons. This continuous link can be displayed with the adiabatic connection Hamiltonian [128–130]:

$$\hat{H}^\lambda = \hat{T} + \lambda \hat{V}_{ee} + \hat{V}_{\text{ext}}^\lambda, \quad (34)$$

with the kinetic energy operator \hat{T} of the electrons, the operator for electron-electron interactions \hat{V}_{ee} , and the λ -dependent external potential $\hat{V}_{\text{ext}}^\lambda$. The electron-electron interactions are scaled by the coupling strength parameter λ , a constant between 0 and 1. For $\lambda = 1$, the λ -dependent external potential $\hat{V}_{\text{ext}}^\lambda$ is equal to the external potential V_{ext} of the physical system (eq. (9)) and the resulting Hamiltonian \hat{H}^1 corresponds to the Hamiltonian of the physical, fully-interacting system. In case of $\lambda = 0$, the Hamiltonian of the non-interacting system \hat{H}^0 is obtained.

For all λ , $\hat{V}_{\text{ext}}^\lambda$ ensures that the density ρ^λ of the ground state $|\Psi_0^\lambda\rangle$, which is given as the eigenfunction of \hat{H}^λ

$$\begin{aligned} \hat{H}^\lambda |\Psi_0^\lambda\rangle &= E_0^\lambda |\Psi_0^\lambda\rangle \\ \langle \Psi_0^\lambda | \hat{H}^\lambda | \Psi_0^\lambda \rangle &= E_0^\lambda, \end{aligned} \quad (35)$$

is equal to the ground-state density ρ of the physical system for all λ :

$$\rho^\lambda = \rho^1 = \rho. \quad (36)$$

This means \hat{V}_{ext}^0 is equal to the Kohn-Sham potential V_s (eq. (14)) and the non-interacting ground state corresponds to the KS determinant $|\Psi_0^0\rangle = |\Phi_0^{\text{KS}}\rangle$, which is constructed as a single Slater determinant of the Kohn-Sham orbitals (eq. (13)).

The derivative of eq. (35) with respect to the coupling strength parameter λ considering the Hellmann-Feynman theorem [131] is given by

$$\begin{aligned} \frac{\partial E^\lambda}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \langle \Psi_0^\lambda | \hat{H}^\lambda | \Psi_0^\lambda \rangle \\ &= \langle \Psi_0^\lambda | \hat{V}_{ee} | \Psi_0^\lambda \rangle + \langle \Psi_0^\lambda | \frac{\partial \hat{V}_{\text{ext}}^\lambda}{\partial \lambda} | \Psi_0^\lambda \rangle. \end{aligned} \quad (37)$$

Using this derivative eq. (37) and the limits of $\hat{V}_{\text{ext}}^\lambda$, the total ground state energy of the physical, interacting system $E_0^1 \equiv E_0$ can be expressed as

$$\begin{aligned} E_0 &= E_0^0 + \int_0^1 d\lambda \frac{\partial E^\lambda}{\partial \lambda} \\ &= E_0^0 + \int_0^1 d\lambda \langle \Psi_0^\lambda | \hat{V}_{ee} | \Psi_0^\lambda \rangle + \int d\mathbf{r} \rho(\mathbf{r}) (V_{\text{ext}}(\mathbf{r}) - V_s(\mathbf{r})), \end{aligned} \quad (38)$$

with the energy expression for the non-interacting system

$$E_0^0 = \langle \Phi_0^{\text{KS}} | \hat{T} | \Phi_0^{\text{KS}} \rangle + \int d\mathbf{r} \rho(\mathbf{r}) V_s. \quad (39)$$

Inserting eq. (39) in eq. (38) finally results in

$$E_0 = \langle \Phi_0^{\text{KS}} | \hat{T} | \Phi_0^{\text{KS}} \rangle + \int d\mathbf{r} \rho(\mathbf{r}) V_{\text{ext}} + \int_0^1 d\lambda \langle \Psi_0^\lambda | \hat{V}_{\text{ee}} | \Psi_0^\lambda \rangle, \quad (40)$$

which allows by comparing to eq. (9) the identification of the exchange-correlation energy E_{xc} as

$$E_{\text{xc}} = \int_0^1 d\lambda \langle \Psi_0^\lambda | \hat{V}_{\text{ee}} | \Psi_0^\lambda \rangle - E_J. \quad (41)$$

With the definition

$$E_{\text{xc}} = E_x + E_c \quad (42)$$

and the fact that the expectation value of \hat{V}_{ee} on a non-interacting system is a sum of the Hartree Repulsion E_J and exact exchange energy E_x

$$\langle \Phi_0^{\text{KS}} | \hat{V}_{\text{ee}} | \Phi_0^{\text{KS}} \rangle = E_J + E_x, \quad (43)$$

an exact formula for the correlation energy is given by

$$E_c = \int_0^1 d\lambda \langle \Psi_0^\lambda | \hat{V}_{\text{ee}} | \Psi_0^\lambda \rangle - \langle \Phi_0^{\text{KS}} | \hat{V}_{\text{ee}} | \Phi_0^{\text{KS}} \rangle. \quad (44)$$

6.1.2 Correlation Energy in Terms of Density Fluctuations

The second quantization formalism [132–134] can be used for the description of many-electron systems. In this approach, particles are created or destroyed in the one-particle state p by the fermionic creation c_i^\dagger and annihilation operator c_i , which satisfy the anti-commutation relations

$$\begin{aligned} \{c_i^\dagger, c_j^\dagger\} &= 0 \\ \{c_i, c_j\} &= 0 \\ \{c_i, c_j^\dagger\} &= \delta_{ij} \end{aligned} \quad (45)$$

to give the correct fermion statistics.

The operators c_i^\dagger and c_i can be linearly combined into field operators

$$\begin{aligned} \hat{\psi}^\dagger(\mathbf{x}) &= \sum_p \phi_p^*(\mathbf{x}) \hat{c}_p^\dagger, \\ \hat{\psi}(\mathbf{x}) &= \sum_p \phi_p(\mathbf{x}) \hat{c}_p, \end{aligned} \quad (46)$$

where $\mathbf{x} = (\mathbf{r}, \sigma)$ is a combined spin-space coordinate.

With the field operators (eq. (46)), the electron-electron interaction is given in the second quantization as

$$\hat{V}_{\text{ee}} = \frac{1}{2} \iint d\mathbf{x}_1 d\mathbf{x}_2 \psi^\dagger(\mathbf{x}_1) \psi^\dagger(\mathbf{x}_2) \hat{v}_{\text{ee}}(\mathbf{x}_1, \mathbf{x}_2) \psi(\mathbf{x}_2) \psi(\mathbf{x}_1), \quad (47)$$

where $\hat{v}_{ee}(\mathbf{x}_1, \mathbf{x}_2) = \hat{v}_{ee}((\mathbf{r}_1, \sigma_1), (\mathbf{r}_2, \sigma_2)) = \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}$ is the electron-electron operator.

Introducing the two-particle density operator in second quantization form

$$\hat{P}(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2} \hat{\psi}^\dagger(\mathbf{x}_1) \hat{\psi}^\dagger(\mathbf{x}_2) \hat{\psi}(\mathbf{x}_2) \hat{\psi}(\mathbf{x}_1), \quad (48)$$

the electron-electron operator in eq. (47) can be rewritten as

$$\hat{V}_{ee} = \iint d\mathbf{x}_1 d\mathbf{x}_2 \hat{v}_{ee}(\mathbf{x}_1, \mathbf{x}_2) \hat{P}(\mathbf{x}_1, \mathbf{x}_2). \quad (49)$$

To obtain an expression for \hat{V}_{ee} in terms of one-particle operators, $\hat{P}(\mathbf{x}_1, \mathbf{x}_2)$ may be rewritten using the one-particle density operator

$$\hat{\rho}(\mathbf{x}) = \hat{\psi}^\dagger(\mathbf{x}) \hat{\psi}(\mathbf{x}) \quad (50)$$

and the fermion anticommutation relations (eq. (45)) to

$$\hat{P}(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2} \left(\hat{\rho}(\mathbf{x}_1) \hat{\rho}(\mathbf{x}_2) - \delta(\mathbf{x}_1 - \mathbf{x}_2) \hat{\rho}(\mathbf{x}_1) \right). \quad (51)$$

Introducing the density fluctuation operator

$$\Delta \hat{\rho}(\mathbf{x}) = \hat{\rho}(\mathbf{x}) - \rho(\mathbf{x}), \quad (52)$$

$\hat{P}(\mathbf{x}_1, \mathbf{x}_2)$ can be further rewritten obtaining

$$\begin{aligned} \hat{P}(\mathbf{x}_1, \mathbf{x}_2) = & \frac{1}{2} \left(\Delta \hat{\rho}(\mathbf{x}_1) \Delta \hat{\rho}(\mathbf{x}_2) + \hat{\rho}(\mathbf{x}_1) \rho(\mathbf{x}_2) + \rho(\mathbf{x}_1) \hat{\rho}(\mathbf{x}_2) \right. \\ & \left. - \rho(\mathbf{x}_1) \rho(\mathbf{x}_2) - \delta(\mathbf{x}_1 - \mathbf{x}_2) \hat{\rho}(\mathbf{x}_1) \right). \end{aligned} \quad (53)$$

Inserting this formulation of $\hat{P}(\mathbf{x}_1, \mathbf{x}_2)$ into eq. (49) and combining it with eq. (44) gives the following expression for E_c considering that all terms except the one with the two density fluctuation operators cancel because the density is independent of λ :

$$E_c = \frac{1}{2} \int_0^1 d\lambda \iint d\mathbf{x}_1 d\mathbf{x}_2 \hat{v}_{ee}(\mathbf{x}_1, \mathbf{x}_2) \left(\langle \Psi_0^\lambda | \Delta \hat{\rho}(\mathbf{x}_1) \Delta \hat{\rho}(\mathbf{x}_2) | \Psi_0^\lambda \rangle - \langle \Phi_0^{\text{KS}} | \Delta \hat{\rho}(\mathbf{x}_1) \Delta \hat{\rho}(\mathbf{x}_2) | \Phi_0^{\text{KS}} \rangle \right). \quad (54)$$

6.1.3 Fluctuation-Dissipation Theorem

The fluctuation-dissipation theorem [135] makes a connection between the density fluctuation at a given value of the coupling strength parameter λ and the response properties of the system. With its help, it is possible to represent the correlation energy (eq. (54)) in terms of density-density response functions.

Linear response functions $\chi(t_1 - t_2)$ in their general form describe the relation between the input (perturbation of the Hamiltonian) $h(t)$ and the response of the system

$$x(t_1) = \int_{-\infty}^t dt_2 \chi(t_1 - t_2) h(t_2) + \dots, \quad (55)$$

if higher order terms of the Volterra expansion [136] for the full nonlinear response (indicated as dots in eq. (55)) are neglected.

Applying the linearly truncated eq. (55) on the density response $\delta\rho(\mathbf{x}_1, t_1)$ to a small perturbation V_1^λ in the external potential V_{ext}^λ gives

$$\delta\rho(\mathbf{x}_1, t_1) = \int_{-\infty}^{\infty} dt_2 \int d\mathbf{x}_2 \chi^\lambda(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2) V_1^\lambda(\mathbf{x}_2, t_2), \quad (56)$$

with the density-density response function

$$\chi^\lambda(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2) = -i\theta(t_1 - t_2) \langle \Psi_0^\lambda | [\hat{\rho}(\mathbf{x}_1, 0), \hat{\rho}(\mathbf{x}_2, t_2 - t_1)] | \Psi_0^\lambda \rangle, \quad (57)$$

where i is the imaginary unit, $\theta(t_1 - t_2)$ the Heaviside step function

$$\theta(t_1 - t_2) = \begin{cases} 0, & t_1 - t_2 < 0 \\ 1, & t_1 - t_2 > 0, \end{cases} \quad (58)$$

which ensures that the response at time t_1 is due to a perturbation at an earlier time, and $[\hat{\rho}(\mathbf{x}_1, 0), \hat{\rho}(\mathbf{x}_2, t_2 - t_1)]$ is the commutator of the time dependent one-particle density operator

$$\hat{\rho}(\mathbf{x}, t) = e^{i\hat{H}_0 t} \hat{\rho}(\mathbf{x}) e^{-i\hat{H}_0 t}, \quad (59)$$

with Hamiltonian $\hat{H}_0 \equiv \hat{H}^\lambda|_{\lambda=1}$ of the physical, unperturbed system.

Since the density-density response function in eq. (57) is only dependent on the time difference, $\chi^\lambda(\mathbf{x}_1, \mathbf{x}_2, t_1 - t_2)$ can be transformed into the imaginary frequency domain by Fourier transform. Insertion of the completeness of the electronic states $1 = \sum_n |\Psi_n^\lambda\rangle\langle\Psi_n^\lambda|$ at any λ then yields χ^λ in Lehmann representation [137]

$$\chi^\lambda(\mathbf{x}_1, \mathbf{x}_2, \omega) = \sum_{n \neq 0} \lim_{\eta \rightarrow 0} \left(\frac{\langle \Psi_0^\lambda | \hat{\rho}(\mathbf{x}_1) | \Psi_n^\lambda \rangle \langle \Psi_n^\lambda | \hat{\rho}(\mathbf{x}_2) | \Psi_0^\lambda \rangle}{\omega - \Omega_{0n}^\lambda + i\eta} - \frac{\langle \Psi_0^\lambda | \hat{\rho}(\mathbf{x}_2) | \Psi_n^\lambda \rangle \langle \Psi_n^\lambda | \hat{\rho}(\mathbf{x}_1) | \Psi_0^\lambda \rangle}{\omega + \Omega_{0n}^\lambda - i\eta} \right), \quad (60)$$

where Ω_{0n} are the respecting excitation energies, and the term $i\eta$ with the infinitesimal real number η arises from the integral representation of the Heaviside step function. Within the zero-temperature fluctuation-dissipation theorem [129, 138–140] the expectation value of $\Delta\hat{\rho}(\mathbf{x}_1)\Delta\hat{\rho}(\mathbf{x}_2)$ is an integral of the imaginary part of the frequency dependent response function along the real positive axis (eq. (61)).

$$\langle \Psi_0^\lambda | \Delta\hat{\rho}(\mathbf{x}_1)\Delta\hat{\rho}(\mathbf{x}_2) | \Psi_0^\lambda \rangle = - \int_0^\infty \frac{d\omega}{\pi} \text{Im}(\chi^\lambda(\mathbf{x}_1, \mathbf{x}_2, \omega)) \quad (61)$$

$$\begin{aligned} &= - \int_0^\infty \frac{d\omega}{\pi} i\chi^\lambda(\mathbf{x}_1, \mathbf{x}_2, \omega) \\ &\stackrel{u \equiv i\omega}{=} - \int_0^\infty \frac{du}{\pi} \chi^\lambda(\mathbf{x}_1, \mathbf{x}_2, iu) \end{aligned} \quad (62)$$

Using the residue theorem, the integration along the real axis can be related to the numerically more stable integration along the imaginary axis (eq. (62)).

With eq. (62), the ground-state correlation energy can be written in terms of χ^λ

$$E_c = -\frac{1}{2} \int_0^1 d\lambda \int_0^\infty \frac{du}{\pi} \iint d\mathbf{x}_1 d\mathbf{x}_2 \hat{v}_{ee}(\mathbf{x}_1, \mathbf{x}_2) (\chi^\lambda(\mathbf{x}_1, \mathbf{x}_2, iu) - \chi^0(\mathbf{x}_1, \mathbf{x}_2, iu)). \quad (63)$$

6.1.4 Random-Phase-Approximation

The calculation of the correlation energy according to eq. (63) requires the evaluation of density-density response functions. For the non-interacting case, $\lambda = 0$, the density-density response function is given as the frequency-dependent, non-interacting Kohn-Sham response function by

$$\chi^0(\mathbf{r}_1, \mathbf{r}_2, \omega) = \sum_i^{N_{\text{occ}}} \sum_a^{N_{\text{virt}}} \sum_{\sigma_1 \sigma_2} \left(\frac{\phi_i^*(\mathbf{r}_1, \sigma_1) \phi_a(\mathbf{r}_1, \sigma_1) \phi_i(\mathbf{r}_2, \sigma_2) \phi_a^*(\mathbf{r}_2, \sigma_2)}{\omega - (\varepsilon_a - \varepsilon_i)} - \frac{\phi_i^*(\mathbf{r}_2, \sigma_2) \phi_a(\mathbf{r}_2, \sigma_2) \phi_i(\mathbf{r}_1, \sigma_1) \phi_a^*(\mathbf{r}_1, \sigma_1)}{\omega + (\varepsilon_a - \varepsilon_i)} \right), \quad (64)$$

where ϕ_i and ϕ_a are occupied and virtual Kohn-Sham spin orbitals with the respective orbital energies ε_i and ε_a .

