

# Non-standard computational approaches applied to molecular systems

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Julius-Maximilians-Universität Würzburg



Julius-Maximilians University, Würzburg  
Faculty of Chemistry  
Institute of physical and theoretical chemistry

**author:** Dustin Kaiser  
from Würzburg, Germany

**examinant:** Prof. Dr. Bernd Engels

**submission:** March 2021, Würzburg

Eingereicht bei der Fakultät für Chemie und Pharmazie am

03. März 2022

Gutachter der schriftlichen Arbeit

1. Gutachter: Prof. Dr. Bernd Engels

2. Gutachter: Prof. Dr. Volker Engel

Prüfer des öffentlichen Promotionskolloquiums

1. Prüfer: Prof. Dr. Bernd Engels

2. Prüfer: Prof. Dr. Volker Engel

3. Prüfer: Prof. Dr. Ingo Fischer

4. Prüfer: Prof. Dr. Jens Pflaum

5. Prüfer: Jun.-Prof. Dr. Ann-Christin Pöppler

Datum des öffentlichen Promotionskolloquiums

09. Juni 2022

Doktorurkunde ausgehändigt am

---

*I am very seldom interested in applications. I am more interested in the elegance of a problem. Is it a good problem, an interesting problem?*

*Claude Shannon*

The present work was conducted at the Faculty of Chemistry of the Julius-Maximilians University, Würzburg under the supervision of Prof. Dr. Bernd Engels.

This semi-cumulative thesis contains excerpts of peer-reviewed scientific publications. Please note that the indexation of the references in these excerpts is not identical to the indexation of the references from the additional chapters in this thesis.



## I. Abstract

In this thesis, several contributions to the understanding and modeling of chemical phenomena using computational approaches are presented. These investigations are characterized by the usage of non-standard computational modeling techniques, which is necessitated by the complex nature of the electronic structure or atomic fluctuations of the target molecules.

In Sections 3, 6, 5 & 4 of this work, multiple biradical-type molecules and their spectroscopic properties were modeled. The complicated open-shell electronic structure of these short-lived molecular species may present a challenge for Density Functional Theory (DFT) based modeling. In the course of the investigation, it is found that especially the impact of correct molecular geometries on the computationally predicted absorption properties may be critical. In order to find the correct minimum geometries, Multi-Reference (MR) methods often have to be invoked.

The impact of geometry relaxation and geometry changes in general on the excitonic properties of Perylene Bisimide (PBI) dimers were investigated. The results of this study are given in Section 2. Oftentimes, these geometry factors are neglected in Organic Semiconductor (OSC) modeling as an approximation. This present investigation suggests that this approximation is not always valid, as certain regimes are identified where geometrical parameters have critical impact on the localization and energetic properties of excitons.

In Section 7 the chemical mechanism of the Triazolinedione (TAD) tyrosine bioconjugation reaction is investigated using quantum-chemical methods. To this end, MR computations were invoked and compared with the results of DFT based approaches. By comparison of different conceivable mechanisms and their energetic ordering, the TAD tyrosine bioconjugation is found to proceed by means of a base-mediated electrophilic aromatic substitution reaction.

In the final part of this thesis (Section 8), the  $k^{\text{th}}$  nearest neighbor (kNN) entropy estimation protocol is investigated in detail. The kNN entropy estimator promises highly accurate entropy estimates even for flexible biomolecules with multiple structural minima. However, applications of the estimator in the realm of computational chemistry have been scarce, as an unfavorably high dependence on the number of molecular geometry snapshots has been reported. Our granular investigation of many formal and practical properties of the kNN estimator suggests that the uneven variance of a molecule's vibrational modes is the cause of the observed slow convergence of the

estimator with respect to the number of samples. A rescaling procedure to reestablish fast convergence is suggested and benchmarks are performed.

## II Table of Contents

<b>I Abstract</b>	<b>I</b>
<b>II Table of Contents</b>	<b>III</b>
<b>III List of Figures</b>	<b>V</b>
<b>IV Table of Abbreviations</b>	<b>VII</b>
<b>V Notation and Constants</b>	<b>X</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Aim of the Work . . . . .	1
1.2 Computational Modeling of Excitons in Organic Semiconductors . . . . .	3
1.3 Computational Modeling of Spectroscopic Properties of Biradical Compounds . . . . .	8
1.4 Quantum-Chemical Investigation of Reaction Mechanisms . . . . .	24
1.5 Computational Modeling of Absolute Entropies for Single-Molecule Systems . . . . .	27
<b>2 Geometry Mediated Localization of Excitons</b>	<b>32</b>
<b>3 Spectroscopic Modeling of Methylbismuth</b>	<b>50</b>
<b>4 Pentadiynylidene and Its Methyl-Substituted Derivates</b>	<b>57</b>
<b>5 Photochemistry of the Diphenylacetylene Biradical</b>	<b>67</b>
<b>6 Rationalizing the Femtosecond Dynamics of Diphenylproynylidene</b>	<b>75</b>
<b>7 The Bioconjugation Reaction of TADs with Tyrosine</b>	<b>87</b>
<b>8 Entropy Estimation</b>	<b>100</b>
8.1 Theoretical Foundations . . . . .	100
8.1.1 Entropy Derived from Statistical Assumptions . . . . .	100
8.1.2 Calculating the Entropy from the Partition Function . . . . .	103
8.2 Entropies of Chemical Systems . . . . .	105
8.2.1 Additive Decomposition of Entropic Contributions . . . . .	105
8.3 Quasi-Harmonic Entropy Estimation Approaches based on Frequency Assignment . . . . .	108
8.3.1 Derivation of a Frequency Assignment based Quasi-Harmonic Entropy Estimator . . . . .	110
8.3.2 Orthogonal Collective Modes from Simulation Data: The Principal Component Representation . . . . .	113
8.4 Differential Entropies from Molecular Simulation via Ergodicity . . . . .	115

---

8.5	Properties of Differential Entropies . . . . .	118
8.5.1	The Mathematical Relation of Differential and Discrete Entropies	120
8.5.2	Translation and Scaling of Differential Entropies . . . . .	123
8.5.3	The Mutual Information Expansion of the Differential Entropy .	124
8.5.4	Multivariate Gaussian Entropy as a Strict Upper Bound . . . . .	126
8.6	The kNN Entropy Estimator - Theory and Formal Considerations . . . .	129
8.7	Errors of the kNN Entropy Estimator . . . . .	137
8.7.1	Asymptotic Bias . . . . .	138
8.7.2	Non-Additivity of Marginal Entropies . . . . .	141
8.7.3	Errors at Domain Boundaries . . . . .	153
8.7.4	Errors due to Different Domain Ranges of Individual Dimensions	156
8.7.5	Considerations for Applying the kNN Entropy Estimator to MD Simulation Data . . . . .	175
8.8	Computational Implementation of Entropy Estimation Methods in the CAST Program Package . . . . .	177
8.9	Benchmark of the kNN Entropy Estimator using a GM Model Derived from MD Simulations . . . . .	182
8.9.1	Background & Description of the Benchmark Design . . . . .	182
8.9.2	Computational Details . . . . .	186
8.9.3	Results for the Alanine Dipeptide GM Model . . . . .	187
<b>9</b>	<b>Summary</b>	<b>199</b>
<b>10</b>	<b>Zusammenfassung</b>	<b>201</b>
<b>11</b>	<b>References</b>	<b>204</b>
<b>12</b>	<b>Acknowledgement</b>	<b>218</b>
	<b>Appendix</b>	<b>I</b>

### III. List of Figures

Fig. 1	Processes occurring upon light absorption in bulk heterojunction (BHJ) organic solar cells . . . . .	4
Fig. 2	Layout of modern OSC variants using the bulk heterojunction (BHJ) approach: ternary and tandem cells . . . . .	5
Fig. 3	Schematic representation of a biradical molecule with its two possible electronic states and its associated Lewis mesomeric formula . . . . .	9
Fig. 4	Possible wavefunctions (eigenstates) of a biradical-type molecule given in terms of Frontier Molecular Orbital (FMO) occupation . . . . .	10
Fig. 5	Schematic representation of the Frontier Molecular Orbital (FMO) p-orbitals of a biradical-type molecule . . . . .	10
Fig. 6	Creation of pentadiynylidene ( $\text{HC}_5\text{H}$ ) and its derivatives from precursor molecules <i>in situ</i> . . . . .	18
Fig. 7	Structural isomers of pentadiynylidene ( $\text{HC}_5\text{H}$ ) . . . . .	20
Fig. 8	Conformations of interest of diphenylacetylene ( $\text{Ph}-\text{C}_2-\text{Ph}$ ) . . . . .	22
Fig. 9	Creation of diphenylpropynylidene ( $\text{Ph}-\text{C}_3-\text{Ph}$ ) from precursor molecules . . . . .	23
Fig. 10	Reaction of the TAD reagent with a tyrosine amino acid residue of a protein . . . . .	26
Fig. 11	Spatial probability density of a classical harmonic oscillator . . . . .	111
Fig. 12	Divergence of the classical harmonic entropy in the quantum regime . . . . .	120
Fig. 13	Quantization of a continuous probability distribution . . . . .	121
Fig. 14	Geometric depiction of the kNN entropy estimation algorithm's approach . . . . .	132
Fig. 15	Numerical values of different proposed bias corrections for the kNN entropy estimator . . . . .	140
Fig. 16	Exemplification of the errors of the kNN entropy estimate at the domain boundary . . . . .	155
Fig. 17	Percentage of points at the domain boundary for a multivariate gaussian with exponentially decreasing standard deviations . . . . .	158
Fig. 18	Overestimation of the kNN entropy estimate for two dimensions with uneven variances . . . . .	159
Fig. 19	Multivariate gaussian entropy benchmark with exponentially decreasing standard deviations for $N_{\text{samples}} = 10000$ . . . . .	160
Fig. 20	Multivariate gaussian entropy benchmark with exponentially decreasing standard deviations for $N_{\text{samples}} = 125$ . . . . .	161
Fig. 21	Multivariate gaussian entropy benchmark with exponentially decreasing standard deviations for $d = 10$ . . . . .	165
Fig. 22	Multivariate gaussian entropy benchmark with exponentially decreasing standard deviations for $d = 6$ . . . . .	166
Fig. 23	Percentage of points at the domain boundary for a multivariate gaussian with linearly decreasing standard deviations . . . . .	167

---

Fig. 24	Multivariate gaussian entropy benchmark with linearly decreasing standard deviations for $d = 10$ . . . . .	168
Fig. 25	Scheme explaining the conception of the Gaussian-Mixture (GM) Model for the alanine dipeptide . . . . .	184
Fig. 26	Lewis structure of the alanine dipeptide . . . . .	187
Fig. 27	Spectrum of the Principal Component Analysis (PCA) eigenvalues across all alanine dipeptide Molecular Dynamics (MD) replica . . .	188
Fig. 28	Histogrammed probability density obtained from the simulation data of the alanine dipeptide (two principal components with highest associated variance) . . . . .	189
Fig. 29	Histogrammed probability density obtained from the Gaussian-Mixture (GM) model of the alanine dipeptide (two principal components with highest associated variance) . . . . .	190
Fig. 30	Convergence of the multivariate gaussian estimate of the differential entropy of the Gaussian-Mixture (GM) model system for the alanine dipeptide . . . . .	191
Fig. 31	kNN entropy estimates for the five dimensional Gaussian-Mixture (GM) model of the alanine dipeptide, using the $\ell_2$ distance norm . .	193
Fig. 32	kNN entropy estimates for the five dimensional Gaussian-Mixture (GM) model of the alanine dipeptide, using the $\ell_\infty$ distance norm .	194
Fig. 33	Convergence of the full dimensional kNN entropy estimate for the five dimensional Gaussian-Mixture (GM) model of the alanine dipeptide, using the $\ell_2$ distance norm . . . . .	195

## IV. Table of Abbreviations

<b>BHJ</b>	bulk heterojunction
<b>BLAS</b>	Basic Linear Algebra Subprograms
<b>CAS</b>	complete active space
<b>CASPT2</b>	2 <sup>nd</sup> Order Expansion of the Complete Active Space Self Consistent Field Wavefunction
<b>CASSCF</b>	Complete Active Space Self Consistent Field
<b>CAST</b>	Conformational Analysis and Search Tool
<b>CC</b>	Coupled-Cluster
<b>CCSD</b>	Coupled-Cluster Singles Doubles
<b>CCSD(T)</b>	Coupled-Cluster Singles Doubles with approximate Triples
<b>CT</b>	Charge-Transfer
<b>DFT</b>	Density Functional Theory
<b>DNA</b>	deoxyribonucleic acid
<b>DOF</b>	Degree of Freedom
<b>dPCA</b>	dihedral Principal Component Analysis
<b>ECP</b>	Effective Core Potential
<b>EPR</b>	Electron Paramagnetic Resonance
<b>FC</b>	Franck Condon
<b>FEP</b>	Free Energy Perturbation
<b>FMO</b>	Frontier Molecular Orbital
<b>GM</b>	Gaussian-Mixture
<b>HF</b>	Hartree-Fock
<b>HO</b>	Harmonic Oscillator
<b>HOMO</b>	highest occupied molecular orbital
<b>IE</b>	Ionization Energy
<b>kNN</b>	k <sup>th</sup> nearest neighbor
<b>LAPACK</b>	Linear Algebra Package
<b>LUMO</b>	lowest unoccupied molecular orbital

<b>MD</b>	Molecular Dynamics
<b>MI</b>	Mutual Information
<b>MIE</b>	Mutual Information Expansion
<b>MIST</b>	Mutual Information Spanning Tree
<b>MM</b>	Molecular Mechanics
<b>MO</b>	molecular orbital
<b>MP2</b>	Second-Order Møller-Plesset Perturbation Theory
<b>MR</b>	Multi-Reference
<b>ms-TPE</b>	mass-selected Threshold Photoelectron
<b>N/A</b>	not available
<b>NM</b>	Normal Mode
<b>NMA</b>	Normal Mode Analysis
<b>NVE</b>	Microcanonical ensemble; total number of particles (symbol: N) & system's volume (symbol: V) & total energy (symbol: E) are constant
<b>NVT</b>	Canonical ensemble; total number of particles (symbol: N) & system's volume (symbol: V) & temperature (symbol: T) are constant
<b>OSC</b>	Organic Semiconductor
<b>PBI</b>	Perylene Bisimide
<b>PCA</b>	Principal Component Analysis
<b>PCM</b>	Polarizable Continuum Method
<b>PES</b>	Potential Energy Surface
<b>PGA</b>	Principal Geodesic Analysis
<b>PMF</b>	Potential of Mean Force
<b>QHA</b>	Quasi-Harmonic Approximation
<b>QM</b>	quantum mechanical
<b>RRHO</b>	Rigid-Rotor Harmonic-Oscillator
<b>SCS</b>	Spin Component Scaling
<b>SI</b>	Supplementary Information
<b>SOMO</b>	singly occupied molecular orbital
<b>TAD</b>	Triazolinedione



**TDDFT** Time-Dependent Density Functional Theory

**TI** Thermodynamic Integration

**TPEPICO** Threshold Photoelectron-Photoion Coincidence

**TPES** Threshold Photoelectron Spectroscopy

**WHAM** Weighted Histogram Analysis Method

**ZPVE** zero-point vibrational energy

## V. Notation and Constants

$\vec{a}$  vector

$(\vec{a})_i$   $i$ -th entry of a vector

$\mathbf{A}$  matrix

$\mathbf{A}^t$  transposed

$\hat{\mathbf{A}}$  estimator

# cardinality

$|\mathbf{A}|$  determinant

$A[f(x)]$  functional

$a \equiv b$   $a$  is defined to be equivalent to  $b$

$a \stackrel{!}{=} b$   $a$  must be equivalent to  $b$

$\forall$  for all

$\langle \rangle$  mean

$\nabla$  gradient (*nabla* operator)

$\mathbb{E}$  expectation value

$\delta(x)$  Dirac delta distribution

$\delta_{ij}$  Kronecker delta

$h$  Planck constant

$R$  universal gas constant

$N_A$  Avogadro constant

$\hbar$  reduced Planck constant

$\gamma_E$  Euler-Mascheroni constant

# 1. Introduction

## 1.1. Aim of the Work

Theoretical analysis of chemical reactions and molecules *in silico* has become a widely accepted procedure and has been routinely used in recent years and decades.<sup>[1]</sup> Frequently, quantum chemical Density Functional Theory (DFT) based methods are used to obtain geometries, pathways and energies, as they enable fairly accurate calculations at feasible computational costs.<sup>[2]</sup>

Nevertheless, it is widely known that in some cases conceptually simple quantum chemical approaches (such as DFT) fail in the prediction of pathways, geometries, properties, or energies. These cases include, but are not limited to, bond-dissociation processes, spectroscopy, excitonic properties, transition states and biradical molecules.<sup>[3]</sup> In some of these cases, Multi-Reference (MR) methods have to be used in order to correctly describe the electronic structure at hand. MR methods are often not only computationally demanding, but also require the researcher to specify further parameters relating to the electronic structure of the molecule at hand. Thus, these calculations cannot be automatically run in a black-box manner. Obtaining accurate and chemically meaningful insights from MR calculations most often implies detailed analysis and consideration of the attained numerical results.<sup>[4]</sup>

For other properties, such as the thermodynamic parameters like the entropy and heat capacity, the accuracy of common quantum chemical approaches also depends on the energy landscape described by the Potential Energy Surface (PES) of a molecule: Flexible biopeptides often fluctuate between multiple structural minima in aquatic solutions at room temperature.<sup>[5,6]</sup> The computational identification of these structural minima is a challenging problem by itself.<sup>[7]</sup> For these kinds of molecules, single-conformational quantum chemical approaches may be conceptually flawed due to the neglect of other contributing conformational structures, irrespective of the quality of the quantum chemical method used to calculate the electronic structure.

This thesis strives to investigate chemical and spectroscopic properties of molecules which - for one reason or another - may be unfeasible to be tackled by routine single-structure DFT computations. The unique electronic structure of open-shell biradical compounds is one example where the electronic structure of the chemical species often demands the usage of sophisticated quantum chemical approaches. The same is true for transition state structures occurring in mechanistic pathways of the Triazolinedione (TAD) tyrosine bioconjugation reaction. The highly accurate 2<sup>nd</sup> Order

Expansion of the Complete Active Space Self Consistent Field Wavefunction (CASPT2) MR approach was used to assess the validity of DFT computations in this case. For the Perylene Bisimide (PBI) dimer, other factors than the computational accuracy of the molecular electronic structure are relevant for reliable *in silico* modeling. We investigated especially the impact of structural relaxation on excitonic properties, which are of high interest in applications of organic photovoltaics and Organic Semiconductors (OSCs). Due to the computational complexity involved in a proper treatment of these intra-molecular degrees of freedom (DOFs), their influence was neglected in modeling studies of excitonic molecular properties in the past, such as in [8, 9]. Our detailed investigation proves that correct computational predictions require the inclusion of these structural effects as well. It has to be mentioned that the closed-shell electronic structure and the singly excited states of the PBI dimer system may well be treated with modern DFT based approaches, unlike the previously mentioned biradicaloid compounds. This was shown by exhaustive prior benchmark studies.<sup>[10–12]</sup> For all chemical systems mentioned up until now, entropic contributions have been fully neglected in the computational investigations. This is an approximation which works well for the rigid PBI molecule as well as the small biradical compounds, as the comparisons of experimental and theoretical results included in this thesis highlights. When macromolecules, such as enzymes, are to be investigated, entropic effects can be very significant and the neglect of them is not justified anymore.<sup>[13]</sup> Nevertheless, reliable methods for the accurate and efficient calculation of single-molecule entropies are scarce: Molecules found in predominantly one conformational minimum may be treated using the simple Rigid-Rotor Harmonic-Oscillator (RRHO) approximation, which performs well in these cases. Flexible molecules need novel approaches, and no commonly applied method has been established until now.<sup>[13]</sup> From the various entropy estimation approaches which are currently being pursued scientifically, we have chosen to re-investigate the  $k^{\text{th}}$  nearest neighbor (kNN) entropy estimator, which promises high-accuracy due to its non-parametric nature. Although the method is cautiously being used in some research efforts, it is of a complicated nature and its mathematical structure presents some difficulties.<sup>[14,15]</sup> In this work, we give a full derivation of the estimator from the exact expression of the quantum-mechanical entropy for a statistical system. Each approximation is highlighted and the effects of these approximations are mentioned. In a second step, we highlight some pitfalls of the kNN entropy estimator. A benchmark system built from Gaussian-Mixture (GM) models is devised and the performance of different variants of the kNN estimator are

evaluated.

In summary, the focus of this work lies especially on the investigation of complex chemical phenomena using sophisticated theoretical and computational approaches. We strive to make clear for which chemical systems the usage of non-standard computational approaches becomes necessary and how they differ from routine approaches.

## 1.2. Computational Modeling of Excitons in Organic Semiconductors

Due to climate change, robust and efficient renewable energy sources need to be developed and established in a timely manner.<sup>[16,17]</sup> Solar cells are widely known representatives of this new generation of more environmentally friendly energy generation methods, and the class of organic solar cells appears to have many advantages over the more established silicon-based solar cells. These include potentially lower manufacturing costs, increased material flexibility and transparency, and more energy-efficient (or CO<sub>2</sub>-neutral) solar cell manufacturing.<sup>[18,19]</sup> These advantages imply that once organic solar cells are able to deliver economical electricity yields, they will become ubiquitous. However, this is not yet the case. After the realization of the first organic solar cell by Tang in 1985,<sup>[20]</sup> their performance has been continuously improved. Nevertheless, the efficiencies of 18% achieved so far (see [21] and [22]) are not sufficient to compete with existing silicon-based solar-panels from an economical perspective. These low efficiencies arise because the conversion of light to electricity involves several very complex processes (see Figure 1), and severe losses can occur in all processes.<sup>[18,19]</sup>

Research on organic bilayer solar cells showed that the coupled electron-hole pairs created upon absorption of light are very short-lived.<sup>[23,24]</sup> These electron-hole pairs are commonly referred to as excitons, which may be treated as a quasi-particle. To remedy this, the bulk heterojunction (BHJ) concept was introduced, in which acceptor and donor layers are mixed. This ensures that the nearest interface between donor and acceptor layer is never far from the location of exciton formation.<sup>[25,26]</sup> Figure 1 shows the schematic layout of a BHJ solar cell. To further increase the efficiency of solar cells, tandem cells as well as ternary cell architectures have been developed.<sup>[27,28]</sup> The latter contain three instead of two different chemical compounds paired with each other. The third compound is usually an additional donor-type molecule which absorbs light at different wavelengths and thus creates excitons from spectral frequencies which

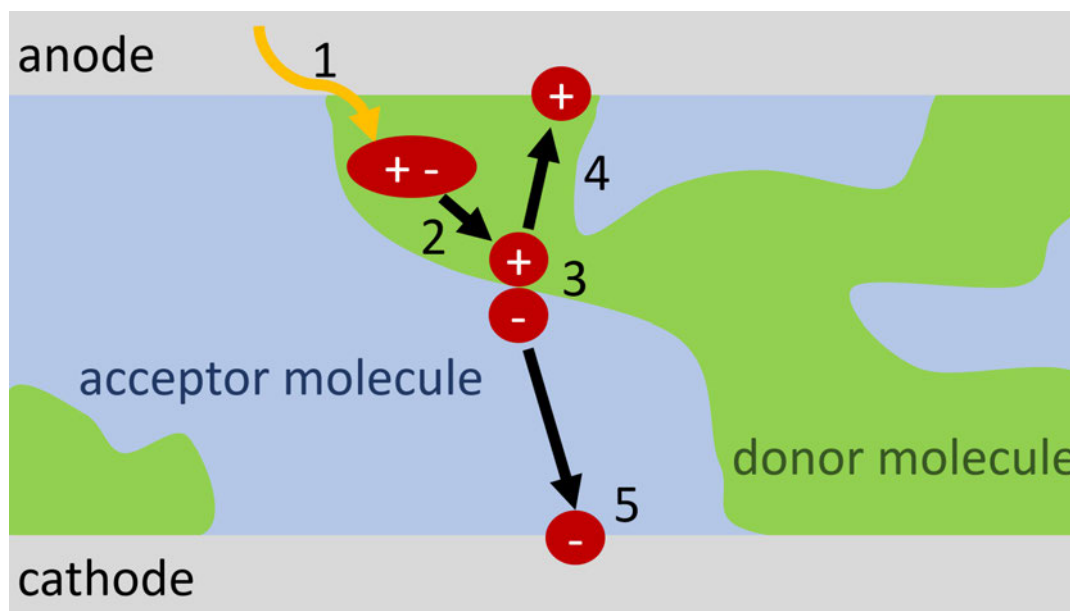
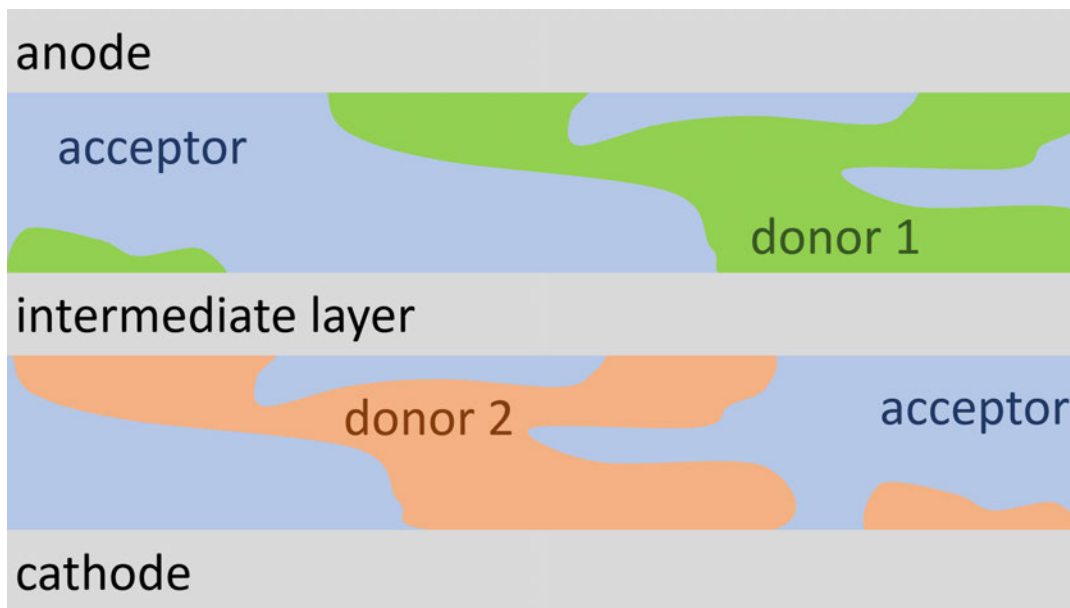


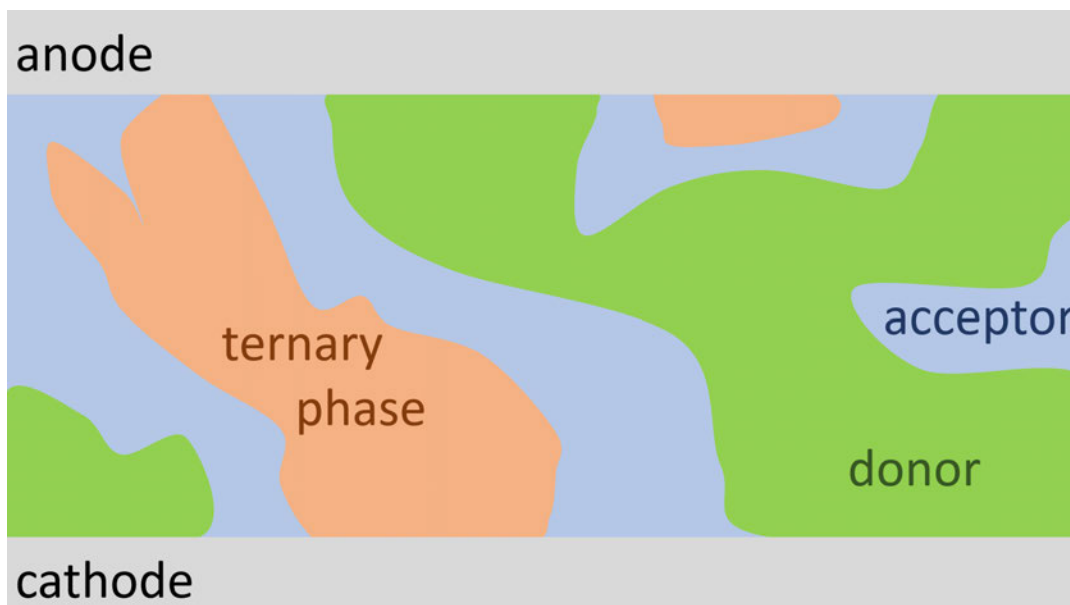
Figure 1: The physico-chemical processes occurring upon light absorption in bulk heterojunction (BHJ) organic solar-cells are depicted. 1: Absorption of light by the donor material. 2: Formation of an exciton. 3: Diffusion of the exciton to the donor-acceptor-interface. 4 & 5: Charge-Carrier Diffusion to the anode and cathode of the cell.

would otherwise not be convertible. Unlike tandem cells, which have two photoactive layers, ternary BHJ solar cells retain only one photoactive layer (built from three photoactive substances), which makes their manufacturing easier. The third molecular substance in ternary BHJ cells might either be of donor- or acceptor-type. Both solar-cell layouts are shown in Figure 2, which should be contrasted with the classical BHJ layout depicted in Figure 1. Fullerenes were in the spotlight as acceptor-molecules for a long time.<sup>[29,30]</sup> In recent years, the investigation of so-called non-fullerene acceptors has intensified again, as these acceptor-molecules are often easier to tune towards the desired end-product specifications.<sup>[31]</sup>

Theoretical studies on organic solar cells underline the importance of the interface between acceptor and donor for the efficiency of the organic solar cell.<sup>[32,33]</sup> It must be designed in such a way that the excitons actually reach the interface and can efficiently change into a Charge-Transfer (CT) complex there. At the same time, the design of the interface must ensure that both charges separate completely from the CT complex into a charge-separated state. The donor and acceptor moieties should be prevented from returning to the ground state of the molecules by recombination of the charge carriers.<sup>[19,34]</sup> The relative time scales of the competing physical processes are strongly influenced by the morphology and structure of the interface. Investigations at the



(a) Schematics of a tandem solar cell layout



(b) Schematics of a ternary solar cell layout

Figure 2: Two layout of modern OSC variants using the BHJ approach are shown. While both designs use three distinct molecular substances, they are only fully mixed in the ternary solar cell. In case of the tandem solar cell, an intermediate layer separates two bilayer BHJ mixtures.

interface are complicated by the fact that the various processes are occurring almost simultaneously. For this reason they are difficult to differentiate experimentally.<sup>[35]</sup> Charge separation may occur with low efficiency because the associated process at the interface is inefficient or because the exciton does not reach the interface at all. Therefore, special attention must be paid to the interplay of processes.

Within current research, the theoretical description of interfaces in organic solar cells has been particularly advanced by the work of Brédas and coworkers.<sup>[36–41]</sup> One focus of their work is the influence of single-molecule properties on the efficiency of the processes leading from photon absorption to electric current generation (see Figure 1). In a recent research paper, Barbatti and Thiel emphasize the influence of the size, position and orientation of the molecules on the photophysics.<sup>[42]</sup> Experimental work by Wang *et al.* showed that the size of the donor-acceptor network has a significant influence on the efficiency of charge separation.<sup>[43]</sup> This is of high relevance for performing accurate theoretical simulations of the excitonic processes of organic solar-cells, as this finding highlights that a suitable aggregate size needs to be taken into consideration. The work of Heremans confirms predictions of Brédas *et al.* on energy and charge transport, which emphasizes the dependence of transport properties at the interface (and thus cell performance) on the crystallographic environment.<sup>[44,45]</sup> Beljonne and Castet, like Heremans, claim that especially so-called “multiscale” methods give good results. These methods capture a moiety of interest using very accurate theoretical approaches, while the surrounding area around this center of interest is captured using a less accurate but computationally faster numerical method.<sup>[46]</sup> However, depending on the method used, theoretical predictions may also be in conflict. In contrast to Brédas, who highlights the great importance of “hot” charge transfer excitons in charge separation, Van Voorhis postulates the “cold” exciton breakup mechanism for the formation of free charge carriers at the interfaces of systems he investigated.<sup>[47,48]</sup> The discrepancy of both approaches underlines that processes in organic semiconductors are still not adequately understood. Bäessler did early pioneering work on charge transport and exciton diffusion in disordered organic semiconductors.<sup>[49,50]</sup> An important recent contribution by Bäessler and Köhler addresses this controversial question of whether charge-transfer systems decay preferentially by a hot or cold mechanism. Bäessler and Köhler conclude that “cold” charge-transfer states are more significant.<sup>[51]</sup>

Andrienko has conducted various theoretical investigations of organic-semiconductor interfaces in recent years. His work especially highlights the influence of electrostatic forces resulting from certain long-range arrangements of molecules in layers of or-



organic semiconductors. Andrienko *et al.* describe that these forces can be exploited to enhance charge separation at interfaces. From this work, the first rules of thumb for the efficiency of organic solar cells are emerging, indicating that a compromise between different energetic and structural parameters must always be found.<sup>[52,53]</sup> In review articles, Andrienko describes the state of research and important unresolved issues in detail.<sup>[54,55]</sup>

In earlier theoretical investigations of Engels and coworkers on small-molecule organic semiconductor compounds, material-dependent exciton diffusion lengths could be attributed to relaxation processes in which two or more monomers move relative to each other. As a consequence, the exciton may be trapped by this motion.<sup>[56,57]</sup> The predicted properties for different organic molecules may be influenced by the morphology of their crystal structures. An aggregate-based approach has become characteristic for the investigations of Engels and coworkers, as their numerical simulations showed the importance of exciton delocalization over aggregates.<sup>[58]</sup> The so-called "dimer method" uses a dimer as the smallest conceivable molecular aggregate. Although it is often-times important to consider larger aggregates to correctly reproduce experimental findings and spectroscopic data,<sup>[59]</sup> the dimer method serves as a conceptually simple model system. It helps to rationalize the impact of monomer movement on absorption and emission. It may also efficiently account for the relevant photo-induced relative motions which influence the relaxation and movement of the excitonic quasi-particles. In more recent work, a program has been developed that allows all of the processes described in Figure 1 to be captured simultaneously. Previous theoretical studies mostly focused on the characterization of one or two processes.<sup>[8,9]</sup>

In the course of the previous research of Engels and coworkers, the inter-monomer movement and its impact has been analyzed in detail using quantum chemical methods. Furthermore, the impact of the choice of the basis-set and the level-of-theory has been benchmarked.<sup>[10,12,60]</sup> One often-neglected aspect of the modeling of exciton pathways in molecular aggregates is the impact of intra-molecular vibrations and movements on the exciton.

In this work, we conduct a detailed survey of the PES of the PBI molecule. To this end, we investigate the impact of geometrical distortions on the excitonic properties of a molecular dimer using interpolation techniques. Detailed results are presented in Chapter 2, where the associated publication (see [61] "Geometry relaxation-mediated localization and delocalization of excitons in organic semiconductors: A quantum

chemical study”) is reprinted. Our results can especially highlight the importance of intra- and inter-molecular DOFs for exciton localization. For weak coupling strengths between monomers, a double-well potential of the PES is found for the first excited singlet state. A harmonic potential is predicted for the second excited singlet state. This is in line with considerations derived from a Pseudo Jahn-Teller type interaction,<sup>[62]</sup> but can be rationalized using Marcus Theory as well.<sup>[63]</sup> The exciton is found to be mainly localized on one PBI monomer for the first excited state. Due to symmetry reasons, the excitonic state is delocalized over both PBI monomers for the second electronic excited singlet state. For strong couplings between the monomers, the exciton remains delocalized in the first three excited singlet states. The character of these states, i.e. whether they possess charge-transfer character, was evaluated using transition density matrix analysis.<sup>[64]</sup> For strong inter-monomer couplings, the energetically lowest two singlet excited states show no charge-transfer character and are thus so-called “Frenkel” states,<sup>[65]</sup> while the third excited state does. The coupling strength can be influenced and controlled by the placement of the two monomers with respect to each other. This is done by variation of their longitudinal shift and their inter-monomer distance. These findings hold for computational modeling in vacuum as well as in a solvent environment.

### 1.3. Computational Modeling of Spectroscopic Properties of Biradical Compounds

Biradicals are open-shell molecules with two unpaired electrons located in degenerate or nearly degenerate molecular orbitals. They occur in combustion processes,<sup>[66]</sup> astrochemistry,<sup>[67,68]</sup> and as highly reactive transient intermediate species.<sup>[69-71]</sup> A molecule with an even number of electrons usually possesses a closed-shell singlet ground state. Due to certain electronic properties, this rule-of-thumb may also be violated and molecular ground-states may be open-shell biradical singlet or triplet states, or so-called charge separated states (see Figure 3). The group of biradical molecules is further subdivided into the subgroup of “true” biradicals, which have unpaired electrons at two different centers in energetically degenerate molecular orbitals (MOs), and so-called biradicaloids with a closed-shell singlet ground state but low-lying open-shell electronically excited states. Further chemical species that are sometimes grouped in the category of biradical-type molecules are carbenes with a triplet or open-shell singlet ground state, and molecules with long-lived, open-shell, or

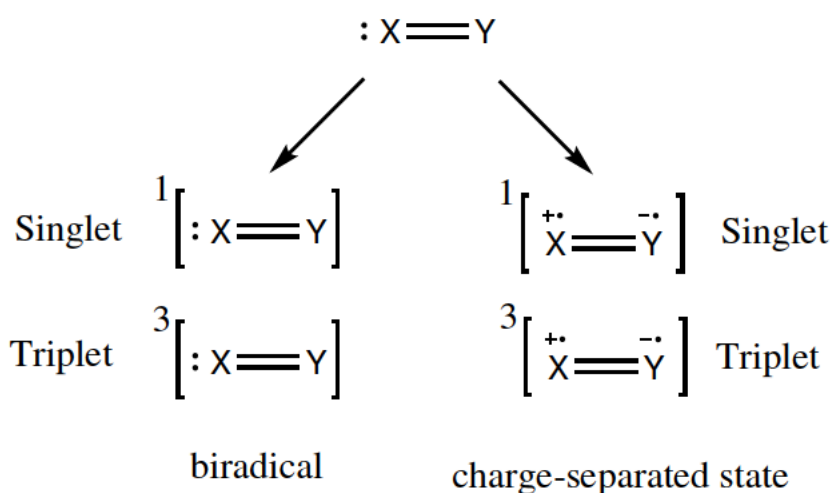


Figure 3: Schematic representation of a biradical-type molecule with its two possible electronic states and associated Lewis mesomeric formula.

charge-separated states. A common characteristic of all molecules in these mentioned categories is that the energetic distances between the electronic ground state and the excited state are small.<sup>[72]</sup> This implies that highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) orbitals may show couplings for energetic reasons or that the LUMO may be partially populated according to the Boltzmann distribution at non-zero temperatures. The physical and chemical properties as well as the geometry of biradical molecules is often determined by the presence of unpaired electrons.<sup>[73]</sup> Figure 3 shows possible electronic configurations that a biradical can adopt. In the valence bond picture, several of the Lewis structures outlined often contribute to the wave function of the respective ground and excited states. Which of the states shown is dominantly occupied also depends on the environment. Thus, neutral biradical states (left side) dominate in the gas phase, while charge-separated states are found in polar solvents. This is due to the fact that these electronical configurations are stabilized and energetically lowered by interactions with the solvent.<sup>[73-75]</sup> If the degeneracy of the molecular orbitals is removed in a biradical, a closed-shell ground state with an energetically very low excited triplet state results. These are precisely the class of molecules termed biradicaloids by Michl and Koutecky.<sup>[76]</sup> A true biradical can be converted into a biradicaloid by adjusting the chemical substituents or ambient effects.<sup>[73,77]</sup>

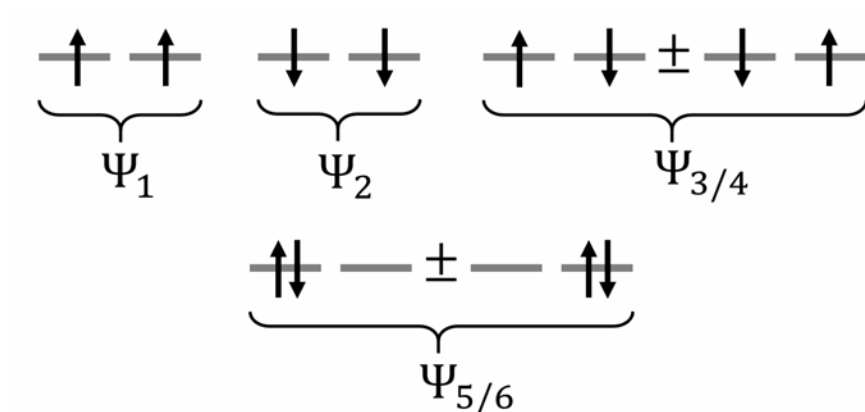


Figure 4: Possible wavefunctions (eigenstates) of a biradical-type molecule are given in terms of Frontier Molecular Orbital (FMO) occupation. For  $\Psi_{3/4}$ , a linear combination of the two possible determinants has to be used in order to yield wavefunctions which are eigenfunctions of the  $\hat{S}^2$  operator. For  $\Psi_{5/6}$ , a linear combination of two determinants has to be considered due to their energetic degeneracy.

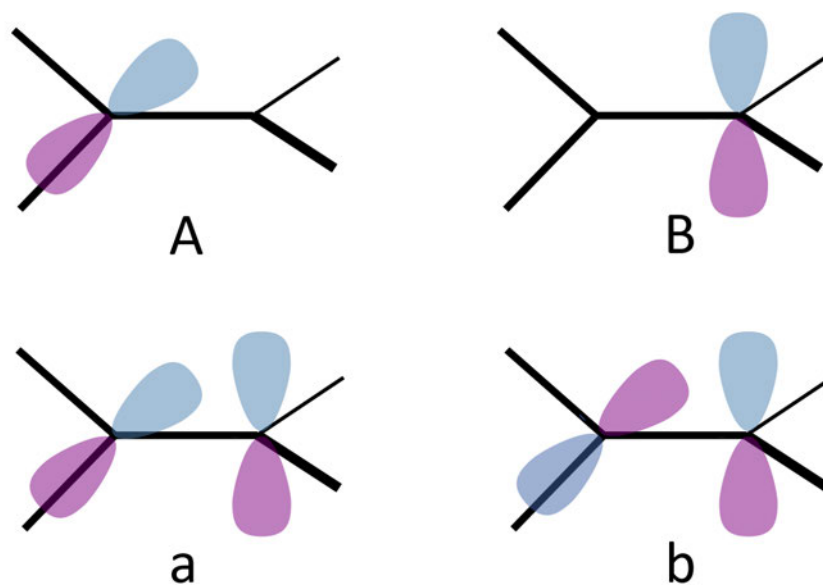


Figure 5: Schematic representation of the Frontier Molecular Orbital (FMO) p-orbitals of a biradical-type molecule. A localized and a delocalized picture is given. Localized orbitals are denoted with upper-case letters A and B, while delocalized orbitals are denoted with the corresponding lower-case letters. Further discussion is given in the text.

A more detailed, Frontier Molecular Orbital (FMO) based description of the electronic structure shall now be presented. The explanations are given in the spirit of Michl and Bonačić-Koutecký (see [76, 78]), and follow the “Discussion of the 4 · 4 CI model to describe biradicaloid systems” given in the Supplementary Information of [79]. Figure 4 shows all six possible determinants that can result from distributing two electrons in two degenerate spatial FMOs. The discussion can proceed using either delocalized or localized spatial orbitals (see Figure 5). While localized spatial orbitals are perhaps more common in mechanistic chemical discussion, delocalized orbitals resemble the results of numerical quantum chemical calculations, such as Hartree-Fock (HF) MOs, more closely. The localized and delocalized orbitals may easily be inter-converted (see Equation 2).

The determinants given in Figure 4 each correspond to one electronic eigenstate of the molecule. Three of these states have singlet multiplicity, and three have triplet multiplicity. While the triplet states are energetically degenerate, the singlet-states might have different energies, depending on their interaction. It should be noted that spin-orbit coupling is fully neglected in this discussion. The explicit mathematical expressions of the wavefunctions of the states are given as

$$\begin{aligned}
 \Psi_1 &= \frac{1}{\sqrt{2}} \underbrace{\alpha(1)\alpha(2)}_{\text{symmetric}} \underbrace{[\phi_1(r_1)\phi_2(r_2) - \phi_1(r_2)\phi_2(r_1)]}_{\text{antisymmetric}} \\
 \Psi_2 &= \frac{1}{\sqrt{2}} \underbrace{\beta(1)\beta(2)}_{\text{symmetric}} \underbrace{[\phi_1(r_1)\phi_2(r_2) - \phi_1(r_2)\phi_2(r_1)]}_{\text{antisymmetric}} \\
 \Psi_4 &= \frac{1}{2} \underbrace{([\phi_1(r_1)\phi_2(r_2) - \phi_1(r_2)\phi_2(r_1)])}_{\text{antisymmetric}} \cdot \underbrace{[\alpha(1)\beta(2) + \alpha(2)\beta(1)]}_{\text{symmetric}} \\
 \Psi_3 &= \frac{1}{2} \underbrace{([\phi_1(r_1)\phi_2(r_2) + \phi_1(r_2)\phi_2(r_1)])}_{\text{symmetric}} \cdot \underbrace{[\alpha(1)\beta(2) - \alpha(2)\beta(1)]}_{\text{antisymmetric}} \\
 \Psi_5 &= \frac{1}{2} \underbrace{([\phi_1(r_1)\phi_1(r_2) - \phi_2(r_1)\phi_2(r_2)])}_{\text{symmetric}} \cdot \underbrace{[\alpha(1)\beta(2) - \alpha(2)\beta(1)]}_{\text{antisymmetric}} \\
 \Psi_6 &= \frac{1}{2} \underbrace{([\phi_1(r_1)\phi_1(r_2) + \phi_2(r_1)\phi_2(r_2)])}_{\text{symmetric}} \cdot \underbrace{[\alpha(1)\beta(2) - \alpha(2)\beta(1)]}_{\text{antisymmetric}},
 \end{aligned} \tag{1}$$

where  $\alpha$  and  $\beta$  are the spin-functions and  $\phi(r)$  designates the spatial one-electron functions. Tables 1 and 2 state the matrix elements of the Hamiltonian of this model

system in the basis of the three singlet states written in Equation 1 and one arbitrary triplet state. Localized orbitals are denoted with upper-case letters A and B, while delocalized orbitals are denoted with lower-case letters. The energy contributions and associated matrix elements calculated using localized orbitals are given in the set of Equations 2. The conversion conventions for evaluations of the expressions using delocalized orbitals are given in Equations 2 (adapted from the SI in [79]) as well.

Table 1: Non-relativistic Hamiltonian matrix elements for the three singlet states and one arbitrary triplet state of a model biradical. The matrix elements are given in terms of integrals between localized orbitals. For the corresponding mathematical expressions, see Equations 1 and 2.

	$^1 A^2 + B^2\rangle$	$^1 A^2 - B^2\rangle$	$^1 AB\rangle$	$^3 AB\rangle$
$^1 A^2 + B^2\rangle$	$E_{\text{triplet}} + 2(K'_{AB} + K_{AB})$	$\delta_{AB}$	$\gamma_{AB}$	0
$^1 A^2 - B^2\rangle$	$\delta_{AB}$	$E_{\text{triplet}} + 2K'_{AB}$	$\gamma_{AB}^-$	0
$^1 AB\rangle$	$\gamma_{AB}$	$\gamma_{AB}^-$	$E_{\text{triplet}} + 2K_{AB}$	0
$^3 AB\rangle$	0	0	0	$E_{\text{triplet}}$

$$\begin{aligned}
K_{AB} &= \iint A(1) B(1) \frac{1}{r_{12}} A(2) B(2) d\tau_1 d\tau_2 = K'_{ab} \\
(AA | AB) &= \iint A(1) A(1) \frac{1}{r_{12}} A(2) B(2) d\tau_1 d\tau_2 \\
h_{AB} &= \int A(1) h(1) B(1) d\tau_1 \\
J_{AB} &= \iint A(1) A(1) \frac{1}{r_{12}} B(2) B(2) d\tau_1 d\tau_2 = \delta_{ab} \\
E_{\text{triplet}} &= E_0 - (K'_{AB} + K_{AB}) \\
E_0 &= h_{AA} + h_{BB} + \frac{J_{AA} + J_{BB}}{4} + \frac{J_{AA}}{2} \\
2K'_{AB} &= \frac{1}{2} \left( \frac{J_{AA} + J_{BB}}{2} - J_{AB} \right) = 2K_{ab} \\
\delta_{AB} &= h_{AA} - h_{BB} + \frac{J_{AA} - J_{BB}}{2} = \gamma_{ab} \\
\gamma_{AB}^- &= (AA | AB) - (BB | BA) \\
\gamma_{AB} &= 2h_{AB} + (AA | AB) + (BB | BA)
\end{aligned} \tag{2}$$

Table 2: Non-relativistic Hamiltonian matrix elements for the three singlet states and one arbitrary triplet state of a model biradical. The matrix elements are given in terms of integrals between delocalized orbitals. For the corresponding mathematical expressions, see Equations 1 and 2.

	$^1 a^2 + b^2\rangle$	$^1 ab\rangle$	$^1 a^2 - b^2\rangle$	$^3 ab\rangle$
$^1 a^2 + b^2\rangle$	$E_{\text{triplet}} + 2(K_{ab} + K'_{ab})$	$\gamma_{ab}$	$\delta_{ab}$	0
$^1 ab\rangle$	$\gamma_{ab}$	$E_{\text{triplet}} + 2K_{ab}$	$\gamma_{ab}^-$	0
$^1 a^2 - b^2\rangle$	$\delta_{ab}$	$\gamma_{ab}^-$	$E_{\text{triplet}} + 2K'_{ab}$	0
$^3 ab\rangle$	0	0	0	$E_{\text{triplet}}$

Using the interconversion relationships given in the set of Equations 2 and noting that the expressions

$$\begin{aligned}
 K'_{ab} \pm K_{ab} &= K'_{AB} \pm K_{AB} \\
 (K'_{ab} - K_{ab})^2 - \gamma_{ab}^- &= (K'_{AB} - K_{AB})^2 - \gamma_{AB}^- \\
 \delta_{ab}^2 + \gamma_{ab}^2 &= \delta_{AB}^2 + \gamma_{AB}^2
 \end{aligned} \tag{3}$$

hold, the interconversion of the matrix elements in Tables 1 and 2 is possible. It is important to note that  $E_0$  and  $E_{\text{triplet}}$  are invariant regarding the choice of localized or delocalized orbitals. Using the explicit matrix elements, one may analyze which contributions will lead to a singlet or triplet ground-state in a biradical-type molecule. If all couplings between the singlet-states ( $\delta, \gamma$ ) are assumed to be zero, a triplet ground-state is predicted. If these contributions are non-zero, a singlet-state might be lowered energetically by a sufficient amount to become the ground-state. In any case (localized or delocalized MOs), the  $^1|a^2 + b^2\rangle$  and  $^1|A^2 + B^2\rangle$  singlet states are the energetically highest singlet states and will thus not be the ground-state of the molecule. In the case of localized orbitals,  $K'_{AB}$  is larger than  $K_{AB}$  which leads to the  $^1|AB\rangle$  state being energetically more favorable than the  $^1|A^2 - B^2\rangle$  state. The reason for this is that the exchange integral  $K_{AB}$  depends on the spatial distance between the radical-carrying atomic centers. If these atoms are far apart, the contribution of  $K_{AB}$  is small or negligible. For delocalized orbitals, the opposite is true and the  $^1|a^2 - b^2\rangle$  state lies below the  $^1|ab\rangle$  state. This can be explained by the switching of the  $K_{ab}$  and  $K'_{AB}$  contributions, as laid out before. In summary, the energetic ordering of a biradical molecule's electronic states depends on the interaction of the spin-centers and thus their separation. The singlet-triplet gap becomes smaller with increasing distance

between the radical-carrying atoms.<sup>[76,79]</sup>

Due to their very reactive nature, biradicals are challenging targets for experimental spectroscopic investigations. One of the central issues of biradical-type molecules concerns their ground-state spin multiplicity: Whether the molecule has a triplet or an open-shell singlet ground state or is a closed-shell singlet realized with a low-lying triplet state (biradicaloid) depends on multiple structural and electronic parameters. The quantitative details of the electronic states' energetic ordering has been discussed in terms of matrix elements in the previous paragraph. Some basic rules with regards to structural parameters are also known:<sup>[73]</sup> In aromatic hydrocarbons a linkage of the radical centers in *meta*-position leads to triplet ground states, whereas an *ortho*- or *para*-linkage leads to singlet states. However, small perturbations - induced by substituents or the solvent environment - may lead to the breakdown of these rules-of-thumb. For ideal biradicals which possess two energetically degenerate FMOs, Hund's rule can be applied and a triplet ground-state should result. As the stabilization of the triplet state with respect to the singlet state is proportional to the overlap of the involved FMOs, violations of Hund's rule for biradicals are known.<sup>[73]</sup> These cases can be rationalized using the framework devised by Michl and Bonačić-Koutecký.<sup>[76,78]</sup>

Experimental techniques for investigating biradicals include PES and Electron Paramagnetic Resonance (EPR). EPR spectroscopy can be used to analyze triplet species biradicals. It is predominantly performed with the substrate embedded in a matrix at (ultra)cold temperatures (<100 K).<sup>[73]</sup> Photoelectron spectroscopy may be performed in one of its more modern variants such as Threshold Photoelectron Spectroscopy (TPES) and Threshold Photoelectron-Photoion Coincidence (TPEPICO), which offer better resolution by an improved signal-to-noise ratio.<sup>[80,81]</sup> The usage of high intensity synchrotron radiation instead of laboratory light sources, as well as velocity focusing techniques, enhances the results further.<sup>[82]</sup>

Computational investigations of the target molecules can assist the interpretation of spectroscopic measurements of biradical(oid) molecules by providing estimates of the Ionization Energy (IE). Furthermore, the vibrational progression of the spectral peak can be modeled using computational techniques. Comparison of the fit of the approximate vibrational progression of a molecule obtained computationally with the experimental data can aid the identification of the correct isomeric molecule which was transiently present in the experiment. To this end, Franck Condon (FC) simulation may be performed. As this technique was used to identify transient biradical molecules in this work, a short introduction to the theoretical foundations of the methods is now



given (following a derivation given in [83]):

Validity of the Born-Oppenheimer approximation is assumed and the wavefunction is written in a product *ansatz*. The intensity of a quantum transition is proportional to the matrix element of the transition dipole moment

$$M_{i \rightarrow j} = \int \Psi_j^* \vec{\mu} \Psi_i \, d\vec{x}. \quad (4)$$

Inserting  $\Psi(R, r) = \Omega(R) \cdot \Phi(\bar{R}, r)$  into Equation 4 with  $\vec{\mu} = \vec{\mu}_R + \vec{\mu}_r$  yields two terms

$$M_{i \rightarrow j} = \langle \Omega_j(R) | \Omega_i(R) \rangle \langle \Phi_j(r, \bar{R}) | \vec{\mu}_r | \Phi_i(r, \bar{R}) \rangle + \langle \Phi_j(r, \bar{R}) | \Phi_i(r, \bar{R}) \rangle \langle \Omega_j(R) | \vec{\mu}_R | \Omega_i(R) \rangle. \quad (5)$$

Due to the orthonormality of the electronic wavefunctions  $\Phi_i(r, \bar{R})$  and  $\Phi_j(r, \bar{R})$ , it holds that

$$\langle \Phi_j(r, \bar{R}) | \Phi_i(r, \bar{R}) \rangle = 0. \quad (6)$$

Now, the Franck-Condon approximation is invoked by setting  $\bar{R} = R_0$ , where  $R_0$  are the nuclear coordinates in the equilibrium geometry of the molecule. In

$$M_{i \rightarrow j} = \langle \Omega_j(R) | \Omega_i(R) \rangle \langle \Phi_j(r, R_0) | \vec{\mu} | \Phi_i(r, R_0) \rangle \quad (7)$$

we call the integral  $\langle \Omega_j(R) | \Omega_i(R) \rangle$  the Franck-Condon factor. Within in the assumed approximations, the intensity of a spectral transition is thus directly proportional to the square of the Franck-Condon factor. If the nuclear potential is assumed to be harmonic, the nuclear wavefunctions  $\Omega_i(R)$  can be taken to be identical to the solutions of the quantum harmonic oscillator for a given force constant. The Franck-Condon factors can then be evaluated in a straight-forward manner (for mathematical details see [84–86]). If the factors are computed for every conceivable vibronic spectral transition, the vibrational progression of an electronic transition may be simulated. Fully integrated software packages for these simulations are available.<sup>[83,87]</sup>

Some general findings for computational modeling of biradical molecules are now discussed. Previous research has shown the impact of especially geometrical considerations may be large:<sup>[80]</sup> Correct equilibrium geometries for biradicaloid-molecules and their associated ionized species might be difficult to estimate computationally. Commonly it is desirable that structural optimization procedures are performed using less-accurate quantum chemical approaches, as the geometry optimization takes significantly more computational effort than a computation of the electronic structure

of one given set of atomic positions. Nevertheless, for the case of the cation of the *ortho*-benzyne biradical molecule, it was found that different computational methodologies might hint towards different minimum structures. In this case, a  $C_{2v}$  symmetric planar minimum geometry of the cation, which was previously proposed in a scientific report,<sup>[88]</sup> did not yield predictions of the ionization spectrum or IE in agreement with experimental data. Only when performing a computationally costly full optimization procedure using the MR CASPT2 approach without symmetry restrictions, a lower-lying equilibrium geometry with  $C_2$  symmetry can be identified. Full numerical hessian calculations then showed the  $C_{2v}$  structure to be a saddle-point of the energy landscape, with only the novel  $C_2$  geometry showing no negative second derivatives and therefore proving to be an actual minimum structure. Properties computed from both structures differ significantly. For the *ortho*-benzyne cation, a proper computational FC absorption spectrum prediction which agrees with experimental data could be made only once the newly proposed equilibrium geometry is combined with the first two excited states. Furthermore, the zero-point vibrational energy (ZPVE) has to be taken into account to estimate reliable energy differences between the electronic states.<sup>[80]</sup>

In this work a number of investigations of biradical-type molecules are presented. These molecules have been analyzed in a joint computational and experimental research effort. A communication on the first organometallic bismuthinidene biradical (see [89] “Methylbismuth: an organometallic bismuthinidene biradical”) is reprinted in Chapter 3. Therein, we report the experimental characterization of the compounds BiMe, BiMe<sub>2</sub>, and BiMe<sub>3</sub>. These reactive molecules were generated from precursor compounds using pyrolysis. They were subsequently investigated with the TPEPICO spectroscopic technique, utilizing synchrotron radiation. The preparation of the mentioned substances is known to be especially challenging, since low-valent molecules tend to undergo degradation quickly.<sup>[89]</sup> In the past, bismuth compounds with the metal in the oxidation state of +1 could only be prepared by incorporating additional Lewis bases for stabilization.<sup>[90–92]</sup> The investigation of these bismuth species promises important insights for photo-physical and catalytic applications of organobismuth compounds.<sup>[89]</sup> While the before-mentioned Lewis base stabilized organobismuth compounds show singlet ground states, we find a triplet ground-state multiplicity for BiMe by combining experimental data with computational models. Although it is clear that relativistic effects are pronounced for heavy metal molecules containing bis-

muth, our computational modeling show that these effects can well be captured using DFT calculations in the Effective Core Potential (ECP) approximation. We conducted scalar-relativistic computations on the organobismuth compound as well, using an all-electron triple-zeta valence basis-set. The results of both approaches are similar and in good agreement with the experimentally measured IEs. We assign the Bismuthinidene BiMe in a triplet ground-state with  $C_{3v}$  symmetry as the lowest energy structure. This is a surprising finding, as for the structural homologues with  $E = N, P$ , the methylene species  $HE=CH_2$  was found to be energetically favored.<sup>[93–96]</sup> The singly occupied molecular orbitals (SOMOs) of BiMe can be reasonably described as  $p_x$  and  $p_y$  orbitals centered on the bismuth atom. Other contributions to the MOs are small. Ionization from one of these MOs leads to Jahn-Teller distortion and symmetry lowering, however, the energy difference between the two possible cationic states is very small and beyond the spectroscopic resolution in this experiment. By creating a Franck-Condon simulation of the vibrational progression upon ionization of BiMe into the two approximately isoenergetic cationic states, the experimental spectrum could be reproduced. Computational IE estimates are in good agreement with experimental findings. Lastly, bond-dissociation energies of the organobismuth compounds under considerations were computed to give further insights into the dissociation of the  $BiMe_3$  molecule.<sup>[89]</sup>

Our study on the biradical molecule pentadiynylidene and its methyl-substituted derivatives (see [81] “Pentadiynylidene and Its Methyl-Substituted Derivates: Threshold Photoelectron Spectroscopy of  $R_1-C_5-R_2$  Triplet Carbon Chains”) is reprinted in Chapter 4 of this work. The experimental investigation of the compounds  $HC_5H$ ,  $HC_5CH_3$ , and  $CH_3C_5CH_3$  has been carried out by the research group of Prof. Dr. Ingo Fischer. They have previously investigated many organic molecules involved in soot formation processes. Among them are the three-carbon-chained analogue molecules of pentadiynylidene (i.e.  $R_1-C_3-R_2$ ).<sup>[97,98]</sup> The focus of the investigation of pentadiynylidene is the accurate determination of IEs and the theoretical rationalization of experimentally obtained mass-selective TPES spectra. Some efforts to characterize the linear five-carbon-chained molecules of type  $R_1-C_5-R_2$  with  $R = H, CH_3$ , which are the target of the present study, have been made in the past: Most of the times, low temperature matrix-isolation techniques were used to render spectroscopy of this highly reactive and thus short-lived class of molecules feasible. As for the simpler chain-analogue  $H-C_3-H$ , molecular point-group symmetry determination is a non-trivial task for

these chemical species.<sup>[97,98]</sup> For H–C<sub>5</sub>–H, experimental EPR measurements<sup>[99,100]</sup> and *ab-initio* coupled-cluster computations<sup>[101,102]</sup> indicate that a linear structure with *D<sub>∞v</sub>* symmetry is the energetically lowest H–C<sub>5</sub>–H isomer. This isomer is furthermore predicted to have a triplet ground-state. The first excited electronic state was determined experimentally using low-temperature matrix-isolation.<sup>[103]</sup> Peyerimhoff and coworkers have determined the excitation energy theoretically at 2.76 eV using the multireference MRD-CI method with a custom Huzinaga–Dunning basis set.<sup>[104]</sup> Zhang *et al.* used a simpler Time-Dependent Density Functional Theory (TDDFT) approach (B3LYP/cc-pVTZ) to determine the excitation energy at 3.63 eV.<sup>[105]</sup> Their CASPT2(8,8) computations place the excitation at 3.02 eV.<sup>[105]</sup> The performance of the B3LYP DFT method has furthermore been tested for vibrational spectra of the ground state of H–C<sub>5</sub>–H.<sup>[106]</sup>

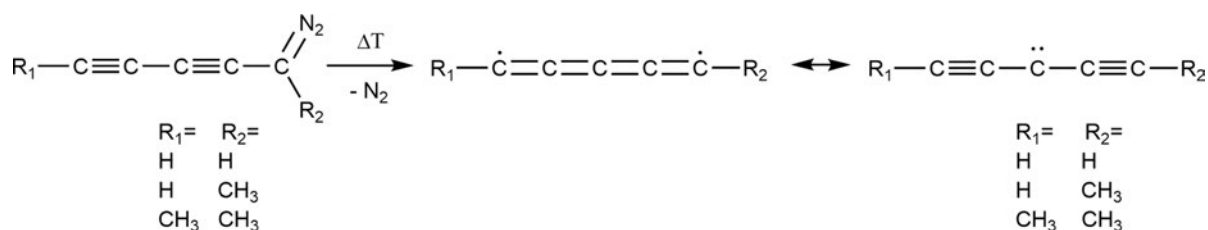


Figure 6: Reaction mechanism of the creation of pentadiynylidene (HC<sub>5</sub>H) and its derivatives H<sub>3</sub>C–C<sub>5</sub>–H and H<sub>3</sub>C–C<sub>5</sub>–CH<sub>3</sub> from precursor molecules *in situ*. Taken from [81].

In the present investigation of the reactive pentadiynylidene molecules, they were generated *via* flash pyrolysis of diazo-precursor-compounds R<sub>1</sub>–C<sub>5</sub>N<sub>2</sub>–R<sub>2</sub> (see Figure 6). As the diazo-precursors themselves are not stable at ambient conditions, their respective sodium-salts were used. Thus, the diazo-compounds were generated *in situ* as well. The synthesis was carried out according to [107, 108]. Although pyrolysis generates both the precursor molecules H–C<sub>5</sub>N<sub>2</sub>–H, H–C<sub>5</sub>N<sub>2</sub>–CH<sub>3</sub>, H<sub>3</sub>C–C<sub>5</sub>N<sub>2</sub>–CH<sub>3</sub> and the target molecules H–C<sub>5</sub>–H, H–C<sub>5</sub>–CH<sub>3</sub>, and H<sub>3</sub>C–C<sub>5</sub>–CH<sub>3</sub>, their spectra can be analyzed separately due to the mass-selective spectroscopy approach. Vibrational bands are visible in the experimental TPES spectra of the diazo precursor molecules. Franck-Condon spectra of the vibrational progression based on DFT calculations of these molecules were carried out using the FCfit 2.8.2 software.<sup>[109]</sup> The vibrational frequencies and associated normal modes were determined at the B3LYP/6-311G(2d,d,p) level of theory. Excitation energies were calculated using the parametric composite CBS-QB3 approach.<sup>[110]</sup> Experimental IEs have not been reported in previ-

ous publications, and have thus been formerly unknown prior to our study. For all three precursor molecules  $\text{HC}_5\text{N}_2\text{-H,H-C}_5\text{N}_2\text{-CH}_3$ , and  $\text{H}_3\text{C-C}_5\text{N}_2\text{-CH}_3$  that are generated *in situ*, the experimental spectra match the theoretical predictions qualitatively and quantitatively. For every substance, the mass-selective spectra show only one absorption peak with its associated vibrational progression. This indicates that the precursor molecules are successfully created in the experimental setup. The structure of the vibrational bands is well reproduced in the Franck-Condon simulations. For the pentadiynylidene molecule itself, the experimental spectra show two significant absorption peaks, one at 8.36 eV, and one at 8.60 eV. Both peaks show clear vibrational progression bands. It is not immediately clear whether the peak at 8.60 eV presents a unique electronic transition or belongs to the vibrational progression of the electronic absorption centered at 8.36 eV. It is furthermore possible that the experimental pyrolysis condition enables the creation of more than one structural isomer of the target compound  $\text{HC}_5\text{H}$ . In order to establish the identity of the peaks in the experimental spectrum, quantum-chemical computations were performed. Relative energies of the  $\text{HC}_5\text{H}$  isomers shown in Figure 7 in both their lowest singlet and triplet states have been computed. Furthermore, computational estimates for the IEs of these isomers were obtained. Both adiabatic and vertical excitations were considered, and all geometric minima of all isomers in their ground- and excited states were evaluated. A wide range of quantum-chemical methodologies was used: The well-known hybrid DFT functionals  $\omega\text{B97xD}$ ,<sup>[111]</sup>  $\text{M05-2X(D3)}$ ,<sup>[112,113]</sup> and  $\text{M06-2X}$ <sup>[114]</sup> were tested. The  $\text{CBD-QB3}$ <sup>[110]</sup> composite method was employed to obtain computational estimates for the adiabatic ionization energies. The *ab initio* methods Coupled-Cluster Singles Doubles (CCSD) and Coupled-Cluster Singles Doubles with approximate Triples (CCSD(T)) as well as the MR CASPT2 (see [115]) method were additionally included in the present work.

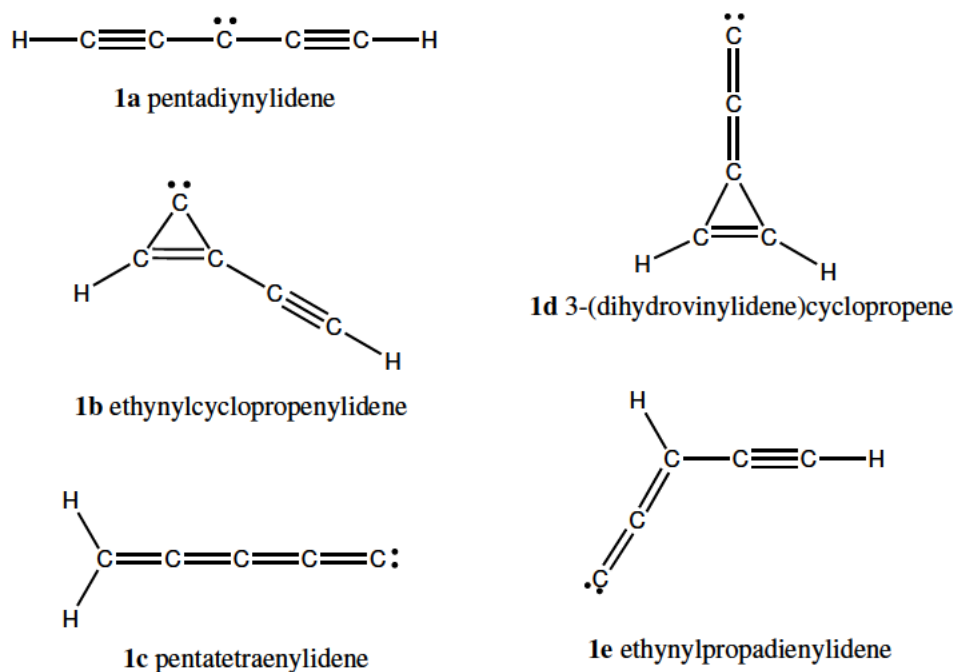


Figure 7: Structural isomers of pentadiynylidene ( $\text{HC}_5\text{H}$ ) considered in the study presented in [81] and Chapter 4 of this work.

These quantum-chemical calculations are not able to unambiguously identify the energetically lowest isomer from the five different structures considered (see Figure 7).  $\omega\text{B97xD}$ ,  $\text{uM05-2X(D3)}$  and  $\text{CASPT2(14,14)}$  predict the lowest triplet state of isomer **1a** to be the thermodynamically most stable configuration.  $\text{uCCSD}$ ,  $\text{uCCSD(T)}$  and  $\text{uM06-2X}$  instead compute the singlet ground state of isomer **1b** as the lowest energy minimum. For the first peak at 8.36 eV, isomer **1a** in its lowest triplet state shows the best agreement of the computational estimates with the experimental value. A Franck-Condon simulation of the vibrational progression shows satisfactory agreement with the peak at 8.36 eV. However, the absorption is unable to account for the second peak at higher energy (8.60 eV). Other combustion side-products with a mass-charge ratio of  $m/z = 62$  might contribute to the spectrum. A literature review revealed several  $\text{C}_4\text{N}$  isomers to be known chemical compounds which might be formed. These species have been investigated before: Ding *et al.* have evaluated various ground-state properties and isomerization pathways of  $\text{C}_4\text{N}$  theoretically in the context of astrochemistry.<sup>[116]</sup> Using computational modeling of  $\text{C}_4\text{N}$  isomers, we find that they do not contribute to the measured mass-selective TPES spectra. Therefore isomers of pentadiynylidene are the most likely cause of the second peak at 8.60 eV in the experimental data. Analysis of the IEs revealed that the computational estimates fluctuate by a large error margin

for isomer **1c** across the different quantum-chemical methods. For instance, the value predicted for IE(S<sub>1</sub>) at the uCCSD level of theory is 9.82 eV, while the CASPT2(14,14) estimate is 8.55 eV. This disagreement, which is not observable for the remaining four other isomers of H–C<sub>5</sub>–H, has its origins in small variations in the predicted equilibrium geometry of the cation **1c**<sup>+</sup>. After careful evaluation of the goodness of the results, we conclusively assign the second peak in the experimental mass-selected TPES spectrum to the ionization of isomer **1c**. However, the reasons and conditions for the formation of isomer **1c** from the precursor molecule are unclear, especially since its ground-state is 0.78 eV higher in energy than the triplet ground-states of isomers **1a** and **1b**. B3LYP DFT calculations of Cong-Jie *et al.* showed isomerization barriers between C<sub>5</sub>H<sub>2</sub> isomers in the gas-phase to be significant (> 2.3 eV).<sup>[117]</sup> Isomerization of **1** during the experimental conditions can for this reason be ruled out. Thus, the formation of the three-ring of compound **1c** would have to occur already in the precursor-molecule, before dissociation of molecular nitrogen. A second precursor isomer at *m/z* = 90 could indeed be detected experimentally when pyrolysis was turned on. The low signal-to-noise ratio made a definite identification of this isomer of the precursor molecule impossible.<sup>[81]</sup> However, the presence of another isomer by itself supports the postulated creation mechanism of **1c** by isomerization of the precursor molecule.

Another biradical molecule investigated in the course of this work is diphenylacetylene, which is also known as “tolane”. The results of this project are presented in Chapter 5 of this work, which is a reprint of a scientific publication (see [118] “A time-resolved photoelectron imaging study on isolated tolane: observation of the biradicalic <sup>1</sup>A<sub>u</sub> state”). The Lewis structures of interest are presented in Figure 8. This molecule has been the focus of a number of spectroscopic investigations in the past: Consideration of a combination of previous experimental and theoretical works leads to the following assumed scheme for the description of photoexcitation processes in diphenylacetylene following initial photon absorption:<sup>[119,120]</sup>



The ground-state of diphenylacetylene was previously characterized by means of X-Ray diffraction.<sup>[121]</sup> It shows D<sub>2h</sub> symmetry. This result has been supported by further spectroscopic investigations.<sup>[122,123]</sup> The A<sub>u</sub> state, which is a dark excited state and not

accessible *via* direct one-photon absorption from the ground state, is assumed to have a different equilibrium geometry than the electronic ground state. Its assumed  $C_{2h}$  symmetric structure shows a trans-bent moiety on the acetylene-unit (see Figure 8).

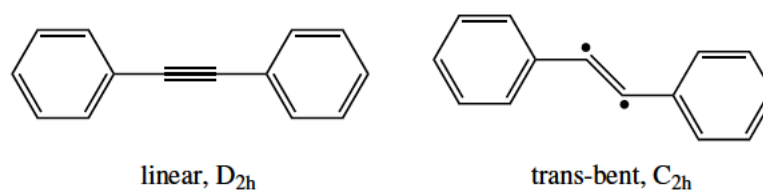


Figure 8: Conformations of interest for diphenylacetylene ( $\text{Ph}-\text{C}_2-\text{Ph}$ ) which have been investigated in the present work. The linear geometry is the equilibrium structure of the triplet ground state, while the trans-bent conformation is adopted by an electronic excited state. This scheme is adapted from [118].

The diphenylacetylene molecule was re-investigated by members of the research group of Prof. Dr. Ingo Fischer. The motivation for this project was that time-resolved spectroscopic data on isolated, cold diphenylacetylene molecules in the gas phase was not previously available. The special focus of this work was the characterization of the likely trans-bent  $A_u$  state. Experimentally, the resonance-enhanced multiphoton ionization technique was used. In order to correctly interpret and assign the measured spectra, existing computational studies<sup>[124–126]</sup> were considered and novel calculations - especially for the ionized molecule - were performed. The combined evidence from the computational and experimental data suggests deactivation of the initial  $B_{1u}$  excited state to the  $A_u$  state. Computational modeling confirms that this electronic state possesses a trans-bent equilibrium structure with  $C_{2h}$  symmetry, as given in Figure 8. This is in agreement with previous computational findings by other authors.<sup>[124–126]</sup> The ionization of the  $A_u$  state is found to proceed *via* an intermediate Rydberg state to a long-lived excited state, which is likely a triplet state reachable through intersystem crossing.<sup>[118]</sup>

The diphenylpropynylidene biradical-type molecule ( $\text{Ph}_2\text{C}_3$ ) was investigated by this author in a further joint experimental-computational research project, whose results are reprinted in Chapter 6 from [127] (“Femtosecond dynamics of diphenylpropynylidene in ethanol and dichloromethane”). Experimentally, the molecule is created using ultraviolet irradiation of the precursor 1,3-diphenyldiazopropyne. DePinto *et al.* and Noro *et al.* have already highlighted the interesting geometrical and electronic aspects of the molecule in 1993, presenting EPR measurements and discussing their



implications.<sup>[128,129]</sup> DePinto *et al.* have then subsequently performed a detailed study using infrared spectroscopy and EPR measurements in 2007.<sup>[130]</sup> To this end, the substance was embedded in an argon matrix at low temperatures. The authors determined the molecule to first spawn in a bent geometry, which then relaxes to a linear conformation with a triplet ground state.<sup>[130]</sup> The creation of diphenylpropynylidene from a precursor molecule and possible conformations adopted by the target molecule are summarized in Figure 9. In the present investigation by this author, a different setting is chosen: The molecule was analyzed in polar- and non-polar solvents using transient absorption in the ultraviolet/visible (UV/Vis) region of the electromagnetic spectrum. The measurements have been performed at an initial temperature of 243 K.

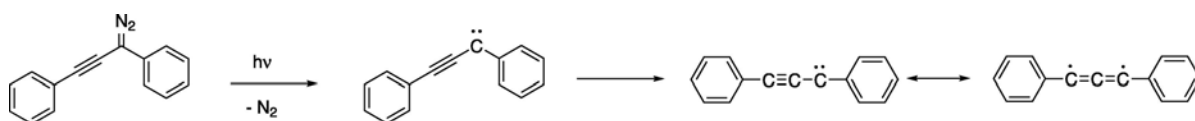


Figure 9: Reaction mechanism of the creation of diphenylpropynylidene (Ph-C<sub>3</sub>-Ph) from precursor molecules. Taken from [127].

The various conformations and states of the target molecule and its precursor were characterized using computational approaches in order to rationalize the experimentally obtained 2D absorption spectra. These results are complimented by novel EPR measurements of the molecule embedded in a solvent matrix at different temperatures, where identical solvents to the transient absorption experiments were used. In this way, the elucidation of the time-dependent relaxation of the excited states, their spin multiplicity and their solvent dependence was possible. The rhombic zero-field splitting parameter was calculated from the EPR spectra. This parameter can be related to the geometry of the molecule and was useful in determining the geometry of diphenylpropynylidene. It was found that at low temperatures (4 K) diphenylpropynylidene is created in a bent geometry from the precursor molecule. Annealing to higher temperatures (70 K) led to the molecule relaxing into a linear geometry which was retained even after subsequent cooling. For the transient absorption spectra, a striking dissimilarity of the time-dependent absorption depending on the solvent was found. For ethanol and dichloromethane with oxygen present, qualitatively similar spectra are obtained. We attribute all absorption bands not to the target molecule diphenylpropynylidene, but to secondary reaction products of diphenylpropynylidene with molecular oxygen or ethanol solvent molecules. The quenching of the nascent biradical diphenylpropynylidene thus happens on an ultrafast timescale and is not observable

in the resolution of the experimental spectroscopic setup. For diphenylpropynylidene in  $\text{Cl}_2\text{C}(\text{CH}_3)_2$  as a solvent in an oxygen-free environment, a qualitatively different absorption spectrum is measured. Comparison with computational estimates for the geometries and excited states for various conformations of diphenylpropynylidene is able to provide an explanation for the measured absorption bands: The diphenylpropynylidene molecule is created in an excited singlet state, namely the first open-shell singlet excited state. This was determined by correlating the MOs of the diazo precursor compound with those of the resulting target molecule  $\text{Ph}_2\text{C}_3$ . A detailed analysis of both the oscillator strengths as well as vertical and adiabatic absorption energies from this excited singlet state to all reachable higher excited states of diphenylpropynylidene showed that the single prominent absorption band in oxygen-free  $\text{Cl}_2\text{C}(\text{CH}_3)_2$  is due to an excitation into the third excited singlet-state. The molecule is subsequently deactivated into the  $T_0$  ground-state, which has a linear geometry. Through detailed analysis incorporating insights from computational modeling, a concise picture of the time-dependent relaxation of diphenylpropynylidene after excitation emerges.

#### 1.4. Quantum-Chemical Investigation of Reaction Mechanisms

The detailed rationalization of the mechanisms of chemical reactions is an important and active area of computational organic chemistry. Calculations may furthermore be used to explain reaction results, predict how a chemical system will behave, and thereby influence and inform the design of synthetic routes.<sup>[131]</sup> Experimental investigation of chemical mechanisms may require sophisticated experimental procedures: The reaction barrier may be determined using Arrhenius' equation

$$k = A e^{\frac{-E_a}{RT}} \quad (9)$$

by measuring the impact of temperature changes.<sup>[132]</sup> Here,  $k$  is the rate constant,  $E_a$  the activation energy (which is related to the energetic height of the transition state),  $T$  is the temperature, and  $R$  is the gas constant. If  $k$  at different temperatures  $T$  is known, Equation 9 may be used to estimate  $E_a$ .<sup>[133]</sup> The identification of transition state structures or intermediates is often harder, as it may require spectroscopic observation of the intermediate species. Computational chemistry approaches can model intermediates as well as transition states directly. If a suitable level-of-theory is chosen, comparison of the species' energetics can lead to the proposal of a suitable minimum-energy reaction path.<sup>[134]</sup> Computational reaction mechanism investiga-

tions commonly begin with determining the relative energies of the reactants, their conceivable transition state, and energetics of the products.<sup>[135]</sup> Houk and coworkers have been especially productive in investigation reaction mechanisms using computational chemistry techniques.<sup>[136–138]</sup>

One important area of research in bioorganic chemistry are bioconjugation reactions. These chemical reactions enable the selective addition of new chemical moieties to certain amino-acids or related biological compounds. The chemical subgroups bound to the biomolecules may introduce a range of desirable properties or changes, such as fluorescence, which can be exploited for spectroscopic investigations.<sup>[139]</sup> Furthermore, the modification of biological macromolecules may lead to an altered function, such as increased or decreased activity, or even altered substrate selectivity. Commonly, the biomolecule is augmented with a carefully engineered reagent which has been produced by means of synthetic organic chemistry. The bioconjugation reaction, coupling the synthetic reagent to the biomolecule, needs to be performed at acceptably mild reaction conditions (temperature in a physiological range, preferably neutral pH, etc.) in order to prevent degradation of the biomolecule.<sup>[140]</sup>

In order to build a set of tailored bioconjugates which serve as tools for modifying biomolecules, it is desirable to develop and identify bioconjugation reagents which will selectively bind to certain biomolecules, such as individual amino acids. Additionally, it is convenient if the synthetic molecule possesses a functional group for easy addition of further chemical functionalities as payloads. The synthetic reagent will then act as a warhead scaffold for selective addition to the target biomolecule, and other functional groups may be coupled to the generic warhead. These may then introduce further desired properties.<sup>[141]</sup>

While for amino acid residues such as lysine and cysteine many reliable bioconjugation protocols have been established,<sup>[142]</sup> the bioconjugation possibilities of tyrosine residues are still emergent and not fully explored. According to Szijj *et al.*, tyrosine bioconjugation offers a range of possibly favorable properties in comparison with other amino acids: In enzymes, tyrosine is rarely exposed to the solvent-accessible surface, which makes it a suitable target for site-selective modification. Lysine shows contrary effects, as it is more abundant and thus a less suitable target for site-selective modifications. Cysteine has a clear drawback stemming from its high reactivity.<sup>[140]</sup> One of the known tyrosine bioconjugation strategies involves the use of the versatile reactant group of diazodicarboxyamides and TADs.<sup>[143]</sup> This synthetic route towards tyrosine bioconjugates has been primarily pursued by the late Barbas III. and

coworkers.<sup>[144,145]</sup> Since then, their work has been utilized to enable the synthesis of DNA-protein conjugates using deoxyribonucleic acid (DNA) as a payload at the TAD reactant.<sup>[146]</sup> Furthermore, TAD reactants have been used in bioorthogonal bioconjugation reactions.<sup>[147]</sup> Seibel and coworkers have utilized a modified variant of the TAD tyrosine bioconjugation to modify levansucrase enzymes. Using this approach in combination with targeted site-directed mutagenesis, they are able to tailor the product spectrum of the enzyme.<sup>[148–150]</sup>

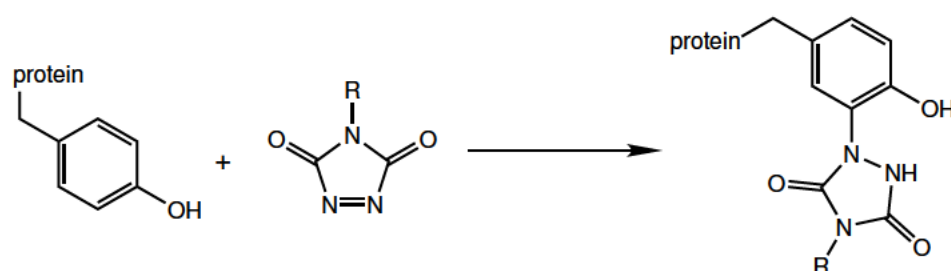


Figure 10: Reaction mechanism of the bioconjugation of the TAD reagent with a tyrosine amino acid residue of a protein. Taken from [151].

As laid out, a significant number of experimental advancements has already been achieved using the versatile TAD bioconjugation reaction. Nevertheless, the reaction mechanism of the bioconjugation has remained unclear. The schematic of the reaction of the TAD reagent with a tyrosine amino acid residue of a protein is showcased in Figure 10. Elucidation of the potential reaction pathways and their associated energetic profiles might help chemists understand the site-specificity of the reaction. The results can be used as a rationale for improving the usability of the TAD bioconjugation. For this reason, we have conducted a detailed quantum chemical study of many conceivable reaction pathways, including pericyclic as well as stepwise mechanisms. High-level MR computations were performed in order to assess and benchmark the usability of less computationally demanding DFT based approaches. The solvent effects were included using implicit solvation techniques as well as by considering the impact of an explicit solvent molecule as a mediator. By combining the results of the computational investigation with experimental findings regarding the impact of the buffer and its pH value, a base-mediated reaction mechanism is proposed. The results are presented in the included report in Section 7, as reprinted from [151].

## 1.5. Computational Modeling of Absolute Entropies for Single-Molecule Systems

The accurate description of the energetics of chemical reactions is one of the most important goals of chemical physics and theoretical chemistry. From thermodynamical arguments, a reaction can be said to be in equilibrium when its associated reaction Gibbs energy is zero:<sup>[132]</sup>

$$\Delta_r G = \left( \frac{\partial G}{\partial \xi} \right)_{p,T} = 0 \quad (10)$$

This implies that the chemical potential associated with the reactants and products is identical. The variable  $\xi$  is the extent of reaction, which is given in mol. For this reason, the Gibbs energy is a fundamental quantity for the description of chemical phenomena. Chemical reactions occur spontaneously if the change in Gibbs energy associated with the reaction is negative at constant temperature and pressure.<sup>[132]</sup> We note that in thermodynamics the impact of reaction timescales is fully neglected. Therefore, by implicitly neglecting kinetic effects, the influence of a possible reaction barrier does not enter these considerations.<sup>[132]</sup> The difference in Gibbs energy is furthermore identical to the maximum amount of non-expansion work which can be extracted from a system. If reactions at constant temperature and volume, i.e. in a canonical thermodynamic ensemble, are considered, the Helmholtz free energy  $A$  plays the same role as the Gibbs energy.<sup>[132]</sup>

The Gibbs energy  $G$  and the Helmholtz energy  $A$  are state functions of a given system and can be written as

$$\begin{aligned} G &= H - TS \\ A &= U - TS, \end{aligned} \quad (11)$$

where  $U$  is the internal energy of a system,  $T$  is the temperature of the system, and  $S$  is the entropy of the system. Thus, in order to state correct predictions about chemical systems in line with thermodynamic considerations, the internal energy of the system is not sufficient. The entropy  $S$  of the chemical system also has to be taken into account in order to evaluate the Gibbs energy.<sup>[132]</sup>

Quantum chemical modeling primarily strives to obtain reliable estimates for the internal energy  $U$  of a molecular system. To this end, sophisticated approximations for the calculation of the electronic structure of the molecule might be invoked. Although much progress in the development of computationally fast and numerically reliable

methodologies for the evaluation of the internal energy have been made, the numerical evaluation of molecular entropies is still an emerging field.<sup>[1]</sup>

The entropies of rigid molecules which only possess one relevant accessible conformation (such as benzene or molecular chlorine etc.) are comparably easy to estimate. If the hessian-matrix of second derivatives with respect to the nuclei of the molecule can be obtained, the harmonic oscillator approximation may be invoked and force constants can be computed. Once these force constants are known, the evaluation of the quantum-mechanical vibrational entropies of each normal mode is straight forward, as analytic mathematical formula are available.<sup>[13]</sup> Common quantum chemical software packages provide this functionality.<sup>[152]</sup>

If one considers larger macromolecules, it will quickly become evident that the assumption of a single harmonic conformational well does not hold, as torsional DOFs lead to further accessible minima. Prime examples of molecular species which may not be accurately treated in the harmonic approximation are polypeptides and polymers, but organic molecules with large side chains may also be troublesome.<sup>[153]</sup> For these systems, novel entropy estimation algorithms need to be devised so that computational researchers may evaluate entropic contributions reliably. Some algorithms working on different assumptions and methodologies have been presented, but to this day no single universally applicable method for the computational estimation of absolute entropy values has emerged.<sup>[1]</sup> Another point of concern for the computational estimation of entropies is Nernst's theorem, known as the third law of thermodynamics. It states that as zero temperature is approached, the entropy of any system will also approach zero. This law cannot be adhered to in the case of classical statistical mechanics. Instead, the absolute entropy diverges towards  $-\infty$  as the temperature approaches zero Kelvin.<sup>[154]</sup>

If - contrary to an absolute entropy - only relative entropies along a pre-designated reaction coordinate are desired, reliable families of methods are available. One of these methods is the Umbrella Sampling approach.<sup>[155]</sup> Therein, the free energy of the reaction along a chosen reaction coordinate is computed by estimating the partition function along this coordinate computationally. The Helmholtz free energy of the reaction along this reaction coordinate  $\xi$  is given as

$$A(\xi) = -\frac{1}{\beta} \ln Q(\xi) \quad (12)$$

with

$$Q(\xi) = \frac{\int \delta[\xi(r) - \xi] e^{(-\beta E)} d^N r}{\int e^{(-\beta E)} d^N r}. \quad (13)$$

In this context,  $A(\xi)$  is often called the Potential of Mean Force (PMF).<sup>[156]</sup> If  $\xi$  is a generalized coordinate rather than a simple Cartesian spacial coordinate, further Jacobian-terms may enter the free energy expression. For an ergodic system, the ensemble average of all conformations is equal to the time-average for infinite sampling timescales.<sup>[157]</sup> This implies, that the partition function  $Q(\xi)$  can in principle be estimated from a simple Molecular Dynamics (MD) trajectory. Nevertheless, this is rarely done as in practice MD simulations have a finite runtime and sampling of higher-energy conformations or transition-states of interested is insufficient.<sup>[158]</sup> Thus, one usually uses non-equilibrium sampling techniques (for example by introducing suitable bias potentials) in order to enhance the sampling of these rare events, which are crucial for obtaining reliable  $Q(\xi)$  estimates.<sup>[159,160]</sup> Umbrella Sampling proceeds by splitting the chosen path along the reaction coordinates into discrete regions (“windows”) and running non-equilibrium MD simulation for each of them: Bias-potentials are added to constrain the motion of the molecule to the chosen subsection of the reaction coordinate. In a second step, an analysis method, such as the Weighted Histogram Analysis Method (WHAM)<sup>[161,162]</sup> or Umbrella Integration<sup>[163]</sup>, needs to be invoked on the simulation data in order to extract the PMF from the biased MD simulation trajectory. It is important to note that the energies of the conformations sampled during the MD simulation runs at no point enter the calculation of the PMF. It is only the relative occurrence probabilities of the conformations from which an approximation to the partition function  $Q(\xi)$  is computed.<sup>[155]</sup> Considering practical numerical applications of the Umbrella Sampling technique, You *et al.* highlight that Umbrella Sampling is able to yield adequate PMF predictions if the biased MD simulations are reasonably converged. Nevertheless, some rare-events in the simulation data, such as abrupt movements during ligand unbinding in the case of You et al, will lead to erroneous PMFs.<sup>[164]</sup>

A conceptually different method to compute free energy differences is Thermodynamic Integration (TI).<sup>[165,166]</sup> In TI, one strives to calculate the change in free energy along a given path as well. However, this path does not need to be a spatial coordinate. Instead, it may be an arbitrary, nonphysical pathway which does not need to be realizable in experimental conditions. To this end, a Hamiltonian  $H_\lambda$  for the molecular system at hand is constructed in such a way that it explicitly depends on a coupling parameter

$\lambda$  where  $1 \geq \lambda \geq 0$ . Thus, states characterized by different values of  $\lambda$  differ in their molecular interaction potentials.<sup>[167]</sup> An analytic expression for the Helmholtz free energy difference between the thermodynamic states  $H_{\lambda=0}$  and  $H_{\lambda=1}$  can be derived as

$$\Delta A = \int_0^1 \left\langle \frac{dU(\lambda, \vec{q})}{d\lambda} \right\rangle_{\lambda} d\lambda. \quad (14)$$

Numerical integration schemes for solving this expression have been devised. The choice in  $\lambda$  is arbitrary in so far as  $H_{\lambda=0}$  might correspond to a system only subject to Lennard-Jones type interaction and  $H_{\lambda=1}$  may be a system described by proper molecular forces. Alternatively, even atoms or functional groups can be inter-morphed in an alchemical sense, where the path from  $\lambda = 0$  to  $\lambda = 1$  might correspond to the substitution of a hydrogen atom for a methyl-group. In the latter case, the method is usually termed Free Energy Perturbation (FEP).<sup>[165,168]</sup> It can be shown that Umbrella Sampling becomes equivalent to thermodynamic integration in the case of infinitely strong bias-functions.<sup>[169,170]</sup>

In the course of this work, we present benchmarks and studies of a certain family of entropy estimation protocols for molecular systems, termed kNN entropy estimators. These algorithms are used in related scientific fields such as information sciences and computer vision as well, as they present many favorable properties.<sup>[171,172]</sup> In principle, they are able to estimate the entropy of arbitrary probability distributions and thus even of flexible molecules with multiple non-harmonic conformational minima. The method operates by processing data obtained from numerical MD simulations, which is an already well-established technique in molecular modeling.<sup>[13,173]</sup>

Nevertheless, the kNN entropy estimators have not yet found widespread application, although they are being used in focused, specialized studies on molecular properties. We were able to identify several potential pitfalls embedded in the theoretical framework of the kNN estimators, which are of high relevance especially for application towards molecular simulation data. These include non-additivity of the estimator for uncorrelated probability distributions, asymptotic bias of the estimator, and its failure for multidimensional data with uneven variances. The impact of each shortcoming is evaluated and correction schemes are proposed where possible. In Section 8, we aim at presenting a thorough presentation of the mathematical framework of the kNN entropy estimators. Implicit and explicit approximations contained in the estimator are highlighted in this derivation, which starts from the formally correct entropy expression for a non-relativistic quantum-mechanical system in a canoni-



cal ensemble. The section concludes with a benchmark of kNN entropy estimators with a low-dimensional benchmark derived from chemical MD simulation data and a benchmark using high-dimensional multivariate gaussian probability densities.

## 2. Geometry Mediated Localization of Excitons

Reprinted from *The Journal of Chemical Physics*, 2020, 153, 224104ff, with the permission of AIP Publishing.

The Journal of  
Chemical Physics

### Geometry relaxation-mediated localization and delocalization of excitons in organic semiconductors: A quantum chemical study

---

Cite as: *J. Chem. Phys.* **153**, 224104 (2020); <https://doi.org/10.1063/5.0028943>  
 Submitted: 09 September 2020 . Accepted: 13 November 2020 . Published Online: 08 December 2020

---

M. Deutsch, S. Wirsing, D. Kaiser, R. F. Fink, P. Tegeder, and B. Engels

---

**COLLECTIONS**

Paper published as part of the special topic on [Excitons: Energetics and Spatio-temporal Dynamics](#)

View Online

Export Citation

CrossMark

---

**ARTICLES YOU MAY BE INTERESTED IN**

[Vibronic and excitonic dynamics in perylene diimide dimers and tetramer](#)  
*The Journal of Chemical Physics* **153**, 224101 (2020); <https://doi.org/10.1063/5.0024530>

[On the quantum origin of few response properties](#)  
*The Journal of Chemical Physics* **153**, 221101 (2020); <https://doi.org/10.1063/5.0027545>

[Photoexcitation dynamics in perylene diimide dimers](#)  
*The Journal of Chemical Physics* **153**, 244117 (2020); <https://doi.org/10.1063/5.0031485>

---

Challenge us.

What are your needs for periodic signal detection?

---

J. Chem. Phys. **153**, 224104 (2020); <https://doi.org/10.1063/5.0028943>

© 2020 Author(s).

**153**, 224104

AIP  
Publishing

# Geometry relaxation-mediated localization and delocalization of excitons in organic semiconductors: A quantum chemical study

Cite as: J. Chem. Phys. 153, 224104 (2020); doi: 10.1063/5.0028943

Submitted: 9 September 2020 • Accepted: 13 November 2020 •

Published Online: 8 December 2020



M. Deutsch,<sup>1</sup> S. Wirsing,<sup>1</sup> D. Kaiser,<sup>1</sup> R. F. Fink,<sup>2</sup> P. Tegeder,<sup>3</sup> and B. Engels<sup>1,a)</sup>

## AFFILIATIONS

<sup>1</sup>Institut für Physikalische und Theoretische Chemie, Universität Würzburg, Emil-Fischer-Str. 42, D-97074 Würzburg, Germany

<sup>2</sup>Institut für Physikalische und Theoretische Chemie, Universität Tübingen, Auf der Morgenstelle 18, 72076 Tübingen, Germany

<sup>3</sup>Physikalisch-Chemisches Institut, Universität Heidelberg, Im Neuenheimer Feld 253, 69120 Heidelberg, Germany

**Note:** This paper is part of the JCP Special Topic on Excitons: Energetics and Spatio-temporal Dynamics.

**a)** Author to whom correspondence should be addressed: [bernd.engels@uni-wuerzburg.de](mailto:bernd.engels@uni-wuerzburg.de)

## ABSTRACT

Photo-induced relaxation processes leading to excimer formations or other traps are in the focus of many investigations of optoelectronic materials because they severely affect the efficiencies of corresponding devices. Such relaxation effects comprise inter-monomer distortions in which the orientations of the monomer change with respect to each other, whereas intra-monomer distortions are variations in the geometry of single monomers. Such distortions are generally neglected in *quantum chemical* investigations of organic dye aggregates due to the accompanied high computational costs. In the present study, we investigate their relevance using perylene-bisimide dimers and diindenoperylene tetramers as model systems. Our calculations underline the importance of intra-monomer distortions on the shape of the potential energy surfaces as a function of the coupling between the monomers. The latter is shown to depend strongly on the electronic state under consideration. In particular, it differs between the first and second excited state of the aggregate. Additionally, the magnitude of the geometrical relaxation decreases if the exciton is delocalized over an increasing number of monomers. For the interpretation of the vibronic coupling model, pseudo-Jahn-Teller or Marcus theory can be employed. In the first part of this paper, we establish the accuracy of density functional theory-based approaches for the prediction of vibrationally resolved absorption spectra of organic semiconductors. These investigations underline the accuracy of those approaches although shortcomings become obvious as well. These calculations also indicate the strength of intra-monomer relaxation effects.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0028943>

## INTRODUCTION

Functionalized polycyclic aromatic molecules are in the focus of a multitude of experimental and theoretical investigations because they are promising materials in the field of organic semiconductors.<sup>1–5</sup> In this context, perylene-based dyes have been frequently investigated as they exhibit favorable properties.<sup>1</sup> In particular, 3,4,9,10-perylene tetracarboxylic acid bisimides (PBIs) have been explored as a replacement for fullerenes in organic photovoltaics as they possess higher electron mobilities<sup>6,7</sup> and high extinction coefficients in the visible region.<sup>8</sup> Furthermore, they are relatively inexpensive and remarkably stable toward light as well as air.<sup>1</sup>

Eventually, their electronic properties and the packing geometries can be tuned by varying the substitution pattern.<sup>9,10</sup> However, despite all these advantages, PBI containing photovoltaic devices show reduced efficiencies in comparison to fullerene analogs.<sup>11</sup> Reasons for these shortcomings were attributed to morphological issues<sup>12</sup> and to trapping processes, which arise due to the formation of excimer states.<sup>13,14</sup> For PBI thin films, a fast relaxation of excitons leading to long-lived immobile states was identified.<sup>14</sup> For  $\alpha$ -perylene, the formation of excimers induced by motions of two monomers with respect to each other was indicated by Raman spectroscopy.<sup>15</sup> Recent investigations on excimer formation by Hoche and co-workers,<sup>16</sup> as well as Kennehan and co-workers,<sup>17</sup> also

indicated a similar mechanism. For amorphous systems, additional traps are due to the strongly varying excitation energies due to the relative orientation of the molecules with respect to each other.<sup>18,19</sup> Actually, it was shown that the energetic disorder is mostly caused by these orientation effects and only to a smaller degree by environmental effects. In this investigation, the energetical disorder predicted by using the Bässler model<sup>20,21</sup> could only be reproduced if a delocalization of excimers is assumed.<sup>22</sup>

A detailed understanding of possible relaxation pathways is very important for the rational design of materials with improved functionality.<sup>1,23</sup> However, while the mentioned studies agree with the fact that exciton trapping occurs rather frequently, the details of the underlying processes are still under debate. Within H-aggregates, trapping can occur because the exciton is initially excited to the higher lying Frenkel state (in a dimer system, this is the  $S_2$  state). If the relaxation to the lowest Frenkel state ( $S_1$ ) is faster than the competing hopping process, the exciton gets trapped because hopping to the higher lying  $S_2$  state is no longer possible due to the energy loss accompanied with the relaxation to the  $S_1$  state. Furthermore, hopping between the lower Frenkel states—which is energetically feasible—is very inefficient due to the vanishing transition dipoles of these states.

Non-negligible contributions of charge transfer (CT) states result in an even more complex situation. Using the dimer  $M-M$  ( $M \equiv$  Monomer) as the most simple aggregate, Frenkel configurations arise from linear combinations of the locally excited  $M^*M$  and  $MM^*$  determinants, while the CT-type configurations stem from linear combinations of the electron exchanging determinants  $M^+M^-$  and  $M^-M^+$ . Due to significant couplings between locally excited- and CT-configurations, the adiabatic states determined by common quantum-chemistry excited state calculations are generally strong mixtures of these configurations. The ratio of Frenkel and CT characters depends on a subtle interplay of the energy difference and the electronic coupling between the diabatic states.<sup>21,24–26</sup> Hence, a rugged potential energy surface (PES) with various conical intersections results, which may provide very efficient decay pathways to lower lying states.<sup>27–30</sup> Going to larger aggregates, the number of CT states increases considerably (trimer: 3 Frenkel, 6 CT; tetramer: 4 Frenkel, 12 CT), resulting in a decreased energy spacing between the states. Consequently, transitions into lower lying states become even more efficient. Note that due to the couplings between states, CT states can influence the position of conical intersections even if they are higher in energy than the Frenkel states.

