Fabian Bohle

Development and Application of efficient automated Workflows for Quantum Chemistry-based property calculations

Dissertation

Development and Application of efficient automated Workflows for Quantum Chemistry-based property calculations

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Fabian Bohle
aus Bad Soden
–Bonn 2025–

Gutachter/Betreuer:	Prof. Dr. Stefan Grimme
Gutachter:	Prof. Dr. Thomas Bredow
Tag der Disputation:	28.04.2025
Erscheinungsjahr:	2025

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

Publications

Parts of this thesis (in order of appearance) have been published in peer-reviewed journals:

1. R. Hahn, F. Bohle, W. Fang, A. Walther, S. Grimme, B. Esser "Raising the Bar in Aromatic Donor-Acceptor Interactions with Cyclic Trinuclear Gold(I) Complexes as Strong π-Donors", *J. Am. Chem. Soc.* **2018**, *140*, 17932-17944.

2. L. Volbach, N. Struch, F. Bohle, F. Topić, G. Schnakenburg, A. Schneider, K. Rissanen, S. Grimme, A. Lützen "Influencing the Self-Sorting Behavior of [2.2]Paracyclophane-Based Ligands by Introducing Isostructural Binding Motifs", *Chem. Eur. J.* **2020**, *26*, 3335-3347.

3. F. Bohle, S. Grimme "Efficient structural and energetic screening of fullerene encapsulation in a large supramolecular double decker macrocycle", *J. Serb. Chem. Soc.* **2019**, *84*, 837-844.

4. H. Neugebauer, F. Bohle, M. Bursch, A. Hansen, S. Grimme " Benchmark Study of Electrochemical Redox Potentials Calculated with Semiempirical and DFT Methods", *J. Phys. Chem. A* **2020**, *124*, 7166-7176.

5. P. Funk, R. B. Richrath, F. Bohle, S. Grimme, A. Gansäuer "Oxidation Under Reductive Conditions: From Benzylic Ethers to Acetals with Perfect Atom-Economy by Titanocene(III) Catalysis", *Angew. Chem. Int. Ed.* **2021**, *60*, 5482-5488.

6. S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher, M. Stahn "Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules", *J. Phys. Chem. A* **2021**, *125*, 4039−4054.

7. F. Bohle, J. Seibert, and S. Grimme "Automated Quantum Chemistry-Based Calculation of Optical Rotation for Large Flexible Molecules", *J. Org. Chem.* **2021**, *86*, 15522-15531.

8. F. Bohle, S. Grimme "Hydrocarbon Macrocycle Conformer Ensembles and ^{13}C-NMR Spectra", *Angew. Chem. Int. Ed.* **2022**, *134*, e202113905.

Further publications:

9. Y.-Q. Zhang, C. Poppel, A. Panfilova, F. Bohle, S. Grimme, A. Gansäuer "S_N2 Reactions at Tertiary Carbon Centers in Epoxides", *Angew. Chem. Int. Ed.* **2017**, *56*, 9719.

10. H. C. Gottschalk, A. Poblotzki, M. A. Suhm, M. M. Al-Mogren, J. Antony, A. A. Auer, L. Baptista, D. M. Benoit, G. Bistoni, F. Bohle, R. Dahmani, D. Firaha, S. Grimme, A. Hansen, M. E. Harding, M. Hochlaf, C. Holzer, G. Jansen, W. Klopper, W. A. Kopp, L. C. Kröger, K. Leonhard, H. Mouhib, F. Neese, M. N. Pereira, I. S. Ulusoy, A. Wuttke, R. A. Mata "The furan microsolvation blind challenge for quantum chemical methods: First steps", *J. Chem. Phys.* **2018**, *148*, 014301.

11. R. Hahn, F. Bohle, S. Kotte, T. J. Keller, S.-S. Jester, A. Hansen, S. Grimme, B. Esser "Donor–acceptor interactions between cyclic trinuclear pyridinate gold(I)-complexes and electron-poor guests: nature and energetics of guest-binding and templating on graphite", *Chem. Sci.* **2018**, *9*, 3477-3483.

12. Y.-Q. Zhang, F. Bohle, R. Bleith, G. Schnakenburg, S. Grimme, A. Gansäuer "Synthesis of 1,3-Amino Alcohols by Hydroxy-Directed Aziridination and Aziridine Hydrosilylation", *Angew. Chem. Int. Ed.* **2018**, *57*, 13528-13532.

13. S. Oswald, N. A. Seifert, F. Bohle, M. Gawrilow, S. Grimme, W. Jäger, Y. Xu, M. A. Suhm "The Chiral Trimer and a Metastable Chiral Dimer of Achiral Hexafluoroisopropanol: A Multi-Messenger Study", *Angew. Chem. Int. Ed.* **2019**, *58*, 5080-5084.

14. J. L. Kuo, T. Gunasekara, A. Hansen, H. B. Vibbert, F. Bohle, J. R. Norton, S. Grimme, P. J. Quinlivan "Thermodynamics of $H^+/H^{\cdot}/H^-/e^-$ Transfer from $[CpV(CO)_3H]^-$: Comparisons to the Isoelectronic $CpCr(CO)_3H$", *Organometallics* **2019**, *38*, 4319-4328.

15. P. Pracht, F. Bohle, S. Grimme "Automated exploration of the low-energy chemical space with fast quantum chemical methods", *Phys.Chem.Chem.Phys.* **2020**, *22*, 7169-7192.

16. H. C. Gottschalk, A. Poblotzki, M. Fatima, D. A. Obenchain, C. Pérez, J. Antony, A. A. Auer, L. Baptista, D. M. Benoit, G. Bistoni, F. Bohle, R. Dahmani, D. Firaha, S. Grimme, A. Hansen, M. E. Harding, M. Hochlaf, C. Holzer, G. Jansen, W. Klopper, W. A. Kopp, M. Krasowska, L. C. Kröger, K. Leonhard, M. M. Al-Mogren, H. Mouhib, F. Neese, M. N. Pereira, M. Prakash, I. S. Ulusoy, R. A. Mata, M. A. Suhm, Melanie Schnell, "The first microsolvation step for furans: New experiments and benchmarking strategies", *J. Chem. Phys.* **2020**, *152*, 164303.

17. S. A. Katsyuba, T. P. Gerasimova, S. Spicher, F. Bohle, S. Grimme, "Computer–aided simulation of infrared spectra of ethanol conformations in gas, liquid and in CCl_4 solution", *J. Comput. Chem.* **2022**, *43*, 279.

Presentations:

1. Poster: “Calculating Association Free Energies of a Highly Charged Host-Guest System“, *Symposium for Theoretical Chemistry*, **2016**, Bochum, Germany.

2. Poster: “Investigating π-acid-base interactions of supramolecular cyclic trinuclear coinage metal complexes“, *World Association of Theoretical and Computational Chemists*, **2017**, München, Germany.

3. Poster: “How to Compute Molecular Recognition Processes in Solution“ *Symposium for Theoretical Chemistry*, **2018**, Halle, Germany.

4. Poster: “Computation of Molecular Recognition Processes in Solution“, *Vrije Universiteit - Theoretical Chemistry Colloquia*, **2017**, Amsterdam, Netherlands.

5. Poster: “Computation of Molecular Recognition Processes in Solution“, *Theoretical Chemistry Colloquia*, **2018**, Bochum, Germany.

6. Poster: “Benchmarking of Conformer Rotamer Ensembles in Solution by Computation of NMR Spectra“, *18th International Conference on Density-Functional Theory and its Applications*, **2019**, Alicante, Spain.

Contents

Publications i

I. Introduction and Theoretical Background 1

1. Introduction 3

2. Theoretical Background 9

2.1. Overview of Methods 9

2.2. Electronic Structure Methods 11

2.2.1. Hartree–Fock Theory and Roothaan–Hall Equations 12

2.2.2. Electron Correlation 15

2.2.3. Kohn-Sham Density Functional Theory 16

2.2.4. Tight-Binding Methods 24

2.2.5. Extended Tight-Binding Methods – The GFN-Method Family 27

2.3. GFN special purpose Force Field 29

2.4. Contributions to Free Energy 30

2.4.1. Gas phase thermodynamic Properties 30

2.4.2. Conformations 31

2.4.3. Effects of solvation 32

2.5. Calculation of Molecular Response Properties 34

2.5.1. Nuclear Magnetic Resonance 34

2.5.2. Optical rotation and electronic circular dichroism 36

II. Efficient Free Energy Calculations for Metal Organic Systems 39

3. Raising the Bar in Aromatic Donor–Acceptor Interactions 43

4. Influencing the Self-Sorting Behavior of [2.2]Paracyclophane 47

5. Efficient structural and energetic screening 51

6. Benchmark Study of Electrochemical Redox Potentials 55

III. Beyond the Single Structure Approach: Automated Workflows for Property Predictions of Nonrigid Systems 59

7. Oxidation Under Reductive Conditions 61

8. Structure Ensembles and Free Energies for Nonrigid Molecules 65

9. Calculation of Optical Rotation for Large Molecules 69

10. Hydrocarbon Macrocycle Conformer Ensembles and ^{13}C-NMR spectra 73

IV. Final Summary and Conclusion 77

Bibliography 85

V. Appendix 113

A1. Supporting Information to Chapter 2 115

A1.1. Extended Tight-Binding Methods 115

A1.1.1. Energy Terms of the GFN*n*-xTB methods 115

A1.1.2. Energies in GFN1–xTB 117

A1.1.3. Energies in GFN2–xTB 118

A1.2. Equations for statistical thermodynamics 120

A1.3. Conductor-like screening model for real solvents (COSMO-RS) 121

A1.4. Structure ensembles and molecular dynamics 123

A2. Paper reproduction belonging to Chapter 3 125

A2.1. Introduction 126

A2.2. Results and discussion 129

A2.3. Conclusions 140

A2.4. Computational Details 141

A3. Supporting Information to Chapter 3 143

A4. Paper reproduction belonging to Chapter 4 147

A4.1. Introduction 147

A4.2. Results and Discussion 148

A4.3. Synthesis 148

A4.4. Coordination studies: solution and gas-phase studies 150

A4.5. Coordination studies: solid-state studies 159

A4.6. Conclusion 162

A4.7. Computational details 162

A5. Supporting Information to Chapter 4 **165**

A6. Paper reproduction belonging to Chapter 5 **173**
A6.1. Introduction 173
A6.2. Computational methods 174
A6.3. Results and discussion 175
A6.4. Conclusion 178

A7. Supporting Information to Chapter 5 **179**

A8. Paper reproduction belonging to Chapter 6 **181**
A8.1. Introduction 181
A8.2. Methods 183
A8.2.1. Computational Details 183
A8.2.2. Calculation of Redox Potentials 184
A8.2.3. Workflow 185
A8.3. Results and Discussion 186
A8.3.1. Ionization Potentials and Electron Affinities 186
A8.3.2. Benchmark Study on Redox Potentials 188
A8.3.3. The *ROP313* Benchmark 188
A8.3.4. Redox Potentials 188
A8.4. Conclusion and Outlook 195
A8.5. Author Information 196

A9. Supporting Information to Chapter 6 **197**

A10. Paper reproduction belonging to Chapter 7 **241**
A10.1. Introduction 241
A10.2. Results and Discussion 242
A10.3. Conclusion 253

A11. Supporting Information to Chapter 7 **255**

A12. Paper reproduction belonging to Chapter 8 **263**
A12.1. Introduction 263
A12.2. Theory 266
A12.2.1. General considerations 266
A12.2.2. Electronic energy 268
A12.2.3. Generating the structure ensemble 270
A12.2.4. Thermostatistical contributions 272
A12.2.5. Conformational and anharmonic terms 273
A12.2.6. Solvation free energy 275

A12.2.7. The CENSO algorithm 277
A12.3. Results and Discussion 282
A12.3.1. Conformational ensembles and optical rotation 283
A12.3.2. Partition Coefficients 284
A12.3.3. pK_a values 285
A12.4. Conclusions and Recommendations 286

A13. Paper reproduction belonging to Chapter 9 289
A13.1. Introduction 289
A13.2. Results and discussion 290
A13.3. Conclusion 302
A13.4. Computational Methods 302

A14. Supporting Information to Chapter 9 305

A15. Paper reproduction belonging to Chapter 10 315
A15.1. Introduction 315
A15.2. Results and Discussion 316
A15.3. Conclusion 326

A16. Supporting Information to Chapter 10 327

A17. List of Abbreviations 360

Acknowledgements 365

Part I.

Introduction and Theoretical Background

1. Introduction

The development of theoretical methods for describing chemical systems has its beginnings in about 1928 with early attempts at solving the Schrödinger equation.[1] The routine application of quantum chemical models has only become possible with the wide availability of computing power and software packages since about the 1970th. From then on, computational simulations have become an essential part of modern chemistry.[2] Over the last 50 years, the research field of chemistry has changed from a mainly experimental science to an interdisciplinary research field in which areas of physics and computer science play an increasingly important role. Current key research topics benefiting from computational chemistry include catalyst[3–9] and materials design[10–16], drug development[17–21], analytical spectroscopy[22–28], energy storage[29–31] and lately the development of very fast machine learning methods.[10,32–39] The course of any chemistry-focused development in basic research or industry involves searching for a desired functionality/property, which may be, for example, good water solubility, high energy density, or strong binding to a receptor. For each of these tasks, a large number of chemical systems exist that are possible candidates, with varying degrees of likelihood of meeting the design requirements.

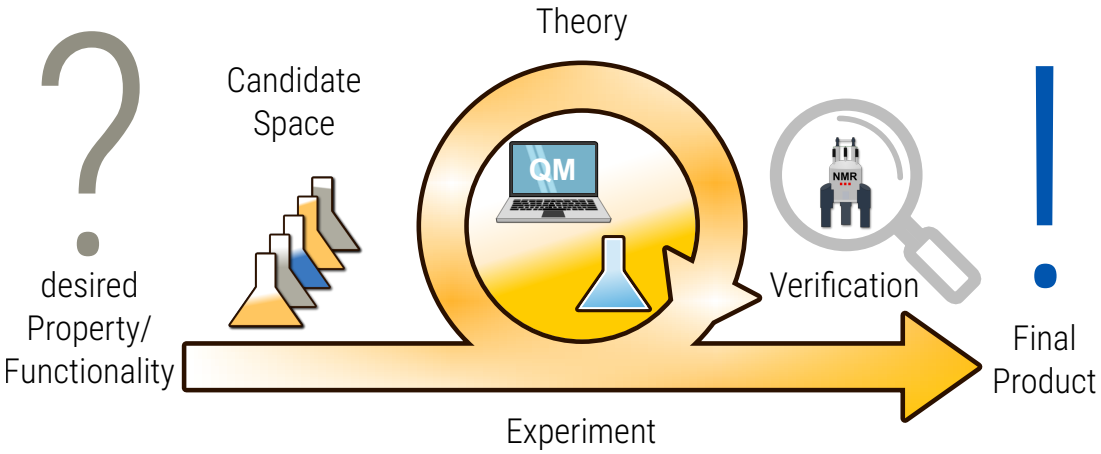


Figure 1.1.: Schematic workflow of chemical research. The process consists of: an initial question, the selection of the possible candidates, the screening, and the validation before the response/-final product. Especially in the screening and validation process, theoretical chemistry can play to its strengths and accelerate these processes.

Figure 1.1 showcases the general flow of compound development by formulating the desired property and choosing a relevant candidate space. Rapid identification of promising candidates saves time and resources in the development process, greatly accelerated by the increasingly routine use of theo-

retical chemistry. In many cases, development is driven by both experiment and theory in a beneficial iterative research process that leads to a more refined understanding with each iteration. A prominent advantage of computational chemistry is that model systems that are difficult to access synthetically or spectroscopically can be investigated. Before the product is finalized, the achieved quality of the desired property is checked or measured. Here, spectroscopy can provide insight into molecular properties, and one goal of this work is to develop a routinely applicable, seamless, and automated multilevel modeling tool for molecular property calculation and prediction.

The central quantity for understanding many processes under standard conditions is the free or Gibbs[40] energy. Its change or difference provides insights into thermodynamic equilibrium or reaction direction. It can be converted into equilibrium constants, including the pharmacologically important octanol-water partition coefficient ($\log K_{OW}$).[41] It refines our understanding of reaction pathways, including barrier heights, and thereby enables optimization of reaction conditions.[42] According to the semi-classical transition state theory (TST), kinetic rate constants (k) are accessible from the activation barrier $\Delta G^{\neq}$.[43] Furthermore, properties can be expressed by the linear free energy relationship (LFER), which includes, for example, the dissociation constants of acids (pKa).[44–46] For systems in thermal equilibrium, their populations can be calculated from the Boltzmann distribution relying on the individual free energies in the canonical ensemble.[40] While predicting gas phase reaction energies at 0 K is nowadays routinely possible,[47] most real-life reactions are performed at finite temperatures and in solution.[48] Due to the vast chemical space and abundance of possible (undiscovered) reactions, accurate modeling or calculation of free energy is an ongoing scientific endeavor.[2,47,49,50] This research is required because of the complexity of ΔG caused by the concurrence of contributions such as the search for the suiting representation of the compound geometry, the thermodynamic influence of temperature, and environmental effects present in the condensed phase.

The understanding of chemical systems is closely related to the principle of structural reactivity,[51] where knowledge of the spatial structure of a compound allows the calculation of molecular (free) energy[52] or reactivity descriptors (e.g., chemical hardness, or electronegativity)[53] followed by the interpretation or prediction of chemical reactivity or properties. Various methods have been developed for theoretical investigations, each with a specific system focus, and selecting the right method for the problem at hand is the most important task of the computational chemist.[54] For this purpose, starting from any input geometry, one can choose from the large pool of wavefunction methods (WFT),[55] Kohn-Sham density functional theory (KS-DFT),[56–58] semiempirical methods[59–61] (SQM) and molecular mechanics force fields[62,63] (FF). The strengths and weaknesses of the methods can be roughly categorized as computational speed, universal applicability, and accuracy or magnitude of method error. The choice of method must be negotiated within this trade-off triangle and is determined by the question of what accuracy must be achieved within the project's scope and the allotted time frame.

Determining a suitable initial geometry that best represents the experimental system is non-trivial. The first attempt is to use available crystal structure or NMR data or to model the geometry by hand. The input structure must be optimized and result in a local minimum on the molecule's potential energy surface (PES). This approach (as shown in Fig. 1.2 **I**) can be called the chemist's intuition and can be valid for very small and/or rigid molecules. The best single representation of the molecule is the

stereoisomer associated with the lowest energy on the PES, also known as the global minimum. The strategy of calculating a molecular property or free energy from the (near)-global minimum geometry (approach **II** in Fig. 1.2) has to be adopted when flexible molecules are involved, and the property under consideration is not sensitive towards structural aspects. The search for spatial isomers with the same covalent bond structure and topology, also called conformers, is a huge undertaking because even relatively small molecules (≈30 atoms) can have many rotatable bonds (i.e., degrees of freedom) and the number of conformers that need to be explored can be enormous. Exploring the "low-energy" chemical space requires an intelligent algorithm that is easy to use and searches the space efficiently. As a second prerequisite, robust (low error-prone) and fast but reasonably accurate computational methods must be used so that the simulations can be routinely applied.

The third conceptual approach (**III**) recognizes that for describing the property of a truly conformationally diverse system, the Boltzmann distribution of conformers is more important than the global minimum alone. This is the supreme discipline, as the search for the complete populated ensemble is inherently difficult. One gains the most insight into the system here, and at the same time, this approach has the highest associated computational cost.

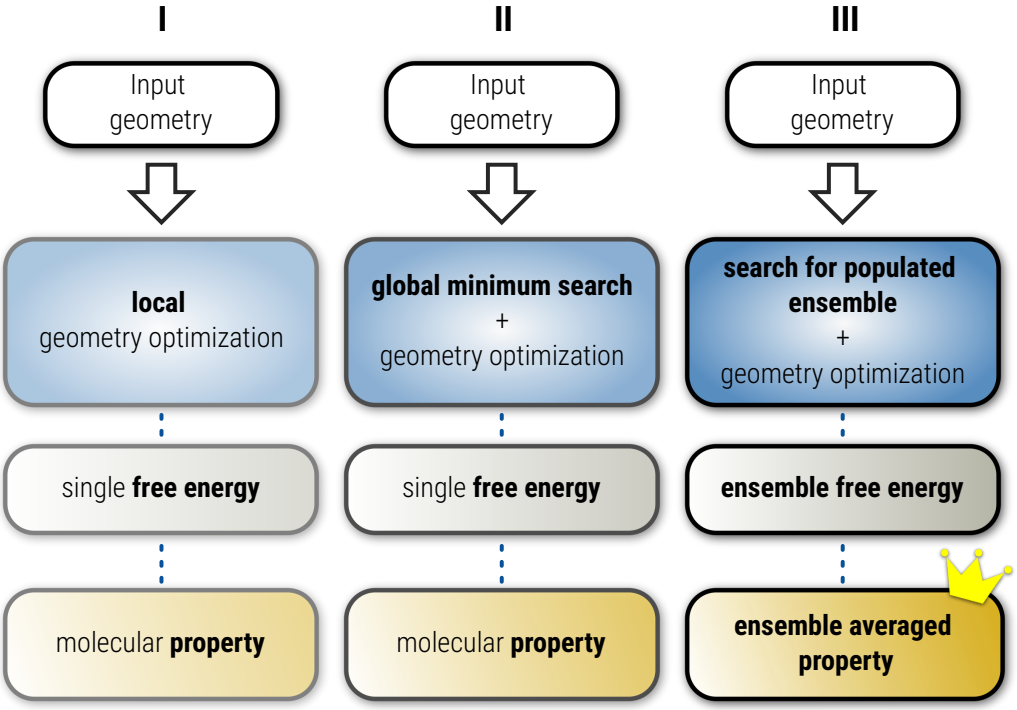


Figure 1.2.: Evolution of computational workflows for molecular property prediction.

Essential in chemistry is transforming molecules through breaking and forming chemical bonds. Here, transition metals, with their diverse coordination patterns and subtle changes in oxidation states, occupy an important position and largely expand the possibilities of synthetic organic chemistry.[64–66] However, describing the electronic structure of transition metals is more complicated than in pure organic chemistry and requires robust methods and often additional contributions from expert scien-

tists.[67] For future-oriented research and fundamental description of chemical systems, there should be no exclusion of "special" elements, and a universal investigation of the whole system should be aimed at. This demands the identification of versatile methods and the formulation of a consistent and automated exploration strategy.

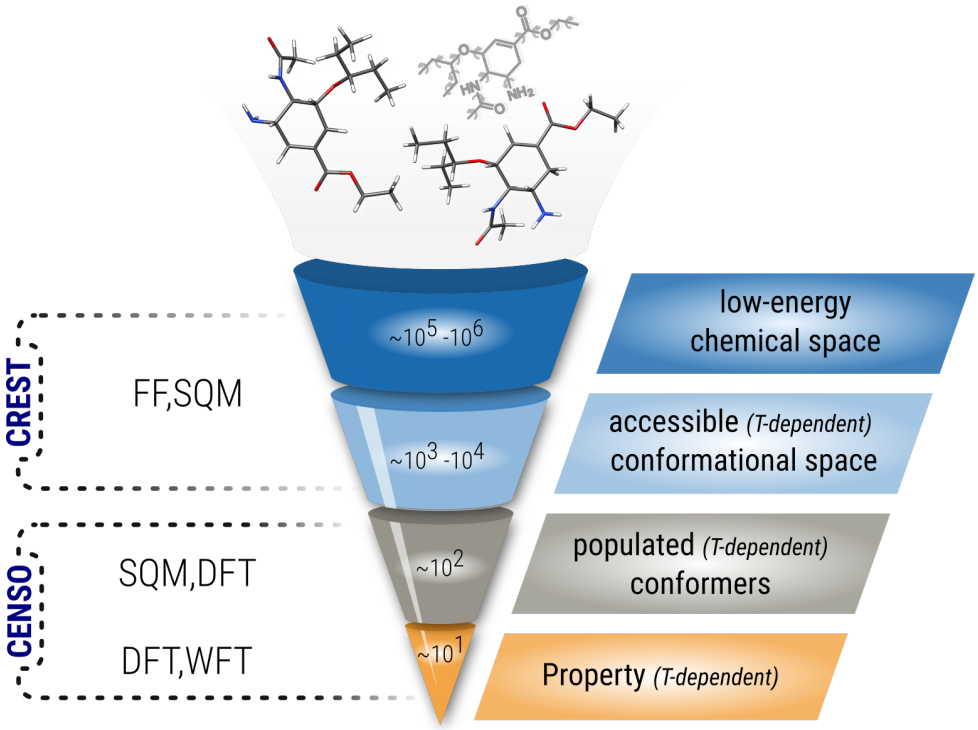


Figure 1.3.: Number of structures examined as part of a routine computational property-prediction project under standard conditions.

This thesis aims to develop and apply an automated universal workflow for calculating or predicting molecular properties. This task requires significant expertise to design, implement, and test a general-purpose computational chemistry workflow. For this purpose, the two programs CREST[68] (abbreviation from Conformer Rotamer Ensemble Sampling Tool) and CENSO[69] (abbreviated from Commandline Energetic Sorting) are applied. As depicted in Fig. 1.3 CREST can be used to sample the low-energy chemical space and condense it to the accessible conformational space. As shown in Fig. 1.3, the number of quantum mechanical calculations that are required to be performed before properties can be predicted is enormous, especially in the exploration of conformational space. Therefore, only fast methods like force fields or SQMs can be used. Together with the requirement for robustness and broad coverage of the periodic table, the newly developed GFN methods close a previous void in method development and are predestined for this application.[61] The GFN family of methods are special purpose methods that focus explicitly on their core goals: good geometries, harmonic vibrational frequencies, and non-covalent interactions. Using a coherent, element-specific design principle, these methods have been consistently parameterized for almost the entire PSE.

The main part of this work is devoted to developing the CENSO workflow, i.e., the efficient sorting

of structural ensembles using higher-level methods. Integral to this development is the availability of fast but consistently accurate methods like the recently published r^{2}SCAN-3c method.

Objectives

Overall, this thesis focuses on two main topics: the first topic is the evaluation and application of efficient newly developed computational approaches to describe Gibbs energies of organic and metal-organic systems. The second topic deals with the creation and investigation of robust automated workflows for property computations of nonrigid compounds. Chapter 2 of Part I provides an overview of the essential theoretical methods employed within this thesis. The following part (Part II) deals with applying different approaches to calculating free energies and properties based on real-world research questions, which are compared with observations from experiments. These include the determination of the strength of supramolecular interactions between organometallic planar π-systems, characterization of Pd^{II} complexes and investigation of their self-sorting behavior, identification of the best binding partner in a large Zn-salen macrocycle, and a benchmark study on the applicability of semiempirical GFN methods for the calculation of organo (-metallic) electrochemical redox potentials. Building on the findings and experiences of the projects above, Part III presents the development and application of an automated workflow for the quantum chemical calculation of the properties of nonrigid molecules. Part IV contains an overview of the realized results and offers an outlook on further goals in the discussed research field.

2. Theoretical Background

The following chapter provides an overview of the theoretical quantum mechanical (QM) methods applied in this thesis for the computational investigation of molecules based on physically motivated concepts. The prediction of macroscopic properties, which can be directly compared to the experiment, is a formidable and challenging task. To achieve a good comparison or even predictive power, all components required to calculate spectroscopic properties must be calculated at the best possible and, at the same time, the most efficient level. The presented theoretical background covers all the steps necessary for calculating macroscopic properties. This entails, first, an overview of modern Kohn-Sham density functional theory (KS-DFT) and fast semiempirical quantum mechanical (SQM) methods. Second, approaches are introduced to describe the thermodynamic contributions to the free energy and the solvation effects required to calculate the Gibbs free energies for describing thermodynamic properties in an equilibrium state. And third, strategies for calculating the molecular properties of optical rotation (OR) and nuclear magnetic resonance (NMR) are discussed. This chapter on the theoretical background is mainly based on Refs. 51,55,61,70,71. Technical standard procedures like structure optimizations or molecular dynamics are not discussed but are outlined in Refs. 51,71.

2.1. Overview of Methods

Various theoretical methods can be used to model different systems. It is the responsibility of the computational chemist to choose a suitable methodology for the respective chemical system and the task at hand. The requirements for the method in a computational chemistry project are to accurately describe the system and physically perform the calculations in the shortest possible time. In addition, the model's error must be within a specified range. A method with an acceptable degree of approximation to adequately represent the system can often already be chosen via the system's size. Figure 2.1 shows a simplified representation of the expected qualitative performance, referring only to the case of molecular systems.

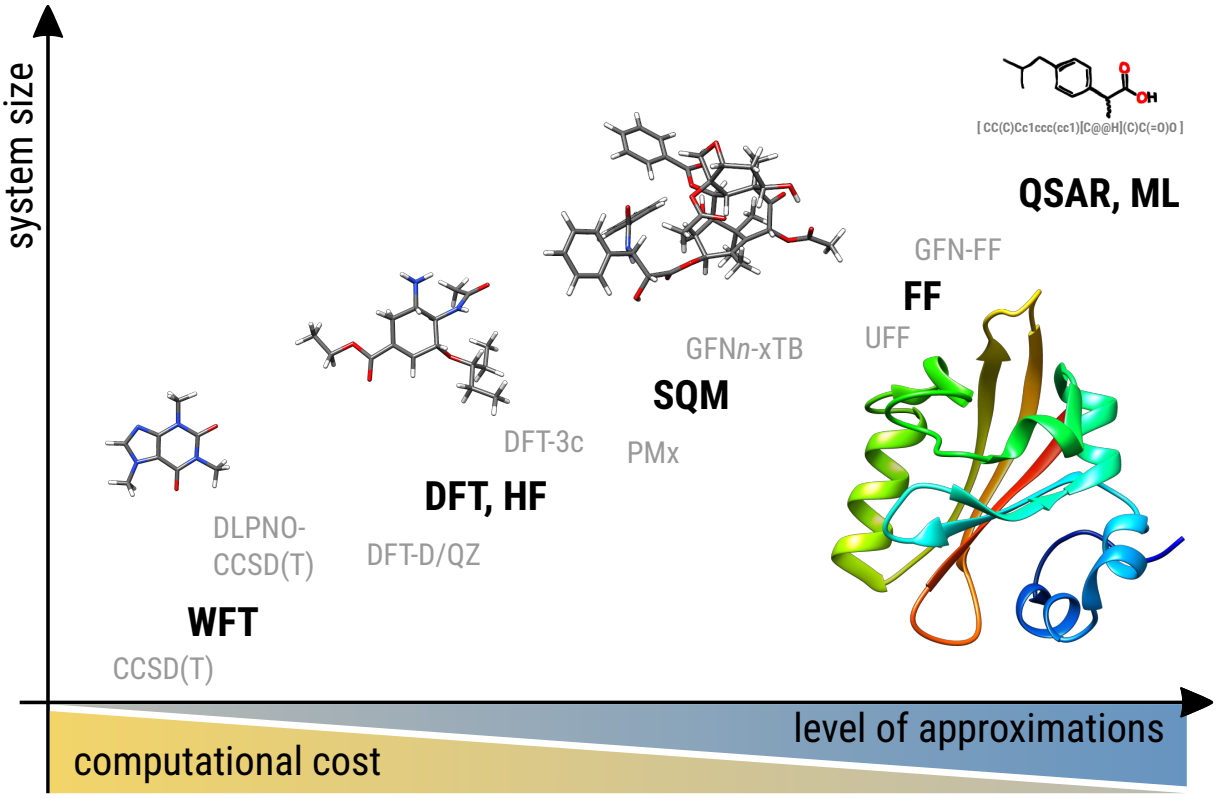


Figure 2.1.: Schematic classification of methods in terms of applicable system size and computational cost, as well as the degree of approximations within the respective models. The shown molecules are caffeine, tamiflu, paclitaxel, photoactive yellow protein[72], and ibuprofen.

However, it must be clearly pointed out that the choice of method-specific settings, the size of the basis set, the hardware, and code optimization can greatly impact the computational cost. The different fields of application, e.g., thermodynamics, kinetics, spectroscopy, place different demands on the methods, e.g., the description of geometries and vibrational frequencies, correct barrier heights, or accurate description of the asymptotic potential. When the system's size does not allow accurate but computationally intensive methods based on high-level wave function theory, faster models are used, which often involve some form of empiricism. The applicability and, more specifically, the accuracy of these methods can only be assured by comparative studies, and the comprehensive GMTKN55 benchmark study provides an excellent overview of the expected error ranges of different (mainly) density functional based models.[73]

Applications such as structural optimization, calculation of harmonic vibrational frequencies, or searching for molecular conformations often require between ten and several hundred thousand single-point and gradient calculations. The large number of required calculations limits the choice of methods, and for this reason, a multi-level approach has become common practice in the computational chemistry community.[69,74–76] This means that not only one method is used for all calculations, but that computationally intensive steps are carried out at an efficient level of theory. In contrast, more accurate and

more expensive calculations are usually applied in the final steps of the projects, which is discussed in more detail in Section A12.2.7.

2.2. Electronic Structure Methods

To study properties of molecular systems, relevant physical conditions have to be translated into a mathematical form, which can represent the system under investigation. The most important information of a system from which many other properties can be derived is the (electronic) energy. In the following, atomic units are used for clarity.[51,70] Molecules and atoms consist of positively charged nuclei (n) in the presence of negatively charged electrons (e). Light particles like fermions cannot accurately be described by classical mechanics and require a quantum mechanical description. The time-dependent (TD), non-relativistic Schrödinger equation (SE)[77]

$$i\hbar\frac{\partial}{\partial t}\Psi_K(t) = \hat{H}\Psi_K(t) \tag{2.1}$$

describes the quantum mechanical treatment between the time-dependent molecular wave function $\Psi_K(t)$ in the state K and the Hamiltonian $\hat{H}$, which is the operator for constructing the total energy of a molecular system. For stationary states, the separation of time and space variables in the Hamilton operator leads to the time-independent SE[51,70], where the phase factor of the wave function is typically neglected.

$$\hat{H}\Psi_K = E_K\Psi_K \tag{2.2}$$

The stationary state K has a constant energy in time and only depends on spatial and spin information. The non-relativistic Hamilton operator is usually expressed in terms of the potential and kinetic energy of the nuclei and electrons:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{ne} + \hat{V}_{ee}\ . \tag{2.3}$$

The Hamiltonian entails operators for the kinetic energy of the nuclei ($\hat{T}_n$) and the electrons ($\hat{T}_e$), the nuclear–nuclear repulsion ($\hat{V}_{nn}$), the nuclear–electron attraction ($\hat{V}_{ne}$) and the electron-electron repulsion ($\hat{V}_{ee}$). Eq. 2.3 can be simplified by invoking the Born–Oppenheimer[78] approximation. Here, the light fermions are treated as quantum mechanical particles in the field of the much heavier nuclei, which are handled classically and enter the SE as parameters. This leads to the formulation of the electronic Hamiltonian, where all terms depending on the nuclei ($\hat{T}_n$, $\hat{V}_{nn}$) are constant, i.e., have no effect

on the operator eigenfunctions. For a system with N electrons and M nuclei the equation is:

$$\hat{H}_e = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} \tag{2.4a}$$

$$= \sum_i^N \hat{h}_i + \hat{V}_{ee} \tag{2.4b}$$

$$= -\frac{1}{2}\sum_i^N \hat{\nabla}_i^2 - \sum_i^N \sum_A^M \frac{Z_A}{|\mathrm{r}_i - \mathrm{R}_A|} + \sum_{i>j}^N \frac{1}{|\mathrm{r}_i - \mathrm{r}_j|} \ . \tag{2.4c}$$

The electronic Hamiltonian $\hat{H}_e$ can be divided into the one-electron part $\hat{h}_i$ describing the kinetic energy of electron i and the Coulomb interaction of the ith electron with the stationary nuclei and the two-electron term $\hat{V}_{ee}$ responsible for the repulsive Coulomb interaction between the electron pairs ij. In Eq. 2.4 Z corresponds to the nuclear charge of atom A, whereas r and R give the position vectors of the respective electrons and nuclei. The electronic energy of the system is obtained by means of the expectation value of the electronic Hamilton operator:

$$\int_{-\infty}^{\infty} \Psi_K^* \hat{H}_e \Psi_K \mathrm{d}r \equiv \left\langle \Psi_K \left| \hat{H}_e \right| \Psi_K \right\rangle \equiv H_{KK} = E_K \left\langle \Psi_K \mid \Psi_K \right\rangle \ , \tag{2.5}$$

where the integration is performed over the entire space r, which is expressed in Dirac's bra and ket notation.[79] By employing an orthonormalized wave function Ψ_K in Eq. 2.5, the overlap integral on the right-hand side vanishes, i.e.,

$$\langle \Psi_K | \Psi_L \rangle = \delta_{KL} \ , \tag{2.6}$$

where the Kronecker delta δ_{KL} returns a value of zero if K ≠ L or one if K = L. Obtaining solutions of the time-independent SE, i.e. exact eigenfunctions of $\hat{H}_e$, is a non-trivial task. A valuable guideline in finding possible solutions is the *variational principle*:

$$\tilde{E}_K = \left\langle \tilde{\Phi}_K | \hat{H}_e | \tilde{\Phi}_K \right\rangle \geq \left\langle \Psi_K | \hat{H}_e | \Psi_K \right\rangle \ , \tag{2.7}$$

which states that the expectation value of a trial wave function ($\tilde{\Phi}_K$) can not yield a lower value than the true wave function Ψ_K.[70]

2.2.1. Hartree–Fock Theory and Roothaan–Hall Equations

Starting from the variational principle, in Hartree–Fock[80,81] theory, the exact ground-state wave function is approximated by a single *Slater determinant* (SD) (Eq. 2.8). The SD wave function of an N electron system exhibits the correct anti-symmetry requirements of the *Pauli principle*[82] with respect to the interchange of any two electrons. The orthonormalized Φ_0 is constructed from one-electron functions,

i.e., molecular orbitals (MOs) ϕ_i, which are the product of a spatial orbital and a spin function.

$$\Psi_K \approx \tilde{\Phi}_0 \equiv \Phi_0(\underline{1}, \underline{2}, \cdots, \underline{\mathrm{N}}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\underline{1}) & \phi_2(\underline{1}) & \cdots & \phi_N(\underline{1}) \\ \phi_1(\underline{2}) & \phi_2(\underline{2}) & \cdots & \phi_N(\underline{2}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\underline{\mathrm{N}}) & \phi_2(\underline{\mathrm{N}}) & \cdots & \phi_N(\underline{\mathrm{N}}) \end{vmatrix} . \tag{2.8}$$

Using the SD as a trial wave function, the next goal is to determine MOs ϕ_i, which yield an energy minimum or stationary state, under the condition that they remain orthogonal and normalized. This constrained optimization can be solved by means of Lagrange multipliers:

$$\mathcal{L} = E - \sum_{i,j}^{N} \lambda_{ij} \left(\langle \phi_i | \phi_j \rangle - \delta_{ij} \right) \tag{2.9a}$$

$$\partial\mathcal{L} = \partial E - \sum_{i,j}^{N} \lambda_{ij} \left(\langle \partial\phi_i | \phi_j \rangle - \left(\langle \phi_i | \partial\phi_j \rangle \right) \right) = 0 \ . \tag{2.9b}$$

These conditions result in Eq. 2.10

$$\hat{f}_i \phi_i = \sum_{j}^{N} \lambda_{ij} \phi_j \ , \tag{2.10}$$

and a simplification by unitary transformation leads to

$$\hat{f}_i \tilde{\phi}_i = \epsilon_i \tilde{\phi}_i \ . \tag{2.11}$$

The pseudo-eigenvalue equation 2.11 is known as the canonical Hartree–Fock equation. Here, the diagonal elements of the transformed Lagrange multipliers matrix λ have the physical interpretation of MO energies, and the set of molecular orbitals ($\tilde{\phi}$) is known as canonical MOs. The Fock operator ($\hat{f}$)

$$\hat{f}_i(\mathrm{r}_1) = \hat{h}_i(\mathrm{r}_1) + \sum_{j}^{N} \left(\hat{J}_{ij}(\mathrm{r}_1) - \hat{K}_{ij}(\mathrm{r}_1) \right) \ , \tag{2.12}$$

is an effective one-electron operator with the MOs as eigenfunctions and orbital energies ϵ_i as eigenvalues. As previously in Eq. 2.4b $\hat{h}_i$ describes the electron-nuclei interaction and kinetic energy of a single electron.

$$\hat{J}_{ij} |\phi_i\rangle = \langle \phi_j | \frac{1}{r_{ij}} | \phi_j | \phi_i \rangle \tag{2.13a}$$

$$\hat{K}_{ij} |\phi_i\rangle = \langle \phi_j | \frac{1}{r_{ij}} | \phi_i | \phi_j \rangle \tag{2.13b}$$

The Coulomb operator $\hat{J}_{ij}$ describes the Coulomb repulsion between all electrons, while the (Pauli) exchange operator $\hat{K}_{ij}$, arising from the anti-symmetry requirement of the wave function, describes the

interaction between electrons of the same spin. Each electron experiences the field of N-1 electrons, which means that a single electron does not have any self-interaction error (SIE), and the interaction for $\langle\phi_i|\hat{J}_{ii}|\phi_i\rangle$ and $\langle\phi_i|\hat{K}_{ii}|\phi_i\rangle$ exactly cancels each other. Since the summation in the Fock operator runs over all occupied MOs, a consequence of using a single Slater determinant, the electron-electron interaction is provided only in an averaged fashion, and for this reason, Hartree-Fock is called a mean-field theory. Following that the Fock operator depends on the molecular orbitals themselves; a solution can only be obtained iteratively in a self-consistent field procedure, where the spatial orbitals have to be systematically varied to arrive at a stationary solution. Up to this point, solving the integro-differential Hartree-Fock equations is very computationally demanding.

A practical approach to expressing unknown MOs by a set of known mathematical functions was introduced by Roothaan and Hall. Each molecular orbital is expanded as a linear combination of atomic orbitals (LCAO).[83,84] The spatial part of MO ϕ_i is approximated by M predefined atomic orbitals or basis functions χ_μ with $C_{\mu i}$ being molecular orbital coefficients.

$$\phi_i(\mathrm{r}_1) = \sum_{\mu}^{M} C_{\mu i}\chi_\mu(\mathrm{r}_1) \ . \tag{2.14}$$

In the limit of a complete basis set (CBS) the LCAO approach is an exact solution. Reformulating the Hartree–Fock eigenvalue problem by expanding MOs in a basis set leads to the Roothaan–Hall matrix equation:

$$\mathrm{FC} = \mathrm{SC}\epsilon \ . \tag{2.15}$$

In Eq. 2.15, the $F_{\mu\nu}$ element of the Fock matrix is given by $\langle\chi_\mu|\hat{f}|\chi_\nu\rangle$, the overlap of basis functions $\langle\phi_i|\phi_j\rangle$ is stored in matrix S, the LCAO coefficients are kept in matrix C and the orbital energies (eigenvalues) are provided in ϵ. By using a predetermined, known set of basis functions, the only unknowns are the elements of the coefficient matrix C, which have to be determined through variational optimization. After determining the MOs via the LCAO coefficients using the *self-consistent field* (SCF) method, the final Hartree-Fock energy can be determined as follows

$$E_{HF} = \sum_{i}^{N}\langle\phi_i|\hat{h}_i|\phi_i\rangle + \sum_{i}^{N-1}\sum_{j>i}^{N}(\langle\phi_i|\hat{J}_{ij}|\phi_i\rangle - \langle\phi_i|\hat{K}_{ij}|\phi_i\rangle) \ . \tag{2.16}$$

By adding the nuclear-nuclear repulsion energy V_{nn} to Eq. 2.16 it yields the total energy E. The Hartree-Fock method converts the solution of the Schrödinger differential equation into the linear algebra Roothaan-Hall eigenvalue equation, for which computers are best suited to solve. The computational cost associated with HF formally scales with $O(N^4)$, where N is the number of basis functions (AOs) used. Computational speed-ups can be achieved by the *resolution of the identity* (RI) or density fitting.[85,86] Here, the product of two basis functions can be expanded by an auxiliary basis set, approximating the calculation of challenging two-electron four-center integrals as products of integrals of lower rank.

2.2.2. Electron Correlation

As a mean-field theory, HF describes non-interacting electrons that experience the average interaction of the remaining N-1 electrons. This averaged interaction already recovers 99 % of the true non-relativistic total energy in the Born–Oppenheimer approximation. The difference between the true energy of fully interacting fermions and the Hartree-Fock energy (in a sufficiently large basis set) is defined as electron correlation energy:[87]

$$E_{corr} = E_{tot} - E_{HF} \tag{2.17}$$

While HF already includes correlated motion of same-spin electrons via the Pauli-exchange (Fermi correlation), the Coulomb repulsion-driven correlation of same-spin electrons (Coulomb correlation) is neglected. Though the magnitude of electron correlation in absolute energies is small, its significance becomes pronounced with respect to relative energies, covalent bonding, and van-der-Waal complexes. Moreover, the long-range part of the electron correlation is necessary to describe attractive London dispersion forces dominant in non-covalently bound systems,[88], which pure Hartree–Fock theory is unable to do. This missing correlation energy can be recovered by post-Hartree-Fock, electron correlated wave function theory (WFT) methods like Rayleigh–Schrödinger perturbation theory[70] (PT), configuration interaction[89,90] (CI) or coupled cluster[91,92] (CC) theory. All of these approaches start from a single determinant and expand the trial wave function by introducing determinants with electrons exited to virtual orbitals.

In perturbation theory, the assumption is that the overall problem is already (approximately) solved and that the true solution to the problem can be expressed as a small perturbation. The Hamilton operator is expressed as the sum of the unperturbed zeroth-order Hamiltonian $\hat{H}_0$ and the product of the perturbation parameter and perturbation operator $\lambda\hat{H}'$. The full perturbation is expressed as a power series, and the termination of the series leads to different orders of perturbation. In the most common PT variant of Møller–Plesset[93] theory, the unperturbed Hamiltonian is chosen as a sum over Fock-operators. MP1, in the first order, provides the HF energy, and electron correlation is described in the second order, where only doubly excited determinants contribute. The MP2 correlation energy is obtained from:

$$E_{corr}^{MP2} = \sum_{j>i}^{occ.} \sum_{b>a}^{virt.} \frac{[(ij|ab) - (ia|jb)]^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}, \tag{2.18}$$

where the indices *i,j* indicate occupied- and *a,b* virtual orbitals respectively. Because of the necessary transformation of the AO to MO basis, MP2 formally scales with $O(N^5)$. Since HF already includes Fermi correlation, different spin-scaled variants of MP2[94] (SCS-MP2) exist, where the components of same-spin- (SS) or opposite-spin (OS) correlation are empirically adjusted to provide a balanced description. The best-known non-perturbative approaches to describing electron correlation are CI and CC. Both approaches solve the Schödinger equation by including additional determinants into the wave function to simulate excitations. An excitation operator $\hat{T}$ is applied onto the HF ground-state Slater

determinant Φ_0 describing exited determinants by exciting electrons into virtual orbitals.

$$\Psi_{CI} = (1 + \hat{\mathrm{T}})\Phi_0 = (1 + \sum_{\lambda}^{\infty} \hat{T}_\lambda)\Phi_0 \tag{2.19a}$$

$$\Psi_{CC} = \exp(\hat{\mathrm{T}})\Phi_0 = \sum_{k=0}^{\infty} \frac{1}{k!}\hat{\mathrm{T}}^k\Phi_0 \tag{2.19b}$$

The CI or CC wave function is determined from higher determinants and optimized expansion coefficients or amplitudes. In the limit of a complete basis set and all possible excitations, the non-relativistic result in the Born–Oppenheimer approximation is exact and known as *FullCI* (FCI). Due to practical limitations, i.e. the sharp rise in computational cost, the expansions have to be truncated. In CI, the excitation operator is applied linearly and as a Taylor expansion in CC. Truncation in CI has the disadvantage that the important property of *size consistency* is lost. Whereas truncating the expansion in coupled cluster has the advantage that higher order excitations are implicitly taken into account to some extent and *size consistency* is maintained. For this reason, the CC variant with single, double, and perturbative connected triple excitations CCSD(T) has established itself as the "gold standard" in quantum chemistry and is often used as a reference method for benchmarking faster, more approximate methods.[73,92]

Method	Scaling	Information
HF	$O(N^4)$	Hartree-Fock, occupied orbitals only
MP2	$O(N^5)$	Møller–Plesset perturbation theory
CISD	$O(N^6)$	Configuration interaction limited to single and double excitations
CCSD(T)	$O(N^7)$	Coupled cluster with singles, doubles and perturbative triples
FCI	$O(N!)$	Exact solution to the SE

Table 2.1.: Formal scaling behavior of QM methods based on their system size N.

However, even with the truncated CCSD(T) theory, the formal computational cost scales as $O(N^7)$ and can only be applied to small systems (see Tab. 2.1). Modern versions like the DLPNO-CCSD(T)[95–97] allow the treatment of somewhat larger systems by efficient localization schemes and the exploitation of generated sparsity while retaining the overall good accuracy.

2.2.3. Kohn-Sham Density Functional Theory

A possible alternative to computationally demanding WFT represents (Kohn–Sham) density functional theory. While in WFT, the energy is obtained from the wave function, in DFT, the electron density $\rho(r)$ of the system is the fundamental quantity.[98] The electron density is an observable exhibiting cusps at the nuclei positions. The slope at the nuclei position provides information on the nuclear charge Z, and integration over the entire electron density returns the number of electrons N in the system. Already in 1927, Thomas and Fermi formulated the idea that ground state properties can be extracted from the observable ρ.[99,100] Hohenberg and Kohn proved in 1964 in their first theorem that a functional exists

that relates the electron density in the ground state with the energy of the system.[52]

$$F_{exact}[\rho(r)_{exact}] = E_{exact} \tag{2.20a}$$

$$F_{exact}[\rho(r)_{trial}] \geq E_{exact} \tag{2.20b}$$

According to the second Hohenberg-Kohn theorem (Eq. 2.20b), the electron density obeys the variational principle, i.e., the density that minimizes the total energy is the exact ground-state density. However, the Hohenberg-Kohn theorems do not provide any information for generating and or systematically improving the density functional. As the exact functional is unknown and may be prohibitively expensive, the development of approximate functionals plays an important role in ongoing research. In analogy to the electronic Hamilton operator, the density-dependent contributions to the final energy can be partitioned and formulated as:

$$F[\rho] = T_e[\rho] + V_{ne}[\rho] + V_{ee}[\rho] \tag{2.21a}$$

$$= T_e[\rho] + V_{ne}[\rho] + J[\rho] + K[\rho] \quad . \tag{2.21b}$$

The contributions are the kinetic energy of the electron $T_e[\rho]$, the interaction of the electrons with the nuclei $V_{ne}[\rho]$ and the electron-electron interaction $V_{ee}[\rho]$ which can be split into a Coulomb part $J[\rho]$ and an exchange part $K[\rho]$. This purely density-driven approach has the desirable advantage of depending on only three spatial variables - a huge reduction from the 3N electron variables in WFT - and is often referred to as orbital-free DFT (OF-DFT).[101] A representative of the OF-DFT class is the Thomas-Fermi-Dirac model:[99,100,102]

$$\begin{aligned} F_{TFD}[\rho] = \frac{3}{10}\left(3\pi^2\right)^{\frac{2}{3}} \int \rho^{\frac{5}{3}}(r)dr - \sum_{a}^{nuclei} \int \frac{Z_a(R_a)\rho(r)}{|Ra - r|}dr \\ + \frac{1}{2}\int\int \frac{\rho(r)\rho(r')}{|r - r'|}drdr' - \frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho^{\frac{4}{3}}(r)dr \quad . \end{aligned} \tag{2.22}$$

However, the accuracy, even in refined modern versions, is too low to be usable for atoms and molecules.[103,104] Here, the largest source of error is found in the kinetic energy functional and the interelectronic description in V_{ee}. Kohn and Sham (KS) proposed the idea of using a fictitious system of non-interacting quasi-particles in a mean-field potential, which has the same density as the real system.[105] The electron density approximated from a Slater determinant is given by:[57]

$$\rho_{approx}(r) = \sum_{i}^{N} \langle \psi(r) | \psi(r) \rangle = \sum_{i}^{N} |\psi(r)|^2 \quad . \tag{2.23}$$

This strategy has the advantage of being similar to the Hartree-Fock method, using (known) formulas, but it also reintroduces an orbital (KS-orbitals) dependence with $3N$ instead of 3 electron coordinates.[106]

In the orbital expression of the electron density, the Kohn-Sham-DFT energy can be written as:

$$E_{KS}[\rho] = T_S[\rho] + V_{ne}[\rho] + J[\rho] + E_{XC}[\rho] \tag{2.24a}$$

$$= -\frac{1}{2}\sum_{i}^{N}\langle\psi_i|\nabla^2|\psi_i\rangle + V_{ne}[\rho] + J[\rho] + E_{XC}[\rho] \ . \tag{2.24b}$$

The kinetic energy T_S of non-interacting electrons in HF is exact and already recovers the largest part of the kinetic energy of the fully interacting system. Here, the $E_{XC}[\rho]$ is introduced where the remaining difference between the non-interacting and exact kinetic energy is absorbed along with the correlation and exchange energy:

$$E_{XC}[\rho] = (T[\rho] - T_S[\rho]) + (V_{ee}[\rho] - J[\rho]) \tag{2.25a}$$

$$= \int \rho(r)\epsilon_{xc}[\nabla^\gamma\rho(r)]dr \ . \tag{2.25b}$$

A more general formulation of the (semi-)local exchange-correlation functional $E_{XC}[\rho]$ is shown in Eq. 2.25b, where ϵ_{xc} represents the energy per particle (energy density) and Δ^γ indicates the dependence on the electron density or its derivatives. Often, ϵ_{xc} is divided into two separated expressions, thereby being able to treat the correlation and exchange part differently, even though in the exchange-correlation hole picture, only the combination has a physical meaning. Finding orbitals that minimize the energy from Eq. 2.24 under the orthonormal constraint leads to the Kohn–Sham pseudo eigenvalue equations:[51]

$$\hat{f}_i^{KS}[\rho]\psi_i = \epsilon_i\psi_i \ , \tag{2.26}$$

with the Kohn–Sham operator,

$$\hat{f}_i^{KS}[\rho] = \hat{h}_i[\rho] + \sum_{j}^{N}\hat{J}_{ij}[\rho] + v_{XC}[\rho] \tag{2.27}$$

$$v_{XC} = \frac{\delta E_{XC}[\rho]}{\delta\rho} \ . \tag{2.28}$$

The exchange-correlation potential $v_{XC}[\rho]$ is the functional derivative with respect to the electron density. The similarity to the Fock operator is evident, and the same strategies as in HF for solving Eq. 2.26 can be applied. There are some theoretical similarities between HF and KS-DFT (hereafter DFT) in that both are single-determinant approaches based on the LCAO/Roothaan-Hall formalism. However, DFT has the potential to go beyond the mean-field theory and potentially provide the exact total energy, including electron correlation, if the exchange-correlation potential is known.[51,57] Further, the (exact) exchange-correlation potential influences the KS-orbitals, which have different shapes compared to HF-orbitals. Unoccupied KS-orbitals are created in the field of N-1 electrons, whereby orbital energy differences better correspond to excitation energies.[107] A pressing drawback of DFT is that there exists no guideline for systematic improvement like in WFT, and as a result, many different

density functional approximations (DFA) have been introduced over the years.[57,73] Among the most important design strategies to find suitable formulations of v_{XC} is the satisfaction of physically imposed constraints or a high degree of empiricism by fitting to experimental properties.[56,108,109] In this thesis, we use the term "DFT" synonymously for approximate KS-DFT. In an attempt to categorize the available DFAs, Perdew and Schmidt ranked the functionals by expected accuracy and computational cost ratio in the image of the biblically inspired "Jacob's Ladder", covering the range from the "Hartree World" to the "Heaven" of chemical accuracy.[57] A short overview of each rung will be discussed briefly in the following.

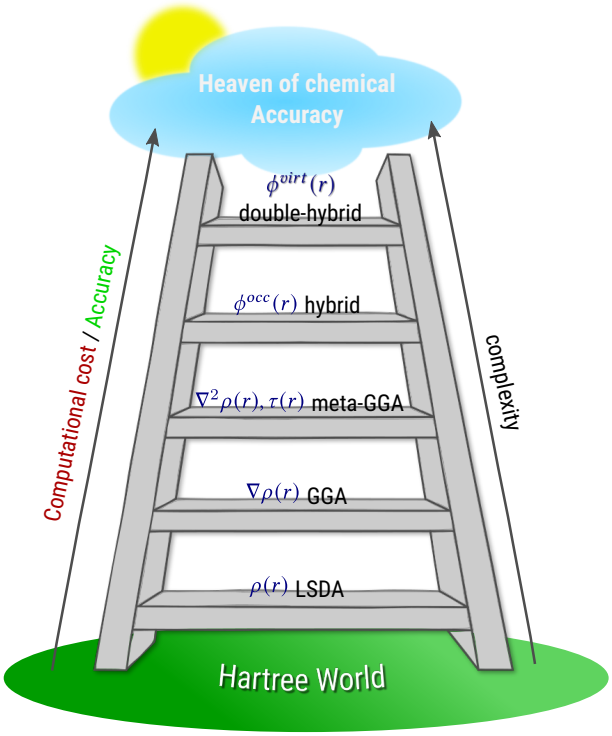


Figure 2.2.: Perdew and Schmidt's "Jacobs ladder" categorizing density functional approximations (DFA).[57] $\rho(r)$ is the electron density, ∇ the Nabla operator, τ the kinetic energy density and $\phi(r)$ is a molecular orbital.

Local (Spin) Density Approximation

The first rung of the "Jacob's ladder"comprises functionals that depend exclusively on the local density value of ρ at the position r. This is well defined only for the uniform electron gas (UEG), where the electron density is assumed to vary slowly (i.e., no cusps). The universal case in which the ρ_α and ρ_β spin densities can be different is called the Local Spin Density Approximation (LSDA), which is equal to the Local Density Approximation (LDA) for closed-shell systems. The exchange functional has been

proposed by Dirac and also finds use in the orbital-free Thomas-Fermi-Dirac model (see Eq. A1.5).[102,110]

$$E_X^{LSDA}[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \left(\rho_\alpha(\mathrm{r})^{\frac{4}{3}} + \rho_\beta(\mathrm{r})^{\frac{4}{3}}\right) \mathrm{dr} \ , \tag{2.29}$$

Electron correlation functionals for the UEG[110] have been realized by analytic interpolation formulas based on Monte-Carlo calculation results which are known as VWN[111] (Vosko, Wilk, and Nusair) or PW[112] (Perdew and Wang). LSDA has found successful application in the physics community for describing metallic systems where the UEG approximation is justified. For molecular systems, LDA tends to overbind and is usually too inaccurate.

Generalized Gradient Approximation

Since the electron density of molecular systems deviates considerably from the spatially uniform assumption in the UEG, the local change in density, i.e. the gradient $\nabla\rho$, must be taken into account for an improved description. In the generalized gradient approximation (GGA), this can be realized by modification of the LSDA with an enhancement factor $F_{XC}^{GGA}[\rho, \nabla\rho]$ which is density and gradient dependent.

$$E_{XC}^{GGA}[\rho] = \int \varepsilon_{XC}^{LSDA}[\rho(\mathrm{r})]\, F_{XC}^{GGA}[\rho(\mathrm{r}), \nabla\rho(\mathrm{r})]\, \mathrm{dr} \ . \tag{2.30}$$

Due to the additional consideration of density changes in the immediate vicinity of the local density at position r, this method is called semi-local. There are many different enhancement factors for the exchange and correlation part, and the best known are the Perdew-Burke-Ernzerhof[113] (PBE), B88[114] Becke exchange, and the Lee-Yang-Parr[115] (LYP) correlation functionals. Compared to LSDA the description of molecules is improved, particularly the description of molecular geometries. Unfortunately, the semi-local contribution leads to an incomplete integration of the Fermi hole. This introduces the self-interaction error (SIE) i.e., a Coulomb interaction of an electron with itself, which results in an over-delocalized description of the electron density.[116]

Meta-Generalized Gradient Approximation

The next logical improvement leads to the meta-generalized gradient approximations (meta-GGA), which contain higher-order derivatives of the electron density $\nabla^2\rho$ in addition to the basic GGA framework. This new rung on the Jacobs ladder is typically for reasons of numerical stability, evaluated by the orbital kinetic energy density $\tau(r)$ instead of the Laplacian of the electron density.[117]

$$\tau(r) = \frac{1}{2} \sum_{\sigma=a,b} \sum_{i}^{N_\sigma} |\nabla\psi_i(r)|^2 \tag{2.31}$$

Here, $\tau(r)$ is calculated from the KS-orbitals ψ_i, which are shaped by the effective potential. Incorporating even more information from the electron density results in only a minor improvement over GGAs.

Popular versions of meta-GGAs are the Tao-Perdew-Staroverov-Scuseria[118] (TPSS) or the strongly constrained and appropriately normed[108] (SCAN) functional.

Hybrid DFT

The fourth rung in the Jacob's ladder classification refers to hybrid DFT methods. Here, the concept is to construct a connection between the exact Hartree–Fock exchange and exchange incorporated in the (true) density functional. This approach is motivated by the Adiabatic Connection[119] (AC) relating the non-interacting to the fully-interacting actual system. By assuming a linear correlation between the two limits and approximating the fully interacting system by meta-GGA exchange-correlation, the empirical global hybrid DFT formula is derived:[120]

$$E_{XC}^{hybrid}[\rho] = a_X E_X^{HF} + (1 - a_X) E_X^{DFT} + E_C^{DFT} \quad . \tag{2.32}$$

The non-local Fock-exchange E_X^{HF} is evaluated from occupied KS-orbitals, and the fitting parameter a_X determines the amount of exact exchange used. Representative global hybrid functionals are B3LYP[120,121] with 20 % Fock exchange, PBE0[122] with 25 % or BHLYP[120] with 50 % exact exchange. The positive impact of introducing exact exchange is a reduction of the SIE occurring in semi-local functionals and all the associated shortcomings, such as insufficient barrier heights, over-delocalization, and false asymptotic behavior of the V_{XC} potential. The correct $1/r$ asymptotic behavior is not realized even with global hybrids, and to address this inconvenience, a partitioning of the electron-electron exchange into a short- and long-range regime can be performed by an error function:[123–126]

$$\frac{1}{r_{12}} = \underbrace{\frac{1 - erf(\omega r_{12})}{r_{12}}}_{short-range} + \underbrace{\frac{erf(\omega r_{12})}{r_{12}}}_{long-range} \tag{2.33}$$

where the parameter ω manages the interpolation. By means of this classification, the ranges can be described by different exchange functionals, and typically, the short range is described by a density functional and the long-range part by exact Fock exchange. This approach provides improved excitation energies for charge transfer states by repairing the underestimated gaps in global hybrids between occupied and virtual orbitals caused by SIE.[107,127,128]

Double Hybrid DFT

The fifth and highest level of the Jacob's ladder is claimed by the so-called double hybrid density functionals (DHDF).[129–131] Following the concept of hybrid functionals, i.e., using occupied KS-orbitals for supplementation of the exchange potential, the DHDFs further make use of the virtual KS-orbital space by replacing parts of the correlation energy with a perturbation MP2-like energy or random phase approximation treatment, where the general expression is given by:

$$E_{XC}^{DHDF}[\rho] = a_X E_X^{HF} + (1 - a_X) E_X^{DFT} + b_C E_C^{DFT} + (1 - b_C) E_C^{PT2} \quad . \tag{2.34}$$

Generally, DHDFs provide a very good description of thermo-chemistry and of overall basic molecular properties.[73] The drawback is computational size dependence, which increases with the MP2-[93] or RPA-like[132,133] correlation treatment to a formal scaling of $O(N^5)$ and slower convergence with the size of the basis set compared to non-DHDFs.

Dispersion corrected DFT

A major drawback of mean field methods (MF) like HF or semi-local KS-DFT is their inability to describe the long-range attractive part of van-der-Waals interaction, also known as London dispersion (LD).[73,134,135] This severely limits the accuracy of these methods when describing chemical (non-covalent) systems, as dispersion is always present and very important for correct inter- and intra-molecular description. In the spirit of addressing systematic deficiencies, such as those seen in the development of the Jacob's Ladder, correction schemes have been developed to include LD. From perturbation theory, the dispersion energy of two atoms A and B in the asymptotic limit (large R) is known from Eisenschitz and London since 1930:[136,137]

$$E_{disp}^{AB} \approx -\frac{3}{2}\frac{I_A I_B}{I_A + I_B}\alpha_A^0 \alpha_B^0 R^{-6} = -C_{6,approx}^{AB} R^{-6} \quad , \tag{2.35}$$

where α^0 is the static dipole polarizability, I represents the atomic ionization potential and the pair specific C_6 coefficient determines the interaction strength of A and B. Several dispersion corrections have since been developed,[138–140] but we limit ourselves to the DFT-D dispersion correction used in this thesis.[141] It has been found that semi-classical (pairwise) dispersion corrections to the energy can be simply added to the mean field method energy.[134]

$$E_{tot} = E_{MF} + E_{disp} \tag{2.36}$$

The widely used D3 correction[134,142] owes its success to the fact that it depends only on the input geometry, requires no further input from a self-consistent calculation, i.e., imposes no density or charge requirements, significantly improves the results, and the computation time is negligible compared to the SCF mean-field calculation.

$$E_{disp}^{\mathrm{D3},AB} = -\sum_{n=6,8} E_{disp}^{n} = -\sum_{AB}\sum_{n=6,8} s_n \frac{C_n^{AB}}{R_{AB}^{(n)}} f_{damp}^{(n)}(R) \tag{2.37}$$

The D3 approach includes the most important leading order dipole-dipole $E_{DISP}^{(6)}$ and dipole-quadrupole $E_{DISP}^{(8)}$ interactions. In Eq. 2.37 the pairwise dispersion energy is calculated from pre-calculated $C_{(n)}^{AB}$ nth order dispersion coefficients, the parameter s_n is adjusted to the underlying exchange-correlation functional as it can already account for some short-range interaction and a damping function which attenuates the interaction at short distances R[143]

$$f_{damp,BJ}^{(n)}(R) = \frac{R^n}{R^n + (a_1\sqrt{C_8^{AB}/C_6^{AB}} + a_2)^n} \, . \tag{2.38}$$

The most sophisticated attenuation function was suggested by Becke and Johnson to dampen the dispersion interaction to a lower constant value at short distances, which is more physical than the earlier proposal of an attenuation to zero.[144,145] The combination of the zero or Becke-Johnson damping function with the D3 model is referred to as *D3(0)* or *D3(BJ)*. The flexibility to describe different chemical environments is achieved by the coordination number-dependent generation of the pairwise $C_{(n)}^{AB}$ coefficients. The coordination number reflects the chemical environment and, using carbon as an example, simulates the different hybridization states. In addition to the pairwise two-body interactions, three-body contributions can be investigated by consideration of the Axilrod-Teller-Muto (ATM) term, which is not further discussed here.[146,147]

The DFT-D4 model was recently introduced to replace the previous DFT-D3 dispersion correction.[148–150] The main improvement over D3 is that atomic charge dependent reference polarizabilities $\alpha(i\omega)$ and modified electronegativity dependent coordination numbers (CN) are employed to calculate $C_{(n)}^{AB}$ coefficients. Basically, in addition to the pure geometry dependence, D4 contains information about the electronic density and systematically improves the number of pre-computed reference polarizabilities for different CNs. The improvements achieved by dispersion corrections are considerable and, in the case of DFT-D methods, involve low computational costs and must now be regarded as standard procedures.

Composite scheme DFA

Since their introduction in 2013, composite methods referred to as "-3c methods" have found a steadily growing number of users. While, in principle, basis set converged results from a dispersion-corrected DFT calculation are usually sufficient, their performance is often too slow to be routinely applied to large systems or in very large numbers. The "-3c" (three corrections) composite schemes are created with the design goals of low computational cost while simultaneously retaining a balanced and consistent description, i.e., no systematic deviations.[151–156] Central to the “3c“ strategy are relatively small, optimized, and balanced AO basis sets, which reduce the computational effort compared to methods with an almost converged number of basis functions. Errors that occur due to the design strategy, such as basis set superposition error (BSSE), or that are inherent to the established underlying DFAs (HF), e.g., missing long-range London dispersion, are corrected for by specifically adjusted classical potentials. These include the D3 or D4 dispersion corrections, the geometric counterpoise (gCP)[157,158] correction for the treatment of intra- and intermolecular BSSE and, in the cases of HF-3c[153] and B97-3c[155], a short-ranged basis (SRB) incompleteness correction term that rectifies systematically overestimated lengths of covalent bonds. For some 3c methods, minimal changes are made in the exchange-correlation part, compensating for errors by slight reparameterization. An overview of existing 3c methods and the corrections used in each case is shown in Table 2.2.

	HF-3c[153]	PBEh-3c[154]	B97-3c[155]	r^{2}SCAN-3c[156]
year of creation	2013	2015	2018	2021
Exact exchange/ %	100	42	0	0
AO basis set	MINIX	def2-mSV(P)	def-mTZVP	def-mTZVPP
gCP	✓	✓	×	✓
SRB	✓	×	✓	×
Dispersion correction	D3	D3	D3	D4
modified functional form	×	✓	✓	×

Table 2.2.: Overview of available "-3c" composite methods/DFAs and their ingredients.

Due to the easy handling and robust performance related to geometries, conformational energies, and general thermochemistry, the “3c“ methods are perfectly suited for use in computational studies.

2.2.4. Tight-Binding Methods

Accelerating calculations has always been a developmental goal and can be achieved by introducing empirical approximations into *ab initio* models. Hartree-Fock-based semi-empirical quantum mechanical methods were introduced as early as the early 1970s.[59] Here, the well-known *zero-differential-overlap* (ZDO) approximation, which leads to the neglect of the frequent three- and four-center two-electron integrals was one of the developments that sparked prominent SQM methods[60] like PM6[159] or PM7[160] that are still in use today. In SQM methods, some universal approximations are introduced to keep the computational cost down. The most important of these is that only valence electrons are treated explicitly, which means the core electrons are assumed to be frozen, and - very important - only a minimal basis set is used.

As already described in chapter 2.2.3, DFT represents an improvement over HF since it can potentially provide exact results at a similar cost and can partially describe static and dynamic correlation via the effective potential. Due to the rapid development of KS-DFT and the simultaneous increase in computing power, Seifert, Elstner, and Frauenheim revived interest in SQM methods and developed the density functional tight-binding (DFTB) models.[59,161–165] In this thesis, the GFN*n*-xTB extended tight binding methods[166,167] have been used extensively, and their linkage to DFT is explained below.[61] The following equations have been adapted from Refs. 61,166,167. The (extended)-tight binding models are semi-empirical approximations to KS-DFT, and their energy expression can be derived from semi-local or, in this case, from general non-local correlation DFA,

$$E_{tot} = V_{nn} + \sum_{i}^{N_{MO}} n_i \int \psi_i^*(\mathrm{r}) \left[\hat{T}_e + \hat{V}_{ne} + v_{XC}^{LDA}[\rho] + v_C^{NL}[\rho(\mathrm{r}), \rho(\mathrm{r}')]\right] \psi_i(\mathrm{r})\mathrm{d}\mathrm{r} \ , \tag{2.39}$$

containing a (NL) correlation kernel

$$v_C^{NL}[\rho(\mathrm{r}), \rho(\mathrm{r}')] = \frac{1}{2} \int \left(\frac{1}{|\mathrm{r} - \mathrm{r}'|} + \Phi_C^{NL}(\mathrm{r}, \mathrm{r}') \right) \rho(\mathrm{r}')\mathrm{d}\mathrm{r}' \ . \tag{2.40}$$

Here, V_{nn} is the nuclear-nuclear repulsion, Ψ_i are molecular spatial orbitals with the corresponding occupation numbers n_i, $\hat{T}(r)$ is the kinetic energy operator, and $V_{ne}(r)$ is the Coulomb electron-nuclear attraction. $v_{XC}^{LDA}[\rho(r)]$ describes the semi-local XC energy per particle and the v_C^{NL} operator defines the electron-electron Coulomb energy and non-local (NL) correlation with the $\Phi_C^{NL}(\mathrm{r},\mathrm{r}')$ kernel. In DFTB and KS-DFT, the electron density is the main quantity from which the energy of the system can be derived. In the tight binding concept, the energy of a system is expanded in a Taylor series around a reference density of neutral, spherical atomic densities (ρ_0):

$$E[\rho] = \sum_{k=0}^{\infty} \frac{1}{k!} E^{(k)}[\rho_0, (\delta\rho)^k] \tag{2.41a}$$

$$= E^{(0)}[\rho_0] + E^{(1)}[\rho_0, \delta\rho] + E^{(2)}[\rho_0, (\delta\rho)^2] + E^{(3)}[\rho_0, (\delta\rho)^3] + \dots , \tag{2.41b}$$

where the density fluctuations are confined to the valence orbitals. Reformulating Eq. 2.39 as a function of the density in terms of reference density ρ_0 generated from spherical, neutral atomic reference densities $\rho_0 = \sum_A \rho_0^A$ and density difference $\Delta\rho$ from $\rho = \rho_0 + \Delta\rho$ leads to:

$$E_{tot} = E_0^H + \Delta E^H + E_{XC}^{LDA}[\rho] + E_C^{NL}[\rho, \rho'] \ . \tag{2.42}$$

Here, the last two terms contain the local and non-local exchange-correlation energies, which are formulated as a function of the electron density. The energy E_0^H depending only on the reference density is expressed by:

$$E_0^H = V_{nn} + \sum_i^{N_{MO}} n_{0,i} \int \psi_i^*(\mathrm{r}) \left[\hat{T}_e + \hat{V}_{ne} + \frac{1}{2} \int \frac{1}{|\mathrm{r} - \mathrm{r}'|} \rho_0(\mathrm{r}')\mathrm{d}\mathrm{r}' \right] \psi_i(\mathrm{r})\mathrm{d}\mathrm{r} \ . \tag{2.43}$$

And the energy ΔE^H depending on the density fluctuations $\Delta\rho$ is provided by:

$$\Delta E^H = \sum_i^{N_{MO}} \Delta n_i \int \psi_i^*(\mathrm{r}) \left[\hat{T}_e + \hat{V}_0 + \frac{1}{2} \int \frac{1}{|\mathrm{r} - \mathrm{r}'|} \Delta\rho(\mathrm{r}')\mathrm{d}\mathrm{r}' \right] \psi_i(\mathrm{r})\mathrm{d}\mathrm{r} \ , \tag{2.44}$$

with the reference potential $\hat{V}_0$:

$$\hat{V}_0 = \sum_A^{nuclei} \left(\int \frac{1}{|\mathrm{r} - \mathrm{r}'|} \rho_0^A(\mathrm{r}')\mathrm{d}\mathrm{r}' - \frac{Z_A}{|\mathrm{r} - \mathrm{R}_A|} \right) \ . \tag{2.45}$$

The first fundamental approximation in the tight binding methods is the truncation of the total energy expansion, which can be done at different orders in the density fluctuation. This leads to different levels of approximations, and the most sophisticated models usually truncate after the third order.[163,166–168] The following provides a brief summary of each classification term. The zeroth order system is composed of *non-interacting*, neutral, spherical atomic reference densities. By design, only the E_0^H term

remains, which is written as

$$E^{(0)}[\rho_0] \approx \sum_{A}^{nuclei} E_A[\rho_0^A] + \frac{1}{2}\sum_{A,B}^{nuclei}\left(E_{rep}[\rho_0^A,\rho_0^B] + E_{disp}[\rho_0^A,\rho_0^B]\right) \tag{2.46a}$$

$$\approx \sum_{A}^{nuclei} E_A[\rho_0^A] + \left(E^{(0)}_{rep,pair} + E^{(0)}_{disp,pair}\right) \ . \tag{2.46b}$$

The zeroth order energy is determined from the sum of the total atomic energies (E_A), which can be precomputed from ρ_0^A at infinite distance and the terms of the inter-atomic pairwise repulsion $E^{(0)}_{rep,pair}$ and (non-local) London dispersion attraction $E^{(0)}_{disp,pair}$. At first order, density fluctuations allow changes in the electronic energy, but the electrostatic potential is kept constant, and no Coulomb terms arise

$$E^{(1)}[\rho_0,\delta\rho] = \Delta E^H + \frac{\partial}{\partial\rho}\left(E_{XC}^{LDA}[\rho_0] + E_C^{NL}[\rho_0,\rho_0']\right)\delta\rho \tag{2.47a}$$

$$\approx E^{(1)}_{EHT} + E^{(1)}_{disp} \ . \tag{2.47b}$$

The second term on the right-hand side represents the first order non-local correction $E^{(1)}_{disp}$ describing the long-range density fluctuation $\delta\rho$ interaction with ρ_0. The first term, known as the extended Hückel theory (EHT)[169,170] contribution, is of high importance since it is primarily responsible for the correct description of covalent bonding. When expressing the EHT treatment in a (minimal) basis set, the speed advantage of tight-binding methods over *ab initio* methods becomes clear since the EHT operator $\hat{H}_0$ is a formal one-electron operator and no two-electron integrals need to be evaluated, based on:

$$E^{(1)}_{EHT} = \sum_{i}^{N_{MO}} \langle\psi_i|\hat{H}^{EHT}|\psi_i\rangle = \sum_{i}\sum_{\mu\nu}^{N_{AO}} n_i C_{\mu i} C_{\nu i} H^{EHT}_{\mu\nu} = \sum_{\mu\nu} P_{\mu\nu} H^{EHT}_{\nu\mu} \ . \tag{2.48a}$$

The EHT energy is given as the sum of changes in the atomic valence energies, computed from the density matrix $P_{\mu\nu}$, which itself is defined from the reference density and the density fluctuations $P_{\mu\nu} = P^0_{\mu\nu} + \delta P_{\mu\nu}$. Truncating the density expansion after the first order has the advantage of robustness associated with non-iterative methods.[164,168,171] In the second order, density fluctuations between atoms are included, and interatomic electrostatic, one-center exchange-correlation, and non-local (NL) correlation terms emerge

$$E^{(2)}[\rho_0,(\delta\rho)^2] = \Delta E^H + \frac{\partial^2}{\partial\rho\partial\rho'}\left(E_{XC}^{LDA}[\rho_0] + E_C^{NL}[\rho_0,\rho_0']\right)\delta\rho\delta\rho' \tag{2.49a}$$

$$\approx E^{(2)}_{ES+XC} + E^{(2)}_{disp} \tag{2.49b}$$

$$\approx \frac{1}{2}\sum_{AB}^{nuclei} q_A q_B \gamma_{AB} + E^{(2)}_{disp} \ . \tag{2.49c}$$

Within the GFN*n*-xTB tight binding methods the $E^{(2)}_{ES+XC}$ term is approximated by a Mataga–Nishimoto–Ohno–Klopman[172–174] damped Coulomb interaction γ_{AB} between the charges $q_{A,B}$ (see Eq. 2.49c), which are derived from a Mulliken population analysis[175] i.e., depending on the overlap $S_{\mu\nu}$ and density ma-

trix $P_{\mu\nu}$. This dependence of the tight binding Hamiltonian on the atomic or shell charges requires an iterative solution referred to as self-consistent charge (SCC) instead of an SCF procedure.

Third-order density fluctuation terms are accounted for by the more sophisticated tight-binding schemes and contain contributions from the third derivatives of local and non-local XC energies:

$$E^{(3)}[\rho_0, (\delta\rho)^3] = \frac{\partial^3}{\partial\rho\partial\rho'\partial\rho''}\left(E_{XC}^{LDA}[\rho_0] + E_C^{NL}[\rho_0, \rho_0']\right)\delta\rho\delta\rho'\delta\rho'' \tag{2.50a}$$

$$\approx \frac{1}{3}\sum_{AB}(q_A)^2 q_B \Gamma_{AB} \ . \tag{2.50b}$$

In DFTB3, these semi-local XC effects are included and mainly stabilize atoms with relatively high charge and can absorb deficits due to missing diffuse functions in the AO minimal basis set. The savings of tight binding methods in computational costs compared to KS-DFT result from the approximations to the various order terms through the use of empirical potentials. Introducing empirical potentials requires the use of element-wise or pair-wise parameters,[161,163], which must be determined from theoretical (*ab initio*) or experimental data. In many cases, however, the parameters are only available for a small number of element/element pairs, which prevents universal use compared to the generally applicable KS-DFT. This was amended by the development of GFN methods within the last few years, starting with the development of stda-xtb in 2016.[28]

2.2.5. Extended Tight-Binding Methods – The GFN-Method Family

The extended GFN*n*-xTB tight-binding methods approximate KS-DFT using only element and global parameters. This distinguishes these methods from the established DFTB3 tight-binding method, which uses element-pair-specific parameters, resulting in a difficult parameterization procedure and only a limited coverage of the periodic table of elements. Due to the design strategy, all GFN methods are consistently available for all elements up to radon (Z=86). In contrast to previous approaches, the GFN parameters are not precomputed using *ab initio* methods but optimized using a large and diverse training set. During the optimization of the parameters, emphasis was placed on a good description of **G** geometries, (vibrational) **F** frequencies, and **N**on covalent interactions, leading to the acronym of the special purpose **GFN***n*-xTB methods. The *extended* refers to the broad availability of parametrized elements, a small, optimized AO basis set, and general improvements to the underlying theory. For brevity, the condensed GFN-terms are discussed here, and the individual expressions are listed in Appendix A1.1.1. All GFN tight binding variants use predefined STO-mG type minimal basis sets for the valence electrons, and the Slater-type orbitals (STO) are approximated by spherical Gaussian-type atomic orbital (GTO) basis functions.

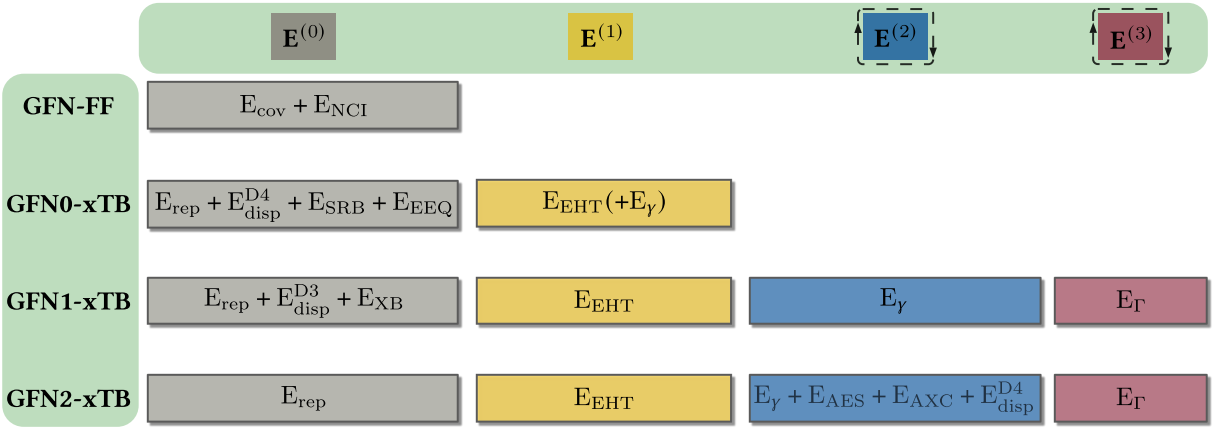


Figure 2.3.: Components of GFN-FF and GFNn-xTB tight binding methods categorized by the corresponding contributions from the density expansion order. Dashed arrows indicate components that require a self-consistent charge procedure.

The GFN development started with GFN1-xTB[166], which shares similarities with the older DFTB3-D3 method. At zeroth order, the overlap of the atomic reference densities leads to the pairwise empirical repulsion term E_{rep}, the D3 model captures the long-range dispersion energy, and halogen-bonds are accounted for via the E_{XB} correction (see Fig. 2.3). At first order in the density expansion, the extended Hückel term (E_{EHT}) is responsible for covalent bonding. The short-range interaction E_γ deals with the isotropic electrostatic and XC energy resulting from the second-order density expansion (see Eq. 2.49c). The third order on-site term E_γ is an electrostatic/XC correction stabilizing high partial charges.

GFN2-xTB[167] is the second GFN tight-binding version, which goes beyond the monopole approximation and also includes anisotropic dipole and quadrupole expressions for the electrostatic E_{AES} and XC E_{XC} energies. These short-range damped cumulative atomic multipole moments (CAMMs)[176] provide an anisotropic, physically more accurate description that eliminates the need for the halogen bond correction required in GFN1-xTB. The London dispersion is described by a charge-dependent modified version of the D4 model[167], which is evaluated self-consistently using partial charges from the Mulliken population. The highest third-order term E_γ in GFN2-xTB is formulated in a shell-specific expression, and the GFN2-xTB tight binding model is the most sophisticated within the GFN family.

The latest development called GFN0-xTB[171] is mentioned here only for the sake of completeness. Compared with GFN1- and GFN2-xTB it is the first non-self-consistent model and truncates the density expansion after first order. The extended Hückel term, augmented with a short-range bond correction E_{SRB}, describes the covalent bonding situation. At zeroth order, the dispersion energy is obtained from a non-self-consistent D4 version, the repulsion energy is again treated by a classical potential, and the classical, non-QM charge equilibrium model[177,178] (EEQ) produces atomic charges.

Apart from the basis set setup, the GFN-versions share additional similarities. At the heart of the tight binding models is the rather complex extended Hückel term (see Eq. A1.8), which provides the necessary flexibility for the variety of different chemical bonding situations. This flexibility strategy is also evident in the electrostatic treatment, which is facilitated by shell-dependent parameters that

naturally enter the E_{EHT} term. Most GFN-parameters are needed to describe covalent bonds expressed by the EHT term, but its functional form eliminates the need for the difficult-to-parameterize element-pair-wise parameters. Another valuable component of all GFNn-xTB methods is the electronic entropy term G_{Fermi} (see Eq. A1.5). This Fermi-smearing[179] technique, in combination with fractional occupation number (FON) treatments for the orbitals, helps with better convergence of the self-consistent charge procedure in xTB. As an effect, static correlation can be simulated (to a degree), which helps, for example, in the description of bond breaking.

Overall, the GFN versions complement each other in their areas of application, and they are characterized by their universal applicability and robustness.[61] For the evaluation of systems in the condensed phase, the implicit solvation models GBSA[166] and ALPB[180] (see Chapter 2.4.3) are available within xTB. The significant computational speed gain compared to KS-DFT, with reasonable and consistent accuracy, makes these methods ideal tools for studying either large systems or systems with many degrees of freedom.

2.3. GFN special purpose Force Field

The GFN-FF[61,181] is the logical development of the GFN methods in the molecular mechanics domain. In a force field, the energy is provided as a function of the nuclei coordinates, and electrons are not considered. This requires some approximations in the form of empirical functions and parameters that must be fitted to ab initio computational data, which for the GFN-FF method are B97-3c equilibrium geometries and harmonic vibrational frequencies. As with the tight-binding versions, an element and global parametrization procedure is applied, and no pair-specific parameters are needed. This strategy makes the FF available for all elements up to Z=86, including transition-metal elements, which are considered 'more difficult.' Combined with the broad parameterization, an automated setup that only requires Cartesian coordinates facilitates the uncomplicated use case. The total GFN-FF energy expression is given by

$$E_{GFN-FF} = E_{cov} + E_{NCI} \tag{2.51a}$$

$$E_{cov} = E_{bond} + E_{bend} + E_{tors} + E_{rep}^{bond} + E_{abc}^{bond} \tag{2.51b}$$

$$E_{NCI} = E_{IES} + E_{disp} + E_{HB} + E_{XB} + E_{rep}^{NCI} \quad , \tag{2.51c}$$

here, E_{cov} describes the bonded FF energy, where the covalent part is divided into asymptotically correct bonding, angular and torsional expressions. The bonding term allows bond cleavage, advancing the GFN-FF into the realm of dissociative FFs. Three-body effects are contained in the E_{abc}^{bond} bonding correction term. The non-covalent energy expression E_{NCI} contains isotropic electrostatics described by the electronegativity-equilibrium-charge (EEQ) model, a simplified D4 dispersion correction scheme, as well as charge-dependent hydrogen-bonding (HB) and halogen-bonding corrections. Repulsive terms, i.e., for the compression of bonds, are given separately for bonded and unbonded interactions. For large systems or long simulation times, the GFN force field provides a robust non-electronic GFN variant that scales formally as $O(N^3)$, qualifying the method for the study of large PES.

2.4. Contributions to Free Energy

2.4.1. Gas phase thermodynamic Properties

Until now, only single molecules and their electronic energies have been discussed. However, in typical chemical reactions, a large number of molecules ($\approx 10^{20}$) is present, and the connection from a single molecule (microscopic system) to a macroscopic sample is described by statistical thermodynamics.[51] From the *partition function Q*, macroscopic observables like the system volume V or temperature T and elementary thermodynamic functions like the Gibbs free energy G, the enthalpy H and the entropy S can be derived, according to:

$$\mathbf{G} = H - TS = kTV\left(\frac{\partial lnQ}{\partial T}\right)_V + klnQ \tag{2.52a}$$

$$\mathbf{H} = U + PV = kT^2\left(\frac{\partial lnQ}{\partial T}\right)_V + kTV\left(\frac{\partial lnQ}{\partial V}\right)_T \tag{2.52b}$$

$$\mathbf{S} = kT\left(\frac{\partial lnQ}{\partial T}\right)_V + klnQ \ . \tag{2.52c}$$

To extract these quantities from Q, information about all possible quantum states of the system is required. For very small systems, i.e., up to three atoms, it is possible to solve the SE, while for larger molecules, the energy levels of a compound in a single conformation can be calculated from the Rigid-Rotor Harmonic-Oscillator (RRHO) approximation. Assuming an isolated molecule, i.e., an ideal gas state, the RRHO approximation can be solved exactly, and the total energy of the isolated molecule can be approximated by a sum of separable energy components. The partition function Q_{tot} can be written as the product of terms involving electronic, translation, rotation, vibration, and conformational states,[182]

$$Q_{tot} = Q_{elec}Q_{trans}Q_{rot}Q_{vib}Q_{conf} \ . \tag{2.53}$$

The separation is a reasonable approximation for many systems but can become problematic if two energy surfaces come close to each other and the Born-Oppenheimer approximation breaks down. As the contributions of enthalpy and entropy depend on the logarithm of Q, the respective quantities can be split according to their degrees of freedom

$$H_{tot} = H_{elec} + H_{trans} + H_{rot} + H_{vib} + H_{conf} \tag{2.54a}$$

$$S_{tot} = S_{elec} + S_{trans} + S_{rot} + S_{vib} + S_{conf} \ . \tag{2.54b}$$

This decoupling allows a more in-depth study of the individual contributions and the introduction of suitable approximations based mainly on the magnitude of energy differences between the energy levels within each term. Here, for example, the electronic partition function is described in good approximation by the ground state of the system since the energy difference of the excited states is usually considerably larger than the ratio of the Boltzmann constant times temperature kT. The rotational degrees of freedom are described within the rigid-rotor approximation, mainly determined from the

molecule's geometry, and the vibrational partition function is calculated by approximating molecular vibrations by harmonic oscillators. For brevity, the individual expressions for the enthalpy and entropy terms are reported in Appendix A1.2. For flexible compounds, the description of conformational terms is important and is treated similarly to the electronic distribution function (see next chapter 2.4.2).

The Gibbs free energy at constant temperature describes the thermodynamic potential and is minimized as the system reaches a chemical equilibrium. Following an established and tested protocol,[50,183] the Gibbs energy of a single-structure

$$G_{tot} = E_{tot} + G^{T}_{trv} \quad , \tag{2.55}$$

can be obtained as the sum of the total energy (E_{tot}), the translation, rotation and vibrational contributions to free energy (G^{T}_{trv}) (see Eq. 2.56) and in case of systems in solution the free energy of solvation has to be added (δG^{T}_{solv}) which is further discussed in Sec 2.4.3. G^{T}_{trv} is calculated in the Rigid-Rotor Harmonic-Oscillator (RRHO) approximation.[183,184]

$$G^{T}_{trv} = (H_{trans} + H_{rot} + H_{vib}) - T(S_{trans} + S_{rot} + S_{vib}) \tag{2.56}$$

In addition, the thermostatistic contribution to the Gibbs energy also includes the zero-point vibrational energy (ZPVE), which describes the quantum mechanical energy embodied in the nuclear motions at zero temperature. The ZPVE correction is approximated as half the sum over the vibrational modes (RRHO) ν_i of the molecule

$$ZPVE = \frac{1}{2} \sum_{i}^{modes} \nu_i \quad . \tag{2.57}$$

2.4.2. Conformations

Rigid molecules are usually accurately described by a single dominant conformation, and the partition function for a single rigid molecule is sufficient for describing the macroscopic system. However, for flexible compounds, the property of interest cannot be accurately represented by a single conformation and is replaced by an equilibrium of multiple structures called a thermodynamic ensemble.

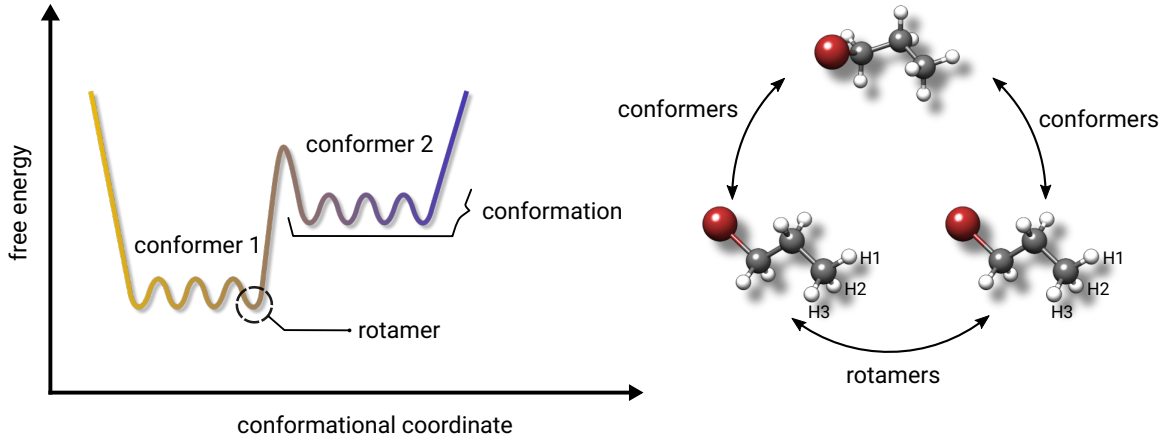


Figure 2.4.: Schematic representation of rotamers and conformers on a potential energy surface.

A schematic representation of stereoisomers populating an ensemble is depicted in Fig. 2.4. The nomenclature used in this thesis defines conformers as stereoisomers that share the same covalent topology and have a unique potential energy minimum. Rotamers are formed by bond rotation or an inversion process that leads to an exchange of nuclei and to minima with identical energies. A conformer with all its associated rotamers is termed a conformation. With the assumption of separable degrees of freedom, the free energy of a structure ensemble is given by

$$G_{tot,ensemble} = \sum_i p_i G_i + G_{conf} \quad , \tag{2.58}$$

with the largest contribution usually coming from the average Gibbs free energy, expressed by the product of the Boltzmann factors p_i and the free energy of the ensemble member i summed over all conformers. The Boltzmann factors at the temperature T can be obtained according to

$$p_i = \frac{exp(-\frac{G_i}{kT})}{\sum_j exp(-\frac{G_j}{kT})} \tag{2.59}$$

while k is the Boltzmann constant and G_i the free energy of conformer i. G_{conf} in Eq. 2.58 contains the conformational free energy of the structure ensemble, which requires knowledge about all populated minima on the potential energy surface. Exploration of the potential energy surface is a task that should be the first step of any computational project and can be very costly due to a large number of associated single-point and gradient calculations. A brief overview of the chemical space exploration workflow as used in the *crest*[68] program used in this thesis can be found in the Appendix A1.4. For detailed information on how to compute G_{conf}, see Ref. 185 and Part A12.2.7. The application of the calculation of molecular (thermodynamic) properties from structural ensembles (SE) is shown in part A12.2.7.

2.4.3. Effects of solvation

Most experiments are conducted in solution, and the effects of the environment on the system must also be considered in computational studies. In most cases, the focus is on the properties of the solute,

which makes a detailed description of the solvent unnecessary as long as the influence of the solvent on the solute is described with sufficient accuracy. This can be achieved by explicitly including solvent molecules in the QM calculation and sampling the phase space or by using computationally efficient implicit solvation models, which were applied in this thesis. Continuum solvation models like the Analytical Linearized Poisson-Boltzmann (ALPB)[180] method available in the xtb code treat the solvent as a uniform polarizable medium described by a dielectric constant ϵ. The solvation free energy δG_{solv} is partitioned into polar (electrostatics) and non-polar (cavity creation and dispersion) contributions.

$$\delta G_{polar}^{ALPB} = -\frac{1}{2}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right)\frac{1}{1+\alpha\beta}\sum_{A=1}^{N}\sum_{B=1}^{N} q_A q_B\left(\frac{1}{f_{AB}} + \frac{\alpha\beta}{\mathcal{A}_{det}}\right) \tag{2.60}$$

The polar part is given in Eq. 2.60, where ϵ_{out} and $\epsilon_{in} = 1$ contain the dielectric constant of the solvent and solute respectively, $q_{A/B}$ hold atomic partial charges and f_{AB} is the interaction kernel, depending on the element- and environment-specific Born radii and the distance R_{AB}. α is considered as a parameter, and β details the quotient of the solute and solvent dielectric constant. The electrostatic size $\mathcal{A}_{det}$ connects the electrostatic energy with the overall global shape of the molecule.[186] The free energy of solvent-solvent cavity formation and solvent-solvent dispersion interactions is described by the non-polar (npol) contribution by means of a solvent-accessible surface area (SASA):

$$\delta G_{npol}^{SA} = \sum_{A}^{N} \gamma_A \sigma_A \quad . \tag{2.61}$$

Here, γ_A is the element-specific surface tension parameter, and σ_A is the SASA of atom A, which is evaluated on an angular Lebedev grid.[187] The important hydrogen bonding (HB) interaction can not be recovered when treating the solvent in an averaged implicit fashion. Therefore, it is approximated by a separate contribution:

$$\delta G_{polar}^{HB} = \sum_{A}^{N} G_A^{HB} \approx \sum_{A}^{N} -g_A^{HB} q_A^2 \frac{\sigma_A}{4\pi (R_A^{surf})^2} \quad , \tag{2.62}$$

where g_A^{HB} is the element-specific HB strength, the SASA contribution describes the probability of finding a solvent molecule interaction site close to atom A, and the hydrogen-acceptor and donor dipole moments are approximated by atomic site charges. The total solvation free energy as obtained from the ALPB model in xtb is calculated from the sum of all polar and non-polar contributions.

$$\delta G_{solv} = \delta G_{polar}^{ALPB} + \delta G_{polar}^{HB} + \delta G_{npol}^{SA} \quad . \tag{2.63}$$

The parameters in ALPB have been fitted to reproduce a combination of experimental data[188] and theoretical values using COSMO-RS[189] as a reference. Solvation effects are very important for the relative energetic ordering of conformer ensembles (see, e.g., section A12.2.7), which necessitates the availability of an efficient solvation model such as ALPB for fast methods (SQM and FF) in the ensemble generation steps.[180]

COSMO-RS provides a higher degree of accuracy of solvation models, which is associated with a higher computational cost and is best used in combination with DFT methods.[48] For the sake of brevity, the derivation of COSMO-RS is discussed in Appendix A1.3.

2.5. Calculation of Molecular Response Properties

2.5.1. Nuclear Magnetic Resonance

NMR has become a routine, noninvasive, and invaluable analytical method for interpreting experimental data.[190] Its significance stems from the ability to differentiate nuclei with respect to their chemical environment, which provides insight into the molecule's geometry. This is possible due to the magnetic moment $\mu_K = \gamma_K m_K$ of the atomic nuclei K with spin I_K, which depends on the magnetic quantum number m_K and the gyromagnetic ratio γ_K. In a static external magnetic field, B_0 the *nuclear Zeeman splitting* due to the interaction of the nuclear magnetic moment with B_0 cancels the degeneracy of energy levels. Through electronic motion and magnetically induced electronic current, an additional magnetic field is generated, and the locally experienced field can be expressed by the *chemical shielding or screening* tensor σ^K and the external magnetic field as $B_{local} = (1 - \sigma^K)B_0$. The resulting energy difference in the presence of a magnetic field is then expressed by:

$$\Delta E = -m_K(1 - \sigma^K)B_0 \ . \tag{2.64}$$

The 3x3 chemical shielding tensor describes the screening of the applied field strength through the induced magnetic field from the electron density surrounding the nucleus K, which is very finely tuned to the chemical environment. Normally, the trace of σ is discussed, which accounts for the rotational motion of the molecule when compared to NMR experiments in the liquid phase:

$$\sigma^K_{iso} = \frac{Tr(\sigma^K)}{3} = \frac{\sigma^K_{11} + \sigma^K_{22} + \sigma^K_{33}}{3} \ . \tag{2.65}$$

For a universal comparison of chemical shielding, independent of the spectrometer and thereby operated different B_0 fields, the chemical shifts δ are discussed relative to a given shielding from a reference compound (usually TMS for the ^1H and ^{13}C nuclei):

$$\delta_K = \sigma^K_{ref,iso} - \sigma^K_{iso} \ , \tag{2.66}$$

and the dimensionless shift is usually provided in ppm. A brief outline of the theory used to calculate chemical shielding constants is given, and more detailed information can be found in Refs. 191–195. The shielding tensor σ^K is a second-order response property that depends on the external magnetic field B and the magnetic spin moment m_K of nucleus K:

$$\sigma^K_{ij} \approx \left.\frac{\partial^2 E}{\partial B_j \partial m_{Ki}}\right|_{B=m_K=0} \quad i, j = 1, 2, 3 \ . \tag{2.67}$$

Since the external magnetic field interacts with the magnetic field generated by the motion of the electrons, this perturbation must be reflected in the kinetic momentum operator of the one-electron Hamiltonian:

$$h(r, B, m) = \frac{\pi^2}{2} - V(r) \tag{2.68a}$$

$$\pi = -i\nabla + A(r, B, m) \tag{2.68b}$$

The generalized momentum operator π incorporates the vector potential A(r,B,m), which is associated with the magnetic field B_0 and is given by

$$A(r, B, m) = \frac{1}{2} B \times (r - R_0) + \alpha^2 \sum_K \frac{m_K \times (r - R_K)}{|r - R_K|^3} \tag{2.69}$$

where r and R_K are the coordinates of the electron and nucleus K, α is the fine-structure constant, and R_0 represents the gauge origin of the vector potential. The first term on the right-hand side of Eq. 2.69 is concerned with the external field, and the second term considers the contribution from the nuclear magnetic moments.

By using the generalized momentum operator in the molecular SE, Ramsey [196] derived expressions for the diamagnetic and paramagnetic contributions to the shielding tensor. However, a computationally more convenient expression for variational solutions like HF or DFT (when the XC-kernel is assumed to be independent of B or m) to the mixed second derivative is given by: [197,198]

$$\sigma_{ij}^K = \underbrace{\sum_{\mu\nu} P_{\mu\nu} \langle \mu | \frac{\partial^2 h_{\mu\nu}}{\partial B_j \partial m_{Ki}} | \nu \rangle}_{diamagnetic} + \underbrace{\sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial B_j} \langle \mu | \frac{\partial h_{\mu\nu}}{\partial m_{Ki}} | \nu \rangle}_{paramagnetic} \tag{2.70a}$$

$$= \sigma_{ij}^{K,dia} + \sigma_{ij}^{K,para} \ . \tag{2.70b}$$

Here $P_{\nu\mu}$ are elements of the density matrix in the atomic orbital representation, and $h_{\mu\nu}$ provide the one-electron Hamiltonian matrix elements. The first term on the r.h.s. of Eq. 2.70 represents the diamagnetic, and the second term is the paramagnetic part of the shielding tensor. The perturbed densities $\frac{\partial P_{\mu\nu}}{\partial B_{x,y,z}}$ necessary for the second order property is obtained by solving three sets of 'Coupled Perturbed Self-Consistent Field' equations (CPSCF). Removing the gauge orient dependence in the calculation of chemical shielding tensors is usually achieved by employing Gauge Independent/Invariant Atomic Orbitals (GIAO) [197,199], which are complex, local gauge dependent basis functions:

$$X_\mu^{GIAO} = e^{-iA_K \cdot r} \varphi_\mu^K \tag{2.71a}$$

$$A_K = \frac{1}{2} B \times (R_K - R_0) \ . \tag{2.71b}$$

Here, the field-independent Gaussian type atomic orbital φ_μ is multiplied by a complex phase factor, which includes the local representation A_K of the external field. When the generalized momentum operator is applied, the gauge origin R_0 is shifted to R_μ the point where the orbital φ_μ is centered, and

the gauge origin dependence also vanishes for matrix elements involving GIAOs.[200]

2.5.2. Optical rotation and electronic circular dichroism

Optical activity is an inherent property of chiral compounds, which are able to rotate plane-polarized light. Circular dichroism and circular birefringence are manifestations of optical activity and depend on the molecular structure and wavelength of the incident light. Linear polarized light can be expressed as a superposition of left and right circular polarized light with equal amplitude. Optical rotation or circular birefringence occurs because the real part of the refractive index for circularly polarized light beams is different in a chiral environment, resulting in a relative phase shift of the left and right circularly polarized light waves, which is observed as an overall rotation of the initial plane of polarization. Circular dichroism, i.e. the different absorption of left and right circularly polarized light, often expressed as the difference between molar extinction coefficients $\Delta\epsilon(\lambda) = \epsilon_L(\lambda) - \epsilon_R(\lambda)$, has its origin in the imaginary part of the refractive index, which causes an individually measurable exponential decrease in the amplitude of circularly polarized light of different directions.[201]

$$\tilde{n} = n - i\kappa \tag{2.72}$$

Equation 2.72 describes the complex refractive index $\tilde{n}$, where the real part n expresses the phase velocity and the imaginary part κ, also called extinction coefficient, describes the degree of attenuation of the electromagnetic wave. In 1929, Rosenfeld established the theoretical basis for the molecular dipole moment induced by a frequency-dependent electromagnetic field, which reads:[201–203]

$$\vec{\mu}_{ind}(t) = \overleftrightarrow{\alpha}(-\omega;\omega)\cdot\vec{E} + \overleftrightarrow{G}(-\omega;\omega)\cdot\frac{\partial\vec{B}}{\partial t} + \cdots \quad , \tag{2.73}$$

with $\vec{E}$ and $\vec{B}$ as the time-dependent electric and magnetic field vectors, $\overleftrightarrow{\alpha}$ as the electric dipole polarizability tensor, and $\overleftrightarrow{G}$ as the mixed electric dipole - magnetic dipole polarizability tensor, which is related to the refractive index.[201,204] The most important quantity for the calculation of OR is the electric dipole - magnetic dipole polarizability tensor, which in the sum-over-states expression is given by:

$$G_{\zeta\zeta}(\omega) = -4\pi \sum_{n\neq 0} \frac{1}{\omega_n^2 - \omega^2} \times \Im(\langle\Psi_0|\hat{m}_\zeta|\Psi_n\rangle\langle\Psi_n|\hat{\mu}_\zeta|\Psi_0\rangle) \quad , \tag{2.74}$$

where $\hat{m}_i$ and $\hat{\mu}_j$ are Cartesian components of the magnetic and electric dipole operator, Ψ_0 and Ψ_n are the ground and excited electronic wave functions, ω_n represents the excitation energy of state n and ω is the angular frequency of the excitation radiation. Since the perturbation of a system by an electromagnetic field in the UV-Vis region is weak, the G-tensor or similar quantities can be calculated from linear response theory instead of solving the full time-dependent Schrödinger equation, and excited-state wave functions do not need to be dealt with explicitly. This is facilitated by solving the standard TD-DFT equation[27,205–207] for the three-component linear electric dipole response vectors $\mathbf{X}_\zeta(\omega)$ and

$\mathbf{Y}_\zeta(\omega)$:

$$\left[\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B} & \mathbf{A}\end{pmatrix} - \omega\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\right]\begin{pmatrix}\mathbf{X}_\zeta(\omega)\\ \mathbf{Y}_\zeta(\omega)\end{pmatrix} = -\begin{pmatrix}\mu_\zeta\\ \mu_\zeta\end{pmatrix} \quad , \tag{2.75}$$

with $\mu_\zeta = \langle\Phi_a|\hat{\mu_\zeta}|\Phi_i\rangle$, where *i,j* refers to occupied and *a,b* to unoccupied molecular orbitals. The orbital rotation Hessians A and B in the adiabatic local density approximation (ALDA)[208] are calculated from:

$$\mathbf{A}_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + 2(ia|jb) - a_x(ij|ab) + (1-a_x)(ia|f_{XC}|jb), \tag{2.76}$$

and

$$\mathbf{B}_{ia,jb} = 2(ia|bj) - a_x(ib|aj) + (1-a_x)(ia|f_{XC}|bj) \quad , \tag{2.77}$$

where the response of the density functional within the ALDA is given by the terms containing f_{XC} which is evaluated from:

$$f_{XC}(r_1, r_2) = \frac{\delta^2 E_{XC}[\rho]}{\delta\rho(r_1)\delta\rho(r_2)}. \tag{2.78}$$

The electric dipole - magnetic dipole polarizability tensor can then be calculated from the linear response vectors and the magnetic moment m_ζ, where ζ reflects the Cartesian component (x,y,z) in the MO representation:[209]

$$G_{\zeta\zeta}(\omega) = -\omega^{-1}\Im\langle\langle m_\zeta; \mu_\zeta\rangle\rangle_\omega = -\omega^{-1}\Im(m_\zeta(X_\zeta - Y_\zeta)) \; . \tag{2.79}$$

The typically reported specific rotation $[\alpha]_{\tilde{\nu}}$ in $^\circ/(dm(g/cm^3))$ can be computed from the G tensor or, more precisely, from the orientation average β:

$$\beta = \frac{Tr(G_{\zeta\zeta}(\omega))}{3} \tag{2.80a}$$

$$[\alpha]_{\tilde{\nu}} = 1.343\times10^{-4}\frac{\beta(\omega))\tilde{\nu}^2}{M} \tag{2.80b}$$

where M is the molecular weight in g/mol, $\tilde{\nu}$ is the wavenumber in cm^{-1}, usually corresponding to the sodium D-line. The most important quantity for electronic circular dichroism (ECD) in the UV-Vis spectral range is the rotational strength R, which is the analog of the oscillator strength in UV spectroscopy and is obtained from the imaginary part of the scalar product of the electric and magnetic dipole transition vectors

$$R_{0n} = \Im(\langle\Psi_0|\hat{\mu}|\Psi_n\rangle\langle\Psi_0|\hat{m}|\Psi_n\rangle) \tag{2.81a}$$

$$= \Im(\mu_{0n}\cdot m_{0n}) \quad . \tag{2.81b}$$

Since the electronic response of a molecular system depends only on the perturbation frequency ω and is independent of the nature of the perturbation itself, it can be described by equation 2.75, where the magnitude of the external perturbation (right-hand side) is set to zero. The resulting general eigenvalue equation is also known as the linear response TD-DFT equation,[206,210] which yields the excitation energies ω_{0n} for the transition from the ground to the excited state *n*, i.e., the left-hand side of Eq. 2.75

goes to zero when the perturbation ω matches the excitation energy of the system. In the framework of frequency-dependent linear response theory, this is equivalent to finding the poles of $G(\omega)$ at real frequencies. The rotatory strength can be calculated from the residues of $G(\omega)$, which in practice is constructed from the eigenvectors[211] $\mathbf{X}$ and $\mathbf{Y}$ via the electric

$$\mu_{0n}^{\zeta} = \langle 0|\mu_{\zeta}|n\rangle = \sqrt{2}\sum_{ai}\mu_{ai}^{\zeta}(X_{ia}^{n} + Y_{ia}^{n}) \quad , \tag{2.82}$$

and magnetic transition moments and Eq. 2.81 respectively

$$m_{0n}^{\zeta} = \langle 0|m_{\zeta}|n\rangle = \sqrt{2}\sum_{ai}m_{ai}^{\zeta}(X_{ia}^{n} - Y_{ia}^{n}) \quad . \tag{2.83}$$

The final ECD spectra can be calculated from the line spectrum:

$$\Delta\epsilon(E) = \frac{1}{2.297\times 10^{-39}\sqrt{2\pi\sigma}}\sum_{n}R_n\Delta E_n \exp[-(\frac{E-\Delta E_n}{2\sigma})]^2 \quad , \tag{2.84}$$

thereby expressing the spectrum as a sum of Gaussian shape-functions, simulating the vibrational and solvent-induced broadening, where σ is the width of the band at 1/e height and ΔE_n and R_n are the excitation energies and the rotatory strength of transition n.[212]

Since both OR and ECD are based on the calculation of the rotational strength by means of the magnetic transition dipole moment, which is not invariant with a change of the gauge origin R, two possibilities have been explored to enforce an origin invariance. They either use gauge-dependent basis functions (GIAOs) or calculate the rotational strength in the dipole velocity formalism, where the origin dependence vanishes.[203,213]

Part II.

Efficient Free Energy Calculations for Metal Organic Systems

The Gibbs energy is a central thermodynamic quantity that provides information about how much work can be extracted from a system at constant temperature and pressure. In chemical reactions, it provides information about the direction of the reaction and the chemical equilibrium. An important subclass of reactions are supramolecular host-guest associations. This type of non-covalent interaction is present in every living being, as in the binding of proteins to ligands, the recognition of antigens, or the binding of drugs, and all of these processes occur in a way that minimizes the free energy of the entire system.[214] Part II of this work focuses on applying and evaluating efficient computational procedures involving DFT-3c and/or SQM methods for describing free energies and derived properties in metalorganic systems. The availability of the newly developed GFN*n*-xTB[61] methods enables the study of larger systems with almost any element composition. It renders investigations of limited cut-out models and their arbitrariness obsolete. The methods are tested on real chemical systems and compared with the experimental results and higher-level DFT calculations.

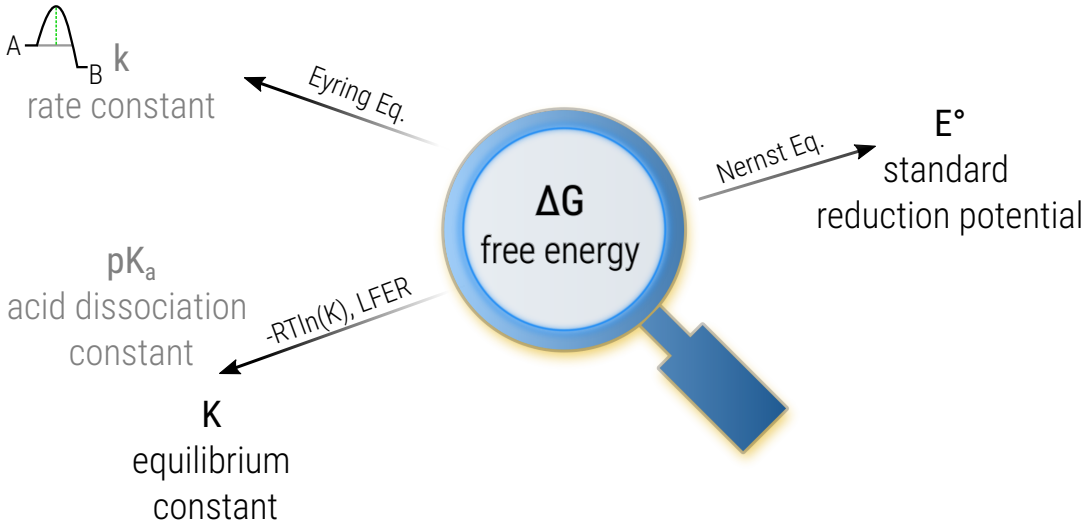


Figure 2.5.: Gibbs energy and derived properties like the kinetic rate constant k via the Eyring equation, pK_a acid dissociation constants e.g. via the linear free energy relationship, equilibrium or association constants or standard reduction potentials E° by means of the Nernst equation.

Chapter 3 is concerned with optimizing strong noncovalent interactions of cyclic trinuclear gold(I) complexes. The strength of the supramolecular donor-acceptor interactions of the planar gold donor molecule is probed with 11 aromatic acceptors, each with a varying degree of π-acidity. Association free energies of these complexes are calculated on PBEh-3c optimized geometries, using a proven DFT-D/large basis set approach. The quantum mechanical results are compared to experimental NMR and UV-Vis measurements, and the data can serve as a design guide, for example, in chemosensing applications.[215] Chapter 4 investigates the self-sorting behavior of isostructural ligands with nitrile and isonitrile moieties with a [2.2]paracyclophane backbone. Understanding self-sorting processes can provide valuable supramolecular design principles important for assembling functional structures, i.e., molecular machines.[216,217] Initially, the absolute configuration of the employed ligands is assigned through

crystal structure analysis and by comparison of simulated and measured electronic circular dichroism (ECD) spectra. Here, the fast GFN1-xTB tight-binding method provides snapshots from a molecular dynamics simulation on top of which UV-Vis and ECD spectra are obtained using the sTDA[218] approach. The thermodynamic stability of differently sized aggregates and their association free energies are computed by the fast GFN2-xTB method with implicit solvation and benchmarked against accurate DFT reference values (PBE0-D4/def2-QZVP + COSMO-RS) and observations from the experiment. The following Chapter 5 presents a screening study identifying the best binding fullerene isomer for a large (roughly 500 atoms) Zn^{II}-macrocycle. Screening studies aim to identify the most promising binding partners rapidly. However, this is only possible with fast methods like GFN2-xTB that retain good accuracy and consistently treat all relevant physical effects. The chapters described so far follow the approach **I** described in Fig. 1.2, i.e., creating the investigated structures manually. The last Chapter of Part II concerns itself with the efficient evaluation of electrochemical redox potentials using semiempirical quantum mechanical methods. Chapter 6 presents a new benchmark set termed ROP313 (Reduction and Oxidation Potentials 313), containing experimental reference values for organic and organometallic reduction and oxidation reactions in several solvents. This project uses the CREST algorithm to identify the most stable structural isomer of the investigated compounds by a conformer search. This approach (approach **II** in Fig. 1.2) generates low-energy conformers based on physical principles. It presents a more evolved approach than the manual generation of structures, which depends very much on the experience and chemical intuition of the scientist. The reduction and oxidation potentials are calculated from the Nernst equation using the Gibbs energy of the corresponding half-cell reaction at standard conditions. SQM of the GFN*n*-xTB and PM*x* families are evaluated in their performance against experimental data and the performance of selected state-of-the-art density functionals from different rungs of Jacob's ladder. This study provides an error estimate for the newly developed GFN methods valuable for the prediction[219] of red/ox potentials and the screening of large compound libraries.

Part II showcases the universal applicability of DFT-3c methods and GFN*n*-xTB SQM for describing noncovalent interactions and Gibbs free energies.

3. Raising the Bar in Aromatic Donor–Acceptor Interactions with Cyclic Trinuclear Gold(I) Complexes as Strong π-Donors

Raiko Hahn,[*] Fabian Bohle,[†] Wenwen Fang,[‡§¶] Andreas Walther,[‡§¶] Stefan Grimme[†] and Birgit Esser[*§]

Received 16th of August 2018 , Published 27th of November 2018

Reprinted (adapted) with permission from[‖]
Hahn, R.; Bohle, F.; Fang, W.; Walther, A.; Grimme, S.; and Esser B. *J. Am. Chem. Soc* **2018**, *140*, 17932–17944.
— Copyright © 2018, American Chemical Society.
DOI 10.1021/jacs.8b08823

Own manuscript contribution

- Performing the calculations
- Interpretation of the results
- Writing the manuscript

[*]Institute for Organic Chemistry, University of Freiburg, Albertstraße 21, 79104 Freiburg, Germany
[†]Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany
[‡]Institute for Macromolecular Chemistry, University of Freiburg, Stefan-Meier-Straße 31, 79104 Freiburg, Germany
[§]Freiburg Materials Research Center (FMF), University of Freiburg, Stefan-Meier-Straße 21, 79104 Freiburg, Germany
[¶]Freiburg Center for Interactive Materials and Bioinspired Technologies (FIT), University of Freiburg, Georges-Köhler-Allee 105, 79110 Freiburg, Germany
[‖]Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

The full paper is reproduced in Appendix A2. Aromatic donor-acceptor interactions are versatile templates for constructing complex molecular systems in material science and biology.[220–222] The noncovalent aromatic interactions between the planar π-systems are often categorized as weak, especially when compared with highly directional hydrogen bonding. This study investigates aromatic supramolecular donor-acceptor complexes with a focus on enhancing their binding strength. Cyclic trinuclear complexes (CTCs) of gold(I) coordinated with nitrogen-containing ligands, i.e., pyridinate, imidazolate, or carbeniate are known to have an electron-rich π-system. When combining these CTCs with electron-deficient π-acceptors, binding is observed, and donor-acceptor complexes are formed. In a previous study (see Ref. 223), we observed binding constants up to $3.0 \times 10^4 Lmol^{-1}$ for the formation of 1:1 donor-acceptor complexes between gold(I)-pyridinate CTCs and organic π-acceptors. This corresponds to a binding free energy of -6.1 kcal mol^{-1}, which constitutes a strong noncovalent association in the class of interacting π-systems.

The structure of gold(I)-CTC molecules and the arrangement of electron density above and below the molecular plane create two binding sites for $\pi - \pi$ interactions. The influence of higher order aggregates, such as, for example, 1:2 (donor:acceptor), was probed using two acceptors differing in the number of vacant binding sites. The binding strength of these complexes was studied experimentally by NMR titrations. The experimental design was chosen to add the acceptor to the donor, preventing the formation of 2:1 donor:acceptor complexes, which are not considered further. Using a bridged napthalene diimide-based acceptor, only one binding site is available, whereas the unbridged acceptor offers two free binding sites. A comparison of the measured binding constants derived from a nonlinear fit indicates that a 1:2 (donor:acceptor) binding model best describes the system. Still, the overall interaction strength is primarily characterized by 1:1 aggregation. In all cases, the K_{12} (A:D:A) aggregation is significantly lower than K_{11}, which can be explained by the anticooperative lowered basicity of the donor within the D:A complexes. Based on these results, quantum chemical calculations were performed to investigate the binding strength of these complexes using 1:1 (donor:acceptor) geometries and neglecting higher-order aggregates to a reasonable approximation.

The geometries of the 1:1 complexes were generated using a black-box strategy, avoiding manual intervention. The intermolecular force field xTB-IFF[224] uses an automated rigid-body global minimum energy search and generates low-energy docked complex geometries. For this, xTB-IFF relies on input from GFN-xTB calculations of the donor and acceptor fragments. All generated geometries are further optimized using the DFT-based PBEh-3c/COSMO scheme while accounting for the solvent's electrostatic effects. Based on the complex and fragment geometries, Gibbs energies were calculated in a multilevel approach.

The binding strength or association free energy is obtained from the difference between the Gibbs energy of the complex and the sum of the Gibbs energies of the acceptor and donor fragments. Using the multilevel approach, all contributions to the Gibbs energy are described with computationally efficient methods and are composed of the sum of the electronic energy, the thermostatistical correction at the temperature of the experiment, and the solvation free energy. The electronic energy was calculated using hybrid DFT at PBE0-D3(BJ)/def2-QZVP level in a large QZ basis combined with the D3(BJ) correction accounting for London dispersion. B97-3c[155] harmonic vibrational frequencies were used to

calculate the thermodynamic contributions to the free energy within the modified rigid-rotor-harmonic oscillator. The solvation free energy was obtained using COSMO-RS[189].

Even though the alkane chains improving the solubility of acceptor and donor were neglected in the model geometries for computational efficiency, the predicted and experimental association free energies for the 1:1 napthalene diimide-based acceptor and gold(I)-CTC aggregates are in very good agreement. Further screening of the non-covalent interaction strength of all donor-acceptor combinations with three CTC gold(I) donors containing imidazolate, pyridinate, and carbenate ligands was investigated with seven organic and three π-acidic pyrazolate CTC acceptors. The calculated theoretical predictions are compared with experimental results from NMR and UV-Vis titration studies, and it is confirmed that the calculations reproduce the experimental trend. The absolute binding strength tends to be underestimated by the computational workflow for donor-acceptor complexes with weaker interactions. Further, it predicts stronger interactions compared to the experiment when in both the donor and acceptor coinage metals are present. This behavior can be attributed to the use of implicit solvation, which might be unable to model all solvent effects for these planar complexes, where metal atoms are directly exposed to the solvent. A strong solvent effect is also observed in UV-Vis titration experiments in different solvents. It corroborates the finding that accurately describing solute-solvent interactions is critical for these systems. For all the complexes studied, the calculations showed that the largest attractive contribution to the binding free energy was due to London dispersion. Overall, the investigated cyclic trinuclear complexes exhibit strong binding when compared with donor-acceptor complexes where only the π-system is relevant. NMR and UV/VIs absorption spectroscopy titrations revealed that the strongest free energy of binding within the studied compounds is ΔG_a = -10.1 kcal mol^{-1}, which highlights the existing potential of the group of cyclic trinuclear complexes for application in other research areas. In summary, the synthesis and characterization of cyclic trinuclear gold(I) complexes exhibiting strong π-donor properties towards electron-deficient aromatic acceptor molecules is presented, utilizing NMR, UVVis, STM, and quantum chemical calculations.

4. Influencing the Self–Sorting Behavior of [2.2]Paracyclophane – Based Ligands by Introducing Isostructural Binding Motifs

Lucia Volbach,[*] Niklas Struch,[*†] Fabian Bohle,[‡] Filip Topić,[§¶] Gregor Schnakenburg,[‖] Andreas Schneider,[*] Kari Rissanen,[§] Stefan Grimme,[‡] and Arne Lützen[*]

This article is part of the Chem. Eur. J. "2020 Supramolecular Chemistry | Hot Paper" collection.