Assuming real molecular orbitals

$$\phi_i^*(\mathbf{r}_1, \sigma_1) = \phi_i(\mathbf{r}_1, \sigma_1), \quad (65)$$

eq. (64) simplifies to

$$\chi^0(\mathbf{r}_1, \mathbf{r}_2, \omega) = \sum_i^{N_{\text{occ}}} \sum_a^{N_{\text{virt}}} \sum_{\sigma_1 \sigma_2} \frac{2(\varepsilon_a - \varepsilon_i) \phi_i(\mathbf{r}_1, \sigma_1) \phi_a(\mathbf{r}_1, \sigma_1) \phi_i(\mathbf{r}_2, \sigma_2) \phi_a(\mathbf{r}_2, \sigma_2)}{\omega^2 - (\varepsilon_a - \varepsilon_i)^2}. \quad (66)$$

An expression for the coupling-strength dependent density-density response function χ^λ can be derived by time-dependent density functional theory [141]. Here, χ^λ is related to the non-interacting response function eq. (64) by a Dyson-type equation:

$$\chi^\lambda(\mathbf{r}_1, \mathbf{r}_2, \omega) = \chi^0(\mathbf{r}_1, \mathbf{r}_2, \omega) + \iint d\mathbf{r}_3 d\mathbf{r}_4 \chi^0(\mathbf{r}_1, \mathbf{r}_3, \omega) f_{\text{Hxc}}^\lambda(\mathbf{r}_3, \mathbf{r}_4, \omega) \chi^\lambda(\mathbf{r}_4, \mathbf{r}_2, \omega), \quad (67)$$

with the frequency-dependent Hartree, exchange, and correlation kernel, which can be decomposed into a Coulomb part and an exchange-correlation (xc) kernel

$$f_{\text{Hxc}}^\lambda(\mathbf{r}_1, \mathbf{r}_2, \omega) = \lambda \hat{v}_{\text{ee}}(\mathbf{r}_1, \mathbf{r}_2) + f_{\text{xc}}^\lambda(\mathbf{r}_1, \mathbf{r}_2, \omega), \quad (68)$$

with the electron-electron operator $\hat{v}_{\text{ee}}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}$.

For evaluation of eq. (67), f_{xc}^λ has to be approximated since the expression is not explicitly known. The simplest approximation one could think of, is to neglect the xc kernel by setting it to zero which is made within the random-phase-approximation (RPA) [142–145],

$$f_{\text{xc}}^{\lambda, \text{RPA}}(\mathbf{r}_1, \mathbf{r}_2, \omega) = 0. \quad (69)$$

Applying this approximation to eq. (67) in matrix notation and solving for $\chi^{\lambda, \text{RPA}}(\omega)$ yields

$$\chi^{\lambda, \text{RPA}}(\omega) = \left(1 - \chi^0(\omega) \lambda \mathbf{v} \right)^{-1} \chi^0(\omega). \quad (70)$$

Switching to matrix notation gives for the integral in eq. (63)

$$\begin{aligned} & \iint d\mathbf{x}_1 d\mathbf{x}_2 \hat{v}_{\text{ee}}(\mathbf{x}_1, \mathbf{x}_2) \left(\chi^\lambda(\mathbf{x}_1, \mathbf{x}_2, iu) - \chi^0(\mathbf{x}_1, \mathbf{x}_2, iu) \right) \\ &= \sum_{\mathbf{r}_1, \mathbf{r}_2} \sum_{\sigma_1, \sigma_2} \hat{v}_{\text{ee}}(\mathbf{r}_1, \mathbf{r}_2) \left(\chi^\lambda((\mathbf{r}_1, \sigma_1), (\mathbf{r}_2, \sigma_2), iu) - \chi^0((\mathbf{r}_1, \sigma_1), (\mathbf{r}_2, \sigma_2), iu) \right) \\ &= \sum_{\mathbf{r}_1, \mathbf{r}_2} \hat{v}_{\text{ee}}(\mathbf{r}_1, \mathbf{r}_2) \left(\chi^\lambda(\mathbf{r}_1, \mathbf{r}_2, iu) - \chi^0(\mathbf{r}_1, \mathbf{r}_2, iu) \right), \\ &= \text{Tr} \left(\mathbf{v} \left(\chi^\lambda(iu) - \chi^0(iu) \right) \right), \end{aligned} \quad (71)$$

where \mathbf{v} and $\chi^\lambda(iu)$ denote matrices of $\hat{v}_{ee}(\mathbf{r}_1, \mathbf{r}_2)$ and $\chi^\lambda(\mathbf{r}_1, \mathbf{r}_2, iu)$, with

$$\chi^\lambda(\mathbf{r}_1, \mathbf{r}_2, iu) = \sum_{\sigma_1, \sigma_2} \chi^\lambda((\mathbf{r}_1, \sigma_1), (\mathbf{r}_2, \sigma_2), iu). \quad (72)$$

Together with eq. (70), the correlation energy within RPA can now be expressed as

$$\begin{aligned} E_c^{\text{RPA}} &= - \int_0^1 d\lambda \int_0^\infty \frac{du}{2\pi} \text{Tr}(\mathbf{v}(\chi^{\lambda, \text{RPA}}(iu) - \chi^0(iu))) \\ &= - \int_0^1 d\lambda \int_0^\infty \frac{du}{2\pi} \text{Tr}((1 - \chi^0(iu)\lambda\mathbf{v})^{-1}\chi^0(iu)) \end{aligned} \quad (73)$$

Analytical integration over the coupling-strength parameter λ finally yields

$$E_c^{\text{RPA}} = \int_0^\infty \frac{du}{2\pi} \text{Tr}(\ln(\mathbf{1} - \chi^0(iu)\mathbf{v}) + \chi^0(iu)\mathbf{v}) \quad (74)$$

in the real-space basis. Expressing eq. (74) in terms of the Hartree kernel matrix \mathbf{V} , with the matrix elements

$$V_{ia,jb} = (ia|jb) = \iint d\mathbf{r}_1 d\mathbf{r}_2 \phi_i^*(\mathbf{r}_1)\phi_a(\mathbf{r}_1)\hat{v}_{ee}(\mathbf{r}_1, \mathbf{r}_2)\phi_j^*(\mathbf{r}_2)\phi_b(\mathbf{r}_2), \quad (75)$$

where, as in the following, the indices i, j and a, b denote occupied and virtual molecular orbitals, and the Kohn-Sham polarization propagator in the canonical orbital space $\Pi^0(iu)$, with

$$\Pi_{ia,jb}^0(iu) = \delta_{ij}\delta_{ab} \frac{-2(\varepsilon_a - \varepsilon_i)}{u^2 + (\varepsilon_a - \varepsilon_i)^2} \quad (76)$$

yields an expression of the RPA correlation energy in the molecular orbital space:

$$E_c^{\text{RPA}} = \int_0^\infty \frac{du}{2\pi} \text{Tr}(\ln(\mathbf{1} - \Pi^0(iu)\mathbf{V}) + \Pi^0(iu)\mathbf{V}). \quad (77)$$

6.2 Efficient Calculation of the RPA Correlation Energy

The RPA correlation energy formulated according to eq. (77) exhibits a scaling behavior of $\mathcal{O}(N^6)$ with respect to the system size. For accessing the RPA correlation energy for larger molecular systems, for which the calculations according to the formulation in eq. (77) are too expensive, the RPA correlation energy has been reformulated to obtain formulations with lower scaling behaviors.

Some important concepts for reducing the scaling behavior of the RPA correlation energy are briefly summarized in the following.

6.2.1 The Resolution-of-the-Identity

The resolution-of-the-identity (RI) is a frequently used approach in quantum chemistry and physics [146–157], in which the identity is spanned by a complete orthonormal set $\{\Phi_m\}$

$$1 = \sum_P |\Phi_P\rangle\langle\Phi_P|. \quad (78)$$

With the use of RI, a four-center, two-electron integral in atomic orbital space

$$(\mu\nu|\lambda\sigma) = \iint d\mathbf{r}_1 d\mathbf{r}_2 \phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)\hat{v}_{ee}(\mathbf{r}_1, \mathbf{r}_1)\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2) \quad (79)$$

can be factorized in two- and tree-center integrals by twofold insertion of eq. (78) to yield

$$(\mu\nu|\lambda\sigma) = \sum_{P,Q} (\mu\nu P)(P|Q)(Q\lambda\sigma), \quad (80)$$

with

$$(\mu\nu P) = \int d\mathbf{r} \phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})\Phi_P(\mathbf{r}), \quad (81)$$

$$(P|Q) = \iint d\mathbf{r}_1 d\mathbf{r}_2 \hat{v}_{ee}\Phi_P(\mathbf{r}_1)\Phi_Q(\mathbf{r}_2). \quad (82)$$

As in practical applications, however, finite auxiliary basis sets are used for the expansion in eq. (78), the equality is no longer given. This might lead to a poor approximation of the four-center integral in eq. (80).

To reduce the error introduced by an incomplete auxiliary basis set $\{\Phi_P\}$ in context of representing a charge distribution $|\lambda\sigma\rangle$, this incomplete set can be fitted with the use of coefficients $C_P^{\lambda\sigma}$

$$|\widetilde{\lambda\sigma}\rangle = \sum_P C_P^{\lambda\sigma} |P\rangle. \quad (83)$$

This procedure is also known as "density fitting" [158–160]. The appropriate set of parameters is found by minimization of the deviation

$$\Delta_{\lambda\sigma} = |\lambda\sigma\rangle - |\widetilde{\lambda\sigma}\rangle \quad (84)$$

with respect to the operator m_{12} according to

$$0 \stackrel{!}{=} \frac{\delta}{\delta C_P^{\lambda\sigma}} (\Delta_{\mu\nu} |m_{12}| \Delta_{\lambda\sigma})$$

$$C_Q^{\mu\nu} = \sum_P (\mu\nu |m_{12}| P)(P |m_{12}| Q)^{-1}, \quad (85)$$

where the superscript " -1 " denotes the inverse of the whole matrix. With the approximated charge densities (eq. (83)) using the optimized coefficients (eq. (85)), the four-center integral in eq. (79) can be expressed as

$$(\mu\nu|\lambda\sigma) \approx (\widetilde{\mu\nu}|\widetilde{\lambda\sigma}) = \sum_{P,Q,R,S} (\mu\nu |m_{12}| P)(P |m_{12}| Q)^{-1}(Q|R)(R|m_{12}|S)^{-1}(S|m_{12}|\lambda\sigma). \quad (86)$$

A common choice for m_{12} is the Coulomb operator, which yields more accurate approximate four-center integrals than the overlap metric $m_{12} = \delta_{r_{12}}$ but is less local [149]. The attenuated Coulomb metric $m_{12} = \text{erfc}(\omega|\mathbf{r}_1 - \mathbf{r}_2|)\hat{v}_{ee}$ has shown to represent a compromise between accuracy and spatial locality [149, 153].

RI was first introduced by Furche and coworkers [156] to the RPA energy expression. Their formulation can be derived by simply approximating the four-center integrals in

eq. (77) within the RI approximation using the Coulomb metric. The introduction of RI and cyclic permutation within the trace yields the RI-RPA expression

$$E_c^{\text{RI-RPA}} = \int_0^\infty \frac{du}{2\pi} \text{Tr} \left(\ln(\mathbf{1} - \phi_{\text{RI}}^0(iu)\mathbf{C}) + \phi_{\text{RI}}^0(iu)\mathbf{C} \right), \quad (87)$$

with

$$\phi_{\text{RI}}^0(iu)_{PQ} = \sum_{i,j,a,b} (P|m_{12}|ia)\Pi_{ia,jb}^0(iu)(jb|m_{12}|Q), \quad (88)$$

$$C_{PQ} = \sum_{R,S} (P|m_{12}|R)^{-1}(R|S)(S|m_{12}|Q)^{-1}. \quad (89)$$

This formulation reduces the effective scaling behavior of $\mathcal{O}(N_{\text{occ}}^3 N_{\text{virt}}^3)$ in eq. (77) to formal $\mathcal{O}(N_{\text{aux}}^2 N_{\text{occ}} N_{\text{virt}})$ scaling with N_{occ} occupied and N_{virt} unoccupied spatial orbitals and N_{aux} auxiliary basis functions.

In the formulations of Schurkus *et al.* [151] and Luenser *et al.* [153] a RI decomposition of four-center integrals with the overlap or attenuated Coulomb metric was used, rather than the Coulomb metric. The formal linear scaling of these formulations arises not only from the RI decomposition but also from the formulation within the atomic orbital space, which is illustrated in the following section.

6.2.2 RPA Correlation Energy in the Atomic Orbital Formalism

Within the LCAO (linear combination of atomic orbitals) ansatz [161], molecular orbitals are expressed in terms of atom centered basis functions (atomic orbitals, AOs)

$$\phi_i(\mathbf{r}) = \sum_{\mu} C_{\mu i} \phi_{\mu}(\mathbf{r}). \quad (90)$$

At this point it is noted that the atom-centered basis functions are labeled with the Greek letters $\mu, \nu, \lambda, \sigma$, while occupied and unoccupied MOs are designated with the Latin letters i, j and a, b , respectively.

Using eq. (90), the integral in eq. (75) may be expressed as

$$(ia|jb) = \sum_{\mu,\nu,\lambda,\sigma}^{N_{\text{basis}}} C_{\mu i} C_{\lambda a} C_{\nu j} C_{\sigma b} (\mu\nu|\lambda\sigma). \quad (91)$$

The complex conjugation is omitted in eq. (91) as the MOs are assumed to be real.

As the charge distributions $|\mu\nu\rangle$ decay exponentially with the distance between ϕ_{μ} and ϕ_{ν} , there are for each ϕ_{μ} only a constant number of AOs ϕ_{ν} , which give a significant overlap. This makes clear, that in the limit of large molecular systems, there is only a linear number of significant charge distributions.

This observation reduces the number of significant integrals in eq. (91) from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^2)$ [162, 163], which can be calculated selectively using integral screening techniques [164, 165].

For the formulation of the RI-RPA correlation energy (eq. (87)) in terms of AOs, only the non-interacting density-density response function (eq. (88)) has to be reformulated.

For this purpose, Schurkus *et al.* used a contracted double-Laplace transform [151] which was shown to be equivalent to a cosine transform [154]

$$\chi_{\text{RI}}^0(iu) = \int_{-\infty}^{+\infty} d\tau \cos(u\tau) \chi_{\text{RI}}^0(i\tau). \quad (92)$$

Using optimized weights and roots for the numerical double-Laplace transformation allows for the calculation of the non-interacting density-density response function in the imaginary time domain according to [154]

$$\chi_{MN}^0(i\tau) = \sum_{\mu,\nu,\lambda,\sigma}^{N_{\text{basis}}} G_{\mu\nu}^0(-i\tau) B_{\nu\lambda}^M G_{\lambda\sigma}^0(i\tau) B_{\sigma\mu}^N, \quad (93)$$

with the three-center integrals

$$B_{\nu\lambda}^M = (\nu\lambda|m_{12}|M), \quad (94)$$

and the one-particle Green's function

$$\mathbf{G}^0(i\tau) = \theta(-i\tau) \underline{\mathbf{G}}^0(i\tau) + \theta(i\tau) \overline{\mathbf{G}}^0(i\tau) \quad (95)$$

$$\underline{G}_{\mu\nu}^0(i\tau) = \sum_i^{N_{\text{occ}}} C_{\mu i} C_{\nu i} e^{-(\varepsilon_i - \varepsilon_F)\tau}$$

$$\overline{G}_{\mu\nu}^0(i\tau) = - \sum_a^{N_{\text{virt}}} C_{\mu a} C_{\nu a} e^{-(\varepsilon_a - \varepsilon_F)\tau},$$

where ε_F denotes the Fermi level.

The time determining step of this atomic orbital based formulation of RI-RPA is the calculation of eq. (92), which can be implemented achieving an effective linear scaling behavior [151].

6.3 Range-Separated DFT

Even though local or semilocal density-functionals have an excellent accuracy to computational cost ratio, these functionals have some limitations. There is the relatively weak performance for thermochemistry [166], incorrect dissociation limits [167, 168], instability of anions [169, 170], or the poor description of localized states [171, 172], to mention only some of the shortcomings. In the following, some approaches that alleviate these problems are given.