Photo-induced relaxation processes in PBI aggregates represent one example for the complicated interplay of various electronic states and geometrical relaxations and distortions.<sup>29–32</sup> Model calculations predict that excimer formation is the reason for the measured strong red shift in the emission spectra of PBI dimers in the minimum structure of the ground state. This red shift is obtained by twisting one of the two monomers of an eclipsed arrangement by  $30^\circ$  about the symmetry axis parallel to the molecular plane [ $R_L = R_T = 0 \text{ \AA}$ ,  $R_z = 3.4 \text{ \AA}$ , and  $\varphi = 30^\circ$  according to the dimer structure parameters in Fig. 3(a)].<sup>33</sup> According to these model calculations, the excimer formation starts with an exciton relaxation from the initially populated  $S_2$  Frenkel state to the lower lying  $S_1$  Frenkel state. However, this first step is not a direct transfer but is mediated by a CT state, which crosses the initially populated  $S_2$

state, resulting in a fast population transfer. The transition between the transiently populated CT state and the  $S_1$  Frenkel state is non-radiative and also very efficient due to the strong coupling between both states. The final emission then takes place from the  $S_1$  state.<sup>29–32</sup> The computations show that intra-monomer (e.g., variations of interatomic distances) and inter-monomer (e.g., the torsional angle between both monomers) relaxation effects have to be taken into account. The relaxation effects not only strongly influence the efficiencies of the relaxation processes but also induce strong red shifts in the emission. On the basis of these processes, it was possible to assign the known spectra of PBI aggregates as well as femtosecond time-resolved experiments.<sup>30,31,33</sup> Recently, the model was also successfully employed to explain the ambient-stable, bright, steady-state photoluminescence from long-lived excitons of H-aggregated PBI crystals.<sup>34</sup> Another important example for localization effects is due to intra-chain dynamics.<sup>35,36</sup>

While intra-monomer relaxations leading to the CT geometry were included in the description of the PBI trapping processes, possible geometry relaxations of the Frenkel states themselves were omitted. Their importance became apparent through our monomer computations, which indicated relaxation effects of about 0.3 eV–0.4 eV. This is in the range of the Davydov splittings found for aggregates of these organic semiconductors. The pronounced impact of geometry relaxation raises the question of how their explicit consideration for Frenkel states may change predictions for aggregate clusters. The inclusion of such relaxation effects might not only induce a change in the energetic position of the excited states but may also affect the individual character of the electronic states. For example, the nuclear relaxation of a single monomer in an aggregate could lead to a localization of the exciton on that monomer. The energetics associated with this situation is sketched in Fig. 1 using a dimer as the simplest model for a molecular aggregate. For clarity, we neglect possible influences of CT states. Figure 1 depicts the coupling between the two localized  $S_1$  states of the monomers if both monomers adopt the same geometry (blue) or if one monomer adopts a structure leading to a lower energy of its localized  $S_1$  state (red). If both monomers adopt the same geometry (blue), the exciton is completely delocalized over both monomers due to simple symmetry considerations. Going to the case given in red, one monomer is relaxed to the

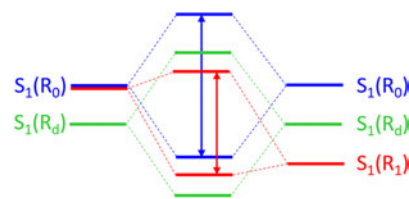


FIG. 1. Comparison of different relaxations. Blue: Both monomers remain in the geometry of the ground state of the monomer ( $R_0$ ). Red: One monomer is relaxed to the geometry of the excited state of the monomer ( $R_1$ ) while the second one remains in the ground state geometry of the monomer ( $R_0$ ). Green: Both monomers relaxed by the same amount ( $R_d =$  optimal geometry for  $S_1$ -state of dimer).

optimal geometry of its localized  $S_1$  state ( $R_1$ ), while the other remains in the geometry of the ground state ( $R_0$ ). Because the energy of the localized  $S_1$  state in the optimal  $S_1$  geometry is lower, both interacting localized states differ not only in their geometry but also in their associated energies. Due to these inequalities, the splitting decreases, and the exciton starts to localize on the monomer in the  $S_1$  geometry for the lower Frenkel state. For the exciton in a dimer, it could also be that both monomers relax by the same amount (green situation,  $R_d$  is the optimal geometry for the  $S_1$  state of the dimer). Then, the energy of the exciton would also decrease, but the exciton remains delocalized over the whole dimer. Which situation is most appropriate to characterize the molecular aggregate depends on the relative energies of the resulting Frenkel states. It might also be that the exciton becomes localized for the lowest Frenkel state but stays delocalized for the second Frenkel state. Since CT states are always delocalized, the degree of delocalization of an exciton may vary strongly during complex relaxation processes involving several electronic states.

Localization effects in self-assembled perylene helices were investigated by Segalina and co-workers.<sup>37</sup> Their study offers valuable insights into the behavior of such aggregates, but localization effects resulting from geometrical changes in monomers were omitted. Such localization effects were investigated by Talipov and co-workers.<sup>38</sup> They investigated relaxation effects in polychromophoric assemblies using covalently linked cofacially arrayed polyfluorenes as model systems. Their computations show that for these systems, excitons always localize on dimers irrespective of the number of chromophores. This behavior is explained by very strong geometry relaxation effects of the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO) in this specific case. Thus, it remains unclear whether these systems can serve as real model systems.

Possible models were already worked out by Fulton and Gouterman who successfully described these excitonic effects with vibronic coupling theory.<sup>39,40</sup> They also showed that this theory is formally equivalent with the pseudo Jahn–Teller (pJT) framework and similar to the potential energy model used in Marcus theory. The vibronic coupling theory was later employed by Diehl and co-workers<sup>41</sup> to assign the spectra for a homologous series of oligo(*p*-phenylene) bridged PBI dimers with intermolecular center-to-center distances ranging from 1.3 nm to 2.6 nm. For these dimer systems, the reorganization effects are in the range of 0.3 eV–0.4 eV, while the coupling between monomers is small by a factor of 5 or even more. Consequently, a double minimum potential arises in which the exciton is either localized on one or the other monomer. In simulations, the monomer on which the exciton remains adopts the  $S_1$  equilibrium geometry, while the other monomer stays in the  $S_0$  structure. For the fully delocalized situation, both monomers were assumed to adopt the  $S_0$  geometry (blue case in Fig. 1).

Dreuw and co-workers employed the pJT framework to model relaxation effects in CO and in the benzene dimer.<sup>42,43</sup> In both cases, they varied inter-monomer geometries of the dimers to modify the coupling between both monomers. They showed that for larger monomer distances, the coupling is weaker than the relaxation, which results in a double-minimum potential. For smaller distances, a parabolic shape is found because the coupling is stronger than the relaxation. More details in relation to our study will be given below.

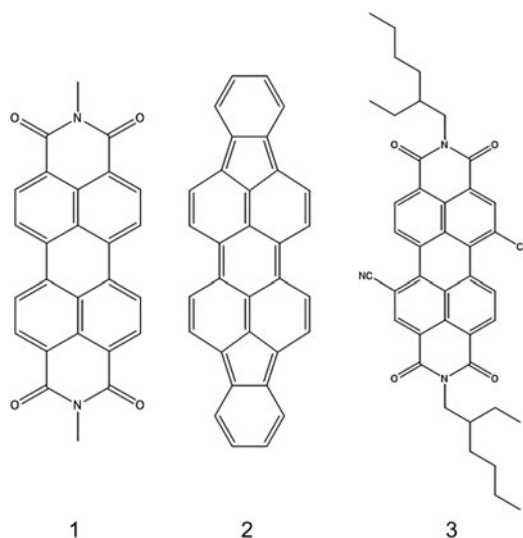


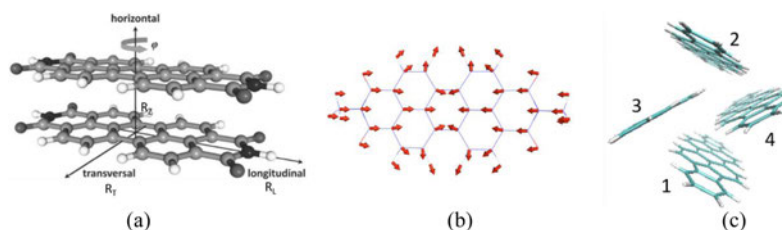
FIG. 2. Lewis structures of the investigated monomer systems. (1) 3,4,9,10-perylene tetracarboxylic acid bisimides (PBIs), (2) diindenoperylene (DIP), and (3) dicyanoperylene-bis(dicarboximide) (PDIR-CN<sub>2</sub>).

In the present study, we will extend these previous investigations of the interplay of geometry relaxation and exciton localization. Whether an exciton is delocalized or localized depends on the subtle interplay between the coupling strength between the units of the aggregate and the decrease in the associated energy due to the intramonomer relaxation of one unit into its  $S_1$  geometry. While the former favors a delocalization of the exciton, the latter induces its localization. We start the investigations with perylene dimers [Figs. 2 and 3(a)] and vary the inter-monomer geometries (e.g., the longitudinal shift or the distance) and the intra-monomer geometry of one monomer. In a second step, we investigate the diindenoperylene (DIP) tetramer as an example for larger clusters. Variations in the electronic character are monitored by computing the localization of the exciton.

This paper is organized as follows: After briefly describing the details of the used theoretical approaches, we first establish expected error bars for different quantum chemical approaches by comparing computed vibrationally resolved excitation energies of the PBI monomer (1) (Fig. 2) with measurements of Klebe *et al.*<sup>44</sup> and Wewer and Stienkemeier.<sup>45,46</sup> For DIP (2) and PDIR-CN<sub>2</sub> (3), we extend these investigations to solvent effects. After establishing the accuracies of our approaches, we turn to the dimer and trimer computations to investigate the interplay of exciton localization and delocalization effects.

#### TECHNICAL DETAILS

Multi-reference approaches would yield ideal reference results for benchmarks because they provide very accurate results for



**FIG. 3.** (a). Description of the varied inter-monomer coordinates for the computed PBI dimer. The coordinates include transversal ( $R_T$ ) and longitudinal ( $R_L$ ) shifts, the distance between both monomers  $R_Z$ , and the rotation of the upper monomer around  $\varphi$ . (b) Enlarged sketch of the geometry change between the equilibrium structures of the  $S_0$  to the  $S_1$  states of the PBI monomer. The sum of the variations is  $<0.005$  Å. (c) Arrangement of the DIP tetramer.

excited states,<sup>47–50</sup> potential energy surfaces,<sup>51,52</sup> as well as other molecular properties.<sup>53</sup> For the size of the present system, however, they are computationally too expensive. Hence, we test the accuracy of time-dependent density functional theory (TD-DFT) calculations against SCS-CC2<sup>54,55</sup> in combination with the SVP (split valence plus polarization)<sup>56,57</sup> and the TZVPP basis sets<sup>56,58</sup>. We compared the TD-DFT-based approaches with respect to SCS-CC2 to keep the computational costs feasible. SCS-CC2 predicted adiabatic excitation energies with a mean absolute error (MAE) of 0.05 eV and a standard deviation of 0.06 eV for a test set of 0–0 transitions in medium-sized and large organic molecules.<sup>59</sup> In this work, SCS-CC2 was shown to perform slightly better than its parent approach CC2 (MAE = 0.09 eV). Furthermore, it improves the description of CT states.<sup>60</sup> For the “low-cost” TD-DFT approach, we mainly used the modern range-separated hybrid  $\omega$ B97X-D functional<sup>61</sup> in combination with SVP, TZVP, cc-pVDZ, cc-pVTZ, and def2-SVP basis sets.<sup>56,58,62,63</sup> This method is sufficiently efficient to compute aggregates up to tetramers.<sup>32,64,65</sup> For the benchmark, we optimized the ground state and the first excited states within the given approximation unless stated otherwise. Vibrational effects were computed using the Franck-Condon approximation alone or in combination with the Herzberg-Teller correction as implemented in the Gaussian16 program package.<sup>66</sup> For the vibrational effects, we used not only the standard time-independent approach<sup>67</sup> but also the time-dependent formalism.<sup>68</sup> For excited state calculations in solution, we used the equilibrium and non-equilibrium options in the IEF-PCM (Integral Equation Formalism version of Polarizable Continuum Model) approach implemented in the Gaussian program suite.<sup>69–71</sup> In the non-equilibrium option, only ultrafast solvent processes (e.g., polarization of the electron cloud of the solvent) are considered.<sup>72</sup> In the equilibrium option, also slower effects (e.g., the reorientation of solvent molecules) are included.<sup>73</sup> An equilibrium calculation describes a situation where the solvent has had time to fully respond to the solute. A non-equilibrium calculation is appropriate for processes that are too rapid for the solvent to have time to fully respond, e.g., for the energy of the 0–0 excitation in an absorption spectrum. The vibrational propagation had to be estimated based on frequencies in the equilibrium model due to technical limitations.

To mimic intra-monomer relaxation effects, we modulated the monomer geometry linearly from the monomer ground state

geometry [ $\vec{R}(S_0)$ ] to the monomer geometry in its  $S_1$  state [ $\vec{R}(S_1)$ ] using

$$\vec{R}_\chi = \vec{R}(S_0) + \chi \Delta\vec{R}, \text{ with } \Delta\vec{R} = \vec{R}(S_0) - \vec{R}(S_1). \quad (1)$$

We varied  $\chi$  from  $-0.5$  to  $1.5$  in  $0.5$  steps for each monomer of the dimer. According to Eq. (1),  $\chi = 0.0$  gives  $\vec{R}(S_0)$ , while  $\vec{R}(S_1)$  is obtained with  $\chi = 1.0$ . An enlarged description of the difference between  $\vec{R}(S_0)$  and  $\vec{R}(S_1)$  for PBI is given in Fig. 3(b). The resulting 2D surfaces for energy and properties were obtained by spline interpolation between the resulting 25 points. We only include relaxations to the equilibrium geometry of the  $S_1$  state of the monomer because higher electronic states of the monomer units are normally not relevant for the photo-induced behavior of crystals, thin films, or amorphous systems.<sup>18,19,74</sup> To investigate the interplay between inter-monomer and intra-monomer changes, we computed the intra-monomer variations for the longitudinal shifts ( $R_L = 0.0$  Å,  $0.5$  Å,  $1.0$  Å,  $1.4$  Å,  $1.7$  Å, and  $2.5$  Å) and the distances between both monomers ( $R_Z = 3.1$  Å,  $3.31$  Å,  $3.5$  Å,  $4.0$  Å,  $5.0$  Å, and  $10.0$  Å). Beside the energies, we also computed the oscillator strengths, the  $P_R$  value, and the CT value. The  $P_R$  values give the delocalization of the exciton, i.e.,  $P_R = 2$  denotes the situation in which the exciton is completely delocalized between the two monomers in a dimer, while  $P_R = 1$  characterizes an exciton that solely resides on one monomer. The CT value gives the percentage of CT character, i.e., CT = 0 corresponds to a pure Frenkel state, and CT = 1 corresponds to a pure CT state. To determine the  $P_R$  and CT values, we employed the TheoDRE program package.<sup>75,76</sup>

#### Benchmark for vibrationally resolved absorption spectra of perylene-based organic semi-conductors

In order to computationally model the relevant processes in organic semiconductor thin films, sufficiently large molecular clusters have to be used to mimic the high density of states, their energy shifts, and the important mixing of CT and Frenkel states. Additionally, the experimental absorption and emission spectra exhibit distinct vibrational progressions, i.e., vibrational effects should also be considered. Finally, the influence of the environment on the energy position of the different states can also be important because the



position of conical intersections or even the energetic order of states might change due to polarizable environments. Obviously, the computational modeling of large molecular clusters and very accurate electronic descriptions exclude each other. Computations of such cluster models are only possible with DFT or even simpler methods because the system size excludes costly high-level multi-reference approaches.<sup>77-79</sup> The inclusion of vibrational effects—even solely on the harmonic level—necessitates geometry optimizations and the determination of the Hessian of the excited state, which is generally too costly for more accurate approaches.

Table I estimates the error bars arising from the use of TD-DFT and neglect of the vibrational effects by comparing the computed results for the PBI monomer (1) with measurements of Klebe *et al.*<sup>44</sup> as well as Wever and Stienkemeier.<sup>45</sup> Klebe *et al.* noted that the first band peak of PBI monomers dispersed in, e.g., polystyrene appears at 530 nm (2.339 eV), while Wever and Stienkemeier measured the 0–0 transition of N–N-dimethyl PBI in He-droplets at 486 nm (2.553 eV). The difference of about 0.2 eV within the experimental spectra may result from the different temperatures, solvent effects, or the influence of substituents. Because we performed the computations of single molecules without considering environmental (solvent) effects, we chose the values of Wever and Stienkemeier as our reference, i.e., a 0–0 transition at 2.553 eV. In the following, we focus on the  $S_0 \rightarrow S_1$  transition because the vertical energy of the  $S_2$  state is about 1 eV higher. Furthermore, the corresponding transition moment vanishes.

From our experience, also the range-separated  $\omega$ B97X-D functional should give reliable results for perylene systems.<sup>74,80-82</sup> In contrast, the B3LYP functional that is very often used for the computations of excited states was less accurate and tends to underestimate the excitation energies of CT states.<sup>28,81,83</sup> Comparing computed vertical excitation energies obtained for vacuum with measured 0–0 energies in He-droplets, B3LYP/TZVP deviates by only –0.122 eV. The discrepancies for SCS-CC2/SVP (+0.418 eV) and  $\omega$ B97X-D/cc-pVDZ (+0.287 eV) are considerably larger. The picture changes if vibrational effects are included and the transitions

between the lowest vibrational states of ground and excited states (0–0 transition) are compared. While the discrepancies between the SCS-CC2/SVP and  $\omega$ B97X-D/cc-pVDZ values and their experimental counterparts decrease to +0.143 eV and +0.008 eV, respectively, the values obtained with B3LYP/TZVP are too low by about 0.3 eV.

The error found for SCS-CC2 decreases to –0.010 eV if the larger TZVPP basis is employed, i.e., SCS-CC2 shows a much improved description with larger basis sets. In contrast, for  $\omega$ B97X-D/cc-pVDZ, we find an error compensation because the agreement to the experimental result of 2.553 eV slightly deteriorates if the basis sets are enlarged (Table II). Comparing the results obtained with the cc-pVDZ basis with those computed with the improved aug-cc-pVDZ basis sets, the 0–0 excitation energy decreases to 2.49 eV, i.e., the deviation from the experimental result increases to about 0.06 eV. The same trend is also found for other larger basis sets. Hence, at least for PBI,  $\omega$ B97X-D/cc-pVDZ represents an excellent choice for the description of the electronic states. The agreement is much better than expected because various benchmarks indicate larger error bars for the computations of excitation energies of organic molecules.<sup>84,85</sup>

The variations going from vertical to vibrationally-resolved excited state estimations mainly result from the relaxation energies of the first excited state [ $E_{S_1}(R_{S_1}) - E_{S_1}(R_{S_0})$ ], which amount to 0.14 eV–0.21 eV depending on the employed method. This explains the distinct vibrational progression found in experimental absorption and emission spectra. Figure 3(b) gives an enlarged description of the geometrical structure difference between the equilibrium geometry of the  $S_0$  state to that of the  $S_1$  state. Note that the sum of all variations is less than 0.005 Å. Table I clearly shows that the small deviations between theoretical B3LYP/TZVP vertical energies and experimental 0–0 energies stem from error compensation. In comparison to the more accurate approaches, B3LYP underestimates the excitation energies by about 0.4 eV–0.5 eV. This error is partly compensated by the neglect of the relaxation energy of the excited state (0.15 eV–0.2 eV). For SCS-CC2/SVP, the

TABLE I. Comparison of computed and measured excitation energies for a PBI monomer.

	B3LYP/ TZVP	$\omega$ B97X-D/ cc-pVDZ	SCS-CC2/ SVP	SCS-CC2/ TZVPP	Reference 44	Reference 45
$E_{\text{vert}}$ (eV) <sup>a</sup>	2.431	2.840	2.971	2.810 <sup>b</sup>		
$E_{\text{adi}}$ (eV) <sup>c</sup>	2.289	2.631	2.766	2.613 <sup>b</sup>		
$E_{\text{relax}}$ (eV) <sup>d</sup>	0.143	0.209	0.205	0.197		
ZPE (eV) <sup>e</sup>	0.030	0.070				
0–0 (eV) <sup>f</sup>	2.259	2.561	2.696 <sup>g</sup>	2.543 <sup>g</sup>	2.339	2.553
$E_{\text{corr}}$ (eV) <sup>h</sup>	0.172	0.279	0.275			

<sup>a</sup>Vertical excitation energy.

<sup>b</sup>Single point computation on the SCS-CC2/SVP geometrical structure.

<sup>c</sup>Adiabatic excitation energy.

<sup>d</sup>Relaxation energy of the excited state ( $E_{\text{vert}} - E_{\text{adi}}$ ).

<sup>e</sup>Energy correction due to zero-point vibrational effects.

<sup>f</sup>(0–0) excitation energy.

<sup>g</sup>Using the ZPE of  $\omega$ B97X-D/cc-pVDZ.

<sup>h</sup>Total correction ( $E_{\text{vert}} - E_{0-0}$ ).

**TABLE II.** Influence of functionals and basis sets on the computed excitation energies for the  $S_1$  state of a PBI monomer. For more explanations, see Table I and the text.

	$\omega$ B97X-D/ cc-pVDZ	$\omega$ B97X-D/ aug-cc-pVDZ	$\omega$ B97X-D/ 6-311++G(d,p)	$\omega$ B97X-D/ def2-TZVP	$\omega$ B97X-D/ def2-TZVPP
$E_{\text{vert}}$ (eV) <sup>a</sup>	2.840	2.763	2.801	2.814	2.811
$E_{\text{adi}}$ (eV) <sup>b</sup>	2.631	2.554	2.587	2.596	2.593
0-0 (eV) <sup>c</sup>	2.561	2.494	2.511	2.510	2.512
	B3LYP/ cc-pVDZ	CAM-B3LYP/ cc-pVDZ	M06-2X/ cc-pVDZ	LC- $\omega$ HPBE/ cc-pVDZ	
$E_{\text{vert}}$ (eV) <sup>a</sup>	2.414	2.803	2.820	3.160	
$E_{\text{adi}}$ (eV) <sup>b</sup>	2.297	2.602	2.620	2.878	
0-0 (eV) <sup>c</sup>	2.226	2.531	2.546	2.803	
	B3LYP/ def2-TZVPP	CAM-B3LYP/ def2-TZVPP	M06-2X/ def2-TZVPP	LC- $\omega$ HPBE/ def2-TZVPP	
$E_{\text{vert}}$ (eV) <sup>a</sup>	2.383	2.766	2.789	3.118	
$E_{\text{adi}}$ (eV) <sup>b</sup>	2.262	2.556	2.580	2.831	
0-0 (eV) <sup>c</sup>	2.174	2.468	2.489	2.744	

<sup>a</sup>Vertical excitation energy.<sup>b</sup>Adiabatic excitation energy.<sup>c</sup>(0-0) excitation energy.

vertical excitation energies do not coincide with the experimental 0-0 transition because no error compensation takes place. Remaining deficiencies in the basis set (SVP vs TZVPP) even lead to an overestimation of the vertical energies by about 0.1 eV-0.2 eV. Assuming SCS-CC2/TZVPP results as a reference value, the  $\omega$ B97X-D functional overestimates the vertical excitation energies by less than +0.1 eV. These estimates are in good agreement with previous investigations of perylenetetracarboxylic dianhydride (PTCDA) single crystals.<sup>29,32,64</sup>

Table S1 shows the results of extended computations employing additional functionals and basis sets, and Table II summarizes the influence of the basis set size for  $\omega$ B97X-D and the influence of the functionals for the smallest (cc-pVDZ) and the largest basis sets (def2-TZVPP). Comparing the vertical excitation energies obtained for the def2-TZVPP basis sets, B3LYP (2.383 eV) gives the lowest value. The B3LYP excitation energy differs by about 0.74 eV from the excited state energy obtained for LC- $\omega$ HPBE (3.118 eV). The  $\omega$ B97X-D functional, which seemed to be quite accurate according to the comparisons given in Table I, predicts a value of 2.811 eV, which is 0.045 eV higher than that of CAM-B3LYP (2.766 eV), the long-range corrected version of B3LYP. The M06-2X functional lies in between (2.789 eV). Comparing the 0-0 transitions with the corresponding experimental values (Table I, 2.553 eV),  $\omega$ B97X-D/def2-TZVPP shows the best agreement (-0.041 eV deviation from the experimental value of Ref. 45). CAM-B3LYP/def2-TZVPP and M06-2X/def2-TZVPP give comparable deviations (-0.085 eV and -0.064 eV, respectively). As already discussed above, B3LYP/def2-TZVPP considerably

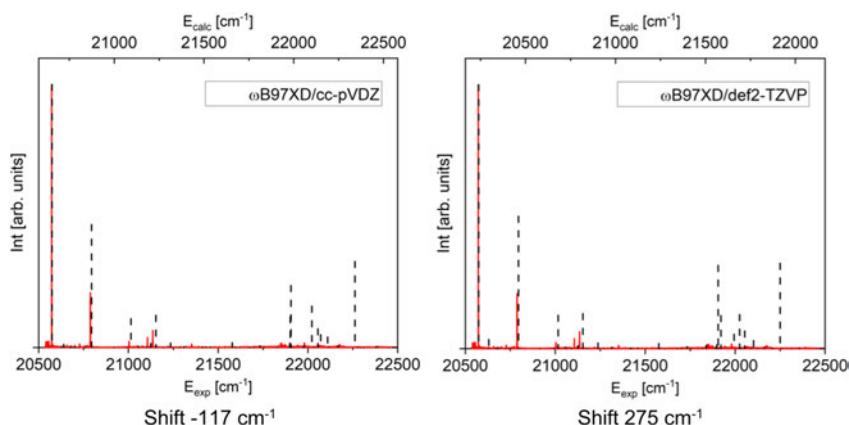
underestimates the experimental value by 0.379 eV, and PBE0 behaves rather similar, while LC- $\omega$ HPBE/def2-TZVPP overestimates the experimental value by 0.191 eV.

Tables II and S1 also show interesting trends with respect to basis sets' selection. Going from larger basis sets to the double-zeta cc-pVDZ basis set, which represent the smallest employed basis sets, the computed 0-0 excitation energies increase irrespective of the used functionals. However, the increments are less than 0.1 eV in all cases. This shows that the neglect of vibrational effects is a more severe approximation than the use of smaller basis sets. The aug-cc-pVDZ basis sets are well adapted to the description of electronically excited states, and indeed, the corresponding 0-0 excitation energies are slightly lower than the ones obtained with the def2-TZVPP basis sets. However, for the most reliable functionals ( $\omega$ B97X-D, CAM-B3LYP, and M06-2X), the predicted values become too low in comparison to the experimental value of 2.553 eV. In summary, the best agreement is found for M06-2X/cc-pVDZ, which only deviates by 56  $\text{cm}^{-1}$  (0.007 eV).  $\omega$ B97X-D/cc-pVDZ underestimates the experimental value by 65  $\text{cm}^{-1}$  (0.008 eV).

Note that while  $\omega$ B97X-D/cc-pVDZ predicts the 0-0 excitation energy rather accurately, a comparison with the SCS-CC2 benchmark values indicates that it seems to overestimate the relaxation energy by about 4%. As indicated by the even larger relaxation energies obtained with larger basis sets (Table II), the choice of the relatively small cc-pVDZ basis set again tends to improve the performance of the method.

Figure 4 depicts a comparison of computationally modeled and experimental absorption spectra. We present the spectrum obtained





**FIG. 4.** Comparison of the experimental (red full lines) and computed (black dashed line)  $S_0$ - $S_1$  absorption spectra of the PBI monomer. The computed spectra were shifted by  $-117\text{ cm}^{-1}$  and  $275\text{ cm}^{-1}$ , respectively, so that the 0-0 lines of experiment and theory coincide. The computed intensities are adjusted with respect to the experimental 0-0 transition intensity; however, the intensities cannot be compared due to experimental reasons. For more information, see the text.

within the  $\omega\text{B97X-D/cc-pVDZ}$  ansatz as an example. To simplify the comparison, the computed spectra were shifted so that the 0-0 lines of experiment and theory coincide. Furthermore, the computed intensities were adjusted with respect to the experimental 0-0 transition intensity. The overall agreement in the energetic positions of the quantized absorption peaks is excellent. As mentioned by Wewer and Stienkemeier,<sup>45</sup> the intensities in their laser induced fluorescence spectra do not agree with absorption spectra, which are simulated in our theoretical approach. It is obvious that the sizes of the employed AO basis sets do not have a relevant influence on the accuracy of the computational prediction.

Table III and Fig. 5 extend our investigations to the DIP molecule. Table III summarizes the computed data and compares the 0-0 band of the  $S_0 \rightarrow S_1$  transition with the energy of the lowest band of the experimental absorption spectrum, measured in acetone at room temperature.<sup>87</sup> Figure 5 compares the stick spectrum and a convoluted version of the computed vibrational transitions with the measured absorption spectrum. The computed data are shifted by 0.04 eV to match the maximum of the energetically lowest peak in the absorption spectrum. More information is given in the supplementary material. For the DIP monomer, the  $S_2$  state is only slightly higher in energy than the  $S_1$  state. Nevertheless, it can be omitted in regard to the assignment of the absorption spectrum as the  $S_0 \rightarrow S_2$  transition is symmetry forbidden.<sup>88</sup>  $S_3$  will also not contribute because it is too high in energy and the computed transition dipole moment vanishes.<sup>88</sup> We combine the  $\omega\text{B97X-D/cc-pVTZ}$  approach with a (non-equilibrium) continuum-solvation approach to account for solvent effects (acetone) in the absorption spectrum. We predict the 0-0 band of the  $S_0$ - $S_1$  excitation at 2.38 eV, which is in excellent agreement with the energetically lowest peak of the experimental absorption spectrum of the monomer (2.35 eV). The corresponding vertical energies are predicted at about 2.7 eV, underlining that the neglect of vibrational effects leads to a blue shift in

the absorption spectrum of 0.32 eV in vacuum and 0.34 eV in acetone. Table III shows that for vacuum and acetone, the strongest contributions to the difference between vertical excitation energies and the 0-0 peak result from the geometrical relaxation into the  $S_1$  geometry (0.24 eV-0.25 eV). Variations in the zero-point energies (ZPEs) of both electronic states are about 0.1 eV, i.e., they further decrease the excitation energy. All other contributions to the

**TABLE III.** Monomer calculations on the  $S_0 \rightarrow S_1$  transition of DIP. All energies are given in eV. For more explanations, see Table I and the text.

Solvent	Vacuum		Acetone		Experiment
	$\omega\text{B97X-D}$	SCS-CC2	Eq.	Non-eq.	
Method	$\omega\text{B97X-D}$	SCS-CC2	$\omega\text{B97X-D}$		
Basis sets	cc-pVDZ	cc-pVTZ			
$E_{\text{vert}}$ (eV) <sup>a</sup>	2.80	2.77	2.72	2.72	
$E_{\text{adi}}$ (eV) <sup>b</sup>	2.59	2.53	2.29	2.47	
$E_{\text{relax}}$ (eV) <sup>c</sup>	0.21	0.24	0.42	0.24	
ZPE (eV) <sup>d</sup>	0.11	0.10	0.09		
0-0 (eV) <sup>e</sup>	2.48	2.43	2.20	2.38 <sup>f</sup>	2.35
$E_{\text{corr}}$ (eV) <sup>g</sup>	0.32	0.34	0.51	0.34	

<sup>a</sup>Vertical excitation energy.

<sup>b</sup>Adiabatic excitation energy.

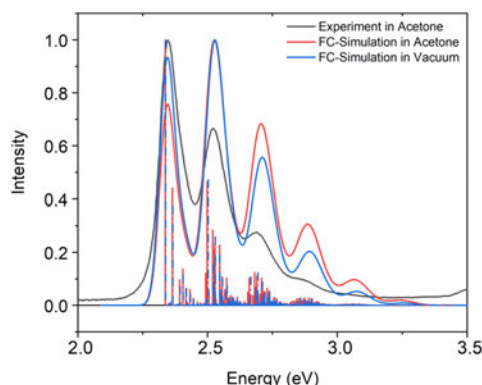
<sup>c</sup>Relaxation energy of the excited state ( $E_{\text{vert}} - E_{\text{adi}}$ ).

<sup>d</sup>Energy correction due to zero-point vibrational effects.

<sup>e</sup>(0-0) excitation energy.

<sup>f</sup>Calculated with the equilibrium ZPE since frequency calculations are not feasible in non-equilibrium solvation.

<sup>g</sup>Total correction ( $E_{\text{vert}} - E_{0-0}$ ).



**FIG. 5.** The monomer  $S_0 \rightarrow S_1$  absorption spectrum of DIP ( $\omega$ B97X-D/cc-pVTZ): The black full line shows the experimental data (in acetone).<sup>86</sup> The computational estimate in acetone is given in red (stick spectrum and its Gaussian convolution, FWHM = 500  $\text{cm}^{-1}$ ). For a better comparison, these theoretical data were shifted by 0.04 eV so that the energetically lowest peak of the theoretical and experimental spectra coincides. The corresponding data in blue give the computational results in vacuum. These datapoints are shifted by 0.14 eV.

0–0 energies are smaller. The reduction in the basis set's size from valence triple-zeta (cc-pVTZ) to valence double-zeta (cc-pVDZ) seems to introduce an error of 0.06 eV. In contrast to PBI, an increase in the basis sets slightly improves the agreement with the experimental values for DIP. Using the wave function based SCS-CC2 approach in combination with the cc-pVTZ basis set, the computed excitation energies change by only about 0.05 eV. In summary, as for PBI, the explicit consideration of the photo-induced relaxation of the geometry (vertical excitation vs adiabatic excitation) introduces the strongest shift in the computed excitation energies.

Figure 5 shows the vibrational structure of the theoretical absorption spectra in acetone (red) or vacuum (blue) at  $T = 0$  K and the experimental spectrum.<sup>86</sup> For the corresponding convolution with a Gaussian, we used a FWHM of 500  $\text{cm}^{-1}$  because it seemed to reflect the overall shape of the bands better than, for example, a FWHM of 270  $\text{cm}^{-1}$ , which is the default option in the Gaussian16 program. Decreasing the FWHM introduces additional features in the envelope of the first band, which are not found in the experimental spectra (Figs. S1 and S2). However, while the computed energy spacings between the first two bands and their widths agree quite nicely with the experimental results, the intensity ratio between the two lowest energy (highest intensity) bands is predicted less well (Fig. 5). The experimental measurements find a decrease in intensity with an increase in energy. In contrast, the lowest band has less intensity than the next higher one in the computed spectrum. This erratic intensity ratio is found for the theoretical spectra in vacuum and even stronger in acetone. Thus, while  $\omega$ B97X-D/cc-pVTZ reproduces the excitation energy quite accurately, it overestimates the relaxation of the DIP molecule due to excitation in the  $S_1$  state.

The amount of which the relaxation of the excited state is overestimated can be assessed by approximating the potentials of the ground end excited states by shifted harmonics. Then, the intensity  $I_n$  of the vibrational line  $n$  is given by a Poisson distribution ( $I_n = \frac{e^{-S} S^n}{n!}$ ), with the Huang–Rys parameter  $S = \frac{1}{2}(\Delta u)^2$ . Here,  $\Delta u$  is the difference between the equilibrium distances in dimensionless coordinates.<sup>89</sup> With the observed intensity patterns, it results that  $\omega$ B97X-D/cc-pVTZ overestimates the structural change in the excited state ( $\Delta u$ ) by about 20%. This seems sufficiently small to employ this model for the qualitative analysis of the excited states shown below. Furthermore, as discussed in the [supplementary material](#), the shape of the theoretical spectrum changes by modifying the FWHM of the Gaussian convolution or by including temperature effects (Fig. S2), while the Herzberg–Teller correction is hardly significant (Fig. S3).

The corresponding data for the PDIR-CN<sub>2</sub> molecule are given in the [supplementary material](#) (Table S2, Fig. S4). We find the same trends as already discussed for PBI and DIP. Similar performance of the quantum chemical models for these molecules could be expected since the electronic excitations are dominated by the perylene core.<sup>28,65</sup> However, in a recent study, we also find comparable relaxation effects for tetracene, which represents a considerably larger molecule.<sup>90</sup>

#### Intra-monomer relaxation effects in aggregate structures

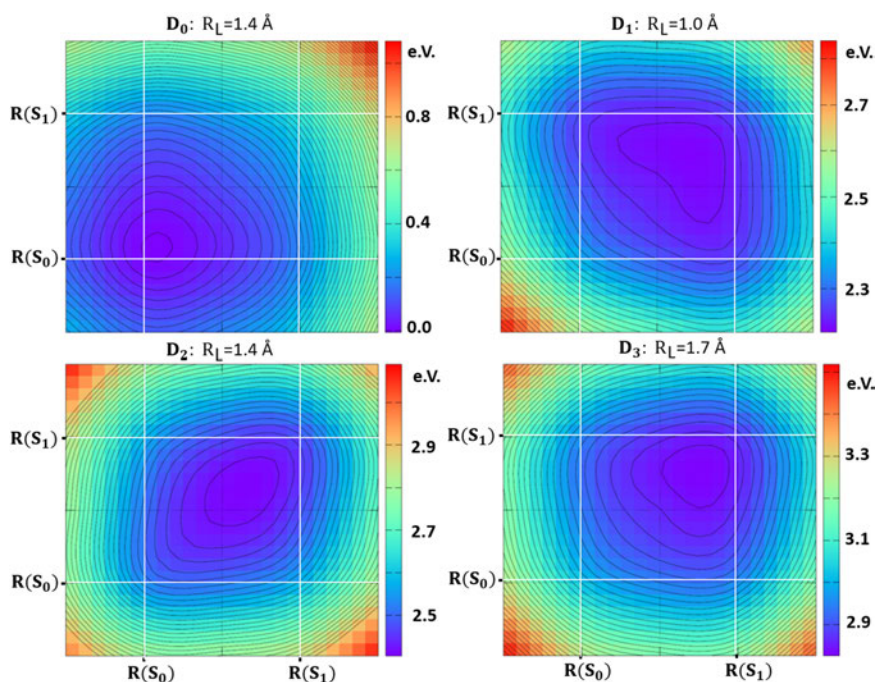
The above-mentioned monomer relaxation effects of 0.2 eV –0.4 eV raise the question how distortions of the monomer geometry (intra-monomer relaxation) influence the photophysical behavior of organic semiconductors. The Davydov splitting in such molecular aggregates is often in a similar range. Previous computational investigations often neglect intra-monomer relaxation effects since full optimizations of aggregates are too time-consuming if sufficiently accurate approaches are employed. Hence, to make the computations feasible, the frozen-monomer approximation has often been used. In this approximation, the geometries of the monomers are fixed to their  $S_0$  equilibrium structures. Using these frozen monomers, the potential energy surfaces are computed as a function of the inter-monomer geometrical parameters to model inter-monomer relaxation effects. The inter-monomer parameters are shown in Fig. 3(a). To investigate how inclusion of intra-monomer effects changes the figure, we utilized the PBI-dimer for model calculations. For this model, we took various relative orientations of the monomers in the dimer and characterized the intra-monomer relaxation by distorting each monomer independently from the equilibrium structure of the ground state of the monomer  $\bar{R}(S_0)$  to the equilibrium structure of the first excited state of the monomer  $\bar{R}(S_1)$ . We performed the computations for the vacuum case, an almost nonpolar environment with a high refractive index ( $\epsilon = 5$  and  $\epsilon_{\text{inf}} = 4.45$ ) and water ( $\epsilon = 78.355$  and  $\epsilon_{\text{inf}} = 1.776$ ) to investigate the influences of polarizable environments. However, since the comparison of the corresponding PESs (Fig. S5) shows very small changes when changing the polarity of the environment, we focus the discussion on the values obtained for the almost nonpolar medium.

Figure 6 summarizes the PESs of the various states as a function of distortions of the monomers. To differentiate between the

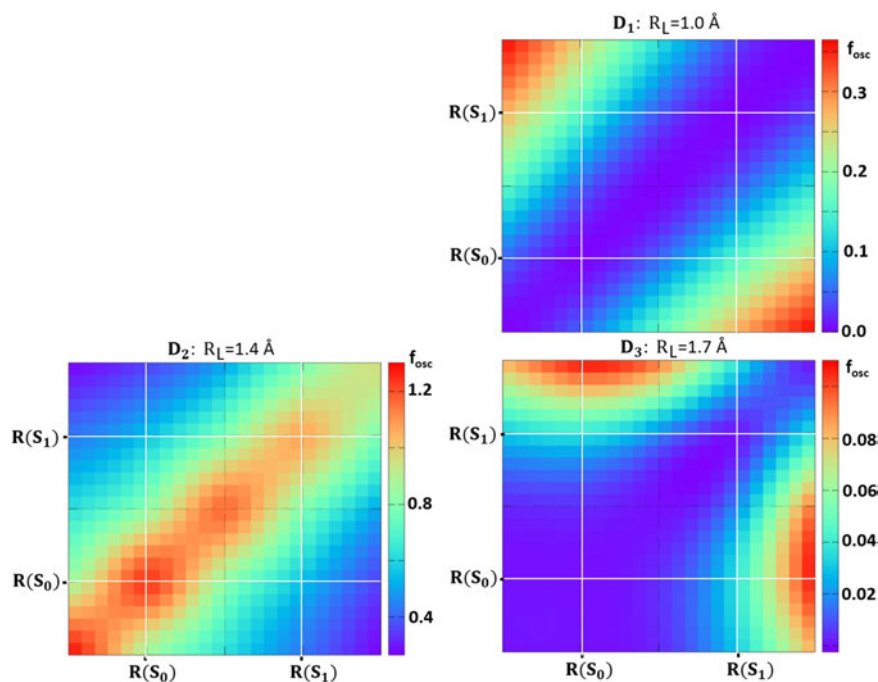
electronic dimer and monomer states, the dimer states are abbreviated as  $D_0$ – $D_3$ , where  $D_0$  represents the singlet ground state of the dimer, while  $D_n$  stands for its  $n$ -th excited singlet state. The white lines are given for a better overview. Within the figures, the abscissa and ordinate depict the distortion of the monomers. Moving along the lowest horizontal white line of the PESs, one monomer remains fixed at its ground state geometry, while the other monomer varies its geometry from  $\bar{R}_{-0.5} = \bar{R}(S_0) + (-0.5)\Delta\bar{R}$  to  $\bar{R}_{1.5} = \bar{R}(S_0) + 1.5\Delta\bar{R}$  [Eq. (1)]. In Eq. (1),  $\Delta\bar{R} = \bar{R}(S_0) - \bar{R}(S_1)$  represents the difference between the equilibrium structures of the monomer in its ground and first excited singlet states. Consequently,  $\bar{R}(S_0)$  is related to  $\chi = 0.0$ , while  $\bar{R}(S_1)$  corresponds to  $\chi = 1.0$ . Along the left vertical white line, the structure of the other monomer changes. The PES must be symmetric due to the symmetry of the system [Fig. 3(a)]. Along the diagonal, both monomers are equally distorted so that the exciton must be completely delocalized over the whole dimer. In the outer diagonal parts, both monomers are differently distorted so that the exciton can localize on one monomer. Single points on the PESs are designated as  $[\bar{R}_A; \bar{R}_B]$  where the structure of the first monomer is  $\bar{R}(S_0) + A\Delta\bar{R}$ , while the second monomer's geometry is  $\bar{R}(S_0) + B\Delta\bar{R}$ .

To make the small variations visible, we choose contour lines with an energy difference of only 0.02 eV. All energies are given with respect to the equilibrium geometry of the ground state, and all dimer structures have a vertical (molecular plane to plane) distance of  $R_Z = 3.31$  Å [Fig. 3(a)]. All calculations were performed with the  $\omega$ B97X-D/def2-TZVP approach. Previous investigations show that the dimer is additionally stabilized by small transversal shifts ( $R_L$ ) or alternatively by a torsion of both monomers with respect to each other.<sup>13,27</sup> Nevertheless, in the scope of the present work, we only vary the longitudinal shift and the distance between the two monomers. Figures 6 and 7 compare the energies and the properties of the various dimer states at their optimal  $R_L$  orientations. This comparison allows us to differentiate between intra- and inter-monomer relaxation effects. Figures 8 and 9 focus on the differences in the behavior of the  $D_1$  and  $D_2$  states for the same inter-monomer orientations (the same  $R_L$  and  $R_Z$  values).

As expected, the minimum of the  $D_0$  surface (Fig. 6, top left) is found if both monomers adopt essentially the monomer ground state geometry. The small elongation of the equilibrium bond distance in the direction of the excited state indicates that the



**FIG. 6.** PES as a function of the monomer distortions computed for a monomer distance of 3.31 Å and the optimal longitudinal shift  $R_L$  for the respective electronic state. The energy color code is given on the left side of each PES. The energy difference between two contour lines is 0.02 eV.  $R(S_0)$  and  $R(S_1)$  denote equilibrium geometries for the monomer in its ground ( $S_0$ ) and first excited states ( $S_1$ ), respectively. The white lines are given for a better overview. For more information, see the text.



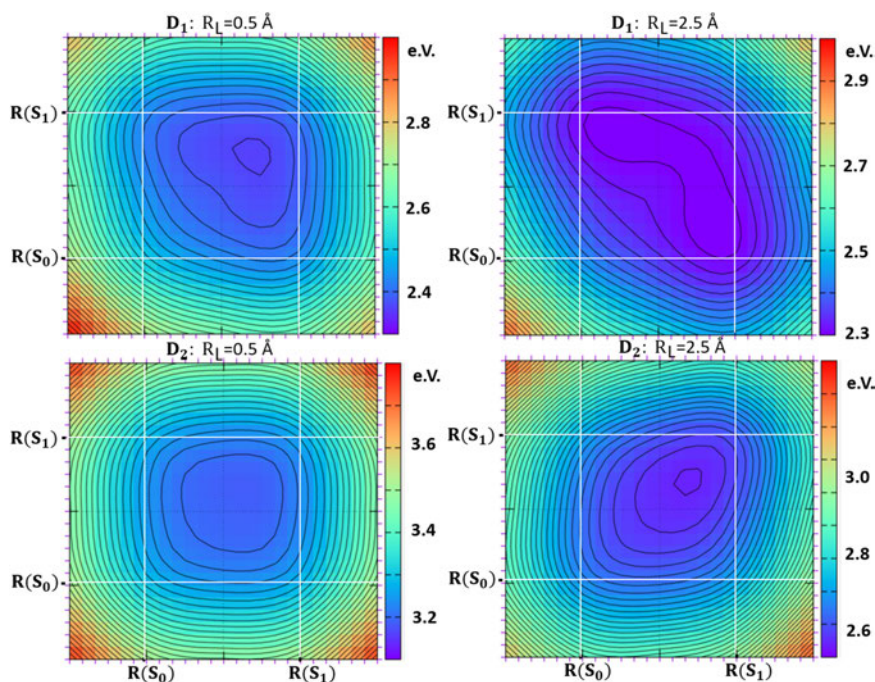
**FIG. 7.** Variation in the oscillator strength  $f_{\text{osc}}$  (a.u.) as a function of the monomer distortions computed for a monomer distance of 3.31 Å and the optimal longitudinal shift  $R_L$  for the given electronic state. The color code for  $f_{\text{osc}}$  is given on the left side of each surface.  $R(S_0)$  and  $R(S_1)$  denote equilibrium structures for the monomer in its ground and first excited states, respectively. The white lines are given for a better overview. For more information, see the text.

interaction of the two monomers tends to promote electrons from the HOMO of one monomer into the LUMO of the other one and vice versa. Thus, both monomers are partially excited due to the interaction with the other one, which introduces the observed departure of the dimer equilibrium geometry from the monomer values. However, the energy lowering with respect to the monomer structure amounts to only about 0.02 eV, which is hardly significant. For the  $D_1$  state, our computations predict different inter- and intra-monomer geometries compared to the ground state. The optimal inter-monomer orientation for the ground state is found at  $R_L = 1.4$  Å, whereas  $R_L = 1.0$  Å gives the lowest computed energy for the first excited state. In the  $D_1$  state, the optimal intra-monomer geometry is at  $[\bar{R}_{0.7}; \bar{R}_{0.7}]$ . This indicates that an exciton that is delocalized over a dimer influences the geometry of the single monomers less than it is found for the  $S_1$  state of the monomer. This is in line with the work of Harbach and Dreuw on the CO dimer.<sup>42</sup> In the CO dimer, the CO distance of the dimer  $D_1$  state is 1.1875 Å, which lies between the distance found for  $S_0$  (1.126 Å) and  $S_1$  (1.222 Å) of the monomer. In terms of the  $\chi$ -parameter, they obtain a value of 0.65, which is surprisingly close to our value of 0.7. Figure 6 also reveals that the  $D_1$ -PES shows a curved valley ranging from

$[\bar{R}_{0.7}; \bar{R}_{0.3}]$  over  $[\bar{R}_{0.7}; \bar{R}_{0.7}]$  to  $[\bar{R}_{0.3}; \bar{R}_{0.7}]$ . In this range, the energy varies by less than 0.02 eV, i.e., all structures are virtually isoenergetic. For the vertical energy of the  $D_1$  state at the structure  $[\bar{R}_{0.0}; \bar{R}_{0.0}]$  and  $R_L = 1.0$  Å, we compute an excitation energy of 2.452 eV. Intermolecular relaxation along the longitudinal axis to  $R_L = 1.0$  Å decreases the excitation energy to 2.360 eV ( $E_{\text{relax}}^{\text{inter}} = 0.09$  eV). Adding the intra-monomer relaxation ( $[\bar{R}_{0.0}; \bar{R}_{0.0}]$  to  $[\bar{R}_{0.7}; \bar{R}_{0.7}]$ ) for  $R_L = 1.0$  Å, the excitation energy drops to 2.225 eV, i.e.,  $E_{\text{relax}}^{\text{intra}}$  amounts to 0.135 eV. Hence, the intra-monomer relaxations are slightly larger than the inter-monomer ones. However, for the former, we only included one degree of freedom, while all possible relaxations are included in the latter. Note that  $E_{\text{relax}}^{\text{intra}}$  computed for the dimer (0.135 eV) is smaller than the variation found for the monomer (Table I; 0.201 eV).

For the  $D_2$  state, inter- and intra-monomer effects differ from those computed for the  $D_1$  state. The optimal  $R_L$  value for  $D_0$  and  $D_2$  is equal, i.e., inter-monomer relaxation effects vanish. Furthermore, the minimum of the  $D_2$ -PES lies along the diagonal, i.e., structures with equally distorted monomers are favored with respect to those with unequally distorted monomers. The intra-monomer relaxation effects for the  $D_2$  state are computed to be 0.13 eV. As





**FIG. 8.** Comparison of the PESs obtained for the longitudinal shifts  $R_L = 0.5$  and  $2.5$  Å. The horizontal distance between both monomers is always  $3.31$  Å. The energy color code is given on the left side of each PES. The energy difference between two contour lines is  $0.02$  eV.  $R(S_0)$  and  $R(S_1)$  denote equilibrium structures for the monomer in its ground and first excited states, respectively. The white lines are given for a better overview. For more information, see the text.

for the  $D_1$  state, the monomers do not fully relax to the  $S_1$  geometry of the monomer. The minimum ranges from about  $[\bar{R}_{0.5}; \bar{R}_{0.5}]$  to  $[\bar{R}_{0.7}; \bar{R}_{0.7}]$  instead.

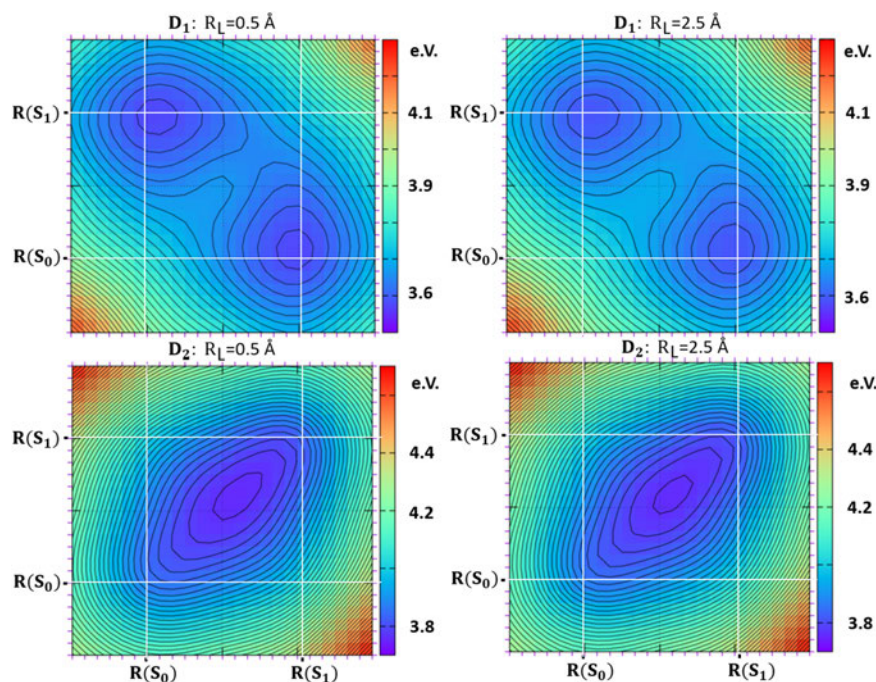
The shape of the intra-monomer PES of the  $D_3$  state seems to be a mixture of shapes of the  $D_1$  and  $D_2$  states. However, while  $D_1$  and  $D_2$  possess mainly Frenkel character, the  $D_3$  represents a CT state. Hence, to describe its intra-monomer relaxation effects, we would have to distort the monomers to the anionic and cationic structures rather than to the  $S_1$  structures of the monomers. Since these additional variations are out of the scope of the present work, we will focus in the following on  $D_1$  and  $D_2$  states of the dimer.

The PES of the  $D_1$  states possesses a broad minimum, which comprises unequally distorted monomers and equally distorted monomers. For the latter, the exciton must be completely delocalized due to symmetry reasons. Since unequally distorted monomers can induce exciton localization, it is of interest to investigate the variation in the  $P_R$  value obtained from the TheoDOR program. The  $P_R$  value differentiates between completely delocalized ( $P_R = 2.0$ ) and localized ( $P_R = 1.0$ ) excitons. Hence, for equally distorted monomers,  $P_R = 2$  is obtained due to the symmetry. The  $P_R$  value for  $[\bar{R}_{0.3}; \bar{R}_{0.7}]$  is about 1.9, showing that excitons remains

mainly delocalized in the range of the curved valley of the  $D_1$  state. For the  $D_2$  state, the minimum runs along the diagonal, and therefore, the excitons are also delocalized.

Another property that differentiates between localized and delocalized excitons is the oscillator strength  $f_{osc}$ . For a completely delocalized exciton,  $f_{osc}$  should vary roughly between twice the monomer value and zero. For completely localized excitons, all oscillator strengths should resemble the monomer values. For example, for an H-aggregate, the  $D_1$  state should be dark ( $f_{osc} = 0$ ) if the exciton is delocalized. If the exciton starts to localize, the oscillator strength of  $D_1$  should increase to the limit of the oscillator strength of the  $S_1$  monomer state. The corresponding data are summarized in Fig. 7. Indeed, for the  $D_1$  state,  $f_{osc}$  vanishes for equally distorted monomers. For unequally distorted ones, the values smoothly increase. For the  $D_2$  state, the opposite is found. The variations found for  $D_3$  resemble the behavior expected for a CT state.

In the vibronic coupling theory, the relative size of the Davydov splitting with respect to the monomer relaxation energy determines whether an asymmetrical distortion of both monomers will take place or not. The monomer relaxation energy within the pJT framework is related to the size of the vibronic interactions.<sup>42,43</sup>



**FIG. 9.** Comparison of the PESs obtained for the longitudinal shifts of  $R_L = 0.5$  and  $2.5$  Å. The horizontal distance between both monomers is always  $5.0$  Å. The energy color code is given on the left side of each PES. The energy difference between two contour lines is  $0.02$  eV.  $R(S_0)$  and  $R(S_1)$  denote equilibrium structures for the monomer in its ground and first excited states, respectively. For more information, see the text.

Figures 8 and 9 investigate these interplays. We compare the situations obtained for  $R_L = 0.5$  Å and  $R_L = 2.5$  Å in Figs. 8 and 9. The horizontal distance between both monomers ( $R_Z$ ) is  $3.31$  Å in Fig. 8. Thus, the coupling is stronger but should vary due to the variation in  $R_L$ . The coupling  $J$  is given as half of the energy difference between  $D_1$  and  $D_2$  ( $\Delta E = 2J$ ) for equally distorted monomers. Using this relationship, for  $R_Z = 3.31$  Å and  $R_L = 0.5$  Å, our computations predict a coupling of  $0.43$  eV for  $[\bar{R}_{0,0}; \bar{R}_{0,0}]$  and  $0.39$  eV  $[\bar{R}_{1,0}; \bar{R}_{1,0}]$ . We compute a coupling of  $0.12$  eV for both monomer relaxations for  $R_Z = 3.31$  Å and  $R_L = 2.5$  Å, i.e., the coupling is indeed smaller. Figure 8 shows that this variation in the coupling influences the shapes of the PESs, but the overall energy variations are small. Figure S6 gives the  $f_{osc}$  values. They show that even for  $[\bar{R}_{0,5}; \bar{R}_{1,0}]$ ,  $f_{osc}$  behaves as expected for an H-aggregate ( $R_L = 0.5$  Å) or a J-aggregate ( $R_L = 2.5$  Å), showing the excitons to be still delocalized.

For  $R_Z = 5.0$  Å, the coupling is only  $0.04$  eV– $0.05$  eV for  $R_L = 0.5$  Å and  $0.03$  eV for  $R_L = 2.5$  Å. The considerably smaller couplings are reflected in the PES shapes, as given in Fig. 9. The PESs of the  $D_1$  state possess two distinct minima at  $[\bar{R}_{1,0}; \bar{R}_{0,0}]$  and  $[\bar{R}_{0,0}; \bar{R}_{1,0}]$  for which one monomer remains at the ground state structure,

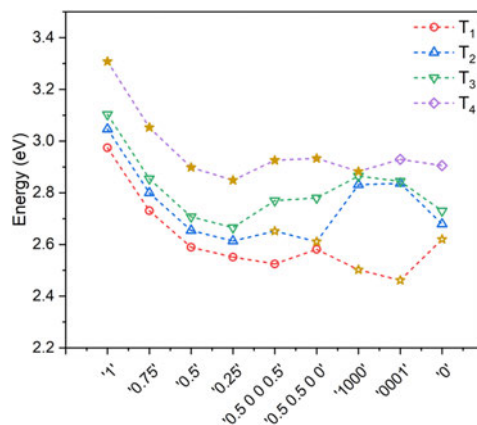
while the other relaxes to the  $S_1$  monomer structure. The exciton is strongly localized in these positions at the relaxed monomer as shown by the computed  $P_R$  value of  $1.08$ . Hence, the system represents a localized exciton on the relaxed monomer, which is weakly perturbed by the non-excited PBI monomer at the distance of  $5$  Å for both minima. For  $R_L = 0.5$  Å, the intra-monomer relaxation energy ( $E_{relax}$ ) is  $0.204$  eV, while  $0.216$  eV is computed for  $R_L = 2.5$  eV. We included environmental effects for these calculations. The corresponding vacuum environment data are  $0.143$  eV for  $R_L = 0.5$  Å and  $0.155$  eV for  $R_L = 2.5$  Å. Both values are smaller than  $E_{relax}$  calculated for the monomer in vacuum ( $0.218$  eV in Table II;  $\omega B97X-D/def2-TZVP$ ). This implies that the disturbed, non-excited monomer slightly decreases the relaxation energy, while a polarizable environment leads to a small increase. Our computations show the well-known double-well potential of the Marcus theory or of a pJT-surface with a strong vibronic coupling for the PESs of the  $D_1$  state. The computations also show that an estimate of the barrier height based on the unrelaxed geometry  $[\bar{R}_{0,0}; \bar{R}_{0,0}]$  leads to an overestimation thereof. The barrier height is estimated at  $0.20$  eV for  $R_L = 0.5$  Å if  $[\bar{R}_{0,0}; \bar{R}_{0,0}]$  is taken as top of the barrier. The value reduces to only about  $0.09$  eV if the optimal arrangement ( $\approx [\bar{R}_{0,7};$

$\bar{R}_{0,7}$ ) is used. The corresponding values for  $R_L = 2.5 \text{ \AA}$  are 0.22 eV (via  $[\bar{R}_{0,0}; \bar{R}_{0,0}]$ ) and 0.08 eV (via  $[\bar{R}_{0,7}; \bar{R}_{0,7}]$ ).

The shapes of the PESs of  $D_1$  and  $D_2$  differ considerably for  $R_Z = 3.31 \text{ \AA}$  and  $5.0 \text{ \AA}$ . Especially the computed shapes for  $5.0 \text{ \AA}$  nicely reflect the 1D potential curves given by Marcus theory for ground and excited states of weakly interacting monomers or that of a pJT effect with strong vibronic coupling. In Fig. 9, the  $D_1$  state exhibits the double-well potential if both minima are connected, while the shape of  $D_2$  along this coordinate is harmonic. A very similar potential was obtained by Garcia-Fernandez *et al.* for the benzene dimer in  $D_{6h}$  symmetry at a distance of  $5 \text{ \AA}$ , emphasizing the generality of our results.<sup>43</sup> As in the Marcus theory picture, the difference between  $D_1$  and  $D_2$  in our model results from the required orthogonality between two solutions of a Hermitian operator such as the electronic Hamilton operator. The exciton is localized on the relaxed monomer for the minimum of the  $D_1$  state at  $[\bar{R}_{0,0}; \bar{R}_{1,0}]$ . Due to the required orthogonality, a localized exciton would reside on the unrelaxed monomer for  $D_2$ . If it is localized on the relaxed monomer, it would populate the  $S_2$  state of the monomer, which is considerably higher in energy. However, if the exciton localized on the unrelaxed monomer, both monomers exhibit unfavorable geometries for this localization. Hence, the symmetrical relaxation of both monomers leading to a delocalized exciton represents the energetically most stable situation for  $D_2$ .

For our investigations on a larger cluster, we selected DIP tetramers in which the monomers take the position of the crystal structure as a first example. While the monomer positions were taken from the experimental crystal structure (CCDC code 642482),<sup>91,92</sup> the intra-monomer coordinates were determined by full optimizations for the  $S_0$  and  $S_1$  states of the DIP monomer. As for the PBI dimer, the intra-monomer geometries were varied by changing the monomer structures linearly from  $\bar{R}(S_0)$  to  $\bar{R}(S_1)$ . The energy variations obtained for the four lowest electronic states of the tetramer ( $T_1$ – $T_4$ ) as a function of varying monomer geometries are collected in Fig. 10. The abscissa gives the distortion of the respective monomers toward  $\bar{R}(S_1)$ . To simplify the abbreviations, we introduce a different nomenclature for the tetramers. If all monomers are equally distorted, the distortion is indicated only once. For example, “1” denotes the structure in which all monomers are in the  $S_1$  geometry, while “0.25” abbreviates the structure in which all monomers are distorted to  $\bar{R}_{0,25}$  ( $[\bar{R}_{0,25}; \bar{R}_{0,25}; \bar{R}_{0,25}; \bar{R}_{0,25}]$  in dimer notation). For differently distorted monomers, the abbreviation denotes which one is distorted (Fig. 3). For the [0.5 0 0 0.5] monomer, 1 and 4 are distorted to  $\bar{R}_{0,5}$ , while 1000 and 0001 abbreviate the clusters in which monomer 1 or 4 adopts the  $S_1$  geometry. For the cluster computations, these geometries have different energies because the corresponding monomers are not transferred into each other by a symmetry operation. Symmetrically equivalent are only monomers 1 and 2 as well as 3 and 4.

As expected, the global minimum of the tetramer ground state  $T_0$  is obtained for “0.” Distortions of the monomers toward  $\bar{R}(S_1)$  increase the ground state energies. For example, for “0.25,” the energy slightly increases by 0.05 eV, while an energy increase of 0.81 eV is obtained for “1.” For “1000,” the energy raises by 0.2 eV. The minimum of the first electronically excited tetramer state  $T_1$  is found for the structure “0001,” i.e., if monomer 4 (Fig. 3) adopts the  $S_1$  geometry, while all other monomers remain in the ground state geometry. The excitation energy with respect to the energy of the



**FIG. 10.** Energies of the four lowest excited states of the DIP tetramer ( $T_1$  in red,  $T_2$  in blue,  $T_3$  in green, and  $T_4$  in purple) as a function of the monomer geometries (for details see text). All energies are given with respect to the energy of the equilibrium geometry of the ground state. \* indicates states with considerable oscillator strengths. A full \* denotes  $f_{osc} > 0.5$ , while an empty one gives  $f_{osc} > 0.2$ .

equilibrium geometry of  $T_0$  ( $E_{adi}$ ) is 2.46 eV, and the corresponding value for “1000” is 2.50 eV. The computed  $P_R$  values of 1.04 and 1.13 indicate that the unequal distortions lead to a nearly complete localization of the exciton on the distorted monomer. If all monomers are equally distorted by a quarter of the difference between  $S_0$  and  $S_1$  (“0.25”),  $E_{adi}$  increases to 2.55 eV. For a tetramer,  $P_R$  varies between 4 (exciton delocalized over all 4 monomers) and 1 (exciton localized on one monomer). For “0.25,” the  $P_R$  value is 3.97. The  $P_R$  value differs from 4 because not all monomers are symmetrically equivalent. The comparison of the “1000,” “0001,” and “0.25” structures shows that for the  $T_1$  state of the DIP tetramer, the localized structures are about 0.1 eV lower in energy than the delocalized ones. The partially localized structures “0.5 0.5 0 0” ( $E_{adi} = 2.58 \text{ eV}$ ) and “0.5 0 0.5 0” ( $E_{adi} = 2.52 \text{ eV}$ ) are similar in energy to the delocalized structures “0.25.” For “0.5 0 0.5 0,” we compute a  $P_R$  value of 2.09, while 3.17 was computed for “0.5 0.5 0 0.” The difference arises because the exciton is delocalized over two neighbors for “0.5 0 0.5 0” but not for “0.5 0.5 0 0.” The structure “0.25” is slightly lower in energy than “0.5” ( $E_{adi} = 2.59 \text{ eV}$ ). If the distortion increases to “0.75” and “1,” the  $E_{adi}$  increases to 2.73 eV and 2.97 eV. In all cases, the  $P_R$  values are near to 4.

As for the PBI dimer, the  $T_2$  state of the DIP tetramer adopts equally distorted structures. The same holds for the  $T_3$  and  $T_4$  states. For the  $T_2$  state, the adiabatic excitation energies of the equally distorted structure “0.25” and of the partially unequally distorted structure “0.5 0.5 0 0” are 2.61 eV. However, the corresponding  $P_R$  values are only around 2, indicating that for “0.25,” the exciton is not completely delocalized as found for the corresponding structure of the  $T_1$  state. A closer inspection, indeed, showed that for the  $T_2$  state, the exciton is only delocalized over the monomers 3 and 4. For  $T_3$ , it is delocalized over the monomers 1 and 2. Such a localization



is possible because only 1 and 2 as well as 3 and 4 are symmetrically equivalent. For  $T_4$ , the exciton is again completely delocalized for “0.25.” Note that the unit cell of the DIP crystal contains two monomers, i.e., also for the crystal, not all monomers are identical. The  $P_R$  value computed for the “0.5 0.5 0 0” structure of the  $T_2$  is also lower than the corresponding value obtained for  $T_1$ . For  $T_3$  and  $T_4$ , the “0.25” structures represent the respective minima (2.67 eV and 2.85 eV). For the  $T_3$  state, this equally distorted structure lies about 0.2 eV below the “1000” or “0001” structures and still 0.1 eV below the structures in which only two monomers are distorted. Even the “0.5” structure is still lower than the unequally distorted structures. For the  $T_4$  state, energy variations are smaller. The “1000” structure is in the same energy range ( $E_{\text{adi}} = 2.88$  eV) as the equally distorted structure “0.25” (2.85 eV), while the partially unequally distorted structures “0.5 0.5 0 0” and “0.5 0 0.5 0” lie slightly higher in energy ( $E_{\text{adi}} = 2.93$ ).

For the  $T_1$  state of the DIP tetramer, the unsymmetrical distortion is clearly favored over the symmetrical one, while for the  $D_1$  state of the PBI dimer, both structures were isoenergetic. This can be explained by the couplings for the DIP tetramer, which are expected to be smaller because the molecular planes of the monomers are tilted with respect to each other. To check if differences in the electronic structure of both molecules also contribute, we performed test computations for DIP dimers in which both monomers lie on top of each other. Based on previous investigations of the minimum structures of such dimers, we choose  $R_L = 1.5$  Å and  $R_Z = 3.6$  Å.<sup>93</sup> For this dimer, the symmetrically distorted [ $\bar{R}_{0.5}$ ;  $\bar{R}_{0.5}$ ] structure ( $E_{\text{adi}} = 2.47$  eV) and unsymmetrically distorted [ $\bar{R}_{0.0}$ ;  $\bar{R}_{1.0}$ ] ( $E_{\text{adi}} = 2.46$  eV) are again virtually identical. The corresponding values for the PBI dimer with  $R_L = 1.4$  Å and  $R_Z = 3.5$  Å are 2.41 eV ([ $\bar{R}_{0.5}$ ;  $\bar{R}_{0.5}$ ]) and 2.35 eV ([ $\bar{R}_{1.0}$ ;  $\bar{R}_{0.0}$ ]), i.e., the variations are similar. Other structures like [ $\bar{R}_{0.0}$ ;  $\bar{R}_{0.0}$ ] and [ $\bar{R}_{1.0}$ ;  $\bar{R}_{1.0}$ ] are higher in energy for DIP ( $E_{\text{adi}} = 2.60$  eV and 2.55 eV) and for PBI ( $E_{\text{adi}} = 2.53$  eV and 2.45 eV). In both dimers, symmetrically distorted structures are favored with respect to unsymmetrical ones in the  $D_2$  state (DIP: [ $\bar{R}_{0.5}$ ;  $\bar{R}_{0.5}$ ]  $E_{\text{adi}} = 2.65$  eV; [ $\bar{R}_{0.0}$ ;  $\bar{R}_{1.0}$ ]  $E_{\text{adi}} = 2.75$  eV; PBI: [ $\bar{R}_{0.5}$ ;  $\bar{R}_{0.5}$ ]  $E_{\text{adi}} = 2.51$  eV; [ $\bar{R}_{0.0}$ ;  $\bar{R}_{1.0}$ ]  $E_{\text{adi}} = 2.69$  eV). This comparison indicates that the favored unsymmetrically distorted monomers for the DIP tetramer mainly result from the different orientations of monomers, while differences between the electronic structure of DIP and PBI are of minor importance.

Comparing the optimized geometries of the first excited states of monomer, dimer, and tetramer, respectively, shows that the distortion relative to the ground state structure decreases. Taking the factor of 1.0 for the difference found for the monomer, for PBI, DIP, and the CO-dimer, the delocalized exciton leads to smaller distortions of 0.7, 0.65, and about 0.5. For the DIP tetramer, the structure in which each monomer is distorted by a factor of “0.25” is already lower in energy than the “0.5” geometry. This trend indicates that for the first excited states, the size of the photo-induced geometry distortions decreases with an increase in delocalization of the exciton. Indeed, the trend is supported by test calculations for a cluster consisting of 14 tetracene. In this system, the vertical excitation energy (factor 0.0) is already lower than the excitation energy to the geometry in which all monomers are equally distorted by a factor of 1/14. Computations to investigate if this is a general trend are under way.

## CONCLUSIONS

In the first part of the present publication, we show that DFT-based descriptions of vibrationally resolved absorption spectra of perylene-based organic semiconductor monomers such as PBI, DIP, and PDIR-CN<sub>2</sub> are very accurate. The computations show that photo-induced geometrical relaxation effects are quite important and have to be included for an accurate description of the spectra. The ZPE also contributes but is of minor influence. For solvent effects, the non-equilibrium option is important. The most intensive band mainly represents the 0–0 transition. Small errors in the ratio of the intensities of the various bands indicate that the photo-induced relaxation effects are slightly overestimated. The Franck-Condon and harmonic approximations are essentially appropriate, but the inclusion of temperature (hot bands) is in some case significant.

The size of the photo-induced relaxation effects found for monomers suggested that the inclusion of intra-monomer relaxation effects is also important for the description of spectra of aggregates. Our investigations of the PBI dimer, indeed, underline their importance. The  $D_1$  state shows the expected double well shape for larger distances for which the coupling is weaker than the intra-monomer relaxation. Therefore, the exciton localizes on one monomer. However, even in this situation, the  $D_2$  PES exhibits the harmonic form, i.e., the exciton remains localized. The strong dependence on the state under consideration results because the wavefunctions of both states have to be orthogonal to each other for each geometry. For the optimal distance between both monomers ( $R_Z = 3.31$  Å), the coupling is sufficiently strong that the excitons remain mainly delocalized for the  $D_1$  as well as for  $D_2$  states, but the differences in the shapes of both PES are obvious. Additionally, the optimal longitudinal shifts ( $R_L$ ) of both states also differ. The results can be modeled by the vibronic coupling, the Marcus theory or by pJT-effect. All models represent a two-state system in which the eigenvalues are determined by an interplay between energy difference of the diagonal elements and size of the outer-diagonal elements. The pJT is closer to the physical nature of the effect. In the delocalized situation for the dimer, both monomers adopt the same structures, which lie somewhere between the  $S_0$  and  $S_1$  structures of the monomer. The relaxation to a localized exciton is then induced by a force, which drives one monomer toward the  $S_1$  geometry and the other toward the  $S_0$  geometry. Obviously, this force leads to the antisymmetric motion, which is also obtained from symmetry considerations. The Marcus theory also allows for a description of these effects; however, the underlying forces remain more abstract as only one coordinate is considered.

For the DIP tetramer, the coupling strength is smaller so that for the lowest electronically excited state  $T_1$ , the localized structure is the global minimum. In this structure, one monomer adopts the equilibrium geometry of the  $S_1$  state of the monomer, while all other monomers remain undisturbed. Structures in which all monomers are equally distorted are more stable for the  $T_2$ – $T_4$  states. However, even for these states, the amount of delocalization varies. For  $T_2$  and  $T_3$ , the exciton is only delocalized over two monomers, while it is completely delocalized for  $T_4$ . This finding emphasizes that intra-monomer relaxation effects strongly depend not only on the ratio of intra-monomer relaxation and coupling but also on the electronic state under consideration. Note that the photo-induced



geometry distortions decrease if the exciton is delocalized over more and more monomers. For dimers, the distortion is smaller by a factor of 0.5–0.7 than the distortion found for the monomer. The distortions shrink to about  $\frac{1}{4}$  if the exciton is delocalized over a tetramer.

#### AUTHORS' CONTRIBUTIONS

M.D., S.W., and D.K. performed all computations and contributed equally to this work. B.E. supervised the calculations. P.T., R.F.F., and B.E. wrote this manuscript. B.E., S.W., and P.T. thank the DFG (B.E. and S.W. in the framework of the GRK2112, P.T. for the project TE479/6-1) and D.K. thanks the Fonds der chemischen Industrie for funding.

#### SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for further information.

#### DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

#### REFERENCES

- F. Würthner, C. R. Saha-Möller, B. Fimmel, S. Ogi, P. Leowanawat, and D. Schmidt, "Perylene bisimide dye assemblies as archetype functional supramolecular materials," *Chem. Rev.* **116**, 962–1052 (2016).
- H. E. Katz, A. J. Lovinger, J. Johnson, C. Kloc, T. Siegrist, W. Li, Y.-Y. Lin, and A. Dodabalapur, "A soluble and air-stable organic semiconductor with high electron mobility," *Nature* **404**, 478–481 (2000).
- S. E. Shaheen, G. E. Jabbour, B. Kippelen, N. Peyghambarian, J. D. Anderson, S. R. Marder, N. R. Armstrong, E. Bellmann, and R. H. Grubbs, "Organic light-emitting diode with 20 lm/W efficiency using a triphenyldiamine side-group polymer as the hole transport layer," *Appl. Phys. Lett.* **74**, 3212–3214 (1999).
- C. W. Tang, "2-layer organic photovoltaic cell," *Appl. Phys. Lett.* **48**, 183–185 (1986).
- M. A. Green and S. P. Bremner, "Energy conversion approaches and materials for high-efficiency photovoltaics," *Nat. Mater.* **16**, 23–34 (2017).
- P. E. Hartnett, A. Timalisina, H. S. S. R. Matte, N. Zhou, X. Guo, W. Zhao, A. Facchetti, R. P. H. Chang, M. C. Hersam, M. R. Wasielewski, and T. J. Marks, "Slip-stacked perylene diimides as an alternative strategy for high efficiency non-fullerene acceptors in organic photovoltaics," *J. Am. Chem. Soc.* **136**, 16345–16356 (2014).
- X. W. Zhan, A. Facchetti, S. Barlow, T. J. Marks, M. A. Ratner, M. R. Wasielewski, and S. R. Marder, "Rylene and related diimides for organic electronics," *Adv. Mater.* **23**, 268–284 (2011).
- C. Huang, S. Barlow, and S. R. Marder, "Perylene-3,4,9,10-tetracarboxylic acid diimides: Synthesis, physical properties, and use in organic electronics," *J. Org. Chem.* **76**, 2386–2407 (2011).
- H. Langhals, S. Demmig, and T. Potrawa, "The relation between packing effects and solid-state fluorescence of dyes," *J. Prakt. Chem.* **333**, 733–748 (1991).
- M. C. R. Delgado, E.-G. Kim, D. A. da Silva Filho, and J.-L. Bredas, "Tuning the charge-transport parameters of perylene diimide single crystals via end and/or core functionalization: A density functional theory investigation," *J. Am. Chem. Soc.* **132**, 3375–3387 (2010).
- S. Shoaee, F. Deledalle, P. S. Tuladhar, R. Shivanna, S. Rajaram, K. S. Narayan, and J. R. Durrant, "A comparison of charge separation dynamics in organic blend films employing fullerene and perylene diimide electron acceptors," *J. Phys. Chem. Lett.* **6**, 201–205 (2015).
- S. Rajaram, R. Shivanna, S. K. Kandappa, and K. S. Narayan, "Nonplanar perylene diimides as potential alternatives to fullerenes in organic solar cells," *J. Phys. Chem. Lett.* **3**, 2405–2408 (2012).
- R. F. Fink, J. Seibt, V. Engel, M. Renz, M. Kaupp, S. Lochbrunner, H.-M. Zhao, J. Pfister, F. Würthner, and B. Engels, "Exciton trapping in  $\pi$ -conjugated materials: A quantum-chemistry-based protocol applied to perylene bisimide dye aggregates," *J. Am. Chem. Soc.* **130**, 12858–12859 (2008).
- I. A. Howard, F. Laquai, P. E. Keivanidis, R. H. Friend, and N. C. Greenham, "Perylene tetracarboxydiimide as an electron acceptor in organic solar cells: A study of charge generation and recombination," *J. Phys. Chem. C* **113**, 21225–21232 (2009).
- B. A. West, J. M. Womick, L. E. McNeil, K. J. Tan, and A. M. Moran, "Influence of vibronic coupling on band structure and exciton self-trapping in  $\alpha$ -perylene," *J. Phys. Chem. B* **115**, 5157–5167 (2011).
- J. Hoche, H.-C. Schmitt, A. Humeniuk, I. Fischer, R. Mitrić, and M. I. S. Röhr, "The mechanism of excimer formation: An experimental and theoretical study on the pyrene dimer," *Phys. Chem. Chem. Phys.* **19**, 25002–25015 (2017).
- E. R. Kennehan, C. Grieco, A. N. Brigeman, G. S. Doucette, A. Rimshaw, K. Bisgaier, N. C. Giebink, and J. B. Asbury, "Using molecular vibrations to probe exciton delocalization in films of perylene diimides with ultrafast mid-IR spectroscopy," *Phys. Chem. Chem. Phys.* **19**, 24829–24839 (2017).
- C. Brückner, F. Würthner, K. Meerholz, and B. Engels, "Structure-property relationships from atomistic multiscale simulations of the relevant processes in organic solar cells. I. Thermodynamic aspects," *J. Phys. Chem. C* **121**, 4–25 (2017).
- C. Brückner, F. Würthner, K. Meerholz, and B. Engels, "Atomistic approach to simulate processes relevant for the efficiencies of organic solar cells as a function of molecular properties. II. Kinetic aspects," *J. Phys. Chem. C* **121**, 26–51 (2017).
- H. Bässler, "Charge transport in disordered organic photoconductors—A Monte-Carlo simulation study," *Phys. Status Solidi B* **175**, 15–56 (1993).
- D. Hertel and H. Bässler, "Photoconduction in amorphous organic solids," *ChemPhysChem* **9**, 666–688 (2008).
- C. Brückner, M. Stolte, F. Würthner, J. Pflaum, and B. Engels, "QM/MM calculations combined with the dimer approach on the static disorder at organic-organic interfaces of thin-film organic solar cells composed of small molecules," *J. Phys. Org. Chem.* **30**, e3740 (2017).
- T. M. Clarke and J. R. Durrant, "Charge photogeneration in organic solar cells," *Chem. Rev.* **110**, 6736–6767 (2010).
- W. L. Liu, S. Canola, A. Köhn, B. Engels, F. Negri, and R. F. Fink, "A model Hamiltonian tuned toward high level *ab initio* calculations to describe the character of excitonic states in perylenebisimide aggregates," *J. Comput. Chem.* **39**, 1979–1989 (2018).
- W. L. Liu, B. Lunkenheimer, V. Settels, B. Engels, R. F. Fink, and A. Köhn, "A general ansatz for constructing quasi-diabatic states in electronically excited aggregated systems," *J. Chem. Phys.* **143**, 084106 (2015).
- N. J. Hestand, R. V. Kazantsev, A. S. Weingarten, L. C. Palmer, S. I. Stupp, and F. C. Spano, "Extended-charge-transfer excitons in crystalline supramolecular photocatalytic scaffolds," *J. Am. Chem. Soc.* **138**, 11762–11774 (2016).
- H.-M. Zhao, J. Pfister, V. Settels, M. Renz, M. Kaupp, V. C. Dehm, F. Würthner, R. F. Fink, and B. Engels, "Understanding ground- and excited-state properties of perylene tetracarboxylic acid bisimide crystals by means of quantum chemical computations," *J. Am. Chem. Soc.* **131**, 15660–15668 (2009).
- V. Settels, W. Liu, J. Pflaum, R. F. Fink, and B. Engels, "Comparison of the electronic structure of different perylene-based dye-aggregates," *J. Comput. Chem.* **33**, 1544–1553 (2012).
- V. Settels, A. Schubert, M. Tafipolski, W. Liu, V. Stehr, A. K. Topczak, J. Pflaum, C. Deibel, R. F. Fink, V. Engel, and B. Engels, "Identification of ultrafast relaxation processes as a major reason for inefficient exciton diffusion in perylene-based organic semiconductors," *J. Am. Chem. Soc.* **136**, 9327–9337 (2014).
- A. Schubert, M. Falge, M. Kess, V. Settels, S. Lochbrunner, W. T. Strunz, F. Würthner, B. Engels, and V. Engel, "Theoretical analysis of the relaxation

dynamics in perylene bisimide dimers excited by femtosecond laser pulses," *J. Phys. Chem. A* **118**, 1403–1412 (2014).

<sup>31</sup>A. Schubert, V. Settels, W. L. Liu, F. Würthner, C. Meier, R. F. Fink, S. Schindlbeck, S. Lochbrunner, B. Engels, and V. Engel, "Ultrafast exciton self-trapping upon geometry deformation in perylene-based molecular aggregates," *J. Phys. Chem. Lett.* **4**, 792–796 (2013).

<sup>32</sup>B. Engels and V. Engel, "The dimer-approach to characterize opto-electronic properties of an exciton trapping and diffusion in organic semiconductor aggregates and crystals," *Phys. Chem. Chem. Phys.* **19**, 12604–12619 (2017).

<sup>33</sup>Z. J. Chen, V. Stepanenko, V. Dehm, P. Prins, L. D. A. Siebbeles, J. Seibt, P. Marquetand, V. Engel, and F. Würthner, "Photoluminescence and conductivity of self-assembled  $\pi$ - $\pi$  stacks of perylene bisimide dyes," *Chem.-Eur. J.* **13**, 436–449 (2007).

<sup>34</sup>S. Samanta, S. K. Ray, S. Deolka, S. Saha, K. R. Pradeep, R. Bhowal, N. Ghosh, and D. Chaudhuri, "Safeguarding long-lived excitons from excimer traps in H-aggregated dye-assemblies," *Chem. Sci.* **11**, 5710–5715 (2020).

<sup>35</sup>W. Popp, M. Polkehn, R. Binder, and I. Burghardt, "Coherent charge transfer exciton formation in regioregular P3HT: A quantum dynamical study," *J. Phys. Chem. Lett.* **10**, 3326–3332 (2019).

<sup>36</sup>R. Binder and I. Burghardt, "First-principles quantum simulations of exciton diffusion on a minimal oligothiophene chain at finite temperature," *Faraday Discuss.* **221**, 406–427 (2020).

<sup>37</sup>A. Segalina, X. Assfeld, A. Monari, and M. Pastore, "Computational modeling of exciton localization in self-assembled perylene helices: Effects of thermal motion and aggregate size," *J. Phys. Chem. C* **123**, 6427–6437 (2019).

<sup>38</sup>M. R. Talipov, M. V. Ivanov, and R. Rathore, "Inclusion of asymptotic dependence of reorganization energy in the modified Marcus-based multistate model accurately predicts hole distribution in poly-*p*-phenylene wires (vol 120, pg 6402, 2016)," *J. Phys. Chem. C* **120**, 10720 (2016).

<sup>39</sup>R. L. Fulton and M. Gouterman, "Vibronic coupling. 1. Mathematical treatment for 2 electronic states," *J. Chem. Phys.* **35**, 1059 (1961).

<sup>40</sup>R. L. Fulton and M. Gouterman, "Vibronic coupling. 2. Spectra of dimers," *J. Chem. Phys.* **41**, 2280 (1964).

<sup>41</sup>F. P. Diehl, C. Roos, A. Duymaz, B. Lunkenheimer, A. Köhn, and T. Basché, "Emergence of coherence through variation of intermolecular distances in a series of molecular dimers," *J. Phys. Chem. Lett.* **5**, 262–269 (2014).

<sup>42</sup>P. H. P. Harbach and A. Dreuw, "A fresh look at excitonically coupled chromophores from a Jahn-Teller perspective," *Chem. Phys.* **377**, 78–85 (2010).

<sup>43</sup>P. Garcia-Fernandez, L. Andjelkovic, M. Zlatar, M. Gruden-Pavlovic, and A. Dreuw, "A simple monomer-based model-Hamiltonian approach to combine excitonic coupling and Jahn-Teller theory," *J. Chem. Phys.* **139**, 174101 (2013).

<sup>44</sup>G. Klebe, F. Graser, E. Hädicke, and J. Berndt, "Crystallochromy as a solid-state effect—Correlation of molecular-conformation, crystal packing and color in perylene-3,4,9,10-bis(dicarboximide) pigments," *Acta Crystallogr., Sect. B: Struct. Sci.* **45**, 69–77 (1989).

<sup>45</sup>M. Wewer and F. Stienkemeier, "Laser-induced fluorescence spectroscopy of *N,N'*-dimethyl 3,4,9,10-perylene tetracarboxylic diimide monomers and oligomers attached to helium nanodroplets," *Phys. Chem. Chem. Phys.* **7**, 1171–1175 (2005).

<sup>46</sup>B. Fimmel, M. Son, Y. M. Sung, M. Grüne, B. Engels, D. Kim, and F. Würthner, "Phenylene ethynylene-tethered perylene bisimide folda-dimer and folda-trimer: Investigations on folding features in ground and excited states," *Chem.-Eur. J.* **21**, 615–630 (2015).

<sup>47</sup>M. Peric, B. Engels, and S. D. Peyerimhoff, "Ab initio investigation of the vibronic structure of the C<sub>2</sub>H spectrum—Calculation of the hyperfine coupling-constants for the 3 lowest-lying electronic states," *J. Mol. Spectrosc.* **150**, 56–69 (1991).

<sup>48</sup>M. Peric, B. Engels, and S. D. Peyerimhoff, "Ab initio investigation of the vibronic structure of the C<sub>2</sub>H spectrum—Computation of the vibrationally averaged values for the hyperfine coupling-constants," *J. Mol. Spectrosc.* **150**, 70–85 (1991).

<sup>49</sup>M. R. Silva-Junior, M. Schreiber, S. P. A. Sauer, and W. Thiel, "Benchmarks of electronically excited states: Basis set effects on CASPT2 results," *J. Chem. Phys.* **133**, 174318 (2010).

<sup>50</sup>M. R. Silva-Junior, M. Schreiber, S. P. A. Sauer, and W. Thiel, "Benchmarks for electronically excited states: Time-dependent density functional theory and density functional theory based multireference configuration interaction," *J. Chem. Phys.* **129**, 104103 (2008).

<sup>51</sup>M. Schmittl, J.-P. Steffen, M. Maywald, B. Engels, H. Helten, and P. Musch, "Ring size effects in the C<sup>2</sup>-C<sup>6</sup> biradical cyclisation of enyne-allenes and the relevance for neocarzinostatin," *J. Chem. Soc., Perkin Trans. 2* **2**, 1331–1339 (2001).

<sup>52</sup>V. Pless, H. U. Suter, and B. Engels, "Ab initio study of the energy difference between the cyclic and linear-forms of the C<sub>6</sub> molecule," *J. Chem. Phys.* **101**, 4042–4048 (1994).

<sup>53</sup>H. U. Suter, V. Pless, M. Ernzerhof, and B. Engels, "Difficulties in the calculation of electron-spin-resonance parameters using density-functional methods," *Chem. Phys. Lett.* **230**, 398–404 (1994).

<sup>54</sup>A. Hellweg, S. A. Grün, and C. Hättig, "Benchmarking the performance of spin-component scaled CC2 in ground and electronically excited states," *Phys. Chem. Chem. Phys.* **10**, 4119–4127 (2008).

<sup>55</sup>O. Christiansen, H. Koch, and P. Jørgensen, "The second-order approximate coupled-cluster singles and doubles model CC2," *Chem. Phys. Lett.* **243**, 409–418 (1995).

<sup>56</sup>A. Schäfer, C. Huber, and R. Ahlrichs, "Fully optimized contracted Gaussian-basis sets of triple zeta valence quality for atoms Li to Kr," *J. Chem. Phys.* **100**, 5829–5835 (1994).

<sup>57</sup>F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, "RI-MP2: Optimized auxiliary basis sets and demonstration of efficiency," *Chem. Phys. Lett.* **294**, 143–152 (1998).

<sup>58</sup>K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, "Auxiliary basis-sets to approximate Coulomb potentials," *Chem. Phys. Lett.* **240**, 283–289 (1995).

<sup>59</sup>N. O. C. Winter, N. K. Graf, S. Leutwyler, and C. Hättig, "Benchmarks for 0–0 transitions of aromatic organic molecules: DFT/B3LYP, ADC(2), CC2, SOS-CC2 and SCS-CC2 compared to high-resolution gas-phase data," *Phys. Chem. Chem. Phys.* **15**, 6623–6630 (2013).

<sup>60</sup>O. Oeltermann, C. Brand, B. Engels, J. Tatchen, and M. Schmitt, "The structure of 5-cyanoindole in the ground and the lowest electronically excited singlet states, deduced from rotationally resolved electronic spectroscopy and ab initio theory," *Phys. Chem. Chem. Phys.* **14**, 10266–10270 (2012).

<sup>61</sup>J.-D. Chai and M. Head-Gordon, "Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections," *Phys. Chem. Chem. Phys.* **10**, 6615–6620 (2008).

<sup>62</sup>T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *J. Chem. Phys.* **90**, 1007–1023 (1989).

<sup>63</sup>E. R. Davidson, "Comment on Dunning's correlation-consistent basis sets—Comment," *Chem. Phys. Lett.* **260**, 514–518 (1996).

<sup>64</sup>D. Bellinger, J. Pflaum, C. Brüning, V. Engel, and B. Engels, "The electronic character of PTCDAs thin films in comparison to other perylene-based organic semi-conductors: Ab initio-, TD-DFT and semi-empirical computations of the opto-electronic properties of large aggregates," *Phys. Chem. Chem. Phys.* **19**, 2434–2448 (2017).

<sup>65</sup>S. Wirsing, M. Hänsel, V. Belova, F. Schreiber, K. Broch, B. Engels, and P. Tegeder, "Excited-state dynamics in perylene-based organic semiconductor thin films: Theory meets experiment," *J. Phys. Chem. C* **123**, 27561–27572 (2019).

<sup>66</sup>M. J. Frisch, Gaussian 16, Revision A.01, Gaussian, Inc., Wallingford, CT, 2016.

<sup>67</sup>J. Bloino, M. Biczysko, F. Santoro, and V. Barone, "General approach to compute vibrationally resolved one-photon electronic spectra," *J. Chem. Theory Comput.* **6**, 1256–1274 (2010).

<sup>68</sup>A. Baiardi, J. Bloino, and V. Barone, "General time dependent approach to vibronic spectroscopy including Franck-Condon, Herzberg-Teller, and Duschinsky effects," *J. Chem. Theory Comput.* **9**, 4097–4115 (2013).

<sup>69</sup>S. Miertus, E. Scrocco, and J. Tomasi, "Electrostatic interaction of a solute with a continuum—A direct utilization of ab initio molecular potentials for the prevision of solvent effects," *Chem. Phys.* **55**, 117–129 (1981).

<sup>70</sup>S. Miertus and J. Tomasi, "Approximate evaluations of the electrostatic free-energy and internal energy changes in solution processes," *Chem. Phys.* **65**, 239–245 (1982).

- <sup>71</sup>J. L. Pascual-Ahuir, E. Silla, and I. Tuñón, "GEPOL—An improved description of molecular-surfaces. 3. A new algorithm for the computation of a solvent-excluding surface," *J. Comput. Chem.* **15**, 1127–1138 (1994).
- <sup>72</sup>B. Mennucci, R. Cammi, and J. Tomasi, "Excited states and solvatochromic shifts within a nonequilibrium solvation approach: A new formulation of the integral equation formalism method at the self-consistent field, configuration interaction, and multiconfiguration self-consistent field level," *J. Chem. Phys.* **109**, 2798–2807 (1998).
- <sup>73</sup>J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," *Chem. Rev.* **105**, 2999–3093 (2005).
- <sup>74</sup>C. Brückner and B. Engels, "Benchmarking singlet and triplet excitation energies of molecular semiconductors for singlet fission: Tuning the amount of HF exchange and adjusting local correlation to obtain accurate functionals for singlet-triplet gaps," *Chem. Phys.* **482**, 319–338 (2017).
- <sup>75</sup>F. Plasser, M. Wormit, and A. Dreuw, "New tools for the systematic analysis and visualization of electronic excitations. I. Formalism," *J. Chem. Phys.* **141**, 024106 (2014).
- <sup>76</sup>F. Plasser, S. A. Bappler, M. Wormit, and A. Dreuw, "New tools for the systematic analysis and visualization of electronic excitations. II. Applications," *J. Chem. Phys.* **141**, 024107 (2014).
- <sup>77</sup>B. Engels and S. D. Peyerimhoff, "The hyperfine coupling-constants of the  $X^3\Sigma^-$  states of NH influence of polarization functions and configuration space on the description of spin polarization," *Mol. Phys.* **67**, 583–600 (1989).
- <sup>78</sup>B. A. Engels, "Detailed study of the configuration selected multireference configuration-interaction method combined with perturbation-theory to correct the wave-function," *J. Chem. Phys.* **100**, 1380–1386 (1994).
- <sup>79</sup>B. O. Roos, R. Lindh, P. A. Malmqvist, V. Veryazov, and P.-O. Widmark, *Multiconfigurational Quantum Chemistry* (John Wiley & Sons, Hoboken, NJ), 216).
- <sup>80</sup>C. Brueckner and B. Engels, "Benchmarking ground-state geometries and vertical excitation energies of a selection of P-type semiconducting molecules with different polarity," *J. Phys. Chem. A* **119**, 12876–12891 (2015).
- <sup>81</sup>C. Walter, V. Kramer, and B. Engels, "On the applicability of time-dependent density functional theory (TDDFT) and semiempirical methods to the computation of excited-state potential energy surfaces of perylene-based dye-aggregates," *Int. J. Quantum Chem.* **117**, e25337 (2017).
- <sup>82</sup>J. Auerswald, B. Engels, I. Fischer, T. Gerbich, J. Herterich, A. Krueger, M. Lang, H.-C. Schmitt, C. Schon, and C. Walter, "The electronic structure of pyracene: A spectroscopic and computational study," *Phys. Chem. Chem. Phys.* **15**, 8151–8161 (2013).
- <sup>83</sup>W. Liu, V. Settels, P. H. P. Harbach, A. Dreuw, R. F. Fink, and B. Engels, "Assessment of TD-DFT- and TD-HF-based approaches for the prediction of exciton coupling parameters, potential energy curves, and electronic characters of electronically excited aggregates," *J. Comput. Chem.* **32**, 1971–1981 (2011).
- <sup>84</sup>A. Fihey and D. Jacquemin, "Performances of density functional tight-binding methods for describing ground and excited state geometries of organic molecules," *J. Chem. Theory Comput.* **15**, 6267–6276 (2019).
- <sup>85</sup>C. Suellen, R. G. Freitas, P.-F. Loos, and D. Jacquemin, "Cross-comparisons between experiment, TD-DFT, CC, and ADC for transition energies," *J. Chem. Theory Comput.* **15**, 4581–4590 (2019).
- <sup>86</sup>U. Heinemeyer, R. Scholz, L. Gisslén, M. I. Alonso, J. O. Ossó, M. Garriga, A. Hinderhofer, M. Kytka, S. Kowarik, A. Gerlach, and F. Schreiber, "Exciton-phonon coupling in diindenoperylene thin films," *Phys. Rev. B* **78**, 085210 (2008).
- <sup>87</sup>E. M. Huber, G. de Bruin, W. Heinemeyer, G. Paniagua Soriano, H. S. Overkleef, and M. Groll, "Systematic analyses of substrate preferences of 205 proteasomes using peptidic epoxyketone inhibitors," *J. Am. Chem. Soc.* **137**, 7835–7842 (2015).
- <sup>88</sup>V. Stehr, R. F. Fink, B. Engels, J. Pflaum, and C. Deibel, "Singlet exciton diffusion in organic crystals based on Marcus transfer rates," *J. Chem. Theory Comput.* **10**, 1242–1255 (2014).
- <sup>89</sup>M. Malagoli, V. Coropceanu, D. A. da Silva Filho, and J. L. Brédas, "A multi-mode analysis of the gas-phase photoelectron spectra in oligoacenes," *J. Chem. Phys.* **120**, 7490–7496 (2004).
- <sup>90</sup>U. Müller, L. Roos, M. Frank, M. Deutsch, S. Hammer, M. Krumrein, A. Friedrich, T. B. Marder, B. Engels, A. Krueger, and J. Pflaum, "Role of intermolecular interactions in the excited-state photophysics of tetracene and 2,2'-ditetracene," *J. Phys. Chem. C* **124**, 19435 (2020).
- <sup>91</sup>S. Kowarik, A. Gerlach, S. Sellner, L. Cavalcanti, O. Kononov, and F. Schreiber, "Real-time X-ray diffraction measurements of structural dynamics and polymorphism in diindenoperylene growth," *Appl. Phys. A* **95**, 233–239 (2009).
- <sup>92</sup>M. A. Heinrich, J. Pflaum, A. K. Tripathi, W. Frey, M. L. Steigerwald, and T. Siegrist, "Enantiotropic polymorphism in di-indenoperylene," *J. Phys. Chem. C* **111**, 18878–18881 (2007).
- <sup>93</sup>C. Brückner, "The electronic structure and optoelectronic processes at the interfaces in organic solar cells composed of small organic molecules—A computational analysis of molecular, intermolecular, and aggregate aspects," Ph.D. thesis, Universität Würzburg, 2016.

## 3. Spectroscopic Modeling of Methylbismuth

Chemical  
Science

EDGE ARTICLE

View Article Online  
View Journal | View IssueCite this: *Chem. Sci.*, 2020, 11, 7562

All publication charges for this article have been paid for by the Royal Society of Chemistry

## Methylbismuth: an organometallic bismuthinidene biradical†

Deb Pratim Mukhopadhyay,<sup>a</sup> Domenik Schleier,<sup>a</sup> Sara Wirsing,<sup>a</sup> Jacqueline Ramler,<sup>b</sup> Dustin Kaiser,<sup>a</sup> Engelbert Reusch,<sup>a</sup> Patrick Hemberger,<sup>b,c</sup> Tobias Preitschopf,<sup>a</sup> Ivo Krummenacher,<sup>b</sup> Bernd Engels,<sup>\*a</sup> Ingo Fischer<sup>id,\*a</sup> and Crispin Lichtenberg<sup>id,\*b</sup>

We report the generation, spectroscopic characterization, and computational analysis of the first free (non stabilized) organometallic bismuthinidene, BiMe. The title compound was generated *in situ* from BiMe<sub>3</sub> by controlled homolytic Bi–C bond cleavage in the gas phase. Its electronic structure was characterized by a combination of photoion mass selected threshold photoelectron spectroscopy and DFT as well as multi reference computations. A triplet ground state was identified and an ionization energy (IE) of 7.88 eV was experimentally determined. Methyl abstraction from BiMe<sub>3</sub> to give [BiMe<sub>2</sub>]<sup>•</sup> is a key step in the generation of BiMe. We reveal a bond dissociation energy of 210 ± 7 kJ mol<sup>-1</sup>, which is substantially higher than the previously accepted value. Nevertheless, the homolytic cleavage of Me–BiMe<sub>2</sub> bonds could be achieved at moderate temperatures (60–120 °C) in the condensed phase, suggesting that [BiMe<sub>2</sub>]<sup>•</sup> and BiMe are accessible as reactive intermediates under these conditions.

