Development of Quantum Chemistry based Workflows for the Theoretical Description of Organic Electronics

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

vorgelegt von
Julia Therese Kohn
aus
Karlsruhe

Bonn 2025

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

Gutachter / Betreuer:	Prof. Dr. Stefan Grimme
Gutachter:	Prof. Dr. Thomas Bredow
Tag der Promotion:	13.01.2025
Erscheinungsjahr:	2025

Affirmation in Lieu of an Oath

I, Julia Therese Kohn, hereby declare that I am the sole author of this dissertation.

The doctoral thesis I submitted is my own work and was prepared without unauthorized outside assistance. I have not included text passages, graphics or other materials from third parties or my own examination papers without identifying them. Only the sources and resources that I have indicated were used. All verbatim and non-verbatim citations from other works are identified in accordance with the citation rules for academic writing. The parts of the thesis that I submitted that have been published (appendices A to F) are indicated at the beginning of Chapters 3 to 8. All other parts have not been published elsewhere. The thesis that I submitted has not yet been submitted in any form as part of an examination/qualification course. I prepared the doctoral thesis that I submitted in accordance with the principles of good research practice. I am aware of the significance and criminal consequences of a false affirmation in lieu of an oath.

My statements are true to the best of my knowledge and belief.

Publications

The following publications serve as part of this thesis.

1. J. T. Kohn, S. Spicher, M. Bursch, and S. Grimme, *Quickstart guide to model structures and interactions of artificial molecular muscles with efficient computational methods*, Chem. Comm. **58**.2 (2022) 258, DOI: 10.1039/d1cc05759f.

2. J. S. Wössner, J. T. Kohn, D. Wassy, M. Hermann, S. Grimme, and B. Esser, *Increased Antiaromaticity through Pentalene Connection in [n] Cyclo-1, 5-dibenzopentalenes*, Org. Lett. **24**.4 (2022) 983, DOI: 10.1021/acs.orglett.1c03900.

3. P. Krämer, J. Kohn, D. A. Hofmeister, M. Kersten, C. Sterzenbach, A. Gres, A. Hansen, S.-S. Jester, S. Grimme, and S. Höger, *Size-Increased All-Phenylene Molecular Spoked Wheels–A Combined Theoretical and Experimental Approach*, Angew. Chem. Int. Ed. (2024) e202411092, DOI: 10.1002/anie.202411092.

4. J. T. Kohn, H. Li, A. M. Evans, J.-L. Bredas, and S. Grimme, *Quantum Chemistry Insight into the Multifaceted Structural Properties of Two-Dimen-sional Covalent Organic Frameworks*, Chem. Mater. **35**.7 (2023) 2820, DOI: 10.1021/acs.chemmater.2c03555.

5. J. T. Kohn, N. Gildemeister, S. Grimme, D. Fazzi, and A. Hansen, *Efficient calculation of electronic coupling integrals with the dimer projection method via a density matrix tight-binding potential*, J. Chem. Phys. **159**.14 (2023), DOI: 10.1063/5.0167484.

6. J. T. Kohn, S. Grimme, and A. Hansen, *A Semi-Automated Quantum-Mechanical Workflow for the Generation of Merocyanine Monolayers and Aggregates for Organic Electronics*, J. Chem. Phys. **161**.12 (2024), DOI: 10.1063/5.0230341.

Before and during the time period of this thesis, the following further studies were carried out.

7. T. A. Schaub, E. A. Prantl, J. T. Kohn, M. Bursch, C. R. Marshall, E. J. Leonhardt, T. C. Lovell, L. N. Zakharov, C. K. Brozek, S. R. Waldvogel, et al., *Exploration of the solid-state sorption properties of shape-persistent macrocyclic nanocarbons as bulk materials and small aggregates*, J. Am. Chem. Soc. **142**.19 (2020) 8763, DOI: 10.1021/jacs.0c01117.

8. M. Bursch, A. Hansen, P. Pracht, J. T. Kohn, and S. Grimme, *Theoretical study on conformational energies of transition metal complexes*, Phys. Chem. Chem. Phys. **23**.1 (2021) 287, DOI: 10.1039/d0cp04696e.

9. M. Hermann, D. Wassy, J. T. Kohn, P. Seitz, M. U. Betschart, S. Grimme, and B. Esser, *Chiral Dibenzopentalene-Based Conjugated Nanohoops through Stereoselective Synthesis*, Angew. Chem. Int. Ed. **60**.19 (2021) 10680, DOI: 10.1002/anie.202016968.

10. T. J. Keller, C. Sterzenbach, J. Bahr, T. L. Schneiders, M. Bursch, J. Kohn, T. Eder, J. M. Lupton, S. Grimme, S. Höger, et al., *Nanopatterns of molecular spoked wheels as giant homologues of benzene tricarboxylic acids*, Chem. Sci. **12**.27 (2021) 9352.

11. J. T. Kohn, M. Bursch, A. Hansen, and S. Grimme, *Computational study of ground-state properties of μ 2-bridged group 14 porphyrinic sandwich complexes*, J. Comp. Chem. **44**.3 (2023) 229, DOI: 10.1002/jcc.26870.

12. S. Rickert, S.-X. L. Luo, J. Bahr, J. T. Kohn, M. Xue, A. Hansen, S. Grimme, S.-S. Jester, T. M. Swager, and S. Höger, *Flexible Phenanthracene Nanotubes for Explosive Detection*, J. Am. Chem. Soc. **146**.5 (2024) 2986, DOI: 10.1021/jacs.3c08131.

13. K. Gratzfeld, A. J. Kny, J. T. Kohn, A. Hansen, S. Grimme, K. Wandelt, and M. Sokolowski, *Tuning the Adsorption Geometry of Structures of Perylene Tetracarboxylic Acid on the Au(111) Electrode by the Potential* (*in preparation*).

Abstract

Organic electronics (OE) have become increasingly significant in daily life. Groundbreaking inventions such as organic solar cells (OSCs), organic light-emitting diodes (OLEDs), and organic field-effect transistors (OFETs) have revolutionized the field of sustainable materials for energy conversion. OE enable the fabrication of molecular circuits with integrated functionalities like sensors and detectors, useful in fields such as molecular computing, nanomedicine, and the Internet of Things (IoT). Compared to conventional inorganic electronics, OE offer advantages regarding their sustainability and reduced production cost due to the use of organic or even plant-based materials, as well as versatility and tunable properties for customized solutions like printable screens, curved photovoltaic windows, intelligent textiles, and pathogen-filtering blood nanorobots.
The chemical space of potential candidates for OE is vast, making the discovery and application of new materials a promising yet tedious process. Supramolecular chemistry is not only a versatile playground for new topologies but also a workhorse in the design of OE materials. Computational chemistry can accelerate the discovery timeline and enhance the understanding of underlying design principles and functionality. However, the complexity and size of OE places severe constraints on this approach. A promising solution is to combine the strengths of experimental and computational approaches in property targeted, yet broadly applicable workflows, requiring both accurate methods like density functional theory (DFT) and fast methods like semiempirical quantum-mechanical (SQM) methods and force fields (FF). This work presents several such workflows tailored to different chemical spaces and target properties in OE that are shortly introduced in the following.

Molecular machines are molecular assemblies that mimic the function of macroscopic machinery such as motors, hinges, or switches. This field is particularly important for fundamental research in nanomechanics and the development of novel therapeutics in nanomedicine. Artificial molecular muscles (AMMs) are a versatile subgroup within this field. The computational prediction of mechanical parameters such as contraction ratio or pull resistance can facilitate the design of novel AMMs. Essential to any simulation is a standardized and reproducible structural model. Therefor, in Chapter 3 a workflow to determine the maximally extended structure of AMMs is presented, which is then verified on a benchmark of contraction ratios of experimentally studied AMMs.
Beyond functional molecular machines, supramolecular chemistry explores topological molecules like hoops, knots, and spiro compounds, as well as molecular analogues of macroscopic objects such as cylinders, bowls, ladders, and wheels. Such macromolecules are typically synthesized from smaller building blocks via C-C coupling reactions, enabling the customization of size, connectivity, and chirality. In this spirit, in Chapter 4 a workflow to examine the formation and stability of different isomers of strained anti-aromatic hoops is described, resolving the experimentally preferred isomers.

In Chapter 5 a workflow for evaluating the flexibility of molecular spoked wheels (MSWs) is presented and the feasibility of specific MSW designs is predicted.
Apart from single molecules or in solution, OE materials can exist as covalent or molecular crystals with varying degrees of order, ranging from amorphous powder to aligned thin films. Covalent organic frameworks (COFs) are part of the highly structured thin film group, yet the inherent polymorphism and high porosity challenge experimental structure determination. In Chapter 6 a workflow for solid state ensemble generation of COFs is presented, aiding in the interpretation of experimentally measured structures and opto-electronic properties.
Semiconducting organic molecular crystals typically exhibit intermediate charge transport rates in the hopping regime. To assess preferred transport pathways, the influence of disorder and defects, and the macroscopic conductivity, several theories were proposed in the past such as generalized Mulliken-Hush (GMH), atomic overlap matrix (AOM), or the dimer projection method (DIPRO). To include thermal disorder, molecular dynamic is the method of choice. However, this requires the reliable evaluation of a substantial number of coupling integrals in a reasonably short time. In Chapter 7 this demand is addressed by implementing DIPRO with a density matrix tight-binding method (PTB) and extended tight-binding methods (GFN-xTB). The DIPRO@PTB workflow applied to a benchmark set of organic homo-dimers and merocyanines yields accurate coupling integrals, with deviations of about 0.07 eV compared to range-separated hybrid DFT, at three to four orders of magnitude lower computational costs.
For practical applications, molecular OE materials are often fabricated as thin films with crystal structures highly dependent on experimental conditions resulting in various polymorphs. As single-structure theoretical descriptions are insufficient and crystal structure prediction (CSP) is computationally demanding and theoretically challenging, a cluster approach is in the focus of Chapter 8. Therein, a workflow is presented based on the automated interaction site screening (aISS) using GFN-xTB methods to generate ensembles of flat and stacked aggregates. The workflow is then applied to build a novel merocyanine benchmark set, tested against highly accurate double-hybrid DFT. Additionally, the convergence of opto-electronic properties with aggregate size is investigated and compared to experimental data.

Contents

CHAPTER 1

Introduction

Organic electronics (OE)[14–17] represent an extremely versatile class of compounds. This field encompasses not only the well-known areas of organic solar cells (OSCs),[18–21] organic light-emitting diodes (OLEDs),[22,23] and organic field-effect transistors (OFETs),[24,25] first discovered between the 1970s and 1980s,[26,27] but also more recent innovations, such as other subtypes of organic photovoltaics (OPVs),[28,29] e.g., dye-sensitized solar cells (DSSCs)[30,31] and polymer solar cells,[32,33] MOFtronics,[34] printable electronics,[35] conductive ink,[36] smart windows,[37] and electronic paper.[38] For the assessment of OE materials, their optical and electrical properties are of particular interest, which include light absorption and emission, charge transport, and outcoupling efficiency.[39–41] The theoretical description of OE materials presents challenges due to the combination of both molecular and bulk properties determining the overall device performance. Additionally, the optoelectronic properties are highly system-dependent, with factors such as defects,[42] disorder,[43] structural complexity,[44] interfaces,[45] and thermal fluctuations[46,47] further complicating the picture. This necessitates the use of high-level theoretical methods and complex multi-step workflows, while also requiring the construction of simplified model systems in order to maintain a system-wide accurate description and understanding.

Supramolecular chemistry involves the design, synthesis, and study of complex structures formed through the assembly of multiple molecular building blocks via covalent and non-covalent interactions.[48–51] This field aims to understand and exploit these interactions to create functional systems with applications in drug delivery,[52] materials science,[53], nanotechnology,[54] and OE. Supramolecular complexes were first discovered in the 1960s and began attracting increased interest in the 1980s with the development of cryptands.[55,56] Notable examples of supramolecular structures include bucky-catchers,[57] filled metal-organic frameworks (MOFs),[58] and molecular machines[59] such as motors[60] or[61] muscles.[1,62] Additionally, mechanically interlocked structures[63] like catenanes[64] or cucurbiturils,[65] as well as topological structures[66] such as knots,[67] molecular spoked wheels (MSWs),[3] and molecular ladders,[68] are significant in this field. These materials find applications in e.g., nanomedicine,[69] nanomechanics,[70] molecular circuits and molecular computers,[71,72] and sensing devices.[12] The quantum-mechanical description of supramolecular complexes presents challenges due to their large size and diverse interaction patterns, especially non-covalent interactions (NCIs) such as π-π stacking .[73] These challenges require a careful trade-off between speed and accuracy in the method selection.

Computational material design (CMD)[74,75], alongside computer-aided drug design, stands as one of the most prominent commercial applications of computational chemistry. Today, there is an

increased focus on OE and supramolecular materials for energy conversion, energy storage, and sustainable alternatives to traditional materials in tech, IT, and construction. The design of new functional materials is a complex task, given the vast chemical space of candidate molecules that needs to be explored and the challenges along the path towards a commercially viable product. Typically, the design process spans multiple scientific disciplines, involving chemistry, physics, material sciences, and production engineering.[76–78] Computational methods, such as quantum chemistry or classical mechanics, provide versatile and cost-saving tools that accelerate material discovery and enhance our understanding of design principles and material properties at a fundamental level.[79,80] CMD targets various material phases, with a primary focus on solids, such as single crystals,[81] molecular crystals,[82] amorphous powders,[83] surfaces,[84] gels and aerogels,[85,86] polymers,[87] glasses,[88] and thin films.[89] It also spans different compound classes, including covalent- and metal-organic frameworks (COFs and MOFs),[4,90] carbon-rich compounds,[3] donor-acceptor systems (e.g., merocyanines, OSC-acceptors),[6,91] and biochemical systems (e.g., peptides, DNA).[92,93] Key tasks of CMD range from structure elucidation (observation),[4] structure-property relationships (analysis),[94] to chemical space exploration (discovery),[95] and material proposals (engineering, prediction).[96,97] Addressing these questions necessitates various computational techniques including density functional theory (DFT),[98,99] semiempirical quantum-mechanical methods (SQM),[100,101] force fields (FF),[102] coarse-graining models,[103] statistical models,[104] and machine learning (ML),[105] often in combination within multilevel workflows and multi-scale models.[106–108] This thesis focuses on first principles CMD for applications in supramolecular chemistry and OE.

In general, the challenges associated with first principles CMD in chemistry can be divided into mainly structural problems and mainly electronic problems (see Figure 1.1).[109–111]

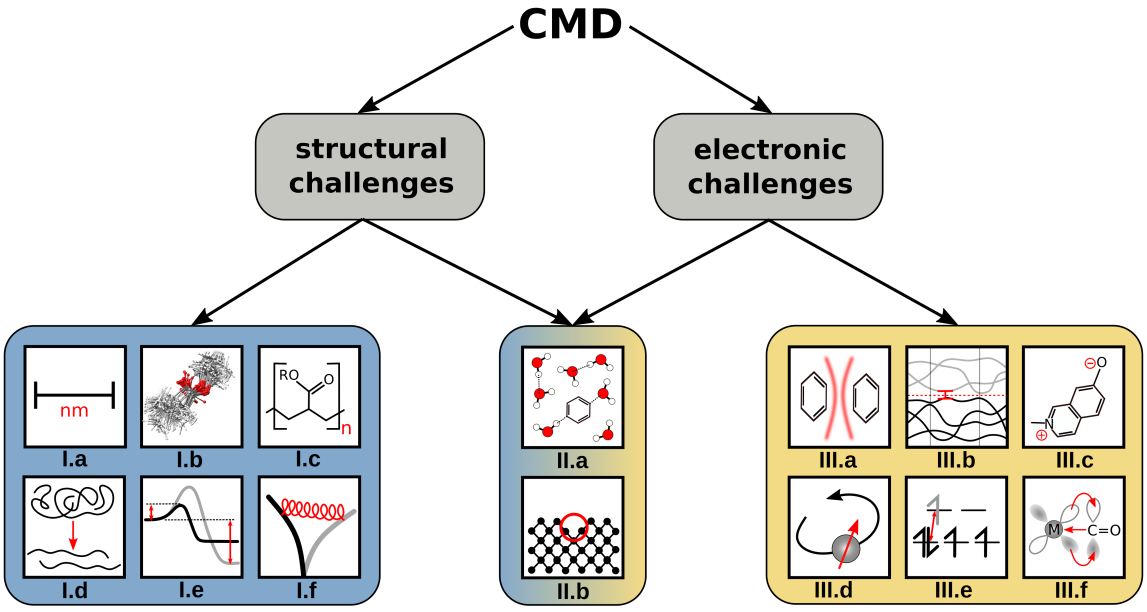

Figure 1.1: Challenges in CMD divided into mainly structural (I.) and mainly electronic (III.) problems. I.a size, I.b conformers, I.c oligomers, I.d aggregation, I.e kinetics, I.f mechanics, II.a solvation, II.b defects, III.a NCIs, III.b band gap, III.c charges, III.d relativity, III.e spin states, III.f special bond situations.

Molecular muscles, for example, are large (I.a) and flexible with many different conformations (I.b) that tend to aggregate (I.d). They are of interest due to their mechanical properties (I.f) such as controlled contraction. However, the contraction mechanism may depend on solvent exchange (II.a), NCIs (III.a), redox reactions (III.c), or metal ion exchange whereas metals can further demand the treatment of open shell character (III.e), or special bonding motifs (III.f). Thus, the choice of appropriate computational methods and structure models to obtain the relevant properties is highly important.

A structure model can be a single molecule, a cluster, a solid, or cutouts thereof (see Figure 1.2). Each model has its advantages and drawbacks depending on the application and property of interest. Molecular models are often used to represent gas-phase and solution-phase chemistry. They are comparably small, easy to build, and allow for straightforward calculations of complex properties such as reaction mechanisms, vibrational modes, and excited states. However, molecular models are not well-suited to describe long-range properties and cooperative effects that arise in the bulk, such as electronic coupling and conductivity, optical splitting, or packing effects. Additionally, the presence of flexible molecules requires the exploration of the conformational space and the choice of either only one appropriate conformer or an ensemble of thermally accessible conformers. Although this is better achievable for molecules than for the bulk, it is still neglected in many investigations. A cluster model considers some bulk interactions while retaining most of the advantages from the computational treatment of molecules and no need yet for periodic boundary conditions. The choice of the right cluster model in terms of size and spatial arrangement is crucial and can significantly impact both quantitative and even qualitative results. Cutouts from a periodic crystal structure ease this choice. A solid-state model captures bulk effects but has limited flexibility for structural modification. Particularly, the initial structure guess is much more complex than for the molecular model and relies frequently on the existence of an experimental crystal structure. Computational crystal structure prediction is, however, a huge challenge on its own.[97] Additionally, not all properties accessible for the molecular case are available for the periodic case at comparative levels of theory.

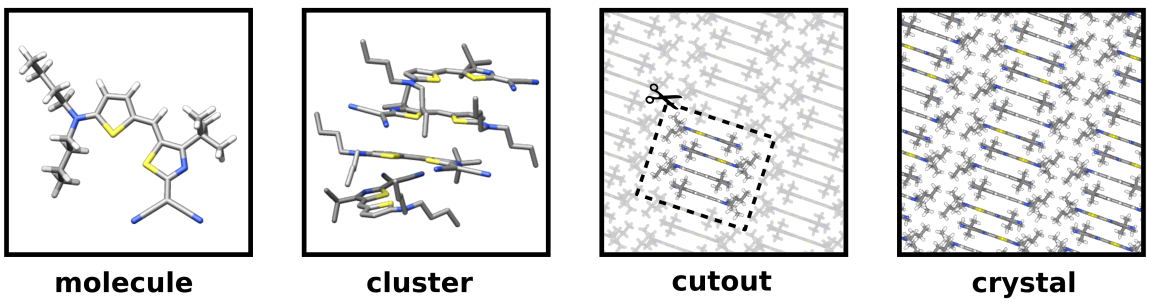

Figure 1.2: Different structure models at the example of a merocyanine with increasing size from left to right.

Aside from the inherent limitations of different structure models, further computational choices have to be made. These include finite temperature, dynamics, solvation, and simplifications such as shortened side chains. Side chains may or may not be significant and have to be chosen carefully. For example, side chains are crucial for the crystal packing of merocyanines, but solubilizing octadecyloxy side chains in supramolecular assemblies usually do not affect their reactivity. Solvation may or may not influence structure formation and properties. Large supramolecular residues do not interact

significantly with the solvents commonly used because they are apolar. On the other hand, e.g., merocyanines exhibit a substantial solvatochromic effect, and porous COFs collapse in the presence of a solvent. The inclusion of temperature effects can also vary in importance. For example, merocyanine aggregates exhibit strong vibronic coupling, allowing many radiationless relaxation pathways of optical excitons, while binding sites or stacking motifs are relatively temperature-independent. Nuclear and electronic dynamics are crucial for time-dependent properties such as charge transport pathways, reactions, or mechanical properties, while other properties like reaction energies or gaps are mostly regarded as a time-independent average.

Appropriate computational theories for atom-resolved CMD vary depending on the question at hand and include DFT, SQM, and FF. Correlated wave function theory (WFT) is usually too expensive for detailed investigations and is better suited to benchmark the accuracy of lower cost quantum chemical methods (e.g., DFT, SQM). DFT is an accurate and broadly applicable method for various properties in CMD, such as reaction barriers or optical spectra. In some cases, robust results can also be obtained for electronically challenging systems such as metal centers or conical intersections.[112] However, the performance, robustness, and ensuing computational costs of individual density functional approximations (DFAs) can vary widely. Additionally, also the employed basis set size has significant influences on the accuracy. Hybrid, double-hybrid (DH), and range-separated hybrid (RSH) DFAs with large basis sets can yield quantitative results for many experimental properties using molecular structure models, but are too costly for considering periodic boundary conditions, dynamics, or systems larger than a few hundred atoms. Generalized gradient approximation (GGA) DFAs allow for the treatment of larger systems, but this comes at the cost of sometimes significantly reduced accuracy. Qualitative results can often be attained by SQM methods at much lower computational costs. SQM methods are able to treat system sizes up to a few thousand atoms. For even larger systems, force fields can be employed. Although they no longer use a quantum chemical description of the electrons but atomistic classical mechanics, they can describe structures and structural dynamics of simpler systems well. They exist in very specialized versions to improve the description of selected parts of the chemical space e.g., FFs for water,[113] ionic liquids,[114] sugars,[115] peptides,[116] DNA,[117] membranes,[118] or polymers,[119] but also as general versions e.g., GFN-FF[120], or UFF.[121] Although electrons are not described explicitly, force fields can access electronic properties such as dipole moment or IR spectra via atomic charges.

Over the last decades, both the efficiency and accuracy were improved as well as combining beneficial approaches and correction schemes from DFT, SQM, and FF in composite methods. Examples are DFT-D,[122] the -3c methods,[123–127] embedding schemes like ONIOM or QM/MM,[128,129] or special-purpose parametrizations like protein FFs.[130] The DFT-3c methods employ a DFA with sometimes modified parametrization and a fixed (sometimes modified) basis set and additional varying corrections and potentials, such as counterpoise correction for the basis set superposition error (BSSE) as in PBEh-3c,[131] or large core effective core potentials as in ωB97X-3c.[123] These methods exhibit performance similar to calculations in a large basis set while achieving significant savings in computation time. Embedding schemes combine DFT, SQM, and FF in different subsystems described by different methods. Although this necessitates a careful choice of the respective regions and interfaces, by applying more accurate methods to critical regions and less expensive methods to peripheral regions, larger systems can be modeled with greater accuracy without a substantial increase in computational cost compared to using a single, less accurate method across the entire system. For example, the reactive metal center and ligands of a protein can be described by DFT, the closest protein side chains by SQM, the rest by a protein FF, and everything embedded in a solvation

continuum model. For CMD, periodic embedding schemes are harder to realize as there are usually no distinct inner and outer regions. An example are point charge embeddings as a representation for e.g., a surface, which would be too costly or challenging to treat explicitly.[132]

To efficiently model experimentally measured properties, the above mentioned methods need to be combined in a tailored workflow. Workflows are automated procedures which provide a series of predefined customized tools with an optimal cost-to-accuracy ratio to serve specific target questions and properties. This is illustrated schematically in Fig. 1.3. Common workflows in first principles CMD include conformer search and ranking for example with the CREST and CENSO programs,[133,134] high-throughput screening for databases,[135,136] automated transition state search,[137] or interaction site screening with genetic algorithms.[138] Many workflows also have become best practices in computational chemistry, such as a multilevel approaches for the calculation of reaction Gibbs free energies or reaction mechanism elucidation.[139] Workflows are special-purpose methods concerning their outcomes but can handle very diverse chemical input. For instance, docking can be performed with a transition metal complex and a ligand, a surface and an adsorbate, or a protein and a drug.

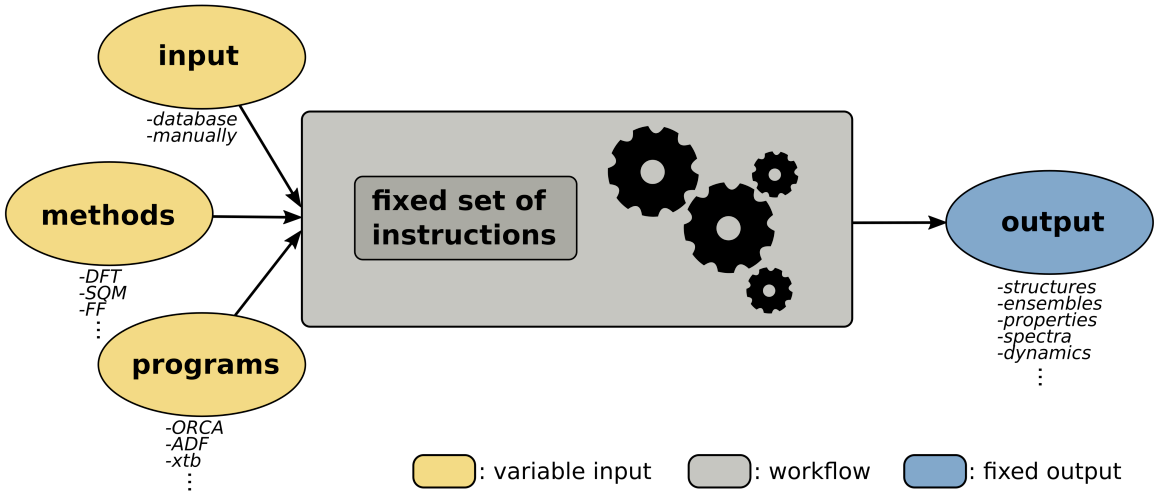

Figure 1.3: Schematic representation of workflow design and function.

In this thesis, several distinct workflows for different target properties and chemical compound spaces are presented. Following the introduction of some theoretical basics in Chapter 2, a workflow for the structure generation of molecular muscles and the investigation of their mechanical properties is detailed in Chapter 3. In Chapter 4 the structural and electronic properties of anti-aromatic carbon-nanohoops are assessed. In Chapter 5 a workflow is described for evaluating the structural, energetic, and mechanical properties of carbon-rich molecular spoked wheels. Subsequently, in Chapter 6 a workflow for determining the structural, energetic, and electronic properties of COFs in the solid state is presented. The focus in Chapter 7 lies on the development of a method to describe charge transfer in merocyanines, while a workflow for generating monolayers and molecular aggregates is introduced in Chapter 8 using merocyanines as an example. Finally, a summary is provided in Chapter 9, placing the findings in a broader context, and discussing impact and future challenges. Appendices A to F contain the original publications presented in Chapters 3 through 8.

CHAPTER 2

Theoretical Background

2.1 Electronic Structure Theory

The heart of computational chemistry is the description of quantum-mechanical systems through mathematical equations. Since analytical solutions of the underlying physical laws are only available for a few simple systems, significant approximations are needed for real world problems. In this chapter some basic concepts of quantum chemistry used throughout this thesis are explained starting with the fundamental description of electronic systems and an introduction into Hartree-Fock and density functional theory, concluding with concepts for more realistic descriptions such as solvation models, dispersion corrections, and semiempirical quantum-mechanical methods. This chapter is based on Refs. [112, 139–144] and for more in-depth descriptions the reader is referred to this original literature.

2.1.1 Schrödinger Equation and Hamiltonian

In classical mechanics, particles are associated with a continuous position, momentum and energy. Quantum mechanics accounts instead for mixed wave-particle properties, most notably the quantized nature of momentum and energy as well as the possibility of superpositions between different states. The equation of motion in quantum mechanics governing the evolution of the wave function in time and space is the time-dependent Schrödinger equation:

$$\hat{H}\Psi(r,t) \;=\; \mathrm{i}\hbar\frac{\partial\Psi(r,t)}{\partial t} \tag{2.1}$$

However, the wave function can be separated into temporal and spatial components by introducing a product ansatz for the overall wave function. This leads to the time-independent Schrödinger equation:

$$\hat{H}\psi(r) \;=\; E\psi(r) \tag{2.2}$$

whereas Ψ is the wave function representing the system depending on the time (t) and space (r) coordinates, ψ describes only the spatial wave function, $\hat{H}$ is the Hamilton operator describing the sum of kinetic and potential energy of the system, and E is the energy. Depending on the system and method both ψ and $\hat{H}$ become approximated later.

The electrostatic interaction of the charged electrons and nuclei is the dominant potential. Meanwhile, the kinetic energy is dominated by the electrons since the nuclei are much heavier and slower. The Born-Oppenheimer (BO) approximation states that systemic contributions can be separated if the particles move on different time scales and thus allows a classical treatment of the nuclei. Thus, the Hamiltonian for a molecular system can be formulated as follows:

$$\hat{H} = \hat{T} + \hat{V} = \hat{T}_e + \hat{T}_\mathrm{N} + \hat{V}_\mathrm{ee} + \hat{V}_\mathrm{NN} + \hat{V}_\mathrm{eN} \tag{2.3}$$

with $\hat{T}$ being the kinetic energy operator of electrons (e) and nuclei (N), and $\hat{V}$ the potential energy operator electron-electron interaction (ee), nuclear-nuclear interaction (NN) and electron-nuclear interaction (eN). The kinetic energy operator is defined as:

$$\hat{T} = -\sum_{i}^{N_\mathrm{el}} \frac{\hbar^2}{2m_i} \nabla_i^2 \tag{2.4}$$

with m_i either the mass of the electron in the case of $\hat{T}_e$, or the mass of the nuclei in the case of $\hat{T}_\mathrm{N}$, and ∇ being the Laplacian operator for spatial differentiation with respect to the coordinates in all three directions of space for each particle. The BO approximation allows straightforward formulations for $\hat{V}_{NN}$ and the one-electron operator $\hat{V}_\mathrm{eN}$ according to Coulomb's law and yields $V_\mathrm{NN} = const$ and $T_\mathrm{N} = 0$:

$$\hat{V}_{IJ} = \sum_{I}^{N_\mathrm{at}-1} \sum_{J=I+1}^{N_\mathrm{at}} \frac{Z_I Z_J e^2}{|R_I - R_J|} \quad ; \quad \hat{V}_{Ij} = -\sum_{I}^{N_\mathrm{at}} \sum_{j}^{N_\mathrm{el}} \frac{Z_I e^2}{|R_I - r_j|} \tag{2.5}$$

with r being the distance between the interacting particles, e the electronic charge, and Z the atomic number, upper case letters I and J are for nuclei and lower case letters i and j for electrons. The classical Coulomb repulsion as part of the two-electron operator $\hat{V}_\mathrm{ee}$ is defined as:

$$\hat{V}_{ij} = \sum_{i}^{N_\mathrm{el}-1} \sum_{j=i+1}^{N_\mathrm{el}} \frac{e^2}{|r_i - r_j|} \tag{2.6}$$

The complete explicit interaction between electrons $\hat{V}_\mathrm{ee}$ is the most complex contribution encompassing both classical contributions like the Coulomb repulsion and non-classical parts such as exchange and correlation. Approximations of the exact wave function and the description of $\hat{V}_\mathrm{ee}$ are at the heart of quantum chemical methods and are discussed in the following.

2.1.2 One-Electron Wave Functions & Basis Sets

The exact expression for the electronic wave function ψ_e is highly complex due to its at least 3N-dimensionality with N the number of electrons. A common approximation is to write the N-electron wave function as a combination of N one-electron wave functions, generally referred to as molecular orbitals (MOs, Φ_i), which are easier to determine and easier to work with. A simple product ansatz does not satisfy the requirements for a valid wave function, especially the Pauli principle, which leads to the construction of Slater determinants instead. A Slater determinant is an anti-symmetrized matrix

product of orbital resolved one-electron wave functions and looks as follows:

$$\Psi \approx \Phi_{\mathrm{SD}}(1,2,\cdots,N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{vmatrix} \tag{2.7}$$

whereas $\phi_i(k)$ denotes the i^{th} MO occupied by the k^{th} electron $\in \{1,\cdots,N\}$. The MOs are orthonormal and products of a spatial and a spin orbital. To obtain a further simplification for the MOs entering the Slater determinant, the linear combination of atomic orbitals (LCAO) ansatz is frequently employed to construct them from simpler AOs:

$$\Phi_i = \sum_a c_{i,a}\chi_a \tag{2.8}$$

whereas $c_{i,a}$ are the MO expansion coefficients and χ_a the atomic orbitals (AOs). These AOs can now be represented either by actual atomic orbitals, or by simpler physically motivated functions such as plane waves, exponentials, polynomials, or Gaussians, depending on the system. Such basis functions have to (approximately) fulfill certain boundary conditions, i.e., square integrability, cusp-conditions and some practical conditions such as efficient evaluation of operator integrals. For molecular problems Gaussian functions have proven to be well suited due to the ease of integration facilitated by the Gaussian product theorem. However, one Gaussian is a very poor physical description, e.g., wrong steepness and wrong cusp behavior as is apparent from the analytically known Slater function solutions to the hydrogen atom. Both the missing cusp at the aufpunkt of the function and the too fast decay are problematic. Thus, a linear combination of multiple primitive Gaussians (PGTOs) is employed to represent each basis function while the coefficients (c_i) and exponents (ζ_i) become optimized in a self-consistent and variational procedure for each atom.

$$\psi^{\mathrm{CGTO}} = \sum_i^{N_{\mathrm{bas}}} c_i \psi_i^{\mathrm{PGTO}}\{\zeta_i\} \tag{2.9}$$

with CGTO the contracted Gaussian type orbitals and PGTO the primitive Gaussian type orbitals. With increasing number of both contracted and primitive basis functions the wave function becomes variationally more accurate but is also more costly to compute. The collection of all contracted Gaussians is then referred to as the basis set.