6.3.1 Long-Range Corrected Density-Functionals

Most of the aforementioned problems arise from the self-interaction error (SIE) [173–175], which is also known as localization problem [176]. In density-functionals which are self-interaction free, the self-interaction of the electrons introduced by the Hartree energy is canceled by the exchange term. This means such a functional satisfies the condition [173]

$$E_{\text{xc}}[\rho_i(\mathbf{r})] + E_{\text{H}}[\rho_i(\mathbf{r})] = 0. \quad (96)$$

In Hartree–Fock (HF) theory, the exchange cancels out the SIE completely. Local density-functionals, however, cannot be corrected so easily by replacing the $E_x^{(\text{semi-})\text{local DFT}}$ functional with the non-local HF exchange functional E_x^{HF} , since it is worthwhile to maintain the error cancellation of consistent $E_x^{(\text{semi-})\text{local DFT}}$ and $E_c^{(\text{semi-})\text{local DFT}}$ pairs [177]. For that reason (global) hybrid density-functionals are constructed that retain a fixed fraction of $E_x^{(\text{semi-})\text{local DFT}}$ (see eq. (17)).

The potential of an exact, SIE free xc-functional, which fulfills eq. (96), should exhibit an $\frac{1}{r}$ asymptotic decay [178, 179]. While global hybrids cure many of the shortcomings of local or semilocal density-functionals, they still do not show the correct long-range behavior as the asymptotic decay of the potential is proportional to the amount of exact exchange. The average potential from Hartree–Fock theory, in contrast, exhibits the correct asymptotic decay. To recover the correct asymptotic decay, the full amount of exact exchange for long-range interactions has to be included.

For this reason, the exchange contribution can be separated into a short-range (sr) part given by a (semi-)local density-functional and a long-range (lr) part given by the exact exchange functional

$$E_x = E_x^{(\text{semi-})\text{local DFT, sr}} + E_x^{\text{HF, lr}}. \quad (97)$$

This range-separation can be achieved by partitioning the electron–electron interactions via the error function and its complementary function [180, 181]

$$\hat{v}_{ee} = \frac{\text{erfc}(\mu|\mathbf{r}_1 - \mathbf{r}_2|)}{|\mathbf{r}_1 - \mathbf{r}_2|} + \frac{\text{erf}(\mu|\mathbf{r}_1 - \mathbf{r}_2|)}{|\mathbf{r}_1 - \mathbf{r}_2|} = \hat{v}_{ee}^{\text{sr}} + \hat{v}_{ee}^{\text{lr}}, \quad (98)$$

or another range-separation function [182, 183] to ensure smooth transition between the description of sr and lr interactions. The range-separation parameter μ defines the range at which the separation takes place. With such range-separated density functionals, the correct long-range behavior is obtained, which remedies some of the problems in (semi-)local DFT.

6.3.2 Range-Separated DFT with the Random-Phase-Approximation

The range-separation scheme in eq. (98) can also be applied to the correlation functional. In doing so, a gain of accuracy can be obtained: (semi-)local density-functionals are even better in the description of the short-range (dynamic) correlation only [180] and the long-range part of the correlation can be described at a higher (*ab initio*) level of theory such as configuration-interaction [180], second-order Møller–Plesset perturbation theory [184, 185], coupled-cluster [186, 187], or RPA [119, 121, 122, 125, 126, 188].

$$E_c = E_c^{(\text{semi-})\text{local DFT, sr}} + E_c^{\text{ab initio, lr}} \quad (99)$$

Having to calculate the long-range part of the correlation energy only with costly *ab initio* methods allows for the usage of smaller one-atomic basis sets compared to calculations of the standard Coulomb interaction, where normally relatively large basis sets are required for an adequate description, since the slowly converging (with the size of the basis set) description of the Coulomb electron–electron cusp is avoided [127].

To obtain an expression of the RI-RPA correlation energy that accounts for the long-range interactions only, as required for eq. (99), the standard Coulomb operator is sub-

stituted by the long-range electron-electron operator (eq. (98)) in eq. (89)

$$C_{PQ}^{\text{lr}} = \sum_{R,S} (P|m_{12}|R)^{-1} (R|\hat{v}_{\text{ee}}^{\text{lr}}|S) (S|m_{12}|Q)^{-1}, \quad (100)$$

which yields the expression

$$E_c^{\text{RI-RPA, lr}} = \int_0^\infty \frac{du}{2\pi} \text{Tr} \left(\ln(\mathbf{1} - \mathbf{\Pi}_{\text{RI}}^0(iu)\mathbf{C}^{\text{lr}}) + \mathbf{\Pi}_{\text{RI}}^0(iu)\mathbf{C}^{\text{lr}} \right). \quad (101)$$

Even though range-separated (rs) DFT in combination with lr RPA has turned out to be advantageous at specific chemical problems, e.g., barrier heights [119, 125], dissociation of rare gas dimers [120–122, 125, 126], atomization energies [119, 125], non-covalent interactions [120, 121], an overview of the performance of this approach over a broad range of chemical properties was absent. **Publication III** fills in the missing pieces with an extensive accuracy benchmark of rs DFT in combination with lr RPA to complete the picture.

Publication III: Range-Separated Density-Functional Theory in Combination with the Random Phase Approximation: An Accuracy Benchmark

A. Kreppel, D. Graf, H. Laqua, C. Ochsenfeld,

”Range-Separated Density-Functional Theory in Combination with the Random Phase Approximation: An Accuracy Benchmark”

J.Chem.Theory.Comput., **2020**, *16*, 2985-2994

Abstract: A formulation of range-separated random phase approximation (RPA) based on our efficient ω -CDGD-RI-RPA [J. Chem.Theory Comput.2018,14, 2505] method and a large scale benchmark study are presented. By application to the GMTKN55 data set, we obtain a comprehensive picture of the performance of range-separated RPA in general main group thermochemistry, kinetics, and noncovalent interactions. The results show that range-separated RPA performs stably over the broad range of molecular chemistry included in the GMTKN55 set. It improves significantly over semilocal DFT but it is still less accurate than modern dispersion corrected double-hybrid functionals. Furthermore, range-separated RPA shows a faster basis set convergence compared to standard full-range RPA making it a promising applicable approach with only one empirical parameter.

Reprinted with permission from:

A. Kreppel, D. Graf, H. Laqua, C. Ochsenfeld,

”Range-Separated Density-Functional Theory in Combination with the Random Phase Approximation: An Accuracy Benchmark”

J.Chem.Theory.Comput., **2020**, *16*, 2985-2994

Copyright 2020 American Chemical Society.

<https://dx.doi.org/10.1021/acs.jctc.9b01294>

Range-Separated Density-Functional Theory in Combination with the Random Phase Approximation: An Accuracy Benchmark

Andrea Kreppel, Daniel Graf, Henryk Laqua, and Christian Ochsenfeld*

Cite This: *J. Chem. Theory Comput.* 2020, 16, 2985–2994

Read Online

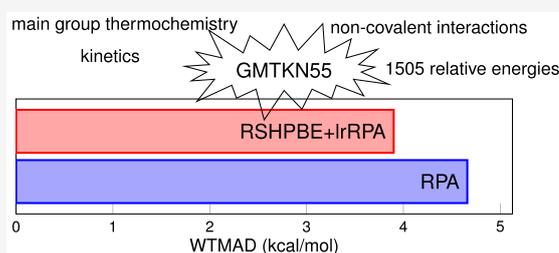
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: A formulation of range-separated random phase approximation (RPA) based on our efficient ω -CDGD-RI-RPA [*J. Chem. Theory Comput.* 2018, 14, 2505] method and a large scale benchmark study are presented. By application to the GMTKN55 data set, we obtain a comprehensive picture of the performance of range-separated RPA in general main group thermochemistry, kinetics, and noncovalent interactions. The results show that range-separated RPA performs stably over the broad range of molecular chemistry included in the GMTKN55 set. It improves significantly over semilocal DFT but it is still less accurate than modern dispersion corrected double-hybrid functionals. Furthermore, range-separated RPA shows a faster basis set convergence compared to standard full-range RPA making it a promising applicable approach with only one empirical parameter.



1. INTRODUCTION

The random phase approximation (RPA)^{1–6} has become an increasingly popular post-Kohn–Sham (KS)⁷ approach. RPA can be considered as a parameter-free density functional and it stands on the fifth and highest rung of the Jacob’s ladder of density-functional theory (DFT).⁸ RPA overcomes several failures of semilocal density functionals, among which one of the most important issues are the poorly described long-range van der Waals interactions.⁹ This means that RPA gives more accurate interaction and cohesion energies.^{10–14} Even though the long-range part of the dispersion interactions is described well, RPA gives a poor approximation for small interelectronic distances.^{3,15–17}

For this reason the idea of treating the short-range interactions with semilocal DFT arose some time ago.^{16,18–20} Recently, a scheme that combines the long-range part of the RPA correlation energy with the short-range part of a density functional via the error function has been established.^{21–23} This range-separated RPA approach has been shown to improve the RPA correlation energy in various cases. One example is the improvement of dissociation curves for rare-gas dimers and alkaline-earth dimers compared to full-range RPA.^{22,23} It also has been shown that the range-separation approach provides accurate interaction energies for a range of noncovalent complexes.^{24,25} Furthermore, the range-separation scheme improves atomization energies and barrier heights of small test sets.²⁶

Here, we present a range-separated RPA method which is based on our efficient linear-scaling ω -CDGD-RI-RPA method^{27–30} in the local atomic orbital space that uses a Cholesky decomposed ground state density (CDGD) and makes use of the resolution-of-the-identity (RI) with the

attenuated ω -Coulomb metric.³¹ The use of our efficient ω -CDGD-RI-RPA algorithm within the range-separation approach enables us to test range-separated RPA on a large scale and to provide a comprehensive picture of the performance of range-separated RPA. Hence, we compare range-separated RPA to full-range RPA for the GMTKN55 data set.³² This large benchmark set comprises 1505 relative energies based on 2462 single-point calculations on molecules with up to 72 atoms and gives a broad overview of general main group thermochemistry, kinetics, and noncovalent interactions.

2. THEORY

Several schemes for range-separated RPA have been proposed so far.^{22,23,23} The formalism of the range-separation scheme used in this work is described by Toulouse et al. in detail in ref 23. Here, we give a brief overview and rather focus on the description of the long-range formulation of our ω -CDGD-RI-RPA method.²⁸ In the subsequent description $\mu, \nu, \lambda, \sigma$ refer to atomic orbitals (AOs) i, j and a, b refer to occupied and virtual molecular orbitals (MOs), respectively, and $\underline{i}, \underline{j}$ refer to Cholesky orbitals. M, N, P, Q denote auxiliary RI functions. Moreover, Einstein’s sum convention³⁴ is used.

Received: December 29, 2019

Published: April 24, 2020



2.1. Range Separation. The separation of the electron–electron interaction into long-range (lr) and short-range (sr) contributions can be achieved by dividing the electron–electron operator v_{ee} into a long-range electron–electron operator v_{ee}^{lr} and a short-range electron–electron operator v_{ee}^{sr} using the error function and its complementary function as

$$v_{ee} = v_{ee}^{lr} + v_{ee}^{sr} = \frac{\text{erf}(\mu r_{12})}{r_{12}} + \frac{\text{erfc}(\mu r_{12})}{r_{12}} \quad (1)$$

where the adjustable range-separation parameter μ defines the range of the separation.

Until now, multiple formulations of short-range PBE were presented in the literature.^{35–37} In this work the range-separated hybrid PBE functional (RSHPBPE) of Goll et al.³⁸ is used, which utilizes the range-separation scheme in eq 1. A detailed description of this functional is given in ref 38. Its energy

$$E^{\text{RSHPBPE}} = E_{\text{H}} + E_{\text{x}}^{\text{PBE,sr}} + E_{\text{x}}^{\text{HF,lr}} + E_{\text{c}}^{\text{PBE,sr}} \quad (2)$$

is composed of the Hartree energy E_{H} , the short-range exchange $E_{\text{x}}^{\text{PBE,sr}}$, and correlation energy $E_{\text{c}}^{\text{PBE,sr}}$ given by the short-range PBE-like functional and the long-range exact exchange energy $E_{\text{x}}^{\text{HF,lr}}$. E^{RSHPBPE} lacks long-range correlation effects and thus can be corrected with the long-range part of the RPA correlation energy $E_{\text{c}}^{\text{RPA,lr}}$ in a *post*-KS calculation:

$$E^{\text{RSHPBPE+lrRPA}} = E^{\text{RSHPBPE}} + E_{\text{c}}^{\text{RPA,lr}} \quad (3)$$

2.2. Long-Range Formulation of the RPA Correlation Energy. The standard full-range RPA total energy within the adiabatic connection formalism³⁹ is given by

$$E^{\text{RPA}} = E^{\text{HF}} + E_{\text{c}}^{\text{RPA}} \quad (4)$$

where E^{HF} is the Hartree–Fock energy evaluated non-self-consistently on the reference orbitals and $E_{\text{c}}^{\text{RPA}}$ is the RPA correlation energy. Using the fluctuation–dissipation theorem together with the RI approximation, the RPA correlation energy can be expressed after coupling-strength integration as^{4–6}

$$E_{\text{c}}^{\text{RPA}} = \frac{1}{2\pi} \int_0^{+\infty} d\omega \text{Tr}[\ln(1 - \mathbf{X}_0(i\omega)\mathbf{V}) + \mathbf{X}_0(i\omega)\mathbf{V}] \quad (5)$$

where

$$V_{MN} = (\mathbf{C}^{-1})_{MP} \tilde{V}_{PQ} (\mathbf{C}^{-1})_{QN} \quad (6)$$

represents the Coulomb operator in the auxiliary basis with

$$C_{MN} = (M|m_{12}|N) \quad (7)$$

$$\tilde{V}_{MN} = (M|v_{ee}(r_{12})|N) \quad (8)$$

and the RI metric m_{12} . In the presented method the attenuated Coulomb metric

$$m_{12} = \frac{\text{erfc}(\omega_{\text{att}} r_{12})}{r_{12}} \quad (9)$$

with $\omega_{\text{att}} = 0.1 a_0^{-1}$ is used, since it has been shown to constitute a good trade-off between accuracy and locality for fitting the full-range Coulomb operator.³¹ \mathbf{X}_0 denotes the noninteracting density–density response function in the zero-temperature case, also represented in the auxiliary basis. For

the sake of efficiency, \mathbf{X}_0 is calculated in the imaginary time domain

$$X_{0,MN}(i\tau) = G_{0,\mu\nu}(-i\tau) B_{\nu\lambda}^M G_{0,\lambda\sigma}(i\tau) B_{\sigma\mu}^N \quad (10)$$

where $G_0(i\tau)$ is the one-particle Green's function

$$\mathbf{G}_0(i\tau) = \Theta(-i\tau)\underline{\mathbf{G}}_0(i\tau) + \Theta(i\tau)\overline{\mathbf{G}}_0(i\tau) \quad (11)$$

$$\underline{G}_{0,\mu\nu}(i\tau) = C_{\mu i} C_{\nu i} \exp(-(\varepsilon_i - \varepsilon_{\text{F}})\tau)$$

$$\overline{G}_{0,\mu\nu}(i\tau) = -C_{\mu a} C_{\nu a} \exp(-(\varepsilon_a - \varepsilon_{\text{F}})\tau)$$

with the Heaviside step function $\Theta(i\tau)$, the MO coefficients $C_{\mu i}$ and $C_{\mu a}$ as well as the MO energies ε_i and ε_a of the occupied and unoccupied MOs, respectively, and the Fermi level ε_{F} . The three-center integrals $B_{\mu\nu}^M$ are given in Mulliken notation by

$$B_{\mu\nu}^M = (\mu\nu|m_{12}|M) \quad (12)$$

The response function of eq 10 is then transformed into the imaginary frequency domain by a contracted double Laplace²⁷ or, equivalently, cosine⁴⁰ transform according to

$$\mathbf{X}_0(i\omega) = \int_{-\infty}^{+\infty} d\tau \cos(\omega\tau)\mathbf{X}_0(i\tau) \quad (13)$$

to perform the final frequency integration.

The main drawback of pure AO formulations is the unfavorable scaling with the size of the basis set compared to MO formulations. To address this problem, pivoted Cholesky decomposition^{41–43} can be applied to density-type matrices^{28,31} in order to obtain local Cholesky vectors/orbitals which can then be used to transform important quantities in the time-determining steps. In the following, pivoted Cholesky decomposition of a given matrix \mathbf{A} is abbreviated by $\mathbf{A} = \mathbf{L}\mathbf{L}^T$.