Received 7th of November 2019, accepted 9th of December 2019

Reprinted under the Creative Commons Attribution License 4.0 (CC BY-NC 4.0)
Volbach, L.; Struch, N.; Bohle, F.; Topić, F.; Schnakenburg, G.; Schneider, A.; Rissanen, K.; Grimme, S.; Lützen, A. *Chem. Eur. J.* **2020**, *26*, 3335–3347.
— Copyright under the Creative Commons Attribution License 4.0
DOI 10.1002/chem.201905070

Own manuscript contribution

- Performing the calculations
- Interpretation of the results
- Writing the manuscript

[*]Kekulé-Institute of Organic Chemistry and Biochemistry, University of Bonn, Gerhard-Domagk Strasse 1, 53121 Bonn (Germany)
[†]current address: Arlanxeo Netherlands B.V. Urmonderbaan 24, 6167 RD Geleen (The Netherlands)
[‡]Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany
[§]Department of Chemistry, University of Jyväskylä P.O. Box 35, 40014 Jyväskylä (Finland)
[¶]current address: Department of Chemistry, McGill University 801 Sherbrooke St. West, Montreal, Qc H3A 0B8 (Canada)
[‖]Institute of Inorganic Chemistry, University of Bonn Gerhard-Domagk Strasse 1, 53121 Bonn (Germany)

The paper is reproduced in Appendix A4. Two isostructural ligands, $\mathbf{L}^{\text{nit}}$ and $\mathbf{L}^{\text{iso}}$, have been synthesized in the laboratory of Prof. Lützen, in which either two nitrile and two isonitrile functional groups are connected to a [2.2]paracyclophane backbone, with the moieties arranged in a pseudo-*meta* substitution pattern. The study involves the synthesis and investigation of the two ligands $\mathbf{L}^{\text{nit}}$ and $\mathbf{L}^{\text{iso}}$ and their behavior in forming metallosupramolecular aggregates in solution upon coordination with *cis*-protected Pd^{II} [(dppp)Pd(OTf)$_2$]. These ligands are interesting because they virtually have the same steric bulk, chirality, shape, and coordination angle of about 120°. The key difference is their nitrile/isonitrile functional groups, which allow us to focus on the metal binding unit's behavior and the binding strength's influence on self-sorting. For the coordination studies, both racemic and enantiomerically pure compounds were required. The (*rac*)-$\mathbf{L}^{\text{nit}}$ was separated by HPLC. The absolute configuration of $\mathbf{L}^{\text{nit}}$ enantiomers was confirmed by comparing the experimental CD spectra with simulated ECD spectra obtained from sTD-DFT calculations, and the assignment was later confirmed by single-crystal X-ray diffraction. The theoretical ECD spectra were calculated following two approaches, from a rigid geometry (sTD-CAM-B3LYP/def2-TZVP//PBEh-3c(COSMO))[128,218] and considering conformational flexibility from a GFN1-xTB/GBSA(CH_3CN) molecular dynamic simulation (sTD-BH-LYP/def2-SVP/COSMO(e = 36.6))[114,115,120] and matched with the first eluting R_p-enantiomer. Fortunately, both enantiomeric forms of $\mathbf{L}^{\text{iso}}$ could be obtained from commercially available optically pure starting materials.

NMR and mass spectrometry could analyze the products of the reaction from stoichiometric amounts of the isostructural ligands with the *cis*-protected palladium complex. They form selective supramolecular aggregates with distinct compositions in noncompeting solvents: $\mathbf{L}^{\text{nit}}$ forms dinuclear rhombic structures, while $\mathbf{L}^{\text{iso}}$ gives trinuclear triangular complexes. In the case of (*rac*)-$\mathbf{L}^{\text{nit}}$, a diastereoselective arrangement in a self-discriminating (also known as social self-sorting) manner is observed, resulting in the exclusive formation of C_{2v} symmetric complexes. Upon coordination with [(dppp)Pd(OTf)$_2$] the constitutional isomer (*rac*)-$\mathbf{L}^{\text{iso}}$ forms a mixture of two trinuclear species: a homochiral [(dppp)$_3$Pd$_3$ {(S_p)-$\mathbf{L}^{\text{iso}}$}$_3$](OTf)$_6$ complex in D_3 symmetry and its enantiomer, and the heterochiral aggregate [(dppp)$_3$ Pd$_3${(S_p)-$\mathbf{L}^{\text{iso}}$}$_2${(R_p)-$\mathbf{L}^{\text{iso}}$}](OTf)$_6$ and its stereoisomer, respectively. The two species form in a 1 to 1.5 homo-to-heterochiral ratio, showing the preference of the homochiral assembly in a less pronounced narcissistic self-sorting manner (also known as self-recognizing).

The difference in the self-sorting behavior of these structurally similar ligands correlates with their binding strength to the Pd^{II} metal ion. Generally, isonitrile ligands possess stronger coordination abilities to metal centers than nitrile ligands, owing to their enhanced σ-donation and superior π-backbonding capabilities. This difference in the electronic structure of the investigated ligands can affect structural parameters like the bending angle of the functional groups. A theoretical analysis of the bending angle and associated deformation energy for acetonitrile and isocycanomethane on GFN2-xTB and PBEh-3c level of theory revealed that lower deformation energies are required for bending isonitrile than nitrile functional groups. This could lead to different bond angles in the resulting complexes, offering a possible explanation for the different compositions of the self-assembly reactions.

Gibbs energies were calculated using high-level quantum chemical (PBE0-D4/def2-QZVP + COSMO-RS//PBEh-3c(DCOSMO-RS)) and efficient GFN2-xTB(GBSA) methods to investigate the complexes'

thermodynamic stability with their hypothetical counterparts, which were not observable in the experiment. When the calculated thermodynamic data of the dinuclear and trinuclear complexes were compared, the trinuclear complexes were found to be thermodynamically more stable for both cases, $\mathbf{L}^{\mathrm{nit}}$ and $\mathbf{L}^{\mathrm{iso}}$. However, for $\mathbf{L}^{\mathrm{iso}}$, this is not consistent with experimental observations, and this suggests that the implicit solvent models do not adequately represent the interactions of the solvent. The influence of single-site solvation at the palladium centers was tested and could not change the original conclusion. The effects of solvent and compound concentrations were also observed in the crystallization of the discrete homo- and heterochiral supramolecular aggregates, which rearranged to form coordination polymers. Nevertheless, the calculated trinuclear $\mathbf{L}^{\mathrm{iso}}$ complex is thermodynamically much more stable than its nitrile counterpart, and the experimental trend established for $\mathbf{L}^{\mathrm{iso}}$ is also correctly reproduced.

The combined experimental and theoretical study of the supramolecular compounds provided new structural and conceptual insights into understanding the self-sorting of racemic mixtures in solution. It was confirmed that the metal binding motif, i.e., the metal-ligand bond strength and the resulting structural parameters, have a crucial influence on the self-sorting behavior, which must be taken into account in the development of self-sorting reactions with high reliability.

5. Efficient structural and energetic screening of fullerene encapsulation in a large supramolecular double decker macrocycle

Fabian Bohle* and Stefan Grimme*

Special Issue Devoted to Prof. emeritus Miljenko Perić

Received 1th of July 2019, Accepted 1st of July 2019, Published 29th of August of 2019

Reprinted under the Creative Commons Attribution License 4.0 (CC BY-NC-ND 4.0)
Bohle F.; Grimme, S. *J. Serb. Chem. Soc.* **2019**, *84*, 837–843.
— Copyright under the Creative Commons Attribution License 4.0
DOI 10.2298/JSC190701079B

Own manuscript contribution

- Performing the calculations
- Interpretation of the results
- Writing the manuscript

*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

The paper is reproduced in Appendix A6. Supramolecular chemistry and, in particular, the concept of host-guest binding are ongoing vital areas of current research.[225] From them, strategies to mimic biological systems are derived and applied in the drug development field.[226] Understanding host-guest systems is of central interest because recognizing specific molecules and creating complementary supramolecular receptor systems is difficult due to often weak intermolecular forces. This study concerns the selective purification of fullerene isomers through higher association in a macromolecular complex. Fullerenes have drawn considerable attention as purest carbon modification and for their properties as semiconductors in electronics,[227] photovoltaic cells[228] or as drug delivery-system in medical applications.[229] Because of their low solubility in organic solvents, purification for further use remains an issue. Several macrocyclic compounds have been investigated as suitable receptors. In 2018, Tanaka *et al.* designed and synthesized a supramolecular coordination cage containing a nano-cavity suitable in size to fit the best-known fullerene C_{60}. The receptor consists of two dibenzothiophenesalen Zn^{II} metallomacrocyclic rings bridged by nitrogen-containing DABCO pillar molecules connecting the square-pyramidal coordinating Zn cations. This double-decker geometry serves as a host molecule with a shape-persistent cavity. The experimentalists observed from ^{1}H-NMR measurements that the receptor selectively recognizes C_{70} over C_{60} through higher association. The association free energy of the 1:1 complex between the receptor and C_{70} in $CHCl_3/CS_2$ was determined to be -3.07 kcal mol^{-1}. The large supramolecular system, which roughly consists of 500 atoms, was investigated by the GFN2-xTB SQM method, and the interaction energy was benchmarked against the higher-level DFT composite scheme B97-3c.[155] The crystal structure of the double-decker receptor is compared to the GFN2-xTB(GBSA)[167] optimized structure, and to reduce flexibility, the R-O-*n*Bu side chains are truncated to methoxy groups in the computational model. The structure is well reproduced with a root mean square deviation(RMSD) of 0.77 Å. Complex geometries are created by docking the guest molecules using the intermolecular force field xTB-IFF to investigate the fullerene-receptor binding.[224] This molecule-specific force field uses "low-cost" QM information at the GFN2-xTB level from the respective monomers, i.e., Mulliken atomic charges, orbital energies, and centers of localized orbitals, to parameterize an intermolecular potential energy function. Using this function, the interaction energy is minimized, providing the lowest aggregate geometry, which is then optimized by GFN2-xTB(GBSA) and used to evaluate the fullerene binding strength. This equilibrium property is computed from the Gibbs energy of association, which can be split into the contributions from the electronic energy (E), thermostatistical correction to free energy (G_{RRHO}) evaluated at room temperature and the solvation contribution (δG_{solv}). The GFN2-xTB(GBSA) SQM method with implicit solvation is in agreement with the experiment and predicts the C_{70} isomer to have a higher association over the C_{60} fullerene. The absolute value of $\Delta G_a(C_{70}) = -7.4$ kcal mol^{-1}has a slight shift, which can be attributed to the mean field treatment from the implicit solvation model and the energy calculation at the tight binding level of theory. The interaction energies between GFN2-xTB(GBSA) and B97-3c/COSMO are in reasonable qualitative agreement. To showcase the capabilities of the fast, tight binding method and study the inclusion process of C_{70} into the receptor, the reaction path was calculated with the growing string method (GSM).[230] From the reaction path, a barrier-free inclusion process is observed with the fullerene in standing orientation inside the cavity. Further analysis with the interaction strain model[231] reveals that almost no molecular deforma-

tion occurs, and the pure interaction energy of the host-guest system well describes the energy along the reaction path. In the search for the best binding guest fullerene for this specific double-decker receptor, 724 fullerene isomers were ranked by free energy at GFN2-xTB(GBSA) level, and the xTB-IFF force field determined the preferred binding site. The complexes were geometry optimized, and the C_{78} fullerene was found to fill the receptor cavity perfectly and has the lowest association free energy of -10.7 kcal mol^{-1}. Overall, using an established computational protocol [183] in combination with the efficient and robust GFN2-xTB(GBSA), we demonstrate that qualitative estimations of Gibbs free energies for large supramolecular complexes can be routinely obtained, even on regular consumer hardware.

6. Benchmark Study of Electrochemical Redox Potentials Calculated with Semiempirical and DFT Methods

Hagen Neugebauer[*], Fabian Bohle [*], Markus Bursch[*], Andreas Hansen[*], and Stefan Grimme[*]

Received 4th of June 2020, Published 7th of August 2020.

Reprinted (adapted) with permission from[‡]
Neugebauer, H.; Bohle, F.; Bursch, M.; Hansen, A. and Grimme, S. *J. Phys. Chem. A* **2020**, *124*, 7166–7176.
— Copyright © 2020, American Chemical Society.
DOI 10.1021/acs.jpca.0c05052

Own manuscript contribution

- Designing the computational workflow
- Interpretation of the results
- Writing the manuscript

[*]Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany
[‡]Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

Figure 6.1.: Table of contents figure to the article [*J. Phys. Chem. A* **2020**, *124*, 7166–7176.].

The paper is reproduced in Appendix A8. A benchmark study for calculating redox potentials by SQM and selected DFT methods emphasizing the GFN*n*-xTB family is presented. The theoretical methods are reviewed based on a newly assembled data set of 313 experimental redox potentials from small to medium-sized organic and organometallic molecules (10-80 atoms) in various solvents. This set, termed ROP313 (Reduction and Oxidation Potentials 313), is further divided into the organic subset OROP and the organometallic subset OMROP for better analysis. The ab intio modeling of redox potentials requires information about the electronic state of the oxidized and reduced species and the correct description of their interactions with the solvent under equilibrium conditions. The significant solvation contribution to free energy can be routinely computed from implicit solvation models, facilitating a separate description of the gas-phase electronic energy. Various theoretical models are available for calculating the electronic energy, ranging from exact wavefunction theory methods that require extended basis sets and long computation times to methods based on density functional theory that allow routine calculation of larger molecules. However, the main selling point of SQM is undoubtedly the considerable reduction in computational time by at least two orders of magnitude compared to an equivalent DFT treatment. For the routine and basic application of SQM, a parameter set covering large ranges of the PSE must be available, which is the case for the GFN*n*-xTB and PM6/7 methods.[159,160] The focus of this study was to analyze the performance of GFN*n*-xTB in calculating redox potentials, which is an off-target property for the SQM family.

The redox potentials can be readily computed from the Nernst equation.

$$E^{\circ}_{R/O} = -\frac{\Delta_r G^{\circ}_{R/O}}{n_e F} - E^{\circ}_{ABS}(\mathrm{REF}), \tag{6.1}$$

where $\Delta_r G^{\circ}_{R/O}$ is the free energy change of the investigated reduction or oxidation process at standard conditions (ideal gas, p=1 bar; ideal-solute solution, c=1 mol L^{-1}; T = 298 K), n_e is the number of electrons which are transferred in the reaction, F is the Faraday constant and the last term is the absolute potential of the reference electrode (REF).[232] Herein, the free energies of the solutes are obtained from a Born-Haber cycle, and the workflow procedure to obtain the required free energies consists of four steps. Initially, the geometries of the oxidized and reduced species are optimized. Secondly, harmonic vibrational frequencies in the rigid-rotor-harmonic-oscillator (RRHO) approximation are calculated,

providing thermostatistical and ZPVE corrections. Opposed to the SQM workflow, the DFT procedure applies a multilevel Ansatz, wherein the geometries and frequencies are calculated on a computationally affordable level (the B97-3c composite method). Different, usually higher level methods obtain the final electronic energies, here: PWPB95-D4[233], PW6B95-D4[234], B97-D3[141], all employing Ahlrichs' large quadruple-ζ def2-QZVPP basis set. In the fourth step, the solvation contribution is computed from continuum solvation models. With this procedure, redox potential can be calculated via the Gibbs free energy difference according to

$$\Delta_r G^\circ_{R/O} = IP + \Delta G^\circ_{RRHO} + \Delta\delta G^\circ_{solv}, \tag{6.2}$$

where IP is the adiabatic electronic ionization potential of the reduced species and $\Delta\delta G^\circ_{solv}$ denotes the difference between reductant and oxidant's solvation free energies. To counteract the SIE error present in DFT and tight-binding, correction shifts were applied to GFN1- and GFN2-xTB.

The employed methods were checked for suitability in reproducing the experimental redox potentials. For the organic subset (OROP), all tested DFAs yield mean absolute deviations (MADs) between 0.22 and 0.27 V, with the double-hybrid functional PWPB95-D4 being the best performer. The similar performance of the PW6B95-D4 hybrid functional and the double hybrid functional, as well as gas-phase IP/EA evaluations of the GMTKN55 benchmark set, are a clear indication that it is not the electronic structure but the accurate description of solvation that limits the theoretical prediction of redox potentials. At a significantly reduced cost, the GGA B97-D3 yields a MAD of 0.27 V, which is closely followed by the GFNn-xTB methods (MAD_{GFN1}=0.31 and MAD_{GFN2}=0.30 V) gaining almost comparable quality, whereas the PMx methods yield MADs twice as large. Comparing the PMx and GFNn-xTB methods, the GFNn-xTB Hamiltonian proves to possess a more flexible description of anionic systems, that results from the reduction of the neutral educt. In the calculation of redox potentials, the correct evaluation of solvation is a dominant source of error observed for higher-level DFT methods. However, for SQM methods, the error of electronic energy contributions also limits their accuracy.

The organometallic OMROP set poses substantial difficulties for all investigated SQMs where MADs are twice as large as for the surveyed DFT methods, and even these show an increased MAD ($\mathrm{MAD}_{PWPB95-D4}$=0.36 V, $\mathrm{MAD}_{PW6B95-D4}$=0.37 V). This is partly because of the complicated electronic structures, with high multiplicities and high negative charges in the OMROP data set. As a result, reliable screening of redox potentials for these systems has yet to be achievable at the SQM level. However, GFN2-xTB performs better and is more robust than all other investigated SQMs, which can be attributed to its anisotropic electrostatic and XC terms in combination with the applied Fermi smearing. On the DFT side, B97-D3/def2-QZVPP outperforms the higher-level DFAs with a MAD of 0.33 V.

GFNn-xTB clearly outperforms the PMx competitors for the OROP subset. While all investigated SQMs struggle with the electronically complicated systems in the metalorganic OMROP subset, GFN2-xTB performs more robustly than the other SQMs. Considering the main advantage of SQM is the substantial reduction in computation time, we recommend using GFN2-xTB for computing (organic) redox potentials in screening applications or for larger molecules.

Part III.

Beyond the Single Structure Approach: Automated Workflows for Property Predictions of Nonrigid Systems

Automated workflows in quantum chemistry represent concepts saving manual labor and/or computing time in regularly occurring repetitive tasks. Accompanied are many benefits like a clearly defined working procedure instead of an error-prone (always differing) manual approach or the processing of larger amounts of data than can routinely be facilitated manually. Part III is concerned with the development and application of automated workflows for the computation of thermochemistry or spectroscopic properties of flexible molecules. Different minima on a molecular PES, resulting, for example, from the rotation of covalent bonds, can have significant differences in their total energy, drastically affecting the observed Boltzmann population. Computing molecular properties, therefore, depends on identifying a rather complete populated structure (conformer) ensemble and determining the accurately averaged free energy or desired property.

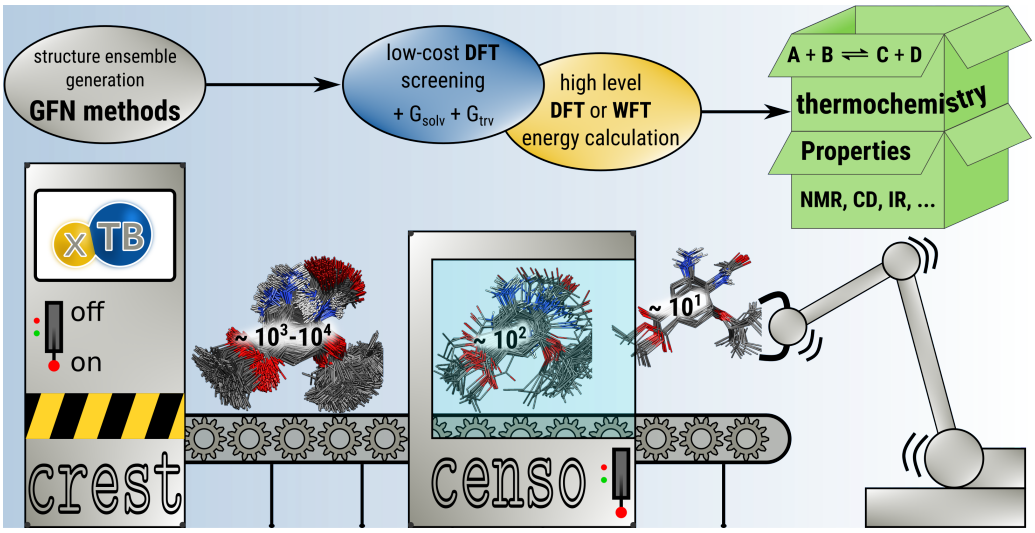


Figure 6.2.: Table of contents figure to the article presented in Chapter 8 [*J. Phys. Chem. A* **2021**, *125*, 4039–4054.].

Figure 6.2 shows the required steps for a property prediction starting from the structure ensemble generation with CREST relying on the GFN methods, followed by the next higher level DFT sorting using the CENSO program and subsequent calculation of the desired molecular properties.
Part III demonstrates the versatility of the CREST + CENSO (i.e., CRENSO) workflows for the quantum chemical calculation of thermodynamic properties such as pK_a and $logK_{OW}$, the investigation of reaction mechanisms, as well as spectroscopic properties such as OR and NMR for nonrigid molecules. The applicability of these workflows in terms of computational cost and robustness is strongly facilitated by the recent development of GFN*n*-xTB and GFN-FF methods, as well as the accuracy of the r^2SCAN-3c composite density functional.

7. Oxidation Under Reductive Conditions: From Benzylic Ethers to Acetals with Perfect Atom-Economy by Titanocene(III) Catalysis

Pierre Funk[*], Ruben B. Richrath[*], Fabian Bohle [†], Stefan Grimme[†] and Andreas Gansäuer[*]

This article is part of the Angew. Chem. Int. Ed. "2021 Oxidation | Hot Paper" collection.

Received 8 October 2020, 27 November 2020

Reprinted under the Creative Commons Attribution License 4.0 (CC BY 4.0)
Funk, P.; Richrath, R. B.; Bohle, F.; Grimme, S.; Gansäuer, A. *Angew. Chem. Int. Ed.* **2021**, *60*, 5482–5488.
— Copyright under the Creative Commons Attribution License 4.0
DOI 10.1002/anie.202013561

Own manuscript contribution

- Performing the calculations
- Interpretation of the results
- Writing the manuscript

[*]Kekulé-Institut für Organische Chemie und Biochemie Universität Bonn Gerhard Domagk Strasse 1, 53121 Bonn, Germany
[†]Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

This cooperation with the workgroup from Prof. Dr. Gansäuer presents a new reaction for synthesizing acetals or hemiaminals starting from respective benzylic ethers or benzylic amines with an attached epoxy group. The paper is reproduced in Appendix A10.

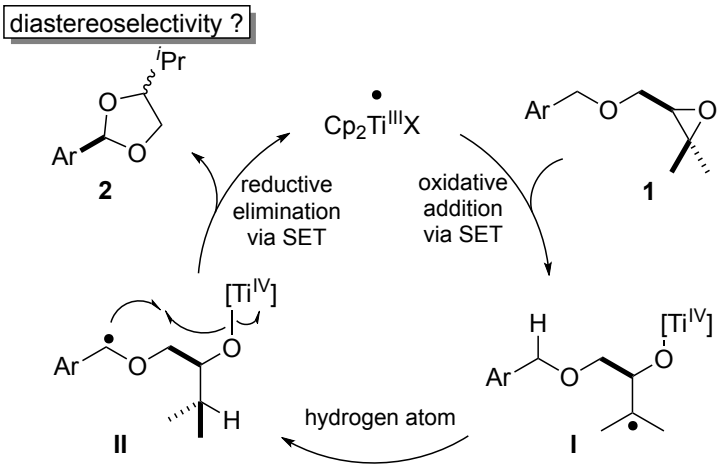


Figure 7.1.: Atom-economic transformation of benzylic ethers into acetals by titanocene(III) catalysis in single-electron steps (SET).

The reaction is catalyzed by activated titanocenes and proceeds in single-electron steps. The titanocene(III) catalyst reacts in an oxidative addition step with the epoxy group of the reactant, forming a strong Ti-O single bond and a carbon-centered radical. The advantage of combining transition metal catalysis and catalysis in single electron steps witnessed in this overall reaction is the efficient generation of radicals while retaining high functional group tolerance and the metal complex's ability to control the selectivities of the reaction proceedings. The oxidative addition is followed by a hydrogen atom transfer from the benzylic position to the previously generated radical, forming a new C-H bond and generating a benzylic radical stabilized by the mesomeric effect. The last step of the catalytic cycle is the reductive elimination, where the product, including a new stereocenter, is formed, and the active catalyst is regenerated. The thermodynamics of the catalytic cycle were investigated with an efficient computational protocol. The conformational space at all stationary states was explored using the CREST[68] code with the fast GFN-FF(GBSA[THF])[181] force field, also taking solvation already at the initial investigation step into account with the GBSA implicit solvation model. To the best of our knowledge, this is the first reaction mechanism exploration, where the conformational influence at the stationary states is investigated in a robust automated procedure using the CREST program. The conformer ensembles are refined concerning energy and geometry optimization at the PBEh-3c/DCOSMO-RS(THF) level, which is realized with the ENSO[235] code. Gibbs energies are evaluated for the lowest lying conformer at each state in the catalytic cycle at the PW6B95-D4/def2-QZVP + COSMO-RS(THF)// PBEh-3c/DCOSMO-RS-(THF) level consisting of the sum of dispersion corrected high-level PW6B95-D4/def2-QZVP single-point energies, thermostatistical contributions at the PBEh-3c/COSMO(THF) level following the mRRHO approach and solvation free energies calculated with

COSMO-RS. The thermodynamic investigation revealed that the experimentally observed highly diastereoselective acetal formation does not stem from thermodynamic control. Furthermore, the investigation of the conformational space identified a conformational relaxation occurring during the H-atom transfer (HAT) reaction with an energy gain of 3.2 kcal mol^{-1}compared to a HAT without a conformer search. This is an important finding, as identifying lower-lying conformers influences all subsequent reactions. The Gibbs energy profile of the reductive elimination reaction and the respective geometries were analyzed to identify why the experimentally observed high diastereoselectivity was observed. The starting conformations for the reductive elimination leading to the *cis* and *trans*-acetal each show two hyperconjugative interactions, but the conformation leading to the *trans*-product additionally exhibits repulsive steric interactions between the substrate and catalyst making this conformation substantially higher in energy (6.5 kcal mol^{-1}) compared to the conformation leading to the *cis*-product. The HAT and conformer inter-conversion reaction barriers were investigated by creating reaction paths using GSM/GFN2-xTB(GBSA[THF]).[230] The reaction paths were refined at the higher PBEh-3c/COSMO(THF) level of theory, transition states were identified, and reaction barriers were obtained. The computed free energy profile of the HAT reaction shows the preferred reaction route to lead to the *cis*-product. Further investigation showed that groups with greater spatial demand at the benzylic position have strong (negative) influence on the diastereoselectivity and an associated lowering of the energetic stability difference of the two conformers leading to the *cis* and *trans*-product, which can be seen in the case of the ketal formation. The thermodynamics of the hemiaminal formation were investigated with the same methodology as described above and showed a lower selectivity as in the ketal formation. Overall, theory and experiment are in good qualitative agreement. We recommend the computational protocol to explore the conformational space at stationary states using the efficient and robust CREST program when investigating mechanistic studies. Here, a new titanocene-catalyzed, atom-economic transformation reaction for synthesizing acetals and hemiaminals from benzylethers and benzylamines with pendant epoxides was presented.

8. Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules

Stefan Grimme*, Fabian Bohle*, Andreas Hansen*, Philipp Pracht*, Sebastian Spicher* and Marcel Stahn*

Received 2nd of February 2021, Published online 10th of March 2021

Reprinted (adapted) with permission from‡
Grimme, S.; Bohle, F.; Hansen, A.; Pracht, P.; Spicher, S.; Stahn, M. *J. Phys. Chem. A* **2021**, *125*, 4039–4054. — Copyright © 2021, American Chemical Society
DOI: 10.1021/acs.jpca.1c00971

Own manuscript contribution

- Writing the CENSO code
- Interpretation of the results
- Writing the manuscript

*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany
‡Reproduced with permission from the American Chemical Society.

This chapter details the theory and workflow required for the efficient quantum mechanical investigation of structure ensembles and their free energies for flexible molecules. The complete paper is reproduced in Appendix A12. Many molecules belong to the class of flexible, i.e., nonrigid molecules whose geometric structure is not adequately described by a single geometry. An ensemble representation of structures in equilibrium corrects the insufficient single-structure assumption. The spatial flexibility, dependent on temperature and physical state (gas, liquid/solution, solid), often originates from several connected covalent single-bonds. They are usually freely rotatable and produce an exponential growth of conformers for increasing system sizes. Conformers are minima on the PES having the same chemical bonding connectivity but differ in their spatial arrangements concerning torsion angles or noncovalent interactions. Exploring the PES in the search for a low-energy conformer ensemble is challenging as the relevant conformations are *a priori* not known, and a huge amount of structures need to be screened. Because of the complex nature of finding minima on the PES, this requires an automated procedure carried out by the CREST code using the fast and robust semiempirical quantum mechanical GFN*n*-xTB and GFN-FF methods. Here, molecular meta-dynamics (MTD) simulations generate conformations that enter a filtering strategy relying on tight geometry optimizations at a uniform theoretical level, enabling energetic and structural comparisons. Thorough testing has shown that errors in the approximate PES at the underlying level of theory and the exhaustive dimensionality can require multiple conformational searches on slightly modified, i.e., more diverse PES for the broad (almost complete) description of the conformational space. This has been realized in the crest_combi script. Identifying relevant conformers under equilibrium conditions is achieved by finding conformers lowest in free energy. The Boltzmann population of the individual conformers can be computed through the Gibbs energy difference, and the conformers' contribution to the overall system can be estimated. Applying the assumption that all degrees of freedom are separable (i.e., electronic, translation, rotation, vibration, solvent) the individual conformers' free energy can be computed from the sum of the gas phase electronic energy (E), the thermostatistical contribution to free energy at temperature T (G^T_{trv}) and in case of solvation, the temperature-dependent solvation free energy (δG^T_{solv}). Since many structures need to be evaluated, all of the aforementioned contributions must be calculated at the best possible and still most efficient level of theory. Therefore, the gas phase electronic energy refinement is performed at DFT level. But even sampling with hybrid density functionals using large basis sets has to be considered too expensive. Providing a balanced description, the "3c" composite methods have established themselves as robust and cost-efficient. Among the "3c" methods, r^{2}SCAN-3c has shown excellent performance concerning conformational energies, thermochemistry, molecular geometries, and technical robustness and is central to the automated refinement of QM calculated (free) energies.[156] The computation of Hessians required for the thermostatistical contribution to free energy in the modified rigid-rotor harmonic-oscillator strategy becomes computationally very demanding for medium to large systems. GFN2-xTB has been shown to be a good alternative to G_{mRRHO} values computed from DFT while being substantially faster at reasonable accuracy,[236] especially when considering relative values as in the case for conformational differences. The applied Single Point Hessian (SPH)[184] approach enables the lower level GFN*n*-xTB calculation of harmonic oscillator vibrational frequencies on a higher level (DFT) geometry. This circumvents problematic cases where the equilibrium geometry on

the lower theoretical level does not match the higher one and the thermostatistical contribution could be evaluated at a mismatching point on the PES. Accounting for solvation effects already in the initial stages is crucial as it affects both geometries and free energies. Implicit solvation models like ALPB or COSMO-RS are fast mean-field methods compared with explicit solvation approaches. They address physically important effects like electrostatics treated by the Coulomb law and non-electrostatics often approximated by a solvent accessible surface area (SASA). These contributions and methodologies are considered in the automated Command-line ENergetic SOrting (CENSO) algorithm. The input structure ensemble is efficiently refined concerning Gibbs energy and molecular geometry by applying three filtering steps. Using robust and efficient methods paired with tested conservative thresholds, relevant conformer candidates are identified quickly, which reduces the computationally costly geometry optimizations at DFT level to the essentials. In the end, a DFT-optimized ensemble ranked by free energy is returned, which can be used for further investigations. There are mainly two modes of application: the "thermo-mode" for accurate thermochemical description (resembling approach II of Fig. 1.2) often dominated by very few low-lying conformers and the overall goal is to identify the global minimum conformation. The "thermo-mode" run-type fields of application are e.g., calculating equilibrium constants. The second run mode is dubbed “property-mode,” where the spectroscopic property often strongly depends on the molecular conformation and only accounting for the single lowest lying one can easily lead to wrong spectral positions and intensities (cf. approach III in Fig. 1.2). CENSO is set up to operate with both run modes and can additionally be used for the automated computation of OR and NMR properties. The capability of this state-of-the-art automated QM workflow is showcased for a small number of molecules for the properties: $\log K_{OW}$, pK_a, and OR. Overall, in this workflow, emphasis is placed on including all relevant physical effects during the structure ensemble creation and higher-level refinement. This automated computational protocol provides relevant populated structures required for highly accurate results while requiring few computational resources.

9. Automated Quantum Chemistry-Based Calculation of Optical Rotation for Large Flexible Molecules

Fabian Bohle[*], Jakob Seibert[*], and Stefan Grimme[*]

Received 20th of August 2021, Published 6th of October 2021

Reprinted (adapted) with permission from[†]
Bohle, F.; Seibert, J.; and Grimme, S. *J. Org. Chem.* **2021**, *86*, 15522–15531.
— Copyright © 2021, American Chemical Society
DOI 10.1021/acs.joc.1c02008

Own manuscript contribution

- Performing and supervising the calculations
- Interpretation of the results
- Writing the manuscript

[*]Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany
[†]Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

Optical rotation (OR) is an important analytical method for assigning absolute or relative configurations. It can even be measured for non-crystallizable compounds and does not require the availability of any functional groups. Knowledge about the OR of a (pure) chiral compound enables the determination of the meaningful enantiomeric excess (*ee*) of a chiral sample, and as such, OR measurements are routinely performed in synthetic laboratories. The full paper is reproduced in Appendix A13. The progress in the field of time-dependent density functional theory (TDDFT) has made the calculation of specific rotations possible even for larger molecules. Many compounds are flexible, and the theoretical description of a structure ensemble is, therefore, more appropriate than considering only a single geometry. The difficulty of computing OR can be attributed to its pronounced sensitivity toward the molecules' three-dimensional structure. It is complicated even further since computed OR values between conformers can change in sign and/or order of magnitude, placing additional demand on the computation of accurate Boltzmann factors and the identification of the true lowest-lying structures. In this study, OR values are computed for larger, flexible, typical drug-like molecules in solution, and for this purpose, we developed and tested an automated QM workflow. Starting from an input geometry, the molecular PES is searched for low-lying minima constituting the structure ensemble. This process is automated by the CREST code that samples the PES using a meta-dynamic algorithm. The effects of the solvent on the PES are treated by the Analytical Linearized Poisson-Boltzmann (ALPB)[180] implicit solvation model, thereby creating solution-phase conformations already in the initial stages. Several different PES are searched to broaden and diversify the conformational search and mitigate potential deficiencies in the underlying semiempirical quantum mechanical and force field methods of the xTB family. The resulting SEs are combined, optimized, and clustered to condense the large ensemble into unique conformations. The "low-level" ensemble is refined at efficient "3c" DFT level using the CENSO[69] workflow. In this automated threshold-based multilevel protocol, gas phase electronic energies, thermostatistical contributions to free energy, and solvation corrections are successively improved. Multiple sorting thresholds drastically reduce the candidates to the relevant contributing ones and limit the number of computationally expensive geometry optimizations. Free energies and Boltzmann populations at the chosen temperature are determined for the DFT-optimized ensemble. TDDFT-based OR calculations are performed for the significantly populated conformers, and an SE averaged OR value is obtained that can be directly compared to the experiment. To facilitate easier usage, the complete workflow, including ensemble generation and higher-level refinement with tested and appropriate thresholds, is combined in the CRENSO "push-the-button" type script. This approach could demonstrate its impressive agreement between the experimental and computed OR values for the small but flexible molecule D-glucose. This challenging example concerning the accurate description of solvation effects validated the usage of implicit solvation models in automated computational workflows. To reduce the influence of the non-deterministic ensemble generation on the final OR value, three independent CRENSO runs with identical settings were averaged. This strategy was applied throughout this benchmark study. We assembled a new benchmark set to test our workflow to compute OR for large nonrigid molecules against experimental data. The set termed "NROR28" comprises 28 nonrigid molecules with system sizes between 28 and 144 atoms, different molecular charges, solvents, and diverse conformational flexibility. When testing the CRENSO workflow on the NROR28 set, the correct sign of the OR values is obtained, and the

average relative deviation of the OR values between theory and experiment is about 72%, which both present a major achievement considering the diversity of these complex and large molecules. The difficulty of computing accurate OR values stems from the high sensitivity toward conformational details, the simultaneous requirement of finding the lowest-lying conformers, and the immense spread in OR values and even signs of individual conformations. An interesting observation of this study was that for most of the NROR28 compounds, the OR value computed from the lowest-lying conformer shows the correct sign, indicating that for a qualitative description, identifying the (dominating) lowest-lying candidate is more important than investigating the whole populated structure ensemble. However, for most of the cases in the NROR28 set, the agreement with the experimental OR values is strongly improved by including the complete ensemble. On average, the ensembles of the nonrigid drug-like molecules only contain roughly eight significantly populated conformers, indicating that even these larger molecules are more rigid than anticipated from their Lewis structures. This suggests that the influence of high-(free)-energy structures in the SE is likely to be dissipated in many cases. We investigated four pairs of flexible diastereomers and could show that the CRENSO workflow can distinguish between the pairs, without any modification, when the difference in OR is sufficiently large. Overall, we could show that these new methods can aid in the computer-assisted assignment of absolute and relative configurations of nonrigid compounds using low computational resources in about 1-2 days of computation time on a standard workstation.

10. Hydrocarbon Macrocycle Conformer Ensembles and ^{13}C-NMR spectra

Fabian Bohle [*], and Stefan Grimme[*]

Received 13th of October 2021, Accepted 31th of January 2022

Reprinted under the Creative Commons Attribution License 4.0 (CC BY-NC 4.0)
Bohle, F.; Grimme, S. *Angew. Chem. Int. Ed.* **2022**, *134*, e202113905.
— Copyright under the Creative Commons Attribution License 4.0
DOI 10.1002/anie.202113905

Own manuscript contribution

- Performing the calculations
- Interpretation of the results
- Writing the manuscript

[*]Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

NMR is a valuable method in the chemist's spectroscopy tool belt, routinely used to elucidate three-dimensional structures in the solution phase. The full paper is reproduced in Appendix A15. Identifying a molecular structure is often the first step in any computational study, followed by reactivity predictions or optimization of chemical properties. When aiming to predict a theoretical NMR spectrum that can be directly compared with the experiment, fast processes on the NMR time scale have to be accounted for. This means that low-barrier conformational processes have to be identified, and the spectrum has to be expressed as an average of the NMR parameters of these conformations. This work investigates a series of conformationally flexible hydrocarbon macrocycles through quantum mechanical calculation of their ^{13}C-NMR spectra using the CRENSO workflow. Compared with the corresponding experimental data, the computed $CDCl_3$ solution phase chemical shifts are gauged. Cycloalkanes from cyclobutane to cycloicosane were chosen as model systems because they are nonrigid with extensive but still explorable PES and can be systematically studied based on their ring size. The basic principle for the calculation of spin-spin coupled NMR spectra is the generation of an almost complete conformer rotamer ensemble (CRE), which is achieved by the availability of the fast GFN*n*-xTB SQM and GFN-FF FF methods that are needed for the efficient search on the PES. In the CRENSO workflow, the CRE is built up by searching on diverse PES and combining the results using a clustering algorithm. The ensemble is efficiently sorted at DFT level, making use of the composite method r^{2}SCAN-3c, which has been shown to provide a good balance between accuracy and computational cost, especially concerning conformational energies.[156] In this threshold-based multilevel approach, structures are optimized, and free energies are computed. This entails the calculation of gas-phase energies, COSMO-RS solvation contributions, and thermostatistical contributions to free energy at the GFN2-xTB/ALPB level of theory. NMR properties like spin-spin coupling constants and chemical shifts can be computed for the significantly populated conformers. From this data, together with the information on chemically and magnetically equivalent atoms provided by CREST, the NMR parameters are properly thermally averaged, the spin Hamiltonian is solved by the anmr[235] code, and a fully coupled spectrum is returned.[194] The variation in ^{13}C NMR chemical shift observed experimentally for the investigated cycloalkanes span a range of ~6 ppm and indicates that conformational effects in the differently large systems play a dominating role. Overall, the experimentally observed chemical shift profile for the cycloalkanes of different sizes is perfectly reproduced by our CRENSO workflow. To remove an almost constant systematic offset between the experiment and the DFT shifts calculated with PBE0/def2-TZVP, linear scaling parameters obtained from a small database of organic molecules can be used. Cycloalkanes can be classified into small, medium, and large ring systems. Depending on their ring size, the flexibility increases, and the observed spectrum is no longer dominated by a single conformer but instead by a structure ensemble in constant equilibrium. Within the larger investigated cycloalkanes the spectrum of cyclononadecane is computed from 205 contributing conformers at 298 K. It could also be shown that only considering the lowest-lying conformer does not yield results consistent with the experiment. A conformational effect is observed between the even and odd-membered hydrocarbon macrocycles, where even-membered rings exhibit smaller (high-field) carbon shifts compared to odd-membered ones, where the placement of the additional methylene group gives rise to increased bond-, angle- and torsion strain. The excellent agreement between theory and experiment has the advantage that the conclusions about the

preferred conformations under the measurement conditions can be derived. Most of the CRENSO calculations for the semi-rigid structures finish within hours to one day of computation time when using 28 cores (Intel(R) Xeon(R) CPU E5-2660 v4 @ 2.00GHz). To showcase that the CRENSO workflow can be routinely applied to more complex compounds, i.e., with organic functional groups, we investigated ^{13}C NMR chemical shifts for a cyclic peptide PGLGF, and two natural products lobophytolins A and (-)-dactylolide. The computational results for these three compounds are impressive, and the excellent agreement is expressed by the mean absolute deviations between the experimental and theoretical carbon chemical shifts ranging from 1.0 to 2.4 ppm. This performance shows that the CRENSO workflow in the context of NMR can compete with existing workflows like the NMR Analysis of Molecular Flexibility in Solution (NAMFIS[237]) or conformational analysis from NMR and density-functional prediction of low-energy ensembles (CANDLE[238]) approaches.

Part IV.

Final Summary and Conclusion

This thesis introduces a new automated Quantum Chemistry-based workflow for making theoretical thermochemistry or spectroscopy property calculations that can be directly compared to the experiment. Multiple technologies must be available and used synergistically to calculate properties such as equilibrium coefficients, NMR chemical shifts, or specific rotations essentially at the push of a button. These properties are computed from thermodynamic ensembles, which must first be created, and relevant contributing structures must be identified based on their Gibbs energy. Therefore, this thesis focuses on the applicability of the recently developed semiempirical methods and DFT-3c composite schemes in multilevel approaches for calculating free energies of organic and organometallic compounds. The GFN*n*-xTB extended tight-binding methods belong to the group of electronic structure methods that are fast and robust enough for routine application in computational studies and are sufficiently accurate. These special-purpose methods for good geometries, frequencies, and non-covalent interactions are parameterized for almost the entire periodic table (up to Z=86), which makes them broadly applicable. The semiempirical nature of the GFN*n*-xTB and GFN-FF methods also impart the necessary robustness so that they can be used in the field of chemical space exploration, for example, with the CREST program or in the evaluation of Gibbs energies of large compound libraries.

Part II focuses on efficiently calculating free energies for organometallic molecular systems. The Gibbs free energy is a key thermodynamic quantity from which insights into chemical equilibrium, reaction direction, and derived macroscopic quantities can be determined. Besides the investigations in benchmark studies[61,75,239], the recently presented efficient GFN*n*-xTB and composite DFT electronic structure methods (DFT-3c) developed by Grimme and coworkers must be tested on real chemical problems.
DFT methods with low computational cost can be applied for electronically complicated systems and are used here to study aromatic donor-acceptor interactions of cyclic trinuclear gold(I) complexes containing pyridinate, imidazolate, or carbeniate ligands. In chapter 3, a cooperation with the workgroup of Prof. Esser is presented, where Gibbs energies of association reactions between supramolecular donor-acceptor molecules were investigated using a multilevel protocol relying on DFT-3c methods. Herein, methods of low computational cost facilitate time-consuming tasks like geometry optimizations or the determination of thermostatistical contributions from harmonic vibrational frequencies. Here, GFN1-xTB was applied for fast optimization of initial geometries, yielding already reasonable complex geometries, which were further refined using the global hybrid PBEh-3c/COSMO density functional scheme additionally taking the electrostatics of solvation effects into account. Free energies were computed from vibrational frequencies at the robust GGA B97-3c level, solvation contributions from COSMO-RS and gas phase energies were obtained with high-level dispersion corrected hybrid DFT in a large basis set (PBE0-D3(BJ)^{ATM}/def2-QZVP). The results obtained with this computational protocol demonstrate that the interaction strength of the donor-acceptor complexes is qualitatively correctly reproduced with gradually increasing π-acidity / π-basicity. A significant solvent influence was observed by comparing the calculated and theoretical association Gibbs energies, which the employed implicit solvation models do not adequately describe.
The following study in Chapter 4 addresses the complex self-sorting behavior in solution of (chiral) isostructural ligands with nitrile and isonitrile moieties bound to a [2.2]paracyclophane backbone upon

coordination with *cis*-protected [(dppp)Pd(OTf)$_2$]. In the laboratory of Prof. Lützen, the ligands $\mathbf{L}^{\mathrm{nit}}$ and $\mathbf{L}^{\mathrm{iso}}$ were synthesized and obtained in racemic and optically pure form. The $\mathbf{L}^{\mathrm{nit}}$ enantiomers could be assigned by comparing simulated and experimental ECD spectra and later on through single crystal XRD analysis. Herein, non-equilibrium structural effects were captured using snapshots from a GFN1-xTB/GBSA(CH_3CN) molecular dynamics simulation on which UV-Vis and ECD spectra were calculated using sTD-BH-LYP/def2-SVP/COSMO,[240] from which the R_p-$\mathbf{L}^{\mathrm{nit}}$ enantiomer could be assigned. The resulting metallosupramolecular aggregates, formed in a solution of noncompetitive solvents, exhibit different compositions and self-sorting for (*rac*)-$\mathbf{L}^{\mathrm{nit}}$ and (*rac*)-$\mathbf{L}^{\mathrm{iso}}$. The thermodynamic stability of experimentally observed di- and trinuclear complexes and their theoretical analogs was investigated on GFN2-xTB/GBSA and verified at the PBE0-D4/def2-QZVP + COSMO-RS//PBEh-3c(DCOSMO-RS) level of theory. Here, GFN2-xTB already supplied suitable initial geometries and provided a significant acceleration in the free energy evaluation, compared to the hybrid DFT multilevel approach and predicted the same trend as the higher-level methods in qualitative observations. It was shown that, among other factors, the binding strength and the binding motif of the ligand with the coordinating metal center significantly influence the self-sorting behavior in solution.

Chapter 5 focuses on the specific purification of fullerene C_{70} in solution achieved by selective encapsulation in a large (about 500 atoms) macrocyclic Zn^{II} complex with an internal cavity. The tight binding method GFN2-xTB was used to evaluate the supramolecular geometries and the Gibbs association energies. Both tasks were performed with satisfactory results; the crystal structures of the Zn macrocycle could be reproduced with a low Cartesian RMSD, and the experimentally observed preferential encapsulation of C_{70} over C_{60} is correctly obtained. Moreover, a predictive screening study for the best binding fullerene was performed based on 724 fullerene isomers, and C_{78} was identified as the one that perfectly fills the space of the inner cavity and has the highest association free energy of all the isomers studied. The results obtained using GFN2-xTB align with experimental trends, enabling qualitative evaluations of free association energies in large supramolecular complexes without substantial computational overhead. This capability is particularly feasible with rapid methods such as GFN2-xTB, which maintain robustness and good accuracy while consistently treating all relevant physical effects. Using the Gibbs energy relationship, macroscopic properties can be expressed for systems in chemical equilibrium. The following chapter 6 deals with calculating organic and organometallic compounds' reduction and oxidation potentials, focusing on semiempirical and selected DFT methods. For this purpose, a new data set termed ROP313, with 313 experimental redox potentials of organic and organometallic small and medium-sized molecules in several solvents, was prepared. For structured analysis, the organic and organometallic subsets were evaluated separately. As it turned out, the intricate electronic structure encountered in the organometallic OMROP subset (often paired with high multiplicities and charges) led to significant deviations from the reference redox potentials for all the investigated semiempirical methods (GFN1&GFN2-xTB as well as PM6 and PM7). However, the robustness of the GFN*n*-xTB methods is striking, especially for GFN2-xTB; all systems could be converged to meaningful results, which can be partially attributed to the Fermi smearing present in GFN*n*-xTB Hamiltonian. Concerning the organic subset, GFN2-xTB reaches DFT accuracy of GGA character at a fraction of the computational cost and outperforms the prominent PM6 and PM7 competitors. The ex-

cellent performance and robustness of GFN2-xTB confirm the developmental progress with respect to previous semiempirical methods and reinforce the applicability for the calculation of redox potentials of large (organic) systems or molecular libraries.

Part II showed real-world application examples based on which experience and understanding could be built to develop an automated workflow. The number of theoretical calculations performed by non-experts is constantly increasing. Within the enormous variety of theoretical methods and associated technical parameters, safe, easy-to-use workflows must be offered that non-experts can operate. In these condensed approaches, a preselection of methods and energy thresholds has to be provided. This requirement has been addressed in Part III, showcasing the development and application of a new quantum chemical workflow.
Chapter 7 presents the mechanistic study of converting benzyl ethers to acetals by titanocene(III) catalysis. The reaction mechanism is studied using high-level dispersion-corrected hybrid density functional theory (PW6B95-D4/def2-QZVP + COSMO-RS(THF)// PBEh-3c/DCOSMO-RS-(THF)). The conformational space in all stationary states is explored using the CREST program and refined at DFT level in an automated procedure. Analysis of the catalytic cycle endorses that the diastereoselectivity of the reaction to the acetal product is not subject to thermodynamic control. The reaction barriers of the stereoselectivity determining reductive elimination reaction were studied using the growing string method (GSM)[230] combined with the robust GFN2-xTB/GBSA and refined at DFT level. It turns out that the diastereoselectivity of the reaction is due to the conformations stabilized by gauche effects, which react to the *cis*/*trans* products. The steric interaction of the titanocene catalyst with the phenyl group of the substrate and the early transition states resembling the conformers, as mentioned earlier, explain the preference of the reaction to the *cis*-acetal. Among other influencing factors, the reaction of substituted benzylamines to hemiaminals was investigated, which follows the same mechanism. However, the much smaller stability difference between the deciding conformers leads to significantly lower diastereoselectivity, which is also observed in this experiment. This study underlines the need for routine conformer searches and automated workflows for energetic ranking. It is necessary for a correct description of reaction mechanisms and efficient work due to the increasing number of species being studied in computational chemistry. Overall, the theoretical and experimental results are in very good agreement, and a jointly established understanding of the reaction mechanism enables more precise reaction control in the synthesis of acetals and hemiaminals.
Chapter 8 presents the CRENSO automated quantum chemical multilevel workflow for determining thermodynamic and spectroscopic properties of nonrigid molecules in solution. Describing molecules with a high number of rotatable bonds can be an uncomfortable problem for computational chemists since it is impossible to determine from their Lewis formula whether a single geometry dominates the property of interest or whether an ensemble, i.e., multiple isomers, is required. For rigorous studies, there is no way around exploring the PES of molecules, which is only routinely feasible with robust and fast methods for molecules with more than 30 atoms. Due to the large number of single-point and gradient calculations required, this task can only be accomplished in an automated fashion. Two main modes of applications have been identified that have different requirements. Application one is the

"thermal" mode, where (if possible) the global minimum is identified on the PES, and its Gibbs energy is determined. In the more complicated case of the "property" mode, the entire ensemble of structures is required in addition to the lowest structure, which is especially necessary for spectroscopic properties. The search of the structure ensemble on the PES is performed using fast but sufficiently accurate GFN2-xTB and GFN-FF methods and the CREST program. In addition to the speed of the methods mentioned above, their robustness and reliable convergence should also be emphasized, which is essential for the first exploration step. The CENSO program sorts and filters the initial ensemble using the meta-GGA composite method r^{2}SCAN-3c, applying energetic safety margins (energy windows) that consider the method error at each filtering step. The desired properties, such as Gibbs energy, OR, or NMR, can be determined for the remaining structures. The presented CRENSO workflow is universally applicable and was developed to process large ensembles (with hundreds of entries) with minimal computational resources. Concerning the different computational approaches for calculating molecular properties, CRENSO represents the most sophisticated computational workflow **III** from Figure 1.2. Its usefulness and applicability have been successfully demonstrated for calculating OR, partition coefficients, and pK_a acid dissociation coefficients.

Chapter 9 describes the investigation of specific rotations $[\alpha]_D$ for typical nonrigid organic drug molecules using the CRENSO workflow. The compound-specific optical or specific rotation is regularly used to evaluate enantiomeric excess or analyze stereochemical relationships. Recently, the theoretical computations of OR were mostly limited to rigid compounds because the OR is very sensitive to small details of the three-dimensional structure. To describe the OR of flexible compounds, all populated conformers must be considered, i.e., a complete ensemble with accurate Boltzmann weights and an efficient but reasonably accurate method to calculate the OR property must be available, which was realized by the CRENSO workflow. The performance of the computational protocol was tested on a newly assembled dataset of flexible compounds in solution termed NROR28, and the computed OR was evaluated against experimentally measured $[\alpha]_D$. For the NROR28 data set, it was demonstrated that the correct OR sign and magnitude were obtained in all cases, which is a tremendous achievement given the high sensitivity of OR. This result proves that the ensembles generated by CRENSO are complete, at least for the populated structures, and that it is possible to obtain Boltzmann weights with sufficient accuracy in a multilevel approach that applies continuum solvation models. An interesting result from the NROR28 evaluation is that the influence of energetically higher lying conformers is often averaged out, and for most cases, the correct OR sign is also correctly reflected by the lowest conformer. However, this does not mean that it is sufficient to consider only the lowest conformer because the example of *D*-glucose already shows that for this comparatively simple molecule, about 30 conformers for the α and β forms must be considered together if a comparison with the experimental results is to be sought. Furthermore, the discrimination between pairs of diastereomers by OR has been successfully demonstrated with the CRENSO protocol that can thus be recommended for the supportive structural elucidation of stereochemical relationships.

The last chapter (chapter 10) of Part III is concerned with the analysis of hydrocarbon macrocycle conformer ensembles and the computation of ^{13}C-NMR spectra and chemical shifts. As in the case of the specific rotation property, calculating NMR spectra requires knowledge of all relevant molecu-

lar conformations. Here, cycloalkanes were chosen as model systems because they have comparatively simple intramolecular interactions, i.e., they do not contain functional groups but still have a fairly large and analyzable potential energy surface. The PES is explored using the fast GFN2-xTB method and the CREST program, which generates the conformer rotamer ensemble (CRE). Unlike OR, NMR requires the knowledge of rotamers to identify all chemically/magnetically equivalent atoms, and the CREST program also generates this information. CENSO automates filtering the initial GFN2-xTB[ALPB] ensemble, and geometry optimizes the contributing conformers. It calculates the Boltzmann populations and the NMR chemical shielding constants from which the NMR chemical shifts are obtained. The CRENSO calculation protocol is tested by comparing calculated and experimental $CHCl_3$ solution chemical shifts of a cycloalkane series from cyclobutane to cycloicosane. The variation in the ^{13}C chemical shift within the cycloalkanes studied is relatively small, about six ppm. Nevertheless, the subtle differences observed between the different ring sizes strongly suggest that conformational effects constitute a significant factor, especially given the high structural similarity between these compounds. The CRENSO protocol reproduces the experimentally obtained shift profile across the different ring sizes, thereby identifying all contributing conformations. In agreement with the experimental data, the cycloalkanes with an odd number of methylene groups show smaller (high-field) chemical shifts, indicating an increased bond, angular, and torsional strain due to an unfavorable positioning of the additional methylene group. Starting with cyclopentadecane, the NMR spectrum is no longer dominated by a single structure but can be well described by an ensemble with many similarly occupied minima. The close agreement between theoretical and experimental results facilitates the deduction of the preferred conformations under the measurement conditions and leads to a better understanding of the investigated compounds based on the structure-reactivity relationship. In addition to the cycloalkane model systems, two macrocyclic natural products and a cyclic peptide were also studied. Again, good agreement was found between the calculated chemical shifts of the carbon atoms and the experiment with low mean absolute deviations of 1.0-2.4 ppm. In all three cases, there was a close agreement between the highest populated conformers found with CRENSO and the experimentally determined crystal structures. The excellent agreement between theory and experiment for the studied macrocycles demonstrated the promising applicability of the CRENSO workflow for automatically calculating chemical shifts of macrocyclic compounds in solution and identifying relevant/dominant conformations.

The applications shown in Part III demonstrate that the choice of DFT-3c and GFN*n*-xTB methods in combination with the selected energy windows are well-balanced and effective, and the results match the experiment or at least reflect the experimental trend. However, this means that the workflow and method development are only partially finalized, and achieving higher consistency between theory and experiment may be possible. This requires several improvements in the complex interplay of several areas. One of the significant difficulties is the correct description of solvation effects for molecular systems in solution. Considering solvent effects is essential, especially in the initial exploration of the PES in search of low-lying conformers. Therefore, further developing more accurate and faster solvent models, predominantly in the SQM area, is essential for notable progress. Other expected developments include new SQM and low-cost DFT methods with low computation time, high robustness, and low error range, especially for organometallic systems. On this basis, more or larger chemical systems could

be investigated. Perhaps more importantly, larger energy windows would be applicable, increasing the confidence that no contributing isomers are mistakenly discarded too early in the energy sorting process. Within a CRENSO run, i.e., the search for popularized conformers/isomers, tens of thousands of single-point calculations are routinely performed. Much of the resulting data is not further used beyond the pure conformer search and could be processed as training data for neural networks. Through a Δ-machine learning[241] potential, the PES of a lower level SQM could be corrected, and a PES with higher accuracy, e.g., close to DFT level, could be obtained. Nonetheless, the integration and confluence of the methods employed in this study constitute progress toward enhancing the predictive capabilities of quantum chemical approaches and closing the divide between theoretical simulations and experimental real-world observations.

In summary, the results shown in this thesis expand the QC toolbox and reduce the complexity of the software available to computational chemists. With the CREST/CENSO technology presented here, the study of complex, nonrigid molecular systems and the calculation of molecular properties such as OR, pKa, logP, and NMR is routinely possible, with low computational resources, "practically at the push of a button," which can be used for predictive purposes or direct comparison with experiments. The presented workflows may facilitate further development of environmental chemistry, drug design, and biochemistry, leading to a much-accelerated exchange between experiment and theory.

Bibliography

[1] Council, N. R. *Mathematical Challenges from Theoretical/Computational Chemistry*; The National Academies Press: Washington, DC, 1995.

[2] Houk, K. N.; Liu, F. *Acc. Chem. Res.* **2017**, *50*, 539–543.

[3] Rogge, S. M. J.; Bavykina, A.; Hajek, J.; Garcia, H.; Olivos-Suarez, A. I.; Sepúlveda-Escribano, A.; Vimont, A.; Clet, G.; Bazin, P.; Kapteijn, F.; Daturi, M.; Ramos-Fernandez, E. V.; Llabrés i Xamena, F. X.; Van Speybroeck, V.; Gascon, J. *Chem. Soc. Rev.* **2017**, *46*, 3134–3184.

[4] Guan, Y.; Ingman, V. M.; Rooks, B. J.; Wheeler, S. E. *J. Chem. Theory Comput.* **2018**, *14*, 5249–5261.

[5] Crawford, J.; Sigman, M. *Synthesis* **2019**, *51*, 1021–1036.

[6] Young, T. A.; Silcock, J. J.; Sterling, A. J.; Duarte, F. *Angew. Chem. Int. Ed.* **2021**, *60*, 4266–4274.

[7] Poree, C.; Schoenebeck, F. *Acc. Chem. Res.* **2017**, *50*, 605–608.

[8] Soyemi, A.; Szilvási, T. *Dalton Trans.* **2021**, *50*, 10325–10339.

[9] Foscato, M.; Jensen, V. R. *ACS Catal.* **2020**, *10*, 2354–2377.

[10] Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. *Nat. Commun.* **2017**, *8*, 872.

[11] Okesola, B. O.; Mata, A. *Chem. Soc. Rev.* **2018**, *47*, 3721–3736.

[12] Pollice, R.; dos Passos Gomes, G.; Aldeghi, M.; Hickman, R. J.; Krenn, M.; Lavigne, C.; Lindner-D'Addario, M.; Nigam, A.; Ser, C. T.; Yao, Z.; Aspuru-Guzik, A. *Acc. Chem. Res.* **2021**, *54*, 849–860.

[13] Zhu, Y. P.; Guo, C.; Zheng, Y.; Qiao, S.-Z. *Acc. Chem. Res.* **2017**, *50*, 915–923.

[14] Rimsza, J. M.; Nenoff, T. M. *ACS Appl. Mater. Interfaces* **2022**, *14*, 18005–18015.

[15] Siahrostami, S.; Villegas, S. J.; Bagherzadeh Mostaghimi, A. H.; Back, S.; Farimani, A. B.; Wang, H.; Persson, K. A.; Montoya, J. *ACS Catal.* **2020**, *10*, 7495–7511.

[16] Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. *APL Mater.* **2013**, *1*, 011002.

[17] Sinko, W.; Lindert, S.; Mccammon, J. A. *Chem. Biol. Drug Des.* **2013**, *81*, 41–49.

[18] Udvarhelyi, A.; Rodde, S.; Wilcken, R. *J. Comput. Aided. Mol. Des.* **2021**, *35*, 399–415.

[19] Montalvo-Acosta, J. J.; Cecchini, M. *Mol. Inform.* **2016**, *35*, 555–567.

[20] Tetko, I. V.; Bruneau, P.; Mewes, H.-W.; Rohrer, D. C.; Poda, G. I. *Drug Discov. Today* **2006**, *11*, 700–707.

[21] Yudin, A. K. *Chem. Sci.* **2015**, *6*, 30–49.

[22] Marcarino, M. O.; Zanardi, M. M.; Cicetti, S.; Sarotti, A. M. *Acc. Chem. Res.* **2020**, *53*, 1922–1932.

[23] Yesiltepe, Y.; Nuñez, J. R.; Colby, S. M.; Thomas, D. G.; Borkum, M. I.; Reardon, P. N.; Washton, N. M.; Metz, T. O.; Teeguarden, J. G.; Govind, N.; Renslow, R. S. *J. Cheminform* **2018**, *10*, 52.

[24] Puzzarini, C.; Bloino, J.; Tasinato, N.; Barone, V. *Chem. Rev.* **2019**, *119*, 8131–8191.

[25] Reimann, M.; Kaupp, M. *J. Phys. Chem. A* **2020**, *124*, 7439–7452.

[26] Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839–1862.

[27] Autschbach, J. *Chirality* **2009**, *21*, E116–E152.

[28] Grimme, S.; Bannwarth, C. *J. Chem. Phys.* **2016**, *145*, 054103.

[29] Kunkes, E. L.; Simonetti, D. A.; West, R. M.; Serrano-Ruiz, J. C.; Gärtner, C. A.; Dumesic, J. A. *Science* **2008**, *322*, 417–421.

[30] Cheng, L.; Assary, R. S.; Qu, X.; Jain, A.; Ong, S. P.; Rajput, N. N.; Persson, K.; Curtiss, L. A. *J. Phys. Chem. Lett.* **2015**, *6*, 283–291.

[31] Agarwal, G.; Doan, H. A.; Robertson, L. A.; Zhang, L.; Assary, R. S. *Chem. Mater.* **2021**, *33*, 8133–8144.

[32] Nandy, A.; Duan, C.; Taylor, M. G.; Liu, F.; Steeves, A. H.; Kulik, H. J. *Chem. Rev.* **2021**, *121*, 9927–10000.

[33] Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.

[34] Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F. *J. Chem. Phys.* **2020**, *153*, 124111.

[35] Axelrod, S.; Gómez-Bombarelli, R. *ArXiv* **2020**, *abs/2012.08452*.

[36] Mater, A. C.; Coote, M. L. *J. Chem. Inf. Model.* **2019**, *59*, 2545–2559.

[37] von Lilienfeld, O. A.; Burke, K. *Nat. Commun.* **2020**, *11*, 4895.

[38] Willett, P. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 46–56.

[39] Smith, J. S.; Isayev, O.; Roitberg, A. E. *Chem. Sci.* **2017**, *8*, 3192–3203.

[40] Gibbs, J. W. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*; Cambridge Library Collection - Mathematics; Cambridge University Press, 2010.

[41] Salthammer, T.; Grimme, S.; Stahn, M.; Hohm, U.; Palm, W.-U. *Environ. Sci. Technol.* **2022**, *56*, 379–391.

[42] Zell, D.; Bursch, M.; Müller, V.; Grimme, S.; Ackermann, L. *Angew. Chem. Int. Ed.* **2017**, *56*, 10378–10382.

[43] Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107–115.

[44] Pracht, P.; Grimme, S. *J. Phys. Chem. A* **2021**, *125*, 5681–5692.

[45] Abraham, M. H.; Ibrahim, A.; Zissimos, A. M.; Zhao, Y. H.; Comer, J.; Reynolds, D. P. *Drug Discov. Today* **2002**, *7*, 1056–1063.

[46] Wells, P. R. *Chem. Rev.* **1963**, *63*, 171–219.

[47] Grimme, S.; Schreiner, P. R. *Angew. Chemie Int. Ed.* **2018**, *57*, 4170–4176.

[48] Hellweg, A.; Eckert, F. Brick by Brick Computation of the Gibbs Free Energy of Reaction in Solution Using Quantum Chemistry and COSMO-RS. 2016; https://arxiv.org/abs/1610.07370.

[49] Skyner, R. E.; McDonagh, J. L.; Groom, C. R.; van Mourik, T.; Mitchell, J. B. O. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6174–6191.

[50] Jensen, J. H. *Phys. Chem. Chem. Phys.* **2015**, *17*, 12441–12451.

[51] Jensen, F. *Introduction to Computational Chemistry*; Wiley: Chichester, UK, 2017.

[52] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.

[53] Deglmann, P.; Schäfer, A.; Lennartz, C. *J. Quantum Chem.* **2015**, *115*, 107–136.

[54] Bursch, M.; Mewes, J.-M.; Hansen, A.; Grimme, S. *ChemRxiv* **2022**,

[55] Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley: Chichester, UK, 2002.

[56] Becke, A. D. *J. Chem. Phys.* **2014**, *140*, 18A301.

[57] Perdew, J. P.; Schmidt, K. *AIP Conf. Proc.* **2001**, *577*, 1–20.

[58] Pribram-Jones, A.; Gross, D. A.; Burke, K. *Annu. Rev. Phys. Chem.* **2015**, *66*, 283–304.

[59] Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. *Chem. Rev.* **2016**, *116*, 5301–5337.

[60] Thiel, W. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 145–157.

[61] Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, e1493.

[62] Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

[63] Warshel, A.; Kato, M.; Pisliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.

[64] Beller, M., Bolm, C., Eds. *Transition Metals for Organic Synthesis, Second Revised and Enlarged Edition*; WILEY-VCH Verlag GmbH & Co. KGaA, 2004.

[65] Zhou, H.-C.; Kitagawa, S. *Chem. Soc. Rev.* **2014**, *43*, 5415–5418.

[66] Doustkhah, E.; Hassandoost, R.; Khataee, A.; Luque, R.; Assadi, M. H. N. *Chem. Soc. Rev.* **2021**, *50*, 2927–2953.

[67] Vogiatzis, K. D.; Polynski, M. V.; Kirkland, J. K.; Townsend, J.; Hashemi, A.; Liu, C.; Pidko, E. A. *Chem. Rev.* **2019**, *119*, 2453–2523.

[68] Pracht, P.; Bohle, F.; Grimme, S. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.

[69] Grimme, S.; Bohle, F.; Hansen, A.; Pracht, P.; Spicher, S.; Stahn, M. *J. Chem. Phys. A* **2021**, *125*, 4039–4054.

[70] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover Publications: New York, 1996.

[71] Cramer, C. *Essentials of Computational Chemistry: Theories and Models*; Wiley, 2005.

[72] Pande, K.; Tenboer, J.; Schmidt, M. Femtosecond Structural Dynamics Drives the Trans/Cis Isomerization in Photoactive Yellow Protein: Dark structure of photoactive yellow protein. 10.2210/pdb5hd3/pdb, 2016.

[73] Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.

[74] Jacob, C. R.; Beyhan, S. M.; Bulo, R. E.; Gomes, A. S. P.; Götz, A. W.; Kiewisch, K.; Sikkema, J.; Visscher, L. *J. Comput. Chem.* **2011**, *32*, 2328–2338.

[75] Bursch, M.; Hansen, A.; Pracht, P.; Kohn, J. T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2021**, *23*, 287–299.

[76] Sure, R.; Grimme, S. *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801.

[77] Schrödinger, E. *Phys. Rev.* **1926**, *28*, 1049–1070.

[78] Born, M.; Oppenheimer, R. *Ann. Phys.* **1927**, *389*, 457–484.

[79] Dirac, P. A. M. *Math. Proc. Cambridge* **1939**, *35*, 416–418.

[80] Hartree, D. R. *Math. Proc. Cambridge* **1928**, *24*, 89–110.

[81] Fock, V. *Z. Physik* **1930**, *61*, 126–148.

[82] Pauli, W. *Phys. Rev.* **1940**, *58*, 716–722.

[83] Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69–89.

[84] Hall, G. G. *Proc. R. Soc. Lond. A* **1951**, *205*, 541–552.

[85] Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652–660.

[86] Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

[87] Löwdin, P.-O. *Phys. Rev.* **1955**, *97*, 1509–1520.

[88] Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, UK, 2016.

[89] Sherrill, D. C.; Schaefer, H. F. In *The Configuration Interaction Method: Advances in Highly Correlated Approaches*; Löwdin, P.-O., Ed.; Advances in Quantum Chemistry; 1999; Vol. 34; pp 143–269.

[90] Szalay, P. G.; Müller, T.; Gidofalvi, G.; Lischka, H.; Shepard, R. *Chem. Rev.* **2012**, *112*, 108–181.

[91] Bartlett, R. J. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359–401.

[92] Bartlett, R. J.; Musiał, M. *Rev. Mod. Phys.* **2007**, *79*, 291–352.

[93] Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.

[94] Grimme, S.; Goerigk, L.; Fink, R. F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 886–906.

[95] Riplinger, C.; Neese, F. *J. Chem. Phys.* **2013**, *138*, 034106.

[96] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. *J. Chem. Phys.* **2013**, *139*, 134101.

[97] Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. *J. Chem. Phys.* **2016**, *144*, 024109.

[98] Parr, R. G.; Yang, W. *Annu. Rev. Phys. Chem.* **1995**, *46*, 701–728.

[99] Thomas, L. H. *Math. Proc. Cambridge Philos. Soc.* **1927**, *23*, 542–548.

[100] Fermi, E. *Z. Physik* **1928**, *48*, 73–79.

[101] Chai, J.-D.; Weeks, J. D. *Phys. Rev. B* **2007**, *75*, 205122.

[102] Dirac, P. A. M. *Math. Proc. Cambridge* **1930**, *26*, 376–385.

[103] García-Aldea, D.; Alvarellos, J. E. *J. Chem. Phys.* **2007**, *127*, 144109.

[104] Laricchia, S.; Constantin, L. A.; Fabiano, E.; Della Sala, F. *J. Chem. Theory Comput.* **2014**, *10*, 164–179.

[105] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.

[106] Engel, E. *Density functional theory : an advanced course*; Springer: Berlin Heidelberg New York, 2011.

[107] Baerends, E. J.; Gritsenko, O. V.; van Meer, R. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16408–16425.

[108] Sun, J.; Remsing, R. C.; Zhang, Y.; Sun, Z.; Ruzsinszky, A.; Peng, H.; Yang, Z.; Paul, A.; Waghmare, U.; Wu, X.; Klein, M. L.; Perdew, J. P. *Nat. Chem.* **2016**, *8*, 831–836.

[109] Mardirossian, N.; Head-Gordon, M. *J. Chem. Theory Comput.* **2016**, *12*, 4303–4325.

[110] Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.

[111] Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.

[112] Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244–13249.

[113] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868; erratum *Phys. Rev. Lett.* **78**, 1396 (1997).

[114] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

[115] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

[116] Bao, J. L.; Gagliardi, L.; Truhlar, D. G. *J. Phys. Chem. Lett.* **2018**, *9*, 2353–2358.

[117] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111*, 10439–10452.

[118] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

[119] Harris, J. *Phys. Rev. A* **1984**, *29*, 1648–1659.

[120] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

[121] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

[122] Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

[123] Tsuneda, T.; Hirao, K. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 375–390.

[124] Gill, P. M.; Adamson, R. D.; Pople, J. A. *Molecular Physics* **1996**, *88*, 1005–1009.

[125] Brémond, É.; Pérez-Jiménez, Á. J.; Sancho-García, J. C.; Adamo, C. *J. Chem. Phys.* **2019**, *150*, 201102.

[126] Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

[127] Karolewski, A.; Kronik, L.; Kümmel, S. *J. Chem. Phys.* **2013**, *138*, 204115.

[128] Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51–57.

[129] Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.

[130] Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 4786–4791.

[131] Brémond, E.; Ciofini, I.; Sancho-García, J. C.; Adamo, C. *Acc. Chem. Res.* **2016**, *49*, 1503–1513.

[132] Paier, J.; Janesko, B. G.; Henderson, T. M.; Scuseria, G. E.; Grüneis, A.; Kresse, G. *J. Chem. Phys.* **2010**, *132*, 094103.

[133] Grimme, S.; Steinmetz, M. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20926–20937.

[134] Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. *Chem. Rev.* **2016**, *116*, 5105–5154.

[135] Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

[136] London, F. *Zeitschrift für Phys.* **1930**, *63*, 245–279.

[137] Eisenschitz, R.; London, F. *Zeitschrift für Phys.* **1930**, *60*, 491–527.

[138] Vydrov, O. A.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *133*, 244103.

[139] Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 073005.

[140] Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. B* **2010**, *82*, 081101.

[141] Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

[142] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

[143] Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[144] Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 024101.

[145] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.

[146] Muto, Y. *Proc. Phys. Math. Soc. Japan* **1943**, *17*, 629.

[147] Axilrod, B. M.; Teller, E. *J. Chem. Phys.* **1943**, *11*, 299–300.

[148] Caldeweyher, E.; Bannwarth, C.; Grimme, S. *J. Chem. Phys.* **2017**, *147*, 034112.

[149] Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. *J. Chem. Phys.* **2019**, *150*, 154122.

[150] Bursch, M.; Caldeweyher, E.; Hansen, A.; Neugebauer, H.; Ehlert, S.; Grimme, S. *Acc. Chem. Res.* **2019**, *52*, 258–266.

[151] Caldeweyher, E.; Brandenburg, J. G. *J. Phys.: Condens. Matter* **2018**, *30*, 213001.

[152] Brandenburg, J. G.; Caldeweyher, E.; Grimme, S. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15519–15523.

[153] Sure, R.; Grimme, S. *J. Comput. Chem.* **2013**, *34*, 1672–1685.

[154] Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. *J. Chem. Phys.* **2015**, *143*, 054107.

[155] Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S. *J. Chem. Phys.* **2018**, *148*, 064104.

[156] Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. *J. Chem. Phys.* **2021**, *154*, 064103.

[157] Kruse, H.; Grimme, S. *J. Chem. Phys.* **2012**, *136*, 154101.

[158] Brandenburg, J. G.; Alessio, M.; Civalleri, B.; Peintinger, M. F.; Bredow, T.; Grimme, S. *J. Phys. Chem. A* **2013**, *117*, 9282–9292.

[159] Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173–1213.

[160] Stewart, J. J. P. *J. Mol. Model.* **2013**, *19*, 1–32.

[161] Elstner, M.; Seifert, G. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2014**, *372*, 20120483.

[162] Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947–12957.

[163] Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 931–948.

[164] Seifert, G.; Porezag, D.; Frauenheim, T. *Int. J. Quantum Chem.* **1996**, *58*, 185–192.

[165] Elstner, M. *Theor. Chem. Acc.* **2006**, *116*, 316–325.

[166] Grimme, S.; Bannwarth, C.; Shushkov, P. *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

[167] Bannwarth, C.; Ehlert, S.; Grimme, S. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.

[168] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.

[169] Hoffmann, R. *J. Chem. Phys.* **1963**, *39*, 1397–1412.

[170] Imamura, A. *J. Chem. Phys.* **1970**, *52*, 3168–3175.

[171] Pracht, P.; Caldeweyher, E.; Ehlert, S.; Grimme, S. *ChemRxiv preprint* **2019**, https://doi.org/10.26434/chemrxiv.8326202.v1.

[172] Ohno, K. *Theor. Chim. Acta* **1964**, *2*, 219–227.

[173] Klopman, G. *J. Am. Chem. Soc.* **1964**, *86*, 4550–4557.

[174] Nishimoto, K.; Mataga, N. *Zeitschrift für Phys. Chemie* **1957**, *12*, 335–338.

[175] Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833–1840.

[176] Köster, A. M.; Leboeuf, M.; Salahub, D. R. In *Molecular Electrostatic Potentials*; Murray, J. S., Sen, K., Eds.; Theoretical and Computational Chemistry; Elsevier, 1996; Vol. 3; pp 105 – 142.

[177] Ghasemi, S. A.; Hofstetter, A.; Saha, S.; Goedecker, S. *Phys. Rev. B* **2015**, *92*, 045131.

[178] Mortier, W. J.; Ghosh, S. K.; Shankar, S. *J. Am. Chem. Soc.* **1986**, *108*, 4315–4320.

[179] Mermin, N. D. *Phys. Rev.* **1965**, *137*, A1441–A1443.

[180] Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. *J. Chem. Theory Comput.* **2021**, *17*, 4250–4261.

[181] Spicher, S.; Grimme, S. *Angew. Chem. Int. Ed.* **2020**, *59*, 15665–15673.

[182] Irikura, K.; Frurip, D. J. *Computational thermochemistry: prediction and estimation of molecular thermodynamics*; American Chemical Society, 1998.

[183] Grimme, S. *Chem. Eur. J.* **2012**, *18*, 9955–9964.

[184] Spicher, S.; Grimme, S. *J. Chem. Theory Comput.* **2021**, *17*, 1701–1714.

[185] Pracht, P.; Grimme, S. *Chem. Sci.* **2021**, *12*, 6551–6568.

[186] Sigalov, G.; Fenley, A.; Onufriev, A. *J. Chem. Phys.* **2006**, *124*, 124902.

[187] Lebedev, V. I.; Laikov, D. N. *Doklady Mathematics* **1999**, *59*, 477–481.

[188] Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *Univ. Minnesota, Minneap.* **2020**,

[189] Klamt, A. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1338.

[190] Tampieri, A.; Szabó, M.; Medina, F.; Gulyás, H. *Phys. Sci. Rev.* **2021**, *6*, 20190086.

[191] Schreckenbach, G.; Ziegler, T. *J. Phys. Chem.* **1995**, *99*, 606–611.

[192] Helgaker, T.; Jaszuński, M.; Ruud, K. *Chem. Rev.* **1999**, *99*, 293–352.

[193] Gauss, J. *Molecular Properties. In Modern Methods and Algorithms of Quantum Chemistry, Proceedings, 2nd ed., Winterschool, Feb. 21–25, 2000, Forschungszentrum Jülich, Germany; Grotendorst, J.*; John von Neumann Institute for Computing: Jülich, German: Jülich, 2000; pp 541–592.

[194] Günther, H. *NMR spectroscopy : basic principles, concepts, and applications in chemistry*; Wiley-VCH: Weinheim, 2013.

[195] Kaupp, M. *Calculation of NMR and EPR parameters : theory and applications*; Wiley-VCH: Weinheim, 2004.

[196] Ramsey, N. F. *Phys. Rev.* **1950**, *78*, 699–703.

[197] Ditchfield, R. *Molecular Physics* **1974**, *27*, 789–807.

[198] Mulder, F. A. A.; Filatov, M. *Chem. Soc. Rev.* **2010**, *39*, 578–590.

[199] Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251–8260.

[200] Stoychev, G. L.; Auer, A. A.; Izsák, R.; Neese, F. *J. Chem. Theory Comput.* **2018**, *14*, 619–637.

[201] Pecul, M.; Ruud, K. *Adv. Quantum Chem.*; Academic Press, 2005; Vol. 50; pp 185–212.

[202] Rosenfeld, L. *Z. Physik* **1929**, *52*, 161–174.

[203] Crawford, T. D. *Theor. Chem. Acc.* **2006**, *115*, 227–245.

[204] Barron, L. D. *Molecular light scattering and optical activity*; Cambridge University Press: Cambridge, UK New York, 2004.

[205] Casida, M. E. *Journal of Molecular Structure: THEOCHEM* **2009**, *914*, 3–18.

[206] Yu, H. S.; Li, S. L.; Truhlar, D. G. *J. Chem. Phys.* **2016**, *145*, 130901.

[207] Autschbach, J.; Ziegler, T. *J. Chem. Phys.* **2002**, *116*, 891–896.

[208] Bauernschmitt, R.; Ahlrichs, R. *Chem. Phys. Lett.* **1996**, *256*, 454–464.

[209] de Wergifosse, M.; Seibert, J.; Grimme, S. *J. Chem. Phys.* **2020**, *153*, 084116.

[210] Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009–4037.

[211] Warnke, I.; Furche, F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 150–166.

[212] Diedrich, C.; Grimme, S. *J. Phys. Chem. A* **2003**, *107*, 2524–2539.

[213] Grimme, S.; Furche, F.; Ahlrichs, R. *Chem. Phys. Lett.* **2002**, *361*, 321–328.

[214] Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*, 2nd ed.; John Wiley & Sons, Ltd: Hoboken, NJ, 2009.

[215] Zheng, J.; Yang, H.; Xie, M.; Li, D. *Chem. Commun.* **2019**, *55*, 7134–7146.

[216] Jiang, W.; Schalley, C. A. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10425–10429.

[217] He, Z.; Jiang, W.; Schalley, C. A. *Chem. Soc. Rev.* **2015**, *44*, 779–789.

[218] Bannwarth, C.; Grimme, S. *Comput. Theor. Chem.* **2014**, *1040-1041*, 45–53.

[219] Truhlar, D. G.; Cramer, C. J.; Lewis, A.; Bumpus, J. A. *J. Chem. Educ.* **2004**, *81*, 596.

[220] Gleiter, R. *Aromaticity and other conjugation effects*; Wiley-VCH: Weinheim, 2012.

[221] Chen, Z.; Lohr, A.; Saha-Möller, C. R.; Würthner, F. *Chem. Soc. Rev.* **2009**, *38*, 564–584.

[222] Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chemie Int. Ed.* **2003**, *42*, 1210–1250.

[223] Hahn, R.; Bohle, F.; Kotte, S.; Keller, T. J.; Jester, S.-S.; Hansen, A.; Grimme, S.; Esser, B. *Chem. Sci.* **2018**, *9*, 3477–3483.

[224] Grimme, S.; Bannwarth, C.; Caldeweyher, E.; Pisarek, J.; Hansen, A. *J. Chem. Phys.* **2017**, *147*, 161708.

[225] Steed, J.; Atwood, J. L. *Supramolecular chemistry*; Wiley: Chichester, UK, 2009.

[226] Cragg, P. J. *Supramolecular chemistry*; Springer: Dordrecht, Netherlands, 2010.

[227] Singh, T.; Marjanović, N.; Matt, G.; Günes, S.; Sariciftci, N.; Montaigne Ramil, A.; Andreev, A.; Sitter, H.; Schwödiauer, R.; Bauer, S. *Org. Electron.* **2005**, *6*, 105–110.

[228] Gñes, S.; Neugebauer, H.; Sariciftci, N. S. *Chem. Rev.* **2007**, *107*, 1324–1338.

[229] Kazemzadeh, H.; Mozafari, M. *Drug Discovery Today* **2019**, *24*, 898–905.

[230] Zimmerman, P. M. *J. Chem. Phys.* **2013**, *138*, 184102.

[231] van Zeist, W.-J.; Bickelhaupt, F. M. *Org. Biomol. Chem.* **2010**, *8*, 3118–3127.

[232] Ho, J.; Coote, M.; Cramer, C.; Truhlar, D. In *Organic Electrochemistry*; Hammerich, O., Speiser, B., Eds.; CRC Press: Boca Raton, 2016; Chapter 6, pp 229–259.

[233] Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 291–309.

[234] Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656–5667.

[235] Grimme, S.; Bannwarth, C.; Dohm, S.; Hansen, A.; Pisarek, J.; Pracht, P.; Seibert, J.; Neese, F. *Angew. Chemie Int. Ed.* **2017**, *56*, 14763–14769.

[236] Spicher, S.; Grimme, S. *J. Phys. Chem. Lett.* **2020**, *11*, 6606–6611.

[237] Cicero, D. O.; Barbato, G.; Bazzo, R. *J. Am. Chem. Soc.* **1995**, *117*, 1027–1033.

[238] Nguyen, Q. N. N.; Schwochert, J.; Tantillo, D. J.; Lokey, R. S. *Phys. Chem. Chem. Phys.* **2018**, *20*, 14003–14012.

[239] Bursch, M.; Hansen, A.; Grimme, S. *Inorg. Chem.* **2017**, *56*, 12485–12491.

[240] Bannwarth, C.; Seibert, J.; Grimme, S. *Chirality* **2016**, *28*, 365–369.

[241] Schmitz, G.; Godtliebsen, I. H.; Christiansen, O. *J. Chem. Phys.* **2019**, *150*, 244113.

[242] Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

[243] Klamt, A.; Diedenhofen, M. *J. Phys. Chem. A* **2015**, *119*, 5439–5445.

[244] van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chemie Int. Ed.* **1990**, *29*, 992–1023.

[245] Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.

[246] Van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Simul.* **1988**, *1*, 173–185.

[247] Hoeben, F. J. M.; Jonkheijm, P.; Meijer, E. W.; Schenning, A. P. H. J. *Chem. Rev.* **2005**, *105*, 1491–1546.

[248] Ariga, K. *Supramolecular Chemistry - Fundamentals and Applications : Advanced Textbook*; Springer-Verlag Berlin Heidelberg: Berlin, Heidelberg, 2006.

[249] Lehn, J.-M. *Supramolecular chemistry* ; Wiley-VCH: Weinheim New York, 1995.

[250] Grimme, S. *Angew. Chemie Int. Ed.* **2008**, *47*, 3430–3434.

[251] Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *J. Chem. Soc., Perkin Trans. 2* **2001**, 651–669.

[252] Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525–5534.

[253] Martinez, C. R.; Iverson, B. L. *Chem. Sci.* **2012**, *3*, 2191.

[254] Foster, R. *Organic charge-transfer complexes*; Academic Press: London, 1969.

[255] Das, A.; Ghosh, S. *Angew. Chemie Int. Ed.* **2014**, *53*, 2038–2054.

[256] Mohr, F. *Gold chemistry : applications and future directions in the life sciences*; Wiley-VCH: Weinheim, 2009.

[257] Burini, A.; Mohamed, A. A.; Fackler, J. P. *Comments Inorg. Chem.* **2003**, *24*, 253–280.

[258] Abdou, H. E.; Mohamed, A. A.; Fackler, J. P.; Burini, A.; Galassi, R.; López-de Luzuriaga, J. M.; Olmos, M. E. *Coord. Chem. Rev.* **2009**, *253*, 1661–1669.

[259] Tekarli, S. M.; Cundari, T. R.; Omary, M. A. *J. Am. Chem. Soc.* **2008**, *130*, 1669–1675.

[260] Omary, M. A.; Mohamed, A. A.; Rawashdeh-Omary, M. A.; Fackler, J. P. *Coord. Chem. Rev.* **2005**, *249*, 1372–1381.

[261] Burini, A.; Fackler, J. P.; Galassi, R.; Grant, T. A.; Omary, M. A.; Rawashdeh-Omary, M. A.; Pietroni, B. R.; Staples, R. J. *J. Am. Chem. Soc.* **2000**, *122*, 11264–11265.

[262] Olmstead, M. M.; Jiang, F.; Attar, S.; Balch, A. L. *J. Am. Chem. Soc.* **2001**, *123*, 3260–3267.

[263] Rawashdeh-Omary, M. A.; Omary, M. A.; Fackler, J. P.; Galassi, R.; Pietroni, B. R.; Burini, A. *J. Am. Chem. Soc.* **2001**, *123*, 9689–9691.

[264] Mohamed, A. A.; Rawashdeh-Omary, M. A.; Omary, M. A.; Fackler, Jr., J. P. *Dalt. Trans.* **2005**, 2597–2602.

[265] Elbjeirami, O.; Rashdan, M. D.; Nesterov, V.; Rawashdeh-Omary, M. A. *Dalt. Trans.* **2010**, *39*, 9465–9468.

[266] Omary, M. A.; Rawashdeh-Omary, M. A.; Gonser, M. W. A.; Elbjeirami, O.; Grimes, T.; Cundari, T. R.; Diyabalanage, H. V. K.; Gamage, C. S. P.; Dias, H. V. R. *Inorg. Chem.* **2005**, *44*, 8200–8210.

[267] Rasika Dias, H. V.; Palehepitiya Gamage, C. S. *Angew. Chemie Int. Ed.* **2007**, *46*, 2192–2194.

[268] Rasika Dias, H. V.; Gamage, C. S. P.; Keltner, J.; Diyabalanage, H. V. K.; Omari, I.; Eyobo, Y.; Dias, N. R.; Roehr, N.; McKinney, L.; Poth, T. *Inorg. Chem.* **2007**, *46*, 2979–2987.

[269] Omary, M. A.; Elbjeirami, O.; Gamage, C. S. P.; Sherman, K. M.; Dias, H. V. R. *Inorg. Chem.* **2009**, *48*, 1784–1786.

[270] Burini, A.; Fackler, J. P.; Galassi, R.; Macchioni, A.; Omary, M. A.; Rawashdeh-Omary, M. A.; Pietroni, B. R.; Sabatini, S.; Zuccaccia, C. *J. Am. Chem. Soc.* **2002**, *124*, 4570–4571.

[271] Klivansky, L. M.; Hanifi, D.; Koshkakaryan, G.; Holycross, D. R.; Gorski, E. K.; Wu, Q.; Chai, M.; Liu, Y. *Chem. Sci.* **2012**, *3*, 2009.

[272] Brynn Hibbert, D.; Thordarson, P. *Chem. Commun.* **2016**, *52*, 12792–12805.

[273] Al Kobaisi, M.; Bhosale, S. V.; Latham, K.; Raynor, A. M.; Bhosale, S. V. *Chem. Rev.* **2016**, *116*, 11685–11796.

[274] Bhosale, S.; Sisson, A. L.; Talukdar, P.; Fürstenberg, A.; Banerji, N.; Vauthey, E.; Bollot, G.; Mareda, J.; Röger, C.; Würthner, F.; Sakai, N.; Matile, S. *Science* **2006**, *313*, 84–86.

[275] Kobaisi, M. A.; Bhosale, R. S.; El-Khouly, M. E.; La, D. D.; Padghan, S. D.; Bhosale, S. V.; Jones, L. A.; Antolasic, F.; Fukuzumi, S.; Bhosale, S. V. *Sci. Rep.* **2017**, *7*, 16501.

[276] Das, A.; Ghosh, S. *Chem. Commun.* **2016**, *52*, 6860–6872.

[277] Zhao, Y.; Sakai, N.; Matile, S. *Nat. Commun.* **2014**, *5*, 3911.

[278] Miros, F. N.; Zhao, Y.; Sargsyan, G.; Pupier, M.; Besnard, C.; Beuchat, C.; Mareda, J.; Sakai, N.; Matile, S. *Chem. Eur. J.* **2016**, *22*, 2648–2657.

[279] Míšek, J.; Vargas Jentzsch, A.; Sakurai, S.-i.; Emery, D.; Mareda, J.; Matile, S. *Angew. Chemie Int. Ed.* **2010**, *49*, 7680–7683.

[280] Zhang, Y.; Zhu, W.; Liu, Y.-L.; Wang, H.; Wang, K.; Li, K.; No, J. H.; Ayong, L.; Gulati, A.; Pang, R.; Freitas-Junior, L.; Morita, C. T.; Oldfield, E. *ACS Med. Chem. Lett.* **2013**, *4*, 423–427.

[281] Chaignon, F.; Falkenström, M.; Karlsson, S.; Blart, E.; Odobel, F.; Hammarström, L. *Chem. Commun.* **2007**, 64–66.

[282] Kishore, R. S. K.; Ravikumar, V.; Bernardinelli, G.; Sakai, N.; Matile, S. *J. Org. Chem.* **2008**, *73*, 738–740.

[283] Akamatsu, M.; Matile, S. *Synlett* **2016**, *27*, 1041–1046.

[284] Semeniuc, R. F.; Baum, R. R.; Veach, J. J.; Wheeler, K. A.; Pellechia, P. J. *Inorganica Chim. Acta* **2013**, *400*, 228–238.

[285] Thordarson, P. *Chem. Soc. Rev.* **2011**, *40*, 1305–1323.

[286] www.supramolecular.org. www.supramolecular.org.

[287] Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029–5036.

[288] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[289] Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224–2235.

[290] Janiak, C. *J. Chem. Soc. Dalt. Trans.* **2000**, 3885–3896.

[291] Vadehra, G. S.; Maloney, R. P.; Garcia-Garibay, M. A.; Dunn, B. *Chem. Mater.* **2014**, *26*, 7151–7157.

[292] Perepichka, D. F.; Perepichka, I. F.; Ivasenko, O.; Moore, A. J.; Bryce, M. R.; Kuz'mina, L. G.; Batsanov, A. S.; Sokolov, N. I. *Chem. Eur. J.* **2008**, *14*, 2757–2770.

[293] Pavelyev, V. G.; Parashchuk, O. D.; Krompiec, M.; Orekhova, T. V.; Perepichka, I. F.; van Loosdrecht, P. H. M.; Paraschuk, D. Y.; Pshenichnikov, M. S. *J. Phys. Chem. C* **2014**, *118*, 30291–30301.

[294] Newman, M. S.; Lutz, W. B. *J. Am. Chem. Soc.* **1956**, *78*, 2469–2473.

[295] Dias, H. V. R.; Diyabalanage, H. V. K.; Eldabaja, M. G.; Elbjeirami, O.; Rawashdeh-Omary, M. A.; Omary, M. A. *J. Am. Chem. Soc.* **2005**, *127*, 7489–7501.

[296] Rasika Dias, H.; Polach, S. A.; Wang, Z. *J. Fluor. Chem.* **2000**, *103*, 163–169.

[297] Prentice, G. M.; Pascu, S. I.; Filip, S. V.; West, K. R.; Pantoş, G. D. *Chem. Commun.* **2015**, *51*, 8265–8268.

[298] Antony, J.; Sure, R.; Grimme, S. *Chem. Commun.* **2015**, *51*, 1764–1774.

[299] McDougald, R. N.; Chilukuri, B.; Jia, H.; Perez, M. R.; Rabaâ, H.; Wang, X.; Nesterov, V. N.; Cundari, T. R.; Gnade, B. E.; Omary, M. A. *Inorg. Chem.* **2014**, *53*, 7485–7499.

[300] Reichardt, C. *Solvents and solvent effects in organic chemistry*; Wiley-VCH: Weinheim, 2011.

[301] TURBOMOLE V7.0.2 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.

[302] Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346.

[303] Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074–5085.

[304] Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

[305] TURBOMOLE V7.2.1 2017, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.

[306] COSMOtherm, C3.0, release 1601, COSMOlogic GmbH & Co KG, `http://www.cosmologic.de`.

[307] Eckert, F.; Klamt, A. *AIChE J.* **2002**, *48*, 369–385.

[308] von Hopffgarten, M.; Frenking, G. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 43–62.

[309] Wu, A.; Isaacs, L. *J. Am. Chem. Soc.* **2003**, *125*, 4831–4835.

[310] Osowska, K.; Miljanić, O. Š. *Synlett* **2011**, *2011*, 1643–1648.

[311] Safont-Sempere, M. M.; Fernández, G.; Würthner, F. *Chem. Rev.* **2011**, *111*, 5784–5814.

[312] Lal Saha, M.; Schmittel, M. *Org. Biomol. Chem.* **2012**, *10*, 4651–4684.

[313] He, Z.; Jiang, W.; Schalley, C. A. *Chem. Soc. Rev.* **2015**, *44*, 779–789.

[314] Taylor, P. N.; Anderson, H. L. *J. Am. Chem. Soc.* **1999**, *121*, 11538–11545.

[315] Krämer, R.; Lehn, J. M.; Marquis-Rigault, A. *Proc. Natl. Acad. Sci.* **1993**, *90*, 5394–5398.

[316] Shivanyuk, A.; Rebek, J. *J. Am. Chem. Soc.* **2002**, *124*, 12074–12075.

[317] Kim, T. W.; Hong, J.-I.; Lah, M. S. *Chem. Commun.* **2001**, *9*, 743–744.

[318] Smith, D. K. *Chem. Soc. Rev.* **2009**, *38*, 684–694.

[319] De Greef, T. F. A.; Smulders, M. M. J.; Wolffs, M.; Schenning, A. P. H. J.; Sijbesma, R. P.; Meijer, E. W. *Chem. Rev.* **2009**, *109*, 5687–5754.

[320] Sato, K.; Itoh, Y.; Aida, T. *Chem. Sci.* **2014**, *5*, 136–140.

[321] Liu, M.; Zhang, L.; Wang, T. *Chem. Rev.* **2015**, *115*, 7304–7397.

[322] Crego-Calama, M., Reinhoudt, D. N., Eds. *Supramolecular Chirality*; Topics in Current Chemistry; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; Vol. 265.

[323] Keene, F. R. *Chirality in supramolecular assemblies: causes and consequences*; John Wiley & Sons, Ltd: Chichester, West Sussex, 2016.

[324] Jędrzejewska, H.; Szumna, A. *Chem. Rev.* **2017**, *117*, 4863–4899.

[325] Kitamura, M.; Okada, S.; Suga, S.; Noyori, R. *J. Am. Chem. Soc.* **1989**, *111*, 4028–4036.

[326] Aspinall, H. C.; Bickley, J. F.; Dwyer, J. L. M.; Greeves, N.; Kelly, R. V.; Steiner, A. *Organometallics* **2000**, *19*, 5416–5423.

[327] Wang, R.; Hong, M.; Yuan, D.; Sun, Y.; Xu, L.; Luo, J.; Cao, R.; Chan, A. S. C. *Eur. J. Inorg. Chem.* **2004**, *2004*, 37–43.

[328] Takacs, J. M.; Hrvatin, P. M.; Atkins, J. M.; Reddy, D. S.; Clark, J. L. *New J. Chem.* **2005**, *29*, 263–265.

[329] Claessens, C. G.; Torres, T. *J. Am. Chem. Soc.* **2002**, *124*, 14522–14523.

[330] Burchell, T. J.; Puddephatt, R. J. *Inorg. Chem.* **2005**, *44*, 3718–3730.

[331] Mizumura, M.; Shinokubo, H.; Osuka, A. *Angew. Chemie Int. Ed.* **2008**, *47*, 5378–5381.

[332] Claessens, C. G.; Sánchez-Molina, I.; Torres, T. *Supramol. Chem.* **2009**, *21*, 44–47.

[333] Weilandt, T.; Kiehne, U.; Schnakenburg, G.; Lützen, A. *Chem. Commun.* **2009**, 2320–2322.

[334] Weilandt, T.; Kiehne, U.; Bunzen, J.; Schnakenburg, G.; Lützen, A. *Chem. Eur. J.* **2010**, *16*, 2418–2426.

[335] Henkelis, J. J.; Carruthers, C. J.; Chambers, S. E.; Clowes, R.; Cooper, A. I.; Fisher, J.; Hardie, M. J. *J. Am. Chem. Soc.* **2014**, *136*, 14393–14396.

[336] Caulder, D. L.; Raymond, K. N. *Angew. Chemie Int. Ed. Engl.* **1997**, *36*, 1440–1442.

[337] Masood, M. A.; Enemark, E. J.; Stack, T. D. P. *Angew. Chem. Int. Ed.* **1998**, *37*, 928–932.

[338] Albrecht, M.; Schneider, M.; Röttele, H. *Angew. Chemie Int. Ed.* **1999**, *38*, 557–559.

[339] Vincent, J.-M.; Philouze, C.; Pianet, I.; Verlhac, J.-B. *Chem. Eur. J.* **2000**, *6*, 3595–3599.

[340] Lützen, A.; Hapke, M.; Griep-Raming, J.; Haase, D.; Saak, W. *Angew. Chem. Int. Ed.* **2002**, *41*, 2086–2089.

[341] Albrecht, M.; Fröhlich, R. *Bull. Chem. Soc. Jpn.* **2007**, *80*, 797–808.

[342] Kiehne, U.; Weilandt, T.; Lützen, A. *Org. Lett.* **2007**, *9*, 1283–1286.

[343] Arnold, P. L.; Buffet, J.-C.; Blaudeck, R. P.; Sujecki, S.; Blake, A. J.; Wilson, C. *Angew. Chemie Int. Ed.* **2008**, *47*, 6033–6036.

[344] Bunzen, J.; Bruhn, T.; Bringmann, G.; Lützen, A. *J. Am. Chem. Soc.* **2009**, *131*, 3621–3630.

[345] Dalla Favera, N.; Kiehne, U.; Bunzen, J.; Hytteballe, S.; Lützen, A.; Piguet, C. *Angew. Chemie Int. Ed.* **2010**, *49*, 125–128.

[346] Gütz, C.; Hovorka, R.; Struch, N.; Bunzen, J.; Meyer-Eppler, G.; Qu, Z.-W.; Grimme, S.; Topić, F.; Rissanen, K.; Cetina, M.; Engeser, M.; Lützen, A. *J. Am. Chem. Soc.* **2014**, *136*, 11830–11838.

[347] Jarzebski, A.; Tenten, C.; Bannwarth, C.; Schnakenburg, G.; Grimme, S.; Lützen, A. *Chem. Eur. J.* **2017**, *23*, 12380–12386.

[348] Struch, N.; Frömbgen, C.; Schnakenburg, G.; Lützen, A. *Eur. J. Org. Chem.* **2017**, *2017*, 4984–4989.

[349] Isaacs, L.; Witt, D. *Angew. Chemie Int. Ed.* **2002**, *41*, 1905–1907.

[350] Kim, H.-J.; Moon, D.; Lah, M. S.; Hong, J.-I. *Angew. Chem. Int. Ed.* **2002**, *41*, 3174–3177.

[351] Hwang, I.-W.; Kamada, T.; Ahn, T. K.; Ko, D. M.; Nakamura, T.; Tsuda, A.; Osuka, A.; Kim, D. *J. Am. Chem. Soc.* **2004**, *126*, 16187–16198.

[352] Burchell, T. J.; Puddephatt, R. J. *Inorg. Chem.* **2006**, *45*, 650–659.

[353] Kamada, T.; Aratani, N.; Ikeda, T.; Shibata, N.; Higuchi, Y.; Wakamiya, A.; Yamaguchi, S.; Kim, K. S.; Yoon, Z. S.; Kim, D.; Osuka, A. *J. Am. Chem. Soc.* **2006**, *128*, 7670–7678.

[354] Maeda, C.; Kamada, T.; Aratani, N.; Osuka, A. *Coord. Chem. Rev.* **2007**, *251*, 2743–2752.

[355] Ronson, T. K.; Fisher, J.; Harding, L. P.; Hardie, M. J. *Angew. Chemie Int. Ed.* **2007**, *46*, 9086–9088.

[356] Schober, K.; Zhang, H.; Gschwind, R. M. *J. Am. Chem. Soc.* **2008**, *130*, 12310–12317.

[357] Gütz, C.; Hovorka, R.; Schnakenburg, G.; Lützen, A. *Chem. Eur. J.* **2013**, *19*, 10890–10894.

[358] Gütz, C.; Hovorka, R.; Stobe, C.; Struch, N.; Topić, F.; Schnakenburg, G.; Rissanen, K.; Lützen, A. *Eur. J. Org. Chem.* **2014**, 206–216.

[359] Meyer-Eppler, G.; Topić, F.; Schnakenburg, G.; Rissanen, K.; Lützen, A. *Eur. J. Inorg. Chem.* **2014**, *2014*, 2495–2501.

[360] Hovorka, R.; Hytteballe, S.; Piehler, T.; Meyer-Eppler, G.; Topić, F.; Rissanen, K.; Engeser, M.; Lützen, A. *Beilstein J. Org. Chem.* **2014**, *10*, 432–441.

[361] Tateishi, T.; Kojima, T.; Hiraoka, S. *Commun. Chem.* **2018**, *1*, 20.

[362] Kai, S.; Kojima, T.; Thorp-Greenwood, F. L.; Hardie, M. J.; Hiraoka, S. *Chem. Sci.* **2018**, *9*, 4104–4108.

[363] Anhäuser, J.; Puttreddy, R.; Lorenz, Y.; Schneider, A.; Engeser, M.; Rissanen, K.; Lützen, A. *Org. Chem. Front.* **2019**, *6*, 1226–1235.

[364] Mukherjee, P. S.; Das, N.; Stang, P. J. *J. Org. Chem.* **2004**, *69*, 3526–3529.

[365] Ward, M. D. *Chem. Commun.* **2009**, 4487.

[366] Jain, V. K.; Jain, L. *Coord. Chem. Rev.* **2010**, *254*, 2848–2903.

[367] Yoshizawa, M.; Fujita, M. *Bull. Chem. Soc. Jpn.* **2010**, *83*, 609–618.

[368] Chakrabarty, R.; Mukherjee, P. S.; Stang, P. J. *Chem. Rev.* **2011**, *111*, 6810–6918.

[369] Inokuma, Y.; Kawano, M.; Fujita, M. *Nat. Chem.* **2011**, *3*, 349–358.

[370] Amouri, H.; Desmarets, C.; Moussa, J. *Chem. Rev.* **2012**, *112*, 2015–2041.

[371] Debata, N. B.; Tripathy, D.; Chand, D. K. *Coord. Chem. Rev.* **2012**, *256*, 1831–1945.

[372] Pollock, J. B.; Cook, T. R.; Stang, P. J. *J. Am. Chem. Soc.* **2012**, *134*, 10607–10620.

[373] Uehara, K.; Oishi, T.; Hirose, T.; Mizuno, N. *Inorg. Chem.* **2013**, *52*, 11200–11209.

[374] Sun, Q.-F.; Sato, S.; Fujita, M. *Angew. Chemie Int. Ed.* **2014**, *53*, 13510–13513.

[375] Grimme, S. *Chem. Eur. J.* **2004**, *10*, 3423–3429.

[376] Bachrach, S. M. *J. Phys. Chem. A* **2011**, *115*, 2396–2401.

[377] Dodziuk, H.; Szymański, S.; Jaźwiński, J.; Ostrowski, M.; Demissie, T. B.; Ruud, K.; Kuś, P.; Hopf, H.; Lin, S.-T. *J. Phys. Chem. A* **2011**, *115*, 10638–10649.

[378] Wolf, H.; Leusser, D.; Jørgensen, M. R. V.; Herbst-Irmer, R.; Chen, Y.-S.; Scheidt, E.-W.; Scherer, W.; Iversen, B. B.; Stalke, D. *Chem. Eur. J.* **2014**, *20*, 7048–7053.

[379] Moffat, J. B. *Chem. Phys. Lett.* **1978**, *55*, 125–130.

[380] Jacopozzi, P.; Dalcanale, E. *Angew. Chemie Int. Ed. Engl.* **1997**, *36*, 613–615.

[381] Mayr, A.; Mao, L.-F. *Inorg. Chem.* **1998**, *37*, 5776–5780.

[382] Fochi, F.; Jacopozzi, P.; Wegelius, E.; Rissanen, K.; Cozzini, P.; Marastoni, E.; Fisicaro, E.; Manini, P.; Fokkens, R.; Dalcanale, E. *J. Am. Chem. Soc.* **2001**, *123*, 7539–7552.

[383] Date, R. W.; Iglesias, E. F.; Rowe, K. E.; Elliott, J. M.; Bruce, D. W. *Dalt. Trans.* **2003**, *10*, 1914–1931.

[384] Zuccaccia, D.; Pirondini, L.; Pinalli, R.; Dalcanale, E.; Macchioni, A. *J. Am. Chem. Soc.* **2005**, *127*, 7025–7032.

[385] Schaly, A.; Rousselin, Y.; Chambron, J.; Aubert, E.; Espinosa, E. *Eur. J. Inorg. Chem.* **2016**, *2016*, 832–843.

[386] Kuznetsov, M. L. *Russ. Chem. Rev.* **2002**, *71*, 265–282.

[387] Bondarenko, L.; Dix, I.; Hinrichs, H.; Hopf, H. *Synthesis* **2004**, *16*, 2751–2759.

[388] Allgeier, H.; Siegel, M. G.; Helgeson, R. C.; Schmidt, E.; Cram, D. J. *J. Am. Chem. Soc.* **1975**, *97*, 3782–3789.

[389] Meyer-Eppler, G.; Sure, R.; Schneider, A.; Schnakenburg, G.; Grimme, S.; Lützen, A. *J. Org. Chem.* **2014**, *79*, 6679–6687.

[390] Hooft, R. W. W.; Straver, L. H.; Spek, A. L. *J. Appl. Crystallogr.* **2008**, *41*, 96–103.

[391] Stang, P. J.; Cao, D. H.; Saito, S.; Arif, A. M. *J. Am. Chem. Soc.* **1995**, *117*, 6273–6283.

[392] Avram, L.; Cohen, Y. *Chem. Soc. Rev.* **2015**, *44*, 586–602.

[393] Engeldinger, E.; Armspach, D.; Matt, D. *Angew. Chem. Int. Ed.* **2001**, *40*, 2526–2529.

[394] Shearer, J.; Jackson, H. L.; Schweitzer, D.; Rittenberg, D. K.; Leavy, T. M.; Kaminsky, W.; Scarrow, R. C.; Kovacs, J. A. *J. Am. Chem. Soc.* **2002**, *124*, 11417–11428.

[395] Ugi, I. *Isonitrile chemistry*; Academic Press: New York, 1971.

[396] Mayr, A.; Guo, J. *Inorg. Chem.* **1999**, *38*, 921–928.

[397] Koiso, N.; Kitagawa, Y.; Nakanishi, T.; Fushimi, K.; Hasegawa, Y. *Inorg. Chem.* **2017**, *56*, 5741–5747.

[398] Konaka, H.; Wu, L. P.; Munakata, M.; Kuroda-Sowa, T.; Maekawa, M.; Suenaga, Y. *Inorg. Chem.* **2003**, *42*, 1928–1934.

[399] Genuis, E. D.; Kelly, J. A.; Patel, M.; McDonald, R.; Ferguson, M. J.; Greidanus-Strom, G. *Inorg. Chem.* **2008**, *47*, 6184–6194.

[400] Niranjana, P.; Pati, A.; Porwal, S. K.; Ramkumar, V.; Gharpure, S. J.; Chand, D. K. *CrystEngComm* **2013**, *15*, 9623–9633.

[401] Wang, Z.; Begum, S.; Krautscheid, H. *Cryst. Growth Des.* **2016**, *16*, 5836–5842.

[402] Brunsveld, L.; Folmer, B. J. B.; Meijer, E. W.; Sijbesma, R. P. *Chem. Rev.* **2001**, *101*, 4071–4098.

[403] Dobrawa, R.; Würthner, F. *J. Polym. Sci. Part A Polym. Chem.* **2005**, *43*, 4981–4995.

[404] Yashima, E.; Maeda, K.; Furusho, Y. *Acc. Chem. Res.* **2008**, *41*, 1166–1180.

[405] Yang, L.; Tan, X.; Wang, Z.; Zhang, X. *Chem. Rev.* **2015**, *115*, 7196–7239.

[406] Yashima, E.; Ousaka, N.; Taura, D.; Shimomura, K.; Ikai, T.; Maeda, K. *Chem. Rev.* **2016**, *116*, 13752–13990.

[407] Shanmugaraju, S.; Hawes, C. S.; Savyasachi, A. J.; Blasco, S.; Kitchen, J. A.; Gunnlaugsson, T. *Chem. Commun.* **2017**, *53*, 12512–12515.

[408] Neese, F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1327.

[409] `https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/stda/stda`.

[410] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

[411] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[412] Kroto, H. W.; Heath, J. R.; O'Brien, S. C.; Curl, R. F.; Smalley, R. E. *Nature* **1985**, *318*, 162–163.

[413] Yan, W.; Seifermann, S. M.; Pierrat, P.; Bräse, S. *Org. Biomol. Chem.* **2015**, *13*, 25–54.

[414] Diederich, F.; Gómez-López, M. *Chem. Soc. Rev.* **1999**, *28*, 263–277.

[415] Kawano, S.-i.; Fukushima, T.; Tanaka, K. *Angew. Chemie Int. Ed.* **2018**, *57*, 14827–14831.

[416] Sure, R.; Grimme, S. *Chem. Commun.* **2016**, *52*, 9893–9896.

[417] Bursch, M.; Neugebauer, H.; Grimme, S. *Angew. Chemie Int. Ed.* **2019**, *58*, 11078–11087.

[418] Ásgeirsson, V.; Bauer, C. A.; Grimme, S. *Chem. Sci.* **2017**, *8*, 4879–4895.

[419] Grimme, S. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.

[420] Zhang, Y.-Q.; Bohle, F.; Bleith, R.; Schnakenburg, G.; Grimme, S.; Gansäuer, A. *Angew. Chemie Int. Ed.* **2018**, *57*, 13528–13532.

[421] Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. *Acta. Cryst.* **2016**, *B72*, 171–179.

[422] Tománek, D. *Guid. Through Nanocarbon Jungle*; Morgan & Claypool Publishers: San Rafael, CA, 2014.

[423] Prier, C. K.; Rankic, D. A.; MacMillan, D. W. C. *Chem. Rev.* **2013**, *113*, 5322–5363.

[424] Connelly, N. G.; Geiger, W. E. *Chem. Rev.* **1996**, *96*, 877–910.

[425] Wang, W.; Sprenkle, V. *Nat. Chem.* **2016**, *8*, 204–206.

[426] Pelzer, K. M.; Cheng, L.; Curtiss, L. A. *J. Phys. Chem. C* **2017**, *121*, 237–245.

[427] Zoski, C. *Handbook of Electrochemistry*; Elsevier Science: Amsterdam, 2007.

[428] Marenich, A. V.; Ho, J.; Coote, M. L.; Cramer, C. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15068–15106.

[429] Isegawa, M.; Neese, F.; Pantazis, D. A. *J. Chem. Theory Comput.* **2016**, *12*, 2272–2284.

[430] Ho, J. *Phys. Chem. Chem. Phys.* **2015**, *17*, 2859–2868.

[431] Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974–12980.

[432] Fu, Y.; Liu, L.; Yu, H.-Z.; Wang, Y.-M.; Guo, Q.-X. *J. Am. Chem. Soc.* **2005**, *127*, 7227–7234.

[433] Roth, H. G.; Romero, N. A.; Nicewicz, D. A. *Synlett* **2016**, *27*, 714–723.

[434] Castro, L.; Bühl, M. *J. Chem. Theory Comput.* **2014**, *10*, 243–251.

[435] Konezny, S. J.; Doherty, M. D.; Luca, O. R.; Crabtree, R. H.; Soloveichik, G. L.; Batista, V. S. *J. Phys. Chem.* **2012**, *116*, 6349–6356.

[436] Kuo, J. L.; Gunasekara, T.; Hansen, A.; Vibbert, H. B.; Bohle, F.; Norton, J. R.; Grimme, S.; Quinlivan, P. J. *Organometallics* **2019**, *38*, 4319–4328.

[437] Maier, S.; Thapa, B.; Raghavachari, K. *Phys. Chem. Chem. Phys.* **2020**, *22*, 4439–4452.

[438] Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.

[439] Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Weber, W.; Steiger, R.; Scholten, M.; Thiel, W. *J. Chem. Theory Comput.* **2016**, *12*, 1082–1096.

[440] Li, G.; Zhang, X.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 8643–8653.

[441] Riccardi, D.; Schaefer, P.; Yang,; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458–6469.

[442] Bhattacharyya, S.; Stankovich, M. T.; Truhlar, D. G.; Gao, J. *J. Phys. Chem. A* **2007**, *111*, 5729–5742.

[443] Jin, H.; Goyal, P.; Das, A. K.; Gaus, M.; Meuwly, M.; Cui, Q. *J. Phys. Chem. B* **2016**, *120*, 1894–1910.

[444] Gieseking, R. L. M.; Ratner, M. A.; Schatz, G. C. *J. Phys. Chem. A* **2018**, *122*, 6809–6818.

[445] Vakulin, I.; Bugaets, D.; Zilberg, R.; Maistrenko, V. *Electrochim. Acta* **2019**, *294*, 423–430.

[446] Constantin, V.-A.; Bongard, D.; Walder, L. *Eur. J. Org. Chem.* **2012**, *2012*, 913–921.

[447] Jovanović, I. N.; Jadreško, D.; Miličević, A.; Hranjec, M.; Perin, N. *Electrochim. Acta* **2019**, *297*, 452–462.

[448] Rzepa, H. S.; Suñer, G. A. *J. Chem. Soc., Chem. Commun.* **1993**, 1743–1744.

[449] Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. *Nat. Mater.* **2013**, *12*, 191–201.

[450] Jinich, A.; Sanchez-Lengeling, B.; Ren, H.; Harman, R.; Aspuru-Guzik, A. *ACS Cent. Sci.* **2019**, *5*, 1199–1210.

[451] Wilbraham, L.; Berardo, E.; Turcani, L.; Jelfs, K. E.; Zwijnenburg, M. A. *J. Chem. Inf. Model* **2018**, *58*, 2450–2459.

[452] Freire, R. O.; Simas, A. M. *J. Chem. Theory Comput.* **2010**, *6*, 2019–2023.

[453] Dutra, J. D. L.; Filho, M. A. M.; Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. *J. Chem. Theory Comput.* **2013**, *9*, 3333–3341.

[454] Bohle, F.; Grimme, S. *J. Serb. Chem. Soc.* **2019**, *84*, 837–844.

[455] Koopman, J.; Grimme, S. *ACS Omega* **2019**, *4*, 15120–15133.

[456] Dohm, S.; Bursch, M.; Hansen, A.; Grimme, S. *J. Chem. Theory Comput.* **2020**, *16*, 2002–2012.

[457] Sterling, C. M.; Bjornsson, R. *J. Chem. Theory Comput.* **2019**, *15*, 52–67.

[458] Tabor, D. P.; Gomez-Bombarelli, R.; Tong, L.; Gordon, R. G.; Aziz, M. J.; Aspuru-Guzik, A. Theoretical and Experimental Investigation of the Stability Limits of Quinones in Aqueous Media: Implications for Organic Aqueous Redox Flow Batteries. 2018; https://doi.org/10.26434/chemrxiv.6990053.v2.

[459] Semiempirical extended tight-binding program package xtb, https://github.com/grimme-lab/xtb. Accessed: 2019-12-19 For help getting started, visit https://xtb-docs.readthedocs.io/en/latest/contents.html.

[460] MOPAC2016, HTTP://OpenMOPAC.net/, Version 19.179L, J. J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA.

[461] Neese, F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 73–78.

[462] Pracht, P.; Grimme, S. Conformer-Rotamer Ensemble Sampling Tool based on the xtb Semiempirical Extended Tight-Binding Program Package `crest`. https://github.com/grimme-lab/crest, Accessed: 2021-01-15.

[463] Katsyuba, S. A.; Zvereva, E. E.; Grimme, S. *J. Phys. Chem. A* **2019**, *123*, 3802–3808.

[464] Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

[465] Lide, D. *CRC Handbook of Chemistry and Physics*, 85th ed.; CRC Press: Boca Raton, 2004.

[466] Baik, M.-H.; Friesner, R. A. *J. Phys. Chem. A* **2002**, *106*, 7407–7412.

[467] Hughes, T. F.; Friesner, R. A. *J. Chem. Theory Comput.* **2012**, *8*, 442–459.

[468] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[469] Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[470] Zheng, J.; Xu, X.; Truhlar, D. G. *Theor. Chem. Acc.* **2011**, *128*, 295–305.

[471] Gansäuer, A.; Fleckhaus, A.; Lafont, M. A.; Okkel, A.; Kotsis, K.; Anoop, A.; Neese, F. *J. Am. Chem. Soc.* **2009**, *131*, 16989–16999.

[472] Gansäuer, A.; Hildebrandt, S.; Vogelsang, E.; Flowers II, R. A. *Dalt. Trans.* **2016**, *45*, 448–452.

[473] van der Vlugt, J. I. *Chem. Eur. J.* **2019**, *25*, 2651–2662.

[474] Huang, H.-M.; Garduño-Castro, M. H.; Morrill, C.; Procter, D. J. *Chem. Soc. Rev.* **2019**, *48*, 4626–4638.

[475] Zard, S. *Radical reactions in organic synthesis*; Oxford University Press: Oxford, 2003.

[476] Renaud, P., Sibi, M. P., Eds. *Radicals in Organic Synthesis*; Wiley: Weinheim, 2001.

[477] Curran, D. P.; Porter, N. A.; Giese, B. *Stereochemistry of Radical Reactions: Concepts, Guidelines, and Synthetic Applications*; Wiley-VCH: Weinheim, 1996.

[478] Gansäuer, A.; Rinker, B.; Pierobon, M.; Grimme, S.; Gerenkamp, M.; Mück-Lichtenfeld, C. *Angew. Chemie Int. Ed.* **2003**, *42*, 3687–3690.

[479] Gansäuer, A.; Rinker, B.; Ndene-Schiffer, N.; Pierobon, M.; Grimme, S.; Gerenkamp, M.; Mück-Lichtenfeld, C. *Eur. J. Org. Chem.* **2004**, 2337–2351.

[480] Dzik, W. I.; Xu, X.; Zhang, X. P.; Reek, J. N. H.; de Bruin, B. *J. Am. Chem. Soc.* **2010**, *132*, 10891–10902.

[481] Lu, H.; Dzik, W. I.; Xu, X.; Wojtas, L.; de Bruin, B.; Zhang, X. P. *J. Am. Chem. Soc.* **2011**, *133*, 8518–8521.

[482] Gansäuer, A.; Klatte, M.; Brändle, G. M.; Friedrich, J. *Angew. Chemie Int. Ed.* **2012**, *51*, 8891–8894.

[483] Henriques, D. S. G.; Zimmer, K.; Klare, S.; Meyer, A.; Rojo-Wiechel, E.; Bauer, M.; Sure, R.; Grimme, S.; Schiemann, O.; Flowers, R. A.; Gansäuer, A. *Angew. Chemie Int. Ed.* **2016**, *55*, 7671–7675.

[484] Hao, W.; Wu, X.; Sun, J. Z.; Siu, J. C.; MacMillan, S. N.; Lin, S. *J. Am. Chem. Soc.* **2017**, *139*, 12141–12144.

[485] Huang, H.-M.; McDouall, J. J. W.; Procter, D. J. *Nat. Catal.* **2019**, *2*, 211–218.

[486] Yao, C.; Dahmen, T.; Gansäuer, A.; Norton, J. *Science* **2019**, *364*, 764–767.

[487] Lankelma, M.; Olivares, A. M.; de Bruin, B. *Chem. Eur. J.* **2019**, *25*, 5658–5663.

[488] Lang, K.; Torker, S.; Wojtas, L.; Zhang, X. P. *J. Am. Chem. Soc.* **2019**, *141*, 12388–12396.

[489] Hu, Y.; Lang, K.; Li, C.; Gill, J. B.; Kim, I.; Lu, H.; Fields, K. B.; Marshall, M.; Cheng, Q.; Cui, X.; Wojtas, L.; Zhang, X. P. *J. Am. Chem. Soc.* **2019**, *141*, 18160–18169.

[490] Zhou, M.; Lankelma, M.; van der Vlugt, J. I.; de Bruin, B. *Angew. Chemie Int. Ed.* **2020**, *59*, 11073–11079.

[491] Nugent, W. A.; RajanBabu, T. V. *J. Am. Chem. Soc.* **1988**, *110*, 8561–8562.

[492] Daasbjerg, K.; Svith, H.; Grimme, S.; Gerenkamp, M.; Mück-Lichtenfeld, C.; Gansäuer, A.; Barchuk, A.; Keller, F. *Angew. Chemie Int. Ed.* **2006**, *45*, 2041–2044.

[493] Gansäuer, A.; Barchuk, A.; Keller, F.; Schmitt, M.; Grimme, S.; Gerenkamp, M.; Mück-Lichtenfeld, C.; Daasbjerg, K.; Svith, H. *J. Am. Chem. Soc.* **2007**, *129*, 1359–1371.

[494] Stateman, L.; Nakafuku, K.; Nagib, D. *Synthesis* **2018**, *50*, 1569–1586.

[495] Brocks, J. J.; Beckhaus, H.-D.; Beckwith, A. L. J.; Rüchardt, C. *J. Org. Chem.* **1998**, *63*, 1935–1943.

[496] Blanksby, S. J.; Ellison, G. B. *Acc. Chem. Res.* **2003**, *36*, 255–263.

[497] Huang, X.; Groves, J. T. *J. Biol. Inorg. Chem.* **2017**, *22*, 185–207.

[498] Zaragoza, J. P. T.; Yosca, T. H.; Siegler, M. A.; Moënne-Loccoz, P.; Green, M. T.; Goldberg, D. P. *J. Am. Chem. Soc.* **2017**, *139*, 13640–13643.

[499] Gansäuer, A.; Behlendorf, M.; von Laufenberg, D.; Fleckhaus, A.; Kube, C.; Sadasivam, D. V.; Flowers, R. A. *Angew. Chemie Int. Ed.* **2012**, *51*, 4739–4742.

[500] Gansäuer, A.; von Laufenberg, D.; Kube, C.; Dahmen, T.; Michelmann, A.; Behlendorf, M.; Sure, R.; Seddiqzai, M.; Grimme, S.; Sadasivam, D. V.; Fianu, G. D.; Flowers II, R. A. *Chem. Eur. J.* **2015**, *21*, 280–289.

[501] Gansäuer, A.; Hildebrandt, S.; Michelmann, A.; Dahmen, T.; von Laufenberg, D.; Kube, C.; Fianu, G. D.; Flowers, R. A. *Angew. Chemie Int. Ed.* **2015**, *54*, 7003–7006.

[502] Richrath, R. B.; Olyschläger, T.; Hildebrandt, S.; Enny, D. G.; Fianu, G. D.; Flowers, R. A.; Gansäuer, A. *Chem. Eur. J.* **2018**, *24*, 6371–6379.

[503] Mühlhaus, F.; Weißbarth, H.; Dahmen, T.; Schnakenburg, G.; Gansäuer, A. *Angew. Chemie Int. Ed.* **2019**, *58*, 14208–14212.

[504] Zhang, Z.; Richrath, R. B.; Gansäuer, A. *ACS Catal.* **2019**, *9*, 3208–3212.

[505] Weißbarth, H.; Mühlhaus, F.; Gansäuer, A. *Synthesis* **2020**, *52*, 2940–2947.

[506] Chakraborty, T. K.; Chattopadhyay, A. K.; Ghosh, S. *Tetrahedron Lett.* **2007**, *48*, 1139–1142.

[507] Inch, T. D.; Williams, N. *J. Chem. Soc. C* **1970**, 263–269.

[508] TURBOMOLE V7.4.1 2019, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.