In the limit of infinitely many basis functions (complete basis set limit CBS), an arbitrary one-electron wave function can be described exactly. Errors arise from the use of an incomplete basis set in the form of the basis set incompleteness error (BSIE) and the basis set superposition error (BSSE). The BSIE simply describes the physically incorrect description of the wave function because too few basis functions are used to represent it. BSIE becomes smaller with increasing basis set size and in principle vanishes at the CBS limit. However, the CBS limit is impossible to reach but can be approximated by either sufficiently large basis sets or extrapolation methods. Practically, a basis set is called converged if the energy change falls below a certain threshold. A CBS extrapolation requires usually two energy calculations with different basis sets. Then, the difference is fitted to an

extrapolation function and the energy at the CBS limit is obtained. However, different extrapolation functions are employed for different energetic contributions, i.e., total energy or correlation energy, due to their different convergence behavior.[145] The BSSE describes the unphysical use of neighboring basis functions when the basis set on either fragment is incomplete. In finite basis sets fragments in an aggregate can borrow spatially close-lying unoccupied basis functions from a different fragment to variationally lower the energy of the system. Since the single unaggregated fragment lacks these functions, an imbalance in BSIE between the fragment and the aggregate arises. BSSE does not occur for complete basis sets. The BSSE can be corrected with a counterpoise scheme by introducing ghost functions, which are basis functions without atoms:

$$E_{AB}^{\mathrm{CP}} = (E_A(AB) - E_A(A)) + (E_B(AB) - E_B(A)) \tag{2.10}$$

$$E_{AB}^{\mathrm{corr}} = E_{AB}(AB) - E_{AB}^{\mathrm{CP}} \tag{2.11}$$

whereas the subscripts denote the fragments A and B and the letters in brackets denote the basis set of the respective fragments. BSSE is especially pronounced for interaction energies of intermolecular fragments but can also occur intramolecularly.

2.1.3 Hartree-Fock

Hartree-Fock theory (HF) is one of the earliest quantum-mechanical methods to treat chemically relevant multi-electron systems. It employs a single Slater determinant as wave function and an effective mean-field Hamiltonian. By minimizing the energy with respect to the coefficients in the LCAO expansion of the MOs according to the variational principle, the HF or Roothaan-Hall equation can be written in matrix form as follows:

$$\mathbf{FC} = \mathbf{SC}\varepsilon \tag{2.12}$$

The Roothaan-Hall equation is a generalized eigenvalue problem. Since the Fock matrix **F** depends through the density matrix on the coefficients, only an iterative solution is possible (self consistent field SCF). After convergence of the SCF, the HF energy can be written as follows:

$$E = \sum_{i=1}^{N} h_i + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(J_{ij} - K_{ij}) + V_{\mathrm{NN}} \tag{2.13}$$

with the Coulomb $\hat{J}$ and exchange $\hat{K}$ operators and $\hat{h}$ the one-electron operator. HF is a mean-field method which neglects the explicit pairwise interaction of electrons and replaces them with the interaction of each electron with the mean-field of the remaining electrons. The difference to the exact energy includes thus all effect due to explicit electron-electron interaction which is termed correlation:

$$E_{\mathrm{corr}} = E_{\mathrm{exact}} - E_{\mathrm{RHF}} \tag{2.14}$$

While the correlation energy is rather small in absolute terms, it is often essential to describe chemistry even qualitatively. Examples for errors due to missing electron correlation are too high barriers in bond breaking, unbound NCI complexes, high-spin favored transition metal complexes, or charge transfer states. Generally, HF theory yields qualitatively correct results for weakly correlated systems

but is inferior to DFT for more complex systems in terms of accuracy.

2.1.4 Electron Correlation

Mean-field models are an approximation to the multi-electron wave function. However, without a specific construction this simplification neglects electron correlation which is a crucial part of many chemical phenomena. The separation of electron correlation is not a physical phenomenon but arises purely from the chosen wave function and its approximations. HF theory accounts for same spin correlation (mostly Fermi correlation) due to the inclusion of the Pauli principle into the Slater determinant but not for opposite spin correlation (Coulomb correlation). A schematic dissociation curve of H_2 with RHF is shown in Fig. 2.1.

The static correlation arises from the poor description of the many-electron system by only one

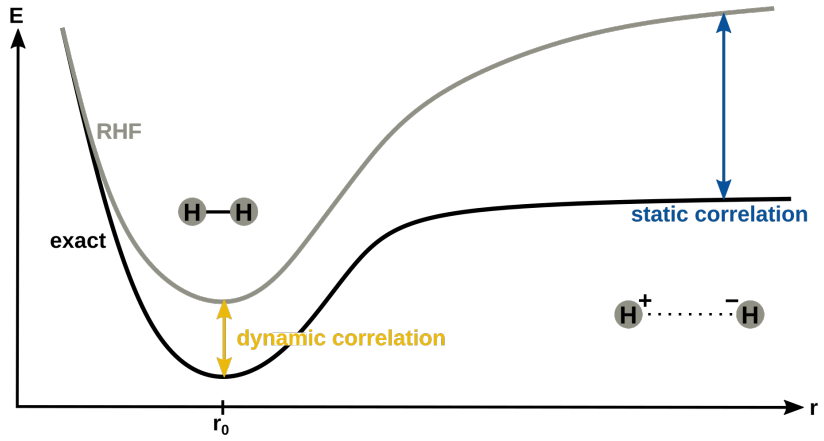

Figure 2.1: Schematic illustration of the dissociation curves of H_2 with restricted HF and the exact solution with markers for correlation sensitive areas.

determinant (single reference method) and can be addressed by a multi-reference treatment through the inclusion of several strongly contributing determinants. The dynamic correlation arises from the missing Coulomb correlation in the HF approximation that overestimates the probability of finding electrons in close proximity and thus increases the potential energy unphysically. This can be addressed by moving from the interaction of one electron with the average of all others to the explicit interactions, e.g., by adding further degrees of freedom in the representation of the wave function through excited determinants.

Post-HF methods such as coupled cluster (CC) or configuration interaction (CI), DFT, or unrestricted HF account for correlation. Improvements of the wave function description, e.g., through the inclusion of linear combinations of excited Slater determinants by row expansion or perturbation theory, lead to CI, CC, or Møller-Plesset perturbation (MP) methods.

Configuration interaction employs a linear combination of the ground and excited state Slater determinants to create the CI wave function. To restrict the summation to only those combinations yielding valid wave functions, i.e., eigenfunctions of the S^2 operator, spin-adapted linear combinations are used instead, also referred to as configuration state function (CSF). The number of CSFs may be

smaller than the number of Slater determinants, reducing the computational effort:

$$\Psi_{\mathrm{CI}} = \Psi_0 + \sum_{i}^{\mathrm{occ}} \sum_{a}^{\mathrm{virt}} c_i^a \phi_i^a + \sum_{ij}^{\mathrm{occ}} \sum_{ab}^{\mathrm{virt}} c_{ij}^{ab} \phi_{ij}^{ab} + \sum_{ijk}^{\mathrm{occ}} \sum_{abc}^{\mathrm{virt}} c_{ijk}^{abc} \phi_{ijk}^{abc} + \ldots \tag{2.15}$$

$$\Psi_{\mathrm{CI}} = (1 + \hat{T}) |\Psi_0\rangle = \Psi_0 + T_1\Psi_0 + T_2\Psi_0 + T_3\Psi_0 + \ldots \tag{2.16}$$

where $\hat{T}$ are excitation operators for single (S), double (D), triple (T), ... excitations. The CI coefficients, e.g., c_{ijk}^{abc}, are usually determined variationally in an iterative procedure. If all excitations are included, this is called full CI (FCI). The correlation energy is solely determined by the double excitations:

$$E_{\mathrm{corr}} = \sum_{abcd} c_{ab}^{cd} \langle \Psi_0 | H | \Psi_{ab}^{cd} \rangle \tag{2.17}$$

with Ψ_0 the ground state wave function, Ψ_{ab}^{cd} the doubly excited wave function, and c_{ab}^{cd} the corresponding CI coefficients. However, all excitations are crucial to determine the right coefficients. The number of CSFs to consider increases with the highest included excitation operator, i.e., CISD scales ON^6, FCI scales highly exponentially. Accordingly, FCI is currently only feasible for very small systems. Much more common are truncated CI methods that stop after the D- or T- or Q-excitation or that truncate the orbital space. However, truncated CI methods are not size-consistent, meaning that the sum of energies of all particles is not the same as the energy of the total system. There are many variants of CI such as multi-reference CI which employs more than one Ψ_0, or multi-configurational CI (MC-SCF) where orbitals are optimized alongside the c_i coefficients. Limiting the MC-SCF procedure to a selected orbital space, e.g., only valence electrons, where FCI can be performed, is called complete active space (CAS) SCF.

In contrast to CI where the excitation operators are expanded linearly, the CC excitation operators are developed in a non-linear approach, i.e., exponentially. Without truncation CI and CC converge to the same solution. The CC wave function is built as follows:

$$\Psi_{\mathrm{CC}} = e^T |\Psi_0\rangle = (1 + T + \frac{1}{2}T^2 + \frac{1}{6}T^3 + \ldots + \frac{1}{N!}T^N) |\Psi_0\rangle \tag{2.18}$$

where T is the excitation operator for S, D, and T excitations ($T = T_1 + T_2 + T_3 \ldots$), also called cluster operator which is expanded in a Taylor series in the second part. Contrary to CISD where doubles are the highest excitation included, CCSD includes in principle also all higher order excitations through the Taylor expansion that can be constructed from multiple applications of the T_1 and T_2 operators (denoted disconnected excitations). Thus, CC is size-consistent and therefore preferred over truncated CI for most applications. CC including singles doubles and pertubative triples (CCSD(T)) is usually regarded as the gold standard reference method in quantum chemistry feasible up to few hundreds of atoms with the right local approximation schemes.

An alternative approach to calculate the correlation energy is to see it as a small perturbation to the mean-field electronic structure and apply perturbation theory. Accordingly, the solution for the perturbed system can be derived from the solution of the unperturbed system by adding a small perturbation Hamiltonian. Møller-Plesset perturbation theory builds up onto HF and perturbs the Hamiltonian ($\mathbf{H}_0$) with a better expression for the electron correlation (perturbation operator $\mathbf{V}$). The perturbation parameter λ is expanded in a power series, finally yielding nth order corrections to the

total energy (MPn):

$$\mathbf{H} = \mathbf{H}_0 + \lambda \mathbf{V} \tag{2.19}$$

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} \ldots \tag{2.20}$$

E(MP0) is the sum of orbital energies (Hartree theory), E(MP1) equals E(HF), and E(MP2) is the first contribution to the correlation energy employing a doubly excited determinant:

$$E^{(2)} = \sum_{ijab} \frac{|\langle ij||ab\rangle|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{2.21}$$

with ε being the orbital energies of the occupied (i,j) and virtual (a,b) orbitals. It accounts for roughly 80-90% of the correlation energy and scales conveniently to the power of five. MP2 is for example employed to enhance the description of correlation in DFT, leading to the class of double-hybrid density functional approximations which will be discussed in Chapter 2.1.5. MP3 and MP4 account for even more correlation than MP2 but scale with up to the power of six and seven and are less frequently used. Disadvantages of MPn are that the correlation energy can be overestimated in an oscillating manner with the nth order and the perturbation series may show severe convergence problems at higher orders. MPn suffers from the similar methodical drawbacks as HF and is not suited to correct for static correlation as in near-degenerate or dissociating systems.

2.1.5 Density Functional Theory

In contrast to all above presented methods, DFT attempts to obtain the energy of a system not based on the wave function but rather the electron density ρ. According to the 1^{st} Hohenberg-Kohn theorem there exists a direct mapping between the energy and a density functional. However, the exact form of this functional is unknown. In an arbitrary partitioning of the total energy a functional F can be expressed as follows:

$$F[\rho] = T[\rho] + V_{\mathrm{Ne}}[\rho] + V_{\mathrm{ee}}[\rho] + E_{\mathrm{x}}[\rho] + E_{\mathrm{c}}[\rho] \tag{2.22}$$

whereas V_{ee} (sometimes also denoted $J[\rho]$) and V_{Ne} are the only known components. The major problem of using the electron density is that the functional for the kinetic energy is unknown, since it cannot immediately be deduced from the operator. To address this issue an auxiliary wave function is introduced, the Kohn-Sham (KS) determinant, analogous to a Slater determinant. It describes a non-interacting system with the same electron density as the reference system. DFT as it is practically applied, refers almost exclusively to KS-DFT. All non-classical contributions and the residual error of the kinetic energy approximation are absorbed into the exchange correlation functional. Thus, the choice of E_{xc} is crucial leading to the development of a wide variety of different density functional approximations (DFAs).
Potentially, DFT is faster than HF, scaling cubically with the system size in the case of purely semilocal DFAs and resolution-of-identity (RI) approximated Coulomb integrals. Thus, modern DFT programs are able to calculate systems with more than 1000 atoms. In contrast to HF, DFAs suffer from the self-interaction error (SIE) since its account for Fermi correlation is incomplete, allowing the Coulomb interaction of an electron with itself. Although DFT is solved by variational minimization of the energy, it is not systematically improvable in the way post-HF methods improve by including

higher-order excitations in CI or CC. Additionally, most standard DFAs employ a high degree of empiricism and have the wrong asymptotic behavior, i.e., do not account for static correlation and miss out on long-ranged dispersion interactions. Dispersion corrections are treated in Chapter 2.1.6. In general, DFT overdelocalizes and overpolarizes molecules, reaction barriers are underestimated, electrons are too weakly bound, excitations are at too low energies. All these drawbacks are addressed with hybrid DFAs because HF overlocalizes, overestimates barriers, and binds electrons too strongly. For applications, DFT is preferred over WFT as it is usually faster, can be applied to larger systems, allows for easier customization to special problems, and a lot of correction schemes are available. DFAs can be classified according to the components used to construct the energy functional compared to Hartree theory. This is commonly represented with Jacob's ladder as shown in Fig. 2.2.

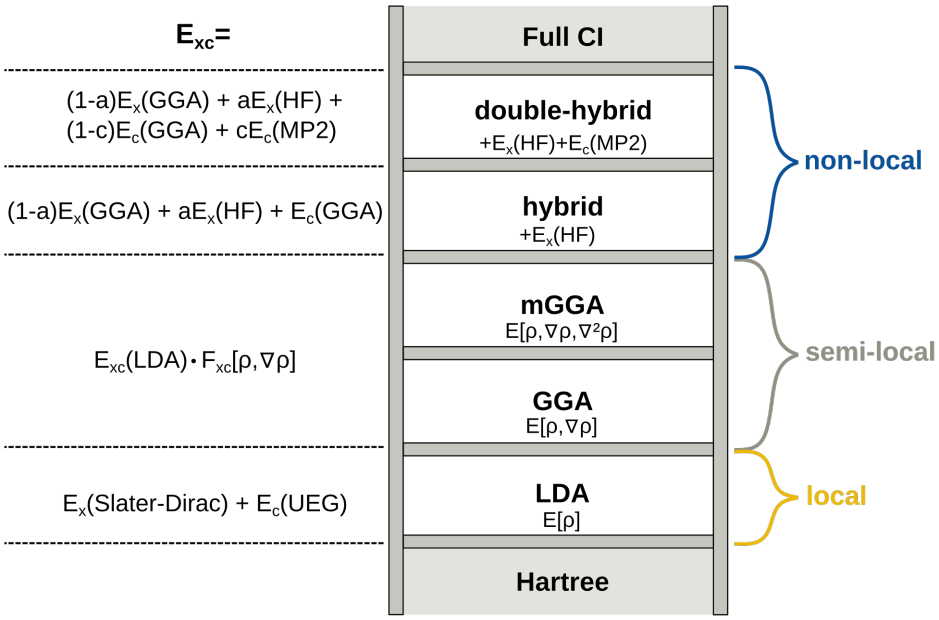

Figure 2.2: Schematic illustration of Jacob's ladder about which information are included by different DFA categories and how they calculate the exchange-correlation.

A higher rung in Jacob's ladder often means a better performance but also increased computational effort due to the higher scaling of the more complex ingredients. The lowest rung in Jacob's ladder, the local density approximation (LDA), employs e.g., Slater-Dirac exchange and for the correlation e.g., the uniform electron gas formulation. LDA functionals tend to overbind and they are rarely used in the study of molecular systems. However, LDA is quite popular in solid state physics. The generalized gradient approximation (GGA) uses the gradient of the electron density ($\nabla\rho$) to describe the semi-local environment better. It employs the LDA exchange-correlation and a scaling factor that depends on ρ and $\nabla\rho$. Different GGA DFAs differ in the specific form of the scaling factor and the use of reference data for the determination of parameters. meta-GGAs (mGGA) additionally depend on higher order derivatives of ρ. They usually yield better geometries but not necessarily better energies than GGAs. Hybrid DFAs partially substitute the GGA exchange by Fock exchange and thus include non-local features, correct for the SIE, and increase the overall accuracy. However, the scaling increases by one order of magnitude to the power of four due to the inclusion of Fock

exchange which is less amenable to the RI approximation. Different functionals vary in the amount and specific form of admixed Fock exchange and a property specific choice is required, i.e., high amounts for excitation spectra or barriers and low amounts for transition metal chemistry. Thermochemistry, transition states, NCIs, dissociation, and barriers can be described well by hybrid DFAs. A special variant are range-separated hybrids (RSHs) where the amount of Fock exchange depends on the interelectronic distance, i.e., high amounts in the long-range and low amounts in the short-range. The correct long-range asymptotics improve charge transfer and NCIs while the short-range DFT includes some static correlation. Optimally tuned (OT) RSHs adjust the range-separation parameter ω specifically for each system, rather than relying on the globally fitted ω-value. This system-specific tuning enhances the accuracy of exchange-dependent properties, such as excited states or optical spectra. The uppermost rung of Jacob's ladder is represented by double-hybrid DFAs which not only employ Fock exchange but also substitute partial amounts of the (m)GGA correlation by e.g., MP2 correlation to include virtual orbital information. They scale to the power of five and are feasible up to few hundreds of atoms. If CC is too costly as reference method, usually double-hybrids are used.

2.1.6 Dispersion Correction

Dispersion is a non-covalent long-ranged correlation effect arising from the interaction of two induced multipoles. It is predominantly attractive although higher order interactions and multibody terms also contribute smaller usually repulsive parts. Dispersion is highest between large, easy polarizable contact surfaces like in organic aromatic molecules or small apolar molecules like rare or organic gases. Dispersion is not considered in standard DFT or HF since both are purely based on the occupied orbitals. However, the inclusion of electron correlation accounts for it through the encoding of virtual orbital information by admixture of excited determinants in WFT based methods. Neglecting dispersion leads to underbound or not at all bound NCI complexes with a major impact on all calculated properties to the point of being qualitatively incorrect. To address this severe issue, several dispersion corrections have been developed, among other dispersion corrected DFT (DFT-D with D ranging over D2,[144] D3,[146] D4[141]), Tkatchenko-Scheffler model (TS),[147] exchange-dipole moment (XDM),[148] or multibody approaches such as many body dispersion correction (MBD)[149] or Axilrod-Teller-Muto three body correction (ATM).[150,151] Alternatively, specific density functional methods such as van-der-Waals density functionals (vdW-DF)[152] and VV10[153] include long-range correlation explicitly into the design of the correlation functional.
Here, the semi-classical DFT-D models are discussed in more detail. They are based on a multipole-expansion of the dispersion interaction. The DFT-D3 dispersion is based on the following formula:

$$E_{\text{disp}}^{\text{D3}} = -\frac{1}{2}\sum_{i,j}\sum_{n} s_n \frac{C_n^{ij}}{R_{ij}^n} f_{\text{damp}}(R_{ij}) \tag{2.23}$$

where ij are atom pairs, C_n is the nth order dispersion coefficient ($n = 6, 8$), s_n is a global DFA specific parameter, and f_{damp} a distance dependent damping function. The C_6 coefficients dominate the long-range dispersion while higher orders contribute more to the medium- and short-range dispersion. These atomic coefficients depend on the cardinal number, ionization potential, and dipole polarizability. They are commonly derived from similar first-principles expressions like in vdW-DFs, i.e., from time dependent DFT. This C_6-based approach is the basis for many pair-wise additive dispersion correction

schemes rooted in a second order perturbation theory ansatz. Advantages are the easy and quick post hoc usage and its suitability even for force fields because no underlying electronic structure method is needed as C_6 coefficients are pre-computed.
D3 improves upon the D2 expression by a dependence of the C_6 coefficients on a coordination number based weighting of reference systems, and by including C_8 coefficients. This treatment accounts for locally different chemical environments, correct asymptotics, and enhances the description of complex hybridization and coordination patterns, e.g., in transition metal complexes. Furthermore, different damping schemes on the basis of functional specific cutoff radii are available enhancing the short-range description of dispersion contribution, e.g., zero damping, BJ damping[122,154].
D4, the successor model to D3, tackles two problems of the C_6-based methods. It introduces anisotropy and treats cations and anions differently. Therefore, D4 employs an atomic partial charge dependent scaling of the polarizabilities by chemical hardness and not only a dependence on the coordination number but also on electronegativity differences. The D4 formula reads as follows:

$$E_{\text{disp}}^{\text{D4}} = -\frac{1}{2}\sum_{i,j}\sum_{n=6,8} s_n \frac{C_n^{ij}}{R_{ij}^n} f_{\text{damp}}^n(R_{ij}) - \frac{1}{6}\sum_{i,j,k} s_9 E_{\text{ATM}}^{ijk} f_{\text{damp}}^9(\bar{R}_{ijk}, \theta_{ijk}) \tag{2.24}$$

with s being DFA specific global scaling factors, C_n being the fitted coefficients that describe dipole-dipole (C_6), dipole-quadrupole (C_8), and triple-dipole (C_9) interactions. The E_{ATM}^{ijk} energy is described further below. Similar to D3, D4 employs empirical parameters only in the damping function, retaining calculated values for the C_6 coefficients. To achieve the new electronegativity dependent scaling of the pre-computed dynamic polarizabilities, a new charge model was introduced, called electronegativity equilibrium model (EEQ).[155] The EEQ results in slightly higher computational costs for D4 compared to D3 but a better and more chemically motivated description of charged species. However, it is still applicable to large systems and force field methods.
Also the TS dispersion correction is based on the calculation of a C_6 coefficient. Unlike D3, the coefficients are derived from the electron density.
VV10 is a vdW-DFA and a non-local addendum to the correlation part of the functional attenuated with a damping function to converge to the local solution. When applied post-SCF it uses the pre-calculated final electron density to derive the non-local correlation energy which is simply treated additively. Although a self-consistent treatment is more accurate, it is also more computationally expensive due to the constant re-evaluation of the additional double space integral.
Three body dispersion corrections are especially important for large systems with many atomic triangles formed in the molecular structure and can easily amount to 10% of the pair-wise dispersion energy.[156] The ATM correction contributes a repulsive energy, is angle (θ) and distance (R) dependent, and usually applied on top of a pair-wise correction. The characteristic parameter is the C_9 coefficient derived from the combination of pair-wise C_6 coefficients.

$$E_{\text{ATM}}^{ijk} = C_9^{ijk} \frac{3\cos(\theta_i)\cos(\theta_j)\cos(\theta_k) + 1}{(R_{ij}R_{ik}R_{jk})^3} \tag{2.25}$$

MBD treats the coupled fluctuating dipoles of the atoms as harmonic oscillators and calculates the eigenvalues of their interaction matrix to account for in principle all many body dispersion effects up to infinite order.
Summarizing, for meaningful results a dispersion correction should generally be employed. Some

DFT or SQM methods favor the use of certain dispersion corrections or are already distributed with a dispersion correction included. A three body correction should be employed additionally for systems exceeding the size of several dozens of atoms. It should be noted that some methods already account for dispersion implicitly either through parametrization or other electron correlation effects included. In this case, the blind application of a dispersion correction might lead to overcompensation and instead mindful calibration e.g., in the form of scaling or damping functions, is needed.

2.1.7 Semi-empirical Tight-binding Methods

Density functional tight-binding (TB) methods are developed as a Taylor series of the density functional expression in orders of the density fluctuation relative to the reference density of neutral non-interacting atoms (ρ_0). Depending on where this expansion is truncated, different TB methods result, e.g., first order truncation yields a non-selfconsistent TB ansatz, second order truncation results in DFTB2, third order in DFTB3. The GFN1 and GFN2 methods are based on a third-order expansion in the framework of DFTB3:

$$E[\rho] = E^{(0)}[\rho_0] + E^{(1)}[\rho_0, \delta\rho] + E^{(2)}[\rho_0, (\delta\rho)^2] + \ldots \tag{2.26}$$

$$E_{\text{tot}} = E_{\text{el}} + E_{\text{rep}} + E_{\text{elstat}} + E_{\text{disp}} + E_{\text{xc}} + \ldots \tag{2.27}$$

The idea behind the TB expansion is that the smaller relative contributions compared to the reference system allow for efficient approximation, saving computation time. The most fundamental is the expression of all integrals by approximate integrals involving at most two-center integrals, i.e., mostly overlap integrals. Based on the energy expressions in different orders of the TB expansion, Hamiltonian contributions can be defined, leading again to a Roothaan-Hall type equation:

$$\mathbf{HC} = \mathbf{SC}\epsilon \tag{2.28}$$

However, the elements of $\mathbf{H}$ need to be approximated. The density dependency of $E[\rho]$ is substituted by a charge dependency $E[q]$ which is easier to calculate through monopole approximations. Thus, no longer a self-consistent field is evaluated but the charges are iterated self-consistently. The elements of $\mathbf{H} = \langle \phi_\mu | H_0 | \phi_\nu \rangle$ are obtained through fitting of usually atom-pair wise parameters with respect to DFT reference calculations. In the xTB methods notably only element-specific parameters are used. These parameters can depend on various chemical properties such as chemical hardness, coordination numbers, or polarizability. Through the fitting procedure properties of the underlying reference method can implicitly be accounted for, i.e., partly compensation of BSIE in the frequently employed minimal basis, SIE, or exchange. Additionally, TB methods, as well as most other SQM methods, usually employ further corrections such as hydrogen bond correction or dispersion correction. Although originally developed for solid state physics, modern TB methods aim for a more general application and are capable to describe large systems sized up to a few thousands of atoms with most elements contained.

GFN1-xTB and GFN2-xTB are TB methods derived from third order truncations specifically designed for geometries, frequencies, and non-covalent interactions (GFN). They employ the D3 and D4 dispersion correction to account for E_{disp}, use a minimally polarized basis set, and avoid pair-potentials for the benefit of atom-specific parameters to reduce the complexity of the parametrization. In contrast to its predecessor GFN1-xTB, GFN2-xTB goes beyond the monopole charge approximation and

employs multipoles up to quadrupoles to account for aniosotropic electrostatic effects and anisotropic exchange of higher order than GFN1-xTB, rendering the use of a halogen bond correction as in GFN1-xTB obsolete. Additionally, D4 provides a refined dispersion treatment. Shortcomings of the GFN methods are for example their poor treatment of low-bandgap systems and underestimation of electronic gaps due to the underlying GGA-like theory, a slight tendency to overbinding of charge-transfer complexes, or rearrangement of some transition metal ligand coordination motifs. Due to the convenient integration into the `xtb` program suite the GFN methods can easily be included into workflows and other program architectures enabling high-quality investigations of large systems that can no longer be treated DFT methods or high-throughput screening where DFT methods would be too slow. Overall, the GFN*n*-xTB methods are more generally applicable and in many cases more accurate than previous HF-based SQM methods such as parametric methods (PM*x*), Austin models (AM*x*), or other neglect of differential overlap (NDO) methods. Special purpose xTB variants exist, e.g., for spin polarization,[157] optical spectra,[158] machine-learning assisted differentiable xTB,[159] or a non-self-consistent variant,[160] enhancing the versatility of these methods even further.

2.1.8 Implicit Solvation

Solvation plays a crucial role in chemistry as most reactions, especially biochemical ones, take place in solution. Computationally, solvation can be modeled explicitly by including complete solvent molecules in the simulation, or implicitly by representing the solvent as a continuum. Explicit solvation is very costly and usually not feasible for large solutes due to both the larger system size and the need for extensive sampling of the solvent ensemble. Alternative approaches include micro-solvation, where only few explicit solvent molecules are considered to capture the most important interactions, or embedding schemes with a combined explicit and implicit solvent treatment. Implicit solvation has some benefits beyond the reduced computational cost, mainly the inherent inclusion of sampling effect through the use of macroscopic experimental constants such as the relative solvent permittivity (dielectric constant ε). In the following, the theory behind implicit solvation is reviewed.
Implicit solvation[161] describes the solvent as uniform and polarizable continuum characterized solely by its dielectric constant. The solute is placed into a cavity in this continuum. Thus, the solvation free energy contains the energy for forming this cavity (ΔG_{cavity}), the dispersion interaction between solvent and solute (ΔG_{disp}), and the electrostatic energy arising from the polarization of the continuum through the solute (ΔG_{elstat}):

$$\Delta G_{\text{solv}} = \Delta G_{\text{elstat}} + \Delta G_{\text{cavity}} + \Delta G_{\text{disp}} \tag{2.29}$$

The solute can be described quantum-mechanically or classically depending on the underlying method (DFT, SQM, force field). The continuum can be described by a stationary dielectric constant or a dielectric parameter that depends on the distance from the solute (r) or the frequency of electronic movement. The latter approaches are far less frequently employed.[162] The cavity may be a simple spherical or ellipsoid cavity or a more system-specific molecule-shaped solvent accessible surface either generated from the atomic radii scaled with an empirical factor or from properties of the wave function such as electron density isosurfaces. While there are several accurate methods available for the description of ΔG_{elstat}, the contributions of ΔG_{cavity} and ΔG_{disp} are usually treated separately. Followingly, some theories for the calculation of ΔG_{elstat} are presented.[161]
The Poisson model relates an arbitrary charge density ρ to the induced electrostatic potential ϕ via

the solvent-specific dielectric constant ε. The Boltzmann model describes the thermal fluctuations of the density ρ (often referring to further solvated electrolytes) to the electrostatic potential ϕ. The combined Poisson-Boltzmann model can be used to describe the response of a (electrolyte) solution to the introduction of a (solute) density. The working equation[163]:

$$\nabla(\varepsilon(r)\nabla\phi(r)) \; - \; A \; \cdot \; \sinh(B \; \cdot \; \phi(r)) \; = \; C \; \cdot \; \rho(r) \tag{2.30}$$

contains three parameters A, B, and C depending on the temperature, solute charge density, and solvent ion strength. For the borderline cases of a weakly polar/charged solute or a strongly screening (high ε) solvent of low ion strength, this equation can be linearized efficiently by e.g., eliminating the *sinh* function through a Taylor expansion. Such an approach is for example used with the ALPB solvation model which will be described below. In general, PB models are computationally demanding due to the need for solving coupled differential equations in three dimensional space. However, the electrostatic component of ΔG_{solv} is accurately described and enables a reliable implicit treatment of strong polarization effects.

The Polarizable Continuum Model (PCM)[164] family is based on the Poisson model but substitutes the three-dimensional differential equations of the volume polarizability for the calculation of ΔG_{elstat} through a two-dimensional surface charge treatment that is less computationally demanding. This simplification is possible because any density fluctuation in a cavity embedded in a solvent, results in change of the electrostatic potential on the surface. The response of the outside dielectric embedding can be exactly reformulated based on apparent surface charges. Thus, the PB equation can be solved exactly on a surface, under certain conditions such as no density protruding into the dielectric. Within those limitations, exact models are available even for realistically sized systems. However, often times slight approximation such as the conductor like screening model COSMO or CPCM are applied to speed up calculations with minor losses in accuracy. For fast semiempirical models, more drastic approximations are required to make it sufficiently fast. One example is the Born model, discussed in the following.

The Born model approximates the cavity by a sphere and expands the solvent density in a multipole expansion retaining only the monopoles. Thus, it assumes a uniform charge distribution consistent with its initial derivation for the solvation free energy of simple ions. Hence, it misses any anisotropy in the solute density and also polarization effects. This leads in the simplest form to a mere screening of the Coulomb interaction between the surface/ionic charges and the monopole:

$$\Delta G_{\text{elstat}}(q) \; = \; -(1 - \varepsilon^{-1}) \, \frac{q^2}{2r} \tag{2.31}$$

This circumvents the costly solution of integral equations and matrix inversions needed in the PCM methods. The Born model is a strongly simplified description of real systems thus, more accurate Born-based models have been developed. There are several extensions to the basic Born model such as the Onsager model which accounts for dipoles in a spherical cavity or the Kirkwood model which accounts for multipole terms in the density expansion and ellipsoid cavities. Another highly practical and widely used extension is the generalized Born (GB) model that employs a more molecule-shaped cavity by using spheres for each atom with specific van-der-Waals, Born, or Bondi radii, effective radii of the solvent-shielded atomic partial charges. The resulting anisotropic cavity can now account for polarization, while maintaining close to the same computational efficiency as the underlying Born theory. However, one of the persisting drawbacks of the PB model class is the symmetric

treatment of positive and negative charges, i.e., especially for negatively charged species the ΔG_{solv} are less accurate. Correction schemes like the Dynamic Radii Adjustment for Continuum Solvation (DRACO),[165] address this issue effectively.
Combined with terms accounting for ΔG_{disp} and ΔG_{cavity}, various solvation models are frequently employed for the calculation of ΔG_{solv}, including GBSA,[166] ALPB,[167] COSMO and COSMO-RS,[168] and SMD.[169] An overview of their specific characteristics will be presented in the following.
The Generalized Born / Surface Area (GBSA) model employs the GB model and enhances it with a position dependent dielectric constant for the solvent description and a solute surface dependent description of ΔG_{disp}. This model is fast and usually applied for SQM or force field calculations of large (bio)molecules. The Analytically Linearized Poisson-Boltzmann (ALPB) model employs a linearized PB model with a molecule-shaped cavity plus eventually a treatment of higher order multipoles, improved smoothing between the solute and solvent region, and a hydrogen-bond correction term. It is slower than GBSA but more accurate and still able to be efficiently used in combination with SQM methods.
DFT is frequently combined with PCM models such as the COnductor-like Screening MOdel for real solvents (COSMO-RS). As mentioned above, the continuum is approximated as a perfect conductor ($\varepsilon = \infty$) to simplify the evaluation. The extension from COSMO to COSMO-RS includes non-electrostatic ΔG_{solv} contributions via the difference between a gas- and (COSMO-based) solvent-phase density. Additionally, statistical thermodynamics to account for more of ΔG_{disp} are employed. On the other hand, Solvent Model Density (SMD) is a hybrid model with its roots in the GB model. It includes non-electrostatic contributions by an empirical relation to the atomic surface tensions and acidity parameters. Generally, a high accuracy for ΔG_{solv} requires extensive empirical fitting to experimental data to model its complex dependency on the geometric and electronic structure. Both, COSMO-RS and SMD, provide reasonably accurate values for ΔG_{solv} at least in the regime of apolar, neutral molecules.

2.2 Workflows

A workflow is a specific set of calculations designed to obtain specific properties using predefined methods, settings, and procedures. These can take the form of best practice manuals, automated black-box programs or run types, and semi-automated scripts that integrate multiple programs. Workflows play a crucial role in ensuring the reproducibility and comparability of data, which is fundamental to any science. They preserve expertise and facilitate knowledge transfer between researchers. By breaking down complex problems into smaller, well-understood components, workflows make challenging tasks more accessible. Importantly, a well-designed workflow does not require the user to fully understand it to apply it effectively, i.e., black-box nature. The following subsections present first some fundamental components of quantum-mechanical workflows of general importance and then some specifically used for generating structure ensembles. This chapter is based upon Refs. [112], [133], and [138]. For more comprehensive insights, the reader is referred to this literature.