Since the one-particle Green's function in the negative imaginary time domain is invariant with respect to projection onto the occupied space, eq 10 can equivalently be expressed as

$$X_{0,MN}(i\tau) = \text{Tr}(\mathbf{P}\mathbf{S}\mathbf{G}_0(-i\tau)\mathbf{S}\mathbf{P}\mathbf{B}^M\mathbf{G}_0(i\tau)\mathbf{B}^N) \quad (14)$$

Cholesky decomposition of the ground state density matrix \mathbf{P} and cyclic permutation within the trace result in

$$X_{0,MN}(i\tau) = \text{Tr}(\mathbf{L}^T\mathbf{S}\mathbf{G}_0(-i\tau)\mathbf{S}\mathbf{L}\mathbf{L}^T\mathbf{B}^M\mathbf{G}_0(i\tau)\mathbf{B}^N\mathbf{L}) \quad (15)$$

and allow the dimensions of the important quantities to be reduced yielding

$$X_{0,MN}(i\tau) = G_{0,j\bar{i}}(-i\tau) B_{i\nu}^M G_{0,\nu\mu}(i\tau) B_{\mu\bar{j}}^N \quad (16)$$

where we defined

$$G_{0,j\bar{i}}(-i\tau) = (\mathbf{L}^T\mathbf{S})_{j\mu} G_{0,\mu\nu}(-i\tau) (\mathbf{S}\mathbf{L})_{\nu\bar{i}} \quad (17)$$

$$B_{i\nu}^M = (\mathbf{L}^T)_{i\mu} B_{\mu\nu}^M \quad (18)$$

The final and most expensive step in the calculation of the response function is then given by

$$X_{0,MN}(i\tau) = B_{j\mu}^M(i\tau) B_{\mu\bar{j}}^N \quad (19)$$

with

$$B_{j\mu}^M(i\tau) = G_{0,j\bar{i}}(-i\tau) B_{i\nu}^M G_{0,\nu\mu}(i\tau) \quad (20)$$

The evaluation of eq 19 formally scales as $\mathcal{O}(N_{\text{aux}}^2 N_{\text{basis}} N_{\text{occ}})$ but can be implemented in an asymptotically linear-scaling fashion using sparse matrix algebra.

To account for the long-range part of the RPA correlation energy only, as required by the presented range-separated functional, the standard Coulomb operator in eq 8 is substituted by the long-range electron–electron operator defined in eq 1 to obtain

$$\tilde{V}_{MN}^{\text{lr}} = (M|v_{\text{ee}}^{\text{lr}}(r_{12})|N) \quad (21)$$

and hence

$$\mathbf{V}_{MN}^{\text{lr}} = (\mathbf{C}^{-1})_{MP} \tilde{V}_{PQ}^{\text{lr}} (\mathbf{C}^{-1})_{QN} \quad (22)$$

This long-range Coulomb operator in the auxiliary basis \mathbf{V}^{lr} is then used in the final expression for the long-range RPA correlation energy according to

$$E_c^{\text{RPA,lr}} = \frac{1}{2\pi} \int_0^{+\infty} d\omega \text{Tr}[\ln(1 - \mathbf{X}_0(i\omega)\mathbf{V}^{\text{lr}}) + \mathbf{X}_0\mathbf{V}^{\text{lr}}] \quad (23)$$

In our standard full-range RPA algorithm, the trace of the matrix logarithm is evaluated using Cholesky decomposition of \mathbf{V} in combination with the Mercator series for $\ln(1 + x)$ according to

$$\text{Tr}[\ln(1 + \mathbf{X}_0(i\omega)\mathbf{V})] = \text{Tr}[\ln(1 + \mathbf{L}^T \mathbf{X}_0(i\omega)\mathbf{L})] \quad (24)$$

$$= 2 \ln \left(\prod_n L'_m \right) \quad (25)$$

where we absorbed the minus sign into the response function and abbreviated the Cholesky decomposition of $1 + \mathbf{L}^T \mathbf{X}_0(i\omega)\mathbf{L}$ by L' . In the presented range-separated RPA algorithm, Cholesky decomposition of the long-range Coulomb operator \mathbf{V}^{lr} has turned out to be problematic in some cases due to very small negative eigenvalues occurring as a reason for numerical inaccuracies. Therefore, Cholesky decomposition of \mathbf{V}^{lr} is avoided by evaluating the trace of the matrix logarithm according to

$$\text{Tr}[\ln(1 + \mathbf{X}_0(i\omega)\mathbf{V}^{\text{lr}})] \quad (26)$$

$$= \text{Tr}[\ln(1 + (\mathbf{V}^{\text{lr}})^{1/2} \mathbf{X}_0(i\omega) (\mathbf{V}^{\text{lr}})^{1/2})]$$

$$= 2 \ln \left(\prod_n L_m \right) \quad (27)$$

where this time \mathbf{L} stems from Cholesky-decomposing $1 + (\mathbf{V}^{\text{lr}})^{1/2} \mathbf{X}_0(i\omega) (\mathbf{V}^{\text{lr}})^{1/2}$. Another alternative avoiding Cholesky decomposition of \mathbf{V}^{lr} is, of course, to simply evaluate the matrix logarithm via diagonalization, which works in any case but comes along with an increased computational cost.

3. COMPUTATIONAL DETAILS

All calculations were performed using the FermiONs++ program package.^{44–46} The self-consistent range-separated hybrid DFT calculations were performed using the short-range PBE functional of ref 38, which was implemented in a development-version of libxc,⁴⁷ and long-range exact exchange. This approach is referred to as “RSHPBPE” in the following. The long-range RPA correlation correction to the RSHPBPE energy was calculated based on these RSHPBPE reference

orbitals using the long-range formulation of the ω -CDGD-RI-RPA method as described above. This range-separated RPA approach is termed “RSHPBPE+lrRPA”. For all range-separated calculations a range-separation parameter of $\mu = 0.5 \text{ a}_0^{-1}$ was used (see also discussion below), unless stated otherwise. Full-range RPA calculations performed on PBE^{48,49} reference orbitals are simply named “RPA” in the following.

All calculations on the GMTKN55 were performed with the Ahlrichs-type split-valence triple- ζ basis set def2-TZVP⁵⁰ and the corresponding auxiliary basis set.⁵¹ The basis set was augmented by diffuse functions for the WATER27, G21EA, AHB21, and IL16 subsets in the same way as for the original calculations on the GMTKN55⁵² to ensure best possible comparability to already existing results of other density functionals. In the WATER27 test set, Dunning’s diffuse s and p functions were applied to oxygen; diffuse s and p functions were applied to non-hydrogen atoms and diffuse s functions to hydrogen in the G21EA, AHB21, and IL16 sets.

Effective-core potentials⁵⁰ were used to replace the core electrons of heavy elements in the HEAVYSB11, HEAVY28, and HALS9 subsets.

For all molecules in the singlet state, closed-shell calculations were performed.

4. RESULTS AND DISCUSSION

4.1. Choice of the Basis Set. Several investigations on the basis set dependence of RSHPBPE+lrRPA indicated that within the range-separated framework a smaller number of basis functions is required for convergence of the RPA energy with respect to the basis set size.^{22,23,26,52} This convergence behavior is caused by the expected exponential convergence of the long-range part of the RPA correlation energy⁵³ and the replacement of the relatively slowly converging short-range part of the RPA correlation by faster converging PBE. As the studies concerning basis set behavior of range-separated RPA rely on a small number of molecular systems, we investigate here the basis set convergence of range-separated RPA energies compared to full-range RPA energies using a larger set of molecules.

We compared RSHPBPE+lrRPA to full-range RPA on the BH76 (barrier heights), BH76RC (reaction energies), and S22 (noncovalent interactions) test sets for different basis sets (detailed results can be found in the Supporting Information). For full-range RPA a rather pronounced basis set dependence can be observed (Figure 1) as the MAD decreases significantly for each of the three subsets going from the triple- to quadruple- ζ basis. The MADs for RSHPBPE+lrRPA, in contrast, vary at most in a range of 0.17 kcal/mol going from def2-TZVP to the larger quadruple- ζ basis set and thus can be considered as sufficiently converged with the def2-TZVP basis set. Further, we want to note that the introduced error by fitting the long-range Coulomb operator with the short-range Coulomb metric is, like for fitting the full-range Coulomb operator, orders of magnitude below the orbital basis set error and the intrinsic error of RPA (see Table S2, Supporting Information). Therefore, the dependence of the results on the quality of the auxiliary basis is assumed to be similar to that of standard RI-RPA which was investigated in ref 5.

Even though the results of full-range RPA are clearly not yet converged with the triple- ζ basis sets, we compare both methods using def2-TZVP as we want to have a fair comparison for practical usage. This means using a basis set that is affordable for many applications. For the performance of

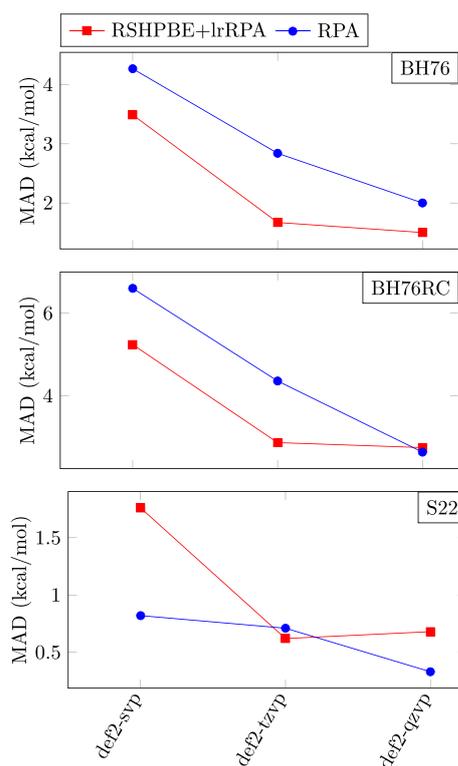


Figure 1. Basis set dependence of the mean absolute deviation (MAD) in kcal/mol for the BH76, BH76RC, and S22 data sets of range-separated RSHPBPE+lrRPA (red) and full-range RPA (blue).

full-range RPA with larger basis sets we refer the interested reader to already existing benchmarks.^{6,54–57}

4.2. Choice of the Range-Separation Parameter. Prior studies investigating the range-separation parameter μ in range-separated methods revealed that its optimal value lies around $0.5 a_0^{-1}$. These prior studies comprise the investigation of the enthalpies of formation for a series of molecules with a combination of srLDA and lrHF exchange⁵⁸ and calculations on atomization energy and barrier height data sets with range-separated RPA.²⁶

It is worth noting here that in the limit $\mu \rightarrow \infty$ the results of RSHPBPE+lrRPA do not converge to the results of conventional full-range RPA based on PBE reference orbitals. In fact, the lrRPA $_{\mu \rightarrow \infty}$ correlation energy formally corresponds to the full-range formulation, but the RSHPBPE (see eq 2) reference orbitals converge to HF orbitals rather than PBE orbitals for $\mu \rightarrow \infty$. This means that RSHPBPE+lrRPA $_{\mu \rightarrow \infty}$ is equal to full-range RPA using HF reference orbitals (RPA@HF, see Figure 2). In the limit of $\mu \rightarrow 0$ the lrRPA correlation energy approaches 0. Thus, RSHPBPE+lrRPA $_{\mu \rightarrow 0}$ approaches the energy of the RSHPBPE reference orbitals, which are identical to those of full-range PBE in the case of $\mu \rightarrow 0$.

To investigate whether a range-separation parameter of $0.5 a_0^{-1}$ is indeed an appropriate choice for a broader range of molecules and properties of molecular systems, we complemented these studies by calculations on the BH76RC, BH76, and S22 data sets with varying range-separation parameter in RSHPBPE+lrRPA. The results (Figure 2, detailed results can be found in the Supporting Information) reveal that the optimal value for μ slightly varies depending on the examined property or system. While for the BH76 and S22 test sets the optimum

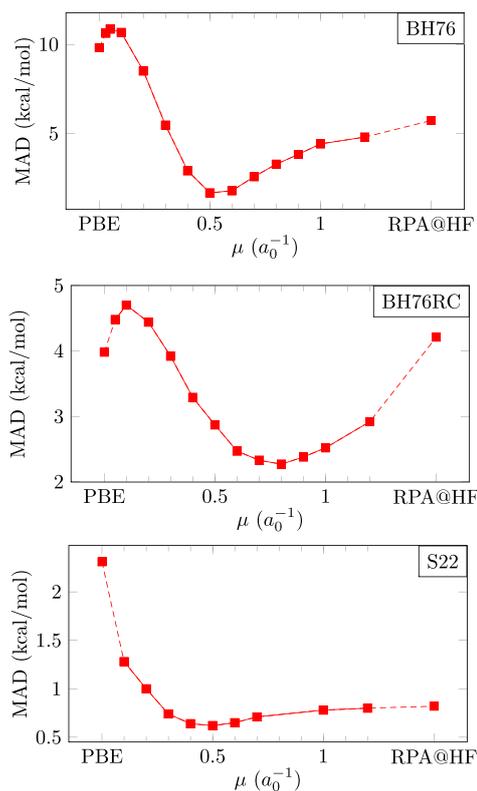


Figure 2. Mean absolute deviation (MAD) for the BH76, BH76RC, and S22 data sets as a function of the range-separation parameter μ for range-separated RSHPBPE+lrRPA calculations using the def2-TZVP basis set. In the limit of $\mu \rightarrow \infty$ RSHPBPE+lrRPA converges to standard RPA evaluated on HF reference orbitals (RPA@HF) and for $\mu \rightarrow 0$ it corresponds to PBE.

of μ lies at $0.5 a_0^{-1}$, it is shifted to a slightly higher value of $0.8 a_0^{-1}$ for the BH76RC test set. A shift of the optimal value of the range-separation parameter to a larger value has also been observed for calculations on reaction energies with a range-separated RPA variant.⁵⁹

Since the results show a quite distinct dependence of the optimal range-separation parameter on the molecular system, we decided to investigate the parameter for an even broader range of molecular systems. We therefore created the set “RAND2x55” which contains two randomly chosen items of each subset of the GMTKN55. The detailed list of contained relative energies can be found in the Supporting Information (Table S1). The absolute values of the relative energies $|\Delta E|$ contained in this test set vary significantly as these describe completely different chemical properties. Items with larger $|\Delta E|$ are expected to give a larger absolute deviation, which in turn leads to a larger change between different μ values. In order to consider each item of the RAND2x55 in the same way for obtaining an optimal range-separation parameter, the absolute deviations of every item are weighted using the weighting factors of weighting scheme 1 of ref 32 for the respective subset. The weighted MADs of the RAND2x55 subset show that there is a broad minimum around $\mu = 0.45 a_0^{-1}$ (see Figure 3) with a deviation of maximally 0.1 kcal/mol in the MADs over the range $\mu = 0.4 a_0^{-1}$ to $\mu = 0.55 a_0^{-1}$. On average, RSHPBPE+lrRPA seems to be quite robust with respect to the choice of μ , reassuring us that the

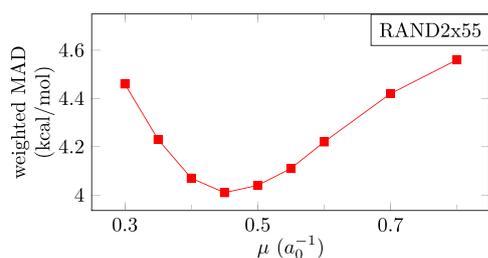


Figure 3. Weighted mean absolute deviation (MAD) for the RAND2 \times 55 data set as a function of the range-separation parameter μ for range-separated RSHPBPE+lrRPA calculations using the def2-TZVP basis set.

choice of $\mu = 0.5 a_0^{-1}$ in previous studies^{23,25,26,58} is reasonable. For this reason a range-separation parameter of $0.5 a_0^{-1}$ was used in the following.

4.3. Results of the GMTKN55 Data Set. The subsets included in the GMTKN55 data set can be grouped into five categories. The first category “basic + small” targets basic properties and reaction energies for small systems. The subsets of the second category “iso + large” comprise reaction energies for large systems and isomerizations. In the third category “barriers”, barrier height test sets are united. The last two subcategories “intermol. NCIs” and “intramol. NCIs” focus on inter- and intramolecular noncovalent interactions, respectively.

As shown in Table 1, RSHPBPE+lrRPA yields a weighted mean absolute deviation according to weighting scheme 1 of

Table 1. Comparison of the WTMAD-1 for the GMTKN55 obtained by RSHPBPE+lrRPA and Full-Range RPA to Density Functionals Grouped by the Rank of the Jacob's Ladder

RSHPBPE+lrRPA	3.86 ^a
RPA	4.72 ^a
GGA	10.70 ^b
meta-GGA	7.31 ^b
hybrid	6.56 ^b
double-hybrid	3.60 ^b

^adef-TZVP basis set, this work. ^bdef2-QZVP basis set and no empirical dispersion correction. Average value taken from ref 32.

ref 32 (WTMAD-1) of 3.86 kcal/mol for the total GMTKN55 data set. With this result RSHPBPE+lrRPA is among the 15% best density functionals tested in ref 32 using the def2-QZVP basis set (see Figure 4) and can be ranked in between the average hybrid and average double-hybrid density functional (see Table 1). It has to be further stressed that the compared (MP2-based) double-hybrid functionals are, due to the inclusion of exchange terms, computationally more expensive than the here presented RPA methods.