[509] NBO 6.0. E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, C. R. Landis, and F. Weinhold, Theoretical Chemistry Institute, (Theoretical Chemistry Institute, University of Wisconsin, Madison, **2013**); *http://nbo6.chem.wisc.edu/* .

[510] Eliel, E. L.; Wilen, S. H. *Stereochemistry of organic compounds*; John Wiley and Sons: New York, 1994.

[511] Ehlert, S.; Bannwarth, C.; Grimme, S. Semiempirical Extended Tight-Binding Program Package `xtb`. `https://github.com/grimme-lab/xtb`, Accessed: 2021-01-15. See also xTB Documentation available at https://xtb-docs.readthedocs.io/en/latest/.

[512] Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 91–100.

[513] Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. TURBOMOLE V7.5.1 2021, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from *http://www.turbomole.com.*

[514] Decherchi, S.; Cavalli, A. *Chem. Rev.* **2020**, *120*, 12788–12833.

[515] Funk, P.; Richrath, R. B.; Bohle, F.; Grimme, S.; Gansäuer, A. *Angew. Chem. Int. Ed.* **2021**, *60*, 5482–5488.

[516] Spicher, S.; Abdullin, D.; Grimme, S.; Schiemann, O. *Phys. Chem. Chem. Phys.* **2020**, *22*, 24282–24290.

[517] Pracht, P.; Grant, D. F.; Grimme, S. *J. Chem. Theory Comput.* **2020**, *16*, 7044–7060.

[518] Eckert, F.; Klamt, A. *J. Comput. Chem.* **2006**, *27*, 11–19.

[519] Bochevarov, A. D.; Watson, M. A.; Greenwood, J. R.; Philipp, D. M. *J. Chem. Theory Comput.* **2016**, *12*, 6001–6019.

[520] Haworth, N. L.; Wang, Q.; Coote, M. L. *J. Phys. Chem. A* **2017**, *121*, 5217–5225.

[521] Yepes, D.; Neese, F.; List, B.; Bistoni, G. *J. Am. Chen. Soc.* **2020**, *142*, 3613–3625.

[522] Řezáč, J.; Hobza, P. *Chem. Rev.* **2016**, *116*, 5038–5071.

[523] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. *J. Chem. Phys.* **2013**, *139*, 134101.

[524] Řezáč, J.; Bim, D.; Gutten, O.; Rulisek, L. *J. Chem. Theory Comput.* **2018**, *14*, 1254–1266.

[525] Sharapa, D. I.; Genaev, A.; Cavallo, L.; Minenkov, Y. *ChemPhysChem* **2019**, *20*, 92–102.

[526] Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2012**, *8*, 141–151.

[527] Wales, D. J. *Annu. Rev. Phys. Chem.* **2018**, *69*, 401–425.

[528] Hawkins, P. C. D. *J. Chem. Inf. Model.* **2017**, *57*, 1747–1756.

[529] Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.

[530] Sure, R.; Antony, J.; Grimme, S. *J. Phys. Chem. B.* **2014**, *118*, 3431–3440.

[531] Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 305–317.

[532] Decherchi, S.; Cavalli, A. *Chem. Rev.* **2020**, *120*, 12788–12833.

[533] Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502–16513.

[534] Stratmann, R. E.; Burant, J. C.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1997**, *106*, 10175–10183.

[535] Bootsma, S., Andrea N.; Wheeler *ChemRxiv* **2019**, https://doi.org/10.26434/chemrxiv.8864204.v5.

[536] Spicher, S.; Bursch, M.; Grimme, S. *J. Phys. Chem. C* **2020**, *124*, 27529–27541.

[537] Chang, K.-C.; Minami, T.; Koutnik, P.; Savechenkov, P. Y.; Liu, Y.; Anzenbacher Jr, P. *J. Am. Chem. Soc.* **2014**, *136*, 1520–1525.

[538] Suárez, D.; Díaz, N. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5*, 1–26.

[539] Li, Y.-P.; Bell, A. T.; Head-Gordon, M. *J. Chem. Theory Comput.* **2016**, *12*, 2861–2870.

[540] Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016–2027.

[541] Lüttschwager, N. O. B.; Wassermann, T. N.; Mata, R. A.; Suhm, M. A. *Angew. Chem. Int. Ed.* **2013**, *52*, 463–466.

[542] Born, M. *Z. Physik* **1920**, *1*, 45–48.

[543] Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

[544] Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

[545] Lange, A. W.; Herbert, J. M. *J. Chem. Theory Comput.* **2012**, *8*, 1999–2011.

[546] Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6532–6542.

[547] Klamt, A. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 699–709.

[548] Işık, M.; Bergazin, T. D.; Fox, T.; Rizzi, A.; Chodera, J. D.; Mobley, D. L. SAMPL6 - log P Prediction. Drug design data resource site, https://drugdesigndata.org/about/sampl6/logp-prediction, Accessed on 06.11.2020.

[549] Loschen, C.; Reinisch, J.; Klamt, A. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 385–392.

[550] Sure, R.; el Mahdali, M.; Plajer, A.; Deglmann, P. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 473–492.

[551] PubChem database. https://www.ncbi.nlm.nih.gov/.

[552] Grimme, S.; Bahlmann, A.; Haufe, G. *Chirality* **2002**, *14*, 793–797.

[553] Marchesan, D.; Coriani, S.; Forzato, C.; Nitti, P.; Pitacco, G.; Ruud, K. *J. Phys. Chem. A* **2005**, *109*, 1449–1453.

[554] Kundrat, M. D.; Autschbach, J. *J. Phys. Chem. A* **2006**, *110*, 12908–12917.

[555] Tam, M. C.; Abrams, M. L.; Crawford, T. D. *J. Phys. Chem. A* **2007**, *111*, 11232–11241.

[556] Oseltamivirphosphate, vendor site, Merck. https://www.sigmaaldrich.com/catalog/product/sigma/sml1606 Accessed: 2021-01-22.

[557] O'Neil, M. The Merck index: An encyclopedia of chemicals, drugs, and biologicals. Whitehouse Station, N.J.: Merck., https://www.rsc.org/merck-index, 2006; Accessed: 2021-01-22.

[558] Rivaroxaban, vendor site, AK Scientific Inc. https://aksci.com/item_detail.php?cat=J10374, Accessed: 2021-01-22.

[559] Turner, J. A. E. e-Pesticide Manual Ver. 5.1. British Crop Protection Council, https://www.bcpc.org/product/bcpc-online-pesticide-manual-latest-version, Accessed: 2021-01-22.

[560] Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525–616.

[561] van der Spoel, D.; Manzetti, S.; Zhang, H.; Klamt, A. *ACS Omega* **2019**, *4*, 13772–13781.

[562] Kundi, V.; Ho, J. *J. Phys. Chem. B* **2019**, *123*, 6810–6822.

[563] Cousins, I.; Mackay, D. *Chemosphere* **2000**, *41*, 1389–1399.

[564] Finizio, A.; Vighi, M.; Sandroni, D. *Chemosphere* **1997**, *34*, 131–161.

[565] Chemical substances quality review, Center for Drug Evaluation and Research. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/209803,209805,209806Orig1s000ChemR.pdf, Accessed: 2021-01-22.

[566] Pracht, P.; Wilcken, R.; Udvarhelyi, A.; Rodde, S.; Grimme, S. *J. Comput.-Aided. Mol. Des.* **2018**, *32*, 1139–1149.

[567] Li, N. C.; Miller, G. W.; Solony, N.; Gillis, B. T. *J. Am. Chem. Soc.* **1960**, *82*, 3737–3739.

[568] Burke, D.; Henderson, D. *Br. J. Anaesth.* **2002**, *88*, 563–576.

[569] Smith, S. W. *Toxicol. Sci.* **2009**, *110*, 4–30.

[570] Friedman, L.; Miller, J. G. *Science* **1971**, *172*, 1044–1046.

[571] Brooks, H. W.; Guida, C. W.; Daniel, G. K. *Curr. Top. Med. Chem.* **2011**, *11*, 760–770.

[572] Barron, L. D., Ed. *Molecular Light Scattering and Optical Activity*; Cambridge University Press, 2009.

[573] Polavarapu, P. L., Ed. *Chiral Analysis*, 2nd ed.; Elsevier Science, 2018.

[574] Cheeseman, J. R.; Frisch, M. J.; Devlin, F. J.; Stephens, P. J. *J. Chem. Phys. A* **2000**, *104*, 1039–1046.

[575] Grimme, S. *Chem. Phys. Lett.* **2001**, *339*, 380–388.

[576] Stephens, P. J.; McCann, D. M.; Cheeseman, J. R.; Frisch, M. J. *Chirality* **2005**, *17*, S52–S64.

[577] Kongsted, J.; Pedersen, T. B.; Jensen, L.; Hansen, A. E.; Mikkelsen, K. V. *J. Am. Chem. Soc.* **2006**, *128*, 976–982.

[578] Crawford, T. D.; Stephens, P. J. *J. Chem. Phys. A* **2008**, *112*, 1339–1345.

[579] Srebro, M.; Govind, N.; De Jong, W. A.; Autschbach, J. *J. Chem. Phys. A* **2011**, *115*, 10930–10949.

[580] Autschbach, J.; Ziegler, T.; Van Gisbergen, S. J.; Baerends, E. J. *J. Chem. Phys.* **2002**, *116*, 6930–6940.

[581] Tam, M. C.; Crawford, T. D. *J. Chem. Phys. A* **2006**, *110*, 2290–2298.

[582] Kapitán, J.; Baumruk, V.; Kopecký, V.; Bouř, P. *J. Chem. Phys. A* **2006**, *110*, 4689–4696.

[583] Wiberg, K. B.; Wang, Y. G.; Wilson, S. M.; Vaccaro, P. H.; Jorgensen, W. L.; Crawford, T. D.; Abrams, M. L.; Cheeseman, J. R.; Luderer, M. *J. Phys. Chem. A* **2008**, *112*, 2415–2422.

[584] Kundrat, M. D.; Autschbach, J. *J. Chem. Theory Comput.* **2009**, *5*, 1051–1060.

[585] Egidi, F.; Barone, V.; Bloino, J.; Cappelli, C. *J. Chem. Theory Comput.* **2012**, *8*, 585–597.

[586] Egidi, F.; Giovannini, T.; Del Frate, G.; Lemler, P. M.; Vaccaro, P. H.; Cappelli, C. *Phys. Chem. Chem. Phys.* **2019**, *21*, 3644–3655.

[587] Osol, A.; Anderson, J. T.; Hoover, J. E. *Remington's pharmaceutical sciences*; Mack Publishing Company, 1975.

[588] Agents Classified by the IARC Monographs, Volumes 1–129 – IARC Monographs on the Identification of Carcinogenic Hazards to Humans. monographs.iarc.who.int/agents-classified-by-the-iarc/, Accessed: 2020-12-20.

[589] O'Neil, M., Ed. *The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*, 15th ed.; The Royal Society of Chemistry: New Jersey: Merck, 2013.

[590] MacBean, C., Ed. *e-Pesticide Manual*, 15th ed.; British Crop Protection Council: Alton, UK, 2009.

[591] AKSci.com. www.aksci.com, Accessed: 2020-12-20.

[592] chemietek.com. www.chemietek.com, Accessed: 2020-12-20.

[593] The crenso script is available from : https://github.com/grimme-lab/CRENSO.

[594] da Silva, C. O.; Mennucci, B.; Vreven, T. *J. Org. Chem.* **2004**, *69*, 8161–8164.

[595] Hesse, M.; Meier, H.; Zeeh, B. *Spektroskopische Methoden in der organischen Chemie*; Thieme electronic book library; Thieme, 2005.

[596] D'Cunha, R.; Crawford, T. D. *J. Chem. Phys. A* **2021**, *125*, 3095–3108.

[597] Crawford, T. D.; Tam, M. C.; Abrams, M. L. *J. Phys. Chem. A* **2007**, *111*, 12057–12068.

[598] Srebro, M.; Autschbach, J. *J. Chem. Theory Comput.* **2012**, *8*, 245–256.

[599] Pecul, M.; Ruud, K. In *Response Theory and Molecular Properties (A Tribute to Jan Linderberg and Poul Jørgensen)*; Jensen, H., Ed.; Advances in Quantum Chemistry; Academic Press, 2005; Vol. 50; pp 185–212.

[600] Polavarapu, P. L.; Donahue, E. A.; Shanmugam, G.; Scalmani, G.; Hawkins, E. K.; Rizzo, C.; Ibnusaud, I.; Thomas, G.; Habel, D.; Sebastian, D. *J. Phys. Chem. A* **2011**, *115*, 5665–5673.

[601] Gussem, E. D.; Bultinck, P.; Feledziak, M.; Marchand-Brynaert, J.; Stevens, C. V.; Herrebout, W. *Phys. Chem. Chem. Phys.* **2012**, *14*, 8562–8571.

[602] Giorgio, E.; Roje, M.; Tanaka, K.; Hamersak, Z.; Sunjic, V.; Nakanishi, K.; Rosini, C.; Berova, N. *J. Org. Chem.* **2005**, *70*, 6557–6563.

[603] Turk, J. A.; Visbal, G. S.; Lipton, M. A. *J. Org. Chem.* **2003**, *68*, 7841–7844.

[604] Qiu, S.; Tehrani, K. A.; Sergeyev, S.; Bultinck, P.; Herrebout, W.; Mathieu, B. *Chirality* **2016**, *28*, 215–225.

[605] Ribe, S.; Kondru, R. K.; Beratan, D. N.; Wipf, P. *J. Am. Chem. Soc.* **2000**, *122*, 4608–4617.

[606] TURBOMOLE V7.5 2020, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from https://www.turbomole.org.

[607] COSMOtherm, Release 19; COSMOlogic GmbH & Co. KG, http://www.cosmologic.de.

[608] All software and scripts are available from : https://github.com/grimme-lab.

[609] Mennucci, B. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 386–404.

[610] grimme-lab online documentation. https://xtb-docs.readthedocs.io/en/latest/contents.html.

[611] grimme-lab online documentation. https://xtb-docs.readthedocs.io/en/latest/crest.html.

[612] grimme-lab online documentation. https://xtb-docs.readthedocs.io/en/latest/CENSO_docs/censo.html.

[613] Filby, P. S.; Rayat, S. *J. Org. Chem.* **2018**, *83*, 1790–1796.

[614] Yang, L.-P.; Zhang, L.; Quan, M.; Ward, J. S.; Ma, Y.-L.; Zhou, H.; Rissanen, K.; Jiang, W. *Nat. Commun.* **2020**, *11*, 2740.

[615] Barton, D. H. R. *Science* **1970**, *169*, 539–544.

[616] Anet, F. A. L.; Wagner, J. J. *J. Am. Chem. Soc.* **1971**, *93*, 5266–5268.

[617] Anet, F. A. L.; Cheng, A. K.; Wagner, J. J. *J. Am. Chem. Soc.* **1972**, *94*, 9250–9252.

[618] Anet, F. A. L.; Cheng, A. K. *J. Am. Chem. Soc.* **1975**, *97*, 2420–2424.

[619] Rubin, B. H.; Williamson, M.; Takeshita, M.; Menger, F. M.; Anet, F. A. L.; Bacon, B.; Allinger, N. L. *J. Am. Chem. Soc.* **1984**, *106*, 2088–2092.

[620] Dale, J. *Acta Chem. Scand.* **1973**, *27*, 1115–1129.

[621] Schleyer, P. v. R.; Williams, J. E.; Blanchard, K. R. *J. Am. Chem. Soc.* **1970**, *92*, 2377–2386.

[622] Grant, D. *Encyclopedia of nuclear magnetic resonance*; John Wiley: Chichester, 1996.

[623] Casarini, D.; Lunazzi, L.; Mazzanti, A. *Eur. J. Org. Chem.* **2010**, 2035–2056.

[624] Klages, J.; Neubauer, C.; Coles, M.; Kessler, H.; Luy, B. *ChemBioChem* **2005**, *6*, 1672–1678.

[625] Sternberg, U.; Tzvetkova, P.; Muhle-Goll, C. *Phys. Chem. Chem. Phys.* **2020**, *22*, 17375–17384.

[626] Farès, C.; Lingnau, J. B.; Wirtz, C.; Sternberg, U. *Molecules* **2019**, *24*, 4417.

[627] Bogaerts, J.; Atilaw, Y.; Peintner, S.; Aerts, R.; Kihlberg, J.; Johannessen, C.; Erdélyi, M. *RSC Adv.* **2021**, *11*, 4200–4208.

[628] Begnini, F.; Poongavanam, V.; Atilaw, Y.; Erdelyi, M.; Schiesser, S.; Kihlberg, J. *ACS Med. Chem. Lett.* **2021**, *12*, 983–990.

[629] Andersson, H.; Danelius, E.; Jarvoll, P.; Niebling, S.; Hughes, A. J.; Westenhoff, S.; Brath, U.; Erdélyi, M. *ACS Omega* **2017**, *2*, 508–516.

[630] Onur, A. http://disconsoft.com/, accessed 14.12.2021.

[631] Atasoylu, O.; Furst, G.; Risatti, C.; Smith, A. B. *Org. Lett.* **2010**, *12*, 1788–1791.

[632] Begnini, F.; Poongavanam, V.; Over, B.; Castaldo, M.; Geschwindner, S.; Johansson, P.; Tyagi, M.; Tyrchan, C.; Wissler, L.; Sjö, P.; Schiesser, S.; Kihlberg, J. *J. Med. Chem.* **2021**, *64*, 1054–1072.

[633] Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. *Nat. Rev. Drug Discov.* **2008**, *7*, 608–624.

[634] Bohle, F.; Seibert, J.; Grimme, S. *J. Org. Chem.* **2021**, *86*, 15522–15531.

[635] Anet, F. A. L.; Krane, J. *Isr. J. Chem.* **1980**, *20*, 72–83.

[636] Burke, J. J.; Lauterbur, P. C. *J. Am. Chem. Soc.* **1964**, *86*, 1870–1871.

[637] Fritz, H.; Logemann, E.; Schill, F.; Winkler, T. *Chem. Ber.* **1976**, *109*, 1258–1268.

[638] Chemical Shift Repository, https://cheshireNMR.info.

[639] Dragojlovic, V. *ChemTexts* **2015**, *1*, 1–30.

[640] Wiberg, K. B. *Angew. Chem. Int. Ed.* **1986**, *25*, 312–322.

[641] Winnik, M. A. *Chem. Rev.* **1981**, *81*, 491–524.

[642] Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419–1427.

[643] Zaretsky, S.; Hickey, J. L.; St. Denis, M. A.; Scully, C. C.; Roughton, A. L.; Tantillo, D. J.; Lodewyk, M. W.; Yudin, A. K. *Tetrahedron* **2014**, *70*, 7655–7663.

[644] Zaretsky, S.; Scully, C. C. G.; Lough, A. J.; Yudin, A. K. *Chem. Eur. J.* **2013**, *19*, 17668–17672.

[645] Li, S.-W.; Cuadrado, C.; Yao, L.-G.; Daranas, A. H.; Guo, Y.-W. *Org. Lett.* **2020**, *22*, 4093–4096.

[646] Larsen, E. M.; Wilson, M. R.; Zajicek, J.; Taylor, R. E. *Org. Lett.* **2013**, *15*, 5246–5249.

[647] Prota, A. E.; Bargsten, K.; Zurwerra, D.; Field, J. J.; Díaz, J. F.; Altmann, K.-H.; Steinmetz, M. O. *Science* **2013**, *339*, 587–590.

[648] The `xtb` program package, `crenso` and all used scripts are available free of charge for academic use via the GrimmeLabs Github page: https://github.com/grimme-lab.

[649] Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. *J. Phys. Chem. A* **1999**, *103*, 7357–7363.

[650] Wiberg, K. B.; Pratt, W. E.; Bailey, W. F. *J. Org. Chem.* **1980**, *45*, 4936–4947.

[651] Rice, K. C.; Wasylishen, R. E. *Org. Magn. Res.* **1976**, *8*, 449–452.

[652] Stothers, J. B. *Carbon-13 NMR spectroscopy*; Academic Press: New York, 1972.

[653] Wiberg, K. B. *J. Comput. Chem.* **1984**, *5*, 197–199.

[654] Ibrahim, M. R.; Von Ragué Schleyer, P. *J. Comput. Chem.* **1985**, *6*, 157–167.

[655] Rablen, P. R. *Chemistry* **2020**, *2*, 347–360.

[656] Moriarty, R. M. *Topics in Stereochemistry*; John Wiley & Sons, Ltd, 1974; pp 271–421.

[657] Dunitz, J. D.; Schomaker, V. *J. Chem. Phys.* **1952**, *20*, 1703–1707.

[658] Aston, J. G.; Schumann, S. C.; Fink, H. L.; Doty, P. M. *J. Am. Chem. Soc.* **1941**, *63*, 2029–2030.

[659] Barton, D. H.; Cookson, R. C. *Q. Rev. Chem. Soc.* **1956**, *10*, 44–82.

[660] Wiberg, K. B. *J. Org. Chem.* **2003**, *68*, 9322–9329.

[661] Dorofeeva, O. V.; Mastryukov, V. S.; Allinger, N. L.; Almenningen, A. *J. Phys. Chem.* **1985**, *89*, 252–257.

[662] Anet, F. A.; Wagner, J. J. *J. Am. Chem. Soc.* **1971**, *93*, 5266–5268.

[663] Pawar, D. M.; Smith, S. V.; Mark, H. L.; Odom, R. M.; Noe, E. A. *J. Am. Chem. Soc.* **1998**, *120*, 10715–10720.

[664] Pawar, D. M.; Brown, J.; Chen, K. H.; Allinger, N. L.; Noe, E. A. *J. Org. Chem.* **2006**, *71*, 6512–6515.

[665] Dos Santos, H. F.; Franco, M. L.; Venâncio, M. F.; Ferreira, D. E.; Anconi, C. P.; Rocha, W. R.; De Almeida, W. B. *Int. J. Quantum Chem.* **2012**, *112*, 3188–3197.

[666] Dunitz, J. D.; Shearer, H. M. M. *Helv. Chim. Acta* **1960**, *43*, 18–35.

[667] Anet, F. A.; Cheng, A. K.; Wagner, J. J. *J. Am. Chem. Soc.* **1972**, *94*, 9250–9252.

[668] Anet, F. A.; Rawdah, T. N. *J. Am. Chem. Soc.* **1978**, *100*, 7166–7171.

[669] Anet, F. A. L.; Rawdah, T. N. *J. Am. Chem. Soc.* **1978**, *100*, 7810–7814.

[670] Anet, F. A. L.; Cheng, A. K. *J. Am. Chem. Soc.* **1975**, *97*, 2420–2424.

[671] Shah, A. V.; Dolata, D. P. *J. Comput. Aided Mol. Des.* **1993**, *7*, 103–124.

Part V.

Appendix

A1. Supporting Information to Chapter 2

Appendix A1 contains:

- Outline of energy terms in the GFN*n*–xTB schemes
- Equations for thermostatistical contributions
- Derivation of the conductor-like screening model for real solvents
- Structure ensembles and molecular dynamics

A1.1. Extended Tight-Binding Methods

A1.1.1. Energy Terms of the GFN*n*-xTB methods

The equations shown in the following are adapted from Refs. 61,166,167. The overall energy expressions of the GFN*n*-xTB methods are:

$$E_{GFN1-xTB} = E_{rep} + E_{disp}^{D3} + E_{XB} + E_{EHT} + E_{\gamma} + E_{\Gamma} + G_{Fermi} \tag{A1.1}$$

$$E_{GFN2-xTB} = E_{rep} + E_{EHT} + E_{disp}^{D4'} + E_{\gamma} + E_{AES} + E_{AXC} + E_{\Gamma} + G_{Fermi} \tag{A1.2}$$

Energy Terms present in all GFN*n*-xTB methods

The methods of the GFN family have common components, which are addressed here before the individual GFN-version-specific expressions are discussed. The wave functions are constructed from partially polarized, (mainly) minimal valence basis sets containing spherical Gaussian-type atomic orbital (GTO) functions. Slater type orbitals (STOs) ϕ_{μ}^{Slater} are approximated from contracted GTOs ϕ_{μ}, which themselves are constructed from three to six primitive GTOs χ_z^{μ} and the corresponding contraction coefficients $d_{z\mu}$

$$\phi_{\mu}^{Slater}(r) \approx \phi_{\mu}(r) = \sum_{z}^{N_{prim}^{\mu}} d_{z\mu}\chi_z^{\mu}(r) \tag{A1.3}$$

Molecular orbitals ψ_j are created by linear combinations using this atomic orbital (AO) basis

$$\psi_j(r) = \sum_{\mu}^{N_{AO}} C_{\mu j}\phi_{\mu}(r) \quad . \tag{A1.4}$$

The Roothaan-Hall-type generalized eigenvalue equation (cf. Section 2.2.1) provides the energy expression with respect to the orbital coefficients. The xTB methods use a non-orthogonal basis, and

no zero-differential overlap (ZDO) approximations are invoked. All GFN*n*-xTB methods use a finite electronic temperature treatment[179] at low temperatures ($T_{el} = 300K$), i.e., this so-called "Fermi smearing" allows fractional orbital occupations and introduces static correlation effects into a formal single reference method. This electronic entropy term, which allows dissociation of covalent bonds, is given by

$$G_{\text{Fermi}} = k_B T_{el} \sum_i \sum_{\sigma=\alpha,\beta} \left[n_{i\sigma} \ln\left(n_{i\sigma}\right) + \left(1 - n_{i\sigma}\right) \ln\left(1 - n_{i\sigma}\right) \right] \quad , \tag{A1.5}$$

and is included in the GFN total energy expression (cf. Eq. A1.1). Within the G_{Fermi} expression, k_B is the Boltzmann constant, and T_{el} is an electronic temperature. The fractional occupation number $n_{i\sigma}$ of the spin-MO $\psi_{i\sigma}$ is computed from the Fermi-distribution

$$n_{i\sigma} = \left(\exp\left[\left(\epsilon_i - \epsilon_F^{\sigma}\right)/k_B T_{el}\right] + 1\right)^{-1} \quad , \tag{A1.6}$$

where the occupation is dependent on the energy ϵ_i of the spatial MO ψ_i and the Fermi level ϵ_F^{σ} within the respective spin (α or β) orbital space. The total occupation of the spatial MO ψ_i can be calculated from the sum of $n_{i\alpha}$ and $n_{i\beta}$. The xTB Hamiltonians are spin-independent and are formally spin-restricted. Thus, the Fermi smearing introduces a spin dependence and can be used for open-shell systems. The repulsion energy in all GFN methods is calculated from a classical atom-pairwise approach

$$E_{rep} = \frac{1}{2} \sum_{A,B} \frac{Z_A^{eff} Z_B^{eff}}{R_{AB}} e^{-\sqrt{\alpha_A \alpha_B}\left(R_{AB}\right)^{k_f}} \quad . \tag{A1.7}$$

Here, Z^{eff} are element-specific fitting parameters that roughly correspond to the effective nuclear charges screened by the core reference density ρ_0^{core}. R_{AB} is the inter-nuclei distance, k_f is treated as a global parameter and the α exponents are element-specific parameters. At the heart of typical tight-binding methods is the extended Hückel (EHT) expression, through which the formation of covalent bonds is made possible

$$E_{EHT} = \sum_{\mu\nu} P_{\mu\nu} H_{\nu\mu}^{EHT} \quad . \tag{A1.8}$$

The entries of the EHT Hamiltonian matrix are constructed from

$$H_{\mu\nu}^{EHT} = \frac{1}{2} \; K_{AB}^{ll'} \; S_{\mu\nu} \left(H_{\mu\mu} + H_{\nu\nu}\right) \; X(EN_A, EN_B) \; \Pi(R_{AB}, l, l') \; Y(\zeta_l^A, \zeta_{l'}^B) \quad . \tag{A1.9}$$

The summation in Eq. A1.8 runs over the AOs indicated by the indices μ, ν, and the same nomenclature is used in the EHT Hamiltonian entries. Atoms are denoted by A, B, and l, l' indicate shells. Also, $\mu \in l(A)$ and $\nu \in l'(B)$. $K_{AB}^{ll'}$ is a shell specific scaling constant, $S_{\mu\nu}$ is the overlap matrix element between the AOs ϕ_μ and ϕ_ν. $X(EN_A, EN_B)$ is a function depending on the electronegativities of atoms A and B, and $Y(\zeta_l^A, \zeta_{l'}^B)$ is a GFN2-xTB specific shell-exponent dependent term. The shell and distance

dependent term $\Pi(R_{AB}, l, l')$ is

$$\Pi(R_{AB}, l, l') = \left(1 + k_{A,l}^{poly}\left(\frac{R_{AB}}{R_{cov,AB}}\right)^{\frac{1}{2}}\right)\left(1 + k_{B,l'}^{poly}\left(\frac{R_{AB}}{R_{cov,AB}}\right)^{\frac{1}{2}}\right) \tag{A1.10}$$

wherein k_l^{poly} are shell and element specific fitted parameters and R_{cov} are summed covalent radii of atoms A and B. The complex form of Eq. A1.9 allows for a great deal of flexibility to describe various covalent bonding situations without needing conceptually undesirable element-pair-specific parameters.

A1.1.2. Energies in GFN1–xTB

In GFN1-xTB the diagonal elements of the EHT Hamiltonian matrix $H_{\mu\mu}$ are given by

$$H_{\mu\mu}^{GFN1} = h_A^l\left(1 + k_{CN,l}CN_A\right) \tag{A1.11}$$

and describe the chemical environment by a coordination number dependent function CN_A [142,143] which adjusts the shell- and element-specific parameter h_A^l together with the global angular momentum specific parameter $k_{CN,l}$. The electronegativity dependent component $X(EN_A, EN_B)$ is

$$X(EN_A, EN_B) = \left(1 + k_{EN}\Delta EN_{AB}^2\right) \ , \tag{A1.12}$$

herein, ΔEN is the difference of the electronegativities and k_{EN} a global parameter.
The isotropic second-order shell-dependent electrostatic and exchange-correlation tight-binding energy is given by

$$E_\gamma^{GFN1} = \frac{1}{2}\sum_{A,B}\sum_{l\in A}\sum_{l'\in B} q_l\, q_{l'}\, \gamma_{AB,ll'} \ , \tag{A1.13}$$

where q_l represents Mulliken [175] partial shell charges. The two-electron integrals are approximated by a Mataga-Nishimoto-Ohno-Klopman damped Coulomb operator $\gamma_{AB,ll'}$

$$\gamma_{AB,ll'} = \frac{1}{\sqrt{R_{AB}^2 + \eta_{AB,ll'}^{-2}}} \tag{A1.14}$$

where the short-range damping function contains an element specific hardness parameter η

$$\eta_{AB,ll'} = 2\left(\frac{1}{\eta_A(1+\kappa_A^l)} + \frac{1}{\eta_B(1+\kappa_B^{l'})}\right)^{-1} \tag{A1.15}$$

and a shell-dependent scaling parameter κ. The third order on-site ES/XC contribution is given as

$$E_\Gamma^{GFN1} = \frac{1}{3}\sum_A (q_A)^3\Gamma_A \ , \tag{A1.16}$$

with q_A being Mulliken atomic partial charges and Γ_A as an element-specific parameter. The dispersion energy E_{disp}^{D3} is calculated by the D3 dispersion correction model[142,143] see Eq. 2.37. The last remaining term is the geometry-dependent halogen-bond (XB) correction.

$$E_{XB}^{GFN1} = \sum_{AB}^{N_{XB}} f_{damp}^{AXB} k_X \left[\left(\frac{k_{XR} R_{cov,AX}}{R_{AX}} \right)^{12} - k_{X2} \left(\frac{k_{XR} R_{cov,AX}}{R_{AX}} \right)^{6} \right] \left[\left(\frac{k_{XR} R_{cov,AX}}{R_{AX}} \right)^{12} - 1 \right]^{-1} \tag{A1.17}$$

Eq. A1.17 contains the global parameters k_{XR} and k_{X2}, the halogen-specific parameter k_X and an angular three-body damping function f_{damp}^{AXB}.

A1.1.3. Energies in GFN2–xTB

The elements of the Hamiltonian matrix $H_{\mu\nu}^{EHT}$ follow the same basic framework (Eq. A1.9) as in GFN1 and have the same electronegativity dependent scaling function $X(EN_A, EN_B)$, but in addition an AO exponent ζ dependent term

$$Y(\zeta_l^A, \zeta_{l'}^B) = \left(\frac{2\sqrt{\zeta_l^A \zeta_{l'}^B}}{\zeta_l^A + \zeta_{l'}^B} \right)^{\frac{1}{2}} \tag{A1.18}$$

is introduced to approximate the effects of one-electron kinetic energy integrals as in HF or KS-DFT. Diagonal elements of the Hamiltonian matrix are formulated by

$$H_{\mu\mu}^{GFN2} = h_A^l - \delta h_{CN'_A}^l CN'_A \,, \tag{A1.19}$$

where h_A^l and $\delta h_{CN'_A}^l$ are shell- and element-specific parameters and CN'_A is a modified[167] version of the D3 coordination number.

The isotropic second order electrostatic energy E_γ is identical to Eq. A1.13, but is augmented with a modified damping function

$$\eta_{AB,ll'} = \frac{1}{2} \left[\eta_A (1 + \kappa_A^l) + \eta_B (1 + \kappa_B^{l'}) \right] \,. \tag{A1.20}$$

GFN2-xTB goes beyond the sole description of the second-order isotropic electrostatic and exchange-correlation energy and additionally applies anisotropic electrostatic (AES) and exchange-correlation

(AXC) terms defined by

$$
\begin{aligned}
E_{AES} =& E_{q\mu} + E_{q\Theta} + E_{\mu\mu} && \text{(A1.21)}\\
=& \frac{1}{2}\sum_{A,B}\{f_3(R_{AB})\left[q_A(\boldsymbol{\mu}_B^T \mathrm{R}_{BA}) + q_B(\boldsymbol{\mu}_A^T \mathrm{R}_{AB})\right]\\
&+ f_5(R_{AB})[q_A \mathrm{R}_{AB}^T \boldsymbol{\Theta}_B \mathrm{R}_{AB} + q_B \mathrm{R}_{AB}^T \boldsymbol{\Theta}_A \mathrm{R}_{AB}\\
&- 3(\boldsymbol{\mu}_A^T \mathrm{R}_{AB})(\boldsymbol{\mu}_B^T \mathrm{R}_{AB}) + (\boldsymbol{\mu}_A^T \boldsymbol{\mu}_B) R_{AB}^2]\} && \text{(A1.22)}\\
E_{AXC} =& \sum_A \left(f_{XC}^{\mu_A} |\boldsymbol{\mu}_A|^2 + f_{XC}^{\Theta_A} ||\boldsymbol{\Theta}_A||^2\right) \ . && \text{(A1.23)}
\end{aligned}
$$

Here, the cumulative atomic multipole moments (CAMM)[176] describe the anisotropic interaction between the atomic charge q_A, the dipole moment μ_A, and quadrupole moment Θ_A. The distance-dependent scaling function $f_n(R_{AB})$ damps the terms at short distances and thereby avoids divergence issues. $f_{XC}^{\mu_A}$ and $f_{XC}^{\Theta_A}$ occurring in the anisotropic exchange-correlation term are fitted, element-specific parameters. The dispersion energy at second order is described by a modified self-consistent charge version of the D4[148,149] model given by

$$
\begin{aligned}
E_{disp}^{D4'} =& -\sum_{A>B}\sum_{n=6,8} s_n \frac{C_n^{AB}(q_A, CN_A, q_B, CN_B)}{R_{AB}^n} f_{damp,BJ}^{(n)}(R_{AB})\\
&- s_9 \sum_{A>B>C} \frac{(3\cos(\theta_{ABC})\cos(\theta_{BCA})\cos(\theta_{CAB}) + 1)\, C_9^{ABC}(CN_{cov}^A, CN_{cov}^B, CN_{cov}^C)}{(R_{AB}R_{AC}R_{BC})^3} \qquad \text{(A1.24)}\\
&\times f_{damp,zero}^{(9)}(R_{AB}, R_{AC}, R_{BC}) \ .
\end{aligned}
$$

Here the dispersion energy depends on the atomic charges from a Mulliken population analysis, the geometry-dependent coordination number, and the Becke-Johnson $f_{damp,BJ}^{(n)}$ and zero damping $f_{damp,zero}^{(9)}$ functions. The highest (third) order on-site IES and IXC term in GFN2-xTB is similar to its GFN1 equivalent (Eq. A1.16), but is shell-dependent and is affected by a global shell-specific parameter K_l^Γ

$$
E_\Gamma^{GFN2} = \frac{1}{3}\sum_A \sum_{l\in A} (q_l)^3 K_l^\Gamma \Gamma_A \ . \qquad \text{(A1.25)}
$$

Compared to its predecessor, the improved physics in GFN2-xTB makes empirical XB-corrections unnecessary.

A1.2. Equations for statistical thermodynamics

Table A1.1.: Symbols that are used for derivation of thermostatistical quantities:

Q	partition function
g_i	degeneracy of state i
ϵ_i	energy of state i
k (k_B)	Boltzmann's constant
T	the temperature
M	mass of the molecule
h	Planck's constant
V	volume of molecule confined in a cubic box (molar volume of an ideal gas)
R	the gas constant
N_A	Avogadro's number
I_A, I_B, I_C	moment of inertia around principle axes of rotation A, B, and C
ν_i	the vibrational frequency for mode i
σ	rotational symmetry number

The total energy of an isolated molecule is described by the product of the individual partition functions according to:

$$Q_{elec} = \sum_i g_i e^{-\epsilon_i/kT} \tag{A1.26a}$$

$$Q_{trans} = (2\pi MkT)^{\frac{3}{2}} h^{-3} V \tag{A1.26b}$$

$$Q_{rot} = \left(\frac{8\pi^2 kT}{h^2}\right)^{\frac{3}{2}} \frac{\sqrt{\pi I_A I_B I_C}}{\sigma} \tag{A1.26c}$$

$$Q_{vib} = \prod_i^{modes} \frac{e^{-h\nu_i/2kT}}{1 - e^{-h\nu_i/kT}} \quad . \tag{A1.26d}$$

By performing the differentiation in Eq. 2.52 the enthalpy and entropy terms can be calculated.

Enthalpy expressions

$$H_{tot} = H_{elec} + H_{trans} + H_{rot} + H_{vib} \tag{A1.27a}$$

$$H_{elec} = RT\frac{\sum_i g_i(\epsilon_i/kT)e^{-\epsilon_i/kT}}{\sum_i g_i e^{-\epsilon_i/kT}} \tag{A1.27b}$$

$$H_{trans} = \frac{5}{2}RT \tag{A1.27c}$$

$$H_{rot}^{non-linear} = \frac{3}{2}RT \tag{A1.27d}$$

$$H_{rot}^{linear} = RT \tag{A1.27e}$$

$$H_{vib} = RT\sum_i^{modes}\left(\frac{h\nu_i}{kT}\right)\frac{e^{-h\nu_i/kT}}{1-e^{-h\nu_i/kT}} \quad . \tag{A1.27f}$$

The zero-point vibrational energy (ZPVE) inherent to all atoms or molecules at zero Kelvin is expressed by $\sum_i^{modes} = 1/2h\nu$ over all vibrational modes and is included in the first part of H_{vib} (see Eq. A1.27f).

Entropy expressions

$$S_{tot} = S_{elec} + S_{trans} + S_{rot} + S_{vib} \tag{A1.28a}$$

$$S_{elec} = R\ln\left(\sum_i g_i e^{-\epsilon_i/kT}\right) + R\frac{\sum_i g_i(\epsilon_i/kT)e^{-\epsilon_i/kT}}{\sum_i g_i e^{-\epsilon_i/kT}} \tag{A1.28b}$$

$$S_{trans} = \frac{5}{2} + Rln\left(\frac{V}{N_A}\left(\frac{2\pi MkT}{h^2}\right)^{\frac{3}{2}}\right) \tag{A1.28c}$$

$$S_{rot}^{non-linear} = R\left[\frac{3}{2} + \ln\left(\frac{\sqrt{\pi}}{\sigma}\left(\frac{8\pi^2 kT}{h^2}\right)^{\frac{3}{2}}\sqrt{I_A I_B I_C}\right)\right] \tag{A1.28d}$$

$$S_{rot}^{linear} = R\left[1 + \ln\left(\frac{8\pi^2 IkT}{\sigma h^2}\right)\right] \tag{A1.28e}$$

$$S_{vib} = R\sum_i^{modes}\left[\left(\frac{h\nu_i}{kT}\frac{e^{-h\nu_i/kT}}{1-e^{-h\nu_i/kT}}\right) - \ln\left(1-e^{-h\nu_i/kT}\right)\right] \tag{A1.28f}$$

A1.3. Conductor-like screening model for real solvents (COSMO-RS)

COSMO-RS [189] builds upon the implicit conductor-like screening model (COSMO), which approximates and simplifies the dielectric medium as a perfect conductor with $\epsilon = \infty$. The molecule is embedded in a cavity surrounded by the dielectric continuum, and a series of surface segments describe the cavity. The electrostatic potential at the contact surface between the conductor and the solute is completely

screened by the conductor polarization charge density, leading to:[242]

$$\mathbf{A}\vec{q} + \left(\frac{\epsilon - 1}{\epsilon + 0.5}\right)\mathbf{B}\vec{Q} = 0 \quad , \tag{A1.29}$$

where $\vec{q}$ contains the surface segment charges, and $\vec{Q}$ contains the charge distribution of the molecule in the cavity. The interaction of the surface charge segments with the charge distribution of the molecule is described by matrix B. The segment-segment interactions are summarized in matrix A, and the ϵ dependent scaling function multiplied to B enables the approximation for the finite dielectric. The surface screening charges computed from $\vec{q} = -A^{-1}B\vec{Q}$ can be used in a SCF procedure, which yields modified solute charge densities and updated surface charges.

In COSMO-RS theory, the condensed phase is approximated by non-compressible fluids. Macroscopic properties of a liquid can be described by statistical thermodynamics, which for COSMO-RS is condensed to the thermodynamic average of all possible configurations of independently interacting pairwise surface segments. This avoids the time-consuming task of sampling the phase space and finding all the associated interaction geometries. From a preceding COSMO calculation, a histogram of the surface polarization charge density called σ-profile is calculated, from which the σ-potential can be determined by integrating over all available partners σ' in solvent S:

$$\mu_S(\sigma) = -\frac{kT}{a_{eff}} ln \int \underbrace{P_S(\sigma)}_{\sigma-profile} \; exp\left\{ -\frac{a_{eff}}{kT}\left(e_{int}(\sigma, \sigma') - \mu_S(\sigma')\right)\right\} d\sigma \quad . \tag{A1.30}$$

Here, $\mu_S(\sigma)$ is a function characterized by the solvent-specific σ-profile and describes how strongly the solvent S is attracted to a surface segment with polarity σ. The parameter a_{eff} determines the size of the surface fragments and the interaction operator e_{int} is associated with the energetic cost of forming a contact between σ and σ' that involves both electrostatics and hydrogen-bonding, while $\mu_S(\sigma')$ represents the cost of releasing a surface segment σ' from another contact. The chemical potential of a molecule X solvated in S is calculated from the segment chemical potentials over the surface of solute X and a combinatorial correction term,

$$\mu_S^X = \int P^X(\sigma)\mu_S(\sigma)d\sigma + \mu_{comb,S}^X \quad . \tag{A1.31}$$

The solvation free energy is calculated from the difference of the solute potential in the solvent S and the gas phase, according to:

$$\mu_{gas}^X = E_{vac}^X - E_{COSMO}^X + \mu_{vdW} - \mu_{ring} + \mu_{ref} \tag{A1.32a}$$

$$\delta G_{solv}^T(X) = \mu_S^X(T) - \mu_{gas}^X(T) \quad . \tag{A1.32b}$$

To incorporate the effects of COSMO-RS in geometry optimizations or properties depending on the density or wave function response of the solvent, the Direct-COSMO-RS (DCOSMO-RS) was developed. Converting the COSMO-RS free solvation energy expressions into a contribution in the Fock

matrix provides a solvent-specific response that, most importantly, also takes hydrogen bonding into account.[243]

A1.4. Structure ensembles and molecular dynamics

Exploration of the relevant low-energy chemical space is usually performed by molecular dynamics (MD) or Monte Carlo simulations.[244] MD simulations, starting from an initial set of coordinates and velocities, reproduce the physical evolution of configurations by a series of finite time steps (on the order of femtoseconds). In typical Born-Oppenheimer MD simulations, the trajectories are calculated numerically using Newton's second equation

$$\underbrace{-\nabla_I E_{tot}}_{\vec{F}_I} = M_I \frac{d^2 \vec{R}_I}{dt^2} \tag{A1.33}$$

where M_I is the nuclear mass, R_I the spatial position and $\vec{F}_I$ defines the force acting on nucleus I. In the general procedure, at each step, the forces acting on the atoms are calculated from the molecular energy calculations and combined with the positions and velocities of the atoms to produce new propagated positions and velocities. The positions of the nuclei are updated, and the cycle is repeated. Several algorithms, such as Verlet[245] or Leap-Frog[246], solve the equations of motion numerically. However, since the time steps are very small and the simulation length is limited by the computational cost of the molecular mechanics or electronic structure method, only a very local region of the PES can be explored.

In so-called meta-dynamics (MTD), the minima on the PES are filled by history-dependent potentials, which facilitate overcoming large reaction barriers (see Fig. A1.1A). The biasing potential is determined from a set of collective variables (CV) which describe the overall system. For fast chemical space exploration, the Cartesian root-mean-square deviation (RMSD) has proven useful as a CV, as it does not require manual adjustment and is a good descriptor of chemically similar structures. The biasing potential is expressed by

$$E_{bias}^{RMSD} = \sum_{i=1}^{n} k_i exp(-\alpha \Delta_i^2) \tag{A1.34}$$

where n is the number of reference structures, the sign and the magnitude of k_i determines the pushing ($k_i > 0$) or pulling ($k_i < 0$) strength associated with the reference structure i and α defines the width of the biasing potential. The collective variable Δ is computed between the structures from the simulation and a reference according to

$$\Delta_i = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (r_j - r_j^{ref,i})^2} \tag{A1.35}$$

with r_j being a component of the three-dimensional atomic Cartesian vector which is compared to

the related entry $r_j^{ref,i}$ of the reference geometry i and N represents the number of atoms within the molecule. Within this work, chemical space exploration is performed with the *crest*[68] program, which contains a conformational search workflow that uses MTDs in combination with the force-field and semiempirical quantum mechanical methods implemented in *xtb*.

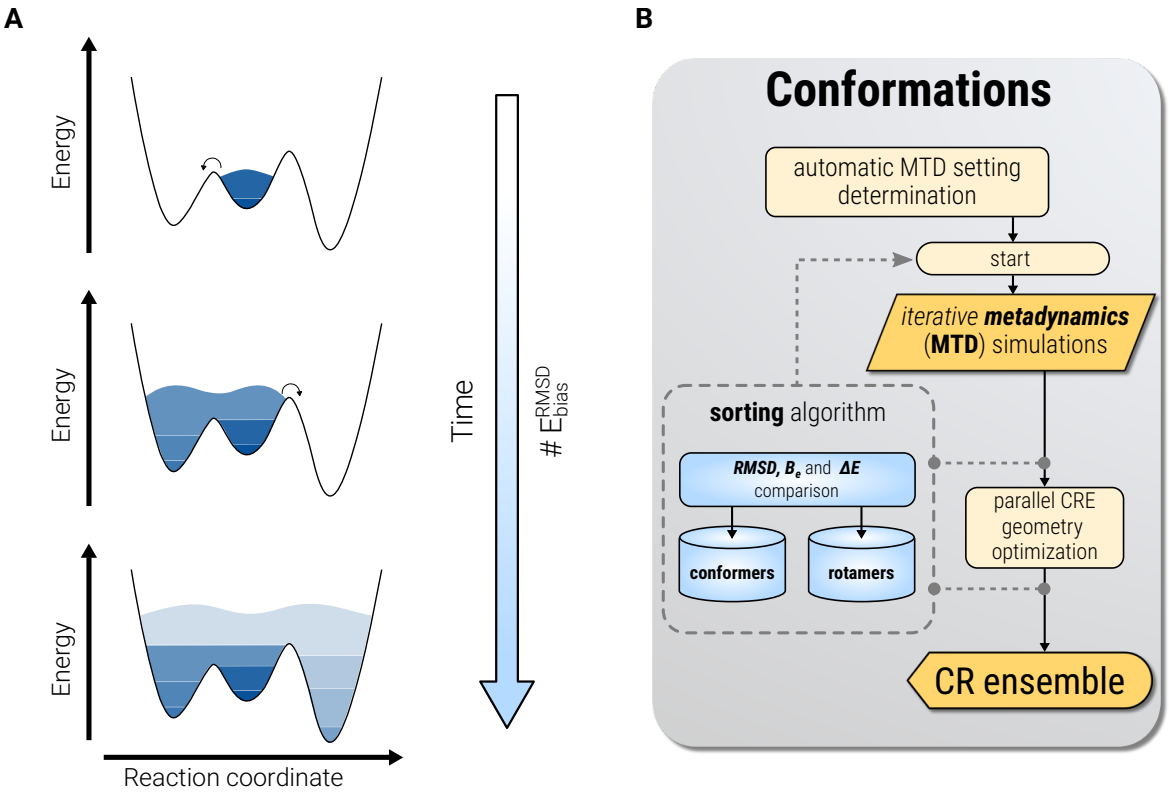
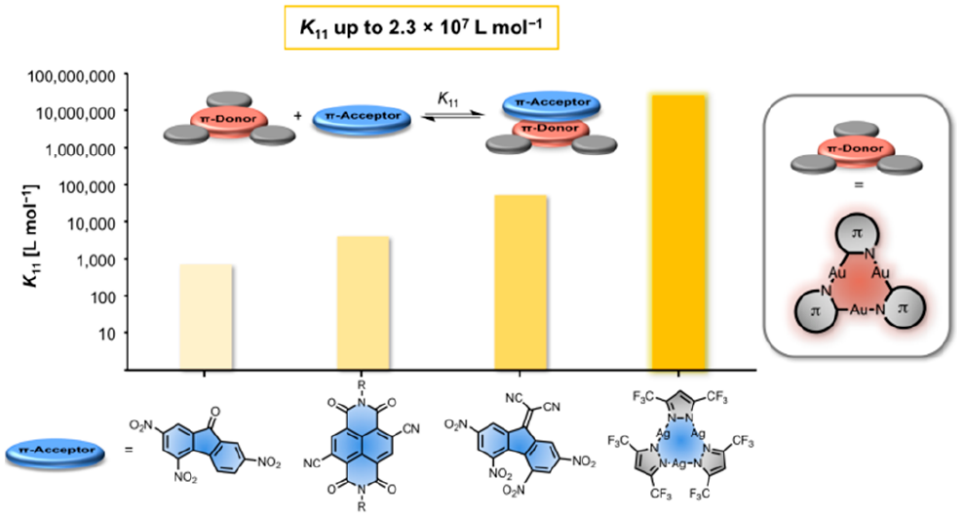


Figure A1.1.: (A) Schematic representation of a one-dimensional PES filled by cumulative biasing potentials. Over time, even larger barrier heights can be overcome, and the minima of the PES can be accessed. (B) Simplified representation of the CREST algorithm for generating CREs. Figure adapted from Ref. 68.

The workflow is based on several MTD simulations with different bias parameters and normal molecular dynamics simulations, which recover minima separated by lower barriers. Snapshots from the resulting trajectories are fully geometry optimized. The resulting structures are sorted based on their RMSD, rotation constant B_e, spatial components, and energy difference ΔE. Structures are sorted into conformers, rotamers, and (discarded) identical structures. The workflow is presented in Figure A1.1B, and further information can be found in Ref. 68 and Part A12.2.7.

A2. Raising the Bar in Aromatic Donor–Acceptor Interactions with Cyclic Trinuclear Gold(I) Complexes as Strong π-Donors

Abstract Aromatic donor-acceptor interactions are of high importance in supramolecular chemistry, materials science and biology. Compared to other noncovalent interactions, such as hydrogen bonding, the binding is often weak. Here we show that strong donor-acceptor interactions between planar aromatics with binding free energies down to -10.1 kcal mol^{-1} and association constants of up to 2.34×10^{7} L mol^{-1} for 1:1 complexes can be realized using cyclic trinuclear complexes of gold(I) with pyridinate, imidazolate, or carbeniate ligands. Data were obtained through NMR and UV/vis absorption spectroscopic studies and supported by quantum chemical calculations for a variety of acceptors. By using a specifically designed bridged naphthalene diimide-based acceptor with only one binding site, we furthermore show that a 1:2 (donor:acceptor) binding model is best suited to quantify the donor and acceptor/complex equilibrium. Scanning electron microscopy on selected donor-acceptor pairs shows crystalline supramolecular assemblies. We anticipate this study to be relevant for the future design of supramolecular systems and chemical sensors and the determination of binding energies between planar donors and acceptors.



A2.1. Introduction

Noncovalent interactions between planar arenes[220] or between π-systems and cations or anions play an important role in many areas of chemistry, materials science and molecular biology.[221,222,225,247–251] These interactions are determined by a variety of non-covalent forces, which include dispersion interactions, electrostatic interactions, attractive and repulsive orbital interactions and solvophobic effects.[250,252,253] Aromatic donor-acceptor (D-A) interactions with face-centered stacking orientation are often associated with a partial charge transfer (CT) from donor to acceptor.[254,255] D-A interactions have been exploited in the design of supramolecular assemblies and interlocked molecules.[255] Cyclic trinuclear complexes (CTCs) of gold(I) with heterocyclic ligands such as pyridinate, imidazolate, or carbeniate (**1**, **2**, and **3** in Fig. A2.1, respectively) possess an electron-rich π-system, resulting in a high π-basicity.[256–258] The electron density is mainly located above and below the trinuclear gold core, as indicated by the red coloring in Fig. A2.1.[259] These Au(I) CTCs are able to form complexes with π-acceptors (or π-acids),[257,260] which has been shown through X-ray crystallography and changes in solvoluminescence[261–269] as well as NMR experiments.[223,270]

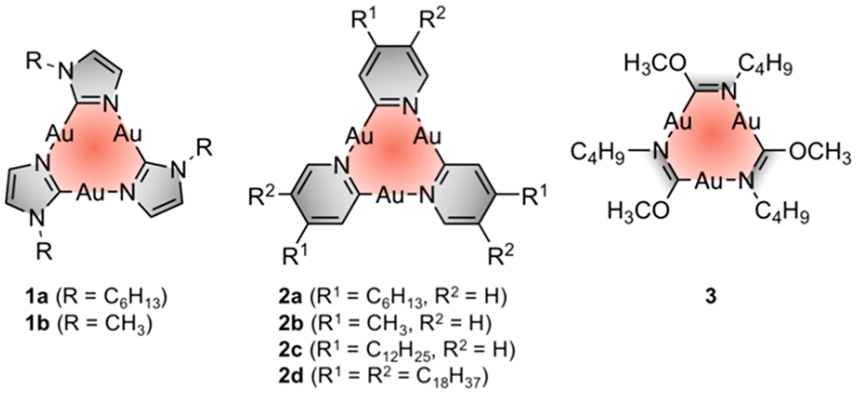


Figure A2.1.: Cyclic trinuclear complexes (CTCs) of gold(I) with imidazolate (**1**), pyridinate (**2**), and carbeniate (**3**) ligands. The area of high electron density above and below the trinuclear gold core is indicated by the red color.

We recently reported the first binding studies between pyridinate Au(I) CTCs of type **2** (Fig. A2.1) and pyrazolate CTCs **4-6** (see Chart A2.2 below) as acceptors employing NMR titrations and high-level quantum chemical calculations.[223] The experiments revealed remarkably high binding constants for the formation of 1:1 complexes of up to 2.83 $\times 10^5$ L mol^{-1} , corresponding to free binding energies of down to -7.5 kcal mol^{-1}. This places them above known D-A-complexes,[255,271] where the majority of the interaction stems from the π-systems.

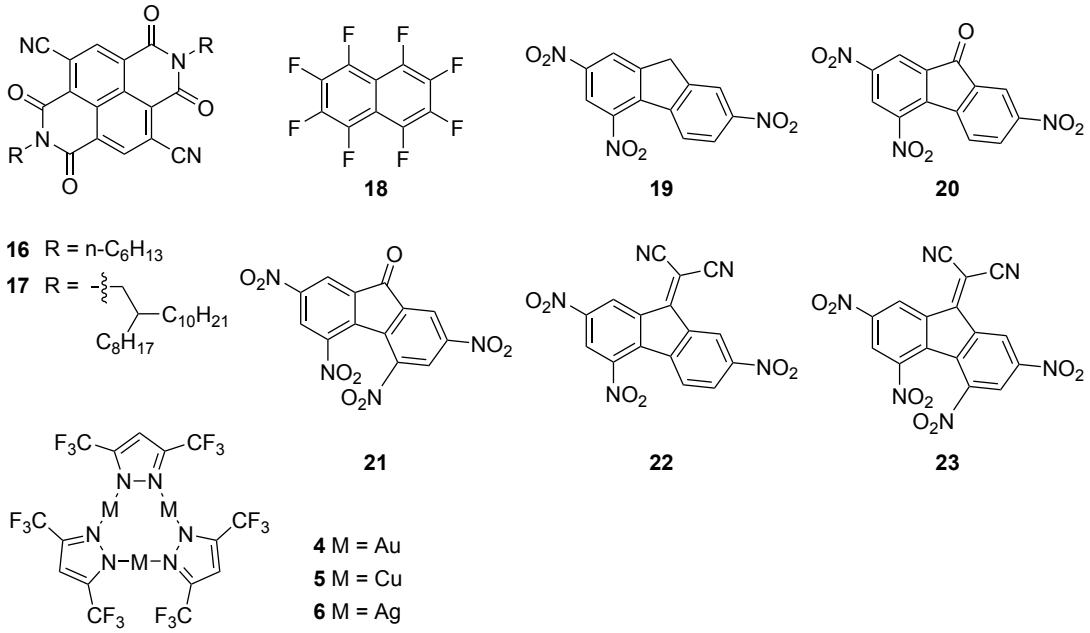


Figure A2.2.: Acceptors investigated in this study.

We furthermore demonstrated the suitability of such gold(I) CTCs in the design of chemical graphene-based sensors for electron-poor aromatics through coadsorption experiments on highly oriented pyrolytic graphite using scanning tunneling microscopy.[223] Due to the planarity of the gold(I) CTCs and most π-acceptors, both the donor and the acceptor possess two equal binding sites on each face of the molecule. For this reason, a quantitative analysis of the π-π interaction is difficult to achieve, since the formation of not only 1:1 but also 2:1 and 1:2 complexes and higher order structures is possible, as shown in SchemeA2.3a. Analyses using Job's method of continuous variation had indicated a 1:1 stoichiometry. However, this method has been shown to provide unreliable results.[272]

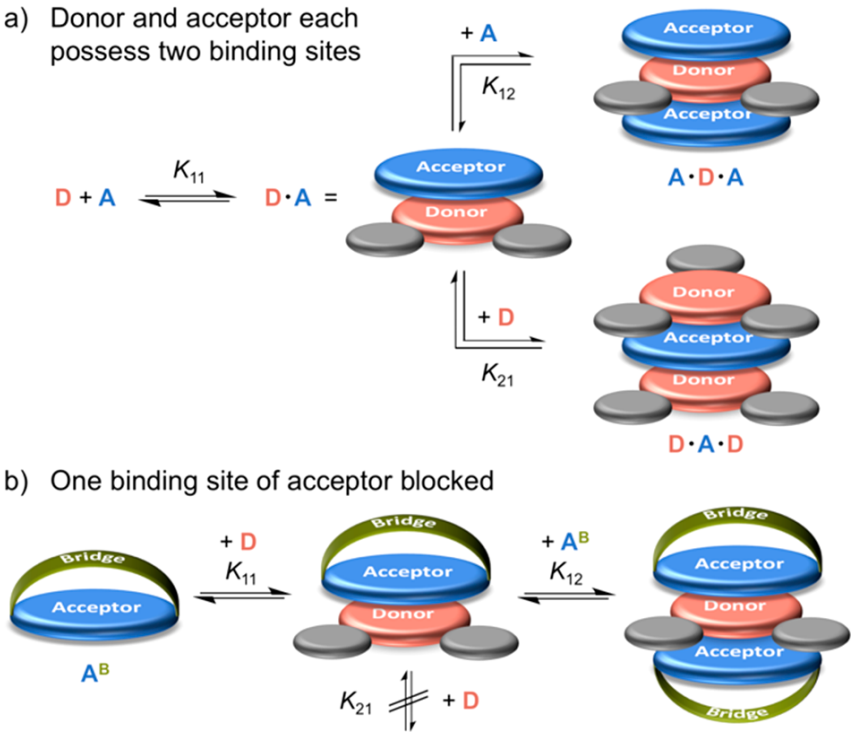


Figure A2.3.: Equilibria of Complex Formation between a Gold(I) CTC as Donor **D** and an Acceptor **A**

In the NMR titration experiments in our previous study we had considered the formation of 1:1 (**D**·**A**) and 1:2 (**A**·**D**·**A**) complexes, while 2:1 (**D**·**A**·**D**) complexes had been neglected based on our experimental design.[223] We herein set out to probe this binding model by using a π-acceptor, which possesses only one binding site, hence the formation of higher order structures is inhibited. This can be achieved by covalently attaching a bridge to one side of the π-system, resulting in bridged acceptor $\mathbf{A}^B$ as shown in Scheme A2.3b. Here, only the formation of complexes $\mathbf{D}{\cdot}\mathbf{A}^B$ and $\mathbf{A}^B{\cdot}\mathbf{D}{\cdot}\mathbf{A}^B$ is possible, but no complex $\mathbf{D}{\cdot}\mathbf{A}^B{\cdot}\mathbf{D}$ or higher order structures can be formed.

To obtain a bridged acceptor with sufficient binding strength, the naphthalene diimide (NDI) core is a good π-acceptor framework, which has been extensively used in supramolecular chemistry.[273–276] It allows attachment of a bridge to one side, and its acceptor strength can be increased through electron-withdrawing substituents on the naphthalene core. Our design of acceptor $\mathbf{A}^B$ was inspired by previous reports of bridged NDIs.[277,278] Here, we used a different covalent bridge and synthesized bridged NDI **7**, carrying two sulfone groups on the NDI core to increase acceptor strength (Figure A2.4).[278,279] The bridge was chosen such that it would impart the least possible sterical hindrance for binding of the donor to the unshielded binding site of the NDI core and that its length would be sufficient to ensure planarity of the latter.

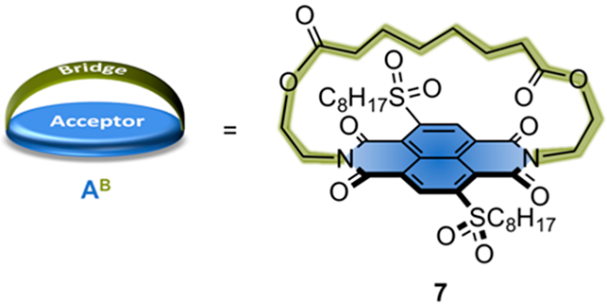


Figure A2.4.: Bridged naphthalene diimide acceptor **7** synthesized and used in this study.

As gold(I) CTC, we chose imidazolate complex **1a** (Fig. A2.1), since it possesses a higher stability toward light and moisture and is synthetically easier to access compared to pyridinate CTCs **2**, which we had previously employed.[223] The hexyl groups at the imidazolate ligands served to ensure good solubility for NMR titrations. After testing the binding model using the bridged acceptor **7** as well as the nonbridged NDI **8** (Scheme A2.6) for comparison, we performed binding studies using NMR titrations with a range of acceptors (Chart A2.2). For selected acceptors, we also report on binding studies with carbeniate Au(I) CTC **3**. We furthermore compared the results of binding studies using NMR titrations with those using UV/vis absorption spectroscopy employing all three donors, **1a**, **2**, and **3**. Quantum chemical calculations provided interaction free energies for all donor-acceptor combinations and allowed assessing the strength of the noncovalent binding. Scanning electron microscopy on donor **1a** with selected acceptors showed crystalline supramolecular assemblies.

A2.2. Results and discussion

Synthesis. The synthesis of imidazolate gold(I) CTC **1a** proceeded as shown in Scheme A2.5. Successive reaction of 1-hexylimidazole **9**[280] with *n*-BuLi and chloro(dimethylsulfide)-gold(I) afforded **1a** in 48% yield.

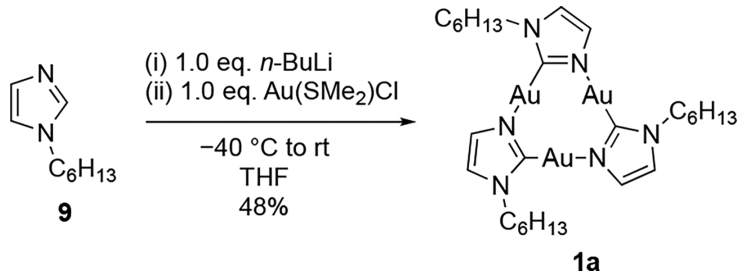


Figure A2.5.: Synthesis of Imidazolate Gold(I) CTC **1a**

The synthesis of bridged, sulfone-functionalized NDI acceptor **7** and nonbridged NDI **8** for comparison (Scheme A2.6) commenced from **10**, which was synthesized according to the literature.[281,282] Nucleophilic replacement of the bromides by octyl sulfide afforded **11**,[283] from which base-mediated ester hydrolysis and dehydration led to dianhydride **12**.[278] En route to bridged NDI **7**, reaction of **12** with

aminoethanol yielded **13**.[284] Double esterification with octanedioyl dichloride afforded cyclophane **14**, which was oxidized in the last step using *meta*-chloroperoxybenzoic acid to furnish bridged acceptor **7**. For the synthesis of nonbridged NDI **8** reaction of **12** with 2-octyldodecan-1-amine led to NDI **15**. Final oxidation with *meta*-chloroperoxybenzoic acid afforded **8**.

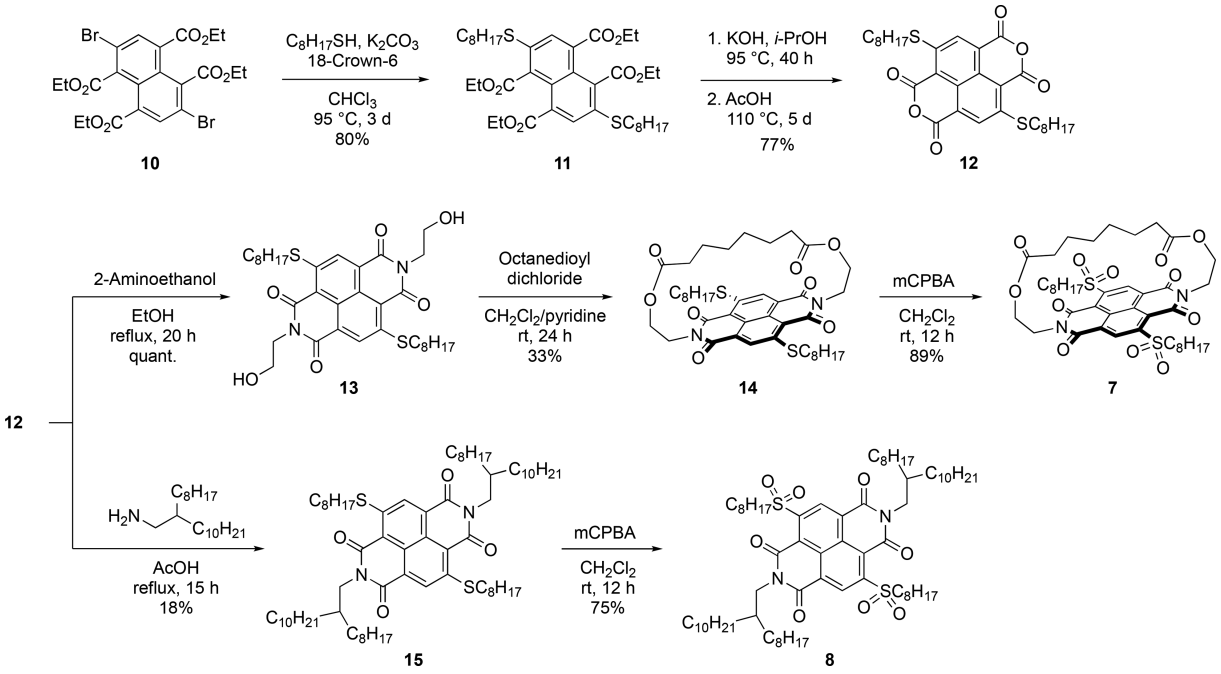


Figure A2.6.: Synthesis of Bridged NDI-Acceptor **7** and Non-Bridged NDI **8**

Probing the Binding Model with Bridged Acceptor 7 and Donor 1. Acceptors **7** and **8** were used for NMR spectroscopic binding studies with imidazolate Au(I) CTC **1a** as donor. The titrations were run such that the acceptor was added to the donor. We used a 1:2 (**D**/**A**) binding model,[285] which considered the formation of 1:1 and 1:2 complexes, but no other or higher order stoichiometry. The latter could be formed from donor **1a** and nonbridged **8** (**A**), but for bridged **7** ($\mathbf{A}^B$), only the 1:1 and 1:2 stoichiometry are possible. Hence a comparison of the results achieved with both acceptors, **7** ($\mathbf{A}^B$) and **8** (**A**), should allow probing the 1:2 binding model. An online tool provided by Thordarson and co-workers[272,286] served to fit the results of the titration experiments using nonlinear regression.

All measurements revealed a fast exchange between donor and acceptor in relation to the NMR time scale, which resulted in one set of signals for the donor and for the acceptor (for NMR spectra, see the Supporting Information). The titration experiments yielded the binding constants and free energies of binding listed in Table A2.1. The binding constants K_{11} and K_{12} and the free binding energies ΔG_{11} and ΔG_{12} lie in a similar range for both bridged acceptor **7** and nonbridged acceptor **8** (first and third entries in Table A2.1).

acceptor	order of addition	K_{11}^{b}	ΔG_{11}^{b}	K_{12}^{c}	ΔG_{12}^{c}
7 ($\mathbf{A}^{(B)}$)^{d}	$\mathbf{A}^{(B)}$ to **D**	60.3	-2.44	2.73	-0.60
7 ($\mathbf{A}^{(B)}$)^{e}	**D** to $\mathbf{A}^{(B)}$	48.7	-2.32		
8 (**A**)^{d}	**A** to **D**	45.3	-2.27	7.43	-1.20

Table A2.1.: Experimental Binding Constants K and Free Binding Energies ΔG of Imidazolate CTC **1a** (as Donor **D**) with NDIs **7** and **8** as (Acceptors $A^{(B)}$)^{a}. ^{a} Binding constant K (L mol^{-1}) from ^{1}H NMR spectroscopic titrations in $CDCl_3$ at 300 K, average of two experiments; Δ G (kcal mol^{-1}). ^{b} For complex **D**·**A** $^{(B)}$. ^{c} For complex **A** $^{(B)}$ · **D**·$\mathbf{A}^{(B)}$. ^{d} 1:2 (**D**/**A**) binding model. ^{e} 1:1 binding model.

To probe whether the order of addition during the titration had an influence and to prove that the bridge efficiently blocked one binding site of NDI **7**, we performed experiments, where donor **1a** (**D**) was added to bridged acceptor **7** ($\mathbf{A}^{B}$); hence, **D** was present in excess. This, however, did not significantly alter the values K_{11} and ΔG_{11} (middle entry in Tab. A2.1). These results show that the 1:2 (**D**/**A**) binding model, where higher order complexes in the case of planar acceptors are neglected, is appropriate to be used for evaluating the binding between a planar donor and acceptor, as depicted in Scheme A2.3a.

Quantum mechanical calculations provided molecular geometries and binding free energies for the 1:1 complexes between **1b** and **7**∗ or **8**∗ (Tab. A2.2). In donor **1b**, the alkyl side chains were replaced by methyl groups for reasons of better computational efficiency (Fig. A2.1).

acceptor	ΔE^{b}	$\Delta E^{disp,c}$	ΔG_{RRHO}^{d}	$\Delta\delta G_{solv}^{e}$	$\Delta G_{calc,11}$
7* ($\mathbf{A}^{(B)}$)	0.2	-27.9	18.4	7.3	-2.0
8* (**A**)	-0.8	-26.8	18.9	6.3	-2.5

Table A2.2.: Calculated Binding Free Energies ΔG_{calc11} of Imidazolate CTC **1b** (as Donor **D**) with NDI-Acceptors **7**∗ and **8**∗^{a}. ^{a} Δ G_{calc11} (kcal mol^{-1}) obtained as the sum of electronic interaction energies ΔE, dispersion correction ΔE_{disp}, thermostatistical contributions ΔG_{RRHO} (298 K) and solvation free energies $\Delta\delta G_{solv}$. ^{b} PBE0/QZ. ^{c} D3(BJ)^{ATM}. ^{d} B97–3c. ^{e} In $CHCl_3$.

In the NDI-acceptors, the octyl groups were replaced by methyl groups, and the branched substituents on the nitrogen atoms in **8** by isobutyl groups, denoted with an asterisk (**7**∗ and **8**∗). The DFT-based composite scheme PBEh-3c[154] together with the COSMO solvent model was used to fully optimize the geometries. On these structures, the binding free energies were computed as the sum of three contributions: PBE0[122,287]-D3[142](BJ),[144,145] as a dispersion-corrected hybrid functional using the large def2-QZVP[288] basis set yielded refined electronic gas phase binding energies ΔE. Thermostatistical contributions to free energies ΔG_{RRHO} (298 K) were evaluated at the B97-3c[155] level, while COSMO-RS[289] provided solvation free energies ($\Delta\delta G_{solv}$(X)) in a black-box fashion. Further details on the computational protocol is provided in the Experimental Section and in the Supporting Information. The calculated values stand in very good agreement with the experimental ones. Unbridged system **8**∗ shows only slightly stronger binding to the donor than bridged **7**∗. The calculated structure of the

1:1 complex between **1b** and bridged acceptor **7**∗ is shown in Fig. A2.7a. The distance between the trinuclear gold center in donor **1b** and the NDI core in **7**∗ amounts to ca. 3.43 Å, indicating a strong interaction, if a typical range of 3.3-3.8 Å for $\pi-\pi$ interactions is assumed.[290] This distance also lies close to values reported for solid-state structures of D-A adducts between CTCs and organic π-acids.[263–265] In order to accommodate the bulky SO_2-Alkyl substituents on the NDI core, one imidazolate ligand is bent away from the main plane of the trinuclear gold core in **1b**. This is also the case for the complex with nonbridged NDI-acceptor **8**∗ (Fig. A2.7b).

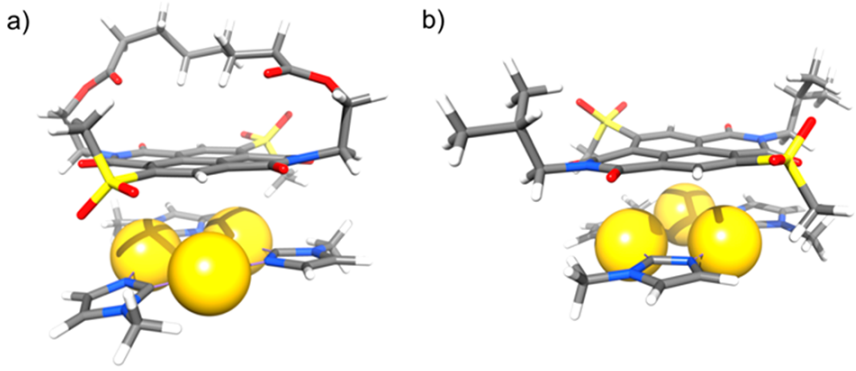


Figure A2.7.: Calculated structures (PBEh-3c/COSMO) of the 1:1 complexes between imidazolate gold(I) CTC **1b** and (a) bridged NDI-acceptor **7**∗ and (b) nonbridged NDI-acceptor **8**∗.

Acceptor Screening Using Imidazolate CTC 1 as Donor. Having confirmed the validity of using a 1:2 (**D**/**A**) binding model to investigate the binding between a planar acceptor and donor, we set out to screen the binding energies between imidazolate Au(I) CTC **1a** and a range of different π-acceptors (Chart A2.2). The syntheses of NDI acceptor **16**,[291] of trinitrofluorene **19**,[292] of fluorenone derivatives **20**,[293] **21**,[294] **22**,[293] and **23**,[293] and of pyrazolate CTCs **4**,[266] **5**,[295,296] and **6**[296] were performed according to literature procedures. The π-acid-base complexes of octafluoronaphthalene **18**[260,264,265] or nitrofluorenones **20** and **21**[262] with imidazolate or carbeniate Au(I) CTCs have been characterized before; however, no binding constants or free energies of binding were determined apart from one singular example.[265] Using strong π-acceptors **4-6**, we previously reported binding constants and free binding energies to π-basic pyridinate Au(I) CTCs **2**.[223] During NMR titrations the acceptor was added to the donor, unless the low solubility of one compound required the opposite order (indicated by a footnote in Tab. A2.3). The NMR solvent was chosen between $CDCl_3$ and C_6D_6 to ensure best possible solubility of both donor and acceptor.

A comparison between these solvents for titrations of donor **3** with acceptors **4-6** had provided similar binding constants. Evaluation of the data using a 1:2 (**D**/**A**) binding model was performed as described above. Table A2.3 lists the binding constants K_{11} and K_{12} and the free binding energies ΔG_{11} and ΔG_{12}. As example, selected ^{1}H NMR spectra from the titration of **1a** with **5** are shown in Fig.A2.8. For the stronger acceptors shown in Tab. A2.3, the binding free energies lie in a range typical for large supramolecular complexes, where London dispersion interactions play a major role.[183,220,225] For example in the case of the benchmark set of realistic host-guest complexes S30L,[76] this value amounts

acceptor	K_{11}^{b}	ΔG_{11}^{b}	K_{12}^{c}	ΔG_{12}^{c}	solvent
16^{d,e}	3.67 x 10^3	-4.89	41.7	-2.22	$CDCl_3$
17^{d,e}	2.34 x 10^3	-4.62	16.3	-1.66	$CDCl_3$
18^{e}	7.3	-1.19			C_6D_6
19	525	-3.73	38.2	-2.17	C_6D_6
20^{e}	673	-3.88	22.0	-1.84	C_6D_6
21^{d}	789	-3.98	17.3	-1.24	C_6D_6
22^{d,e}	5.84 x 10^3	-5.17	242	-3.27	C_6D_6
23^{d}	4.98 x 10^4	-6.45	736	-3.94	C_6D_6
4^{f}	9.08 x 10^3	-5.43	576	-3.79	C_6D_6
5^{e}	3.75 x 10^4	-6.28	3.40 x 10^3	-4.85	C_6D_6
6^{f}	3.20 x 10^5	-7.56	202	-3.17	C_6D_6

Table A2.3.: Experimental Binding Constants K and Free Binding Energies Δ G of Imidazolate CTC **1a** (as Donor **D**) with Different Acceptors **A**^{a}. ^{a} K values (L mol^{-1}) from ^{1} H NMR spectroscopic titrations in $CDCl_3$ at 300 K; ΔG (kcal mol^{-1}). ^{b} For complex **D**·**A**. ^{c} For complex **A**·**D**·**A**. ^{d} The donor **D** was added to the acceptor **A**. ^{e} Average of two experiments. ^{f} Average of three experiments.

to -6.5 kcal mol^{-1} (mean of all experimental free energies). Compared to other donor-acceptor systems, however, where only the interaction of the π-systems is relevant and no other noncovalent forces assist in binding, many of the values shown in Tab. A2.3 are significantly higher.[255,271,297]

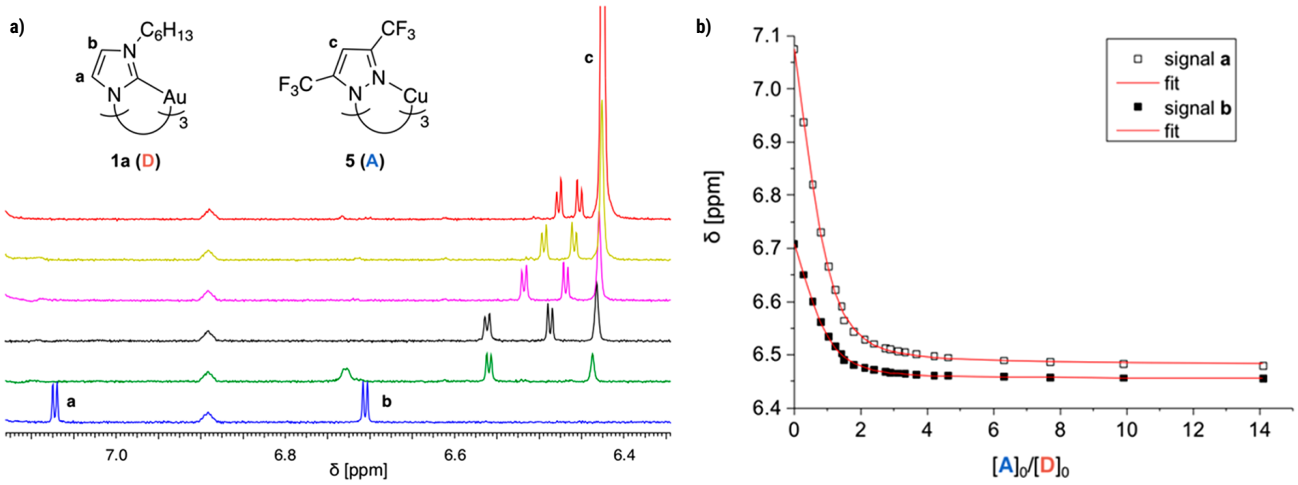


Figure A2.8.: ^{1}H NMR titration of **1a** with **5** (C_6D_6, 300 K). (a) Selected spectra (added equivalents of **5** from bottom to top: 0, 0.8, 1.5, 2.4, 4.2, 14). (b) Plot of data points and curve fitted using nonlinear regression.

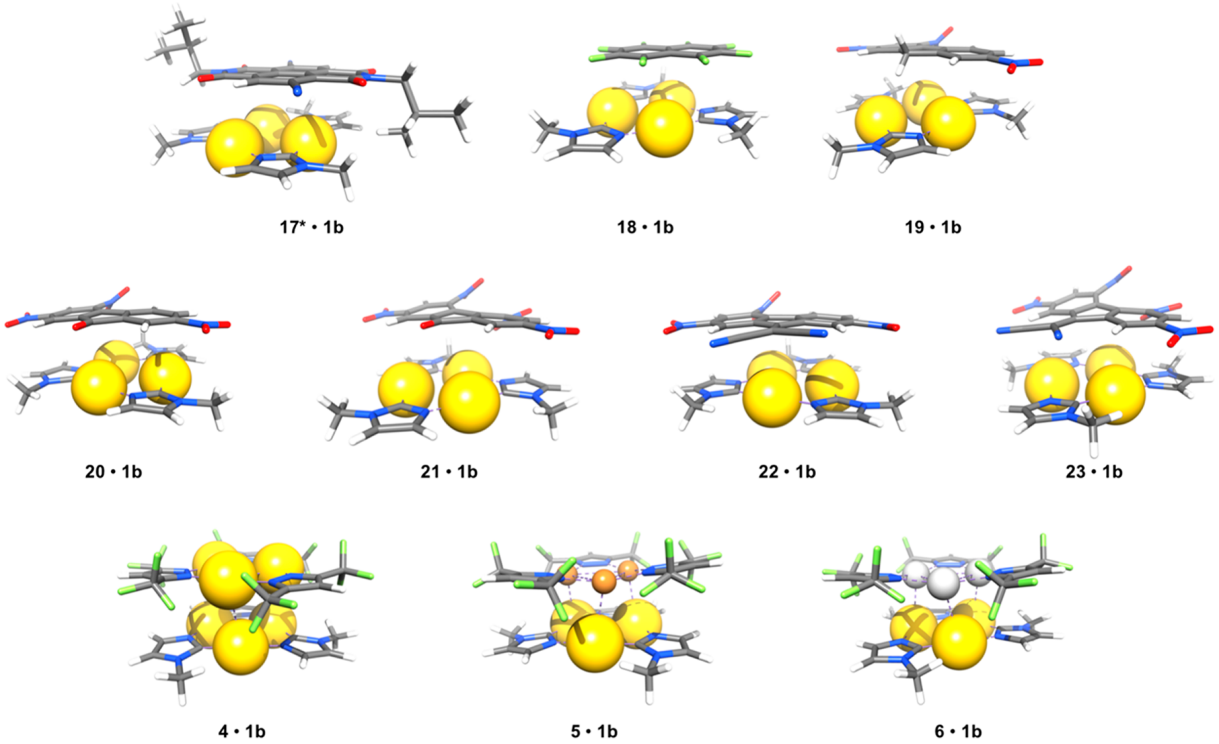


Figure A2.9.: Calculated structures (PBEh-3c/COSMO) of the 1:1 complexes between imidazolate gold(I) CTC **1b** and the acceptors listed in Chart A2.2

The binding constants and free binding energies measured for cyano-substituted NDI acceptors **16** and **17** show that the acceptor strength of the cyano-substituted NDI core is higher than that of the sulfone-substituted core in **7** and **8**. With 3.67×10^3 L mol^{-1} and 2.34×10^3 L mol^{-1}, the K_{11} values between donor **1a** and cyano-substituted NDIs **16** and **17** lie almost 2 orders of magnitude above those for sulfone-NDIs **7** and **8**. This might partly be due to steric effects, since the octyl-sulfone groups in **7** and **8** are sterically more demanding than the cyano groups in **16** and **17**. In the calculated structure of the 1:1 complex between donor **1b** and cyano-substituted NDI-acceptor **17*** (see Fig. A2.9 below), the distance between the trinuclear gold center in **1b** and the NDI core in **17*** amounts to ca. 3.38 Å. This value is smaller than those in the calculated structures of the complexes with sulfone-substituted bridged NDI **7**$*$ (3.43 Å, see above) or nonbridged NDI **8**$*$ (3.53 Å). Furthermore, all three imidazolate ligands in **17**$*$·**1b** roughly remain in one plane with the trinuclear gold core. A comparison between **16** and **17**, which only differ in the length and branching of the alkyl chains but delivered very similar binding constants, shows that the nature of the alkyl chain on the NDI core is of minor importance for its binding ability to **1a**. This is important in view of the comparison between bridged **7** and nonbridged **8**, presented above in Tab. A2.1, which served to probe the binding model employed.

With a binding free energy ΔG_{11} of -1.19 kcal mol^{-1}, octafluoronaphthalene (**18**) was the weakest π-acceptor investigated herein. For this reason, a binding constant K_{12} for an **A**·**D**·**A** complex could not be determined. The results for fluorene **19** and fluorenone derivatives **20-23** demonstrate the increase

in acceptor strength with growing number of nitro substituents and/or replacement of the carbonyl by a cyanomethylene group. The binding constants K_{11} increased by almost 2 orders of magnitude in going from **19** to **23**. With binding free energies ΔG_{11} of -5.17 and -6.45 kcal mol^{-1}, **22** and **23** can be considered strong π-acceptors.

Coinage metal CTCs **4-6** with bis(trifluoromethyl)-pyrazolate ligands are known for their strong π-acidities.[259,266–269] The acceptor character increases in the order Au < Cu < Ag,[266] which is reflected in the values shown in Tab.A2.3. Within this study, Ag(I) CTC **6** was the strongest acceptor investigated, resulting in a binding constant K_{11} of 320 000 L mol^{-1} and a free binding energy ΔG_{11} of -7.56 kcal mol^{-1}. This association constant is significantly higher than the largest value of 26 000 L mol^{-1} reported up to date for a D-A complex between two planar arenes, where no other noncovalent forces assist in binding.[223,271] Compared to previously investigated pyridinate Au(I) CTC **2a**, where we measured binding free energies ΔG_{11} between -6.1 and -7.5 kcal mol^{-1},[223] the values for imidazolate Au(I) CTC **1a** lie in a similar range, but slightly lower for **4**. This observation is in line with calculations, which had predicted a similar π-basicity for **2** and **1**.[266] In all cases, the binding constants K_{12} for the formation of complexes **A**·**D**·**A** are significantly smaller than K_{11}, which is due to a lowering of the electron density of the donor in the **D**·**A** complex. In the statistical case a ratio of 4:1 between K_{11} and K_{12} would be expected;[285] hence, our results indicate an anticooperative situation in all donor-acceptor combinations. In order to assess the effect of temperature on the binding constants, measurements were performed between 10 and 60 °C for acceptors **19-23** and evaluated using the Benesi-Hildebrand method (for data, see the Supporting Information). With increasing temperature, the binding constants are reduced by between 33 and 73%. Quantum chemical calculations, conducted as discussed above, provided association free energies for the 1:1 complexes of **1b** with acceptors **17-23** and **4-6** (Tab. A2.4). In NDI-acceptor **17**, the branched alkyl substituents on the nitrogen atoms were replaced with isobutyl groups for computational efficiency, denoted with an asterisk (**17**∗). The calculated structures of the 1:1 complexes are shown in Fig. A2.9. A reasonable agreement of theoretical and experimental $\Delta G_{calc,11}$ values was observed, and the experimental trend is correctly reproduced. However, it is noteworthy that for the smaller organic acceptors **18-23** the calculations predicted weaker binding than experimentally found. This deviation was strongest for octafluoronaphthalene (**18**) and trinitrofluorene **19** as acceptors, where the calculation yielded an exergonically unfavorable complex formation.

For pyrazolate CTCs **4-6**, the calculations yielded stronger binding free energies than experimentally observed with deviations of up to 4.6 kcal mol^{-1}for the strongest acceptor, **6**. In the calculations, an implicit solvation model was employed, which may be unable to capture all solute-solvent interactions for these planar coinage metal complexes.[298] As we will show later, binding constants determined using UV/vis spectroscopic titrations in different solvents also indicate large solvation effects. The contributions to the binding free energies are specified in Tab. A2.4 and provide additional information on the binding of these noncovalent interactions. In all complexes the largest attractive contribution to the binding free energy stems from London dispersion (ΔE_{disp}), which ranges between 77-93% of the total interaction energy for these systems. This highlights the importance to account for dispersion in noncovalently bound complexes, especially in these planar CTCs.

UV/vis Spectroscopic Binding Studies Using Donors 1-3. Next, we investigated the donor-

acceptor **A**	ΔE^{b}	$\Delta E^{disp,c}$	ΔG^{d}_{RRHO}	$\Delta\delta G_{solv}$	$\Delta G_{calc,11}$
17*	-4.5	-26.4	18.4	8.0^{e}	-4.6
18	-1.2	-15.9	15.1	3.5^{f}	1.6
19	-3.9	-20.3	17.4	7.2^{f}	0.4
20	-4.3	-21.7	17.4	8.4^{f}	-0.2
21	-4.4	-21.7	17.6	8.3^{f}	-0.2
22	-5.0	-25.1	17.9	9.9^{f}	-2.3
23	-6.9	-26.0	18.8	10.8^{f}	-3.3
4	-2.0	-28.3	18.9	5.8^{f}	-5.6
5	-6.3	-29.4	20.2	7.7^{f}	-7.8
6	-8.7	-29.0	18.5	7.1^{f}	-12.2

Table A2.4.: Calculated Binding Free Energies ΔG_{calc11} of Imidazolate CTC **1b** (as Donor **D**) with Different Acceptors **A** a^{a}. ^{a} Δ G_{calc11} (kcal mol^{-1}) obtained as the sum of electronic interaction energies ΔE, dispersion correction ΔE_{disp} , thermostatistical contributions ΔG_{RRHO} (298 K) and solvation free energies $\Delta\delta G_{solv}$. ^{b} PBE0/QZ. ^{c} D3(BJ)^{ATM}. ^{d} B97-3c. ^{e} In $CHCl_3$. ^{f} C_6D_6.

acceptor interactions using UV/vis absorption spectroscopy. This allowed us to elucidate the effect of the solvent and to explain the deviations in binding free energies obtained from experiments and from calculations. All three donors shown in Fig. A2.1, pyridinate Au(I) CTC **2**, imidazolate **1a**, and carbeniate **3**, were used in this final study. The synthesis of **3** was performed according to the literature.[299] For NMR titrations, the choice of solvents had been limited, since high concentrations of donor and acceptor were required, and C_6D_6 was used in most cases. In titrations using UV/vis absorption spectroscopy, however, this was not the case, and we were able to use both benzene for comparison with NMR titrations as well as *n*-hexane as a nonaromatic and nonpolar solvent. It is known that the polarity and nature of the solvent has a significant influence on the binding in D-A systems.[297] In donor-acceptor interactions, a charge-transfer (CT) band can appear in the absorption spectrum at a wavelength bathochromically shifted relative to the absorption of the donor and acceptor.[255] This band can often be assigned to a single-excitation electronic transition from the highest occupied molecular orbital (HOMO) of the donor to the lowest unoccupied molecular orbital (LUMO) of the acceptor. In our case, such a CT-band was visible for interactions of donors **1a**, **2**, and **3** with the strongest acceptors, **4-6**. In *n*-hexane as the solvent, a CT-band appeared for all donors **1a**, **2**, and **3**, while in benzene, this was the case only for stronger donors **1a** and **2** in combination with the strongest acceptor, **6**. Those cases, where a CT band was visible in benzene, allowed us to compare the values with those obtained from NMR titrations in C_6D_6 . Titration experiments were performed by monitoring changes in intensity of the CT-band. Evaluation of the data using a 1:2 (D/A) binding model was performed as described above. The binding constants K_{11} and free binding energies Δ G_{11} are given in Tab. A2.5 together with values from NMR titrations for comparison. Binding constants K_{12} and free binding energies Δ G_{12} can be found in the Supporting Information. Selected absorption spectra of pyridinate CTC **2d**, pyrazolate acceptor **6** and mixtures thereof in benzene and *n*-hexane are shown in Fig. A2.10.

entry	donor **D**	acceptor **A**	K_{11}	ΔG_{11}	solvent	method
1[b,c]	**2a**	**4**	3.01 x10^4	-6.15	C_6D_6	NMR
2	**2c**	**4**	9.73 x10^5	-8.22	*n*-hexane	UV/vis
3[b,c]	**2a**	**5**	1.57 x10^5	-7.13	C_6D_6	NMR
4[d]	**2c/d**	**5**	5.55 x10^6	-9.26	*n*-hexane	UV/vis
5[b,c]	**2a**	**6**	2.83 x10^5	-7.48	C_6D_6	NMR
6	**2d**	**6**	3.21 x10^5	-7.56	C_6H_6	UV/vis
7[d]	**2c/d**	**6**	2.34 x10^7	-10.11	*n*-hexane	UV/vis
8[c]	**1a**	**4**	9.08 x10^3	-5.43	C_6D_6	NMR
9	**1a**	**4**	9.56 x10^5	-6.84	*n*-hexane	UV/vis
10[e]	**1a**	**5**	3.75 x10^4	-6.28	C_6D_6	NMR
11	**1a**	**5**	1.50 x10^5	-8.72	*n*-hexane	UV/vis
12[c]	**1a**	**6**	3.20 x10^5	-7.56	C_6D_6	NMR
13	**1a**	**6**	4.00 x10^5	-7.69	C_6H_6	UV/vis
14	**1a**	**6**	6.42 x10^6	-9.34	*n*-hexane	UV/vis
15	**3**	**4**	1.54 x10^3	-4.37	C_6D_6	NMR
16	**3**	**5**	3.66 x10^3	-4.89	C_6D_6	NMR
17	**3**	**5**	9.13 x10^4	-6.81	*n*-hexane	UV/vis
18	**3**	**6**	2.18 x10^4	-5.96	C_6D_6	NMR

Table A2.5.: Experimental Binding Constants K_{11} and Free Binding Energies ΔG_{11} of Au(I) CTCs **2**, **1a**, and **3** as Donors **D** with Pyrazolate CTCs as Acceptors **A**[a] [a] K_{11} (L mol^{-1}) from ^{1}H NMR and UV/vis absorption spectroscopic titrations at 300 K; ΔG_{11} (kcal mol^{-1}). [b] From Ref 223. [c] Average of three experiments. [d] Average of experiments with **2c** and **2d**. [e] Average of two experiments.

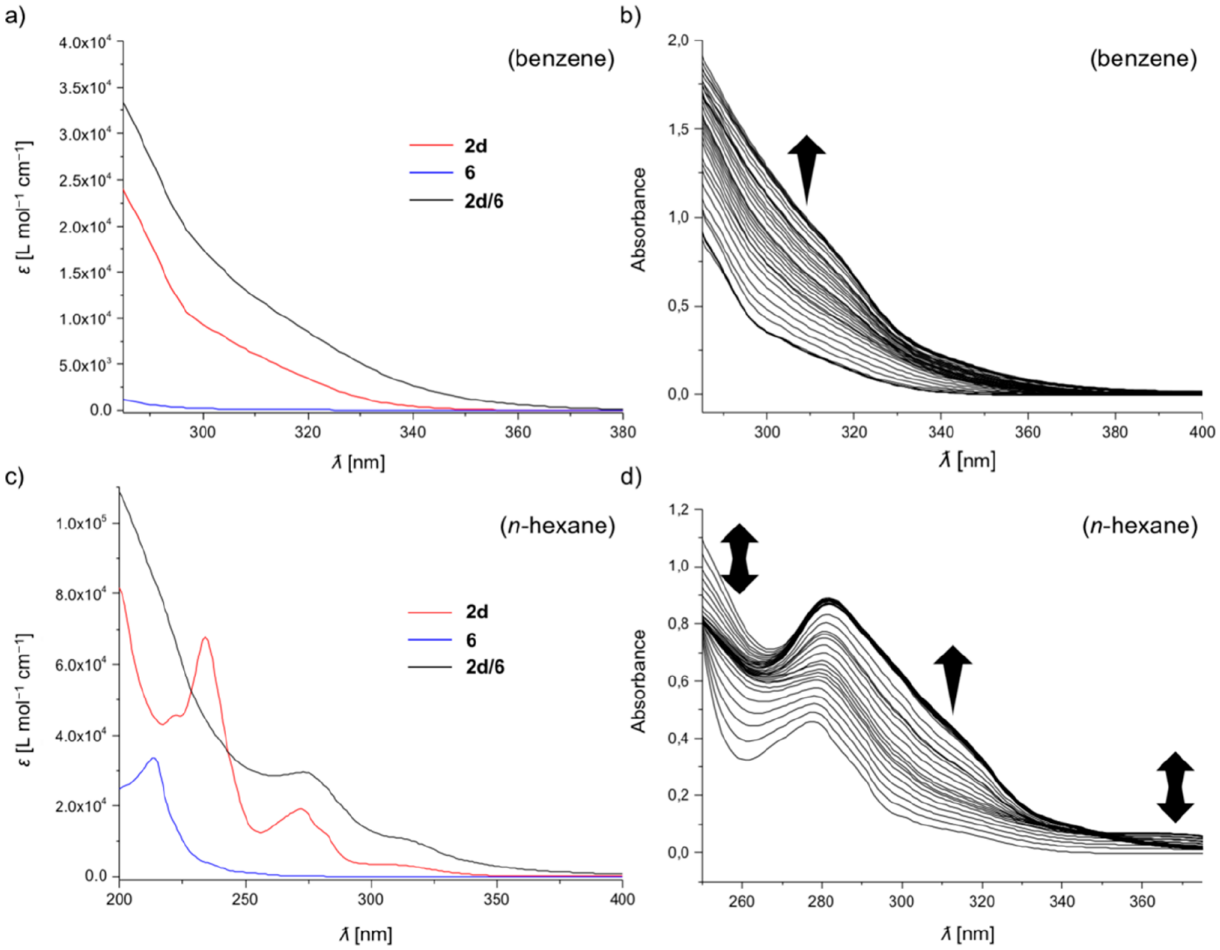


Figure A2.10.: UV/vis absorption spectra of **2d**, **6** and a 1:1 mixture thereof in benzene (a) and *n*-hexane (c) and during titration of **2d** with **6** in benzene (b) and *n*-hexane (d). The arrows indicate changes in signal intensity during the titration.

Further spectra can be found in the Supporting Information. Upon addition of acceptor **6** to donor **2d**, a CT band appeared at 319 nm in benzene and at 318 nm in *n*-hexane (Fig. A2.10b and d). We assume that benzene exhibits stronger interactions with the acceptors compared to *n*-hexane and thereby lowers their acceptor strength or availability for binding. Such interactions in D-A-type complexes between pyrazolate CTCs **4** and **6** and aromatics such as benzene,[267] toluene,[266,268] or naphthalene[269] have been shown with X-ray crystallography. Hence, all binding constants obtained from benzene are lower compared to those from *n*-hexane. Comparing entries 5 and 6 as well as entries 12 and 13 in Tab. A2.5 shows that both methods, titrations using ^{1}H NMR and UV/vis absorption spectroscopy, provided very similar binding constants and free binding energies, when the same solvent was used. This proves the validity of both methods. The choice of solvent, however, had a strong effect. The binding constants obtained in *n*-hexane were 1 or 2 orders of magnitude larger than those in benzene for all donor-acceptor combinations given in Tab. A2.5. This shows that the noncovalent binding was significantly stronger in *n*-hexane compared to that in benzene with an increase in binding free energy of

acceptor **A**	ΔE^{b}	$\Delta E^{disp,c}$	ΔG^{d}_{RRHO}	$\Delta\delta G_{solv}$	$\Delta G_{calc,11}$
4	-2.5	-25.7	19.2	4.6^{e}	-4.4
5	-5.9	-26.8	20.4	6.2^{e}	-6.2
5	-5.9	-26.8	20.4	5.4^{f}	-7.0
6	-8.4	-26.1	19.0	5.7^{e}	-9.9

Table A2.6.: Calculated Binding Free Energies ΔG_{calc11} of Carbeniate CTC **3**∗ (as Donor **D**) with with Pyrazolate CTCs 4–6 as Acceptors **A** a[a]. [a] ΔG_{calc11} (kcal mol^{-1}) obtained as the sum of electronic interaction energies ΔE, dispersion correction ΔE_{disp} , thermostatistical contributions ΔG_{RRHO} (298 K) and solvation free energies $\Delta\delta G_{solv}$. [b] PBE0/QZ. [c] D3(BJ)^{ATM}. [d] B97-3c. [e] In C_6H_6. [f] *n*-Hexane.

up to 2.6 kcal mol^{-1} (difference of the values from entries 6 and 7). These results demonstrate that the deviation between calculated and experimental (NMR) binding free energies could be due to the use of an implicit solvent model allowing for a computationally efficient evaluation of the important δG_{solv}. These models depend on the parametrization, which is normally good for pure organic molecules, but might be limited in describing a direct interaction of the coinage metals in the planar CTCs with the solvent. The binding constants measured by UV/vis absorption spectroscopy in *n*-hexane ranged from 9.13×10^{4} L mol^{-1} for carbeniate Au(I) CTC **3** as the weakest donor to 2.34×10^{7} L mol^{-1} for pyridinate Au(I) CTC **2** as the strongest π-base. In benzene as a more polar solvent,[300] we obtained lower association constants of up to 4.0×10^{5} L mol^{-1}. As mentioned above, C_6D_6 and $CDCl_3$ provided similar values. These association constants lie more than 1 order of magnitude above the highest value of 2.6 x 10^{4} L mol^{-1} reported up to date for a D-A complex (measured in CH_2Cl_2),[271] which is solely based on the interaction of the π-systems and lacks additional assistance from other noncovalent interactions.[255] Quantum chemical calculations were performed (as discussed above) to provide association free energies for the 1:1 complexes of carbeniate CTC **3** with acceptors **4-6** (Tab. A2.6). For computational efficiency, the butyl groups were replaced by methyl groups, indicated with an asterisk (**3***). For acceptor **5**, the binding free energy was calculated in benzene and *n*-hexane for comparison with the experiments. While for the weakest acceptor, **4**, an excellent agreement between experiment and theory was found (-4.37 vs -4.4 kcal mol^{-1}), the calculations overestimated the binding free energies for stronger acceptors **5** and **6**. However, the influence of the solvent is clearly visible when comparing the values calculated in benzene and *n*-hexane for **5** (middle rows in Table 6). While the calculated binding free energy in benzene of -6.2 kcal mol^{-1} significantly deviates from the experimental value of -4.89 kcal mol^{-1}, the value in *n*-hexane is well-reproduced (-7.0 vs -6.81 kcal mol^{-1} for theory and experiment, respectively). Hence the implicit solvent model used in the calculations provided more accurate results for the nonpolar and nonaromatic solvent *n*-hexane compared to benzene or chloroform.

Scanning Electron Microscopy on D-A Assemblies. Lastly, we performed scanning electron microscopy (SEM) to assess whether the assembly of donor and acceptor created any superstructures. We chose donor **1a** and acceptors **22** and **23** for this study.

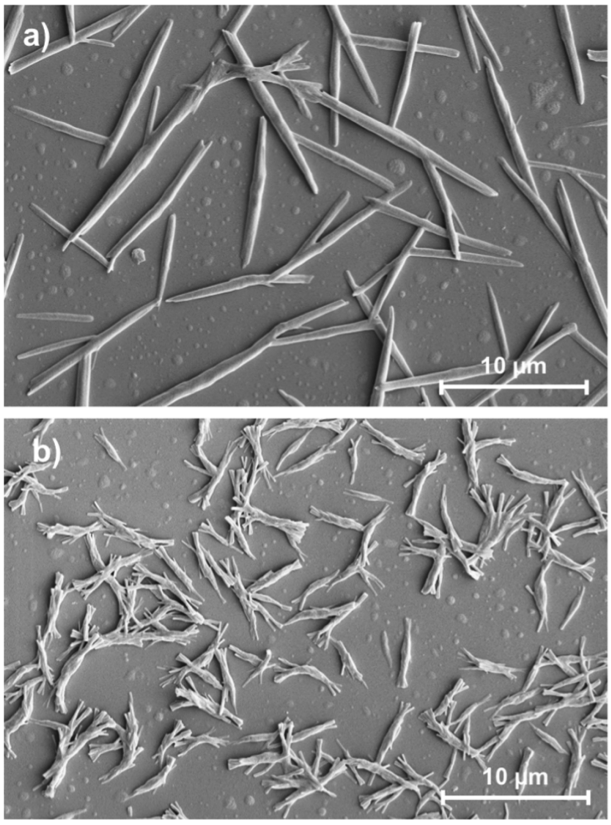


Figure A2.11.: Scanning electron microscopic images of 1:1 mixtures of donor **1a** and acceptors **22** (a) and **23** (b), deposited from acetone on a silicon wafer substrate.

The 1:1 mixtures of both **1a** and **22** as well as **1a** and **23** form crystalline supramolecular assemblies when deposited from acetone (Fig. A2.11a,b). For **1a** and **22**, these were needle-shaped with a length of 13.9 ± 3.2 μm and a diameter of 1.0 ± 0.17 μm. **1a** and **23** display slightly twisted needletype structures and stronger lateral aggregation with a tendency to form spherulite structures. The pure compounds **1a**, **22**, and **23**, on the other hand, did not form crystalline structures or formed distinctly different morphologies, which underscores the successful coassembly (see Supporting Information). When deposited from benzene, a 1:1 mixture of **1a** with **22** as well as **23** formed the same kind of superstructures; however, these were more highly aggregated due to a lower solubility of these compounds in benzene compared to acetone (see the Supporting Information).

A2.3. Conclusions

In summary, we have shown that cyclic trinuclear complexes of gold(I) with pyridinate, imidazolate or carbeniate ligands exhibit strong interactions as donors with a range of planar π-acceptors. Using NMR and UV/vis absorption spectroscopic titrations we determined binding free energies down to -10.1 kcal mol^{-1}and association constants of up to 2.34×10^7 L mol^{-1}, which places them 3 orders of magnitude above known donor-acceptor complexes. To probe the binding model employed we designed and synthesized a bridged naphthalene diimide-based acceptor, which possesses only one binding site.

A comparison with the corresponding nonbridged acceptor showed that nonlinear regression using a 1:2 (donor/acceptor) binding model is best suited to quantify the interaction between these planar donor and acceptor molecules. We observed the type of solvent and its descriptions in DFT calculations to exhibit a significant influence on the free energies of binding. Scanning electron microscopy (SEM) showed that these donor/acceptor pairs form crystalline supramolecular assemblies. Furthermore, quantum chemical calculations showed that London dispersion provides the largest attractive contribution to the binding free energies in these systems. We anticipate that the strong donor-acceptor interactions presented herein will be relevant for the design of supramolecular systems, as already seen in SEM images, and chemical sensors, and that they can serve as a benchmark for donor-acceptor interactions.

A2.4. Computational Details

The semiempirical GFN-xTB[166] method was used for pre-optimizations, and subsequent reoptimization was performed with PBEh-3c/COSMO and the program ancopt as implemented in xtb code.[166] Input geometries of the complexes were created with the intermolecular force-field xTB-IFF[224] or prepared by hand and optimized as described above. Calculations of harmonic frequencies were conducted at the B97-3c level. Single-point energies in the gas phase were calculated with the global hybrid density functional PBE0 in combination with the def2-QZVP[288] basis set and the D3(BJ)^{ATM} [146,147] dispersion correction. All DFT calculations were conducted using the TURBOMOLE 7.0.2 (and 7.2.1 for B97-3c) program package.[301] The resolution of the identity (RI) for the Coulomb integrals[85] was always applied using matching default auxiliary basis sets.[86] For the integration of the exchange-correlation energy the numerical quadrature grid m4[302] was always applied. The solvation model COSMO-RS[289,303] was used as implemented in COSMOtherm employing the BP_TZVP_C30_1601 parametrization. To obtain solvation free energies, two single-point calculations were performed with BP86[110,114]/def-TZVP[304] (one in the gas phase and one in an ideal conductor with $\epsilon = \infty$) and then used as input for the COSMOtherm program.

A3. Supporting Information to Chapter 3

Appendix A3 contains:

- Additional technical details of the calculations
- energy decomposition analysis

Computational Details

The calculation of free energies begins with the optimization of the input-geometries. These were created by hand and in case of the complexes, additionally prepared with the intermolecular force-field xTB-IFF.[224] The geometries are preoptimized with the robust semiempirical GFN-xTB[166] method and subsequently reoptimized with PBEh-3c[154]/COSMO. For optimizations the program *ancopt* was used as implemented in the *xtb* code. The convergence criteria for converged geometries are: $E_{thr} = 5\times10^{-6} E_h$ for the energy change and, $G_{thr} = 1\times10^{-3} E_h$ /Bohr for the total deviation of the gradient. The composite scheme PBEh-3c consists of a variant of the global hybrid density functional PBE0, in combination with a modified double-zeta (DZ) basis set,[288] a geometrical counterpoise correction[157] (gCP) and the D3-dispersion correction[142] (including the three body contribution $D3^{ATM}$[147]). Free energies are calculated for each structure in an established multilevel ansatz[183], consisting of three parts. In part one high level single-point energies E are calculated with PBE0[122,287] -D3(BJ)^{ATM}[144,145]/def2-QZVP[288] in the gas phase. The global hybrid density functional PBE0 is used together with a large QZ basis set and the D3 correction with three body term and Becke-Johnson damping. In part two the calculation of harmonic vibrational frequencies is conducted at the B97-3c[155] level. To this end the structures have to be reoptimized with B97-3c. The frequencies are applied unscaled to obtain the thermostatistical correction G_{RRHO}. To reduce the error of the harmonic approximation for low-lying vibrational frequencies as well as numerical noise in the calculations, the RRHO-scheme (an interpolation between the rigid-rotor- to the harmonic-oscillator is applied at low-lying frequencies (every vibrational mode below 100 cm^{-1}). In part three the solvation contribution to free energy is calculated with COSMO-RS[289,303] at 298.15 K, 1 bar pressure and in the applied solvent. Incorporated in the solvation free energy is the volume work $RT\ln(V_{Ideal})$ going from an ideal gas at 1 bar pressure to an ideal gas at 1 mol/L. The free energy of one compound is obtained as the sum of all three contributions:

$$G = E + G^T_{RRHO} + \delta G^T_{solv}(X) \quad . \tag{A3.1}$$

Binding free energies are calculated from the difference of the free energy of the complex and the monomers:

$$\Delta G_{calc} = G_{complex} - (G_{Monomer1} + G_{Monomer2}) \quad . \tag{A3.2}$$

Further Technical Details

All DFT calculations were conducted with the TURBOMOLE 7.0.2 program package,[301] except for all B97-3c calculations which were performed with TURBOMOLE 7.2.1.[305] The resolution-of-identity (RI) approximation for Coulomb integrals was generally applied[85] using the matching default auxiliary basis sets.[86] The integration of the exchange-correlation contribution was performed on the numerical quadrature grids *m4*. The default convergence criteria for single-points [$10^{-7} E_h$] was applied. The conductor like screening model for real solvents (COSMO-RS) was used as implemented in COSMOtherm2016[306,307] employing the BP_TZVP_C30_1601 parametrization. For each solvation free energy value two single-point calculations with BP86[110,114]/def-TZVP (one in the gas-phase and one in an ideal conductor) have to be performed on the optimized structures. These calculations are then used as input for the COSMOtherm program. The solvents used in the calculations are $CHCl_3$, Benzene and *n*-hexane.

Energy Decomposition Analysis (EDA)

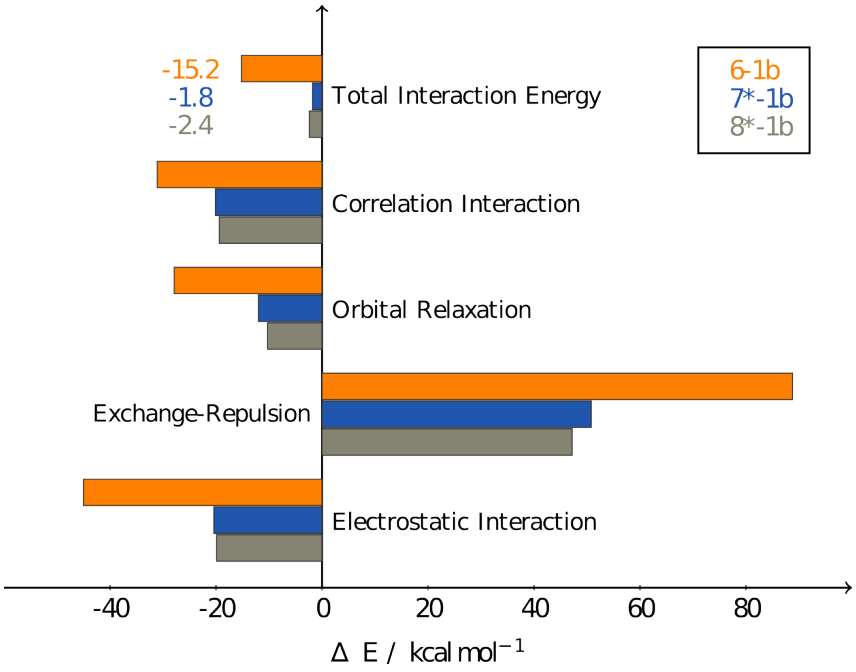


Figure A3.1.: EDA of bridged (**7***·**1b**) and unbridged (**8***·**1b**), as well as stacked CTC complex **6**·**1b**.

Energy decomposition analysis is a useful method for the interpretation of chemical bonds, expressed in separate energy terms.[308] Here we compare the noncovalent interaction in the complexes **7***·**1b** and **8***·**1b** at the level of PBE0/QZ. The additive D3(BJ) dispersion correction has to be treated separately in an EDA calculation. Comparing **7***·**1b** and **8***·**1b** only a small difference in the total interaction energy is visible. The correlation and electrostatic interactions don't differ much between both complexes. The exchange-repulsion term is larger in the bridged system and is only partially compensated for by the orbital relaxation term, which leads to an overall slightly less attractive interaction energy, compared to the unbridged complex (**8***·**1b**). The EDAs have been calculated with ridft as implemented in TURBOMOLE.7.2.1, using *grid 4* and *c1* symmetry.

A4. Influencing the Self–Sorting Behavior of [2.2]Paracyclophane – Based Ligands by Introducing Isostructural Binding Motifs

Abstract Two isostructural ligands with either nitrile ($\mathbf{L}^{nit}$) or isonitrile ($\mathbf{L}^{iso}$) moieties directly connected to a [2.2]paracyclophane backbone with pseudo-*meta* substitution pattern have been synthesized. The ligand itself ($\mathbf{L}^{nit}$) or its precursors ($\mathbf{L}^{iso}$) were resolved by HPLC on a chiral stationary phase and the absolute configuration of the isolated enantiomers was assigned by XRD analysis and/or by comparison of quantum-chemical simulated and experimental electronic circular dichroism (ECD) spectra. Surprisingly, the resulting metallosupramolecular aggregates formed in solution upon coordination of [(dppp)Pd(OTf)$_2$] differ in their composition: whereas $\mathbf{L}^{nit}$ forms dinuclear complexes, $\mathbf{L}^{iso}$ exclusively forms trinuclear ones. Furthermore, they also differ in their chiral self-sorting behavior as (*rac*)-$\mathbf{L}^{iso}$ undergoes exclusive social self-sorting leading to a heterochiral assembly, whereas (*rac*)-$\mathbf{L}^{iso}$ shows a twofold preference for the formation of homochiral complexes in a narcissistic self-sorting manner as proven by ESI mass spectrometry and NMR spectroscopy. Interestingly, upon crystallization, these discrete aggregates undergo structural transformation to coordination polymers, as evidenced by single-crystal X-ray diffraction.

A4.1. Introduction

Mastering the self-sorting behavior in multicomponent mixtures is of utmost importance to cope with the immense structural complexity that such systems offer.[309–313] Geometrical (self-)complementarity concerning shape and size but also chirality have been used to achieve self-sorting either in a narcissistic[314] (also known as self-recognition[315]) or social[316] (also called self-discrimination[317]) manner. Among the features mentioned above, high-fidelity self-sorting is most challenging with enantiomers that are usually highly similar competitors due to their equal size and shape and the chirality as the only difference. Although successful chiral self-sorting is well established in the solid state due to strong interactions and cooperative binding mechanisms found in crystalline matter,[318–321] self-sorting of racemic mixtures in solution is still a true challenge.[322–324] Nevertheless, both scenarios, the heterochiral social[317,325–335] and the homochiral narcissistic self-sorting[336–363] have been reported in metallosupramolecular chemistry. These studies revealed that there are several possibilities to affect and fine-tune the self-sorting behavior, for example, by varying the overall size, the opening angle and the steric bulk of bridging rigid V-shaped ligands.

Based on structural findings in those studies, we wanted to introduce yet another variable concerning only the metal binding unit, to get further insight into these processes and factors that might rule their outcome. Herein, we report on the comparison of two different coordination motifs with regard to self-sorting by introducing two ligands with equal size, shape and coordination angles. The ligands are designed to form supramolecular architectures in solution and in the solid state upon coordination to *cis*-protected palladium(II)-ions,[364–371] giving rise to $[(dppp)_nM_nL_n]$ complexes, which have been proven a simple, yet successful, platform for investigating changes in the self-sorting behavior.

A4.2. Results and Discussion

The design is based on an analogy to a *meta*-substituted benzene core furnished with the commonly used 4-pyridyl moiety. In the literature, there are several examples for directly bound and elongated ligands with for example, ethynyl-spacers coordinating to palladium(II).[372–374] To study chiral self-sorting behavior, we decided to use a pseudo-*meta*-substituted [2.2]paracyclophane backbone as a chiral substitute for the central *meta*-functionalized benzene ring, which also arranges its functional groups at an angle of about 120° designed to construct discrete coordination architectures. Additionally, this scaffold comes with some adaptability due to a possible twist of approximately 12° between the aromatic decks.[375–378] Considering that we wanted to study two different ligands with essentially equal size and shape, we decided to use nitrile and isonitrile groups (bond length of the CN bond in nitriles 1.5 Å and in isonitriles 1.7 Å, respectively),[379] which are formally isoelectronic to carbon monoxide and have been employed previously to construct metallosupramolecular architectures.[334,380–385] It is worth mentioning, though, that multitopic oligoisonitrile ligands have —to the best of our knowledge— not been used for the self-assembly of discrete cyclic or cage-like metallosupramolecular aggregates so far. Both nitriles and isonitriles represent very slim metal-binding motifs. However, they differ in the relative strength of the metal–ligand interaction because the isonitriles bind—similarly to pyridines—much stronger to metal ions like palladium(II) ions than the corresponding nitriles.[386] For our purpose, we synthesized two ligands with either two nitrile or isonitrile moieties attached directly to the [2.2]paracyclophane backbone in a pseudo-*meta*-4,15-substitution pattern. This orientates the functional groups at an angle of about 120° and positions the binding sites spatially close to the planar chiral center, a design intended to give rise to discrete coordination architectures and enable the study of the influence of binding strength on self-sorting.

A4.3. Synthesis

The synthesis of the two different ligands $\mathbf{L}^{nit}$ and $\mathbf{L}^{iso}$ (Scheme A4.1) started with the commercially available unsubstituted [2.2]paracyclophane **1** (Scheme A4.2) which can be easily functionalized through the bromination method developed by Hopf *et al.* to yield 4,15-dibromo[2.2]-paracyclophane **2** and its achiral regioisomer 4,16-dibromo-[2.2]paracyclophane, which could be separated by recrystallization.[387]

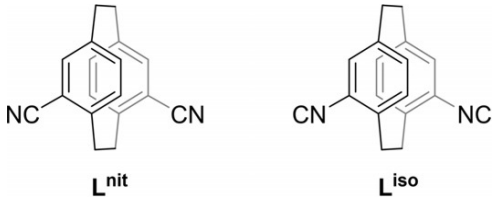


Figure A4.1.: Pseudo-*meta* functionalized [2.2]paracyclophane ligands $\mathbf{L}^{\mathrm{nit}}$ and $\mathbf{L}^{\mathrm{iso}}$.

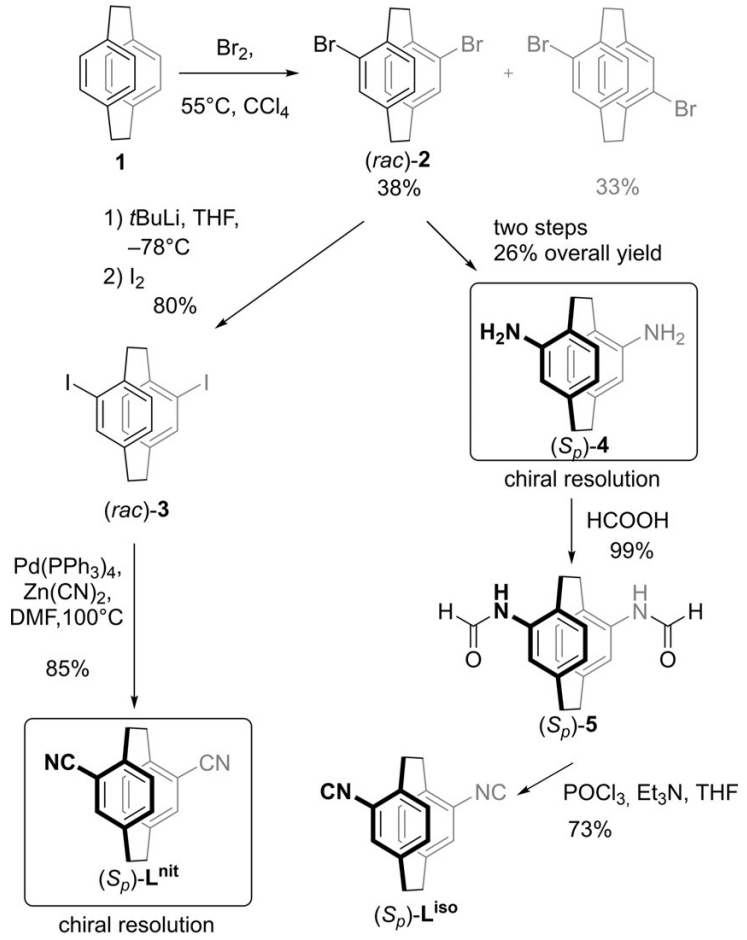


Figure A4.2.: Synthesis of ligands $\mathbf{L}^{\mathrm{nit}}$ and $\mathbf{L}^{\mathrm{iso}}$.

The synthesis of the racemic ligand $\mathbf{L}^{\mathrm{nit}}$ was already reported by Cram and co-workers in 1975[388] but it has neither been chirally resolved nor used for metal-complexation studies so far. We obtained the ligand $\mathbf{L}^{\mathrm{nit}}$ by an alternative route achieving higher yields. Specifically, we performed the already reported bromine–lithium exchange in THF at -78 °C first by using *t*BuLi and then by quenching the dilithiated species with iodine to access the diiodinated [2.2]paracyclophane **3** in 80 % yield.[389] Next, we synthesized $\mathbf{L}^{\mathrm{nit}}$ by palladium-catalyzed cyanation reaction providing the desired product in 85 % yield. Considering that we needed both the racemic and the enantiomerically pure compounds for our studies, we next resolved ligand $\mathbf{L}^{\mathrm{nit}}$ by HPLC on a Chiralpak IB stationary phase both on the

analytical and semi-preparative scale (see Supporting Information). The absolute configuration could be elucidated by comparison of experimentally recorded CD spectra of the resolved material with theoretically simulated ones obtained from simplified time-dependent (sTD)-DFT-calculations.[218] This assignment was corroborated by determination of the Flack parameter by single crystal X-ray diffraction and subsequent analysis of the Bijvoet differences by using Bayesian statistics (vide infra).[390] Both techniques proved that the (R_p)-enantiomer was the first to elute (see Supporting Information). The new ligand $\mathbf{L}^{iso}$ was obtained in a two-step synthesis with excellent yields by using the previously reported pseudo-meta-4,15-diamino[2.2]paracyclophane[389] **4** as the starting material. First, the 4,15-diformamido[2.2]paracyclophane **5** was obtained by dissolving the starting material in an excess of formic acid and sonicating the solution in an ultrasonic bath for 6 h. After purification by filtration through a short silica column and treatment with $POCl_3$ in a 1:1 mixture of THF and triethylamine, ligand $\mathbf{L}^{iso}$ was obtained in 73 % yield. Fortunately, pseudo-*meta*-4,15-diamino[2.2]paracyclophane **4** is available in enantiomerically pure form, because all attempts to resolve ligand $\mathbf{L}^{iso}$ directly by HPLC using chiral columns in a sufficient manner with regard to enantiomeric purity failed. Nevertheless, the absolute configuration of the resolved enantiomers could be confirmed by single crystal X-ray diffraction and subsequent analysis of the Bijvoet differences by using Bayesian statistics (vide infra and Supporting Information)[390]. In this way, we had access to both ligands in racemic and optically pure form, which allowed us to compare their self-sorting behavior in the self-assembly of metallosupramolecular complexes upon coordination to *cis*-protected palladium(II) ions in solution and in the solid state.