Received 28th April 2020  
Accepted 2nd June 2020

DOI: 10.1039/d0sc02410d

rsc.li/chemical-science

## Introduction

Low valent molecular compounds of group 15 elements, E R, with the central atom E in the oxidation state of +1 are highly reactive, electron deficient species (E = N, Bi; R = monoanionic ligand).<sup>1</sup> In principle, they can adopt either singlet or triplet electronic ground states, which strongly influences their physical and chemical properties (Scheme 1). Thus, understanding their (electronic) structures is key to rationalizing the reactivity of these species, which represent important intermediates in fundamental transformations such as insertion and ring expansion reactions.<sup>2</sup> The isolation of such compounds, in which the group 15 atom is bound to only one additional atom, is extremely challenging and has only recently been achieved for the lightest congeners in landmark contributions (E = N, P, Scheme 1a).<sup>2,3</sup> Access to the heavier homologues is increasingly difficult due to their tendency to undergo degradation reactions such as oligomerization, disproportionation, and bond

activation processes. For the heaviest group 15 element bismuth, mononuclear organometallic compounds with the metal atom in the oxidation state of +1 (so called bismuthinidene)s could so far only be accessed by stabilization through adduct formation with Lewis bases (Scheme 1b).<sup>4–9</sup> Very recently, the intriguing properties of such bismuthinidene)s have been exploited for (electro)catalytic and photophysical applications: a Bi<sup>i</sup>/Bi<sup>III</sup> redox couple has been proposed to be a key component in bismuth catalyzed transfer hydrogenation reactions with ammonia borane.<sup>10</sup> In addition, Bi<sup>i</sup> oxalates have been suggested to be involved in the electrochemical reduction of carbon dioxide at bismuth electrodes.<sup>11</sup> When embedded in host materials such as zeolites, [Bi<sup>i</sup>]<sup>•</sup> centers show intriguing photophysical properties such as ultrabroad near infrared emission.<sup>12,13</sup> Without any exceptions, the isolable, Lewis base stabilized bismuthinidene)s reported to date show singlet ground states. This situation changes, when non stabilized, (so far) non isolable, electron deficient molecular bismuthinidene)s, Bi R, are targeted. To date, only inorganic species Bi X have been reported and were accessed by comproportionation (e.g. at high temperature in the melt) or by reaction of Bi atoms with reactive species such as F<sub>2</sub> in the gas phase (X = H, F, I, AlCl<sub>4</sub>).<sup>14–18</sup> Importantly, these synthetic routes did not allow to study the reactivity of bismuthinidene)s, they did not allow to work in conventional solvents at moderate temperature, and they did not allow to access compounds with tunable substituents R, such as organometallic bismuthinidene)s Bi R (R = alkyl, aryl; Fig. 1c).

<sup>a</sup>Institute of Physical and Theoretical Chemistry, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany. E-mail: bernd.engels@uni-wuerzburg.de; ingo.fischer@uni-wuerzburg.de

<sup>b</sup>Institute of Inorganic Chemistry, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany. E-mail: crispin.lichtenberg@uni-wuerzburg.de

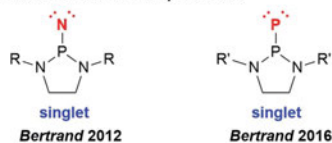
<sup>c</sup>Laboratory for Femtochemistry and Synchrotron Radiation, Paul Scherrer Institut (PSI), CH-5232 Villigen, Switzerland. E-mail: patrick.hemberger@psi.ch

† Electronic supplementary information (ESI) available. CCDC 1991253. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/d0sc02410d

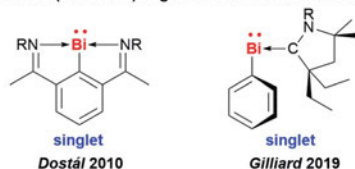


Edge Article

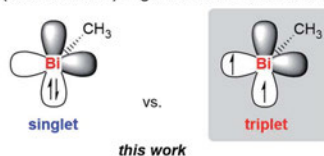
## a) Isolable Nitrene and Phosphinidene



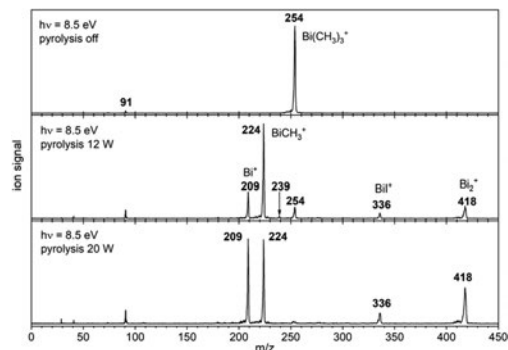
## b) Isolable (Stabilized) Organometallic Bismuthinidenes



## c) Free (Non-stabilized) Organometallic Bismuthinidene

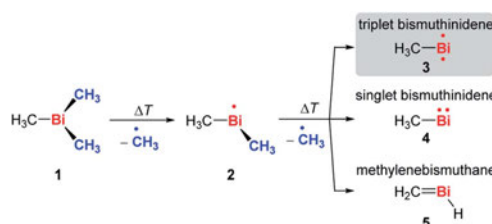


**Scheme 1** Low valent group 15 compounds: (a) isolable nitrene and phosphinidene (singlet ground state). (b) stabilized organometallic bismuthinidenes (singlet ground state; occupied bismuth 6s orbital omitted for clarity); (c) free (non-stabilized) singlet and triplet species that may be envisaged for BiMe (only two bismuth p orbitals shown; p orbital used for Bi-Me bonding and occupied s orbital are omitted for clarity). R = 2,6-Me<sub>2</sub>-C<sub>6</sub>H<sub>3</sub>; 2,6-iPr<sub>2</sub>-C<sub>6</sub>H<sub>3</sub>; R' = 2,6-[4-tBu-C<sub>6</sub>H<sub>4</sub>]<sub>2</sub>-CH<sub>2</sub>. 4-Me-C<sub>6</sub>H<sub>3</sub>.



**Fig. 1** Photoionization mass spectra recorded at 8.5 eV without pyrolysis (top trace), low pyrolysis power ( $T \approx 470$  K, center trace) and medium pyrolysis power ( $T \approx 600$  K, bottom trace). The signal at  $m/z$  91 is due to a background signal from previous experiments.

Here we report the generation and characterization of the first free (*i.e.* non-stabilized) organometallic bismuthinidene methyl bismuth, BiMe, in a gas phase reaction with implications for related reactions in condensed phase at moderate temperature. In

View Article Online  
Chemical Science

**Scheme 2** Controlled, stepwise abstraction of CH<sub>3</sub> radicals from 1 in the gas phase by flash pyrolysis.

addition, dimethylbismuth, [BiMe<sub>2</sub>]<sup>•</sup> was studied to determine the Bi-CH<sub>3</sub> bond dissociation energy in BiMe<sub>3</sub>.

## Results and discussion

In order to access the elusive class of non-stabilized organometallic bismuthinidenes, we aimed at a “top-down approach”, more specifically the homolytic cleavage of Bi-C bonds from the well-defined organometallic precursor BiMe<sub>3</sub> (1) as shown in Scheme 2. The long-known BiMe<sub>3</sub> has been investigated from many different perspectives,<sup>19–21</sup> including its application in chemical vapor deposition, where the abstraction of all its methyl groups is targeted.<sup>22,46</sup>

In a recent study, it was demonstrated that the abstraction of a single methyl group can be achieved by dissociative photoionization of 1, *i.e.* the Bi-C bond cleavage of the [BiMe<sub>3</sub>]<sup>•+</sup> cation in the gas phase.<sup>23</sup> Under these conditions, only a single methyl group was abstracted, yielding the cation [BiMe<sub>2</sub>]<sup>+</sup>. Here we attempt a thermally induced, controlled and stepwise abstraction of methyl groups from neutral BiMe<sub>3</sub>. Thus, a sample of BiMe<sub>3</sub> diluted in Ar was pyrolysed in a microreactor and analyzed by photoelectron photoion coincidence spectroscopy (PEPICO) using synchrotron radiation.<sup>24</sup> This method permits to record photoion mass selected threshold photoelectron spectra (ms TPE) for each species by correlating ions and electrons produced in a single photoionization event. Isomer-selective information is then obtained from an analysis of the photoelectron spectrum based on computations.

Fig. 1 shows mass spectra under various pyrolysis conditions. Without pyrolysis (top trace) only the parent ion 1<sup>•+</sup> is visible, thus dissociative photoionization is irrelevant under our experimental conditions. Already at low pyrolysis power (center trace) a stepwise methyl loss down to atomic Bi occurs, associated with formation of 2 and one of the products 3–5. The small intensity of  $m/z = 239$  ([Bi(CH<sub>3</sub>)<sub>2</sub>]<sup>+</sup>) compared to  $m/z = 224$  ([BiCH<sub>3</sub>]<sup>+</sup>) indicates that cleavage of the second methyl group is more facile than the first. In addition, Bi<sub>2</sub> is visible at  $m/z = 418$ , due to dimerization of bismuth atoms (see ESI, Fig. S4†). In some experiments a further peak appeared at  $m/z = 478$  and is most likely due to Me<sub>2</sub>Bi-BiMe<sub>2</sub>. Note that CH<sub>3</sub> is not observed due to its ionization energy of 9.83891 eV.<sup>25,26</sup> Traces of BiI from the synthesis are also present in the spectrum.

When the pyrolysis power is further increased (bottom trace) the precursor is fully converted and experimental conditions are suitable for studying the molecule formed after loss of two methyl groups. Three structures are possible for  $m/z = 224$ , the bismuthinidene **3** (triplet)/**4** (singlet) or the methyl enebismuthane **5**. Threshold photoelectron spectroscopy provides structural isomer selective information through comparison with Franck Condon simulated or reference spectra. Fig. 2 represents the ms TPE spectrum of a species with the composition  $\text{BiCH}_3$  at  $m/z = 224$ . The first major band at 7.88 eV is assigned to the adiabatic ionization energy ( $\text{IE}_{\text{ad}}$ ). It is followed by several smaller bands that are ca. 40–50 meV apart. Simulations based on DFT and multi reference calculations were carried out for **3**, **4** and **5**. While DFT often provides very accurate geometries and frequencies even for molecules with complicated electronic structures,<sup>48</sup> it is in many cases less accurate for the computation of energy surfaces or excitation energies.<sup>49</sup> Hence, the  $\omega\text{B97X-D3}$ <sup>50</sup> functional was employed for frequency computations, but the multi reference NEVPT2<sup>51</sup> approach was used to determine geometrical changes and to compute ionization energies as well as the energy difference between the two relevant states of the  $[\text{BiCH}_3]^+$  cation (*vide infra*). Both methods (DFT and multi reference) were combined with a scalar relativistic approach and with the SARC ZORA TZVP basis set which allows for an all electron treatment of bismuth<sup>52</sup> (for computational details, see ESI†). Bismuthinidene **3** with a  $C_{3v}$  symmetry and an  $X^3E(T_0)$  triplet ground state is the lowest energy structure ( $\Delta E = 0 \text{ kJ mol}^{-1}$ ). The computations on the NEVPT2 level show a very good agreement with the experimentally determined IE of 7.88 eV ( $\text{IE}_{\text{calc}} = 7.98 \text{ eV}$  for the  $X^+ 2A'' \leftarrow X^3E$  transition (*vide infra*), as compared to singlet bismuthinidene **4** ( $\Delta E = +0.78 \text{ eV}/+75 \text{ kJ mol}^{-1}$ ;  $\text{IE}_{\text{calc}} = 7.21 \text{ eV}$ ) and methylenebismuthane **5** ( $\Delta E = +0.91 \text{ eV}/+88 \text{ kJ mol}^{-1}$ ;  $\text{IE}_{\text{calc}} = 8.68 \text{ eV}$ ; for energy values obtained through DFT calculations see ESI†). This shows, that the ground state (electronic) structure of species with the sum formula  $\text{ECH}_3$  are fundamentally different, depending on the choice of the element E. For  $E = \text{Bi}$  the triplet bismuthinidene **3** is energetically favored and observed (*vide supra*), whereas for the lighter congeners ( $E = \text{N}, \text{P}$ ), the formation of the methylene species  $\text{HN}=\text{CH}_2$  and  $\text{HP}=\text{CH}_2$  has been determined to be more favorable.<sup>27, 32, 53</sup>

Upon photoionization of **3**, one electron is removed from either of the two degenerate SOMOs, which correspond in first approximation to the  $p_x$  and  $p_y$  orbital on the Bi center. The computations for the ionic ground state,  $[\text{BiCH}_3]^+$  ( $3^{**}$ ), yielded a shortening of the Bi–C bond (from 2.27 Å in **3** to 2.21 Å in  $3^{**}$ ) and a deformation of the methyl group with a tilt angle of 4° relative to the Bi–C axis. This was ascribed to antibonding interactions between the unpaired electron and the bonding electrons of the two C–H groups, which are approximately aligned with the singly occupied p type orbital of bismuth (Table S18†). This leads to a loss of the  $C_3$  axis and a reduction to  $C_s$  symmetry. As a consequence of this Jahn–Teller distortion, the  $^2E$  state in the  $C_{3v}$  symmetric cation splits into a  $X^+ 2A''$  and a  $A^+ 2A'$  component. The computations indicate an energy difference of only 50 meV between the two states at the geometry of the  $X^+ 2A''$ . Thus, transitions into both states contribute

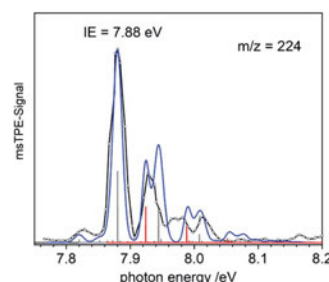


Fig. 2 Mass selected threshold photoelectron spectrum of  $m/z = 224$ . The simulation based on **3** as the carrier (blue line) fits the experiment well. Transitions into the  $X^+ 2A''$  ground state of the ion are given as grey bars, transitions into the  $A^+ 2A'$  state as red bars.

to the spectrum and have to be included in the simulation in Fig. 2 (blue curve). Transitions from the  $T_0$  ground state of neutral **3** into the  $X^+ 2A''$  ionic ground state are given as grey bars, while transitions into the  $A^+ 2A'$  excited state are shown as red bars. In addition, there is a vibrational structure evident in the spectrum with a spacing of around 50 meV, including a hot band transition at 7.83 eV, which is assigned to the Bi–C stretching motion (corresponding to  $\nu'' \approx 50 \text{ meV}$ ). Vibrational activity is expected due to the reduction of the Bi–C bond lengths in the cation by  $> 0.06 \text{ Å}$  (see ESI† for all geometry parameters). Franck Condon simulations based on computations of isomer **5** further support the triplet bismuthinidene **3** as the carrier of the spectrum (Fig. S3†). First, the computed IE of **5** is 0.80 eV higher than the experimental value, and second, a more pronounced vibrational progression with a maximum intensity for a transition into an excited vibrational state would be expected for **5**. Thus a contribution of **5** to the spectrum can be ruled out.

The computed IE of **3** (for the lowest state of  $3^{**}$ ) and the relative energies for the two states of the cation  $3^{**}$  are in excellent agreement with the corresponding experimental data. This indicates that the neglected spin–orbit effects<sup>33</sup> do not play a key role for the determination of the ionization energies, possibly because they are similar in magnitude for all involved states. While the experimental and calculated IEs nicely agree, deviations were observed in the shape of the spectra and were ascribed to the flatness of the potential energy surface (PES) of the cation. Two factors mainly contribute to the flatness of this PES: (i) the  $X^+ 2A''$  state is threefold degenerate due to facile rotation around the Bi–C bond. (ii) The shape of the PES going from the equilibrium geometry of the  $X^+ 2A''$  towards the equilibrium geometry of the  $A^+ 2A'$  state is expected to be non-harmonic. Efforts to obtain a better description of the surface were so far unsuccessful due to strong correlations of the various internal coordinates, so that high dimensional surfaces would be necessary for an appropriate description.

To gain additional information on the formation of BiMe (**3**) by stepwise abstraction of methyl groups from  $\text{BiMe}_3$  (**1**), an ms TPE spectrum of  $[\text{BiMe}_2]^+$  (**2**) ( $m/z(2^+) = 239$ ) was recorded at

## Edge Article

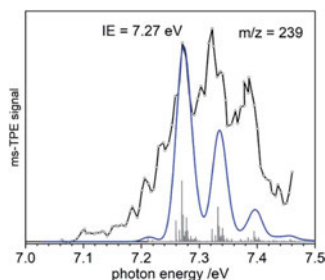


Fig. 3 Ms TPE spectrum of  $\text{Bi}(\text{CH}_3)_2$ ,  $m/z = 239$ . The vibrational progression is due to the symmetric bismuth-carbon stretch and its combination with torsional motion. The blue line represents the simulated spectrum based on 2.

a low pyrolysis temperature (Fig. 3; cf. Fig. 1). Simulations of the spectrum based on DFT calculations indicate  $C_{2v}$  symmetry for both  $2$  and  $2^+$  as well as a  $X^1A_1 \leftarrow X^2B_1$  transition in  $[\text{BiMe}_2]^+$  ( $2$ ). The first major band at 7.27 eV is assigned to the IE, in excellent agreement with the computed value of 7.35 eV at the NEVTP2 level of theory. A vibrational progression with a spacing of 60 meV is visible and is dominated by the symmetric Bi-C stretching mode in the cation. Additional torsional modes of the  $\text{CH}_3$  groups may lead to a broadening of the bands. While there is a good agreement between simulation and experiment, the vibrational intensities (including hot bands) are somewhat underestimated in the simulations. This indicates a slightly larger change of the Bi-C bond length upon ionization than the computed shortening of 0.05 Å (see ESI† for all geometry parameters).

Bond dissociation energies (BDEs) can be determined *via* thermochemical cycles that combine appearance energies (AE) and ionization energies. The zero Kelvin appearance energy  $\text{AE}_{0\text{K}}$  for the abstraction of the first methyl radical from  $1$  has been determined with very high accuracy,  $\text{AE}_{0\text{K}}(\text{Bi}(\text{CH}_3)_3, \text{Bi}(\text{CH}_3)_2^+) = 9.445 \pm 0.064$  eV.<sup>23</sup> According to computations, the methyl loss in the cation is a simple homolytic bond cleavage without a reverse barrier. Combined with the IE of  $2$ , the  $\text{Me}_2\text{Bi}-\text{CH}_3$  bond dissociation energy in  $1$ ,  $\text{BDE}(\text{Me}_2\text{Bi}-\text{CH}_3)$ , can be derived:

$$\text{BDE}(\text{Me}_2\text{Bi}-\text{CH}_3) = \text{AE}_{0\text{K}}(\text{BiMe}_3, [\text{BiMe}_2]^+) - \text{IE}([\text{BiMe}_2]^+)(1)$$

From eqn (1) a value of  $210 \pm 7$  kJ mol<sup>-1</sup> is obtained for  $\text{BDE}(\text{Me}_2\text{Bi}-\text{CH}_3)$ . The bond dissociation energy of the first Bi- $\text{CH}_3$  bond in  $\text{BiMe}_3$  ( $\text{BDE}(\text{Me}_2\text{Bi}-\text{CH}_3)$ ) can be expected to be the highest of the three Bi- $\text{CH}_3$  BDEs in this molecule<sup>34</sup> and is thus crucial for any type of reaction initiation *via* Bi- $\text{CH}_3$  homolysis. However, the value of  $\text{BDE}(\text{Me}_2\text{Bi}-\text{CH}_3)$  has never been determined explicitly in the primary literature. Based on previous investigations into the thermal decomposition of  $1$ ,<sup>34,36</sup> an estimation of  $\text{BDE}(\text{Me}_2\text{Bi}-\text{CH}_3)$  can be made, which yields a value of 182 kJ mol<sup>-1</sup> as the upper limit that would be possible for this parameter.<sup>37</sup> Thus, our results substantially revise the bond dissociation energy of the first Bi-Me bond in  $1$ , which is

View Article Online  
Chemical Science

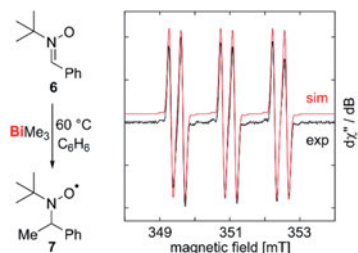


Fig. 4 Abstraction and spin trapping of methyl radicals from  $\text{BiMe}_3$ .

key to the radical chemistry of  $1$  and related bismuth compounds. The calculated isodesmic reaction provides a value of 226 kJ mol<sup>-1</sup> in good agreement with the experimental findings.

Our correction of the BDE in a fundamentally important organometallic compound such as  $\text{BiMe}_3$  raised the question, whether homolytic Bi-C bond cleavage is possible at moderate reaction temperatures in the condensed phase, making this process relevant for synthetic chemistry under conventional experimental conditions. To test for methyl radical abstraction from  $\text{BiMe}_3$ , a benzene solution of  $\text{BiMe}_3$  and the radical trap  $6$  was heated to 60 °C and subsequently analyzed by EPR spectroscopy (Fig. 4). Indeed, a resonance was detected with  $g_{\text{iso}} = 2.006$  and  $a(^{14}\text{N}) = 41.6$  MHz,  $a(^1\text{H}) = 9.63$  MHz,  $a(^{13}\text{C}) = 12.9$  MHz, indicating the formation of  $7$  by methyl radical transfer.<sup>38,54,55</sup>

In order to gain further hints at the generation and subsequent trapping of  $\text{BiMe}$  in the condensed phase, neat  $\text{BiMe}_3$  was

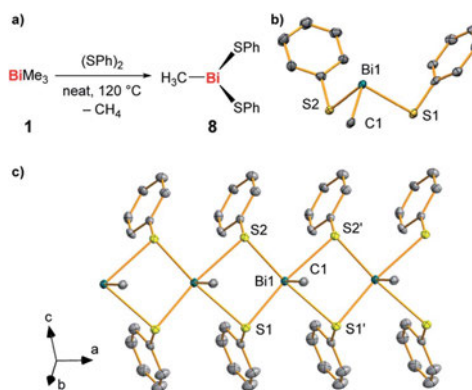


Fig. 5 (a) Synthesis of  $\text{BiMe}(\text{SPh})_2$  ( $8$ ) from  $\text{BiMe}_3$  and  $(\text{SPh})_2$  with methane as a detected by-product. (b and c) Molecular structure of  $8$  in the solid state with one formula unit shown in (b) and a cutout of the coordination polymer shown in (c). Displacement ellipsoids are shown at the 50% probability level. Hydrogen atoms and one set of split positions of disordered atoms are omitted for clarity. For detailed discussion of structural parameters see ESI.† Selected bond lengths (Å) and angles (°): Bi1-C1, 2.208(10); Bi1-S1, 2.736(2); Bi1-S2, 2.699(2); C1-Bi1-S1, 89.6(3); C1-Bi1-S2, 89.4(3); S1-Bi1-S2, 93.18(7).

## Chemical Science

View Article Online

Edge Article

reacted with stoichiometric amounts of (PhS)<sub>2</sub> at 120 °C (Fig. 5a). BiMe(SPh)<sub>2</sub> (**8**)<sup>39, 41</sup> was isolated as the product of this reaction in 41% yield and fully characterized (Fig. 5b and c; for details see ESI†). Methane was detected in the headspace of the reaction by IR spectroscopy (Fig. S5†), suggesting the appearance of methyl radicals in this reaction.

This together with literature reports<sup>42</sup> supports the potential of BiMe to act as a transient reactive species in this reaction. The thioether MeSPh was also detected, suggesting that methyl radical attack at sulfur or  $\sigma$  bond metathesis/disproportionation sequences may also be operative as a parallel reaction pathway.<sup>41</sup> In the solid state, compound **8** forms a coordination polymer through bridging coordination modes of the thiolate ligands (Fig. 5c). This is a unique structural feature within the small number of literature known compounds BiR(SR')<sub>2</sub> (R, R' = aryl, alkyl; for details see ESI†).<sup>42, 45</sup>

## Conclusions

In conclusion, we have generated methylbismuth (BiMe), a fundamental organometallic compound and the first example of an organometallic non stabilized bismuthinidene. BiMe was accessed *via* controlled thermal homolysis in the gas phase. The title compounds shows a triplet (biradical) ground state and an ionization energy of 7.88 eV, as revealed by combination of photoelectron spectroscopy and computations. The homolytic dissociation of the first Me<sub>2</sub>Bi-CH<sub>3</sub> bond in BiMe<sub>3</sub> is crucial to the radical chemistry of this compound (and related species) and is the initiating step in the formation of BiMe. Our results reveal a Me<sub>2</sub>Bi-CH<sub>3</sub> homolytic bond dissociation energy of 210 ± 7 kJ mol<sup>-1</sup>, *i.e.* the previously reported value is revised by more than +15% (+28 kJ mol<sup>-1</sup>). Nevertheless, reactions in the condensed phase demonstrate that the abstraction of methyl radicals from BiMe<sub>3</sub> is possible at moderate reaction conditions and suggest that BiMe may act as an intermediate in reactions with suitable trapping reagents. Future research will be directed towards the generation of non stabilized bismuthinidenes and their exploitation in synthetic chemistry.

## Methods and experimental

Details of experimental conditions for synchrotron experiments, for calculations with DFT and multi reference methods, for the preparation of compounds **1** and **8** (including two methods for the preparation of **1**<sup>19</sup>), and for IR and EPR spectroscopic experiments are given in the ESI†.

All calculations were performed with the ORCA program package, version 4.1.1 and 4.2.<sup>56</sup>

The spectroscopic experiments were carried out at the VUV beamline of the Swiss Light Source (SLS) at the Paul Scherrer Institute, Villigen/CH. In most experiments the photon energy was scanned in 5 meV steps and calibrated using autoionization resonances in Ar. The ionization energies reported in the main paper are accurate to within ±20 meV and were corrected for the Stark shift by the extraction field (8–9 meV). Note that in some

experiments 10 meV steps were used. A detailed description of the beamline is given in the literature.<sup>47</sup>

[MeBi(SPh)<sub>2</sub>] (**8**).<sup>39</sup>

Neat trimethyl bismuth (100 mg, 0.394 mmol) and diphenyl disulfide (85.9 mg, 0.393 mmol) were heated to 120 °C for 3 d. After cooling to ambient temperature, a green solid was obtained, which was dissolved in benzene (1.5 mL) and layered with *n* pentane (1.5 mL). After 16 h at ambient temperature, the product could be isolated by filtration and dried *in vacuo* as yellow needles. Yield: 70.8 mg, 0.160 mmol, 41%.

<sup>1</sup>H NMR (500 MHz, C<sub>6</sub>D<sub>6</sub>):  $\delta$  = 0.94 (s, 3H, CH<sub>3</sub>), 6.86 (dd, 2H, <sup>3</sup>J<sub>HH</sub> = 7.4 Hz, <sup>3</sup>J<sub>HH</sub> = 7.5 Hz, *p* C<sub>6</sub>H<sub>5</sub>), 7.03 (t, 4H, <sup>3</sup>J<sub>HH</sub> = 7.6 Hz, *m* C<sub>6</sub>H<sub>5</sub>), 7.48 (d, 4H, <sup>3</sup>J<sub>HH</sub> = 7.8 Hz, *o* C<sub>6</sub>H<sub>5</sub>) ppm.

<sup>13</sup>C NMR (126 MHz, C<sub>6</sub>D<sub>6</sub>):  $\delta$  = 40.40 (br, CH<sub>3</sub>), 127.29 (s, *p* C<sub>6</sub>H<sub>5</sub>), 128.60 (s, *m* C<sub>6</sub>H<sub>5</sub>), 135.62 (s, *o* C<sub>6</sub>H<sub>5</sub>), 136.06 (s, *ipso* C<sub>6</sub>H<sub>5</sub>) ppm.

Elemental analysis: anal. calc. for [C<sub>13</sub>H<sub>13</sub>BiS<sub>2</sub>] (442.35 g mol<sup>-1</sup>): C 35.30, H 2.96, S 14.50; found: C: 35.12, H 2.90, S 14.40.

## Conflicts of interest

There are no conflicts to declare.

## Author contributions

DPM, DS, ER and PH carried out the synchrotron radiation experiments, JR carried out all synthetic work, SW and DK performed the computations, IK performed the ESR and TP the IR spectroscopy. IF, BE and CL planned the project, contributed to the interpretation of the results and wrote the manuscript.

## Acknowledgements

Funding by the DFG (funding to CL, IF (FI575/13 1) and through GRK2112), by the Swiss Federal Office for Energy (BFE Contract Number SI/501269 01), and the FCI (PhD and Liebig scholarships to DK and CL) are gratefully acknowledged. CL thanks Prof. Holger Braunschweig for continuous support. The spectroscopic experiments were performed at the VUV beamline of the Swiss Light Source, located at the Paul Scherrer Institute (PSI).

## Notes and references

- 1 L. Dostál, *Coord. Chem. Rev.*, 2017, **353**, 142–158.
- 2 F. Dielmann, O. Back, M. Henry Ellinger, P. Jerabek, G. Frenking and G. Bertrand, *Science*, 2012, **337**, 1526–1528.
- 3 L. Liu, D. A. Ruiz, D. Munz and G. Bertrand, *Chem*, 2016, **1**, 147–153.
- 4 P. Šimon, F. de Proft, R. Jambor, A. Růžička and L. Dostál, *Angew. Chem., Int. Ed.*, 2010, **49**, 5468–5471.
- 5 I. Vranova, M. Alonso, R. Lo, R. Sedlak, R. Jambor, A. Ruzicka, F. De Proft, P. Hobza and L. Dostál, *Chem. Eur. J.*, 2015, **21**, 16917–16928.



- 6 I. Vranova, M. Alonso, R. Jambor, A. Ruzicka, M. Erben and L. Dostál, *Chem. Eur. J.*, 2016, **22**, 7376–7380.
- 7 G. C. Wang, L. A. Freeman, D. A. Dickie, R. Mokrai, Z. Benkó and R. J. Gilliard, *Chem. Eur. J.*, 2019, **25**, 4335–4339.
- 8 C. Lichtenberg, *Angew. Chem., Int. Ed.*, 2016, **55**, 484–486.
- 9 A. M. Arif, A. H. Cowley, N. C. Norman and M. Pakulski, *J. Am. Chem. Soc.*, 1985, **107**, 1062–1063.
- 10 F. Wang, O. Planas and J. Cornella, *J. Am. Chem. Soc.*, 2019, **141**, 4235–4240.
- 11 M. C. Thompson, J. Ramsay and J. M. Weber, *Angew. Chem., Int. Ed.*, 2016, **55**, 15171–15174.
- 12 H. T. Sun, Y. Matsushita, Y. Sakka, N. Shirahata, M. Tanaka, Y. Katsuya, H. Gao and K. Kobayashi, *J. Am. Chem. Soc.*, 2012, **134**, 2918–2921.
- 13 H. T. Sun, Y. Sakka, N. Shirahata, Y. Matsushita, K. Deguchi and T. Shimizu, *J. Phys. Chem. C*, 2013, **117**, 6399–6408.
- 14 E. H. Fink, K. D. Setzer, D. A. Ramsay, M. Vervloet and J. M. Brown, *J. Mol. Spectrosc.*, 1990, **142**, 108–116.
- 15 J. D. Corbett, *J. Am. Chem. Soc.*, 1958, **80**, 4257–4260.
- 16 P. Kuijpers and A. Dymanus, *Chem. Phys. Lett.*, 1976, **39**, 217–220.
- 17 E. H. Fink, K. D. Setzer, D. A. Ramsay and M. Vervloet, *Chem. Phys. Lett.*, 1991, **179**, 95–102.
- 18 R. A. Lynde and J. D. Corbett, *Inorg. Chem.*, 1971, **10**, 1746–1749.
- 19 K. Schäfer and F. Hein, *Z. Anorg. Allg. Chem.*, 1917, **100**, 249–303.
- 20 S. Schulz, A. Kuczkowski, D. Bläser, C. Wolper, G. Jansen and R. Haack, *Organometallics*, 2013, **32**, 5445–5450.
- 21 J. M. Herbelin, R. Klingberg, D. J. Spencer, M. A. Kwok, H. Bixler, R. Ueunten, R. Cook and W. Hansen, *Opt. Commun.*, 1981, **36**, 475–476.
- 22 S. W. Kang, K. M. Jeon, J. S. Shin, J. R. Chun, Y. H. Kim, S. J. Lee and J. Y. Yun, *Chem. Vap. Deposition*, 2013, **19**, 61–67.
- 23 B. Hornung, A. Bodi, C. I. Pongor, Z. Gengeliczki, T. Baer and B. Sztáray, *J. Phys. Chem. A*, 2009, **113**, 8091–8098.
- 24 T. Baer and R. P. Tuckett, *Phys. Chem. Chem. Phys.*, 2017, **19**, 9698–9723.
- 25 B. K. Cunha de Miranda, C. Alcaraz, M. Elhanine, B. Noller, P. Hemberger, I. Fischer, G. Garcia, H. Soldi Lose, B. Gans, L. A. Viera Mendez, S. Boye Peronne, S. Douin, J. Zabka and P. Botschwina, *J. Phys. Chem. A*, 2010, **114**, 4818–4830.
- 26 A. M. Schulenburg, C. Alcaraz, G. Grassi and F. Merkt, *J. Chem. Phys.*, 2006, **125**, 104310.
- 27 H. Bock and R. Dammel, *Chem. Ber.*, 1987, **120**, 1961–1970.
- 28 F. Holzmeier, M. Lang, K. Hader, P. Hemberger and I. Fischer, *J. Chem. Phys.*, 2013, **138**, 214310.
- 29 S. J. Kim, T. P. Hamilton and H. F. Schaefer III, *J. Phys. Chem.*, 1993, **97**, 1872–1877.
- 30 S. Lacombe, D. Gonbeau, J. L. Cabioch, B. Pellerin, J. M. Denis and G. Pfister Guillouzo, *J. Am. Chem. Soc.*, 1988, **110**, 6964–6967.
- 31 S. Y. Liang, P. Hemberger, J. Levalois Grützmacher, H. Grützmacher and S. Gaan, *Chem. Eur. J.*, 2017, **23**, 5595–5601.
- 32 S. Y. Liang, P. Hemberger, N. M. Neisius, A. Bodi, H. Grützmacher, J. Levalois Grützmacher and S. Gaan, *Chem. Eur. J.*, 2015, **21**, 1073–1080.
- 33 G. Herzberg, *Molecular Spectra and Molecular Structure*, Krieger, Malabar/FL, 1966.
- 34 S. J. Price and A. F. Trotman Dickenson, *Trans. Faraday Soc.*, 1958, **54**, 1630–1637.
- 35 L. H. Long and J. F. Sackman, *Trans. Faraday Soc.*, 1954, **50**, 1177–1182.
- 36 S. J. W. Price, in *Comprehensive Chemical Kinetics*, ed. C. H. Bamford and C. F. H. Tipper, Elsevier, 1972, vol. 4, pp. 197–257.
- 37 The average bond dissociation energy for a Bi–CH<sub>3</sub> bond in **1** has been reported to be 33.8 kcal mol<sup>-1</sup> (141 kJ mol<sup>-1</sup>) (ref. 35). The sum of the bond dissociation energies associated with the cleavage of the second and the third Bi–CH<sub>3</sub> bond in **1** (BDE(MeBi–CH<sub>3</sub>) plus BDE(Bi–CH<sub>3</sub>)) was estimated to be at least 57.8 kcal mol<sup>-1</sup> (242 kJ mol<sup>-1</sup>) (ref. 34). This gives a maximum value of 43.6 kcal mol<sup>-1</sup> (182 kJ mol<sup>-1</sup>) for the bond dissociation energy of the first Bi–CH<sub>3</sub> bond in **1**.
- 38 D. L. Haire, U. M. Oehler, P. H. Krygsmann and E. G. Janzen, *J. Org. Chem.*, 1988, **53**, 4535–4542.
- 39 In a different synthetic approach, compound **8** has previously been obtained from reaction of *in situ* generated [MeBi(OEt)<sub>2</sub>] with two equiv. HSPH: M. Wieber and U. Baudis, *Z. Anorg. Allg. Chem.*, 1976, **423**, 40–46.
- 40 A. G. Davies and S. C. W. Hook, *J. Chem. Soc. B*, 1970, 735–737.
- 41 M. Wieber and I. Sauer, *Z. Naturforsch., B: Anorg. Chem., Org. Chem.*, 1984, **39**, 1668–1670.
- 42 P. Simon, R. Jambor, A. Ruzicka and L. Dostál, *Organometallics*, 2013, **32**, 239–248.
- 43 M. Dräger and B. M. Schmidt, *J. Organomet. Chem.*, 1985, **290**, 133–145.
- 44 K. M. Anderson, C. J. Baylies, A. H. M. Monowar Jahan, N. C. Norman, A. G. Orpen and J. Starbuck, *Dalton Trans.*, 2003, 3270–3277.
- 45 P. Šimon, R. Jambor, A. Růžička and L. Dostál, *J. Organomet. Chem.*, 2013, **740**, 98–103.
- 46 For reactions of the related precursor Sb<sub>2</sub>Me<sub>4</sub> under near chemical vapor deposition conditions see: N. Bahlawane, F. Reilmann, S. Schulz, D. Schuchmann and K. Kohse-Höinghaus, *J. Am. Soc. Mass Spectrom.*, 2008, **19**, 1336–1342.
- 47 M. Johnson, A. Bodi, L. Schulz and T. Gerber, *Nucl. Instrum. Methods Phys. Res., Sect. A*, 2009, **610**, 597–603.
- 48 (a) E. Welz, J. Böhnke, R. D. Dewhurst, H. Braunschweig and B. Engels, *J. Am. Chem. Soc.*, 2018, **140**, 12580–12591; (b) J. Böhnke, T. Dellermann, M. A. Celik, I. Krummenacher, R. D. Dewhurst, S. Demeshko, W. C. Ewing, K. Hammond, M. Heß, E. Bill, E. Welz, M. I. S. Röhr, R. Mitrić, B. Engels, F. Meyer and H. Braunschweig, *Nat. Commun.*, 2018, **9**, 1197.
- 49 V. Settels, W. Liu, J. Pflaum, R. F. Fink and B. Engels, *J. Comput. Chem.*, 2012, **33**, 1544–1553.
- 50 (a) J. D. Chai and M. Head Gordon, *J. Chem. Phys.*, 2008, **128**, 084106; (b) S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

- 51 (a) C. Angeli, R. Cimiraqlia, S. Evangelisti, T. Leininger and J. P. Malrieu, *J. Chem. Phys.*, 2001, **114**, 10252–10264; (b) C. Angeli, R. Cimiraqlia and J. P. Malrieu, *Chem. Phys. Lett.*, 2001, **350**, 297–305; (c) C. Angeli, R. Cimiraqlia and J. P. Malrieu, *J. Chem. Phys.*, 2002, **117**, 9138–9153.
- 52 D. A. Pantazis and F. Neese, *Theor. Chem. Acc.*, 2012, **131**, 1292.
- 53 It should be noted that methyl nitrene ( $\text{N-CH}_3$ ) and methyl phosphinidene ( $\text{P-CH}_3$ ) also have triplet ground states. But the structures with methylene (rather than methyl) functional groups, *i.e.*  $\text{HN=CH}_2$  and  $\text{HP=CH}_2$ , are energetically favored over  $\text{N-CH}_3$  and  $\text{P-CH}_3$ . See ref. 27–32 and A. G. Kutateladze, *Computational Methods in Photochemistry*, CSC Press Taylor and Francis Group, 2019.
- 54 Bismuth containing radical species were not detected under these conditions. It has been reported that even persistent dialkyl bismuth radicals may not be detectable by EPR spectroscopy, which has been ascribed to fast relaxation as a result of spin orbit coupling: S. Ishida, F. Hirakawa, K. Furukawa, K. Yoza and T. Iwamoto, *Angew. Chem., Int. Ed.*, 2014, **53**, 11172–11176.
- 55 Trapping of alkyl radical species in reactions of bismuth compounds has been reported: J. Ramler, I. Kruppenacher and C. Lichtenberg, *Angew. Chem., Int. Ed.*, 2019, **58**, 12924–12929.
- 56 (a) F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 73–78; (b) F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1327.



## 4. Pentadiynylidene and Its Methyl-Substituted Derivates

THE JOURNAL OF  
PHYSICAL CHEMISTRY A

Cite This: *J. Phys. Chem. A* 2019, 123, 2008–2017

Article

pubs.acs.org/JPCA

### Pentadiynylidene and Its Methyl-Substituted Derivates: Threshold Photoelectron Spectroscopy of $R_1-C_5-R_2$ Triplet Carbon Chains

Engelbert Reusch,<sup>†</sup> Dustin Kaiser,<sup>†</sup> Domenik Schleier,<sup>†</sup> Rachel Buschmann,<sup>‡</sup> Anke Krueger,<sup>‡</sup> Thomas Hermann,<sup>†</sup> Bernd Engels,<sup>\*,†,§</sup> Ingo Fischer,<sup>\*,†,§</sup> and Patrick Hemberger<sup>\*,§,¶</sup>

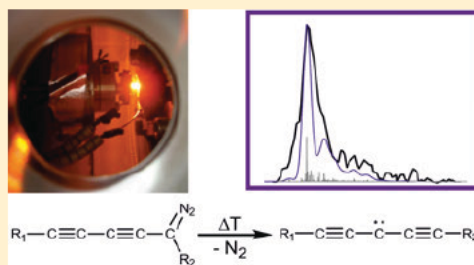
<sup>†</sup>Institute of Physical and Theoretical Chemistry, University of Würzburg, Am Hubland D-97074, Germany

<sup>‡</sup>Institute of Organic Chemistry, University of Würzburg, Am Hubland D-97074, Germany

<sup>§</sup>Laboratory for Femtochemistry and Synchrotron Radiation, Paul Scherrer Institut (PSI), CH-5232 Villigen, Switzerland

Supporting Information

**ABSTRACT:** Mass-selective threshold photoelectron spectroscopy in the gas phase was employed to characterize the dialkynyl triplet carbenes pentadiynylidene ( $HC_5H$ ), methylpentadiynylidene ( $MeC_5H$ ), and dimethylpentadiynylidene ( $MeC_2Me$ ). Diazo compounds were employed as precursors to generate the carbenes by flash pyrolysis. The  $R_1-C_5-R_2$  carbon chains were photoionized by vacuum ultraviolet (VUV) synchrotron radiation in photoelectron photoion coincidence (PEPICO) experiments. High-level ab initio computations were carried out to support the interpretation of the experiments. For the unsubstituted pentadiynylidene ( $R_1 = R_2 = H$ ) the recorded spectrum yields an adiabatic ionization energy ( $IE_{ad}$ ) of  $8.36 \pm 0.03$  eV. In addition, a second carbene isomer, 3-(didehydrovinylidene)cyclopropane, with a singlet electronic ground state, was identified in the spectrum based on the  $IE_{ad}$  of  $8.60 \pm 0.03$  eV and Franck Condon simulations. We found that multireference computations are required to reliably calculate the  $IE_{ad}$  for this molecule. CASPT2 computations predicted an  $IE_{ad} = 8.55$  eV, while coupled-cluster computations significantly overestimate the IE. The cyclic isomer is most likely formed from another isomer of the precursor present in the sample. Stepwise methyl-substitution of the carbene leads to a reduction of the IE to  $7.77 \pm 0.04$  eV for methylpentadiynylidene and  $7.27 \pm 0.06$  eV for dimethylpentadiynylidene. The photoionization and dissociative photoionization of the precursors is investigated as well.



#### INTRODUCTION

In this paper we report threshold photoelectron spectra of the triplet carbene pentadiynylidene,  $HC_5H$ , and its methyl-substituted derivatives,  $R_1-C_5-R_2$ , as depicted in Figure 1 (1, 2, and 3). While chains with even  $n$  like acetylene and di- and triacetylene are stable closed-shell molecules, odd  $n$  carbon chains are highly reactive biradicals or triplet carbenes in the electronic ground state of the neutral. Nevertheless, they occur as intermediates under extreme conditions, for example in combustion processes, the interstellar medium, or in planetary atmospheres. In space they are hard to detect by radioastronomy due to their small ( $HC_3H$ ) or absent ( $HC_5H$ ) permanent dipole moment;<sup>1–4</sup> however, they have been suggested to be involved in the atmospheric chemistry of Titan, the largest moon of Saturn.<sup>5</sup> In combustion, especially in acetylene containing fuels,<sup>6,7</sup> odd  $n$  carbon chains like propargylene ( $n = 3$ ) or pentadiynylidene ( $n = 5$ ) dimerize efficiently and are thus possible intermediates toward the formation of polycyclic aromatic hydrocarbons (PAHs), soot, or fullerenes.<sup>8</sup>

Recently, we investigated propargylene ( $n = 3$ ) in the gas phase, reported a threshold photoelectron spectrum,<sup>9</sup> and studied the photodissociation dynamics by velocity map imaging and by surface hopping computations.<sup>10</sup> In contrast,  $n = 5$  carbon chains have so far been investigated predominately in low temperature matrices. Infrared, UV/vis and EPR spectra in Ar or  $N_2$  matrix are available for 1, 2, and in part for 3.<sup>11–13</sup> While propargylene favors a slightly bent geometry with  $C_2$  symmetry,<sup>14</sup> a linear geometry with  $D_{\infty h}$  symmetry was found for 1,<sup>15,16</sup> as confirmed by coupled cluster calculations.<sup>17</sup> In a valence bond picture, pentadiynylidene can be described by two resonance structures (see Figure 1), one corresponding to a 1,5-biradical and a second dominating one to a diethynylcarbene. Computations indicated that the corresponding triplet state is lower in energy than the singlets. A third structure with the carbene center on a single terminal carbon atom does not contribute significantly.<sup>14,17</sup> For the

Received: December 20, 2018

Revised: January 31, 2019

Published: February 18, 2019

ACS Publications © 2019 American Chemical Society

2008

DOI: 10.1021/acs.jpca.8b12244  
*J. Phys. Chem. A* 2019, 123, 2008–2017

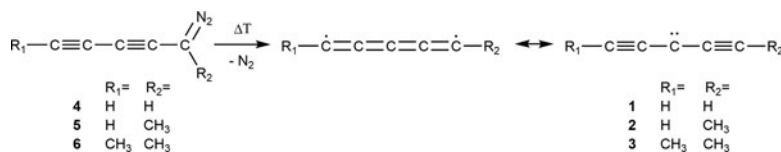


Figure 1. Pentadiynylidene 1, methylpentadiynylidene 2, and dimethylpentadiynylidene 3 were generated by flash pyrolysis from 1-diazo-penta-2,4-diyne 4, 1-diazo-hexa-2,4-diyne 5, and 2-diazo-hepta-3,5-diyne 6.

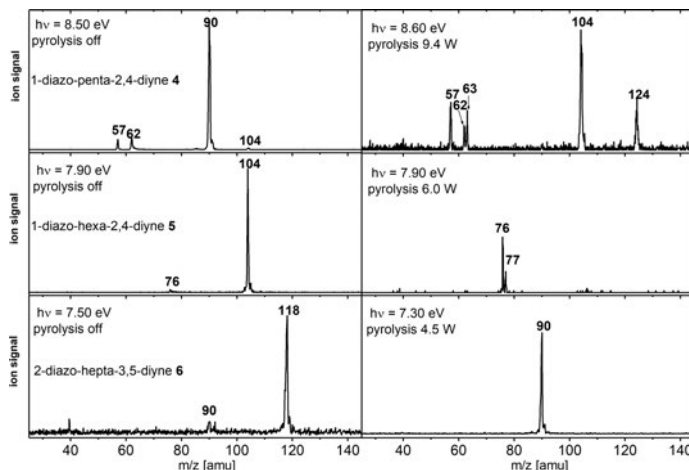


Figure 2. From top to bottom: Mass spectra of 1-diazo-penta-2,4-diyne 4, 1-diazo-hexa-2,4-diyne 5, and 2-diazo-hepta-3,5-diyne 6 without pyrolysis (left-hand column) and with pyrolysis (right-hand column). Efficient conversion to the carbenes is visible upon pyrolysis. Note that for each compound a photon energy only slightly above the IE of the precursor was employed.

methyl-substituted compound 2 the dominant carbene character has also been confirmed.<sup>12</sup> Further interest in 2 and 3 is due to the increased stability compared to 1 and the additional photochemistry observed.<sup>12</sup>

Several  $\text{C}_5\text{H}_2$  isomers exist, and it has been computed that all but pentadiynylidene possess a singlet ground state.<sup>17–20</sup> A number of them have been investigated by Fourier transform microwave spectroscopy,<sup>21–23</sup> but electronic transitions have only been investigated computationally. For pentadiynylidene a weak band at around 400–500 nm was computed to correspond to the  $\text{A } ^3\Sigma_u^- \rightarrow \text{X } ^3\Sigma_g^-$  transition.<sup>24,25</sup> In contrast, for the cation  $\text{HC}_5\text{H}^+$ ,  $\text{I}^+$  and the cation of the isomer pentatetraenylidene electronic absorption spectra were recorded.<sup>26</sup> However, the ionization energy (IE) of 1 has not yet been determined.

Here, we employ photoelectron spectroscopy to explore the photoionization of 1, 2, and 3, using threshold photoelectron-photoion coincidence techniques (TPEPICO).<sup>27</sup> This method permits the photoion mass-selected threshold photoelectron spectra (ms-TPES) to be recorded for each molecule present in a reaction mixture by correlating ions and electrons; thus, contributions from the precursor or possible side products with different molecular mass can be excluded, in contrast to conventional photoelectron spectroscopy. Spectra of numerous reactive intermediates have been recorded by TPEPICO, including reactive species like allyl,<sup>28</sup> propargyl,<sup>29</sup> benzyl,<sup>30</sup> fulvenallenyl,<sup>31</sup>  $c\text{-C}_3\text{H}_2$ ,<sup>32</sup>  $\text{CH}_x$ ,<sup>33–35</sup>  $\text{CF}_3$ ,<sup>36</sup>  $\text{C}_3\text{H}$ ,<sup>37</sup> and HBBH.<sup>38</sup> The method also permits to distinguish different

isomers as shown for example for picolyl,<sup>39</sup> xylyl,<sup>40,41</sup> and xylylenes.<sup>42</sup> Since TPEPICO recently evolved into a versatile chemical analysis tool to probe elusive intermediates in reactive environments such as model flames,<sup>43–45</sup> catalytic reactors,<sup>46,47</sup> or in kinetics experiments<sup>48,49</sup> clean and isomer-specific spectroscopic data of radicals and carbenes are much-needed.

## METHODS

The experiments were carried out at the x04db beamline of the Swiss Light Source (SLS) storage ring in Villigen, Switzerland. The CRF-PEPICO instrument has been described in detail elsewhere; thus, only a brief overview on the setup<sup>50</sup> and the beamline<sup>51</sup> is given here. The molecules were ionized by vacuum ultraviolet (VUV) synchrotron radiation provided by a bending magnet and collimated onto a 150 lines  $\text{mm}^{-1}$  plane grating monochromator. The photon energy resolution of  $E/\Delta E$  is about  $1.5 \times 10^3$  and higher order radiation was absorbed by a  $\text{MgF}_2$  window. The energy was calibrated using the  $11s'$   $13s'$  argon autoionization resonances in first and second order. All spectra are corrected for photon flux.

The PEPICO endstation contains velocity map imaging (VMI) spectrometers for both photoelectrons and photoions.<sup>51–53</sup> Cations and electrons were extracted vertically in opposite directions, accelerated by electric fields, which are designed to allow for simultaneous Wiley McLaren space focusing as well as velocity map imaging conditions. The charged particles are detected by position sensitive delay-line

2009

DOI: 10.1021/acs.jpca.8b12244  
J. Phys. Chem. A 2019, 123, 2008–2017

anode (Roentdek DLD40) detectors. Cations and electrons were correlated employing a multiple start/multiple stop data acquisition scheme and thus analyzed in coincidence.<sup>54</sup> Threshold photoelectrons were selected with a resolution of 5–10 meV and the contribution of hot electrons was subtracted as described in the literature.<sup>55</sup> Averaging times were between 60 and 150 s per data point.

As indicated in Figure 1, pentadiynylidene **1**, methylpentadiynylidene **2**, and dimethylpentadiynylidene **3** were generated by flash pyrolysis of diazo precursors (1-diazo-penta-2,4-diyne **4**, 1-diazo-hexa-2,4-diyne **5**, and 2-diazo-hepta-3,5-diyne **6**).<sup>56</sup> These precursors were synthesized following a procedure similar to the one described in the literature.<sup>11,57</sup> Since **4**, **5**, and **6** are unstable under ambient conditions, we employed the sodium salts of the corresponding tosylhydrazones and synthesized the diazo compounds in situ. The sodium salts were heated to 60–70 °C in an in-vacuum sample container to produce the diazo compounds directly in the gas phase. For part of the discussion below it is important to point out that due to this approach we could not determine the purity of the diazo compound. Subsequently the vapors were picked up by Ar carrier gas and expanded through a 100  $\mu\text{m}$  nozzle directly into the pyrolysis reactor, an electrically heated silicon carbide (SiC) tube of 1 mm inner diameter.<sup>56</sup> Here the diazo compounds were pyrolyzed to the carbenes **1**, **2**, and **3**. The resulting jet was skimmed and directed toward the ionization volume of the CRF-PEPICO endstation.

DFT ( $\omega$ B97xD,<sup>58</sup> M05-2X(D3),<sup>59,60</sup> and M06-2X<sup>61</sup>), CBS-QB3,<sup>62</sup> and coupled cluster (CCSD) calculations were performed using the Gaussian 16 program package.<sup>63</sup> The single-reference methods were used in their unrestricted variant in all cases. CASPT2 evaluations<sup>64,65</sup> used MOLCAS 7.8<sup>66</sup> with the active space being iteratively built from pseudonatural orbitals of smaller spaces, starting at (2,2). All reported stationary geometries were evaluated to be true minima of the potential energy surface by computation of the molecular Hessian using the respective method, except for the numerical CASPT2(10,10) optimization of compound **9**<sup>7</sup> (see below for definition). Values of the  $\mathcal{T}_1$  diagnostic are obtained from converged CCSD amplitudes.<sup>67</sup>

The TPE spectra of the precursors were modeled by Franck-Condon simulations at 0 K with the program FCfit<sup>68</sup> version 2.8.20, while ms-TPE spectra of the carbenes **1**, **2**, and **3** were simulated at 600 K using the program ezSpectrum, which takes hot and sequence band transitions into account.<sup>69</sup>

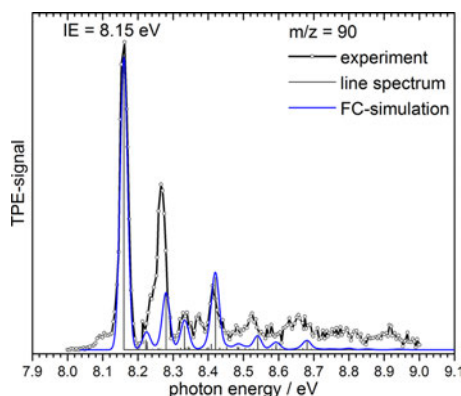
## RESULTS AND DISCUSSION

**Mass Spectra.** Figure 2 shows selected mass spectra with and without pyrolysis. Without pyrolysis (left-hand column) the precursor signals of **4**, **5**, and **6** dominate the spectrum. In addition, a small peak originating from dissociative photoionization ( $\text{N}_2$  loss) is observed in all three spectra. The mass spectra recorded with pyrolysis turned on (right-hand column) show a complete conversion of the precursor even at low pyrolysis power and generation of the carbenes **1** ( $m/z = 62$ , top right), **2** ( $m/z = 76$ , center right), and **3** ( $m/z = 90$ , bottom right). For 1-diazo-penta-2,4-diyne **4** (top row) further peaks appear at  $m/z = 57$  and 104, identified as styrene by ms-TPEs. Both originate from previous experiments, are independent of the pyrolysis temperature, and will thus not be discussed further. Signals at  $m/z = 63$  and 124 are due to  $\text{C}_3\text{H}_3$  (H-addition to the carbene) and the dimerization product of **1**,  $\text{C}_{10}\text{H}_4$ . A ms-TPE spectrum of the dimer,

identified as tetraethynylethene ( $\text{C}_{10}\text{H}_4$ ) is presented in the SI. While efficient dimerization of **1** has been observed before,<sup>11,70</sup> signals for the dimers of **2** and **3** are small. Stabilization of the carbene by the  $\text{CH}_3$  group and/or steric effects might be responsible. In the pyrolysis of **6** a small peak appears at  $m/z = 91$ . It originates from the  $^{13}\text{C}$  isotope of  $m/z = 90$  as evident from the relative peak intensities in the mass spectrum and not from a H atom addition product.

**Photoionization of Precursors **4**, **5**, and **6**.** In the next step photoionization and dissociative photoionization of the precursors **4**, **5**, and **6** were investigated without pyrolysis. As the loss of  $\text{N}_2$  in the cation was computed to be associated with a reverse barrier, no binding energies could be derived. Therefore, the breakdown diagrams are only discussed in the SI.

The threshold photoelectron spectrum of **4** (black line with open circles) is depicted in Figure 3. Based on CBS-QB3



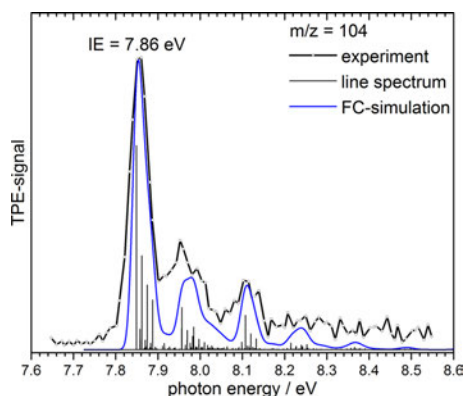
**Figure 3.** TPE spectrum of 1-diazo-penta-2,4-diyne **4** and corresponding Franck-Condon simulation for the precursor (blue line: stick spectrum convoluted with a Gaussian of 20 meV fwhm) shows a reasonable agreement. The ionization energy of **4** is determined to be  $8.15 \pm 0.01$  eV.

calculations of the neutral **4** and the cation  $\mathbf{4}^+$ , a Franck-Condon simulation (blue line) was carried out. Hot and sequence band transitions were ignored; nevertheless, the simulation shows a good agreement with the experiment. While the band positions are well represented, there are some deviations in the band intensities. The first major band at  $8.15 \text{ eV} \pm 0.01 \text{ eV}$  is assigned to the adiabatic ionization energy  $\text{IE}_{\text{ad}}$  and is in excellent agreement with the computed value at 8.14 eV. The error is derived from the full width at half-maximum (fwhm) of the band. Further peaks are found at 8.27 (+970  $\text{cm}^{-1}$ ) and 8.41 eV (+2100  $\text{cm}^{-1}$ ). The first one corresponds to the fundamental of the C–N stretching mode with a computed value of 1001  $\text{cm}^{-1}$ . The second band (computed value 2172  $\text{cm}^{-1}$ ) is a combination of the C–C and the N–N stretching motions. At 8.35 eV the H–C–N bending vibration (1437  $\text{cm}^{-1}$ ) also contributes to the TPE spectrum. This is in line with the most pronounced change in geometry upon ionization affecting the length of the C–N bond,  $R_{\text{C–N}}$  and the C–C bond,  $R_{\text{C–C}}$  next to the diazo ( $\text{N}_2$ ) group. While  $R_{\text{C–C}}$  decreases from 1.40 to 1.35 Å,  $R_{\text{C–N}}$  increases from 1.32 to 1.36 Å. All other geometry parameters change only slightly and in particular  $R_{\text{N–N}}$  and  $R_{\text{C}\equiv\text{C}}$  remain almost constant. All



geometry parameters are summarized in the Supporting Information. Note that the intensity of the band at  $+970\text{ cm}^{-1}$  (around  $8.25\text{ eV}$ ) is significantly underestimated by the simulation. Contributions of another isomer may be a plausible explanation for this observation. Indeed, an isomer of **4** with the  $\text{N}_2$  at the central carbon atom possesses a computed  $\text{IE}_{\text{ad}}$  of  $8.20\text{ eV}$  (CBS-QB3). Interestingly, when the pyrolysis was turned on, the spectrum shown in Figure 3 disappeared and a second band appeared around  $8.70\text{ eV}$ , see Figure S2. Obviously an isomer of  $\text{C}_5\text{H}_2\text{N}_2$  is formed at elevated temperature, most likely from the tosyl hydrazone salt employed as the sample. This will be further discussed below.

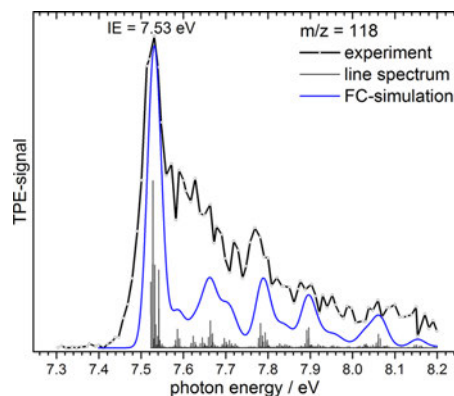
The threshold photoelectron spectrum of **5** is shown in Figure 4. The Franck Condon simulation (blue line) agrees



**Figure 4.** TPE spectrum of 1-diazo-hexa-2,4-diyne **5**. An  $\text{IE}_{\text{ad}}$  of  $7.86 \pm 0.02\text{ eV}$  can be extracted from the first major peak. For comparison, the Franck Condon simulation (blue line: stick spectrum convoluted with a Gaussian of  $25\text{ meV}$  fwhm) is given.

again well with the experimental data and the simulated intensities match the experimental ones. However, the spectrum is broadened due to the torsion of the methyl group, which is not properly represented in the simulation that is based on the harmonic approximation. It is included in the simulations by simply using a Gaussian with a larger fwhm of  $25\text{ meV}$ . Again, the first major band at  $7.86 \pm 0.02\text{ eV}$  represents the  $\text{IE}_{\text{ad}}$  in agreement with the computed value of  $7.81\text{ eV}$ . Two features are visible in the experimental TPEs at  $7.95$  and  $8.10\text{ eV}$ , which correspond to an excitation of the  $\text{C-N}$  stretching mode ( $897\text{ cm}^{-1}$ ) and a combination of  $\text{CC}$  and  $\text{NN}$  stretching vibrations ( $2153\text{ cm}^{-1}$ ), similar to what was found for **4**. Further combination bands with the methyl torsional mode occur ( $106\text{ cm}^{-1}$ ), which lead to an additional broadening of the experimental features. Similar to **4** we also find changes in  $R_{\text{C-C}}$  and  $R_{\text{C-N}}$  upon ionization. The latter increases from  $1.32$  to  $1.37\text{ \AA}$ , while the former decreases from  $1.40$  to  $1.36\text{ \AA}$ . In addition, the  $\text{C-C}$  bond to the methyl group decreases from  $1.52$  to  $1.50\text{ \AA}$ , while the remaining bond lengths differ only slightly; see the SI for details. Note that the simulations performed for **5** do not differ from the experiment as found for **4** for the band at  $+970\text{ cm}^{-1}$ .

Figure 5 depicts the TPE spectrum of **6** (black line) and the corresponding Franck Condon simulation (blue line). From the Franck Condon (FC) fit a reasonable agreement with the experimental data is visible. An  $\text{IE}_{\text{ad}}$  at  $7.53 \pm 0.03\text{ eV}$  is



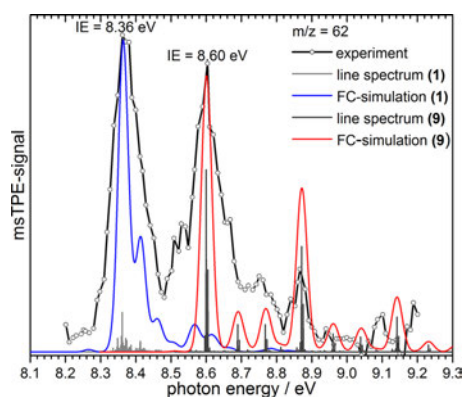
**Figure 5.** TPE spectrum of 2-diazo-hepta-3,5-diyne **6**. An  $\text{IE}_{\text{ad}}$  of  $7.53 \pm 0.03\text{ eV}$  can be extracted from the first peak. For comparison, the Franck Condon simulation (blue line: stick spectrum convoluted with a Gaussian of  $35\text{ meV}$  fwhm) is given.

derived from the first major peak in the experiment, agreeing again well with the value of  $7.48\text{ eV}$ , calculated on the CBS-QB3 level of theory. Beside the first major band, two broad features around  $7.65$  and  $7.79\text{ eV}$  are visible in the experimental spectrum and are accounted for by the simulation. Those bands can again be assigned to a  $\text{C-N}$  stretching ( $1135\text{ cm}^{-1}$ ) mode and to a combination of  $\text{CC}$  and  $\text{NN}$  stretching vibrations at  $2123\text{ cm}^{-1}$ . A low frequency methyl torsion ( $39\text{ cm}^{-1}$ ) is responsible for the peak broadening. In **6** we find again that  $R_{\text{C-N}}$  increases from  $1.31$  to  $1.37\text{ \AA}$  and  $R_{\text{C-C}}$  decreases from  $1.40$  to  $1.36\text{ \AA}$  upon ionization, almost identical to compounds **4** and **5**, confirming the small influence of the methyl group on the structure of the carbon chain. As visible in the mass spectra in Figure 2 dissociative photoionization sets already in at around  $7.5\text{ eV}$ , therefore all electrons were collected.

**Photoionization of the Carbenes 1, 2, and 3.** After the photoionization and the dissociative photoionization of the precursors **4**, **5**, and **6** have been investigated, the pyrolysis was turned on to study the three carbenes pentadiynylidene **1**, methylpentadiynylidene **2**, and dimethylpentadiynylidene **3**.

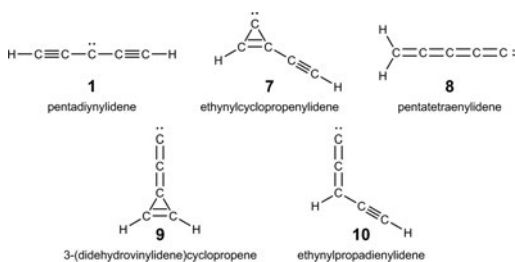
Figure 6 depicts the ms-TPE spectrum of  $m/z = 62$  recorded upon pyrolysis of **4**. Due to the low signal/noise ratio, 5-point smoothing was employed. Two intense bands are observed at  $8.36$  and  $8.60\text{ eV}$ , and in addition, several smaller bands are identified at  $8.53$ ,  $8.75$ ,  $8.87$ , and  $9.10\text{ eV}$ .

CBS-QB3 computations predict an ionization energy of  $8.41\text{ eV}$  for **1**, in excellent agreement with the band at  $8.36\text{ eV}$ . However, the FC-simulation (blue line) shows little vibrational activity in the ion, as expected for ionization from a nonbonding  $p$ -type orbital. Only the  $\text{CC}$  stretching vibration at  $783\text{ cm}^{-1}$  along with some hot- and sequence band transitions contribute to the band, which is well reproduced by a vibrational temperature of  $600\text{ K}$ . Thus, the second band at  $8.60\text{ eV}$  cannot be reasonably assigned to ionization from the triplet ground state ( $T_0$ ) of pentadiynylidene **1**. For the  $S_1$  state of **1** on the other hand an  $\text{IE}$  of  $7.70\text{ eV}$  was computed, in strong disagreement with any of the major bands in the spectrum. A possible source for this band is dissociative photoionization of the precursor. We rule this explanation out for two reasons. First, the precursor was completely converted



**Figure 6.** Ms-TPE spectrum of  $m/z = 62$  (black line with open circles). A Franck Condon simulation including pentadiynylidene **1** (blue line) and 3-(didehydrovinylidene)cyclopropene **9** (red line) agrees well with the experimental spectrum. From the two major peaks  $IE_{ad}$  of  $8.36 \pm 0.03$  eV for **1** and  $8.60 \pm 0.03$  eV for **9** can be extracted.

in the pyrolysis, and second, the kinetic energy distribution of the ions in the images with pyrolysis on shows that the band originates from a species in the molecular beam and not from fragment ions. The ion images with pyrolysis on and off are depicted in Figure S7. Interestingly the band at 8.60 eV becomes smaller, when the temperature of the sample container is lowered at a constant pyrolysis temperature. Since the band cannot be due to dissociative ionization, we computationally investigated the other  $C_5H_2$  isomers depicted in Figure 7 as potential carriers of the band at 8.60 eV. Initial



**Figure 7.** Five  $C_5H_2$  isomers: Pentadiynylidene **1**, ethynylcyclopropenylidene **7**, pentatetraenylidene **8**, 3-(didehydrovinylidene)-cyclopropene **9**, and ethynylpropadienylidene **10**.

CBS-QB3 computations did not yield satisfactory agreement for any of the isomers. To validate the IE's predicted by these computations, we systematically investigated them on a using a variety of computational methods, ranging from density functional theory (DFT) up to CASPT2(14,14). The results are summarized in Tables 1 and 2. While Table 1 shows the relative energies of the neutral isomers, Table 2 compares the adiabatic IE's. For most isomers the deviation from the CBS-QB3 results is small, as expected due to the accuracy of the approach.

The only isomers with a predicted ionization energy in the range of the second band at 8.60 eV are **7** and **9**. As the  $S_0$

ground-state of **7** is almost isoenergetic with the  $T_0$  of **1**, it might be formed in the pyrolysis and was therefore investigated first. Although the predicted IE's do not vary much with the computational method and are not too far off the experimental peak, all methods consistently predict an IE that is at least hundred meV higher than the measured IE of 8.60 eV. Franck Condon (FC) simulations of the TPE-spectrum of **7** are given in the SI. They qualitatively agree for all methods but show a pronounced vibrational progression and are thus in disagreement with the experimental data. In fact they strongly resemble the spectrum obtained for  $c-C_3H_2$ ,<sup>32</sup> because in both cases ionization occurs at the carbene center. We therefore conclude that the second band of  $m/z = 62$  in Figure 6 does not originate from the ionization of isomer **7**.

We now focus on isomer **9**: Interestingly, for the singlet ground-state of **9** the various computational approaches predict vastly different ionization energies (see Table 2). They result from small variations in the predicted equilibrium geometry of  $9^+$  and problems in the convergence of the geometry optimization due to a very flat potential energy surface (see SI Table S9). For a CCSD(T) calculation of the cation with the geometry being obtained at the  $\omega B97xD$  level of theory, the  $\mathcal{T}_1$  diagnostic, which is a measure for the multireference character of a wave function,<sup>67</sup> has a value of 0.16. This indicates that a single-reference method is unable to predict the equilibrium geometry correctly.<sup>71–73</sup> Further informations are given in the SI. Therefore, we optimized the geometry of  $9^+$  by CASPT2(10,10)/cc-pVDZ. Table S7 in the SI compares the single-reference approaches to the CASPT2 geometry and shows that density functional methods are able to reproduce the CASPT2 geometry, while CCSD fails to predict a correct geometry for  $9^+$ . Using the  $\omega B97xD$  optimized geometry for the cation, the ionization energy of **9** from the singlet ground state is predicted to be 8.55 eV at the CASPT2(14,14)/cc-pVTZ level of theory (including ZPE-corrections taken from  $\omega B97xD$  Hessians). Using the CASPT2(10,10) optimized geometry an IE of 8.58 eV is obtained. This is in excellent agreement with the second band in the spectrum at 8.60 eV. Furthermore, FC-simulations including transitions from the  $T_0$  of **1** and the  $S_0$  of **9** (red line in Figure 6) show very good agreement with the experimental spectrum. Vibrational transitions are assigned to a progression of the CC stretch between the ring and the adjacent carbon chain ( $2260\text{ cm}^{-1}$ ) located at 8.88 and 9.14 eV. In addition, the small feature at 8.77 eV can be described by a C C C ring bending vibration at  $1400\text{ cm}^{-1}$ .

This leads to the question how **9** is formed, although it is  $0.78\text{ eV}$  ( $75\text{ kJ mol}^{-1}$ ) less stable compared to **1**. A previous theoretical investigation predicted all isomerization barriers between  $C_5H_2$  isomers to be larger than  $2.3\text{ eV}$ ,<sup>74</sup> as confirmed by our own DFT calculations. These barriers are significant, and therefore rearrangements to higher lying isomers are unlikely under mild pyrolysis conditions. The best explanation relies on the in situ preparation of the diazo precursor **4** from the sodium salt of the tosyl hydrazone. This is supported by our investigations of  $C_5H_2N_2$ : Here a second  $C_5H_2N_2$  isomer appears in the ms-TPEs when the pyrolysis is turned on, see Figure S2. Hence, we suggest that the in situ generation of the diazo-precursor is responsible for the appearance of other  $C_5H_2N_2$  isomers, which upon pyrolysis may yield the singlet carbene **9**.

**Table 1. Relative Energies  $E_{\text{rel}}$  of the  $\text{C}_5\text{H}_2$  Isomers in Their Lowest Singlet and Triplet States Using a Variety of Quantum-Chemical Approaches<sup>a</sup>**

molecule	u $\omega$ B97xD/ 6-311++G(d,p)	uM05-2X(D3)/ 6-311++G(d,p)	uM06-2X/ 6-311++G(d,p)	uCCSD/ cc-pVTZ	uCCSD(T)/cc-pVTZ// uCCSD/cc-pVTZ	CASPT2(14,14)/cc-pVTZ// u $\omega$ B97xD/6-311++G(d,p)
1 $E_{\text{rel}}$ ( $S_1$ )	0.41	0.50	0.55	1.07	0.87	0.70
1 $E_{\text{rel}}$ ( $T_0$ )	0.0	0.0	0.0	0.0	0.0	0.0
7 $E_{\text{rel}}$ ( $S_0$ )	0.04	0.01	-0.16	-0.08	-0.15	0.01
7 $E_{\text{rel}}$ ( $T_1$ )	2.02	2.05	1.93	2.16	2.06	2.03
8 $E_{\text{rel}}$ ( $S_0$ )	0.72	0.73	0.67	0.65	0.40	0.72
8 $E_{\text{rel}}$ ( $T_1$ )	1.50	1.56	1.57	1.57	1.51	1.84
9 $E_{\text{rel}}$ ( $S_0$ )	0.78	0.74	0.63	0.80	0.67	0.78
9 $E_{\text{rel}}$ ( $T_1$ )	3.09	3.13	3.04	<i>b</i>	<i>b</i>	3.15
10 $E_{\text{rel}}$ ( $S_0$ )	0.81	0.82	0.75	0.62	0.47	0.67
10 $E_{\text{rel}}$ ( $T_1$ )	1.93	2.49	1.25	1.68	1.72	1.76

<sup>a</sup>The energy of the thermodynamically most stable isomer 1 in its  $T_0$  ground state is set to zero. Relative energies are given in eV. <sup>b</sup>Geometry optimization does not converge.

**Table 2. Ionization Energies (IEs) of the  $\text{C}_5\text{H}_2$  Isomers in Their Singlet (S) and Triplet (T) Ground State Using a Variety of Quantum-Chemical Approaches<sup>a</sup>**

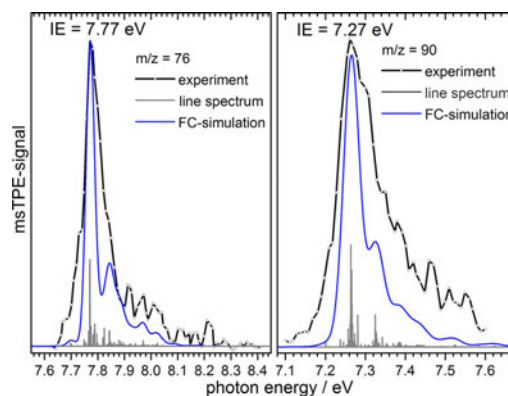
molecule	u $\omega$ B97xD/ 6-311++G(d,p)	uM05-2X(D3)/ 6-311++G(d,p)	uM06-2X/ 6-311++G(d,p)	uCCSD/ cc-pVTZ	uCCSD(T)/cc-pVTZ// uCCSD/cc-pVTZ	CASPT2(14,14)/cc-pVTZ// u $\omega$ B97xD/6-311++G(d,p)	CBS/ QB3
1 IE( $S_1$ )	8.07	8.16	8.04	7.36	7.36 <sup>c</sup>	7.55	7.70
1 IE( $T_0$ )	8.43	8.68	8.60	8.30	8.26 <sup>c</sup>	8.20	8.41
7 IE( $S_0$ )	8.82	8.93	8.92	8.78	8.77 <sup>c</sup>	8.70	8.90
7 IE( $T_1$ )	6.96	7.01	6.94	6.65 <sup>b</sup>	6.68 <sup>c</sup>	6.82	6.95
8 IE( $S_0$ )	9.57	9.32	9.32	9.25	9.40 <sup>c</sup>	9.10	9.44
8 IE( $T_1$ )	8.49	8.56	8.49	8.40	8.36 <sup>c</sup>	8.06	8.47
9 IE( $S_0$ )	8.85	9.02	8.80	9.82	9.96 <sup>c</sup>	8.55	8.80
9 IE( $T_1$ )	6.64	6.73	6.49	<i>b</i>	<i>b</i>	6.28	7.08
10 IE( $S_0$ )	9.74	9.94	9.68	9.71	8.92 <sup>c</sup>	9.62	9.84
10 IE( $T_1$ )	8.64	8.30	9.20	8.67	8.61 <sup>c</sup>	8.55	8.71

<sup>a</sup>Zero-point vibrational energy corrections are included. Energy differences are given in eV. <sup>b</sup>Geometry optimization does not converge. <sup>c</sup>ZPE contribution from u $\omega$ B97xD/6-311++G(d,p).

Note that for the sake of completeness, we also computed IE's for the isomers of  $\text{C}_4\text{N}$ , which also have  $m/z = 62$ , are known to be stable<sup>75,76</sup> and might be formed during pyrolysis. However, the computations ruled out any contributions to the spectrum. Details are therefore only given in the SI (Figure S4).

For comparison and to study the influence of a methyl group on the carbene, we recorded ms-TPE spectra of the methyl-substituted carbene 2 and the dimethyl-substituted one 3, which are presented in Figure 8. Again 5-point smoothing was employed to account for the small signal intensity and the subsequently low signal/noise ratio. The spectrum of  $\text{H}_3\text{C}$   $\text{C}_5\text{-H}$  (Figure 8, left-hand column, average over five spectra) is dominated by a broad band at 7.77 eV, which was assigned to the ionization energy of methylpentadiynylidene 2. From the fwhm of the major peak, error bars of  $\pm 0.04$  eV are assumed. CBS-QB3 calculations yield a value of 7.80 eV for the  $\text{IE}_{\text{ad}}$  again in excellent agreement with the experimental value. The geometry change upon ionization is small and the bond length of  $\text{R}_{\text{C}-\text{CH}_3}$  decreases only from 1.45 to 1.42 Å.

The FC-simulation at an estimated temperature of 600 K shows a very good agreement with the experiment. The peak at 7.77 eV is broadened due to hot and sequence band transitions and contributions of the CC stretching vibration ( $609 \text{ cm}^{-1}$ ). The fact that the spectrum of 2 shows only one dominant peak, confirms that the second peak in the ms-TPE spectrum of 1 is due to a second isomer and not due to an unaccounted vibrational or electronic transition.



**Figure 8.** ms-TPE spectrum of  $m/z = 76$  and Franck Condon simulation of methylpentadiynylidene 2 (blue line) is shown in the left-hand column. The ionization energy is determined to be  $7.77 \pm 0.04$  eV. In the right-hand column the ms-TPE spectrum of  $m/z = 90$  (black line with open circles) and Franck Condon simulation of dimethylpentadiynylidene 3 (blue line) is pictured. An ionization energy of  $7.27 \pm 0.06$  eV can be extracted.

The right-hand side of Figure 8 depicts the corresponding TPE spectrum of dimethylpentadiynylidene 3  $m/z = 90$ , which constitutes an average of ten spectra. The FC-simulation



represents the experimental data quite well. Two methyl rotors contribute to the ms-TPE spectrum of 3, leading to an even more pronounced band broadening, which does not allow to resolve individual transitions in the experiment. The value of  $IE_{ad} = 7.27 \pm 0.06$  eV extracted from the band maximum is in excellent agreement with the computed value of 7.25 eV. The shoulder on the high energy side is due to excitation of a C–C stretch (computed value  $512\text{ cm}^{-1}$ ), which is associated with torsional excitations at  $33\text{ cm}^{-1}$ .

The  $IE_{ad}$  of the three carbenes are summarized in Table 3. As all three molecules can reasonably well be described as

Table 3. Ionization Energies (IE) of the Reactive Molecules Studied in the Present Work<sup>a</sup>

Molecule	$IE_{exp}$ [eV]	$IE_{theo}$ [eV]
1	$8.36 \pm 0.03$	8.41 <sup>b</sup>
9	$8.60 \pm 0.03$	8.55 <sup>c</sup>
2	$7.77 \pm 0.04$	7.80 <sup>b</sup>
3	$7.27 \pm 0.06$	7.25 <sup>b</sup>

<sup>a</sup>Computed IEs are given for comparison. <sup>b</sup>CBS-QB3 computations. <sup>c</sup>CASPT2 computations.

triplet carbenes, ionization occurs from a *p*-type orbital. A comparison shows that the methyl groups lower the ionization energy of 1 by 590 and 1090 meV upon substitution of H by  $\text{CH}_3$ . Since the radical cations are electron-deficient, a methyl group increases the electron density of the cation  $2^+$  and  $3^+$  and thus stabilizes the cation. A comparison of the IE's shows that the effect is slightly larger for the first methyl group, but still substantial for the second one. The stabilization is slightly lower compared to even  $\text{C}_n$  chains with e.g.  $n = 4$ . Butadiyne has an IE = 10.17 eV, which is lowered by 670 and 1270 meV, respectively upon replacing the H atoms subsequently by methyl groups.<sup>77</sup> The IE decrease is also observed for the diazo compounds, but is only around 300 meV per methyl group. In the diazo compounds the lone pair of the nitrogen atoms already stabilizes the cations, so the additional contribution of the methyl groups is smaller. The increased stability of the neutral chain upon methyl addition is reflected in two further observations: First, a dimer is only observed for 1, but not for 2 and 3. Second, the mass spectra in Figure 2 show a more intense peak for  $\text{C}_5\text{H}_3$  than for 1, while for 2 the H-addition is less important and for 3 it is absent. Most likely the steric influence of the methyl groups slows down secondary reactions.

## CONCLUSIONS

The three triplet carbenes pentadiynylidene 1,  $\text{C}_5\text{H}_2$  methylpentadiynylidene 2,  $\text{H}_3\text{C-C}_5\text{H}$  and dimethylpentadiynylidene 3,  $\text{H}_3\text{C-C}_5\text{CH}_3$  were characterized by mass-selected threshold photoelectron spectroscopy. The carbenes were generated by pyrolysis of the diazo precursors 4, 5, and 6. An ionization energy of  $8.36 \pm 0.03$  eV was determined for the  $\text{C}_5\text{H}_2$  isomer pentadiynylidene. A second peak appears in the spectrum at 8.60 eV that cannot be assigned to 1. Since the precursor was completely converted in the pyrolysis and the kinetic energy distribution of the ions shows that the signal originates from a species present in the molecular beam, dissociative ionization can be ruled out. With the aid of computations the peak at 8.60 eV was assigned to isomer 9, 3-(didehydrovinylidene)cyclopropene. Interestingly the calculated IE depends strongly on the computational level and

varies over almost 1.5 eV. We found that a proper description of the ionic ground state  $9^+$  requires a multireference approach. We consequently computed an IE of 8.58 eV by CASPT2, in agreement with the peak at 8.60 eV. It is noticeable that the CBS-QB3 approach, which usually performs well for ionization energies, overestimated the IE by 0.20 eV, while the CCSD(T) value was even more than 1 eV off. As the barriers between the isomers are substantial, isomer 9 is most likely not formed from 1 in the pyrolysis, but rather from an isomer of the precursor,  $\text{HC}_5\text{H}(\text{N}_2)$ .

Substitution of H by  $\text{CH}_3$  leads to a stepwise reduction of the IE to  $7.77 \pm 0.04$  eV for methylpentadiynylidene and  $7.27 \pm 0.06$  eV for dimethylpentadiynylidene, due to the stabilizing effect of the electron-deficient cation by a methyl group. No indication of a second isomer is present in the ms-TPE spectra of 2 and 3. While carbene 1 tends to dimerize to tetraethynylethene, 2 and 3 are more stable due to the additional methyl group and show no dimerization product.

In addition, the photoionization of the precursors 1-diazopenta-2,4-diyne 4, 1-diazo-hexa-2,4-diyne 5, and 2-diazo-hexa-3,5-diyne 6 has been explored. Ionization energies and vibrational frequencies have been determined. The dissociative photoionization was also investigated and breakdown diagrams have been recorded. However, computations indicate the presence of a reverse barrier for dissociation in the ion, so no bond dissociation energies could be experimentally determined.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.8b12244.

Further computational information on geometries of all molecules as well as additional spectra (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: bernd.engels@uni-wuerzburg.de.

\*E-mail: ingo.fischer@uni-wuerzburg.de.

\*E-mail: patrick.hemberger@psi.ch.

### ORCID

Bernd Engels: 0000-0003-3057-389X

Ingo Fischer: 0000-0001-8978-4013

Patrick Hemberger: 0000-0002-1251-4549

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft, Research Training School GRK2112 "Molecular Biradicals" and by Contract FIS75/13-1. We are also grateful for the financial support by the Swiss Federal Office for Energy (BFE Contract Number SI/501269-01). The experiments were conducted at the VUV beamline of the Swiss Light Source storage ring, Paul Scherrer Institute (PSI). We thank Max Baumann and Marius Gerlach for their experimental support. Some computations were carried out at the Linux-cluster of the Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften (LRZ).

2014

DOI: 10.1021/acs.jpca.8b12244  
J. Phys. Chem. A 2019, 123, 2008–2017

## REFERENCES

- (1) Lovas, F. J. NIST Recommended Rest Frequencies for Observed Interstellar Molecular Microwave Transitions - 2002 Revision. *J. Phys. Chem. Ref. Data* **2004**, *33*, 177–355.
- (2) Ridgway, S. T.; Hall, D. N. B.; Kleinmann, S. G.; Weinberger, D. A.; Wojslaw, R. S. Circumstellar Acetylene in the Infrared Spectrum of IRC + 10° 216. *Nature* **1976**, *264*, 345–346.
- (3) Kim, S. J.; Caldwell, J.; Rivolo, A. R.; Wagener, R.; Orton, G. S. Infrared Polar Brightening on Jupiter: III. Spectrometry from the Voyager 1 IRIS Experiment. *Icarus* **1985**, *64*, 233–248.
- (4) José, C.; Ana, M. H.; Tielens, A. G. G. M.; Juan, R. P.; Fabrice, H.; Michel, G.; Waters, L. B. F. M. Infrared Space Observatory's Discovery of C<sub>4</sub>H<sub>2</sub>, C<sub>6</sub>H<sub>2</sub>, and Benzene in CRL 618. *Astrophys. J.* **2001**, *546*, L123–L126.
- (5) Hebrard, E.; Dobrijevic, M.; Benilan, Y.; Raulin, F. Photochemical Kinetics Uncertainties in Modeling Titan's Atmosphere: A Review. *J. Photochem. Photobiol., C* **2006**, *7*, 211–230.
- (6) Boullart, W.; Devriendt, K.; Borms, R.; Peeters, J. Identification of the Sequence CH(Π) + C<sub>2</sub>H<sub>2</sub> → C<sub>3</sub>H<sub>2</sub> + H (and C<sub>3</sub>H + H<sub>2</sub>) Followed by C<sub>3</sub>H<sub>2</sub> + O → C<sub>2</sub>H + HCO (or H + CO) as C<sub>2</sub>H Source in C<sub>2</sub>H<sub>2</sub>/O/H Atomic Flames. *J. Phys. Chem.* **1996**, *100*, 998–1007.
- (7) Taatjes, C. A.; Klippenstein, S. J.; Hansen, N.; Miller, J. A.; Cool, T. A.; Wang, J.; Law, M. E.; Westmoreland, P. R. Synchrotron Photoionization Measurements of Combustion Intermediates: Photoionization Efficiency of C<sub>3</sub>H<sub>2</sub> Isomers. *Phys. Chem. Chem. Phys.* **2005**, *7*, 806–813.
- (8) Goroff, N. S. Mechanism of Fullerene Formation. *Acc. Chem. Res.* **1996**, *29*, 77–83.
- (9) Steinbauer, M.; Lang, M.; Fischer, I.; Cunha de Miranda, B. K.; Romanzin, C.; Alcaraz, C. The Photoionisation of Propargylene and Diazopropyne. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17956–17959.
- (10) Giegerich, J.; Petersen, J.; Mitric, R.; Fischer, I. Photo-dissociation Dynamics of Propargylene, HCCCH. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6294–6302.
- (11) Bowling, N. P.; Halter, R. J.; Hodges, J. A.; Seburg, R. A.; Thomas, P. S.; Simmons, C. S.; Stanton, J. F.; McMahon, R. J. Reactive Carbon-Chain Molecules: Synthesis of 1-Diazo-2,4-pentadiyne and Spectroscopic Characterization of Triplet Pentadiynylidene (H-C≡C-C≡C-H). *J. Am. Chem. Soc.* **2006**, *128*, 3291–3302.
- (12) Thomas, P. S.; Bowling, N. P.; McMahon, R. J. Spectroscopy and Photochemistry of Triplet Methylpentadiynylidene (Me-C≡C-C≡C-H). *J. Am. Chem. Soc.* **2009**, *131*, 8649–8659.
- (13) Thomas, P. S.; Bowling, N. P.; Burrmann, N. J.; McMahon, R. J. Dialkynyl Carbene Derivatives: Generation and Characterization of Triplet tert-Butylpentadiynylidene (t-Bu-C≡C-C≡C-H) and Dimethylpentadiynylidene (Me-C≡C-C≡C-Me). *J. Org. Chem.* **2010**, *75*, 6372–6381.
- (14) Seburg, R. A.; Patterson, E. V.; McMahon, R. J. Structure of Triplet Propynylidene (HCCCH) as Probed by IR, UV/VIS, and EPR Spectroscopy of Isotopomers. *J. Am. Chem. Soc.* **2009**, *131*, 9442–9455.
- (15) Bernheim, R. A.; Kempf, R. J.; Gramas, J. V.; Skell, P. S. Electron Paramagnetic Resonance of Triplet Alternant Methylenes. Propargylene and Homologs. *J. Chem. Phys.* **1965**, *43*, 196–200.
- (16) Bernheim, R. A.; Kempf, R. J.; Reichenbecher, E. F. Electron Paramagnetic Resonance of Triplet Alternant Methylenes. *J. Magn. Reson.* **1970**, *3*, 5–9.
- (17) Seburg, R. A.; McMahon, R. J.; Stanton, J. F.; Gauss, J. Structures and Stabilities of C<sub>3</sub>H<sub>2</sub> Isomers: Quantum Chemical Studies. *J. Am. Chem. Soc.* **1997**, *119*, 10838–10845.
- (18) Fan, Q.; Pfeiffer, G. V. Theoretical Study of Linear Polyatomic Carbon and Polyatomic Carbon Dihydrogen (C<sub>n</sub>(n = 6–10) and HC<sub>n</sub>H (n = 2–10)) Molecules. *Chem. Phys. Lett.* **1989**, *162*, 472–478.
- (19) Cooper, D. L.; Murphy, S. C. Ab Initio Geometries for C<sub>2n+1</sub>H, C<sub>2n+1</sub>H<sup>+</sup>, and C<sub>2n+1</sub>H<sub>2</sub> Species for n = 1, 2, 3. *Astrophys. J.* **1988**, *333*, 482–490.
- (20) Thimmakonda, V. S.; Karton, A. The Quest for the Carbene Bent-Pentadiynylidene Isomer of C<sub>3</sub>H<sub>2</sub>. *Chem. Phys.* **2018**, *515*, 411–417.
- (21) Gottlieb, C. A.; McCarthy, M. C.; Gordon, V. D.; Chakan, J. M.; Apponi, A. J.; Thaddeus, P. Laboratory Detection of Two New C<sub>3</sub>H<sub>2</sub> Isomers. *Astrophys. J.* **1998**, *509*, L141–L144.
- (22) McCarthy, M. C.; Travers, M. J.; Kovacs, A.; Chen, W.; Novick, S. E.; Gottlieb, C. A.; Thaddeus, P. Detection and Characterization of the Cumulene Carbenes H<sub>2</sub>C<sub>5</sub> and H<sub>2</sub>C<sub>6</sub>. *Science* **1997**, *275*, 518–520.
- (23) Travers, M. J.; McCarthy, M. C.; Gottlieb, C. A.; Thaddeus, P. Laboratory Detection of the Ring-Chain Molecule C<sub>3</sub>H<sub>2</sub>. *Astrophys. J.* **1997**, *483*, L135–L138.
- (24) Mavrandonakis, A.; Muehlhaeuser, M.; Froudakis, G. E.; Peyerimhoff, S. D. The Electronic Spectrum of Linear Pentadiynylidene in Comparison with Isomeric Ethynylcyclopropenylidene. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3318–3321.
- (25) Zhang, C.; Cao, Z.; Wu, H.; Zhang, Q. Linear and Nonlinear Feature of Electronic Excitation Energy in Carbon Chains HC<sub>2n+1</sub>H and HC<sub>2n</sub>H. *Int. J. Quantum Chem.* **2004**, *98*, 299–308.
- (26) Steglich, M.; Fulara, J.; Maity, S.; Nagy, A.; Maier, J. P. Electronic Spectra of Linear HC<sub>3</sub>H and Cumulene Carbene H<sub>2</sub>C<sub>5</sub>. *J. Chem. Phys.* **2015**, *142*, 244311.
- (27) Baer, T.; Tuckett, R. P. Advances in Threshold Photoelectron Spectroscopy (TPES) and Threshold Photoelectron Photoion Coincidence (TPEPICO). *Phys. Chem. Chem. Phys.* **2017**, *19*, 9698–9723.
- (28) Fischer, I.; Schufliker, T.; Deyerl, H.-J.; Elhanine, M.; Alcaraz, C. Photoionization and Dissociative Photoionization of the Allyl Radical, C<sub>3</sub>H<sub>5</sub>. *Int. J. Mass Spectrom.* **2007**, *261*, 227–233.
- (29) Hemberger, P.; Lang, M.; Noller, B.; Fischer, I.; Alcaraz, C.; Cunha de Miranda, B. K.; Garcia, G. A.; Soldi-Lose, H. Photoionization of Propargyl and Bromopropargyl Radicals: A Threshold Photoelectron Spectroscopic Study. *J. Phys. Chem. A* **2011**, *115*, 2225–2230.
- (30) Savee, J. D.; Zador, J.; Hemberger, P.; Sztaray, B.; Bodi, A.; Osborn, D. L. Threshold Photoelectron Spectrum of the Benzyl Radical. *Mol. Phys.* **2015**, *113*, 2217–2227.
- (31) Steinbauer, M.; Hemberger, P.; Fischer, I.; Bodi, A. Photoionization of C<sub>7</sub>H<sub>6</sub> and C<sub>7</sub>H<sub>5</sub>: Observation of the Fulvenallenyl Radical. *ChemPhysChem* **2011**, *12*, 1795–1797.
- (32) Hemberger, P.; Noller, B.; Steinbauer, M.; Fischer, I.; Alcaraz, C.; Cunha de Miranda, B. K.; Garcia, G. A.; Soldi-Lose, H. Threshold Photoelectron Spectroscopy of Cyclopropenylidene, Chlorocyclopropenylidene, and Their Deuterated Isotopomers. *J. Phys. Chem. A* **2010**, *114*, 11269–11276.
- (33) Zhu, Y.; Wu, X.; Tang, X.; Wen, Z.; Liu, F.; Zhou, X.; Zhang, W. Synchrotron Threshold Photoelectron Photoion Coincidence Spectroscopy of Radicals Produced in a Pyrolysis Source: The Methyl Radical. *Chem. Phys. Lett.* **2016**, *664*, 237–241.
- (34) Coudert, L. H.; Gans, B.; Holzmeier, F.; Loison, J. C.; Garcia, G. A.; Alcaraz, C.; Lopes, A.; Roder, A. Experimental and Theoretical Threshold Photoelectron Spectra of Methylene. *J. Chem. Phys.* **2018**, *149*, 224304.
- (35) Gans, B.; Holzmeier, F.; Kruger, J.; Falvo, C.; Roder, A.; Lopes, A.; Garcia, G. A.; Fittschen, C.; Loison, J.-C.; Alcaraz, C. Synchrotron-Based Valence Shell Photoionization of CH Radical. *J. Chem. Phys.* **2016**, *144*, 204307.
- (36) Dossmann, H.; Garcia, G. A.; Nahon, L.; de Miranda, B. K. C.; Alcaraz, C. Comprehensive Vacuum Ultraviolet Photoionization Study of the CF<sub>3</sub><sup>•</sup> Trifluoromethyl Radical Using Synchrotron Radiation. *J. Chem. Phys.* **2012**, *136*, 204304.
- (37) Garcia, G. A.; Gans, B.; Kruger, J.; Holzmeier, F.; Roder, A.; Lopes, A.; Fittschen, C.; Alcaraz, C.; Loison, J.-C. Valence Shell Threshold Photoelectron Spectroscopy of C<sub>3</sub>H<sub>x</sub> (x = 0–3). *Phys. Chem. Chem. Phys.* **2018**, *20*, 8707–8718.
- (38) Schleier, D.; Humeniuk, A.; Reusch, E.; Holzmeier, F.; Nunez-Reyes, D.; Alcaraz, C.; Garcia, G. A.; Loison, J.-C.; Fischer, I.; Mitric,

- R. Diborene: Generation and Photoelectron Spectroscopy of an Inorganic Biradical. *J. Phys. Chem. Lett.* **2018**, *9*, 5921–5925.
- (39) Reusch, E.; Holzmeier, F.; Constantinidis, P.; Hemberger, P.; Fischer, I. Isomer-Selective Generation and Spectroscopic Characterization of Picolyl Radicals. *Angew. Chem., Int. Ed.* **2017**, *56*, 8000–8003.
- (40) Hemberger, P.; Trevitt, A. J.; Ross, E.; da Silva, G. Direct Observation of Para-Xylylene as the Decomposition Product of the Meta-Xylyl Radical Using VUV Synchrotron Radiation. *J. Phys. Chem. Lett.* **2013**, *4*, 2546–2550.
- (41) Hemberger, P.; Trevitt, A. J.; Gerber, T.; Ross, E.; da Silva, G. Isomer-Specific Product Detection of Gas-Phase Xylyl Radical Rearrangement and Decomposition Using VUV Synchrotron Photoionization. *J. Phys. Chem. A* **2014**, *118*, 3593–3604.
- (42) Steglich, M.; Custodis, V. B. F.; Trevitt, A. J.; daSilva, G.; Bodi, A.; Hemberger, P. Photoelectron Spectrum and Energetics of the Meta-Xylylene Diradical. *J. Am. Chem. Soc.* **2017**, *139*, 14348–14351.
- (43) Taatjes, C. A.; Hansen, N.; Osborn, D. L.; Kohse-Hoinghaus, K.; Cool, T. A.; Westmoreland, P. R. Imaging<sup>2</sup> Combustion Chemistry via Multiplexed Synchrotron-Photoionization Mass Spectrometry. *Phys. Chem. Chem. Phys.* **2008**, *10*, 20–34.
- (44) Oßwald, P.; Hemberger, P.; Bierkandt, T.; Akyildiz, E.; Kohler, M.; Bodi, A.; Gerber, T.; Kasper, T. In Situ Flame Chemistry Tracing by Imaging Photoelectron Photoion Coincidence Spectroscopy. *Rev. Sci. Instrum.* **2014**, *85*, 025101.
- (45) Kruger, J.; Garcia, G. A.; Felsmann, D.; Moshhammer, K.; Lackner, A.; Brockhinke, A.; Nahon, L.; Kohse-Hoinghaus, K. Photoelectron Photoion Coincidence Spectroscopy for Multiplexed Detection of Intermediate Species in a Flame. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22791–22804.
- (46) Hemberger, P.; Custodis, V. B. F.; Bodi, A.; Gerber, T.; van Bokhoven, J. A. Understanding the Mechanism of Catalytic fast Pyrolysis by Unveiling Reactive Intermediates in Heterogeneous Catalysis. *Nat. Commun.* **2017**, *8*, 15946.
- (47) Paunović, V.; Hemberger, P.; Bodi, A.; López, N.; Pérez-Ramírez, J. Evidence of Radical Chemistry in Catalytic Methane Oxidation. *Nat. Catal.* **2018**, *1*, 363–370.
- (48) Osborn, D. L.; Zou, P.; Johnsen, H.; Hayden, C. C.; Taatjes, C. A.; Knyazev, V. D.; North, S. W.; Peterka, D. S.; Ahmed, M.; Leone, S. R. The Multiplexed Chemical Kinetic Photoionization Mass Spectrometer: A new Approach to Isomer-Resolved Chemical Kinetics. *Rev. Sci. Instrum.* **2008**, *79*, 104103.
- (49) Schleier, D.; Constantinidis, P.; Faßheber, N.; Fischer, I.; Friedrichs, G.; Hemberger, P.; Reusch, E.; Sztaray, B.; Voronova, K. Kinetics of the  $a\text{-C}_3\text{H}_5 + \text{O}_2$  Reaction, Investigated by Photoionization Using Synchrotron Radiation. *Phys. Chem. Chem. Phys.* **2018**, *20*, 10721–10731.
- (50) Johnson, M.; Bodi, A.; Schulz, L.; Gerber, T. Vacuum Ultraviolet Beamline at the Swiss Light Source for Chemical Dynamics Studies. *Nucl. Instrum. Methods Phys. Res., Sect. A* **2009**, *610*, 597–603.
- (51) Sztaray, B.; Voronova, K.; Torma, K. G.; Covert, K. J.; Bodi, A.; Hemberger, P.; Gerber, T.; Osborn, D. L. CRF-PEPICO: Double Velocity Map Imaging Photoelectron Photoion Coincidence Spectroscopy for Reaction Kinetics Studies. *J. Chem. Phys.* **2017**, *147*, 013944.
- (52) Bodi, A.; Hemberger, P.; Gerber, T.; Sztaray, B. A New Double Imaging Velocity Focusing Coincidence Experiment:  $i^2$ PEPICO. *Rev. Sci. Instrum.* **2012**, *83*, 083105.
- (53) Bodi, A.; Johnson, M.; Gerber, T.; Gengeliczki, Z.; Sztaray, B.; Baer, T. Imaging Photoelectron Photoion Coincidence Spectroscopy with Velocity Focusing Electron Optics. *Rev. Sci. Instrum.* **2009**, *80*, 034101.
- (54) Bodi, A.; Sztaray, B.; Baer, T.; Johnson, M.; Gerber, T. Data Acquisition Schemes for Continuous Two-Particle Time-of-Flight Coincidence Experiments. *Rev. Sci. Instrum.* **2007**, *78*, 084102.
- (55) Sztaray, B.; Baer, T. Suppression of hot Electrons in Threshold Photoelectron Photoion Coincidence Spectroscopy using Velocity Focusing Optics. *Rev. Sci. Instrum.* **2003**, *74*, 3763–3768.
- (56) Kohn, D. W.; Clauberg, H.; Chen, P. Flash Pyrolysis Nozzle for Generation of Radicals in a Supersonic Jet Expansion. *Rev. Sci. Instrum.* **1992**, *63*, 4003–4005.
- (57) Bowling, N. P.; Burrmann, N. J.; Halter, R. J.; Hodges, J. A.; McMahan, R. J. Synthesis of Simple Diynals, Diynones, Their Hydrazones, and Diazo Compounds: Precursors to a Family of Dialkynyl Carbenes (R1-C≡C-C-C≡C-R2). *J. Org. Chem.* **2010**, *75*, 6382–6390.
- (58) Chai, J. D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
- (59) 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.
- (60) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.
- (61) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two new Functionals and Systematic Testing of four M06-Class Functionals and 12 other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (62) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A Complete Basis Set Model Chemistry. VI. Use of Density Functional Geometries and Frequencies. *J. Chem. Phys.* **1999**, *110*, 2822–2827.
- (63) Frisch, M. J.; et al. *Gaussian 16*, revision B.01; Gaussian, Inc.: Wallingford, CT, 2016.
- (64) Andersson, K.; Malmqvist, P. Å.; Roos, B. O. Second-Order Perturbation Theory with a Complete Active Space Self-Consistent Field Reference Function. *J. Chem. Phys.* **1992**, *96*, 1218–1226.
- (65) Pulay, P. A Perspective on the CASPT2 Method. *Int. J. Quantum Chem.* **2011**, *111*, 3273–3279.
- (66) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P. Å.; Neogrády, P.; Pedersen, T. B.; Pitoňák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Velyazov, V.; Lindh, R. MOLCAS 7: The Next Generation. *J. Comput. Chem.* **2010**, *31*, 224–247.
- (67) Lee, T. J.; Taylor, P. R. A Diagnostic for Determining the Quality of Single-Reference Electron Correlation Methods. *Int. J. Quantum Chem.* **1989**, *36*, 199–207.
- (68) Spangenberg, D.; Imhof, P.; Kleinermanns, K. The  $S_1$  State Geometry of Phenol determined by Simultaneous Franck-Condon and Rotational Constants Fits. *Phys. Chem. Chem. Phys.* **2003**, *5*, 2505–2514.
- (69) Mozhayskiy, V. A.; Krylov, A. I. ezSpectrum v.3.0, 2009, <http://iopshell.usc.edu/downloads>.
- (70) Bowling, N. P.; McMahon, R. J. Eneidyne Isomers of Tetraethynylethene. *J. Org. Chem.* **2006**, *71*, 5841–5847.
- (71) Engels, B. Estimation of the Influence of the Configurations Neglected within Truncated Multi-Reference CI Wavefunctions on Molecular Properties. *Chem. Phys. Lett.* **1991**, *179*, 398–404.
- (72) Pless, V.; Suter, H. U.; Engels, B. Ab Initio Study of the Energy Difference between the Cyclic and Linear Forms of the  $C_6$  Molecule. *J. Chem. Phys.* **1994**, *101*, 4042–4048.
- (73) Kaiser, D.; Reusch, E.; Hemberger, P.; Bodi, A.; Welz, E.; Engels, B.; Fischer, I. The Ortho-Benzynes Cation is Not Planar. *Phys. Chem. Chem. Phys.* **2018**, *20*, 3988–3996.
- (74) Cong-Jie, Z.; Li-Ling, Z.; Ze-Xing, C.; Qian-Er, Z. Isomerization Mechanisms of  $C_4H_2$  on the Triplet and Singlet Potential Energy Surfaces. *Chin. J. Chem.* **2003**, *21*, 117–120.
- (75) Ding, Y.-H.; Liu, J.-L.; Huang, X.-R.; Li, Z.-S.; Sun, C.-C.  $C_4N$ : The first  $C_4N$  Radical with Stable Cyclic Isomers. *J. Chem. Phys.* **2001**, *114*, 5170–5179.
- (76) McCarthy, M. C.; Fuchs, G. W.; Kucera, J.; Winnewisser, G.; Thaddeus, P. Rotational Spectra of  $C_3N$ ,  $C_6N$ , and the Isotopic Species of  $C_3N$ . *J. Chem. Phys.* **2003**, *118*, 3549–3557.