The results grouped by category (see Table 2 and Figure 5) show that RSHPBPE+lrRPA is not as good as the average double-hybrid density functional for “basic + small” and “barriers” but is significantly better for NCIs. However, the deficiencies of double-hybrid density functionals in describing noncovalent interactions can be compensated by the inclusion of the empirical “D3” dispersion correction of Grimme.^{60,61}

RSHPBPE+lrRPA gives a slightly better result than full-range RPA (WTMAD-1 of 3.86 kcal/mol vs 4.72 kcal/mol) for the complete GMTKN55 test set. Furthermore, RSHPBPE+lrRPA

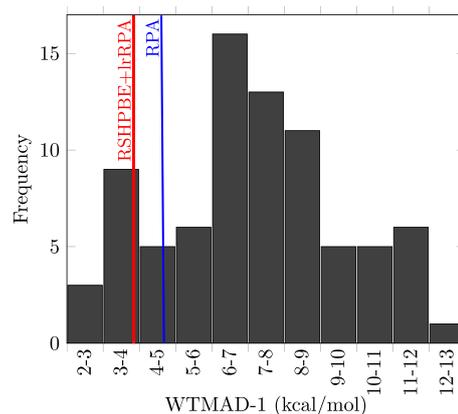


Figure 4. Histogram showing the WTMAD-1 distribution for all tested density functionals without empirical dispersion correction (def2-QZVP) in ref 32 on the total GMTKN55 test set. The red and blue lines illustrate where RSHPBPE+lrRPA and full-range RPA def2-TZVP are placed among the density functionals according to the WTMAD-1.

Table 2. WTMAD-1 Values in kcal/mol for the GMTKN55 Test Set and Its Categories^a

	RSHPBPE	PBE	RSHPBPE+lrRPA	RPA	average double-hybrid	
					no D3	D3
GMTKN55	8.33	8.17	3.86	4.72	3.60	2.05
basic + small	4.92	5.56	3.48	5.41	2.21	1.87
iso. + large	4.97	7.38	3.76	3.10	3.40	2.50
barriers	5.72	7.64	3.56	2.63	1.43	1.59
intermol. NCIs	13.87	10.41	4.27	6.54	5.90	2.02
intramol. NCIs	13.13	11.64	4.40	4.16	5.17	2.39
all NCIs	13.55	10.94	4.33	5.52	5.59	2.18

^aAll calculations were performed using the def2-TZVP basis set. Values for the average double-hybrid functional with and without Grimme's D3 dispersion correction^{60,61} were obtained using the def2-QZVP basis set and are taken from ref 32.

performs more stably over all categories. The WTMAD-1 of RSHPBPE+lrRPA is for all categories about the same and does not show as high fluctuations as the full-range variant. In both cases, range-separated and full-range, the RPA correlation energy on average improves the results of the respective Kohn–Sham reference calculations, RSHPBPE and PBE.

The improvement of the RPA approaches over the respective Kohn–Sham reference is most prominent for the categories concerning noncovalent interactions. Within the subsets of “intermol. NCIs” and “intramol. NCIs” the improvement is most obvious for the IDISP subset which targets intermolecular dispersion interactions (see Table 3). This is not surprising, as RSHPBPE and PBE do not account for any dispersion interactions. Moreover, the remarkably high MAD of RSHPBPE+lrRPA for the WATER27 (hydrogen bonds) subset has to be noted. Apparently, this test set is quite sensitive to the basis set size as all tested methods have a significant deviation in the MAD between the def2-TZVP and def2-QZVP results (see Table 3, values in brackets). This means that for this test set the results of all studied methods, including the references RSHPBPE and PBE, are not sufficiently

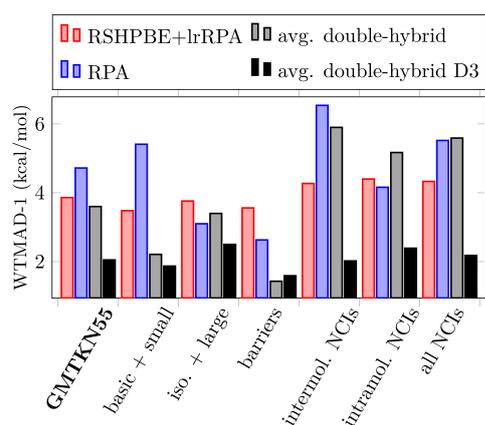


Figure 5. Graphical representation of the WTMAD-1 values for the GMTKN55 test set and its categories. The def2-TZVP basis set was used for RSHBPBE+lrRPA and full-range RPA (this work). The average WTMAD-1s for all tested double-hybrid functionals in ref 32 with (avg. double-hybrid D3) and without (avg. double-hybrid) Grimme's D3 dispersion correction^{60,61} were obtained using the def2-QZVP basis set and are taken from ref 32.

converged with respect to the basis set size at triple- ζ level and are thus not reliable.

For all noncovalent interactions (all NCIs, Table 2), RSHBPBE+lrRPA has a slightly lower WTMAD-1 compared to full-range RPA. This is in line with the observation of Zhu et al.²⁵ that range-separated RPA improves interaction energies of weakly interacting intermolecular complexes. Also, several studies suggest^{17,23,52,62} that a range-separated RPA approach improves interaction energies in rare-gas dimers which we can confirm by the results of the RG18 subset (Table 3).

For reaction barrier heights, varying results for RSHBPBE+lrRPA were obtained. In fact, RSHBPBE+lrRPA has a slightly lower MAD in some reaction barrier height subsets but has also a remarkably higher MAD for the two subsets, PX13 and WCPT18, containing reaction barriers of proton-transfer and -exchange reactions, where water–water interactions, which are also present in the WATER27 test set, play a crucial role. This suggests that the results of PX13 and WCPT18 might also be not sufficiently converged with respect to the basis set size at the triple- ζ level. This is one of the reasons why we have not observed a significant improvement in the description of reaction barrier heights for RSHBPBE+lrRPA over full-range RPA, contradicting the finding of Mussard et al.²⁶ Another reason might be the larger test volume investigated in our present work.

For the category “iso. + large”, a slightly inferior performance of RSHBPBE+lrRPA compared to the full-range variant is observed (3.76 kcal/mol vs 3.10 kcal/mol). In this category, the MADs for the MB16-43 (decomposition of artificial molecules) and DARC (Diels–Alder reaction energies) subsets stand out in particular (Table 3). For the DARC test set the difference in the MADs between RSHBPBE+lrRPA and full-range RPA is remarkable. It should be noted that the errors for this rather specialized test set are mainly systematic as all relative energies contained in this test set describe one single property: the relative stability of a C–C σ bond vs a C–C π bond. The low MAD of full-range RPA arises from a fortuitous error cancellation for this very specific type of reactions. PBE significantly underestimates the relative stability of C–C σ bonds (signed error +6.12 kcal/mol), and the addition of the

full-range RPA correlation compensates this deficiency nearly exactly (signed error +0.48 kcal/mol). In contrast, RSHBPBE already overestimates the relative strength of C–C σ bonds (signed error –1.27 kcal/mol), so that the addition of the long-range RPA correlation results in an even stronger comparative overbinding of σ bonds (signed error –6.79 kcal/mol). However, this error is not unusually large compared to other functionals. The average MAD for all double-hybrid functionals tested in ref 32 without empirical dispersion correction is 4.62 kcal/mol. We also tested the influence of the basis set on this specific test set employing the def2-QZVP basis set instead. The differences in the MADs of RSHBPBE+lrRPA and full-range RPA, however, were found to be smaller than 1 kcal/mol, i.e., this test set is not dominated by basis set incompleteness errors.

For the MB16-43 test set large MADs are not unusual due to the large average of absolute energy differences $|\Delta E|$ of 414.73 kcal/mol. The result of RSHBPBE+lrRPA for this test set is as good as the average result of all double-hybrid functionals tested in ref 32 with 22.91 kcal/mol (without empirical dispersion correction). The MAD of full-range RPA, however, is exceptionally large displaying the deficiency of standard full-range RPA to describe the strength of covalent bonds which is well-known concerning atomization energies.^{54,63–65}

RSHBPBE+lrRPA seems to have an improved performance in basic properties as compared to full-range RPA (“basic + small”, Table 2 and Figure 5). This difference in the WTMAD-1s arises from the stable performance of RSHBPBE+lrRPA compared to the varying results of standard RPA. Here, especially the noticeable high MADs of the W4-11 (atomization energies), SIE4x4 (self-interaction-error related problems), and ALKBDE10 (dissociation energies of group-1 and -2 diatomics) subsets stand out. The obtained results for the atomization energies subset W4-11 are in line with those of Mussard et al.,²⁶ who also observed that range-separated RPA gives more precise atomization energies than the full-range variant. It has to be noted that the large MADs of full-range RPA for atomization energies and dissociation energies arise from the systematical underbinding of standard full-range RPA caused by deficiencies in the description of short-range correlation.^{54,63,65} The poor performance of standard RPA for the self-interaction-error related problems is also not surprising as it is a well-known deficiency of direct RPA. However, the range-separation approach somewhat alleviates this problem, as indicated by the significantly better performance of RSHBPBE+lrRPA in the SIE4x4 test set, confirming the findings of previous work on range-separated RPA.^{33,66} In this context, range-separated RPA may also be regarded as a cost-effective alternative to beyond RPA methods.^{29,67–72}

5. CONCLUSION

In this work we presented a range-separated RPA method, RSHBPBE+lrRPA, based on our efficient linear-scaling ω -CDGD-RI-RPA algorithm.²⁸ Investigations on the basis set dependence revealed that energies obtained by this range-separated method converge faster with respect to the basis set size than full-range RPA energies. For most systems, RSHBPBE+lrRPA yields reliable results with the def2-TZVP basis set. The weaker basis set dependence compared to full-range RPA and the fact that the presented RSHBPBE+lrRPA method is exactly as efficient as the underlying ω -CDGD-RI-RPA algorithm opens up the possibility for efficiently applying

Table 3. Detailed List of the Mean Absolute Deviation in kcal/mol for All Subsets of the GMTKN55 Data Base^a

set	description	RSHPBPE	PBE	RSHPBPE+lrRPA	RPA
Basic Properties and Reaction Energies for Small Systems					
W4-11 ^b	total atomization energies	15.34	14.69	6.94	27.06
G21EA	adiabatic electron affinities	6.43	2.80	3.66	3.39
G21IP	adiabatic ionization potentials	5.09	3.91	4.29	3.41
DIPCS10	double-ionization potentials of closed-shell systems	6.15	4.59	2.94	6.32
PA26	adiabatic proton affinities (incl. of amino acids)	2.53	1.92	1.29	3.88
SIE4x4	self-interaction-error related problems	4.64	23.73	8.63	22.19
ALKBDE10	dissociation energies in group-1 and -2 diatomics	6.19	4.93	4.83	25.00
YBDE18	bond-dissociation energies in ylides	6.99	5.68	2.56	5.28
AL2x6	dimerization energies of ALX _x compounds	6.27	4.04	1.79	2.82
HEAVYSB11	dissociation energies in heavy-element compounds	12.53	4.34	4.97	6.66
NBPRC	oligomerizations and H ₂ fragmentation of NH ₃ /BH ₃ systems	2.62	2.77	1.95	2.53
ALK8	dissociation and other reactions of alkaline compounds	7.09	3.05	3.69	7.79
RC21	fragmentations and rearrangements in radical cations	2.71	6.03	4.09	2.79
G2RC	reaction energies of selected G2/97 systems	5.48	7.50	5.67	7.04
BH76RC	reaction energies of the BH76 set	2.38	3.98	2.87	4.51
FH51	reaction energies in various (in-) organic systems	3.27	4.03	3.31	3.40
TAUT15	relative energies in tautomers	1.18	1.91	0.90	1.19
DC13	13 difficult cases for DFT methods	12.76	10.00	8.49	10.47
Reaction Energies for Large Systems and Isomerization Reactions					
MB16-43	decomposition energies of artificial molecules	49.92	24.24	21.72	60.96
DARC	reaction energies of Diels–Alder reactions	1.61	6.39	6.79	0.92
RSE43	radical-stabilization energies	0.46	3.16	0.53	0.48
BSR36	bond-separation reaction of saturated hydrocarbons	8.43	8.15	0.90	1.88
CDIE20	double-bond isomerization energies in cyclic systems	1.00	1.90	0.69	0.46
ISO34	isomerization energies of small and medium-sized organic molecules	1.70	1.95	1.51	1.43
ISOL24	isomerization energies of large organic molecules	4.74	6.71	3.79	2.01
C60ISO	relative energies between C ₆₀ isomers	23.05	10.48	7.55	7.71
PArel	relative energies in protonated isomers	1.05	1.76	1.05	0.97
Reaction Barrier Heights					
BH76	barrier heights of hydrogen transfer, heavy atom transfer, nucleophilic substitution, unimolecular, and association reactions	3.17	9.82	1.67	2.84
BHPERI	barrier heights of pericyclic reactions	10.74	4.18	1.85	0.73
BHDIV10	diverse reaction barrier heights	5.10	8.24	1.39	1.89
INV24	inversion/racemization barrier heights	3.39	2.95	2.11	1.21
BHROT27	barrier heights for rotation around single bonds	0.90	0.54	0.70	0.75
PX13	proton-exchange barriers in H ₂ O, NH ₃ , and HF clusters	5.07	13.16	7.67	2.36
WCPT18	proton-transfer barriers in uncatalyzed and water-catalyzed reactions	3.59	9.66	3.19	1.68
Intermolecular Noncovalent Interactions					
RG18	interaction energies in rare-gas complexes	0.51	0.36	0.14	0.41
ADIM6	interaction energies of <i>n</i> -alkane dimers	4.54	3.37	1.24	0.30
S22	binding energies of noncovalently bound dimers	3.01	2.31	0.62	0.71
S66	binding energies of noncovalently bound dimers	2.57	1.94	0.72	0.42
HEAVY28	noncovalent interaction energies between heavy element hydrides	1.30	0.49	0.45	0.65
WATER27	binding energies in (H ₂ O) _{<i>n</i>} , H ⁺ (H ₂ O) _{<i>n</i>} , and OH ⁻ (H ₂ O) _{<i>n</i>}	2.27 (5.08)	9.06 (2.84)	11.64 (5.70)	0.89 (3.86)
CARBH12	hydrogen-bonded complexes between carbene analogues and H ₂ O, NH ₃ , or HCl	0.63	1.45	0.59	2.07
PNICO23	interaction energies in pnictogen-containing dimers	1.77	0.86	0.53	1.43
HAL59	binding energies in halogenated dimers (incl. halogen bonds)	1.94	1.36	0.37	1.62
AHB21	interaction energies in anion-neutral dimers	1.22	1.10	1.52	1.33
CHB6	interaction energies in cation-neutral dimers	1.76	1.34	1.68	0.87
IL16	interaction energies in anion–cation dimers	4.29	1.77	0.66	0.95
Intramolecular Dispersion Interactions					
IDISP	intramolecular dispersion interaction	10.72	10.62	2.81	2.63
ICONF	relative energies in conformers of inorganic systems	0.79	0.41	0.43	0.46
ACONF	Relative energies of alkane conformers	0.92	0.58	0.19	0.06
AMINO20x4	Relative energies in amino acid conformers	0.62	0.47	0.27	0.35

Table 3. continued

set	description	RSHPB	PBE	RSHPB+lrRPA	RPA
Intramolecular Dispersion Interactions					
PCONF21	relative energies in tri- and tetrapeptide conformers	3.14	3.20	0.68	1.04
MCONF	relative energies in melatonin conformers	1.83	1.64	0.19	0.64
SCONF	Relative energies of sugar conformers	0.75	0.70	0.49	0.24
UPU23	relative energies between RNA-backbone conformers	2.51	1.87	0.83	0.55
BUT14DIOL	relative energies in butane-1,4-diol conformers	0.19	0.54	0.60	0.14

^aAll calculations were performed using the def2-TZVP basis set. For the WATER27 test set aug-def2-QZVP results are given in brackets. ^bWithout the atomization energy of C₂.

range-separated RPA onto relevant systems with several hundred atoms, as illustrated for the ω -CDGD-RI-RPA method in ref 28 where the largest system comprised 902 atoms.

Investigations on the range-separation parameter μ revealed a shallow minimum between $0.4 a_0^{-1}$ and $0.55 a_0^{-1}$, which is in good agreement with previous findings of $\mu = 0.5 a_0^{-1}$ to be optimal.^{21,23,25,26,38,62}

To give a comprehensive picture of the performance of RSHPB+lrRPA we compared this method to standard RPA on the GMTKN55 data set³² and placed it among previously tested density functionals. The results for GMTKN55 show that RSHPB+lrRPA yields stable results for a broad range of thermochemical and kinetic properties as well as noncovalent interactions. Although the overall performance of RSHPB+lrRPA is comparable to that of full-range RPA, it shows less variance in the WTMAD-1s of the subcategories. It was found that the range-separation approach especially gives better results compared to those of the full-range variant for atomization energies (W4-11), problems that are prone to the self-interaction-error (SIE4x4), and systems containing group-1 and -2 elements (ALKBDE10, ALK8).