A4.4. Coordination studies: solution and gas-phase studies

The metallosupramolecular aggregates were formed upon mixing stoichiometric amounts of ligand (each either in racemic or enantiomerically pure form) with [(dppp)Pd(OTf_2)][391] in different solvents and characterized by NMR techniques and, where possible, by ESI-MS measurements. Coordination of ligand $\mathbf{L}^{nit}$ can be observed in, for example, nitromethane. The ^{1}H NMR spectrum of the complex formed from enantiomerically pure ligand $\mathbf{L}^{nit}$ shows only one set of sharp, significantly shifted signals, indicating the successful formation of a single discrete homochiral metallosupramolecular aggregate of D_2 symmetry (Fig. A4.3).

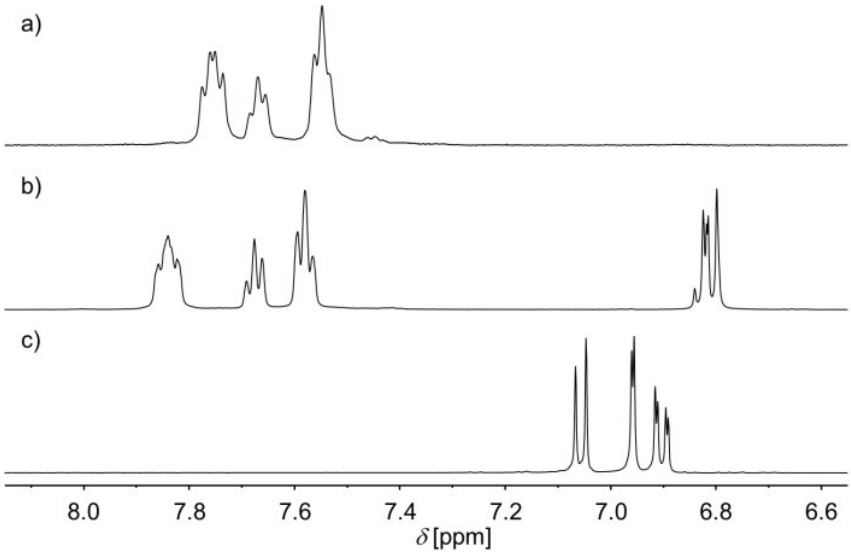


Figure A4.3.: ^{1}H NMR (499 MHz in CD_3NO_2 at 298 K) of a) [(dppp)Pd(OTf)$_2$], b) a 1:1 mixture of [(dppp)Pd(OTf)$_2$] and (S_p)-$\mathbf{L}^{\mathrm{nit}}$, and c) ligand $\mathbf{L}^{\mathrm{nit}}$.

Further proof of the exclusive formation of only one aggregate was obtained by ^{1}H DOSY experiments (Fig. A4.4): all signals belong to a single species with a diffusion coefficient D = 4.42 × 10^{-6} $cm^2 s^{-1}$, which translates to a hydrodynamic diameter d = 16.0 Å, according to the Stokes–Einstein equation. This is considerably larger than the ligand $\mathbf{L}^{\mathrm{nit}}$ alone, but matches very well the calculated value[392] of the hydrodynamic diameter of a dinuclear rhomb (about 16 Å, see Fig. A4.6a). Unfortunately, despite our best efforts, we have not been able to obtain any mass spectra of this kind of complexes formed from bis(nitrile) ligands, presumably due to the weak binding between the palladium atoms and the nitrogen donors.

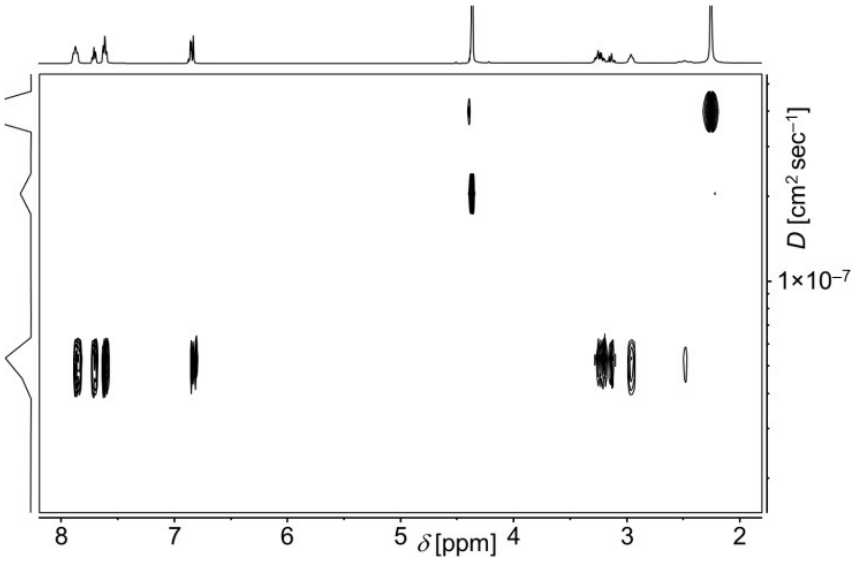


Figure A4.4.: ^{1}H DOSY NMR Spectra (499 MHz in CD_3NO_2 at 293 K) of the aggregate obtained from a 1:1 mixture of [(dppp)Pd(OTf)$_2$] and (S_p)- $\mathbf{L}^{\mathrm{nit}}$.

Nevertheless, we went on to analyze the behavior of the racemic ligand (*rac*)-$\mathbf{L}^{\mathrm{nit}}$ by recording a similar set of NMR experiments. Comparing them with the spectra of the free ligand and the homochiral complex formed from the optically pure ligand revealed that a single metallosupramolecular complex is formed which is different from the homochiral dinuclear complex as indicated by significant shifts (Figure A4.5).

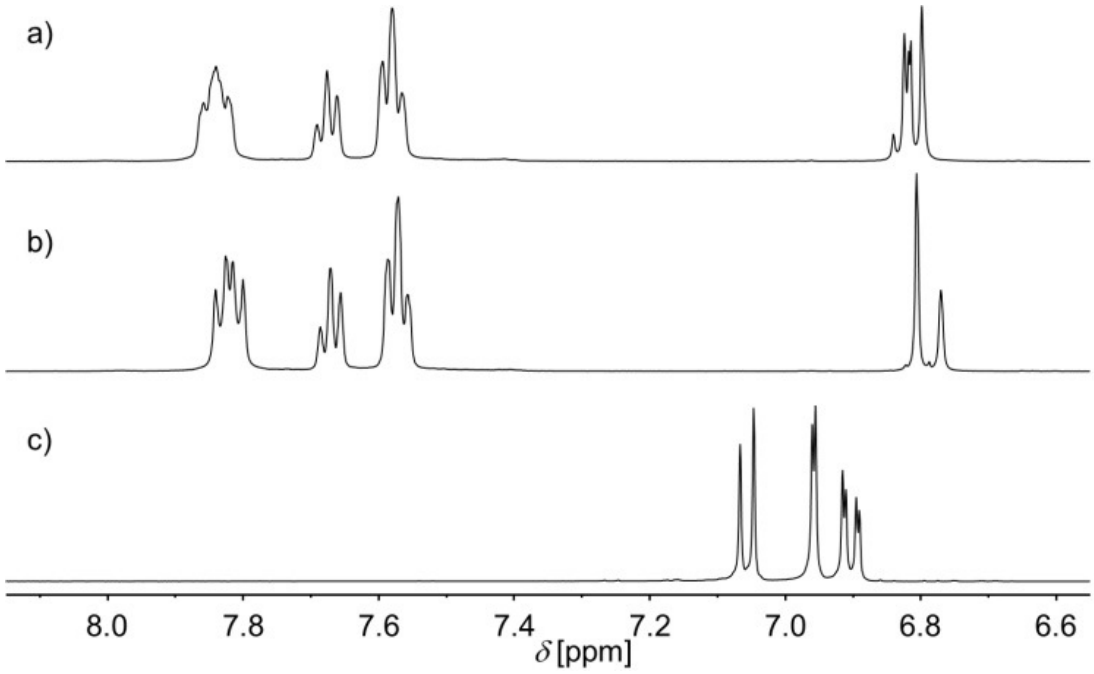


Figure A4.5.: ^{1}H NMR (499 MHz in CD_3NO_2 at 298 K) of a) 1:1 mixture of [(dppp)Pd(OTf)$_2$] and (S_p)-$\mathbf{L}^{\mathrm{nit}}$, b) a 1:1 mixture of [(dppp)Pd(OTf)$_2$] and (*rac*)-$\mathbf{L}^{\mathrm{nit}}$, and c) ligand $\mathbf{L}^{\mathrm{nit}}$.

Considering that the ^{1}H DOSY experiment (see Supporting Information) again indicate the formation of a dinuclear assembly, we conclude that the self-assembly process involving (*rac*)-$\mathbf{L}^{\mathrm{nit}}$ is still highly diastereoselective, but results in the exclusive formation of the heterochiral C_{2v} symmetric complex (Fig. A4.6b). This interpretation in favor of a diastereoselective self-sorting in a self-discriminating manner was further corroborated by comparison of ^{31}P NMR spectra, which show a single signal in each case (see Supporting Information). Compared to [(dppp)Pd(OTf)$_2$], the signals of both the homochiral [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{\mathrm{nit}}$}$_2$](OTf)$_4$ and the heterochiral [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{\mathrm{nit}}$}{($R_p$)-$\mathbf{L}^{\mathrm{nit}}$}](OTf)$_4$ are high-field shifted, indicating the selective formation of the desired dinuclear metallo-supramolecular aggregates.

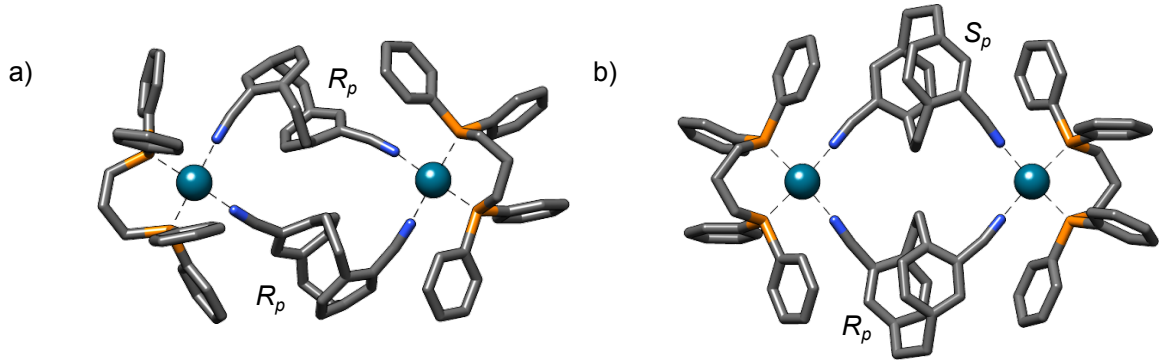


Figure A4.6.: PBEh-3c/DCOSMO-RS(NO_2CH_3) optimized structures of the cationic units of dinuclear metallosupramolecular rhombs: a) homochiral $[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{nit}\}_2](OTf)_4$ and b) heterochiral $[(dppp)_2Pd_2\{(S_p)\text{-}\mathbf{L}^{nit}\}\{(R_p)\text{-}\mathbf{L}^{nit}\}](OTf)_4$ (hydrogen atoms are omitted, color scheme: carbon–grey, nitrogen–blue, phosphorous–orange, palladium–petrol, see Supporting Information for further details).

However, the ^{31}P signals differ in such a way that the signal observed for the heterochiral complex is shifted even further to high field compared to that of its homochiral stereoisomer. This difference is not a result of an anion template effect because the signals of the triflate anions in the ^{19}F NMR spectra were found at the same shift for both dinuclear aggregates $[(dppp)_2Pd_2\{(S_p)\text{-}\mathbf{L}^{nit}\}_2](OTf)_4$ and $[(dppp)_2Pd_2\{(S_p)\text{-}\mathbf{L}^{nit}\}\{(R_p)\text{-}\mathbf{L}^{nit}\}](OTf)_4$, although being clearly different from that of the starting material $[(dppp)Pd(OTf)_2]$ (see Supporting Information). To complete these studies, it is important to mention that essentially the same behavior was observed when chlorinated solvents like dichloromethane were used, but acetonitrile and acetone both proved to be too competitive as they (expectedly) prohibited the complexation (see Supporting Information).[393,394] Having established the highly diastereoselective self-sorting during the self-assembly of ligand $\mathbf{L}^{nit}$ upon coordination to $[(dppp)Pd(OTf)_2]$, we turned our attention to ligand $\mathbf{L}^{iso}$. Like $\mathbf{L}^{nit}$, enantiomerically pure $\mathbf{L}^{iso}$ also gives rise to a single metallosupramolecular species upon coordination to the *cis*-protected Pd^{II} complex fragment. Surprisingly, however, analysis of the DOSY spectra clearly indicate that the aggregate formed from enantiomerically pure $\mathbf{L}^{iso}$ has a diameter of d = 22.6 Å, thus being considerably larger than the dinuclear homochiral complex obtained by self-assembly of optically pure $\mathbf{L}^{nit}$ and $[(dppp)Pd(OTf)_2]$ (see Supporting Information). Fortunately, complexes of palladium(II) ions and isonitrile ligands are more stable than their nitrile analogues,[386] and hence, can be studied by ESI-MS (Fig. A4.7). These studies confirm the formation of trinuclear complexes with enantiomerically pure and racemic $\mathbf{L}^{iso}$, as indicated by signals at m/z = 1463.2 which can be assigned to $\{[(dppp)_3Pd_3(\mathbf{L}^{iso})_3](OTf)_4\}^{2+}$, or the signals at m/z = 811.2 and 925.1 which can both be assigned to superimposed signals arising from triply charged ions $\{[(dppp)_3Pd_3(\mathbf{L}^{iso})_3](Cl)_3\}^{3+}$ and $\{[(dppp)_3Pd_3(\mathbf{L}^{iso})_3](OTf)_3\}^{3+}$ and their singly charged fragments $\{[(dppp)Pd(\mathbf{L}^{iso})](X)\}^{+}$ (X = Cl or OTf), respectively.* Hence, we can conclude that the enantiomerically pure ligand $\mathbf{L}^{iso}$ successfully forms a discrete homochiral trinuclear metallosupramolecular aggregate

*Please note that the chloride stems from abundant chloride in the ESI source of the spectrometer.

of D_3 symmetry upon coordination to [(dppp)Pd(OTf)$_2$].

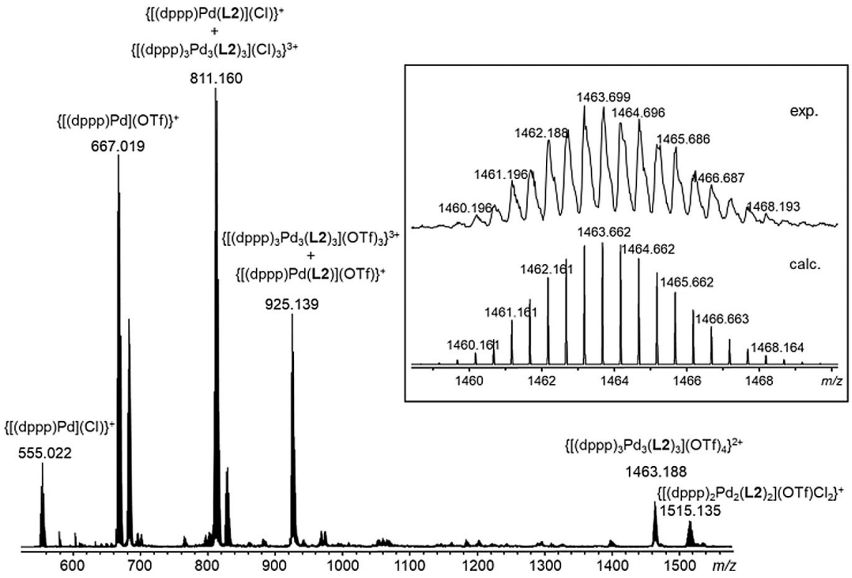


Figure A4.7.: Positive ESI-MS spectrum of the 1:1 mixture of [(dppp)Pd(OTf)$_2$] and (*rac*)-$\mathbf{L}^{\mathrm{iso}}$ showing signals of trinuclear 3:3 aggregates and their fragments. The inset shows the comparison of the experimentally determined isotope pattern of the $\{[(\mathrm{dppp})_3\mathrm{Pd}_3(\mathbf{L}^{\mathrm{iso}})_3](\mathrm{OTf})_4\}^{2+}$ ion and the calculated one.

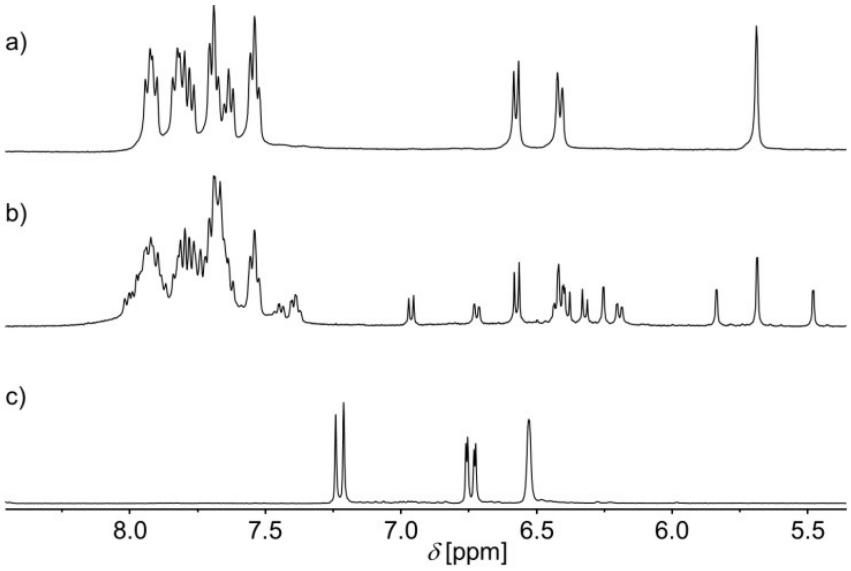


Figure A4.8.: ^{1}H NMR (499 MHz in CD_3NO_2 at 293 K) of a) $[(\mathrm{dppp})_3\mathrm{Pd}_3\{(S_p)\text{-}\mathbf{L}^{\mathrm{iso}}\}_3](\mathrm{OTf})_6$, b) the complexes formed from a 1:1 mixture of [(dppp)Pd(OTf)$_2$] and (*rac*)-$\mathbf{L}^{\mathrm{iso}}$, and c) ligand (*rac*)-$\mathbf{L}^{\mathrm{iso}}$.

Next, we turned our attention to (*rac*)-$\mathbf{L}^{\mathrm{iso}}$. Considering that ESI-MS already revealed the formation of trinuclear complexes, NMR spectroscopic experiments were performed to study its chiral self-sorting

behavior. Interestingly, the ^{1}H NMR spectrum of a 1:1 mixture of [(dppp)Pd(OTf)$_2$] and (*rac*)-$\mathbf{L}^{iso}$ in [D$_3$]nitromethane looks rather complicated (Fig. A4.8). 2D-Correlated spectra confirmed the formation of two different species and ^{1}H DOSY experiments nicely show that these are both of very similar size giving rise to a diffusion coefficient D = 3.20 × 10^{-6} cm^2s^{-1}. This value is again considerably smaller than that observed for the dinuclear complexes formed from ligand $\mathbf{L}^{nit}$ (Fig.A4.9), but nicely matches that of the homochiral trinuclear complex obtained from enantiomerically pure $\mathbf{L}^{iso}$. Therefore, we can assign the two species formed as the homochiral trinuclear complex [(dppp)$_3$Pd$_3${(S_p)-$\mathbf{L}^{iso}$}$_3$](OTf)$_6$ and its enantiomer, and the heterochiral complex [(dppp)$_3$Pd$_3${(S_p)-$\mathbf{L}^{iso}$}$_2$ {(R_p)-$\mathbf{L}^{iso}$}](OTf)$_6$ and its enantiomer, respectively (Fig.A4.10).

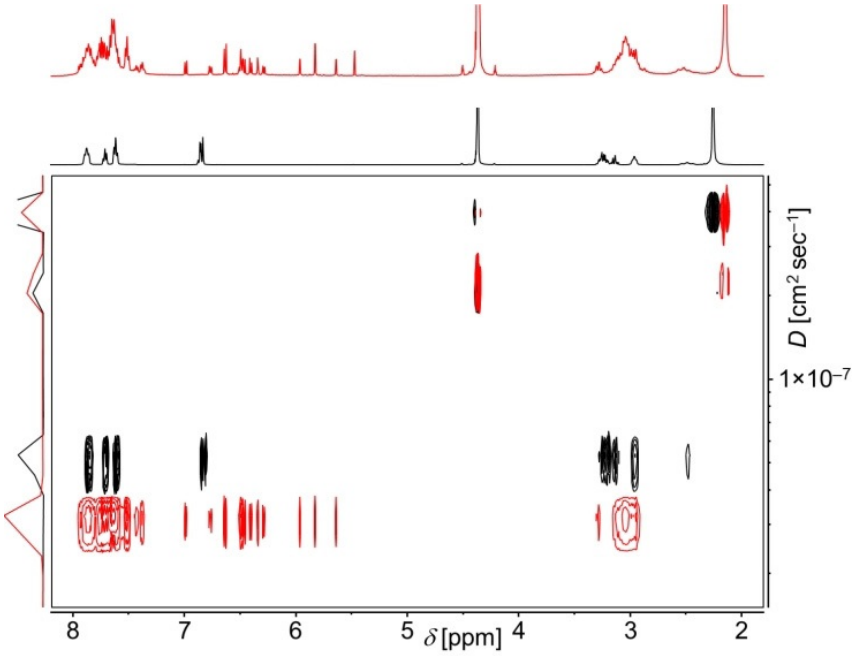


Figure A4.9.: Superimposed ^{1}H 2D DOSY NMR spectra (499 MHz in CD_3NO_2 at 298 K) of [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{nit}$}{(R_p)-$\mathbf{L}^{nit}$}](OTf)$_4$ obtained upon mixing of [(dppp)Pd(OTf)$_2$] and (*rac*)-$\mathbf{L}^{nit}$ (black) and the mixture of [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{iso}$}$_2${(R_p)-$\mathbf{L}^{iso}$}](OTf)$_6$ and the [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{iso}$}$_3$](OTf)$_6$ and their enantiomers, respectively, obtained upon mixing of [(dppp)Pd(OTf)$_2$] and (*rac*)-$\mathbf{L}^{iso}$ (red).

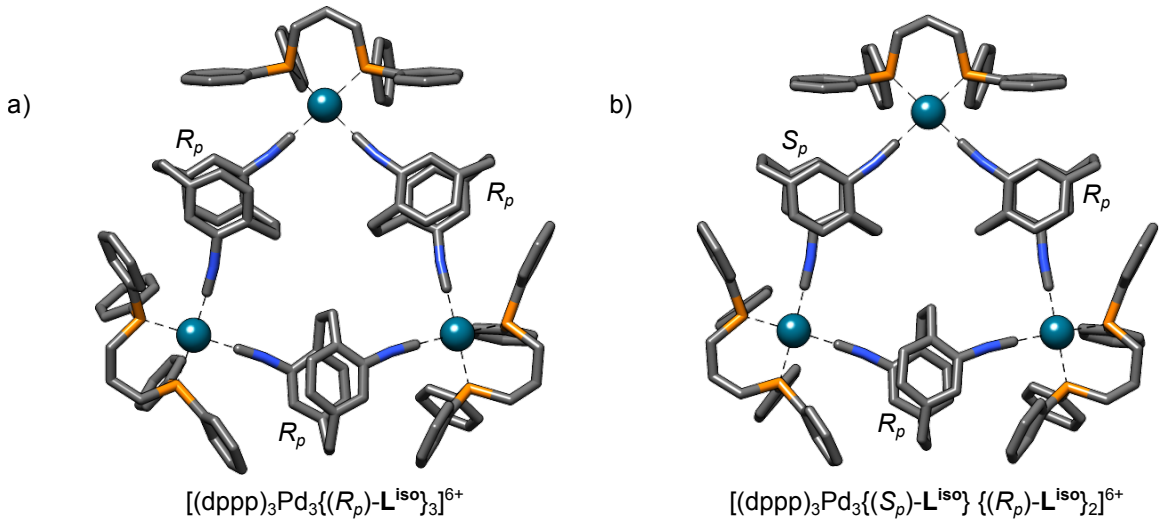


Figure A4.10.: PBEh-3c/DCOSMO-RS(NO_2CH_3) optimized structures of the cationic units of trinuclear metallosupramolecular triangles: a) the homochiral $[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{iso}\}_3](OTf)_6$ and b) heterochiral $[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{iso}\}_2\{(S_p)\text{-}\mathbf{L}^{iso}\}](OTf)_6$ (hydrogen atoms are omitted, color scheme: carbon–grey, nitrogen–blue, phosphorous–orange, palladium–petrol, see Supporting Information for further details).

This is also in agreement with the analysis of the ^{31}P NMR spectra (see Supporting Information). Whereas the homochiral complex $[(dppp)_3Pd_3\{(S_p)\text{-}\mathbf{L}^{iso}\}_3](OTf)_6$ only gives rise to a single signal that is significantly shifted compared to the signal observed for $[(dppp)Pd(OTf)_2]$, the mixture of the racemic homo-chiral and heterochiral trinuclear complexes gave rise to four signals that are again shifted compared to that of $[(dppp)Pd(OTf)_2]$. Like in the case of $\mathbf{L}^{nit}$, we did not observe any anion template effect during the self-assembly of $\mathbf{L}^{iso}$ as we observed only single signals in each of the recorded ^{19}F NMR spectra, which were considerably shifted compared to that of $[(dppp)Pd(OTf)_2]$ (see Supporting Information). Integration of the signals in the ^{1}H and ^{31}P NMR spectra revealed that the homochiral and heterochiral aggregates do not just form statistically in a 1:3 homo-to-heterochiral ratio but rather in a 1 to 1.5 homo-to-heterochiral ratio in an equilibrated solution (see Supporting Information). This translates to a twofold amplified formation of the homochiral assemblies in a narcissistic self-sorting manner.[†] Thus, the bis(isonitrile) ligand $\mathbf{L}^{iso}$ does not only show a different behavior regarding the composition of the metallosupramolecular aggregate formed upon self-assembly with $[(dppp)Pd(OTf)_2]$, but it also shows a contrary self-sorting behavior compared to its almost isostructural bis(nitrile) analogue $\mathbf{L}^{nit}$ (although the self-sorting with $\mathbf{L}^{iso}$ is by far not as perfect as with $\mathbf{L}^{nit}$). This feature is truly remarkable, having in mind the close structural similarity of the two ligands. Obviously, though, these two ligands differ in their dipole moments and their binding strength towards palladium, considering that isonitrile ligands generally coordinate to the metal center much more strongly than nitrile ligands due

[†]It should be noted that similar ratios were observed in acetonitrile and acetone and mixtures of those with chlorinated solvents indicating that the self-assembly process and the self-sorting are not too solvent dependent as long as the solvent is somehow polar, because we could not observe the formation of discrete aggregates in pure chlorinated solvents like dichloromethane or chloroform (see Supporting Information).

to their better σ-donation and superior π-backbonding ability.[383,386,395] Consequently, the (M-)N-C-C and (M-)C-N-C bond angles in nitrile and isonitrile complexes, respectively, can differ because the nitriles tend to be linear, whereas considerable bending of isonitrile groups has been reported.[386] This is also reflected by the results of quantum chemical calculations, which were performed on acetonitrile and isocycanomethane, respectively, to estimate the energetic differences of the bending between the two groups (Fig. A4.11). This might explain the differences in the self-assembly behavior with regard to the different composition, although this needs to be verified with other ligand scaffolds of similar rigidity in the future.

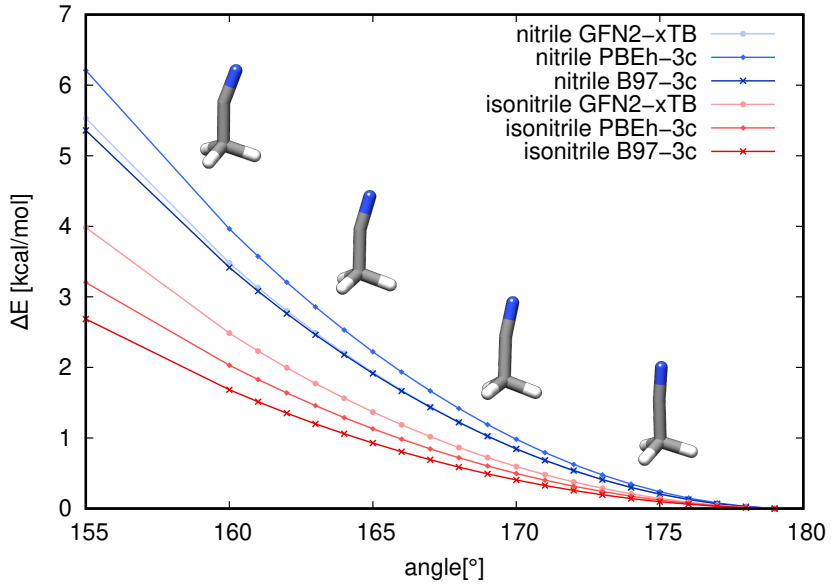


Figure A4.11.: Acetonitrile [N-C-C] and isocyanomethane [C-N-C] bending angle and energy correlation plots obtained when employing different quantum chemical approaches.

A reason for the different self-sorting behavior might be that the heterochiral 2:2 assembly obtained from $\mathbf{L}^{\mathrm{nit}}$ is very symmetric which might favor its formation. However, the heterochiral 3:3 assembly of $\mathbf{L}^{\mathrm{iso}}$ is much less symmetric than its homochiral diastereomer, and hence, might be less favored. However, this is only an empirical explanation based on our experience with similar complexes, and hence, should be regarded with great care.

Nevertheless, we performed quantum chemical calculations to get further hints on the reasons for the striking differences of the self-assembly behavior of our two ligands with regard to complex composition because this is the only way to compare the experimentally observed dinuclear palladium(II) complex of $\mathbf{L}^{\mathrm{nit}}$ with the corresponding hypothetical trinuclear one, and the experimentally observed trinuclear complex of $\mathbf{L}^{\mathrm{iso}}$ with the hypothetical dinuclear one, respectively.

complex	ΔE (PBE0/def2-QZVP)	ΔE_{disp} (D4)	ΔG_{RRHO} (GFN2-xTB/ GBSA[CH_3CN])[a]	$\Delta\delta G_{solv}$ (COSMO-RS[NO_2CH_3])	ΔG_a
$[(dppp)_2Pd_2\{(R_p)\text{–}\mathbf{L}^{nit}\}_2]^{4+}$	-8.2	-37.5	56.8	-66.0	-54.9
$[(dppp)_3Pd_3\{(R_p)\text{–}\mathbf{L}^{nit}\}_3]^{6+}$	102.7	-64.8	89.7	-230.3	-102.7
$[(dppp)_2Pd_2\{(R_p)\text{–}\mathbf{L}^{iso}\}_2]^{4+}$	-41.4	-35.5	60.0	-70.5	-87.4
$[(dppp)_3Pd_3\{(R_p)\text{–}\mathbf{L}^{iso}\}_3]^{6+}$	44.4	-62.6	93.9	-237.6	-161.9

Table A4.1.: Association free energies (ΔG_a) of di- and trinuclear $\mathbf{L}^{nit}$ and $\mathbf{L}^{iso}$ complexes (all energies and free energies in kcal mol^{-1}). [a] Acetonitrile was used as solvent, since nitromethane is not parametrized for the implicit solvation model GBSA and the dielectric constant of acetonitrile is very close to that of nitromethane.

complex	ΔG(GFN2-xTB/ GBSA(CH_3CN))[(a)(c)]	$\Delta\Delta G^{(b)(c)}$ ($2 \cdot M_3L_3 - 3 \cdot M_2L_2$)	ΔG PBE0-D4/def2-QZVP (COSMO-RS(CH_3CN))[(a)]	$\Delta\Delta G^{(b)}$ ($2 \cdot M_3L_3 - 3 \cdot M_2L_2$)
$[(dppp)_2Pd_2\{(R_p)\text{–}\mathbf{L}^{nit}\}_2]^{4+}$	-99.8	-26.3	-54.9	-40.9
$[(dppp)_3Pd_3\{(R_p)\text{–}\mathbf{L}^{nit}\}_3]^{6+}$	-162.9		-102.7	
$[(dppp)_2Pd_2\{(R_p)\text{–}\mathbf{L}^{iso}\}_2]^{4+}$	-149.3	-42.8	-87.4	-61.7
$[(dppp)_3Pd_3\{(R_p)\text{–}\mathbf{L}^{iso}\}_3]^{6+}$	-245.3		-161.9	

Table A4.2.: Free energies of di- and trinuclear $\mathbf{L}^{nit}$ and $\mathbf{L}^{iso}$ complexes (all energies and free energies in kcal mol^{-1}). (a) ΔG_a = association reaction of the complexes. (b) $\Delta\Delta$G = ($2 \cdot M_3L_3 - 3 \cdot M_2L_2$) reaction comparing thermodynamic stability of di- and trinuclear complexes. (c) Acetonitrile was used as solvent, since nitromethane is not parametrized for the implicit solvation model GBSA and the dielectric constant of acetonitrile (ϵ=36) is very close to the one of nitromethane (ϵ=38).

Thermodynamic data of the tri- and di-nuclear complexes are given in Tables A4.1 and A4.2. The semiempirical quantum mechanical GFN2-xTB[167] and high-level PBE0-D4/QZ results show the trinuclear complexes of both binding motifs to be thermodynamically more stable than the dinuclear ones. Considering the efficiency of the GFN2-xTB method compared to the DFT calculations (speed-up of about a factor of 1000 or more), the observed differences in the computed free energies are acceptable and we can recommend this method in general for screening purposes in similar systems, given also that the computed structures are fairly reliable. However, the consistently computed higher stability of trinuclear complexes of $\mathbf{L}^{nit}$, compared to the dinuclear ones, is in stark contrast to the experimental findings. Although the influence of explicit single solvent molecules at the palladium centers was explored in the computations and did not change the initial findings (see Supporting Information), explicit single molecule solvation and implicit solvation models are still inadequate to describe the full effect of solvation on the metallosupramolecular aggregates. Obviously, one would have to take into account a whole ensemble of solvent molecules and study its interaction with the dinuclear and trinuclear complexes in order to get the same trend as observed experimentally. Currently, however, this is clearly beyond the boundaries of this theoretical approach. Nevertheless, it is important to note that the trinuclear complex of the isonitrile ligand $\mathbf{L}^{iso}$ is not only strongly preferred compared to its po-

tential trinuclear analogue of the nitrile ligand $\mathbf{L}^{\mathrm{nit}}$, but it is also much more stabilized compared to the corresponding dinuclear complex as reflected by an energy difference that the solvent effects should not be able to compensate for.

A4.5. Coordination studies: solid-state studies

We were also able to obtain some single crystals of our ligands (see Supporting Information) and some of their complexes suitable for X-ray analysis. Enantiomerically pure $\mathbf{L}^{\mathrm{nit}}$ crystallizes in the orthorhombic space group $P2_12_12_1$ upon evaporation of its dichloromethane solution. The absolute configuration assigned by the comparison of experimental and quantum chemically simulated electronic circular dichroism (ECD) spectra could be confirmed by the analysis of the Flack parameter and the Bijvoet differences by using Bayesian statistics derived from the single-crystal XRD analysis of $\mathbf{L}^{\mathrm{nit}}$ and its coordination complexes. A torsion angle of 119.81° between the two binding units was measured. Enantiomerically pure $\mathbf{L}^{\mathrm{iso}}$ crystallizes in the same orthorhombic space group $P2_12_12_1$ upon evaporation of solvent from a solution in dichloromethane. As $\mathbf{L}^{\mathrm{iso}}$ was prepared from enantiomerically pure building blocks **4** its stereochemistry was already known,[389] but the XRD analysis of its single crystals as well as of its coordination complex further confirmed the stereochemical assignment. As anticipated, a dihedral angle of 117.94° between the coordinating isonitrile groups and the centers of the phenyl rings in the paracyclophane backbone was found (vide supra). Hence, the difference in torsion angle between the two ligands amounts to 1.87°. Diffusion of cyclohexane into a solution of $[(\mathrm{dppp})_2\mathrm{Pd}_2\{(R_p)\text{-}\mathbf{L}^{\mathrm{nit}}\}_2](\mathrm{OTf})_4$ in dichloromethane and subsequent slow evaporation of the solvent afforded single crystals suitable for XRD analysis. Surprisingly, however, addition of the nonpolar solvent and concentration of the solution caused a massive structural rearrangement resulting in a homochiral coordination polymer $[(\mathrm{dppp})\mathrm{Pd}\{(R_p)\text{-}\mathbf{L}^{\mathrm{nit}}\}]_n(\mathrm{OTf})_{2n}$ crystallizing in the orthorhombic space $P2_12_12_1$. The observed polymerization of discrete complexes is most probably due to concentration effects during the crystal growth.[396,397] Coordination polymers with nitrile ligands are known in literature, but usually contain silver(I) centers.[398–401] The asymmetric unit of $[(\mathrm{dppp})\mathrm{Pd}\{(R_p)\text{-}\mathbf{L}^{\mathrm{nit}}\}]_n(\mathrm{OTf})_{2n}$ contains one ligand $\mathbf{L}^{\mathrm{nit}}$, a dppp-protected palladium ion, two triflate anions and a disordered dichloromethane molecule. Each metal center is coordinated to one dppp and two ligands $\mathbf{L}^{\mathrm{nit}}$, forming a helical coordination polymer. Pd(dppp) moieties form the outer edges while the paracyclophane ligands form the inner core of the coordination polymer Fig. A4.12).

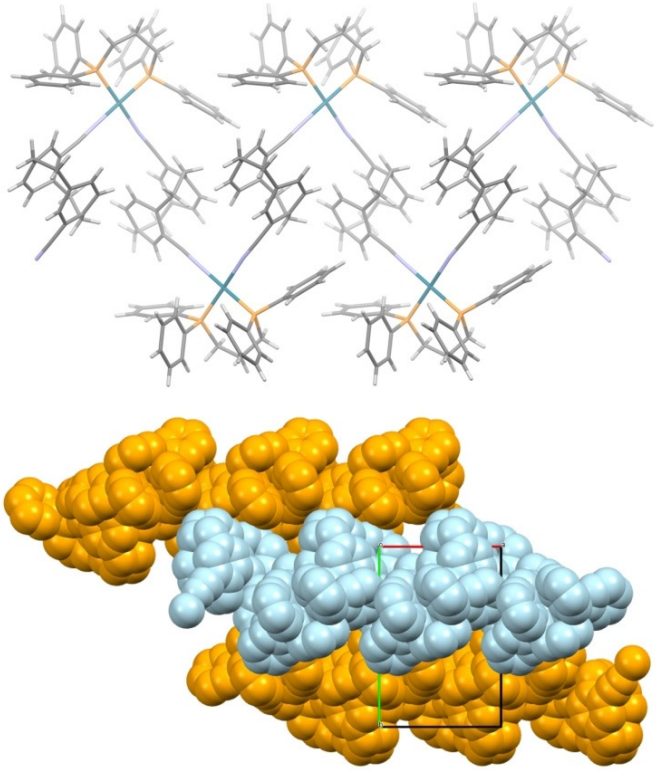


Figure A4.12.: Top: capped sticks representation showing the helical homochiral coordination polymer [(dppp)Pd{(R_p)-$\mathbf{L}^{\mathrm{nit}}$}]$_n(OTf)_{2n}$ (view along the *a* axis, anions and solvent molecules are omitted for clarity, color scheme: grey–carbon, blue–nitrogen, petrol–palladium, orange–phosphorous, white–hydrogen). Bottom: space-filling representation of the coordination polymer [(dppp)Pd{(R_p)-$\mathbf{L}^{\mathrm{nit}}$}]$_n(OTf)_{2n}$ showing the parallel packing of the individual polymer strands (hydrogen atoms, solvent molecules and counter-ions are omitted for clarity).

The polymeric helical strands, generated from the asymmetric unit by a twofold screw axis, are parallel to the crystallographic axis *a*. The triflate anions are placed in pockets formed by the organic ligands above and below the coordination plane of the palladium in apical positions (see Supporting Information). The complete helices are arranged in a parallel and slightly tilted brick-like packing. A structural rearrangement also occurred when cyclopentane was diffused into a solution of heterochiral [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{\mathrm{nit}}$}{($R_p$)-$\mathbf{L}^{\mathrm{nit}}$}](OTf)$_4$ in dichloromethane and the solvents were allowed to slowly evaporate. Here, however, we observed the formation of a heterochiral coordination polymer [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{\mathrm{nit}}$}{($R_p$)-$\mathbf{L}^{\mathrm{nit}}$}]$_n(OTf)_{4n}$ crystallizing in the orthorhombic space group *Fdd*2. The asymmetric unit contains one ligand $\mathbf{L}^{\mathrm{nit}}$, a dppp-protected palladium ion and two disordered triflate anions. Each metal center is coordinated with one ligand of each enantiomer forming a nested-V-like structure (Fig. A4.13). The triflate anions are located between two dppp ligands, again positioned above and below the coordination plane of the palladium centers in apical positions (see Supporting Information). The individual polymer strands are stacked in an interlocked fashion, following the [011] and [01$\bar{1}$] directions in the crystal. Diffusion of benzene into a solution of [(dppp)$_3$Pd$_3${(S_p)-$\mathbf{L}^{\mathrm{iso}}$}$_3$](OTf)$_6$ in

acetonitrile and slow evaporation of the solvent also resulted in the formation of a homochiral coordination polymer [(dppp)Pd{(S_p)-$\mathbf{L}^{iso}$}]$_n$(OTf)$_{2n}$ that turned out to be nearly isostructural to the corresponding homochiral [(dppp)Pd{(R_p)-$\mathbf{L}^{nit}$}]$_n$(OTf)$_{2n}$ coordination polymer. The most significant difference is that solvate acetonitrile, instead of dichloromethane, is packed aligned along [100] with head-to-tail configuration between two parallel polymer chains (see Supporting Information).

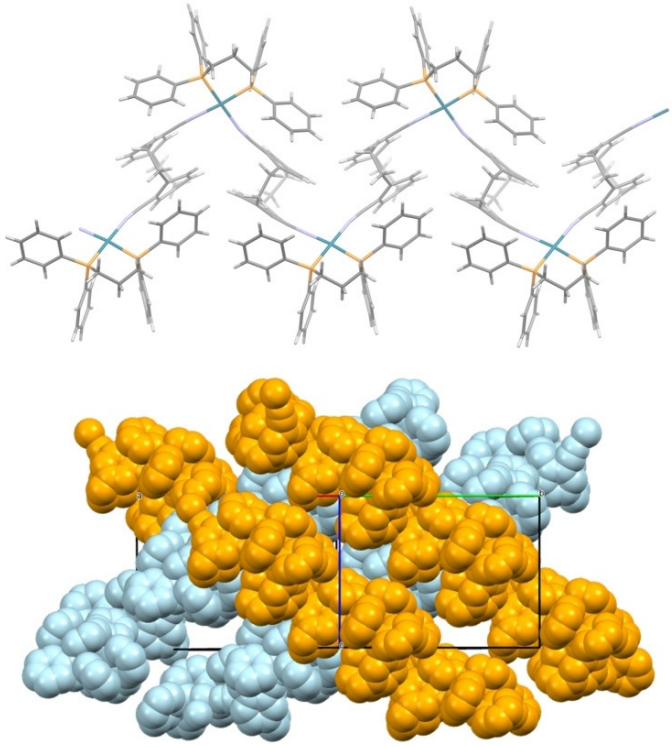


Figure A4.13.: Top: capped sticks representation of the crystal structure of the heterochiral polymer [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{nit}$}{(R_p)-$\mathbf{L}^{nit}$}]$_n$(OTf)$_{4n}$ showing the zig-zag-structure (view along the [011] direction, anions and solvent molecules are omitted for clarity, color scheme: grey–carbon, blue–nitrogen, petrol–palladium, orange–phosphorous, white–hydrogen); bottom: Bottom: space-filling representation of the coordination polymer [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{nit}$}{(R_p)-$\mathbf{L}^{nit}$}]$_n$(OTf)$_{4n}$ showing the interlocked packing pattern of the individual polymer chains (hydrogen atoms, solvent molecules and counter-ions are omitted for clarity).

Despite of our best efforts we did not succeed in obtaining crystals suitable for single-crystal XRD from solutions of the mixture of the four stereoisomers of the trinuclear complex formed from (*rac*)-$\mathbf{L}^{iso}$. Although the formation of helical polymeric structures from discrete cyclic structures, as observed in case of [(dppp)Pd{(S_p)-$\mathbf{L}^{iso}$}]$_n$(OTf)$_{2n}$ and [(dppp)Pd{(R_p)-$\mathbf{L}^{nit}$}]$_n$(OTf)$_{2n}$, is in accordance with earlier observations[329–335,339,341,346,350,351,385] and other self-assembled polymers,[321,330,402–407] the formation of bowl-shaped, concave, yet polymeric, structures as observed in [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{nit}$}{(R_p)-$\mathbf{L}^{nit}$}]$_n$(OTf)$_{4n}$, is more scarce, both on the level of discrete supramolecular structures as well as on the level of self-assembled polymers.

A4.6. Conclusion

We have synthesized two ligands, $\mathbf{L}^{\mathrm{nit}}$ and $\mathbf{L}^{\mathrm{iso}}$ with nitrile and isonitrile metal-binding motifs, respectively, in racemic and enantiomerically pure form and investigated their coordination behavior towards *cis*-protected palladium complex [(dppp)Pd(OTf)$_2$]. Both ligands are almost isostructural concerning the orientation of the nitrile and isonitrile groups (deviation of 1.87°), size and chirality. The only significant difference between the two constitutional isomers is in the bond strength towards the metal ion as isonitrile ligands generally coordinate much more strongly than their nitrile analogues due to better σ-donation and the superior π backbonding ability. We studied the self-assembly and self-sorting by means of NMR spectroscopy, mass spectrometry and single-crystal X-ray diffraction studies. The aggregates of both ligands are formed selectively with regard to their composition, yielding dinuclear rhombs in case of ligand $\mathbf{L}^{\mathrm{nit}}$ and trinuclear triangles in case of ligand $\mathbf{L}^{\mathrm{iso}}$. Interestingly, this difference also leads to a divergence in terms of their chiral self-sorting behavior. Ligand (*rac*)-$\mathbf{L}^{\mathrm{nit}}$ exclusively forms achiral heterochiral dinuclear rhombs [(dppp)$_2$Pd$_2${(S_p)-$\mathbf{L}^{\mathrm{nit}}$}{($R_p$)-$\mathbf{L}^{\mathrm{nit}}$}](OTf)$_4$ in a diastereoselective manner resulting from social self-sorting in non-competitive solvents like nitromethane, dichloromethane or chloroform. In contrast to this, ligand (*rac*)-$\mathbf{L}^{\mathrm{iso}}$ rather shows a twofold preference for narcissistic self-sorting to yield a 1:1.5 mixture of racemic homo- and heterochiral trinuclear complexes. This might be due to the difference between (M-)N-C-C and (M-)C-N-C bond angles in nitrile and isonitrile complexes, respectively, being linear in the case of nitriles, with significant deviation from linearity in the case of the isonitrile groups. Obviously, it is a combination of relative binding strength between the ligands and the metal ions and the interaction with solvent molecules that is responsible for this behavior. Unfortunately, however, simulating this by state-of-the-art quantum chemical calculations is currently not possible. Thus, future studies with other ligands of similar rigidity are needed to establish whether this is a general trend. Single-crystal X-ray diffraction revealed that, upon crystallization, the discrete homo- and heterochiral aggregates rearrange to form coordination polymers. These studies show that the self-sorting behavior is obviously not only dependent on rigid scaffolds with steric crowding or certain bend angles of V-shaped ligands, but even more on the interplay of other important factors, for example, the binding strength and mode towards the coordinated metal centers as well as the resulting structural changes in the metal binding motif.

A4.7. Computational details

All spectra calculations were performed for the R_p-enantiomer. Two approaches were pursued for the calculation of the UV/Vis and ECD spectra of $\mathbf{L}^{\mathrm{nit}}$. First, a single structure of $\mathbf{L}^{\mathrm{nit}}$ was prepared and optimized using the composite method PBEh-3c and the implicit solvation model COSMO(ϵ = 36.6). On the optimized geometry a sTD-CAM-B3LYP/def2-TZVP[128] calculation including proceeding ground state calculation was performed using sTD-DFT[218] as implemented in ORCA 4.0.1.[408] The length form of the electric dipole transition moment is employed for the absorption spectra whereas the velocity form is used for the ECD spectra, which is guaranteeing gauge origin independence. The computed spectra (UV/Vis and ECD) are redshifted by 0.2 eV to roughly match the experimental spec-

trum. In the second approach, conformational flexibility is considered for the spectrum by generating snapshots from a GFN1-xTB/GBSA(acetonitrile)[166] molecular dynamics simulation (at 298 K, using H_{mass} = 4 amu, time step(dt fs^{-1}) = 1.0, md time = 1000 ps, and using the SHAKE algorithm). From the molecular dynamic simulation the equilibration phase was removed and 200 equidistant snapshots were obtained. On these, sTD-BH-LYP/def2-SVP/COSMO(ϵ = 36.6)[114,115,120] calculations were performed using the Turbomole 7.0.2 program package[301] and the sTDA standalone program.[409] The computed spectra (UV/Vis and ECD) are redshifted by 0.4 eV to roughly match the experimental spectrum. Free energies as discussed in the main text were calculated by using the following protocol: all geometries were prepared manually and pre-optimized with the semiempirical quantum chemical method GFN2-xTB/GBSA(CH_3CN).[167] The geometries were subsequently optimized with the composite method PBEh-3c[154] and the implicit solvation model DCOSMO-RS(NO_2CH_3) by using the Turbomole program package. On the optimized geometries free energies are calculated using a multilevel approach.[183] High level single-point energies were calculated with the hybrid density functional PBE0[287] in a large def2-QZVP[288] basis set. The (long-range) London dispersion not accounted for by the density functional was computed with the DFT-D4[149] method at default settings. Solvation contributions to free energy were calculated with COSMO-RS.[289,303] This term includes the volume work RTln(V_{ideal}) to go from an ideal gas at 1 bar to 1 mol L^{-1}. Thermostatistical contributions to free energy were calculated using GFN2-xTB/GBSA(CH_3CN) and the slightly modified RRHO-Scheme (interpolation between the rigid-rotor- to the harmonic-oscillator is applied at low-lying frequencies, see Ref. 183). The total free energies are then calculated as the sum of the single-point energy, dispersion contribution, thermostatistical and solvation contribution. The B3LYP/QZ-SMD[410] single-point calculations were performed with the ORCA 4.1.0 with convergence criteria for single point energies: 1.0×10^{-6} E_h. All DFT (except for the B3LYP[121,411] single-point) calculations were performed by using the TURBOMOLE 7.2.1 program package. The resolution-of-identity (RI) approximation for the Coulomb integrals[85] was generally applied using the matching default auxiliary basis sets.[86] The integration of the exchange-correlation contribution was evaluated on the numerical quadrature grids m4 for PBEh-3c and PBE0. The default convergence criteria for single-point energies were 10^{-7} E_h. For the conductor-like screening model for real solvents (COSMO-RS) free energy, two single-point calculations with BP86/TZ (one in gas-phase and one in an ideal conductor) have to be performed. The output of these calculations is then processed by the COSMOtherm program.[306,307]

Acknowledgments

Financial support from DFG (SFB 813—Chemistry at Spin Centers) and the Academy of Finland (K.R.: projects no. 263256, 292746 and 265328) is gratefully acknowledged. N.S. thanks Evonik Foundation for a doctoral scholarship and the DAAD for a travel grant. F.B. and S.G. gratefully acknowledge financial support by the DFG in the framework of the priority program No. SPP 1807 “Control of Dispersion Interactions in Chemistry”. G.S. thanks Prof. Dr. D. Menche and Prof. Dr. A. C. Filippou for providing X-ray equipment.

A5. Supporting Information to Chapter 4

Appendix A5 contains:

- Additional technical details of the calculations

Computational Details

Free energies of complex formation

complex	ΔE (PBE0/def2-QZVP)	ΔE_{disp} D4	ΔG_{RRHO} (GFN2-xTB/ GBSA(CH_3CN))$^{(a)}$	$\Delta\delta G_{solv}$ (COSMO-RS(NO_2CH_3))	$\Delta\delta G_a$
$[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{nit}\}_2]^{4+}$	-8.2	-37.5	56.8	-66.0	-54.9
$[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{nit}\}_3]^{6+}$	102.7	-64.8	89.7	-230.3	-102.7
$[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{iso}\}_2]^{4+}$	-41.4	-35.5	60.0	-70.5	-87.4
$[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{iso}\}_3]^{6+}$	44.4	-62.6	93.9	-237.6	-161.9

Table A5.1.: (a) Acetonitrile was used as solvent, since nitromethane is not parametrized for the implicit solvation model GBSA and the dielectric constant of acetonitrile (ϵ=36) is very close to the one of nitromethane (ϵ=38). All energies and free energies in kcal mol^{-1}.

- The complexes of Ligand $\mathbf{L}^{iso}$ are more stable than the complexes of Ligand $\mathbf{L}^{nit}$, revealing the stronger interaction of the isonitrile substituted paracyclophane $\mathbf{L}^{iso}$
- The comparison of nitrile ($\mathbf{L}^{nit}$) and isonitrile ($\mathbf{L}^{iso}$) complexes show that the largest difference in the contributions to free energy stems from the interaction energy (e.g. a difference of 58.3 kcal mol^{-1}in interaction energy for the trinuclear complexes), whereas all other contributions are quiet similar.

Free energies of complex formation calculated with a different functional and implicit solvation model

complex	ΔE (B3LYP/def2-QZVP)$^{(a)}$	ΔE_{disp} D4	ΔG_{RRHO} (GFN2-xTB/ GBSA(CH_3CN))$^{(a)}$	$\Delta\delta G_{solv}$ (SMD(NO_2CH_3)	$\Delta\delta G_a$
$[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{nit}\}_2]^{4+}$	3.3	-54.3	56.8	-43.3	-37.5
$[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{nit}\}_3]^{6+}$	124.4	-92.4	89.7	-191.9	-70.3
$[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{iso}\}_2]^{4+}$	-21.9	-50.3	60.0	-52.2	-64.4
$[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{iso}\}_3]^{6+}$	80.7	-88.0	93.9	-206.4	-119.8

Table A5.2.: (a) Energy obtained from single-point/SMD calculation, therefore, ΔE is not directly comparable to a gas phase calculation. (b) Acetonitrile was used as solvent, since nitromethane is not parametrized for the implicit solvation model GBSA and the dielectric constant of acetonitrile (ϵ=36) is very close to the one of nitromethane (ϵ=38). All energies and free energies in kcal mol^{-1}.

- The same trend: $\mathbf{L}^{nit}$ complexes being less attractive compared to $\mathbf{L}^{iso}$ complexes is observed with a different solvation model (SMD) and a different hybrid density functional.

Thermodynamic stability of di- and tri-nuclear complexes

complex	ΔG(GFN2-xTB/ GBSA(CH_3CN))$^{(a)(c)}$	$\Delta\Delta G^{(b)(c)}$ $(2\cdot M_3L_3 - 3\cdot M_2L_2)$	ΔG PBE0-D4/def2-QZVP (COSMO-RS(NO_2CH_3))$^{(a)}$	$\Delta\Delta G^{(b)}$ $(2\cdot M_3L_3 - 3\cdot M_2L_2)$
$[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{nit}\}_2]^{4+}$	-99.8	-26.3	-54.9	-40.9
$[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{nit}\}_3]^{6+}$	-162.9		-102.7	
$[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{iso}\}_2]^{4+}$	-149.3	-42.3	-87.4	-61.7
$[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{iso}\}_3]^{6+}$	-245.3		-161.9	

Table A5.3.: ((a) ΔG = formation reaction of the complexes. (b) $\Delta\Delta$G = $(2\cdot M_3L_3 - 3\cdot M_2L_2)$ reaction comparing thermodynamic stability of di- and trinuclear complexes. (c) Acetonitrile was used as solvent, since nitromethane is not parametrized for the implicit solvation model GBSA and the dielectric constant of acetonitrile (ϵ=36) is very close to the one of nitromethane (ϵ=38). All energies and free energies in kcal mol^{-1}.

- Thermodynamically the trinuclear complexes are preferred over the dinuclear ones. This does not reproduce the experimental observation and can most likely be explained with a strong solvent influence which is not well described by implicit solvation models.
- The semiempirical quantum chemical GFN2-xTB method is in good agreement to the high-level hybrid density functional theory data.

Explicitly solvated complexes (taking into account single solvent molecules in apical positions relative to the coordination plane of the metal ions

complex	ΔE (PBE0/def2-QZVP)	ΔE_{disp} D4	ΔG_{RRHO} (GFN2-xTB/ GBSA(CH_3CN))$^{(b)}$	$\Delta\delta G_{solv}$ (COSMO-RS(NO_2CH_3))	$\Delta\delta G_a$
$[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{nit}\}_2]^{4+}\cdot 2NO_2CH_3$	77.8	-26.2	26.1	-84.7	-7.6
$[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{nit}\}_3]^{6+}\cdot 3NO_2CH_3$	228.7	-44.6	45.9	-260.1	-30.1
$[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{iso}\}_2]^{4+}\cdot 2NO_2CH_3$	42.2	-24.9	29.9	-87.1	-40.0
$[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{iso}\}_3]^{6+}\cdot 3NO_2CH_3$	161.7	-43.7	50.7	-257.6	-88.9

Table A5.4.: (a) $2(M\cdot 2NO_2CH_3)$ + 2L $\rightarrow$ $M_2L_2\cdot$ 2 NO_2CH_3 + 2 NO_2CH_3 or 3(M·2 NO_2CH_3 + 3L $\rightarrow$ $M_3L_3\cdot$3 NO_2CH_3 + 3NO_2CH_3. (b) Acetonitrile was used as solvent, since nitromethane is not parametrized for the implicit solvation model GBSA and the dielectric constant of acetonitrile (ϵ=36) is very close to the one of nitromethane (ϵ=38). All energies and free energies in kcal mol^{-1}.

- Adding explicit nitromethane molecules to shield the Pd cations does not improve the results and the same trend is observed.

Thermodynamic stability of explicitly solvated di- and tri-nuclear complexes

complex	$\Delta\Delta$G(GFN2-xTB/GBSA(CH_3CN))$^{(a)}$ $(2\cdot M_3L_3 - 3\cdot M_2L_2)$	$\Delta\Delta$G(PBE0-D4/def2-QZVP /COSMO-RS(NO_2CH_3)) $(2\cdot M_3L_3 - 3\cdot M_2L_2)$
$[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{nit}\}_2]^{4+}\cdot 2NO_2CH_3$	-22.2	-37.4
$[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{nit}\}_3]^{6+}\cdot 3NO_2CH_3$		
$[(dppp)_2Pd_2\{(R_p)\text{-}\mathbf{L}^{iso}\}_2]^{4+}\cdot 2NO_2CH_3$	-51.9	-58.0
$[(dppp)_3Pd_3\{(R_p)\text{-}\mathbf{L}^{iso}\}_3]^{6+}\cdot 3NO_2CH_3$		

Table A5.5.: (a) Acetonitrile was used as solvent, since nitromethane is not parametrized for the implicit solvation model GBSA and the dielectric constant of acetonitrile (ϵ=36) is very close to the one of nitromethane (ϵ=38). All energies and free energies in kcal mol^{-1}.

Bond dissociation free energies of $[(\mathrm{dppp})_2\mathrm{Pd}]^{2+} \cdot 2$ solvent molecules

complex	ΔG GFN2-xTB/GBSA(CH_3CN)$^{(a)}$
$[(\mathrm{dppp})_2\mathrm{Pd} \cdot 2$ acetonitrile$]^{2+}$	57.9
$[(\mathrm{dppp})_2\mathrm{Pd} \cdot 2$ isocyanomethane$]^{2+}$	81.2
$[(\mathrm{dppp})_2\mathrm{Pd} \cdot 2$ nitromethane$]^{2+}$	51.3

Table A5.6.: The bond dissociation free energies are calculated by: $[(\mathrm{dppp})_2\mathrm{Pd} \cdot 2$ solvent molecules$]^{2+} \rightarrow [(\mathrm{dppp})_2\mathrm{Pd}^{2+} + 2$ solvent molecules. (a) Acetonitrile was used as solvent, since nitromethane is not parametrized for the implicit solvation model GBSA and the dielectric constant of acetonitrile (ϵ=36) is very close to the one of nitromethane (ϵ=38). All energies and free energies in kcal mol^{-1}.

- The bond dissociation free energies at GFN2-xTB/GBSA level of theory show the strong interaction of isonitrile substituted ligands with the palladium metal center.
- The nitrile metal interaction is slightly stronger compared to nitromethane, explaining why the complex formation with ligand $\mathbf{L}^{\mathrm{nit}}$ can occur in nitromethane and probably not in acetonitrile due to competition with the solvent.

Bending angles of $\mathbf{L}^{\mathrm{nit}}$ and $\mathbf{L}^{\mathrm{iso}}$ of undistorted and complex geometries

ligand	Mean (N-C-C) angle [°] ($\mathbf{L}^{\mathrm{nit}}$)	Mean (C-N-C) angle [°] ($\mathbf{L}^{\mathrm{iso}}$)
Free / undistorted	179.8	179.1
Taken from M_2L_2 complex	171.1	166.7
Taken from M_3L_3 complex	175.4	172.1

Table A5.7.: Ligand-deformation measured by the angle between the functional groups in $\mathbf{L}^{\mathrm{nit}}$ and $\mathbf{L}^{\mathrm{iso}}$.

Deformation energies of L^{nit} and L^{iso}

system	ΔE PBE0 -D4/QZ ($\mathbf{L}^{nit}$)	ΔE PBEh-3c ($\mathbf{L}^{nit}$)	ΔE GFN2-xTB/ GBSA(CH_3CN) ($\mathbf{L}^{nit}$)	ΔE PBE0-D4/ QZ ($\mathbf{L}^{iso}$)	ΔE PBEh-3c ($\mathbf{L}^{iso}$)	ΔE GFN2-xTB/ GBSA(CH_3CN) ($\mathbf{L}^{iso}$)
M_2L_2 dinuclear complex	-8.8	-9.6	-6.9	-11.1	-12.2	-9.4
M_3L_3 trinuclear complex	-1.9	-2.3	-1.0	-2.7	-3.1	-2.2

Table A5.8.: Acetonitrile was used as solvent, since nitromethane is not parametrized for the implicit solvation model GBSA and the dielectric constant of acetonitrile (ϵ=36) is very close to the one of nitromethane (ϵ=38). All energies and free energies kcal mol^{-1}. All energies were calculated on PBEh-3c/DCOSMO-RS geometries using the following approach: ΔE = E(free ligand) – E(ligand geometry taken from complex). The bending angles of $\mathbf{L}^{nit}$ and $\mathbf{L}^{iso}$ are listed in Table A5.7.

- From the plot in Figure A4.11 it is apparent that nitrile-groups are less flexible than isonitrile-groups and if distorted exhibit a larger ΔE_{strain} (strain energy).
- isonitrile ligands coordinate more strongly to the metal center, which is reflected in the [C-N-C] angles and larger strain/deformation energy while at the same time showing a more attractive interaction energy than nitrile-ligands.

Modelling of oligonuclear metallosupramolecular aggregates

PBEh-3c/DCOSMO-RS(NO_2CH_3) calculations as implemented in the TURBOMOLE.7.2.1 program package were performed to obtain possible ground state equilibrium geometries of the cationic units of the homo- and heterochiral dinuclear complexes of $\mathbf{L}^{nit}$ and the homo- and heterochiral trinuclear complexes of $\mathbf{L}^{iso}$. PBEh-3c is a global hybrid functional / composite method which provides reliable geometries and the implicit solvation method DCOSMO-RS is employed to be able to optimize the geometries in nitromethane solution.

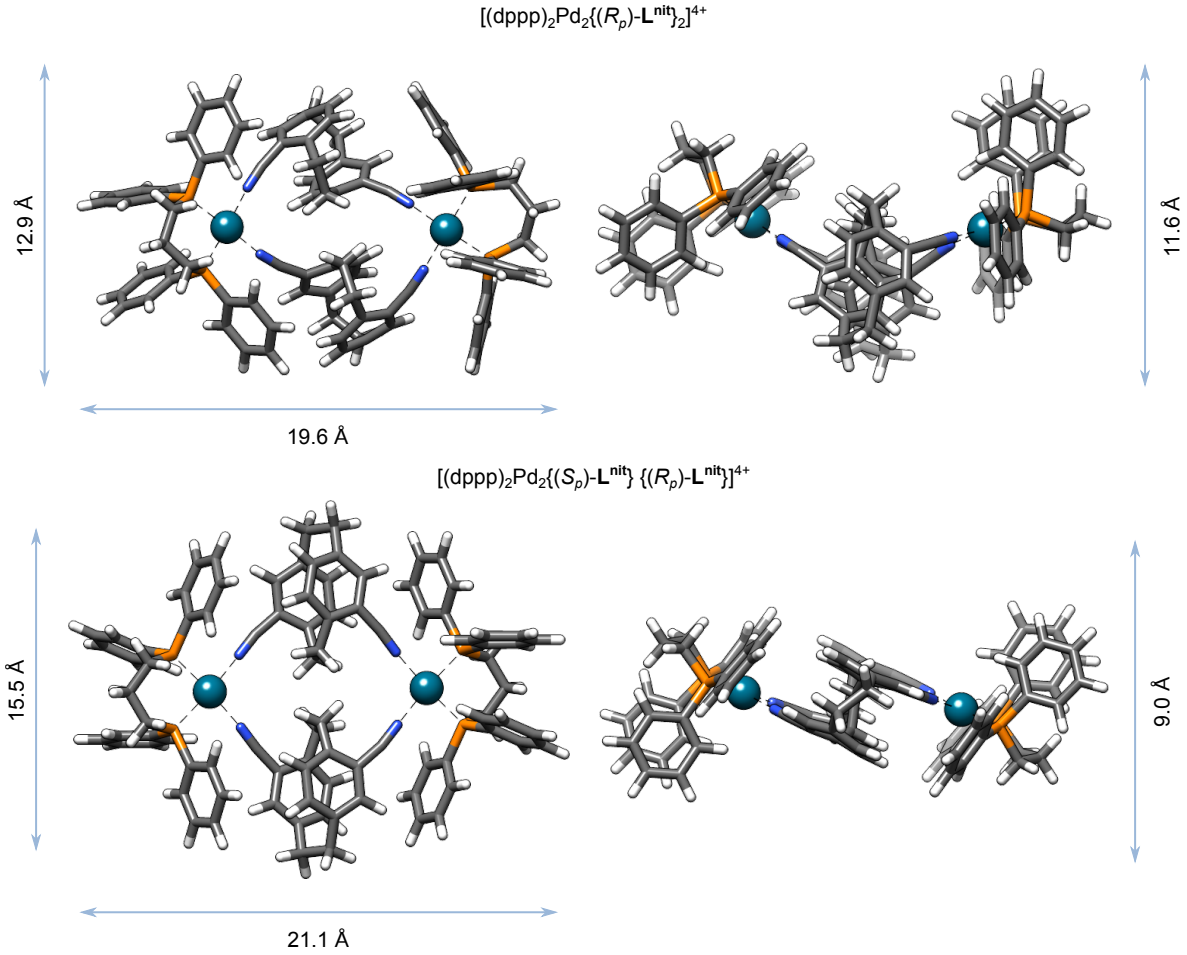


Figure A5.1.: Dinuclear $\mathbf{L}^{\mathrm{nit}}$ complexes.

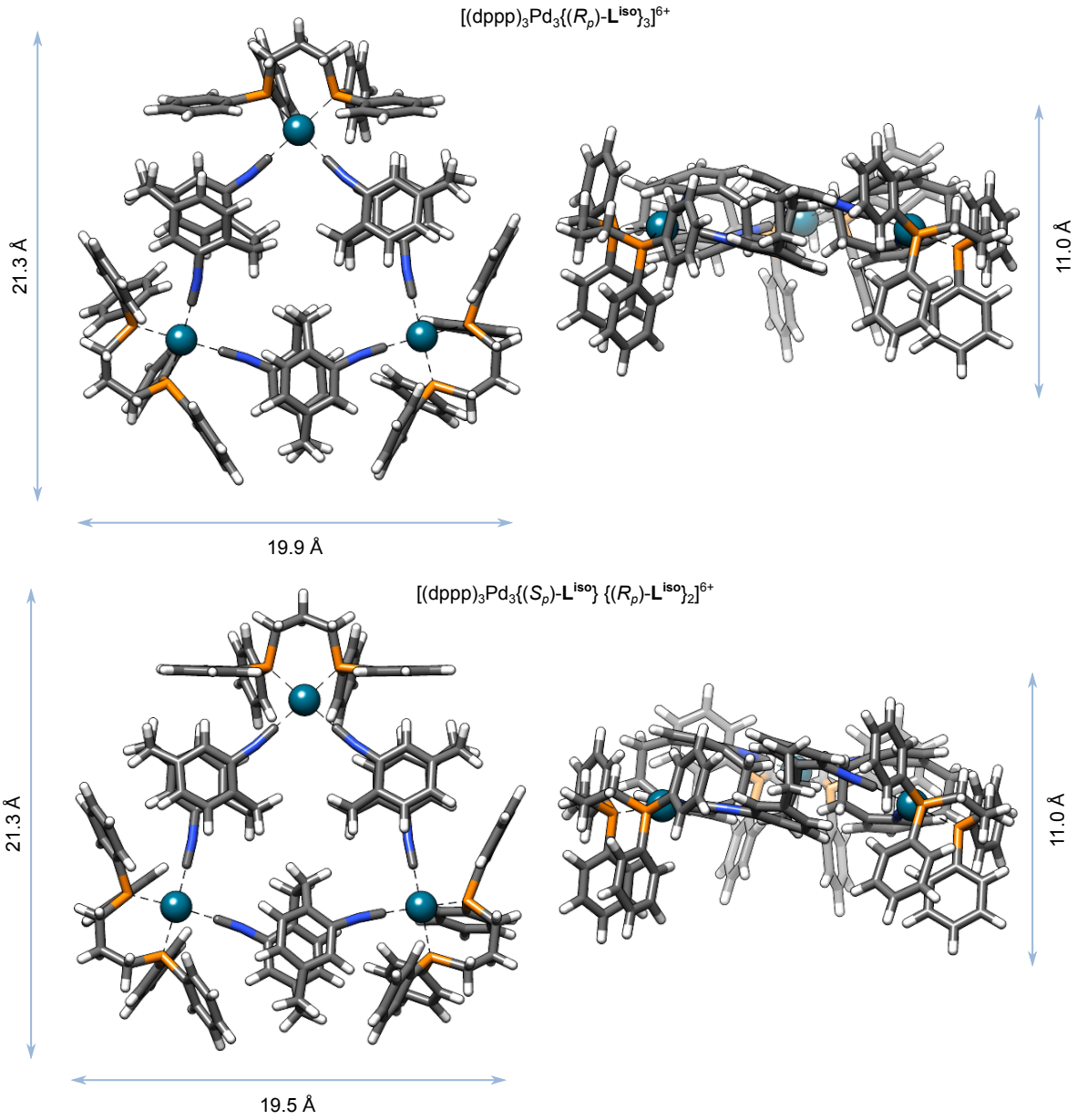


Figure A5.2.: Trinuclear $\mathbf{L}^{iso}$ complexes.

UV-Vis and ECD spectra

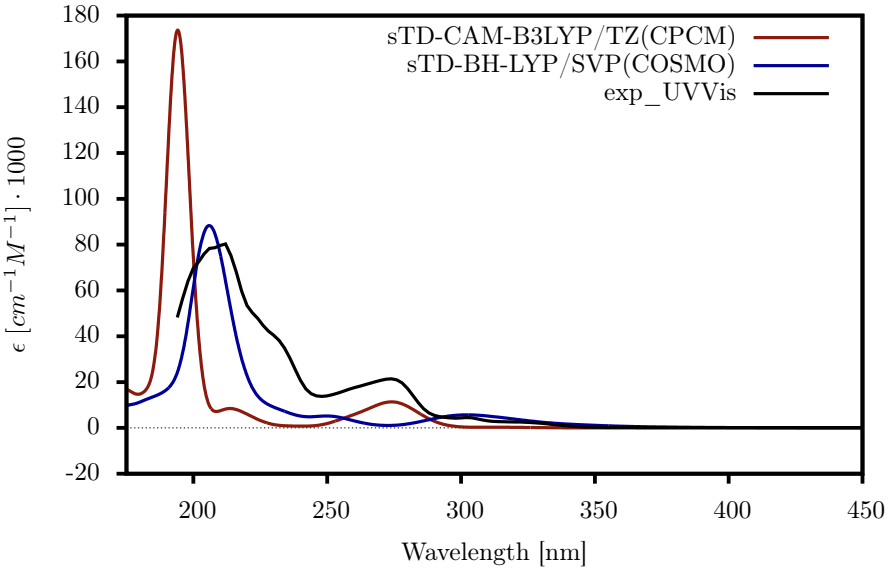


Figure A5.3.: Comparison of calculated and experimental UV-Vis spectra UV-Vis spectra of first eluting enantiomer of $\mathbf{L}^{\mathrm{nit}}$.

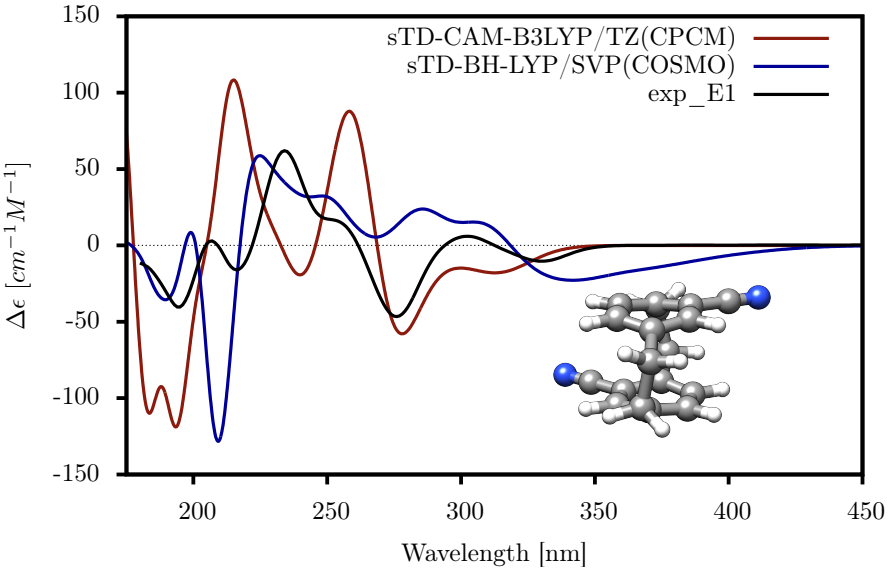


Figure A5.4.: Comparison of calculated and experimental ECD-spectra of (R_p)-$\mathbf{L}^{\mathrm{nit}}$. BH-LYP-D3/SVP(COSMO) shift = −0.4 eV, on 200 snapshots from GFN-xTB(GBSA)-MD; CAM-B3LYP/TZVP(CPCM) shift = −0.2 eV, on single PBEh-3c(COSMO) structure.

A6. Efficient structural and energetic screening of fullerene encapsulation in a large supramolecular double decker macrocycle

Abstract Recently, Tanaka *et al.* have synthesized an organometallic supramolecular double decker macrocycle for encapsulating fullerene C_{70}. We investigate this captivating system consisting of about 500 atoms with the robust quantum mechanical tight binding method GFN2-xTB and evaluate our computational results against an experimentally measured change in association free energy (ΔG_a). Further, GFN2-xTB was used to screen higher fullerene isomers and predict the best binding guest for this specific macrocycle.
Keywords: GFN2-xTB; fullerene; binding free energy.

A6.1. Introduction

Since their discovery in 1985 [412], fullerenes have sparked a wave of new developments in their application [413], as well as the search for understanding and generation of new supramolecular complexes [414]. In 2018 Tanaka *et al.* synthesized a supramolecular double decker cage macrocycle which is able to complexate fullerene C_{70} [415]. Selective purification of fullerenes in solution using macrocyclic supramolecular receptors represents a viable strategy to tackle the low solubility of the unpolar carbon allotropes [415]. Tanakas' receptor is based on two dibenzothiophene-salen Zn^{II}-macrocycles (**1**) (see Fig. A6.1). The double decker macrocyclic structure (**2**) is build up from the macrocycles (**1**) containing each four Zn^{II} metal ions and both macrocycles are joined by four nitrogen-containing pillar molecules (DABCO) connecting the square–pyramidal coordinated Zn cations. Having this large host molecule with a well-defined inner cavity at hand, the experimentalists measured the association of the 1:1 host-guest complex with C_{70} in $CHCl_3/CS_2$. The association constant (K_a) was evaluated by ^{1}H-NMR integration to be about 180 L·mol^{-1}, corresponding to an association free energy (ΔG_a) of −3.07 kcal·mol^{-1}. Further they investigated the complexation of C_{60} and found the receptor (**2**) to be able to selectively recognize C_{70} over C_{60} through higher association.

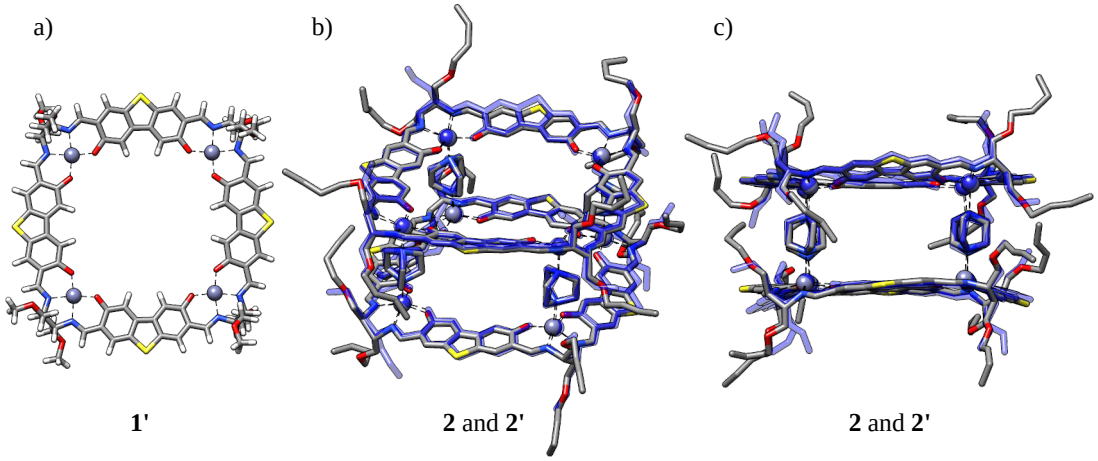


Figure A6.1.: a) Macrocycle **1'** geometry optimized with GFN2-xTB(GBSA); structures b) and c) show the overlay between crystal structure **2** (normal color code) and GFN2-xTB(GBSA) optimized geometry **2'** (blue color code); b) diagonal view; c) front view. In b) and c), hydrogen atoms are omitted for clarity

The large size of supramolecular complexes clearly limits the number of applicable theoretical methods able to calculate accurate association free energies. Our current approach for systems of about a few hundred atoms is consisting of a single-structure, multi-level computational protocol, heavily relying on dispersion corrected density functional theory[76,416]. In this paper, we are investigating the recently published quantum mechanical extended tight binding method GFN2-xTB[167] for routinely calculating binding free energies of large supramolecular complexes in solution. GFN2-xTB compared to its predecessor GFN1-xTB[166] is less empirical and more physically motived, *e.g.*, by including anisotropic second order density fluctuation effects. The robust method is available for all elements up to Z = 86 with target properties being: good geometries, (harmonic) vibrational frequencies and noncovalent interactions of large systems.[417] An implicit solvation model (GBSA) for describing reactions in solution is available for both GFN*n*-xTB methods. Although not specially parametrized for off target properties, they prove to be rather well described and the GFN*n*-xTB methods are currently used in the calculation of mass spectra (QCEIMS),[418] in systematic conformer ensemble generations (CREST)[419] and in finding reaction paths for mechanistic studies (GSM/GFN*n*-xTB)[420].

A6.2. Computational methods

The initial geometries of **1** and **2** were taken from CCDC[421] entries 1546762 and 1546764. All lower and higher fullerenes and their isomers employed in this study are following the "isolated pentagon rule" and were obtained from Tománek[422]. GFN2-xTB was used as implemented in the *xtb* code in the version 6.1. All GFN2-xTB calculations were performed with "verytight" settings and using the GBSA implicit solvation model with chloroform as the solvent. The temperature was set to 298.15 K for the thermostatistical evaluation. The intermolecular force field xTB-IFF[224] was used from its standalone code. B97-3c[155]/COSMO single-point calculations were performed as implemented within the TURBO-

MOLE.7.2.1 program package[305]. More detailed information is provided in the supplementary material to this paper (see Appendix A7).

A6.3. Results and discussion

First we were interested in how well the crystal structure of **2** at the level of GFN2-xTB(GBSA($CHCl_3$)) can be reproduced. For computational efficiency and to reduce the flexibility of the host molecule the R-O-*n*Bu side chains were truncated to R-OMe (**2'**), still retaining the structural properties of the original host molecule. The root mean square deviation (*RMSD*) between the crystal structure and **2'** GFN2-xTB(GBSA) optimized structure equals 0.77 Å. The overlay of both structures showing good correspondence is presented in Fig. A6.1. To investigate the complexes **2'**⊃C_{70} and **2'**⊃C_{60} we first need to obtain the geometry for the best intermolecular interaction. The GFN2-xTB(GBSA) optimized host and guest molecules are docked using the intermolecular force field xTB-IFF[224], which requires a QM potential input generated by GFN*n*-xTB for each monomer. The xTB-IFF-complex-geometry is subsequently reoptimized with GFN2-xTB(GBSA) to yield the final complex geometry. On all optimized geometries, free energies are calculated as the sum of the electronic energy (E) including the D4 dispersion correction[148,149] in a self-consistent formulation, thermostatistical corrections (G_{RRHO}) calculated from vibrational frequencies and the solvation contribution (δG_{solv}) calculated by the implicit solvation model GBSA:

$$\Delta G_a = \Delta E + \Delta G^T_{RRHO} + \Delta \delta G^T_{solv} \tag{A6.1}$$

The association free energy is calculated from the difference of the free energies from the complex, host and guest molecules (Table A6.1).

Table A6.1.: Binding free energies calculated with GFN2-xTB in comparison with experimental values. Bare interaction energies from GFN2-xTB, B97-3c(COSMO) and xTB-IFF are also given. All energies in kcal·mol^{-1}. [a] Calculated with GFN2-xTB(GBSA($CHCl_3$)); [b] taken from experiment; [c] calculated with COSMO $\epsilon = 4.8$.

Name	ΔE^a	ΔG^a_{RRHO}	$\Delta \delta G_{solv}$	ΔG^a_a	$\Delta G^b_{a,exp}$	ΔE B97-3c[c]	ΔE xTB-IFF
2'⊃C_{60}	-38.4	18.2	13.7	-6.4	–	-27.5	-42.4
2'⊃C_{70}	-43.9	19.3	17.3	-7.4	-3.07	-33.0	-50.7
1'⊃C_{60}	-31.1	18.0	9.8	-3.3	–	-13.9	-26.4
1'⊃C_{70}	-32.5	18.6	10.1	-3.8	–	-15.7	-28.2
2'⊃C_{78}	-50.0	19.5	19.8	-10.7	–	-38.1	-58.9

From Table A6.1 it is apparent that our calculated association free energy for **2'**⊃C_{70} is shifted compared to experiment by about 4 kcal·mol^{-1}, which can be partly attributed to the energy being calculated at tight binding level, and the implicit solvation model. For such a large complex and considering the small absolute value, the agreement is in fact satisfying. The next conclusion is that the qualitative ordering of calculated changes in free energies is correct in describing a higher association of C_{70} over

C_{60}. Looking at the contributions to the changes in binding free energy, the attractive contribution mainly stems from London dispersion incorporated in the electronic energy (ΔE) and increases with larger guest size. The thermostatistical correction to the free energy has a rather constant repulsive contribution, and the positive solvation contribution to free energy clearly increases with the guest size as well. The GFN2-xTB interaction energies are benchmarked against B97-3c/COSMO($CHCl_3$) DFT interaction energies. The good qualitative agreement between the semiempirical quantum chemical method and the computationally much more demanding density functional composite scheme is encouraging.

To gain more insight into the **2'**⊃C_{70} complex formation, a reaction path for the C_{70} inclusion process into the inner cavity of **2'** was calculated using the growing string method (GSM)[230] and GFN2-xTB(GBSA) shown in Fig. A6.2.

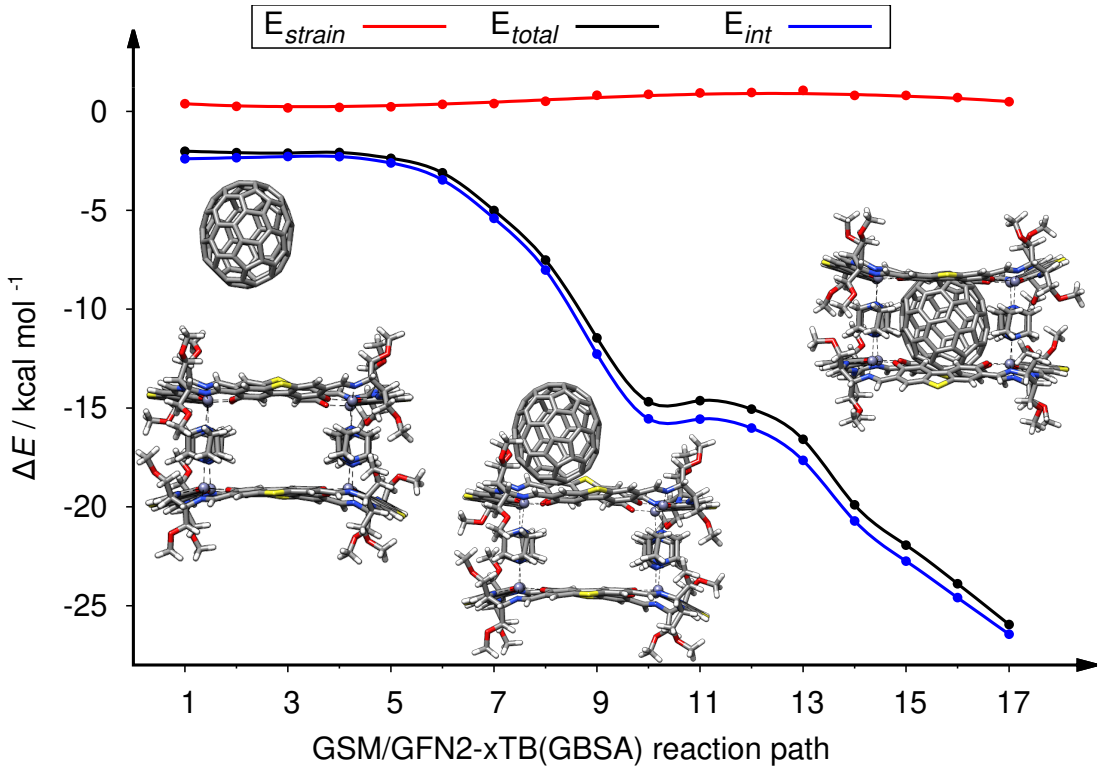


Figure A6.2.: GSM/GFN2-xTB(GBSA) reaction path of C_{70} entering the inner cavity of **2'**.

From the reaction path no transition state could be obtained, suggesting a barrierless reaction. The calculation of the reaction path, with 20 nodes along the string, took 146 h on a single computer node with 4 cores and 3.6 GHz. Dissecting the energy on the reaction path within the interaction strain model[231], shows almost no strain energy, *i.e.*, almost no molecular deformation and the reaction path is mainly described by the interaction energy of the fragments. Important to note is the "standing" orientation of the C_{70} inside the cavity allowing for multiple CH-π interactions of the inward facing protons, especially at the inner rim of the macrocycles **1'**. Knowing that C_{70} binds to the double decker macrocycle, we searched for the best binding higher fullerene using our fast GFN2-xTB method (Fig. A6.3).

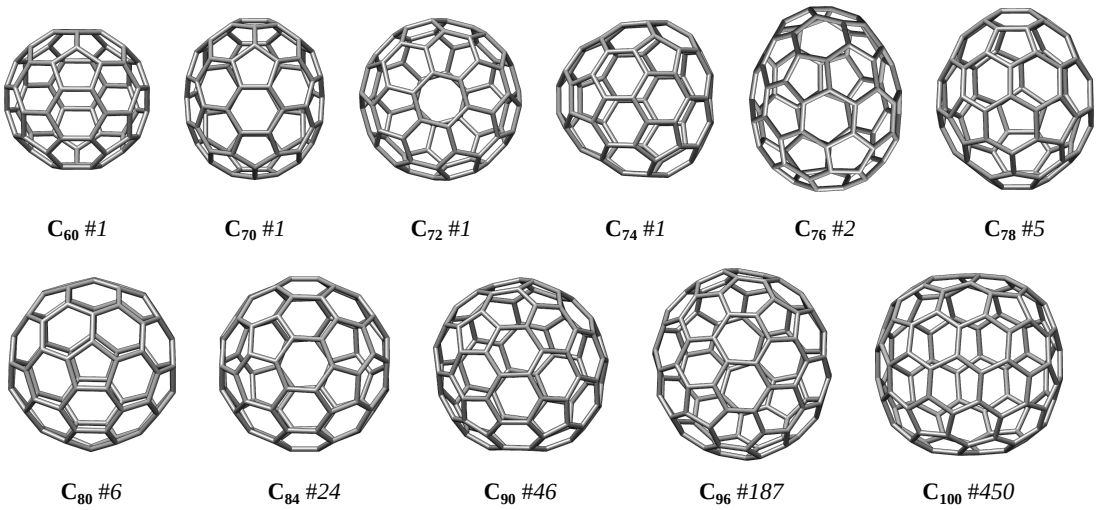


Figure A6.3.: Shown are lower and higher fullerene isomers having the lowest free energy determined by GFN2-xTB(GBSA), within the possible number (#) of isomers obeying the "isolated pentagon rule".

All 724 fullerene isomers were optimized and ranked by free energy obtained from GFN2-xTB(GBSA). The lowest lying isomer was docked with the macrocycle **1'** and the double decker **2'**. The geometries were optimized, and association free energies were calculated as described above. For all complexes up to C_{96} xTB-IFF determines the guest to be docked preferentially inside the cavity. For C_{96} and C_{100} the best guest position is determined to be outside the receptor and therefore these fullerenes were placed inside the cavity manually. Obtaining the free energy for the largest complex**2'**$\supset C_{100}$ took 18 h on 1 node with 14 cores (including optimization and hessian calculation). The resulting association free energies for **1'**$\supset C_x$ and **2'**$\supset C_x$ are shown in Fig. A6.4.

Analysing the associations of the fullerenes with the macrocycle **1'** it is apparent that the free energy increases with the contact surface, e.g., the London dispersion, of the higher fullerenes. The best binding guest within the double decker **2'** is fullerene C_{78} with an association free energy of −10.7 kcal·mol^{-1} . The C_{78} fullerene, in "standing" orientation, perfectly fills the space of the inner cavity and due to its ellipsoidal shape, similar to C_{70}, retains all CH-π interactions of the inward facing protons. For a comparison to C_{70} the contributions to free energy are listed in Table A6.1. Fullerenes with more carbon atoms than C_{78} are too large in size to fill the inner cavity of **2'** which is reflected in an increased Pauli repulsion and higher association free energy. The three-dimensional shape of the guest molecules in the double decker receptor has a large influence on the association which becomes apparent for C_{74} with D_{3h} symmetry.

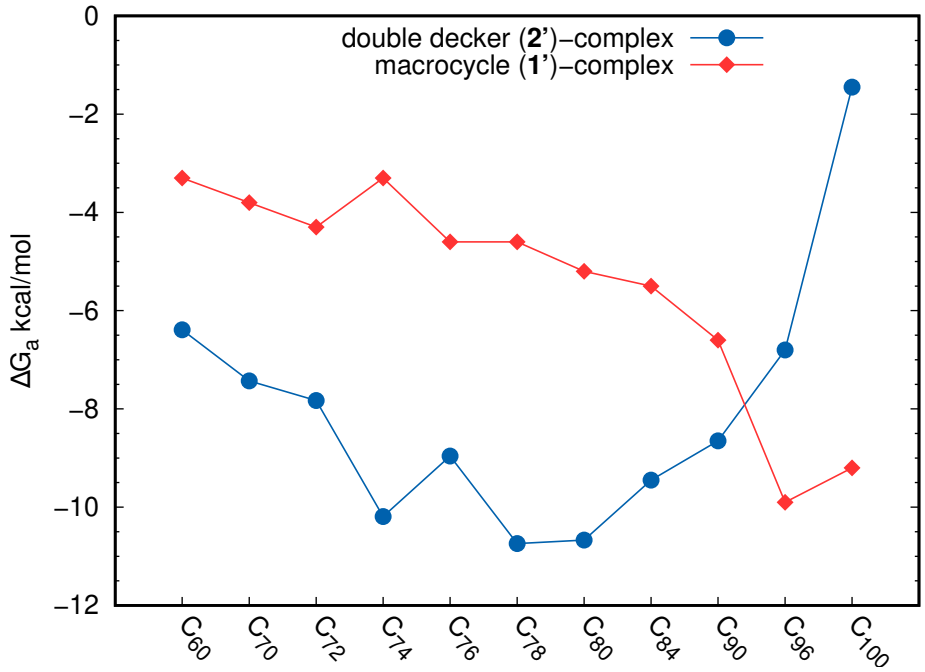


Figure A6.4.: Association free energies of most stable fullerene isomers with the host molecules **1'** and **2'** at GFN2-xTB(GBSA).

A6.4. Conclusion

In this work we demonstrate that GFN2-xTB(GBSA) is capable of reproducing the experimental trend in C_{70} binding to the double decker **2'** supramolecular host by higher association compared to C_{60}. Further we have determined C_{78} to be the best interacting guest for receptor **2'**. Qualitative estimations of association free energies of large supramolecular complexes and efficient screening of various host guest systems are easily possible with GFN2-xTB. The low computational cost of the method allows it to run on a normal consumer hardware highlighting its possibilities in screening applications.

Supplementary material

All geometries in *xmol* and additional computational details which are not mentioned within the manuscript are available electronically from http://www.shd.org.rs/JSCS/, or from the corresponding author on request (see Appendix A7).

Acknowledgments

The authors thank the DFG for funding in the framework of the priority Program No.SPP 1807 "Control of Dispersion Interactions in Chemistry". F. B. thanks Markus Bursch for fruitful discussions. The authors declare no conflict of interest.

A7. Supporting Information to Chapter 5

Appendix A7 contains:

- Additional technical details of the calculations
- Tables with detailed results

Supplementary Material

GFN2-xTB: Version 6.1 of GFN2-xTB[167] was used, as implemented in the *xtb* code. All GFN2-xTB calculations were carried out with the GBSA implicit solvation model and $CHCl_3$ as solvent. For all calculations (optimizations and hessian calculations) the "*very tight*" settings were used (verytight: E_{conv} (energy convergence) = $10^{-7}E_h$; G_{conv} (gradient convergence) = $2 \times 10^{-4}E_h \cdot \alpha^{-1}$; accuracy (for integral cutoffs and SCF criteria) = 0.05). Vibrational frequencies were calculated for thermostatistical correction to free energy and to verify that the optimized geometries are indeed minimum structures on the electronic potential hypersurface (check for no imaginary modes). The thermostatistical contributions to free energy are calculated in the rigid-rotor-harmonic-oscillator approach (RRHO)[183] and include zero point vibrational energies at 298.15 K. To reduce the error of the harmonic approximation for low-lying vibrational frequencies as well as numerical noise in the calculations, the RRHO-scheme (an interpolation between the rigid-rotor (RR)- to the harmonic oscillator (HO) is applied at low-lying frequencies (every vibrational mode below 50 cm^{-1}).

xTB-IFF: The intermolecular force field xTB-IFF[224] is implemented in a standalone code called *xtbiff* and is generated from "*low cost*" quantum mechanical (QM) input data (atomic partial charges, localized molecular orbitals and frontier orbital energies and densities have to be provided), here generated by GFN2-xTB(GBSA($CHCl_3$)). The QM data from each fragment is used to dock the intermolecular fragments and generate the best docking position. The docking itself is run in gas phase, but the provided QM input data is generated within the GBSA implicit solvation model. xTB-IFF returns the intermolecular interaction energy and several docking geometries sorted by their interaction energies. Since only rigid fragments are docked, the best xTB-IFF geometry has to be optimized once more by GFN2-xTB(GBSA).

B97-3c/COSMO reference calculations: B97-3c/COSMO($\epsilon = 4.8$)[155] reference single-point calculations were performed with the TURBOMOLE.7.2.1 program package.[305] The resolution-of-identity (RI) approximation for Coulomb integrals was generally applied[85] using the matching default auxiliary basis sets.[86] The integration of the exchange-correlation contribution was performed on the numerical

quadrature grids *m4*. The default convergence criteria for single-points [$10^{-7}E_h$] was applied. To be able to compare to the GFN2-xTB(GBSA) calculations the implicit solvent model COSMO was applied.

Reaction path calculation: The reaction path is calculated to get a good guess on the transition state structure. To start from reasonable geometries reactants and products are optimized with GFN2-xTB(GBSA($CHCl_3$)). The reaction path is then calculated with the growing string method GSM[230] using GFN2-xTB(GBSA($CHCl_3$)) as the underlying electronic structure method. The reaction path was calculated with 20 nodes on the reaction string. Geometries close to the estimated transition state were picked and hessians were calculated to find the imaginary mode of the transition state. On the examined reaction path no transition state could be obtained (no single imaginary mode was obtainable).

Fullerene isomer name within the manuscript	Fullerene name from Tomaneks'[422] ESI
C_{60}	C_{60}
C_{70}	C_{70}
C_{72}	C_{72-D6d}
C_{74}	C_{74-D3h}
C_{76}	C_{76-D2}
C_{78}	$C_{78-C2v-3}$
C_{80}	$C_{80-D5h-6}$
C_{84}	$C_{84-D2-22}$
C_{90}	$C_{90-C2-45}$
C_{96}	$C_{96-D2-183}$
C_{100}	$C_{100-D2-449}$

Table A7.1.: Lowest lying fullerene isomers determined by GFN2-xTB(GBSA($CHCl_3$) free energy

A8. Benchmark Study of Electrochemical Redox Potentials Calculated with Semiempirical and DFT Methods

Abstract: The calculation of redox potentials by semiempirical quantum mechanical (SQM) approaches is evaluated with a focus on the recently developed GFN*n*-xTB methods. The assessment is based on a data set comprising 313 experimental redox potentials of small to medium-sized organic and organometallic molecules in various solvents. This compilation is termed *ROP313* (Reduction and Oxidation Potentials 313) and divided for analysis purposes into the organic subset *OROP* and the organometallic subset *OMROP*. Corresponding data for a few common density functional theory (DFT) functionals employing extended AO basis-sets and small basis-set DFT composite schemes are computed for comparison. Continuum solvation models are used to calculate the important solvation free energy contribution. The results for *ROP313* show that the GFN*n*-xTB methods provide a robust, efficient and generally applicable workflow for the routine calculation of redox potentials. The GFN*n*-xTB methods outperform the PM*x* competitor for the *OROP* subset (mean absolute deviation from experiment, $MAD_{GFN2\text{-}xTB}$=0.30 V, $MAD_{GFN1\text{-}xTB}$=0.31 V, $MAD_{PM6\text{-}D3H4}$=0.61 V, MAD_{PM7}=0.60 V), almost reaching low-cost DFT quality ($MAD_{B97\text{-}3c}$=0.25 V) at drastically reduced computational cost (2–3 orders of magnitude). All SQM methods perform considerably worse for the *OMROP* subset. Here, the GFN2-xTB still yields reasonable, semiquantitative results slightly better and more robustly than with the PM*x* methods ($MAD_{GFN2\text{-}xTB}$=0.74 V, $MAD_{PM6\text{-}D3H4}$=0.78 V, MAD_{PM7}=0.82 V). The proposed workflow enables large-scale quantum chemical computations of organic and, to a lesser extent, organometallic molecule redox potentials on common laptop computers in seconds to minutes of computation time enabling, e.g., screening of extended compound libraries.

A8.1. Introduction

The redox potential is the key thermodynamic quantity whenever an exchange of electrons between different systems in the condensed phase occurs. In many areas of chemistry, redox potentials are a powerful tool for explaining thermodynamic and kinetic mechanisms, e.g., in photoredox catalysis[423] or organometallic synthesis[424]. Also, the necessity for efficient and environmental friendly energy storage capabilities has increased the interest in redox flow batteries, where knowledge of redox potentials for many molecules is required to identify appropriate electroactive systems[425,426].

Redox potentials are commonly measured by cyclic voltammetry[427]. However, in practice, irreversible half-reactions or highly reactive reagents can make direct measurements difficult or even un-

feasible.[428] Therefore, the calculation of redox potentials is crucial and represents a major challenge to *ab initio* computational chemistry. The accurate modeling of redox processes depends on the correct description of changes in the electronic structure of the target molecule in two oxidation states and their interactions with the solvent. For molecules in solution under equilibrium conditions (which are exclusively considered here), an automatic standard procedure that is based on a quantum mechanical treatment of the redox pair (solutes) combined with a dielectric continuum model of the solvent can be employed for these calculations.[428] For the computation of the solute's electronic energy, various theoretical methods are available. Elaborate wave function theory methods such as CCSD(T)[92] typically yield high accuracy for electronic energies but are not computationally feasible for larger molecules or in extensive screening studies.[429,430] Density functional theory[431] (DFT)-based methods, however, can be routinely applied for systems with a few hundred atoms and offer a good compromise between computational cost and accuracy. There are many examples of DFT computations of redox potentials for organic[432,433] and organometallic[434–436] compounds. Combined WFT/DFT schemes such as the connectivity-based hierarchy (CBH-Redox) model by Raghavachari et al.[437] provide cost effective alternatives to DFT.

Much faster than DFT are semiempirical quantum mechanical (SQM) methods,[59] such as GFN*n*-xTB,[166,167] PM*x*,[159,160] DFTB[164,168] and OM*x*.[438,439] Redox potentials can be calculated relying on SQM methods as demonstrated by the following example applications. DFTB methods were employed in quantum mechanical/molecular mechanical (QM/MM) approaches, where the solvent is sampled by means of molecular dynamics (MD) simulations and redox potentials are calculated via thermodynamic integration.[440–443] From the PM*x* methods, simple orbital based approximations for redox potentials[444,445] as well as workflows based on thermodynamic cycles are known.[446–448] Due to their low computational cost, SQM methods have been employed as tools for high-throughput virtual screening (HTVS)[426,449] in material design.[450,451]

Only a few SQM methods are parameterized for large parts of the periodic table such as the PM6/PM7 (up to Z=85, including lanthanoids *via* PM*x*/Sparkle[452,453]) and the GFN*n*-xTB (up to Z=86) methods. The efficiency, robustness and reasonable performance of the latter has been demonstrated in several applications, e.g., geometry optimizations of diverse lanthanoid[239] and large transition metal-complexes,[417] the screening of interaction energies of supramolecular complexes,[454] the simulation of electron ionization mass spectrometry,[418,455] the screening for ionization potentials (IPs) and electron affinities (EAs) of conjugated polymers,[451] the exploration of low-energy chemical space with metadynamics simulation[68,419] and the effective localization of transition states for complex organometallic reactions.[456] Recently, GFN1-xTB has also been employed in a redox potential calculation procedure, with treatment of explicit solvation.[457] The main advantage of employing SQM methods is a considerable reduction in computation time by at least 2 orders of magnitude compared to corresponding DFT treatments. This speedup enables screening for large compound libraries, e.g., to identify redox pairs suitable for organic aqueous redox flow batteries.[458]

In this work, we mainly investigate whether the good performance of the GFN2-xTB method for its target properties also applies to the calculation of redox potentials by a standard thermodynamic cycle and applying continuum solvation. For this purpose, a benchmark set with two subsets for organic and

organometallic molecules with experimental references is compiled. The performance of various SQM methods is evaluated comparatively with that of DFT methods.

A8.2. Methods

A8.2.1. Computational Details

All quantum chemical calculations were conducted applying the `xtb` 6.2[61,459] (GFN*n*-xTB), MOPAC2016[460] (PM*x*), and ORCA 4.1.2[408,461] (DFT) program packages. `xtb` was generally used as a driver for MOPAC calculations to achieve the best possible comparability of the SQM method results. For organic molecules, an automated conformer search applying GFN2-xTB was carried out for the respective oxidized species using the metadynamics-based iMTD-GC[419] algorithm as implemented in the `crest`[68,462] program. The lowest found conformer was used as initial structure for both species in the redox pair for all methods.

DFT geometry optimizations in the gas phase were conducted applying the B97-3c[155] composite scheme with default convergence criteria and the *SlowConv* keyword as implemented in ORCA. B97-3c is known to provide good geometries and harmonic frequencies at low computational cost, especially for organometallic complexes.[155,463] Final single-point energies were calculated on these geometries with the B97-D3,[141] PW6B95-D4,[234] and PWPB95-D4[233] density functional approximations (DFAs) employing Ahlrichs' large quadruple-ζ def2-QZVPP[288] basis set, *TightSCF*, *VerySlowConv*, and *Grid5 FinalGrid6* settings. The resolution of the identity approximation for Coulomb (RI-J)[85,86] and exchange (RI-JK)[464] integrals in conjunction with matching auxiliary basis sets (*def2/JK* option) was generally applied to speed up DFT calculations. For the MP2 part of PWPB95-D4, the RI approximation was employed using the *def2-QZVPP/C* auxiliary basis set and the frozen core approximation was enabled. The D4[148,149] London dispersion correction scheme was applied for PW6B95-D4 and PWPB95-D4.

GFN*n*-xTB and PM*x* geometry optimizations and energy calculations were conducted with default convergence criteria as implemented in `xtb`. B97-3c optimized geometries were verified as minima on the respective potential energy surface by the absence of imaginary modes larger $10\,i\mathrm{cm}^{-1}$ in the harmonic frequency calculation. Small imaginary frequencies were set real and included in the thermostatistical correction (G°_{RRHO}) obtained by a modified rigid-rotor-harmonic-oscillator approximation (RRHO)[183] at room temperature (T=298.15 K), including zero-point vibrational energy (ZPVE) corrections.

Solvation effects were considered by the implicit solvation models COSMO-RS[289,303] (DFT; `COSMOtherm`[306,307] Version C3.0, release 16.01, 2016 parameterization, parameter file: BP_TZVP_C30_1601.ctd; default Gsolv option), COSMO[242] (PM*x*; as implemented in MOPAC), and the generalized born (GB) model with solvent accessible surface area (SASA) termed GBSA,[61] as implemented in `xtb` for GFN*n*-xTB.

For open-shell systems, the electronic multiplicities (see SI) are supplied for PM*x* and DFT methods. In GFN*n*-xTB calculations the default Fermi smearing option was applied without providing explicit spin states, i.e., they always correspond to a low-spin electronic structure without inclusion of spin-

dependency in the Hamiltonian. This approach indirectly accounts, at least partly, for static correlation and significantly increases the robustness of the method and its black-box fashion applicability. An analogous approach would also be possible for the DFT methods. However, this is not part of this work but will be subject of an upcoming study in our lab. Manually changed α/β-shell occupations for GFNn-xTB did not improve the results (see SI for details). All calculations were performed on Intel® Xeon® E3-1270 v5 @ 3.60 GHz machines with four CPU cores.

A8.2.2. Calculation of Redox Potentials

The redox potential is calculated by the Nernst equation as

$$E^{\circ}_{R/O} = -\frac{\Delta_r G^{\circ}_{R/O}}{n_e F} - E^{\circ}_{ABS}(\mathrm{REF}), \tag{A8.1}$$

where $\Delta_r G^{\circ}_{R/O}$ is the free energy change of the investigated reduction or oxidation process at standard conditions (ideal gas, p=1 bar; ideal-solute solution, c=1 mol L^{-1}; T = 298 K), n_e is the number of electrons which are transferred in the reaction, F is the Faraday constant and the last term is the absolute potential of the reference electrode (REF).[232] Only one-electron transfer reactions were considered, i.e., n_e=1. Differences in the Gibbs free energy were calculated according to

$$\Delta_r G^{\circ}_{R/O} = IP + \Delta G^{\circ}_{RRHO} + \Delta\delta G^{\circ}_{solv}, \tag{A8.2}$$

where IP is the adiabatic electronic ionization potential of the reduced species and $\Delta\delta G^{\circ}_{solv}$ denotes the difference of the solvation free energies of reductant and oxidant. As COSMO only accounts for electrostatic solvation contributions, it yields a solvation energy correction E_{solv} compared to GBSA and COSMO-RS, which both yield solvation Gibbs free energy corrections G°_{solv}. However, because the molecular size and shape of the reductant and oxidant are rather similar, $\Delta\delta G^{\circ}_{solv}$ is dominated by the electrostatic term and the COSMO approximation should be reasonable. For the sake of simplicity, all solvation contributions are not further distinguished in the following. A comparison of the solvation contributions for selected compounds can be found in the SI. The performance of SQM methods in comparison to DFT for the computation of the ΔG°_{RRHO} term has recently been extensively examined[61,236] and found to be sufficent for many practical purposes.

The employed redox potential calculation protocol is based on a thermodynamic Born-Haber cycle, as shown in Figure A8.1.

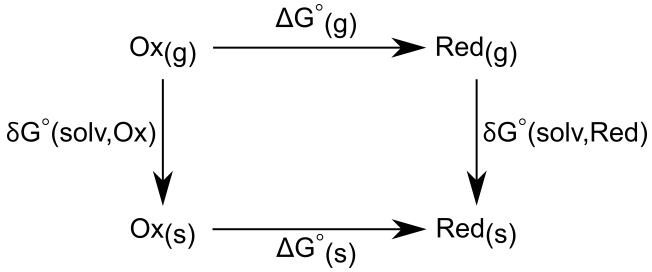


Figure A8.1.: Thermodynamic Born-Haber cycle for the calculation of redox potentials. Subscript (g) indicates that the compound is in gas phase and subscript (s) indicates the solvated state.

A8.2.3. Workflow

The proposed general workflow to calculate redox potentials is depicted schematically in Figure A8.2. It generally includes four steps: (i) geometry optimization of the reductant and oxidant, (ii) harmonic frequency calculations to obtain thermostatistical and ZPVE corrections, (iii) single-point energy calculations, and (iv) calculations of solvation corrections. The DFT and SQM method-based workflows differ in some points. As geometry optimizations and harmonic frequency calculations require a lot of computation time, the DFT approach facilitates a multilevel Ansatz, where both steps are conducted with the efficient B97-3c composite method, while the final single-point energy is computed at a higher level of theory. Optionally, the costly step (ii) in a DFT-only workflow could be replaced by a corresponding SQM calculation of the ΔG°_{RRHO} term, which is, however, not tested here further. The applied density functional approximations (DFAs) were chosen according to their excellent performance for IP/EA calculations in the GMTKN55[73] benchmark study (the smallest mean absolute deviation (MAD) of the respective functional class; the subscript denotes the corresponding benchmark name): B97-D3 (MAD_{G21IP}=0.15 eV, MAD_{G21EA}=0.10 eV), PW6B95-D4 (MAD_{G21IP}=0.12 eV, MAD_{G21EA}=0.06 eV), and PWPB95-D4 (MAD_{G21IP}=0.08 eV, MAD_{G21EA}=0.08 eV)). The solvation corrections in the DFT workflow are calculated employing the COSMO-RS approach.

For the SQM methods, all steps are conducted applying the same SQM level and including the corresponding implicit solvation models. For the GFN*n*-xTB tight binding methods an empirical IP/EA energy shift was employed for the final single-point energy to correct for the self-interaction energy present in the GFN*n*-xTB methods. These shifts are applied consistently to all redox potentials and were determined empirically during the original development of the GFN*n*-xTB methods to be 5.700 V for GFN1-xTB and 4.846 V for GFN2-xTB (*vide infra*). Thus, these shifts were not reoptimized for the present benchmark sets.

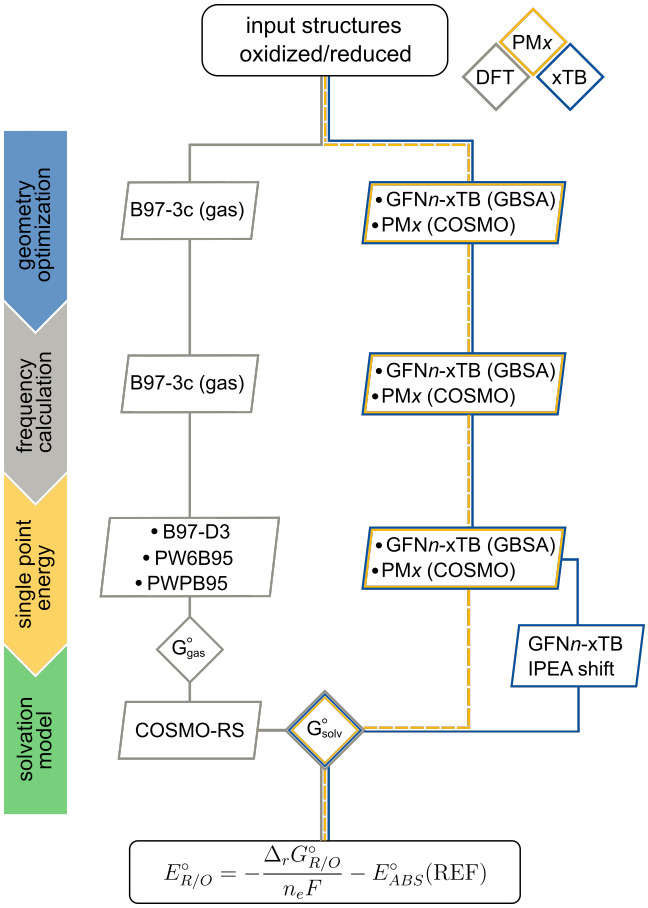


Figure A8.2.: Employed workflows for the calculation of redox potentials by DFT (grey), PM*x* (yellow) and GFN*n*-xTB (blue). Geometry optimizations, vibrational frequency, single point energy, and solvation free energy calculations are carried out for each of the oxidized and reduced species.

A8.3. Results and Discussion

A8.3.1. Ionization Potentials and Electron Affinities

In an initial investigation, the SQM methods were evaluated for the calculation of adiabatic ionization potentials (IPs) and electron affinities (EAs). On the G21IP and G21EA subsets of the GMTKN55 database, all tested SQM methods yield large errors (see SI for details), since the sets are composed mostly of atomic or small molecular systems, which are not the target of most SQM methods. Therefore, benchmark sets with larger, chemically more relevant systems (~ 15 atoms per molecule, 19 IPs, 15 EAs) are chosen for this evaluation. For IPs, the set from a study by Pantazis and co-workers[429] is employed while for EAs selected organic molecule data were taken from Ref. 465 (see SI for details). Note, that the empirical parameterization of the PM*x* methods included IP/EA data in the training sets, while this was not the case for GFN*n*-xTB. Hence, adjusting the IP/EA self-interaction error shift for the GFN*n*-xTB methods on this set of systems as described above and comparison of those (corrected) data

with corresponding (unmodified) PM*x* results is justified. Furthermore, Hartree-Fock-derived methods are not prone to comparable self-interaction errors and thus no shifting procedure was applied for PMx methods. A short derivation of the self-interaction error shift for the GFNn-xTB methods is shown in the SI. The performance of all SQM methods is shown graphically in Figure A8.3.

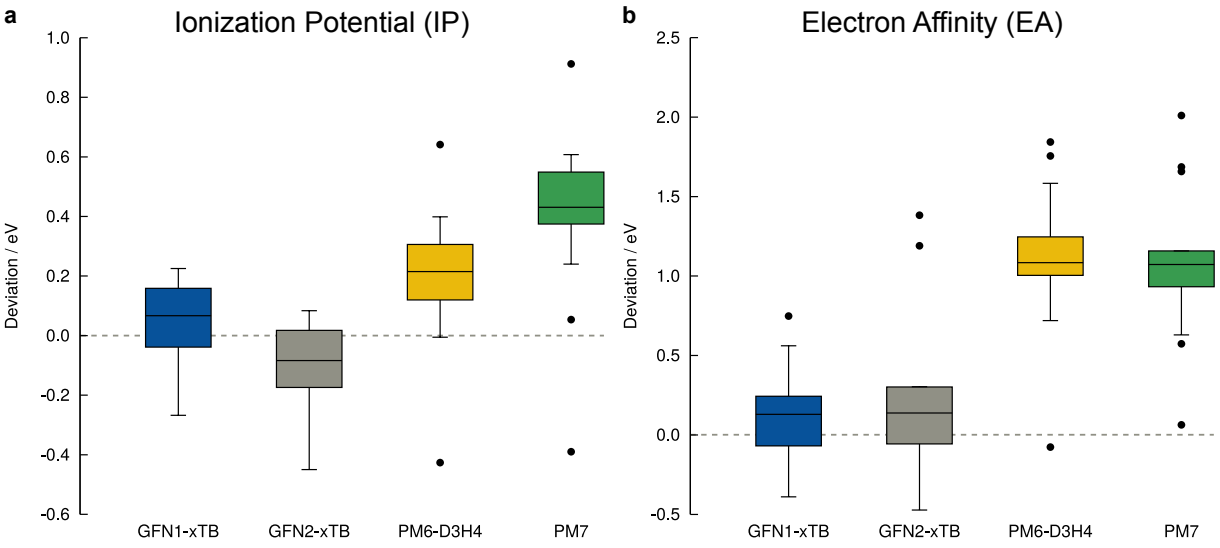


Figure A8.3.: Performance of SQM methods for the calculation of adiabatic ionization potentials (**a**) and adiabatic electron affinities (**b**) of organic compounds. The boxes contain the first and third quartiles, with a horizontal line at the median value. Whiskers extend from the ends of the box to the most distant point whose deviation lies within 1.5 times the interquartile range. Values outside of this range are treated as outliers and are displayed as black dots.

The IP/EA shift for GFN1-xTB and GFN2-xTB properly adjusts the absolute values and hence small mean (signed) deviations are obtained for both sets with these methods. The PM*x* methods overestimate the EAs on average ($MD_{PM6-D3H4}$ =1.09 eV, MD_{PM7}=1.07 eV), while the GFN*n*-xTB methods yield smaller systematic deviations ($MD_{GFN1-xTB}$=0.10 eV, $MD_{GFN2-xTB}$=0.20 eV) for the EA set. For IPs, the GFN*n*-xTB and PM*x* results are comparable, and both methods perform better. Generally, the GFN*n*-xTB methods yield fewer outliers. At this point we want to mention that the description of anionic systems with diffuse electron density is inherently difficult, especially with minimal AO basis sets, as employed by most SQM methods. For the GFN*n*-xTB methods this is presumably implicitly accounted for by the charge derivative of the Hubbard parameter[59,166,167] and also to some extent by the additional basis functions on some elements. Overall, the GFN*n*-xTB methods perform more robustly than PM*x* methods for organic IPs and EAs although neither of these properties was targeted in the parameter fit.

A8.3.2. Benchmark Study on Redox Potentials

A8.3.3. The *ROP313* Benchmark

To test the proposed computational protocol for calculating redox potentials (cf. Figure A8.2), a benchmark set was compiled from the literature hereinafter, referred to as *ROP313*. The set is divided into an organic (*OROP*) and organometallic (*OMROP*) subset, as summarized in Table A8.1. Overall, it includes 313 data points of experimentally determined redox potentials in four different solvents for small to medium-sized molecules (10 – 80 atoms).

Table A8.1.: Key characteristics of the *ROP313* benchmark set compilation. The multiplicity $M = 2S+1$, with S being the total electron spin and the considered range of molecular charges is also given. Both ranges include charges and multiplicities of reduced and oxidized species.

set	#systems	elements	solvents	range of charges	M	ref.
OROP	193	H, C, N, O, S, F, Cl, Br, I, Na, B	MeCN, DMF	−2 to +1	1	433, 466
OMROP	120	H, C, N, O, Si, Ti, V, Cr, Mn, Fe, Co, Ni, Os, Ru	MeCN, CH_2Cl_2, H_2O, DMF	−4 to +3	1 to 6	433, 466, 467

The first main data source (see Ref. 433) for the *ROP313* benchmark set covers more than 180 half-peak potentials of organic molecules measured in acetonitrile by cyclic voltammetry under consistent conditions. It comprises aromatic hydrocarbons, aryl acetylenes, aliphatic and aromatic olefins, and arenes bearing substituents containing heteroatoms, heterocycles, aldehydes, imines, oximes, hydrazones, ketones, amides, carboxylic acids, esters and nitriles. The second data source (Ref. 467) consists of 94 redox potentials of fourth-row transition metal complexes including 6 Ti, 7 V, 20 Cr, 12 Mn, 25 Fe, 18 Co, and 7 Ni complexes ranging from singlet to sextet electronic ground state multiplicity, and molecular charges from −4 to +3. The molecules from Ref. 467 are challenging test cases for all methods due to their high charges, varying multiplicities and partially difficult electronic structure. Further, 28 redox potentials of organic and organometallic molecules, including simple organic aromatic molecules, metallocenes, and complexes with bipyridine and CN ligands, were obtained from Ref. 466. All compound geometries, multiplicities, charges, and solvent information are provided in the SI.

A8.3.4. Redox Potentials

The performance of all considered SQM and DFT methods in reproducing experimental redox potentials was evaluated separately for the *OROP* and *OMROP* subsets. To get a first impression of the benchmark study, Figure A8.4 shows a selection of representative molecules from the *OROP* and *OMROP* subsets together with the respective redox potentials calculated at the GFN2-xTB and PWPB95-D4 level of theory.

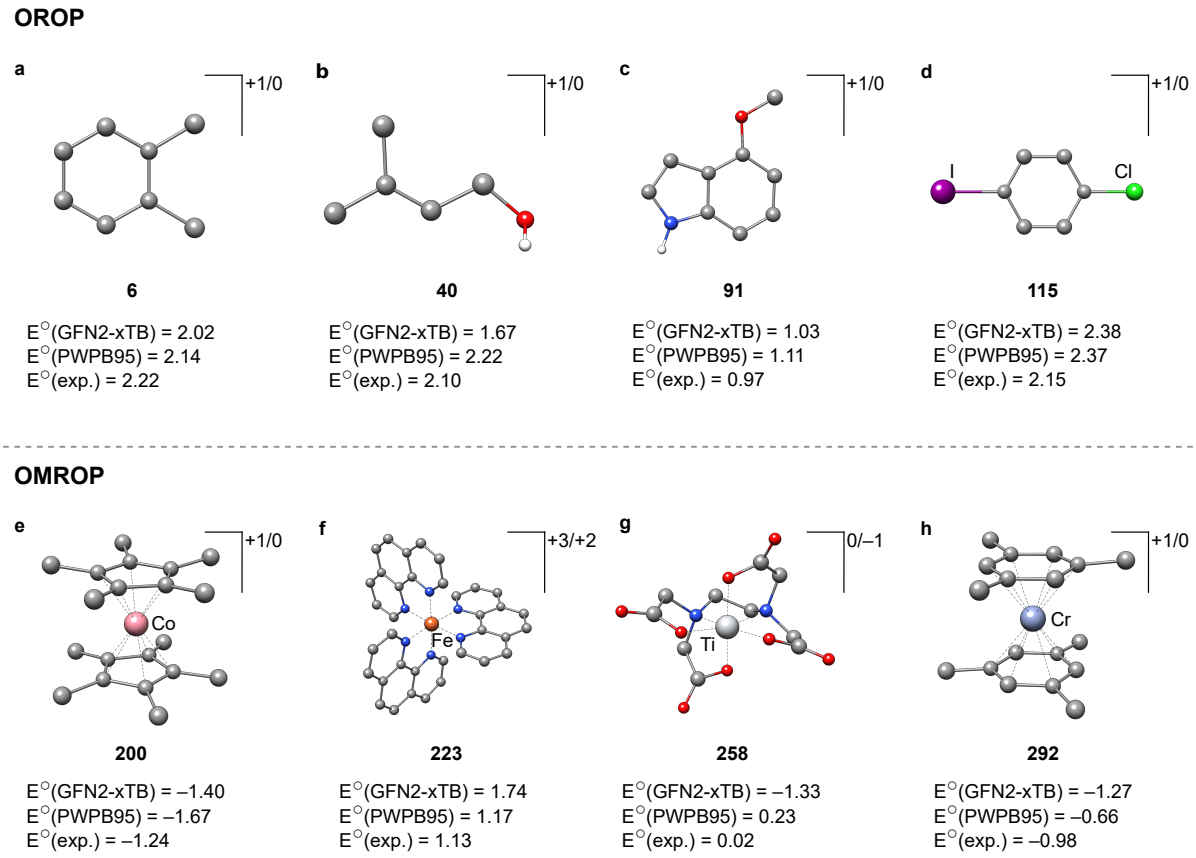


Figure A8.4.: Example of molecules included in the *ROP313* benchmark set with redox potentials in V obtained by DFT and GFN2-xTB as well as the corresponding experimental reference value. Hydrogen atoms bound to carbon are omitted for clarity.

Some redox potentials could not be calculated due to self-consistent field (SCF) convergence failures in the single-point calculations (see Tables A8.3 and A8.4) and hence, two separate statistical evaluations of the results were employed. First, an evaluation including all converged calculations for each tested method is given as indicated by the suffix “all” (e.g., 112 out of 120 calculations converged for B97-3c and were evaluated in the ”all” statistics). Another evaluation excludes all cases for which at least one of the respective SCF iterations did not converge for any of the tested methods thus comparing the same (smaller) number of systems for all methods (termed ”second evaluation” in Table A8.2). Calculated redox potentials were classified as unreasonable and excluded if they exceed a threshold of 4 V. This value is 1 V above the potential limit of the experimental setup with acetonitrile,[427] allowing smaller deviations to enter into the statistics while excluding cases where the results are far from being reasonable. The number of excluded data points and the respective number of redox potentials entering the statistical evaluations is summarized in Table A8.2. While hardly any data point had to be excluded for the organic systems, the tested SQM methods, except for GFN2-xTB, show a larger spread in calculated redox potentials for the organometallic systems, such that more data points are beyond the

physically reasonable 4 V limit and were therefore excluded. The poor convergence of the PW6B95-D4 and PWPB95-D4 single-point calculations for the organometallic systems is striking, which is the main reason for the relatively high number of excluded data points in the second evaluation of the OMROP set. The robustness of the GFN2-xTB method is remarkable, for which no data point has to be excluded in the "all" evaluation of both sets.