2.2.1 Molecular Dynamics

Molecular dynamics (MD) simulations calculate the propagation of nuclear motion in molecules over time mostly treating the nuclei classically based on Newton's equations of motion:

$$M_i \frac{\partial^2 R_i}{\partial t^2} = F_i \tag{2.32}$$

with M_i being the atomic masses, R_i the atomic positions and t the time, and F_i the forces acting onto atom i. Modeling the temporal development of a system is crucial for understanding reactions, mass and charge transport, mechanical properties, molecular movement, phase transitions, molecular recognition, and the behavior of mixtures.[170] Although MD simulations can theoretically be combined with any method able to calculate forces, they are often used with classical force fields due to their substantial computational demands.
Since the forces on all atoms are coupled in complicated patterns, the resulting set of coupled linear differential equations has to be solved by discrete numerical integration. Simplifications include the use of discrete timesteps and truncated Taylor expansions to approximate atomic positions after small time intervals Δt. The choice of timestep size depends on the system and method: small, slow, or rigid systems can tolerate larger timesteps, while flexible, fast-moving, or large systems require smaller timesteps, leading to increased computational cost but greater stability in MD simulations. Several algorithms can be employed to determine atomic positions in the next timestep. Either, they require a memory of two timesteps or calculate additional derivatives trading memory for computational costs. The advantage of MD simulations over static approaches is their ability to account for temperature and pressure, which are derived from the system's degrees of freedom (i.e., kinetic energy) according to the Virial theorem. Temperature and pressure can be controlled through thermostats and barostats, typically by scaling or preserving certain properties such as velocities or heat capacities, or by using thermal reservoirs to extend the system. Additional customization options include adjusting timestep, runtime, temperature, pressure, volume, chemical potential, atomic mass, perturbing potentials, and constraints. However, these parameters may be interdependent and cannot all be controlled simultaneously, e.g., an NVT ensemble keeps the number of particles, volume, and temperature constant but must allow for pressure fluctuations.

MD simulations provide a wealth of insights especially important for CMD, often analyzed using statistical methods. This includes domain analysis, which offers information on spatial proximity (e.g., in solvent shells), structural memory, reactions, and correlated movements. MD simulations can also reveal structural details of processes such as isomerization, folding, disentanglement, host-guest recognition, phase ordering, and phase transitions. Additionally, time-dependent properties like viscosity, mechanical resistance, diffusion, and conductivity are accessible. MD simulations are frequently integrated into multistep workflows and offer versatile applications beyond static single-structure approaches.

2.2.2 Ensemble Effects

If an ensemble of structures is thermally accessible, measured properties depend on the Boltzmann weighted average over all individual structures. Accordingly, to realistically model an experiment, a well balanced ensemble is essential. Single-structure approaches, in contrast, typically capture only isolated snapshots of a system in a given configuration that might be only realized for a fraction of the systems at any given time. These differences between single-structure and ensemble approaches is schematically depicted in Fig. 2.3 for the examples of reaction energies, optical spectra, conformers, mechanical properties, and polymorphs.

In general, an ensemble approach results in a more continuously averaged property rather than a single

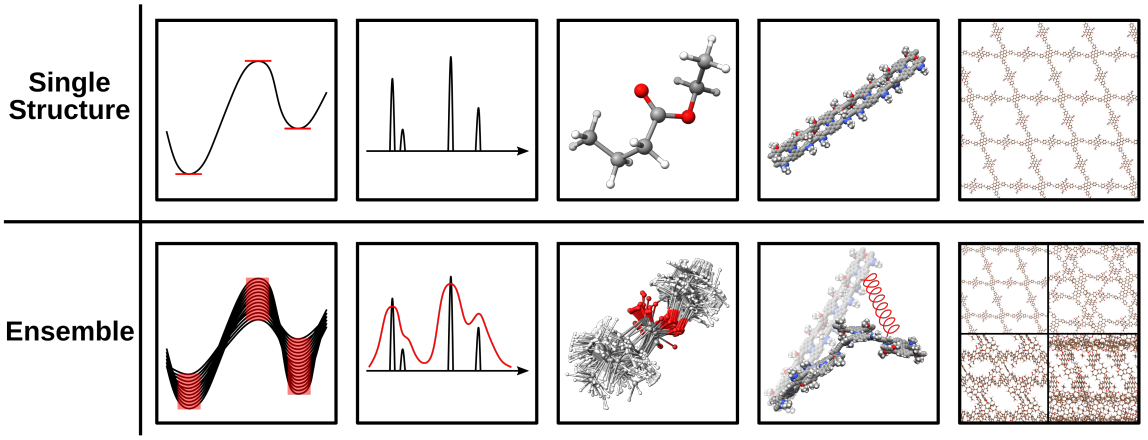

Figure 2.3: Schematic representation of the differences between single structure and ensemble approaches. From left to right examples of reaction energies, optical spectra, conformers, mechanical properties, and polymorphs.

number mostly due to thermal effects. This is most pronounced for flat PES, e.g., large and flexible (organic) structures. For instance, a single-structure approach might produce a singular value for the reaction energy, while an ensemble approach results in a range, with the single-structure value ideally being the lower boundary. In optical spectroscopy, ensemble averaging typically leads to spectral broadening, as different conformers contribute different peak positions. This effect especially applies to spectroscopic techniques like NMR and EPR, both highly sensitive to the underlying structure, where simulated spectra might poorly match experimental data if the ensemble is inadequately represented or if the wrong structures are used.

Mechanical properties are crucial for material design, requiring simulations of how a material responds

to external stress. Protein folding, vital for biochemistry and pharmaceuticals, also depends on modeling motion and reactions to environmental stimuli like solvent. Ensemble approaches are indispensable in both CMD and CADD for making reliable predictions. Moreover, all forms of molecular docking inherently involve ensemble methods, mostly after an initial rigid docking step, including explicit solvation, adsorption, and aggregation. These processes involve generating multiple trial structures, which are iteratively refined or rejected until the desired docking configuration is achieved, reflecting the multidimensional and combinatorial nature of these problems.

2.2.3 Conformer Search

A conformer search is often the first step in molecular investigations. Conformers are thermally populated states of the same molecule, formed through rotations, bends, and ring inversions without breaking covalent bonds. Using the incorrect conformer or missing a Boltzmann averaged ensemble in property calculations can result in significant deviations from experimental outcomes. Computationally, conformational sampling is a global optimization problem on the potential energy surface (PES), searching for all low-lying and particularly the lowest minimum. Several global optimization techniques have proven particularly suitable for computational chemistry, as schematically illustrated in Fig. 2.4 . One of the oldest and most widely used methods for conformer generation involves taking diverse

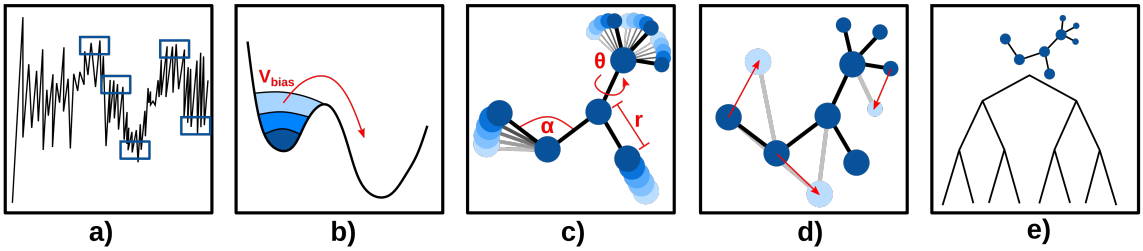

Figure 2.4: Schematic representation of different conformer search algorithms. a) molecular dynamic simulations, b) basin hooping and diffusion methods, c) PES exploration through systematic variation of DOFs, d) stochastic and Monte Carlo methods, e) genetic algorithms.

snapshots from molecular dynamic simulations (Fig. 2.4a) and reoptimizing them. This approach is still applied to very large and flexible molecules, where other methods might be too costly or too narrow. The conformers identified via MD depend heavily on the initial structure, as only barriers below the set temperature can be crossed, preventing the exploration of large parts of the PES. Consequently, MD primarily offers a local exploration of the PES, potentially missing significantly different conformers. Simulated annealing[171] can address this issue by starting at a high temperature to overcome higher barriers and explore a larger portion of the PES. Upon cooling, the molecule becomes trapped in different, hopefully diverse, minima, where local explorations can again be performed.
Diffusion methods offer a more targeted approach to finding global minima rather than trying to explore only the local PES (Fig. 2.4b). These methods perform minima hopping either by filling the current minimum with bias potentials, constrained local derivatives, or by tunneling through barriers. Meta dynamic simulation, as employed in the conformer rotamer ensemble sampling tool (CREST), is an example of the former approach. It performs a conformer search for medium sized molecules and generates conformer ensembles. As a widely used and automated workflow, it requires only input coordinates and can function as a black-box application. CREST employs several parallelized meta

and molecular dynamics to find the best settings (e.g., bias potentials, temperatures) for the given input, exploring both near (MD) and far (metaD) chemical space. This sampling can be performed with semiempirical quantum-mechanical (SQM) methods and force fields from the GFN family. CREST also uses a genetic Z-matrix crossing between the lowest-energy conformers to better account for rotamers and explore new parts of the PES. In the Commandline ENergetic Sorting (CENSO) program,[134] frequently employed together with CREST, the generated conformational ensembles are reoptimized and reranked at either the SQM or DFT level, depending on the user's choice. CREST/CENSO effectively covers a broad range of the PES, providing well-balanced conformational ensembles and global minimum structures.
Another intuitive option for PES exploration is the systematic variation of all degrees of freedom (DOFs) of a molecule (Fig. 2.4c) employing a grid-based search. Scanning all angles, dihedral angles, positions, and inversions is time-consuming due to the exponentially rising number of combinations with the number of flexible bonds, but in principle, it can yield all possible conformers. However, this approach is computationally infeasible for realistically sized molecules and is typically used only for specific regions of the PES by applying constraints to the DOFs.
Monte Carlo (MC) methods (Fig. 2.4d) represent a stochastic, non-deterministic approach to conformer search. An initial geometry is perturbed by random displacement of the DOFs, and the newly generated structure is accepted if its energy is lower than before or according to a probability criterion when the energy is higher, accounting for a Boltzmann distribution of states. Consequently, MC simulations can climb uphill on the PES. Although the unphysical displacement increases PES coverage by generating unintuitive new conformations and tunneling into nearby minima, small step sizes and long simulation times are necessary to maintain high acceptance ratios during the simulation's propagation. Focusing on or excluding certain DOFs is straightforward, allowing for specialized treatment of different compound classes and phases. As MC methods require only energy calculations and not additional gradients or forces, they are easy to implement, computationally efficient, and compatible with various energetic methods.
Genetic algorithms (Fig. 2.4e) start with an ensemble of structures, where residues are allowed to intermix according to genetic criteria through structural crossing. Similar to biological populations, fitter residues (those with lower energy) preferentially mate with their own kind and produce more offspring than higher-energy structures. During this crossing step, mutations, i.e., random distortions of DOFs, can occur. The fittest residues from the offspring and parent ensembles survive and continue to breed until convergence is reached after dozens to hundreds of generations. The genetic criteria used for chemical structures are typically energies, gradients, or forces, aiming for global energy minimization. Genetic algorithms are easy to implement, compatible with different energetic methods, and often used in machine learning applications.
Many of the discussed methods can be enhanced by incorporating memory, preventing double exploration of certain PES areas, or by using simulated annealing to cover more of the PES. They can also be combined with swarms of residues for parallelized PES exploration, traditional minimization methods to transition from point-like to continuous space, and machine learning potentials to generally improve efficiency, reduce computation time, and address intrinsic algorithmic limitations.

2.2.4 Molecular Docking

Docking involves the global optimization of the interaction of two or more molecules. Typical applications are between a protein and a small drug candidate but extending to various combinations

like ligands and catalysts, solutes and solvents, surfaces and adsorbates, sensors and analytes, storage compounds to cages, aggregates, and more. Docking is integral to workflows such as high-throughput screening of candidates for selective binding in Computer-Aided Drug Design (CADD), Computational Materials Design (CMD), and catalysis.
The challenge lies in reducing the dimensions of the problem and efficiently estimating binding energies. Each molecule has inherent DOFs, and the interaction between docking partners introduces additional DOFs related to position, orientation, and relaxation upon interaction. A systematic search across all these DOFs is computationally infeasible due to the exponential scaling of the number in large and flexible systems, as illustrated in Fig. 2.5. Consequently, grid-based searches are often employed, simplifying molecules to their center of mass and principal axes of inertia or using fragment-based representations. Docking partners are typically kept rigid, with relaxation allowed only in subsequent workflow steps. Various algorithms can determine docking positions, including quasi-systematic grid searches, scoring functions, interaction site screening, functional group libraries, genetic algorithms, and machine learning approaches. These algorithms assess structures and energies using quantum-mechanical or classical methods, from force fields to SQM and DFT and recently also ML potentials.[172]

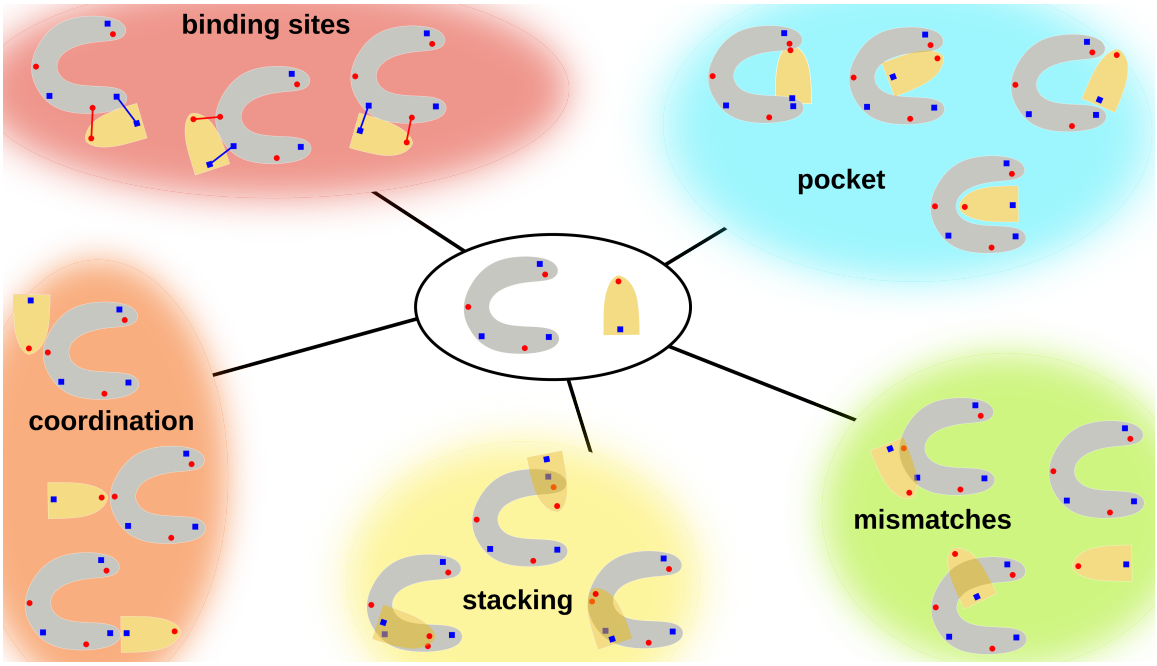

Figure 2.5: Schematic representation of different types of docking positions. The small yellow fragment is docked to the larger gray fragment, red and blue marks represent special molecular features such as recognition units or functional groups.

A scoring function is an empirical often linear approximation used to estimate the binding energy of a host-guest system. Calculating binding energy for each complex individually is often too computationally expensive for screening and ranking interaction sites, so composite approaches are

used instead. A scoring function accounts for different parts of the interaction energy, some are listed in the equation below but the actual contributions considered vary:

$$\Delta G_{\text{scoring}} = a_1 \cdot \Delta E_{\text{vdw}} + a_2 \cdot \Delta E_{\text{el}} + a_3 \cdot \Delta G_{\text{rot}} + a_4 \cdot \Delta G_{\text{H-bonds}} + a_5 \cdot \Delta G_{\text{solv}} \ldots \tag{2.33}$$

The prefactors a_i are fitted to either experimental or computational data. The individual contributions may be approximated with the help of physical concepts or stored data. For example, entropic contributions can be estimated from the DOFs of the structure and solvation contributions from the size of the analyte which both do not require additional calculations or quantum-mechanical operations. This approach is particularly suited for machine learning solutions and fast pre-screening of narrow compound spaces, though the simplifications and omissions involved limit its accuracy and transferability, especially for chemically diverse systems.
The automated interaction site screening module (aISS),[138] implemented into the `xtb` program suite, employs force fields and SQM methods from the GFN family to dock chemically diverse structures such as MOFs, zeolithes, or catalysts in addition to standard organic molecules and biosystems. The goal is to find the global minimum of the interaction energy between two fragments based on the energy expression of the xTB intermolecular force field (xTB-IFF) and employing a sophisticated mixture of global minimization algorithms such as grid-based search and genetic algorithms. The xTB-IFF[173] employs a rigid docking based on quantum-mechanical descriptors of the docking sites, namely localized molecular orbital (LMO) matching. The interaction energy is calculated as follows:

$$E_{\text{IA}} = E_{\text{rep}} + E_{\text{ES}} + E_{\text{disp}}^{\text{D4+ATM}} + E_{\text{ind}}^{\text{Drude}} + E_{\text{CT}} \tag{2.34}$$

with contributions from Pauli repulsion (rep), electrostatics (ES), dispersion (disp) according to the D4 dispersion model and ATM three-body correction,[141,150,151] induction according to the Drude oscillator model,[174] and charge transfer (CT). Unlike scoring functions, these contributions are derived quantum-mechanically. Accordingly, the IFF interaction energy evaluation is slower than a scoring function but more accurate, more chemically motivated, and a general purpose approach suitable for a wide range of very diverse systems. To obtain descriptors like atomic partial charges, LMOs, frontier molecular orbital energies and densities, initial single point calculations of the fragments are conducted at the GFN1-xTB level of theory and are then used to set up the IFF potential automatically. As the docking happens in a rigid way, this information can be reused to speed up any further energy evaluation.
The aISS improves on the docking algorithm of xTB-IFF,[173] going from a basic genetic algorithm to a more space covering, chemically motivated search of docking motifs. Still, all energy evaluations employ the IFF energy expression. A grid-based systematic pre-screening with a charged Krypton atom is conducted to identify promising interaction sites. Starting from the so-generated ensemble, three different pathways are explored, 1) a pocket search employing shape matching and relaxation of the host and guest, 2) angular search, and 3) stack search. The two latter ones undergo a combined refinement through a short genetic optimization with high mutation rates to explore more unknown combinations of orientations and positions in a short amount of time. All found structures from all three searches are subject to a final geometry optimization at the SQM level. Additional options like implicit solvation or constraining potentials are available due to the convenient integration into the `xtb` framework enhancing the versatility of this workflow even more, e.g., to model surface coverage and layer growth.[6]

Libraries of functional groups and interaction motifs are commonly used for bio-organic docking problems. Docking sites, fragments, and interaction energies are pre-computed and stored in databases for quick access upon motif recognition. While this approach saves computation time, it is less flexible and customizable, suitable primarily for specialized applications within a limited chemical space. Genetic algorithms, previously described for conformer search, are equally applicable to docking problems.[175] Machine learning, with its ability to handle complex correlations, is especially effective for docking challenges, as seen in applications like gas uptake in MOFs[176] or CADD.[177]

2.2.5 Molecular Identifiers

Workflows in computational chemistry typically begin with an initial structure guess, which can be obtained through manual input in a molecular editor, from crystallographic data, or by retrieving information from databases. Molecular structures are available in various formats, each containing different types of information, such as 3D structure, topology, crystallography, or specific properties. However, 3D data can be hard to parse or unavailable, especially in high-throughput workflows. To address this, text-based molecular identifiers were developed, providing a more compact and manageable way to represent molecular structures. Molecular identifiers like SMILES,[178] InChI,[179] MolBar,[180] SMARTS,[181] and SLN[182] are widely used, each designed for specific purposes and offering varying levels of detail, ranging from simple molecular formulas to complex stereochemical information. Despite their differences in scope and objectives, these identifiers generally follow a basic process: atoms are numbered, elements and bond orders are recorded, rings are identified, and stereocenters are determined if applicable. This information is then compressed into machine-readable strings, which can, in some cases, be converted back into molecular structures when needed. However, conformational details are usually lost in this process, and backconversion often requires further structural refinement, such as geometry optimization or conformer searches. This is also depicted schematically in Fig. 2.6.

SMILES (Simplified Molecular Input Line Entry System) is the oldest and most widely used text-based molecular identifier. It is human-readable and provides a simple overview of a molecule's connectivity and bond order. However, SMILES has limitations in resolving stereocenters and tautomers and can generate ambiguous keys for complex systems. Variants like canonical SMILES and isomeric SMILES have been developed to address these issues, though challenges remain. SMARTS (SMILES Arbitrary Target Specification) is a specialized variant of SMILES used for defining subsystems, such as in biochemical applications. SMARTS is particularly useful for searching structural motifs, functional groups, and fragments, but it is less human-readable than SMILES. InChI (International Chemical Identifier) offers a more detailed representation, including information on connectivity, bond order, tautomerism, isotopes, stereochemistry, and charge. While InChI is more complex and challenging to generate, especially for large structures, it produces mostly unique and unambiguous keys, making it a reliable identifier. SLN (Sybyl Line Notation) represents a group of molecular identifiers that function more like a chemical code language, capable of describing more than just structural features. SLN is flexible and well-suited for large systems, encoding information about connectivity, bond order, stereocenters, isotopes, charges, and even torsion angles, which allows it to represent different conformers. Furthermore, it allows for the encoding of certain molecular properties like molecular weights or partial charges directly into the string and accept some logical operations for extended substructure search. However, its complexity limits its use in workflows that require interoperability between different programs. MolBar (Molecular Barcode) is another text-based identifier, though its

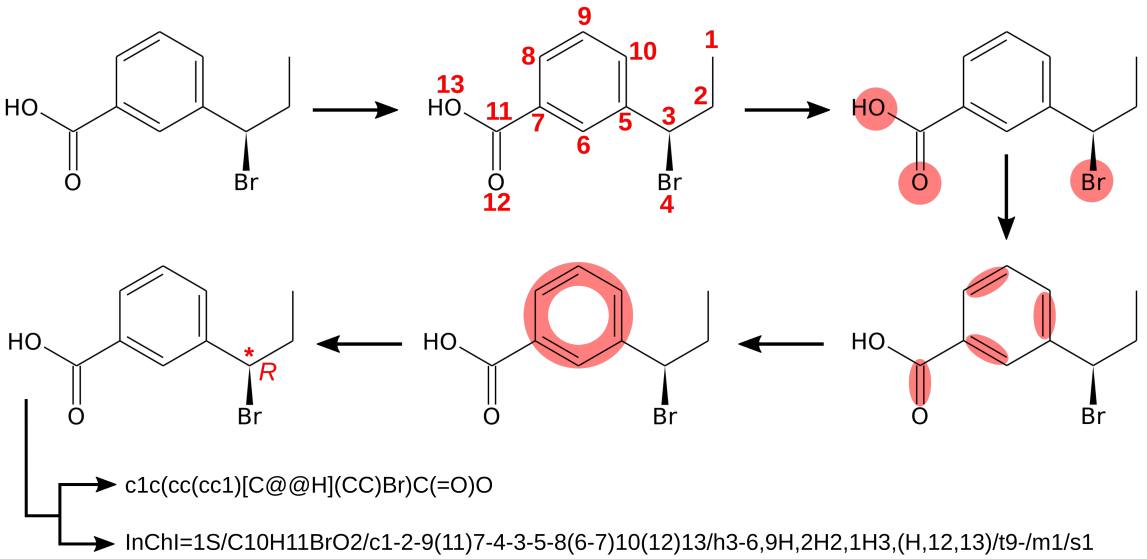

Figure 2.6: Schematic representation of how molecular identification keys are derived. The upper key is in SMILES, the lower one in InChI.

name suggests a classical barcode. MolBar supports connectivity, bond order, tautomerism, isotopes, charge, and even more complex forms of isomerism, such as axial, planar, and helical chirality. This identifier yields unique and unambiguous keys through a fragment-based approach but is limited to 3D structures as a starting point and not invertible.

Textual molecular identifiers offer a convenient way to transfer, compare, and search for molecular structures across different programs and architectures. Their specialized designs target to specific computational applications. However, the dimensional reduction from 3D structures to text strings results in a significant loss of information, making it crucial to carefully select the appropriate identifier for each workflow.

2.3 Organic Electronics

Organic electronics (OEs)[15] are a branch of electronics that focus on the use of organic materials, particularly organic semiconductors, for the construction of electronic devices. These materials offer several advantages over traditional inorganic electronics, including flexibility, light weight, sustainability, and the potential for cost-effective production. Examples of organic electronic devices include organic light-emitting diodes (OLEDs),[23] organic solar cells (OSCs),[21] and organic field-effect transistors (OFETs),[25] all of which will be explored in subsequent sections, following a brief introduction to the theory of charge transport in semiconductors. This chapter is based on Refs. [183] and [184]. For more comprehensive insights, the reader is referred to this literature.

2.3.1 Marcus Theory

Marcus theory, developed in the 1960s, is the major framework for understanding electron transfer energetics and rates.[185] It provides a semi-classical description of electron transfer processes by employing the concept of intersecting parabolic electronic states to calculate transfer barriers and rates. Although Marcus theory was originally formulated for adiabatic electron transfer, it has since been extended to address non-adiabatic processes, such as those encountered in photochemistry. A quantum-mechanical perspective on the transfer process is also available, for instance, through the Marcus-Levich-Jortner theory,[186–188] which incorporates vibrational effects and interactions with the environment. The core equation for the electron transfer rate (k_{ET}) is expressed as:

$$k_{\mathrm{ET}} = \frac{2\pi}{\hbar} J_{AB}^2 \frac{1}{\sqrt{4\pi\lambda k_B T}} \exp\left(\frac{-(\Delta G^0 - \lambda)^2}{4\lambda k_B T}\right) \tag{2.35}$$

Here J_{AB} represents the electronic coupling integrals between states A and B, λ is the reorganization energy, ΔG^0 is the Gibbs free energy change of the transfer reaction, k_B is the Boltzmann constant, $\hbar$ is the reduced Planck constant, and T is the temperature. The DIPRO method on how to obtain J_{AB} is presented in Chapter 7. Typically, when $|\Delta G| \leq \lambda$, the electron transfer rate increases with stronger electronic coupling, lower reorganization energy, more negative Gibbs free energy, and higher temperature. However, when $|\Delta G| > \lambda$, the system enters what is known as Marcus inverted region, where increasing the driving forces no longer accelerates the electron transfer. In this region, ΔG overly stabilizes the reactants, making the reorganization energy, required for state transition, excessively high. The inverted region is particularly relevant to biological systems, such as the respiratory chain, as well as to OE materials.[189]

Marcus theory is illustrated schematically in Fig. 2.7. The theory uses classical parabolic potential energy surfaces along the reaction coordinate, depicting the ground and excited states of a donor-acceptor (D-A) system. A lowering of the energy in the D^+-A^- state results in a lower transition state, represented by the intersection of the D^+-A^- and D-A parabolas. An intersection at the ground state minimum signifies a barrierless transfer, while an intersection on the left side of the D-A parabola indicates the Marcus inverted region. Beyond its direct application to electron transfer processes, Marcus theory is also significant in transition state theory, particularly for reactions involving charge transfer or adiabatic reorganization, such as solvent-solute interactions.[190] The theory plays a crucial role in the calculation of charge transport and conductivity in the hopping regime, the determination of redox rates and the design of redox catalysts in order to optimize OE materials and devices.

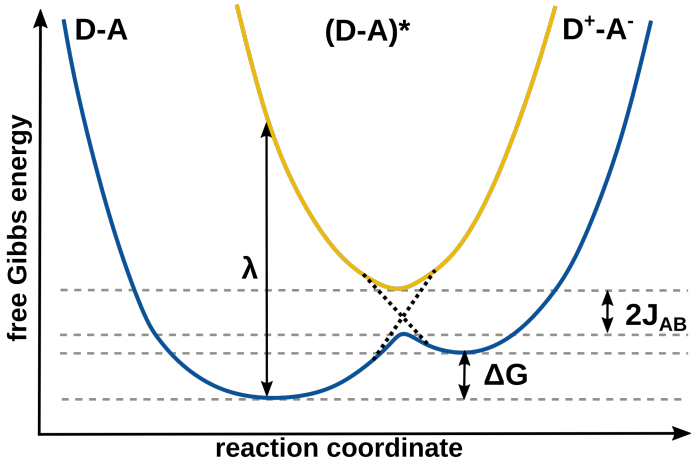

Figure 2.7: Schematic representation of Marcus theory for charge transfer in the normal regime, D stands for donor, A for acceptor, * for excited state, λ is the reorganization energy, ΔG the Gibbs free energy, and J_{AB} is the coupling integral between both states. Solid lines represent the adiabatic transition, dotted lines the diabatic one.

2.3.2 Charge Transport

Charge transport in materials typically manifests as either band transport that decreases with increasing temperature or hopping-type transport that increases with increasing temperature. Band transport usually enables high conductivity while hopping transport results in a lower conductivity. Materials with intermediate conductivity often exhibit a combination of both transport mechanisms, i.e., intermediate transport regime. There are additional types of charge transport if external gradients, such as electric field, doping, or concentration, influence the material. This, or if specific charge carriers like excitons or polarons are regarded, can lead to e.g., drift, diffusion or superconductance. The presence of inhomogeneities in the material, such as disorder, grain boundaries, defects, interfaces, and phase changes, can significantly impede all charge transport types, also referred to as traps.[191]
Band transport occurs when charge carriers move through extended electronic states, which are typically found in highly crystalline materials. In this mechanism, holes and electrons are transported via the conduction and valence bands of conductors and semiconductors, respectively. Band transport efficiency decreases with increasing temperature due to enhanced lattice vibrations, which can trap electrons.
Hopping-type transport dominates in disordered or amorphous materials where electronic states are mostly of localized nature.[192] This type of transport is characterized by low charge mobility due to the energetic barriers between hopping sites. Hopping transport is highly anisotropic, often occurring through preferred channels, such as through-space between stacked conjugated systems. As temperature increases, hopping barriers are easier to overcome, making hops more probable, which in turn increases mobility. However, hopping transport is highly susceptible to trapping effects. Generally, the charge carrier mobility in hopping transport is about three orders of magnitude lower than in band transport, although this can vary significantly depending on the specific system. In OE materials, hopping transport is the predominant regime.

A polaron is a quasi-particle formed when a charge carrier, electron or hole, couples with a lattice vibration in a partially ordered material. Polarons can be categorized as either small or large, depending on their spatial extent.[193] Small polarons are localized, and their transport resembles hopping-type transport, though with lower hopping barriers due to lattice distortion and longer hopping distances. Consequently, polaron transport is faster than simple hopping, with mobility that increases even further with temperature. On the other hand, large polarons are more delocalized, and their transport mechanism resembles band transport but with reduced mobility due to increased phonon scattering and a higher effective mass. For large polarons, an increase in temperature further decreases mobility. Medium polarons exhibit characteristics of hopping and band transport likewise and are assigned to the intermediate transport regime. In summary, polaron transport is generally faster than hopping transport but slower than band transport. The coupling to lattice vibrations offers a more efficient charge transfer mechanism than random hopping, though it remains less efficient than coherent band transport. Exciton transport is discussed in the following subsection.

2.3.3 Excitons

Excitons are Coulomb bound electron-hole pairs that occur upon excitation of an electron in a semiconducting or insulating material. Excitons can be regarded as uncharged bosonic quasi-particles able to spatially transport energy without net charge movement. Excitons were first proposed in the 1930s and experimentally validated about three decades later.[194] Today, excitons are recognized as fundamental to the operation of semiconductor electronics.[195]
The stability of an exciton is primarily determined by the exciton binding energy which is material specific. Higher binding energies correspond to greater stability of the exciton. Exciton binding energies are typically higher in organic materials than in inorganic ones, resulting in lifetime differences about a factor of 10^3 (ps to ns for inorganic semiconductors, ns to μs for organic semiconductors). The strongly bound excitons are commonly called Frenkel excitons[196] and exhibit a very local nature whereas the weakly bound Wannier-Mott excitons[197] are rather delocalized. Other categories focus more on the fundamental properties of the exciton quasi-particle, such as spin and momentum. All theories that apply for bosons and their interactions also apply for excitons, i.e., excited excitons, biexcitons, trions, exciton-particle coupling etc. are conceivable.
Excitons break apart either through charge separation as done in OSCs or through charge recombination which is the working principle of OLEDs. The recombination of electron and hole can occur radiative, under emission of light, or radiation-less, under the release of heat. Exciton transport is distinct from simple hole and electron transport and is characterized by its own set of properties. Transport can happen e.g., via Förster resonance energy transfer (FRET)[198] on the intermediate range and via Dexter energy transfer[199] on the short-range. FRET occurs via non-radiative dipole-dipole coupling of excitons in the host material. The dipole-dipole interactions decrease spatially with $(1/r^3)^2=1/r^6$ and are thus very distance sensitive. Dexter energy transfer relies on exchange of spatially close electrons and holes for the propagation of excitons in the host material, requiring overlap of the wave functions. Exciton transport in materials can be modified via heat (usually higher mobility but also higher probability for radiationless recombination through interaction with phonons), mechanical strain, electric fields, or band structure engineering.[200] The transport is limited by the material quality, e.g., interfaces, grain boundaries, order, phase changes, crystallinity, defects, or impurities, which increase the formation of trap states and lead to a premature charge recombination. Good organic exciton conductors are compound classes like thiophenes, vinylenes, and acenes as proven experimentally.

2.3.4 Organic Solar Cells

The first organic solar cells (OSCs) were developed in the early 1970s using acenes, with a conversion efficiency of less than 0.02%[201] (for comparison, commercially available standard Si solar cells nowadays achieve about 22%). Today, highly optimized OSCs can reach conversion efficiencies of up to approximately 19%. The work mechanism of an OSC is illustrated schematically in Fig. 2.8. Unlike inorganic solar cells, which generate charges directly through the photoelectric effect, OSCs produce excitons upon photoexcitation of the donor material (a). In inorganic solar cells, charges are quickly transported to the electrodes via band transport. In contrast, OSCs employ a slower hopping mechanism for charge transport to the donor-acceptor interface, where excitons are separated (b). The charges are then transported to the electrodes (c), sometimes with the aid of specialized hole or electron-conducting layers integrated into the cell architecture.