(77) Brogli, F.; Heilbronner, E.; Hornung, V.; Kloster-Jensen, E. Die Photoelektronen-Spektren Methyl-Substituierter Acetylene. *Helv. Chim. Acta* 1973, 56, 2171–2178.

2017

DOI: 10.1021/acs.jpca.8b12244  
*J. Phys. Chem. A* 2019, 123, 2008–2017



## 5. Photochemistry of the Diphenylacetylene Biradical



PCCP

PAPER

View Article Online  
View Journal | View IssueCite this: *Phys. Chem. Chem. Phys.*,  
2019, 21, 13157Received 19th April 2019,  
Accepted 31st May 2019

DOI: 10.1039/c9cp02222h

rsc.li/pccp

## A time-resolved photoelectron imaging study on isolated tolane: observation of the biradicalic $^1A_u$ state†

Marco Flock, Lea Bosse, Dustin Kaiser, Bernd Engels and Ingo Fischer \*

Tolane (diphenylacetylene,  $C_{14}H_{10}$ ) was studied by picosecond time resolved photoionisation and photoelectron imaging in a supersonic jet. The low energy part of the  $^1B_{1u} \leftarrow S_0$  REMPI spectrum is very similar to an earlier LIF spectrum, however additional bands were observed at higher energies above the fluorescence cut off. For a number of bands the dynamics were investigated *via* pump-probe photoionisation and photoelectron spectroscopy. Around the  $^1B_{1u}$  origin the lifetimes are in the ns range, but they drop to some 10 ps at higher excitation energies. For the short lived bands at higher energies a sequential two step relaxation to a long lived electronic state was observed proceeding *via* an intermediately populated state with a lifetime of 100–200 ps. By comparison with previous quantum chemical calculations we assign this state as the biradicalic *trans-bent*  $^1A_u$  state that is ionised in a two-photon process.

### 1 Introduction

In this paper we present an experimental analysis of the non-radiative processes in tolane (diphenylacetylene,  $C_{14}H_{10}$ ), a molecule that is employed as a building block for conjugated polymers<sup>1–3</sup> or dendrimers.<sup>4,5</sup> Such systems are promising materials for organic devices like molecular switches<sup>6,7</sup> or light harvesting antennae.<sup>8,9</sup> Therefore, the photophysical properties of tolane are of great interest but they are not yet fully understood. In particular, no time-resolved gas phase studies on tolane are available that can be directly compared with high level computations. The molecule and the coordinate system employed in the current study are depicted in Fig. 1. First jet-cooled fluorescence excitation (LIF) and multi-photon ionisation (REMPI) studies assumed that tolane possesses three close-lying excited electronic states.<sup>10,11</sup> One-photon excitation to the lowest excited  $^1A_g$  state is symmetry forbidden, but the origin of this state was determined to be  $34\,960\text{ cm}^{-1}$  *via* [2+2] two-photon resonant four-photon ionisation.<sup>10</sup> LIF measurements indicated the existence of two more excited states with  $^1B_{1u}$  and  $^1B_{2u}$  symmetry, located at  $35\,248\text{ cm}^{-1}$  and below  $35\,051\text{ cm}^{-1}$ , respectively.<sup>10</sup> In a more recent high-resolution fluorescence excitation study, vibronic bands were assigned to only two electronic states,  $^1A_g$  and  $^1B_{1u}$ .<sup>11</sup> Bands previously assigned to the  $B_{2u}$  electronic state were proposed to be  $b_{1u}$  vibrational modes of the dark  $^1A_g$  state gaining optical activity *via* vibronic coupling with the  $^1B_{1u}$  state.<sup>11</sup> However, in both studies only bands up to around  $700\text{ cm}^{-1}$  above the  $^1B_{1u}$  origin

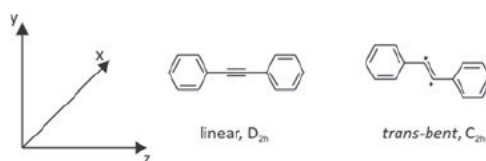


Fig. 1 Molecular structure of tolane in linear ( $D_{2h}$ ) and *trans-bent* ( $C_{2h}$ ) geometry. The orientation of the coordinate system is chosen in agreement with IUPAC recommendations.<sup>38</sup>

were observed, because the fluorescence quantum yield  $\eta_f$  drops significantly at higher excitation energies.<sup>10,11</sup>

This sudden drop of  $\eta_f$  motivated numerous experimental<sup>12–29</sup> and theoretical<sup>29–34</sup> studies on the photophysics of tolane. Temperature-dependent transient absorption (TA) in various solvents yielded a ns lifetime for the  $^1B_{1u}$  state at 123 K, which decreased to around 8 ps at room temperature.<sup>16</sup> A second transient was observed, rising with a time constant comparable to this decay time, indicating a sequential relaxation. The lifetime of this intermediately populated state was determined to be around 200 ps, a deactivation to a long-lived triplet state was suggested.<sup>16</sup> Time-resolved IR<sup>25,26</sup> and Raman<sup>17,21,23,26</sup> studies revealed insights into the excited states geometry changes during the relaxation process. In particular, the vibrational frequency of the central C–C bond was of interest. In the electronic ground state  $S_0$  tolane shows a planar  $D_{2h}$  structure<sup>35,36</sup> (see Fig. 1, left structure) and the central C–C stretch band appears at  $2223\text{ cm}^{-1}$ .<sup>36,37</sup> In the  $^1B_{1u}$  state the wavenumber decreases to  $2099\text{ cm}^{-1}$ , as concluded from time-resolved CARS spectra.<sup>21,23</sup> At long delay times ( $>100\text{ ns}$ ),

*Institute of Physical and Theoretical Chemistry, University of Wuerzburg, 97074 Wuerzburg, Germany. E-mail: ingo.fischer@uni-wuerzburg.de*  
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9cp02222h

another transient appears at  $1970\text{ cm}^{-1}$  that was assigned to the triplet state.<sup>17</sup> Thus the C–C bond strength is lowered in both excited states, but is still in the region of a C–C triple bond. Interestingly no signals from the unidentified intermediate state were found in the C–C stretch region, but transient peaks with lifetimes of 200 ps were observed in the spectral region between  $1550\text{ cm}^{-1}$  and  $1580\text{ cm}^{-1}$ .<sup>21,23</sup> It was therefore concluded that this pronounced shift to lower wavenumbers indicates the presence of a C–C double bond and thus a stilbene like biradicalic *trans bent* structure with  $C_{2h}$  symmetry was suggested (Fig. 1, right hand structure).

A comprehensive theoretical analysis was recently published by Krämer *et al.*<sup>33</sup> Computations revealed an excited singlet state of  $^1A_u$  symmetry showing a *trans bent* minimum energy geometry.<sup>31–34</sup> The potential of this state along the *trans bent* motion crosses the  $^1B_{1u}$  potential and can thus be populated *via* fast internal conversion (IC) after photoexcitation of  $^1B_{1u}$ .<sup>34</sup> This deactivation pathway explains the abrupt drop of fluorescence<sup>10,11</sup> and the strong temperature dependence<sup>16</sup> of the  $^1B_{1u}$  lifetime, because an activation barrier has to be surmounted to reach the conical intersection.<sup>31</sup> Although there is a large body of work on the photophysics in solution as discussed above, no time resolved studies in the gas phase have been carried out so far. It was the goal of the present work to experimentally analyse the role of the *trans bent*  $^1A_u$  state in isolated tolane after  $^1B_{1u}$  excitation and enable a direct comparison with theory. Since a ps laser is employed, zero order vibronic states are excited and their dynamics are studied as a function of excitation energy.<sup>39–42</sup> Photoionisation and photoelectron spectroscopy are chosen as probe methods, because information on the electronic states involved in the relaxation processes can be extracted from the photoelectron kinetic energy as has been discussed in several review articles.<sup>43–45</sup>

## 2 Experimental and computational methods

Diphenylacetylene (tolane) was purchased from Sigma Aldrich and used without further purification. Experiments were conducted in a differentially pumped vacuum chamber. The sample was heated up to  $70\text{ }^\circ\text{C}$  in a home built oven attached to a solenoid pulsed valve, seeded in argon as a carrier gas ( $p_0(\text{Ar}) = 1.2\text{--}1.4\text{ bar}$ ) and expanded into the source chamber through a  $0.3\text{ mm}$  diameter nozzle. After around  $2\text{ cm}$ , the generated supersonic jet passed a conical skimmer with an aperture of  $2\text{ mm}$  and reached the main chamber. For excitation and ionisation, we employed a  $10\text{ Hz}$  Nd:YLF based picosecond laser system from Ekspla, described earlier.<sup>46</sup> The fundamental wavelength of  $1053\text{ nm}$  was frequency tripled and 85% of the third harmonic ( $4\text{--}5\text{ mJ}$ ) were used to produce tuneable pump pulses ( $40\text{--}100\text{ }\mu\text{J}$ ) in an optical parametric generator (OPG). The residual part of the third harmonic ( $351\text{ nm}$ ,  $400\text{--}500\text{ }\mu\text{J}$ ) was sent over a motorised delay stage and used as the probe pulse. Alternatively, probe pulses of  $263.5\text{ nm}$  ( $25\text{--}30\text{ }\mu\text{J}$ ) were obtained *via* frequency doubling of the second harmonic ( $527\text{ nm}$ ,  $0.9\text{--}1.0\text{ mJ}$ ) in a BBO

crystal. The instrument response function (IRF) was found to be around  $4\text{ ps}$  and the bandwidth of the pulses was on the order of  $20\text{ cm}^{-1}$ . The vertically polarised pump and probe pulses were overlapped on a dichroic mirror and slightly focused into the molecular supersonic jet. Rotation of the polarization did not change any of the results reported below. Ions or electrons were accelerated in the charged particle optics of a longitudinal velocity map imaging (VMI) spectrometer<sup>47</sup> and passed a  $50\text{ cm}$  field free flight tube shielded with a  $1.5\text{ mm}$  thick  $\mu$  metal tube before being registered in a VMI detector, consisting of two microchannel plates (Chevron assembly) and a P43 phosphor screen. Ion signals were detected in a home built current monitor. For the REMPI spectra the OPG was scanned in  $0.1\text{ nm}$  steps and each data point was typically averaged over 50 laser shots. In photoelectron VMI experiments, the light emitted from the phosphor screen was imaged onto a progressive scan camera with a  $2/3''$  CCD chip by an achromatic object lens. To obtain a better signal to noise ratio, the applied voltage at the second microchannel plate of the detector was gated by a push/pull switch with a gate width of  $150\text{ ns}$ . The exposure time of the camera was additionally triggered with a gate width of  $1\text{ ms}$ . Each 2D raw image was averaged over up to 5800 laser shots and reconstructed *via* the pBASEX<sup>48</sup> algorithm, employing Legendre polynomials up to the second order. Two digital delay generators were employed for temporal synchronization.

All quantum chemical calculations were performed using the Gaussian 16 program package.<sup>49</sup> Where necessary the orientation of the molecule was chosen based on the IUPAC recommendation<sup>38</sup> (see Fig. 1). The  $z$  axis is directed along the central C–C triple bond and the  $x$  axis is perpendicular to the molecular plane. The geometry of the linear  $D_{2h}$  symmetric structure of tolane was optimised by density functional theory (DFT) employing the  $\omega$ B97xD functional<sup>50</sup> combined with the cc-pVDZ basis set.<sup>51</sup> Electronically excited states were calculated using TD DFT with the same functional and basis set. For the calculation of vibrational frequencies, the excited state structure was subsequently optimised. DFT calculations concerning the  $C_{2h}$  *trans bent* geometry were performed on the (u) $\omega$ B97xD/def2-TZVPP level of theory.<sup>50,52</sup> The optimised structures have been confirmed to be true minima by frequency calculations. CASPT2<sup>53</sup> calculations using Dunning's cc-pVTZ basis set<sup>54</sup> were carried out using MOLCAS7 in  $C_{2h}$  symmetry.<sup>55</sup> The (6,6) active space was built by considering the canonical Hartree–Fock orbitals predicted to contribute most to ionisation and excitation processes by prior SCS ADC(2)/def2-TZVPP calculations.

## 3 Results

### A REMPI experiments

As mentioned above tolane exhibits at least two excited states ( $^1A_g$  and  $^1B_{1u}$ ) in the region of  $4.3\text{--}4.4\text{ eV}$ . While the  $^1B_{1u}$  state is optically active, the totally symmetric doubly excited  $^1A_g$  state is one photon forbidden. Nevertheless, several  $b_{1u}$  vibrational modes become active by vibronic coupling with the  $^1B_{1u}$  state.<sup>11</sup> In a first step we recorded [1+1] REMPI spectra (Fig. 2, black) to compare

PCCP

View Article Online

Paper

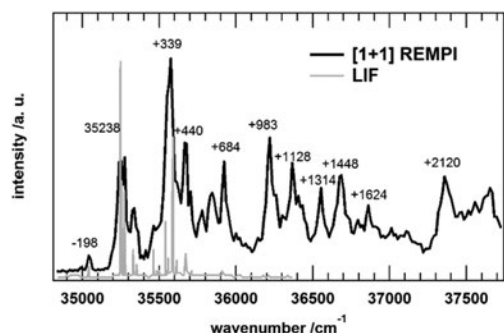


Fig. 2 Picosecond [1+1] REMPI spectrum (black) of toluene in the region of the lowest excited singlet states and fluorescence excitation spectrum (LIF, grey) recorded by Okuyama *et al.*<sup>10</sup> The origin of the  $^1B_{1u}$  transition was assigned to the band at 35 238  $\text{cm}^{-1}$ .

them with the earlier LIF spectra of Okuyama *et al.* obtained with a ns laser (Fig. 2, grey).<sup>10</sup> The authors also reported that the [1+1] REMPI spectrum is identical to their LIF spectrum. As an ionisation energy (IE) of 7.9243 eV has been determined for toluene,<sup>18</sup> two photons are sufficient for ionisation. The most intense bands are summarised in Table 1. Based on quantum chemical calculations, higher energy bands can be tentatively assigned to totally symmetric vibrational modes of the  $^1B_{1u}$  state showing mainly activity on the phenyl rings. As computations reveal no low frequency  $a_g$  modes of the  $^1B_{1u}$  state, those bands might arise from vibronic coupling of the excited  $^1A_g$  and  $^1B_{1u}$  states. Note that all transitions are accompanied by combination bands with the phenyl torsion.<sup>18</sup>

Due to the smaller spectral bandwidth ( $1 \text{ cm}^{-1}$ ) compared to our ps laser setup (*ca.* 20  $\text{cm}^{-1}$ ) the LIF spectrum shows a better resolution and narrower bands than the REMPI spectrum, which cannot resolve the torsional structure. The centre of the first band at 35 238  $\text{cm}^{-1}$  is in good agreement with the reported value of 35 248  $\text{cm}^{-1}$  for the  $^1B_{1u}$  origin.<sup>10</sup> The most remarkable difference between the two spectra is the presence of intense transitions above 35 600  $\text{cm}^{-1}$ , which are absent in

the LIF spectrum. This fluorescence cut off indicates the onset of a non radiative decay channel and a decrease of the excited state lifetime.

In order to investigate the dark  $^1A_g$  state, corresponding mainly to a (HOMO - 1)  $\rightarrow$  LUMO transition,<sup>11</sup> we also recorded [2+2] REMPI spectra and reproduced the spectra recorded by Okuyama *et al.*<sup>10</sup> However, our studies revealed no additional information on this electronic state. The spectrum is therefore given only in the ESI,† Fig. S1.

### B Time resolved measurements

To study the lifetimes of the observed one photon active bands and the underlying dynamics, we excited toluene to several of the bands observed in the REMPI spectrum and ionised them in a [1+1'] REMPI process with a 263.5 nm probe pulse. Recording the photoion signal at different pump-probe delay times yielded time resolved decay traces, a selection depicted in Fig. 3. The traces were fitted by a function including a Gaussian shaped instrument response function (IRF, around 4 ps) and an exponential decay. Bands located up to 700  $\text{cm}^{-1}$  above the  $^1B_{1u}$  origin showed a lifetime on the ns timescale (see Fig. 3, upper trace) in agreement with the high fluorescence quantum yields reported earlier.<sup>10,11</sup> Excitation of bands with excess energies above +700  $\text{cm}^{-1}$  leads to a rapid decrease of lifetime to 162 ps at +983  $\text{cm}^{-1}$  and finally to 35 ps at +1448  $\text{cm}^{-1}$ . The observed lifetimes are also summarised in Table 1. Interestingly, all traces decay to zero monoexponentially, whereas a number of time resolved studies in solution observed a two step decay.<sup>16,21,23,25,26</sup>

A possible reason for a monoexponential decay is a probe energy that is insufficient to ionise the final electronic state. To elucidate possible further steps of the dynamics, we therefore switched to a probe wavelength of 351 nm (third harmonic of Nd:YLF laser), which provides significantly higher pulse energies and thus a higher probability for multiphoton ionisation in the probe step. At the  $^1B_{1u}$  origin, the total energy of a [1+1'] REMPI process (7.90 eV) is only slightly below the IE (7.92 eV<sup>18</sup>), but due to the Stark shift of the IE in the 1000  $\text{V cm}^{-1}$  field, ionisation with only one probe photon is nevertheless possible. Decay traces for selected bands are given in Fig. 4.

Table 1 Summary of the most intense vibrations of isolated toluene and corresponding lifetimes obtained in pump-probe experiments employing 263.5 and 351 nm probe pulses, respectively. Mode assignments are based on computed totally symmetric ( $a_g$ ) vibrational modes of the  $^1B_{1u}$  state scaled with a factor of 0.971 ( $\omega_{B97xD/cc-pVDZ}$ )

Vib. energy ( $\text{cm}^{-1}$ )	Calc. values ( $\text{cm}^{-1}$ )	Tentative assignment	$\tau$ (ps) ( $\lambda_{\text{probe}} = 263.5 \text{ nm}$ )	$\tau_1/\tau_2$ (ps) ( $\lambda_{\text{probe}} = 351 \text{ nm}$ )
-198			> 1 ns	> 1 ns
35 238		$^1B_{1u}$ origin	> 1 ns	> 1 ns
+339			> 1 ns	> 1 ns
+440			> 1 ns	> 1 ns
+684	+684	$\nu_{12}$ ring breathing	> 1 ns	> 1 ns
Cut-off of fluorescence <sup>10,11</sup>				
+983	+968	$\nu_{10}$ ring deform.	162 $\pm$ 10	163 $\pm$ 10
+1128	+1125	$\nu_8$ ring deform.	65 $\pm$ 5	70 $\pm$ 10
				225 $\pm$ 25
+1314	+1322	339 $\text{cm}^{-1}$ + $\nu_{10}$	46 $\pm$ 5	37 $\pm$ 5/188 $\pm$ 20
+1448	+1443	$\nu_6$ ring deform.	35 $\pm$ 5	32 $\pm$ 5/121 $\pm$ 20
+2120	+2119	$\nu_4$ C $\equiv$ C-stretch		18 $\pm$ 2/133 $\pm$ 15
+3106	+3103	$\nu_4 + \nu_{10}$		10 $\pm$ 2/153 $\pm$ 15

Paper

View Article Online

PCCP

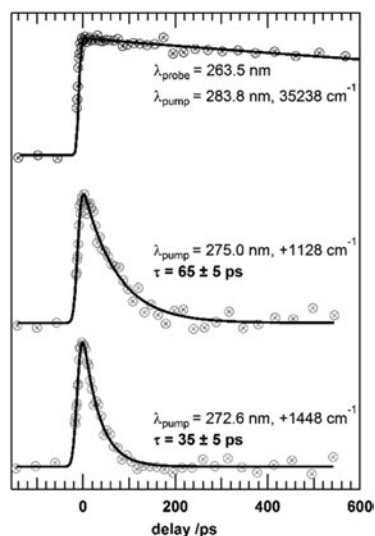


Fig. 3 Decay traces at different excitation energies employing a probe wavelength of 263.5 nm. The lifetime drops monoexponentially at higher excitation energies.

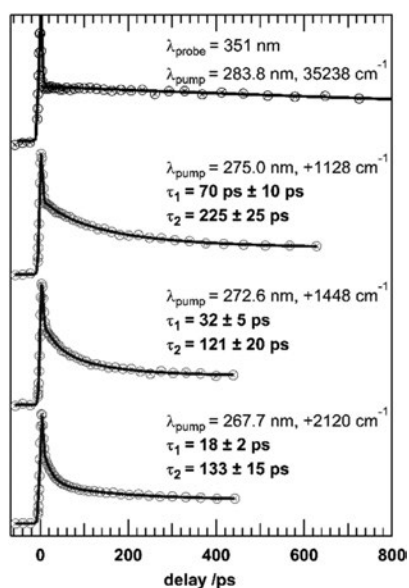


Fig. 4 Decay traces at different excitation energies employing a probe wavelength of 351 nm. The lifetime drops at higher excitation and a biexponential decay with a constant signal offset at long delay times is observed.

Low energy bands show the same time dependence as the traces recorded with 263.5 nm probe pulses, apart from an additional contribution at time zero that decays within the

pulse duration. At higher excitation energies, differences between both probe wavelengths become visible. First, the signal does not drop to zero but a constant signal offset remains, which indicates the population of a long lived state. Second, a biexponential fit function is required to describe the time dependence. The first time constant  $\tau_1$  strongly depends on the excitation wavelength and is in excellent agreement with the values obtained with 263.5 nm probe pulses, as shown in Table 1. The second time constant  $\tau_2$  however lies between 100 and 200 ps and depends only weakly on the excitation energy. The time constants agree well with those recorded in previous solution phase studies<sup>16</sup> and the biexponential decay is in accordance with a sequential relaxation process.<sup>16,21,23</sup>

### C Time resolved photoelectron imaging

To gain further information on the nature of the electronic states involved in the relaxation process, we recorded time resolved photoelectron images (TR PEI) as a function of time delay. Fig. 5 shows the transient photoelectron map recorded at  $\lambda_{\text{pump}} = 267.7 \text{ nm}$  (+2120  $\text{cm}^{-1}$ ), corresponding to excitation of the  ${}^1\text{B}_{1u}$  C–C stretch mode, and  $\lambda_{\text{probe}} = 351 \text{ nm}$ . The most important features are better recognised in the photoelectron spectra shown in the lower part of Fig. 5, which are averaged over a range of delays. An unstructured signal at higher

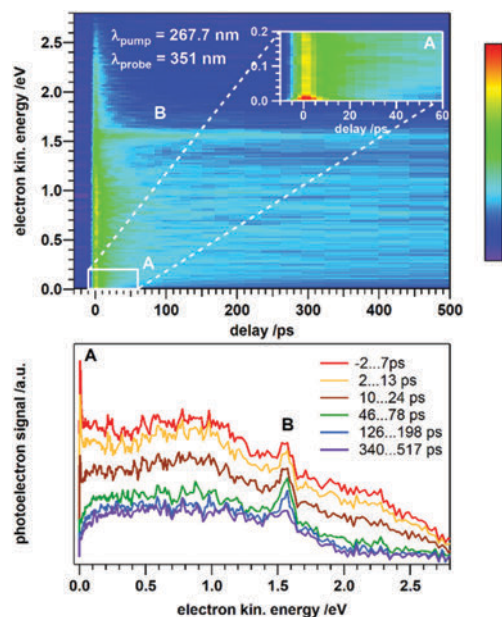


Fig. 5 Transient photoelectron map at an excitation wavelength of 267.7 nm ( ${}^1\text{B}_{1u}$  origin +2120  $\text{cm}^{-1}$ ) employing a probe wavelength of 351 nm. One-colour signals before time zero were subtracted in order to obtain the pure transient signal. The lower trace shows photoelectron spectra at different delay intervals. Two peaks A and B with different temporal behaviour can be resolved.



energies up to 3 eV is present and decays with time. In addition, two prominent peaks appear in the spectra: an intense transient at electron kinetic energies (eKE) close to 0 eV (peak A in Fig. 5) is present at time zero. It is due to threshold electrons from [1+1'] ionisation of the initially excited  $^1B_{1u}$  state. This signal decays quickly within roughly 20 ps. On the same time scale a peak at 1.57 eV gains intensity (peak B in Fig. 5) and maximises at around 60 ps (green trace). Its eKE shows that it originates from a [1+2'] process. The peak decreases at long delay times with  $\tau_2$  and an unstructured long lived transient (purple trace) remains. Its temporal profile shows that it originates from ionisation of a transient state, populated from  $^1B_{1u}$  state within  $\tau_1$ . Note that TR PEI with 263.5 nm probe pulses only showed the first decay associated with  $\tau_1$  and thus provided no information beyond the photoion (PI) signal, see Fig. S2 and S3 in the ESI†

We performed a global fit using the Glotaran<sup>56</sup> software, assuming a sequential two step relaxation to a long lived state, in agreement with previous solution phase studies.<sup>16,21,23,25,26</sup> This fit yields time constants of  $18 \pm 1$  ( $\tau_1$ ) and  $121 \pm 6$  ps ( $\tau_2$ ) in excellent agreement with the values obtained from time resolved REMPI measurements (18 and 133 ps, see Fig. 4, bottom trace). Furthermore, the global fit reveals the evolution associated spectra (EAS) depicted in Fig. 6. In case of a sequential relaxation scheme with increasing lifetimes ( $\tau_1 < \tau_2 < \dots < \tau_n$ ) the EAS can be directly interpreted as the spectrum of the intermediate species.<sup>57</sup> As depicted in Fig. 6, the EAS shows the two distinct peaks A at around 0 eV, originating from [1+1'] ionisation of the initially excited  $^1B_{1u}$  state and B at 1.57 eV, representing ionisation from the intermediate state populated by non radiative deactivation of the  $^1B_{1u}$  state. Analysis of the photoelectron angular distribution provides additional information on the two photon probe process, see Fig. S4 (ESI†). Peak B exhibits a pronounced positive anisotropy of  $\beta = +0.84$ . In this plot of  $\beta$  vs. eKE two further peaks with a similar  $\beta$  appear that are difficult to discern in the photoelectron spectra. In the spectrum recorded

at the  $^1B_{1u}$  origin peak B is absent (Fig. S5, ESI†), confirming that it is not related to ionisation of this state. Note that we recorded TR PEI at different excitation wavelengths, but peaks A and B appeared at same electron kinetic energies (see Fig. S6 and S7 in the ESI†). This issue will be discussed in detail in the next section.

## 4 Discussion

The most obvious difference between the one colour [1+1] REMPI spectrum and the fluorescence excitation spectrum is the absence of transitions above  $700 \text{ cm}^{-1}$  in the latter one. This drop in  $\eta_n$  already indicates a rapid lifetime decrease with excitation energy and thus also an inefficient ionisation with ns laser pulses.<sup>10</sup> This interpretation is supported by the lifetimes of the bands depicted in Fig. 3 and summarised in Table 1. Bands located close to the  $^1B_{1u}$  origin show lifetimes in the ns region while nonradiative relaxation sets in  $0.12 \text{ eV}$  above the origin and higher excited vibrational bands deactivate on a ps time scale (e.g.  $162 \text{ ps}$  at  $+983 \text{ cm}^{-1}$  and  $35 \text{ ps}$  at  $+1448 \text{ cm}^{-1}$ ). Energetically the drop in lifetime coincides with the quenching of fluorescence.

As visible in the EAS, at longer delay times a second peak (B, see Fig. 6) at 1.57 eV grows in with time constant  $\tau_1$  (18 ps), due to ionisation from an intermediate state with a lifetime of 121 ps ( $\tau_2$ ) that is populated by deactivation of the  $^1B_{1u}$  state. Computations identified a curve crossing with the  $^1A_u$  state that has a biradicalic *trans bent* structure<sup>31–34</sup> ( $C_{2h}$  symmetry). A barrier of  $0.17 \text{ eV}$ <sup>31,33</sup> above the  $^1B_{1u}$  minimum energy geometry was calculated for the crossing point, in agreement with measured drop in lifetime. However, there is no information available on the ionisation of the *trans bent*  $^1A_u$  electronic state. Therefore we calculated the IE of *trans bent* tolane and obtained values from 9.46 (DFT) and 9.35 eV (CASPT2), assuming the linear ground state as the reference point. The energies of the electronic states are depicted in Fig. 7.

Based on the computed IE, one 263.5 nm (4.71 eV) probe photon is not sufficient to ionise *trans bent* tolane in a [1+1'] process even at the highest employed pump energy (272.6 nm, 4.55 eV, see Fig. 7) and at least two photons are required for ionisation. At 263.5 nm the probe pulse energy is relatively low (25–30  $\mu\text{J}$ ), multiphoton ionisation is suppressed and the  $^1A_u$  state cannot be ionised. In contrast, the much higher pulse energy at 351 nm (300–350  $\mu\text{J}$ ) enables multiphoton ionisation processes, and provides enough energy to ionise the intermediate *trans bent* state, leading to transient signals in the time resolved TR PEI experiments (see Fig. 5).

As mentioned above, the position of the surprisingly sharp peak B (1.57 eV) in the EAS does not depend on the excitation energy (see Fig. S6, S7 in the ESI†). This indicates that the excited state vibrational energy is conserved in the ionisation step. In principle it is possible that ionisation takes place directly from the vibrationally highly excited *trans bent* state via non resonant two photon absorption. Since the *trans bent*  $^1A_u$  state lies 3.31 eV above the planar ground state,<sup>27</sup> one UV photon (e.g. at 272.6 nm, 4.55 eV) deposits 1.24 eV of vibrational

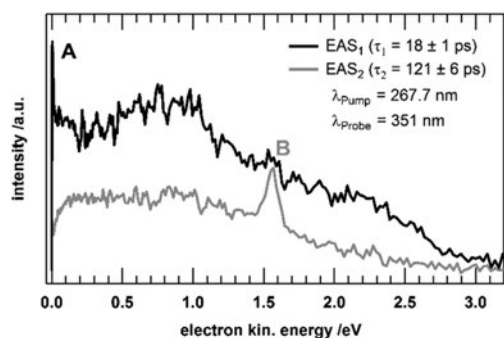
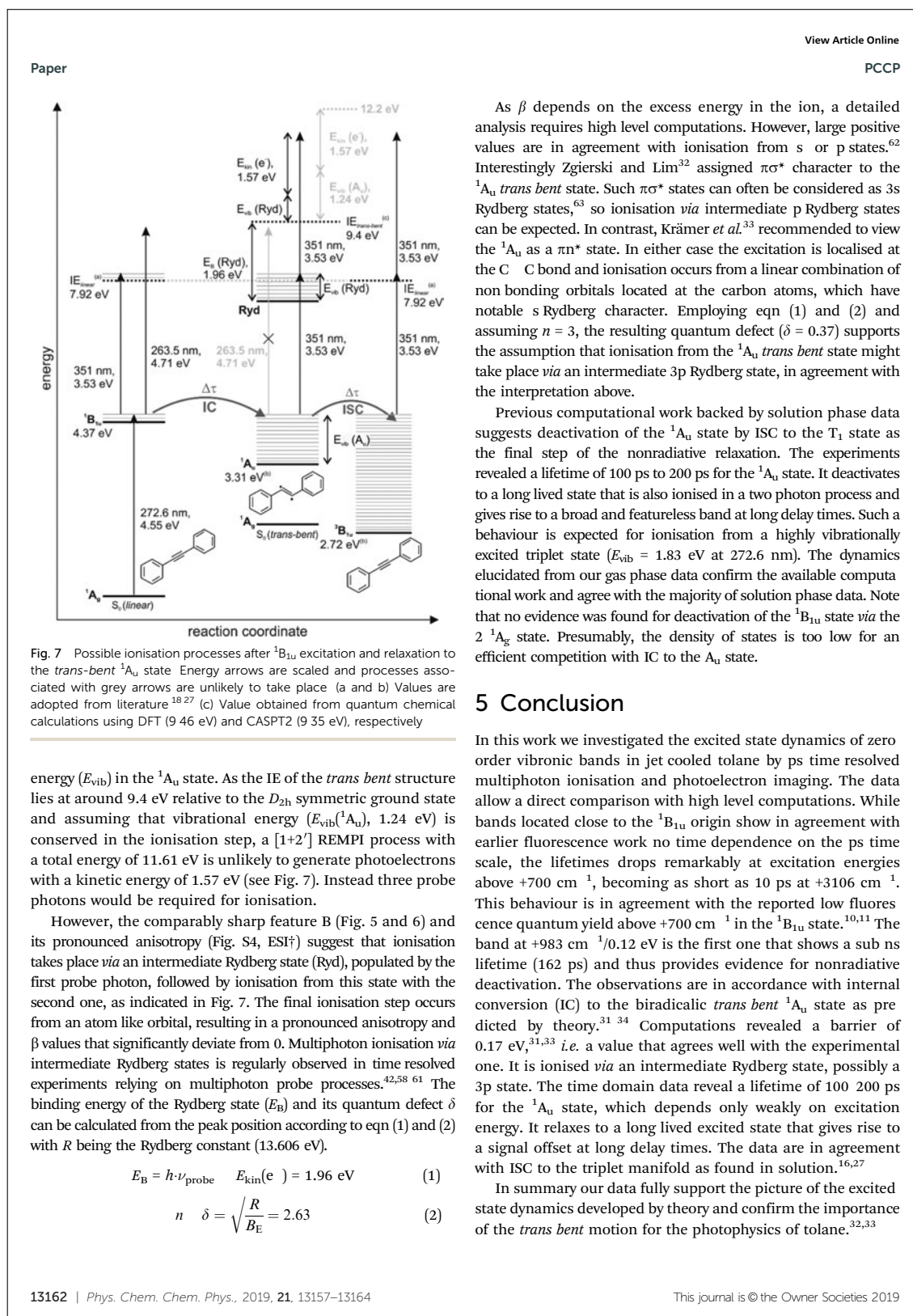


Fig. 6 Evolution associated spectra (EAS) of the species corresponding to time constants  $\tau_1$  (18 ps) and  $\tau_2$  (121 ps), respectively. Spectra were obtained from a global fit of the photoelectron map at an excitation wavelength of 267.7 nm ( $^1B_{1u}$  origin  $+2120 \text{ cm}^{-1}$ ) using the Glotaran<sup>56</sup> software. A two-step sequential relaxation model was assumed.



energy ( $E_{\text{vib}}$ ) in the  $^1A_u$  state. As the IE of the *trans bent* structure lies at around 9.4 eV relative to the  $D_{2h}$  symmetric ground state and assuming that vibrational energy ( $E_{\text{vib}}(^1A_u)$ , 1.24 eV) is conserved in the ionisation step, a [1+2'] REMPI process with a total energy of 11.61 eV is unlikely to generate photoelectrons with a kinetic energy of 1.57 eV (see Fig. 7). Instead three probe photons would be required for ionisation.

However, the comparably sharp feature B (Fig. 5 and 6) and its pronounced anisotropy (Fig. S4, ESI†) suggest that ionisation takes place *via* an intermediate Rydberg state (Ryd), populated by the first probe photon, followed by ionisation from this state with the second one, as indicated in Fig. 7. The final ionisation step occurs from an atom like orbital, resulting in a pronounced anisotropy and  $\beta$  values that significantly deviate from 0. Multiphoton ionisation *via* intermediate Rydberg states is regularly observed in time resolved experiments relying on multiphoton probe processes.<sup>42,58–61</sup> The binding energy of the Rydberg state ( $E_B$ ) and its quantum defect  $\delta$  can be calculated from the peak position according to eqn (1) and (2) with  $R$  being the Rydberg constant (13.606 eV).

$$E_B = h\nu_{\text{probe}} - E_{\text{kin}}(e^-) = 1.96 \text{ eV} \quad (1)$$

$$n - \delta = \sqrt{\frac{R}{E_B}} = 2.63 \quad (2)$$

As  $\beta$  depends on the excess energy in the ion, a detailed analysis requires high level computations. However, large positive values are in agreement with ionisation from s or p states.<sup>62</sup> Interestingly Zgierski and Lim<sup>32</sup> assigned  $\pi\sigma^*$  character to the  $^1A_u$  *trans bent* state. Such  $\pi\sigma^*$  states can often be considered as 3s Rydberg states,<sup>63</sup> so ionisation *via* intermediate p Rydberg states can be expected. In contrast, Krämer *et al.*<sup>33</sup> recommended to view the  $^1A_u$  as a  $\pi n^*$  state. In either case the excitation is localised at the C–C bond and ionisation occurs from a linear combination of non bonding orbitals located at the carbon atoms, which have notable s Rydberg character. Employing eqn (1) and (2) and assuming  $n = 3$ , the resulting quantum defect ( $\delta = 0.37$ ) supports the assumption that ionisation from the  $^1A_u$  *trans bent* state might take place *via* an intermediate 3p Rydberg state, in agreement with the interpretation above.

Previous computational work backed by solution phase data suggests deactivation of the  $^1A_u$  state by ISC to the  $T_1$  state as the final step of the nonradiative relaxation. The experiments revealed a lifetime of 100 ps to 200 ps for the  $^1A_u$  state. It deactivates to a long lived state that is also ionised in a two photon process and gives rise to a broad and featureless band at long delay times. Such a behaviour is expected for ionisation from a highly vibrationally excited triplet state ( $E_{\text{vib}} = 1.83$  eV at 272.6 nm). The dynamics elucidated from our gas phase data confirm the available computational work and agree with the majority of solution phase data. Note that no evidence was found for deactivation of the  $^1B_{1u}$  state *via* the  $^2^1A_g$  state. Presumably, the density of states is too low for an efficient competition with IC to the  $A_u$  state.

## 5 Conclusion

In this work we investigated the excited state dynamics of zero order vibronic bands in jet cooled toluene by ps time resolved multiphoton ionisation and photoelectron imaging. The data allow a direct comparison with high level computations. While bands located close to the  $^1B_{1u}$  origin show in agreement with earlier fluorescence work no time dependence on the ps time scale, the lifetimes drops remarkably at excitation energies above +700  $\text{cm}^{-1}$ , becoming as short as 10 ps at +3106  $\text{cm}^{-1}$ . This behaviour is in agreement with the reported low fluorescence quantum yield above +700  $\text{cm}^{-1}$  in the  $^1B_{1u}$  state.<sup>10,11</sup> The band at +983  $\text{cm}^{-1}$ /0.12 eV is the first one that shows a sub ns lifetime (162 ps) and thus provides evidence for nonradiative deactivation. The observations are in accordance with internal conversion (IC) to the biradical *trans bent*  $^1A_u$  state as predicted by theory.<sup>31–34</sup> Computations revealed a barrier of 0.17 eV,<sup>31,33</sup> *i.e.* a value that agrees well with the experimental one. It is ionised *via* an intermediate Rydberg state, possibly a 3p state. The time domain data reveal a lifetime of 100–200 ps for the  $^1A_u$  state, which depends only weakly on excitation energy. It relaxes to a long lived excited state that gives rise to a signal offset at long delay times. The data are in agreement with ISC to the triplet manifold as found in solution.<sup>16,27</sup>

In summary our data fully support the picture of the excited state dynamics developed by theory and confirm the importance of the *trans bent* motion for the photophysics of toluene.<sup>32,33</sup>

View Article Online

PCCP

Paper

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was financially supported by the Deutsche Forschungsgemeinschaft (DFG), contract FI575/9 2 and GRK 2112. We would like to thank Roland Mitrić for valuable discussions.

## Notes and references

- 1 T. M. Swager, *Acc. Chem. Res.*, 1998, **31**, 201.
- 2 P. F. H. Schwab, M. D. Levin and J. Michl, *Chem. Rev.*, 1999, **99**, 1863.
- 3 J. M. Tour, *Acc. Chem. Res.*, 2000, **33**, 791.
- 4 C. Devadoss, P. Bharathi and J. S. Moore, *J. Am. Chem. Soc.*, 1996, **118**, 9635.
- 5 S. F. Swallen, R. Kopelman, J. S. Moore and C. Devadoss, *J. Mol. Struct.*, 1999, **485**–**486**, 585.
- 6 J. Chen, M. A. Reed, A. M. Rawlett and J. M. Tour, *Science*, 1999, **286**, 1550.
- 7 N. Fuentes, A. Martín Lasanta, L. Álvarez de Cienfuegos, M. Ribagorda, A. Parra and J. M. Cuerva, *Nanoscale*, 2011, **3**, 4003.
- 8 D. Holten, D. F. Bocian and J. S. Lindsey, *Acc. Chem. Res.*, 2002, **35**, 57.
- 9 R. Ziessel, G. Ulrich, A. Haefele and A. Harriman, *J. Am. Chem. Soc.*, 2013, **135**, 11330.
- 10 K. Okuyama, T. Hasegawa, M. Ito and N. Mikami, *J. Phys. Chem.*, 1984, **88**, 1711.
- 11 D. R. Borst, S. Grace Chou and D. W. Pratt, *Chem. Phys. Lett.*, 2001, **343**, 289.
- 12 R. C. Henson and E. D. Owen, *Chem. Commun.*, 1967, 153.
- 13 L. B. Clark, *Chem. Phys.*, 1974, **5**, 484.
- 14 S. Elbel, K. Lienert, A. Krebs and H. T. Dieck, *Liebigs Ann. Chem.*, 1981, 1785.
- 15 M. Gutmann, M. Gudipati, P. F. Schoenart and G. Hohlneicher, *J. Phys. Chem.*, 1992, **96**, 2433.
- 16 Y. Hirata, T. Okada, N. Mataga and T. Nomoto, *J. Phys. Chem.*, 1992, **96**, 6559.
- 17 H. Hiura and H. Takahashi, *J. Phys. Chem.*, 1992, **96**, 8909.
- 18 K. Okuyama, M. C. R. Cockett and K. Kimura, *J. Chem. Phys.*, 1992, **97**, 1649.
- 19 Y. Hirata, T. Okada and T. Nomoto, *Chem. Phys. Lett.*, 1993, **209**, 397.
- 20 Y. Hirata, T. Okada and T. Nomoto, *J. Phys. Chem.*, 1993, **97**, 9677.
- 21 T. A. Ishibashi and H. O. Hamaguchi, *Chem. Phys. Lett.*, 1997, **264**, 551.
- 22 D. Zimdars, R. S. Francis, C. Ferrante and M. D. Fayer, *J. Chem. Phys.*, 1997, **106**, 7498.
- 23 T. A. Ishibashi and H. O. Hamaguchi, *J. Phys. Chem. A*, 1998, **102**, 2263.
- 24 Y. Hirata, *Bull. Chem. Soc. Jpn.*, 1999, **72**, 1647.
- 25 T. a. Ishibashi, H. Okamoto and H. o. Hamaguchi, *Chem. Phys. Lett.*, 2000, **325**, 212.
- 26 T. Nomoto, T. A. Ishibashi, H. Okamoto and H. O. Hamaguchi, *J. Mol. Struct.*, 2005, **735**–**736**, 197.
- 27 J. Saltiel and V. K. R. Kumar, *J. Phys. Chem. A*, 2012, **116**, 10548.
- 28 T. Suzuki, M. Nakamura, T. Isozaki and T. Ikoma, *Int. J. Thermophys.*, 2012, **33**, 2046.
- 29 C. Ferrante, U. Kensy and B. Dick, *J. Phys. Chem.*, 1993, **97**, 13457.
- 30 A. Shimojima and H. Takahashi, *J. Phys. Chem.*, 1993, **97**, 9103.
- 31 Y. Amatatsu and M. Hosokawa, *J. Phys. Chem. A*, 2004, **108**, 10238.
- 32 M. Z. Zgierski and E. C. Lim, *Chem. Phys. Lett.*, 2004, **387**, 352.
- 33 M. Krämer, U. H. F. Bunz and A. Dreuw, *J. Phys. Chem. A*, 2017, **121**, 946.
- 34 C. Robertson and G. A. Worth, *Chem. Phys.*, 2018, **510**, 17.
- 35 A. V. Abramnikov, A. Almenningen, B. N. Cyvin, S. J. Cyvin, T. Jonvik, L. S. Khaikin, C. Romming and L. V. Vilkov, *Acta Chem. Scand., Ser. A*, 1998, **42**, 674.
- 36 Z. Chernia, T. Livneh, I. Pri Bar and J. E. Koresh, *Vib. Spectrosc.*, 2001, **25**, 119.
- 37 G. Baranović, L. Colombo and D. Skare, *J. Mol. Struct.*, 1986, **147**, 275.
- 38 R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1997.
- 39 T. Gerbich, H. C. Schmitt, I. Fischer, J. Petersen, J. Albert and R. Mitrić, *J. Phys. Chem. A*, 2015, **119**, 6006.
- 40 J. Hoche, H. C. Schmitt, A. Humeniuk, I. Fischer, R. Mitrić and M. I. S. Rohr, *Phys. Chem. Chem. Phys.*, 2017, **19**, 25002.
- 41 H. C. Schmitt, M. Flock, E. Welz, B. Engels, H. Schneider, U. Radius and I. Fischer, *Chem. Eur. J.*, 2017, **23**, 3084.
- 42 M. Flock, M. P. Herbert and I. Fischer, *Chem. Phys.*, 2018, **515**, 744.
- 43 A. Stolow, *Annu. Rev. Phys. Chem.*, 2003, **54**, 89.
- 44 A. Stolow, A. E. Bragg and D. M. Neumark, *Chem. Rev.*, 2004, **104**, 1719.
- 45 T. Suzuki, *Annu. Rev. Phys. Chem.*, 2006, **57**, 555.
- 46 J. Auerswald, B. Engels, I. Fischer, T. Gerbich, J. Herterich, A. Krueger, M. Lang, H. C. Schmitt, C. Schon and C. Walter, *Phys. Chem. Chem. Phys.*, 2013, **15**, 8151.
- 47 A. T. J. B. Eppink and D. H. Parker, *Rev. Sci. Instrum.*, 1997, **68**, 3477.
- 48 G. A. Garcia, L. Nahon and I. Powis, *Rev. Sci. Instrum.*, 2004, **75**, 4989.
- 49 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers,

[View Article Online](#)

Paper

PCCP

- K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision B.01*, Gaussian Inc., Wallingford, CT, 2016.
- 50 J. D. Chai and M. Head Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615.
- 51 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007.
- 52 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057.
- 53 K. Andersson, P. Å. Malmqvist and B. O. Roos, *J. Chem. Phys.*, 1992, **96**, 1218.
- 54 R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796.
- 55 F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P. Å. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitoňák, M. Reiher, B. O. Roos, L. Serrano Andrés, M. Urban, V. Veryazov and R. Lindh, *J. Comput. Chem.*, 2010, **31**, 224.
- 56 J. J. Snellenburg, S. Liptenok, R. Seger, K. M. Mullen and I. H. M. van Stokkum, *J. Stat. Softw.*, 2012, **49**, 1.
- 57 I. H. M. van Stokkum, D. S. Larsen and R. van Grondelle, *Biochim. Biophys. Acta, Bioenerg.*, 2004, **1657**, 82.
- 58 C. P. Schick and P. M. Weber, *J. Phys. Chem. A*, 2001, **105**, 3735.
- 59 C. P. Schick and P. M. Weber, *J. Phys. Chem. A*, 2001, **105**, 3725.
- 60 J. L. Gosselin and P. M. Weber, *J. Phys. Chem. A*, 2005, **109**, 4899.
- 61 M. Tsubouchi and T. Suzuki, *J. Phys. Chem. A*, 2003, **107**, 10897.
- 62 E. Bohl, B. Mignolet, J. O. Johansson, F. Remacle and E. E. B. Campbell, *Phys. Chem. Chem. Phys.*, 2017, **19**, 24090.
- 63 M. N. R. Ashfold, G. A. King, D. Murdock, M. G. D. Nix, T. A. A. Oliver and A. G. Sage, *Phys. Chem. Chem. Phys.*, 2010, **12**, 1218.

## 6. Rationalizing the Femtosecond Dynamics of Diphenylpropynylidene

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 254 (2021) 119606



Contents lists available at ScienceDirect  
Spectrochimica Acta Part A: Molecular and  
Biomolecular Spectroscopy

journal homepage: [www.elsevier.com/locate/saa](http://www.elsevier.com/locate/saa)

### Femtosecond dynamics of diphenylpropynylidene in ethanol and dichloromethane

Lea Ress<sup>a,1</sup>, Dustin Kaiser<sup>a,1</sup>, Jeannine Grüne<sup>b</sup>, Marius Gerlach<sup>a</sup>, Engelbert Reusch<sup>a</sup>, Tobias Brixner<sup>a,c</sup>, Andreas Sperllich<sup>b</sup>, Bernd Engels<sup>a</sup>, Ingo Fischer<sup>a</sup>

<sup>a</sup> Institut für Physikalische und Theoretische Chemie, Universität Würzburg, Am Hubland, D 97074 Würzburg, Germany

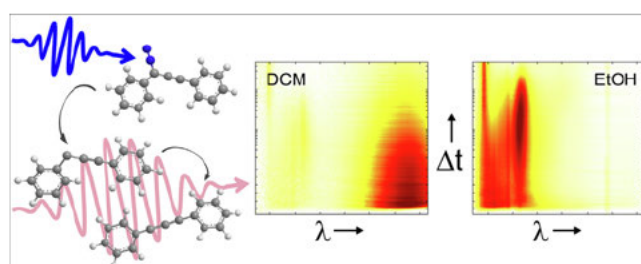
<sup>b</sup> Fakultät für Physik, Universität of Würzburg, Am Hubland, D 97074 Würzburg, Germany

<sup>c</sup> Center for Nanosystems Chemistry (CNC), Universität Würzburg, Theodor Boveri Weg, 97074 Würzburg, Germany

#### HIGHLIGHTS

- We report an experimental and computational investigation of the biradical  $C_6H_5-C_3-C_6H_5$ .
- Its structure and dynamics are investigated by EPR- spectroscopy and fs- transient absorption.
- The interpretation of the data is supported by high-level computations.

#### GRAPHICAL ABSTRACT



#### ARTICLE INFO

##### Article history:

Received 1 December 2020  
Received in revised form 2 February 2021  
Accepted 5 February 2021  
Available online 27 February 2021

#### ABSTRACT

Carbon chains with an odd number of C atoms are reactive intermediates with a high biradical character. Here we report a joint experimental and computational investigation of the dynamics of diphenylpropynylidene,  $C_6H_5-C_3-C_6H_5$  in dichloromethane and ethanol. The biradical is generated by ultraviolet light from 1,3-diphenyldiazopropyne. Electron paramagnetic resonance spectra are recorded to elucidate the spin multiplicity and geometry of the biradical. In both solvents a triplet ground state at 4 K is verified. Transient absorption spectra provide insight into the fate of the biradical. A study in deaerated dichloromethane permits us to follow the photophysics of diphenylpropynylidene and to extract time constants for its vibrational as well as electronic relaxation. In the presence of oxygen, a more complex photochemistry is observed that permits us to derive a model for the reaction of the biradical with  $O_2$ . In ethanol, the spectra recorded in the presence and absence of  $O_2$  are very similar, which can be explained by the similarity of the chromophores of the reaction products.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

#### 1. Introduction

Biradicals, including triplet carbenes, have fascinated chemists for a long time due to their unusual electronic structure and binding properties [1].

The continuing interest in this class of compounds is documented by recent reviews that summarize the theoretical and experimental state of the art [2,3]. This interest is also stimulated by a number of novel discoveries, among them the role of carbonyl oxide biradicals in atmospheric chemistry [4,5] and the importance of biradical character in molecular candidates for singlet fission, a process that has the potential to improve the efficiency of solar cells [6,7]. Recently it was furthermore discovered that boron-containing biradicals are able to fixate and

<sup>1</sup> Authors contributed equally.

E-mail addresses: [brixner@uni-wuerzburg.de](mailto:brixner@uni-wuerzburg.de) (T. Brixner), [andreas.sperllich@uni-wuerzburg.de](mailto:andreas.sperllich@uni-wuerzburg.de) (A. Sperllich), [bernd.engels@uni-wuerzburg.de](mailto:bernd.engels@uni-wuerzburg.de) (B. Engels), [ingo.fischer@uni-wuerzburg.de](mailto:ingo.fischer@uni-wuerzburg.de) (I. Fischer)

<https://doi.org/10.1016/j.saa.2021.119606>

1386 1425/© 2021 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



reduce  $N_2$  [8]. Such boron-based molecules are known to switch from closed-shell to biradicalic structures if the stabilizing cyclic carbene changes from NHC (N-Heterocyclic carbene) to CAAC (cyclic (alkyl)(amino)carbene) [9–11]. Here, insight into the structure and dynamics of excited electronic states is necessary to understand the complex chemistry.

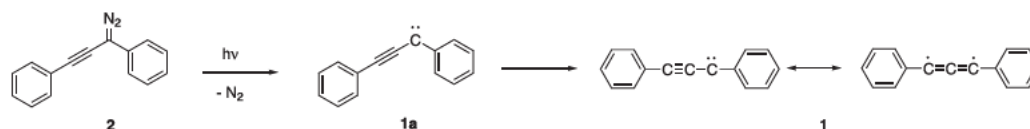
While the ground-state structure of a large number of biradicals has been investigated by theory [12] as well as infrared spectroscopy or electron paramagnetic resonance (EPR) in rare gas matrices [2,13] or supersonic jets [14], there is little information on the excited electronic states. This is due to the small energy gaps and subsequently short lifetimes that require a multireference approach in computations and femtosecond time resolution in the experiment. In the gas phase only the photoinduced dynamics of the  $C_3H_2$  isomers propadienylidene [15] and propynylidene [16], as well as of para-xylylene [17] have been unraveled in joint experimental and computational studies. Little information is also available on the dynamics in solution. Here the reactivity is to a large extent controlled by the singlet–triplet gap  $\Delta E_{ST}$ , which in itself is solvent-dependent. Femtosecond transient absorption of the formation of a few selected carbenes [18] and nitrenes [19,20] and their decay in solution have been investigated before, but detailed studies of the solution-phase chemistry of reactive intermediates, in particular biradicals, are still rare. Only recently femtosecond time-resolved experiments in methanol/acetonitrile solvent mixtures were reported for diphenylmethylcarbene,  $Ph_2C$  [21]. While a triplet ground state was found in the gas phase as well as in acetonitrile, the closed-shell singlet becomes the ground state in methanol. This leads to a very different chemistry of the compound in both solvents.

The solvent dependence of the spin multiplicity and the reactivity are in the center of the present work. Motivated by our longstanding interest in the chemistry and the photoionization of odd- $n$  carbon chains  $R-C_n-R'$  [16,22,23] and of diazo compounds [24,25], we focus here on diphenylpropynylidene **1**, which is obtained by ultraviolet (UV) irradiation from 1,3-diphenyldiazopropyne **2**. As evident from Scheme 1, there are two resonance structures, a biradicalic and a carbenic one. The molecule has been investigated previously by McMahon, Stanton and coworkers, who recorded EPR and IR spectra of **1** in an Ar matrix [26], and found that photolysis leads initially to the bent geometry **1a**. After annealing, a linear geometry was adopted. Here we investigate the ultrafast dynamics of **2** and **1** in solution by transient absorption, guided by computations of excited electronic states of important intermediates and possible reaction products and supported by EPR spectroscopy. As solvents we employ dichloromethane (DCM), an inert non-polar and aprotic solvent that permits to follow the dynamics of the biradical, and ethanol, which is both polar and protic and will react with the nascent biradical. Its reactivity is assumed to be comparable to methanol. A model for the photodynamics in both solvents, is derived.

## 2. Methods

### a) Experimental

Diphenylpropynylidene **1** was generated photolytically from the diazo-compound **2** (see Scheme 1), which was synthesized follow-



**Scheme 1.** Diphenylpropynylidene **1** is generated from 1,3-diphenyldiazopropyne **2** by irradiation with 295 nm or 305 nm light. Note that the biradical resonance structure (far right) is the most appropriate description for the electronic structure of **1**.

ing the procedure given in Ref. [26]. Since the compound is unstable at ambient conditions, it was freshly synthesized before each measurement from the corresponding tosylhydrazone. The solvents dichloromethane (DCM) and ethanol were distilled before use and kept over molecular sieves to minimize water contaminations.

We applied ultrafast transient absorption (TA) spectroscopy in the ultraviolet/visible (UV/Vis) region with a setup that was driven by a regenerative Ti:sapphire amplifier system (Solstice, Spectra-Physics) at a central wavelength of 800 nm and a pulse duration of 120 fs with a repetition rate of 1 kHz. One part of the 800 nm beam was used to generate the  $\sim 70$  fs excitation pulses with a central wavelength of 295 nm and a pulse energy of  $0.3 \mu J$  in a commercial nonlinear optical parametric amplifier (TOPAS White, Light Conversion). The broadband UV/Vis probe pulses were generated by focusing the other part of the fundamental into a linearly moving 5 mm thick  $CaF_2$  window, resulting in a white-light continuum from 320 nm to 665 nm. The pump and probe beams were spatially overlapped in a flow cuvette with a thickness of 200  $\mu m$ , which leads to an optical density between 0.4 and 0.7 at 295 nm for the several samples (see Supplementary Information, SI, Figure S1). To avoid anisotropy effects, the linear polarizations of the two beams were set to the magic angle of  $54.7^\circ$  by rotating an achromatic wave plate in the pump beam. By delaying the pump beam with a mechanical translation stage (M-IMS600, Newport), the pump–probe delay was varied up to 500 ps. Every second pump pulse was blocked by a chopper driven at a frequency of 500 Hz. After passing the sample, the probe pulses were spectrally dispersed by a spectrometer (Acton SP2500i) and detected shot-to-shot by a CCD camera (Pixis 2 K, Princeton Instruments). For each delay time, 2000 absorbance difference spectra were averaged. For each sample, we measured up to three consecutive data sets. To cool down the sample container to  $-30^\circ$ , we used a cooling bath consisting of *o*-xylene and dry ice. The measurement for each sample took about 3 h, after which the cooling bath was thawed. In addition, a color change from red to light yellow of the sample was found. The steady-state absorption spectra of each sample before and after the TA measurements are also shown in SI Figure S1. The color change from red to yellow can be explained by the diminished intensity of the peak around 540 nm after the TA experiment (see insets of Figure S1). Therefore, only spectra obtained with fresh sample will be discussed below. The transient data were evaluated via global analysis [27] with the software package Glotaran [28] based on the R-package TIMP [29].

For EPR measurements, the solution was filled in an EPR quartz tube (Wilmad) with helium atmosphere. EPR experiments were carried out with an X-band spectrometer (Bruker E300) equipped with a continuous-flow helium cryostat (Oxford ESR 900) and a microwave cavity (Bruker ER4104OR, 9.43 GHz) with optical access. Optical irradiation was performed with a 305 nm UV LED of 300 mW power for one minute from each side of the cavity. Detailed specifications such as microwave power, modulation amplitude and sweep duration for each experiment are given in the SI.

### b) Computational

All calculations were performed using the Gaussian16 program suite [30,31]. Reliable predictions about biradicals often deserve

multi-reference-approaches [32,33] but recent investigations showed that DFT also delivers quite accurate insights [34]. The hybrid dispersion-corrected density functional  $\omega$ B97xD was used [35] in combination with the 6-311G(d,p) basis set [36,37]. In all cases, polarizable continuum models (PCM) were employed using the integral equation formalism [38]. We always used the parameters for the respective solvent. For all structures, minimized geometries were verified to be true minima by computation of the molecular hessian. Excited-to-excited-state absorption frequencies and oscillator strengths were obtained using the MultiWFN program suite [39]. Computations of the EPR parameters employing the modern coupled-perturbed (CR) method [40] failed due to convergence problems. Hence, we had to use the quasi-restricted orbital (QRO) method [41]. Because it is only strictly valid for pure GGA DFT functionals we employed the PBE functional [42] in combination with def2SVP basis sets [43,44] for the EPR parameter computations. The computations were performed

on top of the geometries obtained with  $\omega$ B97xD/6-311G(d,p). For these calculations, we used the ORCA4.2 [45] program suite.

### 3. Results

#### a) EPR Spectra

For **1** in Ar matrix a bent geometry **1a** with a triplet ground state was found [26]. However, the ground-state multiplicity can depend on the solvent, as found by Knorr et al. [21] and discussed above. We therefore investigated the spin multiplicity of **1a** at 4 K in frozen ethanol and DCM by irradiating precursor **2** at 305 nm and analyzed the transformation to **1** at higher temperatures.

Fig. 1 shows the resulting spectra in DCM (upper trace, black line) and ethanol (lower trace, black line) upon irradiation at 305 nm. In both solvents, a triplet ground state is found for **1a**. Red traces present EasySpin simulations [46] based on axial (*D*) and rhombic (*E*) zero-field splitting parameters given in Table 1.

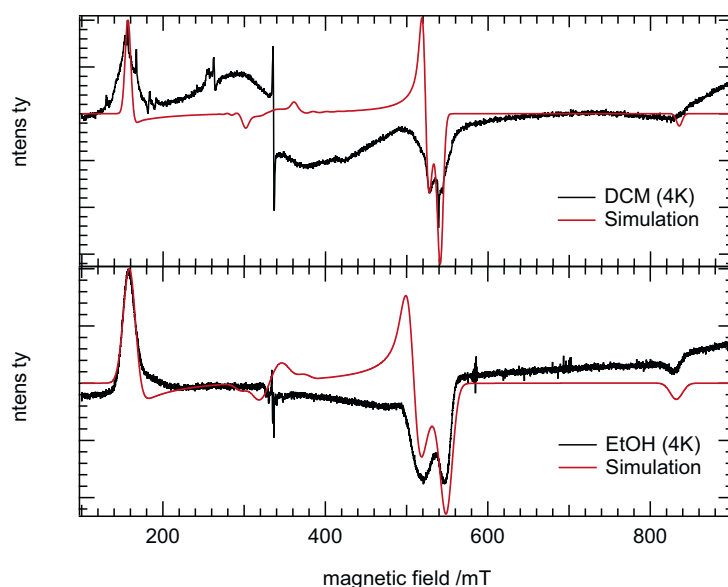


Fig. 1. EPR spectra (black) of **1a** as produced upon irradiation (305 nm) of the precursor **2** at 4 K. In frozen solutions of both DCM (upper trace, black line) and ethanol (lower trace, black line) **1a** is formed in the triplet state. Axial (*D*) and rhombic (*E*) zero-field splitting parameters were determined by simulations (red traces).

Table 1

Experimental ("exp") and calculated ("calc") values of the axial (*D*) and rhombic (*E*) zero-field splitting parameters for **1a** and **1** under various solvent and temperature conditions. WRONG IMAGES FOR GEOMETRY!! PLEASE SEE MANUSCRIPT!

Geometry	Ethanol 4 K	DCM 4 K	DCM > 70 K	Ar 10 K [26]	Ar > 40 K [26]
$D_{exp} / \text{cm}^{-1}$	0.464	0.460	0.460	0.484	0.468
$E_{exp} / \text{cm}^{-1}$	0.010	0.005	< 0.002	0.00423	< 0.0002
$D_{calc} / \text{cm}^{-1}$	-	0.263 <sup>1)</sup>	0.258 <sup>2)</sup>	-	-
$E_{calc} / \text{cm}^{-1}$	-	0.004 <sup>1)</sup>	$1.1 \cdot 10^{-6}$ <sup>2)</sup>	-	-

<sup>1)</sup> Structure optimized for the triplet state for a fixed  $\angle \text{CCC} = 130^\circ$ .

<sup>2)</sup> Linear equilibrium structure of the triplet state.

The axial zero-field splitting value of around  $0.46 \text{ cm}^{-1}$  (13.8 GHz) is similar for both solvents, indicating a comparable delocalization of radical centers in DCM and ethanol. Reports from previous EPR measurements in Ar matrix also show similar axial zero-field splittings (Table 1).[26] As evident from Table 1, a small, but decisive non-zero value of  $0.005\text{--}0.01 \text{ cm}^{-1}$  (150–300 MHz) is found for the rhombic zero-field splitting  $E$ , which indicates a non-axial symmetry. This result confirms that after photolysis, the biradical is formed in the bent geometry 1a, given in Scheme 1. To analyze the transformation to 1, the samples were annealed to different temperatures until EPR spectra reveal a change or disappearance in zero-field splitting parameters, indicating geometrical conversions. When the DCM solution is annealed to 70 K, and subsequently recooled to 4 K for EPR measurement, the triplet ground-state character remains with similar axial zero-field splitting  $D$  as before (see Figure S2). However, the rhombic zero-field splitting parameter  $E$  becomes almost zero ( $<0.002 \text{ cm}^{-1}$ ). This transformation of the  $E$  value presents a transition from previously separated  $X_+$  and  $Y_+$  transitions at around 520 mT (Fig. 1) to a single  $XY_+$  transition (Figure S2).[31] In terms of molecular geometry, this result indicates vibrational relaxation, i.e., excess energy in the bending mode is transferred to the solvent and a linear geometry (1 in Scheme 1) is adopted. Similar results have been obtained in Ar matrix upon annealing to 40 K (Table 1).[26] In contrast, the EPR triplet signal disappears in ethanol after annealing to 80 K (see lower trace of Figure S3), indicating the formation of a product of 1 in a singlet state due to a chemical reaction with the solvent. The same result of a disappearing triplet ground state after anneal-

ing is obtained by EPR measurement in DCM containing  $\text{O}_2$  (see upper trace of Figure S3).

#### b) Transient Absorption Spectra

TA spectra were recorded in DCM and ethanol, both deaerated and open to air, i.e., containing  $\text{O}_2$ . The temperature was varied between  $-30 \text{ }^\circ\text{C}$  and room temperature; however, we concentrate on the low-temperature results below. It was expected that the rapid reaction of open-shell species with  $\text{O}_2$  helps in the band assignment, because bands that originate from open-shell singlet or triplet states will disappear quickly under ambient conditions. The EPR data (vide supra) show that 1a reacts with ethanol at temperatures above 70 K, whereas DCM is expected to be inert and thus allows detection of open-shell species in the absence of oxygen.

Fig. 2 shows the TA maps of the samples prepared in DCM (a, b) and in ethanol (c, d) each without (a, c) and with (b, d) oxygen present. All bands have a positive sign and can thus be assigned to either excited-state or product absorption. A possible ground-state bleach should appear around 320 nm (see steady-state absorption in Figure S1), i.e., at the blue edge of the spectrum. Fig. 3 displays the temporal behavior of the most important bands under the various conditions. All spectra were evaluated with a target analysis based on a sequential kinetic model, yielding five time constants (Table 2). In all spectra, except the spectra in DCM without  $\text{O}_2$ , it was necessary to include two oscillation parameters,  $\nu$ , which describes the

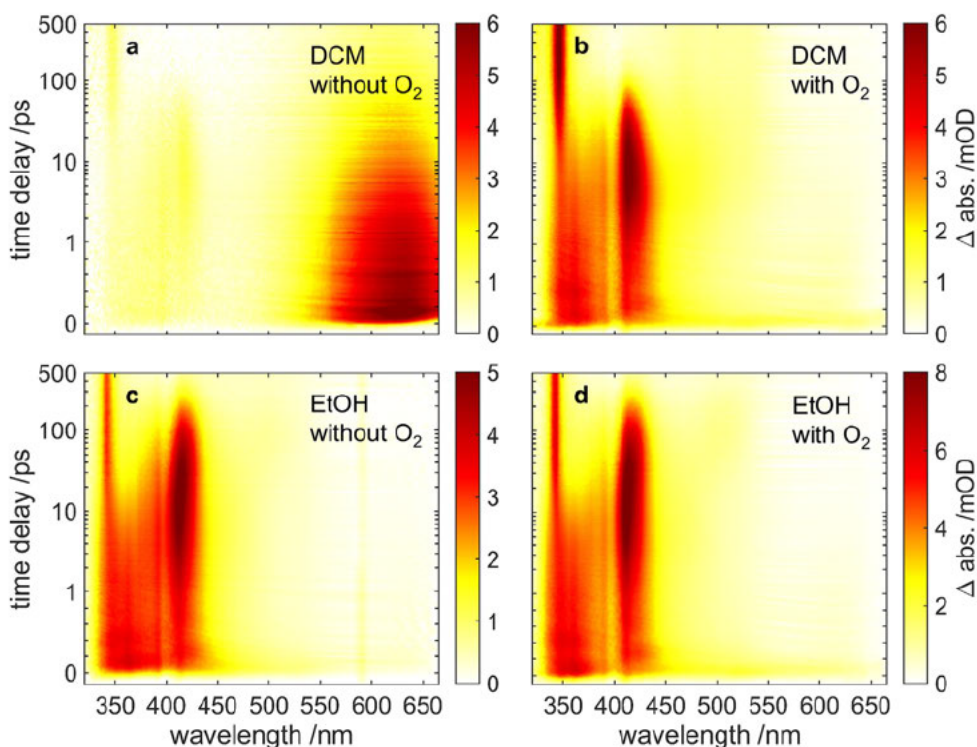
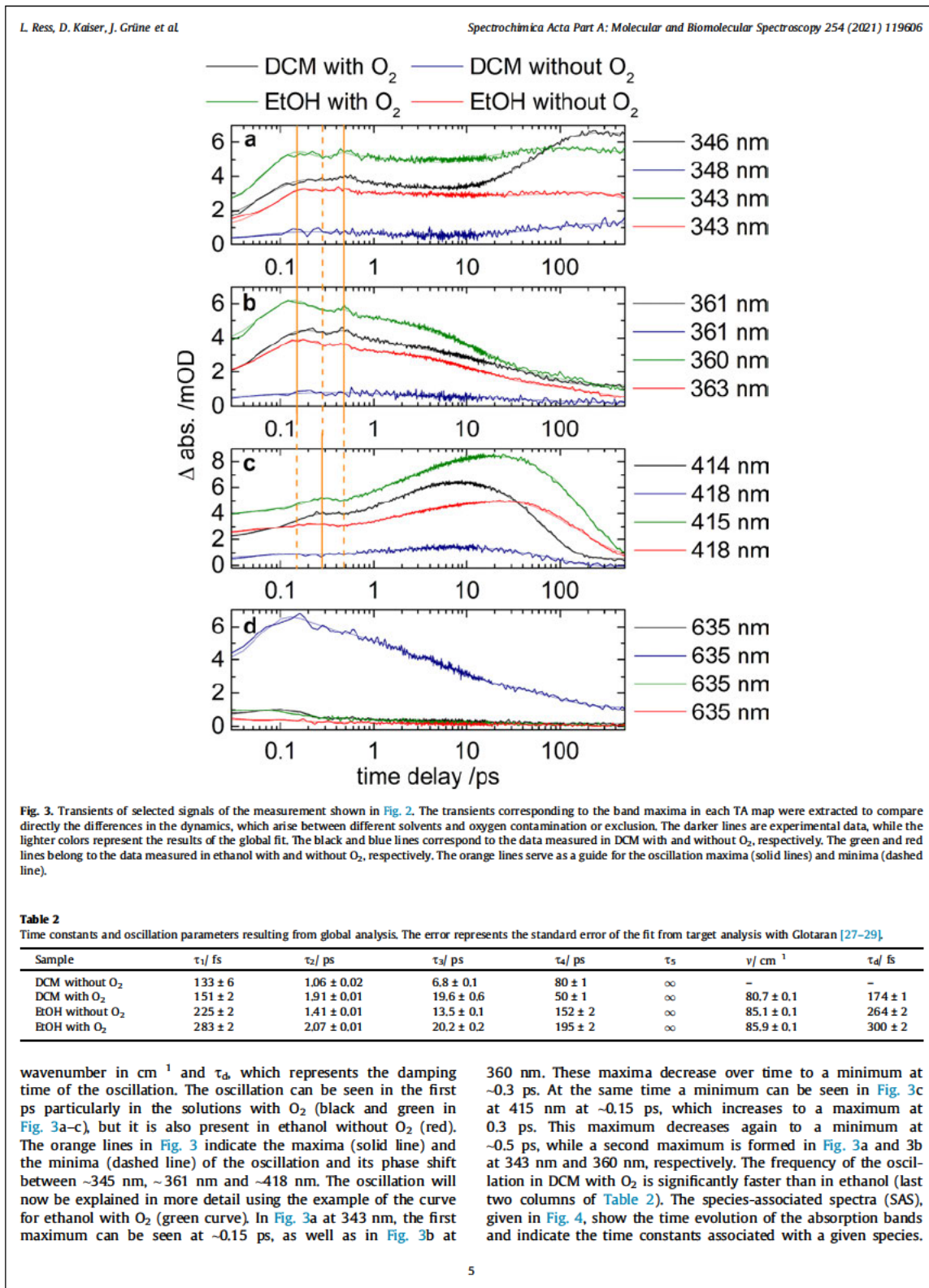
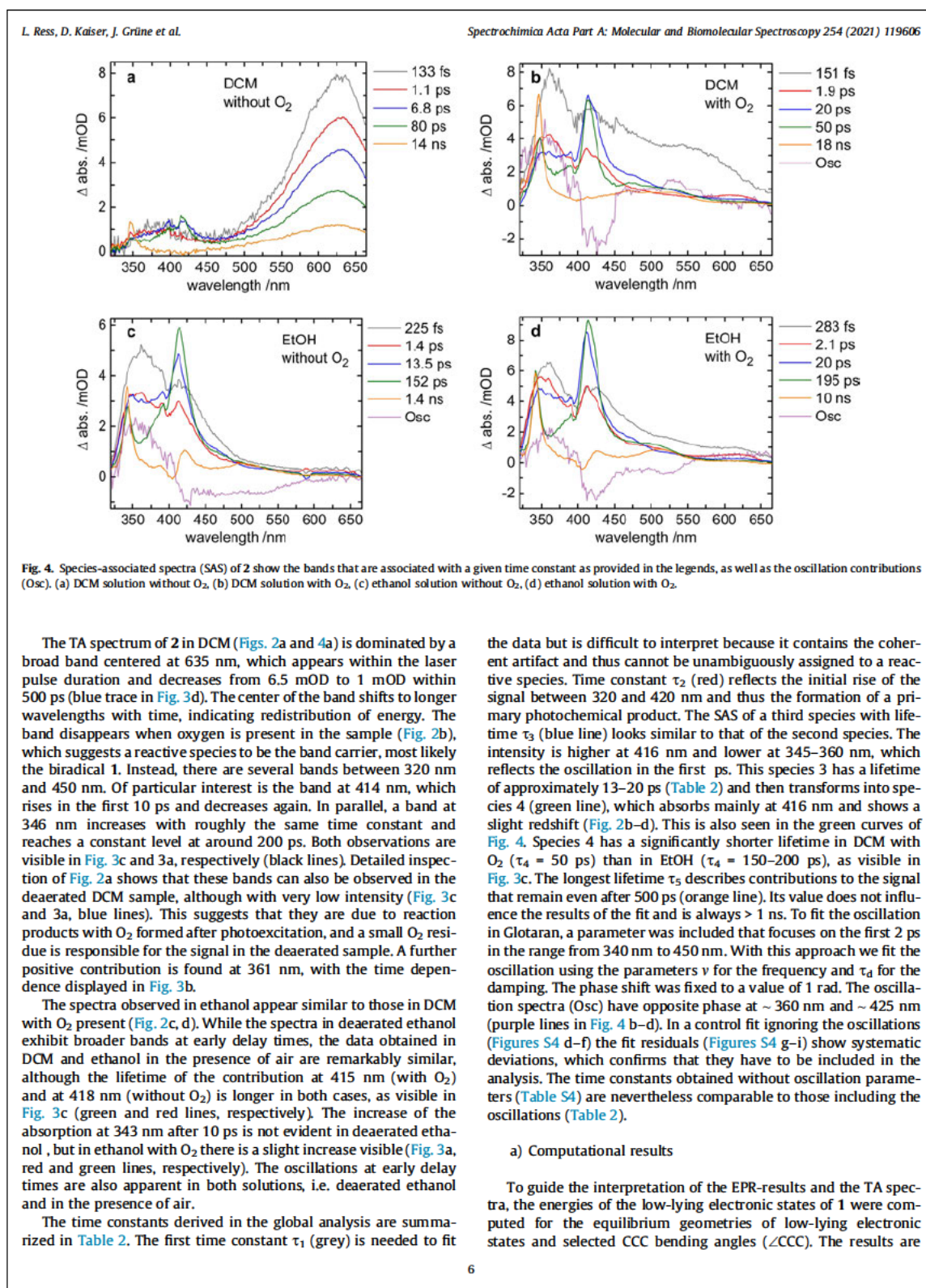


Fig. 2. TA maps of 2 excited at 295 nm at  $30 \text{ }^\circ\text{C}$ . (a) DCM solution without  $\text{O}_2$ , (b) DCM solution with  $\text{O}_2$ , (c) ethanol solution without  $\text{O}_2$ , (d) ethanol solution with  $\text{O}_2$ . The time delay axis is plotted linearly from 0.1 ps to 1 ps and logarithmically from 1 ps to 500 ps.



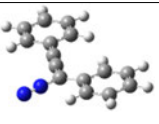
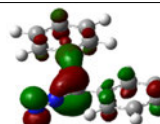
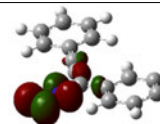
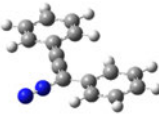
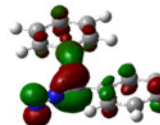
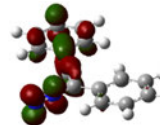
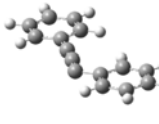
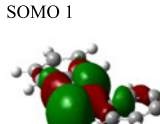
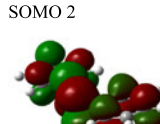








**Table 3**  
Dominant orbital contributions to the first two excited singlet states of the precursor **2** and the dissociation product **1**.

Excitation wavelength / energy	Geometry	Dominant occupied orbital (HOMO)	Dominant virtual orbital
490.1 nm / 2.53 eV			
297.4 nm / 4.17 eV			
Dissociation product		SOMO 1 	SOMO 2 

[47]. When the MOs of the diazo compound are correlated with those of bent  $\text{Ph}_2\text{C}_3$ , one finds that dissociation of the diazo group forms  $\text{Ph}_2\text{C}_3$  in its bent open-shell (OS) excited singlet state  $S_{2,os}$  (see Table 3). Note that spin conservation necessitates formation of **1** in a singlet state, because  $\text{N}_2$  can only be formed in the  $X^1\Sigma_g^+$  state due to its energetically inaccessible lowest triplet state. Thus, an open-shell singlet arises, because the former  $\pi$  electron remains in a carbon-centered orbital, which is out-of-plane with respect to the planar  $\text{Ph}_2\text{C}_3$  fragment (SOMO 1, Table 3), whereas the former C-N  $\sigma$  electron remains in a different orbital in the molecular plane of  $\text{Ph}_2\text{C}_3$  (SOMO 2, Table 3). These two MOs become energetically degenerate at the triplet and open-shell-singlet equilibrium structures.

#### 4. Discussion

We will now relate the experimental data with the computational results to describe the photoinduced dynamics of **2**. In a first step we will assign the simplest TA spectrum recorded in DCM without the presence of oxygen, which shows a single band that ranges approximately from 560 to 660 nm (1.88–2.21 eV). This band appears within  $\tau_2 \approx 1$  ps and decays with a time constant  $\tau_4 \approx 80$  ps.

The 295 nm pump pulse excites the molecule into the  $S_2$  state, which is dissociative along the C–N<sub>2</sub> coordinate. Elimination into a nitrogen molecule and **1** occurs presumably on a sub-100 fs time scale, which cannot be resolved in the present setup, and thus the dissociation of **2** is probably included in the fastest decay time constant  $\tau_1$ . The TA signal disappears, because the precursor leaves the Franck–Condon window when the C–N<sub>2</sub> bond is elongated. Nevertheless,  $\tau_1$  also contains a signal contribution from the coherent artifact, which renders an assignment difficult. As already mentioned, the correlation diagram in Fig. 5 shows that **1** is formed in the  $S_2$  open-shell (OS) singlet state, but in the bent structure **1a** because relaxation to the linear structure is slower than  $\text{N}_2$  dissociation. However, the minimum energy of  $S_{2,os}$  corresponds to a

linear carbon chain with twisted phenyl rings (Figs. 5, 1c). Hence, the lifetime of the bent structure will be only short, because a vibrational relaxation to the linear structure sets in.

Based on the computations, a vertical excitation energy of 2.35 eV is predicted for the  $S_{2,os} \rightarrow S_3$  absorption in **1** at the bent geometry (**1a**). For the corresponding adiabatic excitation energy (red line in Fig. 5) to the equilibrium geometry of the  $S_3$  state a value of 1.96 eV is computed. This is in very good agreement with the peak in the TA spectrum (Fig. 2a) centered at 635 nm (1.95 eV). The molecule is expected to possess a high excess energy and will be highly vibrationally excited after the dissociation of the  $\text{N}_2$  fragment. Note that based on the values given in Table S2 and Fig. 5, the  $S_{1,cs} \rightarrow S_{2,os}$  transition at the bent geometry **1a** is expected at much lower energies, while the  $S_{1,cs} \rightarrow S_3$  transition will appear at much higher energies. Thus  $S_{1,cs}$  can be ruled out as the carrier of the band.

As apparent from Fig. 2a and the red and blue traces in Fig. 4a, the TA spectrum shifts to longer wavelengths (lower energies) with time. This is due to the relaxation of the  $S_{2,os}$ , which is formed with considerable vibrational excitation that is transferred to the solvent. We assign the time constant  $\tau_3 \approx 7$  ps to this vibrational relaxation, i.e., the transition from the bent (**1a**) to the linear biradical **1c** (see Fig. 5). The time constant  $\tau_4 \approx 80$  ps corresponds to non-radiative deactivation of the linear  $S_{2,os}$ . The EPR spectra show that finally the molecule decays into the triplet ground state by intersystem crossing (ISC). Note that  $S_{1,cs}$  is energetically above  $S_{2,os}$  at this geometry, thus a two-step electronic deactivation is excluded.

After assigning the simplest TA spectrum recorded in DCM in the absence of oxygen, we turn to the corresponding EPR results (Table 1 and Table S1). For  $T > 70$  K the experimental EPR data taken in a matrix indicate a linear triplet ground state as the  $E$ -value becomes smaller than  $0.002 \text{ cm}^{-1}$ . In contrast, for  $T = 4$  K an  $E$ -parameter of about  $0.005 \text{ cm}^{-1}$  indicates a bent geometry. The observation of an EPR spectrum confirms a triplet ground state at both geometries. A singlet ground state with a slightly higher

lying triplet state can be excluded because such low temperatures would not allow its population. The different EPR data for 4 K and 70 K may result, because the triplet is trapped in a local minimum with bent geometry from which it cannot escape at 4 K. After heating to 70 K it can surmount the barrier and relaxes to the global linear minimum. Alternatively, the relaxation to the linear geometry might be hampered by the matrix environment at 4 K. For  $T > 70$  K either the matrix environment gets less rigid so that the molecule can relax, or the molecule acquires sufficient thermal energy to surmount the barrier. Astonishingly, while the experimental  $E$  parameter changes going from 4 K to 70 K, the  $D$ -parameter does not change significantly. The findings for 70 K are in line with our computations in the gas phase which predict a triplet ground state with a linear geometry as global minimum of the PES. The computed  $E$ -value of  $1.1 \cdot 10^6 \text{ cm}^{-1}$  agrees well with the experimental result of  $< 0.002 \text{ cm}^{-1}$ . The computed  $D$ -value of  $0.281 \text{ cm}^{-1}$  is about half the experimental value, which is probably due to deficiencies of the employed functionals that often overestimate the delocalization of electrons [31,48].

At 4 K the measured EPR data indicate a triplet ground state with a bent geometry, although such a local minimum was not identified in our computations on the triplet surface. Hence, we expect that the rigidity of the matrix environment hampers the relaxation of the initially formed bent structure to the global minimum of the triplet surface. The computations summarized in Table S1 show that for the equilibrium structure of the  $S_{1,CS}$  the triplet state is about 0.1 eV higher in energy but already an optimization for the triplet at  $\angle CCC = 120^\circ$  reduces the energy difference to 0.01 eV. Interestingly, for  $\angle CCC = 130^\circ$  the triplet lies about 0.2 eV below the singlet. This suggests that at 4 K the  $C_3Ph_2$  relaxes to the triplet state from the initially populated  $S_{2,OS}$ . In the triplet state the  $\angle CCC$  value slightly increases due to the force induced by the shape of the triplet surface, but a complete relaxation to the linear structure is hindered by the matrix environment. The computed  $D$ -values for such structures (e.g.  $\angle CCC = 130^\circ$   $D = 0.263 \text{ cm}^{-1}$  for PBE/defTZVP) are essentially identical to the value obtained for the linear structure ( $D = 0.258 \text{ cm}^{-1}$ ). This explains the similarities of the measured  $D$ -values for 4 K and 70 K.

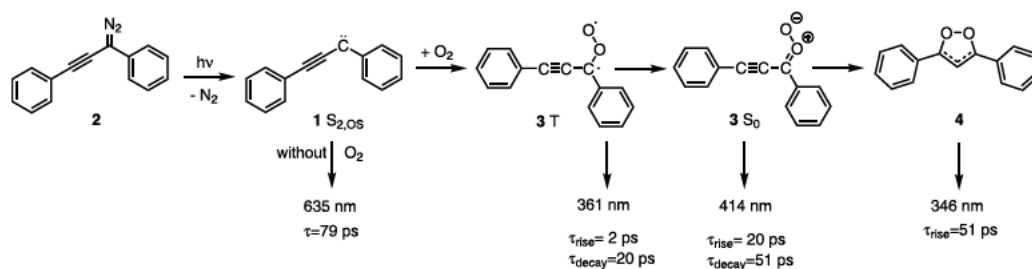
In the presence of air, the band at 635 nm is absent because the open-shell singlet state will react rapidly with an oxidizing agent like  $O_2$ . In the next section we will therefore outline a possible mechanism that explains the fate of **1** in DCM in the presence of oxygen. Two TA bands appear within a time scale described by  $\tau_2$  ( $\approx 2$  ps, Table 2) at  $\approx 414$  and 361 nm. The band at 361 nm decays (but remains non-zero), while a further band centered also around 414 nm grows in intensity with  $\tau_3 = 19.6$  ps and decays again with  $\tau_4 = 50.2$  ps (Fig. 3b,c black curves). Concomitantly with this decay, a band at 346 nm increases in intensity and exhibits a long lifetime (Fig. 3a, black curve). We assume that two different species contribute to this band and assign its late-time part to a reaction product of **1** with  $O_2$ .

Given that the open-shell singlet of **1** is formed in the bent geometry by the pump pulse, oxygen most likely reacts at the carbene center, forming initially the open-chain triplet **3** (biradicalic carbonyl oxide), see Scheme 2. This reaction has been computed to be exothermic by  $\Delta_R H(0 \text{ K}) = -1.09 \text{ eV}$ , thus considerable energy available for further reactions is present in the product. Table 4 characterizes the relevant states of **3** and **4** and indicates that the relaxation from the initially formed  $T_1$  of **3** to the  $S_0$  of **4** can proceed via two pathways, (a) ring closure via  $T_1$  of **4** ( $\Delta E = -0.74 \text{ eV}$ ) followed by ISC, or (b) ISC to  $S_0$  of **3** ( $\Delta E = -2.48 \text{ eV}$ ), followed by ring formation. Both could explain the TA signals found for DCM in air (Fig. 2). The final product for both pathways, the  $S_0$  state of the 1,2-dioxolane **4**, absorbs at about 323 nm with  $f = 1.11$ , we thus assign the peak at 346 nm that grows in with  $\tau_4 = 51$  ps at long delay times to the  $S_0$  state of **4**.

The other two contributions at 361 nm and 414 nm show both a rise and a decay and thus indicate reaction intermediates. The band around 361 nm appears within 2 ps and is thus assigned to the initially formed  $T_1$  state of **3**. The absorption computed at 350 nm with  $f = 0.11$  could be responsible for this band. It will also contribute to the band around 346 nm at early times. The band at 414 nm increases within 20 ps, i.e., the decay time of **3**, and is therefore due to an intermediate sequentially formed from **3**. The  $S_0$  state of **3** could contribute to this band, because we computed an absorption at about 371 nm for this state. The computed strong absorption of the  $T_1$  state of **4** at 335 nm would be shifted by more than 0.7 eV. Furthermore, we do not find an experimental counterpart to the fairly strong absorption at 610–620 nm ( $f = 0.27$  at 614 nm). Therefore, we assume that relaxation proceeds along the pathway outlined in Scheme 2, i.e. via  $S_0$  of **3**.

At early delay times, the broad absorption at 346–361 nm is associated with oscillations corresponding to vibrational motion with a wavenumber of  $81 \text{ cm}^{-1}$ . Two maxima are visible. As outlined above, **1** is formed in a bent geometry upon photolysis, whereas both  $T_0$  and  $S_{2,OS}$  have a linear equilibrium structure with the phenyl groups twisted with respect to each other. For species **3** on the other hand a bent geometry was computed, see Scheme 2. For the phenyl torsion, normal-mode analysis based on the molecular hessian predicts a frequency of  $11 \text{ cm}^{-1}$  in DCM, too low to explain the oscillations. We therefore assign them either to a bending motion of the phenyl group or a change of the C–O–O angle in the triplet state of **3** i.e., a bending-mode wavepacket that is only weakly coupled to other vibrations. The band at 414 nm, which corresponds to the  $S_0$  ground state of **3**, shows one maximum that is out of phase with the contribution at 346–361 nm, suggesting that the wavepacket survives the transition to the ground state at least briefly before it dephases.

We will now switch to a discussion of the TA spectra in ethanol. The spectra recorded in the presence of air ( $O_2$ ) show bands that strongly resemble those in DCM with  $O_2$ . In deaerated solution, bands are broader and band positions are slightly shifted in com-



Scheme 2. Possible reactions of **1** in DCM with and without oxygen.

L. Ress, D. Kaiser, J. Grüne et al.

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 254 (2021) 119606

**Table 4**  
Characterization of the electronic states of compounds **3** and **4**.

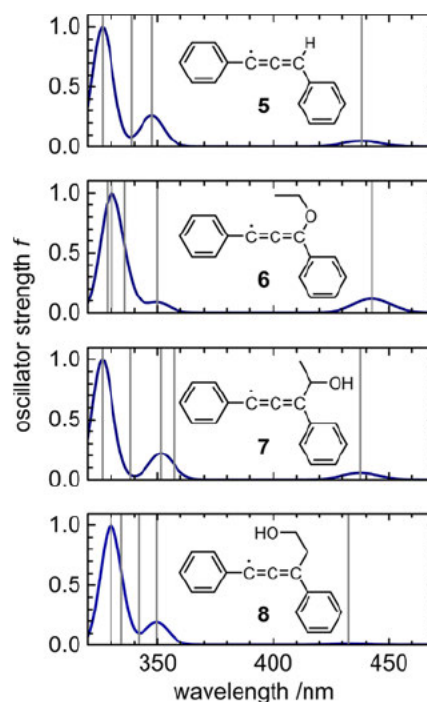
Compound /state	Rel. energy <sup>a)</sup> / eV	Character	$\lambda_{exc}$ /nm	$f^{b)}$
<b>3</b> / $S_0$	0.9	closed shell	487	0.0
			371	0.25
			309	0.84
			1211	0.0
<b>3</b> / $T_1$	2.48	open shell	442	0.09
			350	0.11
			349	0.01
			343	0.03
			325	0.82
			313	0.01
			311	0.0
<b>4</b> / $S_0$	0.0	closed shell	323	1.11
			489	0.0
<b>4</b> / $T_1$	1.74	open shell	1039	0.0
			614	0.27
			587	0.0
			409	0.0
			353	0.0
			342	0.0
			340	0.01
			335	0.67
			334	0.0
			333	0.0

<sup>a)</sup> With respect to the ground state of compound **4**.<sup>b)</sup> Only states with same multiplicity up to 300 nm included.

parison. No absorption that can be reasonably assigned to **1** is detected, which indicates that its  $S_{2,05}$  state reacts rapidly. The increase of the band at 343 nm with time is not as pronounced in ethanol at long delay times and the band at 415 nm displays a longer lifetime. Accordingly,  $\tau_4$  increases to 195 ps. At early times we also see oscillations in ethanol with a period of roughly 2.6 ps (Fig. 3a–c) that can also be assigned to a wave packet along the bending coordinate, following the arguments above given for DCM. The lower wavenumber reflects the different interactions with the solvent in ethanol.

The comparison of the spectra in the two solvents suggests that **1** also reacts with  $O_2$  in the presence of air. In deaerated solution, the time constants are slightly shorter and several bands overlap and extend from 350 to 430 nm. To rationalize the observations, we propose a mechanism that is based on computations guided by the earlier work of Rudolf et al. [24]. In this model, after photochemical generation in the pump step, carbene **1** abstracts a hydrogen atom from an ethanol molecule, leading to  $Ph_2C_3H$  radical **5** and an ethoxy ( $C_2H_5O$ ) radical (not shown). The previous work found a preference for H-abstraction at the hydroxyl group due to kinetic reasons, but H-abstraction at the  $CH_2$  or  $CH_3$  groups is thermodynamically favored. The reaction is so fast that **1** cannot be detected in the experiment. Subsequently, two further reactions are possible: The  $C_2H_5O$  radical might react with another carbene, leading to the products **6–8**, or alternatively a radical/radical recombination of **5** with  $C_2H_5O$  leads to products **9–12** (Table S3).

While for the closed-shell reaction products **9–12** and the possible  $C_2H_5O$  isomers no absorptions are predicted in the measured range, the compounds **5–8** all show absorption that would match observed TA bands. Their vertical absorption spectra are displayed in Fig. 6. All of them show absorption at 440 nm and 350 nm, close to observed TA bands. Isomer **7** is the thermodynamically most stable one and thus the most likely candidate. However, the energy differences are small. It is therefore difficult to assign the bands conclusively, because both SAS corresponding to the time constants  $\tau_3$  and  $\tau_4$  show absorption at 350 nm as well as at 420 nm (Fig. 4). We tentatively suggest that intermediate **5** is rapidly formed and decays in 13.5 ps due to further reactions possibly with the nearby  $C_2H_5O$  radical. Reaction products **6–8** are formed on this time scale and decay with  $\tau_4 = 152$  ps to a closed-shell reaction

Fig. 6. Computed absorption spectra of possible reaction products of **1** with ethanol.

product without significant absorption in the investigated absorption range. Such radical–radical recombination reactions are expected to be very efficient. Due to the larger number of possible products in that case a conclusive assignment is not possible.

As the bands in deaerated ethanol appear at wavelengths that are similar to those in the presence of  $O_2$ , regardless of the solvent, one also has to explore alternative explanations. One possibility would be a reaction of the biradicals with water traces, either in the solvents or in air. However, solvents were dried over molecular sieves and while TA spectra in ethanol appear similar, regardless of whether the experiments were carried out deaerated or under ambient conditions, the spectra recorded in DCM are drastically different. Thus the reaction with water traces is unlikely. A self-reaction of the biradicals should also be visible in deaerated DCM. We therefore propose a different explanation: When the possible reaction products **3** and **4** in DCM/ $O_2$  are compared with **5–8** in ethanol, it becomes evident that the chromophores resemble each other. Thus, similar absorption bands can be expected that are only slightly shifted in wavelength. For products **5–8** this is demonstrated in the computed spectra shown in Fig. 6. The similarity of the TA spectra in ethanol recorded in the presence or absence of oxygen is therefore due to the similar absorption spectra of the reaction products.

## 5. Summary and conclusion

By combining fs transient absorption spectroscopy and EPR spectroscopy with computations, we explored the photochemistry of the biradical diphenylpropynylidene in dichloromethane and ethanol to investigate the solvent dependence of its spin multiplicity and reactivity. The biradical was generated photolytically from

10



1,3-diphenyldiazopropyne by excitation to the  $S_2$  state around 300 nm. EPR spectra at 4 K revealed a triplet ground state in both solvents. Computations indicate that dissociation of the diazo group in the  $S_2$  state initially leads to a biradical in the  $S_{2,os}$  open-shell singlet state. This is in line with the non-zero rhombic zero-field splitting which showed that the biradical is initially formed in a bent geometry. Annealing to 70–80 K led to a linear carbon chain in DCM without oxygen. This agrees with the computations which predict the linear structure as global minimum of the triplet PES. The similarities of the D-parameters measured for 4 K and 70 K is reflected in the theoretical results which show the same trend. The similarity of the D-parameter measured in ethanol and DCM for 4 K indicate that pure solvent effects do not influence the electronic structure of  $C_3Ph_2$  considerably. The picture changes if a reaction with the environment becomes possible.

Transient absorption spectra in deaerated DCM showed a deactivation of the biradical to the  $T_0$  state within 80 ps, accompanied by a fast relaxation of the bent to the linear geometry at early delay times. In the presence of oxygen, a rapid reaction is observed that leads via a carbonyl oxide intermediate to a dioxolane, which does not decay further on the time scale of the experiment. Spectra in ethanol recorded in the presence of air resemble those in DCM strongly, thus the same mechanism should be active, although with slightly different time constants. In deaerated ethanol, reaction with the solvent is assumed to dominate. A number of possible products have been identified, but an unambiguous assignment is difficult because the computed absorption spectra are very similar. In ethanol, the differences between the spectra recorded in the presence and absence of air is small, which is plausible when the similarity of the chromophores of possible reaction products is considered. Oscillations in the TA spectra at short delay times are presumably due to bending-mode wave packets.

To summarize, in an unpolar and non-protic solvent like DCM the photodynamics of the biradical diphenylpropynylidene is similar to the one expected for the isolated molecule. In contrast, in a polar and protic solvent like ethanol, as well as in the presence of air the biradical reacts rapidly. However, in both solvents a triplet ground state is observed, thus no influence of the solvent on the spin multiplicity has been found.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the Research Training School "Molecular biradicals: Structure, properties and reactivity" (GRK 2112). We would like to thank Hans-Christian Schmitt for his contributions to the early phase of this work. The authors are grateful to Prof. W. Kiefer for his support, for many valuable scientific discussions over the years and for his continuing interest in the research of our institute. The interdisciplinary character of this paper reflects the collaborative approach of Wolfgang Kiefer's work, which contributed (and still contributes) to the success of our departments.

#### Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.saa.2021.119606>.