Overall, the results of RSHPB+lrRPA are promising considering that only one empirical parameter was employed. In the future, the method could further be improved by including exchange into the response function, e.g., along the lines of the second order screened exchange (SOSEX) RPA method.^{29,67,68,72} Alternatively, more empirical approaches could be explored in a similar fashion as done by Mardirossian and Head-Gordon,⁷³ i.e., employing more empirical semilocal exchange-correlation functionals (e.g., B97⁷⁴), more complicated range-separation schemes, or adding empirical dispersion interaction corrections.

Due to the lower computational cost compared to standard MP2 and the stable results of range-separated RPA over a broad range of chemical problems, this avenue is in our opinion worth considering for future developments.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.9b01294>.

Details on the "RAND2x55" test set (PDF)

Detailed list of the relative energies for all subsets (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Christian Ochsenfeld – Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), D-81377 Munich, Germany; Max Planck Institute for Solid State

Research, D-70569 Stuttgart, Germany; orcid.org/0000-0002-4189-6558; Email: c.ochsenfeld@fkf.mpg.de

Authors

Andrea Kreppel – Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), D-81377 Munich, Germany

Daniel Graf – Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), D-81377 Munich, Germany

Henryk Laqua – Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), D-81377 Munich, Germany

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jctc.9b01294>

Funding

The authors acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG) within the cluster of excellence (EXC2111) "Munich Center for Quantum Science and Technology", MCQST. C.O. further acknowledges financial support as Max-Planck-Fellow at the MPI-FKF Stuttgart.

Notes

The authors declare no competing financial interest.

■ REFERENCES

- Bohm, D.; Pines, D. A collective description of electron interactions: III. Coulomb interactions in a degenerate electron gas. *Phys. Rev.* **1953**, *92*, 609–625.
- Langreth, D. C.; Perdew, J. P. The Exchange-Correlation Energy of a Metallic Surface. *Solid State Commun.* **1975**, *17*, 1425–1429.
- Langreth, D. C.; Perdew, J. P. Exchange-correlation energy of a metallic surface: Wave-vector analysis. *Phys. Rev. B* **1977**, *15*, 2884–2901.
- Furche, F. Developing the random phase approximation into a practical post-Kohn-Sham correlation model. *J. Chem. Phys.* **2008**, *129*, 114105.
- Eshuis, H.; Yarkony, J.; Furche, F. Fast computation of molecular random phase approximation correlation energies using resolution of the identity and imaginary frequency integration. *J. Chem. Phys.* **2010**, *132*, 234114.
- Eshuis, H.; Furche, F. Basis set convergence of molecular correlation energy differences within the random phase approximation. *J. Chem. Phys.* **2012**, *136*, 084105.
- Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- Perdew, J. P.; Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.* **2000**, *577*, 1–20.

- (9) Andersson, Y.; Langreth, D. C.; Lundqvist, B. I. Van der Waals interactions in density-functional theory. *Phys. Rev. Lett.* **1996**, *76*, 102–105.
- (10) Dobson, J. F.; Wang, J. Successful test of a seamless van der Waals density functional. *Phys. Rev. Lett.* **1999**, *82*, 2123–2126.
- (11) Dobson, J. F.; Wang, J.; Dinte, B. P.; McLennan, K.; Le, H. M. Soft cohesive forces. *Int. J. Quantum Chem.* **2005**, *101*, 579–598.
- (12) Harl, J.; Kresse, G. Cohesive energy curves for noble gas solids calculated by adiabatic connection fluctuation-dissipation theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 045136.
- (13) Nguyen, H. V.; Galli, G. A first-principles study of weakly bound molecules using exact exchange and the random phase approximation. *J. Chem. Phys.* **2010**, *132*, 044109.
- (14) Lebègue, S.; Harl, J.; Gould, T.; Ángyán, J. G.; Kresse, G.; Dobson, J. F. Cohesive properties and asymptotics of the dispersion interaction in graphite by the random phase approximation. *Phys. Rev. Lett.* **2010**, *105*, 196401.
- (15) Singwi, K. S.; Tosi, M. P.; Land, R. H.; Sjolander, A. Electron Correlations at Metallic Densities. *Phys. Rev.* **1968**, *176*, 589–599.
- (16) Kurth, S.; Perdew, J. Density-functional correction of random-phase-approximation correlation with results for jellium surface energies. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 10461–10468.
- (17) Ángyán, J. G.; Liu, R. F.; Toulouse, J.; Jansen, G. Correlation energy expressions from the adiabatic-connection fluctuation-dissipation theorem approach. *J. Chem. Theory Comput.* **2011**, *7*, 3116–3130.
- (18) Perdew, J. P. Local density and gradient-corrected functionals for short-range correlation: Antiparallel-spin and non-RPA contributions. *Int. J. Quantum Chem.* **1993**, *48*, 93–100.
- (19) Savin, A. A Combined Density Functional and Configuration Interaction Method. *Int. J. Quantum Chem.* **1988**, *34*, 59–69.
- (20) Kohn, W.; Meir, Y.; Makarov, D. E. Van der Waals Energies in Density Functional Theory. *Phys. Rev. Lett.* **1998**, *80*, 4153–4156.
- (21) Toulouse, J.; Gerber, I. C.; Jansen, G.; Savin, A.; Ángyán, J. G. Adiabatic-connection fluctuation-dissipation density-functional theory based on range separation. *Phys. Rev. Lett.* **2009**, *102*, 096404.
- (22) Janesko, B. G.; Henderson, T. M.; Scuseria, G. E. Long-range-corrected hybrids including random phase approximation correlation. *J. Chem. Phys.* **2009**, *130*, 081105.
- (23) Toulouse, J.; Zhu, W.; Angyan, J. G.; Savin, A. Range-separated density-functional theory with random phase approximation: detailed formalism and illustrative applications. *Phys. Rev. A: At, Mol, Opt. Phys.* **2010**, *82*, 032502.
- (24) Janesko, B. G.; Henderson, T. M.; Scuseria, G. E. Long-range-corrected hybrid density functionals including random phase approximation correlation: Application to noncovalent interactions. *J. Chem. Phys.* **2009**, *131*, 034110.
- (25) Zhu, W.; Toulouse, J.; Savin, A.; Ángyán, J. G. Range-separated density-functional theory with random phase approximation applied to noncovalent intermolecular interactions. *J. Chem. Phys.* **2010**, *132*, 244108.
- (26) Mussard, B.; Reinhardt, P.; Ángyán, J. G.; Toulouse, J. Spin-unrestricted random-phase approximation with range separation: Benchmark on atomization energies and reaction barrier heights. *J. Chem. Phys.* **2015**, *142*, 154123.
- (27) Schurkus, H. F.; Ochsenfeld, C. Communication: An effective linear-scaling atomic-orbital reformulation of the random-phase approximation using a contracted double-Laplace transformation. *J. Chem. Phys.* **2016**, *144*, 031101.
- (28) Graf, D.; Beuerle, M.; Schurkus, H. F.; Luenser, A.; Savasci, G.; Ochsenfeld, C. Accurate and Efficient Parallel Implementation of an Effective Linear-Scaling Direct Random Phase Approximation Method. *J. Chem. Theory Comput.* **2018**, *14*, 2505–2515.
- (29) Beuerle, M.; Graf, D.; Schurkus, H. F.; Ochsenfeld, C. Efficient calculation of beyond RPA correlation energies in the dielectric matrix formalism. *J. Chem. Phys.* **2018**, *148*, 204104.
- (30) Graf, D.; Beuerle, M.; Ochsenfeld, C. Low-Scaling Self-Consistent Minimization of a Density Matrix Based Random Phase Approximation Method in the Atomic Orbital Space. *J. Chem. Theory Comput.* **2019**, *15*, 4468–4477.
- (31) Luenser, A.; Schurkus, H. F.; Ochsenfeld, C. Vanishing-Overhead Linear-Scaling Random Phase Approximation by Cholesky Decomposition and an Attenuated Coulomb-Metric. *J. Chem. Theory Comput.* **2017**, *13*, 1647–1655.
- (32) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.
- (33) Mezei, P. D.; Kállay, M. Construction of a Range-Separated Dual-Hybrid Direct Random Phase Approximation. *J. Chem. Theory Comput.* **2019**, *15*, 6678–6687.
- (34) Einstein, A. Die Grundlage der allgemeinen Relativitätstheorie. *Ann. Phys.* **1916**, *354*, 769–822.
- (35) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *J. Chem. Phys.* **2001**, *115*, 3540–3544.
- (36) Toulouse, J.; Colonna, F.; Savin, A. Short-range exchange and correlation energy density functionals: Beyond the local-density approximation. *J. Chem. Phys.* **2005**, *122*, 014110.
- (37) Laikov, D. N. Simple exchange hole models for long-range-corrected density functionals. *J. Chem. Phys.* **2019**, *151*, 094106.
- (38) Goll, E.; Werner, H.-J.; Stoll, H.; Leininger, T.; Gori-giorgi, P.; Savin, A. A short-range gradient-corrected spin density functional in combination with long-range coupled-cluster methods: Application to alkali-metal rare-gas dimers. *Chem. Phys.* **2006**, *329*, 276–282.
- (39) Gunnarsson, O.; Lundqvist, B. I. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B* **1976**, *13*, 4274–4298.
- (40) Kaltak, M.; Klimeš, J.; Kresse, G. Low scaling algorithms for the random phase approximation: Imaginary time and Laplace transformations. *J. Chem. Theory Comput.* **2014**, *10*, 2498.
- (41) Koch, H.; Sánchez De Merás, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.* **2003**, *118*, 9481–9484.
- (42) Higham, N. J. Cholesky factorization. *Wiley Interdiscip. Rev. Comput. Stat.* **2009**, *1*, 251–254.
- (43) Harbrecht, H.; Peters, M.; Schneider, R. On the low-rank approximation by the pivoted Cholesky decomposition. *Appl. Numer. Math.* **2012**, *62*, 428–440.
- (44) Kussmann, J.; Ochsenfeld, C. Pre-selective screening for matrix elements in linear-scaling exact exchange calculations. *J. Chem. Phys.* **2013**, *138*, 134114.
- (45) 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.
- (46) Kussmann, J.; Ochsenfeld, C. Hybrid CPU/GPU Integral Engine for Strong-Scaling Ab Initio Methods. *J. Chem. Theory Comput.* **2017**, *13*, 3153–3159.
- (47) Marques, M. A. L.; Oliveira, M. J. T.; Burnus, T. Libxc: a library of exchange and correlation functionals for density functional theory. *Comput. Phys. Commun.* **2012**, *183*, 2272.
- (48) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (49) Perdew, J. P.; Ernzerhof, M.; Burke, K. (ERRATA) Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (50) 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.
- (51) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143–152.
- (52) Ireland, R. M.; Henderson, T. M.; Scuseria, G. E. Long-range-corrected hybrids using a range-separated Perdew-Burke-Ernzerhof

functional and random phase approximation correlation. *J. Chem. Phys.* **2011**, *135*, 094105.

(53) Franck, O.; Mussard, B.; Luppi, E.; Toulouse, J. Basis convergence of range-separated density-functional theory. *J. Chem. Phys.* **2015**, *142*, 074107.

(54) Eshuis, H.; Bates, J. E.; Furche, F. Electron correlation methods based on the random phase approximation. *Theor. Chem. Acc.* **2012**, *131*, 1084.

(55) Ren, X.; Rinke, P.; Scuseria, G. E.; Scheffler, M. Renormalized second-order perturbation theory for the electron correlation energy: Concept, implementation, and benchmarks. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *88*, 035120.

(56) Ruzsinszky, A.; Zhang, I. Y.; Scheffler, M. Insight into organic reactions from the direct random phase approximation and its corrections. *J. Chem. Phys.* **2015**, *143*, 144115.

(57) Grimme, S.; Steinmetz, M. A computationally efficient double hybrid density functional based on the random phase approximation. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20926–20937.

(58) Gerber, I. C.; Ángyán, J. G. Hybrid functional with separated range. *Chem. Phys. Lett.* **2005**, *415*, 100–105.

(59) Heßelmann, A.; Ángyán, J. Assessment of a range-separated orbital-optimized random-phase approximation electron correlation method. *Theor. Chem. Acc.* **2018**, *137*, 155.

(60) 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.

(61) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(62) Toulouse, J.; Zhu, W.; Savin, A.; Jansen, G.; Ángyán, J. G. Closed-shell ring coupled cluster doubles theory with range separation applied on weak intermolecular interactions. *J. Chem. Phys.* **2011**, *135*, 084119.

(63) Furche, F. Molecular tests of the random phase approximation to the exchange-correlation energy functional. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2001**, *64*, 195120.

(64) Dobson, J. F.; Gould, T. Calculation of dispersion energies. *J. Phys.: Condens. Matter* **2012**, *24*, 073201.

(65) Chen, G. P.; Voora, V. K.; Agee, M. M.; Balasubramani, S. G.; Furche, F. Random-Phase Approximation Methods. *Annu. Rev. Phys. Chem.* **2017**, *68*, 421–445.

(66) Mezei, P. D.; Csonka, G. I.; Ruzsinszky, A.; Kállay, M. Construction and Application of a New Dual-Hybrid Random Phase Approximation. *J. Chem. Theory Comput.* **2015**, *11*, 4615–4626.

(67) Freeman, D. L. Coupled-cluster expansion applied to the electron gas: Inclusion of ring and exchange effects. *Phys. Rev. B* **1977**, *15*, 5512–5521.

(68) Grüneis, A.; Marsman, M.; Harl, J.; Schimka, L.; Kresse, G. Making the random phase approximation to electronic correlation accurate. *J. Chem. Phys.* **2009**, *131*, 154115.

(69) Bates, J. E.; Furche, F. Communication: Random phase approximation renormalized many-body perturbation theory. *J. Chem. Phys.* **2013**, *139*, 171103.

(70) Mussard, B.; Rocca, D.; Jansen, G.; Ángyán, J. G. Dielectric Matrix Formulation of Correlation Energies in the Random Phase Approximation: Inclusion of Exchange Effects. *J. Chem. Theory Comput.* **2016**, *12*, 2191–2202.

(71) Dixit, A.; Ángyán, J. G.; Rocca, D. Improving the accuracy of ground-state correlation energies within a plane-wave basis set: The electron-hole exchange kernel. *J. Chem. Phys.* **2016**, *145*, 104105.

(72) Beuerle, M.; Ochsenfeld, C. Short-range second order screened exchange correction to RPA correlation energies. *J. Chem. Phys.* **2017**, *147*, 204107.

(73) Mardirossian, N.; Head-Gordon, M. Survival of the most transferable at the top of Jacob's ladder: Defining and testing the ω B97M(2) double hybrid density functional. *J. Chem. Phys.* **2018**, *148*, 241736.

(74) Becke, A. D. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **1997**, *107*, 8554–8560.

Supporting Information for: Range-Separated Density-Functional Theory in Combination with the Random Phase Approximation: An Accuracy Benchmark

Andrea Kreppel,[†] Daniel Graf,[†] Henryk Laqua,[†] and Christian Ochsenfeld^{*,†,‡}

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

[‡]*Max Planck Institute for Solid State Research, Heisenbergstr. 1, D-70569 Stuttgart,
Germany*

E-mail: c.ochsenfeld@fkf.mpg.de

The detailed results of RSHPBE, PBE, RSHBPE+lrRPA, and full-range RPA on the GMTKN55 obtained in this work can be found in the enclosed file 'GMTKN55.xlsx'.