Table A8.2.: Number (#) of excluded data points and respective number of redox potentials entering the statistical evaluations.

Organic (OROP): # = 193				
method	# > 4 V	# not converged	# in"all" evaluation	# in second evaluation
B97-3c	1	0	192	182
B97-D3	1	0	191	182
PW6B95	2	4	186	182
PWPB95	2	1	189	182
GFN1-xTB	1	2	190	182
GFN2-xTB	0	0	193	182
PM6-D3H4	0	0	193	182
PM7	0	1	192	182
Organometallic (OMROP): # = 120				
method	# > 4 V	# not converged	# in"all" evaluation	# in second evaluation
B97-3c	0	8	112	57
B97-D3	1	0	111	57
PW6B95	1	22	89	57
PWPB95	2	20	90	57
GFN1-xTB	7	8	105	57
GFN2-xTB	0	0	120	57
PM6-D3H4	10	13	97	57
PM7	6	3	111	57

In total, 11 systems of the *OROP* and 63 of the *OMROP* subset, respectively, were excluded from the second evaluation. If not explicitly stated otherwise, the statistical measures for the number of data points given in the last column of the Table A8.2 are discussed in detail below.

Organic Subset (*OROP*)

For the organic subset *OROP* all tested DFAs yield mean absolute deviations (MADs) between 0.22 and 0.27 V (Figure A8.5a; Table A8.3) with the global metahybrid PW6B95-D4 yielding results of almost identical quality as the PWPB95-D4 double-hybrid functional. This and the fact that the MAD of these high-level methods is roughly doubled compared to the that for the IP/EA values in the gas phase, the GMTKN55 benchmark indicates that not the electronic structure part but the solvation contribution to the redox potential is limiting the achievable accuracy in theoretical predictions of the latter. The GGA functional B97-D3 yields slightly worse results with an MAD of 0.27 V but at significantly reduced

computational cost. With the clearly faster B97-3c composite scheme, an even slightly smaller MAD of 0.25 V is obtained. Moreover, the application of minimally augmented, diffuse basis sets was further tested but did not improve the results (see SI).

The GFN*n*-xTB methods yield slightly larger MADs of almost GGA quality with MADs of 0.31 V for GFN1-xTB and 0.30 V for GFN2-xTB (Figure A8.5b; Table A8.3). For the PM*x* methods, these statistical measures are twice as large ($MAD_{PM6\text{-}D3H4}$=0.61 V, MAD_{PM7}=0.60 V). Comparing the results for cationic systems (E_{exp} >0 V), the GFN*n*-xTB and PM*x* methods perform similarly, while for neutral systems (E_{exp} <0 V) the GFN*n*-xTB methods outperform the PM*x* methods, which yield systematically too positive redox potentials. This is in accordance with the observations for the calculations of IPs and EAs (*vide supra*) and indicates that the GFN*n*-xTB Hamiltonian provides a more flexible description of anionic systems, resulting from reduction of the neutral species. A simple shifting procedure as applied for the GFN*n*-xTB methods cannot adequately correct the large systematic error of the PM*x* methods, indicated by the slopes of the linear fit that strongly deviate from an ideal value of unity. Thus, even for an optimal shift specifically determined for the OROP set (0.35 V for PM7, 0.25 V for PM6-D3H4), only minor improvements can be achieved, reducing the MAD_{all}s to 0.50 V for PM7 and 0.58 V for PM6-D3H4. A charge-dependent correction or a linear-scaling approach may enhance the results but is beyond the scope of this work.

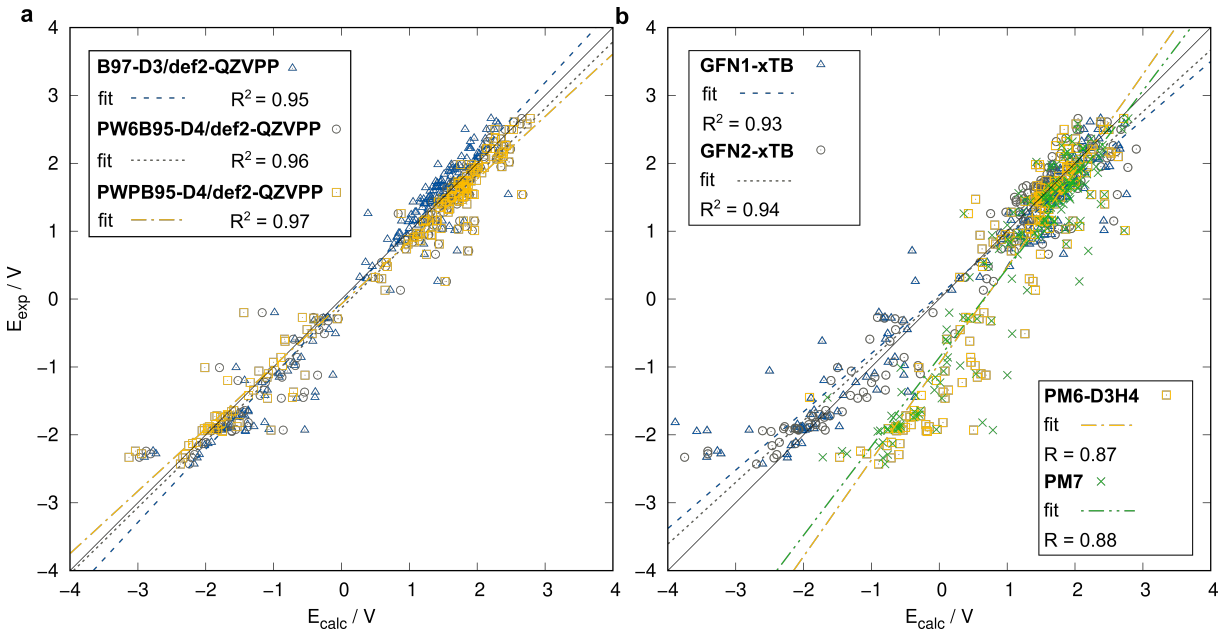


Figure A8.5.: Performance of DFT (**a**) and SQM methods (**b**) on the organic *OROP* data set with reference to experimental redox potentials. Determination coefficients R^2 are given for the corresponding linear fit.

Table A8.3.: Statistics (mean absolute deviation MAD, standard deviation SD, absolute maximum deviation AMAX and determination coefficient R^2) for all tested DFT and SQM methods for the organic *OROP* set (in V).

	Organic (OROP)									
N=182	MD	MAD	SD	AMAX	R^2	MD_{all}	MAD_{all}	SD_{all}	$AMAX_{all}$	$R^2{}_{all}$
B97-3c	-0.03	0.25	0.36	1.29	0.95	-0.03	0.26	0.37	1.29	0.94
B97-D3	0.00	0.27	0.36	1.39	0.95	0.01	0.27	0.36	1.39	0.95
PW6B95	0.15	0.23	0.30	1.29	0.96	0.15	0.23	0.30	1.29	0.96
PWPB95	0.13	0.22	0.31	1.27	0.97	0.13	0.22	0.30	1.27	0.97
GFN1-xTB	0.03	0.31	0.47	2.07	0.93	0.01	0.33	0.50	2.07	0.93
GFN2-xTB	0.02	0.30	0.39	1.42	0.94	0.02	0.30	0.41	1.66	0.94
PM6-D3H4	0.49	0.61	0.69	2.44	0.87	0.52	0.64	0.71	2.56	0.86
PM7	0.52	0.60	0.62	2.72	0.88	0.54	0.62	0.64	2.72	0.88

Apart from the electronic structure methods itself and basis-set effects, the treatment of solvation can play a crucial role for the accurate calculation of redox potentials. There are multiple examples for the calculation of organic redox potentials by DFT in the literature, applying different solvation models. Nicewicz et al. calculated redox potentials using B3LYP[121,411] and M06-2X[468] combined with the 6-31+G basis set and the CPCM solvation model for most of the molecules involved in the organic subset and obtained results similar to those from GFN2-xTB presented here ($MAD_{M06\text{-}2X}$=0.29 V, MAD_{B3LYP}=0.30 V).[433] Pantazis et al. reported similar MAD obtained with CCSD(T)/cc-pV[D/T]Z[92,469] extrapolated energies (MAD=0.28 V) for organic systems in water[429] and identified the solvation contribution as the dominant error source and accuracy limiting factor for the calculation of redox potentials. Therefore, an application of more sophisticated solvation models may improve the results accordingly. However, for the SQM methods the error of the electronic structure part is also significantly contributing. Furthermore, alternative solvation models like COSMO-RS are not fully developed for SQM and computationally more demanding than, e.g., GBSA, thus limiting their application in screening studies.

Organometallic subset (*OMROP*)

For the organometallic *OMROP* subset, all DFT methods yield MADs increased by 22% (B97-D3) to 64% (PWPB95-D4) compared to the *OROP* in a range of 0.33 to 0.37 V (A8.6a, Table A8.4). In contrast to the *OROP* results, B97-D3 outperforms the higher-level DFAs with a MAD of 0.33 V. PW6B95-D4 and PWPB95-D4 yield slightly worse MADs of 0.37 V and 0.36 V, respectively. Here, the triple-ζ basis set B97-3c composite method yields a slightly worse MAD (compared to B97-D3/QZ) of 0.37 V, indicating a moderate basis set dependence of the results which is more pronounced than for the organic set.

The SQM methods yield MADs that are more than twice as large as for assessed DFT methods (Figure A8.6b, Table A8.4), and therefore the linear regression between the calculated and experimentally determined redox potentials is not meaningful here. GFN2-xTB ($MAD_{GFN2\text{-}xTB}$=0.74 V) is still slightly outperforming the tested PMx methods ($MAD_{PM6\text{-}D3H4}$=0.78 V, MAD_{PM7}=0.82 V) and GFN1-

xTB ($MAD_{GFN1\text{-}xTB}$=0.91 V). When considering the "all" evaluation (see Tables A8.2 and A8.4), this difference becomes even more significant. Applying an uniform shift (-0.35 V for PM7, -0.10 V for PM6-D3H4) to the PM*x* methods optimized for the OMROP set (for GFN*n*-xTB no set dependent reevaluation of the applied shifts was applied) reduces the MAD_{all}s slightly to 0.96 V for PM7 and 1.02 V for PM6-D3H4. Generally, the large number of electronically more challenging systems (larger (negative) charges and higher multiplicities) contained in the *OMROP* subset manifests in drastically increased deviations compared to the organic *OROP* subset. The comparably good performance of GFN2-xTB, specifically compared with GNF1-xTB, may be attributed to its anisotropic electrostatic and XC terms not included in the latter, which together with the Fermi smearing is the key to the robust performance of GFN2-xTB also for more difficult systems.

Apart from the complex electronic structures included in the OMROP, two reasons can be suspected to contribute to the substantially worse performance of SQM methods in comparison to DFT for organometallic systems. The OMROP set includes highly negative charged (up to -4) molecules. Thus, as already observed for the IP/EA benchmark set, the minimal basis-sets applied in the SQM methods may not be sufficient for an adequate description of these systems. Furthermore, large charges pose a great challenge to the employed solvation models. It is likely that the more elaborate COSMO-RS model, employed in the DFT-based workflow, describes the important solvation effects better than the GBSA or the COSMO model used in the SQM methods. Improving the performance of implicit solvation models with GFNn-xTB methods is under development in our group.

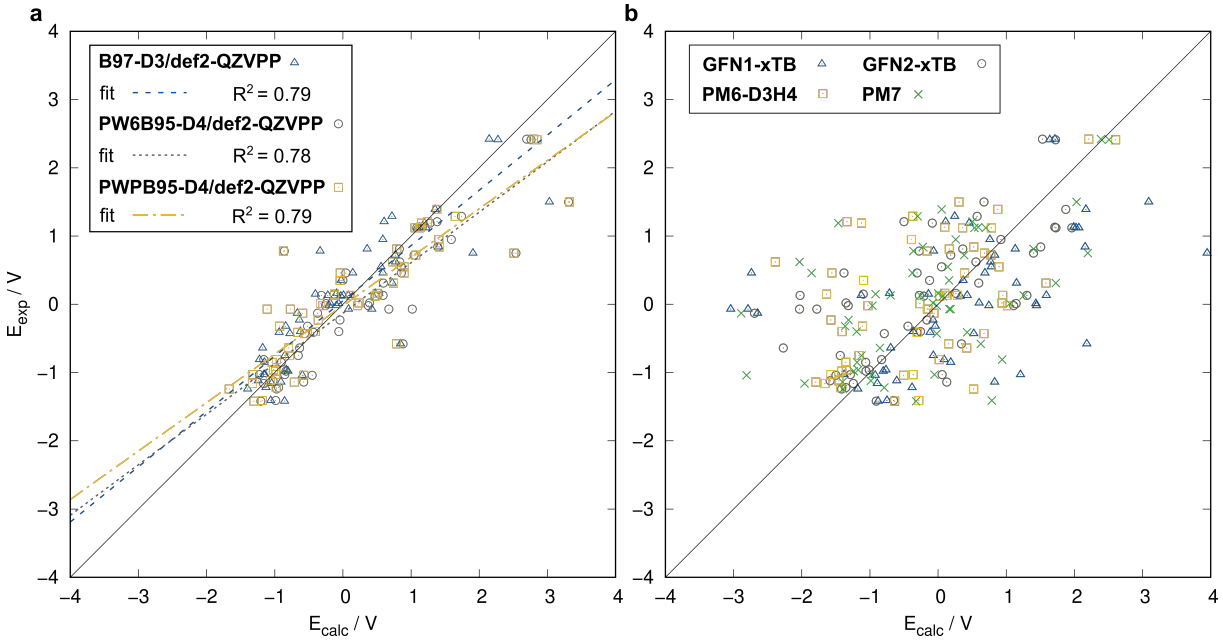


Figure A8.6.: Performance of DFT (**a**) and SQM methods (**b**) on the organometallic *OMROP* set with reference to experimental redox potentials. Determination coefficients R^2 are given for the corresponding linear fit.

Table A8.4.: Statistics (mean absolute deviation MAD, standard deviation SD, absolute maximum deviation AMAX and determination coefficient R^2) for all tested DFT and SQM methods for the organometallic *OMROP* set (in V).

Organometallic (OMROP)										
N=57	MD	MAD	SD	AMAX	R^2	MD_{all}	MAD_{all}	SD_{all}	$AMAX_{all}$	$R^2{}_{all}$
B97-3c	-0.13	0.37	0.48	1.35	0.80	-0.20	0.42	0.48	1.35	0.80
B97-D3	-0.04	0.33	0.47	1.53	0.79	-0.07	0.35	0.45	1.53	0.82
PW6B95	0.20	0.37	0.52	1.81	0.78	0.17	0.33	0.46	1.81	0.83
PWPB95	0.06	0.36	0.55	1.83	0.79	0.03	0.34	0.49	1.83	0.81
GFN1-xTB	0.22	0.91	1.23	3.20	0.26	-0.27	1.03	1.35	3.51	0.30
GFN2-xTB	-0.30	0.74	0.91	2.56	0.42	-0.45	0.73	0.83	2.56	0.53
PM6-D3H4	-0.30	0.78	0.99	3.01	0.26	-0.10	1.03	1.38	4.19	0.05
PM7	-0.27	0.82	1.06	2.76	0.28	-0.44	1.02	1.24	3.54	0.15

Timings

The computation time required for the calculation of redox potentials is analyzed using selected examples from the *OROP* and the *OMROP* subsets (Figure A8.7). The most time-demanding steps are the geometry optimization and the harmonic frequency calculation. As the viability of SQM methods for geometry optimization of large organometallic complexes was already discussed in detail in prior works (see e.g. Ref. 417), we will focus on the other steps. For both examples (compounds **183** and **197**), DFT frequency calculations and single point energy calculations at the chosen levels of theory take 2-3 orders of magnitude longer compared with the GFN*n*-xTB methods. The latter are further significantly faster (about 1 order of magnitude) compared to the PM*x* methods for the energy and frequency calculations. Moreover, the calculation of the solvation contribution with COSMO-RS in the DFT workflow (which is based on a DFT/TZ calculation) takes longer than the entire GFN*n*-xTB treatment including the frequency calculation. Overall, the application of SQM methods decreases the computation time drastically compared to DFT. Reducing the DFT computation times by the choice of a smaller basis-set (e.g. B97-3c) seems to be possible and is considered currently in a multi-level composite Ansatz in our lab.

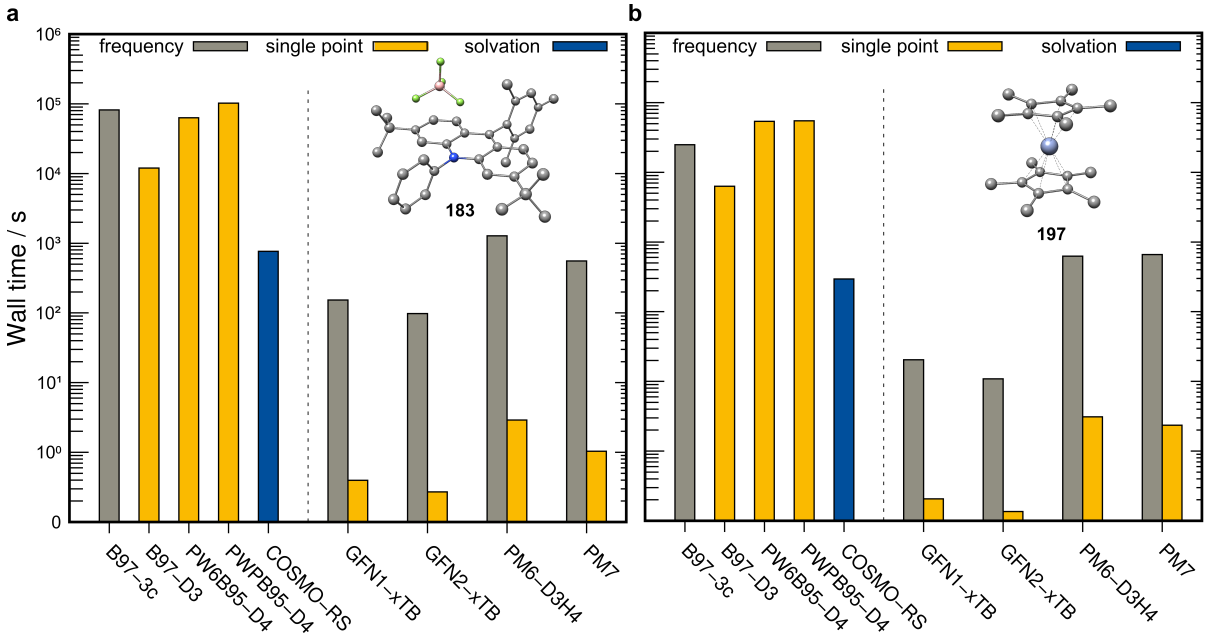


Figure A8.7.: Computational wall time in seconds (on a logarithmic scale) form a typical organic (**a**) and organometallic (**b**) molecule for the single point energy, harmonic frequency, and solvation contributions. SQM and DFT calculations were performed on four Intel® Xeon® E3-1270 v5 @ 3.60 GHz CPU cores.

A8.4. Conclusion and Outlook

A benchmark set termed *ROP313*, including 313 experimentally determined redox potentials of small to medium-sized organic (*OROP* subset) and organometallic (*OMROP* subset) molecules in different solvents, was compiled from the literature. This data set has been employed to asses the performance of SQM and selected DFT methods for the calculation of redox potentials of electrochemically relevant systems. In the organic *OROP* set, the best evaluated DFT method yields a low MAD of 0.22 V (PWPB95-D4/def2-QZVPP), while the semiempirical GFN2-xTB tight binding approach reaches comparable quality with an MAD of 0.30 V for GFN2-xTB (0.31 V for GFN1-xTB) at a fraction of the computational cost. Both tested GGA DFAs, B97-D3/def2-QZVPP and the low-level composite B97-3c method, perform very well with MAD values of only 0.25-0.27 V. Especially for the DFT methods, the majority of the remaining error can be attributed to an insufficient description of solvation effects by the employed implicit continuum models, which to some extent diminishes the advantages of employing more accurate quantum chemical methods. The analysis of the organic *OROP* results shows that GFN2-xTB clearly outperforms the Hartree-Fock based PM*x* methods ($MAD_{PM6\text{-}D3H4}$=0.61 V, MAD_{PM7}=0.60 V).

For the organometallic *OMROP* subset, DFT and SQM methods deviate more from the experimental reference data (increased MAD by a factor of 2-3). Consequently, a reliable yet efficient screening workflow for organometallic complexes with often difficult electronic structures, including high negative charges and high multiplicities, is not yet available at the SQM level of theory. Nevertheless, it is

shown that GFN2-xTB represents an improvement over the PM*x* methods in general. Note that investigating organometallic systems using DFT or the PM*x* methods usually requires individual, manual inspection of the spin states involved. The default application of Fermi smearing in the GFN*n*-xTB methods generally enables a black-box treatment of such systems and further increases the robustness of the method, specifically for GFN2-xTB. For the latter, all systems of the ROP313 could be converged to a meaningful value for the corresponding redox potential, while for the PM6-D3H4 method and especially for the PW6B95-D4 and PWPB95-D4 DFAs, which perform so well on the IP and EA benchmark sets of the GMTKN55 database, this is not the case (almost 20% of the 120 OMROP systems could not be meaningfully evaluated). Whether DFT methods, in particular the promising low-cost composite method B97-3c, can benefit from an analogous treatment will subject of an upcoming investigation in our lab.

Overall, even though the calculation of redox potentials is an off-target property for the GFN*n*-xTB methods, their performance on the *OROP* subset reveals that they are reliable and accurate, at least for organic molecules. Since GFN2-xTB performs slightly better than GFN1-xTB for the *OROP* set and significantly better for the *OMROP* set, we recommend GFN2-xTB as SQM method of choice for screening applications. Its low computational effort and robustness as well as its simple and general applicability establishes GFN2-xTB as an excellent SQM method for the reliable prediction of (organic) redox potentials, which is also applicable for large molecules and virtual high-throughput screening. The presented computational protocol may represent a useful tool whenever knowledge of large amounts of redox potentials is required, e.g., for the development of redox flow batteries[458] or organic electrocatalysts.

A8.5. Author Information

ORCID

Hagen Neugebauer: 0000-0003-1309-0503
Fabian Bohle: 0000-0003-2621-6339
Markus Bursch: 0000-0001-6711-5804
Andreas Hansen: 0000-0003-1659-8206
Stefan Grimme: 0000-0002-5844-4371

Notes

The authors declare no competing financial interest.

Acknowledgements

We thank S. Ehlert for steady support of the xtb code and fruitful discussions. We also thank S. Dohm for his helpful advice. The German Science Foundation (DFG) is gratefully acknowledged for financial support through a Gottfried Wilhelm Leibniz prize to S.G.

A9. Supporting Information to Chapter 6

Appendix A9 contains:

- Additional technical details of the calculations
- coordinates in xmol format

Statistical Means

- For property p (redox potential, IP, EA) with N observations:

$$\delta p = p_{calculated} - p_{reference} \tag{A9.1}$$

MD

$$MD = \frac{1}{N}\sum_{i}^{N} \delta p \tag{A9.2}$$

MAD

$$MAD = \frac{1}{N}\sum_{i}^{N} \mid \delta p \mid \tag{A9.3}$$

SD

$$SD = \sqrt{\frac{\sum_{i}^{N} \mid \delta p - MD \mid^2}{N-1}} \tag{A9.4}$$

Self-Interaction Error Shift

In the self-consistent extended tight-binding framework of the GFN*n*-xTB methods the single proton as well as the single electron have a self-energy. The energy of the proton $E(proton)$ can be calculated from the energy expression of GFN*n*-xTB according to equation A9.5 (see Ref. 166 and Ref. 167 for the full energy expressions).

$$E(proton) = \frac{1}{2}\eta_A(H) + \frac{1}{3}\Gamma_A(H). \tag{A9.5}$$

Here, $\eta_A(H)$ is the chemical hardness for hydrogen and $\Gamma_A(H)$ is the Hubbard parameter for hydrogen, both taken from the corresponding GFN*n*-xTB parameterization. The self-energy of the electron $E(electron)$ is then calculated by subtracting the energy of the proton from the ionization potential of hydrogen $IP(H)$, also taken from the GFN*n*-xTB parameterization:

$$E(electron) = IP(H) - E(proton) \tag{A9.6}$$

The theoretical calculated self energy of the electron for GFN2-xTB is 4.461 eV. The empirical shift of 4.846 eV was evaluated during the development of GFN*n*-xTB methods and performed better than the one obtained from theory for the calculation of ionization potentials.

Ionization Potentials and Electron Affinities

In the following all energies are given in eV.

G21IP and G21EA Set

Table A9.1.: G21IP subset of the GMTKN55 database.[73]

G21IP	GFN1-xTB	GFN2-xTB	PM6-D3H4	PM7	Ref.
h	5.22	5.86	-2.26	-2.26	13.66
li	5.30	4.57	2.72	4.80	5.35
be	8.39	12.57	8.81	7.84	9.32
b	7.84	12.83	5.78	4.04	8.26
c	11.53	22.34	8.76	8.22	11.26
n	13.86	14.06	14.33	14.09	14.54
o	20.47	17.54	9.57	10.54	13.61
f	22.88	19.99	17.55	17.25	17.42
na	2.36	5.02	4.54	5.82	5.14
mg	10.09	8.32	7.46	1.17	7.65
al	2.05	6.43	5.86	7.51	5.98
si	5.65	6.52	7.17	6.75	8.15
p	13.11	9.10	9.71	10.96	10.49
s	15.12	10.48	9.54	9.30	10.36
cl	11.99	13.57	11.29	11.61	12.97
IP59	12.47	11.80	11.73	11.77	12.85
IP60	10.75	11.08	9.06	9.30	10.22
IP61	17.35	15.34	10.41	11.11	13.05
IP62	14.81	13.33	11.40	11.83	12.69
IP63	16.80	15.56	15.23	15.39	16.10
IP64	11.74	10.91	9.66	8.92	11.07
IP65	10.71	8.54	9.49	10.71	10.15
IP66	9.70	8.30	9.26	10.43	9.82
IP67	10.36	9.33	8.92	9.91	9.88
IP68	12.33	10.42	9.08	9.29	10.38
IP70	11.40	11.93	11.57	11.71	12.77
IP71	11.66	11.96	11.20	11.22	11.47
IP72	10.61	10.40	10.05	10.11	10.57
IP73	13.64	13.68	11.76	12.09	14.01
IP74	14.05	13.22	12.70	13.49	15.58
IP75	13.91	12.01	10.44	10.31	12.04
IP76	11.53	8.84	10.69	10.69	10.53
IP77	10.37	8.61	8.93	8.57	9.36
IP78	9.44	11.02	10.72	11.25	11.50
IP79	11.33	12.89	11.47	11.63	12.65
IP80	11.55	11.08	9.73	9.86	11.32

Table A9.2.: G21EA subset of the GMTKN55 database.[73]

G21EA	GFN1-xTB	GFN2-xTB	PM6-D3H4	PM7	Ref.
EAc	-0.96	-1.47	-0.08	-0.63	1.27
EAo	3.95	3.35	0.88	0.99	1.46
EAf	3.15	3.05	3.49	3.36	3.40
EAsi	0.74	-4.82	2.61	-1.54	1.39
EAp	-3.02	1.39	-6.30	3.11	0.75
EAs	0.49	2.18	2.67	2.65	2.08
EAcl	4.14	3.39	3.87	3.84	3.62
EA8n	-0.90	-1.27	0.07	-0.33	1.21
EA9n	0.11	-0.13	-1.09	-0.97	0.58
EA10n	-0.49	-0.65	-1.33	-1.14	0.05
EA11n	0.01	0.88	-0.14	0.08	0.36
EA12n	-0.52	0.45	0.00	0.22	0.73
EA13n	2.26	2.05	1.91	1.69	1.81
EA14n	1.07	-2.90	2.39	0.81	1.27
EA15n	19.65	-2.17	2.12	1.14	1.09
EA16n	118.26	-0.64	1.81	-0.26	1.36
EA17n	-2.38	1.29	-3.02	3.84	1.02
EA18n	-1.90	1.21	0.68	4.63	1.25
EA19n	0.33	1.88	2.60	2.68	2.35
EA20n	2.47	2.32	-0.41	-0.26	0.41
EA21n	-0.61	-0.06	1.36	1.06	-0.01
EA22n	0.91	1.07	2.81	3.11	3.88
EA23n	-1.67	0.42	-3.70	4.03	1.08
EA24n	0.09	2.09	2.46	2.42	1.65
EA25n	1.15	1.75	2.62	2.46	2.37

Table A9.3.: Statistic metrics for the G21IP/EA sets (red colored value from table A9.2 removed).

G21IP	GFN1-xTB	GFN2-xTB	PM6-D3H4	PM7
MD	0.28	0.20	-1.61	-1.53
MAD	1.83	1.57	1.62	1.75
SD	2.71	2.76	2.63	2.87
AMAX	8.43	11.08	15.92	15.92
G21EA	GFN1	GFN2	PM6-D3H4	PM7
MD	0.29	-0.87	-0.73	0.02
MAD	2.30	1.35	1.33	1.15
SD	4.32	1.87	1.98	1.56
AMAX	18.56	6.22	7.05	3.38

SQM IPEA Test Set

Table A9.4.: Ionization potentials of organic compounds (see Ref. 429)

Compound	GFN1-xTB	GFN2-xTB	PM6-D3H4	PM7	Reference
aniline	7.81	7.75	7.94	8.13	7.72
p-methoxyaniline	749.76	7.04	7.52	7.70	7.12
p-methylaniline	7.48	7.33	7.71	7.91	7.46
p-chloroaniline	7.52	7.60	8.05	8.28	7.74
n-methylaniline	7.48	7.39	7.51	7.70	7.32
dimethylamine	8.44	8.18	8.36	8.65	8.24
diethylamine	8.10	7.84	8.21	8.44	8.01
pyrrolidine	8.23	7.95	8.64	8.91	8.00
piperidine	8.01	7.80	8.32	8.57	8.20
anisole	8.31	8.22	8.41	8.59	8.20
indole	7.76	7.78	7.84	8.00	7.76
phenol	8.65	8.57	8.77	8.96	8.49
p-methoxyphenol	7.81	7.69	8.09	8.26	7.71
p-methylphenol	8.22	8.04	8.45	8.64	8.17
p-chlorophenol	8.17	8.29	8.79	9.05	8.44
p-cyanophenol	8.79	8.93	9.18	9.36	9.01
dimethylsulfide	8.89	8.30	8.26	8.30	8.69
dimethyldisulfide	8.19	7.73	8.17	8.77	8.18
thioanisole	7.88	7.70	7.95	7.97	7.92

Table A9.5.: Statistics for the ionization potential set (red colored value from table A9.4 removed).

	GFN1-xTB	GFN2-xTB	PM6-D3H4	PM7
MD	0.03	-0.12	0.20	0.41
MAD	0.13	0.14	0.25	0.45
SD	0.16	0.16	0.21	0.26
AMAX	0.27	0.45	0.64	0.91

Table A9.6.: Electron affinities of organic compounds.[465] Atomic Cartesian coordinates of the corresponding structures are given separately as .trj file.

Compound	GFN1-xTB	GFN2-xTB	PM6-D3H4	PM7	Reference
nitrobenzene	1.00	1.18	2.11	1.94	1.00
hexafluorobenzene	0.13	0.82	2.10	2.18	0.52
2-methyl-nitrobenzene	0.95	1.06	2.01	1.85	0.92
3-nitroanisole	0.99	1.20	2.20	2.01	1.04
azulene	0.48	0.32	1.80	1.87	0.79
cyclooctatetraene	0.48	0.28	0.47	0.61	0.55
anthracene	0.66	0.47	1.61	1.69	0.53
1-nitronaphthalene	1.45	1.53	2.48	2.30	1.23
decafluorobiphenyl	1.57	2.20	2.66	2.83	0.82
perylene	1.13	0.94	2.00	2.02	0.97
pentacene	1.61	1.43	2.36	2.43	1.35
tetracene	1.22	1.04	2.04	2.12	1.04
pentafluoro-nitrobenzene	2.08	2.71	3.28	3.21	1.52
1,4-naphthoquinone	2.05	2.02	2.53	2.38	1.81
3,5-dimethyl-nitrobenzene	0.94	1.05	2.01	1.84	1.21

Table A9.7.: Statistic for the electron affinity set.

	GFN1-xTB	GFN2-xTB	PM6-D3H4	PM7
MD	0.10	0.20	1.09	1.07
MAD	0.24	0.33	1.10	1.07
SD	0.31	0.49	0.46	0.47
AMAX	0.75	1.38	1.84	2.01

Comparison of Selected Solvation Contributions

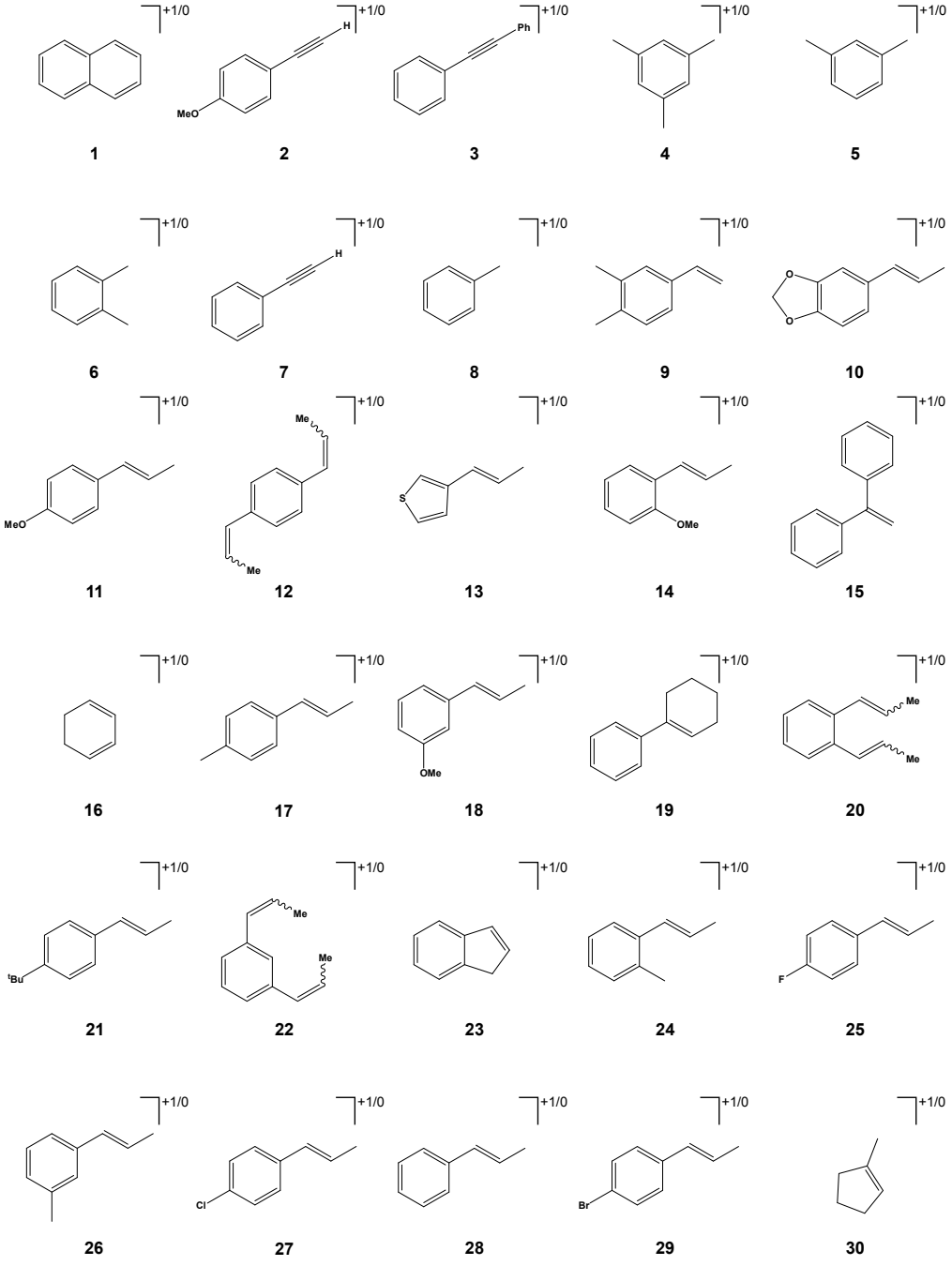
Table A9.8.: Comparison of $\Delta\delta G^{\circ}_{solv}$ contributions of GFN*n*-xTB methods employing the GBSA and of PM*x* methods employing the COSMO implicit solvation model (in V).

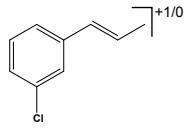
Compound	GFN1-xTB	GFN2-xTB	PM6-D3H4	PM7	COSMO-RS
1	-1.81	-1.82	-1.83	-1.83	-1.70
30	-2.07	-2.07	-2.16	-2.14	-2.02
120	2.76	2.66	3.30	3.29	2.50
150	1.80	1.80	2.69	2.26	1.96
220	-6.57	-6.44	-6.56	-6.50	-6.35
255	1.93	2.37	3.06	2.66	1.88
283	-4.65	-4.46	-4.26	-4.50	-4.42

ROP313

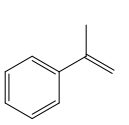
Structural Formulas

OROP

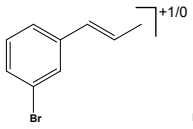




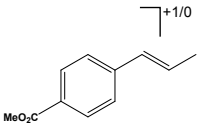
31



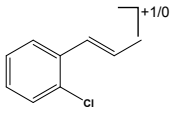
32



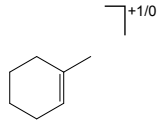
33



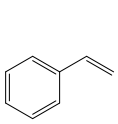
34



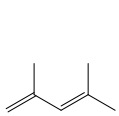
35



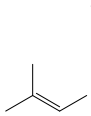
36



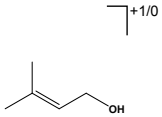
37



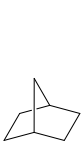
38



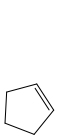
39



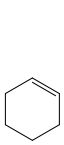
40



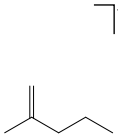
41



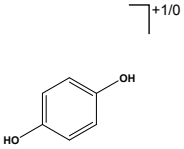
42



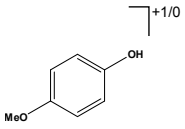
43



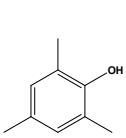
44



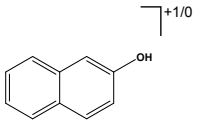
45



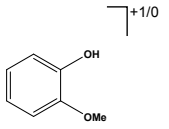
46



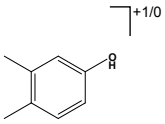
47



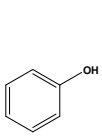
48



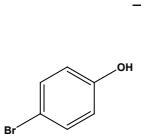
49



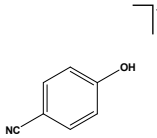
50



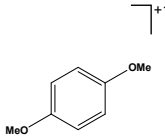
51



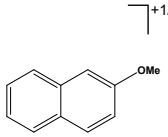
52



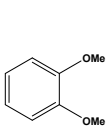
53



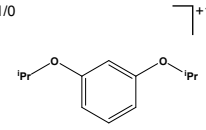
54



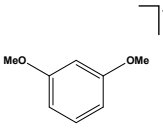
55



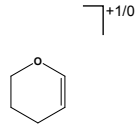
56



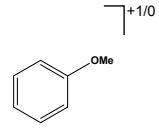
57



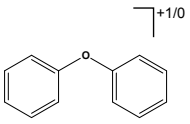
58



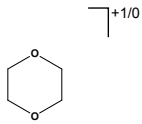
59



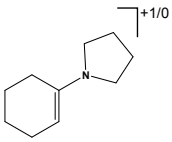
60



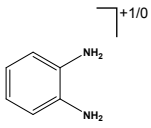
61



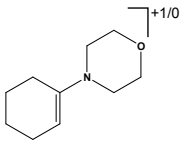
62



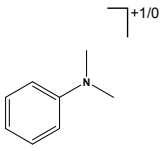
63



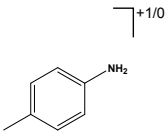
64



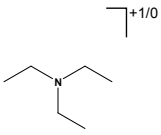
65



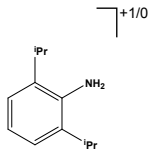
66



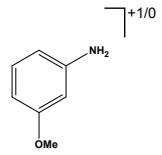
67



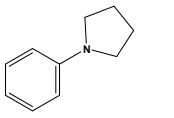
68



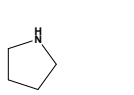
69



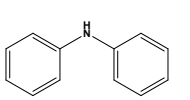
70



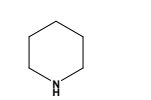
71



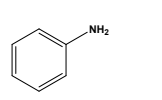
72



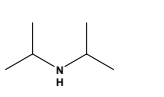
73



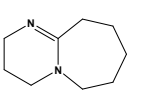
74



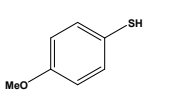
75



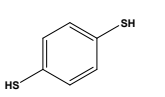
76



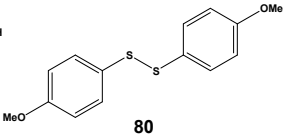
77



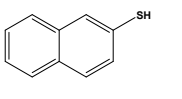
78



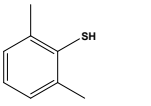
79



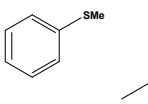
80



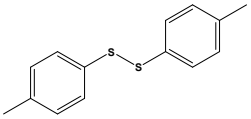
81



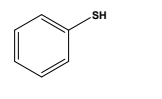
82



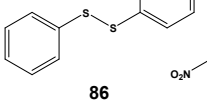
83



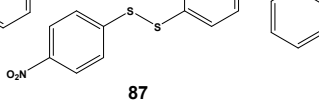
84



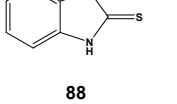
85



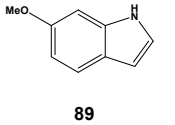
86



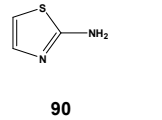
87



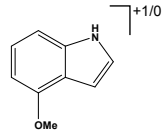
88



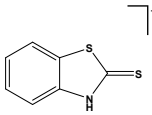
89



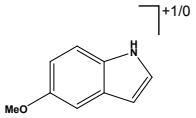
90



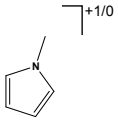
91



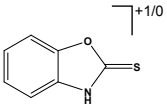
92



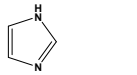
93



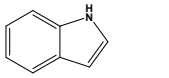
94



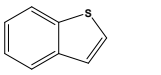
95



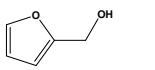
96



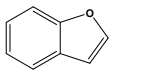
97



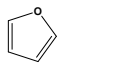
98



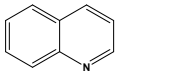
99



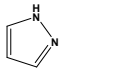
100



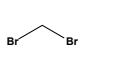
101



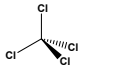
102



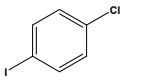
103



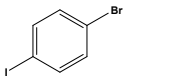
104



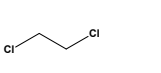
105



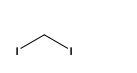
106



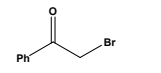
107



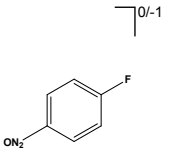
108



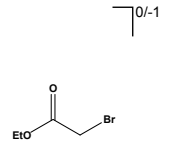
109



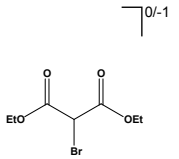
110



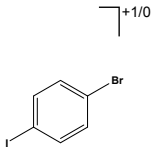
111



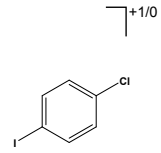
112



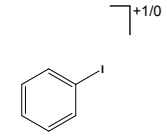
113



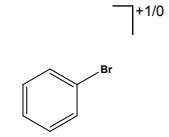
114



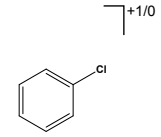
115



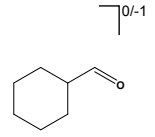
116



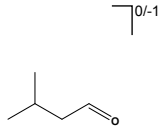
117



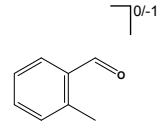
118



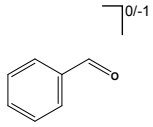
119



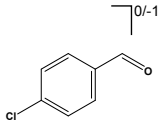
120



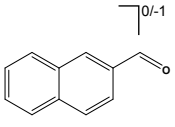
121



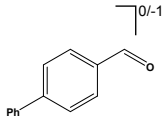
122



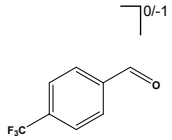
123



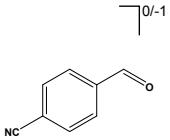
124



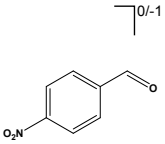
125



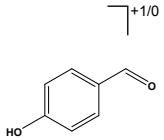
126



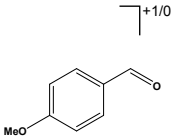
127



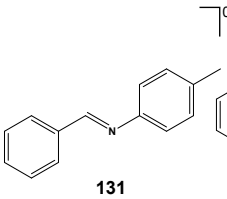
128



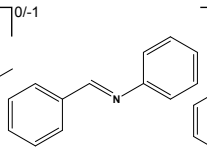
129



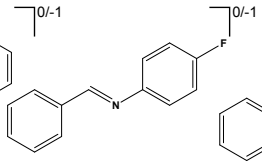
130



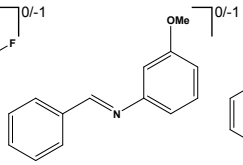
131



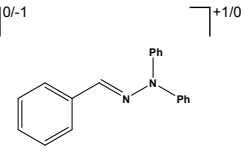
132



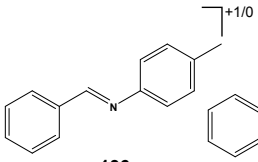
133



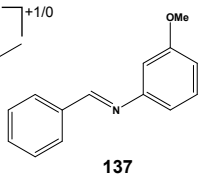
134



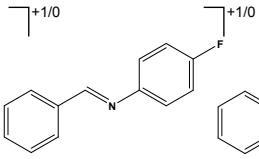
135



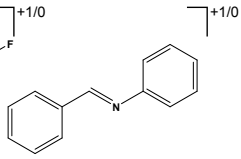
136



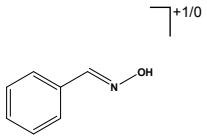
137



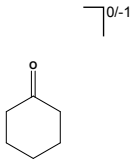
138



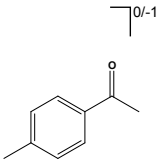
139



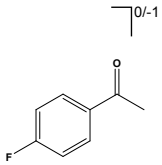
140



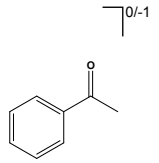
141



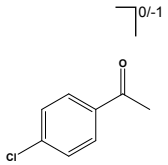
142



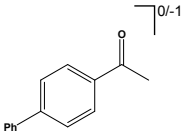
143



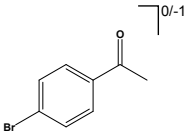
144



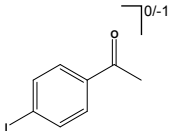
145



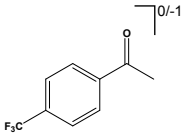
146



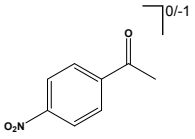
147



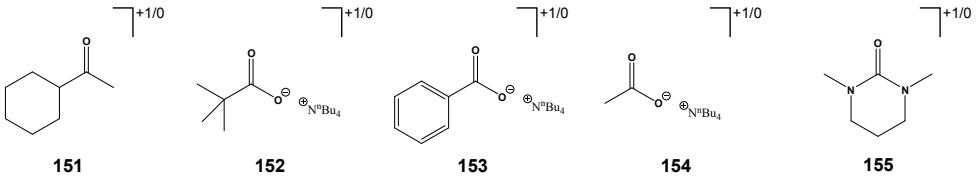
148

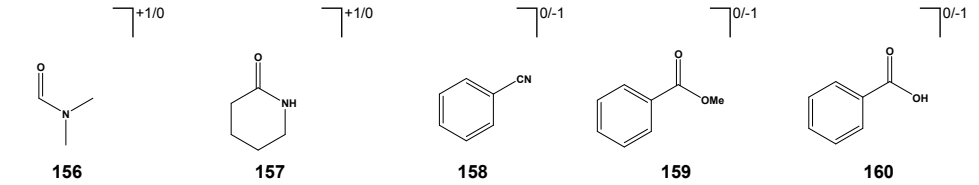


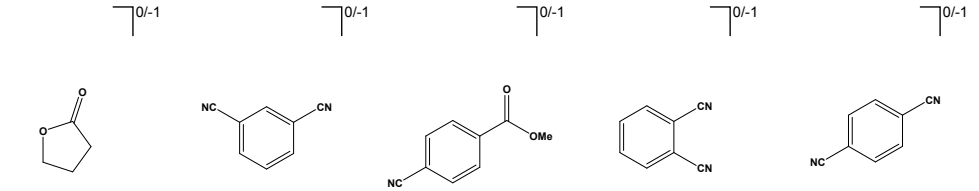
149

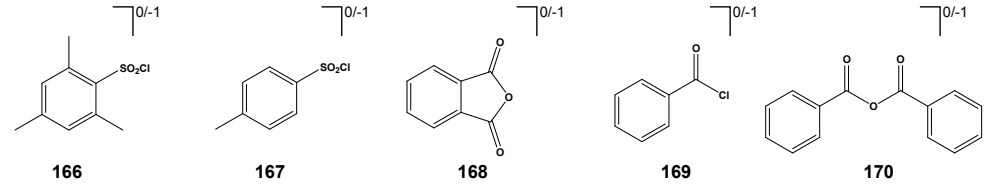


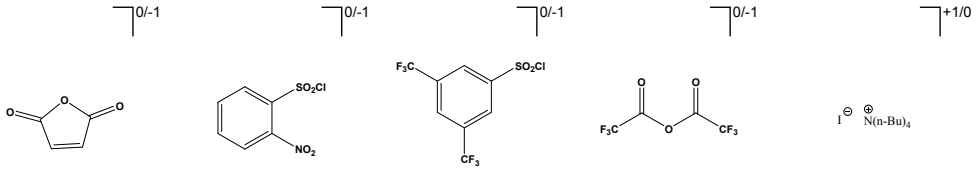
150

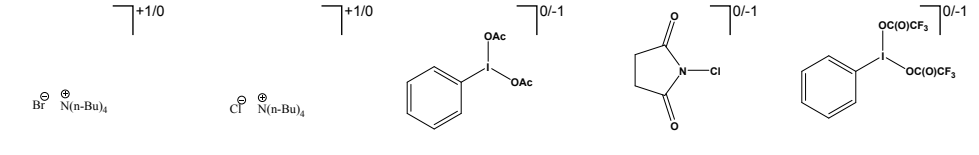

+1/0
+1/0
+1/0
+1/0
+1/0
151
152
153
154
155

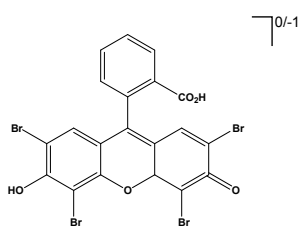

+1/0
+1/0
0/-1
0/-1
0/-1
156
157
158
159
160


0/-1
0/-1
0/-1
0/-1
0/-1
161
162
163
164
165

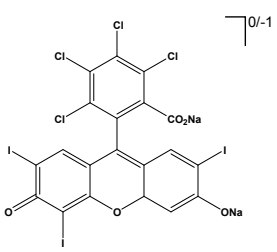

0/-1
0/-1
0/-1
0/-1
0/-1
166
167
168
169
170


0/-1
0/-1
0/-1
0/-1
+1/0
171
172
173
174
175

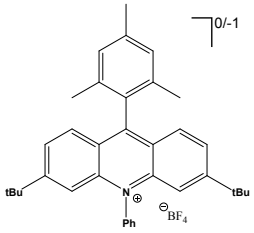

+1/0
+1/0
0/-1
0/-1
0/-1
176
177
178
179
180



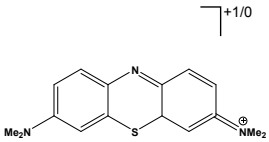
181



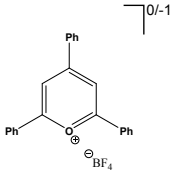
182



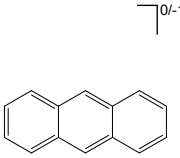
183



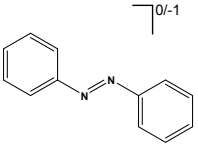
184



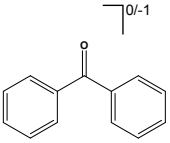
185



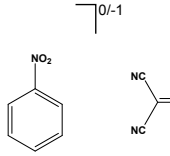
186



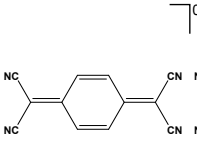
187



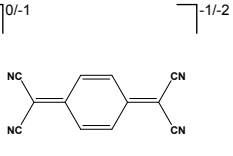
188



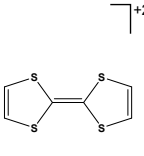
189



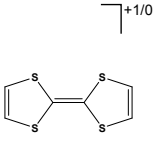
190



191

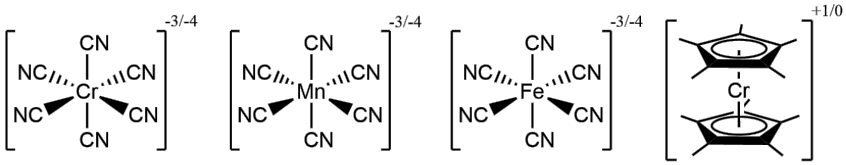


192

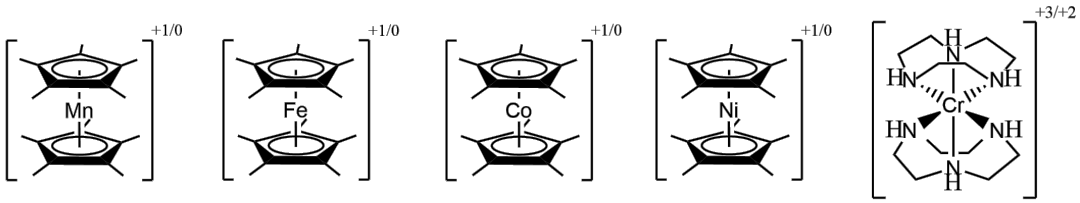


193

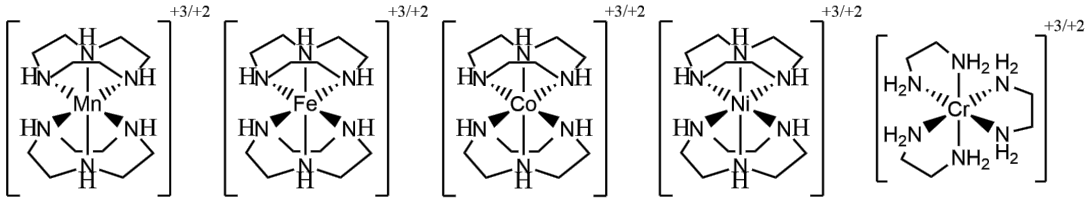
OMROP



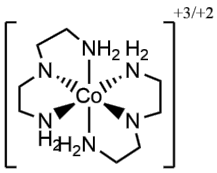
194 **195** **196** **197**



198 **199** **200** **201** **202**



203 **204** **205** **206** **207**



208

$[Cr(tpen)]^{+3/+2}$ **209**

$[Mn(tpen)]^{+3/+2}$ **210**

$[Fe(tpen)]^{+3/+2}$ **211**

$[Co(tpen)]^{+3/+2}$ **212**

$[Ni(tpen)]^{+3/+2}$ **213**

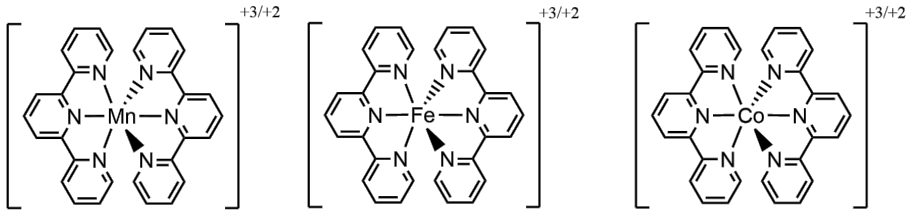
$[Cr(tp)_2]^{+1/0}$ **214**

$[Mn(tp)_2]^{+1/0}$ **215**

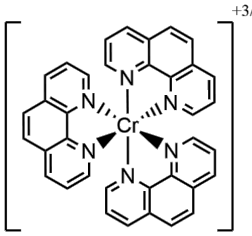
$[Fe(tp)_2]^{+1/0}$ **216**

$[Co(tp)_2]^{+1/0}$ **217**

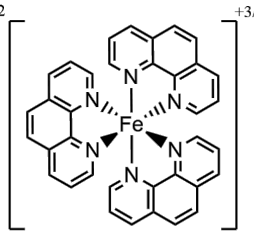
$[Ni(tp)_2]^{+1/0}$ **218**



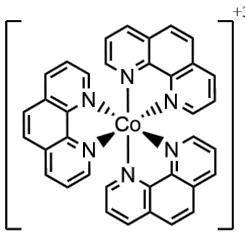
219 **220** **221**



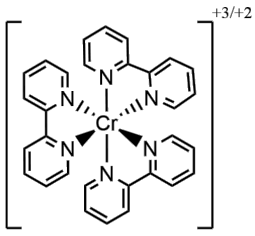
222



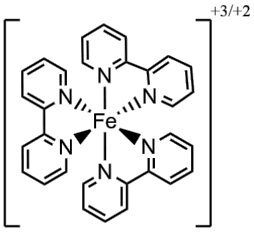
223



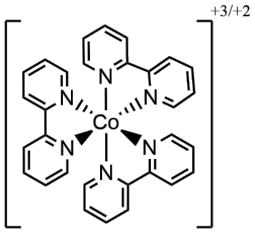
224



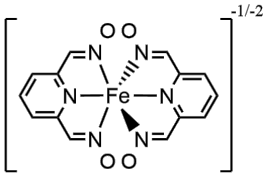
225



226



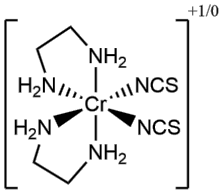
227



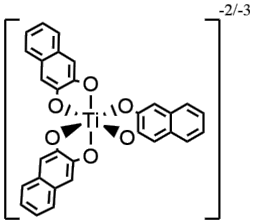
228

$[Fe(py2ox)_3]^{-1/-2}$

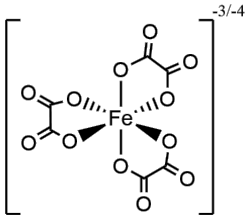
229



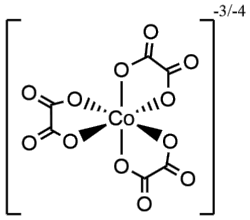
230



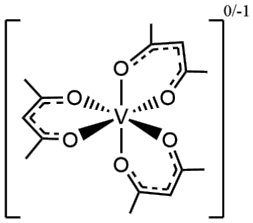
231



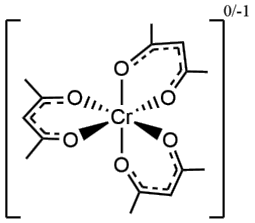
232



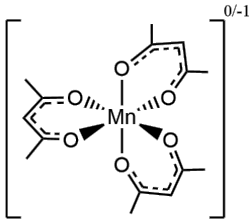
233



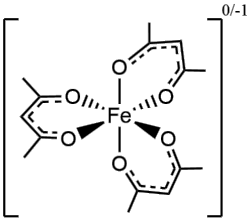
234



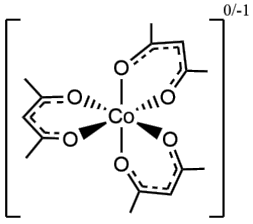
235



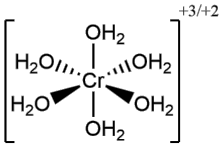
236



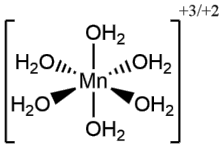
237



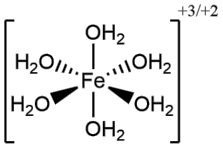
238



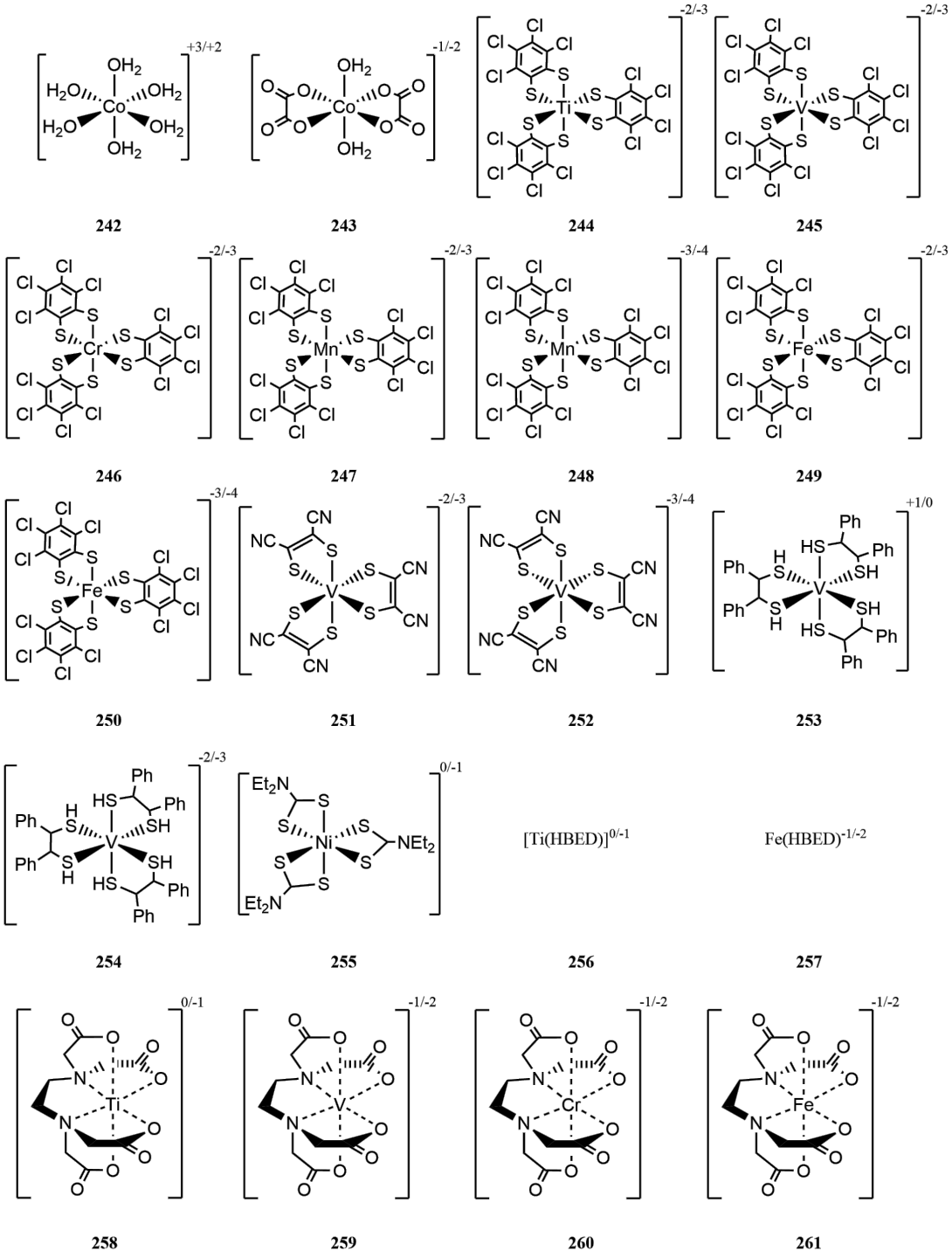
239

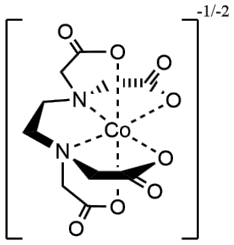


240

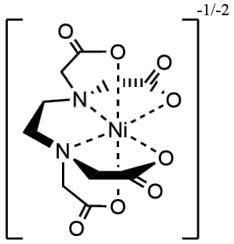


241

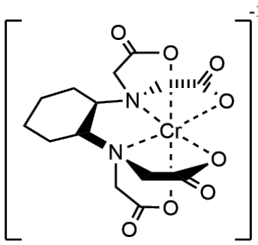

+3/+2
-1/-2
-2/-3
-2/-3
242
243
244
245
-2/-3
-2/-3
-3/-4
-2/-3
246
247
248
249
-3/-4
-2/-3
-3/-4
+1/0
250
251
252
253
-2/-3
0/-1
[Ti(HBED)]0/-1
Fe(HBED)-1/-2
254
255
256
257
0/-1
-1/-2
-1/-2
-1/-2
258
259
260
261



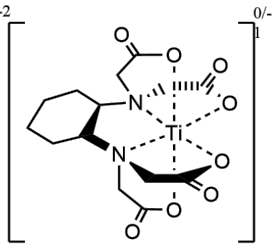
262



263

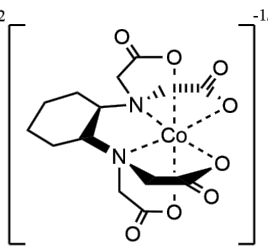


264

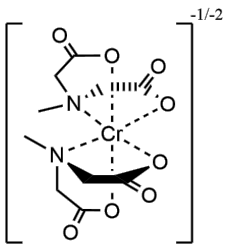


265

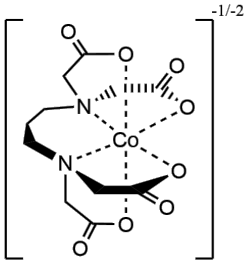
266



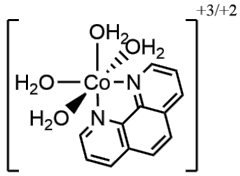
267



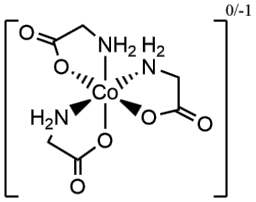
268



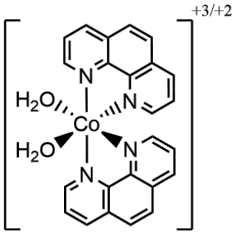
269



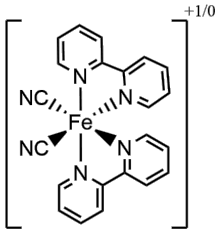
270



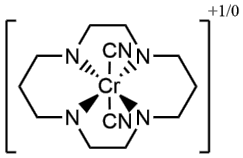
271



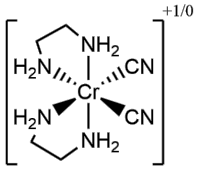
272



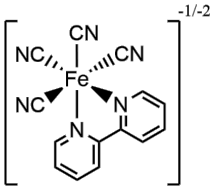
273



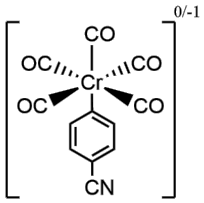
274



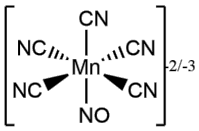
275



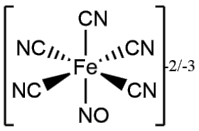
276



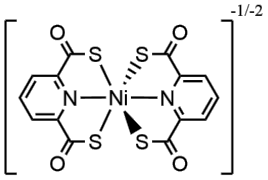
277



278



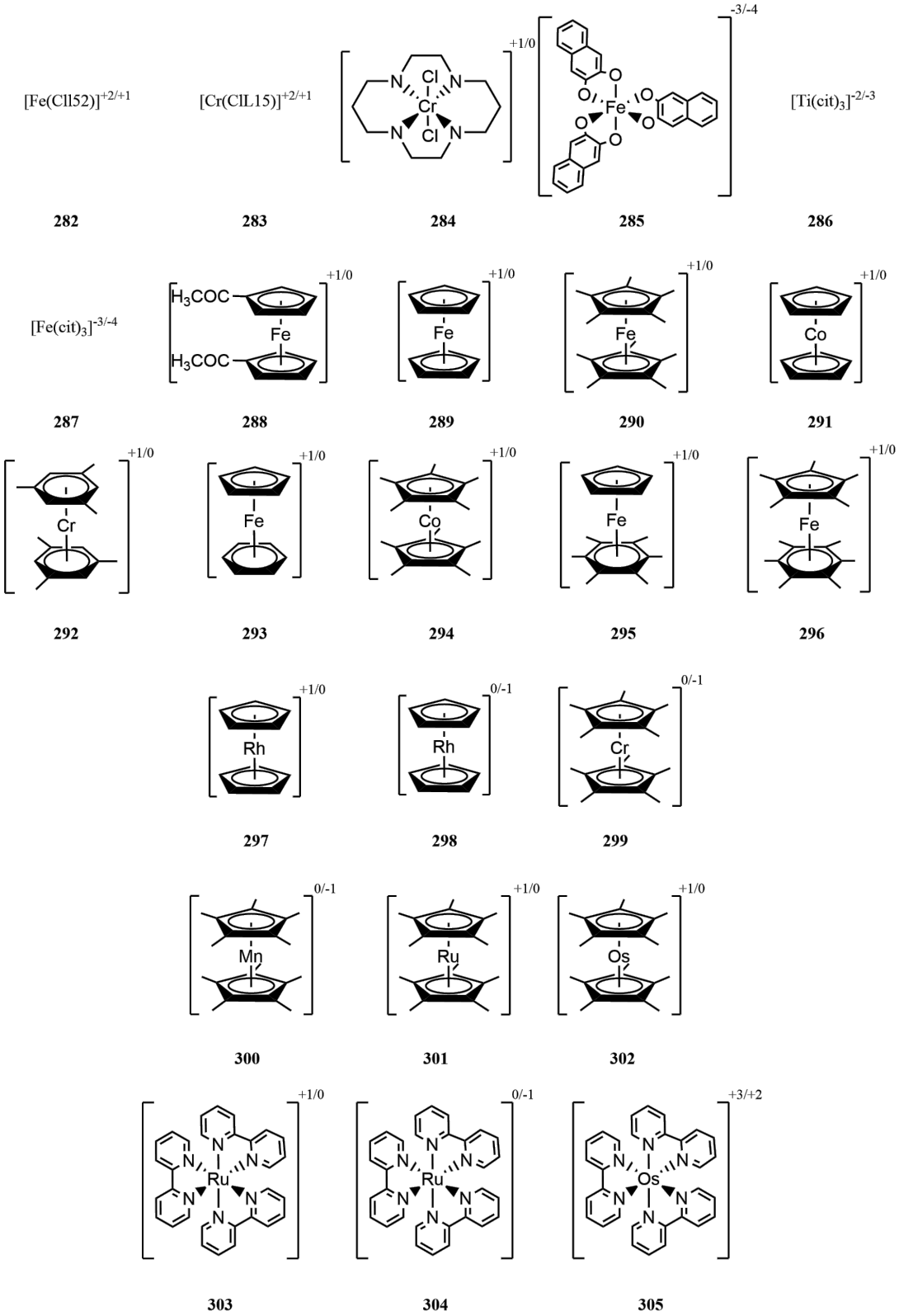
279

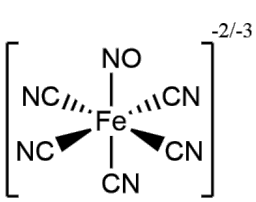


280

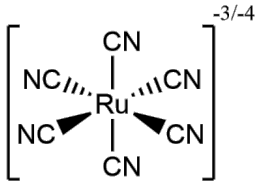
$[Mn(Cll52)]^{+2/+1}$

281

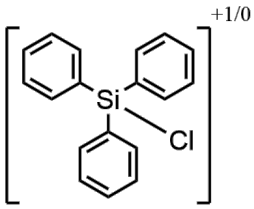

[Fe(Cll52)]+2/+1
[Cr(ClL15)]+2/+1
+1/0
-3/-4
[Ti(cit)3]-2/-3
282
283
284
285
286
[Fe(cit)3]-3/-4
H3COC
Fe
H3COC
+1/0
Fe
+1/0
Fe
+1/0
Co
+1/0
287
288
289
290
291
Cr
+1/0
Fe
+1/0
Co
+1/0
Fe
+1/0
Fe
+1/0
292
293
294
295
296
Rh
+1/0
Rh
0/-1
Cr
0/-1
297
298
299
Mn
0/-1
Ru
+1/0
Os
+1/0
300
301
302
Ru
+1/0
Ru
0/-1
Os
+3/+2
303
304
305



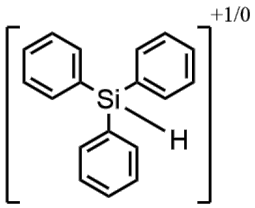
306



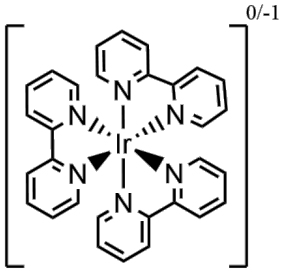
307



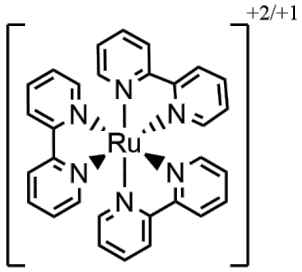
308



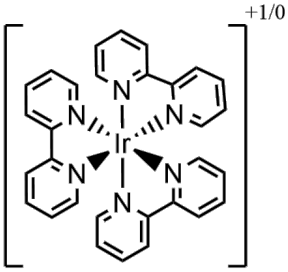
309



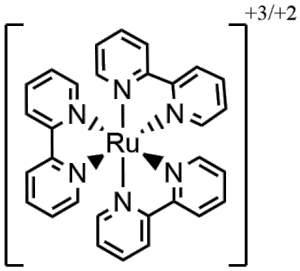
310



311

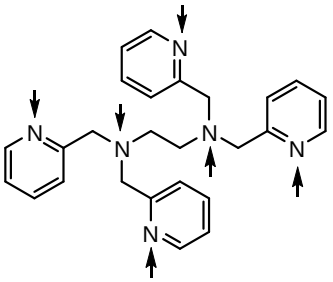


312

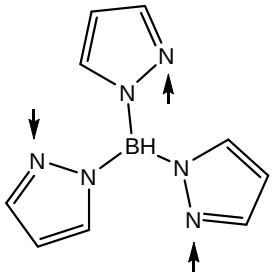


313

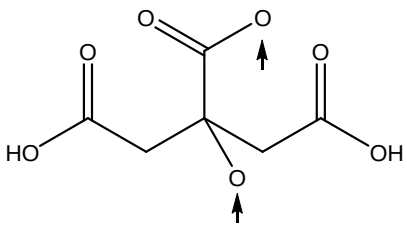
Ligands abbreviated:



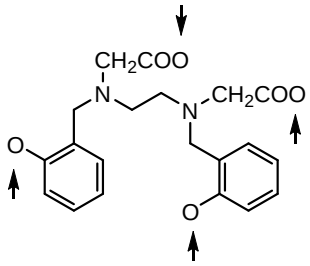
tpen



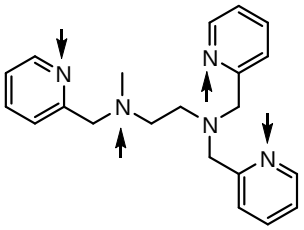
tp



cit



hbed



l52

DFT

In the following, all energies and redox potentials are given in V.

Minimal Augmented Basis Sets

Table A9.9.: Comparison DFT/def2-QZVPP and DFT/ma-def2-QZVPP[470] (all converged results included except outliers with absolute redox potentials larger than 4 V).

OROP					
DFT	N	MD	MAD	SD	AMAX
B97-D3/def2-QZVPP	191	0.01	0.28	0.37	1.39
B97-D3/ma-def2-QZVPP	187	0.04	0.28	0.37	1.45
PW6B95-D4/def2-QZVPP	186	0.15	0.23	0.30	1.29
PW6B95-D4/ma-def2-QZVPP	186	0.17	0.23	0.29	1.29
PWPB95-D4/def2-QZVPP	189	0.13	0.22	0.31	1.27
PWPB95-D4/ma-def2-QZVPP	189	0.14	0.22	0.29	1.27
OMROP					
DFT	N	MD	MAD	SD	AMAX
B97-D3/def2-QZVPP	111	-0.07	0.35	0.45	1.53
B97-D3/ma-def2-QZVPP	89	-0.03	0.39	0.56	2.18
PW6B95-D4/def2-QZVPP	89	0.17	0.33	0.46	1.81
PW6B95-D4/ma-def2-QZVPP	89	0.23	0.38	0.52	1.94
PWPB95-D4/def2-QZVPP	90	0.03	0.34	0.49	1.83
PWPB95-D4/ma-def2-QZVPP	89	0.09	0.36	0.54	1.95

B97-3c and B97-D3

Table A9.10.: Redox potentials and ionization potentials calculated by DFT methods (1=B97-3c, 2=B97-D3/def2-QZVPP, 3=B97-D3/ma-def2-QZVPP) with ΔG°_{RRHO} corrections from B97-3c vibrational frequency calculations and $\Delta\delta G^{\circ}_{solv}$ solvation contributions from COSMO-RS calculations and experimental reference redox potentials (ref.).

Compound	*IP*(1)	ΔG°_{RRHO}	$\Delta\delta G^{\circ}_{solv}$	E(1)	*IP*(2)	E(2)	*IP*(3)	E(3)	ref.
1	7.69	0.00	-1.70	1.56	7.69	1.56	7.93	1.81	1.64
2	7.49	0.00	-1.61	1.46	7.50	1.47	7.78	1.75	1.65
3	7.24	0.00	-1.32	1.50	7.26	1.51	7.55	1.81	1.84
4	8.02	-0.05	-1.68	1.87	8.00	1.85	8.26	2.10	2.07
5	8.17	-0.07	-1.79	1.89	8.15	1.87	8.40	2.12	2.22
6	8.22	-0.08	-1.81	1.91	8.20	1.89	8.44	2.13	2.22

Compound	$IP(1)$	ΔG°_{RRHO}	$\Delta\delta G^{\circ}_{solv}$	E(1)	$IP(2)$	E(2)	$IP(3)$	E(3)	ref.
7	8.31	0.00	-1.79	2.09	8.30	2.09	8.57	2.36	2.27
8	8.48	0.00	-1.93	2.13	8.47	2.12	8.70	2.35	2.36
9	7.52	0.00	-1.60	1.50	7.51	1.49	7.77	1.74	1.15
10	6.86	0.01	-1.57	0.88	6.87	0.88	7.16	1.18	1.15
11	6.93	0.00	-1.50	1.02	6.94	1.03	7.21	1.29	1.24
12	6.87	0.00	-1.40	1.05	6.88	1.06	7.16	1.34	1.43
13	7.49	0.00	-1.74	1.33	7.48	1.32	7.72	1.56	1.49
14	7.08	-0.01	-1.49	1.15	7.09	1.16	7.36	1.44	1.48
15	7.44	0.00	-1.45	1.57	7.44	1.57	7.76	1.89	1.54
16	9.00	0.00	-2.09	2.48	8.98	2.46	9.19	2.68	1.54
17	7.32	-0.02	-1.58	1.31	7.32	1.30	7.57	1.56	1.59
18	7.29	0.00	-1.51	1.36	7.29	1.36	7.62	1.69	1.6
19	7.37	0.00	-1.55	1.40	7.36	1.39	7.63	1.66	1.64
20	7.12	0.00	-1.46	1.24	7.12	1.24	7.43	1.55	1.64
21	7.23	0.00	-1.46	1.35	7.22	1.34	7.49	1.61	1.65
22	7.26	0.00	-1.44	1.39	7.26	1.39	7.61	1.74	1.66
23	7.69	0.00	-1.77	1.50	7.69	1.50	7.92	1.73	1.66
24	7.45	0.00	-1.62	1.41	7.45	1.41	7.70	1.66	1.7
25	7.58	0.00	-1.73	1.43	7.56	1.41	7.83	1.68	1.71
26	7.44	-0.03	-1.59	1.40	7.43	1.39	7.70	1.65	1.73
27	7.54	0.00	-1.64	1.48	7.52	1.46	7.80	1.74	1.73
28	7.59	0.00	-1.69	1.48	7.59	1.48	7.84	1.73	1.74
29	7.50	0.00	-1.59	1.49	7.49	1.48	7.78	1.76	1.75
30	8.10	0.00	-2.02	1.65	8.09	1.65	8.34	1.90	1.75
31	7.76	0.00	-1.70	1.64	7.73	1.61	8.02	1.90	1.89
32	7.84	0.00	-1.73	1.69	7.83	1.67	8.08	1.93	1.91
33	7.72	0.00	-1.65	1.65	7.71	1.64	8.01	1.93	1.92
34	7.69	0.00	-1.56	1.71	7.68	1.70	8.03	2.05	1.94
35	7.69	0.00	-1.65	1.62	7.67	1.59	7.95	1.87	1.94
36	8.06	0.00	-1.94	1.70	8.05	1.69	8.29	1.93	1.94
37	7.98	0.00	-1.82	1.74	7.98	1.74	8.22	1.98	1.97
38	7.68	0.00	-1.82	1.44	7.66	1.42	7.91	1.67	1.98
39	8.22	-0.10	-2.07	1.63	8.21	1.62	8.43	1.84	1.98
40	8.35	0.00	-2.15	1.77	8.35	1.77	8.68	2.10	2.1
41	8.36	0.00	-2.01	1.93	8.34	1.91	8.57	2.14	2.22
42	8.52	0.00	-2.19	1.89	8.52	1.90	8.77	2.15	2.32
43	8.48	0.00	-2.10	1.96	8.48	1.95	8.71	2.19	2.37
44	8.60	0.00	-2.00	2.17	8.57	2.14	8.84	2.41	2.5
45	7.42	0.00	-2.05	0.95	7.46	0.98	7.68	1.21	1.14
46	7.21	0.00	-1.82	0.97	7.24	1.00	7.47	1.24	1.17
47	7.34	-0.03	-1.70	1.19	7.35	1.19	7.58	1.43	1.35
48	7.39	0.00	-1.73	1.23	7.40	1.24	7.66	1.50	1.4
49	7.45	0.00	-1.84	1.19	7.48	1.22	7.73	1.47	1.41
50	7.56	0.00	-1.83	1.31	7.57	1.32	7.80	1.55	1.43
51	8.11	0.00	-2.07	1.61	8.13	1.63	8.34	1.85	1.63
52	7.89	0.00	-1.89	1.58	7.90	1.58	8.16	1.84	1.69
53	8.54	0.00	-2.12	2.00	8.55	2.01	8.83	2.29	2.08
54	7.01	0.00	-1.60	0.99	7.04	1.01	-	-	1.3

Compound	$IP(1)$	ΔG°_{RRHO}	$\Delta\delta G^{\circ}_{solv}$	E(1)	$IP(2)$	E(2)	$IP(3)$	E(3)	ref.
55	7.25	0.00	-1.54	1.28	7.26	1.29	7.53	1.56	1.32
56	7.16	0.00	-1.60	1.14	7.18	1.16	7.43	1.42	1.43
57	7.00	0.00	-1.37	1.21	7.00	1.21	7.32	1.52	1.47
58	7.52	0.00	-1.64	1.46	7.53	1.47	7.85	1.79	1.5
59	7.90	0.00	-2.10	1.37	7.91	1.39	8.13	1.60	1.51
60	7.84	0.00	-1.82	1.60	7.85	1.61	8.08	1.84	1.81
61	7.46	0.00	-1.48	1.57	7.48	1.58	7.82	1.92	1.88
62	8.39	0.00	-2.16	1.80	8.41	1.82	8.79	2.20	2.5
63	6.23	0.02	-1.58	0.24	6.26	0.27	6.44	0.46	0.32
64	6.78	0.00	-1.95	0.41	6.82	0.45	7.02	0.65	0.48
65	6.41	0.00	-1.65	0.34	6.43	0.36	6.67	0.59	0.54
66	6.84	0.00	-1.71	0.71	6.88	0.74	7.07	0.94	0.74
67	7.02	0.00	-1.86	0.74	7.06	0.78	7.24	0.96	0.81
68	7.03	0.00	-1.84	0.77	7.05	0.78	7.24	0.97	0.83
69	6.86	0.00	-1.60	0.85	6.87	0.85	7.07	1.05	0.84
70	7.08	-0.01	-1.77	0.87	7.11	0.90	7.36	1.15	0.86
71	6.71	0.00	-1.64	0.65	6.74	0.68	7.00	0.93	0.88
72	7.64	-0.06	-2.25	0.91	-291.90	-298.63	-296.93	-303.66	0.89
73	6.76	0.00	-1.55	0.79	6.80	0.83	7.05	1.08	0.92
74	7.57	0.00	-2.09	1.05	7.60	1.08	7.80	1.28	0.94
75	7.33	0.01	-2.00	0.91	7.37	0.96	7.55	1.13	0.95
76	7.41	0.00	-1.89	1.10	7.45	1.14	7.67	1.36	1.09
77	7.10	0.00	-1.60	1.09	7.12	1.10	7.34	1.32	1.28
78	7.14	0.00	-1.68	1.05	7.16	1.06	7.42	1.32	1.15
79	7.26	0.00	-1.75	1.09	7.26	1.09	7.54	1.37	1.16
80	6.65	0.00	-1.12	1.11	6.63	1.08	7.06	1.51	1.28
81	7.31	0.00	-1.61	1.28	7.31	1.28	7.59	1.56	1.33
82	7.60	0.00	-1.69	1.49	7.58	1.47	7.83	1.72	1.44
83	7.51	0.00	-1.74	1.35	7.49	1.33	7.72	1.56	1.44
84	7.00	0.00	-1.22	1.35	6.96	1.31	7.32	1.68	1.46
85	7.89	0.00	-1.91	1.56	7.89	1.56	8.12	1.79	1.51
86	7.28	0.00	-1.38	1.47	7.25	1.45	7.63	1.83	1.56
87	8.16	0.00	-1.72	2.02	8.12	1.98	8.62	2.48	2.02
88	7.29	0.00	-1.67	1.19	7.31	1.21	7.53	1.43	0.81
89	6.81	0.00	-1.60	0.79	6.84	0.82	7.08	1.06	0.91
90	7.62	-0.01	-2.10	1.09	7.63	1.10	7.82	1.29	0.95
91	6.79	0.00	-1.56	0.81	6.83	0.84	7.08	1.10	0.97
92	7.57	0.00	-1.64	1.51	7.55	1.49	7.77	1.71	1.03
93	6.97	0.00	-1.58	0.97	7.00	1.00	7.24	1.24	1.04
94	7.76	0.00	-2.02	1.32	7.76	1.32	7.93	1.49	1.04
95	7.73	0.00	-1.72	1.58	7.73	1.59	7.97	1.83	1.07
96	8.62	0.00	-2.38	1.82	8.61	1.81	8.78	1.97	1.15
97	7.36	0.00	-1.79	1.14	7.39	1.17	7.61	1.39	1.16
98	7.75	0.00	-1.75	1.58	7.74	1.56	7.99	1.82	1.62
99	8.16	0.00	-2.17	1.57	8.17	1.57	8.42	1.82	1.73
100	7.93	0.00	-1.86	1.66	7.93	1.65	8.17	1.89	1.74
101	8.61	0.00	-2.35	1.84	8.60	1.82	8.78	2.01	1.94
102	8.20	0.00	-1.74	2.04	8.19	2.03	8.48	2.32	2.14

Compound	$IP(1)$	ΔG°_{RRHO}	$\Delta\delta G^{\circ}_{solv}$	E(1)	$IP(2)$	E(2)	$IP(3)$	E(3)	ref.
103	9.13	0.00	-2.37	2.33	9.11	2.32	9.29	2.50	2.21
104	-	-	-	-	-	-	-	-	-2.48
105	1.25	0.14	2.39	-0.64	1.35	-0.54	1.08	-0.82	-1.93
106	0.66	0.00	2.26	-1.50	0.74	-1.42	0.57	-1.59	-1.94
107	0.68	0.01	2.24	-1.49	0.77	-1.41	0.59	-1.59	-1.95
108	0.88	0.01	2.31	-1.23	1.04	-1.07	0.59	-1.52	-1.92
109	1.05	0.01	2.13	-1.24	1.15	-1.13	0.94	-1.34	-1.82
110	1.41	0.01	2.21	-0.80	1.55	-0.66	1.35	-0.86	-1.46
111	0.94	0.00	2.21	-1.27	1.04	-1.17	1.14	-1.07	-1.23
112	1.12	0.01	2.34	-0.95	1.28	-0.79	1.09	-0.98	-1.06
113	1.67	0.00	2.20	-0.54	1.79	-0.42	1.62	-0.59	-0.62
114	8.10	0.00	-1.66	2.02	8.09	2.01	8.40	2.32	2.1
115	8.18	0.00	-1.71	2.05	8.14	2.01	8.45	2.32	2.15
116	8.30	0.00	-1.78	2.11	8.29	2.09	8.56	2.36	2.17
117	8.56	0.00	-1.87	2.27	8.54	2.25	8.80	2.51	2.5
118	8.69	0.00	-1.94	2.33	8.64	2.28	8.90	2.54	2.61
119	-1.04	0.01	2.53	-2.93	-0.85	-2.73	-0.71	-2.60	-2.28
120	-1.06	0.01	2.50	-2.98	-0.86	-2.78	-0.72	-2.63	-2.24
121	0.20	0.01	2.23	-1.98	0.34	-1.84	0.30	-1.89	-1.94
122	0.26	0.01	2.27	-1.89	0.39	-1.75	0.36	-1.79	-1.93
123	0.54	0.00	2.12	-1.76	0.66	-1.64	0.63	-1.67	-1.85
124	0.66	0.00	2.03	-1.73	0.77	-1.62	0.70	-1.69	-1.73
125	0.80	0.00	1.94	-1.68	0.92	-1.56	0.83	-1.65	-1.72
126	0.98	0.00	1.97	-1.46	1.07	-1.37	1.01	-1.43	-1.66
127	1.31	0.00	1.90	-1.21	1.39	-1.13	1.34	-1.18	-1.42
128	1.82	0.00	1.93	-0.67	1.90	-0.59	1.87	-0.62	-0.86
129	8.41	0.00	-1.97	2.01	8.42	2.03	8.75	2.35	1.95
130	8.16	0.00	-1.77	1.96	8.16	1.97	8.48	2.28	2.06
131	0.52	0.01	1.98	-1.92	0.66	-1.78	-	-	-2.01
132	0.56	0.01	1.99	-1.87	0.70	-1.73	0.63	-1.80	-1.91
133	0.64	0.01	1.91	-1.87	0.76	-1.74	0.69	-1.81	-1.9
134	0.59	0.01	1.97	-1.86	0.72	-1.73	0.67	-1.78	-1.88
135	6.37	0.00	-1.12	0.83	6.40	0.86	6.64	1.10	0.97
136	7.07	-0.03	-1.33	1.29	7.07	1.29	7.35	1.57	1.5
137	7.01	0.00	-1.28	1.30	7.02	1.31	7.37	1.67	1.52
138	7.30	0.00	-1.44	1.43	7.29	1.43	7.61	1.74	1.64
139	7.29	0.00	-1.40	1.47	7.30	1.47	7.58	1.76	1.67
140	7.94	0.00	-1.81	1.71	7.97	1.73	8.23	1.99	1.86
141	-1.06	0.01	2.38	-3.09	-0.87	-2.90	-0.61	-2.64	-2.33
142	0.06	0.01	2.20	-2.16	0.20	-2.02	0.15	-2.06	-2.16
143	0.18	0.01	2.14	-2.09	0.31	-1.97	0.27	-2.00	-2.13
144	0.12	0.01	2.23	-2.06	0.27	-1.92	0.23	-1.96	-2.11
145	0.40	0.01	2.08	-1.93	0.53	-1.80	0.50	-1.84	-1.91
146	0.67	0.00	1.92	-1.83	0.81	-1.69	0.70	-1.80	-1.91
147	0.46	0.15	2.05	-1.76	0.58	-1.64	0.55	-1.67	-1.89
148	0.77	0.12	2.07	-1.46	0.84	-1.38	0.62	-1.60	-1.85
149	0.83	0.00	1.96	-1.64	0.92	-1.54	0.86	-1.60	-1.74
150	1.67	0.00	1.96	-0.79	1.76	-0.71	1.73	-0.74	-0.93

Compound	$IP(1)$	ΔG°_{RRHO}	$\Delta\delta G^{\circ}_{solv}$	E(1)	$IP(2)$	E(2)	$IP(3)$	E(3)	ref.
151	8.38	0.00	-1.73	2.23	8.36	2.20	8.81	2.66	2.66
152	5.83	-0.01	-0.88	0.52	5.70	0.39	6.14	0.84	1.26
153	6.44	-0.09	-0.76	1.17	6.49	1.22	6.98	1.71	1.4
154	6.63	0.00	-0.86	1.34	6.63	1.34	6.98	1.69	1.47
155	7.71	0.00	-1.83	1.45	7.73	1.47	8.13	1.87	1.55
156	8.94	0.00	-2.31	2.21	8.94	2.21	9.14	2.40	2.26
157	8.75	0.00	-2.03	2.29	8.77	2.31	9.10	2.64	2.58
158	-0.06	0.01	2.17	-2.31	0.08	-2.17	0.04	-2.20	-2.43
159	-0.09	0.01	2.24	-2.27	0.05	-2.13	0.00	-2.17	-2.34
160	0.00	0.01	2.18	-2.23	0.14	-2.09	0.11	-2.12	-2.29
161	-1.50	0.17	2.45	-3.29	0.26	-1.54	-4.50	-6.30	-2.13
162	0.85	0.01	1.81	-1.76	0.95	-1.66	-	-	-1.9
163	0.98	0.00	1.93	-1.51	1.07	-1.42	1.02	-1.47	-1.72
164	0.94	0.00	1.92	-1.55	1.03	-1.46	0.99	-1.50	-1.7
165	1.14	0.00	1.82	-1.46	1.22	-1.37	1.19	-1.41	-1.64
166	1.87	0.01	2.30	-0.24	1.72	-0.39	1.58	-0.53	-1.45
167	1.84	0.01	2.34	-0.24	1.69	-0.39	1.57	-0.51	-1.37
168	1.21	0.00	2.06	-1.15	1.27	-1.09	1.25	-1.11	-1.34
169	0.73	0.01	2.15	-1.53	0.84	-1.43	0.77	-1.50	-1.2
170	0.85	0.01	1.92	-1.65	0.94	-1.55	0.72	-1.77	-1.01
171	1.46	0.00	2.20	-0.75	1.52	-0.69	1.56	-0.66	-0.98
172	2.36	0.01	2.12	0.06	2.21	-0.09	2.08	-0.22	-0.51
173	2.44	0.00	1.98	0.00	2.26	-0.18	2.14	-0.30	-0.45
174	1.27	0.14	2.02	-1.00	1.28	-0.99	1.13	-1.14	-0.2
175	6.46	0.00	-0.57	1.46	6.42	1.42	6.55	1.55	0.26
176	6.84	0.00	-0.70	1.72	6.83	1.71	7.01	1.89	0.71
177	6.91	0.00	-0.82	1.67	6.91	1.67	7.18	1.94	1.01
178	2.06	0.18	1.98	-0.20	2.13	-0.13	1.93	-0.33	-1.12
179	1.88	0.00	2.28	-0.26	1.94	-0.20	1.75	-0.39	-0.27
180	3.15	0.01	1.66	0.39	3.17	0.40	3.05	0.29	-0.26
181	1.99	0.00	1.43	-1.00	2.06	-0.93	2.11	-0.88	-1.1
182	-3.47	0.01	6.13	-1.76	-3.40	-1.68	-	-	-0.86
183	2.00	0.15	1.50	-0.77	2.08	-0.69	2.06	-0.70	-0.59
184	4.85	0.09	-0.97	-0.45	4.90	-0.41	5.12	-0.19	-0.32
185	2.53	0.15	1.41	-0.33	2.60	-0.26	2.60	-0.26	-0.28
186	0.43	0.01	1.90	-1.85	0.58	-1.71	0.52	-1.76	-1.92
187	1.06	0.00	1.94	-1.19	1.18	-1.07	1.15	-1.09	-1.36
188	0.58	0.15	2.05	-1.41	0.70	-1.29	0.63	-1.36	-1.88
189	0.86	0.00	2.30	-1.03	0.97	-0.92	1.06	-0.83	-1.15
190	3.52	0.00	1.36	0.69	3.55	0.72	3.71	0.88	0.13
191	-0.70	0.00	4.64	-0.25	-0.64	-0.19	-0.47	-0.02	-0.29
192	10.81	0.05	-5.36	1.31	10.72	1.22	10.89	1.39	0.66
193	6.08	0.02	-1.48	0.43	6.03	0.38	6.21	0.56	0.3
194	-9.63	0.00	13.49	-0.57	-9.40	-0.35	-	-	-1.28
195	-8.95	-0.01	13.68	0.28	-8.71	0.53	-	-	-0.24
196	-8.27	0.00	13.81	1.10	-8.03	1.34	-	-	0.36
197	4.42	0.01	-1.22	-1.24	4.43	-1.22	4.49	-1.16	-0.81
198	4.39	0.24	-1.19	-1.01	4.55	-0.84	4.87	-0.52	-0.32

Compound	$IP(1)$	ΔG°_{RRHO}	$\Delta\delta G^{\circ}_{solv}$	E(1)	$IP(2)$	E(2)	$IP(3)$	E(3)	ref.
199	5.67	-0.01	-1.42	-0.21	5.66	-0.22	5.56	-0.31	0.13
200	3.87	0.28	-1.22	-1.51	3.99	-1.40	4.41	-0.97	-1.24
201	4.73	0.08	-1.40	-1.03	4.83	-0.94	5.34	-0.43	-0.41
202	11.39	0.18	-8.15	-1.02	11.50	-0.91	11.83	-0.58	-1.14
203	13.34	0.01	-8.27	0.64	13.48	0.77	13.54	0.83	0.62
204	12.87	0.00	-8.18	0.25	12.70	0.08	13.21	0.59	0.13
205	11.94	0.00	-8.27	-0.77	12.08	-0.62	12.64	-0.06	-0.4
206	13.15	0.00	-8.21	0.49	13.24	0.59	14.24	1.59	0.95
207	12.18	0.01	-9.05	-1.31	12.35	-1.14	12.64	-0.85	-1.03
208	12.50	0.01	-8.87	-0.81	12.67	-0.64	13.26	-0.04	-0.23
209	10.56	0.07	-6.67	-0.48	10.55	-0.49	10.49	-0.55	-0.71
210	12.34	0.18	-6.90	1.17	12.47	1.14	-	-	1.33
211	-	-	-	-	-	-	-	-	1.01
212	10.98	0.01	-6.89	-0.34	11.09	-0.23	11.58	0.26	0.51
213	12.17	0.00	-6.79	0.94	12.24	1.01	13.23	2.00	1.81
214	3.98	0.00	-1.13	-1.58	4.05	-1.52	-	-	-1.33
215	5.48	0.01	-1.12	-0.07	5.69	0.14	5.60	0.05	0.46
216	5.58	0.04	-1.15	0.03	5.60	0.01	5.70	0.11	0.13
217	4.74	0.01	-1.15	-0.85	4.92	-0.67	5.22	-0.37	-0.13
218	6.07	0.01	-1.14	0.49	6.18	0.61	6.96	1.39	1.21
219	11.74	0.01	-6.44	0.87	11.88	1.01	12.18	1.31	1.28
220	12.07	0.00	-6.35	1.28	12.01	1.22	11.93	1.14	1.12
221	10.81	0.16	-6.44	0.09	10.96	0.23	11.29	0.57	0.31
222	10.41	0.00	-6.03	-0.06	10.38	-0.10	10.22	-0.26	-0.02
223	11.70	0.00	-6.07	1.19	11.63	1.12	11.68	1.16	1.13
224	10.55	0.08	-6.13	0.06	10.63	0.14	11.08	0.59	0.33
225	10.66	0.00	-6.25	-0.03	10.63	-0.06	10.44	-0.25	0.01
226	11.91	0.00	-6.32	1.15	11.85	1.09	11.84	1.08	1.12
227	10.71	0.01	-6.39	-0.11	10.80	-0.02	11.23	0.41	0.31
228	-1.04	0.00	5.29	-0.19	-0.93	-0.08	-0.79	0.06	0.2
229	2.20	0.00	2.16	-0.08	2.25	-0.03	2.29	0.01	0.35
230	4.53	0.26	-1.60	-1.25	4.78	-1.00	5.05	-0.73	-0.9
231	-4.46	0.24	7.04	-1.63	-4.25	-1.42	-	-	-1.12
232	-	-	-	-	-	-	-	-	0.01
233	-7.87	0.01	11.97	-0.33	-7.45	0.10	-	-	0.57
234	1.12	0.00	1.99	-1.32	1.26	-1.19	0.97	-1.48	-1.18
235	0.26	0.01	1.92	-2.26	0.42	-2.10	0.11	-2.40	-1.57
236	1.70	0.14	1.97	-0.63	1.93	-0.40	2.21	-0.12	0.15
237	1.49	0.00	2.01	-0.94	1.65	-0.78	1.87	-0.56	-0.43
238	1.12	0.34	2.05	-0.93	1.51	-0.54	2.15	0.10	-0.11
239	15.22	0.00	-10.07	0.72	15.35	0.85	15.39	0.89	-0.58
240	17.17	0.22	-10.10	2.85	17.35	3.03	17.63	3.31	1.5
241	15.94	0.16	-9.92	1.74	16.11	1.91	16.74	2.54	0.75
242	-	-	-	-	-	-	-	-	1.86
243	-0.58	0.34	4.41	-0.28	-0.05	0.26	0.38	0.69	0.78
244	-2.88	0.00	5.91	-1.41	-2.72	-1.25	-	-	-0.8
245	-3.03	0.00	6.00	-1.46	-2.82	-1.26	-0.38	1.19	-0.75
246	-2.34	0.14	6.05	-0.60	-2.16	-0.42	-	-	-0.12

Compound	$IP(1)$	ΔG°_{RRHO}	$\Delta\delta G^{\circ}_{solv}$	E(1)	$IP(2)$	E(2)	$IP(3)$	E(3)	ref.
247	-2.85	0.01	6.04	-1.25	-2.62	-1.02	-0.60	1.00	-0.7
248	-	-	-	-	-	-	-	-	-0.88
249	-2.26	0.00	5.99	-0.71	-2.09	-0.53	-1.63	-0.08	-0.53
250	-5.50	0.01	8.02	-1.91	-5.33	-1.75	-	-	-1.27
251	-2.73	0.00	6.32	-0.85	-2.55	-0.66	-	-	-0.32
252	-6.17	0.00	9.00	-1.61	-5.92	-1.36	-	-	-1.5
253	6.24	0.01	-0.72	1.08	6.33	1.17	-	-	1.28
254	-3.70	0.21	6.01	-1.92	-3.48	-1.70	-	-	-1.61
255	1.53	0.00	1.88	-1.02	1.68	-0.87	2.44	-0.12	-0.08
256	1.17	0.01	1.93	-1.33	1.33	-1.18	1.88	-0.63	-0.64
257	-1.69	0.18	5.01	-0.95	-1.45	-0.71	-0.90	-0.16	-0.37
258	1.97	0.00	2.15	-0.32	2.10	-0.19	2.69	0.40	0.02
259	-1.82	0.00	5.51	-0.75	-1.67	-0.60	-1.49	-0.42	-1.04
260	-1.88	0.01	5.27	-1.05	-1.64	-0.81	-1.63	-0.80	-0.98
261	-1.31	0.00	5.50	-0.25	-1.09	-0.03	-0.55	0.51	0.12
262	-1.22	0.01	5.10	-0.55	-0.85	-0.18	-0.14	0.53	0.37
263	-0.37	0.00	5.30	0.50	-0.14	0.72	0.91	1.77	1.29
264	1.88	0.00	2.23	-0.33	2.01	-0.20	2.60	0.38	-0.02
265	-1.93	0.01	5.30	-1.07	-1.69	-0.83	-1.70	-0.84	-0.96
266	-1.32	0.00	5.46	-0.30	-1.11	-0.09	-	-	0.04
267	-1.27	0.01	5.07	-0.63	-0.90	-0.26	-0.22	0.42	0.35
268	-2.18	0.01	5.32	-1.29	-1.94	-1.05	-2.02	-1.13	-1.16
269	-1.28	0.01	5.01	-0.70	-0.91	-0.33	-0.16	0.41	0.36
270	14.15	0.01	-8.56	1.16	14.35	1.36	15.25	2.25	1.35
271	1.89	0.01	1.65	-0.90	2.23	-0.55	3.09	0.30	0.2
272	-	-	-	-	-	-	-	-	0.68
273	5.05	0.00	-0.95	-0.34	5.05	-0.33	4.53	-0.86	0.78
274	4.10	0.00	-0.95	-1.28	4.15	-1.24	4.37	-1.02	-1.12
275	4.34	0.00	-1.10	-1.21	4.41	-1.13	4.61	-0.94	-1.22
276	-0.53	0.00	5.77	0.80	-0.46	0.87	-	-	0.54
277	1.65	0.00	1.55	-1.24	1.73	-1.16	1.67	-1.21	-1.03
278	-4.29	0.00	9.68	0.95	-4.15	1.09	-	-	0.6
279	-5.14	0.00	9.45	-0.12	-4.93	0.08	-4.27	0.74	-0.07
280	-0.80	0.00	4.97	-0.26	-0.63	-0.09	0.02	0.55	0.16
281	9.32	0.01	-3.87	1.02	9.50	1.20	9.58	1.28	1.19
282	8.59	0.00	-3.88	0.28	8.67	0.36	9.15	0.83	0.81
283	7.59	0.01	-4.42	-1.26	7.70	-1.15	7.95	-0.90	-0.85
284	4.09	0.01	-0.61	-0.95	4.19	-0.85	4.39	-0.65	-0.75
285	-	-	-	-	-	-	-	-	-0.58
286	-	-	-	-	-	-	-	-	-0.31
287	-5.81	0.01	9.24	-1.00	-5.46	-0.66	-	-	-0.08
288	7.05	0.00	-1.69	0.92	7.03	1.15	6.85	0.97	0.74
289	6.60	0.00	-1.69	0.72	6.60	0.73	6.47	0.59	0.31
290	5.67	-0.04	-1.28	0.16	5.66	0.15	5.57	0.06	-0.1
291	4.96	0.01	-1.77	-0.99	5.09	-0.87	5.49	-0.46	-0.91
292	4.80	0.00	-1.37	-0.76	4.74	-0.82	4.96	-0.60	-0.98
293	4.49	0.18	-1.73	-1.26	4.65	-1.10	4.81	-0.93	-1.36
294	3.88	0.01	-1.27	-1.58	3.98	-1.47	4.40	-1.06	-1.43

Compound	*IP*(1)	ΔG°_{RRHO}	$\Delta\delta G^{\circ}_{solv}$	E(1)	*IP*(2)	E(2)	*IP*(3)	E(3)	ref.
295	3.97	0.01	-1.40	-1.61	4.11	-1.48	4.25	-1.33	-1.57
296	3.57	0.20	-1.25	-1.67	-1918.22	-1923.45	-1944.27	-1949.50	-1.85
297	4.54	0.15	-1.62	-1.12	4.59	-1.06	4.67	-0.98	-1.41
298	0.10	0.20	2.16	-1.73	0.30	-1.53	0.06	-1.77	-2.18
299	-0.07	0.01	1.75	-2.50	0.01	-2.42	-	-	-2.3
300	0.07	0.18	1.73	-2.20	0.21	-2.06	-	-	-2.5
301	6.14	0.04	-1.41	0.58	6.12	0.56	6.42	0.86	0.55
302	6.15	0.04	-1.40	0.61	6.13	0.59	6.41	0.86	0.46
303	4.10	0.00	-0.88	-0.96	4.03	-1.04	3.87	-1.19	-1.49
304	0.82	0.01	1.59	-1.78	0.96	-1.63	-	-	-1.73
305	11.92	0.00	-6.26	1.48	11.85	1.41	11.84	1.39	0.84
306	-5.13	0.10	9.45	0.22	-4.93	0.42	-4.27	1.09	-0.07
307	-	-	-	-	-	-	-	-	-0.03
308	7.92	0.00	-1.19	2.30	7.89	2.27	8.38	2.76	2.41
309	7.78	-0.05	-1.17	2.14	7.79	2.15	8.34	2.70	2.42
310	0.61	0.01	1.71	-2.10	0.76	-1.95	-	-	-2.17
311	6.88	0.00	-3.39	-0.93	6.95	-0.86	6.60	-1.20	-1.42
312	5.96	0.00	-0.81	0.73	5.99	0.77	6.27	1.05	0.72
313	12.14	0.00	-6.28	1.43	12.06	1.35	12.08	1.37	1.39

PW6B95-D4 and PWPB95-D4

Table A9.11.: Redox potentials and ionization potentials calculated by DFT methods (4=PW6B95-D4/def2-QZVPP, 5=PW6B95-D4/ma-def2-QZVPP, 6=PW6B95-D4/def2-QZVPP, 7=PW6B95-D4/ma-def2-QZVPP) with experimental reference redox potentials (ref.).

Compound	*IP*(4)	E(4)	*IP*(5)	E(5)	*IP*(6)	E(6)	*IP*(7)	E(7)	ref.
1	7.93	1.80	7.93	1.81	7.95	1.83	7.95	1.83	1.64
2	7.78	1.75	7.78	1.75	7.83	1.79	7.83	1.80	1.65
3	7.55	1.81	7.55	1.81	7.64	1.90	7.64	1.90	1.84
4	8.26	2.10	8.26	2.10	8.28	2.13	8.28	2.13	2.07
5	8.40	2.12	8.40	2.12	8.42	2.14	8.42	2.14	2.22
6	8.44	2.13	8.44	2.13	8.46	2.15	8.46	2.15	2.22
7	8.57	2.36	8.57	2.36	8.63	2.41	8.63	2.41	2.27
8	8.70	2.35	8.70	2.35	8.71	2.36	8.71	2.36	2.36
9	7.77	1.74	7.77	1.74	7.82	1.80	7.82	1.80	1.15
10	7.16	1.18	7.16	1.18	7.23	1.24	7.23	1.25	1.15
11	7.21	1.29	7.21	1.29	7.26	1.34	7.26	1.35	1.24
12	7.16	1.34	7.16	1.34	7.25	1.43	7.25	1.43	1.43
13	7.72	1.56	7.72	1.56	7.78	1.62	7.78	1.62	1.49
14	7.36	1.43	7.36	1.44	7.42	1.50	7.42	1.50	1.48
15	7.76	1.89	7.76	1.89	7.83	1.96	7.83	1.96	1.54
16	9.19	2.67	9.19	2.68	9.17	2.65	9.17	2.65	1.54
17	7.57	1.56	7.57	1.56	7.64	1.62	7.64	1.62	1.59
18	7.62	1.69	7.62	1.69	7.68	1.75	7.69	1.76	1.6
19	7.63	1.66	7.63	1.66	7.71	1.74	7.71	1.74	1.64
20	7.43	1.55	7.43	1.55	7.53	1.64	7.53	1.64	1.64

Compound	*IP*(4)	E(4)	*IP*(5)	E(5)	*IP*(6)	E(6)	*IP*(7)	E(7)	ref.
21	7.49	1.61	7.49	1.61	7.56	1.68	7.56	1.68	1.65
22	7.61	1.74	7.61	1.74	7.71	1.84	7.71	1.84	1.66
23	7.92	1.72	7.92	1.73	7.96	1.76	7.96	1.77	1.66
24	7.70	1.66	7.70	1.66	7.76	1.72	7.76	1.72	1.7
25	7.83	1.68	7.83	1.68	7.89	1.73	7.89	1.74	1.71
26	7.69	1.65	7.70	1.65	7.76	1.72	7.76	1.72	1.73
27	7.80	1.74	7.80	1.74	7.87	1.81	7.87	1.81	1.73
28	7.84	1.73	7.84	1.73	7.90	1.79	7.90	1.79	1.74
29	7.78	1.76	7.78	1.76	7.85	1.84	7.86	1.84	1.75
30	8.34	1.90	8.34	1.90	8.44	1.99	8.44	1.99	1.75
31	8.02	1.90	8.02	1.90	8.09	1.97	8.09	1.97	1.89
32	8.08	1.93	8.08	1.93	8.13	1.98	8.13	1.98	1.91
33	8.00	1.93	8.01	1.93	8.08	2.01	8.08	2.01	1.92
34	8.03	2.05	8.03	2.05	8.09	2.11	8.09	2.11	1.94
35	7.94	1.87	7.95	1.87	8.01	1.94	8.02	1.94	1.94
36	8.29	1.93	8.29	1.93	8.37	2.01	8.38	2.01	1.94
37	8.22	1.98	8.22	1.98	8.26	2.02	8.26	2.02	1.97
38	7.91	1.67	7.91	1.67	8.00	1.76	8.00	1.76	1.98
39	8.43	1.84	8.43	1.84	8.51	1.92	8.51	1.92	1.98
40	8.68	2.10	8.68	2.10	8.80	2.22	8.80	2.22	2.1
41	8.57	2.14	8.57	2.14	8.63	2.20	8.63	2.20	2.22
42	8.77	2.15	8.77	2.15	8.87	2.25	8.87	2.25	2.32
43	8.71	2.19	8.71	2.19	8.79	2.27	8.79	2.27	2.37
44	8.84	2.41	8.84	2.41	8.90	2.47	8.90	2.47	2.5
45	7.67	1.20	7.68	1.21	7.68	1.21	7.69	1.22	1.14
46	7.47	1.23	7.47	1.24	7.48	1.24	7.48	1.25	1.17
47	7.57	1.42	7.58	1.43	7.60	1.45	7.60	1.45	1.35
48	7.65	1.50	7.66	1.50	7.68	1.52	7.68	1.53	1.4
49	7.72	1.46	7.73	1.47	7.74	1.47	7.74	1.48	1.41
50	7.80	1.55	7.80	1.55	7.82	1.57	7.82	1.57	1.43
51	8.34	1.84	8.34	1.85	8.34	1.84	8.35	1.85	1.63
52	8.15	1.84	8.16	1.84	8.18	1.87	8.19	1.87	1.69
53	8.82	2.28	8.83	2.29	8.86	2.32	8.86	2.32	2.08
54	-	-	-	-	7.29	1.27	7.30	1.27	1.3
55	7.53	1.56	7.53	1.56	7.55	1.58	7.55	1.58	1.32
56	7.43	1.41	7.43	1.42	7.44	1.43	7.45	1.43	1.43
57	7.31	1.52	7.32	1.52	7.33	1.54	7.33	1.54	1.47
58	7.84	1.78	7.85	1.79	7.85	1.79	7.85	1.80	1.5
59	8.13	1.60	8.13	1.60	8.16	1.64	8.17	1.64	1.51
60	8.08	1.84	8.08	1.84	8.09	1.85	8.09	1.85	1.81
61	7.81	1.92	7.82	1.92	7.88	1.98	7.88	1.98	1.88
62	8.78	2.19	8.79	2.20	8.85	2.26	8.85	2.27	2.5
63	6.44	0.46	6.44	0.46	6.49	0.50	6.49	0.50	0.32
64	7.02	0.65	7.02	0.65	7.05	0.68	7.05	0.68	0.48
65	6.67	0.59	6.67	0.59	6.73	0.66	6.74	0.66	0.54
66	7.07	0.94	7.07	0.94	7.09	0.96	7.10	0.96	0.74
67	7.24	0.96	7.24	0.96	7.26	0.98	7.27	0.99	0.81
68	7.23	0.97	7.24	0.97	7.27	1.01	7.27	1.01	0.83

Compound	*IP*(4)	E(4)	*IP*(5)	E(5)	*IP*(6)	E(6)	*IP*(7)	E(7)	ref.
69	7.07	1.05	7.07	1.05	7.09	1.07	7.10	1.08	0.84
70	7.36	1.15	7.36	1.15	7.38	1.17	7.38	1.17	0.86
71	6.99	0.93	7.00	0.93	7.05	0.98	7.05	0.99	0.88
72	-297.23	-303.96	-296.93	-303.66	-301.15	-307.88	-300.86	-307.58	0.89
73	7.05	1.08	7.05	1.08	7.09	1.12	7.10	1.13	0.92
74	7.80	1.28	7.80	1.28	7.85	1.33	7.85	1.34	0.94
75	7.54	1.13	7.55	1.13	7.56	1.14	7.56	1.15	0.95
76	7.67	1.35	7.67	1.36	7.71	1.40	7.71	1.40	1.09
77	7.34	1.32	7.34	1.32	7.41	1.39	7.41	1.39	1.28
78	7.42	1.32	7.42	1.32	7.44	1.34	7.44	1.34	1.15
79	7.54	1.37	7.54	1.37	7.58	1.41	7.58	1.42	1.16
80	7.05	1.51	7.06	1.51	7.12	1.57	7.12	1.58	1.28
81	7.59	1.55	7.59	1.56	7.63	1.60	7.63	1.60	1.33
82	7.83	1.71	7.83	1.72	7.85	1.74	7.85	1.74	1.44
83	7.72	1.56	7.72	1.56	7.75	1.59	7.75	1.59	1.44
84	7.32	1.67	7.32	1.68	7.38	1.73	7.38	1.74	1.46
85	8.12	1.79	8.12	1.79	8.14	1.81	8.14	1.81	1.51
86	7.63	1.83	7.63	1.83	7.66	1.85	7.66	1.86	1.56
87	8.62	2.48	8.62	2.48	8.61	2.47	8.61	2.47	2.02
88	7.53	1.43	7.53	1.43	7.54	1.44	7.54	1.45	0.81
89	7.07	1.05	7.08	1.06	7.09	1.07	7.09	1.07	0.91
90	7.82	1.29	7.82	1.29	7.85	1.32	7.86	1.33	0.95
91	7.08	1.10	7.08	1.10	7.09	1.11	7.10	1.12	0.97
92	7.77	1.71	7.77	1.71	7.79	1.73	7.79	1.73	1.03
93	7.24	1.23	7.24	1.24	7.23	1.23	7.24	1.23	1.04
94	7.93	1.49	7.93	1.49	7.91	1.47	7.91	1.47	1.04
95	7.97	1.82	7.97	1.83	8.00	1.86	8.01	1.86	1.07
96	8.77	1.97	8.78	1.97	8.75	1.95	8.76	1.96	1.15
97	7.60	1.39	7.61	1.39	7.62	1.40	7.62	1.41	1.16
98	7.99	1.81	7.99	1.82	8.01	1.83	8.01	1.84	1.62
99	8.41	1.82	8.42	1.82	8.44	1.85	8.45	1.85	1.73
100	8.16	1.89	8.17	1.89	8.19	1.91	8.19	1.91	1.74
101	8.78	2.01	8.78	2.01	8.78	2.00	8.78	2.01	1.94
102	8.48	2.32	8.48	2.32	8.52	2.36	8.52	2.36	2.14
103	9.29	2.49	9.29	2.50	9.27	2.47	9.27	2.48	2.21
104	-	-	-	-	-	-	-	-	-2.48
105	1.03	-0.86	1.08	-0.82	0.85	-1.04	0.89	-1.00	-1.93
106	0.54	-1.62	0.57	-1.59	0.40	-1.75	0.42	-1.73	-1.94
107	0.56	-1.62	0.59	-1.59	0.43	-1.75	0.45	-1.73	-1.95
108	0.55	-1.56	0.59	-1.52	0.28	-1.83	0.31	-1.79	-1.92
109	0.91	-1.37	0.94	-1.34	0.77	-1.51	0.80	-1.48	-1.82
110	1.32	-0.89	1.35	-0.86	1.12	-1.09	1.15	-1.06	-1.46
111	1.09	-1.12	1.14	-1.07	0.86	-1.34	0.91	-1.29	-1.23
112	1.05	-1.02	1.09	-0.98	0.85	-1.22	0.89	-1.18	-1.06
113	1.60	-0.62	1.62	-0.59	1.38	-0.83	1.40	-0.81	-0.62
114	8.40	2.32	8.40	2.32	8.46	2.38	8.47	2.39	2.1
115	8.45	2.31	8.45	2.32	8.50	2.37	8.50	2.37	2.15
116	8.55	2.35	8.56	2.36	8.59	2.39	8.59	2.39	2.17

Compound	*IP*(4)	E(4)	*IP*(5)	E(5)	*IP*(6)	E(6)	*IP*(7)	E(7)	ref.
117	8.80	2.51	8.80	2.51	8.83	2.53	8.83	2.54	2.5
118	8.90	2.53	8.90	2.54	8.91	2.55	8.91	2.55	2.61
119	-0.93	-2.81	-0.71	-2.60	-1.06	-2.95	-0.94	-2.82	-2.28
120	-0.96	-2.88	-0.72	-2.63	-1.12	-3.04	-0.90	-2.82	-2.24
121	0.27	-1.91	0.30	-1.89	0.13	-2.05	0.15	-2.03	-1.94
122	0.33	-1.82	0.36	-1.79	0.18	-1.97	0.21	-1.94	-1.93
123	0.60	-1.70	0.63	-1.67	0.45	-1.85	0.48	-1.82	-1.85
124	0.68	-1.71	0.70	-1.69	0.52	-1.86	0.55	-1.84	-1.73
125	0.81	-1.68	0.83	-1.65	0.65	-1.83	0.67	-1.81	-1.72
126	0.99	-1.45	1.01	-1.43	0.82	-1.62	0.85	-1.60	-1.66
127	1.32	-1.20	1.34	-1.18	1.15	-1.37	1.17	-1.35	-1.42
128	1.84	-0.65	1.87	-0.62	1.60	-0.89	1.63	-0.86	-0.86
129	8.74	2.35	8.75	2.35	8.76	2.36	8.76	2.37	1.95
130	8.48	2.28	8.48	2.28	8.49	2.29	8.49	2.29	2.06
131	-	-	-	-	0.41	-2.02	0.43	-2.00	-2.01
132	0.61	-1.82	0.63	-1.80	0.46	-1.97	0.48	-1.95	-1.91
133	0.68	-1.83	0.69	-1.81	0.53	-1.98	0.54	-1.96	-1.9
134	0.65	-1.80	0.67	-1.78	0.50	-1.95	0.52	-1.93	-1.88
135	6.64	1.10	6.64	1.10	6.70	1.16	6.71	1.16	0.97
136	7.35	1.56	7.35	1.57	7.42	1.63	7.42	1.63	1.5
137	7.37	1.67	7.37	1.67	7.43	1.72	7.43	1.72	1.52
138	7.60	1.74	7.61	1.74	7.67	1.80	7.67	1.80	1.64
139	7.58	1.76	7.58	1.76	7.65	1.82	7.65	1.83	1.67
140	8.22	1.99	8.23	1.99	8.30	2.07	8.31	2.07	1.86
141	-0.95	-2.98	-0.61	-2.64	-1.10	-3.13	-0.75	-2.78	-2.33
142	0.11	-2.11	0.15	-2.06	-0.03	-2.25	0.01	-2.21	-2.16
143	0.23	-2.04	0.27	-2.00	0.08	-2.19	0.12	-2.16	-2.13
144	0.19	-2.00	0.23	-1.96	0.05	-2.14	0.08	-2.10	-2.11
145	0.46	-1.88	0.50	-1.84	0.31	-2.02	0.35	-1.98	-1.91
146	0.68	-1.82	0.70	-1.80	0.52	-1.98	0.55	-1.95	-1.91
147	0.51	-1.71	0.55	-1.67	0.37	-1.85	0.41	-1.82	-1.89
148	0.60	-1.63	0.62	-1.60	0.46	-1.77	0.48	-1.74	-1.85
149	0.84	-1.63	0.86	-1.60	0.68	-1.78	0.70	-1.76	-1.74
150	1.69	-0.77	1.73	-0.74	1.46	-1.00	1.49	-0.97	-0.93
151	8.81	2.66	8.81	2.66	8.94	2.78	8.94	2.78	2.66
152	6.14	0.84	6.14	0.84	6.19	0.88	6.19	0.88	1.26
153	6.97	1.70	6.98	1.71	7.12	1.85	7.12	1.85	1.4
154	6.98	1.69	6.98	1.69	7.07	1.78	7.07	1.78	1.47
155	8.12	1.87	8.13	1.87	8.19	1.94	8.19	1.94	1.55
156	9.13	2.40	9.14	2.40	9.17	2.43	9.17	2.43	2.26
157	9.09	2.63	9.10	2.64	9.19	2.73	9.20	2.74	2.58
158	0.01	-2.24	0.04	-2.20	-0.12	-2.37	-0.09	-2.34	-2.43
159	-0.03	-2.20	0.00	-2.17	-0.16	-2.34	-0.13	-2.31	-2.34
160	0.07	-2.17	0.11	-2.12	-0.07	-2.30	-0.02	-2.26	-2.29
161	-4.65	-6.45	-4.50	-6.30	-4.86	-6.66	-4.72	-6.52	-2.13
162	-	-	-	-	0.72	-1.89	0.74	-1.87	-1.9
163	1.00	-1.49	1.02	-1.47	0.85	-1.64	0.86	-1.63	-1.72
164	0.97	-1.52	0.99	-1.50	0.81	-1.68	0.83	-1.66	-1.7

Compound	*IP*(4)	E(4)	*IP*(5)	E(5)	*IP*(6)	E(6)	*IP*(7)	E(7)	ref.
165	1.17	-1.43	1.19	-1.41	1.01	-1.59	1.03	-1.57	-1.64
166	1.55	-0.56	1.58	-0.53	1.39	-0.72	1.42	-0.70	-1.45
167	1.54	-0.54	1.57	-0.51	1.37	-0.71	1.41	-0.67	-1.37
168	1.23	-1.14	1.25	-1.11	1.07	-1.29	1.09	-1.27	-1.34
169	0.74	-1.53	0.77	-1.50	0.59	-1.68	0.62	-1.65	-1.2
170	0.70	-1.80	0.72	-1.77	0.49	-2.01	0.51	-1.99	-1.01
171	1.53	-0.68	1.56	-0.66	1.35	-0.87	1.37	-0.84	-0.98
172	2.05	-0.25	2.08	-0.22	1.87	-0.43	1.90	-0.40	-0.51
173	2.11	-0.33	2.14	-0.30	1.93	-0.51	1.96	-0.48	-0.45
174	1.10	-1.17	1.13	-1.14	0.83	-1.44	0.86	-1.41	-0.2
175	6.55	1.55	6.55	1.55	6.53	1.53	6.53	1.53	0.26
176	7.00	1.88	7.01	1.89	6.98	1.86	6.99	1.87	0.71
177	7.17	1.93	7.18	1.94	7.21	1.96	7.22	1.97	1.01
178	1.91	-0.35	1.93	-0.33	1.72	-0.54	1.74	-0.52	-1.12
179	1.72	-0.41	1.75	-0.39	1.56	-0.58	1.58	-0.55	-0.27
180	3.03	0.27	3.05	0.29	2.85	0.09	2.86	0.10	-0.26
181	2.09	-0.90	2.11	-0.88	1.91	-1.07	1.93	-1.06	-1.1
182	-	-	-	-	-	-	-	-	-0.86
183	2.06	-0.71	2.06	-0.70	1.93	-0.84	1.93	-0.83	-0.59
184	5.11	-0.20	5.12	-0.19	4.98	-0.33	4.98	-0.33	-0.32
185	2.59	-0.27	2.60	-0.26	2.46	-0.41	2.46	-0.40	-0.28
186	0.50	-1.78	0.52	-1.76	0.38	-1.90	0.40	-1.89	-1.92
187	1.13	-1.11	1.15	-1.09	0.99	-1.25	1.01	-1.23	-1.36
188	0.61	-1.38	0.63	-1.36	0.45	-1.54	0.47	-1.51	-1.88
189	1.01	-0.88	1.06	-0.83	0.79	-1.10	0.84	-1.05	-1.15
190	3.70	0.87	3.71	0.88	3.47	0.64	3.48	0.65	0.13
191	-0.50	-0.05	-0.47	-0.02	-0.51	-0.06	-0.48	-0.03	-0.29
192	10.88	1.39	10.89	1.39	10.75	1.25	10.75	1.25	0.66
193	6.20	0.55	6.21	0.56	6.25	0.59	6.25	0.60	0.3
194	-	-	-	-	-	-0.69	-	-	-1.28
195	-	-	-	-	-	-	-	-	-0.24
196	-	-	-	-	-	-	-	-	0.36
197	4.49	-1.16	4.49	-1.16	4.68	-0.98	4.68	-0.98	-0.81
198	4.87	-0.52	4.87	-0.52	4.46	-0.93	4.47	-0.93	-0.32
199	5.56	-0.32	5.56	-0.31	5.75	-0.12	5.75	-0.12	0.13
200	4.41	-0.97	4.41	-0.97	3.71	-1.67	3.71	-1.67	-1.24
201	5.34	-0.43	5.34	-0.43	5.10	-0.66	5.10	-0.66	-0.41
202	11.83	-0.58	11.83	-0.58	11.69	-0.71	11.69	-0.71	-1.14
203	13.54	0.83	13.54	0.83	13.44	0.73	13.44	0.73	0.62
204	13.21	0.59	13.21	0.59	13.12	0.50	13.12	0.50	0.13
205	12.64	-0.06	12.64	-0.06	12.30	-0.40	12.30	-0.40	-0.4
206	14.24	1.59	14.24	1.59	14.06	1.41	14.06	1.41	0.95
207	12.64	-0.85	12.64	-0.85	12.51	-0.98	12.51	-0.98	-1.03
208	13.26	-0.04	13.26	-0.04	12.94	-0.36	12.94	-0.36	-0.23
209	10.49	-0.55	10.49	-0.55	10.45	-0.59	10.45	-0.59	-0.71
210	-	-	-	-	-	-	-	-	1.33
211	-	-	-	-	-	-	-	-	1.01
212	11.58	0.26	11.58	0.26	11.28	-0.05	11.28	-0.05	0.51

Compound	*IP*(4)	E(4)	*IP*(5)	E(5)	*IP*(6)	E(6)	*IP*(7)	E(7)	ref.
213	13.23	2.00	13.23	2.00	13.02	1.78	13.02	1.78	1.81
214	-	-	-	-	-	-	-	-	-1.33
215	5.59	0.04	5.60	0.05	5.51	-0.04	5.52	-0.03	0.46
216	5.69	0.10	5.70	0.11	5.70	0.11	5.70	0.11	0.13
217	5.22	-0.37	5.22	-0.37	5.00	-0.59	5.00	-0.59	-0.13
218	6.96	1.38	6.96	1.39	6.82	1.24	6.82	1.24	1.21
219	12.18	1.31	12.18	1.31	-	-	-	-	1.28
220	11.93	1.14	11.93	1.14	11.91	1.13	11.92	1.13	1.12
221	11.29	0.57	11.29	0.57	11.00	0.28	11.00	0.28	0.31
222	10.22	-0.26	10.22	-0.26	10.18	-0.30	10.18	-0.30	-0.02
223	11.68	1.16	11.68	1.16	11.65	1.14	11.65	1.14	1.13
224	11.08	0.59	11.08	0.59	10.82	0.33	10.82	0.33	0.33
225	10.44	-0.25	10.44	-0.25	10.36	-0.33	10.36	-0.33	0.01
226	11.84	1.08	11.84	1.08	11.82	1.06	11.82	1.06	1.12
227	11.23	0.41	11.23	0.41	10.97	0.15	10.97	0.15	0.31
228	-0.82	0.02	-0.79	0.06	-	-	-	-	0.2
229	2.28	0.00	2.29	0.01	2.24	-0.04	2.25	-0.03	0.35
230	5.05	-0.74	5.05	-0.73	4.94	-0.85	4.94	-0.84	-0.9
231	-	-	-	-	-	-	-	-	-1.12
232	-	-	-	-	-	-	-	-	0.01
233	-	-	-	-	-7.01	0.53	-6.79	0.75	0.57
234	0.96	-1.49	0.97	-1.48	0.75	-1.70	0.76	-1.69	-1.18
235	0.11	-2.41	0.11	-2.40	-	-	-	-	-1.57
236	2.20	-0.13	2.21	-0.12	2.21	-0.12	2.21	-0.11	0.15
237	1.86	-0.56	1.87	-0.56	1.87	-0.56	1.88	-0.55	-0.43
238	2.14	0.09	2.15	0.10	-2.11	-4.16	-2.10	-4.15	-0.11
239	15.39	0.88	15.39	0.89	15.29	0.79	15.29	0.79	-0.58
240	17.63	3.31	17.63	3.31	17.65	3.33	17.65	3.33	1.5
241	16.74	2.54	16.74	2.54	16.71	2.51	16.71	2.51	0.75
242	-	-	-	-	-	-	-	-	1.86
243	0.35	0.66	0.38	0.69	-0.08	0.23	-0.05	0.26	0.78
244	-	-	-	-	-	-	-	-	-0.8
245	-2.46	-0.89	-0.38	1.19	-2.85	-1.28	-0.37	1.20	-0.75
246	-	-	-	-	-	-	-	-	-0.12
247	-2.45	-0.85	-0.60	1.00	-2.53	-0.93	-0.80	0.80	-0.7
248	-	-	-	-	-	-	-	-	-0.88
249	-1.67	-0.11	-1.63	-0.08	-1.38	0.18	-1.35	0.21	-0.53
250	-	-	-	-	-	-	-	-	-1.27
251	-	-	-	-	-2.60	-0.72	-2.57	-0.69	-0.32
252	-	-	-	-	-	-	-	-	-1.5
253	-	-	-	-	-	-	-	-	1.28
254	-	-	-	-	-	-	-	-	-1.61
255	2.43	-0.13	2.44	-0.12	2.39	-0.16	2.40	-0.16	-0.08
256	1.87	-0.64	1.88	-0.63	1.70	-0.80	1.71	-0.79	-0.64
257	-1.07	-0.33	-0.90	-0.16	-1.06	-0.31	-0.89	-0.15	-0.37
258	2.68	0.39	2.69	0.40	2.53	0.24	2.54	0.25	0.02
259	-1.52	-0.45	-1.49	-0.42	-1.65	-0.57	-1.62	-0.54	-1.04
260	-1.67	-0.84	-1.63	-0.80	-1.82	-0.99	-1.78	-0.95	-0.98

Compound	*IP*(4)	E(4)	*IP*(5)	E(5)	*IP*(6)	E(6)	*IP*(7)	E(7)	ref.
261	-0.59	0.47	-0.55	0.51	-0.59	0.47	-0.56	0.50	0.12
262	-0.18	0.49	-0.14	0.53	-0.39	0.28	-0.36	0.31	0.37
263	0.88	1.74	0.91	1.77	0.79	1.65	0.82	1.68	1.29
264	2.59	0.37	2.60	0.38	2.43	0.22	2.44	0.23	-0.02
265	-1.73	-0.87	-1.70	-0.84	-1.88	-1.02	-1.85	-0.99	-0.96
266	-	-	-	-	-0.61	0.42	-0.58	0.44	0.04
267	-0.25	0.39	-0.22	0.42	-	-	-	-	0.35
268	-2.05	-1.16	-2.02	-1.13	-2.22	-1.33	-2.18	-1.29	-1.16
269	-0.20	0.38	-0.16	0.41	-0.77	-0.20	-0.74	-0.16	0.36
270	15.24	2.25	15.25	2.25	14.95	1.96	14.95	1.96	1.35
271	3.07	0.28	3.09	0.30	2.87	0.09	2.89	0.10	0.2
272	-	-	-	-	-	-	-	-	0.68
273	4.53	-0.86	4.53	-0.86	4.51	-0.87	4.52	-0.87	0.78
274	4.36	-1.02	4.37	-1.02	4.33	-1.06	4.33	-1.06	-1.12
275	4.60	-0.94	4.61	-0.94	4.56	-0.98	4.57	-0.97	-1.22
276	-	-	-	-	-1.36	-0.03	-0.70	0.63	0.54
277	1.66	-1.22	1.67	-1.21	1.57	-1.32	1.58	-1.31	-1.03
278	-	-	-	-	-4.33	0.90	-4.27	0.96	0.6
279	-4.34	0.67	-4.27	0.74	-6.13	-1.11	-6.06	-1.05	-0.07
280	-0.01	0.53	0.02	0.55	-0.03	0.50	-0.01	0.53	0.16
281	9.58	1.28	9.58	1.28	9.46	1.16	9.46	1.16	1.19
282	9.15	0.83	9.15	0.83	9.11	0.79	9.11	0.79	0.81
283	7.95	-0.90	7.95	-0.90	7.82	-1.04	7.82	-1.04	-0.85
284	4.38	-0.66	4.39	-0.65	4.26	-0.78	4.26	-0.78	-0.75
285	-	-	-	-	-	-	-	-	-0.58
286	-	-	-	-	-	-	-	-	-0.31
287	-	-	-	-	-	-	-	-	-0.08
288	6.85	0.97	6.85	0.97	7.01	1.13	7.01	1.13	0.74
289	6.47	0.59	6.47	0.59	6.62	0.74	6.62	0.74	0.31
290	5.57	0.06	5.57	0.06	5.76	0.25	5.76	0.26	-0.1
291	5.49	-0.46	5.49	-0.46	4.82	-1.13	4.83	-1.13	-0.91
292	4.96	-0.60	4.96	-0.60	5.12	-0.44	5.12	-0.44	-0.98
293	4.81	-0.94	4.81	-0.93	4.50	-1.25	4.50	-1.25	-1.36
294	4.40	-1.06	4.40	-1.06	3.71	-1.75	3.71	-1.75	-1.43
295	4.25	-1.34	4.25	-1.33	4.02	-1.57	4.02	-1.57	-1.57
296	-1946.62	-1951.85	-1944.27	-1949.50	-1969.51	-1974.74	-1967.20	-1972.43	-1.85
297	4.67	-0.99	4.67	-0.98	4.47	-1.18	4.47	-1.18	-1.41
298	0.04	-1.79	0.06	-1.77	-0.14	-1.97	-0.12	-1.95	-2.18
299	-	-	-	-	-	-	-	-	-2.3
300	-	-	-	-	-	-	-	-	-2.5
301	6.42	0.86	6.42	0.86	6.46	0.90	6.46	0.90	0.55
302	6.40	0.86	6.41	0.86	6.45	0.90	6.45	0.90	0.46
303	3.87	-1.20	3.87	-1.19	-	-	-	-	-1.49
304	-	-	-	-	1.00	-1.59	1.02	-1.57	-1.73
305	11.84	1.39	11.84	1.39	11.86	1.42	11.86	1.42	0.84
306	-4.34	1.02	-4.27	1.09	-6.13	-0.77	-6.06	-0.71	-0.07
307	-	-	-	-	-	-	-	-	-0.03
308	8.38	2.76	8.38	2.76	8.46	2.85	8.47	2.85	2.41

Compound	*IP*(4)	E(4)	*IP*(5)	E(5)	*IP*(6)	E(6)	*IP*(7)	E(7)	ref.
309	8.34	2.70	8.34	2.70	8.44	2.80	8.44	2.80	2.42
310	-	-	-	-	-	-	-	-	-2.17
311	6.60	-1.20	6.60	-1.20	6.50	-1.30	6.50	-1.30	-1.42
312	6.27	1.04	6.27	1.05	6.29	1.07	6.30	1.07	0.72
313	12.08	1.37	12.08	1.37	12.09	1.38	12.09	1.38	1.39

SQM

In the following all energies and redox potentials are given in V.