#### References

- [1] V. Bonacic-Koutecky, J. Koutecky, J. Michl, Neutral and charged biradicals, zwitterions, funnels in  $S_1$ , and proton translocation: their role in photochemistry, photophysics, and vision, *Angew. Chem. Int. Ed.* 26 (1987) 170–189.
- [2] M. Abe, *Diradicals*, *Chem. Rev.* 113 (2013) 7011.
- [3] T. Stuyver, B. Chen, T. Zeng, P. Geerlings, F. De Proft, R. Hoffmann, Do diradicals behave like radicals?, *Chem. Rev.* 119 (2019) 11291–11351.
- [4] M.A.H. Khan, C.J. Percival, R.L. Caravan, C.A. Taatjes, D.E. Shallcross, Criegee intermediates and their impacts on the troposphere, *Environ. Sci. Process Impacts* 20 (2018) 437–453.
- [5] O. Welz, J.D. Savee, D.L. Osborn, S.S. Vasu, C.J. Percival, D.E. Shallcross, C.A. Taatjes, Direct kinetic measurements of criegee intermediate ( $CH_2OO$ ) formed by reaction of  $CH_2I$  with  $O_2$ , *Science* 335 (2012) 204–207.
- [6] J.C. Johnson, A.J. Nozik, J. Michl, The role of chromophore coupling in singlet fission, *Acc. Chem. Res.* 46 (2013) 1290–1299.
- [7] T. Zeng, N. Ananth, R. Hoffmann, Seeking small molecules for singlet fission: a heteroatom substitution strategy, *J. Am. Chem. Soc.* 136 (2014) 12638–12647.
- [8] M.-A. Légaré, G. Bélanger-Chabot, R.D. Dewhurst, E. Welz, I. Krummenacher, B. Engels, H. Braunschweig, Nitrogen fixation and reduction at boron, *Science* 359 (2018) 896–900.
- [9] A. Deissenberger, E. Welz, R. Drescher, I. Krummenacher, R.D. Dewhurst, B. Engels, H. Braunschweig, A new class of neutral boron-based diradicals spanned by a two-carbon-atom bridge, *Angew. Chem. Int. Ed.* 58 (2019) 1842–1846.
- [10] C. Saalfrank, F. Fantuzzi, T. Kupfer, B. Ritschel, K. Hammond, I. Krummenacher, R. Bertermann, R. Wirthensohn, M. Finze, P. Schmid, V. Engel, B. Engels, H. Braunschweig, cAAC-stabilized 9,10-diboranthracenes-acenes with open-shell singlet biradical ground states, *Angew. Chem. Int. Ed.* 59 (2020) 19338–19343.
- [11] E. Welz, J. Bohnke, R.D. Dewhurst, H. Braunschweig, B. Engels, Unravelling the dramatic electrostructural differences between n-heterocyclic carbene- and cyclic (alkyl)(amino)carbene-stabilized low-valent main group species, *J. Am. Chem. Soc.* 140 (2018) 12580–12591.
- [12] G. Gryn'ova, M.L. Coote, C. Corminboeuf, Theory and practice of uncommon molecular electronic configurations, *WIRE Comput. Mol. Sci.* 5 (2015) 440–459.
- [13] H.H. Wenk, M. Winkler, W. Sander, One century of aryne chemistry, *Angew. Chem. Int. Ed.* 42 (2003) 502–528.
- [14] F. Hirsch, K. Pachner, I. Fischer, K. Issler, J. Petersen, R. Mitric, S. Bakels, A.M. Rijs, Do Xylylenes Isomerize in pyrolysis?, *ChemPhysChem* 21 (2020) 1515–1518.
- [15] B. Noller, M. Margraf, C. Schröter, T. Schultz, I. Fischer, Excited-state lifetime of propadienylidene, I-C<sub>3</sub>H<sub>2</sub> *Phys. Chem. Chem. Phys.* 11 (2009) 5353.
- [16] J. Giegerich, I. Fischer, J. Petersen, R. Mitric, The photodissociation dynamics of propargylene, HCCCH, *Phys. Chem. Chem. Phys.* 16 (2014) 6294–6302.
- [17] K. Issler, A. Röder, F. Hirsch, L. Poisson, I. Fischer, R. Mitric, J. Petersen, Excited state dynamics and time-resolved photoelectron spectroscopy of paraxylene, *Faraday Discuss.* 212 (2018) 83–100.
- [18] J. Wang, J. Kubicki, H. Peng, M.S. Platz, Influence of solvent on carbene intersystem crossing rates, *J. Am. Chem. Soc.* 130 (2008) 6604–6609.
- [19] J. Kubicki, Y.L. Zhang, S. Vyas, G. Burdzinski, H.L. Luk, J. Wang, J.D. Xue, H.L. Peng, E.A. Pritchina, M. Sliwa, G. Buntinx, N.P. Gritsan, C.M. Hadad, M.S. Platz, Photochemistry of 2-naphthoyl azide. An ultrafast time-resolved UV-Vis and IR spectroscopic and computational study, *J. Am. Chem. Soc.* 133 (2011) 9751–9761.
- [20] J.D. Xue, H.L. Luk, S.V. Eswaran, C.M. Hadad, M.S. Platz, Ultrafast infrared and UV-Vis studies of the photochemistry of methoxycarbonylphenyl azides in solution, *J. Phys. Chem. A* 116 (2012) 5325–5336.
- [21] J. Knorr, P. Sokkar, S. Schott, P. Costa, W. Thiel, W. Sander, E. Sanchez-Garcia, P. Nürnberger, Competitive solvent-molecule interactions govern primary processes of diphenylcarbene in solvent mixtures, *Nature Commun.* 7 (2016) 12968.
- [22] M. Steinbauer, M. Lang, I. Fischer, B.K. Cunha de Miranda, C. Romanzin, C. Alcaraz, The photoionisation of propargylene and diazopropyne, *Phys. Chem. Chem. Phys.* 13 (2011) 17956–17959.
- [23] E. Reusch, D. Kaiser, D. Schleier, R. Buschmann, A. Krueger, T. Hermann, B. Engels, I. Fischer, P. Hemberger, Pentadiynylidene and its methyl-substituted derivatives: threshold photoelectron spectroscopy of  $R_1-C_5-R_2$  triplet carbon chains, *J. Phys. Chem. A* 123 (2019) 2008–2017.
- [24] P. Rudolf, J. Buback, J. Aulbach, P. Nuernberger, T. Brixner, Ultrafast multisequential photochemistry of 5-diazo Meldrum's acid, *J. Am. Chem. Soc.* 132 (2010) 15213–15222.
- [25] M. Lang, F. Holzmeier, I. Fischer, P. Hemberger, Decomposition of diazomeldrum's acid: a threshold photoelectron spectroscopy study, *J. Phys. Chem. A* 118 (2014) 11235–11243.
- [26] J.T. DePinto, W.A. deProphetis, J.L. Menke, R.J. McMahon, Triplet 1,3-Diphenylpropynylidene (Ph-C-C-C-Ph), *J. Am. Chem. Soc.* 129 (2007) 2308–2315.
- [27] I.H.M. van Stokkum, D.S. Larsen, R. van Grondelle, Global and target analysis of time-resolved spectra, *Biochim. Biophys. Acta* 1657 (2004) 82–104.

L. Ress, D. Kaiser, J. Grüne et al.

*Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 254 (2021) 119606

- [28] J.J. Snellenburg, S. Laptinok, R. Seger, K.M. Mullen, I.H.M. van Stokkum, Glotaran: a java-based graphical user interface for the R package TIMP, *J. Stat. Softw.* 49 (2012) 1–22.
- [29] K.M. Mullen, I.H.M. van Stokkum, TIMP: an R package for modeling multi-way spectroscopic measurements, *J. Stat. Softw.* 18 (2007) 1–46.
- [30] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, O. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E. N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, *Gaussian 16 Rev. A.03*, Wallingford, CT, 2016.
- [31] S. Richert, C.E. Tait, C.R. Timmel, Delocalisation of photoexcited triplet states probed by transient EPR and hyperfine spectroscopy, *J. Magn. Reson.* 280 (2017) 103–116.
- [32] H.U. Suter, M.B. Huang, B. Engels, A multireference configuration-interaction study of the hyperfine-structure of the molecules CCO, CNN, and NCN in their triplet ground states, *J. Chem. Phys.* 101 (1994) 7686–7691.
- [33] S. Ruetzel, M. Diekmann, P. Nuernberger, C. Walter, B. Engels, T. Brixner, Multidimensional spectroscopy of photoreactivity, *Proc. Natl. Acad. Sci. USA* 111 (2014) 4764–4769.
- [34] P. Schmid, F. Fantuzzi, J. Klopff, N.B. Schröder, R.D. Dewhurst, H. Braunschweig, V. Engel, B. Engels, Twisting versus delocalization in CAAC- and NHC-stabilized boron-based biradicals: the roles of sterics and electronics, *Chem. Eur. J.* (2020).
- [35] J.D. Chai, M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections, *Phys. Chem. Chem. Phys.* 10 (2008) 6615–6620.
- [36] R. Krishnan, J.S. Binkley, R. Seeger, J.A. Pople, Self-consistent molecular orbital methods 20. Basis set for correlated wave-functions, *J. Chem. Phys.* 72 (1980) 650–654.
- [37] A.D. McLean, G.S. Chandler, Contracted Gaussian-Basis Sets for Molecular Calculations. 1. 2nd Row Atoms, *Z. Phys.* 72 (1980) 5639–5648.
- [38] J. Tomasi, B. Mennucci, R. Cammi, Quantum mechanical continuum solvation models, *Chem. Rev.* 105 (2005) 2999–3093.
- [39] T. Lu, F.W. Chen, Multiwfn: A multifunctional wavefunction analyzer, *J. Comput. Chem.* 33 (2012) 580–592.
- [40] F. Neese, Calculation of the zero-field splitting tensor on the basis of hybrid density functional and Hartree-Fock theory, *J. Chem. Phys.* 127 (2007).
- [41] F. Neese, Importance of direct spin-spin coupling and spin-flip excitations for the zero-field splittings of transition metal complexes: a case study, *J. Am. Chem. Soc.* 128 (2006) 10213–10222.
- [42] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* 77 (1996) 3865–3868.
- [43] A. Schafer, H. Horn, R. Ahlrichs, Fully optimized contracted Gaussian-basis sets for atoms Li to Kr, *J. Chem. Phys.* 97 (1992) 2571–2577.
- [44] A. Schafer, C. Huber, R. Ahlrichs, Fully optimized contracted Gaussian-basis sets of triple zeta valence quality for atoms Li to Kr, *J. Chem. Phys.* 100 (1994) 5829–5835.
- [45] F. Neese, Software update: the ORCA program system, version 4.0, *WIRE Comput. Mol. Sci.* 8 (2018).
- [46] S. Stoll, A. Schweiger, EasySpin, a comprehensive software package for spectral simulation and analysis in EPR, *J. Magn. Reson.* 178 (2006) 42–55.
- [47] T. Piteša, M. Alešković, K. Becker, N. Basarić, N. Došlić, Photoelimination of Nitrogen from Diazoalkanes: Involvement of Higher Excited Singlet States in the Carbene Formation, *J. Am. Chem. Soc.* 142 (2020) 9718–9724.
- [48] D. Hait, M. Head-Gordon, Delocalization errors in density functional theory are essentially quadratic in fractional occupation number, *J. Phys. Chem. Lett.* 9 (2018) 6280–6288.



## 7. The Bioconjugation Reaction of Triazolinediones (TADs) with Tyrosine

JOC

The Journal of Organic Chemistry

Cite This: *J. Org. Chem.* 2018, 83, 10248–10260

Article

pubs.acs.org/joc

### Mechanistical Insights into the Bioconjugation Reaction of Triazolinediones with Tyrosine

Dustin Kaiser,<sup>†</sup> Johan M. Winne,<sup>‡</sup> Maria Elena Ortiz Soto,<sup>§</sup> Jürgen Seibel,<sup>§</sup> Thien A. Le,<sup>†</sup> and Bernd Engels<sup>\*,†,§</sup>

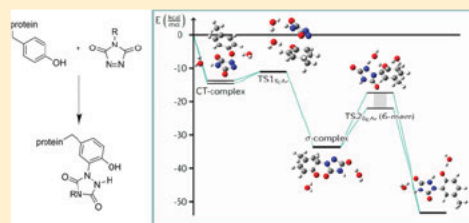
<sup>†</sup>Institute of Physical and Theoretical Chemistry, Julius-Maximilians Universität Würzburg, Emil-Fischer-Str. 42, Würzburg 97074, Germany

<sup>‡</sup>Department of Organic and Macromolecular Chemistry, Ghent University, Krijgslaan 281-S4, B-9000 Ghent, Germany

<sup>§</sup>Institute of Organic Chemistry, Julius-Maximilians Universität Würzburg, Am Hubland, Würzburg 97074, Germany

Supporting Information

**ABSTRACT:** The bioconjugation at tyrosine residues using cyclic diazodicarboxamides, especially 4-substituted 3*H*-1,2,4-triazole-3,5(4*H*)-dione (PTAD), is a highly enabling synthetic reaction because it can be employed for orthogonal and site-selective (multi)functionalizations of native peptides and proteins. Despite its importance, the underlying mechanisms have not been thoroughly investigated. The reaction can proceed along four distinctive pathways: (i) the  $S_NAr$  path, (ii) along a pericyclic group transfer pathway (a classical ene reaction), (iii) along a stepwise reaction path, or (iv) along an unusual higher order concerted pericyclic mechanism. The product mixtures obtained from reactions of PTAD with 2,4-unsubstituted phenolate support the  $S_NAr$  mechanism, but it remains unclear if other mechanisms also take place. In the present work, the various mechanisms are compared using high-level quantum chemistry approaches for the model reaction of 4*H*,3*H*-1,2,4-triazole-3,5(4*H*)-dione (HTAD) with *p*-cresol and *p*-cresolate. In a protic solvent (water), the barriers of the  $S_NAr$  mechanism and the ene reaction are similar but still too high to explain the available experimental observations. This is only possible if the  $S_NAr$  reaction of cresolate with HTAD is taken into account for which nearly vanishing barriers are computed. This satisfactorily explains measured conversion rates in buffered aqueous solutions and the strong activation effects observed upon addition of bases.



#### INTRODUCTION

Triazolinediones, such as 4-phenyl-3*H*-1,2,4-triazole-3,5(4*H*)-dione (PTAD), are very useful reagents in chemistry which are used in thermo<sup>1–16</sup> as well as photochemical reactions.<sup>17–21</sup> PTAD is also an important tool for mass spectrometry analysis, for example, for identifying fatty acids.<sup>22,23</sup> In 2009, Barbas and co-workers extended the application range of cyclic diazodicarboxamides. They published the first report of selective bioconjugation at native tyrosine residues of proteins using cyclic diazodicarboxamides, specifically 4-phenyl-3*H*-1,2,4-triazole-3,5(4*H*)-dione (PTAD).<sup>24</sup> They showed that the reaction is highly selective for tyrosine to the extent that it can be employed for orthogonal and site-selective multifunctionalizations.<sup>24,25</sup> Since then, this clicklike reaction involving triazolinediones has proven to be a promising approach, as it has been employed in biorthogonal bioconjugation<sup>25–27</sup> as well as in macromolecular sciences.<sup>28</sup> Furthermore, the tyrosine click reaction with TADs has been used to prepare DNA–protein conjugates, to couple drugs to monoclonal antibodies, and to prepare glycoconjugates.<sup>29</sup> It was recently shown that diazodicarboxamide bioconjugation

can also be used to redesign and tailor the enzyme levansucrase for the synthesis of novel polysaccharides.<sup>30</sup>

Despite its importance, the mechanism of the reaction between PTAD and tyrosine residues was not investigated in detail; e.g., theoretical characterizations are missing to the best of our knowledge. The most recent discussion about TAD reactions is the comprehensive review by De Bruycker et al.<sup>29</sup> Therein, as in the original report of Barbas, the TAD reaction with phenols is denoted as an “ene reaction” but no detailed discussion (or speculation) about its possible mechanistic steps is provided. Tyrosine derivatives react rapidly (complete in 5 min) with 3 equiv of PTAD in sodium phosphate buffer with pH = 7 and a 1:1 mixture with acetonitrile. There are several key experimental observations that are indicative of significant mechanistic effects that deserve a close investigation. Importantly, the reaction between TAD reagents and phenols in organic solvents is exceedingly slow. For peptide or protein conjugations, TADs are more reactive with free amines and tryptophan residues. In aqueous medium, the adduct formation

Received: June 7, 2018

Published: July 13, 2018

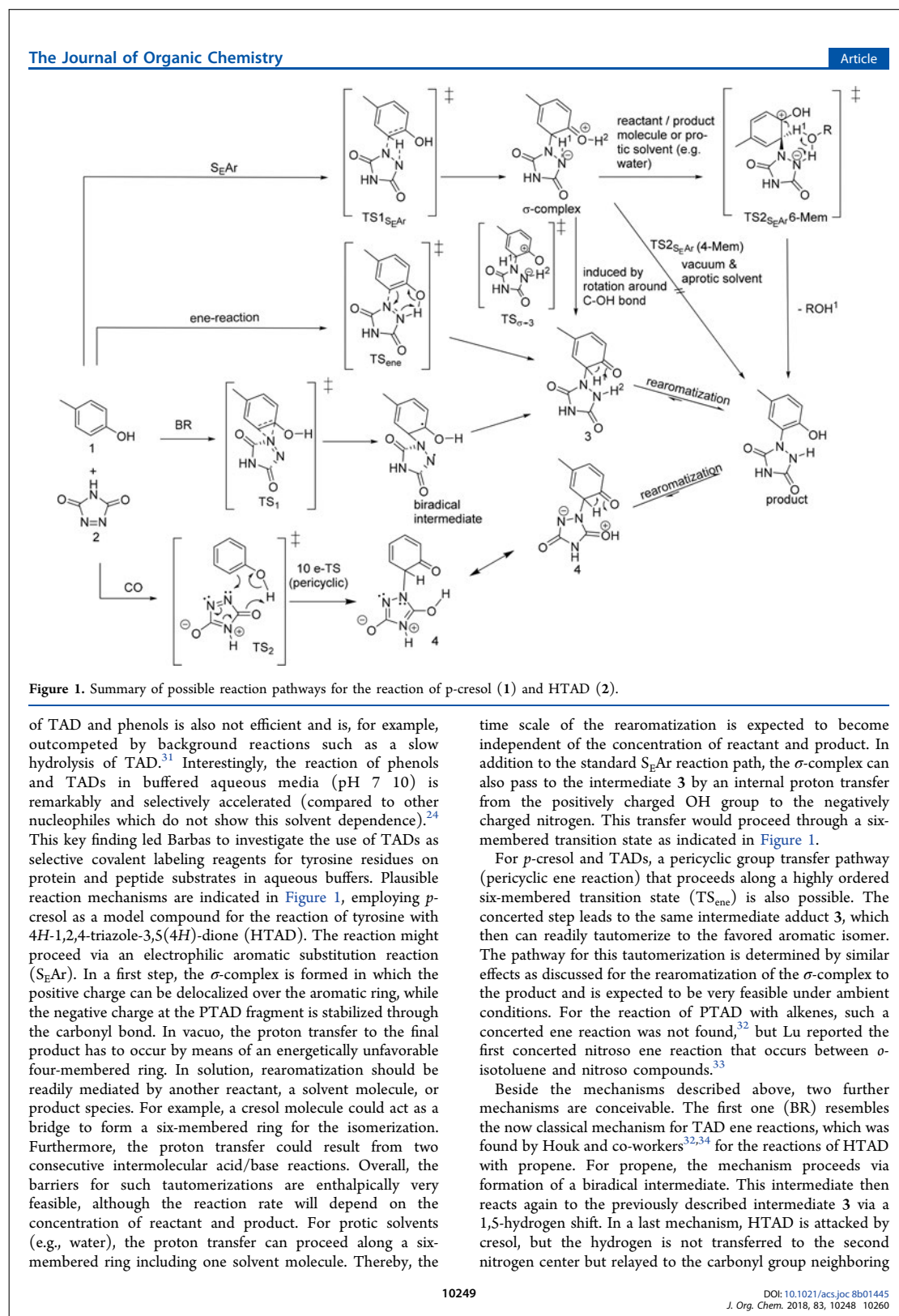


ACS Publications

© 2018 American Chemical Society

10248

DOI: 10.1021/acs.joc.8b01445  
*J. Org. Chem.* 2018, 83, 10248–10260



**Table 1. Relative Energies (kcal/mol) with Respect to the Fragments (to the CT Complex) of the Stationary Points of the Various Reaction Paths of the Reaction of HTAD with *p*-cresol<sup>d</sup>**

	CCSD(T) in vacuo	CCSD(T) + PCM	CCSD(T) + PCM + H <sub>2</sub> O
CT complex	-12.1	-10.0	-9.3 <sup>c</sup> / -8.8 <sup>d</sup>
TS <sub>S<sub>E</sub>AR</sub>	16.2 (28.3)	14.4 (24.4)	14.4 (23.2)
$\sigma$ -complex	16.2 (28.3)	12.6 (22.6)	11.3 (20.1)
TS <sub><math>\sigma</math>-3</sub>	8.2 (20.3)	22.7 (32.7)	21.7 (30.5)
TS <sub>S<sub>E</sub>AR</sub> (4-mem)	nc <sup>b</sup>	nc <sup>b</sup>	nc <sup>b</sup>
TS <sub>S<sub>E</sub>AR</sub> (6-mem)			8.1 (17.4) <sup>c</sup> / 14.8 (23.6) <sup>d</sup>
TS <sub>ene</sub>	14.4 (22.9)	10.8 (20.8)	11.2 (20.0)
TS <sub>Solvent</sub> <sup>e</sup>			15.9 (25.2) <sup>c</sup> / 22.6 (31.4) <sup>d</sup>
3	-24.5 (-16.0)	-20.1 (-10.1)	-18.9 (-10.1)
TS <sub>1</sub>	27.4 (35.9)	24.8 (34.8)	24.6 (33.4)
TS <sub>2</sub>	23.1 (31.6)	21.7 (31.7)	20.1 (28.9)
4	-0.8 (7.7)	-2.8 (8.2)	nc <sup>b</sup>
product	-43.3 (-32.8)	-35.9 (-25.9)	-34.9 (-26.1)

<sup>a</sup>For all calculations, the aug-cc-pVDZ basis sets were used. The geometries were optimized using the  $\omega$ B97xD DFT functional. In addition to the CCSD(T) results, CASPT2(14,14) and  $\omega$ B97xD energetics have been calculated (see Tables S4–S6 in the SI). <sup>b</sup>Not computed. <sup>c</sup>Relative energies with respect to fragments or CT complex which are not stabilized by a water molecule. <sup>d</sup>Relative stability with respect to the fragments (or CT state) which are stabilized by a water molecule. <sup>e</sup>See Figure 6.

to the electrophilic nitrogen center (CO mechanism). Through a kind of pericyclic 10 e-TS<sup>35–37</sup> the proton transfer leads to intermediate 4. The final product is obtained from 4 by a double tautomerization reaction. This mechanism is quite unusual but was already discussed by Houk for the reaction of HTAD with propene.<sup>32</sup>

Previous experimental findings of Barbas and co-workers indicate that the S<sub>E</sub>Ar mechanism most likely takes place.<sup>24,38</sup> While conversions to the product were found for aqueous buffered solutions (NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>/MeCN buffer, pH 7.4), no reactions are found if MeCN or a mixture of MeCN and H<sub>2</sub>O (1:1) are taken as solvents. Additionally, further literature examples showed that the reaction also takes place for phenolate for which no ene reaction is possible.<sup>39</sup> Nevertheless, the experiments do not exclude competing mechanisms in different environments.

To shed some light on this important topic, we computed the various possible mechanisms for the reaction of *p*-cresol with HTAD for different environments in the present work. In addition to vacuum calculations, we also included the influence of a polarizable environment through the polarizable continuum model (PCM). In addition, we characterized the influence of explicit solvent effects by adding water molecules. Water may catalyze the S<sub>E</sub>Ar path considerably because of its hydrogen-bond-donating capacity (Figure 1). To ensure that the mechanisms are described on equal footing, we employed CCSD(T) as well as CASPT2 beside DFT. The multireference approach CASPT2 was used because the intermediates of the stepwise reaction of HTAD with propene possessed severe biradical character.<sup>40</sup> We mainly report CCSD(T) energetics in this report, while comprehensive results of all computational methods can be found in the accompanying Supporting Information. Using these high-level approaches as benchmark, we investigated the applicability of the density functional approach. We extended the investigation toward the reactions of the cresolate anion to study how a basic environment influences the barriers of the S<sub>E</sub>Ar mechanism. In order to investigate which effects determine the energetics, we studied the influence of the aromatic system of cresol by comparing its reaction with the corresponding reactions of vinyl alcohol. For the reaction of TADs with xylene, mesitylene, or anisole,

Breton and Hoke have recently provided conclusive evidence that a radical mechanism takes place.<sup>19,41</sup> To elucidate possible differences between OH and CH<sub>3</sub> substituents, we further compared the reactions of HTAD with vinyl alcohol and propene.

## COMPUTATIONAL METHODS

All geometries were obtained at the B3LYP<sup>42–44</sup>/aug-cc-pVDZ<sup>45</sup> and  $\omega$ B97xD<sup>46</sup>/aug-cc-pVDZ<sup>45</sup> level of theory using the Gaussian09 Rev. E software package.<sup>47</sup> For  $\omega$ B97xD optimizations, structures have previously been optimized at the B3LYP/aug-cc-pVDZ level of theory. In the case of the TAD ene reaction with *p*-cresol and vinyl alcohol, the B3LYP potential energy surface (PES) includes superfluous stationary geometries near transition states. This failure of B3LYP for some organic chemical reactions has previously been reported.<sup>48,49</sup>

For all stationary geometries, frequency calculations were performed to validate them as true minima or first-order saddle points of the PES. For transition states, intrinsic reaction coordinate (IRC) calculations were performed to find the minima connected by the transition state. All energy calculations at the  $\omega$ B97xD, MP2, and CCSD(T) levels of theory were performed using Gaussian09.  $T_1$  diagnostics have been obtained from converged CCSD amplitudes.<sup>50</sup> Single reference approaches are often sufficiently accurate; however, in some cases, multireference approaches are needed to obtain accurate PES,<sup>51</sup> electronically excited states, or even properties.<sup>52</sup> On the basis of the previous studies of Houk and co-workers, for the present study multireference effects could be expected to be important.<sup>32</sup> Hence, we used the Complete Active Space (CAS) approach in which dynamic correlation effects are taken into account in second-order perturbation theory (PT2).<sup>53,54</sup> All CASPT2 calculations were performed using the MOLCAS 7.8 program package.<sup>55</sup> The active space of the CASPT2(14,14) calculations was built iteratively starting from canonical HF orbitals in the [4,4] case and building larger CAS spaces from pseudonatural average orbitals of each smaller CI space. The energies of separated reactants were calculated for both molecules spaced more than 10 Å apart in the case of CASPT2 calculations. Solvent effects were included using the Polarizable Continuum Model (PCM), as implemented in Gaussian09 Rev. E and MOLCAS 7.8. For the CT states and all intermediates we checked for alternating open-shell species using unrestricted singlet and triplet optimizations at the  $\omega$ B97xD<sup>46</sup>/aug-cc-pVDZ<sup>45</sup> level of theory. In no case were relevant open-shell species

observed. Details regarding the influence of the considered open-shell structures are found in Table S7.

## RESULTS AND DISCUSSION

As a first step, we investigated possible paths of the reaction of *p*-cresol with HTAD which are summarized in Figure 1. Within this search, which was performed on the  $\omega$ B97xD/aug-cc-pVDZ level of theory, we could indeed identify the corresponding stationary points of all four different mechanisms. The geometries of the various stationary points are given in Figures S1–S4. Additional single-point energy differences of these geometries were computed using CCSD(T) and CASPT2 to elucidate the influence of the theoretical approach on the relative energies (see the first column of Table 1 and Table S4). All energies are given with respect to the separated reactants. For CASPT2, which is not strictly size consistent,<sup>56</sup> the reactants were computed within a distance of 10 Å.

Our calculations indicate that the fragments form a stable prereactive complex. To check the accuracy of this binding event, we computed the basis set superposition error (BSSE). It is only about 1 kcal/mol at the  $\omega$ B97xD/aug-cc-pVDZ and 5.5 kcal/mol on the CCSD(T)/aug-cc-pVDZ level of theory, showing that the binding is not an artifact of the calculation. To investigate if this complex corresponds to the charge-transfer (CT) complex described previously by Breton and co-workers<sup>18,19</sup> and also by other groups,<sup>6,57</sup> we computed its lowest lying electronically excited states for polar media employing PCM. These computations were performed using the TDDFT approach at the  $\omega$ B97xD/aug-cc-pVDZ level of theory because range-separated functionals are needed for the description of CT states.<sup>58–60</sup> This level of theory predicted electronically excited states at 469 and 425 nm and various states between 350 and 320 nm with quite low or vanishing transition dipole moments. Taking the standard error of 0.1 eV into account, these findings agree nicely with the broad band (430–350 nm) described by Breton and Hoke for MeTAD and substituted benzenes. Hence, in the following passages this complex is named CT complex and the discussion will mainly focus on the relative energies with respect to this CT state. It is formed without any barriers, and due to the subsequent high reaction barriers, its lifetime is sufficiently long for the binding energy to be dissipated before the reaction takes place. The binding of the CT state mainly results due to dispersion effects as shown by comparing the binding energies obtained with B3LYP/aug-cc-pVDZ (0.9 kcal/mol) and with B3LYP-D3/aug-cc-pVDZ (7.6 kcal/mol). This is in line with previous investigations of Scheiner and co-workers.<sup>61</sup> It should also be noted that due to the computed basis set superposition error (BSSE) the relative heights of the reaction barriers are overestimated by 1–5 kcal/mol depending on the theoretical approach.

All hitherto-reported experimental investigations of TAD-tyrosine click reactions are performed in polar protic solutions. Nevertheless, we will briefly discuss the results obtained for vacuum conditions to see the influence of polar aprotic or protic solvent and to investigate differences to the reaction between HTAD and propene which were characterized in vacuo. In the CT complex found for the vacuum computations, both fragments are separated by about 2.75 Å. They are in a face-to-face orientation but slightly shifted with respect to each other so that the N–N bond of HTAD lies above the CO bond of the cresol (see Figure S1 and Figures 3–5). The

computed binding energies of the CT complex vary between 12 kcal/mol for CCSD(T) and +5 kcal/mol for CASPT2-(14,14). Due to the relative orientation of the two fragments with respect to each other two CT complexes with different geometries are possible. In the first one, the OH moiety is placed below the five-membered ring of HTAD while it is oriented away from it in the second geometry. The influence of both orientations on the reaction mechanism will be discussed below because most experiments were performed in solution. In summary, our computations predict that in vacuo the pericyclic ene mechanism is energetically most favorable. The computed barrier heights ( $TS_{\text{ene}}$ ) with respect to the CT complex vary between 21 kcal/mol (CCSD(T)) and about 8 kcal/mol (CASPT2(14,14)). The differences may arise because as indicated by the  $\mathcal{T}_1$ -diagnostic multireference effects seem to be non-negligible. Similar variations are found for all other stationary states.

While the absolute values of the computed barrier heights differ, all approaches agree that the corresponding barriers for the  $S_{\text{E}}\text{Ar}$  mechanism ( $TS1_{S_{\text{E}}\text{Ar}}$ ) are considerably higher than those found for the pericyclic ene reaction (see Table S4). All approaches also agree that the barriers for the two stepwise reactions are even higher than those of the  $S_{\text{E}}\text{Ar}$  barriers. It is important to note that the pericyclic ene reaction might only proceed to the intermediate 3 because the proton transfer necessary to reach the final product will have a quite high barrier in vacuo. This is indicated by the failure of all our attempts to determine the corresponding transition states. We refrain from further investigations because we are mainly interested in the reactions taking place in aqueous solution.

The data given in the second column of Table 1 contains solvent effects within the framework of the PCM continuum approach which includes polarization effects but neglects molecular effects of a given solvent. The structures of the stationary points are given in Figure S2. In the CT complex, with a binding energy varying between 6 and 12 kcal/mol, both fragments are separated by about 3.2 Å (see the Supporting Information); i.e., the separation is slightly larger than for a vacuum. The energy difference between the two possible orientations of the CT complexes is only 0.2 kcal/mol (6.1 vs. 6.3 kcal/mol). This could be expected because the binding energy mainly results from dispersion effects and donor/acceptor interactions which should not be influenced by the orientation of the OH moiety. A switch from one minimum to the other minimum is possible through a rotation of the OH group toward the HTAD fragment or away from the HTAD fragment. To gain insights if the transfer is a fast motion, we computed the barrier of the rotation away from HTAD because the system approaches the transition state of the pericyclic ene reaction if the OH group is rotated toward the HTAD. The barrier of the former rotation is about 4 kcal/mol. The small barrier indicates that the resulting fast equilibrium will not influence which path is taken according to the Curtin–Hammett principle.<sup>62</sup>

As expected, the inclusion of solvent effects stabilizes the stationary points with a large degree of charge separation along the  $S_{\text{E}}\text{Ar}$  pathway. In comparison to the gas phase the  $TS1_{S_{\text{E}}\text{Ar}}$  is stabilized by about 3–4 kcal/mol (relative energy with respect to the CT complex), while the corresponding relative stabilization of the  $\sigma$ -complex is even larger. Starting from the  $\sigma$ -complex, the  $S_{\text{E}}\text{Ar}$  mechanism can proceed along three different paths (Figure 1). The first one proceeds along a four-



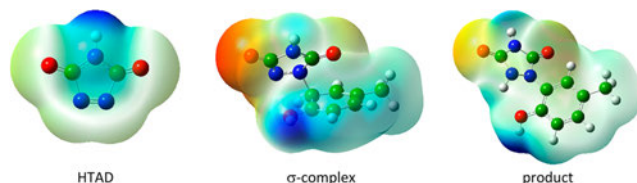


Figure 2. Surface charge densities for the three selected stationary points of the  $S_{EAr}$  reaction path.

membered transition state going directly to the final product. It is expected to be a very unfavorable one. Indeed, a constraint optimization procedure varying the bond length of the C–H bond of the  $\sigma$ -complex incrementally at the  $\omega$ B97xD/aug-cc-pVDZ + PCM level of theory predict a barrier height of 36.7 kcal/mol.

The proton transfer from the formally positively charged OH moiety to the formally negatively charged nitrogen of the HTAD, which represents a connection between the  $\sigma$  complex and intermediate 3, is the second pathway of the  $S_{EAr}$  mechanism. This route passes through a relaxed six-membered ring as indicated in Figure 1. Starting from the situation in which the proton  $H^+$  (see Figure 1) is oriented away from the HTAD moiety the reaction is induced by a rotation of the OH group around the CO bond. Our computations indicate that this rotation possesses a barrier of about 23 kcal/mol while the subsequent proton transfer possess no barrier. This shows that a  $\sigma$ -complex in which the OH group is oriented toward the HTAD will not be stable but will directly proceed to 3. From a closer look on the orientation of the transition state of the pericyclic ene reaction ( $TS_{ene}$ , see Figure S2) it becomes clear that such an orientation of the OH group indeed closely resembles the transition state of the pericyclic ene-mechanism (see below). From 3, the final product is reached through the tautomerization from the keto to the enol form which will be discussed in connection with the pericyclic ene mechanism.

As a final branch in the  $S_{EAr}$  mechanism, the proton transfer leading to the final product can proceed through a six-membered ring which within an aprotic solvent includes a reactant or a product molecule as bridge ( $TS_{S_{EAr}}$  (6-mem)). The barrier height of this proton transfer will be similar to the barrier height including water as a bridge (10–14 kcal/mol), which is shown in Table 1. This is considerably lower than the barrier height of the  $TS_{S_{EAr}}$  transition state (20–25 kcal/mol) so that entropic effects are not expected to shift the relative energy of the  $TS_{S_{EAr}}$  (6-mem) above that of the  $TS_{S_{EAr}}$  transition state. Considering that the final product could also be formed within two consecutive intermolecular acid/base reactions the formation of the  $\sigma$ -complex is predicted to be the rate-determining step of the  $S_{EAr}$  mechanism in aprotic solvents. The energetic and structural properties of the  $S_{EAr}$  mechanisms which were determined for the reaction of TAD and *p*-cresol are showcased in Figure 2.

While the barrier to the  $\sigma$ -complex is considerably lowered by polar solvents, the barrier of the pericyclic ene reaction is influenced to a lesser extent. Consequently, while the pericyclic ene reaction was by far the most favorable pathway in vacuo, the differences shrink for a polar environment. Again, although the computed barriers vary to some extent, all methods agree that the barrier heights of the first step of the ene- and  $S_{EAr}$  mechanism differ only by 3–4 kcal/mol. Additionally, all

methods predict that the  $TS_{ene}$  is lower in energy than  $TS_{S_{EAr}}$ . They also agree that the barriers of the two remaining reaction mechanisms are considerably higher in energy. Within the pericyclic ene mechanism the final product is formed from intermediate 3 by a keto–enol tautomerization. As it will not take place in vacuo, but in solution, the same arguments previously discussed for the final step of the  $S_{EAr}$  mechanism apply. We therefore conclude that this step is not kinetically important as well. Hence, the barrier ( $TS_{ene}$ ) leading to the intermediate 3 is predicted to be the rate-determining step of the pericyclic ene mechanism. This is underlined by the fact that the high barrier (>30 kcal/mol) prevents any backward reactions from 3 at room temperature. This may lead to an initial mixture of ketone and phenol forms in aprotic solvents if the product tautomerization is slow. The mechanistic pathway of the ene reaction is showcased in Figure 3.

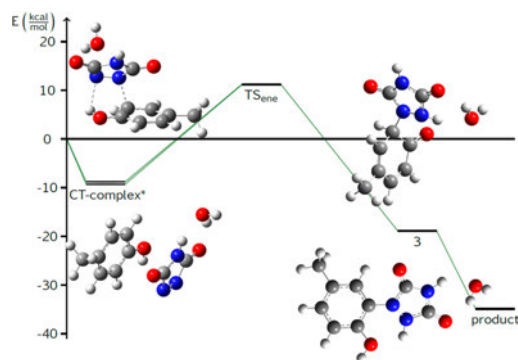
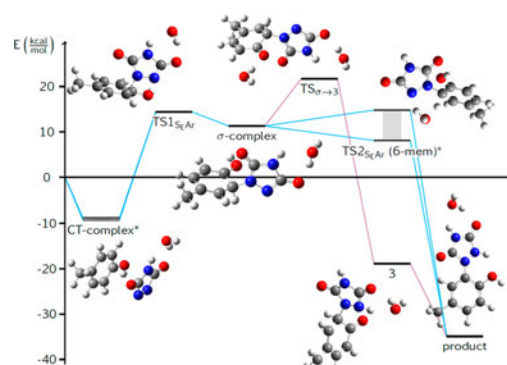


Figure 3. Intermediates and transition states present in the ene reaction pathway of the reaction of HTAD with *p*-cresol. Energies are taken from Table 1 at the CCSD(T) + PCM +  $H_2O$  level of theory. \*For the CT complex, two energy estimates depending on the placement of the explicit water molecule have been calculated. They serve as an upper and lower error bound. For more information, see the text and Table 1.

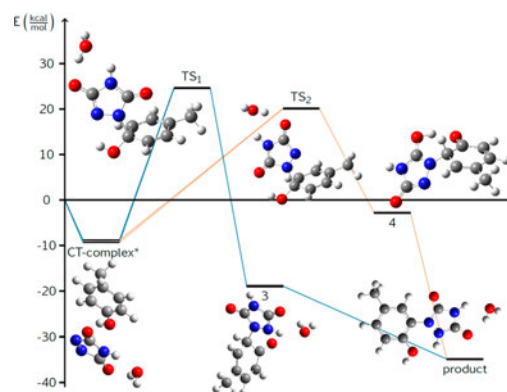
We are mainly interested in the reaction of PTAD derivatives with tyrosine within aqueous solutions. Hence, it is necessary to investigate if a water molecule catalyzes the various mechanisms differently. For the  $S_{EAr}$  mechanism (Figure 1) a protic solvent simplifies the formation of a 6-membered transition structure for the proton transfer connecting the  $\sigma$ -complex with the final product due to the excess of solvent molecules. To estimate this barrier height, we computed the  $TS_{S_{EAr}}$  (6-mem) using a water molecule as model for the bridging molecule. In such calculations, it is very important to balance the stabilization effects of the single water

molecule for the different stationary points. This becomes obvious from variations in the charge distributions for selected stationary points of the  $S_{E}Ar$  mechanism which are summarized in Figure 2. They indicate that the charges of the carbonyl moieties vary considerably along the  $S_{E}Ar$  mechanism. Hence, if the additional water molecule forms a hydrogen bridge to the carbonyl center whose charge varies mostly along the reaction path the reactivity of the HTAD should be influenced. For the  $TS2_{S_{E}Ar}$  (6-mem) transition state, however, the water molecule has to be involved in the 6-membered transition state. Consequently, the relative energy of this transition state is overestimated if the water molecule is connected with the carbonyl group for all other stationary points. The error is reduced because we include polarization effects by the PCM continuum approach. To estimate connected error bars, we give two relative energies for the  $TS2_{S_{E}Ar}$  (6-mem) transition state. As an upper bound for the barrier height of the second step of the  $S_{E}Ar$  mechanism we give the relative energy of the  $TS2_{S_{E}Ar}$  (6-mem) transition state with respect to the fragments HTAD + cresol in which the water molecule is placed at one carbonyl group of HTAD. This can be considered as an upper bound because due to the position of the water molecule the fragments should be more stabilized than the  $TS2_{S_{E}Ar}$  (6-mem) transition state. As a lower bound we give the relative energy of the transition state with respect to the fragment energy without the stabilization through the additional water molecule. This is a lower bound, because only the  $TS2_{S_{E}Ar}$  (6-mem) transition state undergoes some stabilization due to the water molecule. Comparable values are given for the relative energies with respect to the CT complex. They will be discussed below. The same considerations hold for a seven-membered ring of the  $TS_{Solvent}$  (see Figure 6). For all other stationary points, the additional water molecule is always positioned at the carbonyl group possessing the highest negative partial charge. The structures of the various stationary points are given in the Figure S3. The various conceivable mechanisms are visualized including the error estimates in Figures 3–5.

According to  $\omega B97xD$  and CCSD(T), the stability of the CT complex does not change much if a water molecule is added to the carbonyl group of HTAD. Without any water (see second column of Table 1 and Table S5) the computed values vary between 6 kcal/mol ( $\omega B97xD$ ) and 10 kcal/mol (CCSD(T)), while 7 and 10 kcal/mol are computed if one water molecule is added (Table 3). This is expected because the stability of the CT complex results from dispersion effects which are not strongly influenced by the additional water molecule. One might expect that the additional water molecule stabilizes the  $\sigma$ -complex with respect to the CT complex, but these effects are small. Using water as a model for the bridging molecule,  $\omega B97xD$  computations predict the lower and upper bounds for the height of the  $TS2_{S_{E}Ar}$  (6-mem) transition state with respect to the CT complex to be 7 and 12 kcal/mol, respectively (see third column of Table 1 and Table S6). The CCSD(T) approach predicts a higher barrier (17.4 and 23.6 kcal/mol for lower and upper limit, respectively), while CASPT2(14,14) comes to a similar conclusion as  $\omega B97xD$  (7.5 and 14.1 kcal/mol, respectively). All approaches predict that the barrier for the proton transfer connecting the  $\sigma$ -complex with intermediate 3 is higher than the  $TS2_{S_{E}Ar}$  (6-mem) transition state; i.e., the reaction will proceed along the



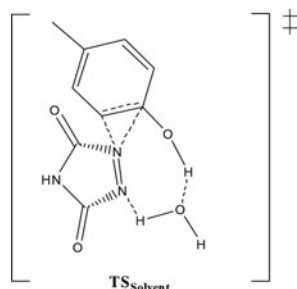
**Figure 4.** Intermediates and transition states occurring for the  $S_{E}Ar$  mechanisms of the bioconjugation reaction. Energies are taken from Table 1 at the CCSD(T) + PCM + H<sub>2</sub>O level of theory. \*For the CT complex and  $TS2_{S_{E}Ar}$  (6-mem) two energy estimates depending on the placement of the explicit water molecule have been calculated. They serve as an upper and lower error bound (gray bar for  $TS2_{S_{E}Ar}$  (6-mem)). For more information, see the text and Table 1.



**Figure 5.** Intermediates and transition states present in the stepwise reaction pathways of the reaction of HTAD with *p*-cresol. Energies are taken from Table 1 at the CCSD(T) + PCM + H<sub>2</sub>O level of theory, except for 4, in which the CCSD(T) + PCM energy difference is displayed. \*For the CT complex, two energy estimates depending on the placement of the explicit water molecule have been calculated. For more information, see the text and Table 1.

latter TS. Summarizing, all theoretical approaches agree that the formation of the  $\sigma$ -complex remains the rate-dependent step in protic solvents because the height of the  $TS1_{S_{E}Ar}$  barrier with respect to the CT complex is considerably higher ( $\omega B97xD$ : 23.2 kcal/mol; CCSD(T): 23.2 kcal/mol; CASPT2 (14,14) 20.6 kcal/mol) than the subsequent barriers.

As expected, the relative transition-state energies of the pericyclic ene mechanism with respect to the CT complex are only slightly influenced by an additional water molecule. We also tested if a bridging water molecule leads to a lower barrier for the ene mechanism ( $TS_{Solvent}$  Figure 6); however, the seven-membered transition state is similarly high or even higher than the six-membered concerted transition state

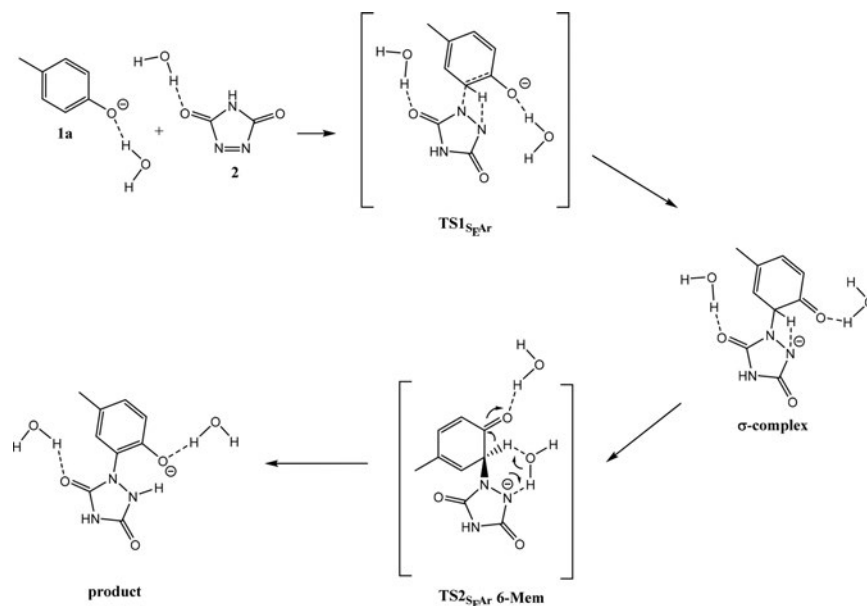


**Figure 6.** Seven-membered transition state of the ene reaction involving one water molecule as bridge.

without an additional water molecule. Consequently, also in protic solvents the barrier for the formation of the intermediate **3** remains so high that only a slow reaction is expected. This agrees well with the available experimental data, as we will discuss below. Finally, the additional water molecule also does not change the barriers heights of the stepwise mechanisms (see [Figure 5](#)) with respect to the CT complex so that they remain considerably less favorable than the pericyclic ene or the  $S_{\text{E}}\text{Ar}$  reaction (see [Figures 3](#) and [4](#)). In summary, for protic solvents the pericyclic ene reaction remains slightly more favorable than the  $S_{\text{E}}\text{Ar}$  reaction; however the computed barriers (20–21 kcal/mol) are quite high so that only very slow reactions are expected.

Experimental data indicate that PTAD also reacts with cresolate or phenolate,<sup>38</sup> for which no ene reaction is possible. To investigate differences between cresol and cresolate we also characterize the  $S_{\text{E}}\text{Ar}$  reactions of *p*-cresolate with HTAD

([Figure 7](#) and [Table 2](#)). The corresponding computed structures are given in [Figure S4](#), and the comprehensive mechanism is visualized in [Figure 8](#). As for the previous calculations, we used one water molecule which is either attached to the carbonyl group or used to form the 6-membered transition state  $\text{TS2}_{S_{\text{E}}\text{Ar}}$  (6-mem). A lower and upper estimate of this barrier are given as well. Because negatively charged oxygen species are often troublesome if solvent effects are only described by a continuum approach,<sup>63</sup> we added another water molecule which for the whole reaction remained at the oxygen center of the cresol moiety. We only employed the  $\omega\text{B97xD/aug-cc-pVDZ}$  approach because the system became too large for the computationally more expensive approaches. Furthermore, the previous computations for cresol have shown that  $\omega\text{B97xD/aug-cc-pVDZ}$  gives the right trends. According to our calculations, all of the barriers nearly disappear ([Table 2](#)). The barrier for the formation of the  $\sigma$ -complex is predicted to be only about 4 kcal/mol with respect to the CT complex. Due to the increased reactivity of the cresolate the formation of the  $\sigma$ -complex is already exothermic and even the  $\text{TS2}_{S_{\text{E}}\text{Ar}}$  (6-mem) transition state is slightly more stable than the CT complex (between 3 and 6 kcal/mol). Nevertheless, it is about 12–16 kcal/mol higher in energy than the  $\sigma$ -complex. Consequently, the second step of the  $S_{\text{E}}\text{Ar}$  mechanism becomes the rate-determining step of the overall reaction. However, this step is still considerably faster than the reaction of the cresol with HTAD because this barrier is 5–9 kcal/mol lower than the lowest barrier in the reaction of the cresol (12–16 vs 21 kcal/mol ( $\omega\text{B97xD}$ ), respectively). Furthermore, even if a large amount of the released energy is dissipated after the formation of the  $\sigma$ -complex, the system will still have a considerable amount of excess energy which



**Figure 7.**  $S_{\text{E}}\text{Ar}$  reaction pathways for the reaction of *p*-cresolate (**1a**) and HTAD (**2**). The water molecules shall only indicate to which centers the hydrogen bond is formed. The computed molecular geometries can be seen in [Figure 8](#) and [Figure S4](#).

10254

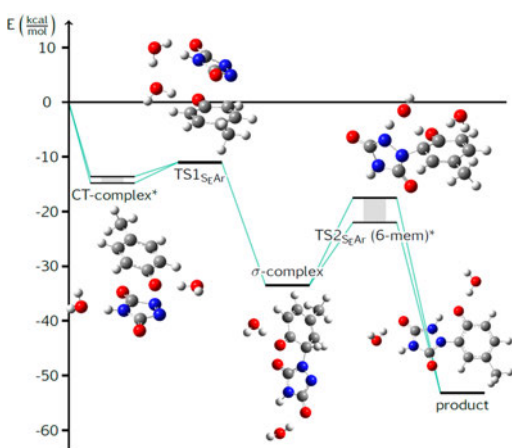
DOI: 10.1021/acs.joc.8b01445  
J. Org. Chem. 2018, 83, 10248–10260



**Table 2. Relative Energies (kcal/mol) with Respect to the Fragments (to the CT Complex) of the Stationary Points of the  $S_EAr$  Mechanism of the Reaction of TAD with *p*-cresolate Including Two Explicit Water Molecules<sup>a</sup>**

CT complex	TS1 <sub>S<sub>E</sub>Ar</sub>	$\sigma$ -complex	TS2 <sub>S<sub>E</sub>Ar</sub> (6-mem)	product
-14.8	-11.0 (3.8)	-33.5 (-18.7)	-17.5 (-2.7) <sup>b</sup>	-53.2 (-38.4)
-13.6			-22.0 (-6.4) <sup>c</sup>	

<sup>a</sup>Calculations were performed at the  $\omega$ B97xD/aug-cc-pVDZ level of theory. Additional polarization effects of the water environment were included via the PCM approach. For more information, see Table 1, Figures 7–8 and the text. <sup>b</sup>Relative energies with respect to the fragments or CT complex (in parentheses) in which the water molecule is positioned at the carbonyl group. <sup>c</sup>Relative energies with respect to fragments or CT complex (in parentheses) which are not stabilized by a water molecule.

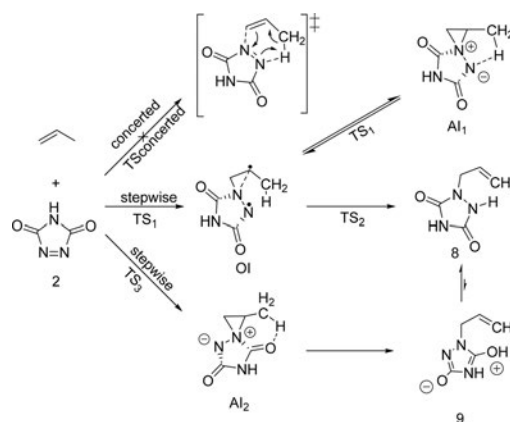


**Figure 8.** Intermediates and transition states present in the  $S_EAr$  reaction of HTAD with the *p*-cresolate anion. Energies are taken from Table 2 at the  $\omega$ B97xD level of theory. Solvent effects are accounted for by the inclusion of two water molecules as well as by using the PCM approach. \*For the CT complex and TS2<sub>S<sub>E</sub>Ar</sub> (6-mem) two energy estimates depending on the placement of the explicit water molecule have been calculated. They serve as an upper and lower error bound (gray bar for TS2<sub>S<sub>E</sub>Ar</sub> (6-mem)). For more information, see text and Table 2.

additionally accelerates the formation of the product. Taking this into account by inserting the computed barriers into the Arrhenius equation, it can be estimated that the rate constant of the cresolate reaction is about  $10^4$ – $10^7$  times larger than that of the cresol reaction. This nicely matches the available experimental results as we will discuss in the conclusion.

**HTAD Reaction with Propene and Vinyl Alcohol.** Our computations predict that for the reaction of HTAD with cresol the ene and  $S_EAr$  mechanism possess barriers of similar heights, while the barriers of the stepwise reactions are considerably higher in energy. In contrast for the reactions of PTAD with xylene, mesitylene or propene radical mechanisms were found.<sup>19</sup> To elucidate the underlying differences we compared the reactions of HTAD with cresol with its reaction with vinyl-alcohol or propene. For vinyl alcohol the same chemical motive as *p*-cresol is involved in the ene reaction but the aromatic ring is lacking. For propene the aromatic ring is missing and additionally the alcohol-moiety is replaced by an alkyl group. The propene reaction was already carefully characterized by Houk and co-workers.<sup>32,34</sup> We simply repeated these computations using CCSD(T), CASPT2 and  $\omega$ B97xD, all in combination with the aug-cc-pVDZ basis to

have consistent data for all systems. To investigate possible solvent effects, we performed optimizations in vacuo and included solvent effects through the PCM continuum model for water. Solvent-dependent changes in the ene reaction of RTAD with alkenes were investigated by various groups.<sup>64–67</sup> The mechanisms computed by Houk and co-workers for the propene reaction are sketched in Figure 9. The corresponding



**Figure 9.** Overview of the possible mechanistic pathways of the ene reaction of TAD with propene, adapted from Houk and Leach.<sup>34</sup>

computed structures of the stationary points are summarized in the SI (Figure S5). To name the various stationary points we adopted the labels coined by Houk and co-workers. For vinyl alcohol we investigated all similar possibilities by replacing the  $CH_2^-$  moiety of the methyl group by an oxygen atom. The results obtained for propene are summarized in Table 3 (in vacuo) and Table 4 (in solvent) while those computed for vinyl alcohol are collected in Table 5.

We could not identify a transition state for the concerted pathway in the case of propene and TAD. Accordingly, we assume that the barrier of the corresponding reaction path is too high. Our computations predict two possible mechanisms. The first reaction mechanism (Figure 9) proceeds to OI via TS<sub>1</sub> and then to compound 8 via TS<sub>2</sub>. Using  $\omega$ B97xD or CCSD(T) TS<sub>1</sub> and TS<sub>2</sub> are higher in energy than OI so that these approaches also predict a two-step mechanism. This is in line with the MP2 computations of Houk and co-workers.<sup>32</sup>

However, in accordance with the Lewis structure given in Figure 9, the  $\mathcal{T}_1$ -diagnostics indicate some multireference character for various stationary points. This effect is especially pronounced for the intermediate OI and the transition state TS<sub>2</sub>. Because this might lead to errors in the single-

**Table 3. Results of Quantum Chemical Calculations of the Reaction of TAD with Propene in Vacuo<sup>a</sup>**

	$\omega$ B97xD	CCSD(T)	CCSD $\mathcal{T}_1$ diagnostic	CASPT2(14,14)
CT complex	-5.3	-5.6	0.016	-6.8
TS <sub>1</sub>	10.3	9.1	0.019	9.6
OI	5.8	2.8	0.024	6.1
TS <sub>2</sub>	0.6	4.3	0.026	0.9
<b>8</b>	-43.3	-34.5	0.016	-30.5
AI <sub>1</sub>	-13	-4.4	0.017	-5.9
TS <sub>3</sub>	15.8	15	0.024	10.4
AI <sub>2</sub>	-11.3	-2.8	0.016	-4.2
<b>9</b>	-11	-3.3	0.027	-13.5

<sup>a</sup>All geometries were optimized at the  $\omega$ B97xD/aug-cc-pVDZ level of theory. For more information, see Table 1 and text. Note that the geometry "OI" is not a minimum on the  $\omega$ B97xD PES. The structure was thus obtained from an unrestricted MP2/aug-cc-pVDZ optimization.

**Table 4. Results of Quantum-Chemical Calculations of the Reaction of TAD with Propene Including Implicit Solvation at the PCM Level of Theory for the Solvent Water<sup>a</sup>**

	CCSD(T)	CASPT2(14,14)
CT complex	-4.4	-5.9
TS <sub>1</sub>	9.9	0.8
OI	3.9	-4.2
TS <sub>2</sub>	4.4	-4.9
<b>8</b>	-34.7	-49.7
AI <sub>1</sub>	-9.7	-9.4
TS <sub>3</sub>	15.2	10.8
AI <sub>2</sub>	-8.4	-12.4
<b>9</b>	-14.3	-32.5

<sup>a</sup>Structures have been obtained at the  $\omega$ B97xD/aug-cc-pVDZ level of theory in vacuo. For more information, see Table 1 and the text. We note that structure OI was not a stationary point on both the  $\omega$ B97xD and uMP2 PES when the PCM implicit solvation model is used.

configuration approaches (CCSD(T) and  $\omega$ B97xD) we also employed the multireference method CASPT2(14,14). In contrast to the single reference approaches, it predicts the OI to be higher in energy than the following transition-state geometry TS<sub>2</sub>. Hence, CASPT2(14,14) predicts that after passing TS<sub>1</sub> the reaction proceeds to compound **8** without any further barrier, i.e., a one-step mechanism with TS<sub>1</sub> as barrier maximum is computed. It is important to note that the absolute differences within the predictions are too small for a definitive answer. CCSD(T) computes the TS<sub>2</sub> to be 1.2

kcal/mol higher than the OI intermediate, while CASPT2-(14,14) calculates it to be 1.2 kcal/mol lower in energy; i.e., the differences in the predictions are only about 3.4 kcal/mol.

All approaches agree that the reaction process along TS<sub>1</sub> becomes more complicated due to the minimum AI. Our computations predict that the path from the reactants to OI but also the path to AI<sub>1</sub> proceed via TS<sub>1</sub> or at least via structurally and energetically very similar transition states. This finding originates from IRC calculations performed at the B3LYP, MP2, and  $\omega$ B97xD levels of theory in vacuo. Employing MP2 or  $\omega$ B97xD, IRC calculations starting from TS<sub>1</sub> connect the reactants with the AI<sub>1</sub> intermediate. Employing B3LYP a connection between the reactants and the OI intermediate is computed. Using  $\omega$ B97xD or unrestricted  $\omega$ B97xD, no OI minimum geometry is located at all, which would be in line with the CASPT2(14,14) computations. The unrestricted MP2 method locates the OI geometry, but only as an intermediate structure along the reaction path to AI<sub>1</sub>. These ambiguous results indicate that the PES of the reaction is very complex and probably possesses various flat minima. This flatness also explains why the different theoretical approaches come to different conclusions. Our computations find no direct transition state between AI<sub>1</sub> and **8**, which supports a previous assumption that the AI<sub>1</sub> is an impasse in the TAD reaction.<sup>34,68,69</sup>

Finally, our computations predict that the second mechanism which proceeds through TS<sub>3</sub> to the intermediate AI<sub>2</sub> and to **9** is less important. All methods agree that its barrier height is about 3.5 kcal/mol higher than the barrier height of the first mechanism running through TS<sub>1</sub>. Note that the approaches predict different barrier heights for both paths but agree in the differences between both paths.

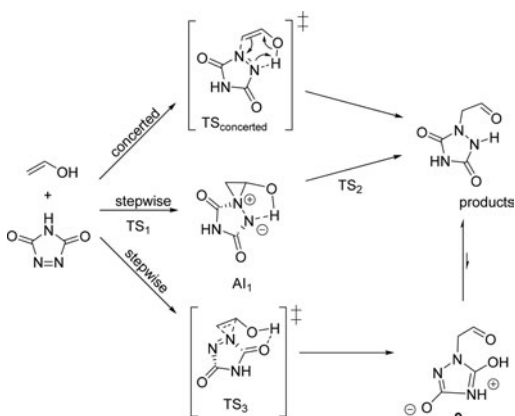
Including solvent effects, the PCM continuum approach does not change the picture. All barriers get smaller, but the relative heights remain. As for a vacuum, single reference approaches predict a two-step mechanism while CASPT2-(14,14) indicates a single-step mechanism. But again, the absolute differences between the various methods are too small for a definitive answer. For a polar environment the ene mechanism can be ruled out because no transition state is found. In summary, our computations agree with previous experimental and theoretical findings that the ene mechanism does not take place for the reaction of HTAD and propene.

To investigate the question of whether the aromatic ring or the OH group leads to the strong stabilization of the ene mechanism, we investigated the reaction between HTAD and vinyl alcohol. The corresponding mechanisms are indicated in Figure 10, while the relative energies of the stationary points

**Table 5. Results of Quantum-Chemical Calculations of the Reaction of TAD with Vinyl Alcohol<sup>a</sup>**

	$\omega$ B97xD	CCSD(T)	CCSD $\mathcal{T}_1$ diagnostic	CASPT2(8,8)	$\omega$ B97xD + PCM	CCSD(T) + PCM	CASPT2(8,8) + PCM
CT complex	-6.7	-7.3	0.017	-7.6	-5.2	-5.8	-7.1
TS <sub>ene</sub>	5.7 (12.4)	5.7 (13.0)	0.021	1.9 (9.5)	5.4 (10.6)	6.0 (11.8)	0.7 (7.8)
TS <sub>1</sub>	10.6	10.5	0.020	8.8	6.6	7.4	-1.9
AI <sub>1</sub>	-10.4	-2.1	0.017	-2.7	-12.9	-4.3	-12.4
TS <sub>2</sub>	-7.6	0.2	0.020	-4.4	-10.8	-1.9	-10.7
TS <sub>3</sub>	11.7	12.3	0.023	8.0	10.7	12.3	6.6
<b>9</b>	-25.8	-17.4	0.019	-16.6	-34.0	-25.2	-26.5
Product	-52.4	-43.9	0.017	-43.0	-52.7	-43.6	-46.5

<sup>a</sup>The PCM implicit solvation model was used to model solvation effects in water. All geometries were optimized at the  $\omega$ B97xD/aug-cc-pVDZ level of theory in vacuo or in PCM. Energies are given in kcal/mol relative to the separated reactants.



**Figure 10.** Overview of the possible mechanistic pathways of the ene reaction of HTAD with vinyl alcohol. To simplify the comparison, the stationary points are labeled according to the reaction with propene.

are summarized in Table 5. It shows that the OH group considerably stabilizes the ene reaction path. For a vacuum, this reaction path possesses a reaction barrier of about 6 kcal/mol, which is about 4 kcal/mol lower in energy than the corresponding barriers of the stepwise reaction paths. If solvent effects are taken into account, however, the barriers of the ene ( $TS_{ene}$ ) and the stepwise reaction ( $TS_1$ ) are nearly the same; i.e., the reaction would proceed along both mechanisms. Considering that for cresol the barrier of the ene reaction was considerably lower (Table 1, CCSD(T) + PCM:  $TS_{ene} = 11$  kcal/mol) than the barrier of the stepwise mechanism (Table 1, CCSD(T) + PCM:  $TS_1 = 25$  kcal/mol) it becomes obvious that the OH group and the aromaticity contribute to the stabilization of the ene reaction path for the reaction of cresol with HTAD. The stationary points of the stepwise reaction paths of vinyl alcohol and HTAD resemble the corresponding points of the HTAD reaction with propene.

## CONCLUSION

Using high-level quantum chemical approaches, we computed various possible reaction paths of the reaction of HTAD with cresol and with cresolate. Solvent effects were included to be able to interpret available experimental results which are mainly obtained in a polar protic solvent (e.g., water). Our computations come to the following conclusions: The stepwise mechanisms which take place in the reaction of PTAD and propene or aromatic hydrocarbons do not play a role of significance for the reaction of PTAD with phenol, cresol or tyrosine. For these compounds, the reaction either proceeds along the pericyclic ene or along the  $S_EAr$  mechanism. The stabilization of the transition state of the ene mechanism, which seems to be very high for the reaction of HTAD with propene, results in the replacement of the CH group by the OH group but is also influenced by the aromatic ring as shown by the comparison of the reaction of vinyl alcohol with HTAD and the reaction of cresol with HTAD.

The computations of the reaction of HTAD with cresol predict that the reaction should be rather slow in aprotic as well as in protic solvents at room temperature because of the prohibitively high activation barriers of 20–21 kcal/mol for the

ene reaction. The barriers of the other possible reaction mechanisms are even higher. The barriers of the ene reaction are slightly lower than those of the  $S_EAr$  mechanism, but the differences are too small for a definitive answer regarding which of the mechanism takes place. Central for the rationalization of the experimental results is our prediction that the barriers of the  $S_EAr$  mechanism for reaction of cresolate with HTAD are considerably lower than the corresponding barriers of the reaction with cresol. The  $\omega B97xD/aug-cc-pVDZ$  approach predicts that the barrier of the rate-determining step of the  $S_EAr$ -reaction ( $TS1_{SEAr}$ ) drops by at least about 5–9 kcal/mol (Table 3 vs Table 4). Additionally, the intermediates have a considerable amount of excess energy because the formation of the  $\sigma$ -complex is already exothermic. Taking this into account the Arrhenius equation predicts that the rate constant of the cresolate reaction is about  $10^4$ – $10^7$  higher than the corresponding cresol rate constant.

This strong difference nicely explains various experimental findings.<sup>24,25,38</sup> The fact that no conversion is found for the reaction in pure MeCN or in a MeCN/H<sub>2</sub>O (1:1) mixture can be explained by the high barriers in aprotic as well as protic solvents. For protic solvents, one has to take into account that PTAD rapidly decomposes in aqueous solution,<sup>15,16</sup> explaining the nonformation of product in water or simple aqueous media (CH<sub>3</sub>CN/H<sub>2</sub>O mixtures). Barbas and co-workers could show that efficient conversion and rapid product formation is only found within different phosphate buffers around pH 7. According to our computational results, this remarkable difference can be traced back to the relative amount of phenolate in these solutions whose reaction rate constant is considerably higher. Assuming a  $pK_a$  value of about 10 for cresol and ignoring possible increased acidity by the formation of CT complex, the ratio of cresolate to cresol is concentration dependent, and will approach  $\sqrt{10^{-pK_a}}$  ( $= 10^{-5}$ ) for concentrated solutions, although a higher degree of protolysis is possible upon dilution. Combining this lower limit for the relative concentration of cresolate anions with the above-mentioned acceleration factor for the reaction constant ( $10^4$ – $10^7$ ), the overall rate of the cresolate reaction is expected to be 0.1–100 times that for its nondeprotonated cresol counterpart. Hence, the measured conversion rates are expected to primarily result from the reaction of cresolate. This also explains the experimental finding that the reaction of PTAD with phenols can be accelerated if a base is added and that the conversion rates at pH 2 are considerably lower than those at pH 8–10.<sup>24,25</sup> It also explains the absence of product formation in neutral water. The reaction product (urazole) is itself also a weak acid ( $pK_a \sim 5$ ), which will further raise the concentration of hydroxonium upon reaction (drop in pH), pushing the protolysis equilibrium of cresol even more to the protonated form, and significantly slowing down the TAD-click reaction, favoring the background hydrolysis reaction of TAD. In a buffered solution, however, the relative concentration of cresolate is kept at a constant, concentration-independent level, as given by the buffer formula, thus fixing the ratio of cresolate to cresol at  $10^{pH-pK_a}$  (i.e., approximately 1/1000 at pH 7) throughout the course of the reaction. Combining this value with the calculated acceleration factor of  $10^4$ – $10^7$ , the reaction of cresolate is predicted to be 10–10,000 times faster than the reaction of cresol. These considerations together with our calculated barriers readily explain all experimental



observations for tyrosine click reactions with TADs. Our computations can also be used to explain the experimental findings that phenol compounds with unblocked para-positions also gives the para-substituted PTAD adduct.<sup>38</sup> With respect to the importance of the various mechanisms our computations predict that for the reaction of cresol with HTAD the pericyclic ene mechanism is within the energy range of the  $S_{\text{E}}\text{Ar}$  mechanism but that the experimentally observed adduct formations ultimately result from reactions of cresolate proceeding only along the  $S_{\text{E}}\text{Ar}$  mechanism. From our theoretical rationalization, the prediction can be made that the selectivity of the bioconjugation of different solvent-exposed tyrosine residues of a protein will not only depend on the steric hindrance but also on the corresponding  $\text{p}K_{\text{a}}$  values of the tyrosines within their ordered microenvironment.  $\text{p}K_{\text{a}}$  values can easily shift over multiple units, giving rise to orders of magnitude difference in reactivity. The presented findings may rationalize yet unexplained selectivities satisfactorily.<sup>30,70,71</sup> Experimental investigations to further substantiate these intriguing predictions are currently underway.

## ■ ASSOCIATED CONTENT

### ● Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.joc.8b01445.

All three-dimensional structures along with comprehensive information on the CASPT2, CCSD(T), and  $\omega\text{B97xD}$  calculations; detailed accounts of our investigation of the CT complex and open-shell intermediates (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: bernd.engels@uni-wuerzburg.de.

### ORCID

Johan M. Winne: 0000-0002-9015-4497

Bernd Engels: 0000-0003-3057-389X

### Notes

The authors declare no competing financial interest.

## ■ REFERENCES

- (1) Cookson, R. C.; Gilani, S. S. H.; Stevens, I. D. R. Diels-Alder Reactions of 4-Phenyl-1,2,4-triazoline-3,5-dione. *J. Chem. Soc. C* **1967**, 0 (19), 1905–1909.
- (2) Jensen, F.; Foote, C. S. Reaction of 4-Phenyl-1,2,4-triazoline-3,5-dione with Substituted Butadienes. A Nonconcerted Diels-Alder Reaction. *J. Am. Chem. Soc.* **1987**, 109 (21), 6376–6385.
- (3) Orfanopoulos, M.; Smonou, I.; Foote, C. S. Intermediates in the Ene Reactions of Singlet Oxygen and N-Phenyl-1,2,4-triazoline-3,5-dione with Olefins. *J. Am. Chem. Soc.* **1990**, 112 (9), 3607–3614.
- (4) Gillis, B. T.; Hagarty, J. D. The Reaction of 4-Phenyl-1,2,4-triazoline-3,5-dione with Conjugated Dienes. *J. Org. Chem.* **1967**, 32 (2), 330–330.
- (5) Hall, J. H.; Jones, M. L. Reactions of azodiones with electron-rich alkenes. 1, 2, 4 Triazoline-3, 5-diones and vinyl ethers. *J. Org. Chem.* **1983**, 48 (6), 822–826.
- (6) Hall, J. H. Interaction of electron-deficient 1,2,4-triazoline-3,5-diones with electron-rich polyalkoxybenzenes. *J. Org. Chem.* **1983**, 48 (10), 1708–1712.
- (7) Kinart, W. J.; Kinart, C. M.; Oszczeda, R.; Tran, Q. T. Studies on the Catalysis by Lithium Perchlorate of Reactions of Aromatic Amines with Diethyl Azodicarboxylate and Naphtalen-2-ol with 4-phenyl-1,2,4-triazoline-3,5-dione. *Catal. Lett.* **2005**, 103 (3), 185–189.
- (8) Zhang, J. W.; Xu, J. H.; Cheng, D. J.; Shi, C.; Liu, X. Y.; Tan, B. Discovery and enantiocontrol of axially chiral urazoles via organocatalytic tyrosine click reaction. *Nat. Commun.* **2016**, 7, 10677.
- (9) Changotra, A.; Das, S.; Sunoj, R. B. Reversing Enantioselectivity Using Noncovalent Interactions in Asymmetric Dearomatization of beta-Naphthols: The Power of 3,3' Substituents in Chiral Phosphoric Acid Catalysts. *Org. Lett.* **2017**, 19 (9), 2354–2357.
- (10) Algar, W. R.; Prasuhn, D. E.; Stewart, M. H.; Jennings, T. L.; Blanco-Canosa, J. B.; Dawson, P. E.; Medintz, I. L. The Controlled Display of Biomolecules on Nanoparticles: A Challenge Suited to Bioorthogonal Chemistry. *Bioconjugate Chem.* **2011**, 22 (5), 825–858.
- (11) Boutureira, O.; Bernardes, G. J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **2015**, 115 (5), 2174–2195.
- (12) Fujita, M.; Matsushima, H.; Sugimura, T.; Tai, A.; Okuyama, T. Asymmetric addition of an electrophile to naphthalenes promoted and stereodirected by alcohol. *J. Am. Chem. Soc.* **2001**, 123 (13), 2946–2957.
- (13) Reaney, M. J. T.; Liu, Y. D.; Taylor, W. G. Gas chromatographic analysis of Diels-Alder adducts of geometrical and positional isomers of conjugated linoleic acid. *J. Am. Oil Chem. Soc.* **2001**, 78 (11), 1083–1086.
- (14) Naik, A.; Alzeer, J.; Triemer, T.; Bujalska, A.; Luedtke, N. W. Chemoselective Modification of Vinyl DNA by Triazolinediones. *Angew. Chem., Int. Ed.* **2017**, 56 (36), 10850–10853.
- (15) Roy, N.; Lehn, J. M. Dynamic Covalent Chemistry: A Facile Room-Temperature, Reversible, Diels-Alder Reaction between Anthracene Derivatives and N-Phenyltriazolinedione. *Chem. - Asian J.* **2011**, 6 (9), 2419–2425.
- (16) Sato, S.; Nakamura, K.; Nakamura, H. Tyrosine-Specific Chemical Modification with in Situ Hemin-Activated Luminol Derivatives. *ACS Chem. Biol.* **2015**, 10 (11), 2633–2640.
- (17) Sasaki, T.; Kanematsu, K.; Hayakawa, K. Studies on Heteroaromaticity. Part XLVIII. Cycloaddition Reactions of 4-Phenyl-1,2,4-triazoline-3,5-dione with Seven-membered Ring Unsaturated Compounds and Photochemical Behaviour of the Adducts. *J. Chem. Soc. C* **1971**, No. 11, 2142–2147.
- (18) Breton, G. W.; Newton, K. A. Further studies of the thermal and photochemical Diels-Alder reactions of N-methyl-1,2,4-triazoline-3,5-dione (MeTAD) with naphthalene and some substituted naphthalenes. *J. Org. Chem.* **2000**, 65 (10), 2863–2869.
- (19) Breton, G. W.; Hoke, K. R. Application of Radical Cation Spin Density Maps toward the Prediction of Photochemical Reactivity between N-Methyl-1,2,4-triazoline-3,5-dione and Substituted Benzenes. *J. Org. Chem.* **2013**, 78 (10), 4697–4707.
- (20) Ulmer, L.; Siedschlag, C.; Mattay, J. Functionalization of 60 fullerene and of 60 fullerene monoadducts by photochemical cycloaddition of 4-methyl-1,2,4-triazoline-3,5-dione. *Eur. J. Org. Chem.* **2003**, 2003 (19), 3811–3817.
- (21) Risi, F.; Alstanei, A. M.; Volanschi, E.; Carles, M.; Pizzala, L.; Aycard, J. P. Photo addition of aliphatic ethers to 4-methyl-1,2,3-triazoline-3,5-dione: Application to the synthesis of functionalized crown ethers and mechanism. *Eur. J. Org. Chem.* **2000**, 2000, 617–626.
- (22) Eyles, D.; Anderson, C.; Ko, P.; Jones, A.; Thomas, A.; Bume, T.; Mortensen, P. B.; Norgaard-Pedersen, B.; Hougaard, D. M.; McGrath, J. A sensitive LC/MS/MS assay of 250H vitamin D-3 and 250H vitamin D-2 in dried blood spots. *Clin. Chim. Acta* **2009**, 403 (1–2), 145–151.
- (23) Christie, W. W. Gas chromatography mass spectrometry methods for structural analysis of fatty acids. *Lipids* **1998**, 33 (4), 343–353.
- (24) Ban, H.; Gavriljuk, J.; Barbas, C. F. Tyrosine Bioconjugation through Aqueous Ene-Type Reactions: A Click-Like Reaction for Tyrosine. *J. Am. Chem. Soc.* **2010**, 132 (5), 1523–1525.
- (25) Ban, H.; Nagano, M.; Gavriljuk, J.; Hakamata, W.; Inokuma, T.; Barbas, C. F. Facile and Stable Linkages through Tyrosine: Bioconjugation Strategies with the Tyrosine-Click Reaction. *Bioconjugate Chem.* **2013**, 24 (4), 520–532.

10258

DOI: 10.1021/acs.joc.8b01445  
*J. Org. Chem.* **2018**, 83, 10248–10260

- (26) Bauer, D. M.; Ahmed, I.; Vigovskaya, A.; Fruk, L. Clickable Tyrosine Binding Bifunctional Linkers for Preparation of DNA Protein Conjugates. *Bioconjugate Chem.* **2013**, *24* (6), 1094–1101.
- (27) McKay, C. S.; Finn, M. G. Click Chemistry in Complex Mixtures: Biorthogonal Bioconjugation. *Chem. Biol.* **2014**, *21* (9), 1075–1101.
- (28) Billiet, S.; De Bruycker, K.; Driessen, F.; Goossens, H.; Van Speybroeck, V.; Winne, J. M.; Du Prez, F. E. Triazolinediones enable ultrafast and reversible click chemistry for the design of dynamic polymer systems. *Nat. Chem.* **2014**, *6* (9), 815–821.
- (29) De Bruycker, K.; Billiet, S.; Houck, H. A.; Chattopadhyay, S.; Winne, J. M.; Du Prez, F. E. Triazolinediones as Highly Enabling Synthetic Tools. *Chem. Rev.* **2016**, *116* (6), 3919–3974.
- (30) Ortiz-Soto, M. E.; Ertl, J.; Mut, J.; Adelman, J.; Le, T. A.; Shan, J.; Tebmar, J.; Schlosser, A.; Engels, B.; Seibel, J. Product-oriented chemical surface modification of a levansucrase (SacB) via an ene-type reaction. *Chemical Science* **2018**, *9* (24), 5312–5321.
- (31) Roy, N.; Lehn, J.-M. Dynamic Covalent Chemistry: A Facile Room-Temperature, Reversible, Diels Alder Reaction between Anthracene Derivatives and N-Phenyltriazolinedione. *Chem. - Asian J.* **2011**, *6* (9), 2419–2425.
- (32) Chen, J. S.; Houk, K. N.; Foote, C. S. The nature of the transition structures of triazolinedione ene reactions. *J. Am. Chem. Soc.* **1997**, *119* (41), 9852–9855.
- (33) Lu, X. Can the nitroso Ene reaction proceed concertedly? *Org. Lett.* **2004**, *6* (16), 2813–2815.
- (34) Leach, A. G.; Houk, K. N. Diels-Alder and ene reactions of singlet oxygen, nitroso compounds and triazolinediones: transition states and mechanisms from contemporary theory. *Chem. Commun.* **2002**, No. 12, 1243–1255.
- (35) Rigby, J. H. [6 + 4] Cycloaddition Reactions. *Organic Reactions* **1996**, *49*, 331.
- (36) Rigby, J. H.; Chouraqui, G. First intramolecular 6 + 4 cycloaddition of tropone and a furan moiety: Rapid entry into a highly functionalized ingenane skeleton. *Synlett* **2005**, No. 16, 2501–2503.
- (37) Kim, D. K.; O'Shea, K. E. The reaction of N-methyl-1,2,4-triazoline-3,5-dione with tetracyclopropylethylene. Formation of an unusual meso-ionic product and its rearrangement to the diazetidone. *J. Am. Chem. Soc.* **2004**, *126* (3), 700–701.
- (38) Ban, H.; Gavriljuk, J.; Barbas, C. F. I. Tyrosine bioconjugation through aqueous ene-like reactions. *J. Am. Chem. Soc.* **2011**, *132*, 1523.
- (39) Barbas, C. F. I.; Ban, H.; Gavriljuk, J.; Tyrosine bioconjugation through aqueous ene-like reactions. Institute, T. S. R., Ed.; US Patent WO/2011/079315, 2011.
- (40) Peric, M.; Engels, B.; Peyerimhoff, S. D. Ab initio investigation of the vibronic structure of the C2H spectrum: Computation of the vibrationally averaged values for the hyperfine coupling constants. *J. Mol. Spectrosc.* **1991**, *150* (1), 70–85.
- (41) Breton, G. W.; Surovic, A. H. Intermediacy of a Persistent Urazole Radical and an Electrophilic Diazenium Species in the Acid-Catalyzed Reaction of MeTAD with Anisole. *J. Org. Chem.* **2016**, *81* (1), 206–214.
- (42) 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: Condens. Matter Mater. Phys.* **1988**, *37* (2), 785–789.
- (43) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A: At, Mol, Opt. Phys.* **1988**, *38* (6), 3098–3100.
- (44) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.
- (45) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96* (9), 6796–6806.
- (46) Chai, J. D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10* (44), 6615–6620.
- (47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian09*, revision E.01; Gaussian: Wallingford, CT, 2009.
- (48) Cheron, N.; Jacquemin, D.; Fleurat-Lessard, P. A qualitative failure of B3LYP for textbook organic reactions. *Phys. Chem. Chem. Phys.* **2012**, *14* (19), 7170–7175.
- (49) Lindh, R.; Malmqvist, P. Å.; Veryazov, V.; Widmark, P.-O. et al. *Multiconfigurational Quantum Chemistry*; John Wiley & Sons, 2016.
- (50) Lee, T. J.; Taylor, P. R. A diagnostic for determining the quality of single-reference electron correlation methods. *Int. J. Quantum Chem.* **1989**, *36* (S23), 199–207.
- (51) Pless, V.; Suter, H. U.; Engels, B. Ab initio study of the energy difference between the cyclic and linear forms of the C6 molecule. *J. Chem. Phys.* **1994**, *101* (5), 4042–4048.
- (52) Engels, B.; Peyerimhoff, S. D. The hyperfine coupling constants of the X 3-sigma-states of NH Influence of polarization functions and configuration space on the description of spin polarization. *Mol. Phys.* **1989**, *67* (3), 583–600.
- (53) Andersson, K.; Malmqvist, P. Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* **1990**, *94* (14), 5483–5488.
- (54) Andersson, K.; Malmqvist, P. Å.; Roos, B. O. Second-order perturbation theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* **1992**, *96* (2), 1218–1226.
- (55) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.-Å.; Neogrády, P.; Pedersen, T. B.; Pitoňák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Veryazov, V.; Lindh, R. MOLCAS 7: The Next Generation. *J. Comput. Chem.* **2010**, *31* (1), 224–247.
- (56) Rintelman, J. M.; Adamovic, I.; Varganov, S.; Gordon, M. S., Multireference second-order perturbation theory: How size consistent is “almost size consistent”. *J. Chem. Phys.* **2005**, *122* (4), 044105.
- (57) Kuhrau, M.; Stadler, R. Low temperature modification of polymers by triazolinediones, 2: On the mechanism of the electrophilic aromatic substitution with 4-substituted 1,2,4-triazoline-3,5-dione. *Makromol. Chem.* **1990**, *191* (8), 1787–1798.
- (58) Settels, V.; Schubert, A.; Tafipolski, M.; Liu, W.; Stehr, V.; Topczak, A. K.; Pflaum, J.; Deibel, C.; Engel, V.; Fink, R. F.; Engels, B. Identification of Ultrafast Relaxation Processes As a Major Reason for Inefficient Exciton Diffusion in Perylene-Based Organic Semiconductors. *J. Am. Chem. Soc.* **2014**, *136*, 9327–9337.
- (59) Walter, C.; Kramer, V.; Engels, B. On the applicability of time-dependent density functional theory (TDDFT) and semiempirical methods to the computation of excited-state potential energy surfaces of perylene-based dye-aggregates. *Int. J. Quantum Chem.* **2017**, *117* (6), e25337.
- (60) Bruckner, C.; Engels, B. A Theoretical Description of Charge Reorganization Energies in Molecular Organic P-Type Semiconductors. *J. Comput. Chem.* **2016**, *37* (15), 1335–1344.
- (61) Liao, M. S.; Lu, Y.; Parker, V. D.; Scheiner, S. DFT calculations and spectral measurements of charge-transfer complexes formed by aromatic amines and nitrogen heterocycles with tetracyanoethylene and chloranil. *J. Phys. Chem. A* **2003**, *107* (42), 8939–8948.
- (62) Bruckner, R. *Reaktionsmechanismen: Organische Reaktionen, Stereochemie, Moderne Synthesemethoden*; Spektrum Akademischer Verlag, 2009.

- (63) Helten, H.; Schirmeister, T.; Engels, B. Model calculations about the influence of protic environments on the alkylation step of epoxide, aziridine, and thiirane based cysteine protease inhibitors. *J. Phys. Chem. A* **2004**, *108* (38), 7691–7701.
- (64) Roubelakis, M. M.; Vougioukalakis, G. C.; Angelis, Y. S.; Orfanopoulos, M. Solvent-Dependent Changes in the Ene Reaction of RTAD with Alkenes: The Cyclopropyl Group as a Mechanistic Probe. *Org. Lett.* **2006**, *8* (1), 39–42.
- (65) Vougioukalakis, G. C.; Roubelakis, M. M.; Alberti, M. N.; Orfanopoulos, M. Solvent-Dependent Changes in the Triazolinedione Alkene Ene Reaction Mechanism. *Chem. - Eur. J.* **2008**, *14* (31), 9697–9705.
- (66) Acevedo, O.; Jorgensen, W. L. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc. Chem. Res.* **2010**, *43* (1), 142–151.
- (67) Acevedo, O.; Squillacote, M. E. A New Solvent-Dependent Mechanism for a Triazolinedione Ene Reaction. *J. Org. Chem.* **2008**, *73* (3), 912–922.
- (68) Squillacote, M. E.; Garner, C.; Oliver, L.; Mooney, M.; Lai, Y.-L. Photochemical Generation of Aziridinium Imines: Are Als Bystanders or Perpetrators in the Ene Reaction? *Org. Lett.* **2007**, *9* (26), 5405–5408.
- (69) Vougioukalakis, G. C.; Orfanopoulos, M. Mechanistic Studies in Triazolinedione Ene Reactions. *Synlett* **2005**, *2005* (05), 0713–0731.
- (70) Hu, Q. Y.; Allan, M.; Adamo, R.; Quinn, D.; Zhai, H. L.; Wu, G. X.; Clark, K.; Zhou, J.; Ortiz, S.; Wang, B.; Danieli, E.; Crotti, S.; Tontini, M.; Brogioni, G.; Berti, F. Synthesis of a well-defined glycoconjugate vaccine by a tyrosine-selective conjugation strategy. *Chemical Science* **2013**, *4* (10), 3827–3832.
- (71) Hu, Q. Y.; Berti, F.; Adamo, R. Towards the next generation of biomedicines by site-selective conjugation. *Chem. Soc. Rev.* **2016**, *45* (6), 1691–1719.

## 8. Entropy Estimation

In this section various entropy estimation methodologies are evaluated and their theoretical foundations are presented. A special focus of this work is on the investigation of entropy estimation approaches applicable to larger, macroscopic chemical systems, such as the  $k^{\text{th}}$  nearest neighbor (kNN) entropy estimator, which is the primary target of the present investigation. To this end, various concepts from classical and statistical thermodynamics as well as Normal Mode Analysis are used.

### 8.1. Theoretical Foundations

#### 8.1.1. Entropy Derived from Statistical Assumptions

The term entropy refers to a measurable state function originally defined within the framework of classical thermodynamics. Rudolph Clausius first used the term (see for example [174]) to describe a function  $S$  whose differential satisfies

$$dS = \frac{\delta q_{\text{rev}}}{T}. \quad (15)$$

As entropy is a state function, it holds that for a reversible cyclic process

$$\oint \frac{\delta q_{\text{rev}}}{T} = 0. \quad (16)$$

For non-cyclic reversible processes, the entropy is evaluated as a path-integral of Equation 15. According to the second law of thermodynamics, the entropy change due to an irreversible process performed on a isolated system is always positive, i.e. the entropy of the system increases.

We now consider chemical systems which can be said to be in statistical equilibrium (or equivalently thermodynamic equilibrium). According to Lanau & Lifshitz (see [175]) this is the case when the system is made up of a large number of macroscopic subsystems whose statistical distribution does not depend on their initial value. Furthermore, the physical quantities associated with the subsystems need to be equal to the mean value across all subsystems “to a high degree of accuracy”. This is equivalent to stating that, “if a closed macroscopic system is observed for a sufficiently long period of time, it will be in a state of statistical equilibrium for much the greater part of this period. If, at any initial instant, a closed macroscopic system was not in a state



of statistical equilibrium [...], it will necessarily enter an equilibrium state".<sup>[175]</sup>

To begin a mathematical derivation of statistical entropy, we consider the time-dependent density matrix  $\mathbf{W}$  of an arbitrary subsystem in the energy representation with entries  $w_{nm}(t)$ . In statistical equilibrium, this matrix becomes diagonal, as the interactions between quasi-closed subsystems tend to zero when the system becomes more macroscopic and statistical. One intuitive example of this emerging macroscopicity is obtained when the number of particles of a system is increased, but the number of subsystems is kept constant (their size may be increased as well). If  $\mathbf{W}$  can be assumed to be diagonal, the entries  $w_n$  can be rationalized as probabilities of being in a certain microstate. These probabilities are a function of the energy alone:

$$w_n = w_n(E) \quad (17)$$

We define  $\Gamma(E)$  as the number of quantum states with an associated energy eigenvalue equal or smaller than  $E$ . The so-called statistical weight  $\Delta\Gamma$  is then given as

$$\Delta\Gamma = \frac{1}{w(\bar{E})} = \frac{d\Gamma(\bar{E})}{dE} \Delta E, \quad (18)$$

where  $\bar{E}$  is the mean energy value of the subsystems. The entropy  $S$  in the spirit of Boltzmann may then be defined as the logarithm of the statistical weight:

$$S = \ln \Delta\Gamma. \quad (19)$$

If  $\Delta E$  is assumed to be "in order of magnitude of the mean fluctuation of energy of the [entire] system", the statistical weight is "the size of the region of phase space in which [a] subsystem will almost always be found".<sup>[175]</sup>

Since subsystems of a macroscopic system are assumed to be independent and thus uncorrelated, it must follow that the probability density of the macroscopic system can be strictly expressed as a product of the individual subsystem's densities:

$$\rho_{\text{macroscopic}} = \prod \rho_{\text{subsystem}} \quad (20)$$

It follows from Liouville's theorem that the probabilities  $\rho_{\text{subsystem}}$  are solely a function of constants of motion. Taking the logarithm of Equation 20 implies that the logarithm of  $\rho_{\text{macroscopic}}$  can be expressed as a linear combination of additive constants

of motion. The seven known additive constants of motion (or equivalently, the additive conservation laws originating from Noether's theorem) are energy, components of the momentum vector and the three components of the angular momentum vector. As this derivation is only strictly valid within non-relativistic mechanics, we neglect further additive conservation laws resulting from Galilean relativity. For more information on this topic, see [176]. Assuming translation and rotation to be negligible as they depend on the chosen reference frame, we find that  $\ln \rho_{\text{subsystem}}$  in statistical equilibrium can be written as

$$\ln \rho_{\text{subsystem}} = \alpha_{\text{subsystem}} + \beta E_{\text{subsystem}}, \quad (21)$$

where  $\alpha_{\text{subsystem}}$  and  $\beta$  are constants. Using the same line of argumentation, the quantum probability  $w_n$  of a subsystem being in a quantum state  $n$  can be written as

$$\ln w_n = \alpha_n + \beta E_n. \quad (22)$$

We now write the mean value of  $\ln w_n$  using  $\bar{E}$  as

$$\langle \ln w_n \rangle = \ln w_n(\bar{E}) = \alpha + \beta \bar{E}. \quad (23)$$

Using Equation 18, the entropy  $S$  may be expressed as a mean-value:

$$S = -\langle \ln w(E_n) \rangle \quad (24)$$

Expressing the mean as a sum over all quantum states, we get

$$S_{\text{QM}} = -k_{\text{B}} \sum_n w_n \ln w_n, \quad (25)$$

where by convention we inserted the Boltzmann constant  $k_{\text{B}}$  to render the unitless expression defined for probabilities into the physical units required by Equation 15. As we will move forward considering various numerical and conceptual approximations for the entropy of molecular systems, we highlight that Equation 25 contains no approximation and is strictly valid for all macroscopic non-relativistic physical systems in statistical equilibrium. This definition is similar to the one given by Shannon in the realm of information theory.<sup>[177]</sup> Jaynes has shown this definition to be reconcilable with the original definition given by Clausius (Equation 15).<sup>[178]</sup>

In classical statistical thermodynamics, the volume of continuous phase space corresponding to one quantum state can be derived to be exactly  $(2\pi\hbar)^d$  with  $d$  being the

number of degrees of freedom (DOFs).<sup>[179]</sup> It follows that in classical thermodynamics

$$S_{\text{Classical}} = \ln \frac{\Delta p \Delta q}{(2\pi\hbar)^d} \quad (26)$$

holds, using

$$\Delta\Gamma_{\text{Classical}} = \frac{\Delta p \Delta q}{(2\pi\hbar)^s}, \quad (27)$$

with  $\Delta p \Delta q$  being the phase space region corresponding to a certain subsystem.<sup>[175]</sup> Throughout this text,  $p$  will be used for momentum contributions and  $q$  for spatial coordinates.

### 8.1.2. Calculating the Entropy from the Partition Function

We now focus on the nature of the probability function  $w_n$  of a quantum system. The function  $w_n$  which maximizes the entropy according to Equation 25, fulfilling the second law of thermodynamics, is sought. Jaynes gives a derivation of  $w_n$  using the method of Lagrangian multipliers in [180].  $\lambda$  and  $\mu$  are the Lagrangian multipliers and the Lagrangian function is given as

$$\mathcal{L} = \left( -k_B \sum_n w_n \ln w_n \right) + \lambda \left( 1 - \sum_n w_n \right) - \mu \left( U - \sum_n w_n E_n \right). \quad (28)$$

One obtains

$$\begin{aligned} 0 &= -k_B \ln w_n + \lambda - \mu E_n - k_B \\ w_n &= e^{-1 + \frac{\lambda - \mu E_n}{k_B}}, \end{aligned} \quad (29)$$

where  $E_n$  is the internal energy of the subsystem  $n$ .<sup>[180]</sup> For the derivation, a canonical ensemble (macroscopic system in contact with a heat bath) is assumed, implying conservation of the average total energy as

$$\langle E \rangle = \sum_n w_n E_n \equiv U. \quad (30)$$

We note that the entropy expressions given in Equation 26 and Equation 25 only hold for a microcanonical ensemble, i.e. a closed system where conservation of the total

energy applies. Changing the notation by introducing the partition function  $Z$  as

$$Z(\mu) = \sum_n e^{-\frac{\mu}{k_B} E_n} \quad (31)$$

yields

$$\begin{aligned} \lambda &= k_B (-\ln Z(\mu) + 1) \\ \langle E \rangle &= -\frac{\partial}{\partial \mu} \ln Z(\mu), \end{aligned} \quad (32)$$

where

$$\sum_n w_n = 1 \quad (33)$$

has been used.  $w_n$  can now be expressed using  $Z$ :

$$w_n = \frac{1}{Z} e^{-\frac{\mu E_n}{k_B}} \quad (34)$$

The entropy  $S$  (Equation 25) can then be written using the partition function:

$$\begin{aligned} S &= -k_B \sum_n w_n \ln w_n \\ &= -k_B \sum_n w_n \left( \frac{-\mu}{k_B} E_n - \ln(Z) \right) \\ &= \mu U + k_B \ln(Z) \end{aligned} \quad (35)$$

The first law of thermodynamics  $dU = TdS - PdV$  (the total energy of an isolated system is conserved; energy can only be supplied as heat or work) is used to obtain an explicit expression for the Lagrangian multiplier  $\mu$  by differentiating Equation 35 with respect to  $U$ :

$$\frac{dS}{dU} = \frac{d(\mu U + k_B \ln(Z))}{dU} = \mu = \frac{1}{T} \quad (36)$$

Equation 31 then becomes

$$Z(\mu) = \sum_n e^{-\frac{E_n}{k_B T}}. \quad (37)$$

In the literature, the fraction  $\frac{1}{k_B T}$  is often termed  $\beta = \frac{1}{k_B T}$  as a shorthand. We may now write  $w_n$  as

$$w_n = \frac{1}{Z} e^{-\frac{E_n}{k_B T}}. \quad (38)$$

The total energy  $U$  may be calculated from the partition function  $Z$  as well:

$$U = \langle E \rangle = \sum_n E_n w_n = \frac{1}{Z} \sum_n E_n e^{-\beta E_n} = -\frac{1}{Z} \frac{\partial}{\partial \beta} Z. \quad (39)$$

Finally, we arrive at

$$S = \frac{U}{T} + k_B \ln Z = \frac{\partial}{\partial T} (k_B T \ln Z) \quad (40)$$

Knowledge of the internal energy  $U$  and the value of the partition function  $Z$  (which yields a dimensionless number) thus permits the exact calculation of the entropy  $S$  of a physical system. The important feature of Equation 40 in comparison to the Shannon-type entropy expression (Equation 25), is that the probabilities  $w_n$  do not need to be known explicitly.

## 8.2. Entropies of Chemical Systems

### 8.2.1. Additive Decomposition of Entropic Contributions

We now introduce some common nomenclature for different entropic contributions which are derived from an additive decomposition of the total entropy into separable components. Additive separability only holds strictly for the translational contribution. If the Born-Oppenheimer approximation is assumed, electronic contributions might be separated as well. Nevertheless, coupling terms between rotational motion and the other contributions are usually assumed to be small in magnitude and thus negligible. The rotational entropy contribution may then be factorized as well. What is left is the so-called configurational entropy, which encompasses vibrational and conformational motion. The total entropy of a molecular system  $S$  is thus commonly decomposed as

$$S = S_{\text{translation}} + S_{\text{rotation}} + S_{\text{vibration}} + S_{\text{electronic}} + S_{\text{nuclear-spin}} + S_{\text{couplings}}, \quad (41)$$

with

$$S_{\text{couplings}} \approx 0. \quad (42)$$

The nuclear-spin contribution  $S_{\text{nuclear-spin}}$  is in general non-zero, but never changes in chemical reactions and is thus irrelevant for chemical purposes.<sup>[181]</sup> The additive decomposition of entropic contributions directly implies that the partition function of

the molecular system can be expressed as a product of the individual contributions:

$$Z = Z_{\text{translation}} \cdot Z_{\text{rotation}} \cdot Z_{\text{vibration}} \cdot Z_{\text{electronic}} \quad (43)$$

The electronic partition function is

$$Z_{\text{electronic}} = \sum_{n=1}^{\text{states}} g_n e^{-\beta E_n}, \quad (44)$$

with the degeneracy of the energy level  $E_n$  given as  $g_n$ . If electronic excited states are energetically well separated and the Born-Oppenheimer approximation is valid, we can approximate

$$Z_{\text{electronic}} \simeq g_{n=1}. \quad (45)$$

The electronic entropy contribution can then be approximated by means of the ground state degeneracy. If  $g_{n=1} = 1$ , then  $S_{\text{electronic}} = 0$ . If the molecular system can be assumed to be found in the electronic ground state at all times, the electronic contribution to the total entropy vanishes.

In order to derive analytical expressions for the contributions  $S_{\text{translation}}$ ,  $S_{\text{rotation}}$ , and  $S_{\text{vibration}}$ , the partition functions for a free particle (translation), a non-linear rigid rotor (rotation), and a harmonic oscillator (vibration) are used in Equation 40. For a full derivation, see [132] or [182]. The following expressions are obtained:

$$S_{\text{translation}} = R \left\{ \frac{5}{2} + \ln \left[ \left( \frac{2\pi M k_B T}{N_A h^2} \right)^{\frac{3}{2}} \frac{RT}{N_A p} \right] \right\} \quad (46)$$

$$S_{\text{rotation}} = N_A k_B \left[ \ln \left( \frac{\sqrt{\pi I_A I_B I_C}}{\sigma_{\text{rot}}} \left( \frac{8\pi^2 k_B T e}{h^2} \right)^{3/2} \right) \right] \quad (47)$$

$$S_{\text{vibration}} = N_A k_B \sum_{i=1}^{3N_{\text{at}}-6} \left( \frac{\frac{\hbar \omega_i}{k_B T}}{\left( e^{\frac{\hbar \omega_i}{k_B T}} - 1 \right)} - \ln \left( 1 - e^{-\frac{\hbar \omega_i}{k_B T}} \right) \right) \quad (48)$$

At a fixed temperature, pressure  $p$  and number of particles, the translational contribution can be determined from the molar mass  $M$  of the molecule alone. The rotational contribution  $S_{\text{rotation}}$  is a function of the temperature  $T$ , the moments of inertia  $I$  and the rotational symmetry number  $\sigma_{\text{rot}}$  rot. For the vibrational term  $S_{\text{vib}}$ , we have



implicitly assumed in Equation 48 that the vibrational models are harmonic. Furthermore, the rotational contribution given in 47 assumes the molecules's rotation to be describable as a rigid rotor. Both mentioned facts need not to be the case and constitute an approximation. Calculating the total entropy of a molecule using Equations 46, 47, and 48 is called the Rigid-Rotor Harmonic-Oscillator (RRHO) approximation:

$$S_{\text{RRHO}} = S_{\text{translation}} + S_{\text{rotation}} + S_{\text{vibration}} \quad (49)$$

For the evaluation of  $S_{\text{vibration}}$  the frequencies  $\omega$  of all normal modes are needed. If the second derivatives with respect to the nuclear coordinates can be calculated, the normal mode frequencies  $\omega$  can be calculated from the Hessian matrix of second derivatives.<sup>[13,181]</sup>

The RRHO approximation abides the third law of thermodynamics only in the sense that - although

$$\lim_{T \rightarrow 0K} S_{\text{rotation}} = -\infty \quad (50)$$

and

$$\lim_{T \rightarrow 0K} S_{\text{translation}} = -\infty, \quad (51)$$

it holds that

$$\lim_{T \rightarrow 0K} S_{\text{vibration}} = \lim_{\omega \rightarrow \infty} S_{\text{vibration}} = -0. \quad (52)$$

This is the case because in the derivation of Equations 46 and 47 the partition functions were devised under the assumption of classical, non-quantized phase space.