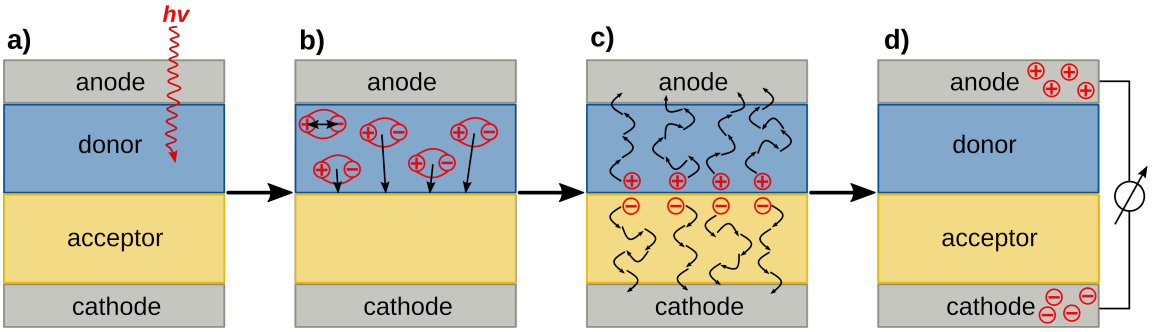

Figure 2.8: Schematic representation of the mechanism of an OSC. a) photoexcitation, b) exciton diffusion to the donor-acceptor interface (or exciton recombination) and subsequent charge separation, c) charge transport to the electrodes, d) current output.

The theoretical efficiency of solar cells is limited by numerous factors. For single-junction inorganic solar cells this is called Shockley-Queisser limit[202] and lies at about 29%. For multi-junction solar cells this limit is much higher and for organic solar cells no theoretical limit has been defined yet. The limitations can roughly be divided into internal and external factors. External limitations comprise the transparency and reflectivity of the electrode and transport layers, the absorption spectrum of the active layer, the layer thickness, surface passivation, impurities and material degradation. This affects any solar cell, inorganic or organic. Internal limitations include thermal losses, where photons with energy beyond the absorption spectrum of the active layer produce heat (which is counterproductive for band transport but somewhat beneficial for hopping transport), and transmission losses from photons with too low energy. Both issues can be mitigated by using multi-junction or tandem cells that cover a broader spectrum. Additional losses occur through charge recombination and subsequent photoemission. This is a more significant problem for OSCs because excitons are more tightly bound and migrate more slowly compared to inorganic cells, where recombination mostly occurs at the surfaces. Unique challenges for OSCs include higher charge separation energy due to tightly bound excitons, low charge mobility due to intrinsic disorder and grain boundaries, which cause slow transport and a high probability of charge recombination, and increased susceptibility to degradation from environmental factors such as light, moisture, temperature, mechanical stress, and pollution.
Both inorganic and organic solar cells are subjects of ongoing research and improvement as the theoretical efficiency limits are not yet reached. OSCs can be enhanced via three different channels,

namely morphology, donors and acceptors, and architecture. Morphology describes the molecular packing, order, defects, crystallinity, and grain boundaries. Architecture describes the combination of materials and constitution of the active layer. Donors and acceptors are chemically different material classes optimized for light-harvesting and exciton generation. Morphological improvement can be achieved via modified fabrication conditions, introduction of templates or matrices for ordered growth, or the use of liquid crystal like materials.
Donor materials for the active layer include for example derivates of thiophenes, carbazoles, pyrroles, indolenes, porphyrins, acenes, fluorenes, and thiazoles, available as small molecule donors and polymers. On the other hand, acceptor materials include for example fullerenes, perylene and naphthalene diimides, thiophenes, indacenes, and coronenes, the majority from the small molecule regime although polymers are conceivable as well. Careful combinations thereof yield customized absorption profiles for e.g., indoor applications. Also more complex molecular designs in the style of donor-acceptor-donor or donor-donor-acceptor are conceivable this way.
The architecture of solar cells in general can contain one, two, or multiple active materials, called homo-junction, hetero-junction, and multi-junction solar cells. All these materials can be distributed within the active layer in three different ways (see Fig. 2.9), flat, ideally dispersed, and randomly intermixed. Thus, a bulk hetero-junction OSC consists out of two materials (one donor and one acceptor) that are randomly mixed in the bulk. Likewise, a flat multi-junction solar cells consists of multiple flat layers of more than two materials that ideally absorb different spectral areas and thus maximize the photon yield. The advantage of flat orientation is the easy processability, however the interface between the phases is small and the transport pathways to the electrodes are long and thus limit the performance. This architecture is most often employed for homo-junction solar cells like perovskite, silicon, or quantum dot cells. Ideally dispersed architectures are hard to obtain and usually require a matrix or template, e.g., porous or MOF-based solar cells, which presents an additional complication, but their interface area is high and the transport of the charge carriers to the electrodes is optimal, guaranteeing a high efficiency. The practical compromise between those two architectures is the randomly mixed bulk, providing an intermediate interface, intermediate transport length, and intermediate processability, as can be seen in e.g., all-polymer, fullerene, or dye-sensitized solar cells.

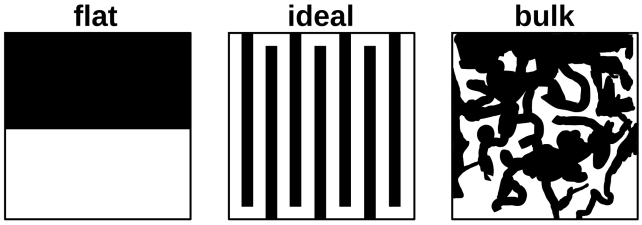

Figure 2.9: Schematic representation of solar cell architectures.

2.3.5 Organic Light Emitting Diodes

The first OLEDs were investigated in the 1980s using an aluminum quinoline derivative mixed with tertiary amines.[26] Today, a variety of compound classes are used in OLED devices, including thiophenes, vinylenes, fluorenes, dendrimers, and thermally activated delayed fluorescence (TADF) emitters, available as either small molecules or polymers. OLEDs operate in a manner fundamentally opposite to that of organic solar cells (OSCs). In OLEDs, an applied voltage injects charges into the active layer, where they migrate and recombine at the interface, resulting in light emission. The performance of OLEDs is measured by their external quantum efficiency (EQE) and internal quantum efficiency (IQE). IQE quantifies how many electrons produce photons, while EQE measures how many of these photons are emitted from the device. Theoretically, IQE can reach 100% when relaxation happens exclusively via combined phosphorescence and fluorescence pathways and not radiationless, but EQE is typically limited to around 50% for LEDs and about 40% for OLEDs. The limitations affecting IQE and EQE are distinct. EQE is constrained by internal total reflection, optical losses, and outcoupling efficiency limits due to destructive interference of light waves and trapping in microcavities. These issues are common to both OLEDs and LEDs but are more pronounced in OLEDs due to the disordered nature of the active layer. Additionally, there is an optimal voltage window for operating OLEDs and LEDs, balancing charge transport velocity and counteracting induced currents. IQE is influenced by the triplet-to-singlet ratio of the material. In fluorescent materials, typically only singlet excitons contribute to emission, while in phosphorescent or TADF materials, both singlet and triplet excitons contribute. Non-radiative recombination of charges can also reduce overall emission efficiency. Unlike in OSCs, the slow charge hopping transport is not a major limiting factor in OLEDs. For OLEDs, a narrow emission spectrum is usually more advantageous, efficient, and easier to fabricate compared to the broad absorption spectrum desired for OSCs as illustrated in Fig. 2.10. White light OLEDs often use a tandem cell configuration.

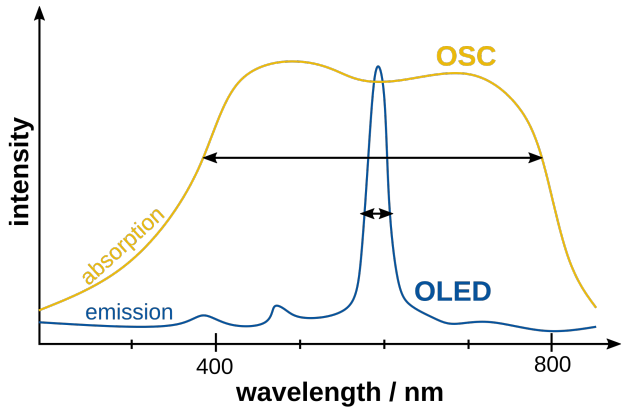

Figure 2.10: Schematic representation of the ideal spectral properties of OLEDs vs. OSCs.

Compared to LEDs, OLEDs offer a higher customization potential, easier production conditions i.e., lower production costs and convenient processing, flexibility and transparency, enhanced color rendering, and better scalability at low voltages, in addition to the environmental benefits common to all OEs.

2.3.6 Organic Field Effect Transistors

Field effect transistors modulate and switch the intensity of electric current flow through an applied electric field, needed for all types of electric circuits and convenient devices. The first OFETs were introduced in the 1980s based on thiophenes.[203] Other widely used compound classes include fluorenes, vinylenes, acenes, and fullerenes, in small molecule and polymeric architectures.[25] The fundamental differences in the structure of OSCs, OLEDs, and OFETs are depicted in Fig. 2.11.
OFETs have a distinct structure compared to OSCs and OLEDs. They require three electrodes: a source and drain electrode connected to the operating circuit, and a gate electrode connected to a control circuit. The gate electrode induces an electric field into the adjacent dielectric layer, which then injects charge carriers into the active layer of the organic semiconductor, forming a conducting channel for the operating current. This channel can be modulated by varying the gate voltage, thereby controlling and switching the current flow in the operating circuit. OFETs can be charge and hole conductors, the limiting factors for their performance are the charge carrier mobility and the voltage thresholds for switching. Compared to inorganic FETs, the charge carrier mobility is up to a factor of about 10^3 lower, resulting in a generally lower performance and slower, less precise switching speed. However, OFETs can be operated at very low voltages and offer the typical production advantages of OEs, i.e., low-cost, ease of processing, customizability, flexibility, or transparency. These characteristics make OFETs particularly beneficial for certain applications, such as large-area or flexible devices, despite their lower performance compared to inorganic FETs.

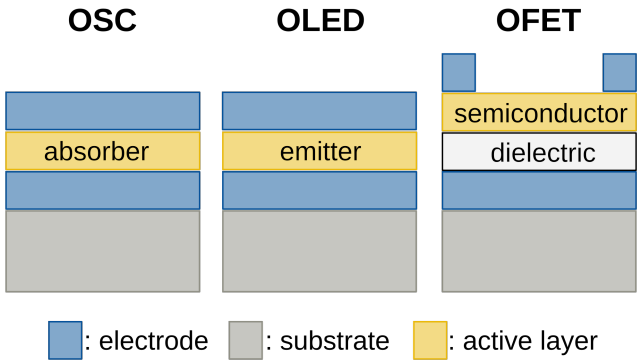

Figure 2.11: Schematic representation of the structure of different devices in organic electronics.

CHAPTER 3

Quickstart Guide to Model Structures and Interactions of Artificial Molecular Muscles with Efficient Computational Methods

Julia Kohn,[†] Sebastian Spicher,[†] Markus Bursch,[†] and Stefan Grimme*[†]

Received: 13 October 2021
Published online: 29 November 2021

Reprinted in Appendix A (adapted) with permission[‡] from
J. T. Kohn, S. Spicher, M. Bursch, and S. Grimme, *Quickstart guide to model structures and interactions of artificial molecular muscles with efficient computational methods*, Chem. Comm. **58**.2 (2022) 258,
DOI: 10.1039/d1cc05759f

Own contributions

- initial idea
- performing of all calculations
- interpretation of the results
- writing of the manuscript

[†]Mulliken Center for Theoretical Chemistry, University of Bonn, Beringstr. 4, D-53115 Bonn, Germany
[‡]Permission requests to reuse material from this chapter should be directed to Royal Society of Chemistry RSC.

Artificial molecular muscles (AMMs) represent an important subgroup of molecular machines,[59,204] imitating movements and functions of macroscopic mechanic machinery like e.g. shuttles,[205] motors,[60] muscles,[62] walkers,[61] or hinges.[206] These molecules are typically investigated in supramolecular chemistry and nanosciences, giving rise to applications as e.g. nanorobots,[207] organic or molecular electronics,[208] or data storage devices.[209] AMMs, also referred to as daisy-chain rotaxanes, typically consist of two interlocked but not covalently bonded arms, a bulky stopper unit (usually tri-^{t}Bu-phenyl- or polysaccharide derivates) to prevent disentanglement, and matching slipping (usually cyclodextrine or crown-ether analogues) and resting unit pairs. The slipping unit is activated upon various chemical stimuli such as light, reduction, or metal-ion exchange and leads to a controlled contraction or expansion of the AMM. The resting or recognition unit is targeted to the specific stimuli, e.g. naphthalene diimide (NDI) units can easily undergo reduction/oxidation, bi- and tridentate N-donating ligands like bipyridine can bind metal ions, whereas *E/Z* isomerizable double bonds are introduced for light switchable AMMs.

Experimental structure determination of this molecule class is challenging. Due to the conformational flexibility it is difficult to crystallize them and obtain real 3D data, and due to the sheer size broad and ambiguous spectra with only connectivity information are caused. Accordingly, computational approaches are needed to support structure elucidation and yield in-depth insight into the structure-property relationships, i.e., contraction ratios and contraction mechanisms. However, as AMMs are typically hundreds to thousands of atoms large and may contain metals, high charges, and challenging structure motifs, a theoretical description is demanding as well. DFT-D methods yield a good description but are too costly for this size.[210,211] On the other hand, many semiempirical (SQM) or force-field (FF) methods are not able to treat metals or non-covalent interactions (NCIs) adequately.

An efficient and easy-to-implement quantum chemistry based workflow was developed for the generation of standardized fully extended AMM structures on the basis of the semiempirical and force field GFN methods. GFN*n*-xTB and GFN-FF[212,213] are also parameterized for metals and NCIs and are thus suited to investigate large AMMs. In the proposed workflow, an initial structure guess is stretched via a semi-relaxed distance scan. The contact positions between slipping and resting units were preserved by molecule specific constraints to avoid artificial overstretching. The most stretched but still chemically undistorted structure is the input to a constrained conformer search at the GFN-FF level using CREST.[133] The structure generation workflow is finalized by a geometry optimization and reranking of the previously obtained ensemble at the GFN2-xTB level,[134,143] applying the same structure retaining constraints as in the former steps.

This workflow was employed to calculate reaction free energies of AMM contraction at various SQM and DFT levels for a small test set of AMMs.[214–217] The test set includes exemplary AMMs of each group of contraction stimuli with a wide chemical variety even containing highly charged species and metal coordination centers. For the only available experimental reference of -0.5 kcal/mol, an outstanding accuracy within ±1 kcal/mol deviation was reached for DFT and GFN2-xTB. Additionally, contraction ratios were calculated. AMMs usually adopt a folded conformation in solution, thus, experimental contraction ratios are only accessible through atomic force microscopy.[218,219] Computational contraction ratios were modeled via MD simulations at GFN-FF and GFN2-xTB level, in good congruence with experimental data. The agreement between SQM and DFT results in this study was reasonable with SQM computation times being two to four orders of magnitude lower, enabling the investigation of even larger systems like molecular elevators or multiply daisy-chained rotaxanes.

[†]The nomenclature has been updated in this publication and in the corresponding Chapters

CHAPTER 4

Increased Antiaromaticity through Pentalene Connection in [*n*]Cyclo-1,5-dibenzopentalenes

Jan S. Wössner,[†] Julia Kohn,[‡] Daniel Wassy,[†] Mathias Hermann,[†] Stefan Grimme*[‡] and Birgit Esser[†§¶]

Received: 16 November 2021
Published online: 14 January 2022

J. S. Wössner, J. T. Kohn, D. Wassy, M. Hermann, S. Grimme, and B. Esser, *Increased Antiaromaticity through Pentalene Connection in [n] Cyclo-1, 5-dibenzopentalenes*, Org. Lett. **24**.4 (2022) 983, DOI: 10.1021/acs.orglett.1c03900

Own contributions

- performing of quantum chemical calculations except for strain energies and optical properties
- interpretation of the quantum chemical results
- contributed to writing the manuscript
- none of the experiments or measurements were conducted by me

[†]Institute for Organic Chemistry, University of Freiburg, Albertstraße 21, 79104 Freiburg, Germany
[‡]Mulliken Center for Theoretical Chemistry, University of Bonn, Beringstr. 4, D-53115 Bonn, 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

Conjugated carbon nanohoops represent a subclass of carbon-rich compounds in supramolecular chemistry. Carbon-rich compounds are organic molecules or polymers with a carbon-to-hydrogen-ratio larger than one. Aromatic nanohoops are especially interesting due to their unique optoelectronic properties,[220,221] host-guest chemistry,[222,223] and cyclic aromaticity induced by the intrinsic curvature of the conjugated π-system. Their properties differ significantly from their flat oligomeric analogues. The prototype of an all-conjugated carbon nanobelt is [*n*]cycloparaphenylene ([*n*]CPP) first synthesized by Vögtle *et.al.* in the 1990s.[224] Substitution of the CPP units by other polyaromatic hydrocarbons (PAHs) allows for targeted modification of optoelectronic properties[225] like e.g., band gap or absorption spectrum.[226,227] In this joint experimental-computational collaboration, anti-aromatic dibenzopentalene units (DBPs) are introduced into the nanohoops while synthesis and optoelectronic properties are investigated. DBP offers two distinct connection possibilities, 2,7- which is along the axis of the conjugated system, and 5,10- which is perpendicular to it and attached to the five-membered rings. In previous work by Esser *et. al.*,[228] enantiomerically pure synthesis of 2,7-joined [*n*]-cyclo-DBP was presented, giving rise to a new type of chirality originating solely from the bending of the achiral DBP precursor within the nanohoop and a high rotation barrier preventing diastereomerization.[9] In this work, the focus lies on the 5,10-joined [*n*]-cyclo-DBPs. In principle, those are also chiral but exhibit much less curvature and lower rotation barriers allowing for isomerization at room temperature.

Synthesis of [*n*]-cyclo-5,10-DBP with *n*=4,5 was performed by Esser *et. al.* via C-C coupling reactions, i.e., Platinum(II) mediated cyclization of the monomeric precursors, giving the products in 1-2% yield. The Pt-intermediates were simulated at the CREST[133]@GFN2-xTB[143](GBSA[166](toluene)) level of theory and showed a puckered ([4]-cyclo-5,10-DBP) and entangled ([5]]-cyclo-5,10-DBP) geometry, respectively, instead of a planar polygonal configuration as commonly expected. These entangled conformations adopt a Ph-Pt-Ph angle of about 80° and thus allow for stabilizing π-π interactions between the DBP units, especially for the entangled [5]-cyclo-5,10-DBP, which was purified in more than twice as much yield compared to [4]-cyclo-5,10-DBP. [3]-cyclo-5,10-DBP has never been observed owing to the fact that a 60° Ph-Pt-Ph angle would be required in the cyclic intermediate to form the desired product and hence, no stabilization through stacking could occur therein.

Experimental NMR spectra confirmed the computationally evaluated structural isomerism which originates from the pattern of differently up or down oriented DBP units and that the rotation barriers are accessible at room temperature. Strain energy calculations showed that the strain arises primarily from the bent DBP-Ph dihedral connections and not from an internal bending of the DBP as in Ref. [9]. UV/vis measurements showed the cyclic conjugation all around the hoop by a red-shift of the DBP band compared to the non-cyclized oligomer, whereas this feature vanished completely for the [*n*]-cyclo-2,7-DBPs. These results undermine that the conjugation in the hoops is more pronounced than in the linear analogues. Furthermore, tailored introduction of substitution patterns at the five-membered rings proofed to be a versatile synthetic tool to tune LUMO properties independently of the HOMO. At last, the degree of anti-aromaticity was investigated via the simulation of nucleus independent chemical shifts (NICS)[229,230] at the B3LYP[231]level of theory employing a Slater-type orbital TZP[232] basis, revealing an enhanced anti-aromaticity compared to the linear reference and the [*n*]-cyclo-2,7-DBPs. Concluding, DBP is a promising building block for carbon-rich materials in supramolecular chemistry with interesting properties such as versatile substitution patterns, pro-chiral nature, pronounced anti-aromatic character, and diverse tuning options for optoelectronics.

[†]The nomenclature has been updated in this publication and in the corresponding Chapters

CHAPTER 5

Size-Increased all-Phenylene Molecular Spoked Wheels - A Combined Theoretical and Experimental Approach

Philipp Krämer,[†] Julia Kohn,[‡] David Hofmeister,[†] Maximilian Kersten,[†] Christopher Sterzenbach,[†] Antonia Gres,[†] Andreas Hansen,[‡] Stefan-Sven Jester,[*†] Stefan Grimme,[*‡] and Sigurd Höger[*†]

Received: 12 June 2024
Published online: 07 August 2024

Reprinted in Appendix C (adapted) with permission[§] from
P. Krämer, J. Kohn, D. A. Hofmeister, M. Kersten, C. Sterzenbach, A. Gres, A. Hansen, S.-S. Jester, S. Grimme, and S. Höger, *Size-Increased All-Phenylene Molecular Spoked Wheels–A Combined Theoretical and Experimental Approach*, Angew. Chem. Int. Ed. (2024) e202411092, DOI: 10.1002/anie.202411092

Own contributions

- performing of all quantum chemical calculations
- interpretation of the quantum chemical results
- contributed to writing the manuscript
- none of the experiments or measurements were conducted by me

[†]Kekulé Institute for Organic Chemistry and Biochemistry, University of Bonn, Gerhard-Domagk-Str. 1, D-53121 Bonn, Germany
[‡]Mulliken Center for Theoretical Chemistry, University of Bonn, Beringstr. 4, D-53115 Bonn, Germany
[§]Permission requests to reuse material from this chapter should be directed to Wiley.

Carbon-rich molecules typically consist of linear (ethynylene, butadiynylene) and cyclic (phenylene, polyaromatic hydrocarbons, PAHs) building blocks. By carefully connecting these components, structures such as e.g., molecular ladders and molecular spoked wheels (MSWs) can be formed, exhibiting remarkable inertness, rigidity, and mechanical stability.[233–235] MSWs feature a central hub, radially connected spokes, an outer macrocycle, and potential side chains. Synthesized with various functionalizations, they enable versatile applications such as liquid crystals and mesoscopic chromophores.[236,237] A key parameter of MSWs is their perimeter, measured in units of *p*-phenylenes (Ph).[234,238] Expanding MSWs laterally with larger precursors to achieve perimeters of 30-Ph or more presents significant challenges. Larger building blocks tend to aggregate and oligomerize, preventing controlled intramolecular bond formation. To mitigate aggregation, spacer side chains and fluorene spokes can be introduced, minimizing contact between the wheels. In a joint experimental-computational effort, quantum chemical simulations were employed to predict the synthetic accessibility of certain MSWs, and these predictions were validated by the successful synthesis of 30-Ph MSWs, followed by imaging via scanning tunneling microscopy (STM).[239]

The synthesis conducted by Höger *et. al.* begins with the formation of anchor-like building blocks. Connected by various cross-coupling methods[240,241] to dumbbells, they undergo cyclotrimerization to form the open precursor of the MSWs. The closure occurs stepwise through Ni(II)-mediated Yamamoto coupling of neighboring arylbromide units.[242] Purification typically yields high amounts (40-70%) of MSWs, with mass spectrometry indicating the formation of minor dimeric and oligomeric byproducts. For the computational investigations, two criteria for successful coupling were established: (1) the strain energy, defined as the reaction energy of the sixth and last coupling step, of the product must be less than 35 kcal/mol, the highest strain measured in previous coupling products, and (2) the carbon atoms to be coupled must approach within 5.4 Å, the distance in the reactive intermediate of the Yamamoto mechanism. Molecular dynamics (MD) simulations at the GFN-FF(ALPB) level[120] were carried out to derive distance distributions for criterion (2) and generated low-energy conformers for a multi-level approach to obtain energies for criterion (1). Additionally, aggregation tendencies were assessed via MD and the Yamamoto coupling mechanism was computed using the growing string method (GSM) and a multilevel approach for reaction energies.[243] The spoke-to-rim ratio (in units of Ph) is critical for predicting successful coupling. Among the synthesis candidates, only those with a 1:1 ratio yielded the desired sixfold coupled product. Candidates with larger ratios formed bowls instead of planar wheels, generating excessive strain energies as predicted by theory.

STM investigations were performed by Höger *et. al.* on a graphene surface and compared to computed STM pictures. While the larger fluorene-spoked 30-Ph MSWs exhibited dense packing along the graphene grid's main orientation axis, it was not possible to obtain images of the smaller 18-Ph MSWs. The side chain length and bulkiness influence the packing density and determine whether C_6 or C_3 symmetric or mixed domains are formed. MD simulations and reaction path search at the GFN2-xTB and GFN-FF levels of theory showed that the fluorene units, which bear alkyl side chains to prevent aggregation during synthesis, cannot rotate freely in the coupled MSW, resulting in non-interconvertible conformers. Consequently, the orientation of the side chains (above or below the MSW plane) is determined during the stepwise formation, making it highly unlikely for all side chains to align in the same direction. However, MSWs can only adsorb to the HOPG surface in side chain-free areas. On the other hand, the fluorene units of the 30-Ph MSWs rotate more freely, allowing for a dynamic distribution of side chains and thus successful STM imaging.

[†]The nomenclature has been updated in this publication and in the corresponding Chapters

CHAPTER 6

Quantum Chemistry Insight into the Multifaceted Structural Properties of Two-Dimensional Covalent Organic Frameworks

Julia Kohn,[†] Hong Li,[‡] Austin Evans[§], Jean-Luc Brédas,[‡] Stefan Grimme,[†]

Received: 29 November 2022
Published online: 20 March 2023

Reprinted in Appendix D (adapted) with permission[¶] from
J. T. Kohn, H. Li, A. M. Evans, J.-L. Bredas, and S. Grimme, *Quantum Chemistry Insight into the Multifaceted Structural Properties of Two-Dimen-sional Covalent Organic Frameworks*, Chem. Mater. **35**.7 (2023) 2820, DOI: 10.1021/acs.chemmater.2c03555

Own contributions

- performing of all calculations
- interpretation of the results
- writing of the manuscript

[†]Mulliken Center for Theoretical Chemistry, University of Bonn, Bonn, 53115, Germany
[‡]Department of Chemistry and Biochemistry, The University of Arizona, Tucson, Arizona, 85721-0088, USA
[§]Department of Chemistry, University of Florida, Gainesville, Florida, 32603, USA
[¶]Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

A key challenge in designing materials for organic electronics (OE) is controlling molecular alignment and macroscopic order.[244,245] Two-dimensional covalent organic frameworks (2D COFs) offer the advantage of easily tailorable molecular synthesis combined with the long-range order of polymers or crystals, making them ideal candidates for OE applications.[246,247] However, due to their sheet-like porous structure, it is challenging to extract reliable structural data.[248] Typically, an eclipsed stacking is assumed, motivating efforts to precisely control interlayer arrangements to engineer specific material properties.[249,250] This work moves beyond a simple single-structure approach to investigate the influence of serrated stacking on optoelectronic properties using semiempirical (SQM) methods with periodic boundary conditions (PBC) as implemented in DFTB+.[251] GFN1-xTB[142] was employed to study an imine-linked COF with quadruple-linked pyrene (TAPPy) and double-linked naphthalenediimide (NDI) arms, known for its good electrical conductivity and tunability.[252,253] All non-rotated bilayer stacking patterns were systematically computationally evaluated and the impact of various structural parameters was assessed, specifically cell vectors, undulation, porosity, and shift, with respect to the relative stability, band structure, IR and X-ray (XRD) spectra. GFN1-xTB, while underestimating electronic properties, provided fast and reliable structural results for more than 120 TAPPy-NDI-COF polymorphs with almost 400 atoms per unit cell, which would otherwise be computationally inaccessible.

2D COFs can exhibit different pore sizes, making them highly relevant for applications such as gas and solvent filtration and storage,[254] carbon capture,[255] electrochemistry,[256] and nanomedicine.[257] Pore size depends on lattice vector lengths and angles, internal staggering, and interlayer interstitial filling. Short cell vectors or significantly distorted cell geometries favor undulation of COF sheets. However, undulation does not prevent large pores, which are governed by the extent of out-of-plane tilting of COF arms. Interstitial filling reduces pore size but increases stabilizing energies due to enhanced dispersion interactions through increased interlayer contact. Porous COFs are synthetically prepared using supercritical CO_2, which has no surface tension.[244,258] Annealing causes pores to collapse, indicating the presence of kinetically and thermodynamically stabilized forms, with reversible switching between them. The calculations confirmed that small-porous conformations are thermodynamically favored, while kinetically favored large-porous conformations can adopt three distinct interlayer shift motifs: small symmetrical shifts, large shifts with over half a unit cell offset, and intermediate shifts with about 30% offset in one direction and up to 10% in the other. The latter was identified computationally as the polymorph contributing predominantly to experimentally measured multilayers. Simulated XRD spectra were able to distinguish between all four structure motifs and showed excellent agreement with experimental spectra.[259] Furthermore, the computational ensemble IR analysis correlated characteristic bands to shifts and structure motifs, suggesting that perfectly eclipsed polymorphs do not significantly contribute to the experimental mixture.[259]

The electronic properties, particularly the position of the conduction band, are dominated by the arrangement of NDI units. Greater overlap between NDI units from neighboring layers results in a smaller electronic band gap and higher charge transport rate. Significant band dispersion occurs only in the interlayer direction. However, these trends were determined by inspection, as an accurate and accessible structural parameter model for differentiating such multidimensional motifs could not be established but could be part of future work. In conclusion, it was demonstrated that this multistructural approach describes TAPPy-NDI-COF more closely than commonly used single structure approaches and better accounts for the inherent polymorphism of 2D COFs.

[†]The nomenclature has been updated in this publication and in the corresponding Chapters

CHAPTER 7

Efficient Calculation of Electronic Coupling Integrals with the Dimer Projection Method via a Density Matrix Tight-Binding Potential

Julia Kohn,[†] Nora Gildemeister,[‡] Stefan Grimme[†], Daniele Fazzi,[§] Andreas Hansen,[†]

Received: 13 July 2023
Published online: 11 October 2023

Reprinted in Appendix E (adapted) with permission[¶] from
J. T. Kohn, N. Gildemeister, S. Grimme, D. Fazzi, and A. Hansen, *Efficient calculation of electronic coupling integrals with the dimer projection method via a density matrix tight-binding potential*, J. Chem. Phys. **159**.14 (2023), DOI: 10.1063/5.0167484

Own contributions

- coding and implementation of DIPRO into `xtb`
- performing of all calculations except for those with Gaussian
- interpretation of the results
- writing of the manuscript

[†]Mulliken Center for Theoretical Chemistry, University of Bonn, Bonn, 53115, Germany
[‡]Department of Chemistry, Greinstrasse 4-6, 50939 Köln, Germany.
[§]Dipartimento di Chimica "Giacomo Ciamician", Via Selmi 2, 40126 Bologna, Italy.
[¶]Permission requests to reuse material from this chapter should be directed to AIP Publishing.

One of the key properties for developing new materials for organic electronics (OE) is their conductivity.[260–262] To investigate conductivity computationally, aside from band-like transport in metals, Marcus theory for hopping transport in semiconductors can be employed, followed by Monte Carlo simulations to assess charge mobility.[189,263] Marcus theory requires charge transfer integrals (J_{ab}), also known as coupling integrals, between different molecular fragments. These integrals can be obtained from various fragmentation methods,[264] such as fragment orbital (FO) DFT, natural transition orbitals (NTOs), or localized molecular orbitals (LMOs),[265,266] combined with coupling integral treatments like projector operator diabatization (POD),[267] generalized Mulliken-Hush (GMH),[268] analytic overlap matrix (AOM),[269] or dimer projection (DIPRO).[270] Exploring all three-dimensional charge transport pathways in a material is very time-consuming, and conformational variety and thermal fluctuations, usually accounted for by molecular dynamics, rendering the calculation of coupling integrals a tedious task. This often limits the screenable chemical space to small organic molecules.[271] In this work, the focus is on the implementation and testing of the DIPRO method in combination with fast semiempirical (SQM) methods to accelerate the evaluation of coupling integrals, thereby enabling large-scale investigations of novel OE materials such as charge transport in metal-organic frameworks (MOFs) or polymer crystals.[272,273]

DIPRO is a post-processing method in principle applicable with any electronic structure method, requiring only three single-point calculations for each of the two fragments (in the dimer geometry) and the supersystem.[274,275] The orbital coefficients of the fragments are projected onto the dimer orbitals and combined to yield an energy-weighted and projected overlap proportional to the macroscopic conductivity. Depending on the underlying method, an empirical scaling factor can be introduced to correct for non-linear relationships between J_{ab} and effective $J_{ab,\text{eff}}$ values.[276,277] Careful testing within this work showed that with a linear scaling improvements of about 60% are feasible. The scaling factor varies slightly for specific structural motifs, the size of the extended conjugated systems, and the distances and angles between the coupling fragments.

Herein, DIPRO was employed with the recently introduced semiempirical non-self-consistent tight-binding potential PTB,[160] which incorporates fitted non-local exchange, enabling it to describe complex systems, like inherently dipolar merocyanines, with the advantages of a hybrid density functional. The approach was tested using the HAB79 benchmark[276] and additionally the novel JAB69 benchmark set was compiled, consisting of medium-sized, organic, planar, co-facially stacked homo-dimers. The DIPRO@PTB coupling integrals were compared to ωB97X-D4/TZ2P[141,278,279] and GMH@NEVPT2[276] reference values. Error estimations showed that different existing implementations of the DIPRO method can already introduce up to 35 meV (10%) error and common SQM methods deviate by up to 100-120 meV (40%) from the reference. In contrast, scaled PTB exhibits significant improvements with a mean absolute deviation (MAD) of only 76 meV, performing comparably to GGA DFT but with far fewer structural outliers and a computational speed-up by a factor of 300. This allowed to test the MEROCY9 benchmark,[189] dimeric cutouts from merocyanine crystals, and a metal-organic cage (MOC) host-guest system with a fullerene acceptor, involving almost 500 atoms. Notably, PTB was able to reproduce both the trend and magnitude of J_{ab} for these challenging systems, with about 50% improvement relative to other SQM methods. Together with its chemical robustness, this renders PTB a suitable SQM method for screening purposes in OE material design via the DIPRO approach.

[†]The nomenclature has been updated in this publication and in the corresponding Chapters

CHAPTER 8

A Semi-Automated Quantum-Mechanical Workflow for the Generation of Monolayers and Aggregates

Julia Kohn,[†] Stefan Grimme[†], Andreas Hansen,[†]

Received: 23 July 2024
Published online: 25 September 2024

Reprinted in Appendix **??** (adapted) with permission[‡] from
J. T. Kohn, S. Grimme, and A. Hansen, *A Semi-Automated Quantum-Mechanical Workflow for the Generation of Merocyanine Monolayers and Aggregates for Organic Electronics*, J. Chem. Phys. **161**.12 (2024), DOI: 10.1063/5.0230341

Own contributions

- coding and implementation
- performing of all calculations except for coupled cluster
- interpretation of the results
- writing of the manuscript

[†]Mulliken Center for Theoretical Chemistry, University of Bonn, Bonn, 53115, Germany
[‡]Permission requests to reuse material from this chapter should be directed to AIP Publishing.