Table A9.12.: Comparison of GFN*n*-xTB methods with and without fixed Multiplicities.

OROP					
GFN	N	MD	MAD	SD	AMAX
GFN1-xTB	190	0.01	0.33	0.50	2.07
GFN2-xTB	193	0.02	0.30	0.41	1.66
GFN1-xTB (fixed M)	191	0.02	0.33	0.50	2.07
GFN2-xTB (fixed M)	193	0.02	0.30	0.41	1.66
OMROP					
GFN	N	MD	MAD	SD	AMAX
GFN1-xTB	105	-0.27	1.03	1.35	3.51
GFN2-xTB	120	-0.45	0.73	0.83	2.56
GFN1-xTB (fixed M)	100	-0.66	1.29	1.50	4.23
GFN2-xTB (fixed M)	119	-0.74	1.03	1.16	4.49

Table A9.13.: Redox potentials calculated by SQM methods (fixed M indicates that multiplicities were supplied to the corresponding GFN*n*-xTB method) and experimental reference redox potentials (ref.).

Compound	E(GFN1)	E(GFN2)	E(GFN1 (fixed M))	E(GFN2 (fixed M))	E(PM6)	E(PM7)	ref.
1	1.85	1.84	1.85	1.84	1.89	2.04	1.64
2	1.73	1.61	1.73	1.61	1.76	1.92	1.65
3	1.64	1.59	1.64	1.59	2.02	2.18	1.84
4	2.12	1.91	2.12	1.91	1.84	2.06	2.07
5	2.20	2.03	2.20	2.03	1.91	2.11	2.22
6	2.17	2.01	2.17	2.01	1.89	2.12	2.22
7	2.29	2.30	2.29	2.30	2.25	2.41	2.27
8	2.35	2.23	2.35	2.23	2.01	2.24	2.36
9	1.75	1.62	1.75	1.62	1.65	1.91	1.15
10	1.30	1.24	1.30	1.24	1.54	1.70	1.15
11	1.35	1.19	1.35	1.19	1.43	1.61	1.24

Compound	E(GFN1)	E(GFN2)	E(GFN1 (fixed M))	E(GFN2 (fixed M))	E(PM6)	E(PM7)	ref.
12	1.30	1.13	1.30	1.13	1.49	1.60	1.43
13	1.49	1.47	1.49	1.47	1.48	1.57	1.49
14	1.44	1.25	1.44	1.25	1.48	1.63	1.48
15	1.79	1.71	1.79	1.71	2.43	2.22	1.54
16	2.76	2.62	2.76	2.62	2.25	2.43	1.54
17	1.57	1.40	1.57	1.40	1.60	1.72	1.59
18	1.66	1.51	1.66	1.51	1.75	1.90	1.60
19	1.64	1.45	1.64	1.45	1.71	1.78	1.64
20	1.43	1.23	1.43	1.23	1.70	1.86	1.64
21	1.59	1.42	1.59	1.42	1.66	1.77	1.65
22	1.56	1.39	1.56	1.39	1.79	1.88	1.66
23	1.85	1.79	1.85	1.79	1.76	1.93	1.66
24	1.65	1.50	1.65	1.50	1.72	1.81	1.70
25	1.79	1.79	1.79	1.79	1.87	1.94	1.71
26	1.69	1.54	1.69	1.54	1.70	1.88	1.73
27	1.57	1.58	1.57	1.58	1.82	1.98	1.73
28	1.75	1.63	1.75	1.63	1.70	1.90	1.74
29	1.63	1.73	1.63	1.73	1.80	1.97	1.75
30	1.85	1.50	1.85	1.50	1.57	1.72	1.75
31	1.79	1.76	1.79	1.76	1.88	2.05	1.89
32	1.96	1.89	1.96	1.89	1.99	2.09	1.91
33	1.78	1.88	1.78	1.88	1.89	1.98	1.92
34	2.00	1.89	2.00	1.89	1.96	2.07	1.94
35	1.68	1.68	1.68	1.68	1.87	2.08	1.94
36	1.93	1.61	1.93	1.61	1.49	1.64	1.94
37	2.06	2.05	2.06	2.05	1.93	2.10	1.97
38	1.66	1.46	1.66	1.46	1.42	1.66	1.98
39	1.96	1.67	1.96	1.67	1.48	1.72	1.98
40	1.95	1.66	1.95	1.66	1.78	1.94	2.10
41	2.22	2.10	2.22	2.10	2.03	2.13	2.22
42	2.10	1.72	2.10	1.72	1.87	1.96	2.32
43	2.20	1.87	2.20	1.87	1.80	1.90	2.37
44	2.34	2.07	2.34	2.07	1.84	2.04	2.50
45	1.62	1.46	1.62	1.46	1.54	1.70	1.14
46	1.53	1.36	1.53	1.36	1.47	1.66	1.17
47	1.61	1.35	1.61	1.35	1.51	1.64	1.35
48	1.64	1.57	1.64	1.57	1.69	1.83	1.40
49	1.65	1.52	1.65	1.52	1.56	1.72	1.41
50	1.75	1.55	1.75	1.55	1.56	1.77	1.43
51	2.14	2.04	2.14	2.04	1.83	2.04	1.63
52	1.93	2.06	1.93	2.06	1.98	2.15	1.69
53	2.32	2.47	2.32	2.47	2.19	2.36	2.08
54	1.45	1.28	1.45	1.28	1.43	1.59	1.30
55	1.59	1.51	1.59	1.51	1.61	1.80	1.32
56	1.56	1.35	1.56	1.35	1.45	1.59	1.43
57	1.67	1.49	1.67	1.49	1.52	1.67	1.47
58	1.87	1.73	1.87	1.73	1.79	1.92	1.50
59	1.89	1.69	1.89	1.69	1.46	1.57	1.51

Compound	E(GFN1)	E(GFN2)	E(GFN1 (fixed M))	E(GFN2 (fixed M))	E(PM6)	E(PM7)	ref.
60	1.99	1.87	1.99	1.87	1.76	1.95	1.81
61	1.75	1.65	1.75	1.65	1.95	2.05	1.88
62	2.13	2.03	2.13	2.03	2.42	2.48	2.50
63	0.60	0.38	0.60	0.38	0.30	0.44	0.32
64	1.04	0.72	1.04	0.72	0.64	0.73	0.48
65	0.70	0.60	0.70	0.60	0.64	0.65	0.54
66	1.02	0.91	1.02	0.91	0.63	0.88	0.74
67	1.33	1.07	1.33	1.07	0.83	1.02	0.81
68	642.38	1.00	1.14	1.00	0.52	0.71	0.83
69	1.39	1.10	1.39	1.10	0.80	1.04	0.84
70	1.50	1.28	1.50	1.28	1.13	1.25	0.86
71	0.94	0.91	0.94	0.91	0.92	1.02	0.88
72	1.75	1.37	1.75	1.37	0.73	0.89	0.89
73	1.20	1.05	1.20	1.05	1.04	1.20	0.92
74	1.55	1.25	1.55	1.25	0.60	0.78	0.94
75	1.55	1.36	1.55	1.36	0.97	1.16	0.95
76	1.59	1.30	1.59	1.30	0.82	1.00	1.09
77	1.28	1.19	1.28	1.19	1.06	1.22	1.28
78	1.24	1.31	1.24	1.31	1.50	1.49	1.15
79	1.06	1.33	1.06	1.33	1.61	1.57	1.16
80	1.02	0.98	1.03	0.99	1.43	1.09	1.28
81	1.39	1.49	1.39	1.49	1.56	1.61	1.33
82	1.54	1.53	1.54	1.53	1.44	1.56	1.44
83	1.56	1.45	1.56	1.45	1.26	1.35	1.44
84	1.28	1.13	1.28	1.14	1.39	1.26	1.46
85	1.72	1.79	1.72	1.79	1.62	1.65	1.51
86	1.46	1.30	1.46	1.30	1.62	1.17	1.56
87	2.08	1.92	2.08	1.92	2.01	1.43	2.02
88	1.25	1.52	1.25	1.52	1.89	1.87	0.81
89	1.24	1.11	1.24	1.11	1.35	1.45	0.91
90	1.65	1.71	1.65	1.71	1.40	1.49	0.95
91	1.20	1.03	1.20	1.03	1.29	1.41	0.97
92	1.48	1.76	1.48	1.76	1.86	1.85	1.03
93	1.36	1.23	1.36	1.23	1.33	1.51	1.04
94	1.85	1.95	1.85	1.95	1.62	1.59	1.04
95	1.65	1.92	1.65	1.92	2.07	2.05	1.07
96	2.42	2.57	2.42	2.57	1.97	2.08	1.15
97	1.57	1.50	1.57	1.50	1.37	1.54	1.16
98	1.78	1.78	1.78	1.78	1.65	1.67	1.62
99	1.99	2.03	1.99	2.03	2.01	1.98	1.73
100	2.02	2.09	2.02	2.09	1.99	2.07	1.74
101	2.45	2.78	2.45	2.78	2.03	2.07	1.94
102	2.12	1.79	2.12	1.79	2.24	2.38	2.14
103	2.66	2.90	2.66	2.90	2.27	2.41	2.21
104	-3.98	-1.81	-3.98	-1.81	0.07	0.20	-2.48
105	-2.80	-1.68	-0.91	-1.68	0.50	0.79	-1.93
106	-3.42	-2.60	-3.42	-2.61	-0.19	-0.87	-1.94
107	-3.57	-2.48	-3.57	-2.48	-0.18	-0.91	-1.95

Compound	E(GFN1)	E(GFN2)	E(GFN1 (fixed M))	E(GFN2 (fixed M))	E(PM6)	E(PM7)	ref.
108	-2.13	-2.07	-2.13	-2.07	-0.03	-0.04	-1.92
109	-3.89	-1.82	-3.89	-1.82	0.05	0.64	-1.82
110	-1.89	-1.20	-1.89	-1.20	0.08	0.04	-1.46
111	-1.23	-1.00	-1.23	-1.00	0.34	0.01	-1.23
112	-2.50	-1.30	-2.50	-1.30	0.08	0.06	-1.06
113	-1.73	-0.73	-1.73	-0.72	0.45	0.11	-0.62
114	1.99	2.53	1.99	2.53	2.64	1.85	2.10
115	1.85	2.38	1.85	2.38	2.54	1.86	2.15
116	2.17	2.60	2.17	2.60	2.50	1.89	2.17
117	2.47	2.68	2.47	2.68	2.50	2.70	2.50
118	2.22	2.47	2.22	2.47	2.38	2.75	2.61
119	-3.22	-3.41	-3.23	-3.41	-1.07	-1.36	-2.28
120	-3.27	-3.41	-3.27	-3.41	-1.16	-1.43	-2.24
121	-2.02	-2.21	-2.02	-2.21	-0.63	-0.87	-1.94
122	-1.98	-2.07	-1.98	-2.07	-0.53	-0.76	-1.93
123	-1.85	-1.82	-1.85	-1.82	-0.41	-0.57	-1.85
124	-1.73	-1.86	-1.73	-1.86	-0.37	-0.59	-1.73
125	-1.73	-1.89	-1.73	-1.89	-0.49	-0.70	-1.72
126	-1.50	-1.43	-1.50	-1.43	-0.27	-0.37	-1.66
127	-1.41	-1.31	-1.41	-1.31	-0.26	-0.35	-1.42
128	-0.65	-0.52	-0.65	-0.52	0.46	0.13	-0.86
129	2.51	2.38	2.51	2.38	2.19	2.31	1.95
130	2.35	2.22	2.35	2.22	2.07	2.23	2.06
131	-2.05	-2.36	-2.05	-2.36	-0.75	-0.68	-2.01
132	-2.02	-2.29	-2.02	-2.30	-0.75	-0.68	-1.91
133	-1.95	-2.13	-1.95	-2.13	-0.61	-0.62	-1.90
134	-2.01	-2.28	-2.01	-2.28	-0.64	-0.62	-1.88
135	0.91	0.90	0.91	0.90	1.16	1.21	0.97
136	1.42	1.11	1.42	1.11	1.42	1.51	1.50
137	1.51	1.22	1.51	1.22	1.49	1.64	1.52
138	1.61	1.44	1.60	1.41	1.49	1.64	1.64
139	1.54	1.28	1.51	1.25	1.44	1.61	1.67
140	2.14	2.09	2.14	2.17	2.17	2.34	1.86
141	-3.43	-3.75	-3.43	-3.75	-1.47	-1.72	-2.33
142	-2.18	-2.46	-2.18	-2.46	-0.77	-0.89	-2.16
143	-2.03	-2.12	-2.03	-2.12	-0.58	-0.69	-2.13
144	-2.12	-2.36	-2.12	-2.36	-0.70	-0.83	-2.11
145	-1.98	-2.09	-1.98	-2.09	-0.58	-0.61	-1.91
146	-1.84	-2.12	-1.84	-2.12	-0.66	-0.76	-1.91
147	-2.05	-1.90	-2.05	-1.90	-0.50	-0.61	-1.89
148	-2.09	-1.92	-2.09	-1.92	-0.20	-0.63	-1.85
149	-1.64	-1.69	-1.64	-1.69	-0.43	-0.38	-1.74
150	-0.75	-0.68	-0.75	-0.68	0.47	0.11	-0.93
151	2.38	2.05	2.38	2.05	2.72	2.79	2.66
152	1.21	0.76	1.21	0.76	0.43	0.36	1.26
153	1.67	1.20	1.67	1.20	2.44	2.38	1.40
154	1.82	1.34	1.82	1.34	0.53	1.93	1.47
155	1.74	1.70	1.74	1.70	1.47	1.58	1.55

Compound	E(GFN1)	E(GFN2)	E(GFN1 (fixed M))	E(GFN2 (fixed M))	E(PM6)	E(PM7)	ref.
156	2.64	2.56	2.64	2.56	1.71	1.80	2.26
157	2.73	2.37	2.73	2.37	2.01	2.04	2.58
158	-2.60	-2.69	-2.60	-2.69	-0.89	-0.80	-2.43
159	-2.24	-2.44	-2.24	-2.44	-0.71	-0.83	-2.34
160	-2.24	-2.39	-2.24	-2.39	-0.59	-0.76	-2.29
161	-3.73	-3.79	-3.73	-3.79	-1.18	-1.50	-2.13
162	-2.16	-1.99	-2.16	-1.99	-0.48	-0.43	-1.90
163	-1.63	-1.61	-1.63	-1.61	-0.27	-0.34	-1.72
164	-1.90	-1.76	-1.90	-1.76	-0.36	-0.30	-1.70
165	-1.84	-1.67	-1.84	-1.67	-0.37	-0.28	-1.64
166	-0.36	-1.07	-0.36	-1.07	-1.91	-1.65	-1.45
167	-0.46	-1.12	-0.46	-1.12	0.06	-0.56	-1.37
168	-0.92	-0.79	-0.92	-0.79	0.14	-0.07	-1.34
169	-1.70	-1.60	-1.70	-1.60	-0.04	-0.17	-1.20
170	-1.48	-1.50	-1.48	-1.50	-0.30	-0.57	-1.01
171	-0.51	-0.31	-0.51	-0.31	0.53	0.31	-0.98
172	-0.76	0.23	-0.76	0.23	1.26	1.07	-0.51
173	-0.47	-0.64	-0.47	-0.64	0.31	0.11	-0.45
174	-0.89	-0.09	-0.89	-0.09	0.66	0.15	-0.20
175	-0.35	0.80	-0.35	0.80	1.38	2.06	0.26
176	-0.40	1.11	-0.40	1.11	1.06	2.20	0.71
177	2.43	1.56	2.43	1.56	2.74	2.72	1.01
178	-0.52	-0.79	-0.52	-0.79	0.69	1.00	-1.12
179	-0.80	-0.91	-0.80	-0.91	0.39	0.35	-0.27
180	-1.10	0.51	-1.10	0.51	1.50	-	-0.26
181	-0.95	-0.28	-0.95	-0.28	0.61	0.54	-1.10
182	-1.96	-0.86	-1.96	-0.86	-0.50	-0.05	-0.86
183	-0.79	-1.02	-0.79	-1.02	0.11	0.10	-0.59
184	-0.52	-0.54	-0.52	-0.54	0.75	0.77	-0.32
185	-0.71	-0.81	-0.71	-0.81	0.45	0.42	-0.28
186	-	-2.19	-	-2.19	-0.43	-0.65	-1.92
187	-	-1.45	-	-1.45	0.18	-0.05	-1.36
188	-1.47	-1.72	-1.47	-1.72	-0.15	-0.57	-1.88
189	-1.07	-0.95	-1.07	-0.95	0.56	-0.06	-1.15
190	0.19	0.65	0.19	0.65	1.42	1.29	0.13
191	-0.83	-0.65	-0.83	-0.65	0.59	0.56	-0.29
192	1.83	1.61	1.83	1.61	1.49	1.15	0.66
193	0.29	0.70	0.29	0.70	1.33	1.07	0.30
194	-3.36	-2.83	-3.31	-2.78	-1.47	-0.65	-1.28
195	-3.33	-1.89	-3.28	-1.84	-1.22	-0.93	-0.24
196	-3.15	-1.78	-3.15	-1.78	-2.13	0.27	0.36
197	0.09	-0.83	0.22	-0.78	0.02	0.94	-0.81
198	-0.04	-0.44	-3.05	-2.76	-1.11	-1.36	-0.32
199	0.43	-0.27	0.43	-0.28	0.28	1.26	0.13
200	-1.18	-1.41	-1.18	-1.42	0.52	-1.40	-1.24
201	-0.12	-0.31	-0.43	-0.59	-0.30	0.43	-0.41
202	0.83	0.13	-1.03	-1.61	-1.79	-1.40	-1.14
203	0.77	0.56	-2.84	-1.99	-2.39	-2.03	0.62

Compound	E(GFN1)	E(GFN2)	E(GFN1 (fixed M))	E(GFN2 (fixed M))	E(PM6)	E(PM7)	ref.
204	1.59	1.30	1.59	1.30	0.95	-0.70	0.13
205	-0.36	-0.26	-0.36	-0.26	-1.41	-1.19	-0.40
206	0.76	0.92	0.17	0.48	-0.39	0.26	0.95
207	1.21	0.06	-0.29	-1.54	-1.58	-0.97	-1.03
208	-0.07	-0.16	-0.07	-0.16	-1.57	-1.31	-0.23
209	19.51	-0.52	-	-0.37	-0.19	0.16	-0.71
210	17.19	0.35	-15.85	-3.16	-1.47	-1.31	1.33
211	-10.20	1.58	-10.20	1.58	0.41	0.14	1.01
212	-9.82	-0.13	-9.82	-0.13	-0.98	-0.26	0.51
213	9.74	1.08	9.63	0.30	-1.75	-1.73	1.81
214	-3.61	-1.57	-3.15	-1.34	-0.54	-1.12	-1.33
215	-2.74	-1.38	-6.95	-4.41	-1.56	-1.86	0.46
216	-1.13	-2.02	-1.13	-2.02	0.13	-0.26	0.13
217	-2.65	-2.69	-2.65	-2.69	-0.06	-2.89	-0.13
218	0.11	-0.50	-0.57	-1.18	-1.34	0.51	1.21
219	1.50	1.37	-2.94	-2.48	-9.84	-0.05	1.28
220	2.06	1.96	2.06	1.96	0.78	0.70	1.12
221	-0.11	-0.14	-0.11	-0.14	-	-	0.31
222	1.44	1.11	1.47	1.13	-1.03	-0.96	-0.02
223	1.99	1.71	1.99	1.45	0.10	0.59	1.13
224	-0.22	-0.36	-0.22	-0.36	-102.67	-1.96	0.33
225	1.44	1.15	1.46	1.16	-0.27	1.05	0.01
226	2.02	1.73	2.02	1.73	0.36	0.53	1.12
227	-0.20	-0.29	-0.20	-0.18	-100.97	-0.85	0.31
228	-0.03	-0.44	-0.03	-0.44	-0.44	-5.00	0.20
229	0.37	-0.12	0.37	-0.12	-1.09	0.16	0.35
230	-0.76	-1.09	-1.55	-2.26	-1.89	-6.37	-0.90
231	0.70	-3.01	0.68	-3.01	-0.11	-1.22	-1.12
232	-0.36	-0.46	-0.31	0.07	2.08	-0.12	0.01
233	-1.44	-1.37	-2.65	-2.52	20.66	-1.73	0.57
234	-1.34	-2.04	-1.39	-1.89	-1.16	-8.05	-1.18
235	-0.78	-1.39	-1.83	-2.80	-1.37	-1.96	-1.57
236	-0.14	-1.10	-1.11	-1.84	-1.64	-0.92	0.15
237	0.52	-0.79	1.18	0.06	0.68	-0.02	-0.43
238	-0.89	-1.81	-2.26	-3.12	-	-1.47	-0.11
239	2.18	0.34	1.43	-0.69	0.16	0.63	-0.58
240	3.09	0.67	2.74	0.69	0.31	2.03	1.50
241	3.95	1.40	5.18	2.52	0.68	2.20	0.75
242	2.04	0.66	0.92	-0.38	0.23	-0.52	1.86
243	-0.06	-0.26	-1.61	-1.63	-	-1.78	0.78
244	1.52	-0.98	1.52	-0.99	-0.59	-0.76	-0.80
245	-2.78	-0.77	-2.90	-0.87	-	-1.34	-0.75
246	-3.09	-0.76	-2.94	-0.75	7.13	-0.33	-0.12
247	-2.77	-0.31	-4.12	-1.13	-	-3.39	-0.70
248	-4.21	-1.41	-4.90	-1.62	-	-3.80	-0.88
249	-2.49	-0.59	-3.47	-1.35	-	7.57	-0.53
250	-3.79	-1.79	-3.11	-0.76	-	-	-1.27
251	-2.72	-0.73	-2.94	-0.92	-1.26	-1.07	-0.32

Compound	E(GFN1)	E(GFN2)	E(GFN1 (fixed M))	E(GFN2 (fixed M))	E(PM6)	E(PM7)	ref.
252	-4.24	-2.03	-4.27	-2.05	-	-2.20	-1.50
253	0.44	0.86	0.44	0.86	-1.78	0.09	1.28
254	-3.36	-2.09	-3.53	-2.20	-	-2.20	-1.61
255	-1.48	-0.14	-1.89	-0.43	0.52	-8.41	-0.08
256	-0.70	-2.27	-0.68	-2.27	0.42	-0.85	-0.64
257	-0.36	-1.07	0.15	-0.30	-	-1.03	-0.37
258	0.60	-1.35	-	-1.35	0.94	-0.08	0.02
259	-0.92	-1.50	-1.51	-1.26	-0.51	-2.81	-1.04
260	-0.80	-1.06	-1.52	-2.13	-1.41	-1.24	-0.98
261	0.32	-0.21	1.01	0.58	0.02	0.07	0.12
262	-1.29	-0.99	-2.45	-2.16	18.04	-1.79	0.37
263	0.24	0.57	-0.47	-0.16	-0.37	-0.30	1.29
264	0.75	-1.32	0.75	-1.33	1.02	-0.02	-0.02
265	-0.77	-1.01	-1.36	-2.12	-1.34	-1.17	-0.96
266	0.31	-0.20	1.02	0.63	0.06	-0.10	0.04
267	-1.27	-0.99	-2.37	-2.17	16.02	-2.14	0.35
268	-0.89	-1.24	-1.63	-2.54	-1.66	-1.96	-1.16
269	-1.37	-1.20	-2.64	-2.34	17.14	-1.98	0.36
270	1.55	0.56	-0.23	-0.84	-100.75	-0.47	1.35
271	-0.72	-0.86	-2.21	-2.00	14.72	-2.68	0.20
272	0.79	0.26	-1.08	-1.28	-	-0.88	0.68
273	-0.07	-0.28	-0.06	-0.28	-0.16	-0.36	0.78
274	-0.61	-1.58	-0.59	-1.31	-1.45	-0.99	-1.12
275	-0.38	-1.35	-0.32	-1.27	-1.43	-0.79	-1.22
276	-1.05	-0.56	-1.05	-0.56	-1.64	0.31	0.54
277	-1.38	-1.29	-1.39	-1.29	-0.36	-0.05	-1.03
278	-1.40	-0.33	-1.40	-0.33	-0.74	-0.21	0.60
279	-3.04	-2.03	-3.04	-2.03	-0.15	0.18	-0.07
280	-0.71	0.26	-1.36	-0.29	0.10	0.00	0.16
281	0.46	-0.08	-2.54	-2.05	-1.12	-1.47	1.19
282	1.13	0.36	1.98	0.90	0.17	1.40	0.81
283	0.19	-1.06	-1.13	-2.35	-1.35	-1.19	-0.85
284	-0.07	-1.43	-1.20	-2.98	-1.16	-1.12	-0.75
285	-1.39	-2.17	-0.92	-1.28	3.61	-	-0.58
286	-2.11	-2.65	-2.11	-2.64	-0.50	-1.46	-0.31
287	-0.09	-0.75	0.39	0.28	0.73	0.27	-0.08
288	-	1.25	-	1.25	2.13	2.10	0.74
289	1.15	0.60	1.15	0.60	1.58	1.73	0.31
290	-	0.21	-	0.21	0.16	1.34	-0.10
291	-	-0.21	-	-0.21	-0.04	-0.79	-0.91
292	-	-1.04	-	-1.04	-1.41	-1.94	-0.98
293	-	-0.74	-	-0.74	1.33	0.04	-1.36
294	-	-1.16	-	-1.16	0.52	-1.51	-1.43
295	-	-1.41	-	-1.41	1.38	-0.32	-1.57
296	-	-1.85	-	-1.85	1.21	-0.61	-1.85
297	-0.75	-0.65	-0.75	-0.65	-0.28	0.79	-1.41
298	-1.57	-1.73	-1.74	-2.02	-	-0.45	-2.18
299	-2.03	-3.41	-1.70	-3.34	-2.55	-2.51	-2.30

Compound	E(GFN1)	E(GFN2)	E(GFN1 (fixed M))	E(GFN2 (fixed M))	E(PM6)	E(PM7)	ref.
300	-2.27	-2.57	-1.45	-1.58	-2.09	-0.88	-2.50
301	0.78	0.05	0.78	0.05	0.89	0.14	0.55
302	0.69	-0.31	0.69	-0.31	0.39	-0.38	0.46
303	-1.19	-1.40	-1.19	-1.40	-0.44	-12.72	-1.49
304	-1.88	-2.30	-1.88	-2.30	0.49	0.83	-1.73
305	2.16	1.51	2.16	1.51	0.52	-0.22	0.84
306	-2.79	-1.78	-2.79	-1.78	-0.16	0.17	-0.07
307	-1.93	-1.70	-1.93	-1.70	1.29	0.51	-0.03
308	1.64	1.72	1.64	1.71	2.60	2.51	2.41
309	1.71	1.53	1.72	1.53	2.21	2.39	2.42
310	-2.19	-2.53	-2.19	-2.52	0.83	-1.42	-2.17
311	-0.88	-0.91	-0.88	-0.91	-0.64	-0.32	-1.42
312	0.83	0.19	0.83	0.18	0.78	0.38	0.72
313	2.17	1.87	2661.97	1.87	0.88	0.05	1.39

A10. Oxidation Under Reductive Conditions: From Benzylic Ethers to Acetals with Perfect Atom-Economy by Titanocene(III) Catalysis

Abstract: Described here is a titanocene-catalyzed reaction for the synthesis of acetals and hemiaminals from benzylic ethers and benzylic amines, respectively, with pendant epoxides. The reaction proceeds by catalysis in single-electron steps. The oxidative addition comprises an epoxide opening. An H-atom transfer, to generate a benzylic radical, serves as a radical translocation step, and an organometallic oxygen rebound as a reductive elimination. The reaction mechanism was studied by high-level dispersion corrected hybrid functional DFT with implicit solvation. The low-energy conformational space was searched by the efficient CREST program. The stereoselectivity was deduced from the lowest lying benzylic radical structures and their conformations are controlled by hyperconjugative interactions and steric interactions between the titanocene catalyst and the aryl groups of the substrate. An interesting mechanistic aspect is that the oxidation of the benzylic center occurs under reducing conditions.

A10.1. Introduction

Catalysis in single electron steps[471,472] or metallo-radical catalysis[473,474] is a concept that merges the concepts of radical chemistry with those of transition metal catalysis. In this manner, the advantages of both fields, such as the ease of radical generation, the high functional group tolerance of radical reactions and the ability of a metal complex to control the pertinent selectivities of transformation can be combined to attractive novel catalytic processes. An essential aspect for the success of catalysis in single electron steps is the availability of oxidative additions and reductions in single electron steps for radical and product generation.[475–477] Both steps can be regarded as homolytic substitution reactions[478,479] and require an efficient shuttling of the catalyst between neighboring oxidation states.[480–490] During the radical translocation, a new bond and radical is formed. This step sets the scene for the ensuing reductive elimination.

A10.2. Results and Discussion

Here, we demonstrate how radical translocation via H–atom abstraction from a benzylic C–H bond by an alkyl radical can be included in a process proceeding via titanocene(III) catalysis in single electron steps (Scheme A10.1).

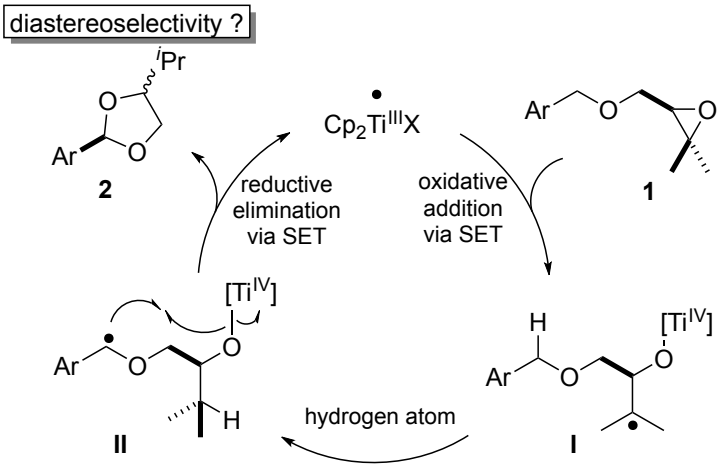
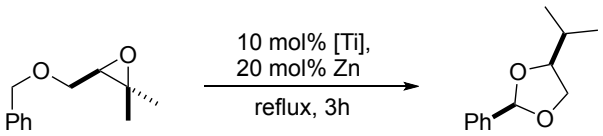


Figure A10.1.: Atom-economic transformation of benzylic ethers into acetals by titanocene(III) catalysis in single-electron steps (SET).

The overall reaction constitutes an atom-economical oxidation at the benzylic position that proceeds under reducing conditions. The oxidative addition occurs by a titanocene(III) mediated epoxide opening [491–493] through electron transfer to yield **I**. From **I**, radical translocation occurs via the pivotal H-atom abstraction* to yield **II**. The driving force for this step arises from the increased strength of the newly formed C–H bond. [495,496] In **II**, the stage is set for the reductive elimination of Cp_2TiX that forms the desired product, acetal **2**. In this step that can be regarded as an organometallic oxygen rebound, a second stereocenter is formed. [497,498] It remains to be seen how selective the ring formation is.

*for a recent review see Ref. 494.



1a **2a**

[Ti]	Solvent	Yield[%]	Conversion [%][a]
$(CH_3C_5H_4)_2TiCl_2$ [b]	THF	2[a]	42
Cp_2TiCl_2 [b]	THF	26[a]	75
$Cp_2Ti(O_2CCF_3)_2$	THF	41, 61[a]	91
$Cp_2Ti(O_3SCH_3)_2$	THF	67	100
$Cp_2Ti(O_3SCH_3)_2$	1,4-dioxane	38[a]	96
$Cp_2Ti(O_3SCH_3)_2$	acetone	33[a]	41

Table A10.1.: Screening of different catalysts and solvents to find the most suitable reaction conditions for the conversion of **1a**. [a] Yields and conversion are determined by quantitative ^{1}H NMR spectroscopy. [b] Requires 50 mol % Collidine*HCl to prevent a decomposition of the catalyst.

Our results for finding suitable reaction conditions for the reaction of **1a** are summarized in Table A10.1. The active titanocene(III) catalysts were prepared by in situ reduction of the titanocene(IV) precatalysts with Zn. Without Zn, no reaction was observed. The use of Mn is possible but less convenient since the reduction of Ti^{IV} requires more time.[491–493,499–505] In THF, Zn-$(CH_3C_5H_4)_2TiCl_2$ and Zn-Cp_2TiCl_2 resulted in the consumption of **1a**. However, hardly any **2a** (2 % and 26 %, respectively) was formed as judged by quantitative ^{1}H-NMR analysis of the reaction mixture. While Zn-$Cp_2Ti(O_2CCF_3)_2$ gave an improved conversion to **2a**, the isolated yield was only 41 %. Zn-$Cp_2Ti(O_3SCH_3)_2$[499–505] (O_3SCH_3= OMs) provided the best catalyst and allowed the isolation of **2a** in 67 % yield. Other solvents than THF lead to inferior results. The superiority of Zn-$Cp_2Ti(O_3SCH_3)_2$ over Zn-$(CH_3C_5H_4)_2TiCl_2$ and Zn-Cp_2TiCl_2 suggests that the reductive elimination is the rate determining step of the catalytic cycle because epoxide opening is faster with more electron rich catalysts whereas reductive eliminations are slowed down. In our case, the reductive elimination constitutes a homolytic substitution at a Ti–O bond by a benzylic radical. Due to the high stabilization of this radical, this step seems particularly difficult.[478,479]

We employed $Cp_2Ti(O_3SCH_3)_2$ that is a stronger oxidant than Cp_2TiCl_2 due to the more electron withdrawing mesylate ligand.[502] In this manner, additional driving force for the reductive elimination is provided. The results summarized in Table A10.1 therefore demonstrate that the tuning of the redox-properties of the titanocenes is a crucial and powerful tool to adjust the efficiency of the steps in the catalytic cycle.[499–505] Remarkably, **2a** is formed as a single diastereoisomer (structure determined by 2D-NMR spectroscopy, see the Supporting Information).[†] To understand this particular feature of the

[†] a) We are not aware of any examples of our reaction or the postulated mechanism in the literature. However, a stoichiometric reaction observed in Ref. 506 may proceed by our mechanism. However, no explanation for formation was given or proposed. We tank W. A Nugent for bringing this reference to our attention. b) The formation of acetals similar to the ones described

reaction and the thermodynamics of the catalytic cycle (Scheme A10.2), these were studied computationally by applying the following protocol.

All initial geometries were prepared manually and meta-dynamics-based conformer searches for all involved geometries were performed using the CREST[68,419] code in combination with the generic force field GFN-FF(GBSA[THF])[181] with implicit solvation as implemented in the freely available xtb code.[61,167‡] In case of metal containing structures, a small harmonic bond lengths constraint was applied during the minimum conformer search in order to avoid dissociation. The conformer ensemble was efficiently refined in multiple steps at DFT level PBEh-3c/DCOSMO-RS(THF)[154] to obtain the lowest lying conformer, facilitated by the ENSO code.[235] After identification of the lowest lying conformer, each of the structures, was used to calculate Gibbs free energies (G) to investigate the thermodynamics of the catalytic cycle shown in Scheme A10.2.

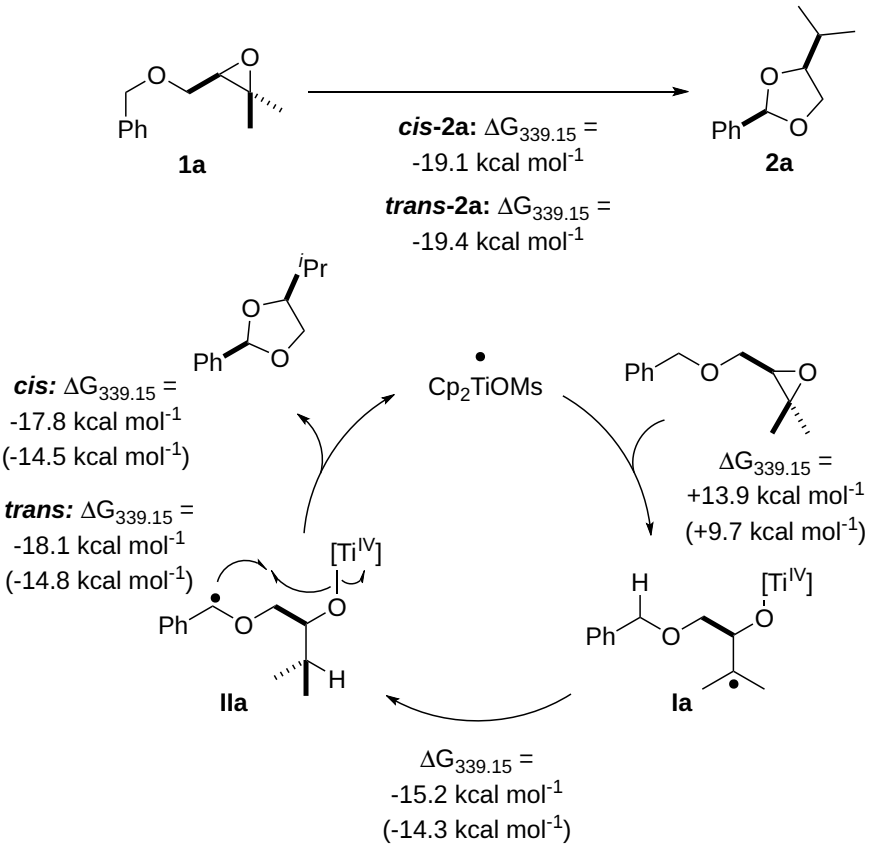


Figure A10.2.: Atom-economic transformation of benzylic ethers into acetals by titanocene(III) catalysis in single electron steps calculated at PW6B95-D4/def2-QZVP + COSMO-RS(THF)// PBEh-3c/DCOSMO-RS-(THF) level. Values within parentheses are calculated for the catalyst Cp_2TiCl.

here with acid catalysis under thermodynamic control give essentially no diastereoselectivity: Ref. 507

‡Semiempirical extended tight-binding program package xtb, *https://github.com/grimme-lab/xtb*. Accessed September 28, **2020**.

Herein, PW6B95-D4/def2-QZVP[149,234,288] dispersion corrected, high-level single-point energies and COSMO-RS[289,303] solvation contributions were calculated on the PBEh-3c + DCOSMO-RS(THF) optimized geometries. Thermostatistical contributions are obtained at the PBEh-3c + COSMO(THF) level using the modified rigid rotor harmonic oscillator (mRRHO) approach.[183,236] The mRRHO ansatz is applied to reduce the error of the harmonic approximation for low lying vibrational frequencies and to reduce the effect of numerical noise in the DFT. The vibrational frequencies are scaled by 0.95 and the thermostatistical analysis is evaluated at 339.15 K. Free energies are calculated as the sum of electronic energy, solvation- and thermostatistical terms. All DFT calculations were performed with TURBOMOLE program package 7.4.1.[508] The resolution of the identity approximation for the Coulomb integrals is generally applied in combination with the default auxiliary basis sets.[86] The numerical integration of the exchange correlation contribution is performed with grids m4 for PBEh-3c and grid m5 for PW6B95. Default convergence criteria are applied for all single-point calculations (Econv 10^{-7} E_h). For geometry optimizations Econv 2x10^{-5} E_h and Gconv 2×10^{-3} E_h $Bohr^{-1}$ are used. The convergence criteria for the transition-state optimizations were set to 5×10^{-6} E_h and 2×10^{-3} E_h $Bohr^{-1}$. The solvation contributions were calculated with the conductor like screening model for real solvents (COSMO-RS) using the BP_TZVP_C30_1601 parameterization and the COSMOtherm2016 package.[307] The solvation contribution incorporates the volume work RTln(V_{ideal}) (T = 339.15 K) for an ideal gas at 1 bar to 1 mol L^{-1}. Geometries and tabulated data are provided in the SI.

The free energy $\Delta G_{339.15}$ of the reaction **1a** to **2a** (-19.1 kcal mol^{-1} for *cis*-**2a** and -19.4 kcal mol^{-1} for *trans*-**2a**) is strongly negative. This is a manifestation of the loss of the epoxide's ring strain.[480–490] The very similar values of $\Delta G_{339.15}$ for *cis*-**2a** and *trans*-**2a** also show that the observed highly diastereoselective acetal formation cannot be under thermodynamic control.

The other important question of the mechanism is the understanding of the high diastereoselectivity of acetal formation in the reductive elimination. To this end, we studied the preferred conformation of the benzylic radical **IIa** and the structures and energies of the transition-states leading to *cis*- and *trans*-**2a**. The oxidative addition of titanocene catalyst Cp_2TiOMs to the epoxide that leads to radical formation is the thermodynamically critical step ($\Delta G_{339.15}$ = + 13.9 kcal mol^{-1} for Cp_2TiOMs and $\Delta G_{339.15}$ =+ 9.7 kcal mol^{-1} for Cp_2TiCl). The more favorable reaction with Cp_2TiCl is in line with an earlier study on the thermodynamics of epoxide opening.[471] Experimental studies on the use of electron deficient catalysts have demonstrated that they lead to substantially lower rate constants (e.g. Cp_2TiOMs: k_{293} = 0.013 M^{-1} s^{-1}) compared to Cp_2TiCl (k_{293} = 1.1 M^{-1} s^{-1}).[502] The HAT step is strongly exergonic in line with the difference in stability of tertiary and benzylic radicals. Here the conformational search for the lowest lying conformer established that the HAT reaction from **Ia** to **IIa** is accompanied by a conformational relaxation. This structural relaxation results in an energy gain of 3.2 kcal mol^{-1}compared to a HAT where no conformational search has been employed (see SI for further information). The identification of a lower lying conformation affects all following reactions in the catalytic cycle.

Finally, the reductive elimination that results in product formation and catalyst regeneration is thermodynamically favorable for both Cp_2TiOMs (***trans***: $\Delta G_{339.15}$ = -18.1 kcal mol^{-1}) and Cp_2TiCl (***trans***: $\Delta G_{339.15}$ = -14.8 kcal mol^{-1}). The results suggest that increased efficiency of Cp_2TiOMs over Cp_2TiCl in the reductive elimination is the reason for the superiority of the former complex as catalyst. We found

that **IIa**-*cis* (leading to *cis*-**2a**, Scheme A10.3) is the most stable conformer of **IIa**. We attribute this stability to two hyperconjugative interactions between the σ-orbitals of C–H bonds and σ^*-orbitals of C–O bonds (the C–O–Ti bond and the C–O–C· bond). This results in the preferred *gauche*-orientation of the two C–O bonds. Additionally, the "upper" Cp-ligand and the -OMs ligand of the catalyst forces the phenyl ring in a position pointing away from "upper" Cp-ligand and the H-substituent of the benzylic radical towards the -OMs ligand. Natural bond orbitals (see Scheme A10.3) and second-order perturbative estimates of donor-acceptor interactions in the NBO basis were calculated using the PBEh-3c/CPCM(THF) density obtained with the ORCA 4.2.1 program package[461] and the NBO6 code.[509] The second conformer **IIa**-*trans* (leading to *trans*-**2a**, Scheme A10.3) that we were able to identify is 6.5 kcal mol^{-1} less favored than **IIa**-***cis*** at 339.15 K. It arises from **IIa**-***cis*** by a rotation around the O–C· bond. The stabilization by the *gauche*-interactions in **IIa**-***cis*** and **IIa**-***trans*** is essentially identical and, therefore, the difference in stability is due to increased repulsive steric interactions between the Ph-group and the "upper" Cp-ligand and the -OMs-ligand. The barrier for the rotation that transforms **IIa**-***cis*** into **IIa**-***trans*** is estimated to be 7.1 kcal mol^{-1} (see SI). From **IIa**-***cis*** and **IIa**-***trans*** we calculated the barriers for the transformation to *cis*-**2a** (16.7 kcal mol^{-1}) and *trans*-**2a** (18.3 kcal mol^{-1}), respectively.

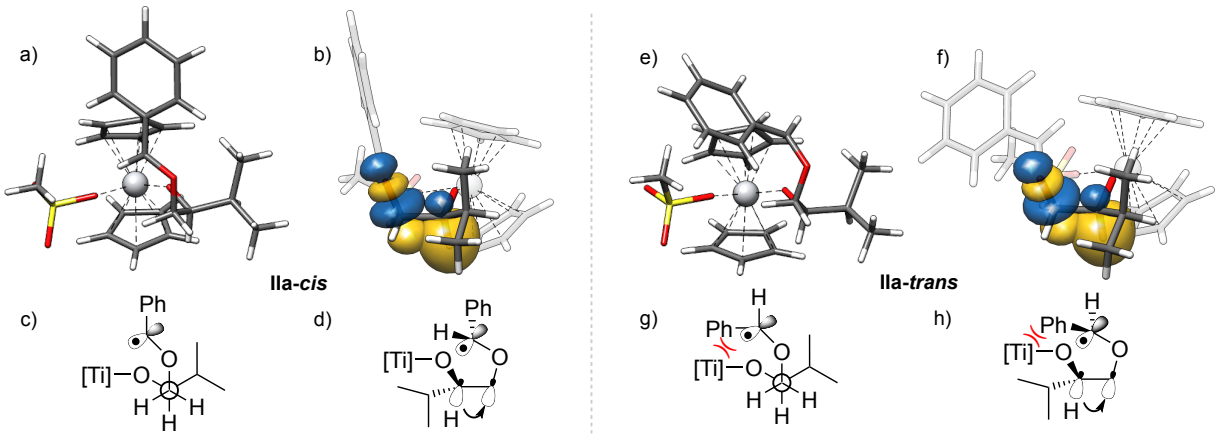


Figure A10.3.: 3D-Structures of the radicals **IIa**-***cis*** (a,b) and **IIa**-***trans*** (e,f) and their stabilization by *gauche* interactions, with side view. The natural bond orbitals (NBO)[509] shown are obtained from a PBEh-3c + CPCM(THF) calculation. In addition Newman projections (c,g) and wedge and dash structures (d,h) of these radicals are given. They show the C–H bond orbital donating into the σ^*-orbital of the C–O bond, visualizing the gauche-effect. Only one of two effective gauche interactions is shown. Isosurface value = 0.05 e$^{1/2}$ bohr$^{3/2}$.

The initial reaction paths were calculated with GSM/GFN2-xTB(GBSA[THF])[61,167,230,456] and subsequently refined at PBEh-3c/COSMO(THF) level. The final transition-state is optimized and verified for having only one imaginary mode. Free energies are calculated using the multilevel approach discussed before (PW6B95-D4/def2-QZVP + COSMO-RS-(THF)// PBEh-3c/ COSMO(THF)). The energy profile of the conversion of the radicals **IIa** to products **2a** is shown in Scheme A10.4. The selectivity of acetal

formation is mainly due to the difference in stability of **IIa-*cis*** and **IIa-*trans***. Both transition-states are "early" with long distances (**TS2a-*trans*** = 285 pm, **TS2a-*cis*** = 242 pm vs. 139 and 140 pm in *trans*-**2a** and *cis*-**2a**) of the forming C–O bonds and are, therefore, structurally similar to **IIa-*cis*** and **IIa-*trans***. Thus, the conformations of radicals **IIa-*cis*** and **IIa-*trans*** and also those of the transition-states **TS2a-*cis*** and **TS2a-*trans*** are controlled by the same two factors. First, the dihedral angles in the **O-CH*i*Pr-CH$_2$-O** chain are determined by hyperconjugative interactions between the σ-orbitals of C–H bonds and σ^*-orbitals of C–O bonds (gauche-interactions). Second, the conformation at the benzylic radical center is determined by the steric interactions of the Ph-ring and the H-substituent with the "upper" Cp-ligand and the OMs-ligand.

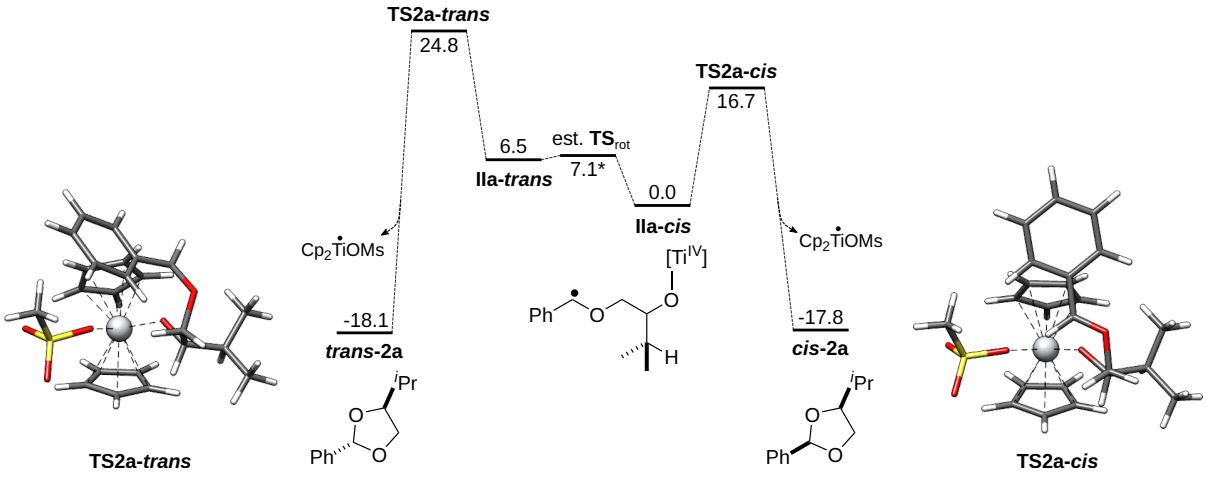


Figure A10.4.: Energy profile of the conversion of **IIa-*cis*** and **IIa-*trans*** to *cis*-**2a** and *trans*-**2a**. All Gibbs free energies are shown in kcal mol^{-1}. The transition-state geometries were obtained with PBEh-3c/COSMO-(THF). The catalytic cycle has been investigated at PW6B95-D4/def2-QZVP + COSMO-RS(THF)// PBEh-3c/ COSMO(THF) level of theory.

This analysis suggests that acetals derived from benzylic ethers (Ar**C**H$_2$O-groups) should be formed with high diastereoselectivity. This is indeed the case as summarized in Table A10.5. Diastereoselectivities of >99: <1 imply that the other isomer could not be detected by ^{1}H-NMR spectroscopy. The *cis*-2,4-disubstituted 1,3-dioxolanes **2a–2i** were smoothly obtained from the substituted benzyl ethers **1a–1i** in reasonable to high yield. Substituents at the *m*- and *p*- positions are readily tolerated. Diastereoselectivity is high (94:6–99:1) for 4-*i*propyl substituted dioxolanes. With a cyclohexyl substituent, the diastereoselectivity is equally high. For the details of the determination of relative configuration by NOESY-spectroscopy and the determination of the diastereomeric ratios by ^{1}H-NMR-spectroscopy see the SI. Our method can also be employed to prepare *cis*-2,4-disubstituted 1,3-dioxolanes in enantiomerically pure form (Scheme A10.6). This experiment proves that no racemization is occurring and provides further support for the proposed mechanism (Scheme A10.2). The transition-state models for the formation of **2a** provide an explanation for the experimentally observed diastereoselectivity with benzyl ethers. Our mechanistic analysis suggests that replacing one-H atom at the benzylic center in **1** with

bulkier alkyl groups should have a profound impact on the reaction. This is indeed the case as exemplified in the synthesis of ketals **4a–4e** (Table A10.7). In all cases, diastereoselectivity was either non-existent or low (50:50–72:38). We note that when applicable the major diastereomer has a *trans*-orientation of the aryl and the *i*Pr group. For details of the structural assignment, see SI. While the experiments were typically performed with 100 mg of the substrate, the reaction of **4e** was performed on a 1 g scale indicating that a larger scale is not problematic.

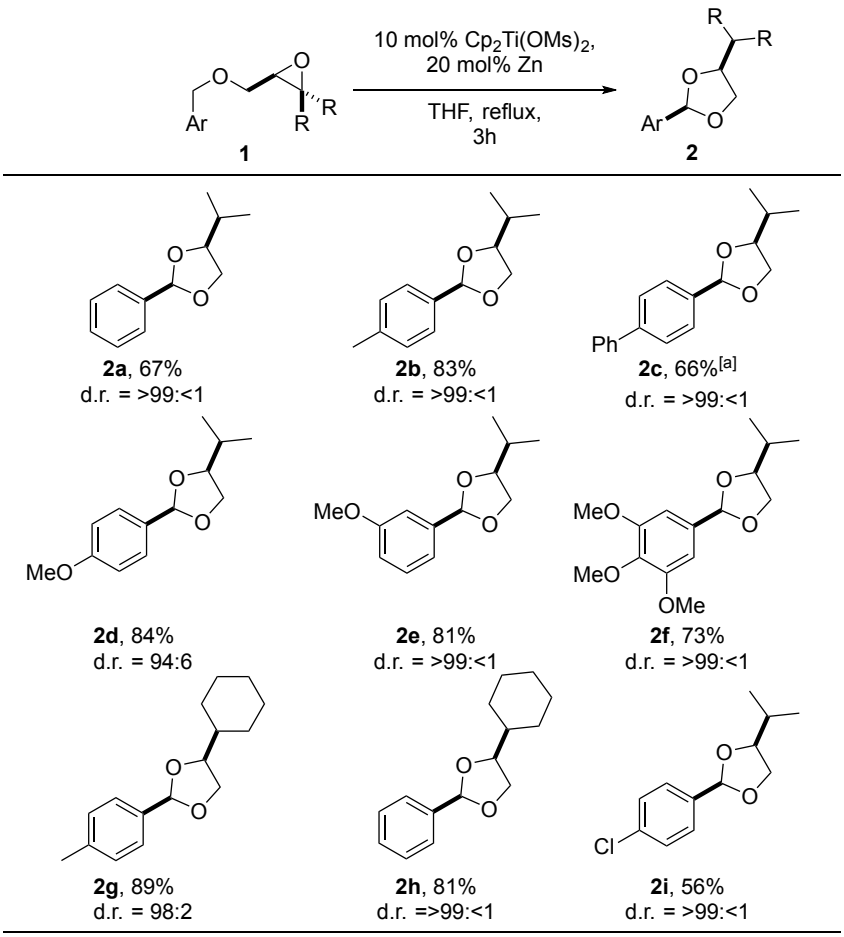


Figure A10.5.: Substrate scope of the 1,3-dioxolane synthesis by catalysis in single-electron steps. Yields are those of the isolated products. The d.r. values were determined by ^{1}H NMR spectroscopy. [a] 19 mol % $Cp_2Ti(OMs)_2$ and 38 mol % Zn are used.

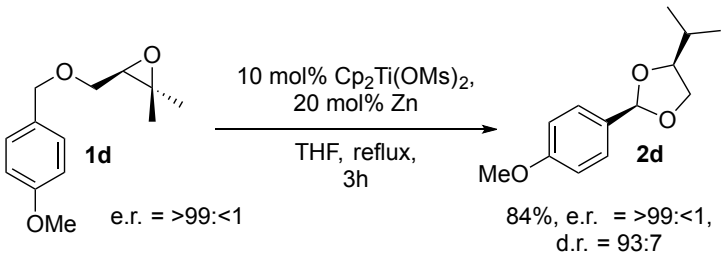


Figure A10.6.: Preparation of enantiomerically pure **2d** from **1d**.

The computed structures and energies of radicals **IVc-*cis*** and **IVc-*trans*** highlight the reasons for the selectivity (Scheme A10.8). In contrast to the situation for radical **II** ($\Delta G_{339.15}$ = 6.5 kcal mol^{-1}), the difference in the stability between the two conformers is low ($\Delta G_{339.15}$ = -0.6 kcal mol^{-1}) and **IVc-*trans*** is more stable than **IVc-*cis***. The small difference of -0.6 kcal mol^{-1}should therefore lead to a slight excess of ***trans*-4c** compared to ***cis*-4c**, which indeed is obtained for synthesis of ketals **4a–4e** (Table A10.7).

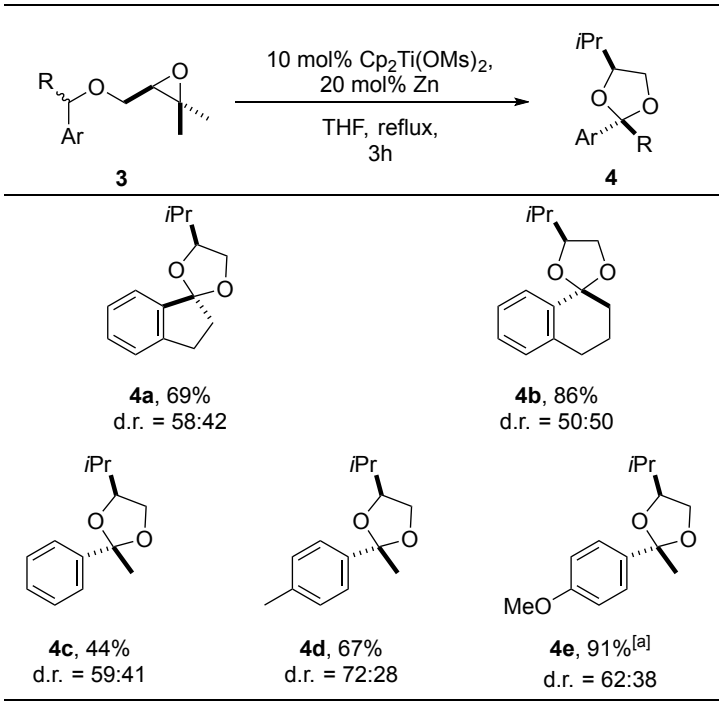


Figure A10.7.: Substrate scope and diastereoselectivity of the 2,2-disubstituted 1,3-dioxolane synthesis by catalysis in single-electron steps. Yields are those of the isolated products. The d.r. values were determined by ^{1}H NMR spectroscopy. [a] The experiment is performed on a 1 g scale.

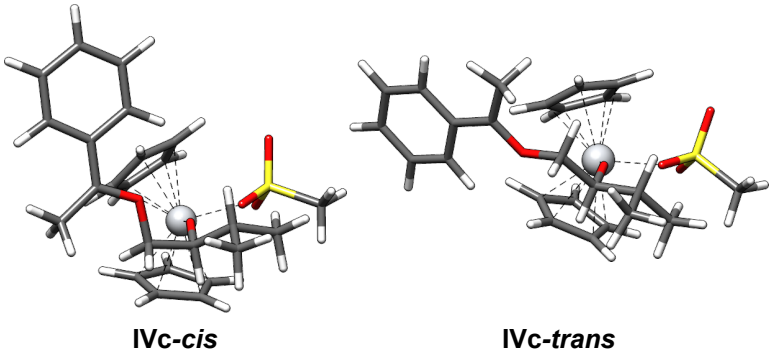


Figure A10.8.: PBEh-3c/DCOSMO-RS(THF) structures of the radicals **IVc-*cis*** and **IVc-*trans***.

Finally, we investigated the synthesis of hemiaminals from suitably substituted benzylamines (Table A10.9). In this manner, it can be established if amines are tolerated under our conditions and how this change in the substitution pattern affects selectivity. Moreover, our method provides a convenient route to hemiaminals and eventually carbonyl compounds in one step from benzyl amines. A practical difficulty associated with the hemiaminals obtained is their lability towards chromatography on SiO_2. The aldehyde derived products **6a** and **6b** could be isolated in good yield. We found that for the synthesis of the ketone derived hemiaminals **6c–6f** dissolving the crude product in diethylether followed by a filtration through a plug of cotton wool was sufficient to isolate the desired products. The thermodynamics of hemiaminal synthesis and of each step of the catalytic cycle (Scheme A10.10) show that hemiaminal formation is slightly less favorable than acetal formation.

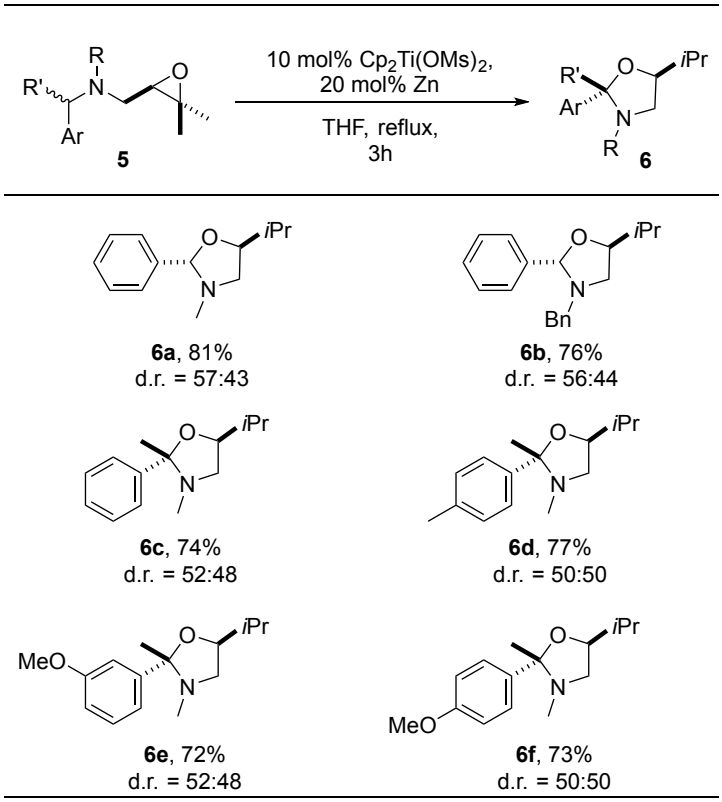


Figure A10.9.: Substrate scope and diastereoselectivity of the hemiaminal synthesis by catalysis in single-electron steps. Yields are those of the isolated products. The d.r. values were determined by ^{1}H NMR spectroscopy.

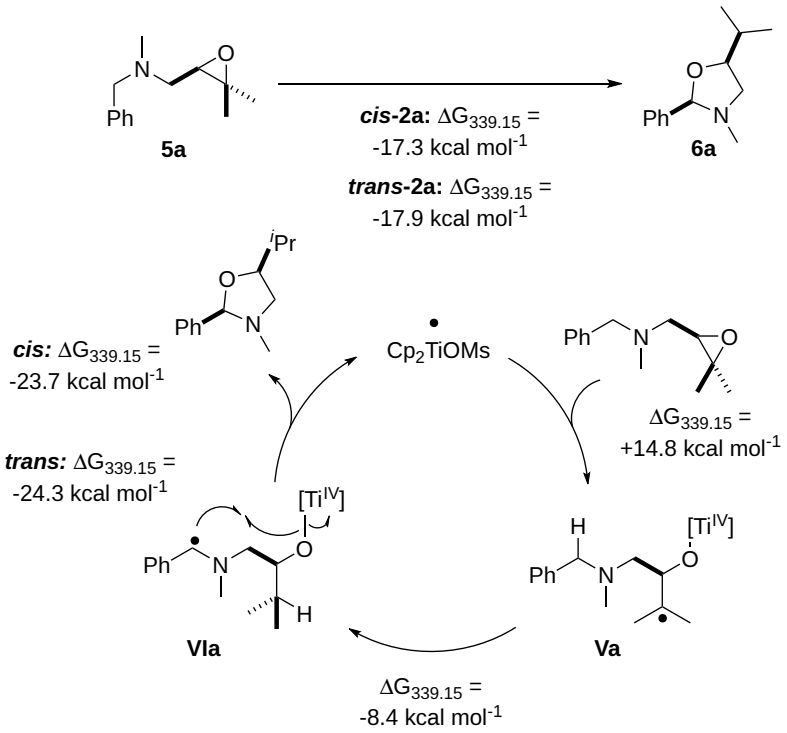


Figure A10.10.: Atom-economic transformation of benzylamines into hemiaminals by titanocene(III) catalysis in single electron steps. Calculated at PW6B95-D4/def2-QZVP + COSMO-RS(THF)// PBEh-3c/DCOSMO-RS(THF) level of theory.

With the oxidative addition being most likely rate-limiting, the calculations suggest that in agreement with the experiment the overall reaction should proceed without problems. Radical translocation by H-atom abstraction is distinctly less attractive (about 8 kcal mol^{-1}) than in acetal formation. It seems that this is not due to differences in the BDE at the benzylic C–H bonds since the calculated BDE in benzylmethylether is only 0.2–0.9 kcal mol^{-1} lower than for benzyldimethylamine for 6 different theoretical methods (see SI fordetails). Additionally, the conformational investigation established that for the HAT from **Va** to **VIa** only a small structural relaxation occurs, which further explains the less attractive value for the HAT reaction compared to the acetal formation (see SI). The reductive heterocycle formation, is substantially more advantageous than for acetals. Radical **VIa-*trans*** is slightly more stable than radical **VIa-*cis*** (0.9 kcal mol^{-1}) and thus a low diastereoselectivity is predicted in agreement with the experiment. This is in line with our analysis of ketal formation. Replacing oxygen with the bulkier alkylated nitrogen results in similar steric interactions with the titanocene as in radicals **IV**. An attractive feature of our method is that from the benzylamines substrates, the corresponding ketones can be prepared when an acidic aqueous work-up is carried out with the crude product (Scheme A10.11).

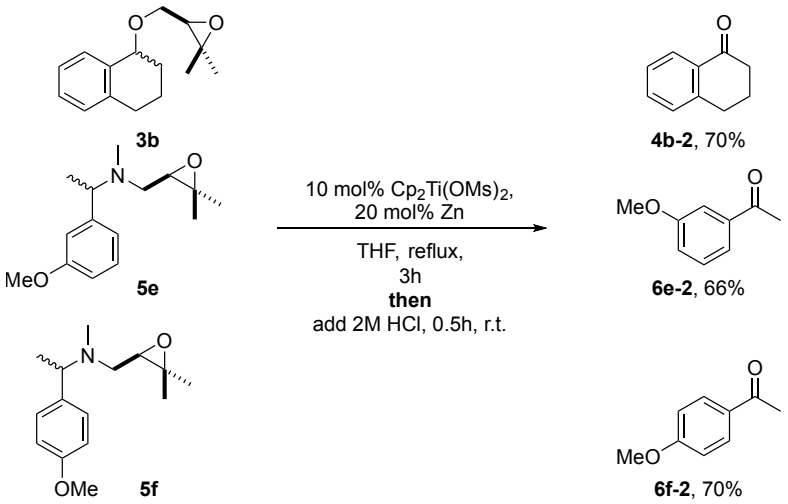


Figure A10.11.: Acetal and hemiaminal synthesis with aqueous acidic work up for the synthesis of carbonyl compounds from benzylic ethers and amines.

A10.3. Conclusion

In summary, we have developed a novel titanocene catalyzed method for the synthesis of acetals and hemiaminals from benzylic ethers and benzylic amines with pending epoxides. The reaction features catalysis in single electron steps, an epoxide opening as oxidative addition, a H–atom transfer to generate a benzylic radical as radical translocation step, and an organometallic oxygen rebound as reductive elimination. The thermochemistry of the reactions was studied at PW6B95-D4/def2-QZVP + COSMO-RS(THF)// PBEh-3c/ DCOSMO-RS(THF) level of theory including full conformational search for all stationary points. Theory and experiment are in good qualitative agreement. The automatic exploration of conformational space by the CREST program was essential for identifying the lowest lying conformations and we recommend our computational approach in general for such mechanistic studies. The stereoselectivity can be deduced from the structure of the lowest lying benzylic radicals. Their conformations are controlled by hyperconjugative interactions and by steric interactions between the catalyst and the aryl groups. A particularly interesting aspect of our mechanism is that the oxidation of the benzylic center occurs under reducing conditions. No external oxidizing agent is required, making it a promising method for highly selective oxidations in natural product and API synthesis. Moreover, the corresponding carbonyl compounds can be directly obtained by an acidic aqueous work-up of the reaction mixture.

Acknowledgements

This work was supported by the DFG (Ga 619/12-2 to A. G.) and in the framework of the “Gottfried Wilhelm Leibniz Prize” to S. G. R. B. R. thanks the Konrad-Adenauer-Stiftung for a fellowship. Open access funding enabled and organized by Projekt DEAL.

Conflict of interest

The authors declare no conflict of interest.

A11. Supporting Information to Chapter 7

Appendix A11 contains:

- Additional technical details of the calculations

geometries[a)]	d(C-O) of forming bond [pm]
IIa-*cis* (radical)	302.123
IIa-*trans* (radical)	313.034
***cis*-2a** (product)	139.873
***trans*-2a** (product)	139.257
TS2a-*cis* (transition-state)	241.568
TS2a-*trans* (transition-state)	284.959
IVc-*cis* (radical)	400.208
IVc-*trans* (radical)	331.991
***cis*-4c** (product)	140.743
***trans*-4c**(product)	140.224
VIa-*cis* (radical)	327.039
VIa-*trans* (radical)	267.178
***cis*-6a** (product)	140.622
***trans*-6a** (product)	140.560

Table A11.1.: a) Extracted from PBEh-3c+DCOSMO-RS(THF) optimized geometries.

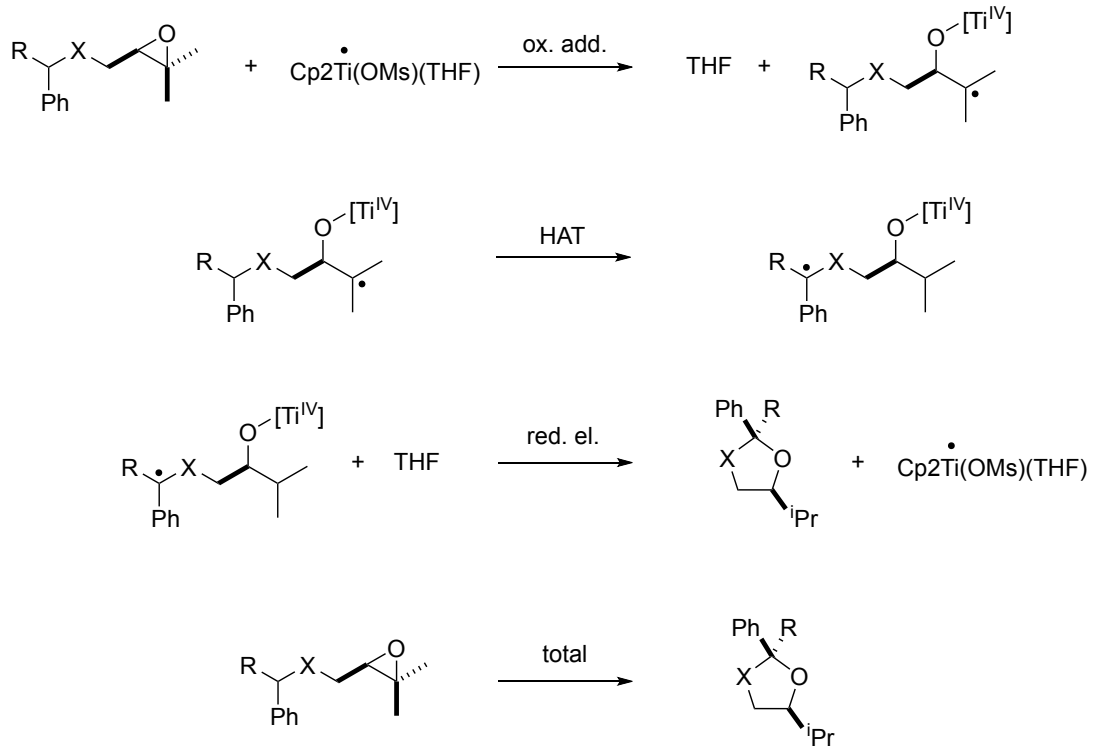


Figure A11.1.: Calculated reactions.

Reaction energies (Acetal 2a with Cp_2TiOMs cat.)

Reaction	ΔG(PW6B95-D4 /def2-QZVP)[a)]	ΔG(PW6B95-D3 /def2-QZVP)[b)]	ΔH(PW6B95-D4 /def2-QZVP)[a)]
oxidative addition	13.9	14.0	10.5
H–atom transfer	-15.2	-15.3	-18.0
reductive elimination → ***cis*-2a**	-17.8	-17.4	-13.6
barrier reductive elimination *cis*	16.7	16.3	15.5
reductive elimination → ***trans*-2a**	-18.1	-17.7	-13.4
barrier reductive elimination *trans*	24.8	24.5	23.6
estimated barrier of **IIa**-*cis* / **IIa**-*trans* interconversion	7.1	7.1	5.8
total reaction (*cis*)	-19.1	-18.7	-21.1
total reaction (*trans*)	-19.4	-19.0	-20.9

Table A11.2.: All free energies and enthalpies in kcal mol^{-1}. a) PW6B95-D4/def2-QZVP + COSMO-RS(THF) // PBEh-3c+DCOSMO-RS(THF), b) PW6B95-D3/def2-QZVP + COSMO-RS(THF) // PBEh-3c+DCOSMO-RS(THF).

Reaction energies (Acetal 2a with Cp_2TiCl cat.)

Reaction	ΔG(PW6B95-D4 /def2-QZVP)[a)]	ΔG(PW6B95-D3 /def2-QZVP)[b)]	ΔH(PW6B95-D4 /def2-QZVP)[a)]
oxidative addition	9.7	9.6	4.6
H–atom transfer	-14.3	-14.2	-15.0
reductive elimination → ***cis*-2a**	-14.5	-14.1	-10.7
reductive elimination → ***trans*-2a**	-14.8	-14.4	-10.5
total reaction (*cis*)	-19.1	-18.7	-21.1
total reaction (*trans*)	-19.4	-19.0	-20.9

Table A11.3.: All free energies and enthalpies in kcal mol^{-1}. a) PW6B95-D4/def2-QZVP + COSMO-RS(THF) // PBEh-3c+DCOSMO-RS(THF), b) PW6B95-D3/def2-QZVP + COSMO-RS(THF) // PBEh-3c+DCOSMO-RS(THF).

Reaction energies (Hemiaminal 6a with Cp_2TiOMs cat.)

Reaction	ΔG(PW6B95-D4 /def2-QZVP)[a)]	ΔG(PW6B95-D3 /def2-QZVP)[b)]	ΔH(PW6B95-D4 /def2-QZVP)[a)]
oxidative addition	14.8	14.6	7.2
H–atom transfer	-8.4	-8.7	-4.3
reductive elimination → *cis*-product	-23.7	-22.7	-21.8
reductive elimination → *trans*-product	-24.3	-23.4	-21.9
total reaction (*cis*)	-17.3	-16.8	-18.9
total reaction (*trans*)	-17.9	-17.5	-19.0

Table A11.4.: All free energies and enthalpies in kcal mol^{-1}. a) PW6B95-D4/def2-QZVP + COSMO-RS(THF) // PBEh-3c+DCOSMO-RS(THF), b) PW6B95-D3/def2-QZVP + COSMO-RS(THF) // PBEh-3c+DCOSMO-RS(THF).

Cis/trans radical isomer stability

Reaction *cis*-isomer - *trans*-isomer	ΔG(PW6B95-D4 /def2-QZVP)[a)]	ΔG(PW6B95-D3 /def2-QZVP)[b)]	ΔH(PW6B95-D4 /def2-QZVP)[a)]
IIa (acetal)	-6.5	-6.5	-5.8
IVc (ketal)	0.6	0.7	1.1
VIa (hemiaminal)	0.9	0.9	0.1

Table A11.5.: All free energies and enthalpies in kcal mol^{-1}. a) PW6B95-D4/def2-QZVP + COSMO-RS(THF) // PBEh-3c+DCOSMO-RS(THF), b) PW6B95-D3/def2-QZVP + COSMO-RS(THF) // PBEh-3c+DCOSMO-RS(THF).

Isomer stability contributions to free energy (Acetal)

	ΔG_{RRHO} (PBEh-3c)	$\Delta\delta G_{solv}$ (THF)	ΔE (PW6B95 /def2-QZVP)	ΔE_{disp} (D4)	ΔG (PW6B95 -D4/def2-QZVP)
(**IIa**-*cis*) -(**IIa**-*trans*)	-0.4	-0.7	-5.9	0.5	-6.5

Table A11.6.: All free energies and energies in kcal mol^{-1}.

Stabilization energies from second-order perturbation estimates of donor-acceptor interactions in the NBO basis using the PBEh-3c density, exemplifying the stereoelectronic gauche effect

geometry	E(2)$\sigma(C-H) \rightarrow \sigma^*(C-H)$	E(2)$\sigma(C-H) \rightarrow \sigma^*(C-H)$
IIa-*cis* (radical)	2.8	3.5
IVc-*cis* (radical)	2.6	3.0
VIa-*cis* (radical)	1.7 $\sigma(C-C) \rightarrow \sigma^*(C-N)$	1.9
TS2a-*cis* (transition state)	1.5	2.6
TS2a-*trans* (transition state)	1.9	2.8

Table A11.7.: All stabilization energies (E(2)) in kcal mol^{-1}. Since in these molecules two different gauche interactions are possible, the weaker effect is shown in the first and the stronger one in the last column.

Bond dissociation energies (BDE) of Benzylmethylether and N,N-Dimethyl-benzylamine, dissociation of a hydrogen atom at the benzylic position

	PW6B95 -D4	B3LYP -D4	PBE0 -D4	ωB97X-D3	PWPB95 -D4	DSD-BLYP -D3(BJ)
Benzylmethylether	80.5	79.7	78.6	81.9	83.1	87.2
N,N-Dimethyl-benzylamine	80.7	80.1	79.3	82.8	83.4	87.7
ΔBDE (N,N-Dimethyl-benzylamine - Benzyl-methylether)	0.2	0.4	0.7	0.9	0.3	0.5

Table A11.8.: All energies in kcal mol^{-1}. The basis set applied is def2-QZVP. Calculation on PBEh-3c/DCOSMO-RS(THF) geometries, thermostatistical contributions are calculated with PBEh-3c/COSMO 339.15 K.

Reaction-path for estimated barrier of IIa *cis/trans* rotation

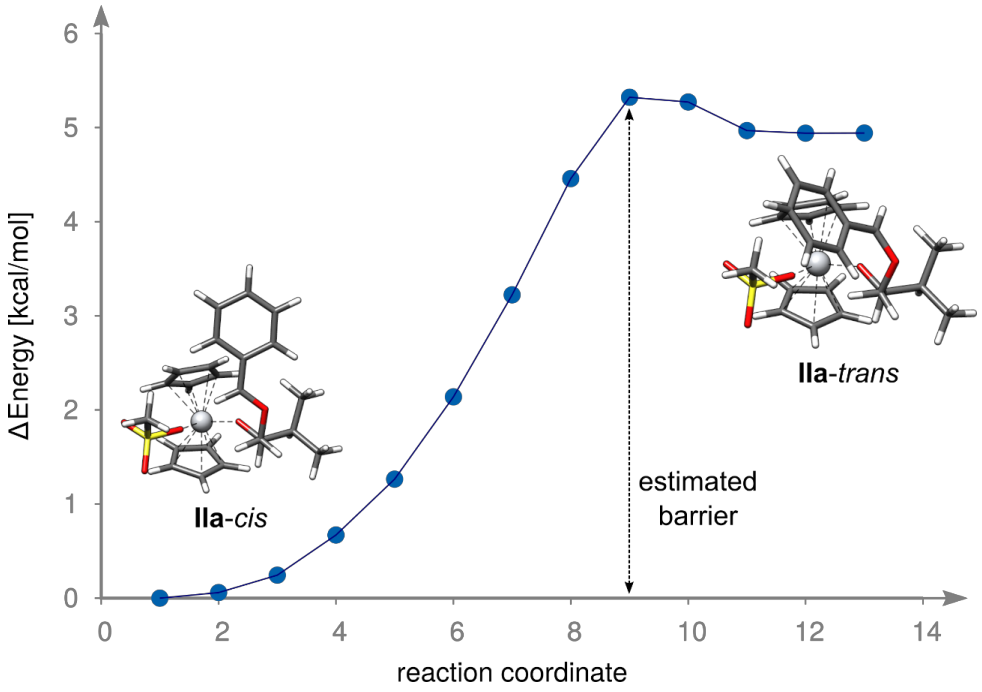


Figure A11.2.: PBEh-3c+COSMO(THF) reaction path of **IIa**-*cis/trans* rotation.

Visualization of geometries

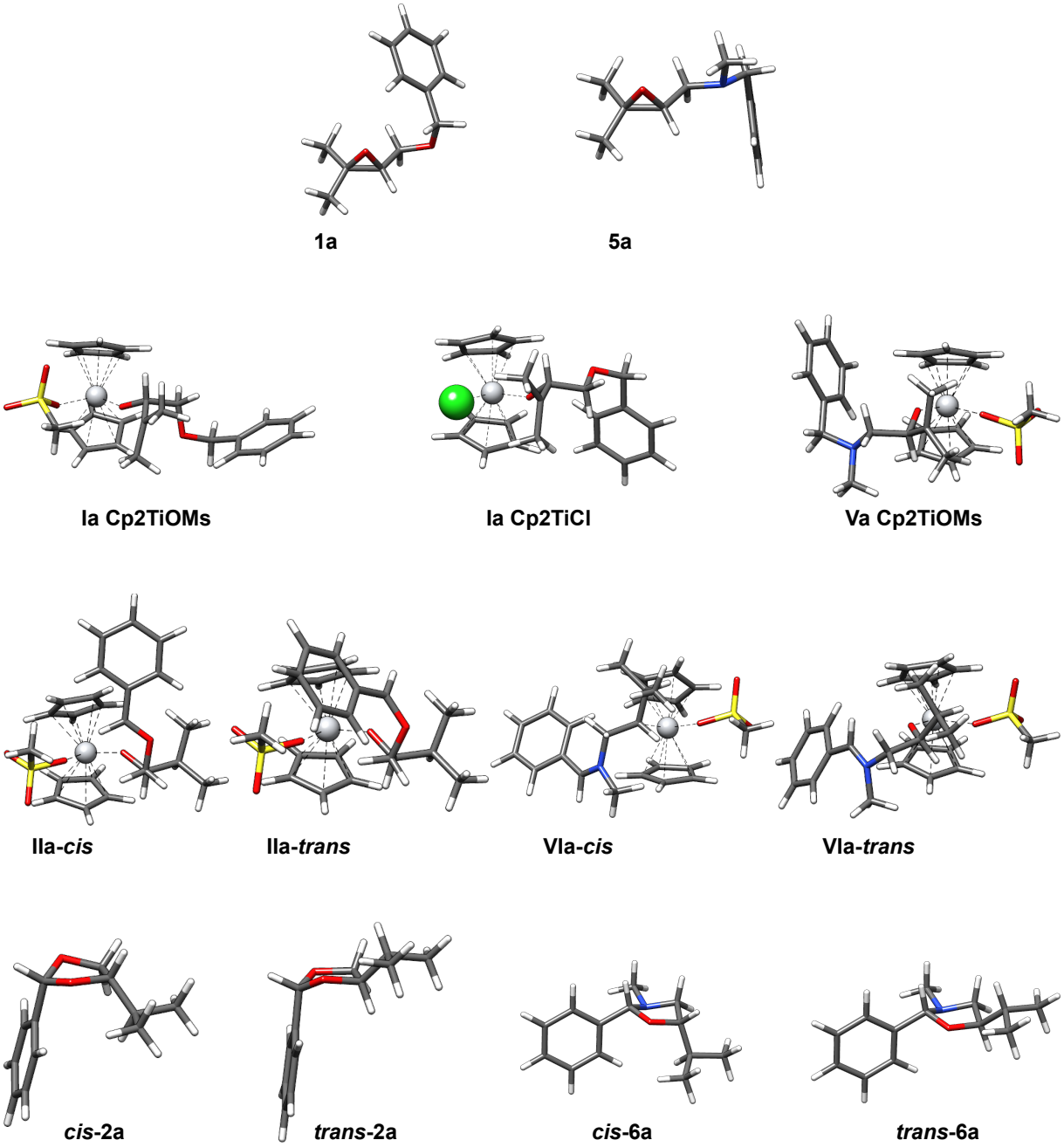


Figure A11.3.: PBEh-3c+DCOSMO-RS(THF) optimized geometries.

Conformational influence

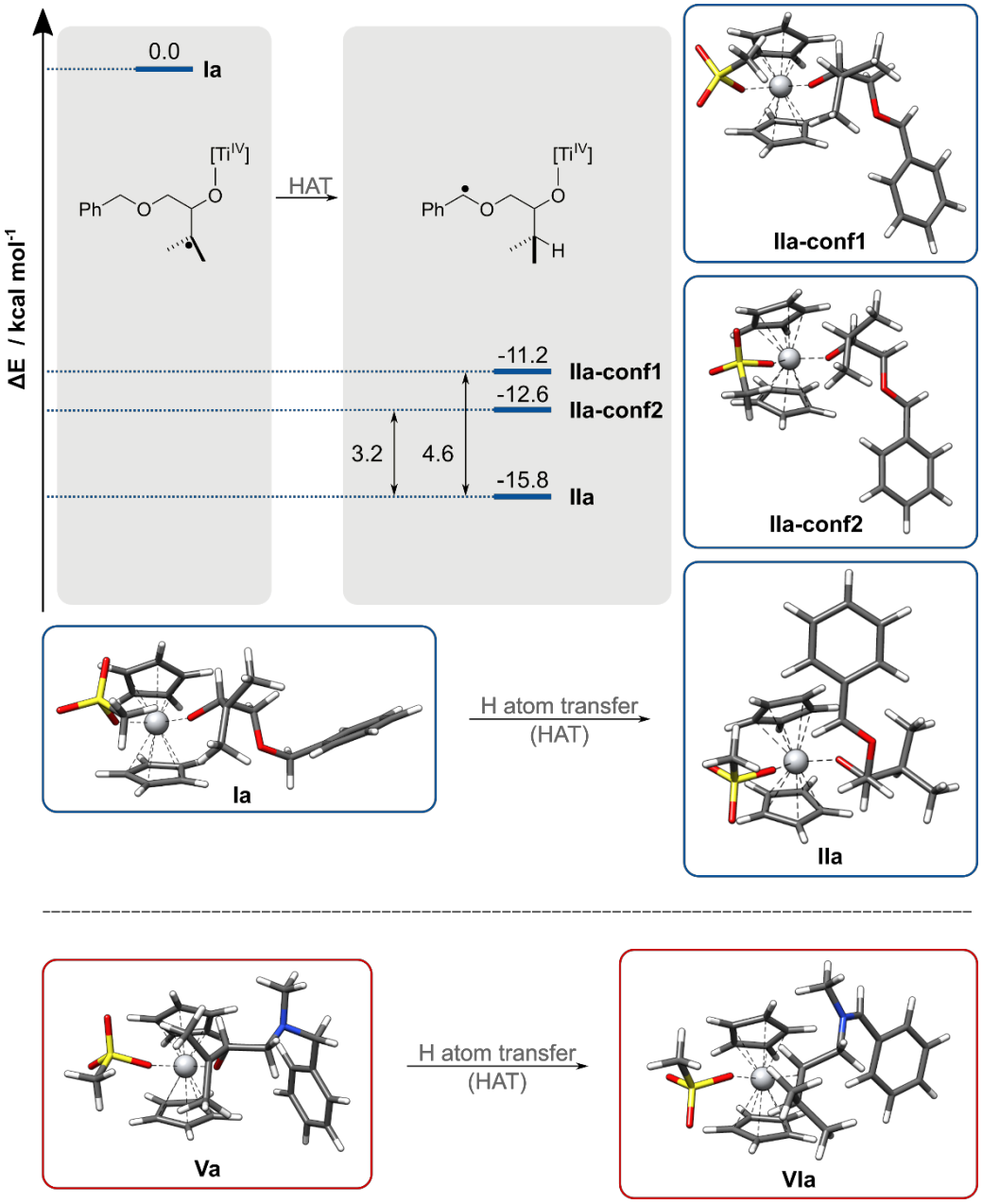


Figure A11.4.: Influence of different conformations on the evaluation of the thermodynamic cycle (see scheme 2 and 4). All energies are evaluated at the PBEh-3c/DCOSMO-RS(THF) level and are given in kcal mol^{-1}.

Applying conformational searches in the elucidation of thermodynamic cycles is essential for identifying lowest energy conformations at the correct position in the catalytic cycle. The necessity can be seen when focusing on the H-atom transfer starting from **Ia** or **Va**. The HAT is an exergonic reaction forming a benzylic radical originating from a tertiary radical. If the computational chemist only investigates the hydrogen atom transfer and does not account for possible conformational relaxation the

evaluated conformation is less stable by 3.2 to 4.6 kcal mol^{-1}. This is based on the energetic (PBEh-3c/DCOSMO-RS(THF)) comparison of structurally similar conformations between **Ia** and **IIa** (e.g. only little relaxation, meaning a simple optimization and no conformational search) with the lowest lying conformation of the benzylic radical **IIa**). The same evaluation for the hemiaminal for comparing structurally close conformations between **Va** and **VIa** against the lowest lying conformation of **VIa** shows only a small difference in stability of 0.3 kcal mol^{-1}. This highlights the strongly deviating structural flexibilities even for similar systems (comparing acetal and hemiaminal) and can only be accounted for by applying robust and efficient conformational searches (for the energetic sorting of the conformer ensembles see Figure A11.5). If the conformational relaxation is not accounted for e.g. for **IIa** the barrier heights of the following reductive elimination reaction would be evaluated from a false starting point. This would wrongfully shift the barrier heights and might result in false statements being made.

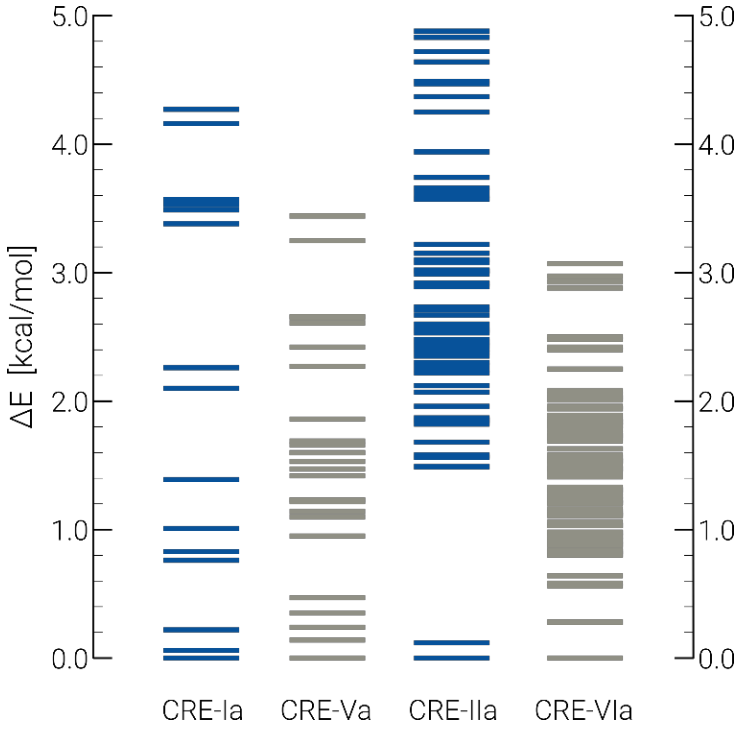


Figure A11.5.: Comparison of the conformer rotamer ensembles of **Ia** and **IIa** as well as **Va** and **VIa** needed for the HAT reaction. All energies are evaluated at the PBEh-3c/DCOSMO-RS(THF) level of theory.

In Figure A11.5 the PBEh-3c/DCOSMO-RS(THF) ranked conformer ensembles are shown for the acetal in blue and the hemiaminal in grey. The conformer ensemble of the hemiaminals (**Va** and **VIa**) are denser compared to the acetal (**Ia** and **IIa**) and by visual inspection the ensembles of the aminal are structurally similar, whereas the acetal seems to be more flexible and structurally different when comparing the tertiary and benzylic radical **Ia** and **IIa**.

A12. Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules

Abstract: The application of quantum chemical, automatic multilevel modeling workflows for the determination of thermodynamic (e.g., conformation equilibria, partition coefficients, pK_a values) and spectroscopic properties of relatively large, nonrigid molecules in solution is described. Key points are the computation of rather complete structure (conformer) ensembles with extremely fast but still reasonable GFN2-xTB or GFN-FF semiempirical methods in the CREST searching approach and subsequent refinement at a recently developed, accurate r^2SCAN-3c DFT composite level. Solvation effects are included in all steps by accurate continuum solvation models (ALPB, (D)COSMO-RS). Consistent inclusion of thermostatistical contributions in the framework of the modified rigid-rotor-harmonic-oscillator approximation (mRRHO) based on xTB/FF computed PES is also recommended.

A12.1. Introduction

A major task of contemporary computational chemistry is the realistic modeling of molecular structure, versatile reactions, and theoretical spectroscopy for analytical (compound identification) purposes[2,47]. One very basic aspect here is determination of the three-dimensional shape of molecules which strongly influences many of their physicochemical as well as macroscopic properties. For rigid molecules, the geometric structure mainly results from the covalent bonding topology and can relatively easily be predicted even by simple theoretical methods. Strong noncovalent interactions (ionic or hydrogen bonding) may also lead to rather rigid structures. In such cases at ambient temperature the structure is well defined and only weakly fluctuates around the corresponding equilibrium (minimum energy) position R_e. However, many fundamentally and practically important systems do not belong to this class but involve a rather flexible spatial arrangement of the atoms that is strongly dependent on temperature and physical state (gas, liquid/solution, solid). This holds for polymers and related soft-matter systems, biomolecules, drugs and even some chemical catalysts. The atomistic origin of this nonrigidity is often the presence of many linearly connected (nonring) covalent single bonds that are rather freely rotatable, leading to an exponentially growing number of conformers with system size. Conformers are characterized by distinct but often shallow potential energy minima regarding specific values of some torsion angles or special (weak) noncovalent interactions but otherwise corresponding to the same chemical bonding framework.[68,510]

As a more specific definition we here employ the term ”rigid” for systems with structure ensembles (SEs) dominated by a single species (i.e., population >98%) at ambient temperature, the more fuzzy wording ”semirigid” for cases in which one structure accounts roughly for only 80-90 % of the SE, and nonrigid for (flexible) systems characterized by several individual species that are needed to account for more than 90 % population.

Compared to the electronic structure problem, which is normally regarded as very complicated, the conformational problem is quite similarly complex or even more challenging. The reason for this lies in the fact that often only a single low-lying electronic state is chemically relevant, which can also be easily narrowed down or estimated and then in principle only refinements have to be calculated by electronic structure, quantum chemistry (QC) methods. The energetically relevant structures in a conformer ensemble, on the other hand, often form an energetic quasi-continuum of which several components are sought and whose composition is initially difficult to estimate.

The efficient generation of thermodynamically correct SEs at a mainly QC theory level for conformers and related low-energy structures has recently been significantly improved by the CREST program[68] (see Sec. A12.2.3). Its automated workflow for systems of almost any chemical composition has been made possible by the introduction of the GFN extended tight-binding (xTB) and force-field (FF) family of methods[61,167,181] combined with a RMSD based biasing potential in meta-dynamics (MTD) simulations.[419]

Here, we concentrate on the efficient, mostly automatized way of refining extended SEs obtained by CREST at higher, normally dispersion corrected DFT, theoretical levels. Relatively large systems with up to hundreds of atoms are our target and hence, a classical description of the nuclear degrees of freedom (DOF) is mandatory. Nuclear quantum effects are added at appropriate steps of the treatment in the framework of a modified rigid-rotor-harmonic-oscillator approximation (mRRHO).[183,185] The general workflow is outlined in Fig. A12.1.

Essentially, the first CREST run on a molecular input structure reduces the extremely huge conformational space (considering may be also tautomers or protomers) with billions of entries to the most important, thermally populated and manageable number of cases (about $10^3 - 10^4$). The size and complexity of the SE after this first filtering step, involving an energetic safety margin (”energy window” to account for method errors), is still enormous involving thousands of structures for a typical drug molecule with 50-100 atoms. This perspective describes efficient algorithms to deal with such SEs without requiring supercomputer resources. Actually, such calculations should be carried out routinely on standard workstation or cluster architectures in hours of computing time.

To carry out the first search step, it is absolutely necessary to have efficient but sufficiently accurate and robust methods available. This is fulfilled by the recently introduced GFN2-xTB[167] semiempirical quantum mechanical (SQM) and GFN-FF[181] approaches, which are consistently available for all elements in the periodic table up to radon (Z=86). They were developed with a special focus on a good description of Geometries, Frequencies and Non-covalent interactions (GFN, see Ref. 61 for a review) and are implemented in the efficient and freely available XTB program.[511] For all DFT or wave function theory calculations we use TURBOMOLE[512,513] or ORCA.[461]

In practice, two main cases of application are identified. The first, dubbed ’thermo-mode’ from now

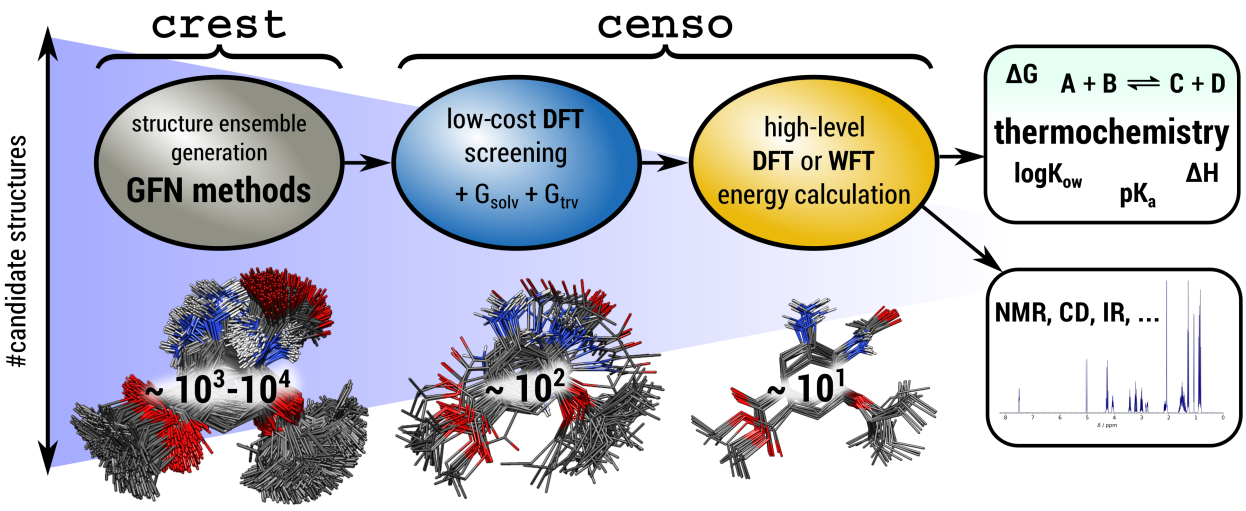


Figure A12.1.: General multilevel workflow. The overlays show structure ensembles (SEs) for the common drug Tamiflu.

on, aims at an accurate thermochemical description. Here, a proper ensemble free energy for, e.g., calculating equilibrium constants (or even rate constants with the help of transition state theory) is the main target. This quantity is often dominated by a few low-lying structures (conformers) and even knowing the free energy of only the lowest structure can be sufficient. In other words, in this mode the main task is first to find with high certainty the structure of lowest free energy within a huge SE ("needle in the haystack" problem). This aspect is not only related to the simulation of the thermodynamics of drug-receptor interactions[514] but also relevant for reaction mechanism calculations.[6,515]

The second application concerns molecular spectroscopy and in fact, the motivation to develop automated SE workflows originated from NMR spectra calculations.[235] In this case, the desired properties (spectral positions and intensities) often strongly depend on the molecular conformation and hence, considering just a single structure is often totally inappropriate. Instead, the procedure has to provide accurate thermal populations of many low-energy structures as input for further property computations. Recent examples from our group where the importance of spectra averaging over a proper SE has been demonstrated are in addition to NMR, electronic circular dichroism[209], EPR[516] and vibrational spectroscopy.[517] In the following we denote this run-type as "property-mode". We are aiming at a composite, multilevel procedure that can treat both modes on equal footing by just adjusting some technical parameters.

In the literature, similar but theoretically and technically less advanced combined quantum chemistry/force-field workflows have been established mainly for acidities (pK_a values) and to a lesser extent partition coefficients ($LogK_{ow}$ values), for spectroscopy, or for reaction mechanisms (see Refs. 18,23,518–521). The procedure proposed here is more complete in the ensemble generation step, applies more accurate quantum chemical and solvation methods, is thermodynamically rigorous, and overall contains little empiricism and is quite generally applicable.

In the next section, a brief description of the underlying thermodynamic theory is given. Then, the components in our multilevel workflow are discussed, namely computation of the (gas phase) electronic energy, generation of the SE, thermostatistical contributions including special conformational and anharmonic terms, and the important solvation contribution. This part is completed with a description of the CENSO DFT reranking procedure. Finally, application examples for optical rotation (OR) calculations of drug molecules, and LogK_{ow}/pK_a values are discussed.

A12.2. Theory

A12.2.1. General considerations

We follow a nondynamical, equilibrium structure approach based on a well-defined molecular free energy and structure for each species (conformer, tautomer, etc.). Molecular meta-dynamics (MTD) simulations are used here (see Sec. A12.2.3) merely to explore the potential energy surface (PES) efficiently and to find favorable, highly populated minima but not for direct computations. An essential feature of our workflow is that every structure in a decision step of the filtering procedure is very tightly geometry optimized so that energy and structure comparisons can be made (c.f. section A12.2.7).

Assuming standard conditions (1 atm gas, 1 M solution), the free energy G of a molecule in a definite minimum energy conformation is given by Eq. A12.1

$$G(T) = E_{gas} + \delta G_{solv}(T) + G_{trv}(T), \tag{A12.1}$$

where E_{gas} is the total molecular gas phase energy, $\delta G_{solv}(T)$ is the temperature dependent free solvation energy (for the solvated case), and G_{trv} is the thermostatistical correction from energy to free energy at absolute temperature T accounting for translation, rotation, and vibration DOF and including the zero-point vibrational and volume work terms. The indicated temperature dependence is dropped from now on and implicitly assumed if applicable. Under equilibrium conditions, any reaction can be described by the corresponding difference of product and reactant free energies denoted if necessary by a superscript with capital letters, e.g., G^A refers to the free energy of species A. The approach is general but we restrict ourselves here for the sake of simplicity mainly to mono-molecular equilibria $A \rightleftarrows B$ (see Fig. A12.2). For rigid molecules this description is already sufficient as long as reasonable approximations for the three terms in Eq. A12.1 are applied, which are discussed in more detail in the Secs. A12.2.2 to A12.2.6.

However, several problems arise with nonrigid systems. First, the concept of molecular structure is replaced by that of an ensemble of structures which are in equilibrium with each other. Second, this SE is *a priori* not known in most cases and hence has to be found employing a PES of reasonable quality. Because the systems of interest are typically large and can be of any chemical composition, precalculated or machine-learned PES are not used here, but instead, it is computed on-the-fly, preferably by quantum chemical methods.

Note that in many current computational chemistry studies the SE is entirely neglected even for relatively flexible systems but is replaced by a single, often assumed to be, lowest (free) energy structure.

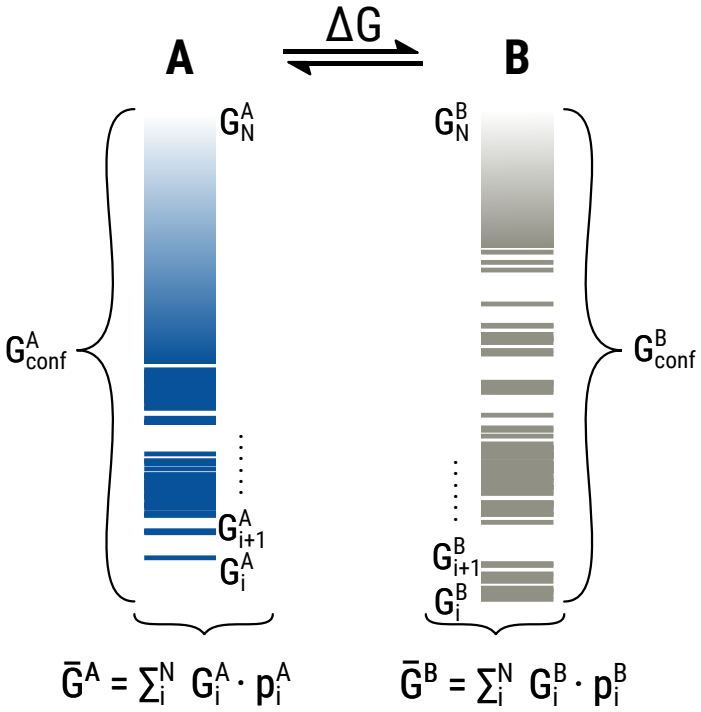


Figure A12.2.: General definitions of molecular and ensemble thermodynamic properties for an example equilibrium.

Under certain conditions discussed below, this is a reasonable approximation, but it generally prevents results from being obtained with high accuracy. Furthermore, as already noted, the important branch of computational spectroscopy requires detailed knowledge of the SE already for the semirigid case.

Under the assumption that all DOF (electronic, translation, rotation, vibration, conformation, solvent) can be separated, the free energy of the SE is obtained as

$$G_{SE} = \overline{G} + G_{conf} \tag{A12.2}$$

where G_{conf} is the conformational free energy part consisting of N distinguishable conformers in the SE. An efficient new procedure[185] for a robust computation of G_{conf} including rotamers (and their symmetry) for each conformer based on low-level PES is briefly discussed in Sec. A12.2.5. The average $\overline{G}$ in Eq. A12.2 is given

$$\overline{G} = \sum_i^N p_i G_i \tag{A12.3}$$

with

$$p_i = \frac{e^{-G_i/kT}}{\sum_i^N e^{-G_i/kT}} \tag{A12.4}$$

where k is the Boltzmann constant and G_i is the molecular free energy of the ensemble member i according to Eq. A12.1. In the example case of $A \rightleftarrows B$, the SE of A and B has to be found so that $\Delta G = G^B_{SE} - G^A_{SE}$. Practically this is done in completely independent computations for A and B even in

the case that the two molecules have the same covalent topology but just differ by the applied solvent (which is the very important case of solution phase partition coefficients like the widely used LogK_{ow}, see Sec. A12.3). These basic thermodynamic considerations are shown schematically in Fig. A12.2. It is important to distinguish between the free energy of an individual R_e structure (Eq. A12.1) and the ensemble value in Eq. A12.2 including many of them.

The physically plausible partitioning of G into the terms in Eqs. A12.1 and A12.2 provides a detailed understanding of the relevant contributions (and errors) to the total free reaction energy ΔG according to

$$\Delta G = \Delta\overline{E}_{gas} + \Delta\overline{\delta G}_{solv} + \Delta\overline{G}_{trv} + \Delta G_{conf} \tag{A12.5}$$

where the overlined quantities are computed in analogy to Eq. A12.3.

In the following subsections, the most important theoretical and technical aspects of the proposed multilevel approach are outlined in more detail. This involves a discussion of typical errors in the contributing terms E_{gas} (choice of geometry and high-level QC method), G_{solv} (choice of solvation model), and G_{trv} (modified RRHO model).

A12.2.2. Electronic energy

Besides a good sampling algorithm and accurate thermostatistical as well as solvation free energy contributions (see below), an efficient, robust and reliable method to calculated E_{gas} in Eq. A12.1 and the derived relative energy for conformational changes ("conformational energy") is needed. Specifically, both strongly directional short-range covalent interactions as well as intramolecular noncovalent interactions (NCI) must be accurately described. Although numerous accurate *or* fast methods are available for this purpose, the simultaneous fulfillment of both requirements along with broad applicability (excluding cases with significant self-interaction error) and numerical stability could not be achieved until very recently with the efficient meta-GGA composite method r^2SCAN-3c (*vide infra*).[156]

The "gold-standard" reference are conformational energies calculated with CCSD(T) extrapolated to the complete basis set (CBS) limit.[522] However, even with local approximations such as, e.g., DLPNO-CCSD(T),[523] this level is computationally too demanding for an exhaustive sampling. This also applies to hybrid and double hybrid DFT approximation/"large basis set"(L) methods, at least if many conformers or systems with more than about 50 atoms are in the focus. Nonetheless, these high-level approaches are important to judge the performance of lower-level yet clearly faster methods via comparison of representative benchmarks comprised of smaller molecules. Of particular importance here are efficient composite DFT, SQM, and FF methods developed for large systems. As will be shown in the following, the '3c' composite DFT methods[155,156] and, on a lower accuracy scale applicable to (very) large molecules, also the GFN family of methods are essentially the only methods available that meet all of the aforementioned criteria (*vide infra*).

In order to give a perspective on the routinely achievable accuracy for conformational energies, in total, more than 750 conformers were assessed. This includes the eight conformational energy benchmark subsets of the GMTKN55 database[73] and two sets (MPCONF196[524] and 37CONF8[525]) comprised of medium-sized molecules (up to about 120 atoms) with (bio)chemical importance. Addition-

ally, the ACONF12 set[156] composed of 12 $C_{12}H_{26}$ alkane chain conformers, which are prototypical for many chemical systems as well a challenging set (TMCONF16[156]) consisting of four rather flexible organometallic complexes of about 50 atoms, are investigated. In order to meaningfully evaluate the deviations from the (DLPNO-)CCSD(T)/CBS reference values, both the range of conformational energies and the accuracy of the reference values are crucial. A high accuracy theory level corresponds to average deviations of about 0.2 kcal mol^{-1}, which would allow to calculate sufficiently accurate thermal populations. With the exception of the MPCONF196 and 37CONF8 sets, which include larger molecules as well as a wider energy range with values up to 30 kcal mol^{-1}, the vast majority of the assessed values are in the practically relevant 0-5 kcal mol^{-1}regime. For the latter, the reference values achieve approximately the required sub-kJ mol^{-1} accuracy, while the lower accuracy of the MCONF196 and 37CONF8 reference values makes the attainability of smaller MAD values than around 0.5 kcal mol^{-1} practically impossible. The statistical evaluation of the performance of selected methods by means of (weighted) mean absolute deviations is summarized in Fig. A12.3.

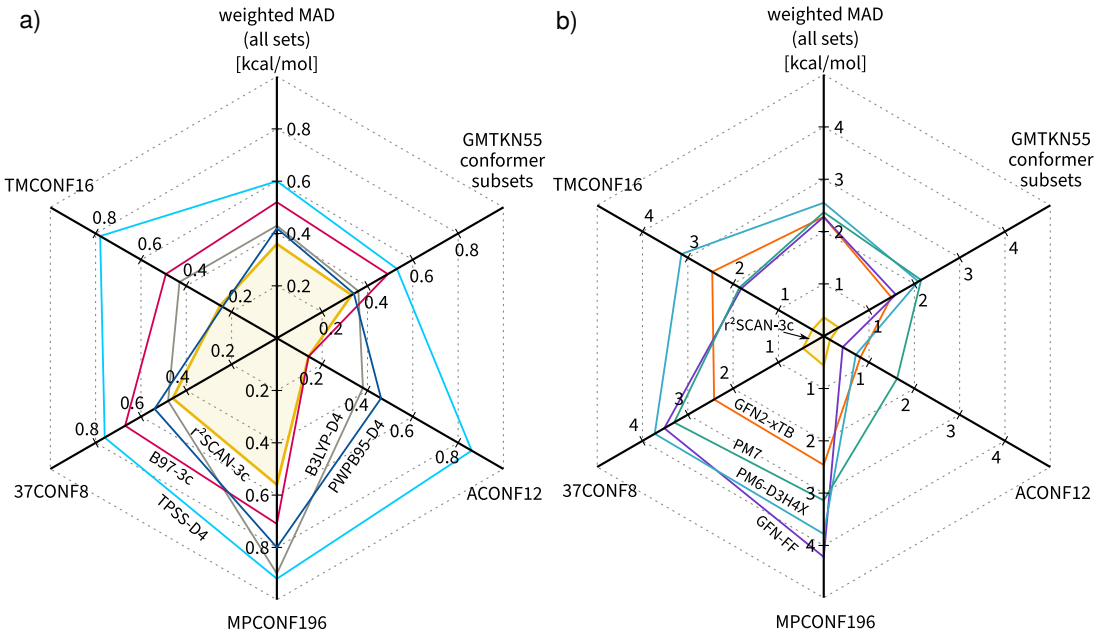


Figure A12.3.: Mean absolute deviations (MADs) of computed conformational energies for respective benchmark sets calculated with a) four DFT/def-QZVP and two '3c' methods with def2-QZVP basis set, and b) three SQM, GFN-FF, and r^2SCAN-3c (filled yellow area). The weighted MAD for the eight GMTKN55 conformer subsets and all sets, respectively, is calculated analogously to the WTMAD-2 defined in Ref. 73.

The yellow highlighted area in Fig. A12.3, which spans the performance of r^2SCAN-3c for the assessed sets, is impressively small. No other of the tested DFT/L methods including hybrids and even double hybrids are competitive here. Of particular importance is the outstanding performance for the ubiquitous alkane chain conformers (ACONF12) as well as for the larger and (bio)chemically relevant molecules of the MPCONF196 and 37CONF8 sets, where r^2SCAN-3c approaches the accuracy of the reference values. The fact that the high accuracy can also be transferred to the organometallic complexes

shows how robust and generally applicable this method is. That this holds for closed-shell organometallic reactions was already shown in the original r^{2}SCAN-3c publication and preliminary investigations indicate that very good results are also obtained for open-shell transition metal reaction energies.