If the molecule under consideration possesses more than one relevant accessible conformations  $\vec{\alpha} = \{\alpha_i\}$ , the RRHO approximation needs to be applied to the equilibrium geometry of each minimum separately. Then, the mixture of conformers approach (see [13]) can be invoked to calculate the total entropy of the molecule by proper weighting the individual contributions and adding a combinatorial term:

$$S_{\text{mixture}} = \bar{S}_{\text{RRHO}} + \Delta S_{\text{mix}} \quad (53)$$

$$\bar{S}_{\text{RRHO}} = \sum_{i=1} \rho_{\alpha_i} S_{\text{RRHO}}(\alpha_i) = \sum_{i=1} \frac{e^{-\frac{E_{\alpha_i}}{k_B T}}}{\sum_{i=1} e^{-\frac{E_{\alpha_i}}{k_B T}}} S_{\text{RRHO}}(\alpha_i) \quad (54)$$

$$\Delta S_{\text{mix}} = k_B \sum_{i=1} -\rho_{\alpha_i} \ln \rho_{\alpha_i} = k_B \sum_{i=1} - \left( \frac{e^{-\frac{E_{\alpha_i}}{k_B T}}}{\sum_{i=1} e^{-\frac{E_{\alpha_i}}{k_B T}}} \right) \ln \left( \frac{e^{-\frac{E_{\alpha_i}}{k_B T}}}{\sum_{i=1} e^{-\frac{E_{\alpha_i}}{k_B T}}} \right) \quad (55)$$

The assumption of all conformations being harmonic wells is implicit in the mixture of conformers approach. In order to compute the frequencies  $\omega_i$  of a molecule, which are necessary to evaluate the RRHO entropy due to Equation 48, the molecular potential  $V(\vec{q})$  is first written as a Taylor expansion truncated at 2<sup>nd</sup> order (harmonic approximation):

$$V(\vec{q}) = V_0 + \sum_{i=1}^{3N_{\text{atoms}}} \left( \frac{\partial V}{\partial q_i} \right)_0 q_i + \frac{1}{2} \sum_{i=1}^{3N_{\text{atoms}}} \left( \frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0 q_i q_j + \mathcal{O}(q^3) \quad (56)$$

We note that at an equilibrium geometry it holds that  $\nabla V = 0$  by definition, and therefore

$$\sum_{i=1}^{3N_{\text{atoms}}} \left( \frac{\partial V}{\partial q_i} \right)_0 q_i = 0. \quad (57)$$

The Hessian matrix  $\mathbf{H}$  is defined as

$$\mathbf{H}_{i,j} = \frac{\partial^2 V}{\partial q_i \partial q_j}. \quad (58)$$

We write the mass-weighted Hessian matrix  $\mathbf{H}_{i,j}^\dagger$  as

$$\mathbf{H}_{i,j}^\dagger = \frac{\mathbf{H}_{i,j}}{\sqrt{m_i} \sqrt{m_j}}, \quad (59)$$

so that Newton's equations of motion become

$$\frac{\partial^2 q_i}{\partial t^2} = - \sum_j \mathbf{H}_{i,j}^\dagger x_j = \omega^2 x_i. \quad (60)$$

Equation 60 constitutes a matrix eigenvalue equation and thus, the square-roots of the eigenvalues of the mass-weighted Hessian matrix  $\mathbf{H}_{i,j}^\dagger$  yields the angular frequencies of the normal modes of a molecule within the harmonic approximation.

### 8.3. Quasi-Harmonic Entropy Estimation Approaches based on Frequency Assignment

For larger chemical systems, both the calculations of the Hessian matrix of nuclear second derivatives (see Equation 60), and the identification of all possible accessible

conformations (see Equation 55) become computationally unfeasible when sufficiently sophisticated computational approaches are used. This is especially the case when the electronic structure of the molecule is taken into consideration using quantum chemical methods. For this reason, other means of calculating the absolute entropy numerically for molecular systems are sought.

Consideration of Equation 48 makes it evident that at a given constant temperature the entropy of harmonic oscillators can be solely determined from their frequency. A family of methods termed quasi-harmonic makes use of this fact in order to estimate absolute entropies. The key idea is to assign a frequency to a molecular vibrational mode not from the diagonalization of the mass-weighted Hessian matrix, but from a Molecular Dynamics (MD) simulation trajectory. The term quasi-harmonic may also encompass the multivariate gaussian-based method proposed by Karplus & Kushick in 1981, which however does not use an actual frequency assignment procedure but is rooted in the spatial domain.<sup>[183]</sup> For this reason, it is not applicable to oscillations in the quantum regime at a given temperature. The first frequency-assignment-based quasi-harmonic approach is the one devised by Schlitter in 1993.<sup>[184]</sup> Therein, a heuristic approximation is proposed as an entropy estimator for MD simulation data which is consistent with the third law of thermodynamics and thus works for quantum modes as well. Karplus and Andricioaei later gave a theoretical underpinning for Schlitter's heuristic derivation.<sup>[185]</sup> Baron *et al.* have generalized Schlitter's original proof, which shows that approximated quasi-harmonic entropies are an upper bound to the true entropy of the system.<sup>[186]</sup>

We now give a derivation of the quasi-harmonic method in spirit of the report published by Numata *et al.* in [154]. The estimator proposed by Schlitter is not discussed in detail. We note that Baron *et al.* have urged caution with respect to the usage of Schlitter's heuristic approximation. They suggest using the exact (quasi-harmonic) entropy formula for molecular modeling purposes instead, which we will derive in the following sections.<sup>[186]</sup>

### 8.3.1. Derivation of a Frequency Assignment based Quasi-Harmonic Entropy Estimator

The spacial probability density function of single classical harmonic oscillator in an NVE ensemble can be calculated analytically as (see [187] and references therein)

$$\rho_{\text{Classical-HO}}(q) dq = \begin{cases} \frac{1}{\pi\sqrt{A^2-q^2}} & \text{if } q^2 < \frac{2E}{\omega^2 m} \\ 0 & \text{else} \end{cases}. \quad (61)$$

$A$  is the maximum amplitude of the oscillator. Conservation of the total energy  $E$  is assumed. The maximum amplitude for a given energy is equivalent to the turning point and can be calculated from the energy as

$$A = \sqrt{\frac{2E}{\omega^2 m}}. \quad (62)$$

We set the mean of the position of the classical harmonic oscillator  $\langle q \rangle = 0$ . The variance of the position  $q$  is calculated as

$$(\sigma[q])^2 = \langle q^2 \rangle - \langle q \rangle^2 = \int_{-A}^A \frac{q^2}{\pi\sqrt{A^2-q^2}} dq = \frac{1}{2}A^2. \quad (63)$$

$\sigma$  denotes the standard deviation, defined as the square-root of the variance. The quasi-harmonic method now uses Equation 63 to assign a maximum amplitude  $A$  to a mode (dimension) of a MD simulation trajectory from its standard deviation. In principle, the maximum and minimum value obtain for one mode in an MD simulation could be used to assign  $A$  directly as well, however, this is not common. A plausible explanation is that the MD simulations under consideration are actually run in an NVT ensemble, utilizing a thermostat-algorithm. Equation 61 instead assumes an NVE ensemble and therefore conservation of the total energy. In the approach of Schlitter,<sup>[184]</sup> it is only *a posteriori* that the NVT simulation data is approximated as a single harmonic-oscillator in a NVE ensemble, where  $E$  is determined from equipartition using

$$\langle E_{\text{potential}}(T) \rangle + \langle E_{\text{kinetic}}(T) \rangle = \langle E_{\text{total}}(T) \rangle \approx k_B T. \quad (64)$$

Once the maximum amplitude and turnings points have been determined, the assignment of a quasi-harmonic frequency  $\omega$  is straight-forward by trivial rearrangement of Equation 62. Combining Equation 64 and Equation 63, the final expression for the

standard deviation of a single classical harmonic oscillator in an NVE ensemble is

$$\sigma [q] = (m\beta\omega^2)^{-\frac{1}{2}}. \quad (65)$$

It is important to explicitly note again that the MD simulation trajectory of the molecule at hand may well exhibit anharmonic motion, which are neglected in quasi-harmonic approaches.

We note that the impact of the temperature has only been considered by means of the equipartition theorem invoked for the total energy (Equation 64). Regarding the probability density  $\rho_{\text{Classical-HO}}(x)$ , a single classical harmonic oscillator with conservation of the total energy is still assumed. Figure 11 gives a plot of the functional form of  $\rho_{\text{Classical-HO}}(x)$ . As we will now show, standard deviation based frequency assignment is identical for both an NVE ensemble with  $\langle E(T) \rangle$  determined from the equipartition theorem using Equation 64, and a NVT ensemble explicitly including thermal effects for non-quantum modes. The following paragraphs will show that

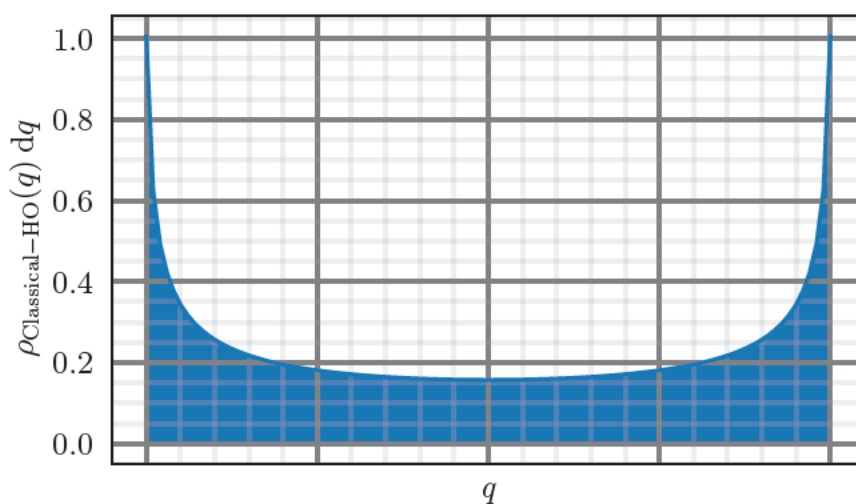


Figure 11: Spatial probability density  $\rho_{\text{Classical-HO}}(q) dq$  of a classical harmonic oscillator.

explicit inclusion of the density of states due to temperature effects using a classical statistical thermodynamics perspective will lead to a different spacial probability density  $\rho_{\text{Classical-HO}}(x)$ , specifically a gaussian distribution. We assume a statistical ensemble of classical Harmonic Oscillators (HOs) at a given temperature  $T$ . The classical statistical partition function of a classical HO can be given analytically (see [188]) as

$$Z_{\text{Classical-HO}} = \frac{1}{\beta\omega\hbar}. \quad (66)$$

The distribution of energies of a statistical ensemble of harmonic oscillators at the temperature  $T = (\beta k_B)^{-1}$  is given by the Boltzmann probability

$$\rho_{\text{Classical-HO}}^{\text{ensemble}}(E) = \frac{e^{-\beta E}}{Z_{\text{Classical-HO}}} = \frac{e^{-\beta E}}{\beta \omega \hbar}. \quad (67)$$

The collective spatial probability density  $\rho_{\text{Classical-HO}}^{\text{ensemble}}(q)$  can be calculated as the weighted mean value of all single spatial densities  $\rho_{\text{Classical-HO}}(q)$  over the Boltzmann probabilities  $\rho_{\text{Classical-HO}}^{\text{ensemble}}(E)$ :

$$\rho_{\text{Classical-HO}}^{\text{ensemble}}(q, \beta) = \int_0^\infty \rho_{\text{Classical-HO}}(q) \frac{e^{-\beta E}}{Z_{\text{Classical-HO}}} \frac{dE}{\hbar \omega} \quad (68)$$

Restricting the integration bounds to physically meaningful energies, this allows one to write

$$\rho_{\text{Classical-HO}}^{\text{ensemble}}(q, \beta) = \int_{\frac{x^2 \omega^2 m}{2}}^\infty \frac{1}{\pi \sqrt{\frac{2E}{\omega^2 m} - q^2}} e^{-\beta E} \beta \hbar \omega \frac{dE}{\hbar \omega}. \quad (69)$$

Using computer algebra software,  $\rho_{\text{Classical-HO}}^{\text{ensemble}}(q, \beta)$  can be determined analytically using symbolic integration and subsequent determination of the limits of the integral by series expansion.<sup>[189]</sup> This yields

$$\rho_{\text{Classical-HO}}^{\text{ensemble}}(q, \beta) = \frac{e^{-\frac{1}{2}mq^2\beta\omega^2} \sqrt{m\beta\omega}}{\sqrt{2\pi}}. \quad (70)$$

The probability density given in Equation 70 is normalized to unity and is a normal distribution in  $q$  for given set of parameters  $\beta, m$ , and  $\omega$ . It has units of meters. Incidentally, it has the same standard deviation as  $\rho_{\text{Classical-HO}}(q)$  (see Equation 61 and 65):

$$\sigma \left[ \rho_{\text{Classical-HO}}^{\text{ensemble}}(q, \beta) \right] = \left( m\beta\omega^2 \right)^{-\frac{1}{2}} \quad (71)$$

Therefore, for the goal of assigning a quasi-harmonic frequency to a mode using its standard deviation, equipartition yields the same result as a full treatment using classical statistical mechanics.

Lastly, Equation 48 can be invoked using the determined  $\omega$  to yield an approximated vibrational entropy contribution for a MD simulation mode. As Equation 48 is derived using the quantum-mechanical partition function of a harmonic oscillator, it obeys the third law of thermodynamics and Nernst's theorem holds. The final expression for



determining a quasi-harmonic vibrational entropy of molecular modes from their MD simulation derived standard derivations  $\sigma_i$  can be determined by combining Equation 71 with Equation 48:

$$S_{\text{quasi-harmonic}} = N_A k_B \sum_{i=1}^{3N_{\text{atoms}}-6} \left( \frac{\frac{\hbar k_B T}{\sigma_i \sqrt{m\beta}}}{e^{\frac{\hbar k_B T}{\sigma_i \sqrt{m\beta}}} - 1} - \ln \left( 1 - e^{-\frac{\hbar k_B T}{\sigma_i \sqrt{m\beta}}} \right) \right) \quad (72)$$

While the quasi-harmonic approximation can in principle be invoked for any set of spacial coordinates, it is useful to invoke collective, decoupled modes in practice. For orthogonal modes, all linear correlation (covariance) is removed. The covariance of two modes or dimensions is defined mathematically as

$$\text{cov}(q_i, q_j) = \langle (q_i - \langle q_i \rangle) (q_j - \langle q_j \rangle) \rangle, \quad (73)$$

with

$$\text{cov}(q_i, q_i) = (\sigma [q_i])^2. \quad (74)$$

Entropic contributions possibly spread out over several cartesian modes might accurately be grouped into one collective mode. A commonly known set of collective modes are the normal modes. These coordinates are the ones in which the representation of the Hessian matrix of second derivatives of the nuclei at a geometry minimum is diagonal. In order to find the transformation matrix which transforms cartesian into normal modes, the second derivatives of all cartesian nuclear coordinates have to be calculated. The eigenvector resulting from the solution of Equation 60 give the transformation matrix. Depending on the form of the nuclear potential, which might explicitly include electronic structure by means of quantum chemical or DFT approaches, the evaluation of the second derivatives may be a computationally demanding task.

### 8.3.2. Orthogonal Collective Modes from Simulation Data: The Principal Component Representation

As laid out, it is favorable for the quasi-harmonic entropy estimation procedure to switch the coordinate representation from the correlated atomic cartesian coordinates to a set of collective modes which are decoupled. One commonly used set of collective

modes derivable from MD simulation data is the principal component representation. This coordinate transform is obtained by a procedure termed Principal Component Analysis (PCA). We now give a brief description of cartesian PCA. Recently, other PCA-inspired techniques operating in the circular space of dihedral angular coordinate have been developed, such as the dihedral Principal Component Analysis (dPCA) approach by Stock and coworkers (see [190–194]), as well as Principal Geodesic Analysis (PGA) approaches (see [195–197]). We restrict the following discussion to regular, cartesian PCA only. For a full derivation see [190] and references therein. The following derivation proceeds in the spirit of the brief PCA description in [154].

We begin by considering the trajectory of an MD simulation, represented as an  $(3 N_{\text{atoms}}) \times N_{\text{snapshots}}$  matrix  $\mathbf{X}$ . This matrix is first mass-weighted, in close analogy to the procedure of calculating normal modes from a cartesian Hessian matrix. This is done by multiplying a diagonal  $(3 N_{\text{atoms}}) \times (3 N_{\text{atoms}})$  matrix  $\mathbf{M}$  containing the square-root of the atomic mass for each cartesian mode from the left. If mass-weighting is omitted, the analogous procedure is termed essential-mode analysis. This technique has been especially pursued by the Berendsen group.<sup>[198]</sup>

$$\mathbf{X}^\dagger = \mathbf{M} \mathbf{X} \quad (75)$$

Then, the  $(3 N_{\text{atoms}}) \times (3 N_{\text{atoms}})$  covariance matrix  $\Sigma$  of  $\mathbf{X}^\dagger$  is calculated in a second step. A matrix-based procedure (according to [199]) may proceed as:

$$\mathbf{X}^\ddagger = \mathbf{X}^\dagger - \left[ \vec{\mathbf{1}} (\vec{\mathbf{1}})^\dagger \right] \mathbf{X}^\dagger (N_{\text{snapshots}})^{-1} \quad (76)$$

$$\Sigma = (\mathbf{X}^\ddagger)^\dagger \mathbf{X}^\ddagger (N_{\text{snapshots}})^{-1} \quad (77)$$

where  $\vec{\mathbf{1}}$  is a column vector containing only the scalar value one with length  $N_{\text{snapshots}}$ . The covariance matrix  $\Sigma$ , which is symmetric, real, and semi-definite, needs to be diagonalized in a last step. The eigenvalues of  $\Sigma$  can be directly related to the standard deviation of the respective principal mode. Concerning MD simulation data, zero eigenvalues and thus zero variances can result if for instance translation and rotation of the molecule have been removed prior to the PCA procedure by proper alignment. In this case, six dimensions, corresponding to translation in  $x$ ,  $y$ , and  $z$  direction and rotation along the three Euler angles, will have zero variance. As we aim at using the

principal component representation to calculate collective modes for quasi-harmonic vibrational entropy analysis, the removal of translation and rotation is mandatory in these cases. As a decomposition algorithm, a singular value decomposition is commonly utilized, as it is able to decompose even semi-definite matrices.

The trajectory in the representation of the principal component collective modes  $\mathbf{X}^{\mathbb{N}}$ , which are orthonormal, can be calculated from the column matrix of eigenvectors  $\mathbf{V}$  of the covariance matrix  $\Sigma$ :

$$\mathbf{X}^{\mathbb{N}} = \mathbf{V}^t \mathbf{X} \quad (78)$$

The singular-values  $\lambda_i$  of  $\Sigma$  yield the variances  $\sigma^2$  multiplied by the effective-mass of the collective principal component mode:

$$\lambda_i = m_{\text{eff}}^i \sigma_i^2 \quad (79)$$

According to [154], the quasi-harmonic frequency assignment procedure for linearly decoupled principal component modes is then given as

$$\omega_i = \sqrt{\frac{k_B T}{m_{\text{eff}}^i \sigma_i^2}} = \sqrt{\frac{k_B T}{\lambda_i}}. \quad (80)$$

The effective mass  $m_{\text{eff}}^i$  can be explicitly calculated as well. For this purpose, we assume eigenvectors to be possibly unnormalized and consider their renormalization explicitly. The mathematical expression then is

$$\frac{1}{m_{\text{eff}}^i} = \sum_k \frac{\mathbf{V}_{k,i}^2}{m_k}. \quad (81)$$

#### 8.4. Differential Entropies from Molecular Simulation via Ergodicity

For larger chemical systems, both the calculations of the Hessian matrix of nuclear second derivatives, and the identification of all possible conformations become computationally unfeasible when sufficiently sophisticated force fields or explicit electronic structure estimation approaches are used. For this reason, other means of calculating the absolute entropy numerically for molecular systems are sought. One common simulation procedure in computational molecular sciences are MD simulations, where the motion of a molecular system is propagated in time. Quantum-mechanical dynamics simulations are possible only for the smallest of systems.<sup>[200]</sup> If quantum approaches

for time propagation of a molecular system are used, they are often referred to as Quantum Dynamics Simulations, while the term MD simulation is mostly reserved for classical, Newtonian or Langevinian propagation. Common algorithms for performing classical molecular simulations include the velocity-verlet algorithm or the leapfrog algorithm (see [201]). The nuclei are propagated according to the Newtonian law  $F = ma$  in a potential  $V$  for which holds  $F = -\frac{dV}{dx}$ . The potential can be taken from quantum chemical or density functional theory-based calculations. For large biomolecules, a mechanical force fields with empirically fitted terms may be used.

In any case, the result of a numerical MD simulation is a trajectory, which is given as the set of structures obtained by means of time-propagation of the potential  $V$  using an approximate integration scheme to solve the Newtonian equations of motion. It is now of interest to extract absolute entropy estimates from this molecular trajectory as an alternative to the RRHO Approximation presented in the previous section.

We will outline the mathematical steps needed in order to built an entropy expression applicable to MD data. The derivation starts from Equation 25, which contains no approximations. If all probabilities of all possible quantum states were known, this equation would give the exact (non-relativistic) entropy for a quantum system. As the nuclei are propagated using purely classical mechanics in MD simulations, we first substitute the sum over discrete quantum-states in Equation 25 with an integration over the continuous classical phase space:

$$S_{\text{Classical}} = - \int_{\epsilon} \rho(E_{\epsilon}) \ln(\rho(E_{\epsilon})) dE \quad (82)$$

The following derivation is loosely based on [202]. If cartesian coordinates are used, the energy  $E_{\epsilon}$  can be rewritten using the Hamiltonian function  $H(p, q)$  as

$$E_{\epsilon}(p, q) = H(p, q) = K(p, m) + U(q), \quad (83)$$

where

$$K(p, m) = K = \frac{1}{2} \sum_{i=1} p_i^2 / m_i. \quad (84)$$

The kinetic energy contribution  $K$  is thus only a function of the classical momenta  $p$  and the masses of the nuclei, while the potential energy  $U$  is solely a function of the coordinates  $q$ . This leads to a factorization of the probability density  $\rho(E_{\epsilon})$  into two

components:

$$\rho(E_\epsilon) = \rho(p, m, q) = \rho_{\text{kinetic}}(p, m) * \rho_{\text{spatial}}(q). \quad (85)$$

This enables one to write the total entropy  $S_{\text{Classical}}$  as a sum:

$$S_{\text{Classical}} = S_{\text{kinetic}}[\rho_{\text{kinetic}}(p, m)] + S_{\text{spatial}}[\rho_{\text{spatial}}(q)] \quad (86)$$

The integral that is the kinetic (momentum) contribution can be written in closed form as

$$\begin{aligned} S_{\text{kinetic}}[\rho_{\text{kinetic}}(p, m)] &= -k_B \int dp \rho_{\text{kinetic}}(p, m) \ln[h^d \rho_{\text{kinetic}}(p, m)] \\ &= k_B \frac{d}{2} [1 + \ln(2\pi m/\beta)] - k_B \ln(h^d), \end{aligned} \quad (87)$$

where  $d$  is the number of DOFs, i.e. the dimensionality of the system. We explicitly note that the contribution of  $-k_B \ln(h^d)$  stems from the multiplicative factor of  $(h)^{-1}$  per DOF added by convention to the classical partition function to render the argument of its logarithm function unitless. The term  $-k_B \ln(h^d)$  is now arbitrarily grouped with the kinetic-contribution, as it is a constant factor and does not depend on the spatial coordinates  $q$ . However, the separation of the total entropy according to Equation 86 is unable to preserve the correct handling of physical units. This means that both  $S_{\text{kinetic}}$  and  $S_{\text{spatial}}$  do not have physically correct units.

As a result, absolute spatial or kinetic entropy values are meaningless, as they depend on the choice of variables and are not invariant to canonical coordinate transformations. However, for cartesian coordinates at a given, finite temperature, entropy changes are physically meaningful, because here the troubling factors cancel:

$$\Delta S_{\text{Classical}} = \Delta S_{\text{spatial}} \quad \text{if } \Delta T = 0 \quad (88)$$

If entropy differences at a constant temperature are of interest, such as the entropy difference between two molecular conformations or a bound/unbound guest-host enzyme-system, the following Equation may be evaluated to calculate  $\Delta S$ :

$$S_{\text{spatial}} = -k_B \int dq \rho_{\text{spatial}}(q) \ln[\rho_{\text{spatial}}(q)] \quad (89)$$

The main idea of MD simulation based entropy estimation is now to take the probability density for the spatial coordinates  $\rho_{\text{spatial}}$  from the observed motions of the

simulation trajectory instead of from the Boltzmann Distribution of all accessible microstates (i.e. molecular conformations). A naive implementation could be to calculate a bin-discretion histogram-version of  $\rho_{\text{spatial}}$ . Hnizdo *et al.* have investigated the performance of a this procedure and found it to be inferior to other comparable estimators.<sup>[203]</sup> For this reason, the histogram-approach will not be considered in more detail in this work.

Equation 89 was derived to express the spatial contribution to the total configurational entropy of a molecule. Due to its integral formation, it is only valid in the regime of classical mechanics, where phase space can be reasonably assumed to be continuous. As mentioned, Equation 89 does not have proper physical units, as the expression  $h^s$ , which solely carries the unit of measurement, is arbitrarily grouped with the momentum contribution (see Equation 87). The mathematical form of Equation 89 is known as a differential entropy. In general, the differential entropy of a continuous random variable  $x$  defined on the domain  $D$  with a probability distribution  $\rho$  is defined as<sup>[204]</sup>

$$S_{\text{differential}}[\rho(x)] = - \int_D \rho(x) \ln \rho(x) dx. \quad (90)$$

The differential entropy is to be contrasted with the discrete Shannon-type entropy of either a discrete random variable ( $\vec{x}$ ) (or, alternatively, of a given sample ( $\vec{x}$ ) drawn from an arbitrary probability distribution):

$$S_{\text{discrete}}[\vec{x}] = - \sum_{i=1} \rho(x_i) \ln \rho(x_i) \quad (91)$$

## 8.5. Properties of Differential Entropies

In this section we will present some important general properties and concepts which apply to differential entropies, and thus to Equation 89.

The continuous nature of phase space in classical mechanics needs to be addressed in more detail: The Shannon-type entropy formula  $S = - \sum_n w_n \ln w_n$  (Equations 91 and 25) gives the entropy for a quantum system exactly, where  $w_n$  is the probability of being in a quantum state with an associated energy eigenvalue  $E_n$  (assuming no degeneracy). Exchanging the summation for an integration over the entire energy range directly results in the differential entropy expression given previously in Equation 82. Equation 82 is identical to the definition of entropy in classical statistics according to Landau & Lifshitz in [175] as well. It is however clear, that Equation 82 is in a physical



sense inferior to Equation 25. For instance, a pure quantum-mechanical eigenstate  $n$  with energy  $E_n$  has  $S = 0$  according to Equation 25, but setting  $\rho(E_n) = \delta E_n$  using the Dirac  $\delta$ -function yields  $S = -\infty$ .<sup>[205]</sup> Thus, differential entropy can become negative and infinite, which is unphysical: It is a direct violation of the third law of thermodynamics, which can be stated by using the Nernst heat theorem: “The entropy of any pure substance in thermodynamic equilibrium approaches zero as the temperature approaches zero”.<sup>[206]</sup>

A less formal and more chemical result of this shortcoming can be seen when considering the standard entropy expressions for a harmonic oscillator, which is solely a function of frequency and temperature. If the quantum-statistical partition function

$$Z_{\text{vibration}} = e^{-\frac{1}{2} \frac{\hbar\omega}{k_B T}} \left( 1 - e^{-\frac{\hbar\omega}{k_B T}} \right)^{-1} \quad (92)$$

is used, the correct behavior of the entropy when moving into the quantum-regime (high frequencies or - equivalently - low temperatures) is obtained as

$$\lim_{\omega \rightarrow \infty} S_{\text{vibration}} = \lim_{T \rightarrow 0K} S_{\text{vibration}} = 0. \quad (93)$$

The explicit entropy expression for classical harmonic oscillators is derived explicitly in [132]. It is defined as

$$S_{\text{classical-HO}} = -k_B \sum_{i=1}^N [\ln(\alpha_i) - 1] \quad (94)$$

with

$$\alpha_i = \frac{\hbar\omega_i}{k_B T}. \quad (95)$$

We recast Equation 48 for quantum harmonic oscillators using  $\alpha_i$  as well and obtain

$$S_{\text{quantum-HO}} = k_B \sum_{i=1}^N \left[ \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln(1 - e^{-\alpha_i}) \right]. \quad (96)$$

If the classical partition function

$$Z = \frac{1}{h^3} \int e^{-\frac{\beta p^2}{2m}} d^3 \vec{p} \int e^{-\frac{\beta kx^2}{2}} d^3 \vec{x}, \quad (97)$$

where identical particles are assumed to be distinguishable, is used,

$$\lim_{\omega \rightarrow \infty} S_{\text{classical-HO}} = \lim_{T \rightarrow 0K} S_{\text{classical-HO}} = -\infty \quad (98)$$

results. The multiplicative factor  $(h^3)^{-1}$  is included by convention to render the partition function unitless. Again, entropies are shown to become negative and thus unphysical within the quantum regime if mathematical expressions derived from classical mechanical assumptions are used. In Figure 12,  $\alpha_i = 1$  can be seen to represent the branching point between the classical and quantum regimes, with  $S_{\text{classical-HO}}$  and  $S_{\text{quantum-HO}}$  agreeing sufficiently for  $\alpha_i < 1$ , but diverging from each other for  $\alpha_i > 1$ .

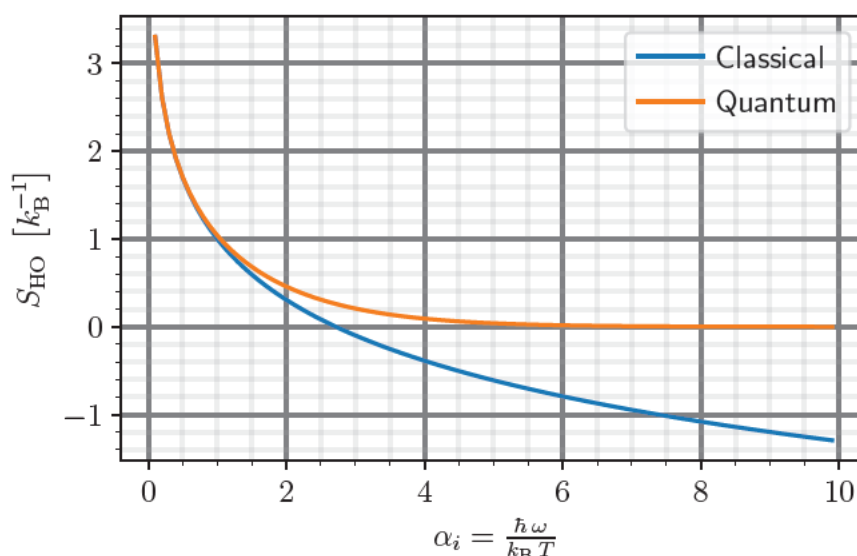


Figure 12: Divergence of the classical harmonic entropy in the quantum regime. The plot is created using expressions derived from statistical mechanics, adapted from [154].

### 8.5.1. The Mathematical Relation of Differential and Discrete Entropies

Equation 90 can be rationalized naively by substituting the summation in Equation 91 with an integration. However, a formal treatment reveals a more complex situation. In the following section, which will investigate the relationship of discrete and continuous (differential) entropy expressions in detail, following the presentation given by Cover and Thomas in [207].

We consider a one-dimensional probability distribution  $\rho(x)$  defined on the domain  $\mathbb{R}$  for the continuous random variable  $x$ . The differential entropy of this distribution

can be evaluated directly from Equation 90. If a certain number  $n$  of samples drawn from  $\rho(x)$  are available as an  $n$ -dimensional vector  $(\vec{x})$ , the discrete Shannon entropy of these samples is given as

$$S_{\text{discrete}}[\vec{x}] = - \sum_{i=1}^n \rho(x_i) \ln \rho(x_i). \quad (99)$$

We now consider the Riemann integral of  $\rho$  by dividing the domain (assumed to be  $\mathbb{R}$  without loss of generality) of  $\rho$  into  $m$  intervals of equal size  $\Delta$ . If the bins are suitably small, we may assume the value of  $\rho$  to be approximately constant inside each bin:

$$\rho(i \cdot \Delta) \approx \rho((i+1) \cdot \Delta) \quad \wedge \quad \int_{i\Delta}^{(i+1)\Delta} \rho(x) dx \approx \rho(i \cdot \Delta) \cdot \Delta \quad (100)$$

From this consideration a discrete version of the continuous random variable  $x$ , termed  $x^\dagger$ , may be defined as

$$x^\dagger = x_i \quad \forall \quad i \cdot \Delta \leq x < (i+1) \cdot \Delta. \quad (101)$$

The discretization procedure is exemplified graphically in Figure 13. As hinted in

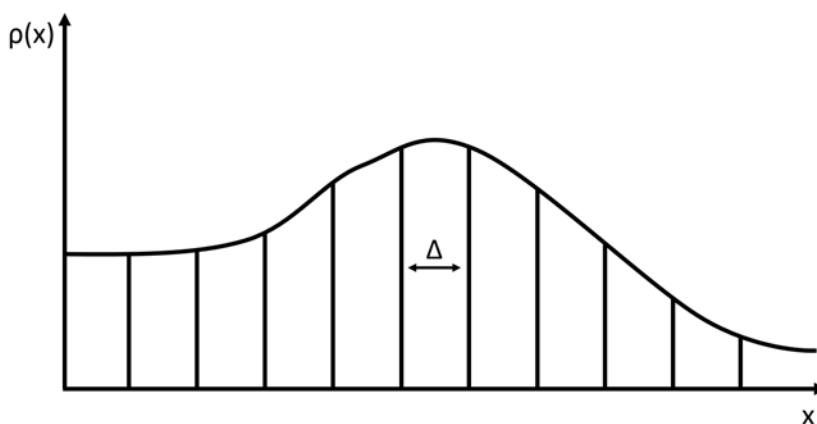


Figure 13: Exemplification of the quantization of a continuous probability density function.

Equation 100, the probability  $\rho_i$  that  $x^\dagger$  has a certain value  $x^\dagger = x_i$  may be evaluated as

$$\rho_i = \int_{i\Delta}^{(i+1)\Delta} \rho(x) dx \approx \rho(i \cdot \Delta) \cdot \Delta. \quad (102)$$

It furthermore holds that

$$\int_{-\infty}^{\infty} \rho(x) dx = 1 = \lim_{\Delta \rightarrow 0} \sum_{i=1}^m x_i \Delta. \quad (103)$$

We now evaluate the discrete Shannon-type entropy (Equation 91) of the discrete random variable  $x^\dagger$  and rearrange using Equation 103:

$$\begin{aligned} S_{\text{discrete}}[x^\dagger] &= - \sum_{i=1}^{\infty} \rho_i \ln \rho_i \\ &= - \sum_{i=1}^{\infty} \Delta \cdot \rho(x_i) \ln(\Delta \cdot \rho(x_i)) \\ &= - \sum_{i=1}^{\infty} \Delta \cdot \rho(x_i) \ln \rho(x_i) - \sum_{i=1}^{\infty} \Delta \cdot \rho(x_i) \ln \Delta \\ &= - \sum_{i=1}^{\infty} \Delta \cdot \rho(x_i) \ln \rho(x_i) - \ln \Delta \end{aligned} \quad (104)$$

Note that as  $\rho$  is defined on the domain  $D = \mathbb{R}$ , the binning procedure yields an infinite amount of discrete bins and thus the summations' upper bounds in Equation 104 are  $\infty$ . If we consider the expression  $-\sum \Delta \cdot \rho(x_i) \ln \rho(x_i)$  to be Riemann integrable, we obtain the definition of the differential entropy (see Equation 90) as the first summand in Equation 104:

$$\begin{aligned} S_{\text{discrete}}[x^\dagger] &= - \sum_{i=1}^{\infty} \rho_i \ln \rho_i \\ &= S_{\text{differential}}[\rho(x)] - \ln \Delta \end{aligned} \quad (105)$$

Thus, the absolute value of the differential entropy is not equal to the absolute value of the discretized, Shannon-type entropy. It differs by a constant offset of  $-\ln \Delta$ , where  $\Delta$  is the bin-width. One peculiar fact arises when the limit of increasingly smaller bin sizes  $\Delta$  is considered as

$$\lim_{\Delta \rightarrow 0} S_{\text{discrete}}[x^\dagger] = S_{\text{differential}}[\rho(x)] - \lim_{\Delta \rightarrow 0} \ln \Delta, \quad (106)$$

with

$$\lim_{\Delta \rightarrow 0} \ln \Delta = -\infty. \quad (107)$$

Increasingly smaller bins, which can be rationalized as considering a higher spatial resolution, lower the absolute value of the differential entropy by a constant, which in the limit of full resolution becomes infinite in magnitude. At this point it is important to note that the thermodynamic spatial entropy derived previously and defined in Equation 89 is a differential entropy derived from a Shannon-type discrete entropy (Equation 25). This implies that when Equation 89 is evaluated from samples containing spatial coordinates, taken from an MD simulation, it will necessarily differ by a constant factor from Equation 25, which is exact within non-relativistic quantum theory. The reason for this is that in the derivation of Equation 82, we naively substituted the summation of Equation 25 with an integration. Thereby, the correction factor  $-\ln \Delta$  was neglected implicitly.

### 8.5.2. Translation and Scaling of Differential Entropies

In this section we describe the behavior of differential entropies under the translation or scaling of their underlying continuous random variable. This has important implications for the evaluation of spatial entropies from molecular simulation data, as scaling of the differential variable corresponds directly to the choice of distance units (e.g. meters versus ångström).

The differential entropy can be identified to be translationally invariant. This follows trivially from its definition in Equation 90, as integration is performed over the entire domain of the random variable. It thus holds that

$$S_{\text{differential}}[\rho(a+x)] = S_{\text{differential}}[\rho(x)]. \quad (108)$$

Perhaps more interestingly, the differential entropy can be shown to be shifted by a constant scalar factor if its continuous random variable is scaled by a constant factor. In one dimension, it holds that

$$S_{\text{differential}}[\rho(a * x)] = S_{\text{differential}}[\rho(x)] + \ln |a|, \quad (109)$$

which generalizes to

$$S_{\text{differential}}[\rho(\mathbf{A} * \vec{x})] = S_{\text{differential}}[\rho(\vec{x})] + \ln |\det(\mathbf{A})| \quad (110)$$

in higher dimensions, using the scaling matrix  $\mathbf{A}$ . The proof is given in [207].

### 8.5.3. The Mutual Information Expansion of the Differential Entropy

A further important property of  $S_{\text{discrete}}$  (Equation 91) and equivalently of differential entropy expressions is that for multi-dimensional probability densities  $\rho$  the sum of marginal (one-dimensional) entropies can be shown to be a strict upper bound for the true, multi-dimensional and possibly correlated entropy:

$$\int d\vec{q} \rho_{\text{spatial}}(\vec{q}) \ln[\rho_{\text{spatial}}(\vec{q})] \leq \sum_{i=1}^d \int dq_i \rho_{\text{spatial}}(q_i) \ln[\rho_{\text{spatial}}(q_i)], \quad (111)$$

with equality when all dimensions are fully uncorrelated.<sup>[207]</sup> Equality in Equation 111 is obtained when it holds that

$$\rho(\vec{q}) = \prod_{i=1}^d \rho(q_i). \quad (112)$$

The approximation of writing a multi-dimensional entropy as the sum of its one-dimensional contributions fully neglects correlation between the different dimensions (modes). This idea can be generalized in a series expansion of the exact, multi-dimensional entropy as a sum of contributions with increasing dimensionality. This expansion is termed the Mutual Information Expansion (MIE). The difference between the exact entropy for two chosen dimensions and the sum of their marginal distributions is called Mutual Information (MI)  $I$ :

$$I(i, j) = S(i) + S(j) - S(i, j). \quad (113)$$

MI captures all information that results from the simultaneous knowledge of two variables, with respect to the knowledge of only each separate variable. Equivalently, it can be stated that MI captures all correlation of two variables, generalizing the previously introduced concept of covariance (Equation 73) which only considered linear correlation. In contrast, MI captures all supralinear correlations as well. If the MI of two variables is identical to zero, they are fully uncorrelated. The concept of MI can be extended to more than two dimension and is then termed higher order MI:

$$I_d = \sum_{i=1}^d (-1)^{i+1} \left[ \sum_{j_1 < \dots < j_i} S(j_1, \dots, j_i) \right]. \quad (114)$$



The summation  $\sum_{j_1 < \dots < j_i} S(j_1, \dots, j_i)$  includes all unique permutations of the different dimensions, whose total number for each summand of the outer sum is  $i$ . To render the mathematical expression clearer, we give explicit examples for  $I_1(i)$ ,  $I_2(i, j)$ ,  $I_3(i, j, k)$ ,  $I_3(i, k, l)$ , and  $I_4(i, j, k, l)$ , where the indices  $i, j, k, l$  denote different dimensions:

$$\begin{aligned}
 I_1(i) &= S(i) \\
 I_2(i, j) &= S(i, j) - S(i) - S(j) \\
 I_3(i, j, k) &= S(i, j, k) - S(i, j) - S(i, k) - S(j, k) + S(i) + S(j) + S(k) \\
 I_3(i, k, l) &= S(i, k, l) - S(i, k) - S(k, l) - S(i, l) + S(i) + S(k) + S(l) \\
 I_4(i, j, k, l) &= S(i, j, k, l) - S(i, j, k) - S(i, j, l) - S(i, l, k) + S(i, j) + S(j, k) + S(k, l) \\
 &\quad + S(i, k) + S(i, l) + S(j, l) - S(i) - S(j) - S(k) - S(l)
 \end{aligned} \tag{115}$$

The MIE is defined as the series expansion of the entropy in terms of higher order MI contributions:

$$S = \sum_{a=1}^d (-1)^{a+1} \left( \sum_{\substack{\text{all possible permutations} \\ \text{of } a \text{ dimensions}}} I_a(\dots) \right) \tag{116}$$

The MIE is exact in its untruncated form. By correcting a marginal sum of one-dimensional entropies for MI, all contributions to the total entropy due to information contained in two separate modes (which was counted twice in the 1<sup>st</sup> order expansion) is removed. This is equivalently called the 2<sup>nd</sup> order MIE. The corrections for MI strictly lower the absolute entropy with respect to the marginal sum of one-dimensional entropies. As mentioned, the latter is itself upper bound to the total entropy. As high-dimensional entropies are notoriously hard to evaluate computationally, the MIE can be truncated at a certain order to yield a computationally feasible approximation for the full-dimensional entropy. This approach has been used in a number of scientific reports.<sup>[14,15,154,208]</sup>

However, recently it has been reported that the MIE may suffer from divergence issues. Namely, in the case of long-range correlation, higher orders of the expansion might contribute largely and thus truncated expansions might yield inaccurate, fluctuating numerical results.<sup>[209]</sup>

### 8.5.4. Multivariate Gaussian Entropy as a Strict Upper Bound

In this section we consider an unknown probability distribution  $\rho$  defined on a domain  $D = \mathbb{R}$  with a given standard deviation

$$\sigma[\rho(x)] = \sqrt{\sum_{i=1}^N \rho_i (x_i - \mu)^2} = \sqrt{\int_D (x - \mu)^2 \rho(x) dx} \quad (117)$$

and mean

$$\mu[\rho(x)] = \int_D x \rho(x) dx. \quad (118)$$

We introduce the Kullback–Leibler divergence (sometimes termed relative entropy, the preferred name by the original author is minimum discrimination information<sup>[210]</sup>), which is measure of (dis)similarity of two probability distributions.<sup>[211]</sup> The Kullback–Leibler divergence for the distributions  $\rho$  and  $\rho^\dagger$  of continuous variables defined on the domain  $D$  is

$$D_{\text{KL}}(\rho, \rho^\dagger) = \int_{-\infty}^{\infty} \rho(x) \ln \left( \frac{\rho(x)}{\rho^\dagger(x)} \right) dx. \quad (119)$$

$D_{\text{KL}}$  is always non negative:

$$D_{\text{KL}}(\rho, \rho^\dagger) \geq 0 \quad (120)$$

Furthermore, the Kullback–Leibler divergence can be related to MI, which was introduced in Section 8.5.3 in Equation 113:

$$D_{\text{KL}}(\rho(i, j), \rho(i) \cdot \rho(j)) = I(i, j) \quad (121)$$

We will now give a proof that a gaussian with the same standard deviation  $\sigma$  as  $\rho$ ,

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{\left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right)} \quad (122)$$

it holds strictly that:

$$S_{\text{differential}}[f(x)] \geq S_{\text{differential}}[\rho(x)] \quad \leftarrow \quad \sigma[\rho(x)] = \sigma[f(x)]. \quad (123)$$

For the proof we assume that both the gaussian function  $f(x)$  and  $\rho(x)$  have been shifted to  $\mu = 0$ . This shifting leaves the differential entropy unchanged for any  $\rho$  due to its property of translation invariance (see Section 8.5.2).

First, we evaluate the differential entropy of the arbitrary gaussian distribution func-

tion  $f(x)$  (proof adapted from [212]):

$$\begin{aligned} S_{\text{differential}}[f(x)] &= -\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \ln\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}\right) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \ln\left(\sigma\sqrt{2\pi} e^{\frac{x^2}{2\sigma^2}}\right) dx \end{aligned} \quad (124)$$

We invoke integration by substitution and plug in  $t = \frac{x}{\sqrt{2}\sigma}$ , considering  $\frac{dt}{dx} = \frac{1}{\sqrt{2}\sigma}$ :

$$\begin{aligned} S_{\text{differential}}[f(x)] &= \frac{1}{\pi} \int_{-\infty}^{\infty} e^{-t^2} \ln(\sigma\sqrt{2\pi} e^{t^2}) dt \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\ln(\sigma\sqrt{2\pi}) + \ln(e^{t^2})) e^{-t^2} dt \\ &= \frac{\ln(\sigma\sqrt{2\pi})}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt + \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt \end{aligned} \quad (125)$$

Recognizing that  $\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$  is a gaussian integral, we write:

$$S_{\text{differential}}[f(x)] = \ln(\sigma\sqrt{2\pi}) + \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt \quad (126)$$

We use integration by parts

$$\int_a^b u(x) \frac{dv(x)}{dx} dx = [u(x)v(x)]_a^b - \int_a^b \frac{du(x)}{dx} v(x) dx \quad (127)$$

to rewrite  $\int_{-\infty}^{\infty} t^2 e^{-t^2} dt$  as

$$S_{\text{differential}}[f(x)] = \ln(\sigma\sqrt{2\pi}) + \frac{1}{\sqrt{\pi}} \left( \left[ -\frac{t}{2} e^{-t^2} \right]_{-\infty}^{\infty} + \frac{1}{2} \int_{-\infty}^{\infty} e^{-t^2} dt \right). \quad (128)$$

It holds that  $\left[ -\frac{t}{2} e^{-t^2} \right]_{-\infty}^{\infty} = 0$ . We solve the integral  $\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$  and obtain a formula for the direct evaluation of the differential entropy of a gaussian solely from

its standard deviation  $\sigma [f(x)]$ :

$$\begin{aligned}
 S_{\text{differential}} [f(x)] &= \ln(\sigma\sqrt{2\pi}) + \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt \\
 &= \ln(\sigma\sqrt{2\pi}) + \frac{\sqrt{\pi}}{2\sqrt{\pi}} \\
 &= \ln(\sigma\sqrt{2\pi}) + \frac{1}{2} \\
 &= -\frac{1}{2} \ln(2\pi e\sigma^2)
 \end{aligned} \tag{129}$$

To commence the proof for Equation 123 (adapted form [213]), we write the Kullback-Leibler divergence of the gaussian  $f(x)$  and  $\rho$  using the analytical expression for the entropy of a normal distribution given in Equation 129:

$$D_{\text{KL}}(\rho(x), f(x)) = \int_{-\infty}^{\infty} \rho(x) \ln\left(\frac{\rho(x)}{f(x)}\right) dx = -S_{\text{differential}}[\rho(x)] - \int_{-\infty}^{\infty} \rho(x) \ln(f(x)) dx \tag{130}$$

Note that  $D_{\text{KL}}(\rho(x), f(x))$  is not symmetric and thus  $D_{\text{KL}}(\rho(x), f(x)) \neq D_{\text{KL}}(f(x), \rho(x))$ . We will now show that  $\int_{-\infty}^{\infty} \rho(x) \ln(f(x)) dx = -S_{\text{differential}}[f(x)]$ :

$$\begin{aligned}
 \int_{-\infty}^{\infty} \rho(x) \ln(f(x)) dx &= \int_{-\infty}^{\infty} \rho(x) \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\right) dx \\
 &= \int_{-\infty}^{\infty} \rho(x) \ln \frac{1}{\sqrt{2\pi\sigma^2}} dx + \ln(e) \int_{-\infty}^{\infty} \rho(x) \left(-\frac{x^2}{2\sigma^2}\right) dx
 \end{aligned} \tag{131}$$

Considering that  $\rho(x)$  is a probability density, which by convention yields one upon integration over its domain, we get

$$\begin{aligned}
 &= -\frac{1}{2} \ln(2\pi\sigma^2) - \ln(e) \frac{\sigma^2}{2\sigma^2} \\
 &= -\frac{1}{2} (\ln(2\pi\sigma^2) + \ln(e)) \\
 &= -\frac{1}{2} \ln(2\pi e\sigma^2) \\
 &= -S_{\text{differential}}[f(x)]
 \end{aligned} \tag{132}$$

Lastly, we use the non-negativity of the Kullback-Leitner divergence

$$D_{\text{KL}}(\rho(x), f(x)) = S_{\text{differential}}[f(x)] - S_{\text{differential}}[\rho(x)] \geq 0 \tag{133}$$

to show that  $\forall \sigma [\rho(x)] = \sigma [f(x)]$ , the differential entropy of the gaussian  $f(x)$  is equal to or larger than the differential entropy  $S_{\text{differential}}$  of  $\rho(x)$ . This proves Equation 123. From this proof it directly follows that the quasi-harmonic entropy determined *via* frequency assignment (Equation 72) is a strict upper bound for molecular vibrational entropies as well. This means that explicit consideration of anharmonicity may only lower the vibrational entropy of a molecular mode.

## 8.6. The kNN Entropy Estimator - Theory and Formal Considerations

We will now solely focus on the estimation of configurational entropies from molecular simulation data (i.e. a finite sample of spatial coordinates). Configurational entropies contain only vibrational and conformational contributions. Rotational and translational contributions are assumed to be removed.

We begin with the Shannon-type entropy formula for a statistical canonical ensemble, which is strictly valid in the realm of non-relativistic quantum physics.<sup>[178]</sup> As outlined in previous sections, we substitute the sum over quantum states with integration over classical continuous phase space (see Equation 82). The derived estimators are thus only valid in the regime of classical physics, quantum effects are neglected. Finally, the spatial contribution is separated from the momentum contribution, which can be evaluated analytically. We are left with Equation 89. The goal is now to derive an estimator for this expression.

First, we note that Equation 89 can be recast as an expectation value  $\mathbb{E}$ :

$$S_{\text{spatial}} = -k_B \int d\vec{q} \rho_{\text{spatial}}(\vec{q}) \ln(\rho_{\text{spatial}}(\vec{q})) = -k_B \mathbb{E}[\ln(\rho_{\text{spatial}}(\vec{q}))] \quad (134)$$

A maximum likelihood estimate of the entropy from  $N$  drawn samples can then be written as:

$$\hat{S}_{\text{spatial}} = -\frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} [\ln(\rho_{\text{spatial}}(\vec{q}_i))] \quad (135)$$

An estimator for  $\rho_{\text{spatial}}(\vec{q})$  from given samples has to be developed. Then, Equation 135 can be evaluated only from samples, without explicit knowledge of  $\rho_{\text{spatial}}(\vec{q})$ . To this end, we want to utilize the distances  $\zeta_{i,k}$  of each  $d$ -dimensional sample point  $\vec{q}_i$  to its  $k^{\text{th}}$  nearest neighbor in the sample quantity.

We provide an explicit example for a one-dimensional case:

If the sample quantity is given as the set  $\vec{q}^* = \{-0.5, 7.0, 1.5, 31.0, 101.7\}$ , the 1<sup>st</sup> nearest neighbor ( $k = 1$ ) of the first sample point  $q_1 = (-0.5)$  is  $q_3 = (1.5)$ . The associated nearest neighbor distance is  $\zeta_{i=1,k=1} = 2.0$ . For the third nearest neighbor ( $k = 3$ ) of  $q_1 = (-0.5)$  we calculate a nearest neighbor distance of  $\zeta_{i=1,k=3} = 31.5$ .

The heuristic for choosing this approach is that for each sample locally, nearest neighbor distances  $\zeta$  will be large whenever  $\rho_{\text{spatial}}(\vec{q}_i)$  is small, and vice versa. By locally we refer to  $\forall q_i \in [\vec{q}_i - dq, \vec{q}_i + dq]$  where  $dq$  is sufficiently small.

We reproduce a derivation of the kNN estimator in the spirit of Lombardi and Pant (see [214]). The  $k^{\text{th}}$  nearest neighbor distance in  $d$ -dimensional space  $\zeta_{d,i,k}$  of a  $d$ -dimensional sample point  $\vec{q}_i$  to its  $k^{\text{th}}$  nearest neighbor defines a region in space. This region is fully characterized by the criterion  $\ell_p(\vec{q} - \vec{q}_i) < \zeta_{d,i,k}$ , with the distance norm  $\ell_p(x)$ :

$$\ell_p(\vec{q}) = \begin{cases} \left\{ \sum_{j=1}^d |q_j|^p \right\}^{1/p} & \forall \quad 1 \leq p < \infty \wedge \vec{q} = (q_1, \dots, q_d) \in \mathbb{R}^d \\ \max_{1 \leq j \leq d} |q_j| & \text{if } p = \infty \wedge \vec{q} = (q_1, \dots, q_d) \in \mathbb{R}^d \end{cases} \quad (136)$$

In Equation 136,  $q_i$  denotes the  $i^{\text{th}}$  element of the vector  $\vec{q}$ . Therefore, a complete notation of  $\zeta$  includes the integer  $p$  of the  $\ell_p$  norm as well, as  $\zeta$  is a function of  $p$ :  $\zeta = \zeta_{i,k}^{(p)}$ . In the remainder of this text superscripts and subscripts of  $\zeta_{i,k}^{(p)}$  will be omitted for notational clarity whenever they are obvious or irrelevant.

For  $p = 2$ , the euclidean distance metric familiar from euclidean geometry is recovered.  $\zeta_{i,k}$  may be defined from a set of  $d$ -dimensional sample points  $Q = \{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n\}$  as the  $(k + 1)^{\text{th}}$  largest element of the  $i^{\text{th}}$  row of the symmetric matrix  $\mathbf{Z}$ , given by

$$\mathbf{Z}_{i,j} = \ell_p(\vec{q}_i - \vec{q}_j). \quad (137)$$

To render things more clearly, we give an explicit set  $Q$  for an arbitrary example of a set of three-dimensional sample points

$$Q_{\text{example}} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 9 \\ 6 \\ 3 \end{pmatrix}, \begin{pmatrix} 7 \\ 7 \\ 7 \end{pmatrix} \right\}. \quad (138)$$



Using the euclidean distance norm ( $\ell_2$ ), the matrix  $\mathbf{Z}_{\text{example}}$  is then given as

$$\begin{pmatrix} 0 & 1 & 15 & 12.12 \\ 1 & 0 & 9.64 & 10.39 \\ 15 & 9.64 & 0 & 4.58 \\ 12.12 & 10.39 & 4.58 & 0 \end{pmatrix}. \quad (139)$$

As part of the exemplification, we give some numerical values for the  $k^{\text{th}}$  nearest neighbor distance  $\zeta$  for the set of sample points  $Q_{\text{example}}$  given in Equation 138:

$$\begin{aligned} \zeta_{i=1,k=2} &= 12.12 \\ \zeta_{i=3,k=1} &= 4.58 \\ \zeta_{i=4,k=3} &= 12.12 \end{aligned} \quad (140)$$

It is also desirable to state a purely mathematical, formal definition of  $\zeta_{i,k}^{(p)}$ . For this purpose, we may select the  $i^{\text{th}}$  column of the matrix  $\mathbf{Z}$  using the elementary vector  $\vec{e}$ , whose entries are defined as  $e_j = \delta_{ij}$ . Then the  $i^{\text{th}}$  column of the matrix  $\mathbf{Z}$  is given as

$$\vec{z} = \mathbf{Z} \vec{e}. \quad (141)$$

Now we are able to define  $\zeta_{i,k}$  by first finding the set of  $k$  smallest values in  $\vec{z}$ . For this reason we define the set  $A = \{z_1, z_2, \dots, z_{\#\vec{z}}\}$  containing all entries of the vector  $\vec{z}$ . Then, using the argmin function known from set theory (see [215]), we define the subset  $B$  of  $A$  containing only its  $k + 1$  smallest elements as

$$B = \operatorname{argmin}_{A' \subset A, \#A' = k+1} \sum_{a \in A'} a. \quad (142)$$

Lastly,  $\zeta_{i,k}$  can be formally defined as the largest element, the supremum, of the set  $B$ :

$$\zeta_{i,k} = \sup B \quad (143)$$

In this way,  $\zeta_{i,k}$  can be defined formally.

We now move on to the derivation of the estimator  $\hat{\rho}_{\text{spatial}}(\vec{q}_i)$ , to be used in Equation 135. The probability mass  $P$  of  $\rho_{\text{spatial}}(\vec{q}_i)$  inside the  $k^{\text{th}}$  nearest neighbor region (see

Figure 14) can be given as

$$P_{\vec{q}_i}(\zeta_{i,k}) = \int_{\ell_p(\vec{q}-\vec{q}_i) < \zeta_{i,k}} \rho_{\text{spatial}}(\vec{q}) d\vec{q}. \quad (144)$$

The expected number of sample points  $\langle k \rangle$  found within the region defined by

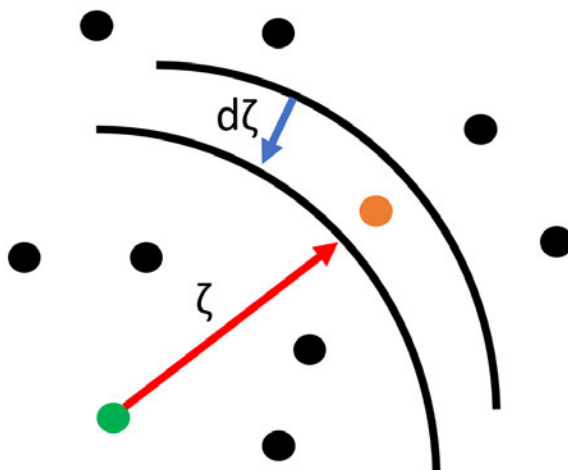


Figure 14: Geometric depiction of the  $k^{\text{th}}$  nearest neighbor algorithm's approach in two dimensions. The orange point is the  $k = 5^{\text{th}}$  nearest neighbor of the green point  $\vec{q}_i$  using the  $\ell_2$  distance norm. This picture is adapted from [214].

$\ell_p(\vec{q}-\vec{q}_i) < \zeta_{i,k}$  is calculated as

$$\langle k \rangle = N_{\text{samples}} * P_{\vec{q}_i}(\zeta_{i,k}). \quad (145)$$

kNN-type entropy estimators now assume that  $\rho_{\text{spatial}}(\vec{q})$  is constant inside  $\ell_p(\vec{q}-\vec{q}_i) < \zeta_{i,k}$ , which implies that  $P_{\vec{q}_i}(\zeta_{i,k})$  can be expressed as

$$P_{\vec{q}_i}(\zeta_{i,k}) = \tau * \rho_{\text{spatial}}(\vec{q}_i), \quad (146)$$

where  $\tau$  is a scalar constant. The original kNN entropy estimator proposed by Kozachenko and Leonenko ([216]), as well as Singh *et al.* ([217]), works under the assumption that

$$P_{\vec{q}_i}(\zeta_{i,k}) = (\zeta_{i,k})^d * c_d * \rho_{\text{spatial}}(\vec{q}_i), \quad (147)$$

where  $d$  is the dimensionality and  $c_d$  is the volume of a  $d$ -dimensional unit ball. Although we will sometimes neglect the superscript for notational clarity, the expression

$c_d$  is actually depending on the chosen  $\ell_p$  norm as well,<sup>[218]</sup>

$$c_d^{\ell_p} = \frac{\left(2\Gamma\left(\frac{1}{p} + 1\right)\right)^d}{\Gamma\left(\frac{d}{p} + 1\right)} \quad (148)$$

with

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx. \quad (149)$$

To clarify, using the euclidean ( $\ell_2$ ) distance norm with  $d = 2$ ,  $c_d^{\ell_2}$  is the area of a unit circle. For  $d = 3$ , it is the volume of a unit sphere. For  $d = 3$  using the maximum ( $\ell_\infty$ ) norm,  $c_d^{\ell_\infty}$  corresponds to the volume of a cube with an edge length of 2. Rearranging Equation 147 yields

$$\frac{P_{\vec{q}_i}(\zeta_{i,k})}{(\zeta_{i,k})^d * c_d} = \rho_{\text{spatial}}(\vec{q}_i). \quad (150)$$

Equation 150 can be inserted into Equation 135:

$$\hat{S}_{\text{spatial}} = -\frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ \ln \left( \frac{P_{\vec{q}_i}(\zeta_{i,k})}{(\zeta_{i,k})^d * c_d} \right) \right]. \quad (151)$$

In Equation 145, we set  $\langle k \rangle$  equal to the number of observed sample points ( $k$ ) inside the region  $\ell_p(\vec{q} - \vec{q}_i) < \zeta_{i,k}$ :

$$\frac{k}{N_{\text{samples}}} = P_{q_i}(\zeta_{i,k}) = (\zeta_{i,k})^d * c_d * \rho_{\text{spatial}}(\vec{q}_i) \quad (152)$$

Rearranging yields an estimator for  $\rho_{\text{spatial}}(\vec{q}_i)$ :

$$\hat{\rho}_{\text{spatial}}(\vec{q}_i) = \frac{k}{(\zeta_{i,k})^d * c_d * N_{\text{samples}}}. \quad (153)$$

The estimator  $\hat{\rho}_{\text{spatial}}$  is asymptotically biased.<sup>[217]</sup> It can be shown that - independently of the form of the true  $\rho_{\text{spatial}}(\vec{q})$  - the probability density  $\rho$  of  $P_{\vec{q}_i}(\zeta_{i,k})$  for each point  $\vec{q}_i$  follows a beta-distribution:<sup>[219]</sup>

$$\rho\left[P_{\vec{q}_i}(\zeta)\right] = \binom{N_{\text{samples}} - 1}{1} \binom{N_{\text{samples}} - 2}{k - 1} \left(P_{\vec{q}_i}^k(\zeta_{i,k})\right)^{k-1} \left(1 - \left(P_{\vec{q}_i}(\zeta_{i,k})\right)\right)^{N_{\text{samples}} - 1 - k} \quad (154)$$

The beta distribution can be rationalized as an extension of the binomial distribution for non-integer values.  $\rho\left[P_{\vec{q}_i}(\zeta)\right]$  can be rationalized as the probability of finding  $k - 1$  sample points in the interval  $\left[0, \zeta_{i,k}\right]$ , one point in  $\left[\zeta_{i,k}, \zeta_{i,k} + d\zeta\right]$  and all remaining  $N_{\text{samples}} - 1$  points farther away. Figure 14 shows this graphically. The value of  $P_{\vec{q}_i}(\zeta)$  is always between one and zero, as it is constituted by an integration over a subdomain of the probability distribution  $\rho_{\text{spatial}}(\vec{q}_i)$  (see Equation 144). The finding that  $P_{\vec{q}_i}(\zeta)$  follows a beta-distribution implies further facts which can be derived from general properties of the binomial distribution. We note that a binomial distribution is related to a beta-distribution for fractional, non-integer parameters as

$$\begin{aligned} \text{Beta}(x, \alpha, \beta) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \\ \text{Binomial}(x, n, k) &= \binom{n}{k} x^k (1-x)^{n-k} \\ \text{Binomial}(x, n, k) &= \frac{\text{Beta}(x, k+1, n-k+1)}{n+1}. \end{aligned} \quad (155)$$

In our case, the parameters  $\alpha$  and  $\beta$  of the beta-distribution are  $N_{\text{samples}} - 1$  and  $k$ , which are conceptually always integers. The expression given in Equation 154 can therefore also be viewed to describe a Bernoulli trial, i.e. a series of stochastic experiments where the outcome is binary (“success”: Finding a sample point inside the given interval, or “failure”: not finding a sample point inside the given interval). We note that the exact value of  $P_{\vec{q}_i}(\zeta_{i,k})$  can be determined from Equation 144. Thus, for a given  $\vec{q}_i$  and therefore a defined nearest neighbor distance  $\zeta_{i,k}$ ,  $\rho\left[P_{\vec{q}_i}(\zeta_{i,k})\right]$  needs to be a Dirac  $\delta$ -distribution with its peak at the exact value. However, since we are not able to calculate neither  $P_{\vec{q}_i}(\zeta_{i,k})$  nor  $\rho_{\text{spatial}}(\vec{q}_i)$  solely from drawn samples without an explicit knowledge of  $\rho_{\text{spatial}}(\vec{q})$ , this approach is not viable. If we recall that the order statistics sampled from a uniform distribution are each following a beta-distribution, we can rationalize the assumptions contained in taking  $P_{\vec{q}_i}(\zeta_{i,k})$  to be beta-distributed as such:

We are to draw  $N_{\text{samples}} - 1$  from a one-dimensional uniform distribution in the range  $[0, 1]$ , and we want to calculate the  $i^{\text{th}}$  largest sample value  $x_i$  of the total  $n = N_{\text{samples}} - 1$  samples drawn from this uniform distribution. The value  $x_i$  is then called the  $i^{\text{th}}$  order statistic. It can be shown that the  $i^{\text{th}}$  order statistic of a uniform distribution, if taken

as a random variable itself, is distributed according to a beta-distribution:

$$\rho[x_i] = (N_{\text{samples}} - 1) \binom{N_{\text{samples}} - 2}{k - 1} (1 - x_i)^{N_{\text{samples}} - 1 - k} x_i^{k-1} \quad (156)$$

This means that the probability of obtaining the number  $x_i$  as the  $i^{\text{th}}$  largest value when drawing  $n$  samples is beta distributed. Equivalently, Equation 144 can thus be seen as the probability distribution of drawing  $x_i = P_{\vec{q}_i}(\zeta_{i,k})$  as the  $k^{\text{th}}$  largest sample when drawing from a uniform probability distribution. Therefore, in the asymptotic limit of kNN entropy estimation for  $N_{\text{samples}} \rightarrow \infty$ , it is assumed that samples drawn are distributed uniformly across the spatial domain. The asymptotic bias of Equation 153 stems from the expectation value of the logarithm of a beta function, as we will show. We recognize the equivalence

$$\sum_{i=1}^{N_{\text{samples}}} \ln(P_{\vec{q}_i}(\zeta)) = \mathbb{E}[\ln P_{\vec{q}_i}(\zeta)]. \quad (157)$$

Therefore

$$\mathbb{E}[\ln P_{\vec{q}_i}(\zeta)] = \int_0^{\infty} \rho[P_{\vec{q}_i}(\zeta)] \ln P_{\vec{q}_i}(\zeta) d\zeta = F(k) - F(N_{\text{samples}}), \quad (158)$$

where  $F(z) \equiv \int_0^{\infty} \left( \frac{e^{-t}}{t} - \frac{e^{-zt}}{1-e^{-t}} \right) dt$  is the digamma function.<sup>[214]</sup> We briefly note that the assumption of beta-distributed  $P_{\vec{q}_i}(\zeta_{i,k})$  was shown not to hold for datapoints in the vicinity of the borders of the distribution's domain by Rahvar and Ardakani, who propose a correction scheme to remedy this shortcoming in numerical applications of the kNN entropy estimator.<sup>[220]</sup> We will give more details on this shortcoming and the proposed correction scheme in Section 8.7.3.

We are now able to simplify Equation 151 by extracting the expressions containing the digamma function  $F$  from the logarithm:

$$S_{\text{spatial}} = k_B \left( -F(k) + F(N_{\text{samples}}) \right) + \frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ \ln \left( (\zeta_{i,k})^d c_d \right) \right]. \quad (159)$$

After trivial rearrangement, the canonical form of the kNN entropy estimator using the approximation introduced by Kozachenko and Leonenko<sup>[216]</sup> is obtained as

$$S_{\text{spatial}} = k_B \left( -F(k) + F(N_{\text{samples}}) + \ln(c_d) \right) + \frac{k_B d}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ \ln(\zeta_{i,k}) \right]. \quad (160)$$

Note the additional terms  $k_B (F(N_{\text{samples}}) - F(k))$ , which are corrections for the asymptotic bias of Equation 153.

Alternative expressions for the kNN entropy estimator, differing only by a constant scalar offset from Equation 160, have been proposed. This is presumably due to different derivations of the asymptotic bias-correcting constants. In his works, Hnizdo uses an expression derived earlier by himself and coworkers.<sup>[217]</sup> It is given as

$$S_{\text{spatial}}^{\text{Hnizdo}} = k_B \left[ \frac{1}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln \frac{N_{\text{samples}} \pi^{\frac{d}{2}} \zeta_{i,k}^d}{k \Gamma\left(\frac{d}{2} + 1\right)} - (L_{k-1} - \ln k - \gamma_E) \right], \quad (161)$$

with  $L_j = \sum_{i=1}^j 1/i$ , ( $L_0 \equiv 0$ ), and  $\gamma_E \approx 0.577216$ . The main difference to our derivation results from his use of the Poisson limit theorem for approximating the digamma function  $F$ . Furthermore, the distance norm for the kNN neighbor distance calculation is taken as the  $\ell_2$  norm in Equation 161. We can relate Equation 161 to Equation 160 by first recasting the digamma Function  $F(x)$  for integer arguments  $x$  as

$$F(x) = L_{x-1} - \gamma \quad \forall x \in \mathbb{N}. \quad (162)$$

In the asymptotic limit the digamma function  $F$  behaves as

$$F(x \rightarrow \infty) \sim \ln x - \frac{1}{2x} + \sum_{n=1}^{\infty} \frac{g(1-2n)}{x^{2n}} = \ln x + \mathcal{O}\left(\frac{1}{x}\right), \quad (163)$$

where  $g(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$  is the Riemann zeta function. We can now rewrite Equation 161 as

$$S_{\text{spatial}}^{\text{Hnizdo}} = k_B \left[ \frac{1}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln \frac{\pi^{\frac{d}{2}} (\zeta_{d,i,k})^d}{\Gamma\left(\frac{d}{2} + 1\right)} - F(k) + F(N_{\text{samples}} \rightarrow \infty) \right], \quad (164)$$

Furthermore, Goria *et al.* derive a similar expression in [221],

$$\begin{aligned} \hat{S}_{\text{spatial}}^{\text{Goria}} &= k_{\text{B}} \left[ \frac{1}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln \frac{(N_{\text{samples}} - 1) \pi^{\frac{d}{2}} (\zeta_{d,i,k})^d}{k \Gamma\left(\frac{d}{2} + 1\right)} - (L_{k-1} - \ln k - \gamma) \right] \\ &= k_{\text{B}} \left[ \frac{1}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln \frac{\pi^{\frac{d}{2}} (\zeta_{d,i,k})^d}{\Gamma\left(\frac{d}{2} + 1\right)} - F(k) + F((N_{\text{samples}} - 1) \rightarrow \infty) \right], \end{aligned} \quad (165)$$

whose modification is reminiscent of Bessel's correction.<sup>[222]</sup> In this present work, we will mainly use with Equation 160. From Equations 164, 165, and 160 it becomes clear that the derivations of Hnizdo *et al.* and Goria *et al.* contain a truncated Taylor Series of the asymptotic expression of the digamma function  $F$ , while the derivation presented before (modified from [214]) contains the digamma function  $F$  exactly and not as a truncated series expansion.

In summary, the kNN entropy estimator is derived by assuming a constant density in a hyperspheric / hypercubic region (depending on the Minkowski metric chosen) spanning from each sample point to its  $k^{\text{th}}$  nearest neighbor. It contains an asymptotic bias resulting from the local regions where constant probability density is assumed being beta-distributed. Different exact and approximate bias corrections have been proposed.

## 8.7. Errors of the kNN Entropy Estimator

In this section, we give a detailed overview of possible sources of error that may be present in applications of the kNN entropy estimator. We do this in order to learn about possible shortcomings and pitfalls of the kNN entropy estimator. A deeper understanding of the properties of the estimator will hopefully lead to results that render the application of the entropy estimator towards molecular data more reliable and easy.

First, we briefly mention the formal error analysis of the kNN density estimator by Fukunaga in [219], where the variance and mean square error are calculated explicitly under some assumptions. Fukunaga states that the parameter  $k$  should be chosen  $k > 2$  in order to avoid a large variance in the density estimate  $P_{q_i}(\zeta_{i,k})$ . Expressions for the calculation of an optimal  $k$  value are suggested by Fukunaga as well. For large sampling sizes, the effect of the parameter  $k$  can be found to be minimal, as the



distance to the  $k^{\text{th}}$  nearest neighbor is approximately equal to the  $(k + 1)^{\text{th}}$  nearest neighbor. Furthermore, within Fukunaga's analysis formal arguments predict that in high dimensionality, the effect of both  $k$  and the number of samples will be negligible.

### 8.7.1. Asymptotic Bias

As outlined in the derivation of the kNN entropy estimator, different variants of the estimator have been proposed (see Equations 160, 164, and 165). Formally, Equation 160 is the most accurate expression, as it does not contain the Poisson limit theorem for approximating the digamma function  $F$ . In the works of Hnizdo and coworkers, which prevail in the field of computational chemistry, a different derivation of the kNN entropy estimator is given by first deriving an asymptotically biased plug-in estimator and subsequently deriving a bias-correcting constant. Such a derivation is given in [217]. The biased estimator derived therein has the form

$$\begin{aligned}\hat{S}_{k, N_{\text{samples}}}^{\text{biased}} &= -\frac{1}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln \left[ \frac{k}{N_{\text{samples}}} \cdot \frac{F\left(\frac{d}{2} + 1\right)}{\pi^{\frac{d}{2}} \zeta_{d,i,k}} \right] \\ &= \frac{d}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln \zeta_{d,i,k} + \ln N_{\text{samples}} - \ln k + \ln \frac{\pi^{\frac{d}{2}}}{F\left(\frac{d}{2} + 1\right)}.\end{aligned}\quad (166)$$

The unbiased estimators may then be defined as

$$\hat{S}_{k, N_{\text{samples}}}^{\text{unbiased}} = \hat{S}_{k, N_{\text{samples}}}^{\text{biased}} - b(k, N_{\text{samples}}). \quad (167)$$

The mathematical expressions for the bias-correction  $b(k, N_{\text{samples}})$  corresponding to Equations 164, 165, and 160 are

$$\begin{aligned}b_{\text{Singh}}(k) &= F(k) - \ln k \\ b_{\text{Goria}}(k, N_{\text{samples}}) &= b_{\text{Singh}}(k) + \ln \left( \frac{N_{\text{samples}}}{N_{\text{samples}} - 1} \right) \\ b_{\text{exact}}(k, N_{\text{samples}}) &= b_{\text{Singh}}(k) + \ln n - F(N_{\text{samples}}).\end{aligned}\quad (168)$$

We note that in the definitions given in the set of Equations 168,  $b_{\text{Singh}}(k)$  only depends on the number of nearest neighbors  $k$ .  $b_{\text{exact}}$  and  $b_{\text{Goria}}$  are functions of  $k$  and  $N_{\text{samples}}$ . We give a plot of the biased kNN entropy estimator in Figure 15a (according to Equation 166). The plot shows, that the biased entropy estimate deviates by

a constant factor from the exact entropy value. The deviation becomes smaller if  $k$  is increased. The bias-corrected kNN entropy estimates are shown in Figure 15b. The kNN entropy estimates that have been unbiased using  $b_{\text{Singh}}(k)$ , and now indeed converge to the exact differential entropy value for all values of  $k$ . The bias-correction function  $b_{\text{Singh}}(k)$  can be seen to converge to zero for increasing values of  $k$  (see Figure 15c).  $b_{\text{Goria}}(k, N_{\text{Samples}})$  and  $b_{\text{exact}}(k, N_{\text{Samples}})$  (eq. 168) additionally depend on the number of data points  $N_{\text{Samples}}$ . For large data sets, as they occur in the analysis of MD simulations, the correction factors due to  $N_{\text{Samples}}$  are close to zero.  $\ln\left(\frac{N_{\text{Samples}}}{N_{\text{Samples}}-1}\right)$  in  $b_{\text{Goria}}(k, N_{\text{Samples}})$  evaluates to 0.01 for  $N_{\text{Samples}} = 100$ , and to 0.001 for  $N_{\text{Samples}} = 1000$ . Similarly,  $\ln(N_{\text{Samples}} - F(n))$  in  $b_{\text{exact}}(k, N_{\text{Samples}})$  evaluates to 0.005 for  $N_{\text{Samples}} = 100$ , and to 0.0005 for  $N_{\text{Samples}} = 1000$ . In Conclusion, we can derive several learnings from the analysis of the bias-corrections presented in this section: When dealing with MD simulation trajectory data sets, the following generally holds:  $N_{\text{Samples}}$  is large (at least  $> 1000$ ), but sampling in the high-dimensional conformational space of the molecular system is nevertheless sparse. For this reason,  $N_{\text{Samples}}$  should be made as large as possible for the available computational resources. In order to render the kNN entropy estimation procedure feasible, small values of  $k$  are favorable. In these cases, Figure 15 shows that one needs to correct for the asymptotic bias, since the value of  $b_{\text{Singh}}(k)$  is largest for small values of  $k$ . The biased estimator of Equation 166 should not be used. Which correction method and thus which kNN entropy estimation flavor is used is not very important, because for large values of  $N_{\text{Samples}}$  all bias-correcting functions  $b$  yield approximately similar values. Explicitly, this shows that the approximate bias-correction  $b_{\text{Singh}}(k)$  performs well and may be used equivalently to the exact bias correction term  $b_{\text{exact}}(k, N_{\text{Samples}})$ .

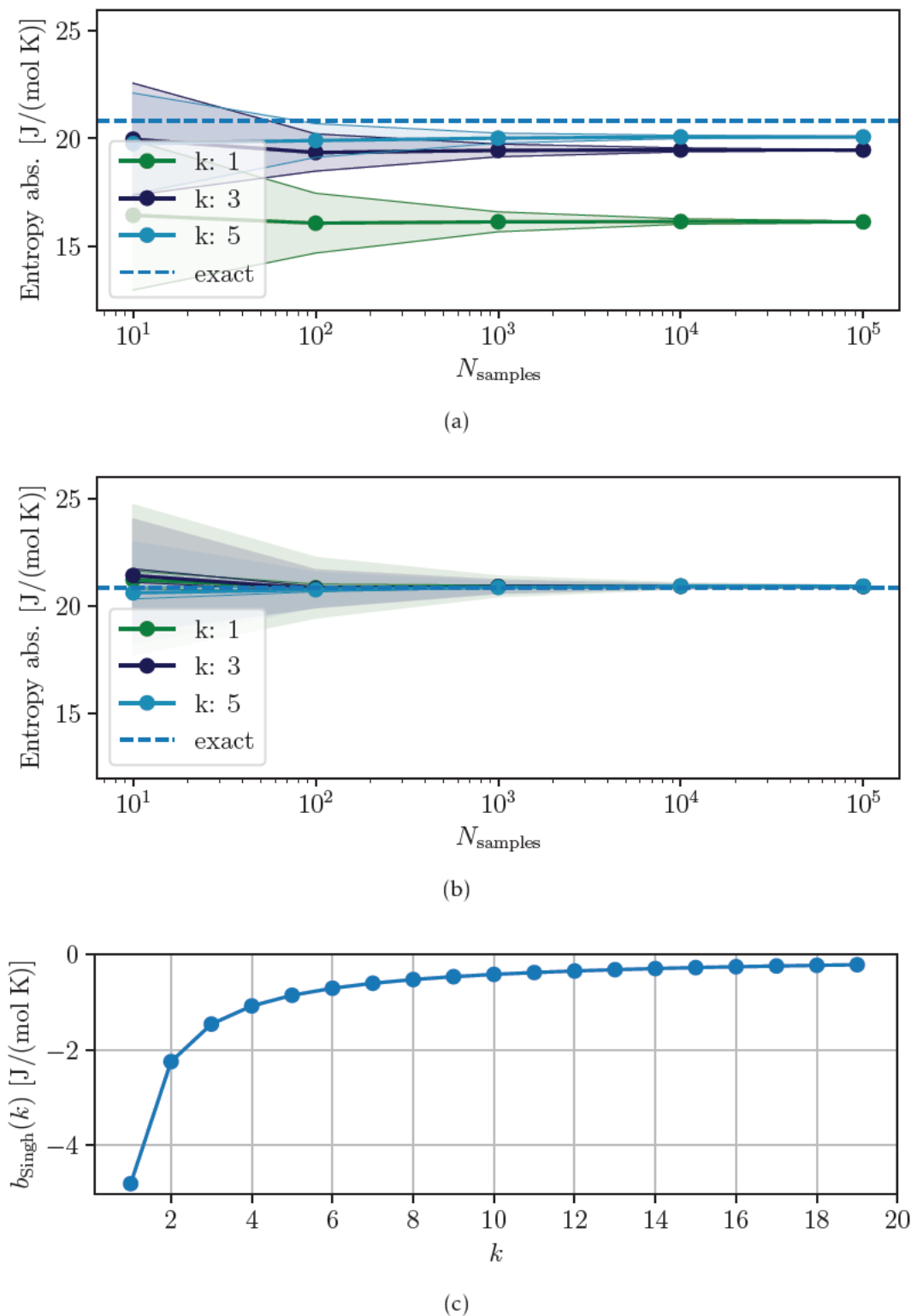


Figure 15: Numerical values of bias-corrections for the kNN entropy estimator, proposed in [217]. Random samples were drawn from a one-dimensional normal distribution with  $\sigma = 3$ . a: biased kNN entropy estimates according to Equation 166, b: unbiased kNN entropy estimates according to Equation 167, c: bias  $b_{\text{Singh}}(k)$  as a function of  $k$ , according to Equation 168. Dashed lines: reference value. Shaded area: statistical uncertainty at confidence  $1\sigma$ .

### 8.7.2. Non-Additivity of Marginal Entropies

In the following, we will derive errors of the kNN estimator (Equation 160) stemming from the assumption of constant density inside  $\ell_p(\vec{q} - \vec{q}_i) < \zeta_k$ . A special focus will be put on the manifestation of errors with increasing dimensionality.

We now assume a high-dimensional  $\rho_{\text{spatial}}(\vec{q})$  to be built from uncorrelated, one-dimensional probability density functions:

$$\rho_{\text{spatial}}(\vec{q}) = \prod_{a=1} \rho_{\text{spatial}}^{d=a}(q_a) \quad (169)$$

We want to check whether the product of the derived density estimators

$$\prod_{a=1} \hat{\rho}_{\text{spatial}}^{d=a}(q_a) \quad (170)$$

is equal to the full-dimensional estimator  $\hat{\rho}_{\text{spatial}}(\vec{q}_i)$ . To this end, we insert the estimators  $\hat{\rho}_{\text{spatial}}$  directly into Equation 135. The derivation of the asymptotic bias constant  $-F(k) + F(N)$  is skipped because the constant is independent of the dimensionality.

For the one-dimensional kNN entropy estimates, we obtain:

$$\begin{aligned} \hat{S}_{\text{kNN}}[\rho_{\text{spatial}}^{d=1}] &= -\frac{k_{\text{B}}}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln(\hat{\rho}_{\text{spatial}}(q_i^{d=1})) \\ &= k_{\text{B}} \left[ -F(k) + F(N_{\text{samples}}) + \ln(c_1) \right] + \frac{k_{\text{B}}}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ \ln(\zeta_{d=1,i,k}) \right]. \end{aligned} \quad (171)$$

For the  $d$ -dimensional estimator, we calculate:

$$\begin{aligned} \hat{S}_{\text{kNN}}[\rho_{\text{spatial}}^d] &= -\frac{k_{\text{B}}}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln(\hat{\rho}_{\text{spatial}}(\vec{q}_i)) \\ &= k_{\text{B}} \left[ -F(k) + F(N_{\text{samples}}) + \ln(c_d) \right] + \frac{k_{\text{B}} d}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ \ln(\zeta_{d,i,k}) \right] \end{aligned} \quad (172)$$

We now assert whether the property of additivity holds for the one-dimensional entropy estimators:

$$\sum_{a=1}^d \hat{S}_{\text{kNN}}[\hat{\rho}_{\text{spatial}}(q_i^{d=1})] \stackrel{!}{=} \hat{S}_{\text{kNN}}[\hat{\rho}_{\text{spatial}}(\vec{q}_i)]. \quad (173)$$

This is equivalent to asserting whether the 1<sup>st</sup> order MIE is identical to the exact full-dimensional entropy, due to all higher order MI terms being necessarily identical to zero. This holds since we have explicitly defined the multi-dimensional probability distribution as a product of only one-dimensional functions. No correlations are contained. Entropy - in both a physico-chemical and information-science context - can be seen to be strictly additive for uncorrelated probability distributions, or, phrased more physically, for non-interacting statistical subsystems. It would thus be desirable for any estimator to preserve this property of additivity, which is intrinsic to the definition of the concept of entropy itself (see [175]).

To this end, we evaluate the estimators' difference  $\Delta_{\text{kNN}}$  with respect to each other. We denote the one-dimensional nearest neighbor distance of dimension  $a$  as  $\zeta_{d=a,i,k}$  and nearest neighbor distances in the full-dimensional space as  $\zeta_{i,k}$ .

$$\begin{aligned} \Delta_{\text{kNN}} &= \sum_{a=1}^d \hat{S}_{\text{kNN}}[\rho_{\text{spatial}}^{d=a}] - \hat{S}_{\text{kNN}}[\rho_{\text{spatial}}] \\ &= \sum_{a=1}^d \left[ k_{\text{B}} (-F(k) + F(N_{\text{samples}})) + k_{\text{B}} \ln c_1 + \frac{k_{\text{B}}}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln(\zeta_{d=a,i,k}) \right] \\ &\quad - \left[ k_{\text{B}} (-F(k) + F(N_{\text{samples}})) + k_{\text{B}} \ln c_d + \frac{k_{\text{B}}}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln(\zeta_{i,k})^d \right] \\ &= \sum_{a=1}^{d-1} \left[ k_{\text{B}} (-F(k) + F(N_{\text{samples}})) \right] + \sum_{a=1}^d \left[ k_{\text{B}} \ln c_1 + \frac{k_{\text{B}}}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln(\zeta_{d=a,i,k}) \right] \\ &\quad - \left[ k_{\text{B}} \ln c_d + \frac{k_{\text{B}} d}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln(\zeta_{i,k}) \right] \end{aligned} \quad (174)$$

$$\begin{aligned}
\Delta_{\text{kNN}} &= k_B (d-1) \left( -F(k) + F(N_{\text{samples}}) \right) + \sum_{a=1}^d [k_B \ln c_1] \\
&\quad - \frac{k_B}{N_{\text{samples}}} \sum_{a=1}^d \left[ \sum_{i=1}^{N_{\text{samples}}} \ln(\zeta_{d=a,i,k}) \right] - k_B \ln c_d - \frac{k_B d}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln(\zeta_{i,k}) \quad (175) \\
&= k_B (d-1) \left( -F(k) + F(N_{\text{samples}}) \right) + k_B \ln(c_1)^d - k_B \ln c_d \\
&\quad + \frac{k_B}{N_{\text{samples}}} \sum_{a=1}^d \left[ \sum_{i=1}^{N_{\text{samples}}} \ln(\zeta_{d=a,i,k}) \right] - \frac{k_B d}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln(\zeta_{i,k})
\end{aligned}$$

We group all summation terms which do not depend on the  $d$ -dimensional  $k^{\text{th}}$  nearest neighbor distance of the point  $j$ , i.e. those that are independent of  $\zeta$ :

$$\gamma(d, k, N_{\text{samples}}) = k_B (d-1) \left( -F(k) + F(N_{\text{samples}}) \right) \quad (176)$$

$$\xi(d) = d \cdot k_B \ln c_1 - k_B \ln c_d \quad (177)$$

We note that, mathematically due to the divergence of the harmonic series,

$$\lim_{N_{\text{samples}} \rightarrow \infty} \gamma(d, k, N_{\text{samples}}) = \infty. \quad (178)$$

This is a nonsensical result, as it implies that the kNN estimator in higher dimensionality will not convergence when  $N_{\text{samples}}$  is increased. Although Equation 178 holds from a formal mathematical point of view, a more detailed discussion of the role of the bias corrections which are grouped in  $\gamma(d, k, N_{\text{samples}})$  will make the role of the term  $\gamma$  more tangible:

The result given in Equation 176 can be rationalized by rethinking the role of the term  $\left( -F(k) + F(N_{\text{samples}}) \right)$  in Equation 160. It is exactly the expectation value  $\mathbb{E} \left[ \ln P_{\hat{q}_i}(\zeta) \right]$ , which we used to extract the unknown variable  $P_{\hat{q}_i}(\zeta)$  from the logarithm in Equation 151. In this sense,  $\gamma(d, k, N_{\text{samples}})$  is not an error term, but arises due to the necessary correction of an asymptotic bias, which is present in the estimator of  $\rho_{\text{spatial}}$  used in the kNN estimator (see Equation 153 and Section 8.7.1). For the kNN entropy estimator, an asymptotic bias-correcting factor is needed in each case; once for the full-dimensional case and one per dimension for the 1<sup>st</sup> order MIE. The nearest neighbor computation introduces the neighbor distance computation functions (i.e. the  $\ell$  norms) as an so-called ordering function on the arbitrary-dimensional probability density function. An order statistic, which we will explain in more detail in the following

paragraphs, is then computable from the nearest neighbor distances. For this reason, every time nearest neighbor distances are evaluated for some (or all) dimensions of the probability density, an asymptotic bias is introduced.<sup>[223]</sup> In summary, although we can calculate  $\gamma(d, k, N_{\text{samples}})$  explicitly, we do not expect it to be present as an error in numerical application of the kNN estimator. From formal arguments it can be seen that  $\gamma(d, k, N_{\text{samples}})$  is in fact not an error, but merely accumulates grouped asymptotic-bias correcting factors, which are strictly necessary for convergence.

After some more rearrangements, a concise expression for the intrinsic dimensionality-related error of the kNN estimator of the multi-dimensional density  $\rho = \prod_{a=1}^d \rho_{\text{spatial}}^{d=a}$  is obtained:

$$\begin{aligned}
\Delta_{\text{kNN}} &= \gamma(d, k, N_{\text{samples}}) + \xi(d) + \frac{k_B}{N_{\text{samples}}} \sum_{a=1}^d \left[ \sum_{i=1}^{N_{\text{samples}}} \ln(\zeta_{d=a,i,k}) \right] - \frac{k_B d}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln(\zeta_{i,k}) \\
&= \gamma(d, k, N_{\text{samples}}) + \xi(d) + \frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ -d \ln(\zeta_{i,k}) + \sum_{a=1}^d \ln(\zeta_{d=a,i,k}) \right] \\
&= \gamma(d, k, N_{\text{samples}}) + \xi(d) + \frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ -\sum_{a=1}^d (\ln(\zeta_{i,k}) + \ln(\zeta_{d=a,i,k})) \right] \\
&= \gamma(d, k, N_{\text{samples}}) + \xi(d) - \frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \sum_{a=1}^d \ln\left(\frac{\zeta_{i,k}}{\zeta_{d=a,i,k}}\right)
\end{aligned} \tag{179}$$

From Equation 179 it becomes clear that there are two distinct sources of errors in the kNN estimator. We state again that  $\gamma(d, k, N_{\text{samples}})$  should not be seen as an error term, but as the grouping the necessary asymptotic corrections. The remaining two terms add up to the total error. Please note that these sources of error have been derived by formal arguments and are thus independent of the error due to finite sampling size, which will additionally be present in any application of the kNN estimator towards MD simulation data.

The second error term has been named  $\xi(d)$ . It is a function of the dimensionality  $d$  and depends on the Minkowski metric ( $\ell$  norm) which is used for the evaluation of the nearest neighbor distances  $\zeta$ . If the maximum norm ( $\ell_\infty$ ) is used,  $c_d = 2^d$ . Furthermore,



$\forall p \in \mathbb{N}$ :  $c_1 = 1$  in all cases. Thus,  $\xi(d)_{\ell_\infty} = 0$ . If the euclidean metric  $\ell_2$  is used, we get

$$\xi(d)_{\ell_2} = k_B \left( \ln c_d - \ln 2^d \right) = k_B \left( \ln \frac{\pi^{\frac{d}{2}}}{\Gamma\left(1 + \frac{d}{2}\right)} - \ln 2^d \right) \neq 0. \quad (180)$$

More specifically, for the generalized distance norm  $\ell_p$ , the volume of a  $d$  dimensional unit ball was given in Equation 148 as

$$c_d^{\ell_p} = \frac{\left(2\Gamma\left(\frac{1}{p} + 1\right)\right)^d}{\Gamma\left(\frac{d}{p} + 1\right)} \quad (181)$$

and therefore

$$\begin{aligned} \xi(d)_{\ell_p} &= k_B \left( \ln c_d^{\ell_p} - \ln 2^d \right) = k_B \left( \ln \frac{\left(2\Gamma\left(\frac{1}{p} + 1\right)\right)^d}{\Gamma\left(\frac{d}{p} + 1\right)} - \ln 2^d \right) \\ &= k_B \ln \frac{\left(\Gamma\left(\frac{1}{p} + 1\right)\right)^d}{\Gamma\left(\frac{d}{p} + 1\right)}. \end{aligned} \quad (182)$$

The term  $\xi(d)_{\ell_p}$  of Equation 182 vanishes for the  $\ell_\infty$  norm. This can be rationalized by noticing that the unit ball in the  $\ell_\infty$  is a hypercube, and the volume of a hypercube is the result of the product of its one-dimensional edge lengths. The edge-lengths correspond to the “volume” of a unit ball for  $d = 1$ . Therefore, it holds for the  $\ell_\infty$  norm only that

$$c_d^{\ell_\infty} = 2^d = \prod_{a=1}^d c_1^{\ell_\infty}. \quad (183)$$

It can be seen that same does not hold for other norms, such as the euclidean  $\ell_2$  norm, since the volume of a hypersphere is not identical to its diameter raised to the  $d^{\text{th}}$  power.

Lastly, there is a contribution to the total error resulting from the  $k^{\text{th}}$  nearest neighbor distances  $\zeta$  themselves. To state this more clearly, the error results from the ratio of the full-dimensional ( $d$ -dimensional) nearest neighbor distance  $\zeta_{i,k}$  raised to the  $d^{\text{th}}$  power and the product of the one-dimensional nearest neighbor distances  $\zeta_{d=a,i,k}$  of

each dimension:

$$\Delta_{\text{kNN}} - \gamma(d, k, N_{\text{samples}}) - \xi(d) = -\frac{k_{\text{B}}}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \sum_{a=1}^d \ln \left( \frac{\zeta_{i,k}}{\zeta_{d=a,i,k}} \right) \quad (184)$$

The expression found in Equation 184 follows from rearrangement of the last term in Equation 179. At this point, it will be helpful to give an analogy for this expression: It is a ratio of two  $d$ -dimensional volumes. If the volumes are identical, the contribution vanishes and no error from the nearest neighbor distances, and thus from the sampled data itself, is made. This is the case if and only if

$$\left( \zeta_{i,k} \right)^d = \prod_{a=1}^d \zeta_{d=a,i,k} \quad (185)$$

holds.

While  $\prod_{a=1}^d \zeta_{d=a,i,k}$  is a hyperrectangle, only becoming a hypercube when all individual neighbor distances  $\zeta_{d=a,i,k}$  are identical,  $\left( \zeta_{i,k} \right)^d$  is by definition a hypercube with fixed edge length. Obviously, Equation 185 does not hold in all cases. More so, it may only be approximately true if all individual domains have a similar magnitude and thus similar variance. Krasov *et al.* give geometrical arguments for this finding in [224] and conclude that the kNN entropy estimator is not strictly additive for this reason.

To render things clearer we give an explicit example: Let us assume a two-dimensional model where  $\rho_{\text{spatial}}^{d=1}$  and  $\rho_{\text{spatial}}^{d=2}$  are uniform distributions, each defined on the domain  $D = [0, 1]$ . The higher-dimensional  $\rho$  built from the uncorrelated marginal probability densities is then  $\rho_{\text{spatial}} = \rho_{\text{spatial}}^{d=1} \cdot \rho_{\text{spatial}}^{d=2}$ . In the case of a uniform distribution, we may assume all one-dimensional nearest neighbor distances for  $k = 1$  to be equal, having a value of  $\zeta_{d,i,k} = \frac{1}{N_{\text{samples}}-1} \forall d$ . In this case, Equation 185 holds if the  $\ell_{p=\infty}$  metric is used. Then, no error due to the nearest neighbor distances is made.

Now let us assume the same situation, but with  $\rho_{\text{spatial}}^{d=2}$  defined on the domain  $D' = [0, 10]$ . Now,  $\zeta_k^{j,d=1} = \frac{1}{N_{\text{samples}}}$ , but  $\zeta_{d=\{1,2\},i,k} = \frac{10}{N_{\text{samples}}-1}$ . Now,  $\prod_{a=1}^d \zeta_{d=a,i,k} = \frac{10}{(N_{\text{samples}}-1)^2}$  be-

comes a hyperrectangle. The value of  $(\zeta_{i,k})^d$  now depends on the  $\ell_p$  metric:

$$\begin{aligned}
 p = \infty: \quad (\zeta_{i,k})^d &= \left( \frac{10}{N_{\text{samples}} - 1} \right)^2 \\
 p = 2: \quad (\zeta_{i,k})^d &= \left( \sqrt{\left( \frac{10}{N_{\text{samples}} - 1} \right)^2 + \left( \frac{1}{N_{\text{samples}} - 1} \right)^2} \right)^2
 \end{aligned} \tag{186}$$

In any case, Equation 185 does not hold and an error according to Equation 184 results. In the case of identical one-dimensional probability density functions  $\rho$ , which have only been scaled to encompass different domains, we are able to remedy this by rescaling the sample points to equal variance prior to the kNN entropy estimation. In accordance with Equation 110, we are able to remove the effect of this scaling by evaluating the shifting constant analytically from the scaling factors. In this manner, the kNN estimator's intrinsic error is avoided and the differential entropy, whose absolute value does depend on the size of the domain, can be recovered. A similar procedure might also be able to remedy the intrinsic error of the kNN estimator for chemical, molecular simulation data. In this case however, modes will not only differ in scale but also in their shape. Correlation might also be present. Thus, the scaling to equal variance should then only be seen as an approximate scheme to render the application of the kNN estimator feasible at low intrinsic errors.

We finish this section on the formal discussion of errors of the kNN estimator by giving some example values of the contributions of  $\gamma$  and  $\xi$ , which can be evaluated directly without the need for empirical samples drawn from a distribution. The magnitude of the contribution  $\gamma(d, k = 1, N_{\text{samples}})$  again showcases the critical need for asymptotic bias corrections in kNN entropy estimation procedures, as previously highlighted in Figure 15.

Table 3: Numerical results for the contribution  $\gamma(d, k = 1, N_{\text{samples}})$ , units:  $\frac{\text{J}}{\text{molK}}$

$N_{\text{samples}}$	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
50	0.00	37.24	74.48	111.73	148.97
100	0.00	43.05	86.09	129.14	172.19
1000	0.00	62.23	124.46	186.69	248.92
5000	0.00	75.61	151.23	226.84	302.46
10000	0.00	81.38	162.76	244.13	325.51

Table 4: Numerical results for the error  $\xi(d)_{\ell_p} = k_B \ln \frac{\Gamma(\frac{1}{p}+1)^d}{\Gamma(\frac{d}{p}+1)}$ , units:  $\frac{\text{J}}{\text{molK}}$ 

$p$	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
1	0.00	-5.76	-14.90	-26.42	-39.81
2	0.00	-2.01	-5.38	-9.78	-15.01
3	0.00	-1.03	-2.82	-5.22	-8.10
4	0.00	-0.63	-1.75	-3.27	-5.12
5	0.00	-0.43	-1.19	-2.25	-3.55
10000 $\approx \infty$	0.00	0.00	0.00	0.00	0.00

As the nearest neighbor distances  $\zeta_{d=1,i,k}$  are directly dependent on the sampling procedure, we cannot give a general estimate of the error induced by the term

$$-\frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \sum_{a=1}^d \ln \left( \frac{\zeta_{i,k}}{\zeta_{d=a,i,k}} \right). \quad (187)$$

However, if we assume every  $\rho_{\text{spatial}}^d$  to be a uniform distribution on a bounded domain  $D_d$  in  $\mathbb{R}$ , where the spatial extend of the domain

$$\max(D_d) - \min(D_d) \quad (188)$$

is related via

$$\max(D_d) - \min(D_d) = b \cdot [\max(D_{d+1}) - \min(D_{d+1})], \quad (189)$$

the one-dimensional neighbor distances can be given explicitly.  $b$  is a constant scalar that describes the scaling of the dimensions with respect to each other. A one-dimensional  $k^{\text{th}}$  nearest neighbor distance may reasonably be assumed to be equidistant and expressed by

$$\zeta_{d,i,k} = \frac{[\max(D_1) - \min(D_1)] \cdot k \cdot b^{(d-1)}}{N_{\text{samples}} - 1} \quad \forall i. \quad (190)$$

Therefore, assuming uniform probability densities  $\rho_{\text{spatial}}^d(\vec{q}_i) = \frac{1}{\max(D_d) - \min(D_d)}$ , we can

calculate  $-\frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \sum_{a=1}^d \ln\left(\frac{\zeta_{i,k}}{\zeta_{d=a,i,k}}\right)$  explicitly as

$$\begin{aligned}
& -\frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \sum_{a=1}^d \ln\left(\frac{\zeta_{i,k}}{\frac{[\max(D_1) - \min(D_1)] \cdot k \cdot b^{(d-1)}}{N_{\text{samples}}^{-1}}}\right) = \\
& -\frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \sum_{a=1}^d \ln\left(\frac{\left\{\sum_{b=1}^d \left|\frac{[\max(D_1) - \min(D_1)] \cdot k \cdot b^{(b-1)}}{N_{\text{samples}}^{-1}}\right|^p\right\}^{\frac{1}{p}}}{\frac{[\max(D_1) - \min(D_1)] \cdot k \cdot b^{(a-1)}}{N_{\text{samples}}^{-1}}}\right) = \quad (191) \\
& -k_B \sum_{a=1}^d \ln\left(\frac{\left\{\sum_{b=1}^d \left|[\max(D_1) - \min(D_1)] \cdot b^{(b-1)}\right|^p\right\}^{\frac{1}{p}}}{[\max(D_1) - \min(D_1)] \cdot b^{(a-1)}}\right).
\end{aligned}$$

Equation 191 may be evaluated numerically for given values of  $d$ ,  $p$ , and  $b$ .