Table S1: Detailed list of the RAND2x55 test set. For each item its number in the original subset (#) is given. The system names correspond to the geometry files of the corresponding test set. The reference values are given in kcal/mol. In the last column, the weighting factor of the corresponding test set in the WTMAD-1 scheme is given.

subset	#	systems	stoichiometry	ref.	w1
W4-11	8	sih si h	-1 1 1	73.921	0.1
W4-11	90	hocl h o cl	-1 1 1 1	166.229	0.1
G21EA	20	EA_20n EA_20	1 -1	9.5	1
G21EA	2	EA_o EA_o-	1 -1	33.7	1
G21IP	36	IP_80 48	1 -1	261.153	0.1
G21IP	22	IP_65 IP_n65	1 -1	234.107	0.1
DIPCS10	2	c2h6 c2h6_2+	-1 1	667.1	0.1
DIPCS10	7	h2s h2s_2+	-1 1	733	0.1
PA26	15	ch3cooh ch3coohp	1 -1	190.9	0.1
PA26	10	h2s h2sp	1 -1	174.3	0.1
SIE4x4	5	he he+ he2+_1.0	1 1 -1	56.9	1
SIE4x4	8	he he+ he2+_1.75	1 1 -1	19.1	1
ALKBDE10	2	beo be o	-1 1 1	106.6	0.1
ALKBDE10	7	lio li o	-1 1 1	82.5	0.1
YBDE18	6	me2s-ch2 me2s ch2	-1 1 1	51.74	1
YBDE18	16	ph3-ch2 ph3 ch2	-1 1 1	60.11	1
AL2x6	4	al2me4 alme2	-1 2	38.4	1
AL2x6	3	al2cl6 alcl3	-1 2	32.5	1
HEAVYSB11	11	br br2	2 -1	53.17	1
HEAVYSB11	4	sh h2s2	2 -1	67.85	1
NBPRC	7	BH3PH3 BH3 PH3	1 -1 -1	-25.2	1
NBPRC	5	nh2-bh2 bz h2	-3 1 3	-48.9	1

ALK8	6	li5_ch li4_c li_h	-1 1 1	66.28	1
ALK8	2	na8 na2	-1 4	53.15	1
RC21	5	3e 3p1 3p2	-1 1 1	57.93	1
RC21	13	6e 6p1 ethylene	-1 1 1	21.21	1
G2RC	10	58 59 57 60	-1 -1 1 1	-27.15	1
G2RC	6	128 13 126 22	-1 -1 1 1	-10.7	1
BH76RC	29	C2H6 NH2 C2H5 NH3	-1 -1 1 1	-6.52	1
BH76RC	13	hmc hcn	-1 1	-15.06	1
FH51	13	2-pentyne H2 trans-2-pentene	-1 -1 1	-44.82	1
FH51	4	C4H9SO2H H2O2 C4H9SO3H H2O	-1 -1 1 1	-82.55	1
TAUT15	9	6a 6b	-1 1	-0.17	10
TAUT15	10	6a 6c	-1 1	-0.87	10
DC13	12	o3 c2h4 o3_c2h4_add	-1 -1 1	-58.7	1
DC13	2	c20cage c20bowl	-1 1	-7.7	1
MB16-43	13	13 H2 CH4 N2 O2 MgH2 S2	-2 -5 4 4 2 2 2	19.8751	0.1
MB16-43	32	32 H2 LiH BH3 N2 F2 AlH3 SiH4 S2	-2 -11 2 6 1 2 2 2 1	685.5818	0.1
DARC	6	ethine chdiene P6	-1 -1 1	-49	1
DARC	3	ethene cpdiene P3	-1 -1 1	-29.9	1
RSE43	42	E44 P1 E1 P44	-1 -1 1 1	-6.7	1
RSE43	13	E15 P1 E1 P15	-1 -1 1 1	-6.4	1
BSR36	26	c2h6 r11 ch4	11 -1 -12	8.93	1
BSR36	21	c2h6 r6 ch4	7 -1 -7	9.78	1
CDIE20	6	R28 P26	-1 1	4	10
CDIE20	20	R60 P60	-1 1	8.6	10
ISO34	20	E20 P20	-1 1	18.12	1
ISO34	24	E24 P24	-1 1	12.26	1

ISOL24	24	i24e i24p	-1 1	15.4	1
ISOL24	9	i9e i9p	-1 1	21.09	1
C60ISO	8	1 9	-1 1	143.96	0.1
C60ISO	7	1 8	-1 1	142.18	0.1
PArel	19	c2cl43 c2cl42	-1 1	2.47	10
PArel	12	sugar0 sugar3	-1 1	3.21	10
BH76	75	C5H8 RKT22	-1 1	39.7	1
BH76	63	h H2S RKT16	-1 -1 1	3.9	1
BHPERI	15	13r_5 13_c2h4 13ts_5a	-1 -1 1	6.5	1
BHPERI	26	09r 00r 09ts	-1 -1 1	31.3	1
BHDIV10	1	ed1 ts1	-1 1	25.65	1
BHDIV10	5	ed5 ts5	-1 1	15.94	1
INV24	3	SO2 SO2_TS	-1 1	60.6	1
INV24	12	Dibenzocycloheptene Dibenzocycloheptene_TS	-1 1	10.3	1
BHROT27	24	ethylthiourea_180 ethylthiourea_TS1	-1 1	10.36	10
BHROT27	22	butadiene_strans butadiene_TS	-1 1	6.3	10
PX13	6	h2o_4 h2o_4_ts	-1 1	26.6	1
PX13	9	hf_2 hf_2_ts	-1 1	42.3	1
WCPT18	8	reac8 ts8	-1 1	28.97	1
WCPT18	7	reac7 ts7	-1 1	32	1
RG18	15	c2h6Ne ne c2h6	-1 1 1	0.24	10
RG18	17	bzNe ne bz	-1 1 1	0.4	10
ADIM6	5	AM6 AD6	2 -1	4.6	10
ADIM6	6	AM7 AD7	2 -1	5.55	10
S22	10	10 10a 10b	-1 1 1	1.448	10
S22	7	7 07a 07b	-1 1 1	16.66	10

S66	2	02A 02B 2	1 1 -1	5.59	10
S66	53	53A 53B 53	1 1 -1	4.36	10
HEAVY28	21	sbh3_nh3 sbh3 nh3	-1 1 1	2.84	10
HEAVY28	11	pbh4_hcl pbh4 hcl	-1 1 1	0.75	10
WATER27	20	OHmH2O OHm H2O	-1 1 1	26.687	0.1
WATER27	3	H2O4 H2O	-1 4	27.353	0.1
CARBH12	1	1O 1O_A 1O_B	-1 1 1	5.37	10
CARBH12	10	2CL 2CL_A 2CL_B	-1 1 1	10.483	10
PNICO23	5	5 5a 5b	-1 1 1	2.86	10
PNICO23	1	1 1a p1b	-1 1 1	1.43	10
HAL59	32	BrBr_FCCH BrBr FCCH	-1 1 1	0.74	10
HAL59	38	BrBr_OCH2 BrBr OCH2	-1 1 1	4.41	10
AHB21	15	15 15A 15B	1 -1 -1	-8.62	1
AHB21	5	5 5A 5B	1 -1 -1	-15.61	1
CHB6	6	27 27A 27B	1 -1 -1	-19.9	1
CHB6	3	24 24A 24B	1 -1 -1	-17.83	1
IL16	1	008 008A 008B	1 -1 -1	-100.41	0.1
IL16	7	187 187A 187B	1 -1 -1	-114	0.1
IDISP	1	antdimer ant	1 -2	-9.15	1
IDISP	4	undecan1 undecan2	1 -1	9.1	1
ICONF	3	N4H6_1 N4H6_2	-1 1	0.13	10
ICONF	4	N4H6_1 N4H6_3	-1 1	2.33	10
ACONF	8	H_ttt H_gtg	-1 1	1.178	10
ACONF	11	H_ttt H_g+x-t+	-1 1	2.632	10
AMINO20x4	59	PRO_xae PRO_xaf	-1 1	4.187	10
AMINO20x4	66	THR_xaq THR_xag	-1 1	3.08	10

PCONF	15	SER_ab SER_aR	-1 1	1.47	10
PCONF	7	99 412	-1 1	2.18	10
MCONF	51	1 52	-1 1	8.75	10
MCONF	41	1 42	-1 1	6.39	10
SCONF	8	C1 C9	-1 1	6.19	10
SCONF	1	C1 C2	-1 1	0.86	10
UPU23	2	2p u1b	-1 1	2.97	10
UPU23	17	2p 7p	-1 1	3.9	10
BUT14DIOL	45	B1 B46	-1 1	3.18	10
BUT14DIOL	21	B1 B22	-1 1	2.74	10

Table S2: Comparison of the WTMAD-1 (kcal/mol) for the RAND2x55 test set using the attenuated Coulomb metric ($\omega = 0.1$) and the standard Coulomb metric to fit the long-range Coulomb operator in the auxiliary basis for two different range-separation values.

μ	ω -Coulomb	Coulomb	Δ
0.45	4.00618	4.00590	-2.74E-04
0.5	4.04240	4.04217	-2.37E-04

A detailed list of all relative energies included in this work can be downloaded from:
https://pubs.acs.org/doi/suppl/10.1021/acs.jctc.9b01294/suppl_file/ct9b01294_si_002.xlsx

7 Conclusion

In publication III, the range-separated RPA method, RSHPBE+lrRPA, was applied to the GMTKN55 benchmark set, whereby a comprehensive overview of the performance of the method in general main group thermochemistry, kinetics and noncovalent interactions was obtained. While it has already been established that range-separated RPA has a better basis set convergence behavior than standard RPA [119, 122, 125–127], we were able to confirm this observation with our calculations on GMTKN55, since a triple- ζ basis set was sufficient for most subsets. However, for a few data sets, such as for the WATER27 test set, a larger base set is required even for RSHPBE+lrRPA.

RSHPBE+lrRPA gives a stable performance over the complete GMTKN55 test set, as it shows less fluctuations in the accuracy between the subsets than standard RPA. According to the results, RSHPBE+lrRPA is comparable with the average double-hybrid without empirical dispersion correction tested in Ref. [118].

Our benchmark revealed that, contrary to expectations, RSHPBE+lrRPA is not particularly good for barrier heights. Here, the average double-hybrid without empirical dispersion correction tested in Ref. [118] gives far better result and even standard RPA is more accurate. This discrepancy between the observed performance of RSHPBE+lrRPA in the description of barrier height compared to the previous study [119] might result from the larger test volume employed in our work. Another reason might be that a triple- ζ basis set for RSHPBE+lrRPA may not be sufficient in some subsets.

In conclusion, RSHPBE+lrRPA has shown to be a promising method based on a stable performance over a broad range of chemical problems using a moderately sized triple- ζ basis set. Further improvement of this method could potentially be achieved by the inclusion of exchange into the response function like in the second order screened exchange (SOSEX) RPA method [189–192]. Alternatively, more empirical approaches could be explored in a similar fashion as done by Mardirossian and Head-Gordon [193], i.e., employing more empirical semilocal exchange-correlation functionals (e.g., B97 [194]), more complicated range-separation schemes, or adding empirical dispersion interaction corrections.

References

- [1] T. Hudlicky, J. W. Reed, *Chem. Soc. Rev.* **2009**, *38*, 3117–3132.
- [2] A. S. Bommarius, M. F. Paye, *Chem. Soc. Rev.* **2013**, *42*, 6534–6565.
- [3] A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, B. Withold, *Nature* **2001**, *409*, 258–268.
- [4] J. M. Choi, S. S. Han, H. S. Kim, *Biotechnol. Adv.* **2015**, *33*, 1443–1454.
- [5] A. Madhavan, R. Sindhu, P. Binod, R. K. Sukumaran, A. Pandey, *Bioresour. Technol.* **2017**, *245*, 1304–1313.
- [6] J. Chapman, A. E. Ismail, C. Z. Dinu, *Catalysts* **2018**, *8*, 238.
- [7] H. Kubinyi, *Curr. Opin. Drug Discov. Dev.* **1998**, *1*, 4–15.
- [8] S. P. Craig, A. E. Eakin, *Vitam. Horm.* **2000**, *58*, 149–169.
- [9] A. Depeursinge, D. Racoceanu, J. Iavindrasana, G. Cohen, A. Platon, P.-A. Poletti, H. Muller, *Artif. Intell. Med.* **2003**, *10*, 787–797.
- [10] M. Marquart, J. Walter, J. Deisenhofer, W. Bode, R. Huber, *Acta Crystallogr. B* **1983**, *39*, 480–490.
- [11] L. Zhou, X. Cheng, B. A. Connolly, M. J. Dickman, P. J. Hurd, D. P. Hornby, *J. Mol. Biol.* **2002**, *27*, 590–609.
- [12] D. W. Banner in *Protein-Ligand Interactions From Molecular Recognition to Drug Design*, WILEY-VCH Verlag GmbH & Co. KGaA, **2003**, Chapter 7.
- [13] K. S. Gates, *Chem. Res. Toxicol.* **2009**, *22*, 1747–1760.
- [14] O. Fleck, O. Nielsen, *J. Cell Sci.* **2004**, *117*, 515–517.
- [15] A. B. Robertson, A. Klungland, T. Rognes, I. Leiros, *Cell. Mol. Life Sci.* **2009**, *66*, 981–993.
- [16] Y.-J. Kim, D. M. Wilson III, *Curr. Mol. Pharmacol.* **2011**, *5*, 3–13.
- [17] S. S. Wallace, *DNA Repair* **2014**, *19*, 14–26.
- [18] J. T. Stivers, *Prog. Nucleic Acid Res. Mol. Biol.* **2004**, *77*, 37–65.
- [19] A. L. Jacobs, P. Schär, *Chromosoma* **2012**, *121*, 1–20.
- [20] E. A. Mullins, A. A. Rodriguez, N. P. Bradley, B. F. Eichman, *Trends Biochem. Sci.* **2019**, *44*, 765–781.
- [21] V. Bailly, W. G. Verly, T. O'Connor, J. Laval, *Biochem. J.* **1989**, *262*, 581–589.
- [22] M. Bhagwat, J. A. Gerlt, *Biochemistry* **1996**, *35*, 659–665.
- [23] S. Sowlati-Hashjin, S. D. Wetmore, *Phys. Chem. Chem. Phys.* **2015**, *17*, 24696–706.
- [24] S. Boiteux, E. Gajewski, J. Laval, M. Dizdaroglu, *Biochemistry* **1992**, *31*, 106–110.
- [25] A. Karakaya, P. Jaruga, V. A. Bohr, A. P. Grollman, M. Dizdaroglu, *Nucleic Acids Res.* **1997**, *25*, 474–479.
- [26] Z. Hatahet, Y. W. Kow, A. A. Purmal, R. P. Cunningham, S. S. Wallace, *J. Biol. Chem.* **1994**, *269*, 18814–18820.

- [27] P. L. McKibbin, A. Kobori, Y. Taniguchi, E. T. Kool, S. S. David, *J. Am. Chem. Soc.* **2012**, *134*, 1653–1661.
- [28] C. Bergonzo, A. J. Campbell, C. D. L. Santos, A. P. Grollman, C. Simmerling, *J. Am. Chem. Soc.* **2011**, *133*, 14504–14506.
- [29] H. Li, A. V. Endutkin, C. Bergonzo, A. J. Campbell, C. de Los Santos, A. Grollman, D. O. Zharkov, C. Simmerling, *Nucleic Acids Res.* **2016**, *44*, 683–94.
- [30] J. Tchou, A. P. Grollman, *J. Biol. Chem.* **1995**, *270*, 11671–11677.
- [31] D. O. Zharkov, R. A. Rieger, C. R. Iden, A. P. Grollman, *J. Biol. Chem.* **1997**, *272*, 5335–5341.
- [32] J. C. Fromme, G. L. Verdine, *Nat. Struct. Biol.* **2002**, *9*, 544–552.
- [33] K. Sadeghian, D. Flaig, I. D. Blank, S. Schneider, R. Strasser, D. Stathis, M. Winnacker, T. Carell, C. Ochsenfeld, *Angew. Chem. - Int. Ed.* **2014**, *53*, 10044–10048.
- [34] I. D. Blank, K. Sadeghian, C. Ochsenfeld, *Sci. Rep.* **2015**, *5*, 10369.
- [35] P. J. Berti, J. A. McCann, *Chem. Rev.* **2006**, *106*, 506–555.
- [36] J. C. Fromme, G. L. Verdine, *J. Biol. Chem.* **2003**, *278*, 51543–51548.
- [37] R. A. Perlow-Poehnelt, D. O. Zharkov, A. P. Grollman, S. Broyde, *Biochemistry* **2004**, *43*, 16092–16105.
- [38] E. I. Zaika, R. A. Perlow, E. Matz, S. Broyde, R. Gilboa, A. P. Grollman, D. O. Zharkov, *J. Biol. Chem.* **2004**, *279*, 4849–4861.
- [39] K. Song, V. Hornak, C. De Los Santos, A. P. Grollman, C. Simmerling, *Biochemistry* **2006**, *45*, 10886–10894.
- [40] K. Song, C. Kelso, C. de los Santos, A. P. Grollman, C. Simmerling, *J. Am. Chem. Soc.* **2007**, *129*, 14536–14537.
- [41] S. Sowlati-Hashjin, S. D. Wetmore, *Biochemistry* **2018**, *57*, 1144–1154.
- [42] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204–220.
- [43] L. D. Moore, T. Le, G. Fan, *Neuropsychopharmacology* **2013**, *38*, 23–38.
- [44] S. Seisenberger, J. R. Peat, T. A. Hore, F. Santos, W. Dean, W. Reik, *Philos. Trans. R. Soc. B Biol. Sci.* **2013**, *368*, 1–11.
- [45] A. Jeltsch, *ChemBioChem* **2002**, *3*, 274–293.
- [46] S. C. Wu, Y. Zhang, *Nat Rev Mol Cell Biol* **2010**, *11*, 607–620.
- [47] R. P. Hausinger, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 21–68.
- [48] W. Li, T. Zhang, J. Ding, *Nucleic Acids Res.* **2015**, *43*, 10026–10038.
- [49] J. Lu et al., *Phys. Chem. Chem. Phys.* **2016**, *18*, 4728–4738.
- [50] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
- [51] Y.-F. He et al., *Science* **2011**, *333*, 1303–1308.
- [52] F. Guo et al., *Cell Stem Cell* **2014**, *15*, 447–459.