The r^{2}SCAN-3c conformational energies are more accurate than any previously considered DFT/"large basis set" method and are computed even 2–3 orders of magnitude faster compared to the latter.[156] r^{2}SCAN-3c may be used consistently for final geometry optimization and energy ranking and hence, it represents the perfect method for a multilevel scheme as depicted in Fig. A12.1. In rare cases, where r^{2}SCAN-3c results may be of questionable accuracy, cross-checking is only meaningful at all with (local)CCSD(T)/CBS or methods with comparable accuracy, which, however, would only be applicable for smaller system sizes. This is indeed a paradigm shift, rendering the computationally more expensive combination of (meta-)GGA/"smaller basis set" for structure optimization and (double)hybrid/"large basis set" for energies almost obsolete.

Though the MADs are significantly larger compared to r^{2}SCAN-3c, the performance of GFN2-xTB and GFN-FF is very respectable considering their orders of magnitude lower computational costs, especially for the latter. For molecules with 500-1000 atoms or even larger, GFN2-xTB and GFN-FF, respectively, provide robust high-speed alternatives to r^{2}SCAN-3c though their accuracy is somewhat limited. PM7[160] performs better than PM6-D3H4X,[159,526] but given the fact that they both are also computationally more expensive than GFN2-xTB and GFN-FF and that their weighted total MADs for all sets is larger compared to the two latter, PM6-D3H4X and PM7 are not recommended. In any case, the significantly lower accuracy of the SQM/FF methods compared to DFT should be taken into account by choosing a suitably large (conservative) energy window in a multilevel (filtering) protocol (see Sec. A12.2.7). However, as will be discussed in the next two subsections, free energies in solution (and not pure gas phase energies evaluated typically in such benchmarks) are the decisive factor for the observable conformer population and besides a reliable E_{gas} contribution, approximate thermostatistical and solvation free energy contributions should be able to keep up with this impressive accuracy.

A12.2.3. Generating the structure ensemble

For any meaningful computational work, exploration of the PES (and the global minimum in particular) is necessary.[527] In the nonrigid molecule case this boils down to the main task of finding the most favorable conformers, for which many computational tools and algorithms exist in the literature.[528] Recently, we have published one of such under the acronym CREST,[68,462] which is an abbreviation for Conformer-Rotamer Ensemble Sampling Tool. The underlying idea of CREST is to generate SEs based on physically plausible MTD simulations, including a history dependent bias potential to heavily accelerate exploration of the PES. A core component of the sampling procedure are atomic Cartesian RMSD based MTD simulations, which were introduced in Ref. 419 for this purpose. The biasing contribution is given by a Gaussian-type potential

$$V_{bias} = \sum_{i}^{n} k_i \exp(-\alpha \Delta_i^2) \,, \tag{A12.6}$$

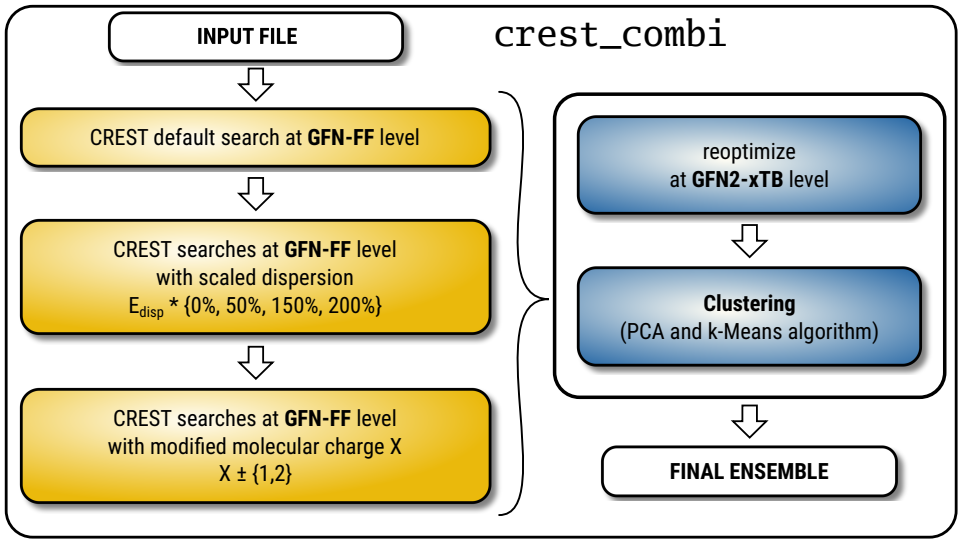


Figure A12.4.: Outline of the CREST_COMBI script.

where k_i and α define the potentials shape and the RMSDs enter as collective variables Δ_i, with n reference structures. During the MTD, structures that are similar to ones from previously explored regions of the PES (i.e., with a small RMSD) receive an energy penalty and therefore are automatically avoided. This way structural changes are achieved much faster during the simulation than with conventional MD and much larger portions of the PES can be explored within the same computational time frame. Structures are, however, only comparable (with respect to their energy and RMSD) upon a postprocessing, full geometry optimization (which is omitted in many existing approaches[528]). A prerequisite for routine application of such an algorithm is use of the efficient GFN methods in the MTD and optimization parts where on the order of 10^5 intermediate geometries are optimized.

A single search in CREST produces a SE that is sufficiently complete regarding low-energy structures (<6-8 kcal mol^{-1}), but usually suffers from errors for the approximate PES at the underlying level of theory. To extend the idea of a broad exploration of the PES aiming at a complete description of the conformational space after postprocessing with DFT (section A12.2.7), a series of multiple conformational searches on slightly modified PES are conducted. More specifically, several simulations are efficiently run at the GFN-FF level of theory with either up- or down-scaled dispersion interactions (0 % – 200 %), and with positive or negative molecular charges (±1, ±2 relative to the input charge). The sequential procedure is outlined in Fig. A12.4. These artificial settings for the intramolecular electrostatic or dispersion interactions alter the PES and will produce new minima that still closely resemble the target molecule. The corresponding searches serve the sole purpose of generating more diverse conformations, i.e., accounting for the uncertainties of the low-level PES. As in the CREST algorithm itself, the different SEs obtained from the searches need to be reranked and compared at a chosen theoretical level in order to be comparable. All generated (merged) GFN-FF structures are usually reoptimized at the unmodified GFN2-xTB level. As a further approximation the obtained ensemble is then clustered, using a combined principle component analysis (PCA) and k-Means clustering algorithm[529] with dihedral angles as descriptors.

The entire procedure including all conformational searches is executed by a script called CREST_COMBI, which yields a final SE to be postprocessed as described in Sec. A12.2.7. Adaption to different workflows regarding computation time by changing the theory level (GFN2-xTB instead of GFN-FF), the number of runs or length of the MTD simulation is straightforward. All conformational searches can be combined with available implicit solvation models (see Sec. A12.2.6) and it is recommended to include solvation effects already in this search step. One drawback of the proposed workflow, which the user should always keep in mind, is that due to the stochastic nature of the MTD runs and the finite precision of the geometry optimization, results from different runs with the same input are not perfectly reproducible and can differ by about a few tenth of a kcal mol^{-1} for G even if the very same global minimum is found. If very high accuracy is required, we simply suggest to average data over several independent runs.

The quality of the ensemble is largely determined by the electronic structure method used in the conformer search with CREST. The CREST_COMBI script tries to fix possible deficiencies by performing multiple searches on different (modified) PESs. Another option is to improve low-level SQM/FF PES by including bias potentials based on accurate DFT conformer energies and using a corrected PES in the conformer MTD search. Such an approach termed Geometric Energy Surface Correction (GESC) has been implemented and can be applied using the RMSD bias potential (Eq. A12.6) where n is the number of reference conformers for which DFT conformational energies have been computed, k_i is the SQM/FF error for conformer i, and α_i is automatically determined from the RMSDs to all other reference structures. The GESC ansatz allows for an automatic structure based PES correction and further details will be published elsewhere. It is used by default in our most complex search level for very high quality SEs ('LEVEL2', see Sec. A12.3).

A12.2.4. Thermostatistical contributions

Statistical thermodynamics enables the transition from energy E to free energy G as given in Eq. A12.1 and this has a major impact for the comparison of theory and experiment[530–532]. Usually, thermostatistical contributions are calculated within the rigid-rotor-harmonic-oscillator (RRHO) approximation[533,534]. Herein, vibrational frequencies are obtained within the HO approximation by a second-order Taylor expansion around R_e while the rotational DOF are described as a rigid-rotor.

As low-lying HO vibrational frequencies (HOVF) are notoriously inaccurate, a correction scheme for G_{trv} in Eq. A12.1 was introduced in 2012 in the context of free energy calculations for large supramolecular complexes[183]. Therein, G_{trv} values were obtained in a low-frequency modified RRHO approximation (mRRHO), i.e., $G_{trv} \approx G_{mRRHO}$. The physically sound rotor-type modification of the standard RRHO term provides finite entropy values for low-frequency modes, which otherwise asymptotically approaches infinity and thus corrects their large and numerically problematic[535] contribution.

Because the computation of HOVF by first-principles methods with sufficiently large AO basis sets becomes prohibitively expensive, faster but still accurate methods are desirable here. The GFN methods were recently shown to yield good G_{mRRHO} values in comparison to DFT calculations[236]. Because of their general applicability to all elements, GFN methods open up new dimensions for the calculation of free energies in supramolecular chemistry, material science and for biomacromolecules[516,536].

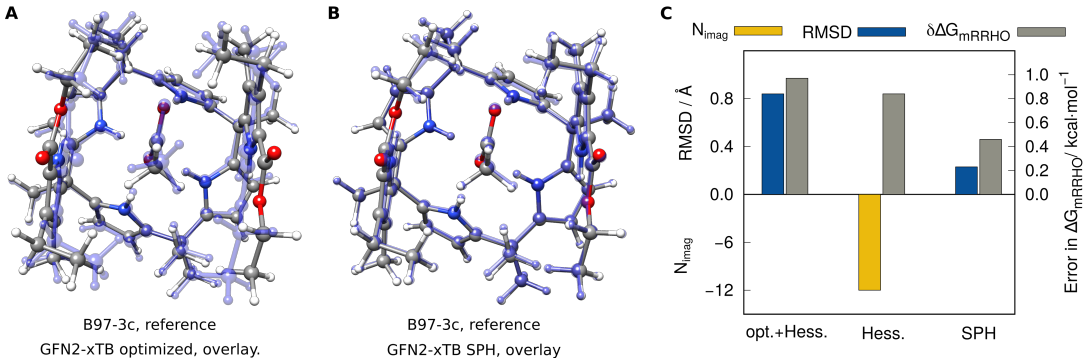


Figure A12.5.: Calculation of the thermostatistical contribution to the free energy for the inclusion of acetate in calix[4]pyrrole. (**A**) B97-3c[155] reference structure and GFN2-xTB optimized structure as blue overlay. (**B**) B97-3c reference and the structure obtained within the SPH approach as blue overlay. (**C**) Thermostatistical contribution calculated with and without reoptimization of the reference geometry, and with the SPH approach in comparison to the DFT result. N_{imag} and RMSD values are added up for host, guest, and complex.

Further improvements have recently been achieved by introducing the *Single Point Hessian* (SPH) approach[184]. Until now, only fully optimized, true equilibrium structures were used to compute the (m)RRHO term which limits the application of fast low-cost methods in the computationally costly HOVF calculation step. When the low-level method (normally FF) provides a qualitatively wrong geometry, inaccurate G_{mRRHO} values are obtained. The often implied assumption, that structural differences between different theoretical levels are minor and can thus be neglected can be invalid. Fig. A12.5A shows an example for the differences between DFT and SQM geometries. With the SPH approach, the DFT reference structure used in the final step of the multilevel procedure is retained in the low-level Hessian calculation by applying the RMSD bias given in Eq. A12.6 (see Fig. A12.5B). Free association energy contributions ΔG_{mRRHO} for acetate binding in calix[4]pyrrole[537] were calculated at the GFN2-xTB level of theory starting from the reference B97-3c structure (see Fig. A12.5C). Calculating HOVF directly from the DFT structure (Hess.) leads to 12 significant imaginary modes, which vanish upon reoptimization (opt.+Hess.). This, however, leads to a large RMSD of 0.84 Å between the DFT and GFN2-xTB structure. The deviations in ΔG_{mRRHO} from the DFT result amount to 0.85 and 0.97 kcal mol^{-1}, respectively. The SPH approach, on the other hand, shows no imaginary modes, an RMSD of just 0.2 Å compared to the DFT reference and the error in ΔG_{mRRHO} diminishes to 0.45 kcal mol^{-1}.

The computation of G_{mRRHO} with the SPH procedure is the default in our standard protocol where it is normally applied for GFN2-xTB frequencies at finally DFT optimized geometries (dubbed SPH(GFN2@DFT)). However, for very large systems SPH(FF@GFN2) calculations can also be recommended.

A12.2.5. Conformational and anharmonic terms

Consideration of a SE has other implications beyond the determination of a global minimum structure. This mainly concerns the calculation of the G_{conf} term in Eq. A12.2, which is related to an entropic contribution, commonly referred to as conformational or configurational entropy S_{conf}[538], as a result

of mixing of multiple, populated conformers according to

$$S'_{conf} = -R \sum_{i}^{N} p_i \ln p_i \ , \tag{A12.7}$$

where the sum is taken over all distinguishable conformers in the ensemble and p_i is the Boltzmann population for the respective structure as given by Eq. A12.4. The formulation in Eq. A12.7 is also known as Gibbs-Shannon entropy. If for the final result the average $\overline{G}$ according to Eq. A12.3 is calculated over the SE, then from S'_{conf} the ensemble free energy $G_{conf} = -TS'_{conf}$ is calculated and can be added directly. Otherwise, for example if G is calculated only for the global minimum structure and is *not* averaged, it is advisable to additionally estimate the average contribution $\overline{S}_{trv}$ via

$$\overline{S}_{trv} = (\sum_{i}^{N} p_i S_{trv,i}) - S_{SPH,ref} \ , \tag{A12.8}$$

at a SQM or FF level and add it to the configurational entropy $S_{conf} = S'_{conf} + \overline{S}_{trv}$. In this case a conformational average must also be included for the enthalpy and $G_{conf} = H_{conf} - TS'_{conf}$. Herein, $S_{SPH,ref}$ refers to the above described SPH/mRRHO term, calculated for the global minimum structure, and $\sum_i^N p_i S_{trv,i}$ is a population average over the whole ensemble at SQM or FF level. This term can be seen as a correction to S_{conf} resulting from variations due to structural diversity. Note again that, for the correct ΔG (Eq. A12.5) the average $\overline{S}_{trv}$ is already included via $\overline{G}_{trv}$.

In the calculation of reaction energies or partition coefficients (Eq. A12.5), the difference between the conformational free energies ΔG_{conf} can be significant if the SE differ strongly in size or energy distribution, e.g., see Fig. A12.2. Typical examples are reactions with a large change of internal DOF, like ring-opening or -closing reactions. The ΔG_{conf} term can become an accuracy bottleneck for critical cases and should be investigated if there are notable differences in the SEs. In practice, this contribution will often be neglected due to high computational effort required for the accurate determination of S_{conf}. A thorough investigation of the influence of G_{conf} is yet necessary, which could easily be done with the here presented tool chains. For the automatic calculation of S_{conf} a special run type has been implemented in CREST, but details are discussed elsewhere.[185]

Nonconformational vibrational anharmonicity effects also occur for the higher-frequency modes and can be accounted for semiempirically by standard linear scaling of the HOVF values as discussed in detail in Ref. 185. In this way for rigid medium size molecules, errors of <1-2% for absolute entropies can be obtained at standard DFT levels (rather 2% error for GFN2-xTB). For non-rigid molecules and when relative entropies (free energies) between structurally similar species are considered (i.e., for conformational SEs), such errors may partially cancel and are mostly not relevant on a 0.2-0.3 kcal mol^{-1} target error level.

For the important entropy term which is strongly influenced by anharmonic and conformational effects, the mRRHO+S_{conf} approach can be well benchmarked with reference to accurate experimental gas phase values as shown in Fig. A12.6, which documents high to very high accuracy of composite DFT as well as GFN2-xTB treatments.

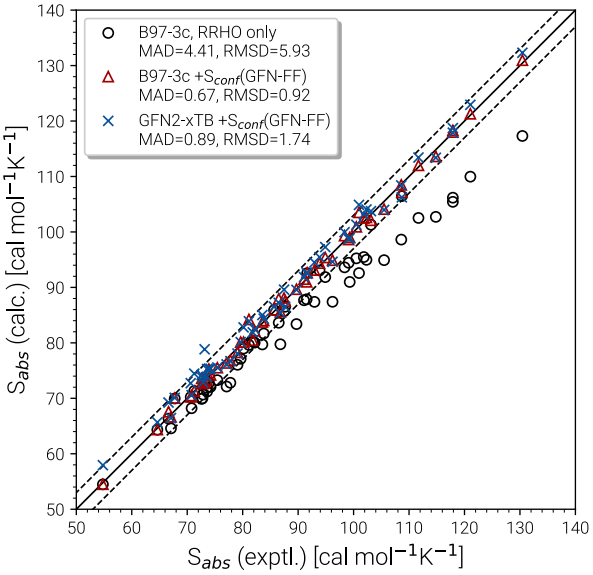


Figure A12.6.: Parity plot for the comparison of experimental and calculated absolute entropies in the gas phase for 58 molecules (extended set from Ref. 539). Experimental values are compared with standard RRHO from B97-3c, with frequency-scaled mRRHO values for B97-3c, and mRRHO GFN2-xTB values. The last two methods are corrected additionally by the conformational term S_{conf} (at GFN-FF level). The dashed lines denote a chemical accuracy range of about 3 cal mol^{-1}K^{-1}.

A12.2.6. Solvation free energy

To generate realistic and accurate structure ensembles with CREST (see Sec. A12.2.3), it is imperative to include solvation effects in the theoretical treatment from the very beginning, see Refs. 540 and 50 for excellent, detailed overviews. The importance of solvation for generating conformer ensembles is exemplified by Fig. A12.7 for the case of hexadecane.

In the gas phase, the folded conformer is lower in energy than the commonly considered linear form, which is in agreement with high-level experimental and theoretical studies[541]. However, in chloroform, the extended conformer having the larger solvent accessible surface area (SASA) is stabilized more by the more negative free solvation energy term, resulting overall in a qualitative change of conformation from the gas to solution phase. This is a common case for nonrigid, not too small (>30 atoms) molecules.

Due to the sheer amount of conformers in an SE (on the order of 10^5 in this example at ambient temperature), a robust and efficient solvent model, that does not inflate the computational cost is needed in the initial steps of a multilevel approach. The GB[542]/SA[543,544] implicit solvation model approximates the solvation free energy by partitioning into electrostatic (ES) and nonelectrostatic contributions.

The short-range ES interactions are described by the Born model applying Coulomb's law which is called the generalized Born model. For atoms with partial charges q_i and q_j in a compound with an internal dielectric constant of ε_{in}, embedded in a medium with dielectric constant of ε_{out}, the most

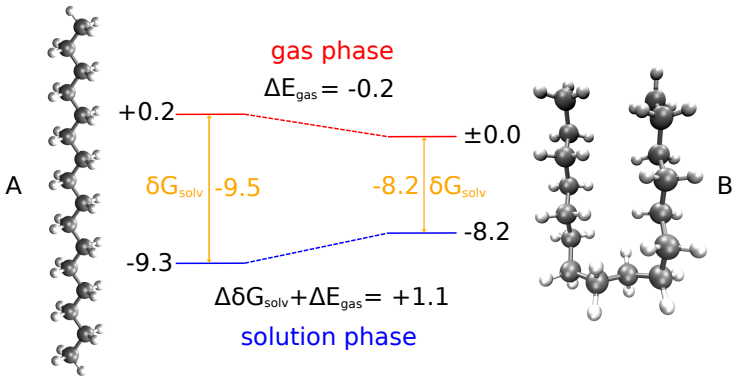


Figure A12.7.: Energy levels (in kcal mol^{-1}) of a linear (A) and a folded (B) conformer of hexadecane in the gas phase and in chloroform. The (free) energies were obtained at the r^{2}SCAN-3c/COSMO-RS level of theory. The energy of conformer B in the gas phase is set to zero. $\Delta\delta G_{solv}$ and ΔE_{gas} refer to the reaction $A \rightarrow B$.

commonly used equation[543] in conventional GB theory is

$$\Delta G_{ES}(q_i, q_j) = -\frac{1}{2}\left(\frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{out}}\right)\sum_{ij}\frac{q_i q_j}{f_{ij}}, \tag{A12.9}$$

with f_{ij} being the interaction kernel. Instead of the canonical form proposed by Still[543], the more recently introduced p16 kernel[545] is used in the latest version of xTB as it is generally more accurate and robust. The non-ES interactions are taken to be proportional to the SASA as

$$\Delta G_{NES} = \gamma_i SASA_i, \tag{A12.10}$$

with $SASA_i$ being the accessible surface area belonging to atom i and γ_i being an element specific parameter, which is determined by fitting to experimental solvation free energies.

The analytical linearized Poisson-Boltzmann (ALPB) model[186], used in this work, is an extension of the GB/SA solvation model for the ES. For an accurate description, the global and element specific parameters used in the ALPB model need to be parameterized to experimental or theoretical data. Experimental reference solvation free energies were obtained from the Minnesota solvation database (MNSOL) version 2012[188,546]. However, this approach is only feasible for a few solvents and elements, namely hexadecane, octanol and water and the elements H, C, N, O, F, Si, P, S, Cl and Br. For other, more rare cases, the parameters were optimized on a fit set containing about 2500 compounds for 67 elements with COSMO-RS reference values (version 2016 with default treatment, see below). Further details will be published separately.

However, while the ALPB model can be used for efficient prescreening of SEs, its accuracy is limited. The use of apparent surface charge dielectric continuum models, like the conductor like screening model (COSMO)[242], can improve the description. COSMO employs simpler conductor boundary conditions (instead of the exact dielectric ones) that hold true for a medium with infinite dielectric permittivity ε and are scaled by a function $f(\varepsilon)$ to approximate the solution for a finite dielectric. One of the

main weaknesses of COSMO, that it shares with other continuum models, is due to the use of a macroscopic static dielectric constant ε for the polarization of a solvent. This assumption seems reasonable for many organic solvents, but treating a highly polar solvent like water as a homogeneous dielectric continuum, might lead to severe errors[547]. Additionally, many of these models do not explicitly describe hydrogen bonding and are generally limited to macroscopically parameterized pure solvents at room temperature, due to the usage of a single ε value.[547] In contrast, partition coefficients like LogK_{ow} (see Sec. A12.3) involve a mixture of octanol and water.

The introduction of the COSMO model for realistic solvation (COSMO-RS)[289,303] made it possible to treat both the solvent and the solute on solid statistical thermodynamic grounds, including solvent mixtures. The core concept is a solvent specific function, that describes the affinity of a solvent to a molecular surface of polarity σ, called the σ-potential. By integrating the σ-potential over the surface of the solute, COSMO-RS can approximate the chemical potential of the solute as a function of temperature and concentration in pure or mixed solvents[547]. With the introduction of direct COSMO-RS (DCOSMO-RS)[243], the COSMO-RS solvent response function was integrated into the electronic DFT calculation enabling geometry optimization in solution consistent to the COSMO-RS (free) energy level.

Because (D)COSMO-RS as relatively high-accuracy models are based on DFT calculations with polarized TZ AO basis sets, they are not routinely applicable in early stages of a multilevel treatment to large SEs. As described in Sec. A12.2.7, they are applied to a largely reduced number of candidates (about $10^2 - 10^3$) optimized at the GFN2-xTB/ALPB level of theory while DFT/DCOSMO-RS is used to finally optimize to the truly lowest ones ($10^1 - 10^2$).

The choice to use COSMO-RS as the solvation model was fueled not least by the fact that in current comparative benchmarks, like the LogK_{ow} SAMPL6[548] challenge, it outperformed all other submissions,[549] including other established continuum methods and even explicit FF-based solvation treatments. This finding is corroborated by our own experience over the years in dozens of applications. We finally show here a comparison of the performance of the ALPB/COSMO-RS models for a typical solvation free energy benchmark. These and all other COSMO-RS calculations in this work were conducted in its 2019-standard parameterization.

Fig. A12.8 shows calculated δG_{solv} values for chloroform in comparison with experimental data for 50 selected compounds from the MNSOL database. ALPB(GFN2-xTB) reaches a mean absolute deviation (MAD) of only 0.96 kcal mol^{-1} and a standard deviation (SD) of 1.17 kcal mol^{-1} with an error range of about 5.0 kcal mol^{-1}, which is sufficiently accurate in the SE generation step. COSMO-RS performs significantly better with a MAD of 0.54 kcal mol^{-1} and a SD of 0.66 kcal mol^{-1}. This accuracy for absolute δG_{solv} leads due to error cancellation to uncertainties for relative (conformational energies) of about 0.1-0.3 kcal mol^{-1} depending on size and complexity of the system and solvent for medium sized (about 50 atoms), neutral molecules. For ions as considered in pK_a calculations, the errors may increase to about 1-2 kcal/mol which can, however, be lowered by explicit, microsolvation techniques.[550]

A12.2.7. The CENSO algorithm

Searching for minima on a PES can only be performed by very fast methods, since a huge number of geometry optimizations have to be performed in our scheme and for this reason, SQM or reliable

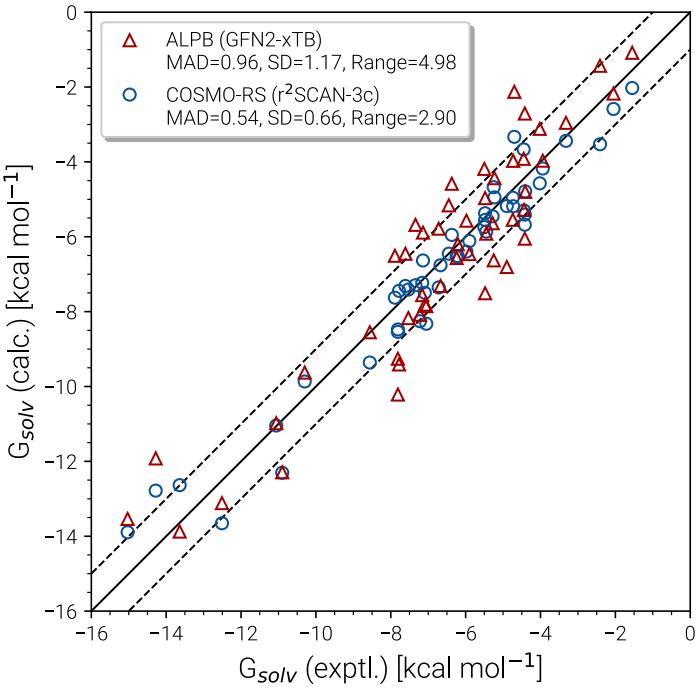


Figure A12.8.: Parity plot showing the correlation of calculated solvation free energies δG_{solv} at the ALPB(GFN2-xTB) and COSMO-RS(r^{2}SCAN-3c) levels with experimental data for chloroform.

FFs of the GFN-family are employed initially. Chemical properties like NMR, IR, pK_a/LogK_{ow} require an accurate description of geometries and free energies. However, the approximations inherent to the SQM/FF methods lead to a non-parallelism of the PES in comparison with high-level DFT. Therefore the SE generated at a lower-level PES by CREST has to be refined at the DFT level, both regarding structure and even more important, (free) energy.

The Command-line ENergetic SOrting (CENSO) algorithm is designed to efficiently refine an input SE from SQM/FF calculations with CREST at a DFT level (see Fig. A12.9) providing a DFT optimized SE, high-level free energies/populations. Calculation of optical rotation (OR) or nuclear magnetic resonance (NMR) spectra can be conducted optionally. Cheaper workflows omit the DFT geometry optimization part and just rerank with DFT. CENSO is divided into several parts to target individual sources of errors and allow for filtering steps at different levels of theory. The algorithmic complexity also results from our aim to treat large ensembles (hundreds of entries) efficiently with rather small computational resources.

The multilevel filtering protocol depends on the use of different thresholds/energy windows applied in each part. Since the final populations are determined by Gibbs free energies G, the sorting scheme has to rely on these values. However, not all contributions to G are easily available in every filtering step, and to make this evident the following naming convention is applied: thresholds with a lower letter g_{thr} represent incomplete free energies, e.g., missing thermostatistical contributions whereas a capital G_{thr} indicates complete free energies (see Eq. A12.1).

In the first very fast prescreening part (*Part0*) the electronic energy is improved compared to the input SQM value by performing dispersion corrected GGA/small-basis set calculations at the B97-D3(0)/def2-SV(P)+gCP level. Definitely high-lying conformers can be discarded after this inexpensive single-point calculation. Solvation effects in *Part0* are treated by ALPB(GFN2-xTB). The ensemble is ranked and sorted by the free energy window threshold $g_{thr,1}$.

The following prescreening part (*Part1*) improves the description by further refining the electronic energy using the new r^{2}SCAN-3c composite method (see Sec. A12.2.2) and by a more accurate account of solvation effects with COSMO-RS (see Sec. A12.2.6). The ensemble is ranked and high-lying entries are removed (threshold $g_{thr,1}$). To obtain complete free energies, G_{trv} contributions must be calculated as well. The computation of a large number of GFN2-xTB/ALPB single-point Hessians (SPH) for G_{mRRHO} (see Sec. A12.2.4) can become computationally costly and therefore the thermostatistical contributions are calculated after the first filtering step in *Part1*. The final sorting in *Part1* is performed applying the (complete) free energy threshold $G_{thr,1}$. This value can be increased as a function of the standard deviation of G_{mRRHO} in the SE indicating high structural diversity and possibly larger errors (so-called fuzzy sorting). At this point, a complete, thermally averaged free energy but for a non-DFT-geometry-optimized ensemble is available which can already be used for thermochemistry (termed 'LEVEL1').

These initial steps are conducted mainly to avoid the costly DFT geometry optimizations of irrelevant conformers in the optimization part (*Part2*) for which r^{2}SCAN-3c/DCOSMO-RS is employed. For this step, a new ensemble optimization algorithm has been implemented where all conformers are batch-wise optimized with a fixed maximum number of eight optimization steps in each optimization meta-cycle. After a meta-cycle is finished, several checks are performed to reduce the number of conformers entering the next iteration. First, a Spearman rank coefficient for the conformer energies is calculated to determine if the PES for all conformers is parallel and if so, the energy window $G_{thr,2}$ within which all conformers are fully optimized is lowered. Second, conformers outside of this window are discarded from the optimization if their gradient norm is below a predefined threshold. Normally, the number of conducted optimizations diminishes in the course of the batch procedure mainly because irrelevant entries can be sorted out more strictly. After all remaining conformers are fully optimized, the free energies are evaluated using r^{2}SCAN-3c/COSMO-RS and GFN2-xTB/ALPB for G_{mRRHO} as in *Part1*. This fully-DFT-optimized SE with its Boltzmann averaged free energy is normally used for further processing (e.g. the calculation of pK_a or LogK_{ow}).

The overall good performance of r^{2}SCAN-3c makes a further refinement unnecessary in many cases. Nevertheless, for systems showing large self-interaction error, a third refinement part (*Part3*) is available where the electronic, thermostatistical (frequency) or solvation (e.g., COSMO-RS-fine) components can be evaluated at a higher (hybrid)DFT or wave function theory level. Furthermore, CENSO can be used to calculate properties for NMR spectra simulation (*Part4*, shielding and coupling constants) or optical rotatory (OR) dispersion (*Part5*).

The CENSO algorithm can be used for both the "thermo-mode" (for identifying the lowest-lying main conformers by using *Part0-2*) and "property-mode" (full high-quality SE using *Part0-2* and possibly *Part4* or *Part5*) by adjusting the most relevant sorting thresholds: $G_{thr,1}$ and $G_{thr,2}$ to either filter out

conformers fast, or the conservative approach with large energy windows for a more complete SE. The number and quality of the CREST runs in the CREST_COMBI script has to be adjusted accordingly, however. The CENSO code is written in PYTHON3 and currently requires (in addition to xTB), either TURBOMOLE or ORCA as QM packages to be installed.

In summary, the CENSO algorithm enables a fast filtering/evaluation of an input SE by identifying efficiently on-the-fly the significantly populated structures with state-of-the-art QC methods for calculation of free energies. This is accomplished by very good r^2SCAN-3c electronic energies, robust and reasonably accurate GFN2-xTB/ALPB/SPH thermostatistical contributions and accurate solvation free energies from COSMO-RS. This can optionally be augmented by the conformational entropy term (S_{conf}) from a separate CREST `-entropy` run.

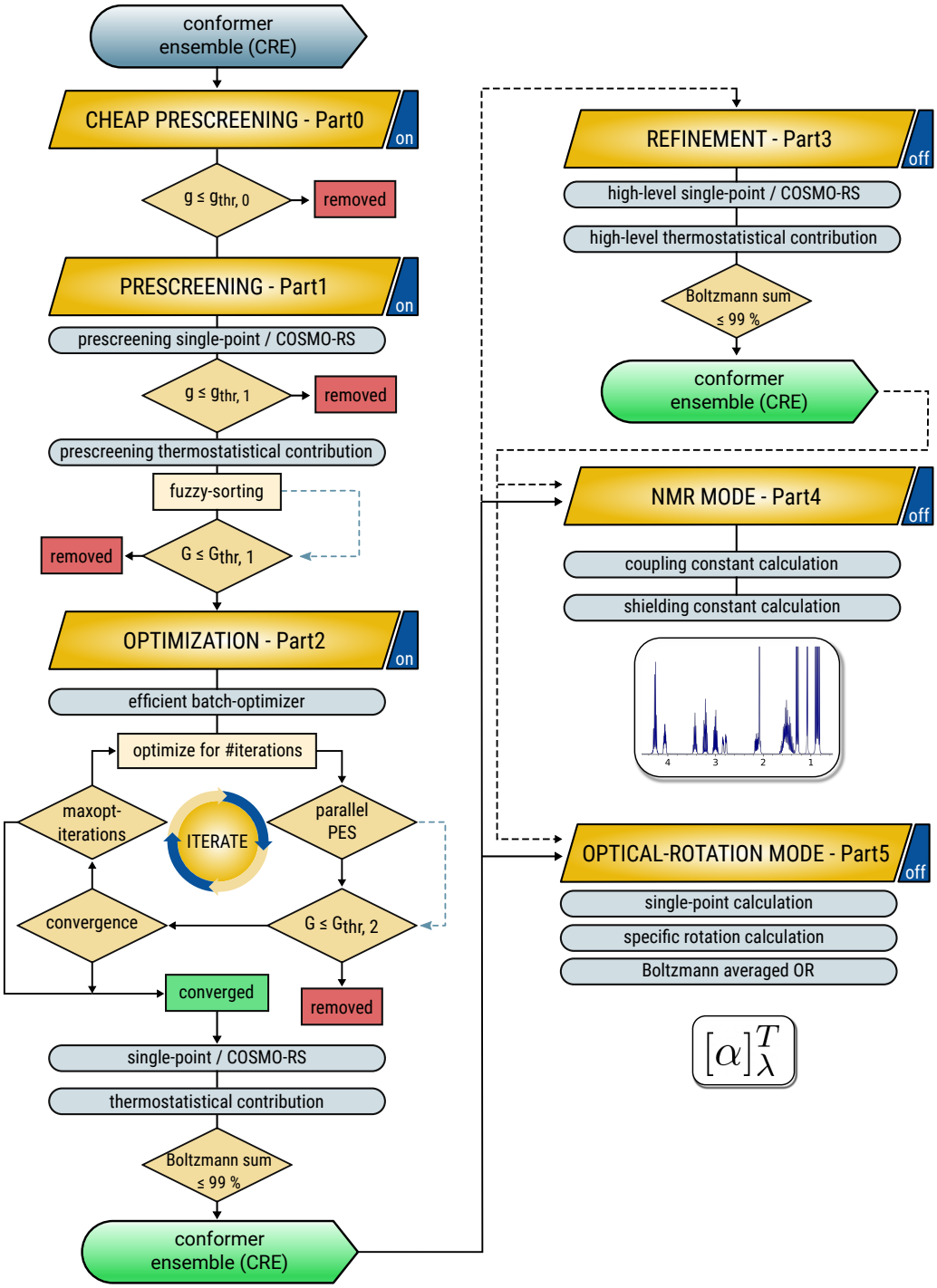


Figure A12.9.: CENSO workflow for structural and free energy refinement of a SE usually taken from a CRES_COMBI run. The prescreening parts (*part0* and *Part1*) re-rank and discard high-lying conformers. The remaining species are efficiently optimized (*Part2)* in a batch-by-batch approach, enabling filtering steps during geometry optimization. Free energies and populations are calculated on DFT optimized geometries and can be used for further processing (e.g. NMR, OR, pK_a, $LogK_{ow}$). Thresholds with a lower letter g represent incomplete free energies missing thermostatistical contributions and capital G free energies with solvation and thermostatistical terms. Parts denoted by on/off can be computed optionally and restartably.

A12.3. Results and Discussion

In the following examples we discuss three XTB/CREST/CENSO approaches of increasing complexity in order to illustrate the effects of the quality of a SE for OR (stereochemical assignment), partition coefficients ($LogK_{ow}$) and pK_a values. They always start from same guessed (PubChem[551]) Cartesian coordinate input structure treated in the following way:

1. GFN2-xTB/ALPB single-structure optimization and single-point r^2SCAN-3c/COSMO-RS calculation without any conformational search (denoted as LEVEL0)

2. Quick CREST_COMBI search at GFN-FF/GFN2-xTB/ALPB level followed by r^2SCAN-3c/COSMO-RS single-point ensemble reranking but no DFT optimization (LEVEL1)

3. Extended CREST_COMBI search including GFN2-xTB parts with GESC, full r^2SCAN-3c/DCOSMO-RS optimization, and final r^2SCAN-3c/COSMO-RS ensemble reranking (LEVEL2)

The computation time roughly increases by 1 order of magnitude for each higher level. No further data manipulation by the user is required and at the end of each run, the final SE and target property are printed out. The procedures are executed by a script with a single Unix command and level/mode arguments (e.g., `crenso -or methanol -l1` or `crenso -kow -l2`). Essentially, `crenso` calls XTB/CREST/CENSO with the appropriate settings.

The protocol was developed targeting organic drug-like molecules with a system size of 50-150 atoms. However, the search level and structural details (flexibility) of the molecule strongly determines the actual size-limit of the methodology. When lower search levels (LEVEL0 and LEVEL1) are applied, larger molecules can be investigated and for the highest search level (LEVEL2) the dominating computational cost stems from the number of DFT geometry optimizations, which is difficult to estimate in advance. For these reasons, it is difficult to give a good guess of the limits of the protocol. Furthermore, it is noted, that the predefined sensible thresholds and energy windows were tested mainly for main-group or organic compounds. It is generally advised to increase them if electronically complicated and/or highly flexible molecules are studied (see Ref. 75 for a recent study of transition metal complex conformations).

The workflow is exemplified for a few typical molecules shown in Fig. A12.10.

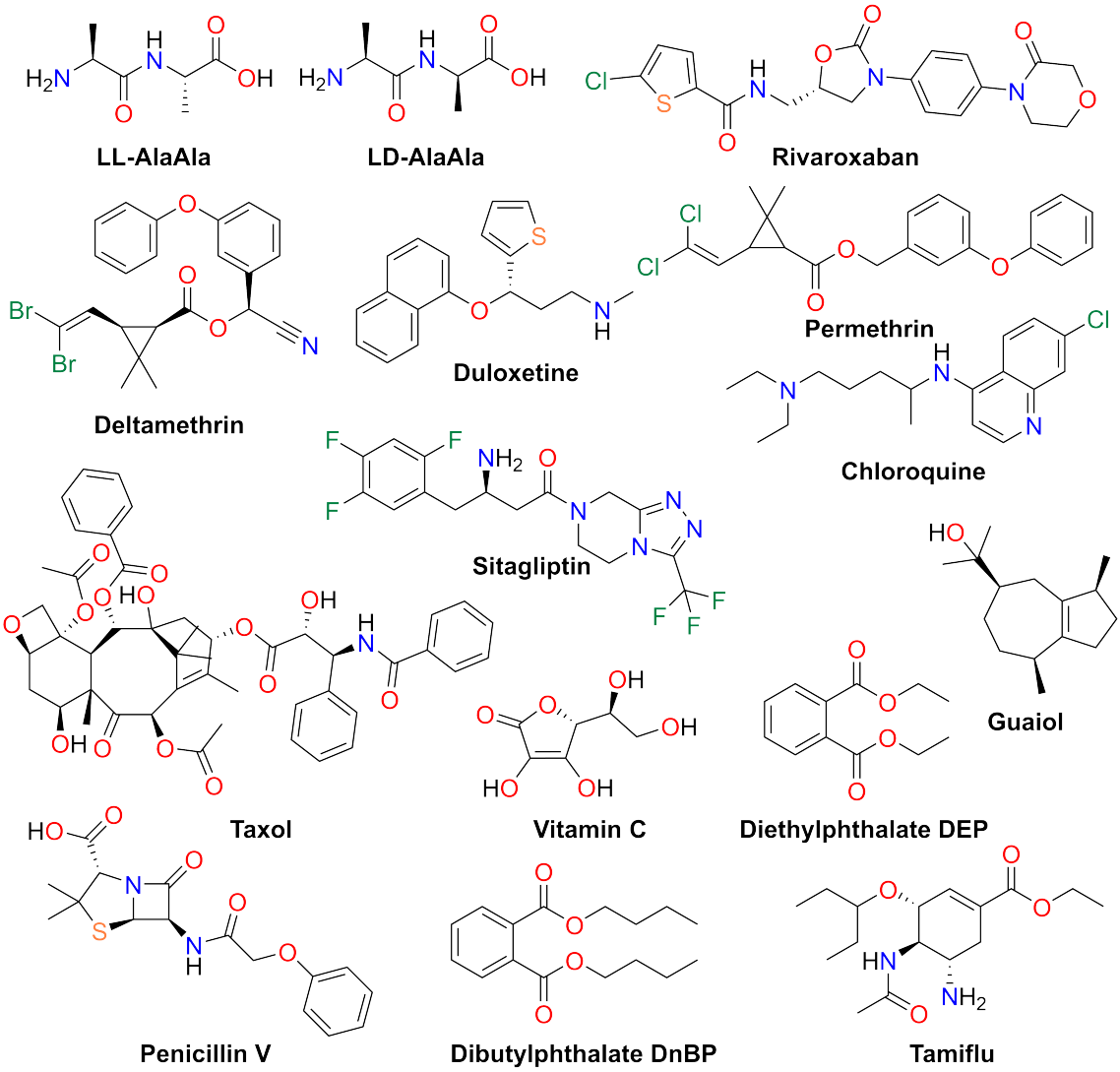


Figure A12.10.: Lewis formulas of the example molecules.

A12.3.1. Conformational ensembles and optical rotation

Previous studies[552–555] underline the importance of considering a conformer ensemble for the reasonable theoretical simulation of chiroptical properties like optical rotation (OR). An established workflow in this regard is the computation of the property for a few conformers and averaging the values weighted by the respective Boltzmann population. While this is rather successful for small nonrigid systems with typically $< 20 - 30$ atoms, such a procedure has not been reported for large drugs where an accurate routine treatment would represent a breakthrough for stereochemical or absolute configuration assignment. The main problem is that OR is extremely sensitive to details of the conformation. Changes of OR sign and value by up to 2 orders of magnitude is not uncommon and hence even obtaining a qualitatively correct OR value requires a SE of very high quality. Furthermore, the competing intra- and intermolecular (solute-solvent) interactions for typical drugs in polar solvents are difficult to

describe.

The Tab. A12.1 shows standard TDDFT computed OR values for four typical drug molecules employing our three workflows with increasing degree of sophistication. It should be noted that for such large and complicated structures, a correspondence in sign and order of magnitude between theory and experiment is an essential achievement.

Table A12.1.: Comparison of calculated and experimental specific optical rotations for six natural product, pesticide and drug molecules.

treatment[a]	specific rotation (in $^\circ cm^2 dmol^{-1}$)					
	Guaiol	Tamiflu	Rivaroxaban	Duloxetin	Deltamethrin	Sitagliptin
LEVEL0	−47.2	−175.9	−83.0	−32.3	−4.2	69.5
LEVEL1	−36.2	−123.8	−105.7	225.3	28.6	45.9
LEVEL2	−30.8	−89.7	−70.6	252.0	52.7	−27.7
exptl.	−35[b]	−(26–36)[c]	−41[d]	213[e]	25[f]	−42[g]

[a] OR values are computed with TDDFT PBE/def2-SVPD/COSMO in the velocity formalism at the sodium-D line wavelength using corresponding geometries of LEVEL1-3 and r^{2}SCAN/def2-SVPD orbitals. [b] in EtOH, Ref. 556 [c] in water, Ref. 556 [d] in MeOH, Ref. 557 [e] in MeOH, Ref. 558 [f] in Toluene, Ref. 559 [g] in water, Ref. 558

For three cases (Guaiol, Tamiflu, Rivaroxaban), the sign of the OR (and therefore the assignment of absolute configuration) is correct at LEVEL0 although the value is massively underestimated for Tamiflu. The OR values computed with proper SEs (LEVEL1/2) generally agree much better to those from experiment and for Duloxetin, Deltamethrin, and Sitagliptin only the higher level calculation yields the correct sign. Unfortunately, the required level can not be predicted in advance by considering the chemical structure. In the case of Sitagliptin, the composition of the SE varies significantly and at the less deep search LEVEL1 an incorrect (dominating) lowest free energy structure is found. How difficult the problem really is can for example be seen from the ten most contributing conformers for Deltamethrin with roughly 31, 12, 7, 7, 6, 6, 5, 5, 5, 5% population and widely varying OR values (in $^\circ cm^2 dmol^{-1}$) of 17.3, -123.9, -108.8, +295.6, +111.1, +83,2, +291.3, +91.8, +13.2, +269.9. Note again that in such difficult cases due to the stochastic nature of the applied algorithms the thermally averaged OR may vary widely requiring additional averaging over several independent runs. In any event, our workflow appears to represent a significant and robust step towards a routinely applicable stereochemical-conformational analysis tool for complex drugs, natural products, and related compounds.

A12.3.2. Partition Coefficients

After a chemical has been released into the environment, its molecular interactions determine the distribution and transport between air, water and soil. The exposure of organisms to chemical substances and the bioaccumulation in these organisms is influenced by the distribution between compartments and subcompartments. The octanol/water partition coefficient K_{ow} (also termed LOGP) is commonly

used to determine whether a molecule accumulates preferably in the hydrophilic or lipophilic phase, which is an indicator for its tendency to bioaccumulate.[560] In general, the logarithmic partition coefficient K_{12} of a molecule between phase 1 (here O: wet 1-octanol) and a phase 2 (W: water) is given by

$$\log K_{12} = \frac{\Delta G_{12} = \overline{G}_2 - \overline{G}_1}{\ln 10 RT} \tag{A12.11}$$

where $\overline{G}_n$ denotes the ensemble free energy of the species in phase n as given by Eq. A12.3. In our approach, we are not approximating the free energy of phase transfer ΔG_{12} by the difference of solvation free energies $\Delta\delta G_n^{solv}$ as usually done in many approximate schemes. Instead, all physical contributions to the ensemble free energy (except S_{conf}) are considered. The SEs for the two phases 1/2 are generated fully independently by separate CREST/CENSO runs. For a recent application of COSMO-RS to the LogK_{ow} problem see Refs. 561,562 and in the context of the SAMPL6 blind challenge see Ref. 549 (and other papers in the corresponding journal issue). The results obtained with different levels of our workflow for some example molecules are given in Tab. A12.2.

Table A12.2.: Comparison of calculated and experimental[a,b] LogK_{ow} values of seven example molecules.

treatment	DEP	DnBP	Tamiflu	Chloroquin	Taxol	Permethrin	Sitagliptin
LEVEL0	2.81	4.71	1.42	5.40	5.46	4.85	1.82
LEVEL1	2.41	4.80	1.22	5.55	7.42	6.83	1.67
LEVEL2	2.49	4.55	1.12	5.70	4.73	6.41	0.93
exptl.	2.45(30)[a]	4.44(37)[a]	1.1[b]	4.6[b]	4(2)[c]	6.5[d]	0.0[e]

[a] Ref. 563 [b] QSAR (XLogP3) estimate from Ref. 551 [c] Average of four QSAR values (ALOGPS, XLogP3, ChemAxon, ACD), see Ref. 551. [d] Ref. 564 [e] Ref. 565

As can be seen from the data in Tab. A12.2, the agreement between experimental (or QSAR estimates) and our QM computed values is generally very good. The influence of the ensemble, i.e., when going from LEVEL0 to LEVEL1, is moderate for four cases but very large for two (Taxol and Permethrin) and hence it should be taken into account if possible. In fact, for the large and relatively complicated drug Taxol, even LEVEL1 is insufficient and full DFT optimizations have to be performed in order to get a reasonable LogK_{ow} value.

A12.3.3. pK_a values

The pK_a value is directly related to the dissociation free energy ΔG_{diss} for the loss of a proton of species A in water according to

$$A^n H \rightarrow A^{n-1} + H^+ \tag{A12.12}$$

reducing the molecular charge from n to $n-1$. The ΔG_{diss} value for this process involves the free energy of the solvated proton which is difficult to compute and hence this value is normally empirically

accounted for by so-called linear free energy relationships (LFER)[566]. Here we take a more advanced version of this concept where the computed ΔG_{diss} for the above process, but with the solvated proton free energy set to zero ($\Delta G'_{diss}$), is inserted into a cubic polynomial of the form

$$pK_a = c_1 + c_2 \log k_{diss} + c_3 (\log k_{diss})^2 + c_4 (\log k_{diss})^3 \,. \tag{A12.13}$$

$c_1 = -1656.7$, $c_2 = 23.185$, $c_3 = -0.11103$, and $c_4 = 0.0001835$ are empirically determined constants (by a fit to about 20 experimental values for rigid small and medium sized organic molecules in the pK_a range from about −10 to +50) and $\log k_{diss} = \frac{\Delta G'_{diss}}{ln10RT}$. The two free energies of A^nH and A^{n-1} in water are obtained exactly as described for the partition coefficients without considering more than a single acidic site (microscopic vs. macroscopic pK_a). Taking more than a single acidic site into account is easily possible but not required for the shown examples.

Table A12.3.: Comparison of calculated and experimental pK_a values of four example molecules.

treatment	pK_a			
	Vitamine C	Penicilin V	LL-AlaAla	LD-AlaAla
LEVEL0	3.39	3.20	4.20	3.50
LEVEL1	4.29	3.33	4.15	3.16
LEVEL2	3.92	2.25	3.55	3.10
exptl.	4.17[a]	2.7[a]	3.30[b]	3.12[b]

[a] Ref. 551 [b] Ref. 567

As can be seen from Tab. A12.3, the agreement between experimental and theoretical data is excellent and similar to the $LogK_{ow}$ values where only simpler neutral cases are involved. Because δG_{solv} is roughly 1 order of magnitude larger for an ion than for a neutral compound, larger errors can be expected here but are actually not observed. Note that even subtle pK_a differences between diastereomers of AlaAla are correctly obtained by our approach. In this case, a substantial improvement is observed when going from LEVEL0/1 to LEVEL2 which is also true for the absolute pK_a value of Penicillin V. Substantial conformational effects even show up in relatively small compounds like vitamin C, where the largest change is found from LEVEL0 to LEVEL1.

A12.4. Conclusions and Recommendations

In summary, we would like to give the interested reader and possible user of these theoretical methods for the description of nonrigid molecules some recommendations and briefly discuss the key points of our and related workflows:

1. The solvation free energy and intramolecular London dispersion energy decisively depend on the shape of the three dimensional molecular structure (e.g. folded vs. unfolded). Both (competing) effects should already be taken into account in the first SE generation (search and filtering) step

at the lowest level of theory, i.e., dispersion corrections and continuum solvation models should be applied. The SE energy window should be large to enough to account for the typical errors of SQM and FF methods for conformational energies (2-3 kcal mol^{-1}). Because the energy range of interest at ambient temperature is also about 2 kcal mol^{-1}, recommended conservative conformational energy windows are 6-8 kcal mol^{-1}. Corresponding values for protomer or tautomer searches are generally larger (10-20 kcal mol^{-1}) and more strongly depend on the system.

2. Dispersion corrected DFT methods with continuum solvation are sufficiently accurate for the subsequent filtering step(s) and the final geometry optimization. Single-point calculations with higher-level methods (hybrid functionals with larger AO basis sets or coupled-cluster theory) on these structures yield more accurate final gas phase energies if necessary. The new r^2SCAN-3c composite DFT method seems to provide a perfect compromise between accuracy and computational effort for systems with up to 100-200 atoms.

3. The efficient COSMO-RS solvation model seems to be the most accurate currently available and even provides a semiquantitative temperature dependence of the solvation free energy. The remaining errors for relative free energies (partition coefficients or conformational energies) of neutral systems may about equally be attributed to the gas phase energy and the solvation part while for ionic systems (e.g., pK_a values) the error of $\Delta\delta G_{solv}$ dominates. In our opinion, full explicit treatments of solvation for nonrigid molecules as an alternative are computationally much too costly or at a very early stage of development.

4. The modified rigid-rotor-harmonic-oscillator approximation (mRRHO) together with the single-point-Hessian (SPH) method is accurate, robust and computationally feasible at a SQM or FF level even for very large systems. Differential effects of the G_{mRRHO} term can be significant (up to $1-2$ kcal mol^{-1}) and this contribution should be included in intermediate filtering steps when aiming at good accuracy. Energetic reordering in a SE by including this term should be expected whenever its structural/conformational diversity is large.

5. The absolute value of the conformational free energy contribution $G_{conf}(298)$ is a few kcal mol^{-1} for flexible systems (e.g., medium sized alkanes) and can efficiently be computed at the GFN-FF level with CREST. However, its effect on reactions when the flexibility does not change much between reactants and products, may be small compared to other sources of error and can often be neglected but should be investigated further before general recommendations can be made. Preliminary tests on some drug molecules indicate that this term may contribute up to 0.5 units in pK_a or $LogK_{ow}$ computations.

6. A number of quantum chemical codes and software tools are freely available to conduct all the required calculations in the proposed efficient, automatic, multi-level modeling workflow. Nevertheless, the problem is and remains complex as indicated also by the complexity of the computer codes/scripts and the time that was needed for their development. Larger ionic or similar systems with strong (explicit) solvation interactions, and extremely flexible, mainly non-covalently interacting aggregates are the challenges for improvement and generalization of the procedures.

We hope that this perspective helps to understand the intricate problem of non-rigid molecules in the condensed phase more clearly and to consider the most relevant effects routinely for good computational chemistry work.

Acknowledgement

This work was supported by the DFG in the framework of the "Gottfried Wilhelm Leibniz-Preis". The authors thank Sebastian Ehlert, Jakob Seibert, Uwe Hohm, Wolf Palm, and Tunga Salthammer for fruitful discussions, suggestions, and technical support.

A13. Automated Quantum Chemistry-Based Calculation of Optical Rotation for Large Flexible Molecules

Abstract: The calculation of the optical rotation (OR, $[\alpha]_D$) for nonrigid molecules was limited to small systems due to the challenging problem of generating reliable conformer ensembles, calculating accurate Boltzmann populations and the extreme sensitivity of the OR to the molecules three-dimensional structure. Herein, we describe and release the `crenso` workflow for the automated computation of conformer ensembles in solution and corresponding $[\alpha]_D$ values for flexible molecules. A comprehensive set of 28 organic drug molecules (28-144 atoms) with experimentally determined values is used in our assessment. In all cases, the correct OR sign is obtained with an overall mean relative deviation of 72 % (mean absolute deviation of 82 °$[dm(g/cm^3)]^{-1}$ for experimental values in the range −160 to 287 °$[dm(g/cm^3)]^{-1}$). We show that routine $[\alpha]_D$ computations for very flexible, biologically active molecules are both feasible and reproducible in about a day of computation time on a standard workstation computer. Furthermore, we observed that the effect of energetically higher-lying structures in the ensemble on the OR is often averaged out and that in 23 out of 28 cases the correct OR sign is obtained by just considering only the lowest free energy conformer. In four example cases, we show that the approach can also describe the OR of pairs of flexible diastereomers properly. In summary, even very sensitive, multifactorial physicochemical properties appear reliably predictable with minimal user input from efficiently automated quantum chemical methods.

A13.1. Introduction

Molecular chirality is an intriguing concept omnipresent in nature and often an essential factor for the biological activity of molecules. The active sites of mostly chiral proteins have a specific three-dimensional structure, which the binding molecule must complement.[568,569] This results, for example, in a very different olfactorial activity of enantiomers with the prototypical case of limonene,[570] and that only drug molecules with proper stereocenters strongly bind at its target[571].

Optical rotatory (OR) dispersion is among the fundamental properties of chiral molecules.[572,573] Together with electronic and vibrational circular dichroism, it is routinely measured for the determination of the absolute configuration of molecules. Matching the experimentally determined and calculated OR of a chiral molecule enables a definite assignment of absolute (and in case of many stereocenters also relative) configuration even for non-crystallizable compounds. In recent years, this has attracted considerable theoretical interest, mostly related to advances in time-dependent density functional theory

(TDDFT) allowing efficient calculations of chiroptical response from first principles also for relatively large systems.[209,574–578]

A multitude of wave function (WFT) or density functional theory (DFT) based modeling of optical activity to date has been conducted for structurally rigid molecules (see, e.g., Ref. 579 and references therein) for which deviations between theoretical and experimental OR values are typically about 30 %.[580] However, most bioactive substances are conformationally flexible in solution and practically exist as a conformational ensemble rather as a single, structurally well-defined species. A common approach to include these effects is the modeling of chiroptical properties by TDDFT calculations on static (geometry optimized) conformer structures and weighting the values with the respective thermal population (Boltzmann factor).[552,553,581–583] However, already in 2009, Kundrat and Autschbach stressed that "the overall accuracy of such modeling is limited by the accuracy of the Boltzmann factors, which can be difficult to gauge."[584] For compounds in solution, an accurate description of the OR is more complicated compared to systems in the gas phase since both the solvation contribution to the conformer population (explicitly or implicitly calculated, see supporting information) and, in the case of strong solute-solvent interaction observed for small molecules, the influence of the solvent on the OR itself can be difficult to describe.[585,586]

One main reason for the difficulty of reliable, routine simulations of the OR for larger flexible molecules is also an extreme sensitivity of the computed value on the underlying three-dimensional structure. Change of sign and order of magnitude of the computed OR between conformers is common,[209,554,584] but it is totally unclear to date how this problem develops with larger systems. On the other hand, a good match of experimental OR and corresponding theoretical ensemble averaged values would enable an indirect conclusion regarding the "effective" structure under the measurement conditions.

A13.2. Results and discussion

The determination of highly populated conformer structures in solution is already a challenging task, mainly due to the lack of robust automatic computational schemes. This problem has recently been addressed with the Conformer-Rotamer Ensemble generation Tool (`crest`[68]), which is based on robust and reasonably accurate semiempirical quantum chemistry or force-field potentials[61] coupled with a meta-dynamics search algorithm[419] and a DFT-based multilevel energetic sorting procedure (`censo`[69]).

Here, we develop and comprehensively test for the first time this approach for the automatic computation of the OR for typical drug molecules in solution. For related recent applications of `crest` see Refs. 75,209,516. In addition to a resulting stereochemical assignment, detailed insight into the conformational behavior of biologically active substances can be obtained from such a procedure. A new test set of typical drug molecules was composed to evaluate the proposed computational protocol. The set contains 28 molecules with system size between 28 and 144 atoms and is termed NROR28 from now on ("NR" for nonrigid). Important properties of the respective molecules and their Lewis structures are summarized in Table A13.2, Figure A13.1 and in SI-Table A14.1. The NROR28 set was assembled with the motivation to provide sufficient diversity of functional groups, solvents, molecular charge, and conformational flexibility. According to the internal flexibility measure in `crest`[68], the

most rigid molecule in this set is oxycodone, while the most flexible is pregabalin (see SI-Table A14.2). Originally, the test set contained one more system (penicillin V). However, it was not possible in this case to achieve qualitative agreement between calculated and experimental OR, even with many tested adjustments to various technical parameters in the calculations. Since in this case incorrect experimental data or an unclear protonation state under the experimental conditions cannot be excluded, penicilin V was not considered in the final set. Due to different sources of the experimental reference data and partly unknown measurement conditions, the discussed specific rotations at the sodium-D wavelength ($[\alpha]_D$) do not always have full benchmark quality (estimated uncertainty of 20-30%) but are nevertheless sufficient for the current purpose. For the molecules where a range of measured OR values is available, the arithmetic mean of this range was taken as reference value. In cases where the experimentally used solvent was not available in the chosen theoretical implicit solvation model, a chemically and physically similar solvent was chosen (see SI-Table A14.1). In the following, all $[\alpha]_D$ values are given in $^\circ[dm(g/cm^3)]^{-1}$ units, are not corrected for the Lorentz solvent factor as suggested in Refs. 207,580, and refer to $T = 298$ K. Protonation states were determined based on experimental conditions and known pKa values.

Molecule[a]	**# atoms**[b]	$[\alpha]_D$, **exptl.**[c]	$[\alpha]_D$, **ref.**[d]	**solvent**[e]	**species**[f]	**Pubchem ID**[g]	**Reference**[h]
amoxicillin	44	248 to 268	258.0	n/a	Zwitterion	33613	[587]
ampicillin	43	287	287.0	H_2O	Zwitterion	6249	[588]
blasticidin S	56	108.4	108.4	H_2O	Zwitterion	170012	[589]
clindamycin	60	214	214	$CHCl_3$		446598	[589]
deltamethrin	47	57.4	57.4	Toluene		40585	[590]
dexibuprofen	33	58 to 60	59.0	EtOH		39912	[591]
duloxetine	40	116 to 125	120.5	MeOH		60835	[591]
ergotamine	78	−160	−160.0	$CHCl_3$		8223	[589]
erythromycin	118	−78	−78.0	EtOH		12560	[589]
esomeprazole	43	30 to 36	33.0	H_2O	Anion	9568614	[591]
ezetimibe	51	−33.9	−33.9	MeOH		150311	[589]
guaiol	42	−12 to −3	−7.5	n/a		227829	[589]
ibrutinib	57	−128	−128.0	$CHCl_3$		24821094	[592]
linezolid	44	−9	−9	$CHCl_3$		441401	[589]
methylergonovine	50	−45	−45.0	n/a		8226	[589]
oxycodone	44	−125	−125.0	H_2O		5284603	[589]
paroxetine	44	−87	−87.0	EtOH		43815	[589]
penicillin	41	269	269.0	MeOH	Anion	2349	[589]
pregabalin	28	10	10.0	H_2O	Zwitterion	5486971	[589]
rapamycin	144	−58.2	−58.2	MeOH		5284616	[589]
ritonavir	98	7 to 10	8.5	MeOH		392622	[591]
rivaroxaban	47	−41	−41.0	DMSO		9875401	[591]
rosuvastatin	61	14.8	14.8	MeOH/H_2O		446157	[589]
sitagliptin	43	−18 to −23	−20.5	H_2O	Cation	4369359	[591]
sofosbuvir	65	40 to 45	42.5	EtOH		45375806	[591]
tamiflu	50	−26 to −36	−31.0	H_2O		65028	[589]
taxol A	113	−49	−49.0	MeOH		36314	[589]
tenofovir	33	−26 to −20	−23.0	HCl(aq)	Cation	464205	[591]

Table A13.1.: Experimental information for the entries of the NROR28 set. [a] Name; [b] Number of atoms; [c] Experimental specific rotation ($[\alpha]_D$, **exptl.**); [d] Unified specific rotation (used as reference, $[\alpha]_D$, ref.); [e] Solvent under measurement conditions; [f] Information on the protonation state; [g] Pubchem ID of the corresponding compound; [h] Reference for the respective entries.

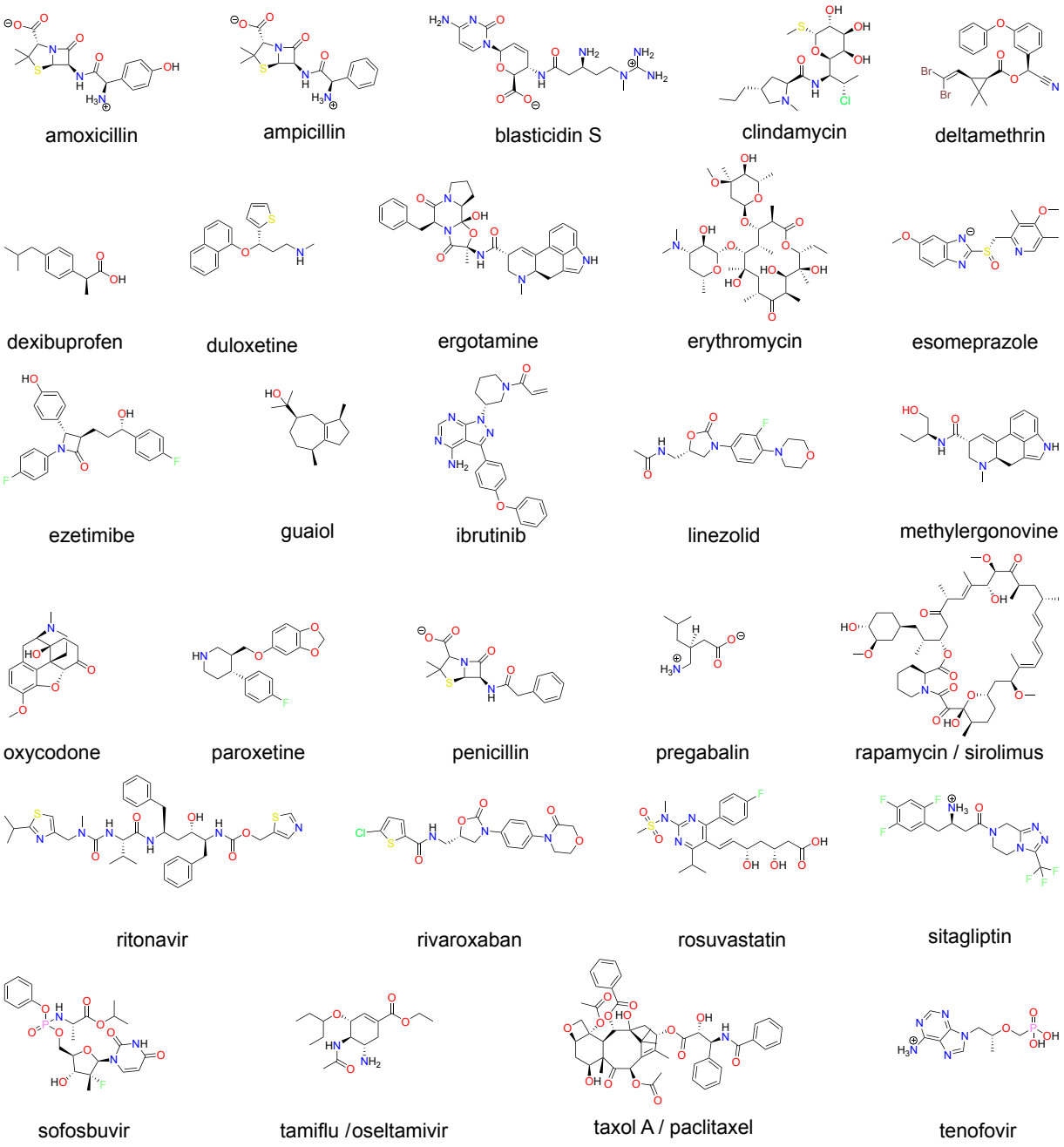


Figure A13.1.: Lewis structures of the investigated molecules in the NROR28 benchmark set.

The calculation of the ensemble averaged OR has been implemented in the `crenso` "push-the-button" type script, and the general workflow is outlined schematically in Figure A13.2. Conceptual and technical aspects of the `crenso` workflow are described in the following paragraph, and a section with a summary of the programs and terminologies used is included in the supporting information. Starting from an initial three-dimensional input geometry, a structure ensemble (SE) consisting of all energetically low-lying conformers is created using the `crest` meta-dynamics based algorithm[419]. A

merged SE from searches on various potential energy surfaces (PES) is used to enrich the diversity of the SE and to account for deficiencies of the underlying "low-cost" methods which are semiempirical quantum mechanical (SQM) and force-field methods from the GFN*n*-xTB family[61,167,181]. Importantly, all calculation steps consistently apply an implicit continuum solvation model. All generated structures are fully geometry optimized on a uniform level of theory (GFN2-xTB/ALPB[69]) and sorted into conformers, rotamers and (discarded) identical structures. For computational efficiency, the ensemble is clustered, using a combination of principle component analysis (PCA) and a k-Means clustering algorithm[185], further reducing the number of candidate conformers. The entire SE generation has been implemented in the `crest_combi` script which simply calls `crest` with different technical settings. The subsequent SE postprocessing at a composite DFT level is realized by the command-line energetic sorting algorithm `censo`. It uses a threshold based multilevel filtering protocol to progressively improve the description of the electronic energy, the solvation contribution, and the thermostatistical contribution to the free energy. Each applied filtering step removes high-lying, irrelevant conformers and reduces the number of computationally demanding DFT geometry optimizations. In the final step, free energies and Boltzmann weights at a chosen temperature are calculated and TDDFT based OR computations are conducted for the significantly populated conformers from which the SE averaged OR is obtained. This value should be compared to the experiment, although for analysis purposes, we here additionally provide a value computed for the structure of the lowest free energy and for ones randomly chosen from the initial SE in order to investigate the sensitivity of the results.

The entire workflow (see Figure A13.2) with sensible predefined technical settings of various thresholds and energy window values (see SI-Table A14.3) was assembled in the `crenso`[593] script. It is applied as a simple unix command with level arguments (e.g., `crenso -or_fine methanol -l3`) and essentially represents a highly improved, much more efficient and reliable further development of our procedure described previously for the automatic computation of NMR spectra[235]. The evaluation of the NROR28 set has shown that the use of the geometric energy surface correction (GESC) introduced in Ref. 69 does not systematically improve the computed OR values at increased computational cost, and hence it is not applied by default anymore and the `crenso` script has been updated accordingly. The total computation time for a typical `crenso` run is composed of about 25 % for `crest_combi` (SE generation), 5 % for initial DFT screening, 65 % for DFT geometry optimization, and 5 % for the OR computation. For information on calculation settings, see the Computational Methods section.

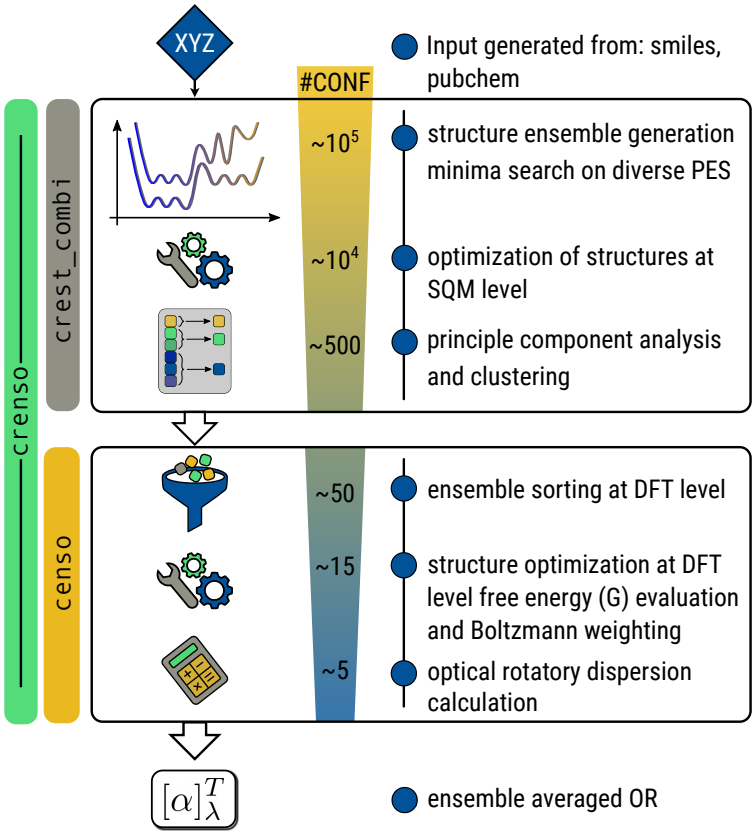


Figure A13.2.: `crenso` workflow for the automated calculation of the OR for flexible molecules. The structure ensemble (SE) is created by meta-dynamics runs, snapshot optimization on several PES and clustering using principle component analysis (realized by `crest_combi`). In the course of the multilevel filtering procedure, the number of candidate conformers is reduced by several orders of magnitude. The large number of structures to be considered initially (#CONF) makes the use of the efficient GFN methods mandatory in the first steps of the workflow. The SE is subsequently refined and efficiently optimized at DFT level. The $[\alpha]_D$ value is calculated for all significantly populated structures (i.e., until the sum of populations equals 99 %) and thermodynamically averaged (performed by `censo`).

Before evaluating the performance of the newly established workflow for large nonrigid molecules, its applicability to a smaller, well-known but still challenging and highly non-trivial case is documented by the textbook example of D-Glucose or D-glucopyranose in water shown in Figure A13.3. This system has been considered theoretically by TDDFT and continuum solvation models in a more manual procedure employing predetermined gas-phase conformers by Mennucci et al.[594].

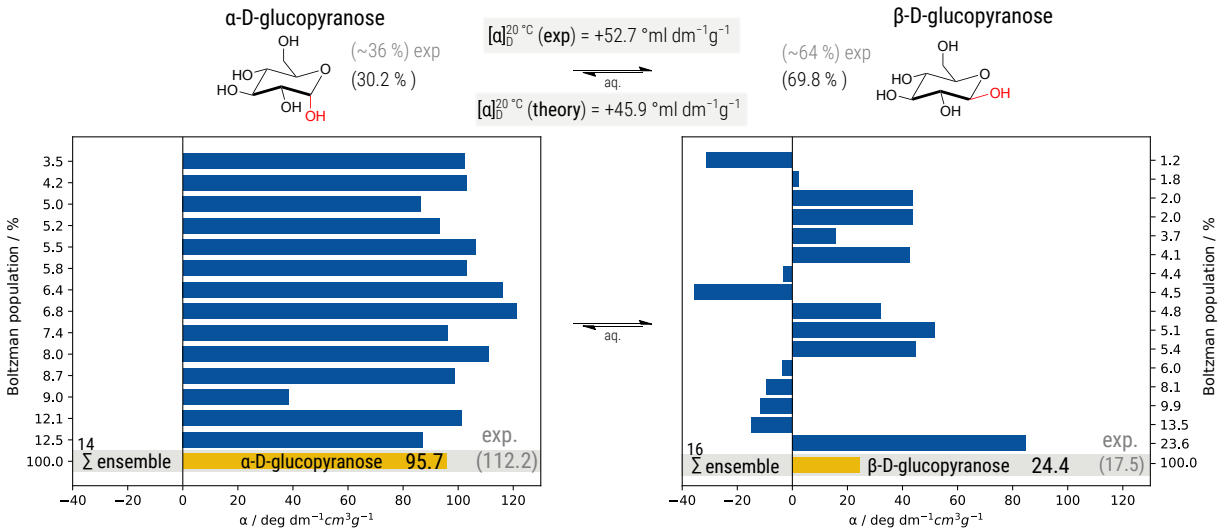


Figure A13.3.: Comparison of calculated (shown is only a single `crenso` run of the three performed) and experimental[595] $[\alpha]_D$ for the α- and β-form of D-glucopyranose in water. The average over three independent `crenso` runs is with 51.7 °$[dm(g/cm^3)]^{-1}$, in excellent agreement with the experiment. For clarity the results for only one of the three runs is shown. Values for the pure forms (in parentheses) as well as populations of the conformers and corresponding OR values are given.

Mutarotation leads to an equilibrium of the α- and β-form which are treated independently in the calculation. Figure A13.3 shows that the conformers of the α-form more uniformly have large positive OR contributions, while the more abundant β-form conformations exhibit strongly varying positive and negative values. The experimental and calculated values for the two pure forms as well as for the equilibrium in water agree very very well, especially when the error in δG_{solv} that is, the solvation contribution to free energy, is considered, which can be larger for systems containing hydrogen bonds. By averaging over three independent runs the effect of the nondeterministic ensemble generation steps is reduced and a technically averaged OR of 51.7 °$[dm(g/cm^3)]^{-1}$ for the equilibrated α/β-mixture is obtained, which is almost spot on with the experimental value[595]. This impressively demonstrates that the competing effects between intra- and intermolecular hydrogen bonding for the various conformers is properly accounted for by the composite DFT treatment with COSMO-RS solvation. The very good result for this difficult case demonstrates that the applied implicit standard solvent models seem to be sufficient for the description of both the conformational free energy and the OR. This also supports the important conclusion that in the context of a routine workflow, the computationally unfeasible explicit solvation treatment with its associated long computation times is not required and that implicit models can be applied. For a brief overview on implicit and explicit solvation, see the supporting information. Seemingly, our automated procedure can identify with high confidence the most populated structures even when none is dominating which is some kind of worst case scenario that may be also common for the larger drug systems discussed below.

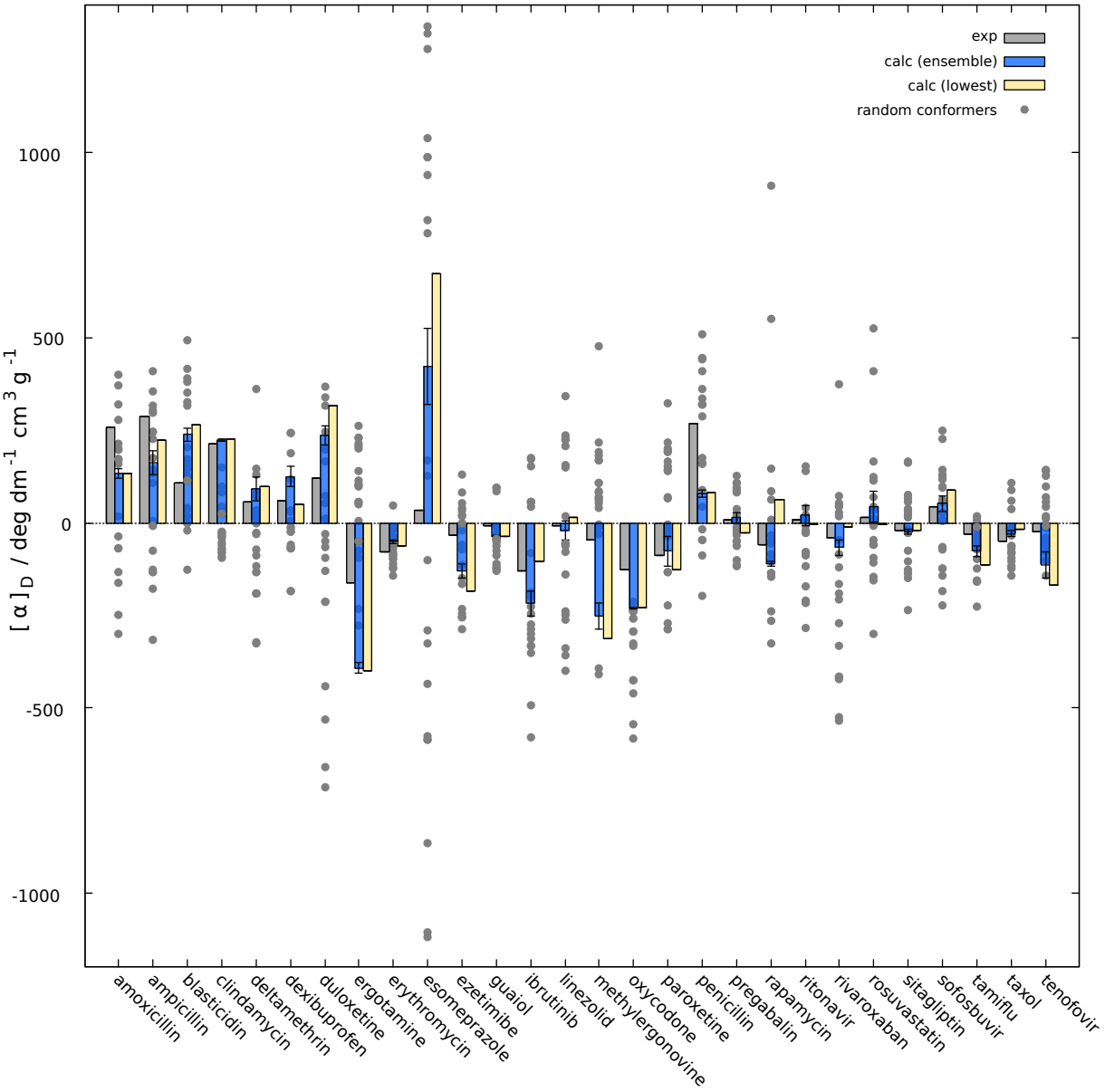


Figure A13.4.: Ordinate shows the $[\alpha]_D$ values of the 28 test molecules: from the experiment (gray), calculated from the weighted ensemble (blue), calculated from the lowest conformer (yellow), and calculated for 20 conformers randomly selected from a `crest` conformer ensemble (gray dots). The error bars indicate the estimated error resulting from the DFT free energy computation.

Molecule[a]	$[\alpha]_D$, **ref.**[b]	$[\alpha]_D$, **SE**[c]	**# confs.** [d]	**pop. [%]**[e]
amoxicillin	258	134.5	3	58.8
ampicillin	287	162.1	2	80.0
blasticidin	108.4	238.4	2	75.0
clindamycin	214	224.1	2	76.2
deltamethrin	57.4	92.6	9	40.8
dexibuprofen	59	125.0	7	31.9
duloxetine	120.5	237.4	8	36.0
ergotamine	-160	-392.1	4	35.5
erythromycin	-78	-50.3	3	63.7
esomeprazole	33	423.5	5	50.4
ezetimibe	-33.9	-129.8	28	20.5
guaiol	-7.5	-35.8	10	24.9
ibrutinib	-128	-216.8	44	8.8
linezolid	-9	-20.7	4	65.7
methylergonovine	-45	-251.3	10	31.4
oxycodone	-125	-229.7	2	97.6
paroxetine	-87	-75.7	3	79.1
penicillin	269	79.5	8	37.3
pregabalin	10	14.0	6	29.8
rapamycin	-58.2	-111.1	1	94.5
ritonavir	8.5	20.5	5	53.6
rivaroxaban	-41	-65.7	9	27.6
rosuvastatin	14.8	45.1	22	17.9
sitagliptin	-20.5	-22.7	4	77.2
sofosbuvir	42.5	52.6	3	78.7
tamiflu	-31	-76.1	6	29.8
taxol A	-49	-28.8	5	60.0
tenofovir	-23	-113.3	3	72.6

Table A13.2.: Comparison of the Reference and Calculated OR for the NROR28 set, (Computed Values are Averaged over Three Independent `crenso` runs). [a] Name; [b] Experimental reference specific rotation ($[\alpha]_D$, **ref.**) [c] Calclated SE averaged OR ($[\alpha]_D$, **SE**); [d] Number of conformers in the final SE; [e] Population of the lowest free energy conformer.

Figure A13.4 compares the experimental and calculated OR values for all NROR28 entries. Because the proposed automatic workflow contains nondeterministic parts as noted above, the reproducibility was evaluated in detail (see SI-Figure A14.3). Three individual `crenso` runs were performed with identical settings and the standard deviation (SD) of the OR is computed. The three values differ insignifi-

cantly with an average SD of 28 $^{\circ}[dm(g/cm^3)]^{-1}$ over the NROR28 set (see the supporting information for details). The choice of three runs is a compromise between computational cost and accuracy that our tests have shown to work well in practice. In critical cases and for small OR values we suggest to proceed in this way, that is, averaging the OR over three `crenso` runs, and these values are discussed in the following.

First of all and most importantly, for all molecules in the test set, the sign of the OR is correctly obtained outright when the final, DFT-based `censo` structure ensemble is considered. This is a major, unprecedented achievement for such a diverse set of large and complicated pharmaceutical molecules. If the one dubious case (penicillin V) would have been also included, this corresponds to a success rate of 28 out of 29 cases (96%). The average relative deviation of the OR values between theory and experiment is about 72 %. Here, because of the large range of reference values from −160 to 287 and the included very small values, a regularized denominator $\frac{1}{2}(OR_{calc.} + OR_{ref.})$ is used in the relative deviation calculation (see the supporting information for the employed equations). The corresponding mean absolute deviation is 82 $^{\circ}[dm(g/cm^3)]^{-1}$, which is a good result considering the mentioned high complexity of the studied molecules and the problem in general. For 22 cases, a single `crenso` run finished within one day (Intel® Xeon® CPU E5-2660 v4 @ 2.00GHz at 28 cores). The final SE for the OR contains on average only eight considered conformers (Table A13.2), which indicates that even large drug molecules are often more rigid than formally derived from the Lewis structure and a topological bond-ring analysis.

The high sensitivity of OR to details of the conformation renders this chiroptical property an important analytical tool for the determination of stereochemical information. However, this sensitivity increases the complexity of the OR prediction problem since both high-quality geometries and the correct low-lying conformers have to be computed. To illustrate this difficulty, 20 conformers for each molecule were randomly selected, i.e., not according to an energy criterion, from a standard GFN-FF/ALPB `crest` run within an energy window of 10 kcal/mol. The corresponding OR values for those individual conformers are included as gray dots in Figure A13.4. They clearly show an extreme spread with sign and value changes by up to two orders of magnitude. This gives an impression how badly a single structure or low-quality SE based approach would perform. The error bar to the theoretical ensemble value represents the standard deviation of the OR calculated under the assumption that each conformer has a normal distributed error in its relative Gibbs energy of $\sigma = 0.4$ kcal/mol which is a reasonable estimate for our DFT treatment (see for example the relative conformer energies test sets ICONF, ACONF, MCONF, and SCONF in the GMTKN55 thermochemistry benchmark[73]). This standard deviation includes estimated errors for solvation (main effect), electronic energy, and thermostatistical corrections. For some systems, such as paroxetine, pregabalin, sitagliptin, sofosbuvir or rivaroxaban, the experimental and computed OR values fully agree when this error bar is taken into account. Such a good agreement in turn means that very likely, all relevant conformations were identified correctly.

Somewhat surprisingly, the OR value calculated for the energetically lowest-lying conformer has the correct sign for all cases except pregabalin, rosuvastatin, ritonavir, rapamycin and linezolid. This means that for a qualitatively correct description – considering the immense spread in OR values among conformers – it is of higher importance to find the energetically lowest conformer than to generate a

rather complete high quality SE. The effect of high-(free)-energy structures in the SE is presumingly washed out in many cases, which is an important finding of this work. Nevertheless, the agreement with the experimental values obtained in this simplified way is significantly worse than with the full ensemble, except for some cases (ampicillin, dexibuprofen, erythromycin, guaiol, itbrutinib, oxycodone, penicilin, sitagliptin), and the overall mean relative deviation is with 337 % also much larger. Note, that the population of lowest-lying conformer averaged over the test set is only 52 % (cf. Table 2).

Within an automated approach with targeted low computational cost, it is difficult to model all physically relevant processes properly. The here neglected vibrational effects, which have been considered in detail for small systems[596], can in principle be accounted for by molecular dynamics simulations, although one would have to constrain the geometries of the individual conformers in order to avoid double counting of nonrigidity effects. Moreover, a detailed investigation of protomers (i.e., differently protonated molecules) and possible tautomers is mandatory for more difficult systems. Whether automated workflows for explicitly solvated conformers can be developed is an important open question, but solvation effects on the population (and less importantly on the OR) are according to our experience the dominating factor for the achievable accuracy. Such developments would presumably bring the accuracy of the theoretical OR values to an even better quantitative level. In eventual more in-depth application studies, it is of course possible to improve on the here applied TDDFT based OR calculation at a higher level of theory, see for example Refs. 27,597–599. In any case, for flexible molecules, the determination of an accurate low-energy ensemble is of crucial importance. If independent CREST runs result in very different ensembles or if the OR values are very dependent on details of the conformation, the search step should be extended manually and carefully controlled.

Finally, we investigated whether it is possible to distinguish between pairs of diastereomers using our new workflow. Most OR applications are performed for enantiomers, but under certain conditions the absolute and relative configuration of diastereomers with two or more stereocenters can be solved[600] although it has to be noted that in most cases the relative stereochemistry is already known[601]. In Fig. A13.5, four examples A-D with diastereomeric pairs and their measured and calculated $[\alpha]_D$ values from a single `crenso` run are shown.

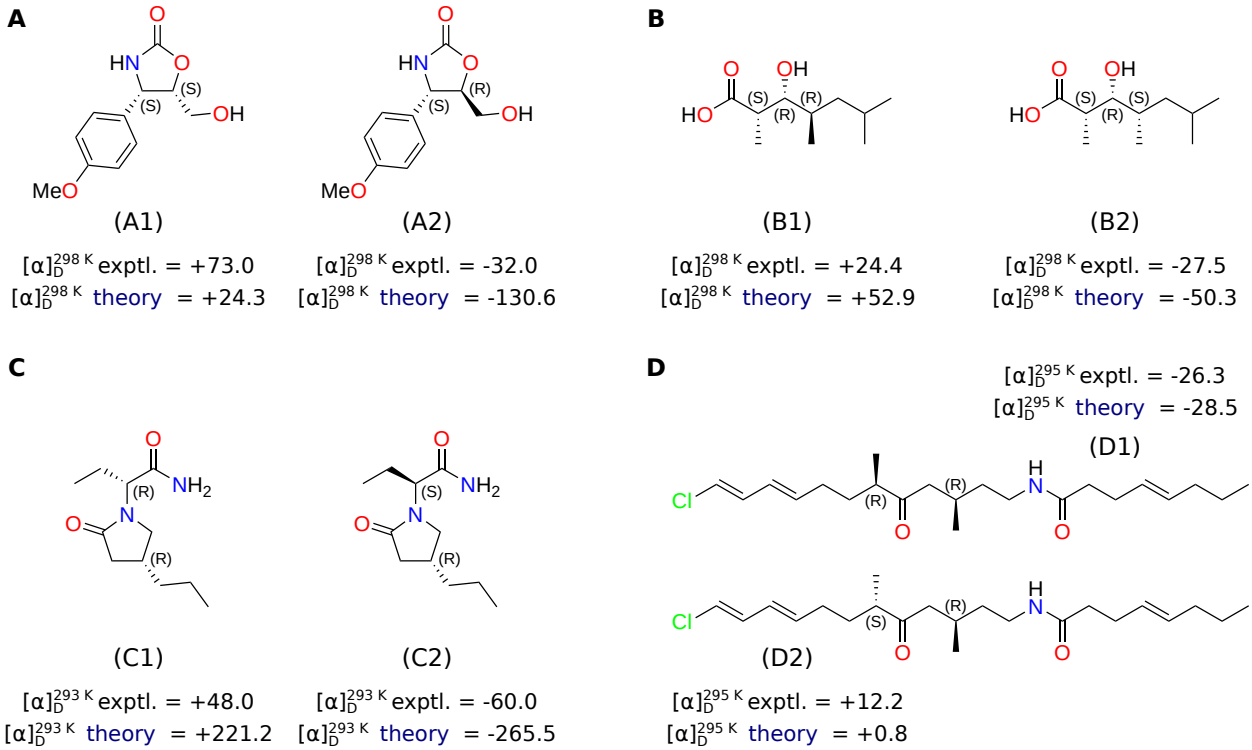


Figure A13.5.: Investigated pairs of diastereomers with measured and caluclated OR: A) (4*S*,5*S*)-*cis*-cytoxazone and (4*S*,5*R*)-*trans*-cytoxazone[602] (in MeCN), B) (*SRR*) and (*SRS*)-3-hydroxy-2,4,6-trimethylheptanonicacid diastereomers[603] (in $CHCl_3$), C) brivaracetam diastereoisomers[604] (in $CHCl_3$), D) diasteromers of pitiamide A[605] (in $CHCl_3$).

The chemical immuno-modulator agents *cis*- and *trans*-cytoxazone [(A1) and (A2), respectively] with experimental OR values of $[\alpha]_D$= 73.0 and $[\alpha]_D$= -32.0, were measured in acetonitrile[602]. These values are in reasonable agreement with the calculated ones of 24.3 and -130.6, respectively, for both diastereomers. The theoretical and experimental[603] values of the heptanocid acid diastereomers in Fig. A13.5B) match very well, both in terms of OR sign overall magnitude. The diastereomers C1 and C2 of the anti seizure drug brivaracetam (see Fig. A13.5C) can be distinguished by theory as well based on the sign of the calculated OR, but the magnitude of the computed values deviates somewhat from experiment[604]. The chlorinated lipid pitiamid A (see Fig. A13.5D) exhibits a very flexible backbone and two chiral carbon centers[605]. The high flexibility and multiple intramolecular interactions motifs make this 62-atom compound a real challenge for any theoretical approach. Nevertheless, the D1 diastereomer shows perfect agreement between experiment and theory whereas the D2 diastereomer exhibits the correct sign in the calculation, but the magnitude of the OR is too small. The shown examples indicate that our computational workflow can distinguish between pairs of diastereomers if the difference in their OR is sufficiently large.

A13.3. Conclusion

In this work, we proposed and evaluated the state-of-the-art `crenso` workflow for the automated calculation of the OR for large, nonrigid, biologically active molecules. The extension to related properties like electronic circular dichroism is straightforward, and tests for NMR spectra are currently undertaken in our lab. With this scheme an ensemble averaged specific rotation value can be obtained on a standard workstation within 1-2 days of computation time for flexible compounds consisting of roughly 50-100 atoms. Because of the very good parallelization efficiency of the algorithm, even larger systems with up to 200 atoms can be treated. For a challenging benchmark set of 28 typical drug molecules (NROR28 set), the sign and magnitude of the experimental $[\alpha]_D$ value is correctly reproduced in every case. This is a remarkable achievement, considering the complexity of optical rotation in particular for nonrigid compounds. Analysis of the conformer contributions indicates that in the majority of cases, the OR when evaluated only for the lowest-lying conformer is qualitatively correct and dominates the ensemble average. This, however, does not mean that a single structure approach may be always appropriate since a sophisticated conformer search still has to be performed to ensure that the correct global conformer minimum was found. Conversely, a good theoretical result in comparison to experiment can be interpreted in such a way that the lowest energy structure is likely at least qualitatively correct. The quality of the computed SE and the influence of Boltzmann factors on the OR could be studied more thoroughly if specific rotations at different temperatures are measured routinely, and we invite experimentalists to address this matter.

For the textbook case of the α- and β-form of D-glucopyranose in water we compute an equilibrium $[\alpha]_D$ value of 51.7 $^\circ[dm(g/cm^3)]^{-1}$ in excellent agreement with the experimental value of 52.7. This example shows that even small changes in the stereochemical structure (i.e., α- vs. β) can drastically change the OR contribution of the individual conformers. Furthermore, even in such seemingly rigid six-membered ring structures, about 15 almost equally contributing conformers need to be considered for an accurate OR calculation.

Although the complete algorithm is quite involved, we have developed a set of default technical parameters for the `crenso` script in this work enabling robust, highly efficient calculations with minimal user input. This extends the quantum chemistry toolbox to a routine treatment of the OR for nonrigid, larger druglike molecules. Importantly, this has been achieved without applying any empirical machine learning technique, and consequently the workflow can be applied out-of-the-box to other substance classes like, e.g., chiral transition metal complexes. The presented case study for a few pairs of flexible diastereomers already indicates that our treatment is expandable without any modifications. We hope that these new methods and workflows enable reliable assignments of absolute and relative configuration and the elucidation of complicated, general sterochemical relationships in the future.

A13.4. Computational Methods

All DFT calculations were conducted with Turbomole 7.5.1.[606] The COSMO-RS[307] solvation contribution to the Gibbs energy was calculated with COSMOtherm19[607] with the *BP_TZVP_C30_1601.ctd*

parameterization. To save computation time, all COSMO-RS calculations use the electron density and energy from the corresponding respective DFT level within the `censo` program. Geometries were optimized with implicit solvation at the r^{2}SCAN-3c/DCOSMO-RS [156,243] level of theory. r^{2}SCAN-3c is the best performing "low-cost" density functional method for general thermochemistry and, in particular, is very accurate for conformational energies where it closely approaches coupled-cluster benchmark accuracy [156]. The final Gibbs energies were calculated as the sum of r^{2}SCAN-3c electronic energy (E), COSMO-RS free solvation energy ($\delta G^T_{solv}(X)$) and the thermostatistical part (G^T_{mRRHO}) at GFN2-xTB/ALPB(solvent) level. Optical rotations were calculated using standard TDDFT in the velocity gauge formalism at the PBE/def2-SVPD[COSMO] level of theory employing r^{2}SCAN/def2-SVPD[COSMO] orbitals. Test calculations for the OR with larger AO basis sets or hybrid density functionals yield rather similar results (see SI-Figure A14.2) at much higher computational cost. For further details on the `crenso` workflow and technical settings see the supporting information and Ref. 69.

Conflict of interest

The authors declare no competing interests.

Supplementary information contains additional computational details and descriptions of abbreviations and programs and the NROR28 geometries. The `xtb` program package, `crenso` and all used scripts are available free of charge for academic use via the GrimmeLabs Github page. [608]

The German science foundation (DFG) is gratefully acknowledged for financial support in the framework of the Gottfried Wilhelm Leibniz prize to S.G. The authors thank P. Pracht and A. Hansen for helpful discussions and comments to the manuscript.

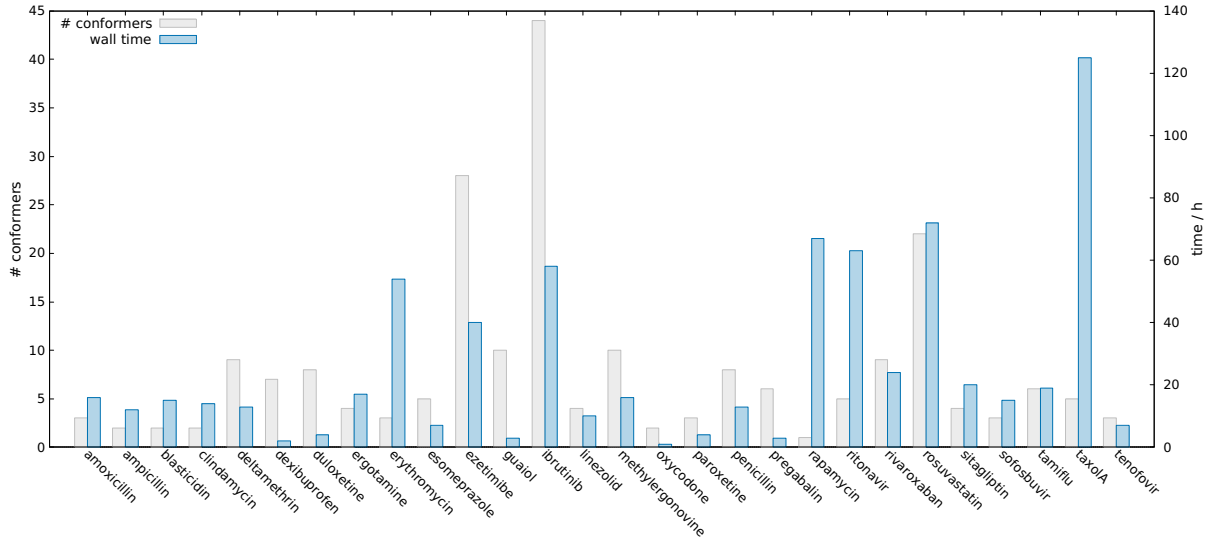
A14. Supporting Information to Chapter 9

Appendix A14 contains:

- Additional technical details of the calculations
- coordinates in xmol format

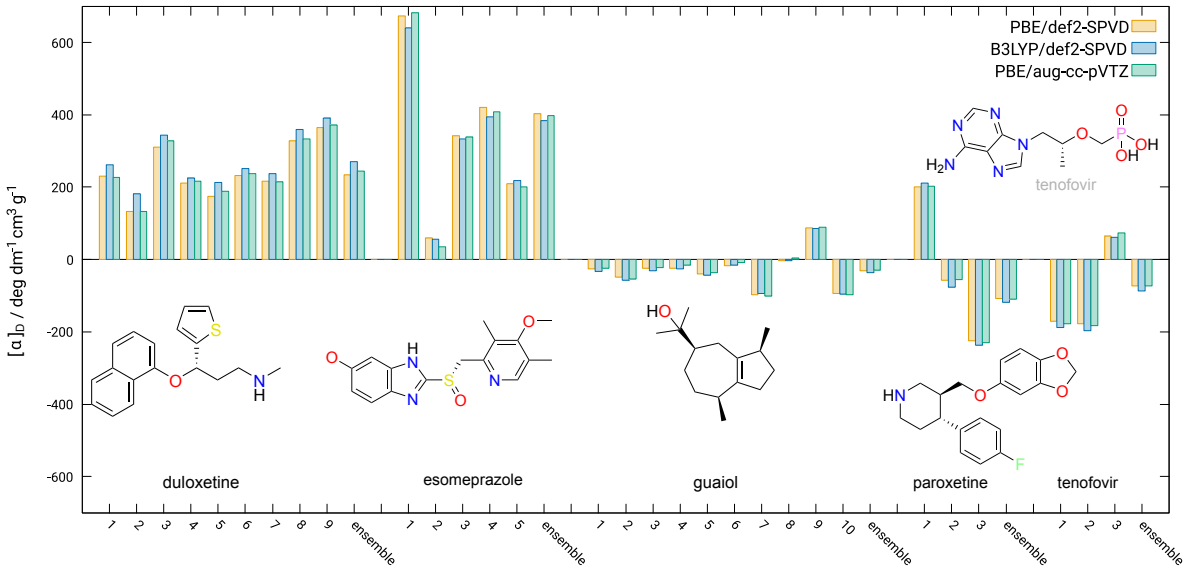
Computational Details

Final number of conformers and computation time



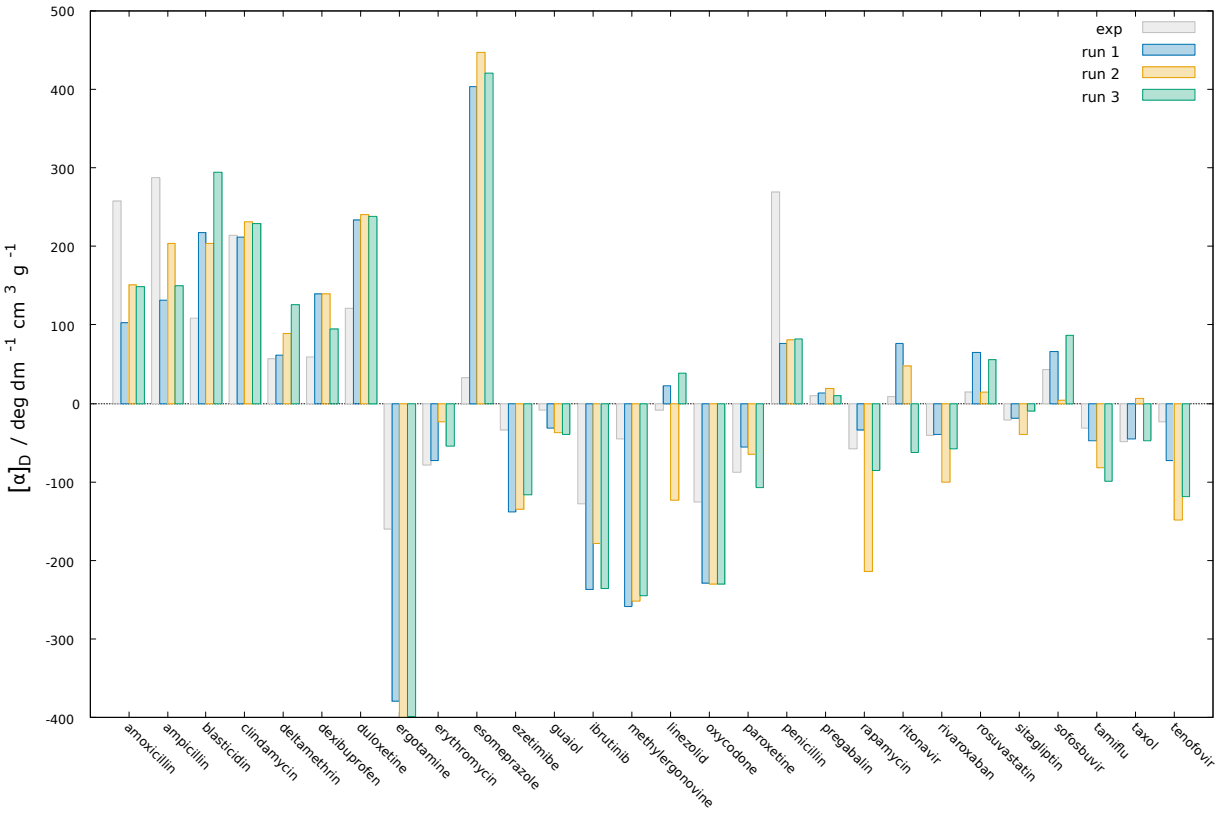
FigureA14.1.: Y-axis1 shows the averaged number of conformers in final ensemble from three independent `crenso` runs (grey bars) and y-axis2 shows the averaged wall time in hours (blue bars) (Intel® Xeon® CPU E5-2660 v4 @ 2.00GHz at 28 cores) of the NROR28 test set.

Comparison and validation of theoretical methods for OR computation



FigureA14.2.: Specific rotation values ($[\alpha]_D$) computed by PBE/def2-SVPD[COSMO] using r^{2}SCAN/def2-SVPD[COSMO] orbitals (yellow), B3LYP/def2-SVPD[COSMO] (blue) and PBE/aug-cc-pVTZ[COSMO] using r^{2}SCAN/def2-SVPD[COSMO] orbitals (green) for a NROR28 subset of five molecules. Solvents are applied as stated in SI-Table A14.1. Numbers refer to individual conformers and the Boltzmann-weighted value is denoted with "ensemble". For all three method combinations within the respective compound the same Boltzmann factors are applied. The B3LYP calculations were performed in the length representation, due to missing implementation of the velocity formalism for hybrid density functionals in TURBOMOLE.

Computed $[\alpha]_D$ for the NROR28 test set for three independent crenso runs



FigureA14.3.: Specific rotation values ($[\alpha]_D$) gathered from experiment[587–592] (grey) and computed from three independent `crenso` runs (blue, yellow, green) of the NROR28 test set.

Solvent information for the NROR28 test set

TableA14.1.: Solvents used in experimental OR measurements and in the calculations. The total charge of the compounds is provided as well.

Molecule	solvent, experiment	solvent, theory	total charge
amoxicillin	n/a	H_2O	0 (zwitterionic)
ampicillin	H_2O	H_2O	0 (zwitterionic)
blasticidin S	H_2O	H_2O	0 (zwitterionic)
clindamycin	$CHCl_3$	$CHCl_3$	0
deltamethrin	Toluene	Toluene	0
dexibuprofen	EtOH	MeOH	0
duloxetine	MeOH	MeOH	0
ergotamine	$CHCl_3$	$CHCl_3$	0
erythromycin	EtOH	MeOH	0
esomeprazole	H_2O	H_2O	-1
ezetimibe	MeOH	MeOH	0
guaiol	n/a	H_2O	0
ibrutinib	$CHCl_3$	$CHCl_3$	0
linezolid	$CHCl_3$	$CHCl_3$	0
methylergonovine	n/a	H_2O	0
oxycodone	H_2O	H_2O	0
paroxetine	EtOH	MeOH	0
penicillin	MeOH	MeOH	-1
pregabalin	H_2O	H_2O	0 (zwitterionic)
rapamycin	MeOH	MeOH	0
ritonavir	MeOH	MeOH	0
rivaroxaban	DMSO	DMSO	0
rosuvastatin	MeOH/H_2O	MeOH	0
sitagliptin	H_2O	H_2O	+1
sofosbuvir	EtOH	MeOH	0
tamiflu	H_2O	H_2O	0
taxol A	MeOH	MeOH	0
tenofovir	HCl(aq)	H_2O	+1

Flexibility measure of investigated compounds

TableA14.2.: The flexibility measure [185] has been extracted from a final `crest` run in the `crenso` script. Rigid molecules exhibit flexibility measures close to zero and highly flexible linear alkanes have values of about 0.9. Flexible drug-like molecules show as a rule of thumb measures between 0.3 and 0.5. Larger values for the standard deviation of the ensemble G_{mRRHO} value indicate a higher structural diversity in the ensemble. The SD(G_{mRRHO}) values have been extracted from the pre-screening filtering part one in `censo`.

Molecule	**no. atoms**	**Flexibility measure**	**SD(G_{mRRHO})**
amoxicillin	44	0.28	0.05
ampicillin	43	0.28	0.21
blasticidin S	56	0.25	0.42
clindamycin	60	0.35	0.20
deltamethrin	47	0.32	0.53
dexibuprofen	33	0.34	0.18
duloxetine	40	0.41	0.29
ergotamine	78	0.23	0.47
erythromycin	118	0.30	0.51
esomeprazole	43	0.37	0.36
ezetimibe	51	0.30	0.41
guaiol	42	0.23	0.35
ibrutinib	57	0.21	0.47
linezolid	44	0.36	0.32
methylergonovine	50	0.29	0.68
oxycodone	44	0.16	0.02
paroxetine	44	0.33	0.29
penicillin	41	0.31	0.22
pregabalin	28	0.58	0.29
rapamycin	144	0.27	2.61
ritonavir	98	0.37	0.94
rivaroxaban	47	0.30	0.63
rosuvastatin	61	0.35	0.48
sitagliptin	43	0.33	0.65
sofosbuvir	65	0.37	0.35
tamiflu	50	0.48	0.48
taxol A	113	0.27	0.66
tenofovir	33	0.46	0.32

Qualitative description of terms and abbreviations

SE = Structural Ensemble refers to a collection of equilibrium geometries (3D-structures) that describes a collection of individual states representing the macroscopic system in the thermodynamic equilibrium. The term "ensemble" originates from statistical mechanics and usually describes a smaller set of possibilities selected from the set of possible states. As such, the term "structural/conformational ensemble" is derived from the canonical ensemble and describes the collection of conformers representing the molecule under the given conditions.

PES = The Potential Energy Surface describes a quantum mechanical system in the Born-Oppenheimer description with its potential energy as a function of the nuclear coordinates. Minima on the PES are for example conformers or rotamers. Dominantly populated conformers decisively describe the macroscopic system. We use the term Conformer-Rotamer-Ensemble (CRE) for the thermodynamic collection of low (free) energy conformers and rotamers.

protomer = In chemistry, a so-called protomer is a molecule which displays tautomerism due to position of a proton (see: https://en.wikipedia.org/wiki/Protomer).

SQM = Semiempirical Quantum Mechanical methods are electronic structure methods that are either based on the Hartree Fock or Density Functional Theory formalism, but introduce approximations and often have empirical parameters fitted to specific properties. The approximations are often the neglect of computational expensive parts, e.g., integral evaluations, which speed up the computation significantly compared to *ab initio* methods. Some examples of SQM methods are: the Hückel method, MNDO, AM1, PMx (x=6,7), DFTB, xTB and many more. Some information can be found in Ref. 59,61.

Explicit and implicit solvation = In computational chemistry the interaction of the solute with the solvent can be described by mainly two approaches: the implicit (continuum) solvation or "explicit" solvation involving individual solvent molecules. Continuum models describe the solvent as a uniform polarizable medium with an associated dielectric constant ϵ. Two examples of implicit models are: the CONductor-like Screening MOdel (COSMO)[242] and the Polarizable Continuum Model (PCM)[609]. The explicit treatment of solvation is more involved as individual solvent molecules are included in the quantum mechanical calculation. The explicit treatment increases the cost significantly, but also describes all interactions between solute and solvent correctly, within the applied level of theory. However, the minimum number of solvent molecules required to properly solvate the investigated compound is difficult to estimate. In addition, the explicit solvent box/cluster around the solute must be equilibrated to represent a system in thermodynamic equilibrium.

Short and qualitative methods description

xTB: xTB is an eXtended Tight Binding program package. Within xTB semiempirical quantum mechanical (SQM) methods like GFN1-xTB and GFN2-xTB are implemented. Their target properties are

Geometries, vibrational **F**requencies and **N**on covalent interactions. The implemented methods are consistentlly parameterized for all elements up to Radon (Z=86). Through the use of a minimal atomic orbital (AO) basis set, as well as an efficient Hamiltonian the GFN*n*-xTB methods are at least two orders of magnitude faster, compared with *ab initio* methods (e.g., DFT). For very fast calculations or when the investigated system size is very large (≈ 1000 atoms) a GFN-force field (GFN-FF) is available in xTB. The combination of systematic availability, computational efficiency and robustness makes these methods ideal for the evaluation of chemical space, where millions of single point and gradient calculations have to be performed. Additional information can be obtained from Ref. 61, the online documentation 610 and the code can be obtained from Ref. 511.

`crest`: The Conformer Rotamer Ensemble Search Tool (CREST) is a sampling program making use of the `xtb` code, i.e., energies and gradients are calculated at semiempirical or force field level of theory. It can be used to search the low energy chemical space, and create/analyse conformer, rotamer ensembles (CRE), in other words: identify a conformer distribution for a target compound. In addition, `crest` can be used to generate protomer or tautomer ensemebles and to find NMR (nuclear exchange) equivalencies. The efficient chemical space exploration is achieved by extensive meta-dynamics (MTD) sampling and information on the use of `crest` can be obtained from Ref. 68,419,611.

`censo`: The Commandline ENergetic SOrting program (CENSO) is used to refine conformer ensembles at a higher (DFT) level. Refinement of conformer ensembles in terms of optimized geometries and, in particular, Gibbs free energies is required because the accuracy of SQM methods is not sufficient when very sensitive properties such as optical rotation are studied. CENSO is designed to process large ensembles of conformers, by applying sophisticated combinations of methods (DFT) in terms of computational cost, accuracy and a predefined reasonable choice of sorting thresholds. CENSO interfaces to fast QM codes (TM, ORCA) and analyses the returned results. Additionally, CENSO can provide DFT-optimized conformer ensembles, free energies and Boltzmann factors and can be used to calculate specific rotation and NMR properties. Information on the use of CENSO can be obtained from Ref. 69,612.

`crest_combi`: crest_combi is a shell script that uses CREST to search conformers on different potential energy surfaces (PES) and combines the results. In this way, a more diverse ensemble of conformers can be obtained, i.e., a larger chemical space is explored, than by a single CREST run.

`crenso`: crenso is a shell script which simplifies the handling of `crest` and `censo` for the calculation of properties like logP, NMR, OR, pKa and can be used by simple unix commands e.g. (*crenso -l3 -or_fine h2o > crenso.out*). Its use is explained in Ref. 69 and can be obtained from: `https://github.com/grimme-lab/CRENSO`.

Statistical means

MD:

$$MD = \frac{\sum_{i}^{N} X_i}{N} \tag{A14.1}$$

SD:

$$SD = \sqrt{\frac{\sum_{i}^{N} (X_i - MD)^2}{N-1}} \tag{A14.2}$$

MAD:

$$MD = \frac{\sum_{i}^{N} |X_i|}{N} \tag{A14.3}$$

relMD(%):

$$relMD = \frac{\sum_{i}^{N} \frac{X_i - X_{i,ref}}{0.5 \cdot (X_{i,ref} + X_i)}}{N} \cdot 100 \tag{A14.4}$$

relMAD(%):

$$relMAD = \frac{\sum_{i}^{N} \left| \frac{X_i - X_{i,ref}}{0.5 \cdot (X_{i,ref} + X_i)} \right|}{N} \cdot 100 \tag{A14.5}$$

Technical settings applied in the `crenso` calculations

Free energies for each conformer are calculated from (see Ref: 236):

$$G = E + G_{mRRHO} + \delta G^{T}_{solv} \tag{A14.6}$$

Where E is the electronic energy, G_{mRRHO} represents the thermostatistical contribution to free energy, including the zero point vibrational energy, calculated in the modified rigid rotor harmonic oscillator framework [183,184] and δG^{T}_{solv} is the solvation free energy for the transfer of a compound in the gas phase into solution under standard conditions.

Table A14.3.: Technical settings as applied in the `crenso` calculation using (*crenso -or_fine solvent -l3*)

	CENSO general information:
CENSO version	1.0.8
Temperature	298.15 K
Solvent	corresponding to test set system
	PART0 - cheap pre-screening
Energy	b97-d/def2-SV(P) + GCP
Energy_settings	scfconv 5, grid 1 using D3(0) instead of D3(BJ)
G_mRRHO	not included
G_solv	alpb[gfn2]
Geometry	GFNn-xTB (input geometry)
Threshold	5.0 kcal/mol
main QM code	TM
	PART1 - pre-screening
Energy	r2scan-3c
Energy_settings	grid m4 scfconv 6
G_mRRHO	GFN2[alpb]-bhess SPH
G_solv	COSMO-RS-normal[r2scan-3c]
Geometry	GFNn-xTB (input geometry)
Threshold	3.0 kcal/mol
main QM code	TM
	PART2 - optimization
Energy	r2scan-3c
Energy_settings	grid m4 scfconv 6
G_mRRHO	GFN2[alpb]-bhess SPH
G_solv	COSMO-RS-normal[r2scan-3c]
Geometry	r2scan-3c[DCOSMORS] @optlevel: lax using grid m3 scfconv 6
Threshold	Opt_limit: 2.5 kcal/mol, Boltzmann sum threshold: 99.0 %
main QM code	TM
	PART5 - OR mode
Energy	using Energy from part2 - optimization
Energy_settings	see part2 - optimization
G_mRRHO	using G_mRRHO from part2 - optimization
G_solv	using G_solv from part2 - optimization
Geometry	r2scan-3c[DCOSMORS] @optlevel: lax
Threshold	Boltzmann sum threshold: 99.0 %
main QM code	TM
Optical rotation	pbe/def2-SVPD[COSMO]_[SCF=r2scan/def2-SVPD[COSMO]] using grid 2 scfconv 6

A15. Hydrocarbon Macrocycle Conformer Ensembles and ^{13}C-NMR spectra

Abstract: NMR as a routine analytic method provides important three-dimensional structure information of compounds in solution. Here we apply the recently released `crenso` computational workflow for the automated generation of conformer ensembles to the quantum mechanical calculation of ^{13}C-NMR spectra of a series of flexible cylcoalkanes up to $C_{20}H_{40}$. We evaluate the computed chemical shifts in comparison with corresponding experimental data in chloroform. It is shown that accurate and properly averaged theoretical NMR data can be obtained in about a day of computation time on a standard workstation computer. The excellent agreement between theory and experiment enables to deduce preferred conformations of large, nonrigid macrocylces under ambient conditions from our automated procedure.

A15.1. Introduction

Conformational analysis and thereby the knowledge about preferred molecular conformations and the three dimensional structure of molecules is of fundamental importance in chemistry.[510] It enables an in-depth understanding of reaction mechanisms[613], molecular recognition,[614] or identification of bioactive states.[21] The investigation of flexible macrocycles already started in the early 20th century and the Nobel price awarded to D. Barton[615] in 1969 highlights the general interest in understanding conformationally rich cyclic compounds. Cycloalkanes represent valuable model systems for studying macrocycle conformations in solution. They meet the requirements of readily available experimental data and exist in systematically increasing ring sizes. The absence of functional groups enables unobstructed investigation of basic conformational effects and minimizes additional influences of strong interactions, e.g., hydrogen bonding. From a theoretical point of view, cylcoalkanes are well suited as basic model compounds due to their relatively simple intramolecular interactions accompanied by an enormously large and rich but still explorable potential energy surface (PES). Hence, many studies on cycloalkanes were performed already in the early 1970s with important contributions by Anet, Allinger, Dale and Schleyer[616–621] to name a few. The reported data mainly consist of crystal structure information, force-field strain-energy calculations, and low-temperature NMR measurements. NMR spectroscopy presents an excellent analytical method for the study of flexible molecules and is nowadays an extremely important routine application in the field.[190,622,623]

A15.2. Results and Discussion

Before presenting our computational approach, we briefly discuss previous works in the field of conformational analysis focusing on macrocyclic compounds in solution. From an experimental viewpoint NMR data like nuclear overhauser effect crosspeak volumes (NOE), ^3J scalar couplings constants and residual dipolar couplings (RDCs)[624–626] are used to analyse structural motifs. This data provides information on interatomic distances, dihedral angles and relative orientational restraints from which a molecular structure can be derived. However, often the structure identification problem is underdetermined and to address this problem combined experimental and theoretical workflows like the NMR Analysis of Molecular Flexibility in Solution (NAMFIS) method have been developed.[237] Bazzo et al. proposed to compute conformer ensembles from Monte Carlo and molecular dynamics (MD) simulations, which can already in their creation process be guided by experimentally derived constraints, i.e., using ^3J dihedral or NOE distance constraints. From the ensemble NAMFIS determines relative populations of conformers that are in agreement with experimental evidence and evaluates the quality of the reproduction of the structural features, without the evaluation of theoretically determined (free) energies. This very much experimentally driven approach has been successfully applied in the investigation of flexible LSD-1 inhibitors,[627] the solution structures of cell permeable macrocycles[628] and the folding propensities of β-hairpin peptides.[629] An improvement upon NAMFIS is the Distribution of Solution Conformations (DISCON) software which re-implemented NAMFIS and adds a clustering algorithm and new solvers.[238,630,631] More recently in 2018 Nguyen et al. introduced the conformational analysis from NMR and density-functional prediction of low-energy ensembles (CANDLE) approach.[238] It is based on an iterative generation of the ensemble through high-temperature CHARMM molecular dynamics simulations followed by energetic sorting and clustering of conformers. DFT optimized gas-phase conformers are created from each cluster, which serve as starting point for room temperature MDs, that are then re-clustered. The final ensemble is optimized at DFT level, energies and chemical shifts are computed and compared directly to the experiment. The authors identified the most complex task to be the exploration of the chemical space, but also noted that errors in the chemical shift calculations can hamper a comparison to experiment.

Currently, enormous efforts are being made to use macrocyclic compounds as drug leads for many difficult but highly important targets.[632,633] The main difficulty is that they are flexible and their solution ensembles and dynamics have to be taken into account, but few accurate methods exist to perform such an analysis routinely. This motivated us to study cycloalkanes as model systems and to propose a state of the art method to analyze the solution conformations of flexible macrocycles.

In 2017 we proposed a theoretical protocol for the automated computation of spin-spin-coupled NMR spectra for flexible compounds.[235] Its' key aspect is the generation of a fairly complete (low-energy) conformer-rotamer ensemble (CRE) by searching on the PES of the fast semiempirical GFN*n*-xTB[61,68] methods. The classification of the ensemble into identical isomers, conformers, and rotamers is based on electronic energy, atomic Cartesian RMSD, and purely structure-based molecular rotational constants using predefined thresholds (for more details see Ref. 68). Subsequent refinement of geometries and free energies at higher DFT theoretical level accurately identifies the most populated conformers

for which NMR properties, i.e., shielding and spin-spin-coupling constants are calculated. Using the assembled data, the NMR parameters are properly thermally averaged, the spin Hamiltonian is solved and a fully coupled spectrum is obtained without any manual intervention which can be directly compared to experiment. The same procedure has recently been successfully applied to calculate the conformationally averaged optical rotation values of 28 large commercial drug molecules[634] as well as other properties of nonrigid compounds.[69]

Here, we investigate the series of cycloalkanes from cyclobutane to cylcoicosane and their conformational behavior in $CHCl_3$ solution at 298 K. The quality of the computed conformational ensemble is investigated by comparison of theoretical and experimental ^{13}C-NMR spectra (^{1}H and ^{13}C chemical shifts are given in the SI). The high sensitivity of NMR to structural details enables a reliable benchmarking of how well theory can describe conformer populations in solution. However, for nonrigid compounds at ambient conditions, the rapid nuclei interchange typically occurring on a sub ps time scale, leads to an average of NMR parameters on the much slower time scale of the experiment which has to be considered theoretically. The relevant thermally accessible ensemble of minimum energy structures generally consists of conformers as well as rotatmers (conformer/rotamer ensemble, CRE). A conformer is characterized by a distinct energy minimum while rotamers arise from bond rotation or inversion process leading to an interchange of nuclei and to minima with identical energies and NMR parameters.[235] For this variety of structures, an averaged NMR spectrum is measured[194] and hence the accurate computation of the CRE is a key for reliable NMR simulations of flexible compounds.

The effect of including or ignoring an ensemble average is demonstrated here for the ^{13}C-NMR spectrum of cycloheptadecane ($C_{17}H_{34}$) in Figure A15.1. The hypothetical spectrum of only one frozen rotamer, i.e., a single geometry depicted in FigureA15.1A a) in the absence of any nuclei exchange, is shown in part B, spectrum a). Here, all 17 carbon atoms are chemically inequivalent and can be distinguished by separate signals.