Structure elucidation is crucial for designing new materials, as it helps understand structure-property relationships and induce controlled changes in target material properties. This task is highly challenging, both experimentally and computationally, due to the vast conformational space that must be explored. Inherent polymorphism, structural disorder, and the strong dependence of material properties and quality on fabrication conditions complicates the modeling even more.[4] Typically, three approaches are used: periodic calculations,[280] the cluster approach,[281] and the single structure approach.[282] In this study, the cluster approach was adopted and a quantum mechanical workflow for generating aggregates and thin films was presented, using merocyanines as an example.

The workflow employs the automated interaction site screening (aISS) algorithm with the semiempirical GFN-xTB methods.[138] The aISS algorithm yields ensembles of high interaction energy aggregates, although other conditions such as surface coverage or electric field, could also be the target. Using a constraining sandwich potential and attractive directing potentials, layerwise and columnar growth modes were implemented, enabling controlled layerwise docking of flat molecules, analogous to atomic layer deposition (ALD). This workflow was then utilized to create the D11A8MERO benchmark set, consisting of tetrameric flat and stacked aggregates from 28 different merocyanines, totaling 512 structures. These clusters closely resemble the structural motifs identified in experimental polymorphic crystal structures.[189] Unlike the single structure approach, the use of a Boltzmann weighted ensemble provides a more diverse and realistic description of the material. Merocyanines are important for organic electronics (OE) applications, such as organic field effect transistors (OFETs)[283] and dye-sensitized solar cells (DSSCs).[30] They are inherently dipolar and exhibit complex stacking motifs that significantly influence their optoelectronic properties. Their computational description is challenging due to intricate charge delocalization and long-range dispersion interactions within the aggregates. Interaction energies for tetramer formation were benchmarked against high-level double hybrid density functional theory (DH DFT) PWPB95-D4/def2-TZVPP.[284] Additionally, coupled cluster (CC) references at the PNO-LCCSD(T)-F12b/aug-cc-pVTZ+D,H=VTZ level[285,286] were computed for a small subset, and the effects of basis set dependencies, Fock exchange, and different dispersion corrections were evaluated. The performance of various semiempirical methods (SQM) and low-cost DFT methods, such as the composite DFT-3c methods[123,124] was assessed. Generally, deviations for the stacked subset were higher than for the flat subset owing to the more complex interactions. Among the SQM methods, GFN2-xTB[143] performed best reaching almost the accuracy of PBEh-3c[126] but at three orders of magnitude lower costs. Among the low-cost DFT methods, ωB97X-3c showed the best performance. Using the GFN-FF,[120] which yielded outstanding results for a force field (FF), i.e., SQM accuracy, aggregate sizes up to several thousands of atoms would be conceivable for future investigations.

Also the influence of the aggregate size on optical and electronic properties using the well-known merocyanine HB238 as a case study was investigated.[287] Angle-resolved absorption spectra were calculated at the OT-ωB97X-D4/def2-SVP(SMD(THF)) level of theory and compared to experimental data. The studies enabled the determination the growth direction of the measured thin film and assign the H- and J-bands to different polarization directions. In accordance with the literature,[288] the calculations showed that at least tetramers are required to qualitatively describe the optical trends correctly. Additionally, the electronic coupling using the dimer projection method (DIPRO) was investigated,[270] with the same conclusion that at least four molecules are needed for a reasonable cluster model.

[†]The nomenclature has been updated in this publication and in the corresponding Chapters

CHAPTER 9

Conclusion and Outlook

Novel materials drive technological progress. Our world would be vastly different without the advent of carved stone, pottery, metallurgy, glass, concrete, or plastics. Modern high-performance materials such as organic electronics (OE) are typically tailored for specific purposes, demanding significant time and cognitive resources for optimization. Computational methods can expedite the discovery and engineering of new materials, allowing us to explore previously unknown material spaces. This interdisciplinary process benefits from diverse perspectives, effectively addressing design challenges. Among bottom-up approaches, computational chemistry is important, providing fundamental insights into processes. However, first principles methods represent only one piece of the computational material design (CMD) puzzle. Addressing complex systems with an optimal cost-to-accuracy ratio usually requires multi-level workflows for both qualitative and quantitative analysis in CMD. The presented thesis focused on the development of computational workflows for supramolecular chemistry[1–3] and OE,[4–6] leveraging DFT, SQM, and FF methods, and concentrated primarily on structural, energetic, and electronic properties.

In Chapter 3 a workflow for the structure generation of artificial molecular muscles (AMMs)[289] and the investigation of their mechanical properties are presented. AMMs are mechanically interlocked supramolecular organic compounds, consisting of several hundreds to thousands of atoms, which mimic the contraction and expansion of human muscles[62] in response to various stimuli such as metal ion exchange, redox reactions, or light.[214–217] The mechanical properties of AMMs, including contraction ratio, contraction mechanism, and contraction strength, are of particular interest in CMD.[218] However, their considerable length and flexibility make it challenging to build standardized and reproducible structure models. The real structure of AMMs is far more coiled than the idealized structures typically depicted. To address this, a workflow utilizing FF and SQM methods was developed. The initial structural guess stretches in a semi-relaxed scan, preserving crucial contact positions. The most stretched but chemically intact structure undergoes a conformer search at the FF level, followed by energetic reranking at the SQM level. The resulting final structure prediction is reproducible. This workflow was validated against experimental data for contraction ratios[219] using a newly developed AMM benchmark set that includes various contraction mechanisms. The availability of this workflow allows different researchers to investigate the same systems collaboratively, with the defined protocol ensuring full data reproducibility and offering an accessible starting point for exploring this class of compounds.

The focus of Chapter 4 lies on the structural and electronic properties of dibenzopentalene (DBP) carbon nanobelts. Such carbon-rich materials hold significant potential for future materials due to their diversity and low environmental impact.[290] The circular belt structure results in curvature of these conjugated moieties, introducing a strong antiaromaticity and a novel type of chirality.[228] Despite their significant strain, these carbon nanobelts are stable and promise applications in organic electronics, which can be tuned by customizing substitution patterns and belt size.[225,227] A dedicated protocol for handling nanobelt structures during and after formation was developed and similarly used in related projects.[9,291] As with the AMMs, the actual shape of the metal-coordinated precursors differs significantly from their usually assumed shape. The precursors are not quasi-circular like the final belts; instead, they adopt a puckered shape for tetrameric belts and an entangled conformation for pentameric belts, maximizing non-covalent interactions between the conjugated moieties. These findings were obtained using the CREST/CENSO workflow,[133,134] followed by a multi-level treatment of the different isomers arising from the orientation pattern of substituted monomers above and below the belt plane. Additional investigations included strain analysis and the assessment of antiaromaticity using nucleus-independent chemical shift (NICS) calculations.[229,230] To date, various carbon nanobelts, differing in their conjugated units, connection patterns, and substitution patterns, have been synthesized. Consistently executing the same analyses according to the presented scheme allows for comparisons and eases to determine design principles for enhancing specific properties and build a compound library for further use in CMD.

In Chapter 5 also supramolecular carbon-rich compounds are explored, focusing this time on molecular spoked wheels (MSWs).[235] Beyond their role in fundamental science, MSWs have practical applications such as surface functionalization for antimicrobial coatings or drug recognition. Their synthesis is straightforward, utilizing click-chemistry[292] and metal-mediated C-C coupling reactions,[242] allowing easy modification and customization, making them intriguing for CMD. MSWs are characterized by their rim perimeter, spoke radius, and the connecting units of the spokes. To predict the feasibility of various combinations prior to synthesis, a protocol based on structural and energetic criteria was developed. FF molecular dynamics[120] provide a distance distribution of the rims to be coupled. If this distance falls within a reasonable range for forming the intermediate transition metal complex, C-C coupling is feasible. To calculate the reaction energies of the last coupling step, a measure for the built-up strain and, a multi-level procedure was employed due to the immense size of the MSWs. Fully coupled MSWs tolerate strain, leading to bowl-like deformations. However, forming such a bowl requires one of the flat intermediates to distort out-of-plane, which is highly unfavorable if the rim-to-spoke length ratio is too small. This procedure also predicts the behavior of more complex wheels with inner and outer rims, potentially saving weeks of synthetic trial and error by revealing that spokes can couple across rims, leading to entangled, non-wheel-like compounds. By investigating numerous MSWs according to these criteria using the presented workflow, design principles for CMD can be derived. The valuable insights gained into these 2D compounds can then also be used to develop 3D molecular architectures.[12]

Most of the presented work utilized molecular or cluster models. In contrast, in Chapter 6 a periodic model of 2D layered covalent organic frameworks (COFs) was employed. COFs are highly modern and versatile compounds, primarily used for gas and liquid storage and filtration due to their inherent porosity.[254] Additionally, their conveniently tunable electronic properties enable applications in energy conversion and organic electronics.[246,247] COFs offer significant customization potential, as central hubs, bridging units, and the coordination number of nodes can all be varied.[253] Consequently, they have become an important compound class subject to machine learning in CMD,

holding promising potential for numerous sustainable future applications.[293] Due to their 2D layered structure, COFs exhibit variable morphologies concerning density, layer undulation, and layer offset, making it challenging to build representative and reproducible structure models for CMD.[248] A SQM-based workflow was developed to generate a solid-state ensemble of various COF polymorphs by systematically altering the stacking pattern and offset. The resulting ensemble was clustered into dominant motifs and characterized according to energetic and electronic properties using DFT. This investigation helped to understand the diversity of COFs in the solid state and how it influences material properties. Furthermore, it highlighted the importance of a polymorphic approach in CMD to model systemic properties more reliably and realistically and proposed a valuable tool for COF conformational space exploration.

The focus of Chapter 7 lies less on the generation of structure models but more on electronic properties. In organic electronics, conductivity is the crucial property,[260,261] as it determines the usefulness of novel materials in e.g., dye-sensitized organic solar cells (DSSCs) and organic field-effect transistors (OFETs). Merocyanines, inherently dipolar donor-acceptor systems with high customization potential, are a well-studied and versatile compound class for this purpose.[30,283] Chapter 7 presented the implementation of the dimer projection method (DIPRO)[270] into the `xtb` program package for the calculation of charge transfer integrals, which allows for estimating the conductivity in molecular organic crystals. Although DIPRO can theoretically be used with any underlying method, DFT is typically too computationally expensive for such investigations due to the need to simulate many charge-hopping pathways through the crystal with large coupling moieties.[189] To test the performance of efficient SQM methods versus DFT the DIPRO@xTB method was applied to a novel benchmark set consisting of various aromatic homo-dimers.[276] With the aid of an empirical scaling factor, the PTB method[160] describes charge transfer integrals more accurately and with less outliers than the tested GGA DFT methods. Additionally, organoelectronic crystalline materials of realistic size were tested successfully in a reasonable amount of time at the example of a small merocyanine benchmark. DIPRO@xTB is a versatile tool that supports CMD for semi-conductive organic materials and can easily be employed in future workflows, like the work presented in Chapter 8, thanks to its seamless integration into the `xtb` program.

Chapter 8 focused on a workflow for generating monolayers and molecular aggregates, using an extensive merocyanine benchmark set as an example. As previously mentioned, an ensemble approach is better suited to describe real systems than a single-structure approach.[4] However, for aggregates and monolayers, a simple conformer search is insufficient because it does not capture different constitutional motifs.[189] To address this issue, a workflow utilizing FF and SQM methods, and employing the automated interaction site screening (aISS) docking module in `xtb`,[138] was developed. By using constraining potentials, both stacked and flat aggregates can be obtained. The generated clusters feature realistic motifs from experimental polymorphic crystal structures,[294] except for strongly slipped-stacked arrangements, which occur in the crystal due to solid-state packing effects that are challenging to replicate in the molecular phase. This workflow was then applied to generate a comprehensive benchmark set of several hundred merocyanine clusters. The performance of SQM and low-cost DFT methods for energies was assessed and compared to high-level coupled cluster references. Additionally, studies on the convergence of optoelectronic properties with the cluster size were conducted, highlighting that tetrameric clusters already represent a good qualitative structure model for the bulk.[288] The workflow is particularly useful for CMD as it provides reproducible starting structures and cluster conformer ensembles for further investigation. It may also help the observation of mechanisms involved in thin film and cluster growth and rearrangement in the future.

In summary, this thesis approached CMD from various directions presenting several important compound classes such as merocyanines, COFs, and carbon-rich compounds in the fields of OE and supramolecular complexes. Systematic and in-depth investigations were conducted to better understand the properties and mechanisms of these materials, often involving the creation of novel benchmark sets and extensive method testing. By focusing on the development and implementation of workflows and protocols for structure generation, manipulation, and property investigation, this work provides valuable tools for the reliable and reproducible study of molecules, molecular clusters, and 2D materials. By covering a range of compound classes, phases, methods, and properties, the thesis aims to offer insights that may appeal to a diverse audience. The workflows have been designed with flexibility in mind, allowing for potential modifications and extensions to accommodate more specialized applications. While this work provides an initial methodological contribution, it is anticipated that future research in computational chemistry, organic electronics, and first-principles CMD might build upon these findings. It is hoped that these contributions, though modest, may assist in advancing understanding in these areas.

Future projects, building on this work, include testing and adapting the novel GP3-xTB method, currently under development, for organoelectronics, with a particular focus on merocyanines (inspired by Chapter 7). Additionally, extending the surface coverage workflow from Chapter 8 to address slipped-stacked aggregates, secondary formation conditions, and scaling up to multilayers is anticipated. Creating a library for carbon nanobelts, MSWs, and molecular ladders, alongside a protocol for the automated calculation of their stiffness, is also envisioned (inspired by Chapters 4 and 5). Further advanced tests with specialized parametrizations of SQM and FF methods for merocyanines are conceivable, as is the development of a new reference method for large supramolecular systems (inspired by Chapter 8). Lastly, investigating the rearrangement mechanism from merocyanine monolayers to bilayers is another exciting direction inspired by Chapter 8. May these endeavors inspire and guide future research in the field.

APPENDIX A

Quickstart Guide to Model Structures and Interactions of Artificial Molecular Muscles with Efficient Computational Methods

Julia Kohn,[†] Sebastian Spicher,[†] Markus Bursch,[†] and Stefan Grimme*[†]

Received: 13 October 2021
Published online: 29 November 2021

Own contributions

- initial idea
- performing of all calculations
- interpretation of the results
- writing of the manuscript

[†]Mulliken Center for Theoretical Chemistry, University of Bonn, Beringstr. 4, D-53115 Bonn, Germany
[‡]Permission requests to reuse material from this chapter should be directed to Royal Society of Chemistry RSC.

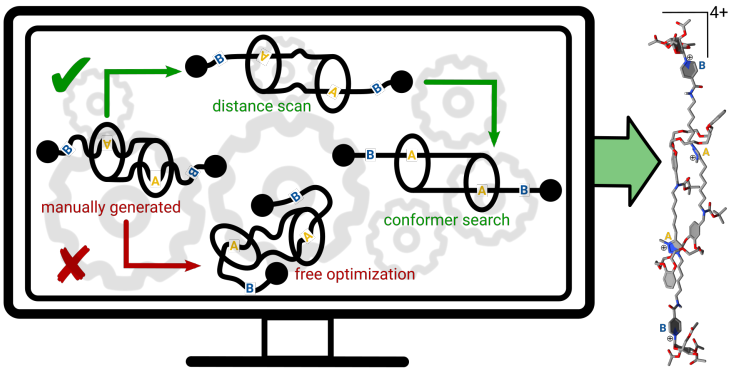

Figure A.1: Table of content graphic (ToC).

Abstract Artificial molecular muscles (AMMs) represent an important subgroup of molecular machines. Their theoretical treatment is challenging due to size, element composition, and complex interaction motifs. Moreover, experimentally determined structures often only yield insights into the covalent connectivity of atoms rather than their three-dimensional structure. Accordingly, a reproducible computational modeling of such structures is complicated. In this work we present a standardized, mostly quantum chemical protocol on how to obtain reliable structures from scratch and to compute contraction free energies for daisy-chain rotaxane AMMs efficiently. In this protocol, the recently developed force-field (GFN-FF) and extended tight-binding methods (GFN*n*-xTB) are employed. For comparison, dispersion-corrected density functional theory (DFT-D) based reference contraction free energies ΔG_c were computed. In one case for which data are available, excellent agreement between theoretical and experimental ΔG_c values within 1-2 kcal $\cdot$ mol^{-1} is obtained.

A.1 Introduction

Molecular machines cover a broad range of structure types, imitating the movements and functions of macroscopic mechanic components. AMMs represent an important subgroup of molecular machines that are subject to interdisciplinary research in the field of supramolecular chemistry and molecular nanosciences. Applications cover, *e.g.*, nanorobots,[207] data storage devices,[209] and molecular electronics.[208] AMMs typically consist of hundreds to thousands of atoms with highly flexible and variable structure motives with many different possible conformations. The underlying contraction mechanisms are often light, charge, solvent, or metal ion induced, and thus, their theoretical treatment is challenging. London dispersion corrected DFT (DFT-D) is among the most popular and frequently applied electronic structure method. It yields accurate results for a variety of chemical systems, including non-covalent interactions.[210,211] Nevertheless, the sheer size of typical AMMs limits the applicability of DFT-D for, e.g., geometry optimizations or molecular dynamics (MD) simulations, due to an extensive computational demand. Alternatively, semiempirical quantum mechanical (SQM) and force-field (FF) methods, such as GFN*n*-xTB/GFN-FF[212], PM*x*[295] or DFTB,[296] can be applied to treat large molecules.

The size of typical AMMs also limits the availability of experimentally determined structure data. Such are usually obtained from, e.g., NMR spectroscopy while more direct X-ray data of the true three-dimensional structure of AMMs is mostly missing. Hence, no experimental input geometries are available for in-depth computational studies and structures have to be generated manually. On the

downside, this is prone to human and technical bias which limits the meaningfulness and reproducibility of such studies. Therefore, we here present an intuitive and easy-to-implement, yet equally standardized and efficient workflow for the generation of disentangled AMM structures for computational studies. Furthermore, the exemplary computation of contraction free energies and geometric contraction ratios is presented.

A.2 Methodology

A.2.1 Computational Details

All calculations were conducted with the `xtb`[297] and `TURBOMOLE`[298,299] program packages. Conformer sampling was performed with the Conformer–Rotamer Ensemble Sampling Tool `CREST`[133] at the GFN-FF[212] level. The implicit analytically generalized Poisson-Boltzmann (ALPB) continuum solvation model[167] was used throughout with chloroform as a typical solvent. Subsequent geometry optimizations, using the command line energetic sorting module for conformer ensembles `CENSO`,[134] were conducted at the GFN2-xTB(ALPB) level. Single-point energies (SPEs) were calculated on various DFT,[124,278,300,301] SQM, and GFN-FF levels. For more details, see *Supplementary Information* (ESI†).

A.2.2 Workflow

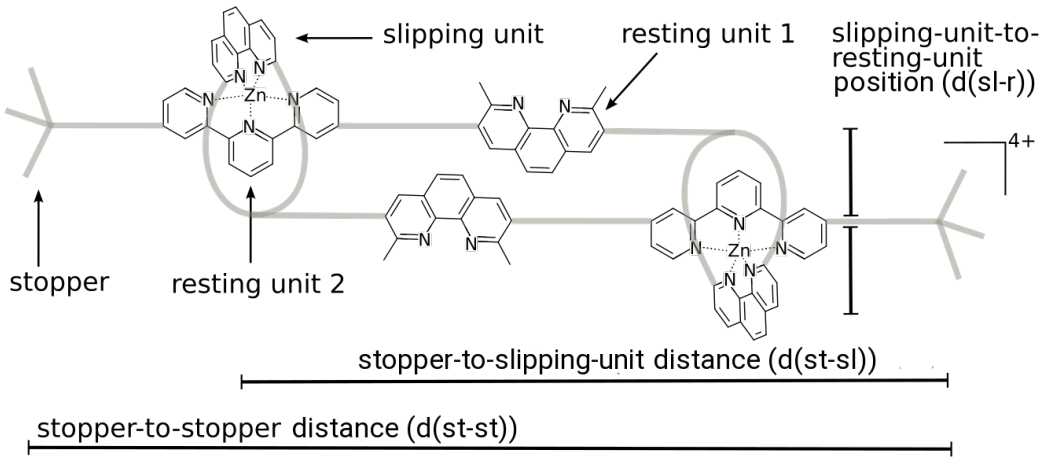

Figure A.2: Schematic representation of a daisy-chain rotaxane AMM and important building blocks and definition of distances.

AMMs usually consist of two non-covalently bound and mechanically interlocked parts called muscle arms. Each arm contains a stopper to prevent unthreading, several resting (or recognition) units that define the contraction state, and a slipping unit as counterpart to the resting units (cf. Fig. A.2). The proposed workflow of the structure generation procedure for such AMMs is shown in Figure A.3. The initial structures of the evaluated AMMs (Fig. A.4) were generated manually using the molecular builder implemented in Avogadro. Various alternatives are available for this step, including computer-aided approaches such as 2D-to-3D conversion.[212] To enforce the disentangling

of the initial structure, a relaxed distance scan, allowing for five optimization cycles at most, of the stopper-to-stopper distance (d(st-st)) was performed at the GFN-FF(ALPB) level. For GFN-FF it is crucial to provide correct integer fragment charges for (possibly ionic) units in order to prevent errors due to artificial charge transfer. In total, d(st-st) is scanned in 200 steps (d(st-st)→d(st-st)+40 Å). At long distances covalent bond dissociation may occur. During the disentangling, the slipping-unit-to-resting-unit arrangement (d(sl-r)) was retained by applying a constrain to at least two independent distances involving muscle-specific structure determine atoms at the slipping and resting units (cf. ESI†). Thereby, a loss of the $\mathbf{X}_{\boldsymbol{long}}$ or $\mathbf{X}_{\boldsymbol{short}}$ conformation by artificial overstretching is avoided. From the scan, the most stretched but still chemically intact form was taken for further evaluation. This structure is input to a conformer sampling conducted with `CREST` at the GFN-FF(ALPB) level to find energetically low-lying conformations. Hereby, prior constraints were kept. Additionally, the stopper-to-slipping-unit distances (d(st-sl)) of the muscle arms were constrained to avoid entanglement in the `CREST` step. The 50 energetically lowest `CREST` conformers were optimized and re-ranked at the GFN2-xTB(ALPB) level to identify the optimum structure. To automatize this step, the `CENSO` program (but without any DFT steps) was employed.

A.3 Results

The presented workflow was applied to four AMMs. The resulting structures from the procedure described above are shown in Fig. A.4. Each AMM represents a different class of daisy-chain rotaxanes, characterized by their contraction mechanism. **1**[214] carries a charge of 4+ and consists of triazolium and amide recognition units, peracylated β-D-mannose stoppers and crown-ether derived slipping units. Its contraction is induced by increasing solvent polarity. **2**[215] consists of α-cyclodextrine slipping units and dihydroxytriazine stoppers but no specific recognition units. Contraction is light-induced through E/Z-isomerization. **3**[216] carries phenantroline and bipyridine recognition units, a crown-ether derived and phenantroline-joined slipping unit and tri(tertbutylphenyl)methane stoppers. The contraction is induced by heating and a metal ion exchange. In this work, we only model the contraction of the Zn^{2+} complex. **4**[217] is built of tri(tertbutylphenyl)methane stoppers, bibenzyl and naphthalene diimide (NDI) resting units and crown-ether derived naphthalene-joined slipping units. It contracts upon twofold reduction of the two NDI units to form a tetra-radical.

For AMMs, the determination of the contraction forces, contraction free energies ΔG_c, and geometric contraction ratios (CR) is of experimental interest.[218] Since AMMs typically show a folded (contracted) conformation in solution[302] the experimental determination of CRs is complicated and can only be achieved by atomic force microscopy or attaching the AMM to a micro cantilever[219]. Due to this lack of experimental data, many property evaluations are done computationally for which well standardized structures are required. Here, the AMM structures generated with the presented protocol are used to calculate contraction energies and ratios.

ΔG_c ($\mathbf{X}_{\mathbf{long}}$→$\mathbf{X}_{\mathbf{short}}$; $\Delta G_c = \Delta E_{el} + \Delta G^{298K}_{mRRHO}$; Tab. A.1) were calculated in the gas phase to prevent obfuscation of the self-interaction error (SIE) by electrostatic models. Thermostatistical contributions were calculated on the GFN2-xTB level at 298 K within the modified rigid-rotor-harmonic-oscillator (mRRHO) approximation.[303] According to its typically robust and good performance for thermochemistry and the absence of SIE, i.e., artificial charge transfer for the charged units, the range-separated hybrid density functional ωB97X-D4 was taken as reference.

ΔG_c for **1** is $\approx$22 kcal $\cdot$ mol^{-1}, favoring the extended conformation. For **2**, the contracted form is

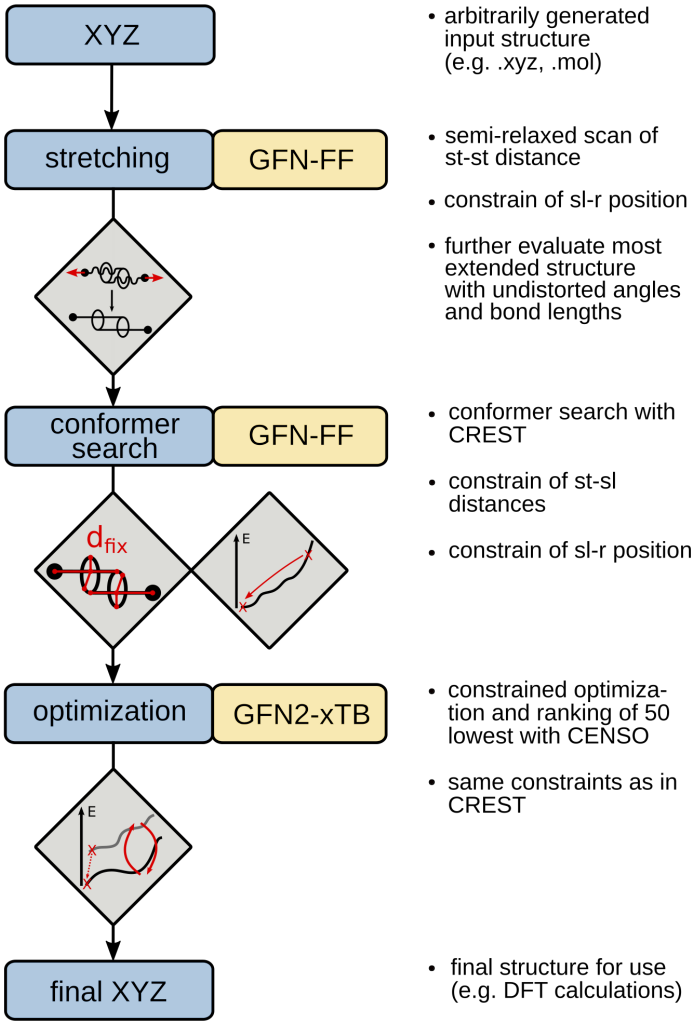

Figure A.3: Proposed structure generation workflow for AMMs. All steps were performed applying the implicit solvation model ALPB.

stabilized by about $-15\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}$ whereas ΔG_c amounts to more than $-200\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}$ for **3**. Compared to ωB97X-D4, the other DFAs yield similar results. For **3**, PBE-D4 underestimates ΔG_c by more than $50\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}$ (22 %). Regarding the SQM methods, both GFN*n*-xTB methods yield qualitatively correct values. Nevertheless, both strongly deviate for **3** similar to PBE-D4. The PM*x* methods also yield reasonable trends but overestimate ΔG_c for **2**. The agreement for **3** is improved over the GFN*n*-xTB methods, probably indicating the beneficial inclusion of Hartree-Fock type exchange (reduction of SIE). GFN-FF performs similar to PM7 while being at least two orders of magnitude faster. For $\mathbf{1}_{long}\rightarrow\mathbf{1}_{short}$ the experimental ΔG_c in solution is $-0.5\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}$.[214] This result is very well reproduced when the solvation free energy $\Delta\delta G_{solv}$ is included employing the COSMO-RS solvent model (see ESI†). To retain computational efficiency, GFN methods use the robust ALPB solvation model. The usage of ALPB results in a shift compared to the gas phase of $-36.9\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}$ for **1** in favor of the short form (see ESI†, $-23.6\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}$ for COSMO-RS).

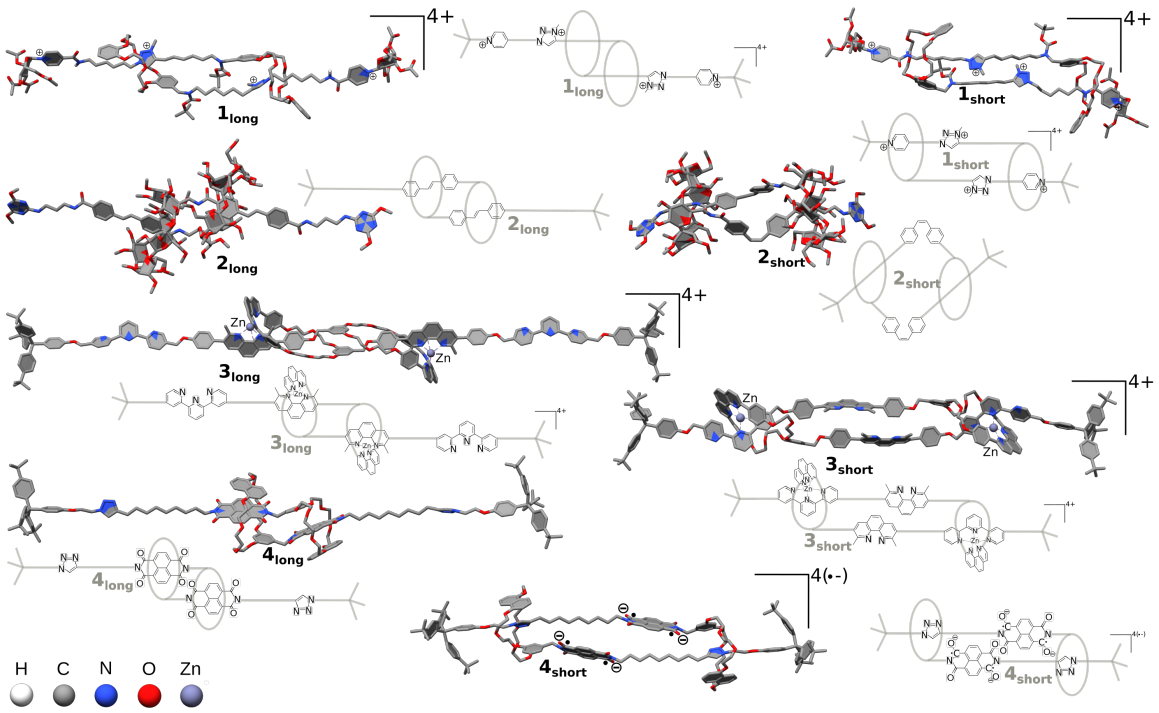

Figure A.4: AMMs and their schematic representation obtained from the protocol described in Fig. A.3. Carbon-bound hydrogens are omitted for clarity.

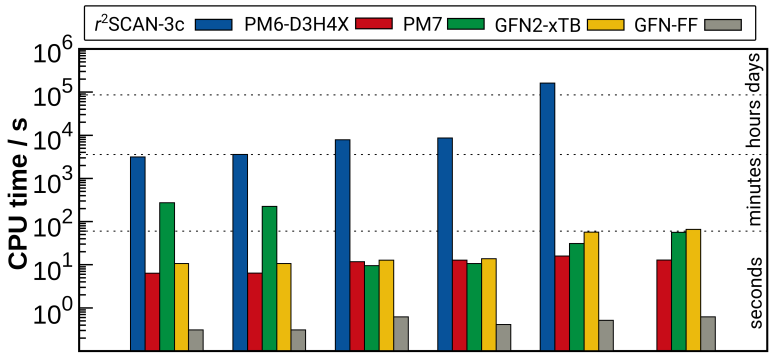

Figure A.5: SPE CPU timings in s for selected methods. All calculations were conducted in parallel on 4 Intel® Xeon® CPU E3-1270 v5 @ 3.60GHz.

The contraction $\mathbf{1}_{long} \rightarrow \mathbf{1}_{short}$ reduces the solvent accessible surface area (SASA). Nevertheless, the electrostatic interaction between the short form and the dielectric continuum is larger than for the extended form. Thus overall, the solvation contribution favors $\mathbf{1}_{short}$. In general, the agreement of the SQM and FF ΔG_c values in solution with the respective DFT values is reasonable regarding their much lower computation time and their almost "black-box" applicability concerning, *e.g.*, SCF convergence (cf. ESI†). Selected SPE CPU timings are depicted in Fig. A.5. The GFN*n*-xTB methods are at least two and GFN-FF more than four orders of magnitude faster compared to DFT, and thus,

Table A.1: $\Delta G_{c,gas}$ in kcal $\cdot$ mol^{-1}. B97X, PBE0, and PBE applied with def2-TZVP basis. ΔG_c + $\Delta\delta G_{solv}$(COSMO – RS(DMSO)) in parentheses.

	$\mathbf{1}_{long}\rightarrow\mathbf{1}_{short}$	$\mathbf{2}_{long}\rightarrow\mathbf{2}_{short}$	$\mathbf{3}_{long}\rightarrow\mathbf{3}_{short}$
ωB97X-D4	22.1 (-1.6)	-15.6	-247.3
PBE0-D4	22.2 (-1.4)	-14.8	-245.7
PBE-D4	25.3 (1.6)	-14.9	-192.2
r^2SCAN-3c	22.8 (-0.9)	-12.7	-215.2
PM6-D3H4X	10.3 (-13.3)	-39.3	-237.0
PM7	-14.7 (-38.3)	-35.8	-259.7
GFN2-xTB	23.8 (0.2)	-11.9	-183.3
GFN1-xTB	20.6 (-3.0)	-15.9	-158.0
GFN-FF	-4.5 (-28.2)	-28.6	-165.7
exp.	-0.5[214]	–	–

can treat much larger systems.
For the applicability of AMMs, the relative distance difference upon contraction, the contraction ratio $\left(CR = 1 - \frac{st-st(X_{short})}{st-st(X_{long})}\right)$, is of importance. CR determination is in general complicated and different computational approaches exist.[219] One is based on MD simulations that yields insight into the molecular movements. To investigate the dynamics of AMMs, MD simulations applying GFN2-xTB and GFN-FF were performed for the energetically lowest structure generated from our protocol and CRs were computed from d(st-st) distributions (Figure A.6). GFN2-xTB and GFN-FF show similar d(st-st) values and the CRs amount to 35 % and 33 %, respectively. These values are in good agreement with the CR from Ref. [217], in which MD simulations under periodic boundary conditions were conducted for a much longer simulation time of 600 ns. The presented approach could also be extended to screening purposes and targeted molecular design tasks for other AMMs.