Table 5: Numerical results for the error term  $-\frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \sum_{a=1}^d \ln\left(\frac{\zeta_{i,k}}{\zeta_{d=a,i,k}}\right)$ , units:  $\frac{\text{J}}{\text{molK}}$ ,  $b = 1$

$p$	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
1	0.00	-11.53	-27.40	-46.11	-66.91
2	0.00	-5.76	-13.70	-23.05	-33.45
3	0.00	-3.84	-9.13	-15.37	-22.30
4	0.00	-2.88	-6.85	-11.53	-16.73
5	0.00	-2.31	-5.48	-9.22	-13.38
10000 $\approx \infty$	0.00	0.00	0.00	0.00	-0.01

Table 6: Numerical results for the error term  $-\frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \sum_{a=1}^d \ln\left(\frac{\zeta_{i,k}}{\zeta_{d=a,i,k}}\right)$ , units:  $\frac{\text{J}}{\text{molK}}$ ,  $b = 0.5$

$p$	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
1	0.00	-12.51	-31.25	-55.49	-85.13
2	0.00	-7.62	-20.68	-39.30	-63.59
3	0.00	-6.42	-18.38	-36.06	-59.48
4	0.00	-6.02	-17.69	-35.12	-58.30
5	0.00	-5.87	-17.45	-34.79	-57.90
10000 $\approx \infty$	0.00	-5.76	-17.29	-34.58	-57.63

At this point, we are able to sum up the error terms for given set of parameters  $d$ ,  $p$ ,  $k$ , under the assumption of uncorrelated one-dimensional uniform probability densities

$$\rho_{\text{spatial}}(\vec{q}) = \prod_{a=1}^d \rho_{\text{spatial}}^{d=a}(q_a) = \prod_{a=1}^d \frac{1}{\max(D_a) - \min(D_a)} \quad (192)$$

with their bounded domains  $D_d$  in  $\mathbb{R}$  again constrained by

$$\max(D_d) - \min(D_d) = b \cdot [\max(D_{d+1}) - \min(D_{d+1})]. \quad (193)$$

Table 7: Numerical results for the error term  $\Delta_{kNN} - \gamma(d, k=1, N_{\text{samples}})$ , units:  $\frac{\text{J}}{\text{molK}}$ ,  $b = 1.0$ ,  $k = 1$

$p$	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
1	0.00	-17.29	-42.30	-72.53	-106.71
2	0.00	-7.77	-19.08	-32.83	-48.46
3	0.00	-4.87	-11.96	-20.58	-30.40
4	0.00	-3.51	-8.60	-14.79	-21.85
5	0.00	-2.73	-6.67	-11.47	-16.93
$10000 \approx \infty$	0.00	0.00	0.00	0.00	0.00

Table 8: Numerical results for the error term  $\Delta_{kNN} - \gamma(d, k=1, N_{\text{samples}})$ , units:  $\frac{\text{J}}{\text{molK}}$ ,  $b = 0.5$ ,  $k = 1$

$p$	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
1	0.00	-18.27	-46.15	-81.91	-124.93
2	0.00	-9.63	-26.06	-49.08	-78.60
3	0.00	-7.45	-21.21	-41.27	-67.58
4	0.00	-6.65	-19.44	-38.38	-63.43
5	0.00	-6.29	-18.64	-37.04	-61.44
$10000 \approx \infty$	0.00	-5.76	-17.29	-34.58	-57.63

Table 7 shows the error for a multidimensional uniform distribution with equal domains. In line with theoretical arguments, the error due to  $\Delta_{kNN} - \gamma(d, k, N_{\text{samples}})$  can be seen to vanish for  $p \rightarrow \infty$ . If the domains of the individual dimensions are uneven (see Table 8 for  $b = \frac{1}{2}$ ), the error does not vanish even for  $p \rightarrow \infty$  and can have substantial magnitude. In Table 8 the absolute error decreases with increasing  $p$ . From these results, it can be stated that the 1<sup>st</sup> order MIE of the kNN entropy should always be performed using the  $\ell_\infty$  norm in order to minimize errors. From the arguments and tables presented above, it seems that although the kNN entropy estimator is suffering from different sources of errors, it might be able to yield the entropy of a one-dimensional uniform probability distribution exactly, since in this case the assumption of uniform density within the region  $[q_i - \zeta, q_i + \zeta]$  is trivially true. Furthermore, the nearest neighbor distances  $\zeta$  can be evaluated analytically in this case. Nevertheless, we will show that even then the kNN entropy estimator remains approximative and

cannot yield the correct differential entropy.

Assume a uniform probability density  $\rho$  with a domain  $D = [0, 1]$ . From [225] it is known that the analytic differential entropy of this distribution is calculated solely from the boundaries of the domain as

$$\hat{S}[\rho] = k_B \ln \frac{1}{1-0} = 0. \quad (194)$$

Now, due to the uniformity of the density, we may again reasonably assume that samples drawn are spaced in an equidistant manner. Then, for a set of  $N_{\text{samples}}$  drawn sample points, their individual values are taken as

$$N_i = \frac{i}{N_{\text{samples}} - 1} \quad (195)$$

and their nearest neighbor distances  $\zeta_{i,k}$  are

$$\zeta_{i,k} = \frac{k}{N_{\text{samples}} - 1}. \quad (196)$$

Now, inserting this into Equation 160 yields

$$S_{\text{kNN}} = k_B \left[ -F(k) + F(N_{\text{samples}}) + \ln(c_d) \right] + \frac{k_B d}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ \ln(\zeta_{i,k}) \right]. \quad (197)$$

Without loss of generality, we take  $k = 1$  and, as we are analyzing the one-dimensional case,  $d = 1$  and thus  $c_d = 2$  independent of the Minkowski metric:

$$S_{\text{kNN}} = k_B \left[ F(N_{\text{samples}}) + \ln 2 + \ln \frac{1}{N_{\text{samples}} - 1} \right] = S_{\text{kNN}}(N_{\text{samples}}). \quad (198)$$

We take the limit of Equation 198 for  $N_{\text{samples}} \rightarrow \infty$  and obtain

$$\lim_{N_{\text{samples}} \rightarrow \infty} S_{\text{kNN}} = k_B \lim_{N_{\text{samples}} \rightarrow \infty} \left( F(N_{\text{samples}}) + \ln 2 + \ln \frac{1}{N_{\text{samples}} - 1} \right) \approx k_B \ln 2 \neq 0. \quad (199)$$

In Equation 199 we have used the series expansion of the digamma function  $F$  and retained only the leading term, as previously discussed with regards to Equation 163. Equation 199 shows that the kNN entropy estimate of the uniform one-dimensional probability density is not identical to the analytic result (compare with Equation



194), not even in the limit of infinite samples. The reason for this is the previously laid out approximation of taking  $P_{q_i}(\zeta_{i,k})$  to be distributed according to a beta distribution. In case of a one-dimensional uniform distribution,  $P_{q_i}(\zeta_{i,k})$  is actually a delta-distribution for each given point  $q_i$ . We show this explicitly by first rewriting Equation 197 without extracting  $P_{q_i}(\zeta_{i,k})$  from the summation (and thereby assuming it to be beta distributed), using Equation 150:

$$\begin{aligned} S_{\text{kNN}} &= -\frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ \ln \left( \frac{P_{\vec{q}_i}(\zeta)}{(\zeta_{i,k})^d * c_d} \right) \right] \\ &= k_B \ln(c_d) - \frac{k_B \cdot d}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ \ln \left( \frac{P_{\vec{q}_i}(\zeta)}{\zeta_{i,k}} \right) \right]. \end{aligned} \quad (200)$$

For a uniform distribution, we may calculate  $P_{\vec{q}_i}(\zeta)$  exactly. Its value is identical for all points  $\vec{q}_i$ , because they are spaced equidistantly. For a sample size of  $N_{\text{samples}}$ , it holds that

$$P_{\text{uniform}} \left( \zeta = \frac{1}{N_{\text{samples}} - 1} \right) = \frac{2}{N_{\text{samples}} - 1}. \quad (201)$$

If we insert Equation 201 in Equation 200, we obtain

$$\begin{aligned} S_{\text{kNN}}^+ &= k_B \ln(c_d) - \frac{k_B \cdot d}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \left[ \ln \left( \frac{\frac{2}{N_{\text{samples}} - 1}}{\frac{1}{N_{\text{samples}} - 1}} \right) \right] \\ &= k_B \ln(2) - \frac{k_B}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \ln(2) = 0. \end{aligned} \quad (202)$$

Thus, the error in Equation 199 stems from assuming  $P_{\vec{q}_i}(\zeta)$  to be beta-distributed. For numerical applications towards MD simulation data. One however cannot operate the kNN entropy estimator without this assumption, as  $P_{\vec{q}_i}(\zeta)$  cannot be calculated exactly, since in practical applications  $\rho_{\text{spatial}}(\vec{q}_i)$  is unknown.

Although this preceding discussion of errors is important to understand the empirical behavior of the kNN estimator, the errors given in Table 8 are in a sense hermetic and not applicable to MD simulation data. Although the conformational and vibrational modes from chemical simulation data might be distributed according to a normal distribution, they are decisively not uniformly distributed. Thus, the errors observed in empirical estimations of differential entropies are of a different nature and can hardly be captured with analytical mathematical derivations.

### 8.7.3. Errors at Domain Boundaries

Rahvar and Ardakani have given a concise discussion of the error of the kNN entropy estimator at the domain boundaries of the underlying probability distribution  $\rho$  in [220]. Specifically, they assert that the failure of the kNN entropy estimator stems from the assumption of constant density in the region given by  $P_{\vec{q}_i}(\zeta_{i,k}) = \int_{\ell_p(\vec{q}-\vec{q}_i) < \zeta_{i,k}} \rho^{\text{spatial}}(\vec{q}) d\vec{q}$  (see Equation 144). The assumption is not valid at the domain boundaries, as the size of the region is overestimated by the estimator there. This is the case whenever the volume  $V = \frac{k}{(\zeta_{i,k})^d c_d N_{\text{samples}}}$  inside of which the density  $\rho$  is assumed to be constant extends farther than the domain of  $\rho$ . This means that the assumption is invalid for those sample points that lie close to the boundary of the domain. This in turn leads to  $P_{\vec{q}_i}(\zeta_{i,k})$  not being distributed according to a beta-distribution, even in the limit of infinite sample points. Thus, Equation 154 is not valid at the boundary. Rahvar and Ardakani have developed a correction scheme for the supremum-norm case ( $\ell_\infty$ ). We will state a novel although straight-forward extension of their correction scheme to the euclidean ( $\ell_2$ ) norm as well. The correction scheme operates by assuming that the minimum/maximum value found in the set of sample points for each individual dimension is identical to the domain boundary of that dimension. For every  $\vec{q}_i$  from the set of drawn sample points  $Q$  which is considered in the kNN entropy estimator, we first check whether it is close to the boundary by asserting if the kNN distance  $\zeta_{i,k}^p$  is larger than the distance to the outer-most sample points in each dimension. To check when to correction needs to be applied, we define an auxiliary function  $\Omega$ :

$$\Omega(\vec{q}_i) = \sum_{a=1}^d \begin{cases} 1 & \text{if } \zeta_{i,k}^p > |\sup \{(\vec{q}_j)_a - (\vec{q}_i)_a \mid \vec{q}_j \in Q\}| \\ 1 & \text{if } \zeta_{i,k}^p > |\inf \{(\vec{q}_j)_a - (\vec{q}_i)_a \mid \vec{q}_j \in Q\}| \\ 0 & \text{else} \end{cases} \quad (203)$$

Rahvar and Ardakani's correction scheme operates by assuming the minimum/maximum value found in the set of sample points for each individual dimension is identical to the domain boundary of that dimension. For every sample point  $\vec{q}_i$  which is considered in the kNN entropy estimator, we first check whether it is close to the boundary by asserting if the kNN distance  $\zeta_{i,k}$  in the chosen  $\ell_p$  norm is larger than the distance to the outer-most sample points in each dimension ( $q_{\min}^a = \inf \{(\vec{q}_j)_a \mid \vec{q}_j \in Q\}$ ,  $q_{\max}^a = \sup \{(\vec{q}_j)_a \mid \vec{q}_j \in Q\}$ ). This is the case if  $\Omega(\vec{q}_i) > 0$  in Equation 203. If this is the case,  $\zeta_{i,k}$  associated with point  $\vec{q}_i$  is discarded and substituted with a newly calculated,

corrected nearest neighbor distance. In order to compute this corrected nearest neighbor distance, we define a new distance  $r_{i,k,a}$  per dimension  $a$  of the multidimensional domain:

$$r_{i,k,d=a} = \min \left\{ (\vec{q}_i)_a + \frac{\zeta_{i,k}}{2}, \sup \left\{ (\vec{q}_j)_a \mid \vec{q}_j \in Q \right\} \right\} \\ - \max \left\{ (\vec{q}_i)_a - \frac{\zeta_{i,k}}{2}, \inf \left\{ (\vec{q}_j)_a \mid \vec{q}_j \in Q \right\} \right\}. \quad (204)$$

Note that

$$r_{i,k,d=a} = \zeta_{i,k} \quad (205)$$

if

$$\zeta_{i,k} < |(\vec{q}_i)_a - (\vec{q}_j)_a| \quad \forall \vec{q}_j \in Q, \quad (206)$$

i.e. when no  $\zeta_{i,k}$  extends beyond the assumed domain boundary for the chosen point  $\vec{q}_i$ . If  $\Omega(\vec{q}_i) > 0$ , we are to apply the correction scheme to the point  $\vec{q}_i$  by substituting its nearest neighbor distance  $\zeta_{i,k}$  with  $\zeta_{i,k}^{\text{corrected}}$ . This is akin to modifying the volume inside of which the kNN entropy estimator locally assumes the probability density to be constant. This corrected volume is obtained as

$$V_{\text{corrected}} = \left( \sum_{a=1}^d r_{i,k,a} \right) * c_d. \quad (207)$$

This turns the original isotropic geometrical volumes (hypersphere or hypercube) into their anisotropic counterparts ( $\ell_2$ : hyperellipsoid,  $\ell_\infty$ : hyperrectangle). Inside of this volume (which defines a region in the multidimensional space), we may more reasonably assume the probability density to be constant during kNN entropy estimation, as regions beyond the assumed domain boundary are now excluded. In the last step of the correction scheme, the edge-length or radius of an equivalent isotropic geometrical figure with identical volume (hypercube / hypersphere) is calculated from  $V_{\text{corrected}}$ . This edge-length or radius is equivalent to the final, corrected nearest neighbor distance  $\zeta_{i,k}^{\text{corrected}}$ :

$$\frac{V_{\text{corrected}}^{\frac{1}{d}}}{c_d} = \left( \sum_{a=1}^d r_{i,k,a} \right)^{\frac{1}{d}} = \zeta_{i,k}^{\text{corrected}}. \quad (208)$$

We give a graphical example of the geometric idea of the correction scheme in two dimensions in Figure 16. In this sketch we take  $N_{\text{samples}} = 8$  and  $k = 2$ . We consider the sample point  $\vec{q}_i$  highlighted in green color. Its associated nearest neighbor distance  $\zeta_{i,k=2}$  in the  $\ell_\infty$  norm is shown in yellow color. The boundary of the domain of the probability distribution, as estimated from the sample points, is shown as a black rectangle. The region inside of which the uncorrected kNN entropy estimator would assume constant probability density for the point  $\vec{q}_i$  is shown as a square with dashed lines. It is clear that the subregion of this square colored in blue is outside the domain of the probability density function. Instead of assuming constant probability density there, the best guess for the value of  $\rho$  in that region is zero. The correction scheme proceeds by determining the outermost sample points in each dimension. Then, the region where constant density is assumed is truncated accordingly, resulting in the gray rectangular region. As the kNN entropy estimator works on isotropic volumes, the red square with identical volume to the gray rectangle is built.  $\zeta_{i,k}^{\text{corrected}}$  is identical to the purple dashed line, with magnitude of half the edge length of the region given by the red square. For all points  $\vec{q}_i$  where  $\Omega(\vec{q}_i) > 0$ ,  $\zeta_{i,k}^{\text{corrected}}$  is different from the original

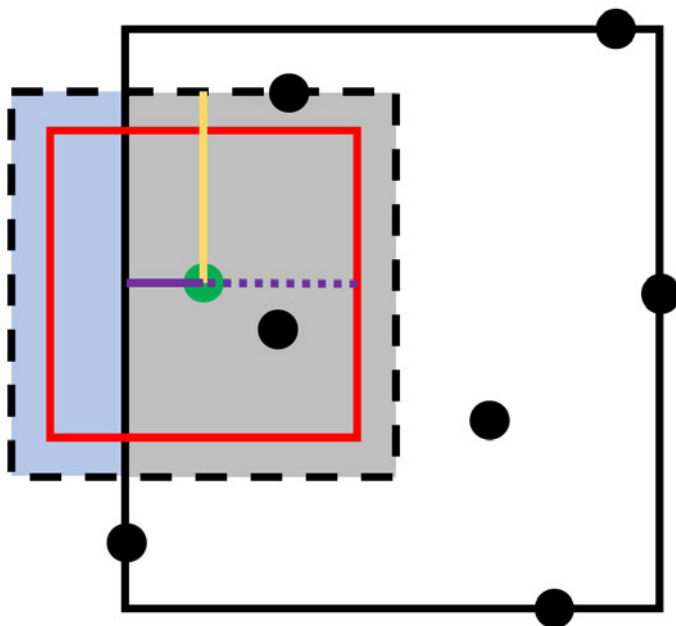


Figure 16: Exemplification of the error of the kNN entropy estimate for two dimensions for sample points at the domain boundary. The  $\ell_\infty$  norm is used in combination with  $N_{\text{samples}} = 8$  and  $k = 2$ . The region of assumed uniform density extends beyond the domain boundary for the sample point highlighted in green in the regular kNN entropy estimation approach. For detailed explanations, see associated paragraph.

nearest neighbor distance  $\zeta_{i,k}$ .  $\zeta_{i,k}^{\text{corrected}}$  can be inserted into the kNN entropy estimator (Equation 160) without further modifications of its mathematical framework. For this reason, the computational implementation of Rahvar and Ardakani's correction scheme is straight forward. While Rahvar and Ardakani have given their derivation of the correction only for the  $\ell_\infty$  norm, the more abstract derivation in this text naturally generalized to all  $\ell_p$  metrics. The impact of the correction scheme will be assessed in more detail in the benchmark presented in Section 8.7.4.

#### 8.7.4. Errors due to Different Domain Ranges of Individual Dimensions

Another separate source of error of the kNN entropy estimator is its application towards multidimensional probability densities where the individual dimensions are uncorrelated but possess different variances or domains. To the best of this author's knowledge, this source of error has not been previously discussed in scientific literature. Conceptually, it is related to the error at domain boundaries discussed in Section 8.7.3. In this section, we will show that the error derived from the uneven variances of individual dimensions might well be the most significant source of error. We propose that this error is the source of the often observed slow convergence of the kNN entropy estimator with respect to  $N_{\text{samples}}$  for MD simulation data.

If the high-dimensional probability density is built from uncorrelated one-dimensional probabilities, the 1<sup>st</sup> order MIE yields the exact entropy of the distribution. The full-dimensional kNN estimator in its uncorrected formulation (Equation 160) operates by assuming a constant local probability density in a hypersphere / hypercube, depending on the  $\ell$  distance norm. If some dimensions have significantly smaller extends (domains or variances) than others, the size of the hypersphere is dominated by those dimensions with a large variance due to the nature of the distance norm. It is then possible that the extend of the hypersphere, inside of which the probability density is assumed constant for all dimensions, extends the range (domain or variance) of the smaller dimensions.

By constructing a model probability density we will now showcase the occurrence and magnitude of the errors introduced by applying the full-dimensional kNN entropy estimator to distributions with uneven variances in its dimensions. The probability density used is a multivariate gaussian with no covariance. First, we consider the case of exponentially decreasing standard deviation. The standard deviations (i.e. the square-roots of the diagonal entries of the multivariate gaussian's covariance matrix)

of each individual dimension  $d_a$  are calculated as

$$\sigma_{d_a}^{\text{exp.-decr.}} = 3 e^{\left(\frac{a-1}{d_{\text{max}}-1}\right)}, \quad (209)$$

where  $d_a = 1$  is the first dimension and the full number of dimensions is  $d_{\text{max}}$ . The exact differential entropy for this test-system can be evaluated by summing the entropic values calculated for the analytic entropy expressions of a single gaussian function of each dimension (see Equation 129). We have drawn samples from the multidimensional distribution and evaluated the kNN entropy from these samples. To this end, the uncorrected estimator (Equation 160) and the correction scheme according to Rahvar and Ardakani (see Section 8.7.3) was used.<sup>[220]</sup> Additionally, we have used the uncorrected estimator on modified data-points which have been rescaled to have identical variances. In accordance with Equation 110, this leads to a constant shift in the absolute differential entropy, which can then be accounted for *a posteriori*. To rephrase, in this newly proposed correction scheme to remedy the error derived from uneven dimensional variances, the data is first rescaled to identical variance in all dimensions. The constant shift introduced by this rescaling is then subtracted as a last step from the kNN entropy estimate to restore the proper magnitude of the entropy relating to the original data. For the rescaling procedure two different approaches were used: In the case of the multivariate-gaussian benchmark probability density, the exact variances of each dimension are known and may be used for rescaling the randomly drawn sample points. Alternatively, the variances can be estimated from the samples themselves, which would be necessary in chemical entropy estimation from MD simulation data, as in this case the exact probability density  $\rho$  is unknown. In order to account for the stochastic nature of the benchmark and the fluctuations in the data, we have repeated the benchmark ten times and plotted the mean entropy values along with the standard deviation of the estimate across ten trial runs for the benchmarks. This means that samples were drawn and entropies were evaluated ten times for each test case and its associated probability density. To summarize, we have performed kNN entropy estimation on the multivariate gaussian probability density

$$\rho^{\text{exp.-decr.}}(\vec{q}) = \prod_{a=1}^d \frac{1}{\sigma_a^{\text{exp.-decr.}} \sqrt{2\pi}} e^{\left(-\frac{1}{2} \frac{((\vec{q})_a)^2}{(\sigma_a^{\text{exp.-decr.}})^2}\right)}. \quad (210)$$

The following entropies have been calculated for the probability density  $\rho^{\text{exp.-decr.}}(\vec{q})$ :

1. Analytic differential entropy.
2. Uncorrected kNN entropy estimate according to Equation 160 using the  $\ell_2$  norm.
3. kNN entropy estimate according to Equation 160 using the  $\ell_2$  norm and including the correction scheme by Rahvar and Ardakani from [220] (see Section 8.7.3).
4. kNN entropy estimate according to Equation 160 using the  $\ell_2$  norm with rescaling procedure using analytic variances  $\left(\sigma_a^{\text{exp.-decr.}}\right)^2$ .
5. kNN entropy estimate according to Equation 160 using the  $\ell_2$  norm with rescaling procedure variances computed from the samples drawn.

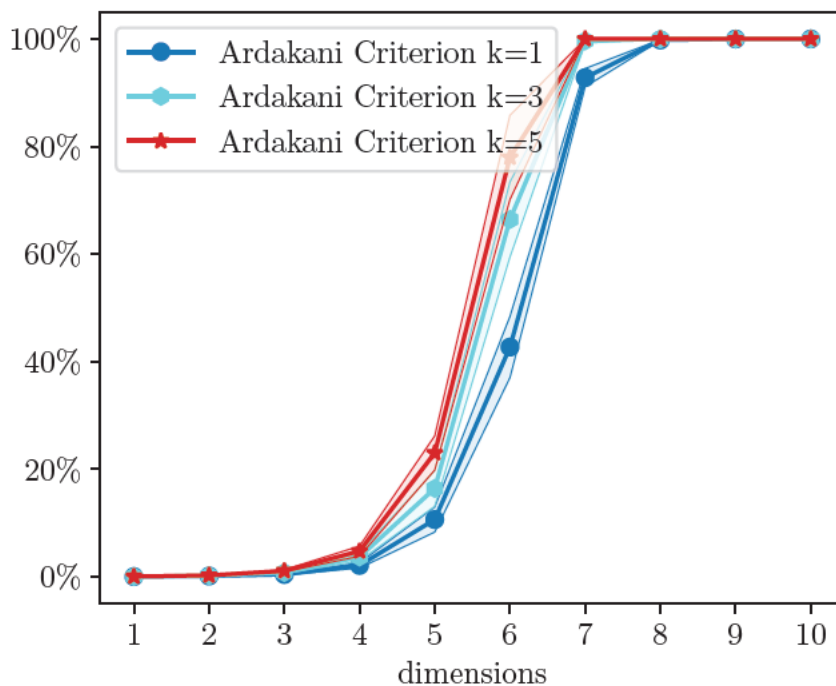


Figure 17: Percentage of points at the domain boundary for a multivariate gaussian with exponentially decreasing standard deviations (see Equation 209) for  $N_{\text{samples}} = 10000$ . Samples have been drawn ten times, and the confidence interval of  $1\sigma$  is given in the plot via the transparently colored area around the mean value of all ten trials. Points are determined to lie at the domain boundary when  $\Omega(\vec{q}_i) \neq 0$  (see Equation 203).

First, it will be interesting to see how many of the sample points drawn from  $\rho^{\text{exp.-decr.}}(\vec{q})$  are determined to lie on the boundary of the domain. For these points,  $\Omega(\vec{q}_i) \neq 0$  (see



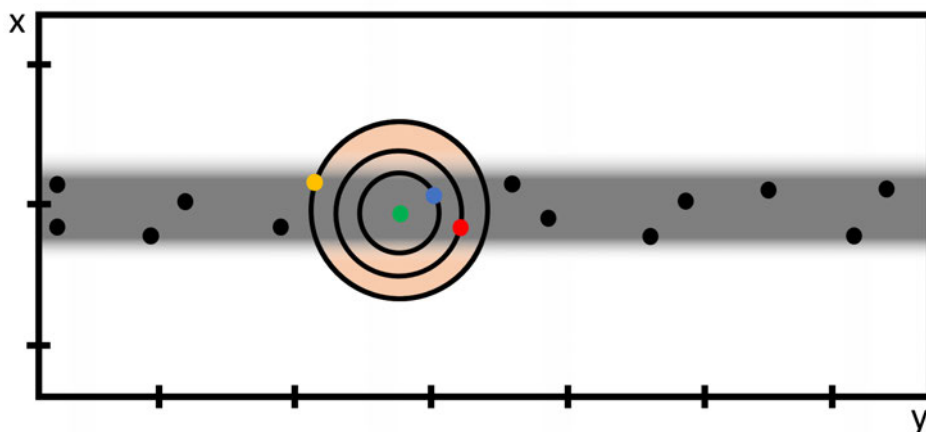


Figure 18: This figure shows the overestimation of the kNN entropy estimate for two dimensions with uneven variances. The  $\ell_2$  norm is used. The radius of the circular region where constant density is assumed may be larger than the extend of any single dimension's domain. The region where constant probability density is assumed in the kNN entropy estimator is overestimated. It extends beyond the domain boundary of the variable  $x$ . Some error in the kNN entropy estimate is then due to the overestimation of the region where constant probability density is assumed, highlighted in red for  $k = 5$ .

Equation 203), and thus the correction scheme is applied. Figure 17 plots

$$\frac{\sum_{i=1}^{N_{\text{samples}}} \Omega(\vec{q}_i)}{N_{\text{samples}}} \quad (211)$$

as a percentage for  $N_{\text{samples}} = 10000$  using the  $\ell_2$  norm for increasing dimensionality in  $\rho^{\text{exp.-decr.}}$ . It is clear that when the number of dimensions increases, the fraction of sample points determined to lie at the boundary of the domain of the probability density function increases. For  $d = 10$ , almost all of the sample points have the correction scheme presented in Section 8.7.3 applied to them. The reason for this is the uneven variance of the individual dimensions, which were set up to decrease steadily. We give a graphical rationalization of the problem at hand for kNN entropy estimation of probability densities with uneven variances in Figure 18. It shows the boundary of the variables' domains as a gray shaded area. The area inside which the kNN estimator assumes uniform probability density is shaded in orange. It can be seen that this region may extend beyond the domains of the variables. This is the case because the  $\ell_2$  distance determining the circles' radii is dominated by the neighbor distance of the dimension with a large extend  $y$ . Figure 19 shows the absolute entropy estimates as a function of the increasing dimensionality up to  $d_{\text{max}} = 10$  for a sample size of  $N_{\text{samples}} = 10000$ . Figure 20 plots the same results for a reduced sampling size

of  $N_{\text{samples}} = 500$ . In both cases, the neighbor parameter  $k = 1$  is used. The reference “with correction” refers to the kNN differential entropy estimate taking into account the correction scheme proposed by Rahvar and Ardakani, laid out in Section 8.7.3. From Figures 19 and 20 it becomes clear that the uncorrected kNN entropy estimator

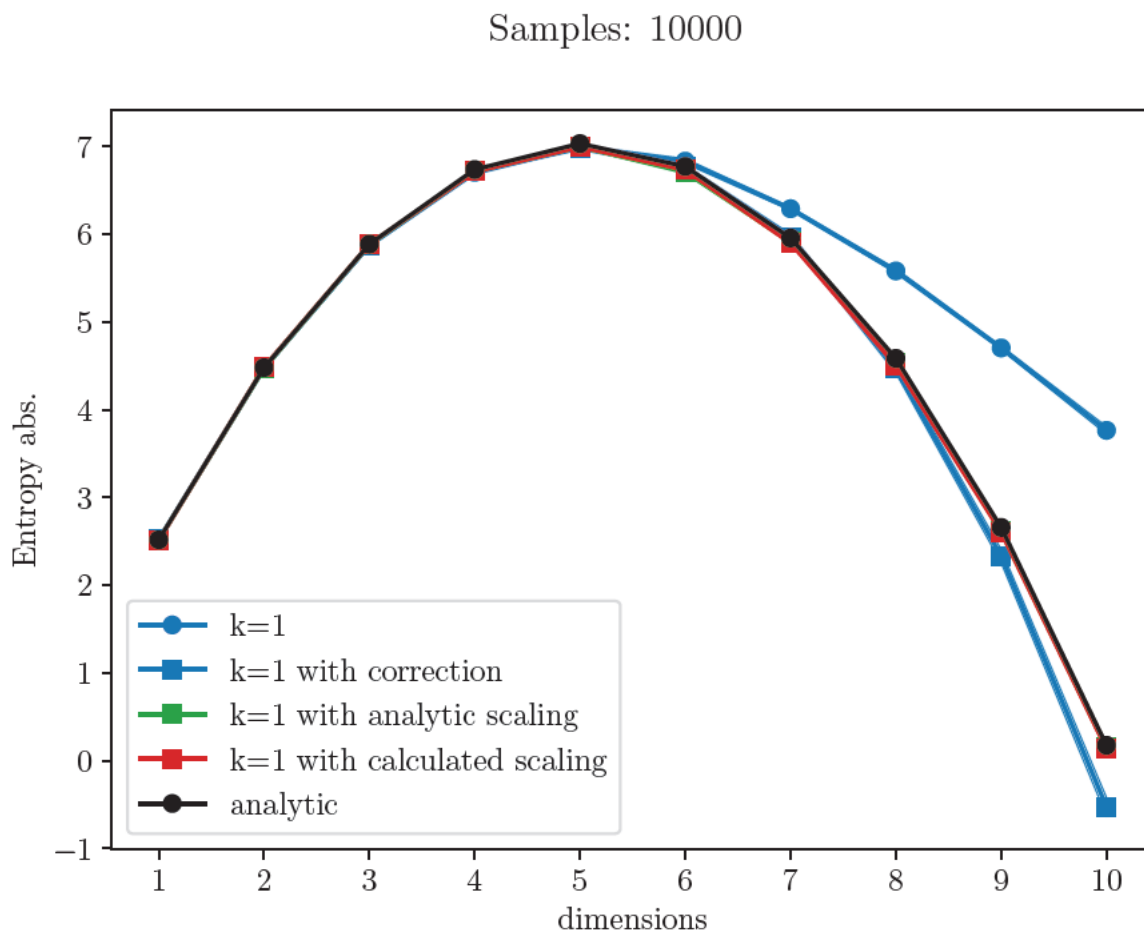


Figure 19: Multivariate gaussian entropy benchmark with exponentially decreasing standard deviations (see Equation 209) for  $N_{\text{samples}} = 10000$ . Samples have been drawn ten times, and the confidence interval of  $1\sigma$  is given in the plot via the transparently colored area around the mean value of all ten trials.

shows slow convergence for probability densities with exponentially decreasing variances. Comparison of both figures show that the benchmark using  $N_{\text{samples}} = 10000$  sample points is better converged in the case of the uncorrected kNN estimator than the version operating on  $N_{\text{samples}} = 500$ . Divergence (or, more precisely, slow convergence) of the uncorrected kNN estimator begins to show when the dimensionality exceeds 5 dimensions. The correction scheme proposed by Ardakani and Rahvar (see Section 8.7.3) is able to remedy this pitfall somewhat: Especially for the benchmark

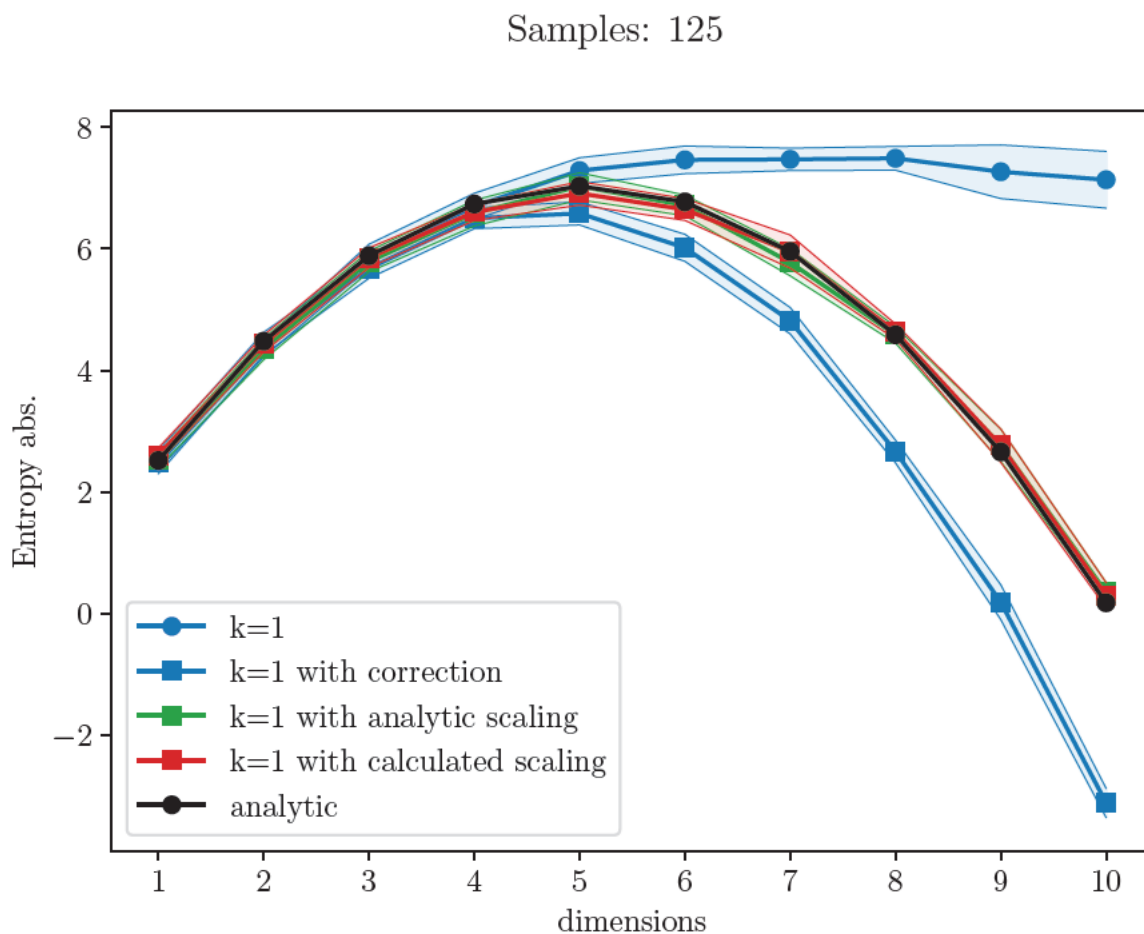


Figure 20: Multivariate gaussian entropy benchmark with exponentially decreasing standard deviations (see Equation 209) for  $N_{\text{samples}} = 125$ . Samples have been drawn ten times, and the confidence interval of  $1\sigma$  is given in the plot via the transparently colored area around the mean value of all ten trials.

using 10000 sample points, the correction can reestablish a significant degree of accuracy. As the correction scheme infers domain boundaries of the probability density function from the available sample points, the domain boundaries are expected to be underestimated by the correction scheme. In line with this reasoning, the corrected kNN entropy estimate approaches the analytic value from smaller values, while the uncorrected estimator converges towards the analytical entropy from above. Nevertheless, the analytic result is only fully restored once the stochastically drawn sample points are rescaled to equal variances before the uncorrected kNN estimator is used. Although to the best of our knowledge this finding has not been concisely reported, Marx and coworkers briefly note in their study on molecular entropy that they also invoked scaling to obtain more accurate kNN entropy estimates.<sup>[14]</sup> We briefly note that as the Gaussian-Mixture (GM) model-based benchmark will lay out in section 8.9 of this work, when variances are similar, the kNN entropy estimator can yield accurate results, even for non-gaussian distributed data. The full numerical results of the benchmarks for the multivariate gaussian probability density with exponentially decreasing standard deviation are given in Tables 9 and 10, as well as Tables 14 and 15 for increasing dimensionality. Figures 21 and 22 show data from the same benchmark with focus on the convergence with respect to increasing sampling size. In Figure 21, the convergence of the kNN entropy estimates is shown for a system with  $d = 10$  dimensions. It is clear that with the uncorrected and the boundary-corrected kNN entropy estimates converge very slowly. Figure 22 shows the same plot for a benchmark system composed of only 6 dimensions: Here, convergence up to numerical accuracy can be observed. Comparison of both plots shows the significantly reduced convergence of the entropy estimator when modes with different variances are estimated. This finding implies the following conclusion: By virtue of the exclusion principle, the slow convergence of the kNN entropy estimator can be traced back to the uneven variances of high-dimensional data. Increasing the number of samples leads to smaller kNN distances  $\zeta$ . The number of samples can be increased until the average distance  $\zeta$  captures the modes with small variances accurately. However, rescaling is able to reestablish convergence of the entropy estimate at drastically reduced sampling sizes. For comparison and validation, we have set up a second multivariate gaussian based benchmark system using linearly decreasing standard-deviations instead of exponentially decreasing standard-deviations, according to

$$\sigma_{d=a}^{\text{lin.-decr.}} = 1 + \left( \frac{2a - 2}{d_{\text{max}} - 1} \right). \quad (212)$$

Table 9: Numerical values of the kNN entropy estimates for the multivariate gaussian system with exponentially decreasing standard deviation with  $d = 5$  dimensions. The highest, lowest and average values for ten independent simulation runs, as well as the standard deviation  $\sigma$  is reported. The analytic differential entropy is  $1.459 \frac{\text{J}}{\text{mol K}}$ .

$N_{\text{samples}}$	125	250	500	1000	2000	5000	10000
5-dimensional model with exponentially decreasing standard deviations							
Uncorrected kNN estimator (Equation 160)							
average	54.842	54.347	53.432	52.694	52.891	52.747	52.620
$\sigma$	1.159	1.177	0.923	0.667	0.363	0.284	0.200
highest	56.910	56.188	55.004	53.808	53.521	53.121	52.821
lowest	52.687	52.338	51.4105	51.207	52.398	52.368	52.272
kNN estimator with rescaling using calculated variances							
average	52.856	52.609	52.226	52.357	52.154	52.303	52.448
$\sigma$	1.444	0.864	1.107	0.42879	0.296	0.198	0.149
highest	56.301	53.721	54.380	53.410	52.715	52.533	52.656
lowest	51.534	50.548	50.746	51.688	51.578	51.827	52.125
kNN estimator with rescaling using analytic variances							
average	51.901	52.118	51.951	52.290	52.341	52.379	52.506
$\sigma$	1.046	1.204	0.583	0.881	0.409	0.255	0.140
highest	54.069	53.869	52.877	53.213	53.146	52.860	52.748
lowest	50.010	49.822	51.184	50.199	51.858	52.039	52.331
kNN estimator with correction scheme (Section 8.7.3)							
average	48.164	51.222	51.409	51.916	52.235	52.379	52.437
$\sigma$	1.100	1.145	0.814	0.484	0.344	0.299	0.135
highest	50.298	53.326	53.469	52.622	52.676	53.031	52.687
lowest	46.820	49.044	50.434	51.186	51.549	51.914	52.291

Table 10: Numerical values of the kNN entropy estimates for the multivariate gaussian system with exponentially decreasing standard deviation with  $d = 10$  dimensions. The highest, lowest and average values for ten independent simulation runs, as well as the standard deviation  $\sigma$  is reported. The analytic differential entropy is  $1.459 \frac{\text{J}}{\text{mol K}}$ .

$N_{\text{samples}}$	125	250	500	1000	2000	5000	10000
10-dimensional model with exponentially decreasing standard deviations							
Uncorrected kNN estimator (Equation 160)							
average	58.526	54.632	49.661	43.980	39.760	34.941	31.304
$\sigma$	3.311	1.896	1.144	1.076	0.471	0.443	0.338
highest	62.010	57.318	52.591	45.553	40.580	35.712	31.948
lowest	52.027	50.568	48.248	41.800	39.099	34.045	30.921
kNN estimator with rescaling using calculated variances							
average	3.740	2.326	1.871	1.391	1.279	1.432	1.121
$\sigma$	1.991	1.012	1.143	0.576	0.464	0.260	0.204
highest	7.230	3.408	3.367	2.378	2.266	1.880	1.292
lowest	1.306	0.411	-0.751	0.274	0.399	0.967	0.602
kNN estimator with rescaling using analytic variances							
average	1.136	3.096	1.920	1.8544	1.403	1.224	1.116
$\sigma$	2.002	1.071	1.124	0.634	0.290	0.335	0.176
highest	5.098	5.285	3.283	2.784	1.96	1.751	1.386
lowest	-1.926	1.746	-0.204	1.019	0.911	0.726	0.846
kNN estimator with correction scheme (Section 8.7.3)							
average	-26.497	-21.157	-16.127	-10.594	-9.153	-6.041	-4.580
$\sigma$	1.171	3.114	1.335	1.325	0.587	0.834	0.863
highest	-23.917	-16.846	-13.284	-7.903	-8.312	-4.622	-2.734
lowest	-28.213	-28.430	-18.281	-12.365	-10.271	-7.366	-5.821

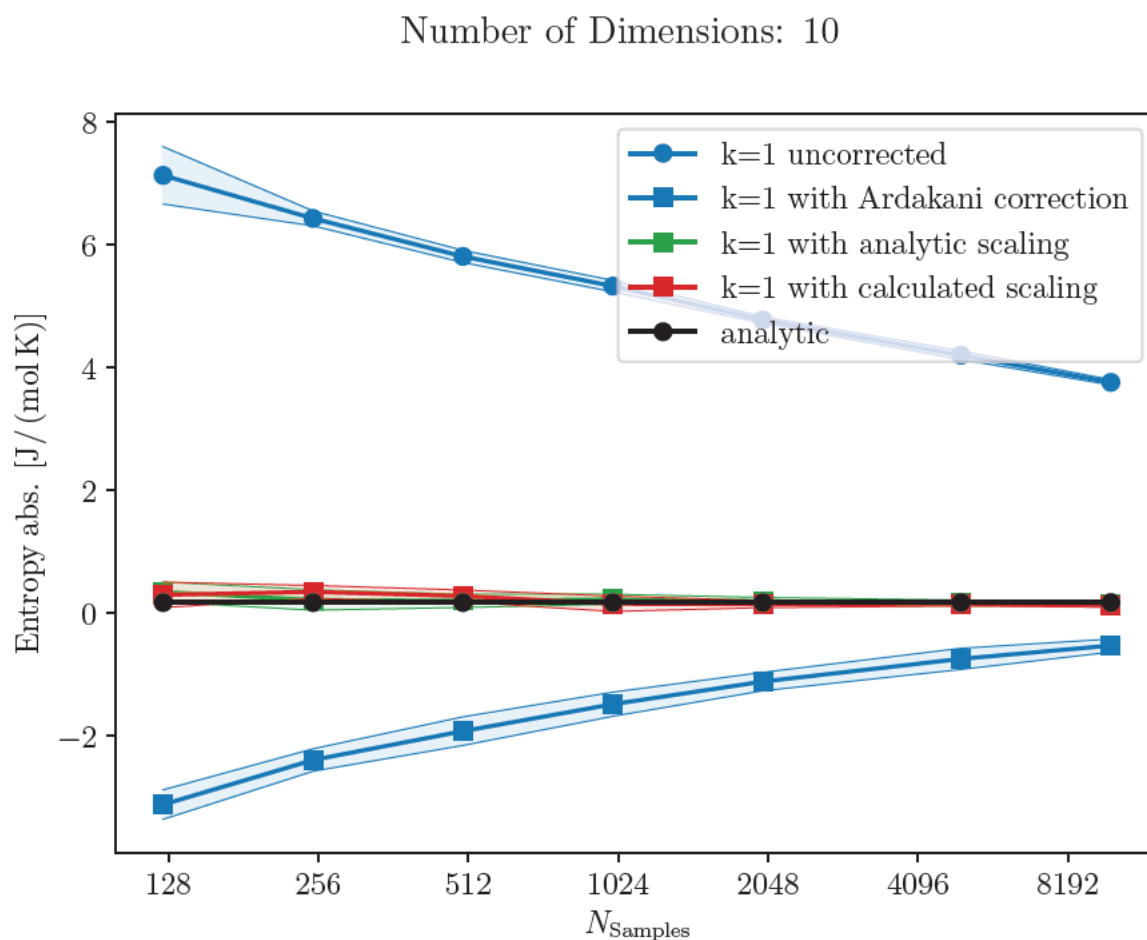


Figure 21: Multivariate gaussian entropy benchmark with exponentially decreasing standard deviations (see Equation 209) for  $d = 10$ . Samples have been drawn ten times, and the confidence interval of  $1\sigma$  is given in the plot via the transparently colored area around the mean value of all ten trials.



Number of Dimensions: 6

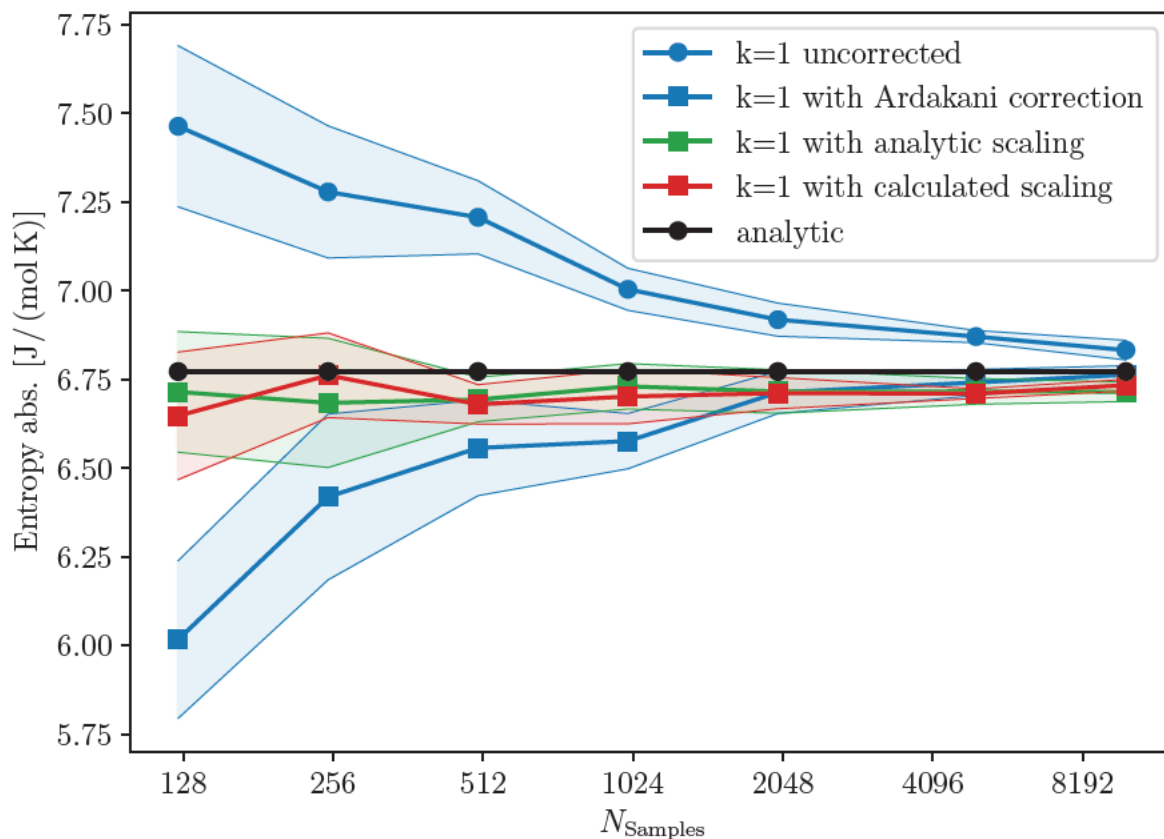


Figure 22: Multivariate gaussian entropy benchmark with exponentially decreasing standard deviations (see Equation 209) for  $d = 6$ . Samples have been drawn ten times, and the confidence interval of  $1\sigma$  is given in the plot via the transparently colored area around the mean value of all ten trials.

The full numerical results of the benchmarks are summarized in Tables 11 and 12 for increasing sampling sizes, as well as Tables 13 and 16 for increasing dimensions. Figure 23 gives the percentage of the sample points drawn that lie at the boundary of

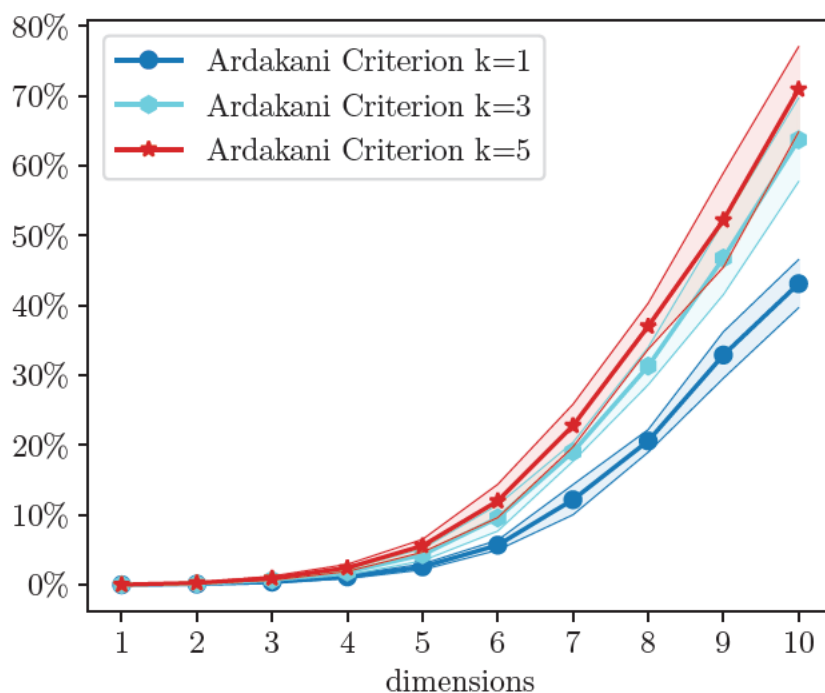


Figure 23: Percentage of points at the domain boundary for a multivariate gaussian with linearly decreasing standard deviations (see Equation 212) for  $N_{\text{samples}} = 10000$ . Samples have been drawn ten times, and the confidence interval of  $1\sigma$  is given in the plot via the transparently colored area around the mean value of all ten trials. Points are determined to lie at the domain boundary when  $\Omega(\vec{q}_i) \neq 0$  (see Equation 203).

the domain of the probability density function. When the dimensionality is increased, the fraction of points determined to be at the boundary rises. Comparison with Figure 17 reveals that the increase depends on the relationship of the dimensional variances, as the number of points at the boundary increases more slowly for a probability density with linearly decreasing variances than for one with exponentially decreasing variances. Figure 24 shows the convergence of this benchmark distribution with increasing sampling size for 10 dimensions. In this case, the convergence of the uncorrected estimator is more rapid, while the corrected estimator undershoots the analytical entropy value if too little samples are used, likely due to the mentioned over-correction. It has to be noted that the scaling-procedure is formally (mathematically) exact, while the boundary-correction scheme is approximate in the sense that the kNN hypersphere is reduced in size according to the minimum and maximum value in each

dimension, and thereby arbitrarily cut.

Number of Dimensions: 10

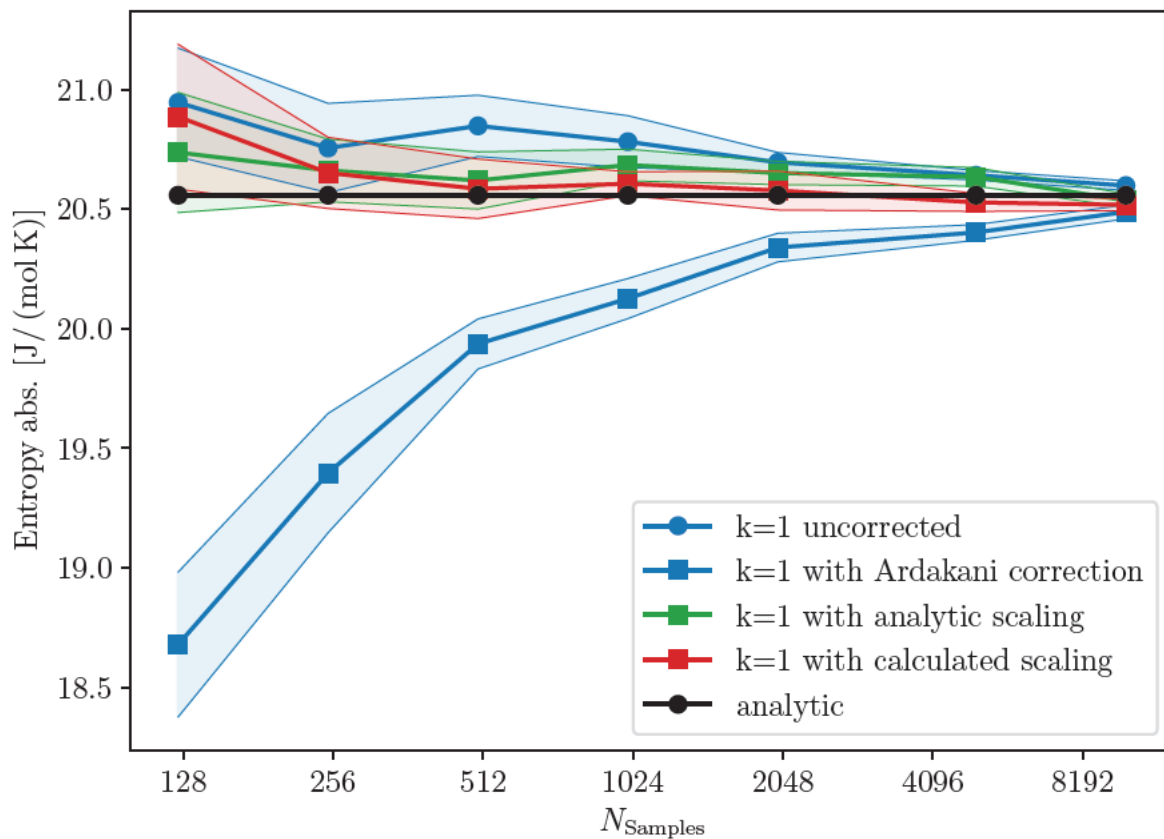


Figure 24: Multivariate gaussian entropy benchmark with linearly decreasing standard deviations (see Equation 212) for  $d = 10$ . Samples have been drawn ten times, and the confidence interval of  $1\sigma$  is given in the plot via the transparently colored area around the mean value of all ten trials.

Table 11: Numerical values of the kNN entropy estimates for the multivariate gaussian probability density with linearly decreasing standard deviations for  $d = 5$  dimensions. The highest, lowest and average values for ten independent simulation runs, as well as the standard deviation  $\sigma$  is reported. The analytic differential entropy is  $170.911 \frac{\text{J}}{\text{mol K}}$ .

$N_{\text{samples}}$	125	250	500	1000	2000	5000	10000
5-dimensional model with linearly decreasing standard deviations							
Uncorrected kNN estimator (Equation 160)							
average	84.508	84.147	84.824	84.525	84.532	84.610	84.566
$\sigma$	1.381	0.992	0.590	0.563	0.299	0.261	0.153
highest	86.803	85.901	85.677	85.828	84.909	85.008	84.797
lowest	82.345	82.159	83.696	83.804	83.928	84.138	84.332
kNN estimator with rescaling using calculated variances							
average	83.323	84.527	84.085	84.390	84.563	84.571	84.594
$\sigma$	1.165	0.653	0.946	0.722	0.320	0.185	0.25
highest	84.981	85.687	85.095	85.547	84.996	84.925	85.065
lowest	81.253	83.618	81.805	83.351	84.078	84.252	84.146
kNN estimator with rescaling using analytic variances							
average	84.496	84.678	84.073	84.466	84.704	84.490	84.522
$\sigma$	1.663	1.130	0.651	0.474	0.306	0.227	0.253
highest	86.766	86.137	84.966	85.391	85.277	84.851	85.110
lowest	81.068	82.788	83.043	83.599	84.351	84.105	84.221
kNN estimator with correction scheme (Section 8.7.3)							
average	81.371	83.358	83.807	83.990	84.468	84.422	84.499
$\sigma$	1.102	1.336	0.801	0.846	0.290	0.217	0.115
highest	83.521	84.869	85.086	85.375	84.920	84.692	84.743
lowest	79.565	79.852	81.900	82.318	83.845	84.099	84.297

Table 12: Numerical values of the kNN entropy estimates for the multivariate gaussian system with linearly decreasing standard deviations and  $d = 10$  dimensions. The highest, lowest and average values for ten independent simulation runs, as well as the standard deviation  $\sigma$  is reported. The analytic differential entropy is  $170.911 \frac{\text{J}}{\text{mol K}}$ .

$N_{\text{samples}}$	125	250	500	1000	2000	5000	10000
10-dimensional model with linearly decreasing standard deviations							
Uncorrected kNN estimator (Equation 160)							
average	173.069	173.704	172.433	172.771	171.943	171.512	171.193
$\sigma$	2.683	1.931	0.832	0.715	0.625	0.372	0.239
highest	178.693	176.844	173.580	173.696	172.820	172.236	171.543
lowest	169.050	170.668	170.558	171.322	170.566	171.053	170.761
kNN estimator with rescaling using calculated variances							
average	172.005	172.684	171.851	171.224	170.563	170.769	170.628
$\sigma$	2.395	1.312	0.896	0.824	0.665	0.242	0.175
highest	175.900	174.080	173.343	172.450	171.428	171.042	170.909
lowest	168.255	169.595	170.305	169.872	169.068	170.422	170.364
kNN estimator with rescaling using analytic variances							
average	172.675	171.188	171.518	170.787	170.878	170.717	170.571
$\sigma$	2.369	1.405	1.366	0.913	0.630	0.292	0.185
highest	176.530	173.599	173.287	171.948	171.991	171.276	170.795
lowest	169.417	169.132	168.549	168.784	170.147	170.295	170.195
kNN estimator with correction scheme (Section 8.7.3)							
average	157.361	162.501	165.251	167.815	169.024	169.614	170.342
$\sigma$	2.074	2.035	0.925	0.722	0.556	0.304	0.223
highest	161.213	167.080	166.756	169.046	169.955	170.184	170.573
lowest	154.205	158.879	164.148	166.565	168.259	169.044	169.948

Table 13: Numerical values of the kNN entropy estimates for the multivariate gaussian probability density with linearly decreasing standard deviations for  $N_{\text{samples}} = 10000$ . The highest, lowest and average values for ten independent simulation runs, as well as the standard deviation  $\sigma$  is reported. The analytic differential entropy in  $\frac{1}{\text{mol K}}$  is reported in the last row.

$d$	1	2	3	4	5	6	7	8	9	10
$N_{\text{samples}} = 10000$ , linearly decreasing standard deviations										
Uncorrected kNN estimator (Equation 160)										
average	20.867	41.233	60.706	79.460	97.340	114.394	130.250	145.262	158.953	171.193
$\sigma$	0.139	0.132	0.137	0.204	0.202	0.126	0.192	0.281	0.160	0.239
highest	21.062	41.444	60.909	79.822	97.635	114.585	130.462	145.618	159.112	171.543
lowest	20.632	41.000	60.453	79.104	96.941	114.196	129.891	144.856	158.601	170.761
kNN estimator with rescaling using calculated variances										
average	20.885	41.132	60.723	79.531	97.350	114.305	130.220	145.056	158.657	170.628
$\sigma$	0.095	0.173	0.180	0.156	0.151	0.208	0.184	0.177	0.205	0.175
highest	21.021	41.485	61.001	79.778	97.556	114.698	130.526	145.322	159.046	170.909
lowest	20.763	40.889	60.396	79.334	97.082	114.000	129.877	144.740	158.391	170.364
kNN estimator with rescaling using analytic variances										
average	20.853	41.148	60.723	79.550	97.322	114.434	130.313	145.162	158.704	170.571
$\sigma$	0.115	0.129	0.191	0.218	0.167	0.208	0.181	0.130	0.183	0.185
highest	21.062	41.364	60.895	79.853	97.561	114.812	130.614	145.428	159.017	170.795
lowest	20.659	40.991	60.338	79.197	97.070	114.083	129.978	144.988	158.366	170.195
kNN estimator with correction scheme (Section 8.7.3)										
average	20.895	41.242	60.735	79.388	97.371	114.265	130.200	144.926	158.416	170.342
$\sigma$	0.115	0.140	0.145	0.187	0.157	0.193	0.188	0.229	0.146	0.223
highest	21.132	41.565	60.921	79.738	97.699	114.761	130.452	145.296	158.583	170.573
lowest	20.739	41.048	60.509	79.125	97.170	114.035	129.888	144.483	158.128	169.948
Analytic differential entropy										
	20.931	41.222	60.820	79.661	97.671	114.755	130.799	145.654	159.119	170.916

Table 14: Numerical values of the kNN entropy estimates for the multivariate gaussian probability density with exponentially decreasing standard deviations for  $N_{\text{samples}} = 125$ . The highest, lowest and average values for ten independent simulation runs, as well as the standard deviation  $\sigma$  is reported. The analytic differential entropy in  $\frac{\text{J}}{\text{mol K}}$  is reported in the last row.

$d$	1	2	3	4	5	6	7	8	9	10
$N_{\text{samples}} = 125$ , exponentially decreasing standard deviations										
Uncorrected kNN estimator (Equation 160)										
average	2.4677	4.4824	5.854	6.6095	7.2252	7.4361	7.3955	7.3305	7.4179	7.0391
$\sigma$	0.1074	0.1733	0.134	0.1204	0.1959	0.1674	0.2805	0.2727	0.3261	0.3982
highest	2.6037	4.7077	6.133	6.8668	7.6183	7.6949	7.9157	7.8818	7.8825	7.4688
lowest	2.3051	4.1781	5.6272	6.4502	6.9754	7.1799	6.9474	6.9846	6.9623	6.2574
kNN estimator with rescaling using calculated variances										
average	2.4763	4.4078	5.754	6.7024	7.011	6.6973	5.9075	4.5635	2.714	0.4498
$\sigma$	0.1519	0.1053	0.1891	0.1891	0.1709	0.2088	0.3468	0.1924	0.3631	0.2398
highest	2.7735	4.6032	5.9774	6.9558	7.4002	7.0772	6.3375	4.9492	3.2539	0.8696
lowest	2.2022	4.2733	5.4734	6.4234	6.7926	6.3181	5.3004	4.2903	2.1044	0.1571
kNN estimator with rescaling using analytic variances										
average	2.5456	4.4478	5.8780	6.7561	6.9218	6.8716	5.8529	4.6581	2.7323	0.1366
$\sigma$	0.1688	0.1730	0.1535	0.1762	0.1808	0.1328	0.1221	0.2510	0.3258	0.2408
highest	2.7675	4.7067	6.1502	7.0719	7.2214	7.085	6.0305	5.1154	3.4705	0.6132
lowest	2.1400	4.0668	5.6285	6.5335	6.6572	6.6746	5.5987	4.2511	2.3476	-0.2317
kNN estimator with correction scheme (Section 8.7.3)										
average	2.5151	4.4113	5.7767	6.5277	6.6028	6.0751	4.7535	2.7111	0.1375	-3.1868
$\sigma$	0.1112	0.1660	0.1679	0.1575	0.1894	0.1918	0.2498	0.2483	0.1718	0.1409
highest	2.7260	4.6676	6.1676	6.7671	6.8871	6.4220	5.1390	3.1654	0.3134	-2.8766
lowest	2.3540	4.1765	5.5978	6.2297	6.2053	5.8135	4.3467	2.3992	-0.3217	-3.3932
Analytic differential entropy										
	2.5176	4.4795	5.8860	6.7369	7.0322	6.7720	5.9562	4.5849	2.6580	0.1755

Table 15: Numerical values of the kNN entropy estimates for the multivariate gaussian probability density with exponentially decreasing standard deviations for  $N_{\text{samples}} = 10000$ . The highest, lowest and average values for ten independent simulation runs, as well as the standard deviation  $\sigma$  is reported. The analytic differential entropy in  $\frac{\text{J}}{\text{molK}}$  is reported in the last row.

$d$	1	2	3	4	5	6	7	8	9	10
$N_{\text{samples}} = 10000$ , exponentially decreasing standard deviations										
Uncorrected kNN estimator (Equation 160)										
average	2.5153	4.4760	5.8736	6.7205	7.0168	6.8375	6.2917	5.5762	4.7343	3.7650
$\sigma$	0.0137	0.0192	0.0184	0.0157	0.0214	0.0140	0.0221	0.0266	0.0225	0.0406
highest	2.5378	4.5170	5.9000	6.7504	7.0484	6.8566	6.3269	5.6255	4.7700	3.8424
lowest	2.4976	4.4516	5.8417	6.6987	6.9875	6.8116	6.2611	5.5135	4.6997	3.7190
kNN estimator with rescaling using calculated variances										
average	2.5184	4.4804	5.8728	6.7188	7.0107	6.7375	5.9045	4.5187	2.6018	0.1348
$\sigma$	0.0189	0.0129	0.0150	0.0227	0.0197	0.0166	0.0162	0.0175	0.0280	0.0245
highest	2.5427	4.5014	5.9004	6.7453	7.0440	6.7598	5.9318	4.5453	2.6522	0.1554
lowest	2.4807	4.4588	5.8444	6.6797	6.9870	6.7113	5.8741	4.4952	2.5503	0.0724
kNN estimator with rescaling using analytic variances										
average	2.5113	4.4681	5.8900	6.7254	6.9949	6.7185	5.8882	4.5294	2.5890	0.1342
$\sigma$	0.0211	0.0184	0.0155	0.0172	0.0280	0.0223	0.0289	0.0141	0.0136	0.0212
highest	2.5475	4.4932	5.9121	6.7470	7.0499	6.7549	5.9351	4.5600	2.6063	0.1667
lowest	2.4868	4.4436	5.8627	6.6928	6.9519	6.6785	5.8406	4.5057	2.5605	0.1017
kNN estimator with correction scheme (Section 8.7.3)										
average	2.5203	4.4771	5.8731	6.7161	6.9980	6.7599	5.9717	4.4596	2.3003	-0.5508
$\sigma$	0.0135	0.0111	0.0226	0.0248	0.0097	0.0267	0.0327	0.0373	0.0879	0.1038
highest	2.5456	4.4885	5.9097	6.7536	7.0188	6.8017	6.0264	4.5298	2.3946	-0.3289
lowest	2.5010	4.4592	5.8410	6.6618	6.9821	6.7185	5.9257	4.3912	2.1140	-0.7001
Analytic differential entropy										
	2.5176	4.4795	5.8860	6.7369	7.0322	6.7720	5.9562	4.5849	2.6580	0.1755



Table 16: Numerical values of the kNN entropy estimates for the multivariate gaussian probability density with linearly decreasing standard deviations for  $N_{\text{samples}} = 125$ . The entropy is given in  $\frac{\text{J}}{\text{mol K}}$ .

$d$	1	2	3	4	5	6	7	8	9	10
$N_{\text{samples}} = 125$ , linearly decreasing standard deviations										
Uncorrected kNN estimator (Equation 160)										
average	21.0614	41.0422	61.1707	78.9898	96.4694	113.7134	129.6119	146.7733	160.2496	173.0692
$\sigma$	1.1068	1.0989	1.5427	1.3301	1.3668	2.2333	1.6478	1.7415	2.235	2.6828
highest	23.6965	42.5257	63.5907	81.0805	98.1997	118.6954	132.9786	150.3008	163.2343	178.6932
lowest	19.8271	39.388	59.0814	76.5674	94.01	109.9311	127.5496	144.9022	156.1788	169.0495
kNN estimator with rescaling using calculated variances										
average	20.4255	42.1372	60.7742	79.4012	96.6096	114.233	131.9806	146.7737	160.0628	172.005
$\sigma$	1.0942	1.1725	1.1851	1.4577	1.1443	2.1105	2.1388	2.1974	1.7471	2.3948
highest	22.1327	44.1409	63.1103	81.7829	98.4313	117.0616	136.8324	151.7346	162.7276	175.9003
lowest	18.8487	40.4164	58.9102	76.9199	94.9937	110.3862	129.1154	143.1823	157.5849	168.2553
kNN estimator with rescaling using analytic variances										
average	20.8498	41.4202	59.7542	79.6756	97.4398	114.4795	131.1729	145.6946	160.4795	172.6746
$\sigma$	1.2849	1.3942	1.1616	1.8424	1.3305	1.0977	1.3879	1.7875	2.5626	2.3691
highest	23.2899	44.5748	60.9923	83.2117	99.7629	116.252	133.2968	148.3573	165.8904	176.5296
lowest	18.9707	39.7042	56.961	76.5062	95.8947	112.2632	128.8406	142.3527	155.7995	169.4166
kNN estimator with correction scheme (Section 8.7.3)										
average	21.672	39.9086	59.954	77.2079	94.8584	111.2342	124.9278	139.206	149.9738	157.3607
$\sigma$	0.8683	1.3702	1.2406	1.7978	1.3493	1.318	1.3592	2.7072	2.1547	2.074
highest	22.7531	42.8886	61.2068	79.6506	96.8454	112.6844	127.9449	146.1232	153.2957	161.2125
lowest	19.3913	37.4899	57.9584	74.3377	91.8957	109.097	123.0237	136.1285	146.8712	154.2047
Analytic differential entropy										
	20.931	41.222	60.820	79.661	97.671	114.755	130.799	145.654	159.119	170.916

In summary, we have shown how the uneven variances of multi-dimensional distributions can impact the accuracy of the kNN entropy estimator and introduce errors. These errors can be remedied by increasing the number of samples used in the entropy estimator, however, convergence is slow. The correction scheme proposed to account for the boundary effect by Rahvar and Ardakani yields faster convergence with respect to the number of samples and more accurate estimates in the case of significantly different variances (i.e. exponentially decreasing standard deviations). Although this correction scheme has been developed to fix the error at the domain boundaries, the rescaling of the isotropic hypersphere where density is assumed constant into a hyperellipsoid (for the  $\ell_2$  norm) is able to correct the error arising from uneven variances as well. In the present benchmark, only the rescaling of the stochastically drawn data points according to Equation 110, and thereby assuring identical variances across all dimensions before performing the kNN entropy estimation, yields exact results. The scaling procedure is easy to apply for this test-distributions, as the exact probability density and its variances are known. For chemical data derived from MD simulations, the application of the scaling procedure will likely prove to be more challenging.

#### 8.7.5. Considerations for Applying the kNN Entropy Estimator to MD Simulation Data

Previous reports on the use of the kNN entropy estimator for configurational entropies showed highly desirable properties: Hnizdo *et al.* applied their estimator to ensembles of samples of dihedral angles, showing that both the kNN estimators yield estimates below the quasi-harmonic entropy estimate.<sup>[226]</sup> Hnizdo *et al.* furthermore considered the higher order MIE of kNN-estimated and histogrammed entropies, discovering favorable properties of the kNN entropy estimate.<sup>[203]</sup> Numata *et al.* used the kNN estimator to improve upon entropies based on the Quasi-Harmonic Approximation (QHA) *a posteriori*. They provide a brief proof-of-concept application, comparing their results with entropies calculated from Normal Mode Analysis (NMA). Numata *et al.* found that the impact of correcting for MI terms is found to be bigger than corrections for anharmonicities.<sup>[154]</sup> The practical application of the “non parametric” kNN estimator somewhat counter intuitively requires specifying several parameters that may influence the accuracy of the estimator: Trajectory length, the number of MD snapshots considered for the entropy calculation ( $N_{\text{samples}}$ ), and the neighborhood-parameter  $k$ . Furthermore, if the system is of high dimensionality, the MIE might be

invoked. In this case, the truncation order of the MIE needs to be specified. If one assumes that the MD simulation at hand has traversed the entire region of relevant phase space, the impact of the trajectory length vanishes. All other parameters are interdependent. The convergence of the kNN estimator with the trajectory length was considered by Hnizdo and coworkers.<sup>[203,226]</sup> They observed near-convergence only for sample sizes larger than 5 million frames for the full-dimensional kNN procedure of a simulation of tartaric acid. For this chemical system with only a small number of DOFs, the sample size required for convergence is disconcertingly high. Hnizdo's findings indicate that estimators using lower  $k$  values converge faster, which is a favorable finding as computational effort increases when using larger values of  $k$ . Grubmüller and coworkers have used an adapted method based in spirit on the kNN entropy estimation protocol, called the Minimally Coupled Subspace Approach and find favorable properties for it.<sup>[15]</sup>

A remedy for the dependence of the estimator's convergence on sampling sizes is thus desirable. Especially for high-dimensional systems, decreasing the number of data points will improve the speed of the nearest neighbor search dramatically. Whether this data-reduction retains accuracy depends on an open question: Does the kNN estimator indeed need millions of samples to reach convergence? Alternatively, it is plausible that fewer sample points may be used if only they are taken from a simulation trajectory faithfully representing the Boltzmann ensemble. In practical application, it is clear that the largest error made is presumably due to the finite runtime of the simulation. After the length of the simulation as well as the number of simulation frames is chosen, one must consider whether the full-dimensional nearest neighbor search is feasible at all for the chemical system at hand. For large macromolecules, this will most likely not be the case due to computational scaling. Expansion schemes such as the MIE or the Mutual Information Spanning Tree (MIST) method nevertheless enables kNN entropy estimation by expanding the full-dimensional entropy as a series of lower-dimensional, (higher order) MI based contributions.<sup>[13]</sup> This series may then be truncated after a certain order. The order of truncation may be chosen arbitrarily. While 1<sup>st</sup> order corrections can be directly related to the concept of correcting modes for anharmonicity in an HO picture, higher order corrections are notoriously hard to rationalize.<sup>[227]</sup> Even worse, recent investigations of Goethe *et al.* showed that for some multivariate gaussian probability densities, the MIE may diverge.<sup>[209]</sup> Although this finding is surely bad news for all MIE-based estimators, it does not necessarily prohibit practical application. The divergence of the sieve method based MIE at high orders

is troublesome primarily from a theoretical point of view. Empirical benchmarks on chemical data are necessary for assessing the performance of truncated kNN-based MIE schemes. In their analysis, Hnizdo *et al.* noted that the 1<sup>st</sup> order MIE of kNN entropies converges much faster and yields an improved upper bound to the true entropy in comparison with the quasi-harmonic entropy estimate.<sup>[226]</sup> For the MIE of the kNN entropy, higher order terms are found to converge slower with regards to sampling size. Finally, an appropriate value for the neighbor identifying parameter  $k$  must be chosen. While Hnizdo *et al.* critically discuss the impact of  $k$  ranging from 1 to 5 without giving general advice, Numata *et al.* suggest using  $k = 4$  without formal justification.<sup>[154]</sup>

The parameters mentioned up until now ( $k$ , the MIE truncation order, simulation length) have been analyzed by the previous works on the kNN estimator at least to some degree. There are still further options implicitly contained in the kNN estimator which have been recognized mainly in information sciences. When literature on the entropy estimator from these related fields is considered, it quickly becomes clear that the  $\ell_2$  euclidean distance norm for the neighbor distance is not the only possible choice.<sup>[214,220]</sup> The  $\ell_\infty$  (maximum) metric may equivalently be used to estimate kNN entropies. In their untruncated theoretical framework, the entropy estimators are invariant to the norm used, up to an additive constant. Furthermore, as mentioned, at least three distinct but different mathematical expressions for the entropy estimator's function have been proposed. They differ solely in the definition of the additive constant which corrects the asymptotic bias intrinsic to the kNN entropy estimator. The one closest to the original estimator by Kozachenko and Leonenko (see [216]) is derived by Gorja *et al.*<sup>[221]</sup> Others differing in additive constants are found in the literature by Hnizdo<sup>[217]</sup> as well as Lombardi and Pant.<sup>[214]</sup>

In order to obtain entropic values comparable with values obtained from RRHO approaches, it is important that the momentum contribution (see Equation 87) is taken into account as well.

## 8.8. Computational Implementation of Entropy Estimation Methods in the CAST Program Package

The kNN entropy estimator has been implemented into the CAST program package for molecular modeling.<sup>[228]</sup> The code is available online and may be distributed freely due to the open-source nature of its license.<sup>[229]</sup> Additionally, several quasi-harmonic

entropy estimation protocols have been made available in the CAST package as well. Vibrational entropy estimation protocols based on computation of the molecular Hessians are not available in CAST, due to CAST currently not being able to handle and operate on molecular Hessians. For this reason, the mixture of conformers approach and the RRHO approach are not implemented. These estimates may however be easily evaluated using standard quantum chemical molecular modeling software, such as the GAUSSIAN16 program package, or others.<sup>[152]</sup>

The CAST package is written in the C++ programming language, which enables high computational speed and flexibility. The CAST program provides object-oriented classes for the abstraction of common chemical moieties such as atoms, molecules, force field parameters, and so on. Furthermore, CAST provides an abstraction layer for different approximative methods for evaluating the energy and gradient of a chemical system. The methods implemented in CAST are therefore fully agnostic with regards to the level of theory used for energy- and gradient computations. The common AMBER or OPLSAA force fields (as implemented in the TINKER program package, see [230, 231]) may be used. Quantum chemical software such as GAUSSIAN16 may be accessed by means of a custom file-reader interface as well. Cost effective semi-empirical quantum chemical approaches may also be employed by interfacing the MOPAC<sup>[232]</sup> or DFTB+<sup>[233]</sup> program suites. Natively, CAST provides code for executing MD simulation runs as well as for using hybrid-approaches such as QM/MM or QM/QM/MM with the available energy interfaces.<sup>[234]</sup> Potential of Mean Force (PMF) computations of the free energy are possible by using the Umbrella Sampling approach in conjunction with the WHAM method. More details on CAST may be found in the reference publication or in the publicly available source code repository.<sup>[228,235,236]</sup>

For the entropy estimation methods implemented in CAST, a MD simulation trajectory in the TINKER ARC or XYZ format needs to be provided. The XYZ format is easily obtainable from interconversion of almost any other file-format using programs such as OpenBabel.<sup>[237]</sup> Alternatively, an MD simulation run can be executed using the MD task provided in CAST itself.<sup>[236]</sup>

The entropy estimation methods are grouped in the task ENTROPY, which needs to be specified in the CAST input file. In a first step, the MD trajectory is processed to create an object of the type `TRAJECTORYMATRIXREPRESENTATION`. This class is wrapper for an  $N_{\text{samples}} \times N_{\text{DOF}}$  matrix. In order to build this matrix from the cartesian MD trajectory, various user-defined input options for truncation and modification of the data are

available: The molecular snapshots may be rotationally aligned prior to the calculation of the `TRAJECTORYMATRIXREPRESENTATION`. The rotational fit is performed using a custom Singular Value Decomposition based implementation of the Kabsch algorithm,<sup>[238–240]</sup> which was parallelized using OpenMP for computational speedup.<sup>[241]</sup> Furthermore, the number of snapshots taken in the entropy estimation algorithms may be reduced for computational speedup by pruning the MD simulation data and only considering every  $n^{\text{th}}$  frame. The cartesian coordinates  $x, y, z$  of every atom are serialized in the `TRAJECTORYMATRIXREPRESENTATION`. This means all are treated on equal footing: Column 1 corresponds to the  $x$ -coordinate of the first atom, column 3 to the  $z$ -coordinate of the first atom, column 4 to the  $x$ -coordinate of the second atom, and so on. Furthermore, the heating and equilibration phases of an MD simulation run, whose cartesian structures may still be present in the MD trajectory data, can be excluded from the entropy estimation by providing a user-defined starting step. The input processing is also able to only consider certain, user-specified atoms in the construction of the `TRAJECTORYMATRIXREPRESENTATION`, such as to exclude explicit solvent molecules or only consider one relevant site of a chemical macro structure. By default, the distances saved in the `TRAJECTORYMATRIXREPRESENTATION` object are by default in the unit Ångstrom, but may be converted to SI units if desired. Some entropy estimation protocols require mass-weighting of the `TRAJECTORYMATRIXREPRESENTATION`, i.e. multiplication of each atomic nuclei's positional coordinate by the square-root of the atomic mass associated with the atom, or for collective coordinates (i.e. linear combinations of atomic coordinates) multiplication with the reduced mass of the respective mode. CAST will perform this step automatically if required, however, it is important to keep in mind that the absolute entropies will differ depending on whether mass-weighting is applied or not. The comparison of absolute entropy values between different entropy estimation protocols is not possible if only one of them operates on mass-weighted coordinates. This can be explained by considering Equation 110: Mass-weighting is similar to scaling of the individual dimensions of the MD data by a constant factor, which will shift the absolute value of the differential entropy expression by a constant factor. This is the case for all entropy estimation approaches based on estimating the differential entropy in the spatial domain. After the built-up of the `TRAJECTORYMATRIXREPRESENTATION` object, the actual entropy estimation procedures and the computations involved therein may proceed. In the course of this work, the following entropy estimation approaches have been implemented into the CAST program package:

- a) Quasi-Harmonic entropy estimation by Karplus and Kushick according to [183]
- b) Quasi-Harmonic entropy estimation by Schlitter according to [184] (therein: Equation 15)
- c) kNN entropy estimation by Hnizdo according to [217] (full-dimensional)
- d) kNN entropy estimation by Hnizdo according to [226] (arbitrary order MIE)
- e) kNN entropy estimation hybrid scheme by Numata *et al.* according to [154]

For method *c* and *d*, both the  $\ell_2$  and the  $\ell_\infty$  distance norm variants have been implemented, with correction terms according to [214] for the  $\ell_\infty$  case. The arbitrary order MIE procedure for method *d* was implemented using OpenMP parallelization<sup>[241]</sup> using recursive function calls. Two linear algebra libraries were integrated into CAST to enable speedup of computational timings on flexible deployments of the code. The `EIGEN` header-only matrix library is written using templated C++ header files and has to be compiled on each machine where CAST is to be executed. Furthermore, CAST is interfaced with the `ARMADILLO` C++ wrapper library which gives access to Basic Linear Algebra Subprograms (BLAS) and Linear Algebra Package (LAPACK) libraries available on the target machine. High-speed replacement libraries such as OpenBLAS<sup>[242]</sup> or Intel's MKL<sup>[243]</sup> may be used as well. Which of the two libraries offers the best execution speed is depending on the machine architecture, computational power of the LAPACK and BLAS packages deployed thereon, as well as the degree of compiler optimization present during the compilation of CAST. CAST uses a internal wrapper-class for interfacing both linear-algebra libraries on equal footing. Substitutions for functionality missing in either `EIGEN` or `ARMADILLO` are provided as part of the linear-algebra-implementation in CAST. All entropy estimation protocols may be run using either of the two libraries.

To enable the kNN entropy calculations presented in this thesis, a novel implementation of the kNN entropy estimation algorithm was developed. For the nearest neighbor search, a brute-force approach was chosen. This is in contrast to the approach used in the ANN package, which was up until now used by some researchers in kNN entropy computations for molecular systems.<sup>[14,154]</sup> The ANN package works by initially building a so-called kd-tree data-structure. Our benchmarks in the CAST program package indicated that the performance loss from building a kd-tree previous to the nearest neighbor search outweighs the performance gain during the search. For this reason, a brute-force approach, calculating every nearest neighbor distance explicitly, might

be favorable for practical applications. By means of straight forward parallelization through the OpenMP framework, additional performance increases on multi-core architectures were achieved. The MIE algorithm in CAST was implemented in a generalized, recursive manner. It enables the evaluations of the MIE at arbitrary truncation orders. We note that for small sampling sizes ( $< 100000$ ), the full-dimensional procedure is often computationally faster than truncated MIE approaches beyond the 1<sup>st</sup> order.

To compensate for numerical issues in the summation of the kNN entropy estimation algorithm, we have implemented our algorithms using Kahan's summation technique. This algorithm was developed to give reduced numerical errors due to finite machine floating-point precision when summing a large amount of numbers.<sup>[244]</sup> The pseudocode of this running error compensation for a set of numbers  $[x_1, x_2, x_3, \dots]$  is as such:

---

```

1  algorithm kahanSumation is
2  input: Set of numbers X = [x1, x2, x3, ...]
3  output: Finite-Precision-Error compensated summation estimate s
4  s ← 0
5  c ← 0
6  for each xi in X do
7  y ← xi - c
8  t ← s + y
9  c ← (t - sum) - y
10 s ← t
11 return s

```

---

Listing 1: Pseudocode of Kahan's summation algorithm.

Algebraically, the variable  $c$  should be identical to zero. However, a finite precision computer implementation accumulates errors due to truncation. CAST will print a naive summation estimate as well as a value computed using Kahan's correction. If these number differ, Kahan's estiamte should be the more reliable one. Often, the relevance of Kahan's correction will depend on the distance unit chosen for the construction of the `TRAJECTORYMATRIXREPRESENTATION`. If SI units are taken, then distances between neighboring molecular structures can be very small numbers, whose summation will accumulate errors due to the truncated floating point nature of the computational implementation. The nearest neighbor distances of the molecular snapshots' coordinates themselves are computed by means of a custom function. This code was optimized for speed and contains as little branching-points as possible. A



pointer to a reserved area of heap space for storage of the nearest neighbor distances is passed in, in order to avoid costly reallocation of heap space when executing the nearest neighbor calculation function in large succession.

## 8.9. Benchmark of the kNN Entropy Estimator using a GM Model Derived from MD Simulations

### 8.9.1. Background & Description of the Benchmark Design

Testing the accuracy of kNN entropy estimators is difficult, as analytical closed-form expressions of entropies are only known for the simplest distributions. Model systems where the absolute entropy can be obtained exactly are needed, in order to compare the performance of approximate estimators. This has been recognized by Goethe *et al.* as well, who used multivariate gaussians to assess the goodness of the MIE, as for these functions an absolute closed-form expression for the entropy is available.<sup>[209]</sup> In Section 8.7.4 of this work, we have utilized uncorrelated multivariate gaussian functions to detangle the error sources intrinsically present in the kNN entropy estimators. This benchmark showed that slow convergence of the estimator likely results from the uneven variances of different dimensions. However, the purpose of the nonparametric kNN entropy estimation is to go beyond the assumption of normal distributions, thereby accounting for anharmonicity and supra-linear correlations, which cannot be captured and removed by a linear coordinate transformation. The dynamics of large chemical systems display multiple minima and therefore deviations from normality in their probability distributions. A multivariate gaussian is thus too simple of an approximation for the underlying molecular probability density, except for very rigid systems with only one conformation to consider. A more reasonable way to obtain a reference probability density function as a closed-form expression is to use a Gaussian-Mixture (GM) model, which is a normalized sum of several multivariate gaussians  $G$  with mean  $\vec{\mu}$  and covariance matrix  $\Sigma$ :

$$\rho_{\text{GM}}(\vec{x}) = \sum_{i=1}^n c_i * G(\vec{\mu}_i, \Sigma_i, \vec{x}) \quad (213)$$

Here,  $G(\vec{\mu}_i, \Sigma_i, \vec{x})$  is a multivariate gaussian probability density whose function is given by

$$G(\vec{\mu}, \Sigma, \vec{x}) = \frac{e^{(-\frac{1}{2}(\vec{x}-\vec{\mu})^t \Sigma^{-1} (\vec{x}-\vec{\mu}))}}{\sqrt{(2\pi)^d |\Sigma|}}. \quad (214)$$

Although there is no closed-form analytical expression for the entropy of a GM Model probability density, the differential entropy expression (as given in Equation 90) may be integrated numerically. Due to the computational curse of dimensionality, even numerical integration schemes become unfeasible for systems with more than a few dimensions. For these reasons, we construct an artificial reference system by reducing the dimensionality of the molecular data before we fit the GM Model. This is done by calculating the principal components of the MD simulation. A common notion when using PCA in computational chemistry is that the first principal components (associated with the modes of highest variance) deviate most from normal distributions.<sup>[154,245]</sup> Empirical findings support this statement.<sup>[246]</sup> These first principal modes may be used to fit a low-dimensional GM Model mode, retaining most chemical information.

An artificial but chemically relevant benchmark environment is therefore generated via dimensionality reduction of MD trajectories. Apart from being derived from chemical simulation data, the GM Model probability density function may be regarded as an arbitrary test-distribution. Key properties of this probability density function are that it is a continuous, low-dimensional, and potentially multimodal function. In a second step, random samples are drawn from the fitted GM model's probability density function. From these samples, kNN entropies are estimated and compared to an accurate entropy value obtained from numerical integration of the analytic entropy expression. By this procedure, the impact of each of the previously discussed parameters on the kNN estimates of configurational entropies can be assessed. Figure 25 showcases the reasoning and work-flow of the described benchmarking procedure. We note that the absolute entropy obtained from integration of the GM Model does not differ significantly from the entropy estimated from the samples of the original MD simulation.

In order to draw samples from the probability distributions given by the GM models, a custom procedure which is only applicable to GM models was used: The method operates by exploiting the fact that GM models are linear combinations of individual multivariate normal distributions. The summation factors  $c_i$  (see Equation 213) are

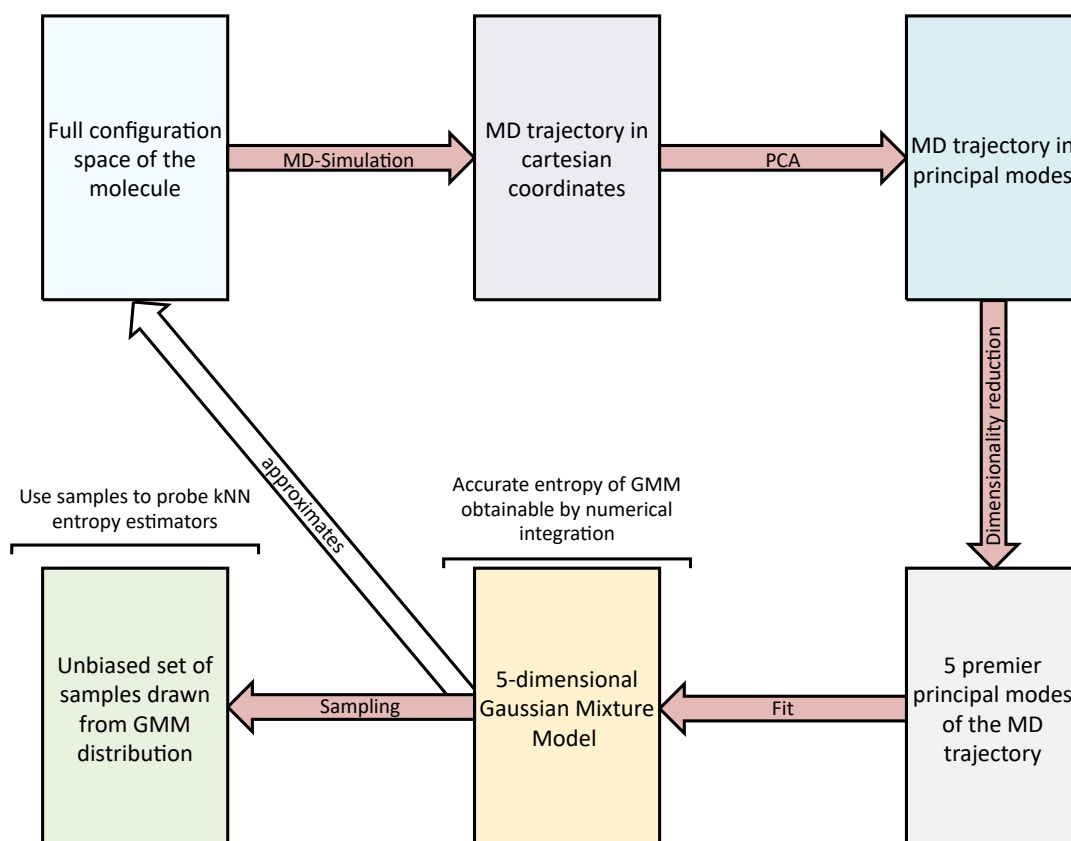


Figure 25: Scheme explaining the conception of the GM Model for the alanine dipeptide. Arrows colored in red represent steps where information is discarded or lost.

normalized and sum to unity:

$$\sum_{i=1}^n c_i = 1. \quad (215)$$

In order to draw samples from the GM Model distribution we define the auxiliary variable  $e_i$  as

$$e(i) = e_i = \sum_{j=1}^i c_j \leq 1. \quad (216)$$

The algorithm works on custom vector- and matrix-datatypes and assumes that functions for drawing both uniformly and normally distributed proceeds as:

---

```

1  algorithm gmSampling is
2  input: d-dimensional gm-model  $G(\vec{\mu}_i, \Sigma_i, \vec{x})$ ,  $\{e_i\}$ 
3  output: d-dimensional sample drawn from GM model
4  r ← uniformRandomNumber with range [0,1]
5  for each i in d do
6  if r smaller or equal to e(i) do
7  ii ← i
8  break
9  # The sample will be drawn from the ii'th normal-distribution
10 # by first drawing from one-dimensional Gaussian with  $\mu=0$  and  $\sigma=1$ 
11 # and then scaling appropriately
12 rv ← []
13 covEigVal ← Diagonal Matrix with Eigenvalues of  $\Sigma_{ii}$ 
14 covEigValSqrt ← square root of covEigVal
15 covEigVec ← Matrix of Eigenvectors of  $\Sigma_{ii}$ 
16 for each i in d do
17 rr ← normalRandomNumber with  $\mu=0$  and  $\sigma=1$ 
18 append rr to vector rv
19 rrv ←  $\vec{\mu}_{ii} + \text{covEigVal} * \text{covEigValSqrt} * \text{rv}$ 
20 return rrv

```

---

Listing 2: Pseudocode of custom GM Model sampling algorithm. Comments are included to make the rationale of the computational steps clearer. Comment lines are marked with an ampersand at the beginning of the line.

We note that this algorithmic approach has been proposed previously in an online discussion on the topic (see [247]).

### 8.9.2. Computational Details

All entropy calculations have been performed with the development version of the CAST program package.<sup>[228]</sup> In order to build the GM Model,  $\rho_{\text{GM}}(\vec{x})$ , publicly available MD simulation trajectories from multiple simulation runs of the alanine dipeptide. The simulations were performed using the GROMACS software package.<sup>[248,249]</sup> Details of the configuration of the MD simulations can be found in their respective original publications.<sup>[250,251]</sup>

In a first step, all solvent molecules (water) were removed from the trajectory data. The simulations were performed using periodic boundary conditions. The artifacts of the periodic boundary conditions were removed by eliminating the so-called boxjump, i.e. making sure the target molecule is always given in proper cartesian coordinates. To this end, conformations with broken-bonds, resulting from the molecule being placed at the edge of the periodic box, have been reestablished. In the next step, translation of the molecule was eliminated by calculating and subsequently aligning the center-of-mass of each simulation frame. Next, we concatenated all 20 available simulation replica (performed using identical parameters) to obtain one combined trajectory representation. Then, the molecular structures making up the trajectory were aligned using Kabsch's procedure, which finds a rotation that renders the root-mean-square-error of the euclidean distance of each conformation minimal with respect to some given reference conformation.<sup>[252]</sup> Kabsch's procedure was implemented in the CAST program package using a rotation-matrix-based procedure, which was shown to be formally equivalent to quaternion-based implementations.<sup>[253,254]</sup> The procedure according to Kabsch is known to work well for reasonably rigid molecules, but fails for conformationally too flexible molecules.<sup>[255]</sup> In our case, we have inspected the results of the alignment visually using computer-graphic-software rendering of the molecule and ensured proper rotational alignment.

In the next step, the principal components of the aligned simulation trajectories were evaluated using the PCA implementation as available in the CAST software, which is modeled after [190]. Every 100th frame of the concatenated super-trajectory was used for the computation of the covariance matrix in the PCA procedure. The trajectory was projected into the principal component representation. GM models were obtained using k-means clustering and expectation maximization as implemented in the armadillo 8.3 linear algebra package.<sup>[256–258]</sup> 20 multivariate gaussian functions were used in the GM models to fit the principal components with largest associated variance.

In a second step, we used the GM models based probability distribution to draw samples from it. Numerical integration of the GM model's entropy was performed using the `CUBATURE` code by Steven G. Johnson (see [259], based on [260] and [261]). As mentioned in Section 8.8, the nearest neighbor search is implemented using a parallelized naïve brute-force approach. In all plots, the functional form of the estimator according to Hnizdo *et al.* (Equation 161) is used.<sup>[226]</sup> The estimators given by Lombardi and Pant<sup>[214]</sup> as well as Gorla *et al.*<sup>[221]</sup> differ only with respect to the correction terms for the asymptotic bias of the estimator. In section 8.7.1 of this work it was shown that the difference in the bias-correction terms is negligible.

### 8.9.3. Results for the Alanine Dipeptide GM Model

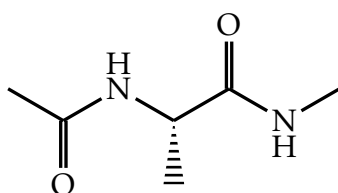


Figure 26: Lewis structure of the alanine dipeptide

In this section, we will compare entropy estimates from different approximations schemes for a GM model probability density built from MD simulations of the alanine dipeptide model. The simulation data was taken from [250, 251]. The alanine dipeptide is a common model compound used for estimating the goodness of computational modeling approaches towards biochemical protein systems, due to its structural simplicity.<sup>[262,263]</sup> Its full chemical IUPAC name is *N*-acetyl-L-alanine-*N*-methylamide. Its chemical structure is shown in Figure 26.

First, we will assess the performance of the quasi-harmonic approach. Then, the full-dimensional kNN estimate as well as truncated MIE approximations will be evaluated. In all cases our focus lies on establishing clear error-bars with respect to the reference value. To obtain an accurate reference entropy, we have performed numerical cubature integration of the entropy of the GM model distribution in a hypercubic region of the first 5 mass-weighted PCA modes. The histogrammed probability density given by the first two PCA modes is shown in Figure 28. As the domain of GM model probability densities is  $(-\infty, \infty)$ , a bounded domain for the integration needs to be specified and the GM model needs to be renormalized accordingly. The integration region was determined by assessing the MD trajectory in the principal component representation, assuring that every sample is included. This defines a hyper-rectangular region, which

was chosen as the domain for numerical integration. The reference value obtained via numerical integration in this case (including renormalization) is  $70.7 \frac{\text{J}}{\text{mol K}}$  for a hyper-rectangular domain of  $[-17 \frac{\text{g}}{\text{mol \AA}}, 17 \frac{\text{g}}{\text{mol \AA}}]$ . The reference value was obtained as the converged numerical integration estimate for increasingly smaller integration grids. It has to be noted that numerical grid-integration is computationally unfeasible for GM models with dimensionality  $d > 5$ . The validity and similarity of the GM

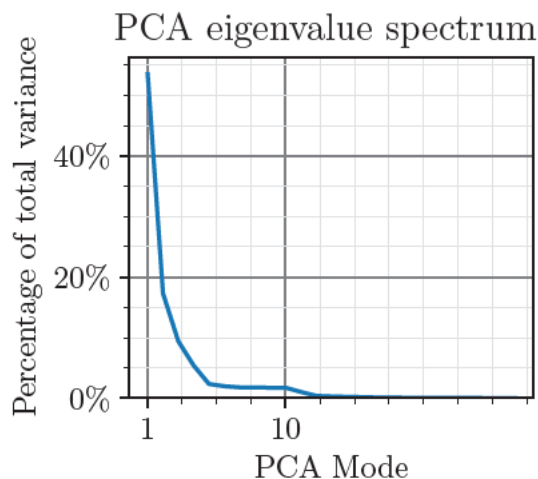


Figure 27: Spectrum of the PCA eigenvalues for a concatenated super-trajectory composed of 20 MD simulation replica of the alanine dipeptide taken from [250, 251].

model can be assessed by visually comparing the histogrammed probability density calculated from samples drawn from the GM model with the density of the PCA modes presented in Figure 29. The plot of the first two principal modes of the concatenated MD trajectories of the alanine dipeptide are given in Figure 28. It is clear that the multimodal and specifically non-gaussian structure of the data is retained faithfully in the GM model. This verifies that the GM benchmark system contains chemically-derived information. The performance of kNN entropy estimators should for this reason be able to be assessed using the model.

First, we investigate the performance of the QHA results for different sample sizes. For our purposes in this hermetic model systems, we will not perform the common QHA method including frequency assignment as proposed by Schlitter or Karplus.<sup>[184,185]</sup> The frequency-assignment-based QHA method can be effectively reduced to the computation of a differential entropy from the covariance matrix within classical statistical mechanics, assuming the probability density to be a gaussian. For this reason, the method corresponding to the QHA approximation for MD simulation data in this GM model benchmark is the calculation of the entropy under the assumption of

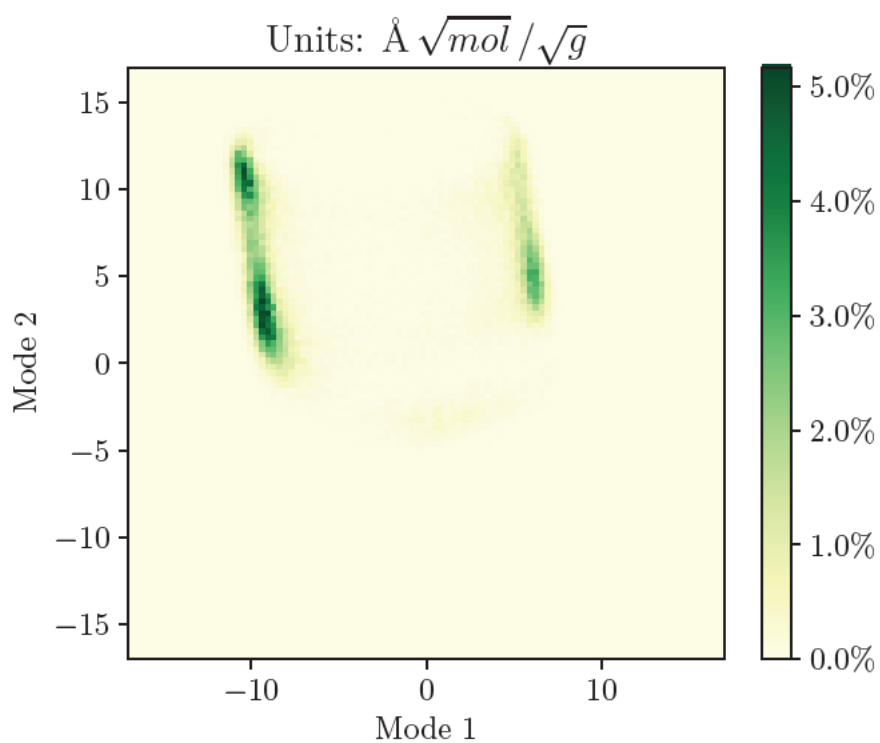


Figure 28: Histogrammed probability density obtained from the simulation data of the alanine dipeptide. Only the two Principal Components with the highest associated variance are plotted.