- [53] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeyer, M. Truss, T. Carell, *Angew. Chem. - Int. Ed.* **2012**, *51*, 6516–6520.
- [54] S. Schiesser et al., *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.
- [55] K. Iwan et al., *Nat. Chem. Biol.* **2018**, *14*, 72–78.
- [56] S. Xu, W. Li, J. Zhu, R. Wang, Z. Li, G.-L. Xu, J. Ding, *Cell Res.* **2013**, *23*, 1296–309.
- [57] X. Sheng et al., *Front. Chem.* **2018**, *6*, 1–9.
- [58] A. Warshel, M. Strajbl, J. Villa, J. Florian, *Biochemistry* **2000**, *39*, 14728–14738.
- [59] S. Raugai, M. Cascella, P. Carloni, *J. Am. Chem. Soc.* **2004**, *126*, 15730–15737.
- [60] N. Wu, Y. Mo, J. Gao, E. F. Pai, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2017–2022.
- [61] P. Beak, B. Siegel, *J. Am. Chem. Soc.* **1976**, *98*, 3601–3606.
- [62] J. K. Lee, K. N. Houk, *Science* **1997**, *276*, 942–945.
- [63] T. C. Appleby, C. Kinsland, T. P. Begley, S. E. Ealick, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2005–2010.
- [64] T. S. Lee, L. T. Chong, J. D. Chodera, P. A. Kollman, *J. Am. Chem. Soc.* **2001**, *123*, 12837–12848.
- [65] K. Shostak, M. E. Jones, *Biochemistry* **1992**, *31*, 12155–12161.
- [66] R. B. Silverman, M. P. Groziak, *J. Am. Chem. Soc.* **1982**, *104*, 6434–6439.
- [67] A. Schön, E. Kaminska, F. Schelter, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Angew. Chem. Int. Ed.* **2020**, 1–5.
- [68] W. F. van Gunsteren, H. J. C. Berendsen, *Angew. Chem. Int. Ed.* **1990**, *29*, 992–1023.
- [69] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- [70] J. Jones, *Proc. R. Soc. Lond. A* **1924**, *106*, 463–477.
- [71] J. E. Lennard-Jones, *Proc. Phys. Soc.* **1931**, *43*, 461.
- [72] P. Li, K. M. Merz, *J. Chem. Inf. Model.* **2016**, *56*, 599–604.
- [73] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- [74] P. A. M. Dirac, *Proc. R. Soc. Lond. A* **1929**, *123*, 714–733.
- [75] L. H. Thomas, *Math. Proc. Cambridge Philos. Soc.* **1927**, *23*, 542–548.
- [76] E. Fermi, *Zeitschrift für Physik* **1928**, *48*, 73–79.
- [77] P. Hohenberg, W. Kohn, *Phys. Rev. B* **1964**, *136*, 864–871.
- [78] W. Kohn, J. Sham, *Phys. Rev.* **1965**, *140*, A1133–A1138.
- [79] J. P. Perdew, K. Schmidt, *AIP Conf. Proc.* **2001**, *577*, 1–20.
- [80] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

- [81] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.
- [82] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822–8824.
- [83] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- [84] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [85] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
- [86] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200–1211.
- [87] R. O. Jones, *Rev. Mod. Phys.* **2015**, *87*, 897–923.
- [88] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227–249.
- [89] D. Bakowies, W. Thiel, *J. Phys. Chem.* **1996**, *100*, 10580–10594.
- [90] H. M. Senn, W. Thiel, *Angew. Chem. - Int. Ed.* **2009**, *48*, 1198–1229.
- [91] P. Sherwood et al., *J. Mol. Struct. THEOCHEM* **2003**, *632*, 1–28.
- [92] H. Lin, D. G. Truhlar, *Theor. Chem. Acc.* **2007**, *117*, 185–199.
- [93] H. Lin, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 3991–4004.
- [94] S. I. Newton in *The Mathematical Principles of Natural Philosophy*, Benjamin Motte, **1729**, p. 19.
- [95] L. Verlet, *J. Phys. D. Appl. Phys.* **1967**, *159*, 98–103.
- [96] W. C. Swope, H. C. Andersen, P. H. Berens, K. R. Wilson, *J. Chem. Phys.* **1982**, *76*, 637–649.
- [97] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comp. Phys.* **1977**, *23*, 327–341.
- [98] D. Marx, J. Hutter in *Mod. Methods Algorithms Quantum Chem. Vol. 3*, (Ed.: J. Grotendorst), John von Neumann Institute for Computing, Jülich, **2000**, pp. 329–477.
- [99] H. C. Andersen, *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- [100] L. V. Woodcock, *Chem. Phys. Lett.* **1971**, *10*, 257–261.
- [101] M. Parrinello, A. Rahman, *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- [102] H. J. Berendsen, J. P. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- [103] J. Nocedal, S. J. Wright in *Numerical Optimization*, Springer-Verlag, **2006**, pp. 135–163.
- [104] H. B. Schlegel, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 790–809.
- [105] J. D. Lipson, *SYMSAC '76 Proc. third ACM Symp. Symb. Algebr. Comput.* **1976**, 260–270.
- [106] C. G. Broyden, *J. Inst. Math. Appl.* **1970**, *6*, 76–90.
- [107] R. Fletcher, *Comput J* **1970**, *13*, 317–322.
- [108] D. Goldfarb, *Math. Comput.* **1970**, *24*, 23.
- [109] D. F. Shanno, *Math. Comput.* **1970**, *24*, 647.

- [110] S. Metz, J. Kästner, A. A. Sokol, T. W. Keal, P. Sherwood, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 101–110.
- [111] G. Mills, H. Jónsson, *Phys. Rev. Lett.* **1994**, *72*, 1124–1127.
- [112] G. Mills, H. Jónsson, G. K. Schenter, *Surf. Sci.* **1995**, *324*, 305–337.
- [113] H. Jónsson, G. Miles, K. W. Jacobsen, *Classical and Quantum Dynamics in Condensed Phase Simulations* **1998**, 385–404.
- [114] ChemShell, a Computational Chemistry Shell.
- [115] D. Sheppard, R. Terrell, G. Henkelman, *J. Chem. Phys.* **2008**, *128*, 1–10.
- [116] J. H., *J. Chem. Phys.* **2000**, *113*, 9901.
- [117] J. C. Kromann, A. S. Christensen, Q. Cui, J. H. Jensen, *PeerJ* **2016**, *2016*, 1–17.
- [118] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, S. Grimme, *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.
- [119] B. Mussard, P. Reinhardt, J. G. Ángyán, J. Toulouse, *J. Chem. Phys.* **2015**, *142*, 1–7.
- [120] B. G. Janesko, T. M. Henderson, G. E. Scuseria, *J. Chem. Phys.* **2009**, *131*, 034110.
- [121] W. Zhu, J. Toulouse, A. Savin, J. G. Ángyán, *J. Chem. Phys.* **2010**, *132*, 244108.
- [122] J. Toulouse, W. Zhu, G. Angy, A. Savin, *Phys. Rev. A* **2010**, *82*, 032502.
- [123] J. Toulouse, W. Zhu, A. Savin, G. Jansen, J. G. Ángyán, *J. Chem. Phys.* **2011**, *135*, 084119.
- [124] J. G. Ángyán, R. F. Liu, J. Toulouse, G. Jansen, *J. Chem. Theory Comput.* **2011**, *7*, 3116–3130.
- [125] R. M. Irelan, T. M. Henderson, G. E. Scuseria, *J. Chem. Phys.* **2011**, *135*, 094105.
- [126] B. G. Janesko, T. M. Henderson, G. E. Scuseria, *J. Chem. Phys.* **2009**, *130*, 081105.
- [127] O. Franck, B. Mussard, E. Luppi, J. Toulouse, *J. Chem. Phys.* **2015**, *142*, 074107.
- [128] J. Harris, R. O. Jones, *Phys. Lett. A* **1974**, *46*, 407–408.
- [129] D. C. Langreth, J. P. Perdew, *Solid State Commun.* **1975**, *17*, 1425–1429.
- [130] O. Gunnarsson, B. I. Lundqvist, *Phys. Rev. B* **1976**, *13*, 4274–4298.
- [131] R. P. Feynman, *Phys. Rev.* **1939**, *56*, 340–343.
- [132] V. Fock, *Zeitschrift für Physik* **1932**, *75*, 622–647.
- [133] C. A. Ullrich in *Time-Dependent Density-Functional Theory*, Oxford University Press, Inc., **2012**, p. 313.
- [134] A. Szabo, N. Ostlund, Szabo and Ostlund - Modern Quantum Chemistry, **1996**.
- [135] R. Kubo, *Rep. Prog. Phys.* **1996**, *29*, 255–283.
- [136] V. Volterra in *Theory Funct. Integr. Integro-Differential Equations*, Madrid 1927 (Spanish), translated version reprinted New York: Dover Publications, **1959**.
- [137] H. Lehmann, *Il Nuovo Cimento* **1954**, *11*, 342–357.
- [138] J. L. Jackson, *Phys. Rev.* **1951**, *83*, 34–40.

- [139] D. C. Langreth, J. P. Perdew, *Phys. Rev. B* **1977**, *15*, 2884–2901.
- [140] F. Furche, T. Van Voorhis, *J. Chem. Phys.* **2005**, *122*, 164106.
- [141] E. Runge, E. K. U. Gross, *Phys. Rev. Lett.* **1984**, *52*, 997–1000.
- [142] D. Bohm, D. Pines, *Phys. Rev.* **1951**, *82*, 625.
- [143] D. Pines, D. Bohm, *Phys. Rev.* **1952**, *85*, 338–353.
- [144] D. Bohm, D. Pines, *Phys. Rev.* **1953**, *92*, 609–625.
- [145] H. Eshuis, J. E. Bates, F. Furche, *Theor. Chem. Acc.* **2012**, *131*, 1084.
- [146] J. L. Whitten, *J. Chem. Phys.* **1973**, *4496*, 4496–4501.
- [147] M. Feyereisen, G. Fitzgerald, A. Komornicki, *Chem. Phys. Lett.* **1993**, *208*, 359–363.
- [148] D. E. Bernholdt, R. J. Harrison, *Chem. Phys. Lett.* **1996**, *250*, 477–484.
- [149] Y. Jung, A. Sodt, P. M. Gill, M. Head-Gordon, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6692–6697.
- [150] E. Epifanovsky, D. Zuev, X. Feng, K. Khistyayev, Y. Shao, A. I. Krylov, *J. Chem. Phys.* **2013**, *139*, 134105.
- [151] H. F. Schurkus, C. Ochsenfeld, *J. Chem. Phys.* **2016**, *144*, 031101.
- [152] H. F. Schurkus, A. Luenser, C. Ochsenfeld, *J. Chem. Phys.* **2017**, *146*, 211106.
- [153] A. Luenser, H. F. Schurkus, C. Ochsenfeld, *J. Chem. Theory Comput.* **2017**, *13*, 1647–1655.
- [154] D. Graf, M. Beuerle, H. F. Schurkus, A. Luenser, G. Savasci, C. Ochsenfeld, *J. Chem. Theory Comput.* **2018**, *14*, 2505–2515.
- [155] F. Furche, *J. Chem. Phys.* **2008**, *129*, 114105.
- [156] H. Eshuis, J. Yarkony, F. Furche, *J. Chem. Phys.* **2010**, *132*, 234114.
- [157] H. Eshuis, F. Furche, *J. Chem. Phys.* **2012**, *136*, 084105.
- [158] E. J. Baerends, D. E. Ellis, P. Ros, *Chem. Phys.* **1973**, *2*, 41–51.
- [159] B. I. Dunlap, J. W. Connolly, J. R. Sabin, *J. Chem. Phys.* **1979**, *71*, 4993–4999.
- [160] H. J. Werner, F. R. Manby, P. J. Knowles, *J. Chem. Phys.* **2003**, *118*, 8149–8160.
- [161] C. C. Roothaan, *Rev. Mod. Phys.* **1951**, *23*, 69–89.
- [162] V. Dyczmons, *Theor. Chim. Acta* **1973**, *28*, 307–310.
- [163] J. Almlöf, K. Faegri, K. Korsell, *J. Comput. Chem.* **1982**, *3*, 385–399.
- [164] M. Häser, R. Ahlrichs, *J. Comput. Chem.* **1989**, *10*, 104–111.
- [165] T. H. Thompson, C. Ochsenfeld, *J. Chem. Phys.* **2019**, *150*, 044101.
- [166] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [167] T. Bally, G. N. Sastry, *J. Phys. Chem. A* **1997**, *101*, 7923–7925.
- [168] Y. Zhang, W. Yang, *J. Chem. Phys.* **1998**, *109*, 2604–2608.
- [169] M. Weimer, F. Della Sala, A. Görling, *Chem. Phys. Lett.* **2003**, *372*, 538–547.

- [170] J. M. Smith, Y. Jami Alahmadi, C. N. Rowley, *J. Chem. Theory Comput.* **2013**, *9*, 4860–4865.
- [171] W. M. Temmerman, Z. Szotek, H. Winter, *Phys. Rev. B* **1993**, *47*, 11533–11536.
- [172] E. Engel, R. N. Schmid, *Phys. Rev. Lett.* **2009**, *103*, 1–4.
- [173] J. P. Perdew, A. Zunger, *Phys. Rev. B* **1981**, *23*, 5048–5079.
- [174] P. Mori-Sánchez, A. J. Cohen, W. Yang, *J. Chem. Phys.* **2006**, *125*, 201102.
- [175] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, V. N. Staroverov, J. Tao, *Phys. Rev. A - At. Mol. Opt. Phys.* **2007**, *76*, 5–8.
- [176] A. J. Cohen, P. Mori-Sánchez, W. Yang, *Science* **2008**, *321*, 792–794.
- [177] O. A. Vydrov, G. E. Scuseria, *J. Chem. Phys.* **2004**, *121*, 8187–8193.
- [178] C. O. Almbladh, U. Von Barth, *Phys. Rev. B* **1985**, *31*, 3231–3244.
- [179] R. Baer, E. Livshits, U. Salzner, *Annu. Rev. Phys. Chem.* **2010**, *61*, 85–109.
- [180] T. Leininger, H. Stoll, H.-j. Werner, A. Savin, *Chem. Phys. Lett.* **1997**, *275*, 151–160.
- [181] H. Iikura, T. Tsuneda, T. Yanai, K. Hirao, *J. Chem. Phys.* **2001**, *115*, 3540–3544.
- [182] R. Baer, D. Neuhauser, *Phys. Rev. Lett.* **2005**, *94*, 2–5.
- [183] Y. Akinaga, S. Ten-no, *Chem. Phys. Lett.* **2008**, *462*, 348–351.
- [184] I. C. Gerber, J. G. Ángyán, *Chem. Phys. Lett.* **2005**, *416*, 370–375.
- [185] E. Goll, T. Leininger, F. R. Manby, A. Mitrushchenkov, H.-J. Werner, H. Stoll, *Phys. Chem. Chem. Phys.* **2008**, *10*, 3353–3357.
- [186] E. Goll, H. J. Werner, H. Stoll, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3917–3923.
- [187] E. Goll, H.-J. Werner, H. Stoll, T. Leininger, P. Gori-giorgi, A. Savin, *Chem. Phys.* **2006**, *329*, 276–282.
- [188] A. Heßelmann, J. Ángyán, *Theor. Chem. Acc.* **2018**, *137*, 1–13.
- [189] D. L. Freeman, *Phys. Rev. B* **1977**, *15*, 5512–5521.
- [190] A. Grüneis, M. Marsman, J. Harl, L. Schimka, G. Kresse, *J. Chem. Phys.* **2009**, *131*, 154115.
- [191] M. Beuerle, C. Ochsenfeld, *J. Chem. Phys.* **2017**, *147*, 204107.
- [192] M. Beuerle, D. Graf, H. F. Schurkus, C. Ochsenfeld, *J. Chem. Phys.* **2018**, *148*, 204104.
- [193] N. Mardirossian, M. Head-Gordon, *J. Chem. Phys.* **2018**, *148*, 241736.
- [194] A. D. Becke, *J. Chem. Phys.* **1997**, *107*, 8554–8560.