A.4 Conclusion

In this work, we presented a structure generation protocol for large daisy-chain AMMs. On the resulting structures, the performance of DFT, SQM and FF methods was tested for the computation of contraction free energies. Due to a lack of experimental three dimensional structures and contraction energies, a reliable and reproducible modeling procedure of AMM geometries is desired. Therefore, we combined relaxed distance scans to stretch the initial structure, subsequent constrained conformer searches (CREST at the GFN-FF(ALPB) level), and final geometry optimizations and energy re-weighting at the GFN2-xTB(ALPB) level, to a standardized protocol. The obtained structures were exemplary used for the determination of contraction free energies and geometric contraction ratios. Compared to range separated hybrid density functionals, the GFN and PM*x* methods provided reasonably accurate ΔG_c values. Further, the SQM methods showed higher robustness in terms of SCF convergence in contrast to DFT, and are generally at least two orders of magnitude faster. For **4**, our MD simulations of the generated AMM structure yield similar CR values as with the computationally more demanding protocol of Ref. [217]. In conclusion, we can recommend the presented workflows for the structure generation and CR computation of AMMs in order to obtain reliable and reproducible

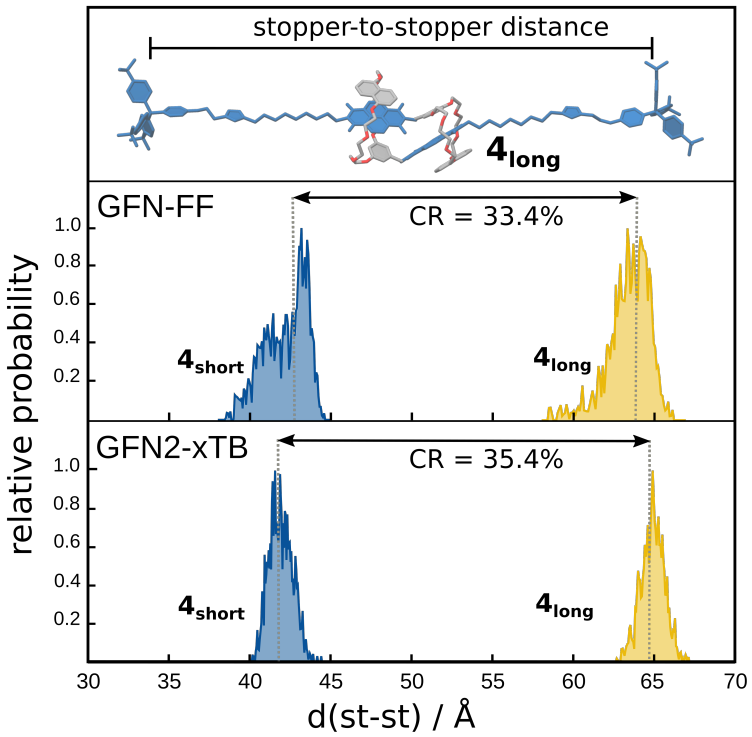

Figure A.6: Structure of **4** with constrained parts in blue. Constraints were applied to keep the extended conformation but to allow slipping, bending and the measurement of an independent d(st-st). Distance distribution function of d(st-st) in Å and CRs for **4** derived from 1 ns MDs. Dotted lines represent the average probability weighted d(st-st).

results. Nevertheless, this is not yet a fully “black-box" procedure and, depending on the system, modifications mainly regarding the constraints may be required. In future studies we will investigate more types of AMMs including elevators, cage and press rotaxanes. As a future goal, an evaluation of the kinetic stability and calculation of the reaction barriers is targeted, by applying semi-automated reaction pathfinders. Furthermore, explicit solvation treatments may describe more realistically the solvent switchability of **1** and contraction forces could be accessible, using fine-tunable electrostatic pulling potentials.

A.5 Acknowledgements

Funding by the DFG (RTG-2591 "TIDE – *Template-designed Organic Electronics*") is acknowledged.

A.6 Data Availability Statement

The data that support the findings of this study are available within the article and its supplementary material (available at https://doi.org/10.1039/d1cc05759f). Any further information is available from the corresponding author upon reasonable request.

APPENDIX B

Increased Antiaromaticity through Pentalene Connection in [*n*]Cyclo-1,5-dibenzo-pentalenes

Jan S. Wössner,[†] Julia Kohn,[‡] Daniel Wassy,[†] Mathias Hermann,[†] Stefan Grimme*[‡] and Birgit Esser[†§¶]

Received: 16 November 2021
Published online: 14 January 2022

J. S. Wössner, J. T. Kohn, D. Wassy, M. Hermann, S. Grimme, and B. Esser, *Increased Antiaromaticity through Pentalene Connection in [n] Cyclo-1, 5-dibenzopentalenes*, Org. Lett. **24**.4 (2022) 983, DOI: 10.1021/acs.orglett.1c03900

Own contributions

- performing of quantum chemical calculations except for strain energies and optical properties
- interpretation of the results
- contributed to writing the manuscript

[†]Institute for Organic Chemistry, University of Freiburg, Albertstraße 21, 79104 Freiburg, Germany
[‡]Mulliken Center for Theoretical Chemistry, University of Bonn, Beringstr. 4, D-53115 Bonn, 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 ACS.

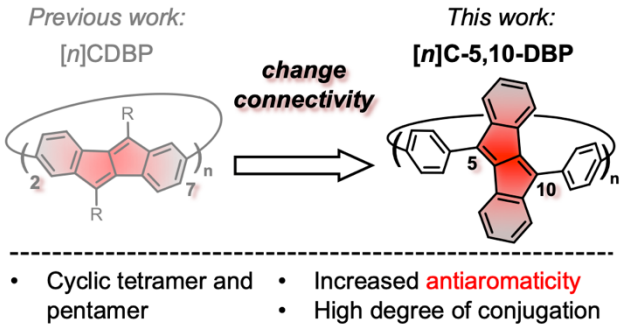

Figure B.1: Table of content graphic (ToC).

Abstract Conjugated nanohoops incorporating non-alternant hydrocarbons have altered optoelectronic properties compared to [*n*]cycloparaphenylenes or other purely aromatic hoops. We synthesized [*n*]cyclo-1,5-dibenzopentalenes (n = 4, 5), in which non-alternant dibenzo[*a*,*e*]pentalenes are connected through their pentalene units. This leads to an increase in antiaromatic character and low-lying LUMO energies. Calculations show puckered or entangled conformations of the precursor macrocyclic Pt-complexes. Our study proves dibenzopentalene as versatile non-alternant building block for conjugated nanohoops with modifiable antiaromaticity and optoelectronic properties.

B.1 Introduction

Curved aromatic hydrocarbons have attracted significant interest[304,305] based on their unique optoelectronic properties,[220,221] aromaticity and host-guest chemistry[222,223] as well as their aesthetic appeal and the challenges associated with their synthesis.[306,307] A strategy to influence the optoelectronic and supramolecular properties of conjugated nanohoops is by varying the aromatic panels they are composed of.[225] Using polycyclic aromatic hydrocarbons (PAHs) often leads to an interesting interplay of conjugation localized within the PAHs or around the hoop. While replacing phenylene units in [*n*]cycloparaphenylenes ([*n*]CPPs) by small PAHs with alternating π-systems usually leads to moderate shifts in the optoelectronics, incorporating non-alternating π-systems effects large changes, even when these are small π-systems.[308,309] An example is dibenzo[*a*,*e*]pentalene (DBP), which, due to its antiaromatic character, displays a relatively high HOMO and low LUMO energy considering the small size of its π-system (16 π-electrons).[310–313] Depending on how DBP is incorporated into more extended π-systems – through the 5,10-positions at the five-membered rings or the 2,7-positions at the six-membered rings (see atom numbering in Scheme B.2) – its antiaromaticity, conjugation and electronic character can be influenced.[226,227] This has been demonstrated in small molecule DBPs[227] but also in DBP-based conjugated polymers, where a change in this connecting pattern altered their electronic properties.[314] Conjugation through the 5,10-positions has a stronger influence on the LUMO energies,[315] while substituents in the 2,7-positions more strongly influence the HOMO energy.[310] Recently we reported on [*n*]cyclodibenzopentalenes ([*n*]CDBPs), in which the DBP units are connected through the 2,7-positions.[228] These hoops showed significant cyclic conjugation and a decreased bandgap due to a large increase in HOMO energy relative to planar reference compounds. The antiaromaticity of the pentalene units in these hoops as well as in other DBP-based hoops connected through the 2,7-positions, however, was slightly decreased.[316–318] To evaluate the influence

of the connection pattern on the antiaromaticity and optoelectronic properties of DBP-based hoops, we now synthesized [*n*]cyclo-1,5-dibenzopentalenes ([*n*]C-1,5-DBPs) with n = 4 and 5, in which the DBP units are connected via their five-membered rings, and evaluated their structural and optoelectronic properties. As opposed to all other DBP-based hoops reported before, we found an increase in antiaromaticity in the pentalene units of the hoops compared to a planar reference compound. The LUMO energies of the [*n*]C-1,5-DBPs are lowered through the 5,10-conjugation.

B.2 Results

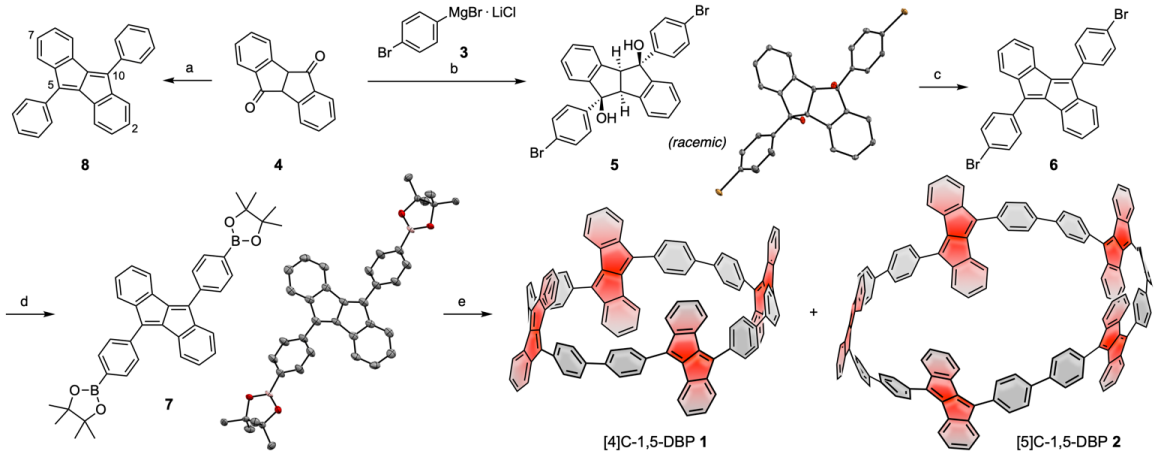

Figure B.2: Synthesis of [*n*]C-1,5-DBPs 1 and 2 and Reference Compound 8[a] and Selected Molecular Structures in the Solid State. [b]

[a] Reactions and conditions: a) PhMgBr, $CeCl_3$, THF, 0 °C to rt, 3 d; then *p*-TsOH · H_2O, toluene, reflux, 20 h, 75% over two steps; b) $CeCl_3$, THF, 0 °C to rt, 3 d, 77%; c) *p*-TsOH · H_2O, toluene, reflux, 20 h, 94%; d) B_2pin_2, Pd(dppf)Cl_2, KOAc, 1,4-dioxane, 80 °C, 2 d, 85%; e) Pt(cod)Cl_2, CsF, THF, 55 °C, 4 d; then PPh_3, toluene, reflux, 2 d (0.8% for **1** and 1.9% for **2**). [b] Displacement ellipsoids are shown at the 50% probability level; hydrogen atoms, are omitted for clarity.

For the synthesis of the [*n*]C-1,5-DBPs **1** and **2** we used the Pt-mediated coupling of boronic esters (Scheme B.2).[319,320] Cerium trichloride-mediated addition of the turbo-Grignard **3** to diketone **4**[321] furnished diol **5** in 77% yield. This addition proceeded diastereoselectively from the convex side of diketone **4**, as the solid-state structure of **5** shows, of which crystals were obtained by slow diffusion of methanol into a chloroform solution. The following acid-catalyzed water elimination afforded purely soluble dibromide **6** in 94% yield. Miyaura borylation furnished bis(boronic ester) **7** in 85% yield as precursor for the Pt-mediated cyclization. As can be seen from its molecular structure in the solid state from single crystals obtained by slow evaporation of a benzene-d_6-solution, this compound is highly planar, explaining its good π-stacking ability. Reaction of **7** with Pt(cod)Cl_2 and cesium fluoride afforded an insoluble product mixture containing the macrocyclic platinum complexes, which was directly subjected to reductive elimination after ligand exchange with triphenylphosphine. This furnished [*n*]C-1,5-DBPs **1** and **2** after successive purification by column chromatography on silica gel and semipreparative gel-permeation chromatography in average yields of 1–2%. The structures of nanohoops **1** and **2** were unambiguously determined using ^{1}H- and ^{13}C NMR spectroscopy and

HRMS. To allow comparing their optoelectronic properties to a planar reference compound, we synthesized **8** from diketone **4** in an analogous way (Scheme B.2, top left). [*n*]C-1,5-DBPs **1** and **2** are soluble in common aromatic or halogenated solvents and are stable as solids and in solution under air and ambient light. Interestingly, the trimeric hoop [3]C-1,5-DBPs was never formed, even when a higher concentration of **7** was used in the macrocyclization reaction.[322] In the case of the 2,7-connected [*n*]CDBPs, we had observed only the tri- and tetrameric nanohoops.[228] This prompted us to investigate the structures of the intermediate cyclic platinum complexes. Calculations at the CREST[133]@GFN2-xTB[143](GBSA[166](toluene)) level of theory showed that the structures of the tetra- and pentameric Pt-complexes as precursors to **1** and **2** are not planar, other than commonly expected (Figure B.3). The tetrameric cycle assumes a puckered shape, while the pentameric cycle takes on an entangled conformation. This results in average Ph-Pt-Ph bond angles of 80° for n = 4 and 81° for n = 5. Both conformations allow for π-π-interactions between the DBP units, in particular the entangled pentamer.

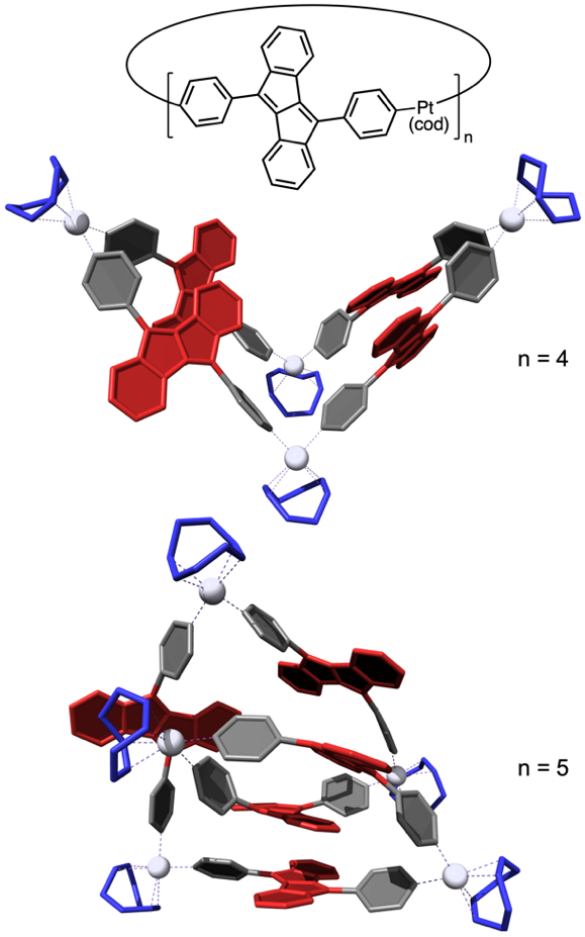

Figure B.3: Structures of the tetra- and pentameric Pt-complexes as precursors to **1** and **2** (CREST@GFN2-xTB(GBSA(toluene))).

The ^{1}H NMR spectra of [*n*]C-1,5-DBPs **1** and **2** show a low-field shift of all signals compared to the planar reference compound **8** (Figure B.3). The three multiplets from the DBP protons (colored in Figure B.3) are less defined in the hoops (Figure B.3B,C) compared to **8** (Figure B.3A). This may indicate a lower symmetry of the hoops than may be expected from the D_n-symmetrical structures shown in Scheme B.2. DFT calculations confirm this assumption in that the different conformational isomers, resulting from a rotation of DBP units through the hoops, have similar energies (within 0–1.7 kcal · mol^{-1}) and that the energetic barriers for rotation of a DBP unit through the hoops is low, hence multiple conformations exist at room temperature (see below). NMR spectra recorded at higher temperatures of up to 343 K for **2** did not reveal significant changes compared to the spectrum shown in Figure B.4B, which indicates that the interconversion of conformers is already fast at room temperature (see SI). Regarding the phenylene units (grey in Figure B.3), with chemical shifts of 7.77 (n = 5) and 7.71 ppm (n = 4) for the centered multiplets, their protons resonate low field-shifted compared to [*n*]CPPs of similar size ([12]CPP: 7.62 ppm,[323] [15]CPP: 7.67 ppm[324]) or the *ortho*-protons in biphenyl (7.63 ppm, all in $CDCl_3$).[325] When comparing [*n*]C-1,5-DBPs **1** and **2** with each other, a high-field shift with increasing ring strain from n = 5 to n = 4 is observed for the phenylene protons, as also reported for [*n*]CPPs or [*n*]CDBPs.[228] This may be caused by an increased quinoidal character of the conjugated π-system in smaller hoop sizes.

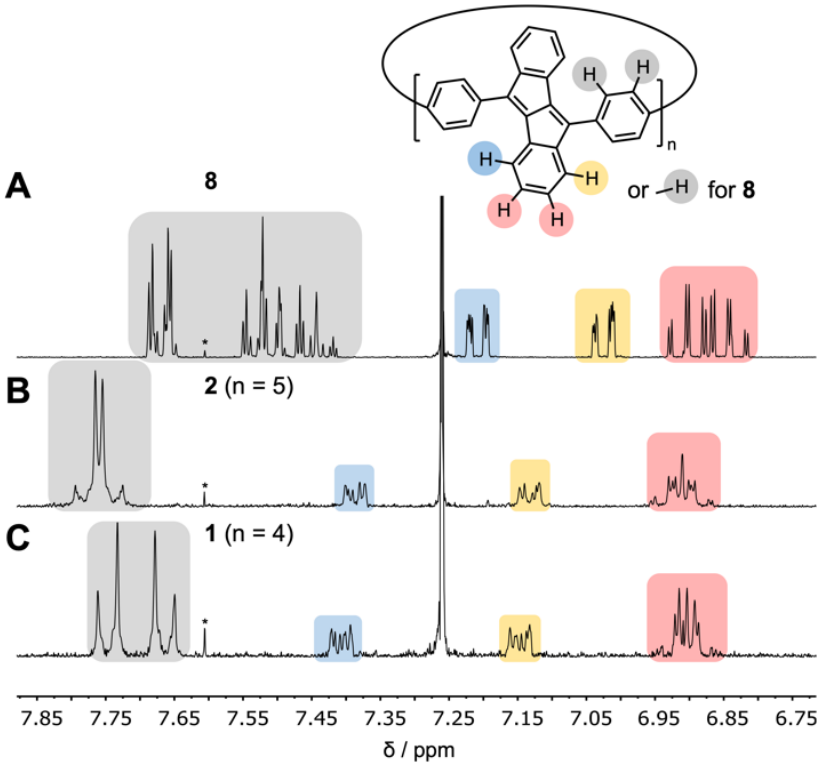

Figure B.4: Selected regions of the ^{1}H NMR spectra of reference compound **8** (A) and [*n*]C-1,5-DBPs **2** (B) and **1** (C) in $CDCl_3$ (300 MHz, rt, signals marked with an asterisk correspond to the ^{13}C-satellite of $CDCl_3$).

To obtain information on the structural isomerism in these nanohoops, we performed DFT calculations at the B97-3c[125] level of theory. The most stable isomers are the AAAB conformation for **1** with one DBP unit rotated compared to the other ones, and the AABAB conformer for **2**. However, all four possible conformations of each nanohoop lie within a small energetic range of $\leq 1.0\ \text{kcal} \cdot \text{mol}^{-1}$ for **1** and $\leq 1.7\ \text{kcal} \cdot \text{mol}^{-1}$ for **2**. From their relative free energies at the GFN2-xTB[143] level of theory we calculated the relative abundance of each isomer to be 18/51/20/11 for AAAA/AAAB/AABB/ABAB in tetrameric **1** and 8/31/5/56 for AAAAA/AAAAB/AAABB/AABAB in pentameric **2** at room temperature (for more information see SI). With $5.5\ \text{kcal} \cdot \text{mol}^{-1}$ for **1** and $5.7\ \text{kcal} \cdot \text{mol}^{-1}$ for **2** the barriers for rotation of a DBP unit are low for both [*n*]C-1,5-DBPs, and free rotation is possible at room temperature (calculated using B3LYP-D4[231]/def2-TZVP (COSMO-RS(CHCl3))[168] level with thermal corrections from GFN2-xTB,[143] see SI for details). These facts explain why the signals in the ^{1}H NMR spectra of **1** and **2** are less well defined than in **8**. The diameters of the [*n*]C-1,5-DBPs amount to 16.8 Å for **1** and 20.0 Å for **2**. With 46.1 and 41.6 $\text{kcal} \cdot \text{mol}^{-1}$, respectively, the strain energies (calculated using B3LYP/6-31G(d)) are similar to those of [*n*]CPPs of similar diameter: [12]CPP with 16.6 Å diameter has a strain energy of 48.1 $\text{kcal} \cdot \text{mol}^{-1}$, and [15]CPP with 20.6 Å diameter of 39.2 $\text{kcal} \cdot \text{mol}^{-1}$.[326] As StrainViz[327] calculations show, the largest constituent is dihedral strain localized in the CC-bonds connecting the phenylene/DBP units within the hoop (see SI).

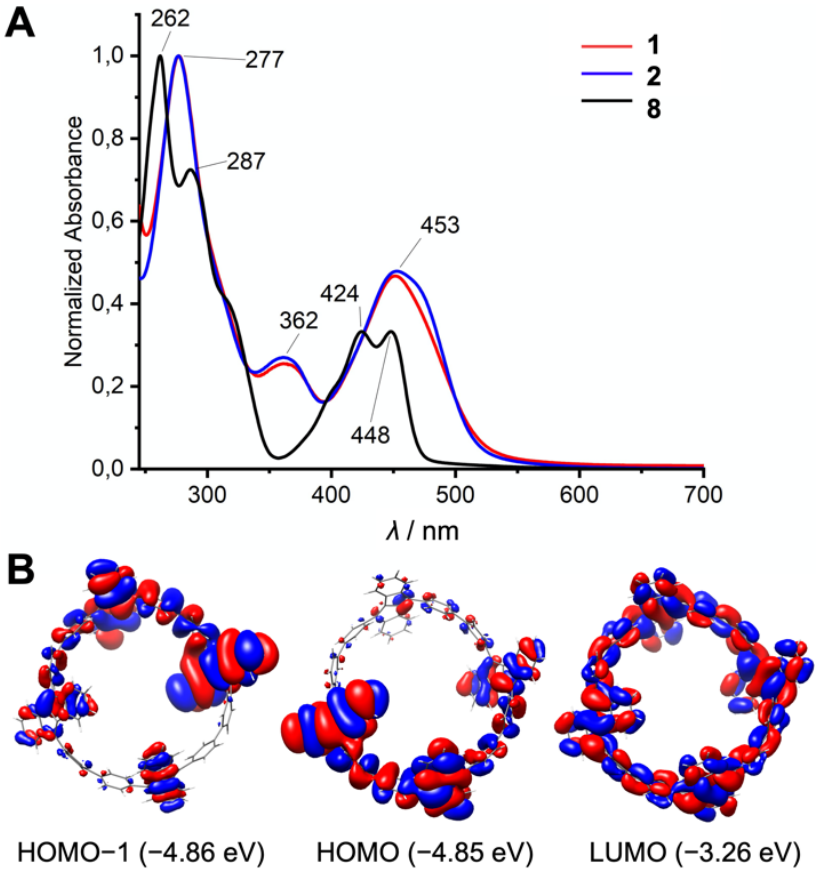

Figure B.5: A) UV/Vis absorption spectra of nanohoops **1** and **2** and reference compound **8** (10^{-5}–10^{-6} M in CH_2Cl_2); B) Frontier molecular orbitals of **1** (AAAB conformer, B97-3c).

The UV/Vis absorption spectra of nanohoops **1** and **2** provide proof for the extended conjugation around the hoops compared to reference compound **8** (Figure B.5). **8** shows the characteristic DBP bands with several maxima in the regions of 250–300 nm and 400–500 nm. In nanohoops **1** and **2**, these characteristic bands are still visible, but red-shifted due to the extended conjugation and less resolved compared to **8**. This distinguishes the [*n*]C-1,5-DBPs from 2,7-connected [*n*]CDBPs, where the characteristic DBP bands had completely disappeared in the absorption spectra.[228] Nanohoops **1** and **2** have optical band gaps of 2.02 and 2.03 eV, respectively, determined from the longest wavelength absorption onsets, which show a red-shift compared to reference compound **8** with a bandgap of 2.11eV. All three compounds were non-emissive, a feature commonly observed in DBP derivatives.[228,310,315]
To assess the extent of cyclic versus linear conjugation in the [*n*]C-1,5-DBPs, we simulated the UV/Vis absorption spectra of linear oligomeric reference compounds (Ph-DBP-Ph)$_n$ as open and H-saturated isomers as well as those of hoops **1** and **2** (see SI). The spectra of the linear oligomers with n = 4 and 5 are very similar in shape to those of the cyclic [*n*]C-1,5-DBPs. However, the longest wavelength absorption maxima are red-shifted in the hoops, for n = 5 by 0.1 eV, compared to the linear pentamer. The absorption onsets reach to 610 nm for the linear reference compounds, but to 700 nm for the hoops due to expressed shoulders above 600 nm in the calculated spectra. This shows that in nanohoops [*n*]C-1,5-DBPs a higher degree of conjugation is present compared to their linear isomers.
The calculated HOMO and LUMO energies of the most stable AAAB and AABAB conformers of hoops **1** and **2** are almost identical with -4.85 eV for the HOMOs and -3.26 respective -3.28 eV for the LUMOs (B97-3c). While the LUMOs are symmetrically distributed over the hoops, HOMO and HOMO-1 are almost degenerate for both **1** and **2** and more localized in one part of the hoop (Figure B.5B for **1** and SI for **2**). Comparing these orbital energies to reference compound **8** as well as the linear oligomers (Ph-DBP-Ph)$_n$ (n = 4, 5) shows that the HOMO energies are hardly affected by the increased conjugation through the five-membered rings (**8**: -4.89 eV; (Ph-DBP-Ph)$_n$: -4.91 eV). The LUMO energies, on the other hand, sink in the order **8** (-3.02 eV), (Ph-DBP-Ph)$_4$ (-3.17 eV) and [4]C-1,5-DBP **1** (-3.26 eV), which demonstrates an additional effect through cyclic conjugation in hoop **1** based on the additional lowering of its LUMO energy compared to linear (Ph-DBP-Ph)$_4$. This observation is in line with previous studies by our group that show that substituents at the five-membered DBP rings significantly influence its LUMO, but not the HOMO energy.[315,321] This stands in contrast to 2,7-connected [*n*]CDBPs, where the LUMO energies were similar to the small molecule reference compound, whereas the HOMO energies were significantly increased by the additional conjugation through the six-membered DBP rings.[228] A similar observation can be made when comparing the electrochemical reduction potentials of 5,10-dimethyl-DBP ($E_{1/2}$ = -2.04 V *vs.* Fc/Fc$^+$)[20] with π-extended 5,10-diphenyl-DBP **8** ($E_{1/2}$ = -1.68 V *vs.* Fc/Fc$^+$),[328] demonstrating a lowering of the LUMO energy of the latter.
To assess the extent of (anti)aromaticity within the DBP units in the conjugated hoops, we performed NICS (xy) scans (nucleus-independent chemical shift[229]) at a height of 1.7 Å above and below the DBP units at the B3LYP[231]/TZP[232] level of theory (Figure B.6).[230] With increasing strain (from **8** to **2** to **1**), the antiaromaticity in the pentalene units significantly increases, reflected in higher NICS values above the 5-membered rings, while the aromaticity of the 6-memembered rings decreases (less negative values). This distinguishes these 5,10-connected [*n*]C-1,5-DBPs from all DBP-based hoops[9,228,317,318] and DBP-phanes[316] investigated before, where the DBP units were incorporated through their 2,7-positions at the six-membered rings, and in which the antiaromaticity decreased in the macrocycles compared to planar DBPs.

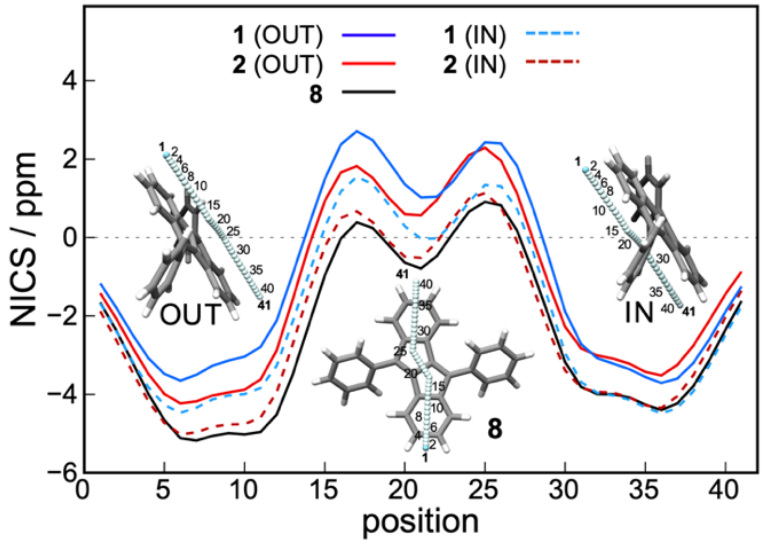

Figure B.6: NICS (xy) scans at a height of 1.7 Å above and below the DBP units in **1** and **2** and reference compounds **8** (B3LYP/TZP).

B.3 Conclusion

In conclusion we herein presented [*n*]cyclo-1,5-dibenzopentalenes ([*n*]C-1,5-DBPs) as conjugated carbon nanohoops constituted of dibenzo[*a*,*e*]pentalenes as non-alternant hydrocarbons. A platinum-mediated cyclization furnished the tetrameric cycle **1** as well as pentameric **2**, and calculations showed the precursor macrocyclic platinum complexes to assume puckered or entangled shapes instead of planar polygons. Compared to a small molecule reference compound and to other DBP-based nanohoops, the antiaromaticity of the DBP units is increased in these [*n*]C-1,5-DBPs, and they feature lowered LUMO energies. We ascribe this to the incorporation of the DBP units into the hoops via their five-membered rings. Calculations and absorption spectroscopy provides proof for the extended conjugation around the hoop in these novel carbon nanostructures. Our study shows how DBP is a versatile non-alternant building block for carbon nanohoops, whose antiaromaticity and optoelectronic properties can by tuned by simply changing the substitution pattern.

B.4 Supporting Information

Materials and methods, synthetic manipulations, NMR spectra (including variable temperature), mass spectra, information on the stereoisomers, single-crystal X-ray diffraction data, details on (TD)DFT calculations, energies and Cartesian coordinates of calculated structures. The Supporting Information is available free of charge on the ACS Publications website. CCDC 2118222 (**5**) and 2043704 (**7**) contain the supplementary crystallographic data for this paper. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/structures.

B.5 Acknowledgements

We thank Dr. M. Keller for NMR experiments, C. Warth for mass spectrometry measurements, and B. Benkmil for recording the X-ray diffraction (all University of Freiburg). Generous support by the German Research Foundation (230408635, 434040413, INST 40/467-1 FUGG and INST 39/1081-1 FUGG) and the state of Baden-Württemberg through bwHPC is gratefully acknowledged.

APPENDIX C

Size-Increased all-Phenylene Molecular Spoked Wheels - A Combined Theoretical and Experimental Approach

Philipp Krämer,[†] Julia Kohn,[‡] David Hofmeister,[†] Maximilian Kersten,[†] Christopher Sterzenbach,[†] Antonia Gres,[†] Andreas Hansen,[‡] Stefan-Sven Jester,[*†] Stefan Grimme,[*‡] and Sigurd Höger[*†]

Received: 12 June 2024
Published online: 7 August 2024

Own contributions

- performing of all quantum chemical calculations
- interpretation of the results
- contributed to writing the manuscript

[†]Kekulé Institute for Organic Chemistry and Biochemistry, University of Bonn, Gerhard-Domagk-Str. 1, D-53121 Bonn, Germany
[‡]Mulliken Center for Theoretical Chemistry, University of Bonn, Beringstr. 4, D-53115 Bonn, Germany
[§]Permission requests to reuse material from this chapter should be directed to Wiley.

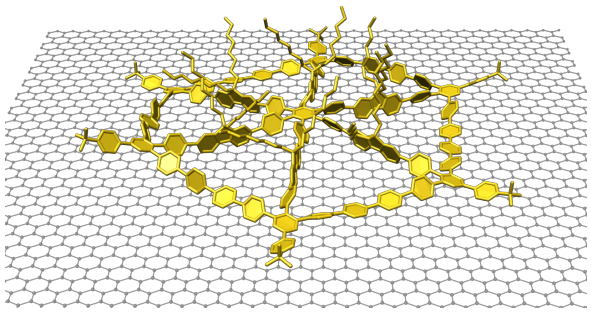

Figure C.1: Table of content graphic (ToC).

Abstract A lateral expansion of molecular spoked wheels (MSWs) based on an all-phenylene backbone is described. The MSWs contain a central hub, six spokes and a rim that is formed by a sixfold *Yamamoto* coupling of the respective non-cyclized dodecabromo precursor yielding MSWs with up to 30 phenylene rings in the perimeter. Attempts to prepare compounds of such size without flexible side groups at the spokes were unsuccessful, most probably due to an aggregation and accompanying oligomerization of the precursors during the cyclization. To overcome these problems, fluorene units are inserted into the spokes. These contain additional alkyl chains and lead to a curvature of the wheels. Quantum chemical calculations on the mechanism of the *Yamamoto* coupling leads to geometry and strain-related criteria for the successful rim closure to the respective MSW. Subsequently, MSWs are prepared with four and even six phenylene units at each edge of the hexagonal wheels. The resulting MSWs are characterized by spectroscopic methods, and additionally some of them are visualized *via* scanning tunneling microscopy (STM).