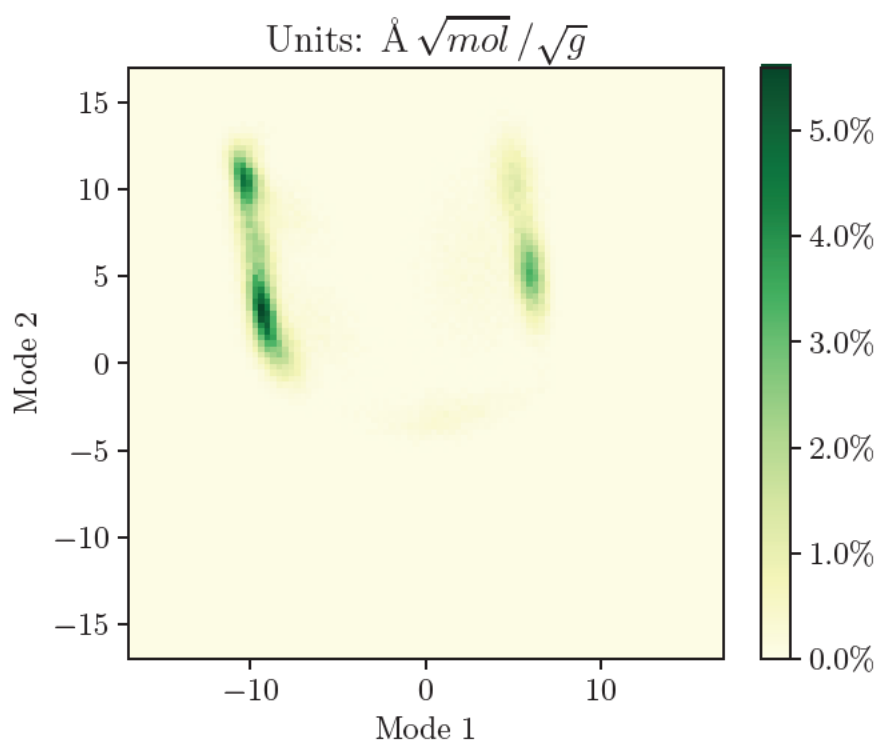


Figure 29: Histogrammed probability density obtained from the GM Model of the alanine dipeptide. Only the two modes with the highest associated variance are plotted.

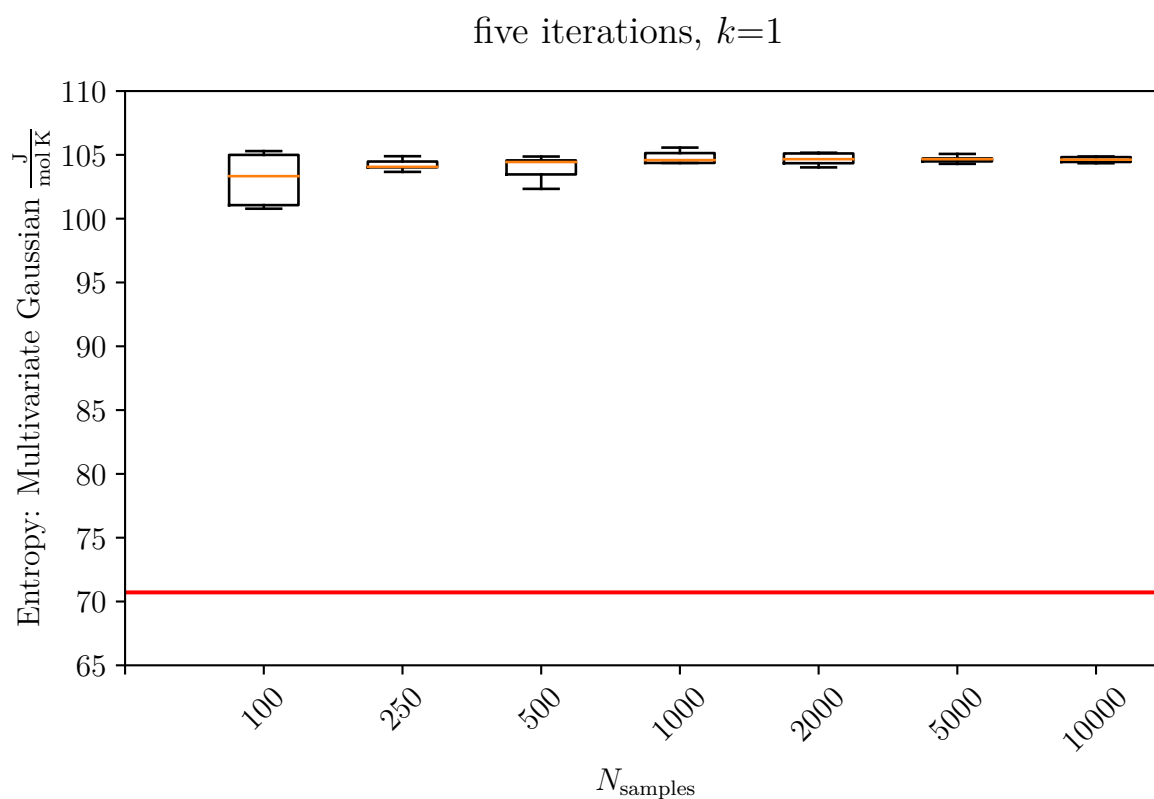


Figure 30: Convergence of the multivariate gaussian estimate of the differential entropy of the GM model system for the alanine dipeptide. A box plot is shown for 5 independent iterations of the benchmark in order to give further insights into the convergence of the quasi-harmonic entropy estimate. The numerical integration estimate of the differential entropy, which we assume to be exact, is plotted as a red line.

each dimension being distributed according to a normal distribution. In this case, a multi-variate version of the formula derived in Section 8.5.4 is used. This is similar in spirit to the original quasi-harmonic entropy estimation method proposed by Karplus and Kushick in 1981 (see [183]), which is itself based on the entropy of a multivariate gaussian rooted strictly in the spatial domain. Since the sampling method chosen for our benchmark (see Listing 2) produces uncorrelated samples, rapid convergence of the standard deviation of each dimension and thus of the multivariate gaussian estimate with respect to sampling size is expected. It is important to keep in mind that the samples from an MD simulation trajectory are correlated in the sense that each subsequent sample originates from a modification of the previous one. Sampling in MD simulations is thus local, and might not capture the entire configuration space accurately for finite MD simulation lengths. The rapid convergence expected in the case of the present GM model benchmark is found as expected, as Figure 30 shows. Figure 30 displays boxplots for five separately drawn and unique sample sizes for each value of  $N_{\text{samples}}$ . The multivariate gaussian estimate converges rapidly, overestimating the cubature integration value by  $34 \frac{\text{J}}{\text{molK}}$ . Only 125 samples are sufficient in this 5-dimensional model to obtain a sufficiently converged, albeit biased, result.

We will now consider the results for the full-dimensional kNN entropy estimates. In case of this study's model system for the alanine dipeptide, the full dimensionality is  $d = 5$  and the evaluation of both exact entropies from cubature integration and kNN entropy estimates is still possible at acceptable computational costs. The parameters to be assessed and benchmarked for the kNN entropy estimator are the number of samples  $N_{\text{samples}}$ , the neighbor parameter  $k$  and the norm ( $\ell_2$  or  $\ell_\infty$ ) used for the calculation of nearest neighbor distances. For the full-dimensional entropy estimate using the euclidean ( $\ell_2$ ) norm, a clear correlation of the error with respect to sampling size  $N_{\text{samples}}$  and  $k$  is visible (see Figure 31b). We only report sample sizes up until 10000 datapoints. For all larger datasets, the kNN entropy estimate is fully converged. Interestingly and perhaps unexpectedly, low numbers of  $k$  seem to provide much better estimates in all cases, but especially for small sampling sizes. We obtain virtually identical results when invoking either the  $\ell_\infty$  norm or the usual euclidean  $\ell_2$  distance norm in the kNN entropy expression (compare Figure 31 and Figure 32). Another remarkable finding is that for uncorrelated and unbiased samples, already 500 samples can be enough to yield an accurate and converged kNN estimation of the entropy. The kNN estimate using the  $\ell_\infty$  norm,  $N_{\text{samples}} = 1000$ , and  $k = 1$  yields an

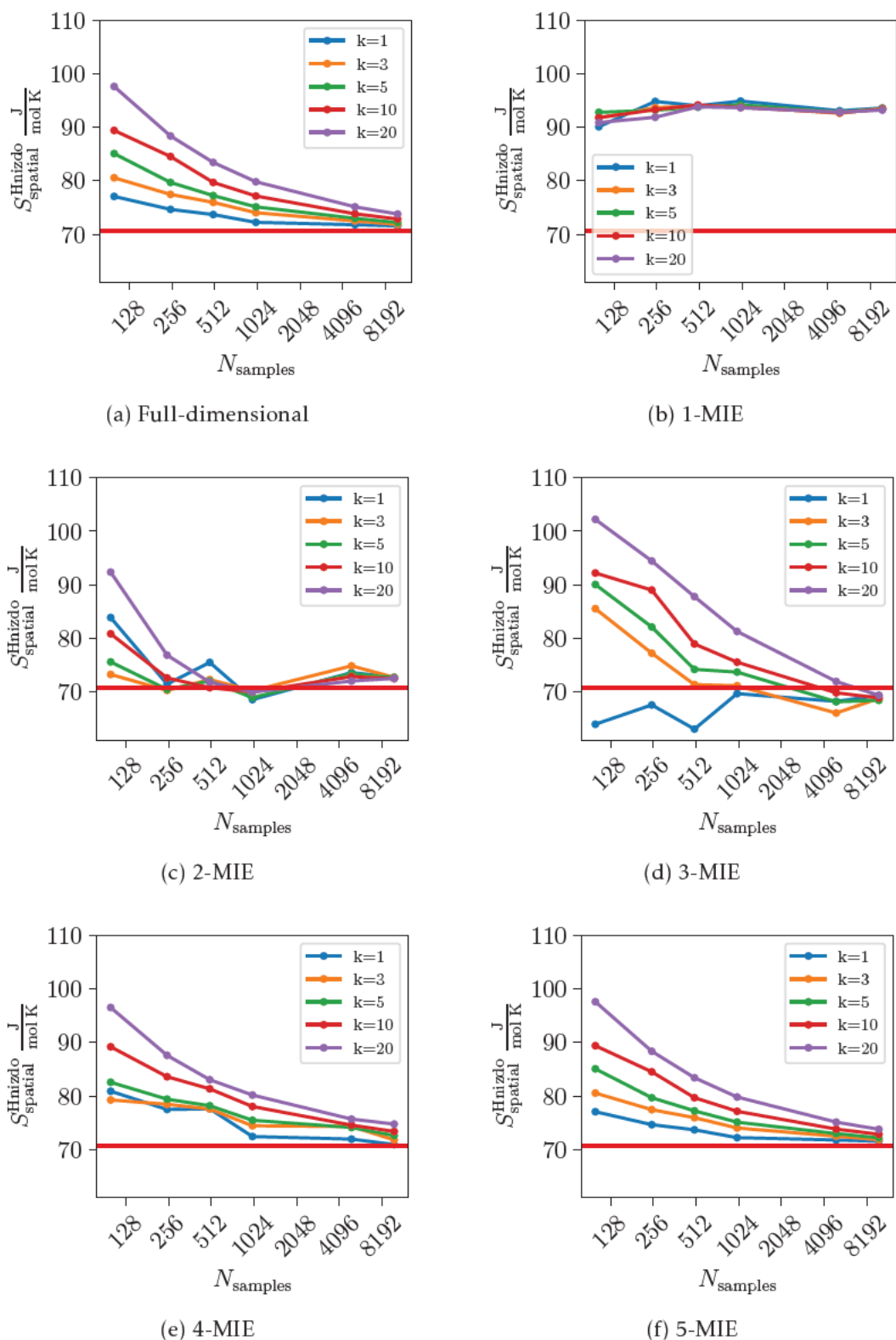


Figure 31: kNN entropy estimates for the 5-dimensional GM model of the alanine dipeptide, using the  $\ell_2$  distance norm. The exact integration value is plotted as a black line.

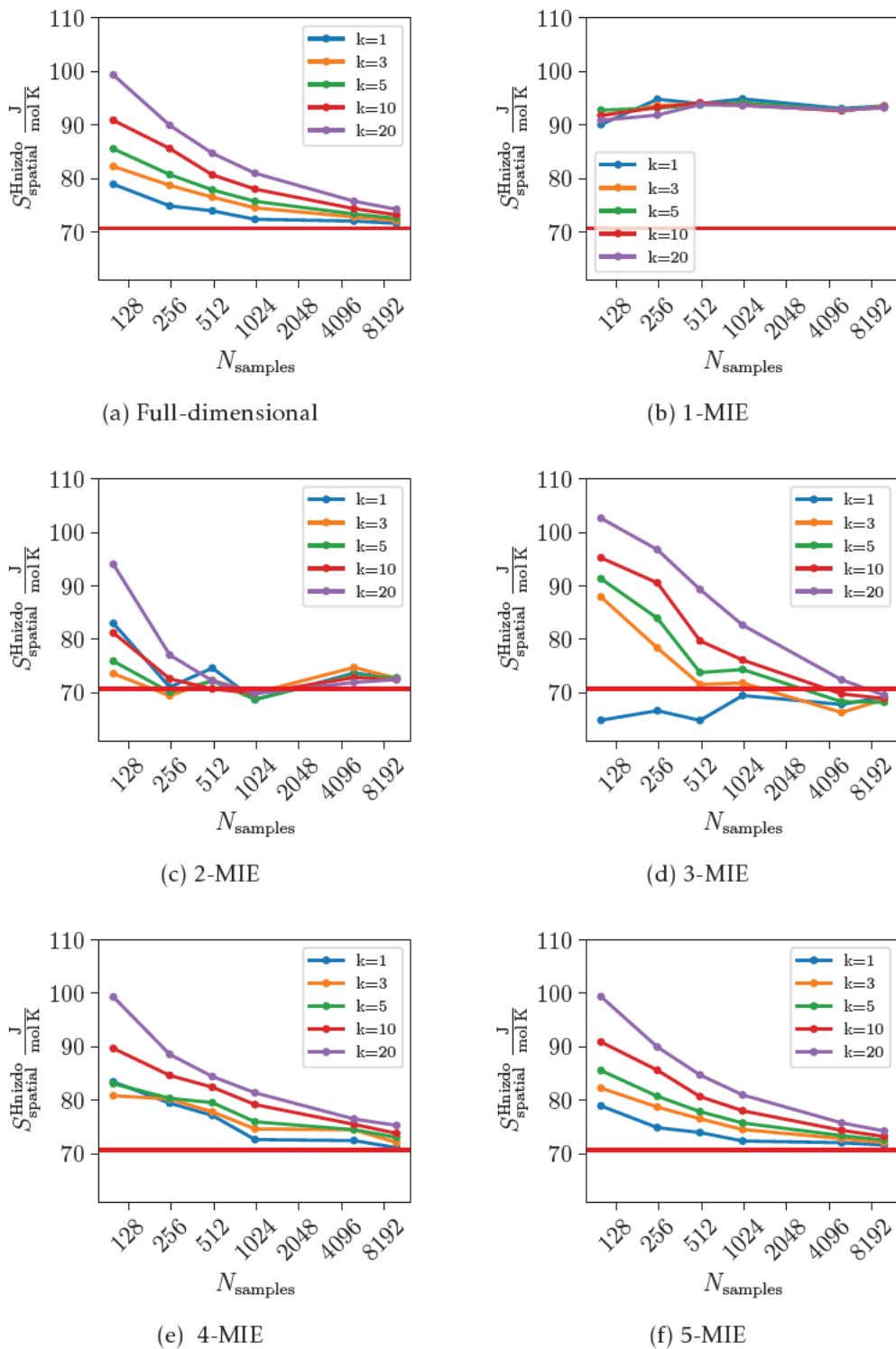


Figure 32: kNN entropy estimates for the 5-dimensional GM model of the alanine dipeptide, using the  $\ell_{\infty}$  distance norm. The exact integration value is plotted as a black line.

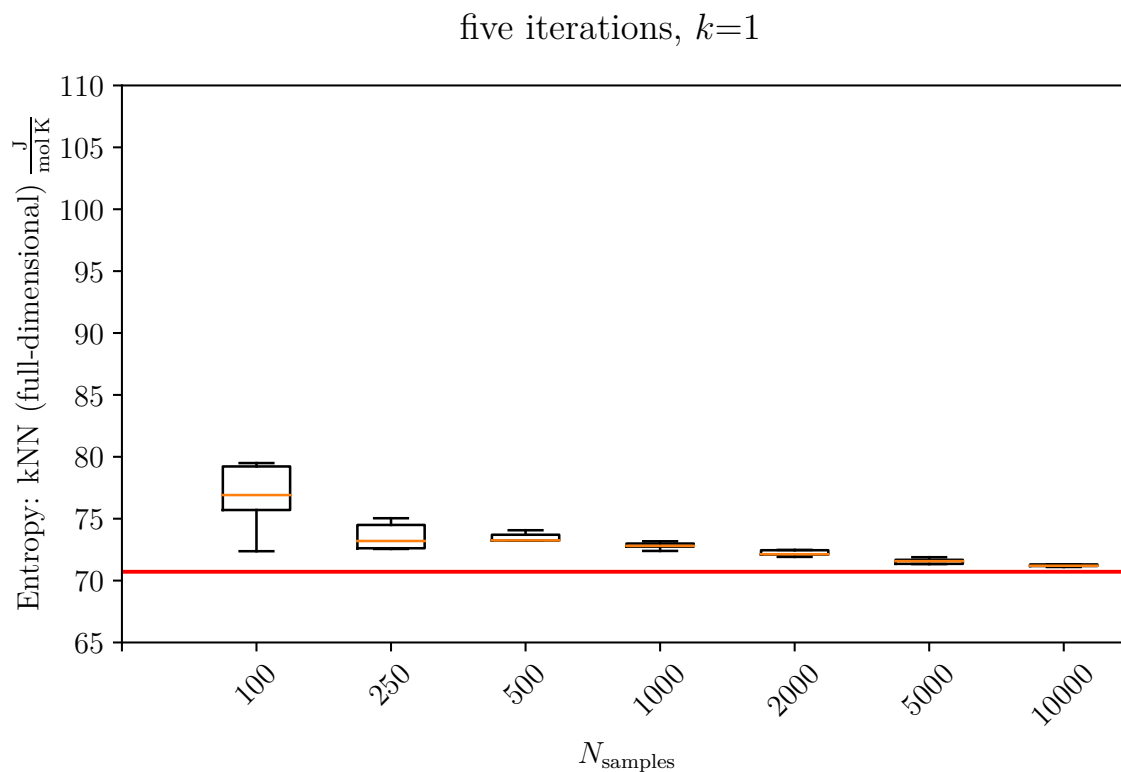


Figure 33: Convergence of the full-dimensional kNN entropy estimate for the 5-dimensional GM Model of the alanine dipeptide, using the  $\ell_2$  distance norm and the bias correction according to Hnizdo. The exact integration value is plotted as a red line.

estimate with a low relative error with respect to the cubature integration estimate. At this point we cannot specify whether this error compensation results because at high  $k$  and  $N_{\text{samples}}$ , numerical errors are accumulated, or if this is an underlying property of the estimator. As previously mentioned, other reports have suggested that the usage of  $k$  values larger than  $k = 1$  yields improved results and smaller errors.<sup>[14,154,219]</sup> In our benchmark, we do not find a justification for choosing  $k > 1$ .

In practice, the full-dimensional kNN entropy can sometimes not be obtained for high-dimensional systems due to the increased computational costs. To remedy this, the MIE can be used. Figure 31b depicts the 1<sup>st</sup> order MIE of the kNN entropy. The physical meaning of this procedure is correcting harmonic modes for anharmonicity separately. The dependence on the number of samples is not significant for the 1<sup>st</sup> order MIE. As seen in Figure 31b, for  $N_{\text{samples}} \geq 500$  a systematically biased kNN estimate is obtained in a converged manner for values of  $k > 5$ . Fluctuations are still present for  $k < 5$ . For smaller sample sizes, larger fluctuations are calculated. When the sample size is sufficient, the impact of  $k$  on the estimate seems to be negligible. It is of interest to check whether one is gaining overall accuracy by using the 1<sup>st</sup> order MIE in comparison to the multivariate gaussian estimate (displayed in Figure 30). This is the case, as for all possible combinations of  $k$  and  $N_{\text{samples}}$ , the 1<sup>st</sup> order MIE entropy estimate is closer to the cubature integration value than the multivariate gaussian estimate. This holds even for estimates using a low number of samples and low  $k$ , where fluctuations are occurring. For the  $\ell_\infty$  norm, all these conclusions hold as well. In line with theory, we observe that the corrections of the 1<sup>st</sup> order MIE are strictly negative and lower the absolute entropy value with respect to the multivariate gaussian estimate. The 1<sup>st</sup> order MIE correction, which is the sum of marginal (1-dimensional) entropies, is known to be a strict upper bound to the entropy.<sup>[226]</sup> In the limit of large  $k$  and  $N_{\text{samples}}$  values, the 1<sup>st</sup> order MIE estimate converges to a systematically biased value, which deviates by approx. 1.8% (at  $N_{\text{samples}} = 10000$  &  $k = 5$ ) from the cubature integration value.

Figure 31c and Figure 32c show the 2<sup>nd</sup> order MIE. Formally, 2<sup>nd</sup> order MIE contributions always are always positive and thus strictly lower the 1<sup>st</sup> order MIE entropy estimate. As the data of our benchmark is based on principal modes of MD simulations, linear correlation between any two dimensions has already been eliminated. Moving from the 1<sup>st</sup> order MIE to the 2<sup>nd</sup> order MIE thus leads to the inclusion of all supralinear correlation between each two dimensions into the entropy estimate. As visible in Figure 31c and Figure 32c, the behavior of the estimator is noisy and

trends are less predictable than the full-dimensional estimator at low values of  $k$  and  $N_{\text{samples}}$ . However, minding the scale of the plot, it becomes clear that both the converged estimate as well as the overall fluctuations are much lower than the 1<sup>st</sup> order MIE results in all cases. These results hold for estimations using the  $\ell_\infty$  distance norm as well. At large values of  $k$  and  $N_{\text{samples}}$ , the estimate converges to a value deviating from the cubature integration estimate by only 0.8%, which is significantly less than the bias of the multivariate gaussian estimate. Numata and Knapp have suggested using the 2<sup>nd</sup> order MIE to gain systematically more accurate entropies than using a multivariate gaussian estimate,<sup>[154]</sup> and the present results prove this to be viable. Since computational costs rise significantly when going to higher orders of the MIE, the 2<sup>nd</sup> order MIE of kNN entropy is a suitable candidate for reliable routine application. The choice of  $k = 4$  proposed previously by Knapp *et al.*,<sup>[154]</sup> in agreement with statements by Fukunaga in [219] that  $k$  should not be chosen too low, is a reasonable choice with respect to calculation time and goodness of the estimate. Nevertheless, in the case of this model system,  $k = 1$  resulted as the most accurate parameter for the estimator, yielding the fastest computation times as well.

Considering the 3<sup>rd</sup> and 4<sup>th</sup> order MIE, it is important to note that the underlying probability density of the GM model has only five dimensions. The utility of high order MIEs for this rather low-dimensional probability density is limited. As showcased in Figure 31d-f and Figure 32d-f, no significant overall accuracy is gained in the higher order MIEs in comparison to the 2<sup>nd</sup> order MIE. For this particular alanine dipeptide probability density, the MIE visibly converges to the (regular) full-dimensional kNN entropy estimate. Divergence of the MIE – as discovered by Goethe *et al.* in [209] – is not observed in our test cases.

The benchmark for the alanine dipeptide indicates that significant accuracy can be gained by using the 2<sup>nd</sup> order MIE of the NN entropy in comparison to the multivariate gaussian estimate. Already the 1<sup>st</sup> order MIE has some impact, reducing the error bar. We note that the increased computational cost of higher order MIE estimates renders their usage troublesome. Reassuringly, the full-dimensional estimate is increasingly regained when including higher order MIE terms in case of this benchmark. The impact of the utilized distance norm ( $\ell_2$  or  $\ell_\infty$ ) in the kNN distance search is found to be negligible and all presented results hold for both the euclidean and the  $\ell_\infty$  norm. The finding that the maximum norm may be used for chemical entropy estimation is good news: Computationally and mathematically, the choice of the maximum norm often leads to simpler frameworks, as hyperelliptic regions in space become hyper-



rectangular when switching from the  $\ell_2$  to the  $\ell_\infty$  norm. Lombardi and Pant have proposed a more sophisticated estimator based on the kNN entropy estimator, which requires the integration of a multivariate gaussian inside a finite multidimensional region. While this is feasible of hyperrectangular regions ( $\ell_\infty$  norm), no proposal has been made for hyperelliptic regions ( $\ell_2$  norm).<sup>[214]</sup>

## 9. Summary

In the course of this work, we have discussed computational approaches for modeling several non-trivial examples of molecular or spectroscopic problems.

First, the spectroscopic properties of the biradical(oid) molecules methylbismuth, diphenylacetylene, pentadiynylidene (including its methyl-substituted derivatives), and diphenylpropynylidene were investigated in a joint experimental and theoretical effort. The results are found in Sections 3, 6, 5 & 4. Computational approaches were used to estimate ionization energies (IEs) of the molecules. Furthermore, the absorption spectra upon ionization were calculated using quantum chemical protocols in the Franck-Condon approximation. For methylbismuth, scalar-relativistic and ECP methods were used. By analysis of the excited electronic states and their geometry-dependent energetic properties, the femtosecond pump-pulse transient absorption spectrum of diphenylpropynylidene was characterized. The photochemistry of diphenylacetylene was investigated by considering the bending-motion of the molecule using the  $\omega$ B97xD DFT functional. For the pentadiynylidene molecule and its methyl substituted derivatives, a potential pitfall of DFT based modeling of organic biradical-type molecules with MR character was identified: We found that DFT based approaches work well for each given conformational structure of a molecule, but tend to yield misleading geometric minima in optimization procedures. Once the correct structure is known and has been obtained, DFT-approaches often yield accurate results even in cases where the electronic structure of the molecule has some MR character. The same finding has been previously established by us for the *ortho*-benzyne cation.<sup>[80]</sup> By comparison of Franck-Condon simulated absorption spectra of different structural isomers of pentadiynylidene, possible isomerization pathways and their resulting molecular species after electron detachment (ionization) in the experiment were considered. Finally, the structural isomers observed in the spectroscopic experiment were assigned by comparison of the computational and experimental data.

In summary, biradical(oid) molecules are seen to be at times difficult to analyze using computational approaches due to their open-shell electronic structure. Computationally efficient DFT methods may lead to pitfalls for these molecules, and comparisons with MR computational approaches are necessary to validate their results. One important finding is that the impact of the correct modeling of the geometric structure is adamant in computational modeling of biradicals.

In a second research project, the impact of molecular geometry on the localization of

excitons in Organic Semiconductor (OSC) materials was investigated (see Section 2). As OSC materials often possess large conjugated and delocalized  $\pi$ -electron systems, their size may well be too large to employ highly accurate methodologies in quantum-chemistry. In this research effort, we highlighted the effect of geometry relaxation in the excited state geometry on the excitonic properties of molecular dimers and larger aggregates. In our study on the Perylene Bisimide (PBI) molecule, the effects can be seen to be of significant magnitude and thus care should be taken if geometrical properties of the excited state are neglected in computational schemes.

The bioconjugation reaction of Triazolinedione-molecules was investigated using quantum chemical approaches (see Section 7). The reaction mechanism has been discussed as an “ene-type” reaction in the literature. Our results identify a base-mediated reaction mechanism as the most plausible candidate from thermodynamical and kinetic considerations. In the course of the work, multiple other reaction mechanisms’ energetics have been calculated as well. For the calculations, results from the high-level single-reference method CCSD(T) and the MR CASPT2 approach were compared. In comparison, the  $\omega$ B97xD DFT method yields sufficient accuracy at reduced computational costs.

Lastly, the kNN entropy estimation procedure for extracting absolute molecular entropies from MD simulation trajectories was investigated in detail (see Section 8). Using artificially-built model systems, we showed that the estimator in principle may perform very well, as it is able to yield accurate entropy estimates in the case of an 5-dimensional GM model of the alanine dipeptide. Additionally, we gave an overview of several sources of error in the kNN entropy estimator and show how they derive from its general mathematical framework. First, we analyzed the different proposed asymptotic bias-correcting functions. Secondly, we discussed the additivity of the kNN entropy estimator for uncorrelated independent probability distributions. We exemplified how the known problem of slow convergence of the kNN entropy estimator stems from uneven variances of the dimensions of the data, and proposed correction-schemes to remedy this.

In summary, this thesis tackled multiple problems in computational modeling of chemical substances, gave an overview over potential problems and pitfalls and suggested ways to mitigate the shortcomings, by applying the approaches to chemically-derived data.

## 10. Zusammenfassung

Im Rahmen dieser Arbeit wurden Berechnungsansätze für die Modellierung mit Standardverfahren näherungsweise berechenbarer molekularer und spektroskopischer Probleme diskutiert. Zunächst wurden die spektroskopischen Eigenschaften der biradikalischen Moleküle Methylbismut, Diphenylacetylen, Pentadiynylidin (sowie dessen methylsubstituierte Derivate) und Diphenylpropynyliden untersucht. Diese Forschungsvorhaben wurden in Zusammenarbeit mit experimentell arbeitenden physikalischen und synthetischen Chemikern durchgeführt. Der Vergleich von sowohl spektroskopisch als auch durch theoretische Rechnungen erlagter Erkenntnisse erlaubt eine genaue Charakterisierung der physiko-chemischen Eigenschaften der Moleküle. Mit Hilfe von computergestützter Modellierung wurden die Ionisierungsenergien und die Absorptionsspektren der Ionisierung in der Franck-Condon-Approximation berechnet. Für Methylbismut wurden skalar-relativistische und Pseudopotenzial-basierte Ansätze zur Berechnung verwendet. Durch Analyse der angeregten elektronischen Zustände und ihrer geometrieabhängigen energetischen Eigenschaften, konnte das transiente Absorptionsspektrum von Diphenylpropynyliden genau charakterisiert und rationalisiert werden. Die Photochemie von Diphenylacetylen wurde unter Berücksichtigung der Biegeschwingungsmode des Moleküls mit Hilfe des  $\omega$ B97xD DFT Funktionals untersucht. Für das Pentadiynyliden-Molekül und seine Methyl-substituierten Derivate konnte eine potenzielle Fehlerquelle bei der DFT-basierten Modellierung organischer Moleküle mit Biradikal-Charakter identifiziert werden: Wir haben festgestellt, dass DFT-basierte Ansätze für eine gegebene Konformation eines Moleküls hinreichend gute Ergebnisse liefert. Diese Methoden neigen jedoch dazu, irreführende falsche geometrische Minima in Optimierungsverfahren vorherzusagen. Sobald die richtige Struktur bekannt ist, liefern selbst DFT-Ansätze oft genaue Ergebnisse, selbst wenn die elektronische Struktur des Moleküls einen gewissen Multireferenzcharakter aufweist. Die gleiche Erkenntnis wurde von uns in einer vorherigen Forschungsarbeit bereits für das *ortho*-Benzin Kation festgestellt und ist mutmaßlich charakteristisch für die Computermodellierung von Biradikalmolekülen.<sup>[80]</sup> Durch Vergleich von Franck-Condon simulierten Absorptionsspektren verschiedener Strukturisomere des Pentadiynylidens wurden mögliche Isomerisierungspfade und deren molekulare Produkte nach der Elektronenabspaltung (Ionisierung) im Experiment evaluiert. Die im spektroskopischen Experiment beobachteten molekularen Spezies konnten durch Vergleich der rechnerischen und experimentellen Daten erfolgreich zugeordnet werden. Zusammenfassend

lässt sich sagen, dass biradikalische Moleküle aufgrund ihrer offenschaligen elektronischen Struktur mitunter nur schwer durch die Routinemethoden quantenchemischer Computersimulationen charakterisiert werden können. Die rechnerisch effizienten DFT-Methoden können bei diesen Molekülen zu fehlerbehafteten Vorhersagen führen, und Vergleiche mit Multireferenz-Rechenverfahren sind notwendig, um Ergebnisse zu validieren. Eine wichtige Erkenntnis ist, dass der Einfluss der korrekten Modellierung der geometrischen Struktur bei der computergestützten Modellierung von Biradikalen von entscheidender Bedeutung ist.

In einem zweiten Forschungsprojekt wurde der Einfluss der Molekulargeometrie auf die Lokalisierung von Exzitonen in organischen Halbleitermaterialien untersucht. Da organische Halbleitermaterialien oft delokalisierte  $\pi$ -Elektronensysteme besitzen, ist ihre Elektronenanzahl meist zu groß um hochgenaue Methoden der Quantenchemie anzuwenden. In dieser Forschungsarbeit untersuchen wir die Auswirkungen der Geometrie-Relaxation im angeregten Zustand auf die exzitonischen Eigenschaften von molekularen Dimeren und größeren Aggregaten. In der als Teil dieser Arbeit inkludierten Forschungsstudie über das PBI-Molekül zeigt sich, dass die Auswirkungen von signifikanter Größenordnung sind und daher Vorsicht geboten ist, wenn die geometrischen Eigenschaften des angeregten Zustands in Berechnungsschemata vernachlässigt werden.

Die Biokonjugationsreaktion von TAD-Molekülen mit der Aminosäure Tyrosin wurde mit quantenchemischen Ansätzen untersucht. Der Reaktionsmechanismus wurde in der Fachliteratur bisher schlicht als En-Reaktion diskutiert. Unsere Ergebnisse identifizieren einen Basen-vermittelten, durch Deprotonierung der Phenol-Gruppe des Tyrosins eingeleiteten Reaktionsmechanismus als den plausibelsten Kandidaten. Diese Rationalisierung stützt sich auf thermodynamische und kinetische Überlegungen auf Basis der quantenchemischen Modellierung. Im Laufe der Arbeit wurden auch die Extrema der Potenzialhyperflächen mehrerer weiterer, verwandter Reaktionsmechanismen berechnet. Für die Berechnungen wurden die Ergebnisse der sehr exakten Single-Reference-Methode CCSD(T) und des MR-Ansatzes CASPT2 verglichen. Die  $\omega$ B97xD DFT-Methode liefert im Vergleich eine ausreichende Genauigkeit bei geringerem Rechenaufwand.

Zuletzt wurde das Verfahren der kNN-Entropieschätzung zur Berechnung absoluter molekularer Entropien aus Molekulardynamiksimulationen im Detail untersucht. Anhand von speziell erzeugten Test-Modellsystemen konnte gezeigt werden, dass die Schätzungsmethode prinzipiell sehr gut funktionieren kann, da sie im Falle eines 5-

dimensionalen GM-Modells des Alanin Dipeptids genaue Entropieschätzungen liefern kann. Darüber hinaus wurde ein Überblick über verschiedene Fehlerquellen im kNN-Entropieschätzer gegeben und gezeigt, wie diese sich allgemein mathematisch ableiten lassen. Zunächst wurden die verschiedenen vorgeschlagenen asymptotischen Korrekturfunktionen analysiert. Zweitens wird die Additivität des kNN-Entropieschätzers für unkorrelierte und unabhängige Wahrscheinlichkeitsverteilungen analysiert. Es wurde veranschaulicht, wie das bekannte Problem der langsamen Konvergenz des kNN Entropieschätzers auf ungleiche Varianzen der Dimensionen der Daten zurückzuführen ist. Korrekturschemata wurden vorgeschlagen, um dieses Problem zu lösen. Zusammenfassend thematisierte diese Arbeit mehrere Probleme bei der computergestützten Modellierung chemischer Fragestellungen. Es wurde ein Überblick über potenzielle Probleme gegeben und Möglichkeiten, die Unzulänglichkeiten bekannter Methoden durch die Anwendung spezialisierter, moderner Ansätze zu mildern, vorgeschlagen.

## 11. References

1. S. Grimme, P. R. Schreiner, *Angewandte Chemie International Edition* **2018**, *57*, 4170–4176.
2. K. Burke, *The Journal of Chemical Physics* **2012**, *136*, 150901.
3. B. O. Roos, R. Lindh, P. A. Malmqvist, V. Veryazov, P.-O. Widmark in *Multiconfigurational Quantum Chemistry*, Hoboken: John Wiley & Sons, **2016**, Chapter 1.
4. P. Pulay, *International Journal of Quantum Chemistry* **2011**, *111*, 3273–3279.
5. D. P. Goldenberg, *Journal of Molecular Biology* **2003**, *326*, 1615–1633.
6. K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, D. E. Shaw, *Journal of the American Chemical Society* **2012**, *134*, 3787–3791.
7. A. Shehu in *Introduction to Protein Structure Prediction*, Hoboken: John Wiley & Sons, Ltd, **2010**, Chapter 19.
8. C. Brückner, F. Würthner, K. Meerholz, B. Engels, *The Journal of Physical Chemistry C* **2017**, *121*, 4–25.
9. C. Brückner, F. Würthner, K. Meerholz, B. Engels, *The Journal of Physical Chemistry C* **2017**, *121*, 26–51.
10. V. Settels, W. Liu, J. Pflaum, R. F. Fink, B. Engels, *Journal of Computational Chemistry* **2012**, *33*, 1544–1553.
11. C. Brueckner, B. Engels, *Journal of Computational Chemistry* **2016**, *37*, 1335–1344.
12. C. Brückner, B. Engels, *The Journal of Physical Chemistry A* **2015**, *119*, 12876–12891.
13. D. Suárez, N. Díaz, *WIREs Computational Molecular Science* **2015**, *5*, 1–26.
14. R. Topolnicki, F. Briec, C. Schran, D. Marx, *Journal of Chemical Theory and Computation* **2020**, *16*, 6785–6794.
15. U. Hensen, O. F. Lange, H. Grubmüller, *PloS ONE* **2010**, *5*, e9179.
16. G. P. Hammond, P. J. Pearson, *Energy Policy* **2013**, *52*, 1–9.
17. M. A. Delucchi, M. Z. Jacobson, *Energy Policy* **2011**, *39*, 1170–1190.
18. V. May, O. Kühn, *Charge and energy transfer dynamics in molecular systems*, Wiley-VCH Verlag GmbH & Co. KGaA, **2008**.
19. A. Köhler, H. Bässler, *Electronic processes in organic semiconductors: An introduction*, Wiley-VCH Verlag GmbH & Co. KGaA, **2015**.
20. C. W. Tang, *Applied Physics Letters* **1986**, *48*, 183–185.

21. M. Zhang, L. Zhu, G. Zhou, T. Hao, C. Qiu, Z. Zhao, Q. Hu, B. W. Larson, H. Zhu, Z. Ma, et al., *Nature communications* **2021**, *12*, 1–10.
22. M. A. Green, E. D. Dunlop, J. Hohl-Ebinger, M. Yoshita, N. Kopidakis, X. Hao, *Progress in Photovoltaics: Research and Applications* **2021**, *29*, 657–667.
23. B. Weinberger, M Akhtar, S. Gau, *Synthetic Metals* **1982**, *4*, 187–197.
24. D. Kearns, M. Calvin, *The Journal of Chemical Physics* **1958**, *29*, 950–951.
25. G Yu, A. J. Heeger, *Journal of Applied Physics* **1995**, *78*, 4510–4515.
26. G. Yu, J. Gao, J. C. Hummelen, F. Wudl, A. J. Heeger, *Science* **1995**, *270*, 1789–1791.
27. T. Ameri, G. Dennler, C. Lungenschmied, C. J. Brabec, *Energy & Environmental Science* **2009**, *2*, 347–363.
28. T. Ameri, P. Khoram, J. Min, C. J. Brabec, *Advanced Materials* **2013**, *25*, 4245–4266.
29. J. Liu, L. Qiu, S. Shao, *J. Mater. Chem. C* **2021**, *9*, 16143–16163.
30. C. J. Brabec, S. Gowrisanker, J. J. M. Halls, D. Laird, S. Jia, S. P. Williams, *Advanced Materials* **2010**, *22*, 3839–3856.
31. J. Hou, O. Inganäs, R. H. Friend, F. Gao, *Nature Materials* **2018**, *17*, 119–128.
32. S. Günes, H. Neugebauer, N. S. Sariciftci, *Chemical Reviews* **2007**, *107*, 1324–1338.
33. X. Yang, J. Loos, *Macromolecules* **2007**, *40*, 1353–1362.
34. H.-G. Wagemann, H. Eschrich, H.-G. Wagemann, *Photovoltaik: Solarstrahlung und Halbleitereigenschaften*, B.G. Teubner Verlag & GWV Fachverlage GmbH: Wiesbaden, **2007**.
35. S. R. Cowan, N. Banerji, W. L. Leong, A. J. Heeger, *Advanced Functional Materials* **2012**, *22*, 1116–1128.
36. J.-L. Brédas, D. Beljonne, V. Coropceanu, J. Cornil, *Chemical Reviews* **2004**, *104*, 4971–5004.
37. V. Coropceanu, J. Cornil, D. A. da Silva Filho, Y. Olivier, R. Silbey, J.-L. Brédas, *Chemical Reviews* **2007**, *107*, 926–952.
38. J. Cornil, D. Beljonne, J.-P. Calbert, J.-L. Brédas, *Advanced Materials* **2001**, *13*, 1053–1067.
39. J.-L. Brédas, J. E. Norton, J. Cornil, V. Coropceanu, *Accounts of chemical research* **2009**, *42*, 1691–1699.
40. J.-L. Brédas, J. P. Calbert, D. da Silva Filho, J. Cornil, *Proceedings of the National Academy of Sciences* **2002**, *99*, 5804–5809.



41. Y. Zhou, C. Fuentes-Hernandez, J. Shim, J. Meyer, A. J. Giordano, H. Li, P. Winget, T. Papadopoulos, H. Cheun, J. Kim, *Science* **2012**, 336, 327–332.
42. D. Fazzi, M. Barbatti, W. Thiel, *The Journal of Physical Chemistry Letters* **2017**, 8, 4727–4734.
43. T. Wang, T. R. Kafle, B. Kattel, W.-L. Chan, *Journal of the American Chemical Society* **2017**, 139, 4098–4106.
44. S. Verlaak, D. Beljonne, D. Cheyons, C. Rolin, M. Linares, F. Castet, J. Cornil, P. Heremans, *Advanced Functional Materials* **2009**, 19, 3809–3814.
45. V. Arkhipov, P. Heremans, H. Bässler, *Applied Physics Letters* **2003**, 82, 4605–4607.
46. F. Castet, G. D’Avino, L. Muccioli, J. Cornil, D. Beljonne, *Physical Chemistry Chemical Physics* **2014**, 16, 20279–20290.
47. L. Shi, C. K. Lee, A. P. Willard, *ACS Central Science* **2017**, 3, 1262–1270.
48. Y. Pan, W. Li, S. Zhang, L. Yao, C. Gu, H. Xu, B. Yang, Y. Ma, *Advanced Optical Materials* **2014**, 2, 510–515.
49. P. Borsenberger, L. Pautmeier, H. Bässler, *The Journal of chemical physics* **1991**, 94, 5447–5454.
50. H. Bässler, *Physica Status Solidi B (Basic Research)* **1993**, 175.
51. H. Bässler, A. Köhler, *Physical Chemistry Chemical Physics* **2015**, 17, 28451–28462.
52. C. Poelking, D. Andrienko, *Journal of the American Chemical Society* **2015**, 137, 6320–6326.
53. C. Poelking, M. Tietze, C. Elschner, S. Olthof, D. Hertel, B. Baumeier, F. Würthner, K. Meerholz, K. Leo, D. Andrienko, *Nature materials* **2015**, 14, 434–439.
54. P. Kordt, J. J. van der Holst, M. Al Helwi, W. Kowalsky, F. May, A. Badinski, C. Lennartz, D. Andrienko, *Advanced Functional Materials* **2015**, 25, 1955–1971.
55. B. Baumeier, F. May, C. Lennartz, D. Andrienko, *Journal of Materials Chemistry* **2012**, 22, 10971–10976.
56. R. F. Fink, J. Seibt, V. Engel, M. Renz, M. Kaupp, S. Lochbrunner, H.-M. Zhao, J. Pfister, F. Würthner, B. Engels, *Journal of the American Chemical Society* **2008**, 130, 12858–12859.
57. W. Liu, V. Settels, P. H. P. Harbach, A. Dreuw, R. F. Fink, B. Engels, *Journal of Computational Chemistry* **2011**, 32, 1971–1981.

58. A. Schubert, V. Settels, W. Liu, F. Würthner, C. Meier, R. F. Fink, S. Schindlbeck, S. Lochbrunner, B. Engels, V. Engel, *The Journal of Physical Chemistry Letters* **2013**, *4*, 792–796.
59. D. Bellinger, J. Pflaum, C. Brüning, V. Engel, B. Engels, *Physical Chemistry Chemical Physics* **2017**, *19*, 2434–2448.
60. C. Brückner, B. Engels, *Journal of Computational Chemistry* **2016**, *37*, 1335–1344.
61. M. Deutsch, S. Wirsing, D. Kaiser, R. F. Fink, P. Tegeder, B. Engels, *The Journal of Chemical Physics* **2020**, *153*, 224104.
62. I. B. Bersuker, *Chemical Reviews* **2021**, *121*, 1463–1512.
63. V. Stehr, R. F. Fink, B. Engels, J. Pflaum, C. Deibel, *Journal of Chemical Theory and Computation* **2014**, *10*, 1242–1255.
64. F. Plasser, *The Journal of Chemical Physics* **2020**, *152*, 084108.
65. C. J. Bardeen, *Annual review of physical chemistry* **2014**, *65*, 127–148.
66. C. F. Goldsmith, G. R. Magoon, W. H. Green, *The Journal of Physical Chemistry A* **2012**, *116*, 9033–9057.
67. A. Rimola, D. Skouteris, N. Balucani, C. Ceccarelli, J. Enrique-Romero, V. Taquet, P. Ugliengo, *ACS Earth and Space Chemistry* **2018**, *2*, 720–734.
68. J Enrique-Romero, S Álvarez Barcia, F. J. Kolb, A Rimola, C Ceccarelli, N Balucani, J Meisner, P Ugliengo, T Lamberts, J Kästner, *Monthly Notices of the Royal Astronomical Society* **2020**, *493*, 2523–2527.
69. P. J. Wagner, P. A. Kelso, R. G. Zepp, *Journal of the American Chemical Society* **1972**, *94*, 7480–7488.
70. K. K. Milnes, M. C. Jennings, K. M. Baines, *Journal of the American Chemical Society* **2006**, *128*, 2491–2501.
71. M. Hatzimarinaki, M. M. Roubelakis, M. Orfanopoulos, *Journal of the American Chemical Society* **2005**, *127*, 14182–14183.
72. D. Doehnert, J. Koutecky, *Journal of the American Chemical Society* **1980**, *102*, 1789–1796.
73. M. Abe, *Chemical Reviews* **2013**, *113*, 7011–7088.
74. A. B. Padias, H. K. Hall, *The Journal of Organic Chemistry* **1987**, *52*, 4536–4539.
75. L. J. Johnston, J. C. Scaiano, *Chemical Reviews* **1989**, *89*, 521–547.
76. V. Bonačić-Koutecký, J. Koutecký, J. Michl, *Angewandte Chemie International Edition* **1987**, *26*, 170–189.
77. T. Stuyver, B. Chen, T. Zeng, P. Geerlings, F. De Proft, R. Hoffmann, *Chemical Reviews* **2019**, *119*, 11291–11351.

78. J. Michl, V. Bonačić-Koutecký, *Tetrahedron* **1988**, *44*, 7559–7585.
79. P. Schmid, F. Fantuzzi, J. Klopff, N. B. Schröder, R. D. Dewhurst, H. Braunschweig, V. Engel, B. Engels, *Chemistry – A European Journal* **2021**, *27*, 5160–5170.
80. D. Kaiser, E. Reusch, P. Hemberger, A. Bodi, E. Welz, B. Engels, I. Fischer, *Phys. Chem. Chem. Phys.* **2018**, *20*, 3988–3996.
81. E. Reusch, D. Kaiser, D. Schleier, R. Buschmann, A. Krueger, T. Hermann, B. Engels, I. Fischer, P. Hemberger, *The Journal of Physical Chemistry A* **2019**, *123*, 2008–2017.
82. T. Baer, R. P. Tuckett, *Phys. Chem. Chem. Phys.* **2017**, *19*, 9698–9723.
83. P. Imhof, PhD thesis, Heinrich-Heine-Universität, Düsseldorf, **2001**.
84. F. Duschinsky, *Acta Physicochim. URSS* **1937**, *7*, 551–566.
85. F. Ansbacher, *Zeitschrift für Naturforschung A* **1959**, *14*, 889–892.
86. S. Koide, *Zeitschrift für Naturforschung A* **1960**, *15*, 123–128.
87. S. Gozem, A. I. Krylov, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2021**, e1546.
88. H. Li, M.-B. Huang, *Phys. Chem. Chem. Phys.* **2008**, *10*, 5381–5387.
89. D. P. Mukhopadhyay, D. Schleier, S. Wirsing, J. Ramler, D. Kaiser, E. Reusch, P. Hemberger, T. Preitschopf, I. Krummenacher, B. Engels, I. Fischer, C. Lichtenberg, *Chem. Sci.* **2020**, *11*, 7562–7568.
90. G. Wang, L. A. Freeman, D. A. Dickie, R. Mokrai, Z. Benkő, R. J. Gilliard Jr., *Chemistry – A European Journal* **2019**, *25*, 4335–4339.
91. I. Vránová, M. Alonso, R. Lo, R. Sedlák, R. Jambor, A. Růžička, F. D. Proft, P. Hobza, L. Dostál, *Chemistry – A European Journal* **2015**, *21*, 16917–16928.
92. P. Šimon, F. de Proft, R. Jambor, A. Růžička, L. Dostál, *Angewandte Chemie International Edition* **2010**, *49*, 5468–5471.
93. F. Holzmeier, M. Lang, K. Hader, P. Hemberger, I. Fischer, *The Journal of Chemical Physics* **2013**, *138*, 214310.
94. S. J. Kim, T. P. Hamilton, H. F. Schaefer, *The Journal of Physical Chemistry* **1993**, *97*, 1872–1877.
95. S. Liang, P. Hemberger, N. M. Neisius, A. Bodi, H. Grützmacher, J. Levalois-Grützmacher, S. Gaan, *Chemistry – A European Journal* **2015**, *21*, 1073–1080.
96. S. Liang, P. Hemberger, J. Levalois-Grützmacher, H. Grützmacher, S. Gaan, *Chemistry – A European Journal* **2017**, *23*, 5595–5601.

97. M. Steinbauer, M. Lang, I. Fischer, B. K. C. de Miranda, C. Romanzin, C. Alcaraz, *Physical Chemistry Chemical Physics* **2011**, *13*, 17956–17959.
98. J. Giegerich, J. Petersen, R. Mitrić, I. Fischer, *Physical Chemistry Chemical Physics* **2014**, *16*, 6294–6302.
99. R. Bernheim, R. Kempf, E. Reichenbecher, *Journal of Magnetic Resonance* **1970**, *3*, 5–9.
100. R. A. Bernheim, R. J. Kempf, J. V. Gramas, P. S. Skell, *The Journal of Chemical Physics* **1965**, *43*, 196–200.
101. V. S. Thimmakondur, A. Karton, *Chemical Physics* **2018**, *515*, 411–417.
102. R. A. Seburg, R. J. McMahon, J. F. Stanton, J. Gauss, *Journal of the American Chemical Society* **1997**, *119*, 10838–10845.
103. J. Fulara, P. Freivogel, D. Forney, J. P. Maier, *The Journal of Chemical Physics* **1995**, *103*, 8805–8810.
104. A. Mavrandonakis, M. Mühlhäuser, G. E. Froudakis, S. D. Peyerimhoff, *Phys. Chem. Chem. Phys.* **2002**, *4*, 3318–3321.
105. C. Zhang, Z. Cao, H. Wu, Q. Zhang, *International Journal of Quantum Chemistry* **2004**, *98*, 299–308.
106. J. M.L. Martin, J. El-Yazal, J.-P. François, *Chemical Physics Letters* **1995**, *242*, 570–579.
107. N. P. Bowling, R. J. Halter, J. A. Hodges, R. A. Seburg, P. S. Thomas, C. S. Simmons, J. F. Stanton, R. J. McMahon, *Journal of the American Chemical Society* **2006**, *128*, 3291–3302.
108. N. P. Bowling, N. J. Burrmann, R. J. Halter, J. A. Hodges, R. J. McMahon, *The Journal of Organic Chemistry* **2010**, *75*, 6382–6390.
109. D. Spangenberg, P. Imhof, K. Kleinermanns, *Phys. Chem. Chem. Phys.* **2003**, *5*, 2505–2514.
110. J. A. Montgomery, M. J. Frisch, J. W. Ochterski, G. A. Petersson, *The Journal of Chemical Physics* **1999**, *110*, 2822–2827.
111. J.-D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
112. Y. Zhao, N. E. Schultz, D. G. Truhlar, *Journal of Chemical Theory and Computation* **2006**, *2*, 364–382.
113. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *The Journal of Chemical Physics* **2010**, *132*, 154104.
114. Y. Zhao, D. G. Truhlar, *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.

115. T. Helgaker, P. Jorgensen, J. Olsen, *Molecular Electronic-Structure Theory*, John Wiley & Sons Ltd.: Chichester, **2000**.
116. Y.-H. Ding, J.-L. Liu, X.-R. Huang, Z.-S. Li, C.-C. Sun, *The Journal of Chemical Physics* **2001**, *114*, 5170–5179.
117. Z. Cong-Jie, Z. Li-Ling, C. Ze-Xing, Z. Qian-Er, *Chinese Journal of Chemistry* **2003**, *21*, 117–120.
118. M. Flock, L. Bosse, D. Kaiser, B. Engels, I. Fischer, *Phys. Chem. Chem. Phys.* **2019**, *21*, 13157–13164.
119. M. Z. Zgierski, E. C. Lim, *Chemical Physics Letters* **2004**, *387*, 352–355.
120. Y. Hirata, T. Okada, N. Mataga, T. Nomoto, *The Journal of Physical Chemistry* **1992**, *96*, 6559–6563.
121. A. Mavridis, I. Moustakali-Mavridis, *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* **1977**, *33*, 3612–3615.
122. A. Abramenkov, A. Almenningen, B. Cyvin, S. Cyvin, T. Jonvik, L. Khaikin, C. Rømming, L. Vilkov, *Acta Chemica Scandinavica A* **1988**, *42*, 674–684.
123. G. Baranović, L. Colombo, D. Skare, *Journal of Molecular Structure* **1986**, *147*, 275–300.
124. M. Krämer, U. H. F. Bunz, A. Dreuw, *The Journal of Physical Chemistry A* **2017**, *121*, 946–953.
125. C. Robertson, G. Worth, *Chemical Physics* **2018**, *510*, 17–29.
126. Y. Amatatsu, M. Hosokawa, *The Journal of Physical Chemistry A* **2004**, *108*, 10238–10244.
127. L. Ress, D. Kaiser, J. Grüne, M. Gerlach, E. Reusch, T. Brixner, A. Sperlich, B. Engels, I. Fischer, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2021**, *254*, 119606.
128. J. T. DePinto, R. J. McMahon, *Journal of the American Chemical Society* **1993**, *115*, 12573–12574.
129. M. Noro, N. Koga, H. Iwamura, *Journal of the American Chemical Society* **1993**, *115*, 4916–4916.
130. J. T. DePinto, W. A. deProphetis, J. L. Menke, R. J. McMahon, *Journal of the American Chemical Society* **2007**, *129*, 2308–2315.
131. M. Elkin, T. R. Newhouse, *Chemical Society Reviews* **2018**, *47*, 7830–7844.
132. P. Atkins, J. De Paula, J. Keeler, *Atkins' Physical Chemistry*, Oxford University Press: Oxford, **2018**.
133. S. Logan, *Journal of Chemical Education* **1982**, *59*, 279–281.

134. G.-J. Cheng, X. Zhang, L. W. Chung, L. Xu, Y.-D. Wu, *Journal of the American Chemical Society* **2015**, *137*, 1706–1725.
135. E. Kraka, D. Cremer, *Accounts of Chemical Research* **2010**, *43*, 591–601.
136. K. N. Houk, F. Liu, Z. Yang, J. I. Seeman, *Angewandte Chemie International Edition* **2021**, *60*, 12660–12681.
137. E. Goldstein, B. Beno, K. N. Houk, *Journal of the American Chemical Society* **1996**, *118*, 6036–6043.
138. P. H.-Y. Cheong, C. Y. Legault, J. M. Um, N. Çelebi Ölçüm, K. N. Houk, *Chemical Reviews* **2011**, *111*, 5042–5137.
139. A. N. Kapanidis, S. Weiss, *The Journal of Chemical Physics* **2002**, *117*, 10953–10964.
140. P. A. Szijj, K. A. Kostadinova, R. J. Spears, V. Chudasama, *Organic & Biomolecular Chemistry* **2020**, *18*, 9018–9028.
141. J. P. M. António, R. Russo, C. P. Carvalho, P. M. S. D. Cal, P. M. P. Gois, *Chemical Society Reviews* **2019**, *48*, 3513–3536.
142. N. Stephanopoulos, M. B. Francis, *Nature Chemical Biology* **2011**, *7*, 876–884.
143. K. De Bruycker, S. Billiet, H. A. Houck, S. Chattopadhyay, J. M. Winne, F. E. Du Prez, *Chemical Reviews* **2016**, *116*, 3919–3974.
144. H. Ban, M. Nagano, J. Gavrilyuk, W. Hakamata, T. Inokuma, C. F. Barbas, *Bioconjugate Chemistry* **2013**, *24*, 520–532.
145. H. Ban, J. Gavrilyuk, C. F. Barbas, *Journal of the American Chemical Society* **2010**, *132*, 1523–1525.
146. D. M. Bauer, I. Ahmed, A. Vigovskaya, L. Fruk, *Bioconjugate Chemistry* **2013**, *24*, 1094–1101.
147. C. McKay, M. Finn, *Chemistry & Biology* **2014**, *21*, 1075–1101.
148. M. E. Ortiz-Soto, J. Ertl, J. Mut, J. Adelman, T. A. Le, J. Shan, J. Teßmar, A. Schlosser, B. Engels, J. Seibel, *Chemical Science* **2018**, *9*, 5312–5321.
149. M. E. Ortiz-Soto, J. R. Porrás-Domínguez, J. Seibel, A. López-Munguía, *Carbohydrate Polymers* **2019**, *219*, 130–142.
150. J. Ertl, M. E. Ortiz-Soto, T. A. Le, J. Bechold, J. Shan, J. Teßmar, B. Engels, J. Seibel, *Chemistry – A European Journal* **2019**, *25*, 6533–6541.
151. D. Kaiser, J. M. Winne, M. E. Ortiz-Soto, J. Seibel, T. A. Le, B. Engels, *The Journal of Organic Chemistry* **2018**, *83*, 10248–10260.
152. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M.

- Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian16 Revision C.01, Gaussian Inc. Wallingford CT, **2016**.
153. R. M. Levy, M. Karplus, J. Kushick, D. Perahia, *Macromolecules* **1984**, *17*, 1370–1374.
154. J. Numata, M. Wan, E.-W. Knapp, *Genome Informatics* **2007**, *18*, 192–205.
155. J. Kästner, *WIREs Computational Molecular Science* **2011**, *1*, 932–942.
156. M. S. Shell, A. Panagiotopoulos, A. Pohorille in *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, Berlin & Heidelberg: Springer-Verlag Berlin Heidelberg, **2007**, pp. 77–118.
157. C. C. Moore, *Proceedings of the National Academy of Sciences* **2015**, *112*, 1907–1911.
158. C. Chipot, M. S. Shell, A. Pohorille in *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, Berlin & Heidelberg: Springer-Verlag Berlin Heidelberg, **2007**, pp. 1–31.
159. M. Mezei, *Journal of Computational Physics* **1987**, *68*, 237–248.
160. V. Babin, V. Karpusenko, M. Moradi, C. Roland, C. Sagui, *International Journal of Quantum Chemistry* **2009**, *109*, 3666–3678.
161. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.
162. M. Souaille, B. Roux, *Computer Physics Communications* **2001**, *135*, 40–57.
163. J. Kästner, W. Thiel, *The Journal of Chemical Physics* **2005**, *123*, 144104.
164. W. You, Z. Tang, C.-e. A. Chang, *Journal of Chemical Theory and Computation* **2019**, *15*, 2433–2443.
165. J. G. Kirkwood, *The Journal of Chemical Physics* **1935**, *3*, 300–313.
166. C. D. Christ, A. E. Mark, W. F. van Gunsteren, *Journal of Computational Chemistry* **2010**, *31*, 1569–1582.

167. T. P. Straatsma, J. A. McCammon, *The Journal of Chemical Physics* **1991**, *95*, 1175–1188.
168. M. J. Mitchell, J. A. McCammon, *Journal of Computational Chemistry* **1991**, *12*, 271–275.
169. J. Kästner, H. M. Senn, S. Thiel, N. Otte, W. Thiel, *Journal of Chemical Theory and Computation* **2006**, *2*, 452–461.
170. T. Yu, H. Guo in *Mechanistic Enzymology: Bridging Structure and Function*, Washington DC: American Chemical Society, **2020**, Chapter 6.
171. L. Faivishevsky, J. Goldberger, *Advances in neural information processing systems* **2008**, *21*, 433–440.
172. W. Oleszkiewicz, P. Kairouz, K. Piczak, R. Rajagopal, T. Trzciński in *Computer Vision – ACCV 2018*, Cham: Springer International Publishing, **2019**.
173. S. Kassem, M. Ahmed, S. El-Sheikh, K. H. Barakat, *Journal of Molecular Graphics and Modelling* **2015**, *62*, 105–117.
174. R. Clausius, *The mechanical theory of heat*, MacMillan and Co: London, **1879**.
175. L. Landau, E. Lifshitz, *Statistical Physics*, Pergamon Press: Oxford, **1969**.
176. N. J. Higham, M. R. Dennis, P. Glendinning, P. A. Martin, F. Santosa, J. Tanner, *The Princeton Companion to Applied Mathematics*, Princeton University Press: Princeton, NJ, **2015**.
177. C. E. Shannon, *ACM SIGMOBILE Mobile Computing and Communications Review* **2001**, *5*, 3–55.
178. E. T. Jaynes, *American Journal of Physics* **1965**, *33*, 391–398.
179. D. McQuarrie, *Statistical Mechanics*, Harper & Row: New York, **1975**.
180. E. T. Jaynes, *Physical Review* **1957**, *106*, 620–630.
181. J. W. Ochterski, *Gaussian Inc* **2000**, <https://gaussian.com/wp-content/uploads/dl/thermo.pdf>, [Online; accessed 2021-10-30].
182. G. Jeschke, *Statistical Thermodynamics*, [E-Book only; accessed 2021-10-30], [https://chem.libretexts.org/Bookshelves/Physical\\_and\\_Theoretical\\_Chemistry\\_Textbook\\_Maps/Book%3A\\_Statistical\\_Thermodynamics\\_\(Jeschke\)](https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Book%3A_Statistical_Thermodynamics_(Jeschke)), **2020**.
183. M. Karplus, J. N. Kushick, *Macromolecules* **1981**, *14*, 325–332.
184. J. Schlitter, *Chemical Physics Letters* **1993**, *215*, 617–621.
185. I. Andricioaei, M. Karplus, *The Journal of Chemical Physics* **2001**, *115*, 6289–6292.



186. R. Baron, W. F. van Gunsteren, P. H. Hünenberger, *Trends in Chemical Physics* **2006**, *11*, 87–122.
187. T. Radożycki, *Molecular Physics* **2016**, *114*, 3112–3126.
188. R. Swendsen, *An Introduction to Statistical Mechanics and Thermodynamics*, Oxford University Press: Oxford, **2020**.
189. Wolfram Research Inc. (Champaign, IL), Mathematica, Version 12.3, **2021**, <https://www.wolfram.com/mathematica>.
190. F. Sittel, A. Jain, G. Stock, *The Journal of Chemical Physics* **2014**, *141*, 014111.
191. F. Sittel, T. Filk, G. Stock, *The Journal of Chemical Physics* **2017**, *147*, 244101.
192. A. Altis, PhD thesis, Johann Wolfgang Goethe-Universität Frankfurt, **2009**.
193. Y. Mu, P. H. Nguyen, G. Stock, *Proteins: Structure Function and Bioinformatics* **2005**, *58*, 45–52.
194. A. Altis, P. H. Nguyen, R. Hegger, G. Stock, *The Journal of Chemical Physics* **2007**, *126*, 244111.
195. K. Sargsyan, J. Wright, C. Lim, *Nucleic Acids Research* **2011**, *40*, e25–e25.
196. P. Fletcher, C. Lu, S. Pizer, S. Joshi, *IEEE Transactions on Medical Imaging* **2004**, *23*, 995–1005.
197. B. Eltzner, S. Huckemann, K. V. Mardia, *The Annals of Applied Statistics* **2018**, *12*, 1332–1359.
198. A. Amadei, A. B. Linssen, H. J. Berendsen, *Proteins: Structure Function and Bioinformatics* **1993**, *17*, 412–425.
199. Website: stattrek.com - Variance-Covariance Matrix, <https://stattrek.com/matrix-algebra/covariance-matrix.aspx>, [Online; accessed 2021-10-30].
200. J. Albert, D. Kaiser, V. Engel, *The Journal of Chemical Physics* **2016**, *144*, 171103.
201. M. A. Cuendet, W. F. van Gunsteren, *The Journal of Chemical Physics* **2007**, *127*, 184102.
202. V. Hnizdo, M. K. Gilson, *Entropy* **2010**, *12*, 578–590.
203. V. Hnizdo, J. Tan, B. J. Killian, M. K. Gilson, *Journal of Computational Chemistry* **2008**, *29*, 1605–1614.
204. J. Beirlant, E. J. Dudewicz, L. Györfi, E. C. Van der Meulen, *International Journal of Mathematical and Statistical Sciences* **1997**, *6*, 17–39.
205. F. Nielsen, R. Nock, *IEEE Signal Processing Letters* **2017**, *24*, 402–406.
206. A. Levy, R. Alicki, R. Kosloff, *Physical Review E* **2012**, *85*, 061126.
207. T. Cover, J. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc.: Hoboken, **2012**.

208. L. P. Heinz, H. Grubmüller, *Journal of Chemical Theory and Computation* **2020**, *16*, 108–118.
209. M. Goethe, I. Fita, J. M. Rubi, *The Journal of Chemical Physics* **2017**, *147*, 224102.
210. S. Ghosh, K. P. Burnham, N. F. Laubscher, G. E. Dallal, L. Wilkinson, D. F. Morrison, M. W. Loyer, B. Eisenberg, S. Kullback, I. T. Jolliffe, J. S. Simonoff, *The American Statistician* **1987**, *41*, 338–341.
211. S. Kullback, R. A. Leibler, *The Annals of Mathematical Statistics* **1951**, *22*, 79–86.
212. Anonymous in Proofwiki - online compendium of mathematical proofs: Differential Entropy of Gaussian Distribution, [https://proofwiki.org/wiki/Differential\\_Entropy\\_of\\_Gaussian\\_Distribution](https://proofwiki.org/wiki/Differential_Entropy_of_Gaussian_Distribution), [Online; accessed 2021-10-30].
213. Anonymous in wikipedia - Differential entropy, [https://en.wikipedia.org/wiki/Differential\\_entropy](https://en.wikipedia.org/wiki/Differential_entropy), [Online; accessed 2021-10-30].
214. D. Lombardi, S. Pant, *Physical Review E* **2016**, *93*, 013310.
215. R. T. Rockafellar, R. J. Wets in Heidelberg: Springer Science & Business Media, **1998**, Chapter 1.
216. L. Kozachenko, N. N. Leonenko, *Problemy Peredachi Informatsii* **1987**, *23*, 9–16.
217. H. Singh, N. Misra, V. Hnizdo, A. Fedorowicz, E. Demchuk, *American Journal of Mathematical and Management Sciences* **2003**, *23*, 301–321.
218. Z.-Y. Huang, B.-W. He, *Journal of Shanghai University (English Edition)* **2008**, *12*, 107–109.
219. K. Fukunaga in *Introduction to Statistical Pattern Recognition (Second Edition)*, Boston: Academic Press, **1990**, Chapter 6.
220. A. A. Rahvar, M Ardakani, *Physical Review E* **2011**, *83*, 051121.
221. M. N. Gorja, N. N. Leonenko, V. V. Mergel, P. L. N. Inverardi, *Journal of Non-parametric Statistics* **2005**, *17*, 277–297.
222. S. So, **2008**, *Why is the sample variance a biased estimator?* Griffith University Technical Report, [https://maxwell.ict.griffith.edu.au/sso/biased\\_variance.pdf](https://maxwell.ict.griffith.edu.au/sso/biased_variance.pdf) [Online; accessed 2021-10-30].
223. S. Wilks in *Mathematical Statistics: A Wiley publication*, New York: John Wiley & Sons Inc., **1963**, Chapter 8.7.
224. A. Kraskov, H. Stögbauer, P. Grassberger, *Physical Review E* **2004**, *69*, 066138.
225. A. V. Lazo, P. Rathie, *IEEE Transactions on Information Theory* **1978**, *24*, 120–122.
226. V. Hnizdo, J. Tan, B. J. Killian, M. K. Gilson, *Journal of Computational Chemistry* **2008**, *29*, 1605–1614.

227. H. Matsuda, *Physical Review E* **2000**, 62, 3096.
228. C. Grebner, J. Becker, D. Weber, D. Bellinger, M. Tafipolski, C. Brückner, B. Engels, *Journal of Computational Chemistry* **2014**, 35, 1801–1807.
229. B. Engels, Github repository of the CAST program package, <https://github.com/AKEngels/CAST>, [Online; accessed 2021-10-30].
230. J. W. Ponder et al., TINKER: Software tools for molecular design, <https://dasher.wustl.edu/>, [Online; accessed 2021-10-30].
231. G. Kaminski, W. L. Jorgensen, *The Journal of Physical Chemistry* **1996**, 100, 18010–18013.
232. J. J. Stewart, *Journal of Computer-Aided Molecular Design* **1990**, 4, 1–103.
233. B. Hourahine, B. Aradi, V. Blum, F Bonafé, A Buccheri, C. Camacho, C. Cevallos, M. Deshayé, T Dumitrică, A Dominguez, et al., *The Journal of Chemical Physics* **2020**, 152, 124101.
234. S. Sauer, Dissertation, Julius-Maximilians Universität Würzburg, **2020**.
235. D. Bellinger, Dissertation, Julius-Maximilians Universität Würzburg, **2017**.
236. J. Becker, Dissertation, Julius-Maximilians Universität Würzburg, **2015**.
237. N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *Journal of Cheminformatics* **2011**, 3, 1–14.
238. E. A. Coutsiás, C. Seok, K. A. Dill, *Journal of Computational Chemistry* **2005**, 26, 1663–1665.
239. W. Kabsch, *Acta Crystallographica Section A* **1976**, 32, 922–923.
240. W. Kabsch, *Acta Crystallographica Section A* **1978**, 34, 827–828.
241. R. Chandra, L. Dagum, D. Kohr, R. Menon, D. Maydan, J. McDonald, *Parallel Programming in OpenMP*, Academic Press: San Diego, **2001**.
242. Z. Xianyi, W. Qian, W. Saar: OpenBLAS: An optimized BLAS library, 2016 <http://www.openblas.net/>, [Online; accessed 2021-10-30].
243. E. Wang, Q. Zhang, B. Shen, G. Zhang, X. Lu, Q. Wu, Y. Wang in *High-Performance Computing on the Intel® Xeon Phi™*, Cham: Springer International Publishing Switzerland, **2014**, Chapter 7.
244. W. Kahan, *Communications of the Association for Computing Machinery* **1965**, 8, 40.
245. O. F. Lange, H. Grubmüller, *The Journal of Physical Chemistry B* **2006**, 110, 22842–22852.
246. Y. Mu, P. H. Nguyen, G. Stock, *Proteins: Structure Function and Bioinformatics* **2005**, 58, 45–52.

247. M. Hamacher, Sampling from Gaussian Mixture Models, <https://medium.com/analytics-vidhya/sampling-from-gaussian-mixture-models-f1ab9cac2721>, [Online; accessed 2021-10-30], 2021.
248. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, 1-2, 19–25.
249. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, *Bioinformatics* **2013**, 29, 845–854.
250. F. Vitalini, A. S. Mey, F. Noé, B. G. Keller, *The Journal of Chemical Physics* **2015**, 142, 02B611\_1.
251. F. Vitalini, F. Noé, B. Keller, *Data in brief* **2016**, 7, 582–590.
252. W. Kabsch, *Acta Crystallographica Section A* **1978**, 34, 827–828.
253. E. A. Coutsiias, C. Seok, K. A. Dill, *Journal of Computational Chemistry* **2005**, 26, 1663–1665.
254. E. A. Coutsiias, C. Seok, K. A. Dill, *Journal of Computational Chemistry* **2004**, 25, 1849–1857.
255. K. L. Damm, H. A. Carlson, *Biophysical Journal* **2006**, 90, 4558–4573.
256. C. Sanderson, R. Curtin, *Journal of Open Source Software* **2016**, 1, 26.
257. C. Sanderson, R. Curtin, *Journal of Open Source Software* **2017**, 2, 365.
258. C. Sanderson, R. Curtin in 11<sup>th</sup> International Conference on Signal Processing and Communication Systems (ICSPCS), IEEE, 2017, pp. 1–8.
259. Cubature - a simple C package for adaptive multidimensional integration of vector-valued integrands over hypercubes, <https://github.com/stevengj/cubature>, [Online; accessed 2021-10-30].
260. A. C. Genz, A. A. Malik, *Journal of Computational and Applied Mathematics* **1980**, 6, 295–302.
261. J. Berntsen, T. O. Espelid, A. Genz, *ACM Transactions on Mathematical Software* **1991**, 17, 437–451.
262. M. Feig, *Journal of Chemical Theory and Computation* **2008**, 4, 1555–1564.
263. P. E. Smith, *The Journal of Chemical Physics* **1999**, 111, 5568–5579.

## 12. Acknowledgement

First, I would like to thank Prof. Dr. Bernd Engels for the possibility to write this thesis as a member of his research group in the field of theoretical chemistry. The discussions and meetings with him always yielded fruitful insights into the investigated chemical systems and quantum-chemistry in general. His support as a supervisor is gratefully acknowledged. Secondly, I would like to thank my co-supervisor Prof. Dr. Volker Engel for many discussion over the years.

Further thanks are addressed to all members of the Engels research group. Their presence made the work on this thesis and especially the coffee breaks even more delightful. I would like to especially mention Dr. Daniel Bellinger, Dr. Thien Anh Le, and Dr. Susanne Sauer with whom I at times shared an office. I can hardly imagine better company. Further special mentions deserve (in alphabetical order) Dr. Charlotte Brückner, Luca Craciunescu, Marian Deutsch, Julian Erdmannsdörfer, Alexander Geis, Dr. Anna Heilos, Christoph Herok, Jessica Meyr, Bastian Michels, Michael Prem, Dr. Eileen Welz, and Sara Wirsing, who aided me in various situations. I would like to especially highlight the administrative assistant Ms. Ursula Ruppel. Her aid in filing bureaucratic paperwork ensured smooth research operations throughout my studies. The cooperation with the working groups of Prof. Dr. Ingo Fischer, Prof. Dr. Tobias Brixner and PD habil. Dr. Andreas Sperlich on the analysis of the biradical-type molecules is gratefully acknowledged. Special thanks deserve Dr. Marco Flock, Jeanine Grüne, Lea Ress, and Dr. Engelbert Reusch, as I cooperated closely with them in order to rationalize and interpret their experimental and spectroscopic results. I am furthermore indebted to Prof. Dr. Johan Winne from Ghent University, who supported our research on the tyrosine bioconjugation with his tremendous expertise in synthetic organic chemistry.

The job position which was provided by Prof. Dr. Bernd Engels for part of my time in his research group is gratefully acknowledged. I am very grateful for further financial support received from the Verband der Chemischen Industrie who sponsored my Kekulé PhD fellowship.

I want to thank the research interns Hieu Nguyen and Jonah Herr who have supported me in my projects. I'd like the furthermore highlight Phillipp Stephan, Vincent Schüssler, and Yannik Zimmermann. The scientific discussions with them have contributed significantly to my research interests. Further gratitude go to Jonah Herr, Simon Rachor, Joshua Weinmann, and Sara Wirsing for proofreading sections of this

work.

I want to thank all members of my family who supported me in my academic endeavors, especially my parents and my grandparents.

Most of all I would like to thank my wife, Julia. Without her continuing support neither this work nor any of my other undertakings would have been possible. Her patience and kindheartedness are unsurpassed.

## Appendix

## Copyright Clearance Notes

This thesis includes articles (co-)written by this author which were published in scientific peer-reviewed journals. The following sections give details on the necessary copyright clearances.

Section 2: *Geometry relaxation-mediated localization and delocalization of excitons in organic semiconductors: A quantum chemical study*, M Deutsch, S Wirsing, D Kaiser, RF Fink\*, P Tegeder\*, B Engels\*, *The Journal of Chemical Physics*, 2020, 153, 224104ff: Reprinted from *THE JOURNAL OF CHEMICAL PHYSICS*, 2020, 153, 224104ff, with the permission of AIP Publishing.

Section 3: *Methylbismuth: an organometallic bismuthinidene biradical*, Deb Pratim Mukhopadhyay, Domenik Schleier, Sara Wirsing, Jacqueline Ramler, Dustin Kaiser, Engelbert Reusch, Patrick Hemberger, Tobias Preitschopf, Ivo Krummenacher, Bernd Engels\*, Ingo Fischer\*, Crispin Lichtenberg\*, *Chemical Science*, 2020, 29, 7562-7568: This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Section 4: *Pentadiynylidene and Its Methyl-Substituted Derivates: Threshold Photoelectron Spectroscopy of  $R_1-C_5-R_2$  Triplet Carbon Chains*, Engelbert Reusch, Dustin Kaiser, Domenik Schleier, Rachel Buschmann, Anke Krueger, Thomas Hermann, Bernd Engels\*, Ingo Fischer\*, Patrick Hemberger\*, *The Journal of Physical Chemistry A*, 2019, 10, 2008-2017: Reprinted (adapted) with permission from *THE JOURNAL OF PHYSICAL CHEMISTRY A*, 2019, 10, 2008-2017. Copyright 2019 American Chemical Society.

Section 5: *A time-resolved photoelectron imaging study on isolated tolane: observation of the biradicalic  $^1A_u$  state*, Marco Flock, Lea Bosse, Dustin Kaiser, Bernd Engels\*, Ingo Fischer\*, *Physical Chemistry Chemical Physics*, 2019, 24, 13157-13164: Reproduction of material from PCCP (Physical Chemistry Chemical Physics). Reproduced from Ref. [118] with permission from the PCCP Owner Societies.

Section 6: *Femtosecond dynamics of diphenylpropynylidene in ethanol and dichloromethane*, Lea Ress, Dustin Kaiser, Jeannine Grüne, Marius Gerlach, Engelbert Reusch, Tobias Brixner\*, Andreas Sperlich\*, Bernd Engels\*, Ingo Fischer\*, *Spectrochimica Acta Part*




A: *Molecular and Biomolecular Spectroscopy*, 2021, 254, 119606ff: This is an open access article distributed under the terms of the Creative Commons CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium.


Section 7: *Mechanistical insights into the bioconjugation reaction of Triazolinediones with tyrosine*, Dustin Kaiser, Johan M Winne, Maria Elena Ortiz-Soto, Jürgen Seibel, Thien A Le, Bernd Engels\*, *The Journal of organic chemistry*, 2018, 17, 10248-10260: Reprinted (adapted) with permission from *THE JOURNAL OF ORGANIC CHEMISTRY*, 2018, 17, 10248-10260. Copyright 2018 American Chemical Society.

## Listings of author contributions

## Geometry relaxation-mediated localization and delocalization of excitons in organic semiconductors: A quantum chemical study



**UNIVERSITÄT  
WÜRZBURG**



Fakultät für  
**CHEMIE UND  
PHARMAZIE**


**Erklärung zur Autorenschaft**

Geometry relaxation-mediated localization and delocalization of excitons in organic semiconductors: A quantum chemical study, Marian Deutsch, Sara Wirsing, Dustin Kaiser, Reinhold F. Fink, Petra Tegeder, Bernd Engels, The Journal of Chemical Physics, 2020, 153, 22, 224104


Detaillierte Darstellung der Anteile an der Veröffentlichung (in %)  
Angabe Autoren/innen (ggf. Haupt- / Ko- / korrespondierender Autor/in) mit Vorname Nachname (Initialen)

Autor	A1	A2	A3	A4	A5	A6	Σ in Prozent
Konzeption und Idee des Forschungsvorhabens	5%	5%	5%	15%	15%	55%	100%
Hauptautorenschaft: Quantenchemische Rechnungen zum Einfluss der Geometrieverzerrung auf PBI-Dimere	0%	0%	100%	0%	0%	0%	100%
Hauptautorenschaft: Benchmark von Anregungsenergien des PBI-Monomers	100%	0%	0%	0%	0%	0%	100%
Hauptautorenschaft: Analyse von DIP & PDIR-CN2 Monomer- & Tetramer-Aggregaten	0%	100%	0%	0%	0%	0%	100%
Verfassen des Teils: „Intra-monomer relaxation effects in PBI aggregate structures“	0%	0%	90%	0%	0%	10%	100%
Verfassen des Teils: „Benchmark for vibrationally resolved absorption spectra of perylene-based organic semi-conductors“	90%	0%	0%	0%	0%	10%	100%
Verfassen des Teils: „Intra-monomer relaxation effects in DIP aggregate structures“	0%	90%	0%	0%	0%	10%	100%
Zusammenführung & Verbinden der Einzelbereiche	10%	10%	10%	20%	20%	30%	100%
Korrektur der Publikation	23%	23%	23%	10%	10%	10%	100%
Diskussion der Ergebnisse	23%	23%	23%	10%	10%	10%	100%

**6 (Bernd Engels), Autorin 1 (Marian Deutsch), Autorin 2 (Sara Wirsing), Autorin 3 (Dustin Kaiser), Autorin 4 (Reinhold F. Fink), Autorin 5 (Petra Tegeder), Autorin 6 (Bernd Engels)**



Julius-Maximilians-  
**UNIVERSITÄT**  
**WÜRZBURG**




Fakultät für  
**CHEMIE UND**  
**PHARMAZIE**

Die Mitautoren der in dieser (teil-)kumulativen Dissertation verwendeten Manuskripte sind sowohl über die Nutzung als auch über die angegebenen Eigenanteile informiert und stimmen dem zu.


**Angabe Autorenschaft: Anwählen Dropdownmenü / Autorenunterschrift oder Angabe Verweis: Kontrollkästchen über Eigenschaften aktivieren !**

<p>_____ Autor/in 1 (Marian Deutsch) Hauptautor/in</p> <p><input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt</p>	<p>_____ Autor/in 2 (Sara Wirsing) Hauptautor/in</p> <p><input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt</p>	<p>_____ Autor/in 3 (Dustin Kaiser) Hauptautor/in</p> <p><input type="checkbox"/> Verweis: E-Mail hinterlegt</p>
<p>_____ Autor/in 4 (Reinhold Fink) Koautor/in</p> <p><input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt</p>	<p>_____ Autor/in 5 (Petra Tegeder) Koautor/in</p> <p><input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt</p>	<p>_____ Autor/in 6 (Bernd Engels) Korrespondenzautor/in</p> <p><input type="checkbox"/> Verweis: E-Mail hinterlegt</p>



Würzburg, 14 Oktober 2021  
(Datum)

  
Prof. Dr. Bernd Engels (Betreuer)

## Methylbismuth: an organometallic bismuthinidene biradical

 Fakultät für <b>CHEMIE UND PHARMAZIE</b>													
<b>Erklärung zur Autorenschaft</b> Methylbismuth: an organometallic bismuthinidene biradical D. P. Mukhopadhyay, D. Schleier, S. Wirsing, J. Ramler, D. Kaiser, E. Reusch, P. Hemberger, T. Preitschopf, I. Krummenacher, B. Engels, I. Fischer, C. Lichtenberg <i>Chem. Sci.</i> , 2020, 11, 7562-7568.													
Detaillierte Darstellung der Anteile an der Veröffentlichung (in %) Angabe Autoren/innen (ggf. Haupt- / Ko- / korrespondierender/r Autor/in) mit Vorname Nachname (Initialen)													
<b>Autor/in 1 (D. P. M.), Autor/in 2 (D. S.), Autor/in 3 (S. W.), Autor/in 4 (J. R.), Autor/in 5 (D. K.), Autor/in 6 (E. R.), Autor/in 7 (P. H.), Autor/in 8 (T. P.), Autor/in 9 (I. K.), Autor/in 10 (B. E.), Autor/in 11 (I. F.), Autor/in 12 (C. L.)</b>													
Autor	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Σ in Prozent
Synthesis of all compounds, including analytical data				100%									100%
X-ray crystallography				40%								60%	100%
Theoretical calculations			60%		40%								100%
IR spectroscopy				30%				70%					100%
EPR spectroscopy				10%					80%			10%	100%
Synchrotron radiation experiments	20%	20%				20%					20%		100%
Conception of the publication										20%	40%	40%	100%
Writing the publication										20%	40%	40%	100%
Correcting the manuscript	8%	8%	8%	8%	8%	8%	8%	8%	8%	8%	8%	8%	100%
Coordination of the publication										33%	34%	33%	100%
<b>Summe</b>	8%	8%	12%	8%	6%	8%	4%	3%	3%	10%	15%	15%	100%








Die Mitautoren der in dieser (teil-)kumulativen Dissertation verwendeten Manuskripte sind sowohl über die Nutzung als auch über die angegebenen Eigenanteile informiert und stimmen dem zu


The co-authors of the manuscripts used in this (partly) cumulative dissertation have been informed about the use of these manuscripts and the individual author contributions listed above and agree with it.

Angabe Autorenschaft: Anwählen Dropdownmenü / Autorenunterschrift oder Angabe Verweis: Kontrollkästchen über Eigenschaften aktivieren !


Autor/in 1 (Deb Pratih Mukhopadhyay) Koautor/in <input checked="" type="checkbox"/> Verweis E-Mail hinterlegt	Autor/in 2 (Domenik Schleier) Koautor/in <input checked="" type="checkbox"/> Verweis E-Mail hinterlegt	Autor/in 3 (Sara Wirsing) Koautor/in <input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt
4 Autor/in 4 (Jacqueline Ramler) Koautor/in <input type="checkbox"/> Verweis: E-Mail hinterlegt	Autor/in 5 (Dustin Kaiser) Koautor/in <input checked="" type="checkbox"/> Verweis E-Mail hinterlegt	Autor/in 6 (Engelbert Reusch) Koautor/in <input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt
Autor/in 7 (Patrick Hemberger) Koautor/in <input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt	Autor/in 8 (Tobias Preitschopf) Koautor/in <input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt	Autor/in 9 (Ivo Krummenacher) Koautor/in <input type="checkbox"/> Verweis E-Mail hinterlegt

 <p>Julius-Maximilians- <b>UNIVERSITÄT WÜRZBURG</b></p>	<p>Fakultät für <b>CHEMIE UND PHARMAZIE</b></p> 	
<p>Autor/in 10 (Bernd Engels) <b>Korrespondenzautor/in</b></p> <p><input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt</p>	<p>Autor/in 11 (Ingo Fischer) <b>Korrespondenzautor/in</b></p> <p><input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt</p>	<p>Autor/in 12 (Crispin Lichtenberg) <b>Korrespondenzautor/in</b></p> <p><input type="checkbox"/> Verweis: E-Mail hinterlegt</p>

## Pentadiynylidene and Its Methyl-Substituted Derivates: Threshold Photoelectron Spectroscopy of R<sub>1</sub>-C<sub>5</sub>-R<sub>2</sub> Triplet Carbon Chains



Julius-Maximilians-  
**UNIVERSITÄT**  
**WÜRZBURG**



Fakultät für  
**CHEMIE UND**  
**PHARMAZIE**

**Erklärung zur Autorenschaft**


Pentadiynylidene and Its Methyl-Substituted Derivates: Threshold Photoelectron Spectroscopy of R<sub>1</sub>-C<sub>5</sub>-R<sub>2</sub> Triplet Carbon Chains, Engelbert Reusch, Dustin Kaiser, Domenik Schliefer, Rachel Buschmann, Anke Krueger, Thomas Hermann, Bernd Engels, Ingo Fischer, Patrick Hemberger, The Journal of Physical Chemistry A, 2019, 10, 2008-2017

Detaillierte Darstellung der Anteile an der Veröffentlichung (in %)  
Angabe AutorInnen (ggf. Haupt- / Ko- / korrespondierender AutorIn) mit Vorname Nachname (Initialen)


**AutorIn 1 (Engelbert Reusch), AutorIn 2 (Dustin Kaiser), AutorIn 3 (Domenik Schliefer), AutorIn 4 (Rachel Buschmann), AutorIn 5 (Anke Krueger), AutorIn 6 (Thomas Hermann), AutorIn 7 (Bernd Engels), AutorIn 8 (Ingo Fischer), AutorIn 9 (Patrick Hemberger)**

Autor	A1	A2	A3	A4	A5	A6	A7	A8	A9	Σ in Prozent
Idee und Konzeption des Forschungsvorhabens	5%	0%	0%	0%	20%	0%	10%	35%	30%	100%
Quantenchemische Rechnungen	5%	95%	0%	0%	0%	0%	0%	0%	0%	100%
Spektroskopie & Synthese	70%	0%	5%	5%	5%	15%	0%	0%	0%	100%
Verfassen der Publikation: Anteil „Theoretische Rechnungen“	5%	75%	0%	0%	0%	0%	15%	5%	0%	100%
Verfassen der Publikation: Anteil „Experimentelle Ergebnisse“	75%	0%	5%	0%	0%	0%	0%	15%	5%	100%
Korrektur der Publikation	10%	10%	0%	0%	20%	0%	20%	20%	20%	100%
Diskussion der Ergebnisse	11%	11%	11%	11%	11%	11%	11%	11%	11%	100%

Die Mitautoren der in dieser (teil-)kumulativen Dissertation verwendeten Manuskripte sind sowohl über die Nutzung als auch über die angegebenen Eigenanteile informiert und stimmen dem zu.



Julius-Maximilians-  
**UNIVERSITÄT**  
**WÜRZBURG**



Fakultät für  
**CHEMIE UND**  
**PHARMAZIE**

Angabe Autorenschaft: Anwählen Dropdownmenü / Autorenunterschrift **oder** Angabe Verweis: Kontrollkästchen über Eigenschaften aktivieren !

Autor/in 1 (Engelbert Reusch)	Autor/in 2 (Domenik Schleier)	Autor/in 3 (Domenik Schleier)
<b>Hauptautor/in</b>	<b>Hauptautor/in</b>	<b>Koautor/in</b>
<input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt	<input type="checkbox"/> Verweis: E-Mail hinterlegt	<input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt
Autor/in 4 (Rachel Buschmann)	Autor/in 5 (Anke Krueger)	Autor/in 6 (Thomas Hermann)
<b>Koautor/in</b>	<b>Koautor/in</b>	<b>Koautor/in</b>
<input type="checkbox"/> Verweis: E-Mail hinterlegt	<input type="checkbox"/> Verweis: E-Mail hinterlegt	<input type="checkbox"/> Verweis: E-Mail hinterlegt
Autor/in 7 (Bernd Engels)	Autor/in 8 (Ingo Fischer)	Autor/in 9 (Patrick Hemberger)
<b>Korrespondenzautor/in</b>	<b>Korrespondenzautor/in</b>	<b>Korrespondenzautor/in</b>
<input type="checkbox"/> Verweis: E-Mail hinterlegt	<input type="checkbox"/> Verweis: E-Mail hinterlegt	<input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt
Würzburg, dd Month yyyy (Datum)		
Prof. Dr. Bernd Engels (Betreuer)		



## A time-resolved photoelectron imaging study on isolated tolane: observation of the biradicalic $^1A_u$ state




### Erklärung zur Autorenschaft

A time-resolved photoelectron imaging study on isolated tolane: observation of the biradicalic  $^1A_u$  state, Marco Flock, Lea Bosse, Dustin Kaiser, Bernd Engels, Ingo Fischer, Physical Chemistry Chemical Physics, 2019, 21, 24, 13157-13164


Detaillierte Darstellung der Anteile an der Veröffentlichung (in %) Angabe Autoren/namen (ggf. Haupt- / Ko- / korrespondierender Autor/in) mit Vorname Nachname (Initialen)

**Autor/in 1 (Marco Flock), Autor/in 2 (Lea Bosse), Autor/in 3 (Dustin Kaiser), Autor/in 4 (Bernd Engels), Autor/in 5 (Ingo Fischer)**

Autor	A1	A2	A3	A4	A5	$\Sigma$ in Prozent
Konzeption & Idee des Forschungsvorhabens	15%	0%	0%	5%	80%	100%
Quantenchemische Rechnungen	0%	0%	90%	10%	0%	100%
Spektroskopie	50%	30%	0%	0%	20%	100%
Verfassen der Publikation: Anteil „experimentelle Ergebnisse“	50%	30%	0%	0%	20%	100%
Verfassen der Publikation: Anteil „theoretische Rechnungen“	5%	0%	90%	5%	0%	100%
Korrektur der Publikation	20%	20%	20%	20%	20%	100%
Diskussion der Ergebnisse	30%	10%	20%	10%	30%	100%



Die Mitautoren der in dieser (teil-)kumulativen Dissertation verwendeten Manuskripte sind sowohl über die Nutzung als auch über die angegebenen Eigenanteile informiert und stimmen dem zu.  
**Angabe Autorenschaft: Anwählen Dropdownmenü / Autorenunterschrift oder Angabe Verweis: Kontrollkästchen über Eigenschaften aktivieren !**



Autor/in 1 (Marco Flock)	Autor/in 2 (Lea Bosse)	Autor/in 3 (Dustin Kaiser)
<b>Hauptautor/in</b>	<b>Koautor/in</b>	<b>Hauptautor/in</b>
<input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt	<input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt	<input type="checkbox"/> Verweis: E-Mail hinterlegt
Autor/in 4 (Bernd Engels)	Autor/in 5 (Ingo Fischer)	
<b>Koautor/in</b>	<b>Korrespondenzautor/in</b>	
<input type="checkbox"/> Verweis: E-Mail hinterlegt	<input type="checkbox"/> Verweis: E-Mail hinterlegt	

Würzburg, 14 Oktober 2021  
(Datum)

Prof. Dr. Bernd Engels (Betreuer)

## Femtosecond dynamics of diphenylpropynylidene in ethanol and dichloromethane



## Erklärung zur Autorenschaft

Femtosecond Dynamics of Diphenylpropynylidene in Ethanol and Dichloromethane, Lea Röss, Dustin Kaiser, Jeannine Grüne, Marius Gerlach, Engelbert Reusch, Tobias Brixner, Andreas Sperlich, Bernd Engels, Ingo Fischer, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021, 119606

Detaillierte Darstellung der Anteile an der Veröffentlichung (in %) Angabe Autorennamen (ggf. Haupt- / Ko- / korrespondierender Autor/in) mit Vorname (Nachname (Initialen))

Autor	A1	A2	A3	A4	A5	A6	A7	A8	A9	$\Sigma$ in Prozent
Performing Quantum-Chemical calculations	0	100	0	0	0	0	0	0	0	100
Interpretation of Quantum-Chemical results	0	80	0	0	0	0	0	20	0	100
Written composition & internal meeting presentations: Quantum-Chemical calculations	0	100	0	0	0	0	0	0	0	100
Performing Transient Absorption experiments	100	0	0	0	0	0	0	0	0	100
Interpretation & Analysis of Transient Absorption results	80	0	0	0	0	10	0	0	10	100
Written composition & internal meeting presentations: Transient Absorption experiments	100	0	0	0	0	0	0	0	0	100
Reconciliation of theoretical and experimental data	40	40	0	0	0	0	0	20	0	100
Performing EPR spectroscopy	0	0	100	0	0	0	0	0	0	100

**Autorin 1 (Lea Röss), Autorin 2 (Dustin Kaiser), Autorin 3 (Jeannine Grüne), Autorin 4 (Marius Gerlach), Autorin 5 (Engelbert Reusch), Autorin 6 (Tobias Brixner), Autorin 7 (Andreas Sperlich), Autorin 8 (Bernd Engels), Autorin 9 (Ingo Fischer)**



Interpretation & Analysis of EPR spectroscopy results	0	0	80	0	0	0	20	0	0	100
Written composition & internal presentation: EPR Spectroscopy	0	0	100	0	0	0	0	0	0	100
Chemical Synthesis	0	0	0	50	0	0	0	0	0	100
Main paper writing: quantum-chemical calculations	0	75	0	0	0	0	0	25	0	100
Main paper writing: spectroscopy	65	0	20	0	0	5	5	0	5	100
Main paper writing: Introduction, Conclusion, Discussion	25	15	10	0	0	5	5	15	25	100
Main paper writing: corrections	50	20	10	0	0	0	0	10	10	100
Main paper writing: coordination	10	5	5	0	0	10	10	10	50	100
<b>Contribution to project in total:</b>	<b>35</b>	<b>35</b>	<b>8</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>12</b>	<b>100</b>

Die Mitautoren der in dieser (teil-)kumulativen Dissertation verwendeten Manuskripte sind sowohl über die Nutzung als auch über die angegebenen Eigenanteile informiert und stimmen dem zu.

**Angabe Autorenschaft: Anwählen Dropdownmenü / Autorenumterschrift oder Angabe Verweis: Kontrollkästchen über Eigenschaften aktivieren !**

Autor/in 1 (Lea Ress)

Hauptautor/in

Verweis E-Mail hinterlegt

Autor/in 2 (Dustin Kaiser)



Hauptautor/in

Verweis: E-Mail hinterlegt

Autor/in 3 (Jeamine Grüne)

Koautor/in

Verweis E-Mail hinterlegt


<p>_____ Autor/in 4 (Marius Gerlach)</p> <p><b>Koautor/in</b></p> <p><input type="checkbox"/> Verweis: E-Mail hinterlegt</p>	<p>_____ Autor/in 5 (Engelbert Reusch)</p> <p><b>Koautor/in</b></p> <p><input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt</p>	<p>_____ Autor/in 6 (Iobias Brxner)</p> <p><b>Koautor/in</b></p> <p><input type="checkbox"/> Verweis: E-Mail hinterlegt</p>
<p>_____ Autor/in 7 (Andreas Sperlich)</p> <p><b>Koautor/in</b></p> <p><input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt</p>	<p>_____ Autor/in 8 (Bernd Engels)</p> <p><b>Koautor/in</b></p> <p><input type="checkbox"/> Verweis: E-Mail hinterlegt</p>	<p>_____ Autor/in 9 (Ingo Fischer)</p> <p><b>Korrespondenzautor/in</b></p> <p><input type="checkbox"/> Verweis: E-Mail hinterlegt</p>

Würzburg, *14. Oktober 2021*  
(Datum)


\_\_\_\_\_  
Prof. Dr. Bernd Engels (Betreuer)



## Mechanistical insights into the bioconjugation reaction of Triazolinediones with tyrosine



Julius-Maximilians-  
**UNIVERSITÄT**  
**WÜRZBURG**



Fakultät für  
**CHEMIE UND**  
**PHARMAZIE**

**Erklärung zur Autorenschaft**

Mechanistical insights into the bioconjugation reaction of Triazolinediones with tyrosine, Dustin Kaiser, Johan M. Winne, Maria Elena Ortiz-Soto, Jürgen Seibel, Thien A. Le, Bernd Engels, The Journal of organic chemistry, 2018, 10248-10260



Detaillierte Darstellung der Anteile an der Veröffentlichung (in %)  
Angabe Autoren/innen (ggf. Haupt- / Ko- / korrespondierende/r Autor/in) mit Vorname Nachname (Initialen)

**Autor/in 1 (Dustin Kaiser), Autor/in 2 (Johan M. Winne), Autor/in 3 (Maria Elena Ortiz-Soto), Autor/in 4 (Jürgen Seibel), Autor/in 5 (Thien A. Le), Autor/in 6 (Bernd Engels),**

Autor	A1	A2	A3	A4	A5	A6	Σ in Prozent
Quantenchemie: Durchführung, Interpretation der Ergebnisse, schriftliche Zusammenfassung der Ergebnisse, Abgleich mit experimentellen Daten	90%	0%	0%	0%	0%	10%	100%
Verfassen der Publikation	75%	10%	0%	0%	5%	10%	100%
Korrektur der Publikation	80%	10%	0%	0%	0%	10%	100%
Diskussion der Ergebnisse	35%	20%	5%	15%	5%	15%	100%
<b>Anteil an der Gesamtveröffentlichung</b>	<b>80%</b>	<b>7%</b>	<b>2%</b>	<b>2%</b>	<b>2%</b>	<b>7%</b>	<b>100%</b>

Die Mitautoren der in dieser (teil-)kumulativen Dissertation verwendeten Manuskripte sind sowohl über die Nutzung als auch über die angegebenen Eigenanteile informiert und stimmen dem zu.

**Angabe Autorenschaft: Anwählen Dropdownmenü / Autorenumterschrift oder Angabe Verweis: Kontrollkästchen über Eigenschaften aktivieren !**

			
Autor/in 1 (Dustin Kaiser) <b>Hauptautor/in</b> <input type="checkbox"/> Verweis: E-Mail hinterlegt	Autor/in 2 (Johan M. Winne) <b>Koautor/in</b> <input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt	Autor/in 3 (Maria Elena Ortiz-Soto) <b>Koautor/in</b> <input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt	Autor/in 4 (Jürgen Seibel) <b>Koautor/in</b> <input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt
Autor/in 5 (Thien A. Le) <b>Koautor/in</b> <input checked="" type="checkbox"/> Verweis: E-Mail hinterlegt	Autor/in 6 (Bernd Engels) <b>Korrespondenzautor/in</b> <input type="checkbox"/> Verweis: E-Mail hinterlegt	Würzburg, 14 Oktober 2021 (Datum)	
		Prof. Dr. Bernd Engels (Betreuer)	