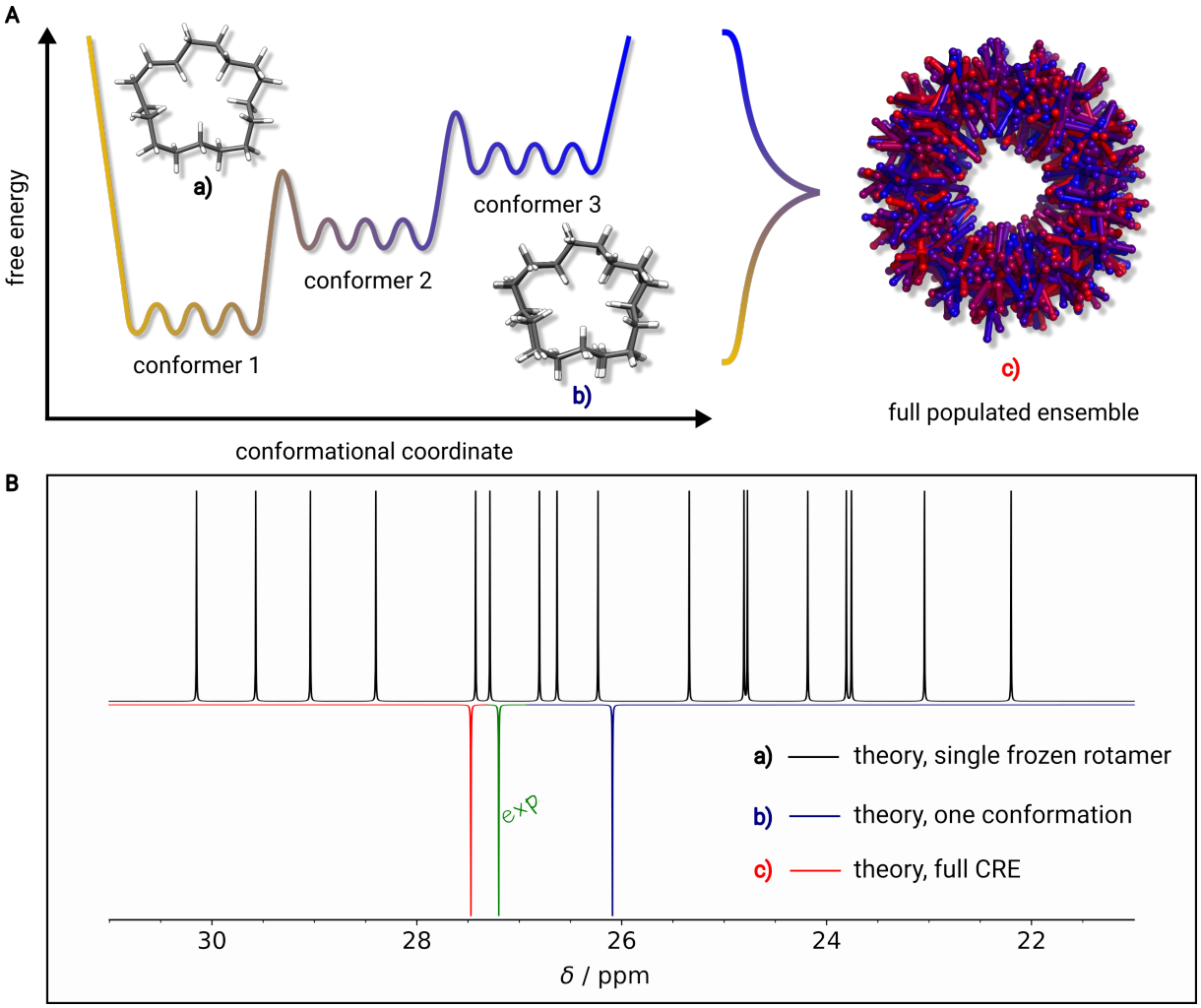


Figure A15.1.: **A** represents a schematic potential energy surface (PES) of cycloheptadecane with three conformers and some energetically degenerate rotamers. **B** shows the theoretical ^{13}C-NMR spectra of cycloheptadecane $C_{17}H_{34}$ in $CHCl_3$ at 298 K obtained for: a) without rotamer averaging, b) including a single (lowest) conformer with all its rotamers and c) including all significantly populated conformers (168) in the ensemble (the colors in the overlay visualize the diversity of contributing conformers). The experimental reference chemical shift is shown in green (see SI-Table A16.1).

As shown in Figure A15.1A b) and B b), the description improves dramatically when all rotamers belonging to one conformer are considered, i.e., correctly only a single singlet signal (ignoring spin-spin coupling as in the experiment) is obtained. For cycloalkanes, inter-conversion takes place through fast pseudo-rotation or inversion processes, whereby all methylene groups become chemically equivalent at ambient temperature.[617] However, a reasonably accurate chemical shift is obtained only by inclusion of the entire populated ensemble as shown in Figure A15.1B, spectrum c) and the corresponding structure ensemble presented in Figure A15.1A c). This example illustrates the complexity of NMR simulations for

flexible compounds even when spin-spin coupling is not involved, since all the populated conformers and rotamers of a compound and all chemically/magnetically equivalent atoms have to be identified.

To simplify the tedious task of generating fully coupled NMR spectra,[235] we have developed and tested a highly improved workflow implemented in a shell script called `crenso`[69] that can be used to calculate a variety of properties, e.g., NMR, optical rotation ($[\alpha]_D$), acid dissociation constants (pKa) or partition coefficients (logP).[69]

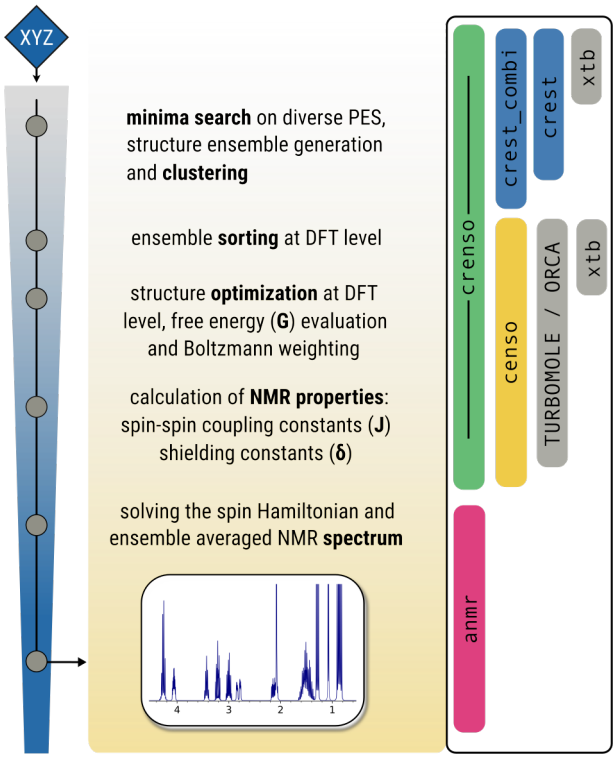


Figure A15.2.: Computational workflow for the automated computation of NMR spectra and therein employed computer codes and scripts. The colored boxes indicate codes that are directly accessed and contain relevant thresholds for generating and sorting the CRE or are related to the generation of the NMR spectrum/properties. The gray colored boxes indicate the necessary infrastructure relevant to the QM description of the model.

The required steps to obtain a theoretical NMR spectrum are illustrated in Figure A15.2 starting from an initial three dimensional molecule structure. First the CRE is created using the `crest_combi` script.[68,69] Hereby the vast chemical space is explored using meta-dynamic[419] runs at different, artificially enriched PES.[69] The huge number of associated energy and gradient calculations required in this endeavor is made possible by efficient and robust quantum mechanical methods of the GFN*n*-xTB family[61] and the GFN-FF force field.[181] Using a special principle component analysis clustering algorithm,[185] the number of conformer candidates is efficiently condensed to the essential structures. This initial ensemble is then treated by a threshold based algorithm termed `censo`, progressively optimizing according to both, lowest total Gibbs energies as well as the solution phase geometries of the conformers. After the identification of dominantly populated ones, spin-spin coupling- and shielding

constants are calculated by standard DFT methods. Finally, the Boltzmann weights, NMR parameters and information on chemically- and magnetically equivalent nuclei are transferred to the `anmr` program, which averages the NMR data over the CRE, solves the spin Hamiltonian, and returns the final spectrum. Because our CRE is of high quality but not complete, i.e., the exact number of rotatmers for each conformer is not known, we assume that all conformer degeneracies are unity. By a randomized variation of populations (explained in the caption of Fig. A15.3) we checked that that this approximation is valid and leads only to insignificant changes of the average (see gray confidence band in Figure A15.3). More relevant is the proper consideration of the molecular symmetry of each conformer in the thermostatistical evaluation of the free energy (molecular rotation/symmetry number) which is in some cases crucial for determining reliable Boltzmann factors (an extreme example is cyclononane,[635] see SI-Figure A16.4).

All DFT calculations were conducted with Turbomole 7.5.1.[606] The COSMO-RS[307] solvation contribution to the Gibbs energy was calculated with COSMOtherm19[607] (*BP_TZVP_19.ctd* parameterization). To save computation time, these calculations use the electron density and energy from the corresponding respective DFT level within the `censo` program. Geometries were optimized with implicit solvation at the r^{2}SCAN-3c/DCOSMO-RS[156,243] level of theory. r^{2}SCAN-3c shows good performance for general thermochemistry and yields highly accurate conformational energies making this efficient composite density functional an ideal choice.[156] The final Gibbs energies were calculated as the sum of r^{2}SCAN-3c electronic energy (E), COSMO-RS solvation free energy ($\delta G^{T}_{solv}(CHCl_3)$) and the thermostatistical correction (G^{T}_{mRRHO})[184] at GFN2-xTB/ALPB($CHCl_3$)[180] level. NMR parameters (shielding constants) are calculated at the PBE0/def2-TZVP[DCOSMO-RS][122,287] level. The entire workflow can be called by simple UNIX commands, e.g., `crenso -nmr chcl3 -l3` and for further details and technical settings see the supporting information and Ref. 69,634.

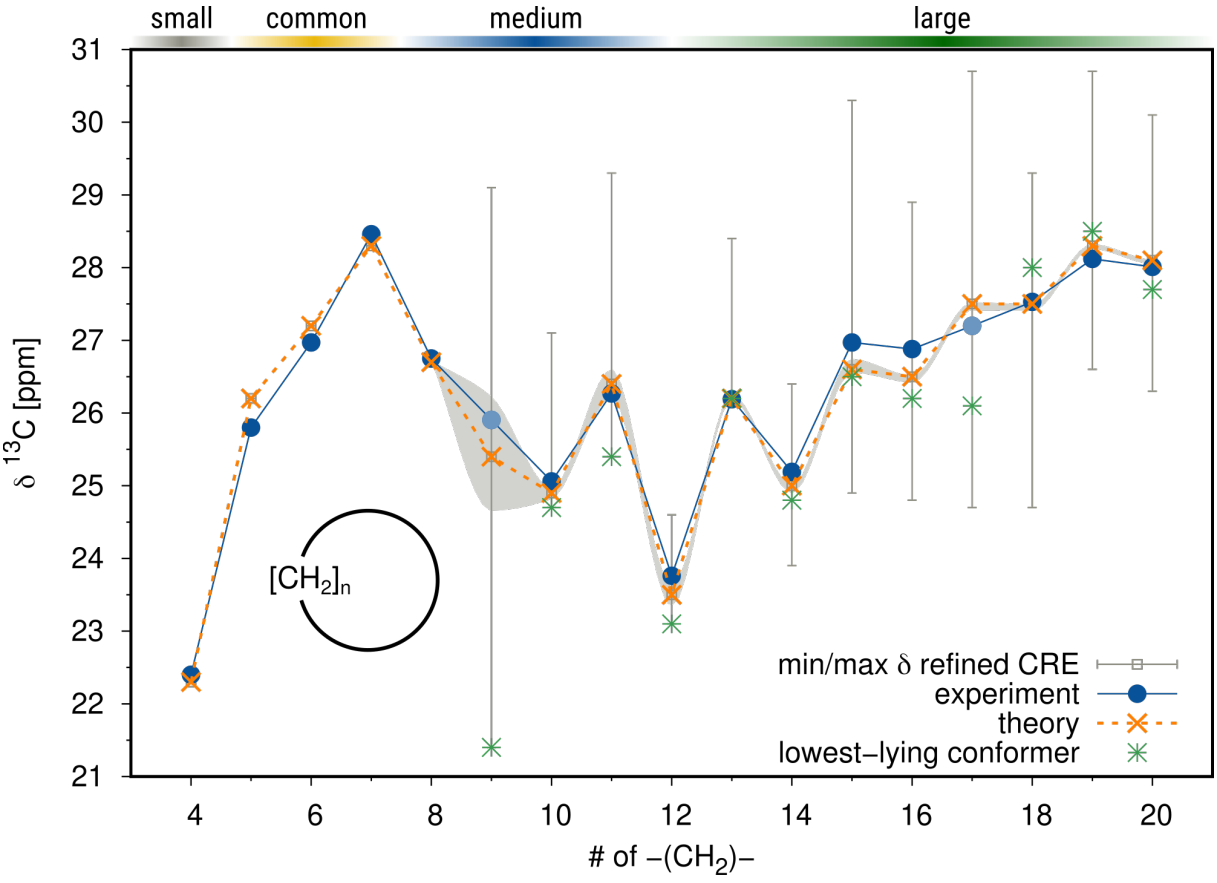
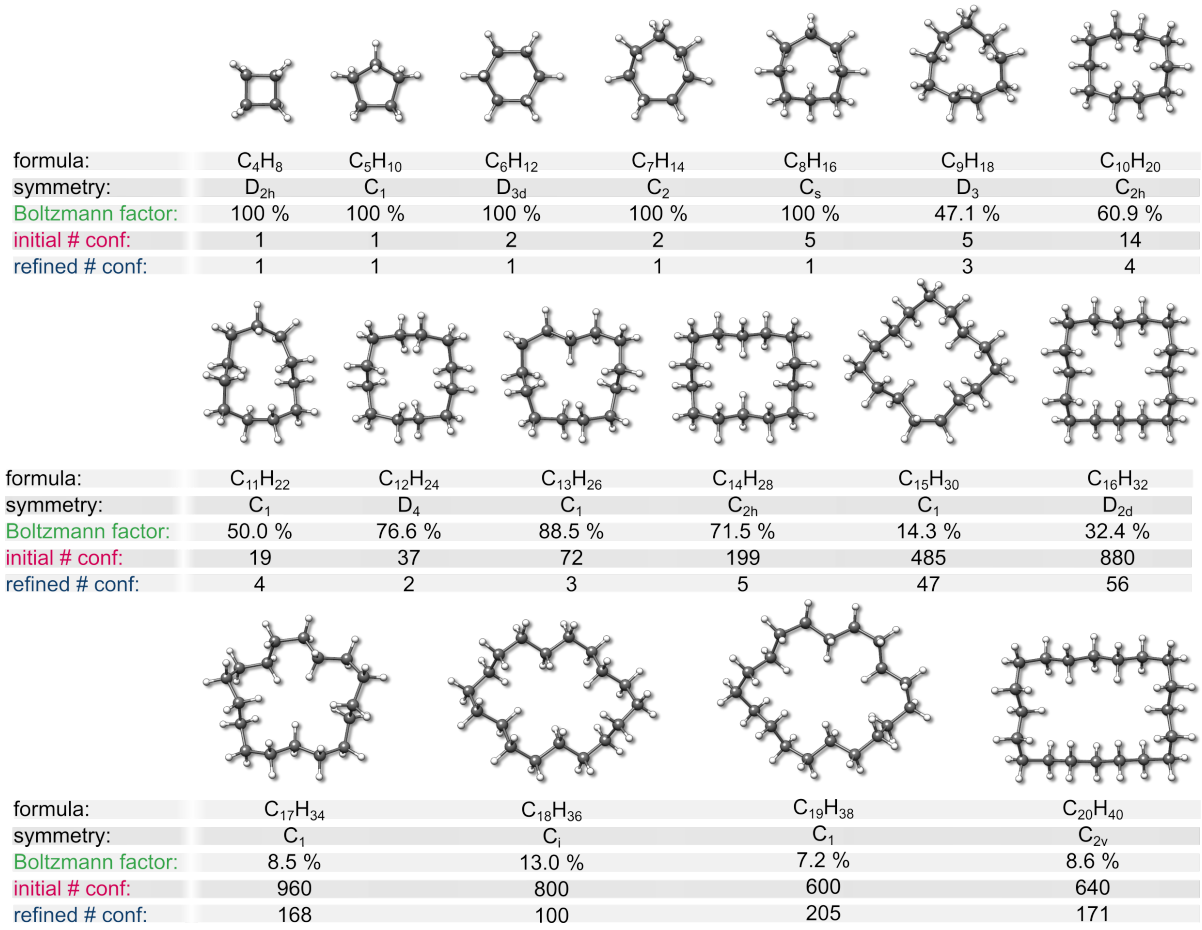


Figure A15.3.: Experimental[636,637] (dark blue points) and theoretical (orange crosses) ^{13}C chemical shifts of cycloalkanes in $CHCl_3$ are plotted against their ring size. Error bars represent the smallest and largest chemical shift computed for the populated structures in the ensemble. Light blue points show extrapolated reference values of systems for which no experiment was available. Green crosses depict values of the highest populated conformer. The gray confidence band represents estimated changes in the shifts when the Boltzmann factors of randomly selected conformers in the CRE are adjusted by changing their free energy by up to 0.4 kcal mol^{-1}, which is a reasonable error estimate for our treatment (see for example the relative conformer energies test sets ICONF, ACONF, MCONF, and SCONF in the GMTKN55 thermochemistry benchmark[73]). The connecting lines are only drawn for visual guidance.

The workflow described in Figure A15.2 is applied to cycloalkanes containing 4-20 carbon atoms, starting from smiles strings or PubChem geometries. A comparison of theoretical and experimental chemical shifts in $CHCl_3$ is shown as a function of ring size in Figure A15.3. The experimental values refer to the fundamental work by Fritz et. al.[637]. Since experimental data in chloroform is not available for cyclononane and cycloheptadecane, their reference values are extrapolated from the next larger and smaller homologous, respectively. This extrapolation is considered appropriate because experi-

mental data measured in other solvents result in significant inconsistencies if the trend over a series of compounds is of main interest.

The conformers of lowest free energy are displayed in Figure A15.4 and since the chemical shifts and structures are interrelated, they are discussed together hereafter. As recommended by Lodewyk et al.,[26] calculated DFT shielding constants $\sigma_{DFT,raw}$ are corrected globally by linear regression to eliminate systematic discrepancies between theory and experiment[26] using the formula $\delta_{DFT,corr.} = 0.942 \times (186.156 - \sigma_{DFT,raw})$ where $\delta_{DFT,corr.}$ is the corrected shift discussed in the following. The linear regression parameters have been determined from a database of small organic molecules with experimental carbon and proton shifts, accessible at the chemical shift repository[638] (see the supporting information for details).



formula:	C_4H_8	C_5H_{10}	C_6H_{12}	C_7H_{14}	C_8H_{16}	C_9H_{18}	$C_{10}H_{20}$
symmetry:	D_{2h}	C_1	D_{3d}	C_2	C_s	D_3	C_{2h}
Boltzmann factor:	100 %	100 %	100 %	100 %	100 %	47.1 %	60.9 %
initial # conf:	1	1	2	2	5	5	14
refined # conf:	1	1	1	1	1	3	4

formula:	$C_{11}H_{22}$	$C_{12}H_{24}$	$C_{13}H_{26}$	$C_{14}H_{28}$	$C_{15}H_{30}$	$C_{16}H_{32}$
symmetry:	C_1	D_4	C_1	C_{2h}	C_1	D_{2d}
Boltzmann factor:	50.0 %	76.6 %	88.5 %	71.5 %	14.3 %	32.4 %
initial # conf:	19	37	72	199	485	880
refined # conf:	4	2	3	5	47	56

formula:	$C_{17}H_{34}$	$C_{18}H_{36}$	$C_{19}H_{38}$	$C_{20}H_{40}$
symmetry:	C_1	C_i	C_1	C_{2v}
Boltzmann factor:	8.5 %	13.0 %	7.2 %	8.6 %
initial # conf:	960	800	600	640
refined # conf:	168	100	205	171

Figure A15.4.: r^{2}SCAN-3c/DCOSMO-RS geometries of the highest populated conformers given with Boltzmann weight of all investigated cycloalkanes up to cycloicosane. The initial number of conformers equals the number of conformers provided by `crest_combi` at the GFN-FF/GFN2-xTB level while the refined number refers to the final ensemble obtained from `censo` at DFT level. The symmetry of the lowest free energy conformer is provided in Schoenflies notation.

The variation of the ^{13}C chemical shift for the investigated cycloalkanes is relatively small with ~6 ppm. However, because their chemical structure is very similar, the observed subtle changes for different ring sizes suggest that conformational effects play a major role and that their account is theoretically challenging. In fact, a fully automated QM-based study of this type has never been performed for the NMR spectra of such large and immensely flexible systems with hundreds of contributing conformers. The proton spectra of the cycloalkanes studied (see SI-Figure A16.1) show only minor changes for the larger cycloalkanes, and thus only the more informative carbon spectra will be discussed.

Cycloalkanes and their properties are usually categorized into small, common, medium and large rings.[639] Cyclobutane is classified as a small ring showing large deviations from the usual tetrahedral angles with corresponding high angle- and torsion-strain, resulting in a low chemical shift.[640] Significantly populated conformers are shown in the supporting information, along with a discussion of conformers identified from previous experimental or computational studies (see the supporting information). Common rings with 5 to 7 carbon atoms exhibit outside facing hydrogens, small deviation from the tetrahedral bonding angle and cyclohexane is usually referred to as a strain-free.[641] These common rings in $CHCl_3$ exist in a single conformation at room temperature (cf. Figure A15.4). The reduced strain (see SI-Figure A16.5) compared to the small rings is evident in the increased shifts reaching a maximum for cycloheptane. For medium-sized rings from octane to undecane, steric interaction occur due to hydrogens facing inside, called transannular interaction. Cyclononane is the first compound with more than a single dominantly populated conformer and the lowest free energy triangular [333] conformation with D_3 symmetry is populated by only 47.1 %. Cyclodecane with an even number of ring atoms can be superimposed on a diamond lattice indicating nearly ideal tetrahedral angles and therefore low strain. However, some transannular interaction between competing inward facing hydrogen atoms can be observed. The cyclododecane conformations are exposed to less stress than in cycloundecane and therefore it exhibits the largest high-field shift within the medium-sized rings (see SI-Figure A16.5). Large rings with more than 12 carbon atoms slowly start to resemble their open-chain alkane analogues.[637] This is also reflected in the geometries of the highly populated conformers, which, for example in $C_{16}H_{32}$ and $C_{20}H_{40}$, adopt rectangular structures and thus the inner methylene groups adopt geometries similar to linear *n*-alkanes. The population of cycododecane is strongly dominated by the square [3333][620] conformation with D_4 symmetry, while other conformers are significantly higher in free energy. Except for the small rings, cyclododecane represents the system with the strongest high-field shift of all investigated compounds. From cyclotridecane on, a trend towards low-field chemical shifts can be observed, which is only significantly interrupted by cyclotetradecane in its preferred rectangular, diamond-lattice type conformation.[620]

Overall, the characteristic experimentally determined shift profile for different ring sizes is almost perfectly reproduced by our theoretical workflow. This is particularly impressive for the larger cycloalkanes where up to almost 1000 conformers are initially considered and many structures contribute to the average. To illustrate the difficulty in the correct description of the ^{13}C-NMR spectrum of larger rings, the range of chemical shifts within the final ensemble is shown by gray error bars in Figure A15.3, spanning a range of ~7 ppm which is similar to the overall variation with rings size. Thus, a randomly chosen structure will most likely not yield a satisfactory result, and the need for an extensive and re-

liable conformer search becomes apparent. Furthermore, our data suggest that the sole consideration of the lowest-lying conformer (represented by green crosses) in Figure A15.3 is also insufficient, which highlights the influence of higher lying conformers in the CRE. Evidently, smaller (high-field) shifts are regularly obtained for rings with an even number of methylene groups than for odd-numbered ones. This also indicates a conformational effect and it can be deduced from the structure ensembles that the odd-numbered homologous exhibit increased bond-, angle- and torsion strain due to unfavorable positioning of the additional methylene group.

From cyclopentadecane on, with 47 low-lying conformations, an increasing number of significantly populated conformers is observed, reaching a maximum of 205 structures at cyclononadecane. In such cases, the spectrum is no longer determined by a single or few dominant structures but rather by an ensemble with dozens of similarly populated minima. In 1990 Saunders et al.[642] compared the results of several conformational space exploration methods for cycloheptadecane and by combining all search results they identified 262 gas-phase conformers within an energy window of 3 kcal mol^{-1}.[642] This is in good agreement with the 168 populated solution phase conformers found here within a free energy threshold of 2.5 kcal mol^{-1}.

The computation time strongly depends on the number of relevant conformers and is mostly spent for geometry optimization (see SI-Figure A16.3 S3). For the most flexible cyclononadecane, the initial conformer search took about two hours and the refinement including NMR property calculation required about 50 hours computation time on 28 cores (Intel(R) Xeon(R) CPU E5-2660 v4 @ 2.00GHz). Most of the calculations for more rigid structure finish within a day or even a few hours computation time.

When comparing our CRENSO workflow to other approaches such as NAMFIS or CANDLE, the main focus is also on generating the most complete (low-energy) conformer ensemble possible. However, unlike NAMFIS, our conformer search explores the SQM/FF PES without experimental constraints that could spuriously influence the search results and the final Boltzmann weights of the solution ensemble are computed *ab initio* and not determined from a fit to experimental data. Similar to CANDLE, we need the experimental chemical shifts of protons or carbon atoms only for the final comparison step, which is more convenient than a full spectral assignment required for NOE or RDCs. With CRENSO, we try to model all relevant physical processes on a as high as computationally possible level and quantum mechanical basis while maintaining an automated workflow with low computational cost. Thus, structures are optimized with implicit solvation and we accurately account for all free energy terms including (intra-molecular) London dispersion with a proper description of chemical shifts. This work extends the toolbox of existing analytical methods for nonrigid macrocyles in solution and in principle these methods can complement each other, i.e., by combining and further completing differently generated ensembles or employing the created ensemble in a NAMFIS like evaluation.

In this communication we propose a robust, automated approach of identifying populated hydrocarbon macrocycle conformations in chloroform solution. The workflow is generally applicable to more complicated compounds with typical functional groups or other solvents.

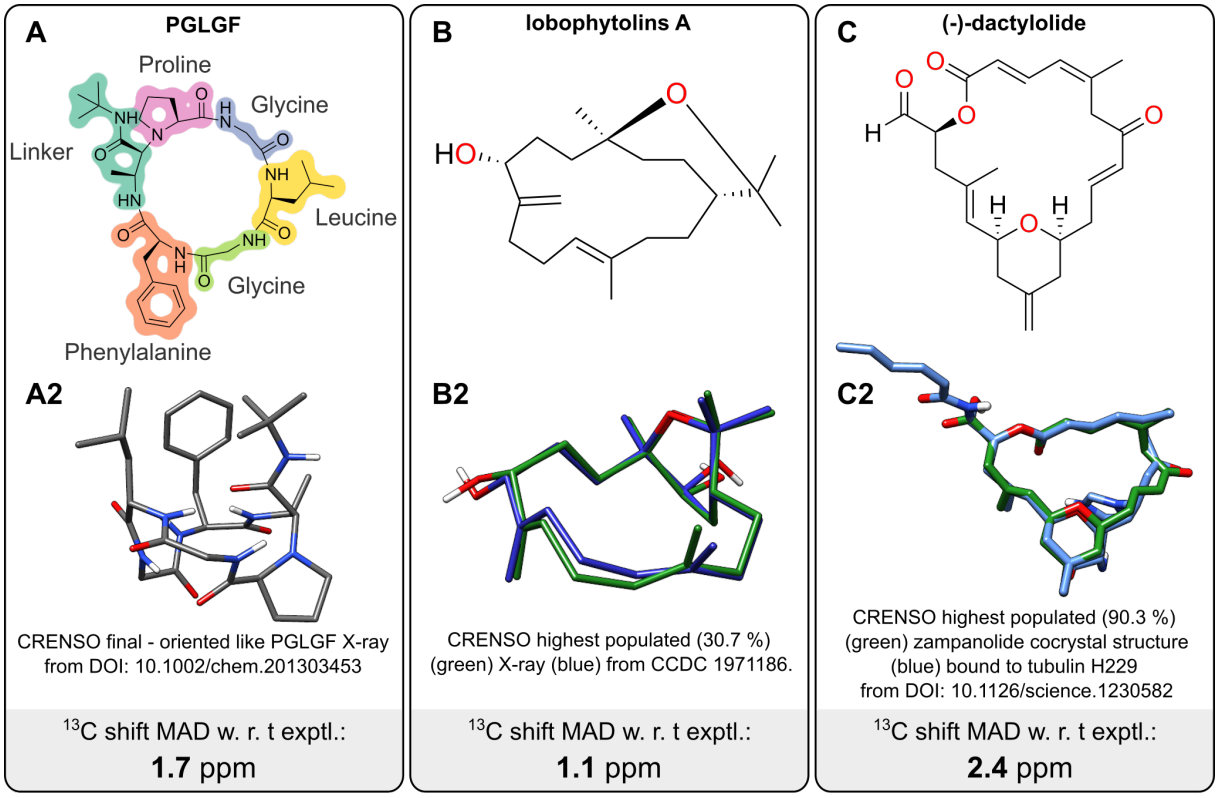


Figure A15.5.: Shown are the results of the CRENSO workflow on three macrocyclic compounds containing functional groups. A displays the cyclic peptide PGLGF[643] whose highest populated conformer is shown in A2 and visually corresponds to the crystal structure of PGLGF from Ref. 644. Part B examines lobophytolin A[645], which has 20 populated conformers and the overlay of the lowest lying conformer and the crystal structure is shown in B2. In C, the Lewis structure of (-)-dactylolide[646] is shown and C2 includes the overlay of the highest populated conformer and the cocyrstal structure of the related natural product zampanolide[647]. For all systems, the mean absolute deviation of the carbon chemical shifts with respect to the experiment is reported, which shows good overall agreement.)

As a proof of principle this is demonstrated here (see Fig. A15.5) for a cyclic peptide PGLGF[643], and the two natural products lobophytolins A[645] and (-)-dactylolide[646]. The results are discussed in more detail the supporting information and show impressive overall agreement expressed by the mean absolute deviations between the experimental and theoretical carbon chemical shifts ranging from 1.0-2.4 ppm. In addition, the experimentally identified structures are found in the final CRENSO ensembles as lowest free energy conformers in all three cases highlighting the accuracy and generality of our approach.

A15.3. Conclusion

Cyclic hydrocarbons were solely chosen as model systems mainly because good experimental reference data are available and the intramolecular interaction motifs are treated reliably even by low-level theoretical methods. For more complicated, e.g., polypeptide or drug structures with more complex PES, the initial search step at the lowest theoretical level becomes crucial which is a topic of ongoing research. Nevertheless, this and a related study[634] provide strong evidence that our methods can explore accurately the conformational space of very flexible macrocycles. This is convincingly shown here by comparison of computed and experimental solution phase ^{13}C-NMR spectra. Herein our recently published `crenso` workflow efficiently samples the PES, refines the structure ensemble and automatically performs all NMR property calculations. We showed that identifying conformer ensembles with dominant populations is essential for an accurate prediction of NMR spectra, and that considering only random or even the lowest lying conformer is insufficient. In turn, the excellent agreement between theory and experiment enables to deduce the preferred conformations under ambient measurement conditions. The calculations with our sophisticated mostly quantum chemical workflow can be performed on a normal workstation computer within hours with minimal user input and all our codes are freely available. This opens a route for highly accurate NMR simulations for more diverse and challenging macrocycles in the future.

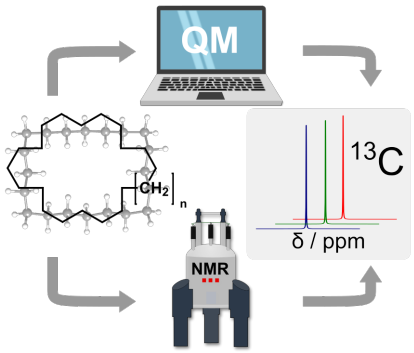
Acknowledgements

The German science foundation (DFG) is gratefully acknowledged for financial support in the framework of the Gottfried Wilhelm Leibniz prize to S.G. The authors thank P. Pracht for helpful discussions.

Conflict of interest

The authors declare no competing interests.

Supplementary information is available for this paper containing additional computational details.[648]



A16. Supporting Information to Chapter 10

Appendix A16 contains:

- Additional technical details of the calculations
- coordinates in xmol format

Linear regression scaling factors for obtaining computed chemical shifts

Chemical shifts ($\delta_{DFT,corr}$) are obtained by subtracting computed DFT chemical shielding constants ($\sigma_{DFT,raw}$) from a fitted reference chemical shielding constant ($\sigma_{fitted_reference}$) (intercept) and applying a scaling factor ($scaling_factor$) (slope), see eq. A16.1. The scaling factor and reference chemical shielding constant are obtained from a linear regression fit between computed chemical shielding constants ($\sigma_{DFT,raw}$) and experimentally measured chemical shifts (δ_{exp}). The linear regression correction is applied to reduce systematic errors between theory and the experiment which are, for example, the result of errors in the DFT treatment, solvation or rovibrational effects. The intercept from the linear regression represents the reference isotropic shielding value of a common standard, typically TMS.

$$\delta_{DFT,corr} = scaling_factor * (\sigma_{fitted_reference} - \sigma_{DFT,raw}) \tag{A16.1}$$

For the linear regression, a database of small organic molecules compiled with experimentally determined ^{1}H and ^{13}C data is taken from the supporting information in Ref. 26, which can also be found at the chemical shift repository[638]. The database is divided into the TESTSET and a smaller PROBESET for cross validation. The employed database containing small organic molecules and experimental references was first developed by Rablen[649] and augmented by Lodewyk, Siebert and Tantillo[26].

The following molecules were used in the the TESTSET:
1,1,1-trichlorethane, 1,1,1-trichloroacetone, 1,1-dichloroethylene, 1-chloro-2-methylpropene, 2-butyne, 2-chloro-2methlypropane, 2-chloropropane, 2-chloropropionitrile, 2-cyclohexenone, 2-cyclopentenone, 2-methyl-2-butene, 2-methyl-2-nitropropane, 2-methylpropanenitrile, 2-methylpropene, 3,3-dimethyl-1-butene, 3,4-dichloro-2_5h-furanone, 3-butyne-2-one, 3-chloro-propene, 4-methylthiazole, acetaldehyde, acetone, acetonitrile, acetylchloride, benzene, butyrolactone, chlorobenzene, chlorodimethylsilane, chloroethane, cyclobutane, cyclobutanone, cyclobutene, cyclohexane, cyclohexanone, cyclopentane, cyclopentanone, cyclopentene, cyclopropane, cyclopropanone, dimethylsulfoxide, fluorobenzene, furan, furfural, isoxazole, methanol, methylacetate, methyltert-butylether, nitrobenzene, nitroethane, nitromethane, N-methylpiperidine, N-methylpyrrole, N-methylpyrrolidine, N,N-dimethylformamide,

oxetane, oxirane, *p*-benzoquinone, propionitrile, pyrazine, pyridazine, pyridine, pyrimidine, tetrahydrofuran, tetrahydropyran, thiirane, thiophene, trans-1,2-dichloroethylene, trichloroethylene, trimethylacetonitrile.

The following molecules were used in the the PROBESET:
1,3,5-triazine, 1,3-dioxane, 2,5-dihydrofuran, 2-methylpyrazine, 2-methylthiophene, allylisothiocyanate, crotonaldehyde, diketene, dimethylacetal, dimethylcyanamide, E-3,4-dimethyl-2-pentene, ethylcyanoformate, ethylenecarbonate, ethylsulfide, fumaronitrile, isoprene, maleicanhydride, malononitrile, methylenecyclobutane, methylisothiocyanate, pivalaldehyde, pyridine, quadricyclane, tert-butylisocyanide.

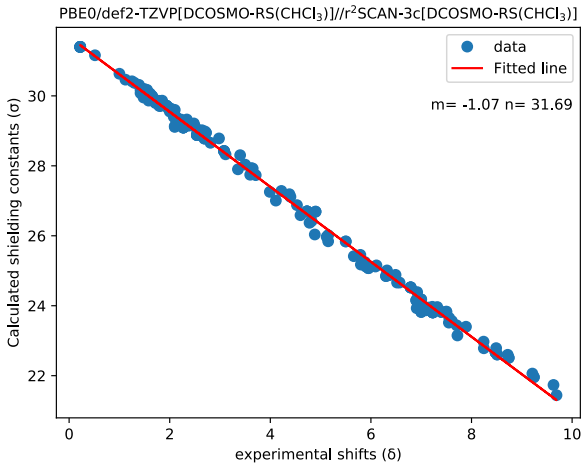
The computed chemical shielding constants of the molecules in the TESTSET and PROBESET were obtained in the following way: Conformers, rotamers, chemical- and magnetical equivalencies were determined in a `crest` run with implicit solvation ($CHCl_3$). The resulting conformer rotamer ensembles (CRE) are sorted by `censo` while improving both the description of free energy (G) and the molecular solution phase structure. Shielding constants are calculated for the dominantly populated conformers. Through consideration of the chemical equivalencies (identified during the conformer generation in CREST) and Boltzmann factors the shielding constants are properly thermally averaged over the CRE. The following method combinations were applied in the `censo` calculations:

Geometry:	r^2SCAN-3c[DCOSMO-RS($CHCl_3$)]
Solvent:	$CHCl_3$
T in K:	298.15
Electronic Energy ΔE:	r^2SCAN-3c
Thermal corrections to ΔG^T_{RRHO} :	GFN2-xTB[GBSA($CHCl_3$)]
Solvation free Energy ($\Delta\delta G_{solv}$):	COSMO-RS Version 19 (Param: ctd = BP_TZVP_19.ctd)
Chemical shielding constants ($\sigma_{DFT,raw}$):	PBE0/def2-TZVP[DCOSMO-RS($CHCl_3$)]

The linear regression parameters are calculated from the TESTSET and analyzed using the PROBESET. Linear regression parameters for ^1H and ^{13}C for the method combination of PBE0/def2-TZVP[DCOSMO-RS($CHCl_3$)]//r^2SCAN-3c[DCOSMO-RS($CHCl_3$)] are evaluated below.

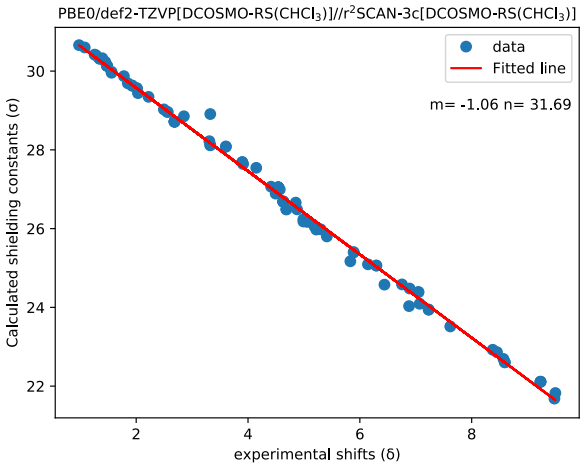
^{1}H-NMR Shifts

^{1}H-NMR Shifts
TESTSET (68 molecules)



R^2:	0.9975
Scaling factor:	0.933
Fitted reference:	31.686
(using TESTSET-scaling factor):	
(and fitted reference)	
unscaled RMSD (δ) :	0.387
scaled RMSD (δ) :	0.129

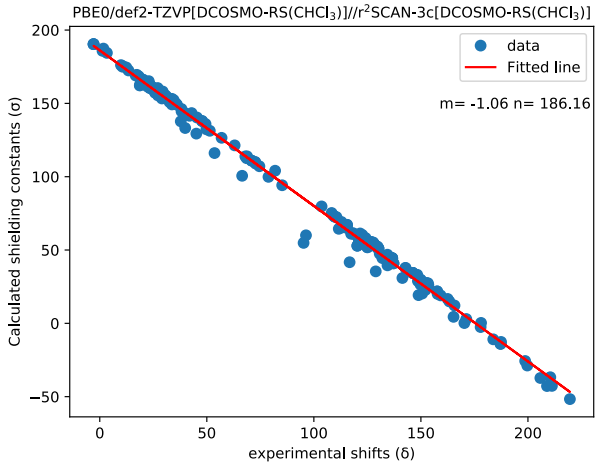
^{1}H-NMR Shifts
PROBESET (24 molecules)



R^2:	0.9963
Scaling factor:	0.945
Fitted reference:	31.690
(using TESTSET-scaling factor):	
(and fitted reference)	
unscaled RMSD (δ) :	0.327
scaled RMSD (δ) :	0.158
(using PROBE-scaling factor):	
(and fitted reference)	
unscaled RMSD (δ) :	0.330
scaled RMSD (δ) :	0.144

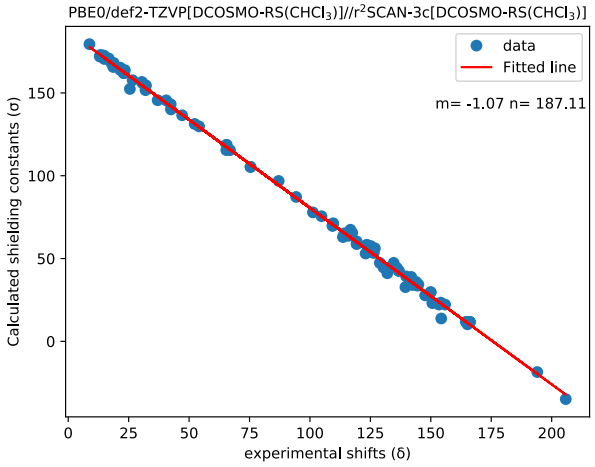
^{13}C-NMR Shifts

^{13}C-NMR Shifts
TESTSET (68 molecules)



R^2:	0.9427
Scaling factor:	0.942
Fitted reference:	186.156
(using TESTSET-scaling factor):	
(and fitted reference)	
unscaled RMSD (δ) :	7.948
scaled RMSD (δ) :	4.217

^{13}C-NMR Shifts
PROBESET (24 molecules)



R^2:	0.9386
Scaling factor:	0.938
Fitted reference:	187.106
(using TESTSET-scaling factor):	
(and fitted reference)	
unscaled RMSD (δ) :	6.637
scaled RMSD (δ) :	2.215
(using PROBE-scaling factor):	
(and fitted reference)	
unscaled RMSD (δ) :	7.391
scaled RMSD (δ) :	2.125

The data points clearly deviating from the fitted line stem from molecules with chlorine atoms.

Cycloalkane chemical shifts

^{13}C-chemical shifts

Table A16.1.: ^{13}C chemical shifts of the investigated cycloalkanes in $CHCl_3$. a) cyclononane chemical shift extrapolated from cyclooctane and cyclodecane, using references Ref. 636,637; b) cycloheptadecane chemical shift extrapolated from cyclohexadecane and cyclooctadecane using Ref. 637; c) unscaled theoretical chemical shifts obtained with tetramethylsilane (TMS) as standard; d) theoretical values by applying linear regression correction ($\delta_{DFT,corr}$) (used in the main text). All computed chemical shifts given in ppm are evaluated at 298.15 K

ringsize (# of -(CH_2)-)	experiment	theory unscaled[c]	theory scaled[d]
4	22.4[650]	25.9	22.3
5	25.8[650]	30.0	26.3
6	27.0[637]	31.1	27.2
7	28.5[636,651]	32.3	28.3
8	26.8[637]	30.6	26.7
9	25.9 [a]	29.2	25.4
10	25.1[637]	28.6	24.9
11	26.3[637]	30.2	26.4
12	23.8[637]	27.1	23.5
13	26.2[637]	30.0	26.2
14	25.2[637]	28.7	25.0
15	27.0[637]	30.5	26.6
16	26.9[637]	30.3	26.5
17	27.2 [b]	31.3	27.5
18	27.5[637]	31.3	27.5
19	28.1[637]	32.2	28.3
20	28.0[637]	32.0	28.1

^1H-chemical shifts

Table A16.2.: ^{1}H chemical shifts of the investigated cycloalkanes in $CHCl_3$. a) chemical shift measured in neat; b) chemical shift measured in $CHCl_3$ at 38 °C [636,637]; c) unscaled theoretical chemical shifts obtained with tetramethylsilane (TMS) as theoretical standard; d) scaled theoretical shifts by applying linear regression correction ($\delta_{DFT,corr}$). All computed chemical shifts given in ppm are evaluated at 298.15 K

ringsize (# of -(CH_2)-)	experiment	theory unscaled[c]	theory scaled[d]
4	1.980[a] [636,652]	1.976	2.021
5	1.506[a] [636,652]	1.586	1.632
6	1.431[b] [637]	1.466	1.512
7	1.526[a] [636,652]	1.566	1.612
8	1.531[b] [637]	1.566	1.612
9	1.527[a] [636,652]	1.536	1.582
10	1.514[b] [637]	1.566	1.612
11	1.441[b] [637]	1.476	1.522
12	1.344[b] [637]	1.386	1.432
13	1.348[b] [637]	1.356	1.403
14	1.331[b] [637]	1.376	1.422
15	1.333[b] [637]	1.356	1.403
16	1.319[b] [637]	1.346	1.393
17	1.316[a] [636,652]	1.356	1.403
18	1.310[b] [637]	1.346	1.393
19	1.307[b] [637]	1.346	1.393
20	1.301[b] [637]	1.336	1.383

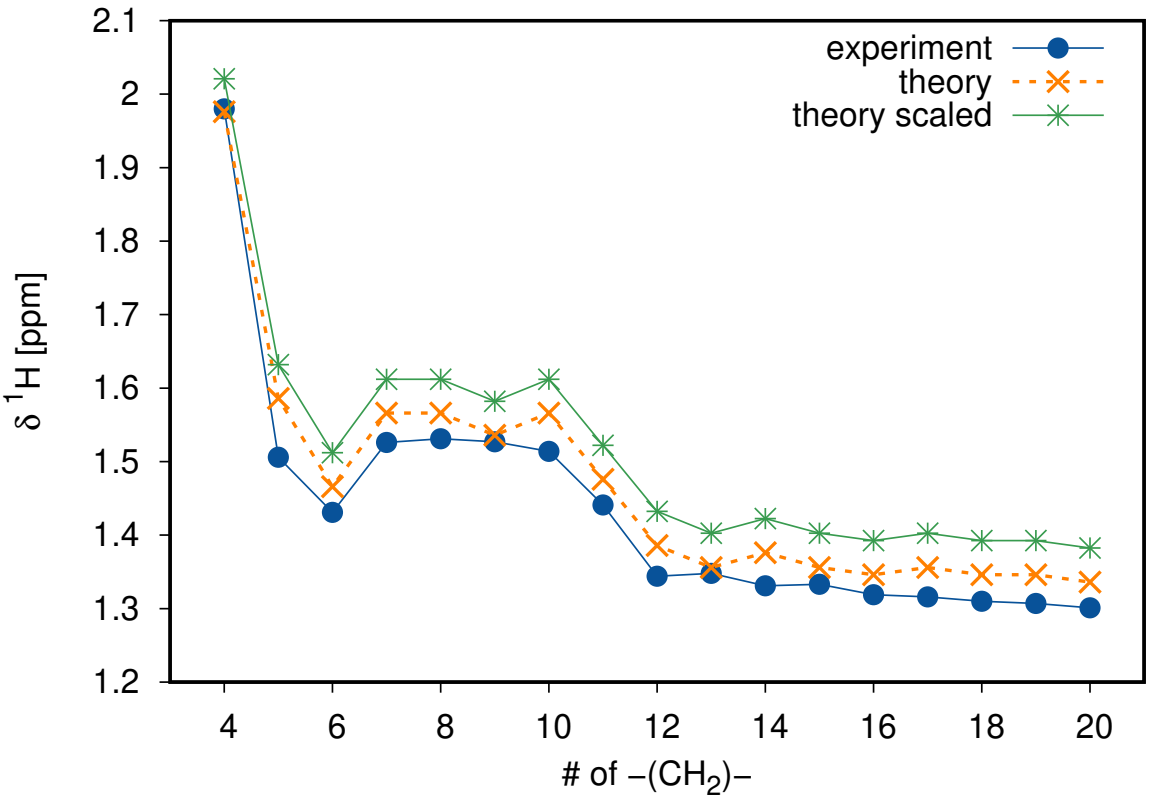


Figure A16.1.: Computed and experimental ^{1}H chemical shifts of the investigated cycloalkanes in $CHCl_3$. Information on experimental data is provided in Table A16.2. Blue points represent the experimental measurements, orange crosses the theoretically derived proton chemical shifts with TMS as reference and computed chemical shifts corrected by linear scaling are shown in green.

Number of conformers in each `crenso` step

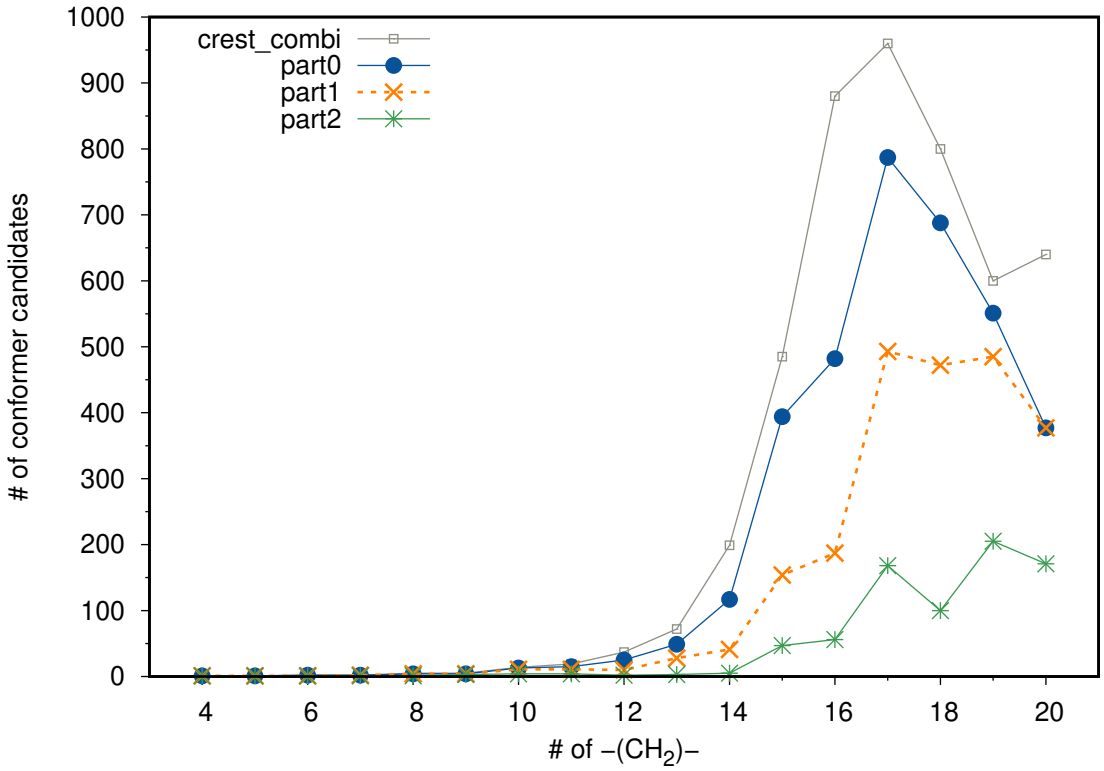


Figure A16.2.: Number of conformer candidates remaining after each of the 'crenso steps': crest_combi (conformer ensemble generation), CENSO:part0 (cheap pre-screening), CENSO:part1 (pre-screening), CENSO:part2 (optimization and free energy (G) evaluation).

Timings

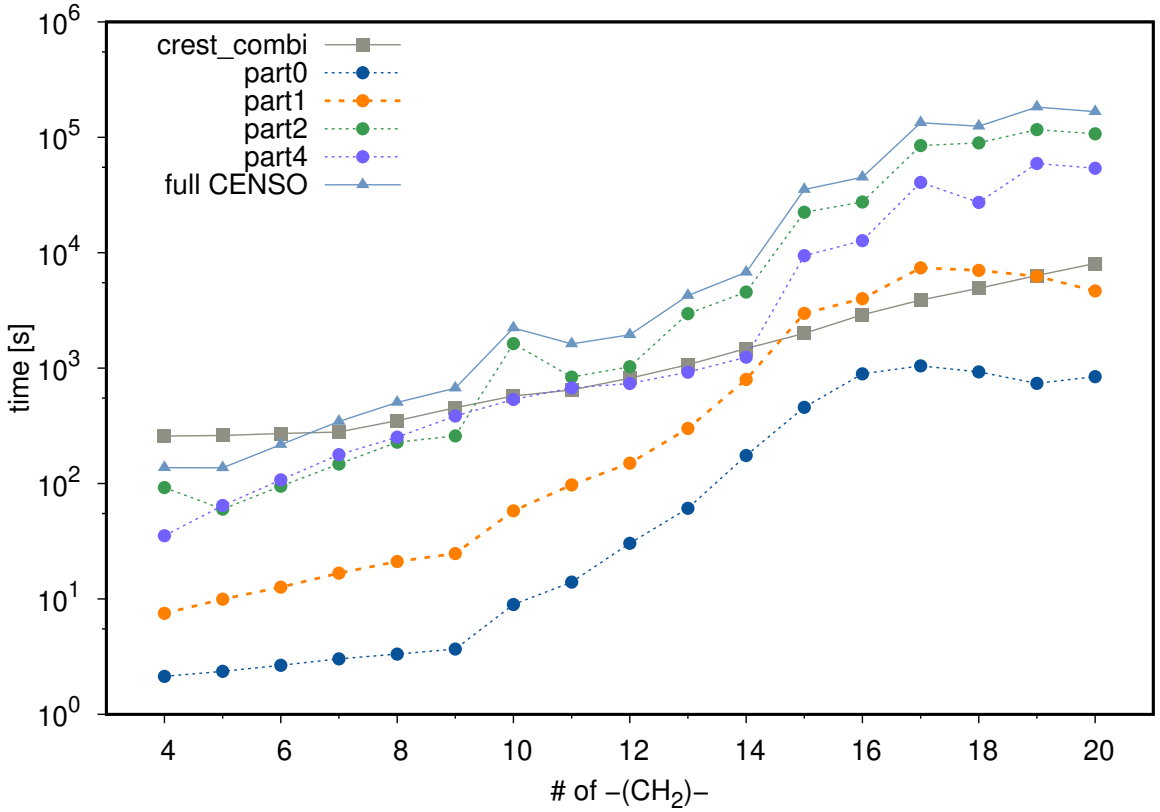


Figure A16.3.: Timings of the sorting steps: crest_combi, CENSO:part0, CENSO:part1, CENSO:part2, CENSO:part4. Evaluated at Intel(R) Xeon(R) CPU E5-2660 v4 @ 2.00GHz using 28 cores (see Table A16.3 for employed threshold-settings).

Influence of symmetry on final chemical shifts

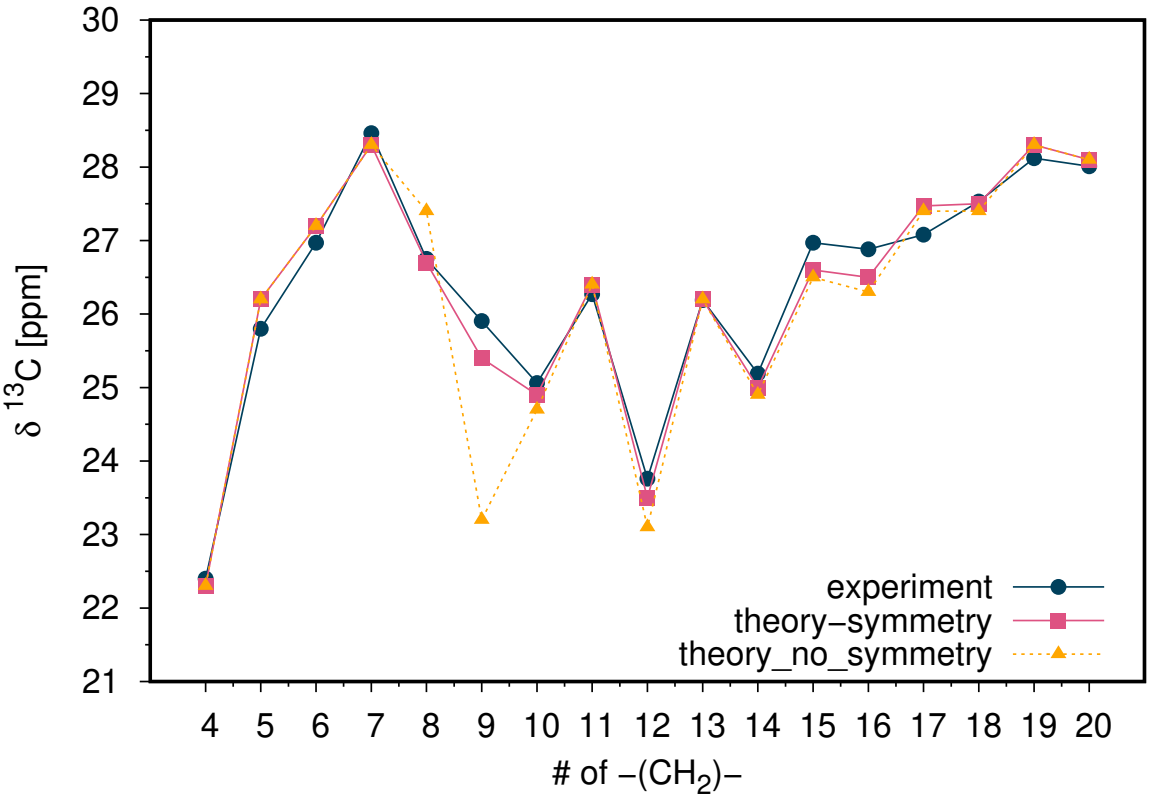


Figure A16.4.: Calculated chemical shifts when symmetry in the thermostatistical $G_{mRRHO}^{298.15K}$ treatment is considered or ignored, i.e., the influence of the Boltzmann factor on the ensemble averaged property.

Technical settings applied in crenso

	CENSO general information:
CENSO version	1.0.8
Temperature	298.15 K
Solvent	$CHCl_3$
	PART0 - cheap pre-screening
Energy	b97-d/def2-SV(P) + GCP
Energy_settings	scfconv 5, grid 1 using D3(0) instead of D3(BJ)
G_mRRHO	not included
G_solv	ALPB[gfn2]
Geometry	GFNn-xTB (input geometry)
Threshold	6.0 kcal/mol
main QM code	TM
	PART1 - pre-screening
Energy	r2scan-3c
Energy_settings	grid m4 scfconv 6
G_mRRHO	GFN2[alpb]-bhess SPH
G_solv	COSMO-RS-normal[r2scan-3c] param= 19-normal
Geometry	GFNn-xTB (input geometry)
Threshold	4.5 kcal/mol
main QM code	TM
	PART2 - optimization
Energy	r2scan-3c
Energy_settings	grid m4 scfconv 6
G_mRRHO	GFN2[alpb]-bhess SPH
G_solv	COSMO-RS-normal[r2scan-3c] param= 19-normal
Geometry	r2scan-3c[DCOSMORS] @optlevel: normal using grid m4 scfconv 6
Threshold	Opt_limit: 2.5 kcal/mol, Boltzmann sum threshold: 95.0 %
main QM code	TM
	PART4 - NMR mode
Energy	using Energy from part2 - optimization
Energy_settings	see part2 - optimization
G_mRRHO	using G_mRRHO from part2 - optimization
G_solv	using G_solv from part2 - optimization
Geometry	r2scan-3c[DCOSMORS] @optlevel: normal
Threshold	Boltzmann sum threshold: 95.0 %
main QM code	TM
NMR shielding constants	pbe0/def2-TZVP[DCOSMO-RS] (TM) using grid m5 scfconv 7
NMR coupling constants	pbe0/def2-TZVP[DCOSMO-RS] (TM) using grid m5 scfconv 7

Table A16.3.: Technical settings as applied in the crenso calculation using (*crenso -nmr $CHCl_3$ -l3*), but with slightly increased thresholds for evaluation purposes.

Ensemble strain energy

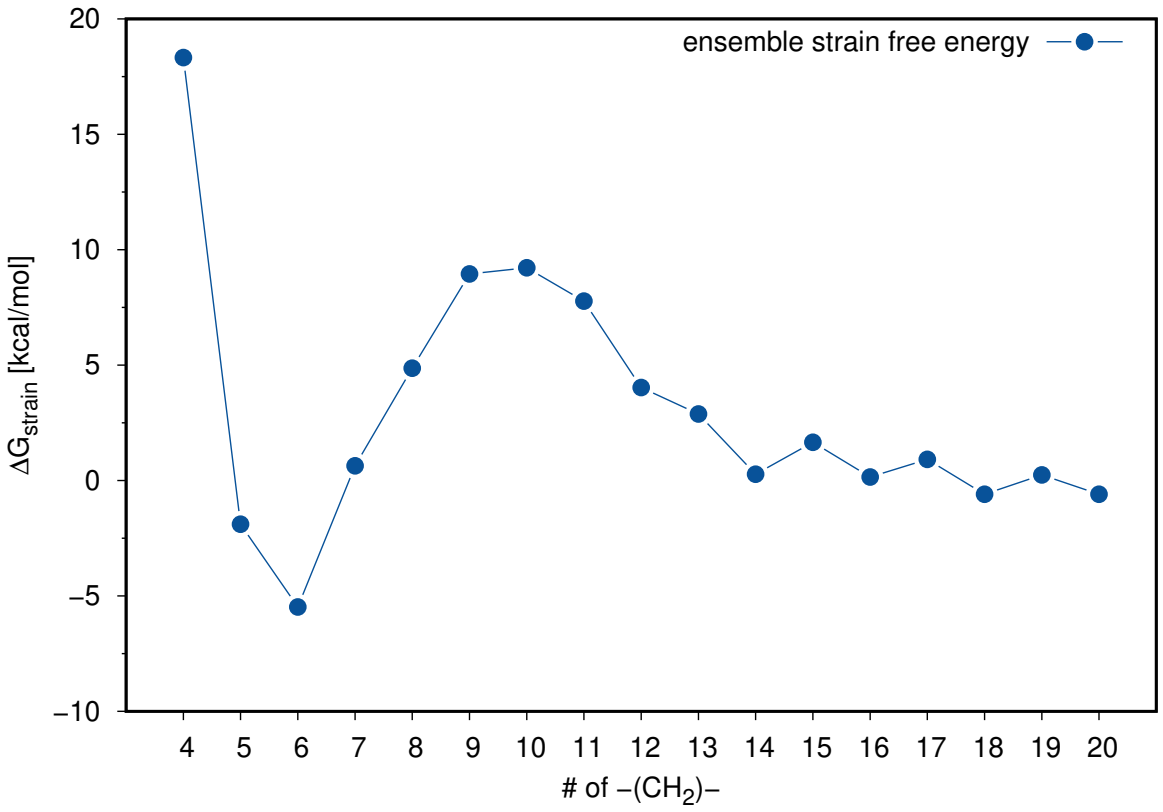


Figure A16.5.: Ensemble averaged strain free energy, calculated with implicit $CHCl_3$ at 298.15 K as a function of the cycloalkane ring size.

The ensemble averaged Gibbs strain energy is calculated using the ensemble free energies of the cycloalkanes (c.f. Table A16.3(part2)). Methlyene is used as a computational group equivalent (pioneered by Wiberg[653] and Schleyer[654]) and the free energy of the methylene group equivalent is obtained from Eq. A16.2:

$$G(CH_2) = (G^{ensemble}_{n-heptane} - G^{ensemble}_{n-propane})/4 \tag{A16.2}$$

The ensemble averaged Gibbs strain energy is calculated from Eq. A16.3 and this approach is adapted from Ref.[655].

$$\Delta G_{strain} = G^{ensemble}_{C_xH_{2x}} - x \cdot G(CH_2) \tag{A16.3}$$

The ensemble averaged Gibbs strain energies are shown in Fig. A16.5 and the values are discussed solely in terms of overall trend, with high values representing greater strain. Cyclohexane exhibits the lowest strain and cyclodecane the largest, if cyclcobutane is disregarded. The overall strain decreases

for larger cycloalkanes, approaching the limit of zero for an *n*-alkane.

Cycloalkane discussion with respect to experimental or theoretical reference

Cyclobutane

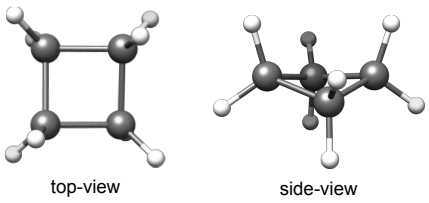


Figure A16.6.: Highest populated conformer of cyclobutane.

Cyclobutane is identified in its puckered (wing-shaped) ring form as a single conformer[656]. This puckered conformation is more stable compared to the higher energy planar conformation and reduces torsional strain (see also Figure A16.5). The puckered conformation has been experimentally identified by Dunitz and Schomaker using electron diffraction.[657]

Cyclopentane

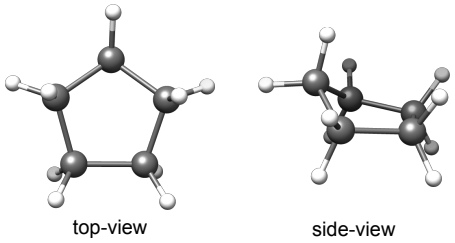


Figure A16.7.: Highest populated conformer of cyclopentane.

Like cyclobutane the dominant conformation of cyclopentane is non planar and is present in a single envelope conformation. The ring strain is reduced considerably compared to cyclobutane and the bond angles come closer to the ideal tetrahedral angle of 109.5°. The envelope conformation was confirmed by Aston, Schumann, Fink and Doty in 1941[658].

Cyclohexane

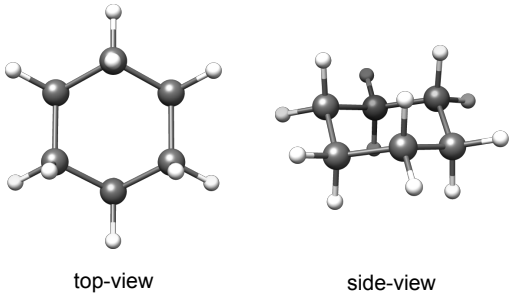


Figure A16.8.: Highest populated conformer of cyclohexane.

The lowest energy conformer of cyclohexane is the chair conformation which is attested by Raman- and IR spectroscopy as well as electron diffraction.[659]. The cyclohexane ring is considered to be strain free and two type of carbon-hydrogen bonds can be identified (axial and equatorial).

Cycloheptane

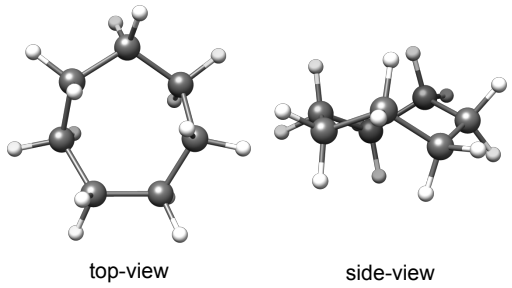


Figure A16.9.: Highest populated conformer of cycloheptane.

Wiberg computed the twist-chair TC conformation to have the lowest free energy at room temperature, which is matched by our CRENSO workflow.[660]

Cyclooctane

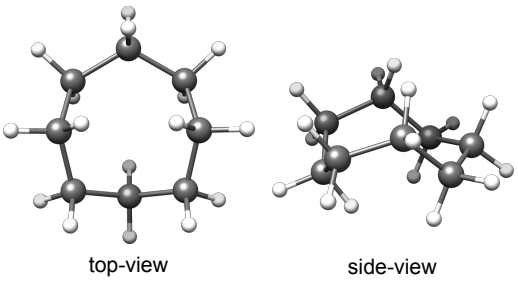


Figure A16.10.: Highest populated conformer of cyclooctane.

Anet and Basus showed that cyclooctane's dominant conformation is the boat-chair (BC) form at room temperature.[661] The BC conformation in C_s symmetry is correctly identified by our algorithm.

Cyclononane

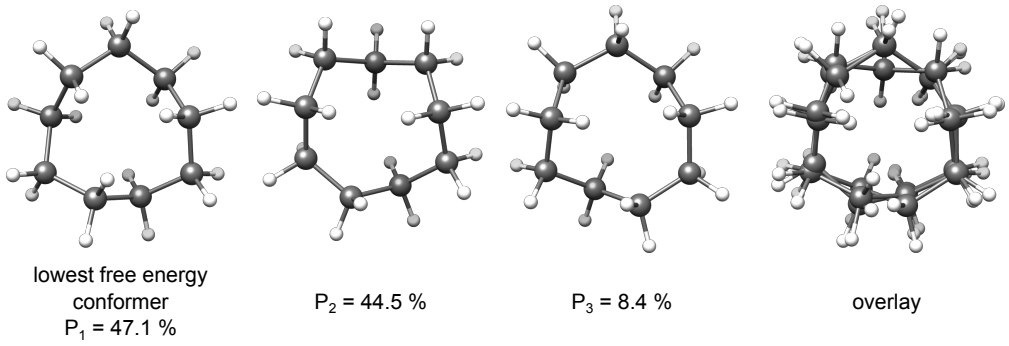


Figure A16.11.: Highest populated conformers of cyclononane.

^{13}C NMR measurements at -162 °C from Anet et al. with two resolved signals confirmed Allinger's strain-energy results that the twist boat chair (TBC) conformation in D_3 symmetry is the dominant conformation.[662] As seen in Fig. A16.11 this is in perfect agreement with our calculations, with the twist-chair-boat (TCB) conformation (P_2) being higher in free energy.

Cyclodecane

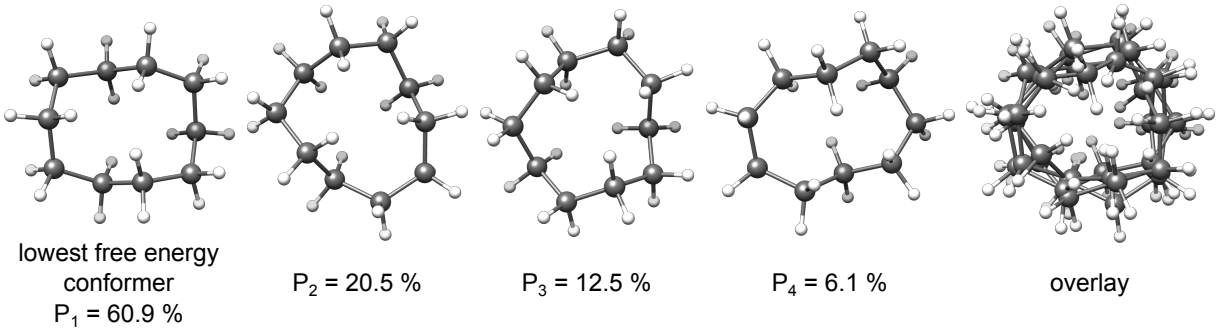


Figure A16.12.: Highest populated conformers of cyclodecane and overlay of all shown structures.

The lowest conformation is the boat-chair-boat (P_1) conformation as identified from NMR measurements[663] and confirmed by theoretical investigation[660]. The second conformer has been identified as the (TBCC) conformation in C_2 symmetry, followed by the (TBC) conformation.[660]

Cycloundecane

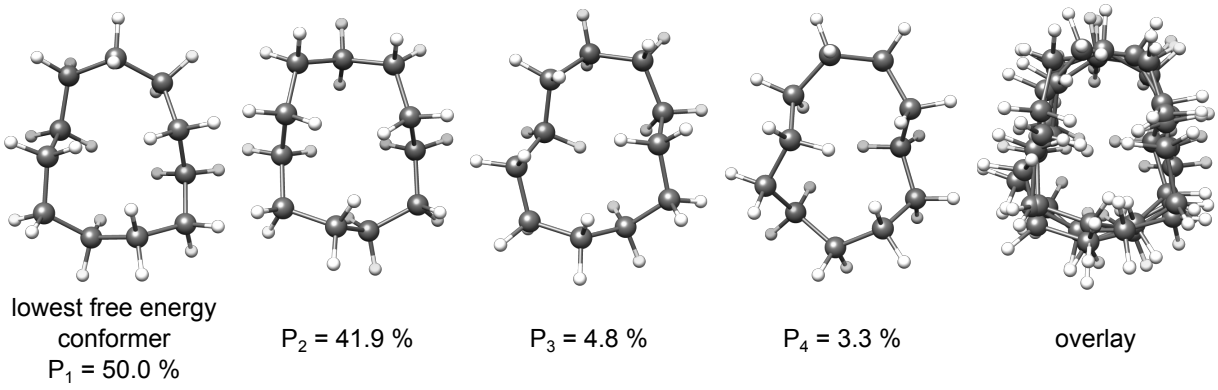


Figure A16.13.: Highest populated conformers of cycloundecane and overlay of all shown structures.

Pawar et al. showed cycloundecane to exist as a mixture of two conformations [12323] (59%) and [335] (41 %) at -183.1 °C.[664,665] In our calculations in $CHCl_3$ at room temperature the ordering inverses and we identify [335] (P_1) as highest populated followed by [12323] (P_2).

Cyclododecane

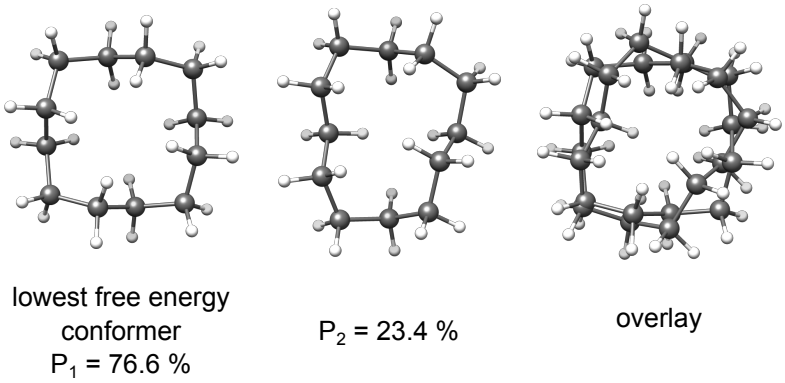


Figure A16.14.: Highest populated conformers of cyclododecane and overlay of all shown structures.

Crystal structure[666] and low temperature carbon NMR measurements[667] reveal a square conformation [3333] in Dale's nomenclature with D_4 symmetry, consistent with the lowest free energy conformation we have identified. Anet identified the [2334] conformation (C_1) through strain-energy calculations as next viable candidate, which is our P_2 second conformer.[668]

Cyclotridecane

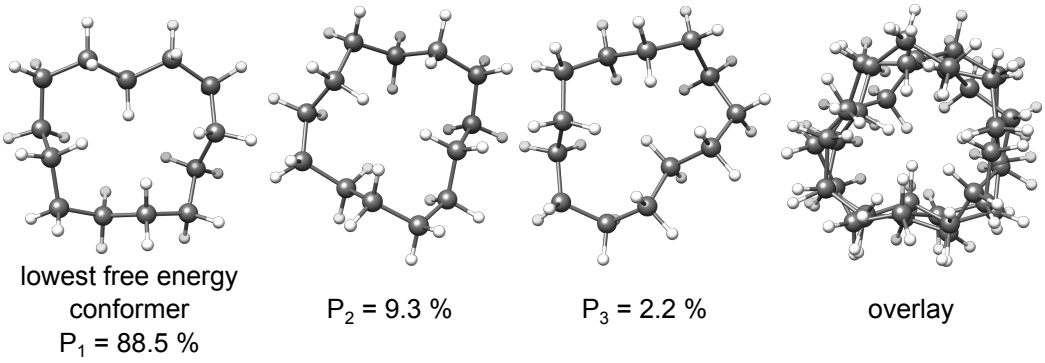


Figure A16.15.: Highest populated conformers of cyclotridecane and overlay of all shown structures.

Conformer identification for cyclotridecane, an odd-membered macrocycle, by NMR measurements is nontrivial as even at -135 °C the pseudo-rotation leads to a single signal. For this reason comparisons were made based on crystal structures of similar compounds, which match our lowest free-energy conformer very well.[619,665] The experimentally (X-ray data) derived population estimate on the dominant conformation is with 80 % very close to our evaluation in $CHCl_3$ P_1= 88.5 %.[619,665]

Cyclotetradecane

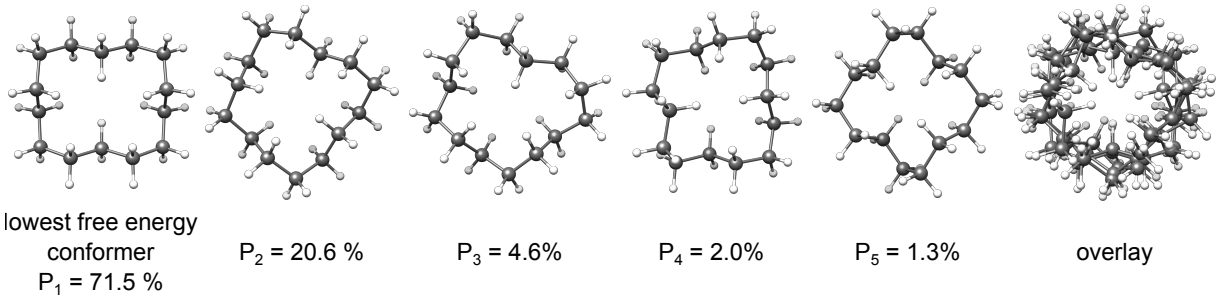


Figure A16.16.: Highest populated conformers of cyclotetradecane and overlay of all shown structures.

Low temperature carbon NMR measurements[667] of cyclotetradecane (-132 °C) validate the rectangular diamond lattice conformation identified as our lowest lying conformer. Followed by the [3344] conformation, correctly identified as the second lowest lying conformer.[639]

Cyclopentadecane

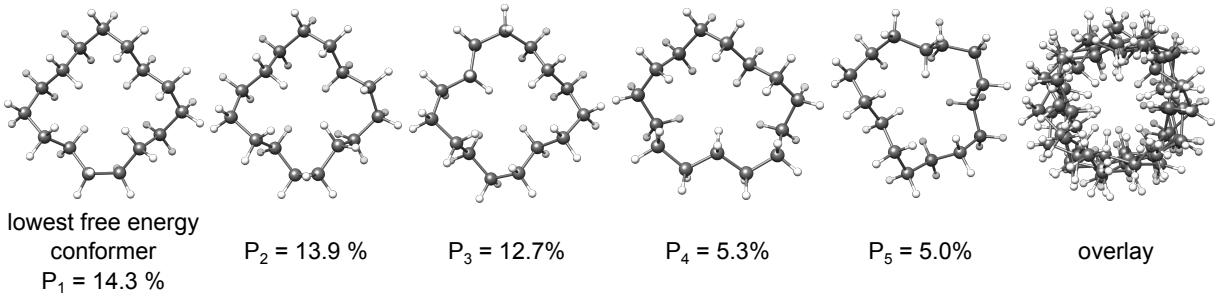


Figure A16.17.: Highest populated conformers of cyclopentadecane and overlay of all shown structures.

Anet et al. predicted several stable cyclopentadecane conformations at room temperature, where the [13443] conformation agrees with our quinquangular lowest free energy conformer.[669] The conformer P_4 closely resembles the predicted [13434] conformation.

Cyclohexadecane

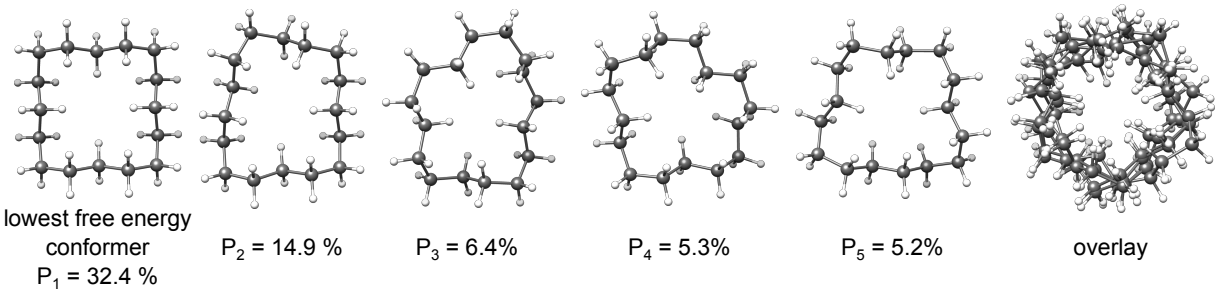


Figure A16.18.: Highest populated conformers of cyclohexadecane and overlay of all shown structures.

In 1974 Anet et al. performed low temperature ^{1}H and ^{13}C NMR measurements as well as iterative strain-energy calculations and concluded that the predominant cyclohexadecane conformation is the [4444] or diamond lattice, square conformation.[670] The rectangular [3535] conformation which Anet et al. also considered for the explanation of the carbon signal ratio (1:2:1 at -152 °C) is also found in our ensemble and is in accord with their strain-energy computations 1.65 kcal mol^{-1} less favorable in free energy compared to the square conformation.

Cycloheptadecane

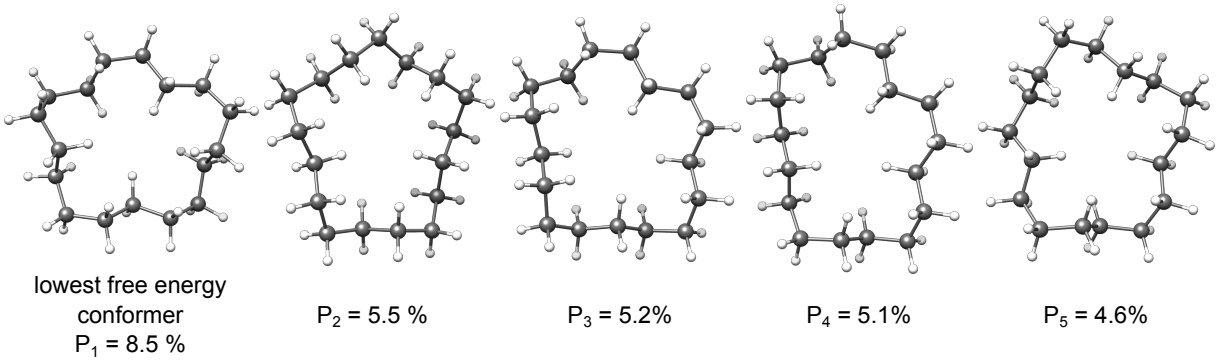


Figure A16.19.: Highest populated conformers of cycloheptadecane.

In a comparison of computational search methods Saunders et al. investigated the potential energy surface of cycloheptadecane[642]. Combining all search results they estimated a Boltzmann population of the most stable conformer to be 8 % at 300 K which is in perfect agreement with our lowest lying conformer and Boltzmann population P_1 = 8.5 % (in $CHCl_3$).

Cyclooctadecane

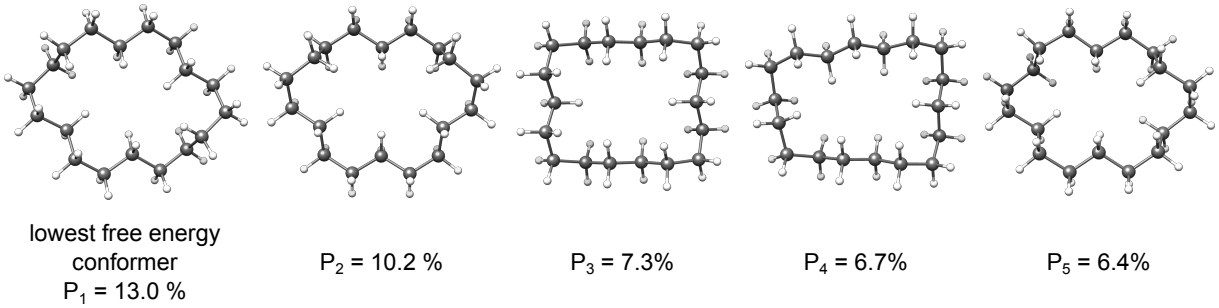


Figure A16.20.: Highest populated conformers of cyclooctadecane.

Investigations of larger cycloalkanes increase in complexity through their flexibility and the large amount of different accessible conformations. Using Monte Carlo computations Shah et al. investigated cyclooctadecane and identified the [3636] (not shown here, but identified in our ensemble), [4545] equal to P_3, and the 'nicked' conformation [234234] (our lowest lying conformer P_1) conformations to be the most stable.[671]

Cyclononadecane

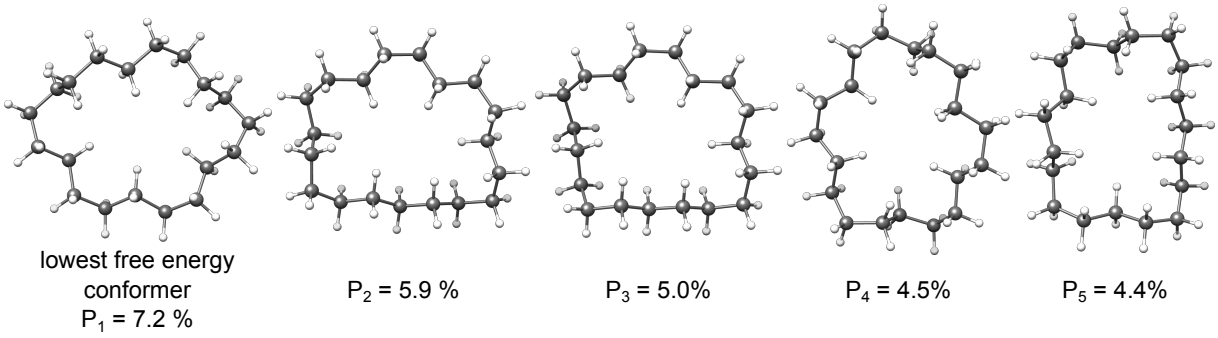


Figure A16.21.: Highest populated conformers of cyclononadecane.

The odd-membered cyclononadecane was also investigated by Shah et al. and all of their shown stable conformers can be found in our ensemble, but in a different order because of our Gibbs free energy sorting (in $CHCl_3$) compared to their gas phase MM2 energies.[671]

Cycloicosane

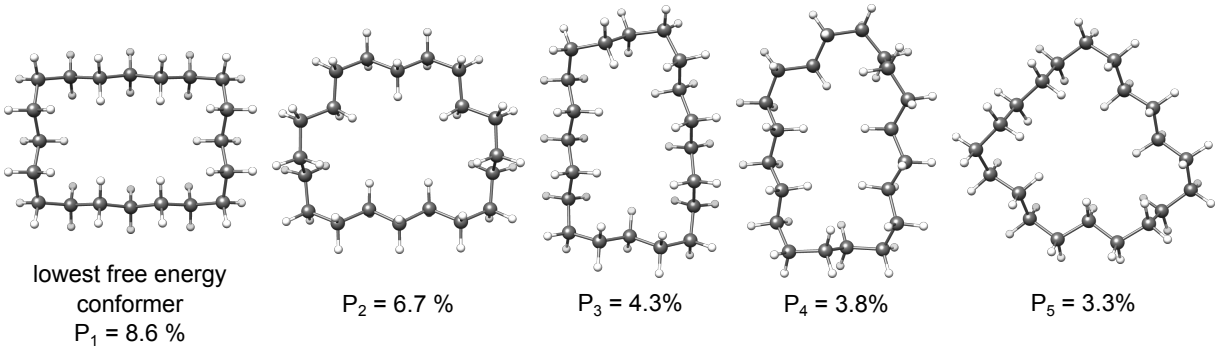


Figure A16.22.: Highest populated conformers of cycloicosane.

Cycloicosane as the largest here investigated cycloalkane has the lowest lying conformer in [7373] conformation (i.e. a rectangle of 4 sides with length of 7, 3, 7 and 3) which matches our P_1 conformer.[671] Like in the case of cyclononadecane other conformers identified and presented by Shah et al. are found in our ensemble but are reordered based on free energy.

Application examples

In all three following application examples, the same technical settings are used as in the cycloalkane studies described in Table A16.3.

Application example I: cyclic peptide PGLGF

The characterization of cyclic peptides is a nontrivial task because cyclization confines the chemical environment, resulting in different chemical shifts compared to the linear peptide analog. Zaretsky et al. proposed and evaluated a computational workflow for the QM prediction of chemical shifts in cyclic peptides,[643] which was refined into the CANDLE workflow in 2018[238]. As a cyclic peptide model they studied PGLGF (see Fig. A16.23), which we evaluate here with our CRENSO algorithm.

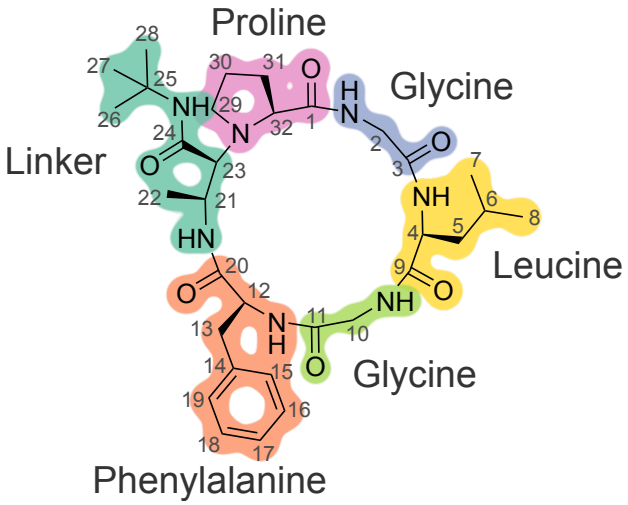


Figure A16.23.: Lewis structure of the cyclic peptide PGLGF with corresponding carbon atom numbering.

In this and the following to examples we employ our default CRENSO workflow consisting of the following steps:

1. conformer ensemble (CRE) generation using meta-dynamics driven by force-field and semi-empirical quantum mechanical (SQM) methods of the GFN-xtb family.
2. clustering of the ensemble to group similar conformers and condense the often large ensembles to the essential candidate conformers
3. refinement and sorting of the CRE at DFT level using CENSO
4. calculation of NMR properties facilitated by CENSO

In all three application examples, the same technical settings are used as in the cycloalkane studies described in Table A16.3.

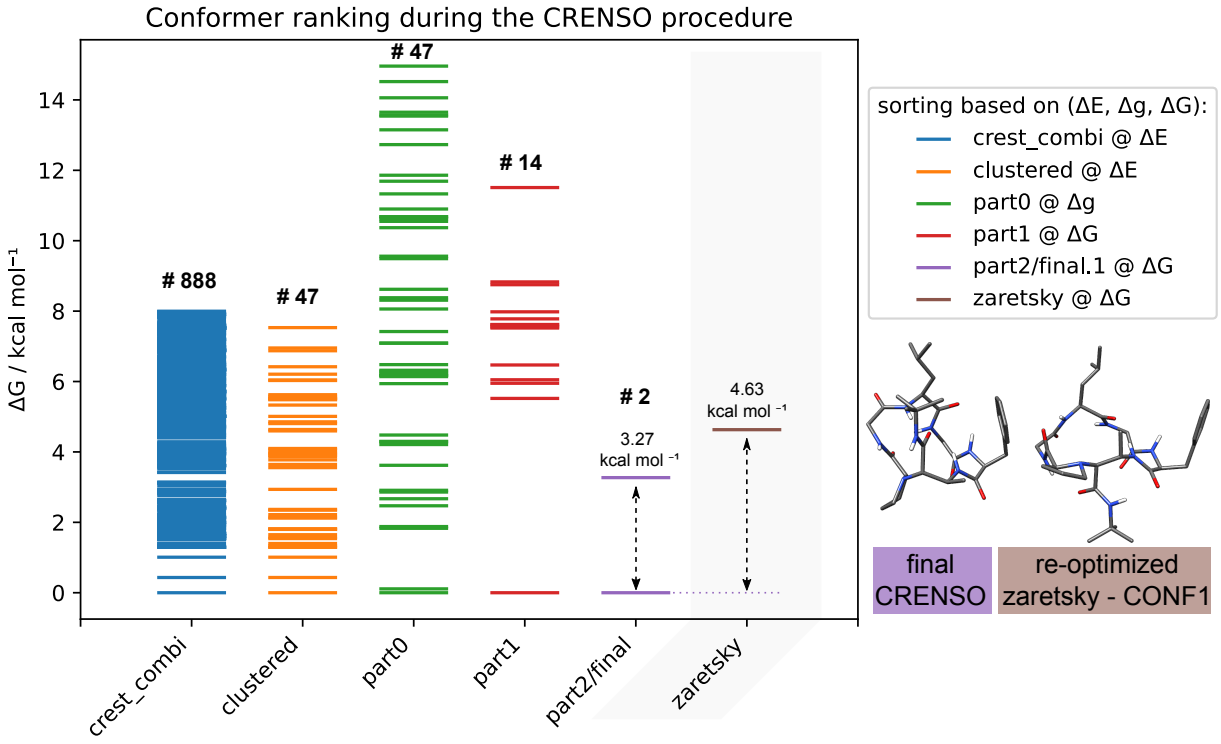


Figure A16.24.: Potential energy surface (PES) evaluation and conformer ensemble ranking. Hydrogen atoms in the shown geometries are omitted for clarity.

Figure A16.24 details the CRE generation and sorting of the 18-membered cycle PGLGF. Within an energy window of 8 kcal mol^{-1} crest_combi identifies 294 conformations. These are clustered using a special principle component analysis clustering algorithm whereby 47 conformers remain. In part0 an evaluation with a fast but improved energy description at DFT level, including a solvation contribution to the free energy, reorders the conformer ensemble compared to the SQM ranking. The sorting of conformers at the DFT level is obviously more distributed compared to the GFN-SQM methods, which is also evident in Part1 where free energies with thermostatistical correction are considered. In part2 conformers are optimized at r2scan-3c level with implicit solvation and the final ensemble is obtained. Here the picture is that a very stable conformer is identified with a second conformer 3.27 kcal mol^{-1} higher in Gibbs energy, i.e., not significantly populated anymore at room temperature. The search was started from Zaretsky's CONF1, which was reoptimized using consistent CRENSO settings and shows to be substantially higher in free energy compared to our identified final conformer.

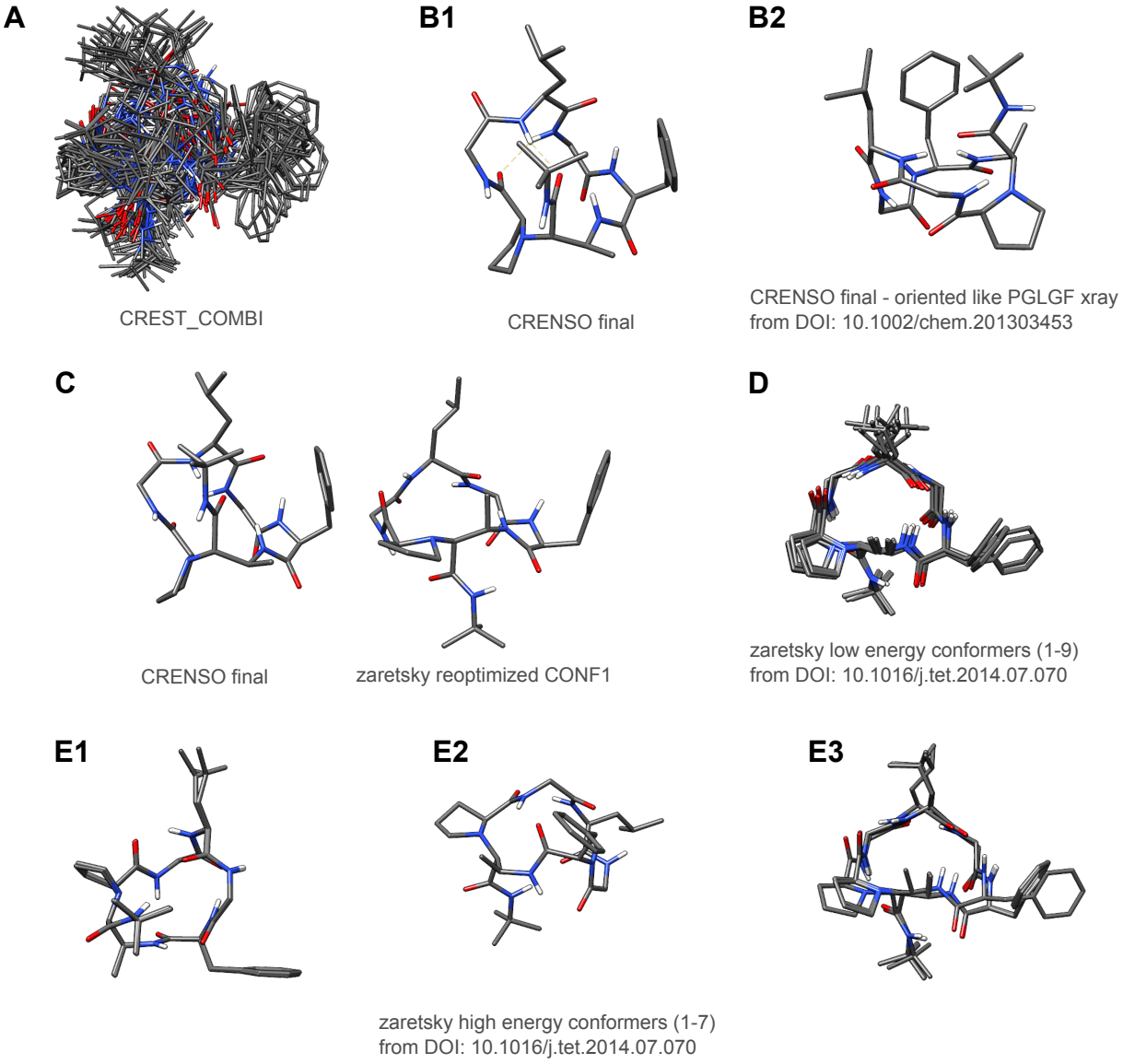


Figure A16.25.: Geometries created by the CRENSO algorithm and in comparison to conformers created by Zaretsky et. al.[643,644]. Hydrogen atoms are omitted for clarity.

The generated geometries and structures of Zaretsky et al. are shown in Fig. A16.25. In A the overlay of the clustered crest_combi ensemble shows a large diversity in the orientation of the ring itself, as well as the phenyl-, isopropyl- and linker groups. B1 and B2 show the final conformation obtained by CRENSO in two orientations, where the B2 geometry is oriented like the PGLGF X-ray structure from Ref. 644 and from visual inspection appears to be a very close match to the crystal structure, with all groups oriented on the same side of the macrocyle. In Fig. A16.25C the CRENSO lowest lying conformation is compared to the reoptimized CONF1 of Zartesky and a large difference in the orientation of the linker group is visible. Parts D and E show the low and high energy conformers as identified by Zaretsky et al.

	residue	position	CRENSO	zaretsky	zaretsky_MSTD
0	Pro1	C=O	-0.47	1.84	1.50
1	Gly2	A	0.74	3.06	2.75
2	Gly2	C=O	-2.82	4.44	7.78
3	Leu3	alpha	-1.46	1.45	0.93
4	Leu3	beta	-2.68	1.26	-1.32
5	Leu3	gamma	0.21	3.40	3.69
6	Leu3	delta	-0.37	1.13	-0.62
7	Leu3	delta	-4.05	1.13	-0.62
8	Leu3	C=O	-3.46	1.03	4.37
9	Gly4	alpha	-0.11	4.05	3.70
10	Gly4	C=O	-0.40	2.10	1.45
11	Phe5	alpha	2.71	4.35	3.66
12	Phe5	beta	-0.17	2.50	2.43
13	Phe5	gamma	-2.95	2.16	-1.08
14	Phe5	delta	-1.78	1.62	-4.53
15	Phe5	epsilon	-3.09	1.93	-4.81
16	Phe5	zeta	-1.97	2.61	-5.41
17	Phe5	epsilon	-3.09	1.93	-4.81
18	Phe5	delta	-1.78	1.62	-4.53
19	Phe5	C=O	-1.29	0.84	4.29
20	Linker6	alpha	-0.63	1.48	1.17
21	Linker6	beta	-0.47	1.83	-1.17
22	Linker6	gamma	-3.44	0.74	-0.27
23	Linker6	Exo-C=O	-1.20	1.01	2.52
24	Linker6	tBuC	0.29	5.00	4.37
25	Linker6	tBuCH	-2.10	2.89	-2.52
26	Linker6	tBuCH	-2.10	2.89	-2.52
27	Linker6	tBuCH	-2.10	2.89	-2.52
28	Pro1	delta	3.57	3.99	-4.28
29	Pro1	gamma	0.20	0.76	-0.29
30	Pro1	beta	1.64	2.98	-2.66
31	Pro1	alpha	-3.33	2.45	1.52
		MAD:	**1.77**	**2.29**	**2.81**

^{13}C chemical shift deviation

-7 7

(calc. - exp.) [ppm]

CRENSO:
crest_combi + CENSO workflow

zaretsky:
Data taken from
10.1016/j.tet.2014.07.070

zaretsky_MSTD:
Data taken from
10.1016/j.tet.2014.07.070

MAD (mean absolute deviation):

$$\frac{1}{n}\sum_{1}^{n}|\delta_{calc}-\delta_{exp}|$$

Figure A16.26.: ^{13}C chemical shift deviations compared to experiment.[643,644]

The calculated isotropic shielding constants from the final CRENSO conformation are converted to

chemical shifts by the scaling factors and reference determined in section A16. The chemical shift deviations compared to the experiment are color coded in Fig. A16.26 and results from Zaretsky et al. who used a similar scaling and a multistandard (MSTD) approach for converting shielding constants to chemical shifts are listed for comparison. The mean absolute deviation (MAD) shows very good overall agreement with the experiment at 1.77 ppm for the CRENSO workflow. The entire CRENSO calculation took roughly 12 hours on 28 cores of an Intel(R) Xeon(R) CPU E5-2660 v4 @ 2.00GHz machine.

Application example II: lobophytolins A

In 2020 Li et al. isolated macrocyclic lobophytolins A from soft coral and assigned the structure using NMR and QM calculations.[645] The lewis structure of lobophytolins A is shown in Fig. A16.29 and is investigated by the same procedure and settings as described in the previous example application section A16.

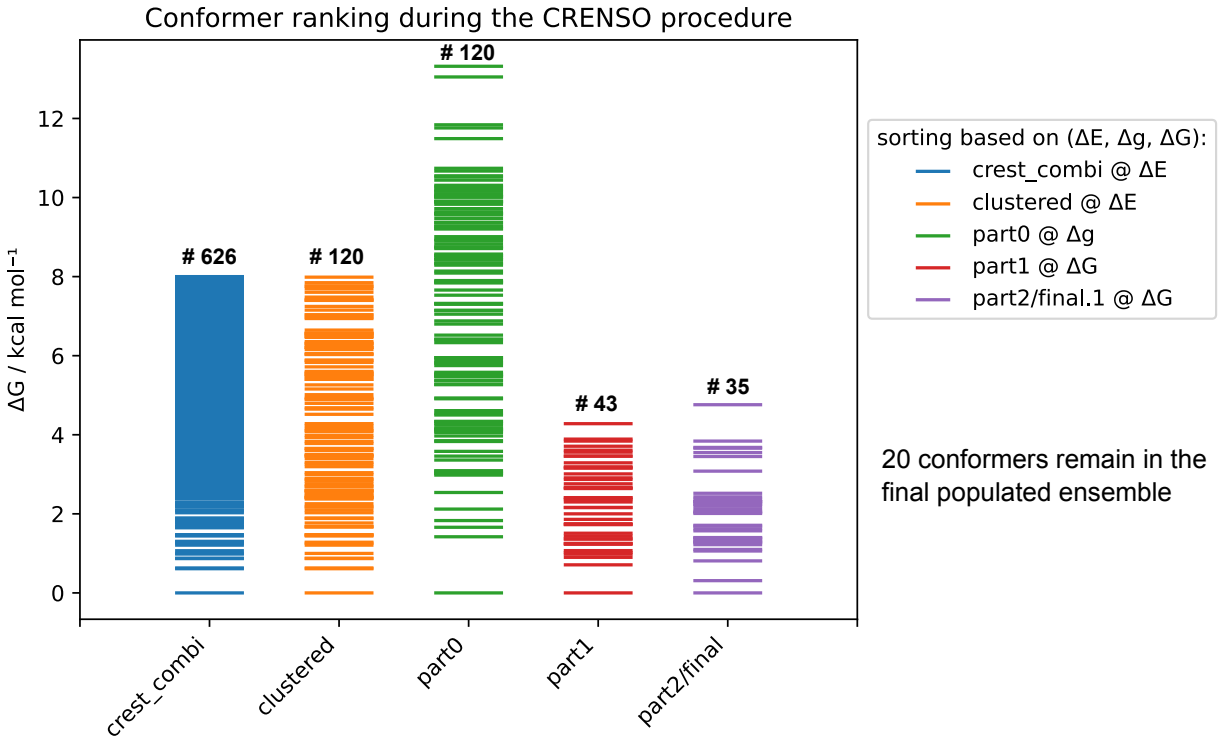


Figure A16.27.: PES evaluation and conformer ensemble sorting.

The evolution of CRE generation and sorting is shown in Fig. A16.27, starting with 626 conformers identified by crest_combi and reduced to 120 after clustering. Free-Energy evaluation at DFT level in part0 and part1 reorder the conformers and show a less flat PES compared to the SQM ranking. In the final populated CRENSO ensemble remain 20 conformers, which are shown in Fig. A16.28.

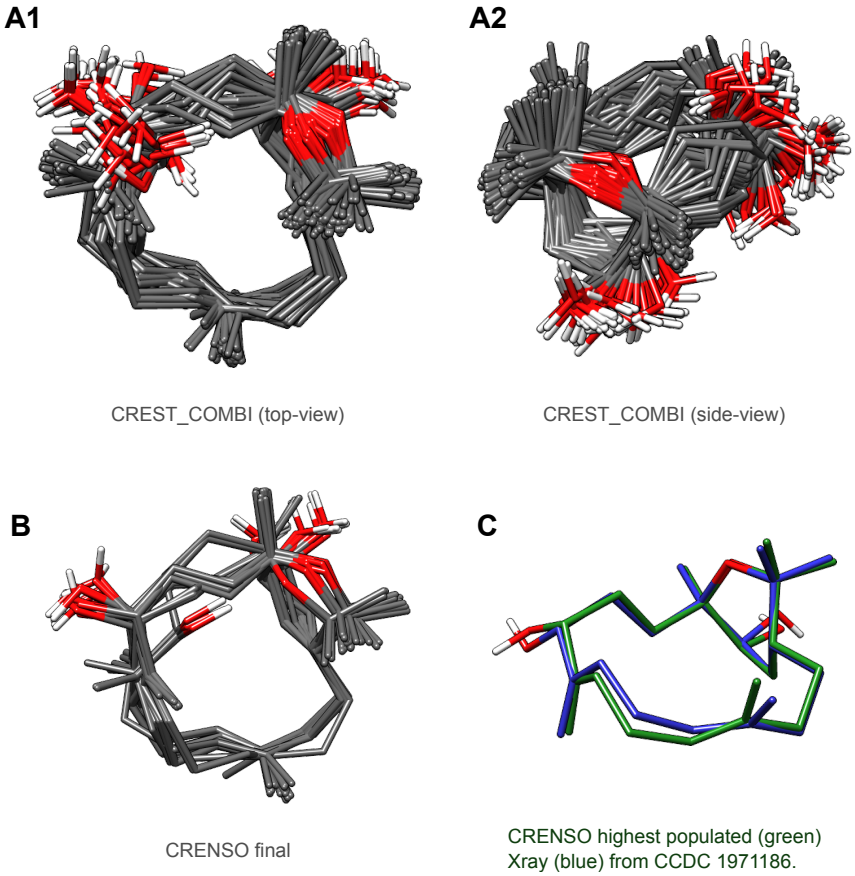


Figure A16.28.: Geometries created by the CRENSO algorithm and in comparison to the crystal structure[645]. Hydrogen atoms are omitted for clarity.

Part A1 and A2 in Fig. A16.28 show the overlay of the clustered crest_combi ensemble. The overlay of the final CRENSO ensemble is presented in B. Fig. A16.28C shows the overlay of the CRENSO lowest lying $CHCl_3$ solution conformer (P_1 = 30.07 %) in green with the crystal structure in blue. The populations of the other conformers are listed in Table A16.4.

#CONF	$\Delta G[kcal \cdot mol^{-1}]$	P [%]
1	0.00	30.07
2	0.31	17.82
3	0.81	7.71
4	1.06	5.07
5	1.10	4.76
6	1.22	3.88
7	1.27	3.58
8	1.30	3.40
9	1.30	3.40
10	1.37	3.04
11	1.39	2.93
12	1.40	2.87
13	1.57	2.17
14	1.62	2.00
15	1.69	1.77
16	1.71	1.71
17	2.01	1.04
18	2.04	0.99
19	2.06	0.95
20	2.13	0.85

Table A16.4.: Free energy differences and Boltzmann populations of the conformers in the final CRENSO ensemble of lobophytolins A.

	position	δ (exp)	isotropic shielding constant	CRENSO
0	1	35.10	149.11	-0.20
1	2	32.00	154.55	-2.20
2	3	70.70	111.76	-0.60
3	4	76.30	105.48	-0.20
4	5	34.20	149.69	0.20
5	6	28.80	155.49	0.10
6	7	74.00	106.47	1.10
7	8	149.70	24.45	2.80
8	9	32.50	151.22	0.40
9	10	23.60	160.23	0.80
10	11	126.90	51.93	-0.40
11	12	134.20	41.81	1.90
12	13	36.90	146.84	0.20
13	14	28.80	155.74	-0.10
14	15	75.80	105.82	-0.10
15	16	30.80	156.17	-2.50
16	17	23.70	163.41	-2.30
17	18	25.20	162.46	-2.90
18	19	108.10	73.10	-1.50
19	20	15.00	171.18	-0.90
			MAD:	**1.06**

^{13}C chemical shift deviation

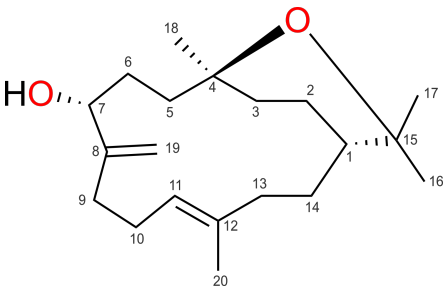
-7 7

(calc. - exp.) [ppm]

CRENSO:
crest_combi + CENSO workflow

MAD (mean absolute deviation):

$$\frac{1}{n}\sum_{1}^{n}|\delta_{calc}-\delta_{exp}|$$



Iobophytolins A

Figure A16.29.: Comparison of ^{13}C chemical shift deviation to experiment.[645]

The calculated isotropic shielding constants from the final CRENSO conformation are converted to chemical shifts by the scaling factors and reference determined in section A16. The agreement of the calculated chemical shifts with experiment are exceptionally good for the CRENSO solution ensemble, with a MAD of 1.06 ppm. The entire CRENSO calculation took roughly 15 hours on 28 cores of an Intel(R) Xeon(R) CPU E5-2660 v4 @ 2.00GHz machine.

Application example III: (-)-dactylolide

Larsen et al. studied the solution conformation behavior of (-)-dactylolide (see Fig A16.32), which acts as a tubulin binding agent and can be used as a chemotherapeutic agent.[646] The $CHCl_3$ solution conformations of synthesised (-)-dactylolide was studied by NMR measurements and computational modeling,

creating a solution ensemble using NMR data in the DISCON (Distribution of Solution Conformations) software.
We applied the default CRENSO workflow as already described in section A16 and the ensemble generation and evaluation is shown in Fig. A16.30.

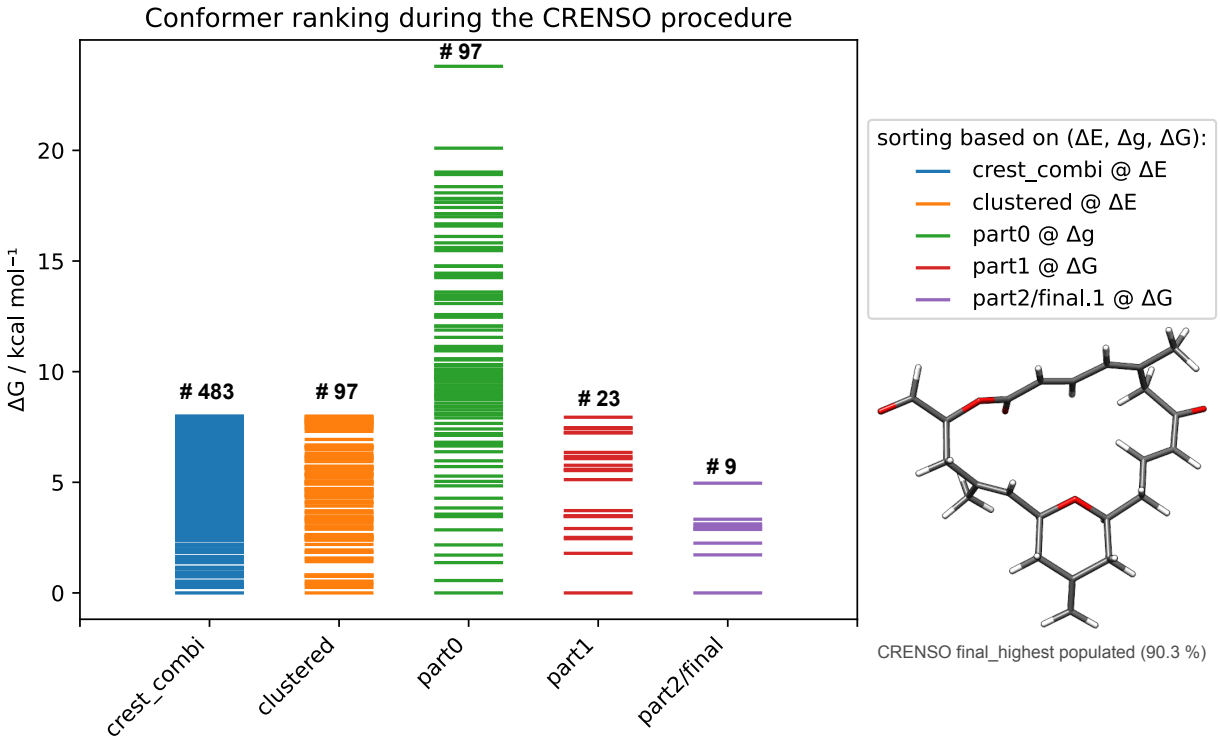


Figure A16.30.: PES evaluation and conformer ensemble ranking. And lowest lying CRENSO solution conformer geometry.

The combined conformer search yields 483 conformers in an energy window of 8 kcal mol^{-1} and 97 clustered conformations are evaluated in the CENSO algorithm. The sorting at DFT level using free energy thresholds drastically reduces the number of relevant conformer candidates and only 9 conformers remain in the final ensemble.

#CONF	$\Delta G[kcal \cdot mol^{-1}]$	P [%]
1	0	90.31
2	1.72	4.94
3	2.25	2.01
4	2.87	0.71
5	2.91	0.66
6	3.02	0.55
7	3.12	0.47
8	3.33	0.33
9	4.96	0.02

Table A16.5.: Free energy differences and Boltzmann populations of the conformers in the final CRENSO ensemble of (-)-dactylolide.

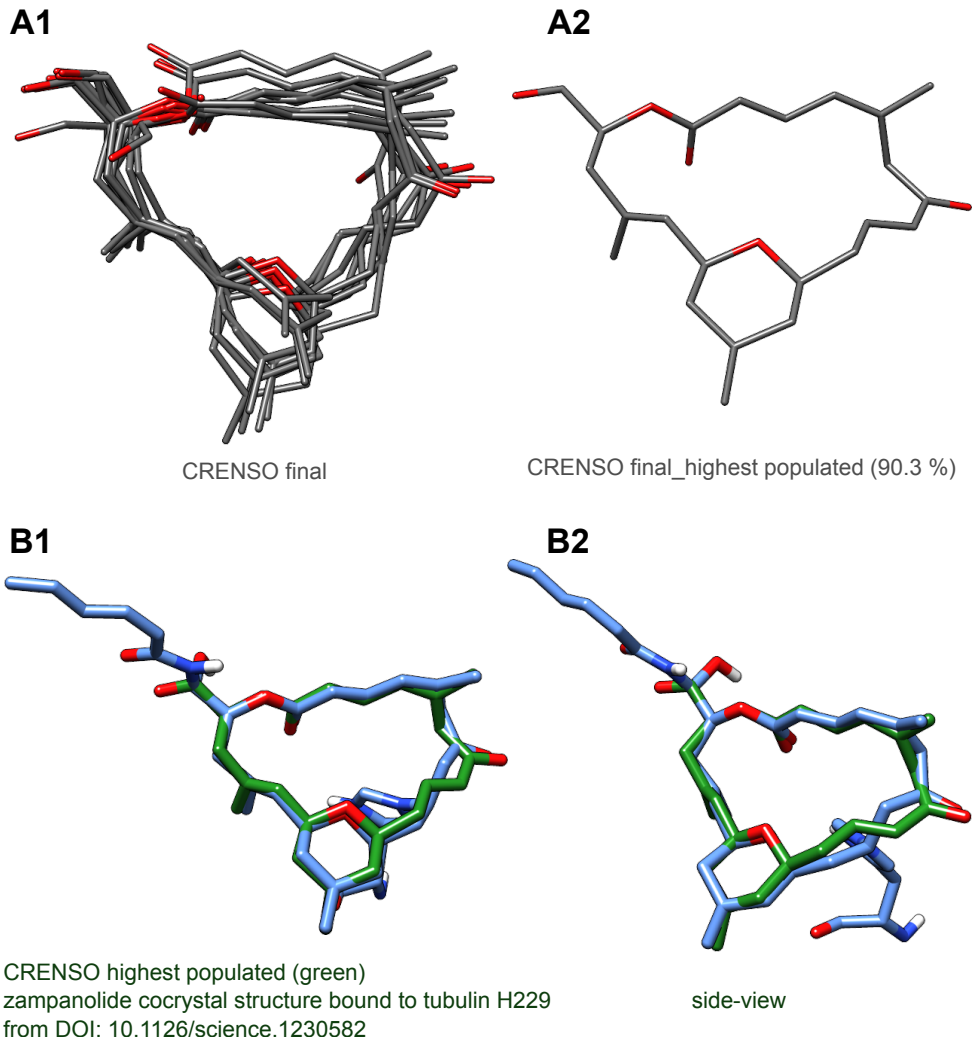


Figure A16.31.: Geometries created by the CRENSO algorithm and in comparison to zampanolide cocrystal structure bound to tubulin H229.[647]

The large population of 90.3 % suggest a very stable conformation (see Fig. A16.31A2). The lowest lying conformation closely resembles the cocrystal structure of zampanolide (a similar polyketide natural product) bound to tubulin H229.[647]

	position	type	δ (exp)	isotropic shielding constant	CRENSO
0	1	sp2	166.50	10.46	-0.90
1	2	sp2	119.90	64.77	-5.50
2	3	sp2	140.60	34.96	1.90
3	4	sp2	125.70	56.50	-3.50
4	5	sp2	144.20	22.76	9.80
5	6	sp3	45.00	138.64	-0.20
6	7	sp2	197.60	-23.98	0.50
7	8	sp2	131.60	51.10	-4.30
8	9	sp2	146.20	22.73	7.90
9	10	sp3	39.90	142.33	1.40
10	11	sp3	76.60	105.20	-0.30
11	12	sp3	41.00	142.51	0.10
12	13	sp2	143.60	30.54	3.10
13	14	sp3	40.60	142.59	0.50
14	15	sp3	75.90	105.46	0.20
15	16	sp2	130.70	49.56	-1.90
16	17	sp2	131.10	44.03	2.90
17	18	sp3	39.90	143.68	0.10
18	19	sp3	75.50	106.32	-0.20
19	13a	sp2	109.50	73.74	-3.50
20	17a	sp3	16.20	170.27	-1.20
21	19a	sp2	199.30	-30.26	4.70
22	5a	sp3	24.30	160.78	-0.40
				MAD:	**2.39**

^{13}C chemical shift deviation

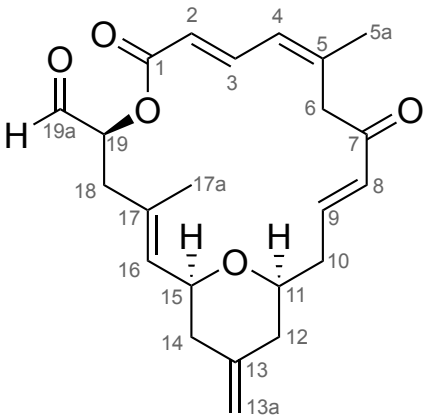
-7 7

(calc. - exp.) [ppm]

CRENSO:
crest_combi + CENSO workflow

MAD (mean absolute deviation):

$$\frac{1}{n}\sum_{1}^{n}|\delta_{calc}-\delta_{exp}|$$



(-)-dactylolide

Figure A16.32.: Comparison of ^{13}C chemical shifts to experiment.[646]

The calculated isotropic shielding constants from the final CRENSO conformation are converted to chemical shifts by the scaling factors and reference determined in section A16. The agreement of the calculated chemical shifts with the experiment is reasonable with an overall MAD of 2.39 ppm. The entire CRENSO calculation took roughly 9 hours on 28 cores of an Intel(R) Xeon(R) CPU E5-2660 v4 @ 2.00GHz machine.

A17. List of Abbreviations

AES	Anisotropic electrostatics
AC	Adiabatic connection
ACONF	Alkane chain conformer benchmark set
ALDA	Adiabatic local density approximation
ALPB	Analytical linearized Poisson–Boltzmann implicit solvation model
AMAX	Absolute maximum deviation
AO	Atomic orbital
ATM	Axilrod-Teller-Muto
AXC	Anisotropic exchange
BDE	Bond dissociation energy
BJ	Becke-Johnson
BO	Bond order
BSIE	Basis set incompleteness error
BSSE	Basis set superposition error
CAMM	Cumulative atomic multipole moments
CANDLE	Conformational analysis from NMR and density-functional prediction of low-energy ensembles
CBS	Comlete basis set
CC	Coupled cluster
CCDC	Cambridge Crystallographic Data Centre
CCSD	Coupled cluster singles and doubles
CD	Circular dichroism
CCE	Clustered conformer ensemble
CE	Conformer ensemble
CENSO	Commandline energetic sorting
CFER	Cubic free energy relationship
CI	Configuration interaction
CN	Coordination number
COSMO	Conductor-like screening model
COSMO-RS	Conductor-like screening model for real solvents
CPCM	Conductor-like polarizable continuum solvation model
CPSCF	Coupled-perturbed self-consistent field theory
CPU	Central processing unit
CRE	Conformer rotamer ensemble
CRENSO	Bash script to automate CREST and CENSO
CREST	Conformer rotamer ensemble sampling tool
CSD	Cambridge Structural Database
CT	Charge transfer
CTC	Cyclic triangular complexes
DCOSMO-RS	Direct conductor-like screening model for real solvents
DFA	Density functional approximation
DFG	Deutsche Forschungsgemeinschaft
DFT	Density functional theory
DFTB	Density functional tight binding
DHDF	Double hybrid density functional
DISCON	Distribution of Solution Conformations
DLPNO	Domain based local pair natural orbital
DOF	Degrees of freedom
EA	Electron Affinity
ECD	Electronic circular dichroism
ECP	Effective core potential
EDA	Energy decomposition analysis
EEQ	Electronegativity equilibrium
EHT	Extended Hückel theory
EN	Electronegativity
FCI	Full configuration interaction
FER	Free energy relationship

FF	Force field
GC	Genetic crossing
GBSA	Generalized Born, solvent accessible surface area
GFN	Geometries, frequencies, non-covalent interactions (referring to GFN*n*–xTB)
gCP	Geometrical counterpoise
gESC	Geometrical energy surface correction
GGA	Generalized gradient approximation
GIAO	Gauge invariant atomic orbital
GTO	Gaussian type orbital
GSM	Growing string method
HAT	Hydrogen atom transfer
HB	Hydrogen bond
HF	Hartree–Fock
HO	Harmonic oscillator
HOMO	Highest occupied molecular orbital
HTVS	High-throughput virtual screening
HVF	Harmonic vibrational frequencies
IES	Isotropic electrostatics
IP	Ionization potential
IR	Infrared
IXC	Isotropic exchange
KS–DFT	Kohn-Sham density functional theory
LCAO	Linear-combination of atomic orbitals
LFER	Linear free energy relationship
LMO	Localized molecular orbital
LP	Lone-pair
LDA, LSDA	Local (spin) density approximation
LUMO	Lowest unoccupied molecular orbital
MAD	Mean absolute deviation
MD	Molecular dynamics
MDL	Maschine learning
MM	Molecular mechanics
MNDO	Modified neglect of diatomic overlap
MO	Molecular orbital
MP	Møller-Plesset
msRRHO	Modified and scaled RRHO
MSFS	Molecule specific frequency scaling (factor)
MTD	Metadynamics
NAMFIS	NMR analysis of molecular flexibility in solution
NCI	Noncovalent interaction
NBO	Natural bond orbital
NDDO	Neglect of diatomic differential overlap
NL	Non-local
NOE	Nuclear Overhauser effect
NMR	Nuclear magnetic resonance (spectroscopy)
OR	Optical rotation
PCA	Principle component analysis
PM	Parametric method (referring to PM*x* methods)
PES	Potential energy surface
PSE	Periodic system of the elements
QFER	Quadratic free energy relationship
QSAR	Qualitative structure activity relationship
QM	Quantum mechanical
RI	Resolution-of-identity
RMSD	Root-mean-square deviation
RRHO	Rigid-rotor harmonic-oscillator approximation
RPA	Random phase approximation
SASA	Solvent accessible surface area
SCC	Self-consistent charge
SCF	Self-consistent field

SE	Schrödinger equation
SD	Standard deviation
SEM	Scanning electron microscope
SET	Single electron transfer
SIE	Self-interaction error
SMD	Solvation model based on density
SPE	Single-point energy
SPH	Single-point hessian
SQM	Semiempirical quantum mechanical
SRB	Short-range bond
STM	Scanning tunneling microscope
STO	Slater type orbital
TDDFT	Time dependent density functional theory
TM	Transition metal
TS	Transition state
TST	Transition state theory
UEG	Uniform electron gas
UV	Ultraviolet
VIS	Visible
WBO	Wiberg bond order
WFT	Wave function theory
xTB	Extended tight-binding
ZPVE	Zero-point vibrational energy

Acknowledgments

At this point, I would first like to thank Prof. Dr. Stefan Grimme for supervising my dissertation and giving me the freedom to work on thematically very different projects. Secondly, I would like to thank Prof. Dr. Thomas Bredow for kindly taking over the second review of my thesis. I want to thank Prof. Dr. Gansäuer for encouraging me in my decision to explore the world of theoretical chemistry. Furthermore, I thank Dr. Philipp Pracht and Dr. Marc de Wergifosse for discussing and proofreading parts of this thesis.

For the wonderful time together filled with exciting academic and non-academic discussions, I thank all current and former colleagues: Dr. Andreas Hansen, Dr. Marc Steinmetz, Dr. Stefan Ehrlich, Dr. Gerit Brandenburg, Dr. Rebecca Sure, Dr. Christoph Bannwarth, Mirko Bauer, Dr. Christoph Bauer, Sebastian Dohm, Jana Pisarek, Dr. Eike Caldeweyher, Dr. Jakob Seibert, Dr. Markus Bursch, Dr. Sebastian Spicher, Dr. Philipp Pracht, Dr. Marc de Wergifosse, Dr. Vilhjalmur Asgeirsson, Dr. Jan Mewes, Dr. Sebastian Ehlert, Dr. Jeroen Koopmann, Hagen Neugebauer, Sarah Schmitz, Julius Stückrath, Johannes Gorges, Thomas Gaesevic, Thomas Rose, Marcel Stahn, Marcel Müller, Christian Hölzer.
Moreover, I would also like to thank all my collaborators with whom I have had the pleasure of working. Special thanks are extended to Prof. Suhm, Prof. Esser, Dr. Raiko Hahn, Prof. Dr. Gansäuer, Prof. Lützen, Dr. Jana Anhäuser, and Dr. Lucia Volbach.

I greatly appreciated the pleasant time spent at the Mulliken Center, which was made possible by the excellent organizational and technical support thanks to Claudia Kronz, Jens Mekelburger, and Dr. Werner Reckien.

Finally, I would like to thank my family for their unwavering support.