C.1 Introduction

Carbon-rich compounds, *i.e.* compounds with a C to H ratio of ≥ 1 have been of high interest since the early days of chemistry. They include also all-carbon compounds (carbon allotropes) and cutouts of those, as well as linear, ladder and cyclic structures, and their combinations.[329–331] The fusion of rigid linear and cyclic structures leads to so-called molecular spoked wheels (MSWs) consisting of a central hub, the rigid spokes, and the surrounding macrocycle.[233–235] The connection of the linear and cyclic units leads to a dramatical enhancement of the rigidity of the MSWs as compared to the rod-only or macrocycle-only analogues.[332] MSWs were synthesized in different sizes featuring a broad scope of functionalization ranging from apolar alkyl chain groups,[236] over bottle brush polymers,[237] to polar carboxylic acids.[238] Although the alkyl side groups reduce the C to H ratio, the compound's rigid backbone still allows them to be termed carbon-rich. The molecule's backbone relies generally on phenylene, ethynylene and/or butadiyne moieties and their combinations. However, some of these MSWs show stability problems when they are stored under ambient conditions (air and room temperature), as they show after a while in gel permeation chromatography (GPC) elugrams the formation of additional peaks with a higher molecular weight. In those cases, we ascribe the slow decomposition to the presence of the ethynylene and/or butadiynylene groups. Recent studies in our laboratory therefore focused on structures made of phenylenes only, which provide more rigidity, stability, and inertness.[10,234,238]

The preparation of the all-phenylene MSWs starts by synthesizing an anchor-shaped unit, consisting

of the spoke and a rim fragment of the respective structure. Two of these anchor-shaped molecules were then connected by an acetylene linker leading to a dumbbell-like structure. The central alkyne is cyclotrimerized under Co-catalysis forming the open framed precursor in which each of the six rim segments is terminated with arylbromide groups on both sides. In the final step, the rim is closed in a sixfold *Yamamoto* coupling[242] yielding the all-phenylene MSW (Figure C.2).

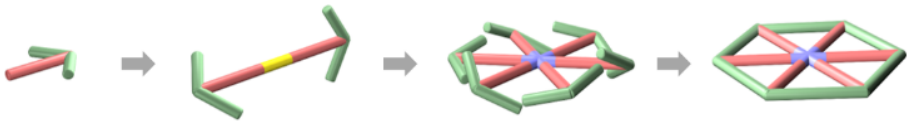

Figure C.2: Schematic synthesis of an all-phenylene molecular spoked wheel.

Most of the all-phenylene based MSWs synthesized in the past consist of six edges, each having a length of four phenylene units. Since the corner phenylenes belong to two edges, the total number of phenylene units in the perimeter is 18 ("18 Ph-MSWs"). We were also able to enlarge this and to prepare a MSW with a total of 24 phenylenes in the perimeter ("24 Ph-MSW") with a yield of 17% in the final *Yamamoto* cyclization.[239] However, this yield was substantially lower than for the respective 18 Ph-MSW (74 %),[238] and the GPC of the crude cyclization product shows the formation of considerable amounts of probably dimers and higher oligomers. Here, we describe the synthesis of MSWs with a larger perimeter and explain in detail why alkyl groups at their spokes are necessary for their successful synthesis. Prior to their preparation, quantum mechanical calculations applying a geometrical and an energetical argument were performed to predict the synthetic success of the final cyclization reaction.

C.2 Results and Discussion

Following the so far successfully applied synthetic protocol, attempts were undertaken to synthesize the next larger MSW with a perimeter consisting of 30 phenylenes ("30 Ph-MSW", Figure C.3). While the synthesis of the MSW precursors with different solubilizing side groups at their corners was straightforward and yields are comparable to the results obtained for the smaller compounds, the cyclization *via Yamamoto* coupling under various conditions completely failed. Investigations of the crude products *via* GPC revealed the formation of species of about twice as large as the desired 30 Ph-MSWs as well as higher molecular mass byproducts in all cases (see supporting information (SI) for synthetic and analytic details).

At first, we could not find a reasonable explanation for the unsuccessful MSW synthesis. Although the larger precursors have a higher flexibility than the precursors for the 18 Ph-MSWs and the 24 Ph-MSWs, the dimensions of the 30 Ph-MSW precursors are still way below the persistence length of poly(*p*-phenylene)s of about 13 nm.[333] However, the absence of 30 Ph-MSWs in the respective crude products of the *Yamamoto* coupling could be explained with a higher aggregation tendency of the larger aromatic structures. During the course of the rim closure this aggregation intensifies and favors intermolecular coupling over the intramolecular ring closure. This also explains why the crude *Yamamoto* cyclization product of the 24 Ph-MSW synthesis contains more dimeric and oligomeric products compared to the respective crude product of the 18 Ph-MSW synthesis (see SI).

In summary, in order to access 30 Ph-MSWs it is essential to inhibit the aggregation of the respective precursors during the final *Yamamoto* coupling. One possible molecule design includes alkyl- or

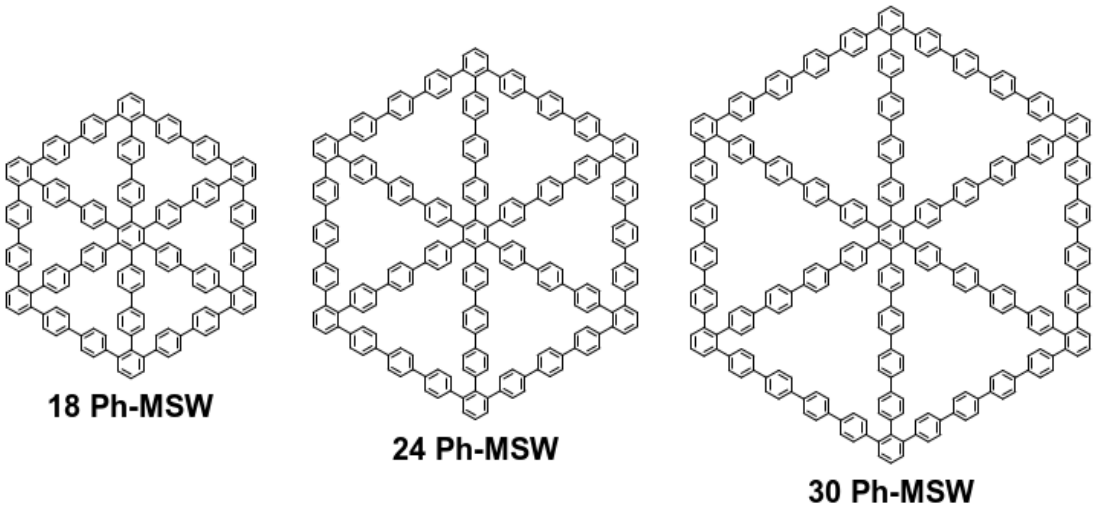

Figure C.3: Backbone structures of a 18 Ph-MSW, 24 Ph-MSW, and 30 Ph-MSW.

alkoxy chains attached to the phenylenes of the spoke units, similar to our previously described phenylene-ethynylene-butadiynylene MSWs.[334]

Alternatively, we identified fluorene units at the spokes of the wheel as an attractive structural motif for several reasons (Figure C.3). First, the 9,9'-positions of the fluorene can be easily alkylated with two alkyl chains aiming to suppress an aggregation of the intermediates by special separation of the molecules during the final *Yamamoto* coupling. Second, the connection of the chains directly to a carbon atom enables a broad scope of functionalization *via* etherification of the rim. Usually, these long alkoxy substituents at the corners of the MSW are introduced after methyl ethers are cleaved at a late stage within the synthetic route (*c.f.* **11c** $\longrightarrow$ **11e**, Figure C.5). This deprotection requires harsh conditions like BBr_3 and would also cleave other ethers already present within the molecule. Direct carbon alkylation gives substrates that are not prone to such conditions and that allow the aforementioned functionalization without further difficulties. This tolerance is proven and exploited in the synthesis of **2c** within this work. Third, the 2,7-connectivity of the fluorene leads, contrary to the 4,4´-connectivity of the biphenyl unit, to a curvature of the intermediates and the final structures with alkyl chains directing to both planes of the molecule, expected to reduce the tendency of all intermediates and the final wheel to aggregate (Figure C.4). Fourth, the 2- and 7-position of the dialkylated fluorene can be addressed regioselectively, by halogenation and subsequent coupling reactions, thus allowing a straightforward synthesis of all required building blocks.

However, the curvature of the new spokes might have a dramatic impact on the synthesis of the wheel structure, since it is *a priori* unclear whether the halogenated carbon atoms of the rim fragments are still in a range so that the intramolecular *Yamamoto* coupling can take place. For this reason, quantum chemical simulations were run to predict the successful synthesis of fluorene MSWs. This theoretical approach goes far beyond purely chemical drawings or ball-and-stick models of the final product structures since it includes all intermediates from the dodecabromide wheel precursor to the final MSW. Figure C.5 describes the respective coupling reaction and the different pathways how the MSW formation can occur.

Most probably, the six *Yamamoto* coupling steps will not run parallel but successively. After the first

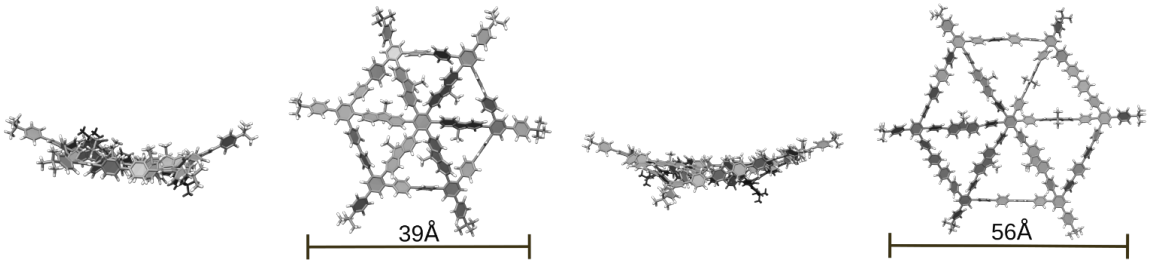

Figure C.4: Geometries of fluorene MSWs **1a** (18 Ph-MSW) and **2b** (30 Ph-MSW), side and top views (alkyl chains at the fluorenes are omitted for clarity).

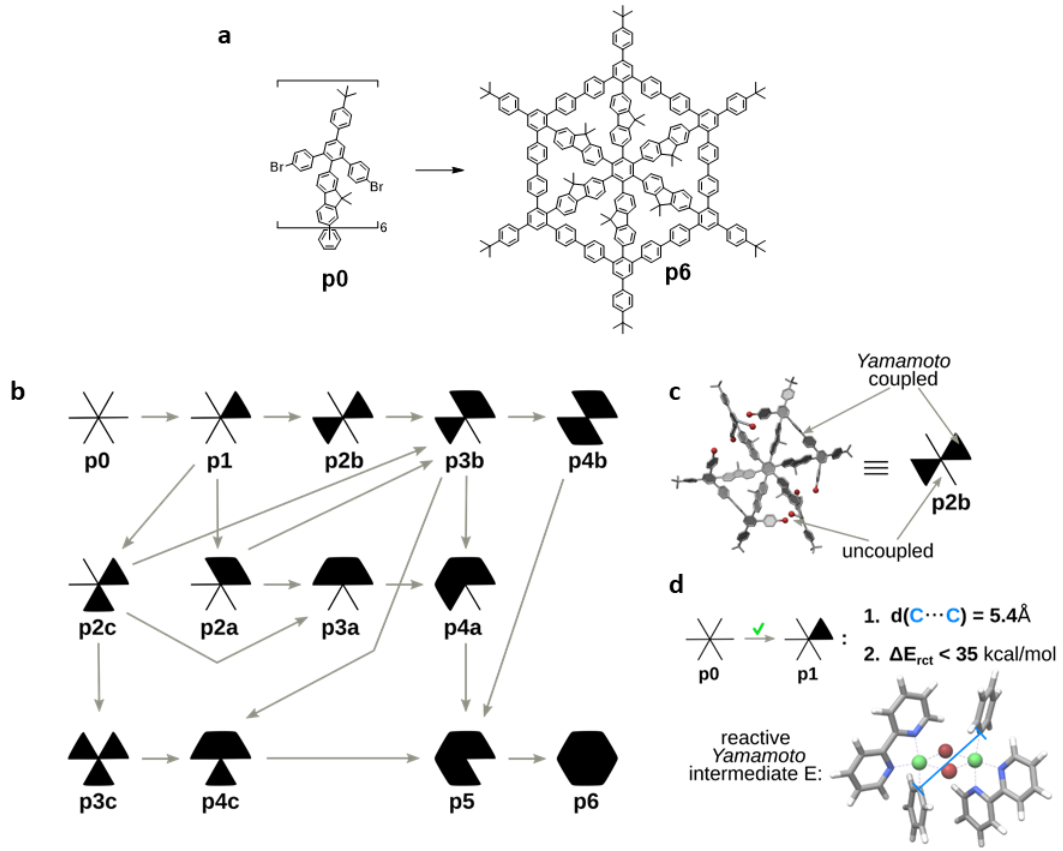

Figure C.5: Proposed final synthetic step of an all phenylene fluorene based 18 Ph-MSW. a) Ring closure of the dodecabromo precursor **p0** to the spoked wheel **p6** by sixfold *Yamamoto* coupling; b) different pathways of the final *Yamamoto* couplings are possible to convert the precursor into the MSW; c) assignment of the schematical representation in b: filled triangles represent coupled rim segments; d) energetic and geometric conditions that must be fulfilled for a successful coupling and the reactive *Yamamoto* intermediate **E** of which the distance condition is derived from, calculated at r^{2}SCAN-3c(DMF) level of theory (details are found in the SI).

ring closure (**p1**), the next ring closure can occur close by (**p2a**), one segment further apart (**p2b**), or at the opposite side (**p2c**, Figure C.5b). From each of these intermediates several pathways for the next

ring coupling steps are possible until **p5** is formed that undergoes the final coupling towards MSW **p6**. In order to address the question if the wheels can be prepared, we performed a computational study on the mechanism of the *Yamamoto* coupling using the growing string method (GSM) as implemented in ORCA V5.0.4 at the ωB97X-3c(DMF)//r^{2}SCAN-3c(DMF) level of theory[123,124,243,335]. In our proposed pathway two neighbored Ar-Ni(II)L_2-Br complexes (L_2 = 2,2´-bipyridine) form the dinuclear [Ar-Ni(II)L_2-Br]$_2$ complex **E** as an intermediate (Figure C.6 and SI for details). In this, the C-C distance for the carbon atoms in the aromatics which are coupled is calculated to 5.4 Å. Therefore, we proposed that a successful *Yamamoto* coupling can take place if 1) the molecular dynamics simulation shows that the aromatic carbon atoms to be coupled can come as close as 5.4 Å, and 2) the strain built up in the coupling step is below about 35 kcal/mol, a value that is calculated for the strain that is built up in the bicyclophane formation by a *Yamamoto* coupling that we described earlier and which gives nevertheless high yields of the respective products even though its undeniable strain (see SI).[336] However, these thresholds might be adjusted in the future.

Figure C.6 shows the distribution of C-C distances of the carbons to be coupled for a fluorene spoke 18 Ph-MSW. In order to simulate the actual bonding situation in the Ni-intermediate, the respective halogens were removed and substituted by "dummy" atoms to saturate the bonding situation. No matter which pathway (Figure C.5b) the ring closure follows, in any case the carbon atoms can come close enough to form the reactive intermediate dinuclear complex **E** (see SI for the proposed mechanism, intermediates, and transition state of the *Yamamoto* coupling reaction). In addition, none of the steps builds up an intermediate with a strain energy of more than 7 kcal/mol, far below our threshold of 35 kcal/mol. Therefore, the quantum chemical calculations propose that the wheel formation should be possible. The fluorene based 18 Ph-MSW **1a** was then synthesized as shown in Figure C.7.

Pyrylium salt **3** was condensed with the corresponding sodium fluoreneacetate **4a** to give the anchor shaped compound **5a**.[337–339] **4a** is not accessible from the alkylated diiodofluorene *via* statistical Pd catalyzed cross-coupling with a Zn enolate.[340] In that case only starting materials and diester could be isolated, an observation also made by others when dihalofluorenes were cross-coupled under Pd catalysis.[341,342] However, a statistical Br-Li exchange on the alkylated dibromofluorene using *n*-BuLi and subsequent quenching of the anion with trimethylsilylchloride (TMS-Cl), coupling of the remaining halogen with a Zn enolate, TMS/I exchange (with ICl) and hydrolysis of the ester made **4a** available. Selective *Sonogashira* coupling of **5a** with cyanopropyldimethylsilyl acetylene (CPDMS-acetylene),[343] *in situ* deprotection of **6a** and repeated selective acetylene coupling with **5a** at the iodo-position gave the dumbbell-shaped acetylene **7a**. The Co-catalyzed acetylene trimerization of **7a** generates the dodecabromide precursor **8a** for the subsequent *Yamamoto* coupling in acceptable 31% yield together with byproducts of lower molecular weight that were not characterized. The subsequent *Yamamoto* coupling gave **1a** (18 Ph-MSW) in 71 % yield after recycling gel permeation chromatography (recGPC). Notably, **1a** is a mixture of several atropisomers since the free rotation of the fluorene groups is at rt sterically hindered. Its successful synthesis was important for at least two reasons. First, despite the non-planarity of the final MSW structure, ring closing *Yamamoto* coupling worked well and confirms that fluorene spokes do not interfere with the proposed synthetic strategy, as proposed by the theoretical predictions. Second, although the periphery contains only *tert*-butyl groups and no long (branched) side chains, the solubility of all intermediates and **1a** is so good that their chromatographic purification can be performed and full characterization in solution was possible, proving that the incorporation of dialkylfluorene spokes guarantees the processability of 18 Ph-MSWs and their precursors.

We also tried to synthesize **1b** with longer side chains by an analog reaction sequence. However,

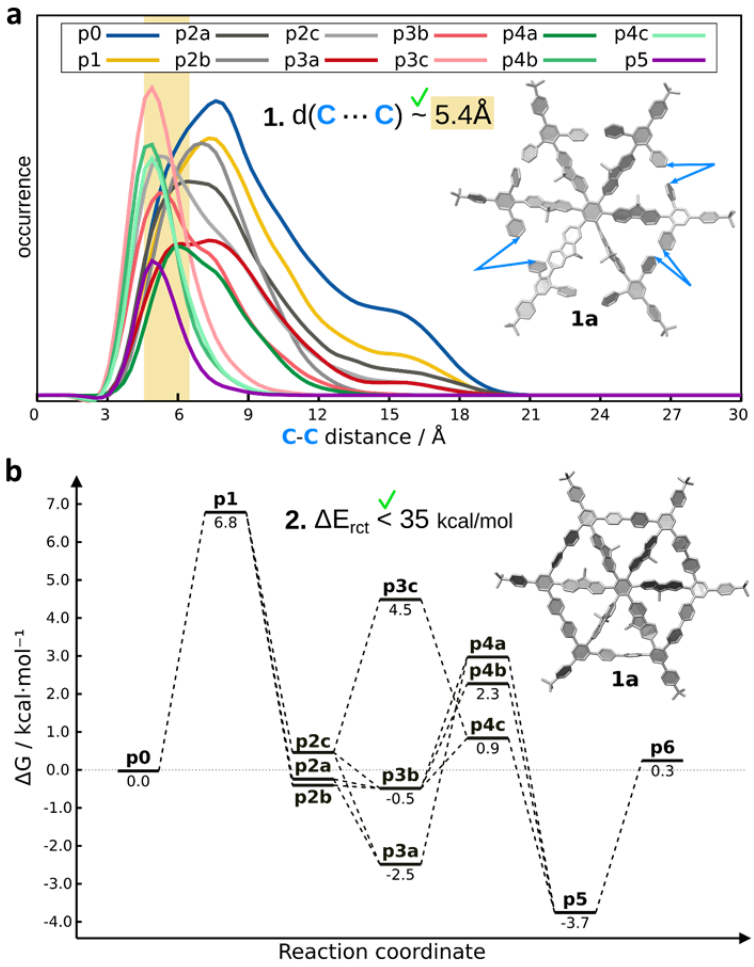

Figure C.6: a) Radial distribution functions (RDFs) of the C-C distances between neighboring brominated spokes of 18 Ph-MSW **1a** during all steps of the sixfold coupling, whereas the yellow area marks the distance range at which we propose a successful coupling; b) reaction free *Gibbs* energies at the r^{2}SCAN-3c(COSMO(THF)) theory level for each path respectively. The results for **2b** and **E3** are given in the SI.

unexpectedly the Co-catalyzed acetylene trimerization of **7b** to **8b** failed under our conditions and gave so tiny product amounts that only a mass spectrometric identification of **8b** within a mixed fraction was possible. The isolated material amounts did neither allow us to perform an NMR spectroscopic characterization nor to go on with the subsequent ring closing reaction. At present, we ascribe that to the higher steric demand of the longer alkyl chains that prevent the successful synthesis of **8b**.[344,345] One way to overcome this sterical crowding at the [2+2+2] cycloaddition reaction center is the introduction of an unsubstituted phenylene spacer between the acetylene and the fluorene units and keeping the rim unaltered (see Scheme S6 in the SI). However, our computational models predicted that the last ring closure in that structure (**p5** ⟶ **p6**) is highly improbable due to 1) the large C-C distance of around 18 Åof the respective coupling positions and 2) the enormous strain build-up of 142

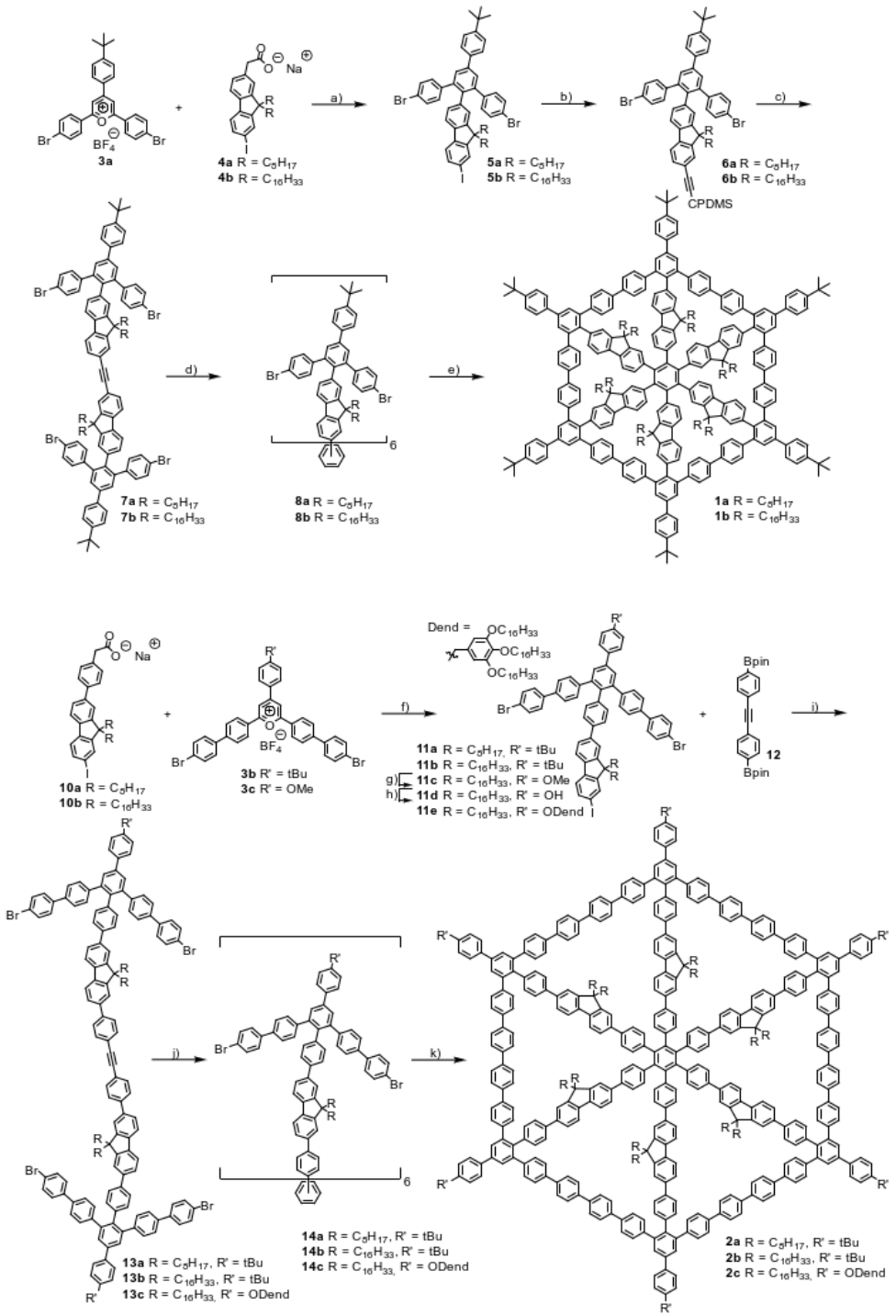

Figure C.7: Synthesis of **1a** and **1b**: a) Benzoic anhydride, 150 degC, 35% (**5a**), 8% (**5b**); b) CPDMS-acetylene, $Pd(PPh_3)_2Cl_2$, PPh_3, CuI, THF/piperidine, rt, 84% (**6a**), 23 % (**6b**); c) $Pd(PPh_3)_4$, CuI, Bu_4NF, THF/piperidine, rt, 45% (**7a**), 26% (**7b**); d) $Co_2(CO)_8$, toluene, 135 degC, 31% (**8a**), 0% (**8b**); e) $Ni(COD)_2$, 2,2´-bipy, COD, THF, 120 degC(MW), 71% (**1a**); synthesis of **2a-c**: f) benzoic anhydride, 150 degC, 26% (**11a**), 41% (**11b**), 56% (**11c**); g) BBr_3, DCM, -78 degC⟶rt, 93 %; h) 3,5-bis(hexadecyloxy)benzyl chloride, Cs_2CO_3, DMF, 100 degC, 75 %; i) **12**, $Pd(PPh_3)_4$, Cs_2CO_3, THF/water, 50 degC, 47% (**13a**), 32% (**13b**), 17 % (**13c**); j) $Co_2(CO)_8$, toluene, 135 degC, 64% (**14a**), 56% (**14b**), 15 % (**14c**); k) $Ni(COD)_2$, 2,2´-bipy, COD, THF, 120 degC(μW), traces (**2a**), 43% (**2b**), 60 %(**2c**).

kcal/mol (see Fig. S4 in the SI). Nevertheless, we took the chance to test this theoretical prediction also experimentally and performed the Yamamoto coupling of the respective wheel precursor (sse SI). Besides higher oligomers that were removed by recGPC, we could only detect, by mass spectrometry, a product in which the last C-C coupling doesn´t seem to occur but rather a protodebrominated product is formed, as proposed by our model. However, the successful Co-catalyzed cyclotrimerization (41% isolated yield) showed that introduction of an additional phenylene ring between the fluorene and the acetylene is a promising way to bring this reaction to work.

As shown in Figure C.4, the synthesis of the larger MSWs **2a**, **2b** and **2c** is based on the similar strategy as for the synthesis of **1a**. In order to generate an unstrained product for which our theoretical calculations predict a successful synthesis (see SI), each spoke segment and each side of the rim is enlarged by two phenylene units as compared to **1a**. However, an additional feature is the synthesis of the dumbbell molecule **13** by a *Suzuki* coupling of **11** and **12** which avoids the tedious separation of the *Glaser* side product always obtained in a *Sonogashira* reaction. The spatial separation of the alkylated fluorene unit from the cyclotrimerization center has an enormous effect on the formation of **14**. While **8a** is formed in only 31 % yield and **8b** could be obtained only in traces, **14a** and **14b** are obtained in 64% and 56% yield, respectively. However, the *Yamamoto* coupling of **14a** to the MSW **2a** offered some problems. Due to the very low solubility of **2a**, it could only be purified in traces and characterized by MALDI-TOF spectrometry. On the other hand, **2b** with longer alkyl chains was obtained in 43% yield and could be fully characterized in solution. This is the first time that we were able to synthesize and characterize 30 Ph-MSWs. The successful synthesis of **2a** and **2b** shows that 30 Ph-MSWs are accessible as long as aggregation during their synthesis is consequently suppressed. Using the same strategy allowed also the decoration of the 30 Ph-MSW by 3,4,5-tris(hexadecyloxy)benzyloxy groups as in the successful synthesis of **2c**. Here, the BBr_3-induced cleavage of the methoxy group worked well without interfering with the alkylated fluorene groups, as described in the introduction.

Scanning tunneling microscopic (STM) investigations of **1a**, **2a**, **2b**, and **2c** were performed at the liquid/solid interface of a solution of the respective compound in either 1-phenyloctane (PHO), 1,2,4-trichlorobenzene (TCB), or octanoic acid (OA) and highly oriented pyrolytic graphite (HOPG). However, no adsorbate structures could be observed for **1a** using any of these solvents. This is most likely due to the fact that **1a** exists at rt as a mixture of several non-interconvertible conformers. The alkyl chains at the non-freely rotating fluorene spokes are statistically distributed to both sides of the disc-like compound, thus preventing adsorption on the HOPG surface. Next, we investigated **2a** using PHO as solvent. At a concentration of $c = 3 \times 10^{-6}$ M in PHO, **2a** forms a two-dimensional (2D) supramolecular nanopattern to which a unit cell of $a = b = (5.1 \pm 0.2)$ nm^2 and $\gamma(a,b) = (59 \pm 2)$deg and one molecule per unit cell is indexed (Figure C.8). The oligo-arylene spokes and rim segments are oriented along the HOPG main axis directions (± 2deg). The unit cell vector a is oriented relative to the HOPG main axis direction d_1 with $\gamma(a,d_1) = (18 \pm 2)$deg. The backbones pack densely, and the octyl side chains are either oriented along the graphite surface in the intramolecular triangular nanopores or point towards the solution phase. Alike nanopatterns were observed for **2b**, however with lower resolution (see SI).

Under similar conditions (using PHO), no adsorbate patterns of **2c** were observed. However, when **2c** in TCB is applied to the HOPG, a hexagonal pattern with domain sizes of several 10×10 nm^2 with a unit cell (red in Fig. 4b, d, f) of $a = b = (13.3 \pm 0.2)$ nm^2 and $\gamma(a,b) = (62 \pm 2)$deg is observed. The unit cell is oriented relative to the HOPG main axis direction d_1 with $\gamma(a,d_1) = (21 \pm 2)$deg, and contains three molecules.

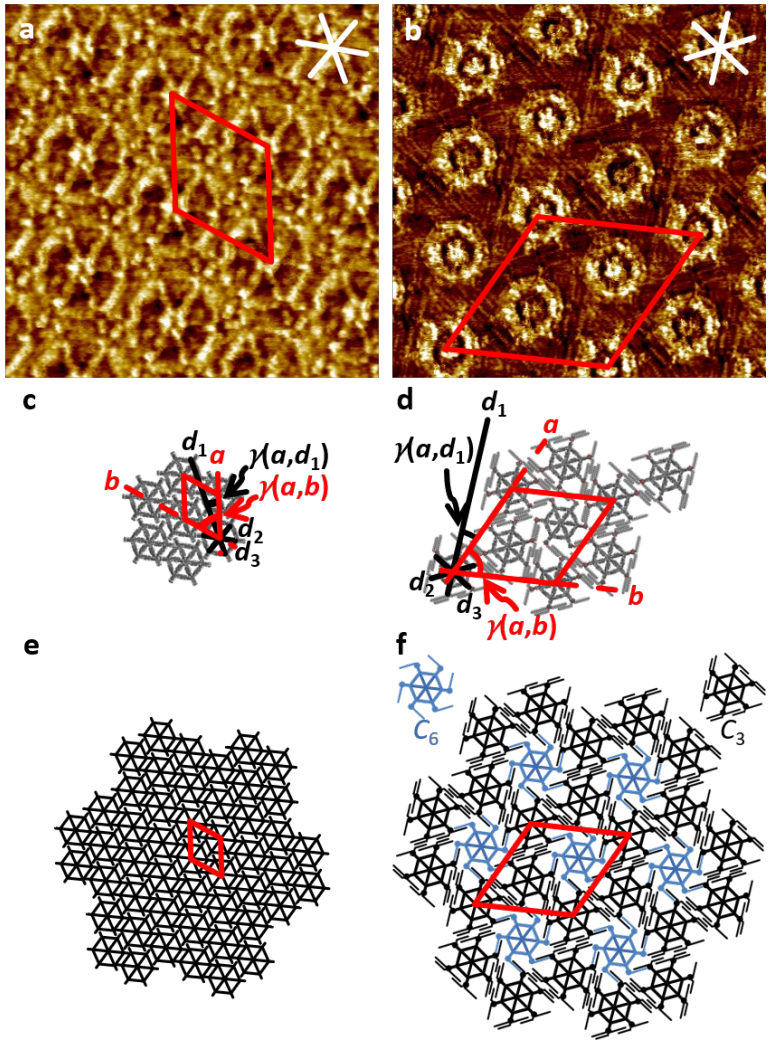

Figure C.8: (a),(b) Scanning tunneling microscopy images, (c),(d) proposed supramolecular models, and (e),(f) schematic models of supramolecular nanopatterns of **2a** ((a),(c),(e)) and **2c** ((e),(d),(f)) at the liquid/solid interface of a solution of the respective compound in (a),(c),(e) 1-phenyloctane or in (b),(d),(f) 1,2,4-trichlorobenzene. Image parameters: (a) $c = 3 \times 10^{-6}$ M, 17 × 17 nm^2, V_S = -2.0 V, I_t = 35 pA; (b) $c = 10^{-4}$ M, 30 × 30 nm^2, V_S = -0.67 V, I_t = 109 pA; to promote the two-dimensional mobility of the molecules on the surface, and thus self-assembly, both samples were thermally annealed for 20s at 80 degC prior to imaging as a routine procedure. Unit cell and additional packing parameters: (a),(c),(e) $a = b$ = (5.1 ± 0.2) nm, $\gamma(a,b)$ = (59 ± 2)deg, $\gamma(a,d_1)$ = (18 ± 2)deg; (d),(e),(f) $a = b$ = (13.3 ± 0.2) nm, $\gamma(a,b)$ = (62 ± 2)deg, $\gamma(a,d_1)$ = (21 ± 2)deg. The packing of **2c** consists of a C_3-symmetric (black) and a C_6-symmetric (blue) conformer (bold black and blue lines indicate backbones, and black and blue thin lines indicate adsorbed alkoxy side chains). Red lines indicate the unit cells. White and black asterisks in the images (a),(b) and supramolecular models (c),(d) indicate the HOPG main axis directions.

The backbones appear (such as the backbones of **2a** and **2b**) in bright color and the intraannular alkyl side chains remain invisible whereas the extraannular alkoxy side chains appear in medium bright color and seperate the backbones from each other. The nanopattern consists of a C_3- and C_6-symmetric conformer in a stoichiometry 2:1. The backbones of the C_3-symmetric conformer align along the HOPG main axis directions, while the backbones of the C_6-symmetric conformer are oriented along the armchair directions of the HOPG.
Regarding the alkoxy side chains of the C_3-symmetric conformer, 15 of the 18 hexadecyloxy side chains are adsorbed on the graphite surface and are oriented along the three HOPG main axes directions, while the three remaining side chains point towards the solution phase and remain invisible. For the C_6-symmetric conformer, six of the 18 hexadecyloxy side chains are adsorbed on the graphite surface and are oriented along the three HOPG main axis directions, while the remaining side chains point towards the solution phase and remain invisible. The 15 side chains of the C_3-symmetric MSW form the rims of a triangular object with small gaps at the corners into which the chains of the C_6-symmetric conformer interdigitate in a porous fashion. Unexpectedly, the nanopattern observed here has an identical distribution of symmetry elements as observed for an MSW in a previous study[238] but the coordination *via* alkyl chains is different as described above. The ability for compounds **2a**, **2b**, and **2c** to adsorb on the graphite and form stable 2D layers is clearly a result of the ability of the fluorene units to rotate rather freely so that in the STM experiment conformations can be populated in which the aromatic backbone comes in close contact with the graphite and all alkyl chains point after the adsorption to the supernatant solution. However, this makes the submolecular resolution of the molecules in the STM images all the more remarkable.

C.3 Conclusion

In summary, we described the synthesis of pure phenylene MSWs with a perimeter of 18 and 30 phenylene rings. The latter size is only accessible when the spokes carry alkyl substituents that prevent an aggregation and intermolecular C-C coupling during the final *Yamamoto* reaction. We tackled this problem by introducing alkylated fluorene units into the spokes. While the fluorene units of the final structures of the 30 Ph-MSWs can rotate freely, the 18 Ph-MSW exists at rt as a mixture of different not interconvertible conformers. However, since it was *a priori* not clear if the introduced curvature in the molecules affects the cyclization, the *Yamamoto* coupling was theoretically treated first. Based on quantum chemical calculations on the mechanism of the *Yamamoto* coupling we proposed a geometric criterion that allows to predict if the C-C coupling can for steric reasons be successful. An energetic criterion was extracted from our previously described bicyclophane synthesis that works with high yields. Experimental results of the MSW synthesis support our model. As far as we know, this is the first time that such criteria were applied for the macrocycle synthesis under *Yamamoto* coupling conditions and we expect that this theoretical prognosis can also be applied for other *Yamamoto* coupling reactions under constrained conditions. Interestingly, despite the two alkyl groups at the fluorene, we were able to prove the final structure of the MSWs also by STM-investigations at the solution/HOPG interface which suggests that freely rotating fluorene units adsorb with their 4- and 5-positions at the graphite and the alkyl groups point into the solution phase.

C.4 Supporting Information

The authors have cited additional references within the Supporting Information.[30-55] The data that support the findings of this study are available in the supplementary material of this article.

C.5 Acknowledgements

The authors gratefully acknowledge support from the Deutsche Forschungsgemeinschaft (DFG, RTG-2591 “Template-designed Organic Electronics-TIDE”, Project ID 418404117 and Project ID 455731873).

APPENDIX D

Quantum Chemistry Insight into the Multifaceted Structural Properties of Two-Dimensional Covalent Organic Frameworks

Julia Kohn,[†] Hong Li,[‡] Austin Evans[§], Jean-Luc Brédas,[‡] Stefan Grimme,[†]

Received: 29 November 2022
Published online: 20 March 2023

Reprinted in Appendix D (adapted) with permission[¶] from
J. T. Kohn, H. Li, A. M. Evans, J.-L. Bredas, and S. Grimme, *Quantum Chemistry Insight into the Multifaceted Structural Properties of Two-Dimen-sional Covalent Organic Frameworks*, Chem. Mater. **35**.7 (2023) 2820, DOI: 10.1021/acs.chemmater.2c03555

Own contributions

- initial idea
- performing of all calculations
- interpretation of the results
- writing of the manuscript

[†]Mulliken Center for Theoretical Chemistry, University of Bonn, Bonn, 53115, Germany
[‡]Department of Chemistry and Biochemistry, The University of Arizona, Tucson, Arizona, 85721-0088, USA
[§]Department of Chemistry, University of Florida, Gainesville, Florida, 32603, USA
[¶]Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

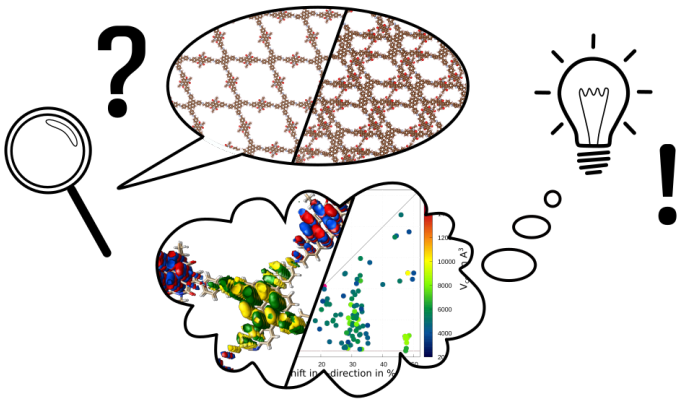

Figure D.1: Table of content graphic (ToC).

Abstract 2D covalent organic frameworks (COFs) possess a wide range of desirable mechanical and optoelectronic properties, rendering them a versatile compound class that is widely investigated in many areas of chemistry. However, experimental structure determination remains challenging as mostly inseparable isomers are measured while computational models mostly account for a very limited number of model structures, making the establishment of comprehensive structure–property relationships difficult. Thus, here, we go computationally beyond the conventional but insufficient description of eclipsed stacked COF building blocks. We emphasize the presence of a variety of stacking and structural motifs and investigate their impact on the electronic properties. Given the very large size of the investigated systems, we employ the fast and accurate tight-binding based semiempirical quantum mechanical (SQM) GFN-xTB method. Taking an imine-linked COF as a representative system, we discuss our structural, energetic, electronic, and spectroscopic results calculated at the periodic GFN1-xTB quantum chemistry level in comparison to experimental findings.

D.1 Introduction

2D covalent organic frameworks (COFs) are highly innovative materials for cutting-edge applications and have been intensively investigated since 2005.[346–350] Among the various recent developments, systematically engineering the crystal structures of 2D COFs for targeted applications has attracted great interest.[244] In particular, the directional bonding in the 2D COFs can lead to tessellated monolayers defined by their monomer subunits which then stack to form bulk crystalline structures.[351] Leveraging this design principle, significant effort has been made to enhance the mechanical,[352,353] optical,[246,354,355] electronic,[247,356,357] thermal,[358,359] and molecular transport properties[360–362] of 2D COFs. While conventional 2D COF design focuses on the in-plane covalent bonding of molecular sheets, these properties are also interconnected to the multi-layer arrangement of 2D COF layers. In addition, practical 2D COF solution syntheses almost exclusively lead to the formation of multi-layer COF materials.[248] This realization mandates a deeper consideration of multi-layer stacking in the 2D COFs and its impact on the emergent properties of this materials class.

The vast majority of experimental 2D COF reports assume that multi-layer sheets stack in an eclipsed fashion. This is in-part due to the experimental challenges with simultaneously determining the in-plane and cross-plane structures of polycrystalline materials with limited (sub-100 nm) crystallographic dimensions. Recently, advances in materials quality have led several experimentalists to note that

non-eclipsed stacking motifs better represent experimental 2D COF structural data.[249,250,363] These findings have further motivated synthetic designs aimed at controlling interlayer arrangements of multi-layer sheets,[245,364,365] which some have preliminarily demonstrated can be used as a design parameter to control physical properties. However, a systematic investigation into the structures of multi-layer stacking and the impact this has on emergent materials properties has yet to be performed. Here, we computationally reevaluate the interlayer structural assignment of a model 2D COF system that we have previously investigated both experimentally and computationally and that is known for its versatility, good electrical conductivity and overall decent performance. The model imine-linked COF, TAPPy-NDI COF, is constructed by condensing 1,3,6,8-tetrakis(4-aminophenyl)pyrene (TAPPy) and 3,7-di(4-formylphenyl)naphthalenediimide) (NDI).[252,258,366] We have previously demonstrated experimentally that this 2D COF has moderate electronic conductivities when doped.[259] Later, we showed that the *cis-/trans-* isomerism of the imine bonds that bind monomer subunits together has an impact on the electronic structure of these materials.[253] However, in both of these reports, we made the assumption that the 2D COF networks stack in an eclipsed fashion.
In this work, our aim is to provide a much more realistic description of the structural features of multi-layer 2D COFs by simulating serrated stacking and its influence on various electronic and spectroscopic properties. We apply the fast and accurate semiempirical quantum mechanical (SQM) GFN1-xTB method, with emphasis on the systematic evaluation of stacking patterns.

D.2 Methods

All calculations were conducted under periodic boundary conditions at the GFN1-xTB[142,212] level of theory as implemented in `DFTB+ V21.2`[251] . The atomistic simulation environment ASE[367] was used for structure modifications. Structure visualization was performed with `Avogadro`[368] and `VESTA`[369,370] and figure generation with `gnuplot`[371] and `inkscape`[372] . Powder diffraction patterns (PXRD) were simulated using `VESTA`.
The reference and starting structure of a monolayer was taken from Ref.[253]. Displaced structures for a bilayer were generated by shifting the adjacent layer in steps of 5% between 0% and 50% in the direction of the crystallographic *a* and *b* axes, resulting in a total of 121 new bilayer unit cells (see Figure E.2).
For all 121 displaced structures, geometry optimizations including lattice parameter optimization and full relaxation of atomic positions, were conducted employing supercell folding with a k-point mesh of 1 1 9 and an energy convergence criterion of 10^{-5} Hartree (E_h). All other program parameters were kept at default values. Band structures and band gaps were evaluated for all optimized geometries using the same settings. Band structures were calculated along the $\Gamma - Y - N - X - \Gamma - Z - F - E - L$ path with a resolution of 30 k-points for $\Gamma - Z$ and 10 k-points for any other section. For PXRD simulations, the $Cu_{K\alpha}$ radiation with a wavelength of 1.54 Å was used. Relative energies are calculated according to $E_i - E_{ref}$ whereas E_{ref} corresponds to the total energy of the optimized reference structure. IR spectra were calculated using `DFTB+ V22.1` in the development version from August 2022[373] and a custom modified version of `xtb V6.5.1`[374]. Experimental PXRD and IR spectra are taken from Ref. [259].

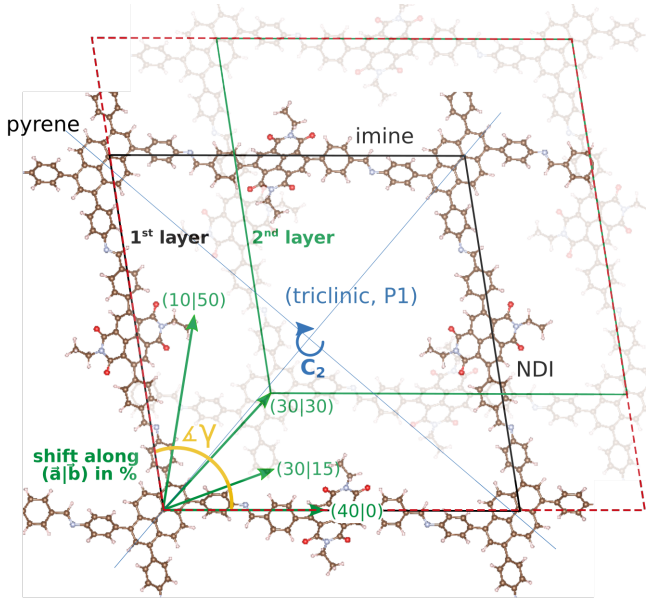

Figure D.2: Structure and unit cell of a TAPPy-NDI monolayer (black). Serrated stacking was obtained by shifting a second identical layer in the direction of the crystallographic a and b axes (green). The unit cell obtained in this way (red) contains 372 atoms in TAPPY-NDI bilayers.

D.3 Results and Discussion

D.3.1 Structural Definitions

The TAPPy-NDI COF is a rhombic 2D network consisting of four-arm TAPPy cores and two-arm NDI linkers. Previous electronic-structure studies[253,259] have shown that its top valence band is associated with contributions from TAPPy, while the bottom conduction band is associated with contributions from NDI. Previous work by Xun *et al.*[253] evaluated the impact of imine bond orientations on the formation of different isomeric structures and their electronic structures. Here, we greatly expand the range of structural configurations by considering continuous slipping of neighboring layers.
The number of unequivalent double-layer unit cell structures generated by shifting the "top" layer with respect to the "bottom" layer can be largely reduced by considering the arbitrariness in the assignment of the "top" and "bottom" layers (i), translational symmetry (ii) and equivalence of the a and b axes (iii) (Figure E.3):

i. $(a \mid b) = (-a \mid -b)$

ii. $(a \mid b) = (100-a \mid b)$

iii. $(a \mid b) = (b \mid a)$

The resulting irreducible and experimentally meaningful shift area is now reduced to 1/8 of the full area. In the following, only this reduced area is used for graphical representation.
The optimized structures were categorized according to the pore size and shape of the layers. Layers can be planar and flat or undulated and intercalated. This is shown schematically in Figure E.4. The reference lattice parameters are given in Table F.3.

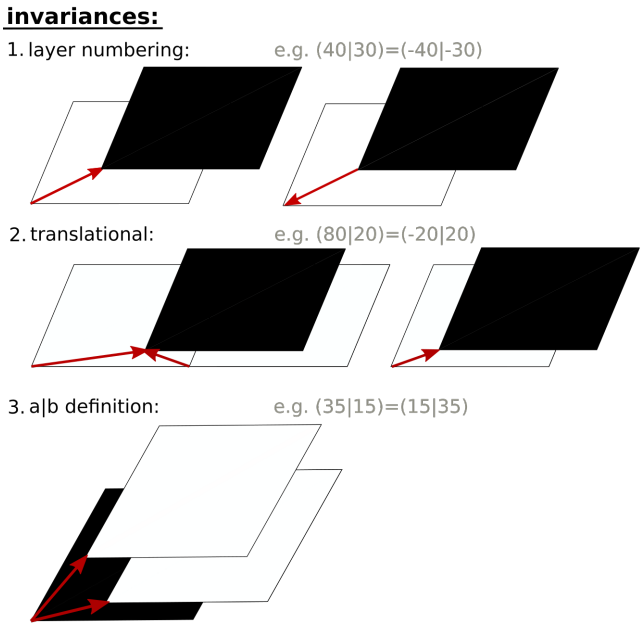

Figure D.3: Illustration of the reduction in the symmetry of shift positions.

angle	in °	length	in Å
α	72.51	a	34.60
β	85.01	b	34.49
γ	98.56	c	8.21

Table D.1: Optimized reference lattice parameters in Å and °.

Undulation occurs when at least a or b is shorter than approximately 33 Å, although this threshold value can increase in the case of very small pores. Layers stay planar when a and b are both longer than 33 Å. Furthermore, the systems may have large pores ($\mathbf{L_P}$), small pores ($\mathbf{S_P}$), or no pores ($\mathbf{N_P}$). The absence of pores occurs when cell volumes are below 4000 Å^3 ; small pores are present for cell volumes below 5000 Å^3 and large pores above this value. Examples for all occurring combinations are given in Figure E.5.

Permanently porous systems are of higher interest for materials chemistry and non-porous systems are rarely investigated. The porosity is usually obtained by a procedure using supercritical CO_2, which has no surface tension and accordingly prevents the pores from collapsing during its evaporation. However, annealing of the porous substrate leads to a complete loss of porosity, as confirmed by N_2-adsorption-desorption experiments. These two steps are reversible.[244,258]

D.3.2 Relative Energies of the Stacking Patterns

Relative energies of the generated COF ensemble are presented in Figure E.6, sorted by structural motif. The average relative energy decreases from class $\mathbf{L_P P}$ over $\mathbf{S_P U}$ to $\mathbf{N_P U}$. There is a linear correlation between relative energy and cell volume, with the smaller V_{cell}, the lower E_{rel}. This

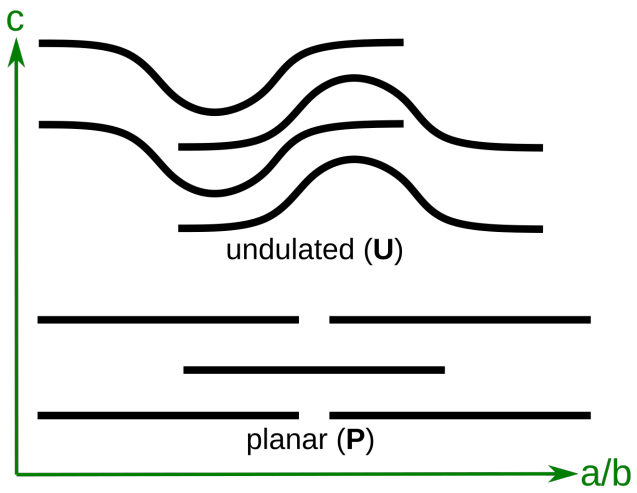

Figure D.4: Schematic representation of undulated (**U**) and planar (**P**) structural motifs.

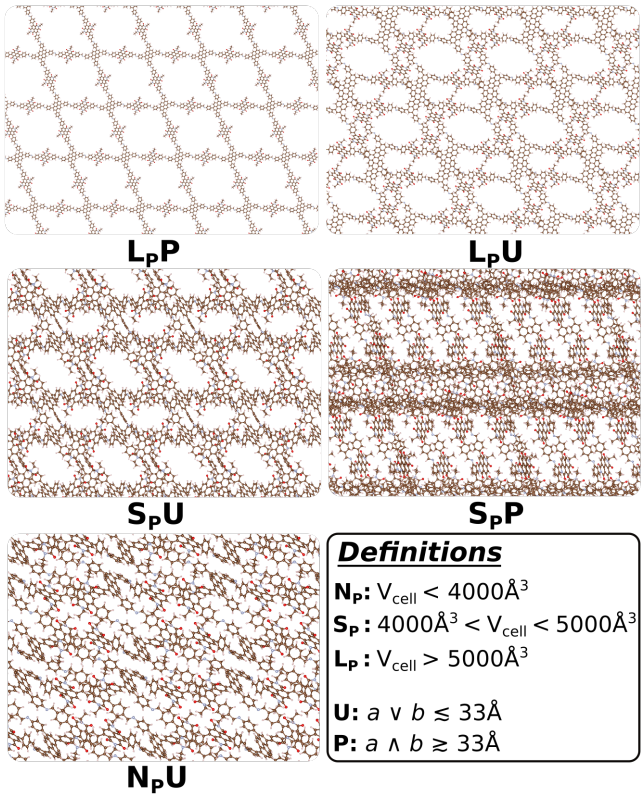

Figure D.5: Representative examples of various structural motifs and their classification, as viewed along c.

energy lowering can mainly be attributed to stronger London dispersion interactions (see ESI). While class $\mathbf{N_PU}$ is thermodynamically most stable, class $\mathbf{L_PP}$ that has the least favorable energy is also experimentally observed, which implies that it must be kinetically stabilized. This assumption is fully supported by the experimental findings mentioned above[259] . There are three distinct areas in which class $\mathbf{L_PP}$ occurs. Small symmetrical shifts up to 12%, a big corridor at 30% shift in one direction, and one high-energy corridor at 50% shift in one direction, see Figure E.6.

Among the structural parameters other than V_{cell}, only angle γ between the a and b axes showed a

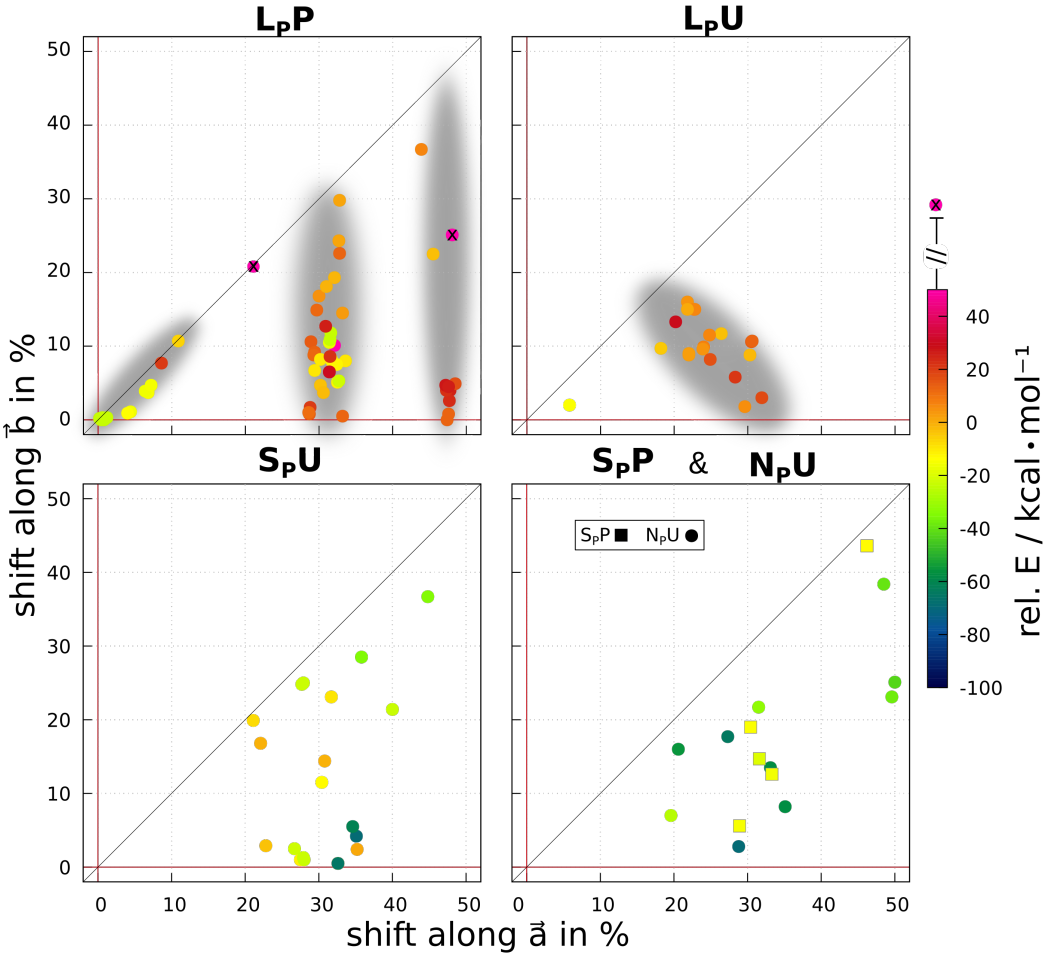

Figure D.6: Relative energies in kcal · mol^{-1} (color coded) with respect to the interlayer shift, sorted by structural motifs. Outliers are marked with crosses while grey shades represent the stability areas of different motifs.

correlation with the shift. This is depicted in Figure E.7. The $\mathbf{N_PU}$ class shows very high γ values resulting in a strongly distorted, stretched parallelogram where intralayer spaces are optimally filled. In contrast, the $\mathbf{S_PP}$ class, which is also small-pored, has γ angles close to 90°. Here, high filling is obtained by what can be referred to as internal staggering, which means staggering not exclusively in the c-direction but also within a plane, e.g., when NDI moieties are strongly twisted versus the COF plane. In general, the more undulated a structure, the higher its γ angle.

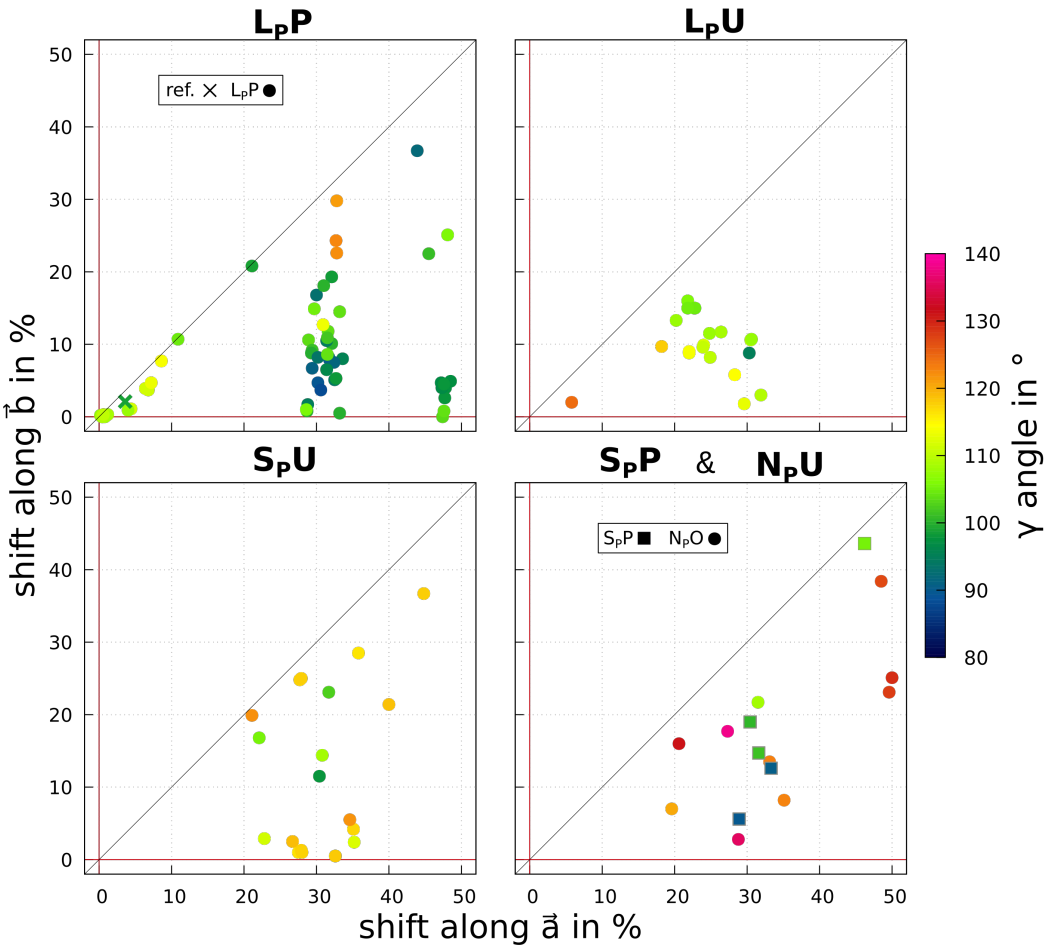

Figure D.7: Evolution of the gamma angle in ° (color coded) as a function of the interlayer shift, sorted by structural motifs. The cross in the upper left panel represents the reference structure.

D.3.3 XRD Analysis

Precise structural determination of multi-layer COF stacking via PXRD is a daunting experimental challenge due to limited understanding of the nature of adsorbates, the ensemble crystallites that are probed, the impact of defects, and the flexible nature of porous organic materials. However, the large single crystals required for single-crystal structural determination, which would overcome these limitations, are not yet widely available. This limitation has led to the ambiguous assignment of stacking motifs in many experimental studies. Here, our simulated PXRD patterns can clearly separate all five assigned structural motifs (see *ESI*). They also allow us to exclude the (50|xx) area of $\mathbf{L_P P}$ from the experimentally contributing ensemble. The best matching pattern is shown in Figure D.8. The occurrence of specific reflections may vary by 2θ ranges of 4°, meaning that a perfect assignment of experimental reflection indices is not feasible. There is no correlation between peak position and shift position but a complex influence of the α, β, and γ angles. It should be pointed out again that the

generation of cumulated PXRD patterns of whole ensembles is very demanding as the weighting of contributing residues is unknown. Vice versa, the extraction of several pertinent structures from the cumulated experimental PXRD pattern is not feasible as well.

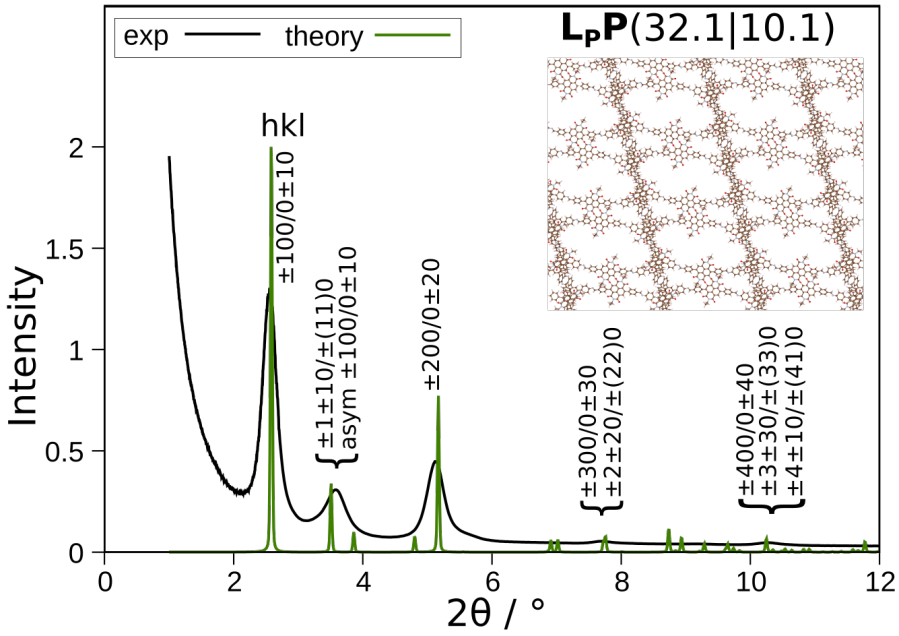

Figure D.8: Comparison of the experimental PXRD pattern from Ref. [259] and the best matching simulated pattern.

D.3.4 Electronic Properties

Producing electronically active synthetic organic layers would merge the advantageous features of polymeric and crystalline (semi-)conductors. However, a reliable understanding of the electronic structure of 2D COFs requires an understanding of their multi-layer arrangement. In the case of TAPPy-NDI, as mentioned above, the NDI units have a dominant contribution to the bottom conduction bands, while TAPPy is the main contributor to the top valence band. In the description of the electronic properties and band gaps, the TAPPy stacking motifs appear to play a minor role compared to the NDI stacking motifs, although in the columnar and woven structures defined in Figure D.9 below, TAPPy stacking leads to some dispersion of the valence band. Thus, this section focuses on the NDI motifs. As the relative positions of the NDI units within the network are not well described by the a/b shift or the structural motifs, a new classification has to be introduced in relation to the NDI structural motifs, as illustrated in Figure D.9. These motifs were determined manually since meaningful and automatically evaluable crystallographic or geometric criteria could not be found. One potentially useful parameter for future investigations could be the "twist" of the NDI units, which informs how NDI moieties are rotated versus the COF planes, with a simultaneous evaluation of TAPPy stacking and twist motifs.
The direct band gap correlates well with the NDI structural motifs, as shown in Figure D.10. The

band structures are depicted in Figure D.11; their overall differences will be difficult to discern experimentally. We note that the generalized gradient approximation (GGA) based semiempirical quantum mechanical (SQM) methods used here strongly underestimate the electronic gaps compared to the experimental value of 1.6 eV as well as the band dispersions, due to a missing treatment of exchange. The only path where significant band dispersion occurs is $\Gamma \longrightarrow Z$, which represents the interlayer direction. The rather flat bands support the assumption of hopping-type electron transport. The degeneracy of the electronic bands is explained by the chemical nature of the TAPPy-NDI COF. Each unit cell comprises two equivalent pyrene units, on which the highest occupied crystal orbitals (HOCO) are localized, and four equivalent NDI units, on which the lowest unoccupied crystal orbitals (LUCO) are located. All these findings are in excellent agreement with previous theoretical studies[253] at the DFT level and experimental results[259].

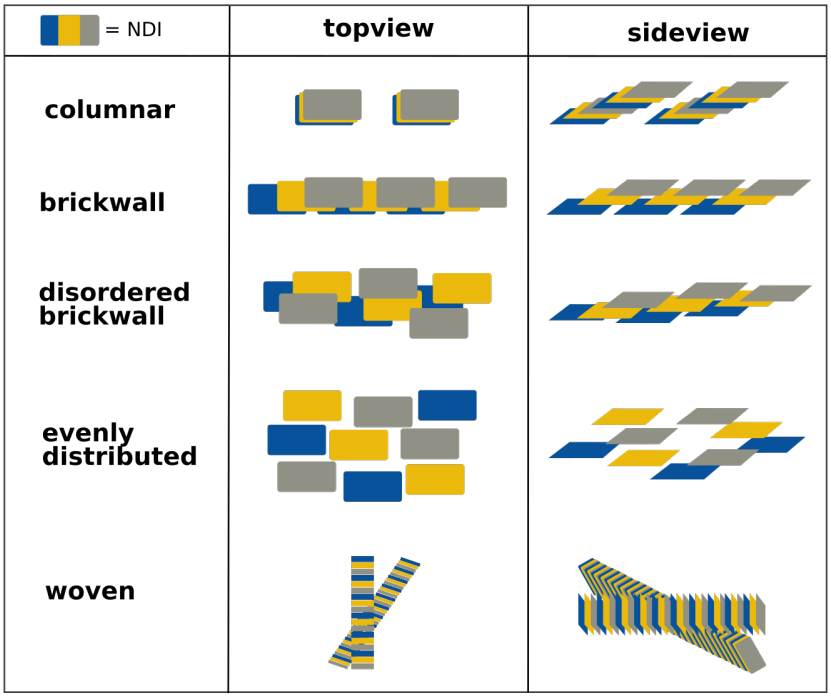

Figure D.9: Schematic representation of the NDI structural motifs.

D.3.5 IR Spectroscopy

Infrared spectroscopy is a critically important tool to evaluate the chemical structure of 2D COFs. Here, the IR spectra were calculated for selected structures and are compared to experimental data in Figure D.12.

The ν(C=N$_{NDI}$), ν(C=N$_{imine}$) and ν(C=O) vibrations can reliably be assigned. C=C vibrations of all kinds are distributed over a 300 cm^{-1} range of frequencies. The overall good quality of the spectra even allows us to exclude the strictly eclipsed $\mathbf{L_P P}$ conformations from the experimentally relevant structures (see the marked offset ν(C=O)). The best matching single spectrum in fact belongs to a $\mathbf{L_P P}$

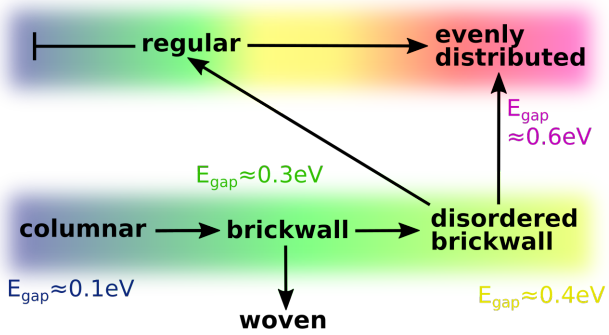

Figure D.10: Qualitative evolution of the band gap as a function of the NDI structural motifs. A motif is referred to as "mixed" if it cannot be clearly assigned to one of the other structural motifs and instead possesses mixed features.

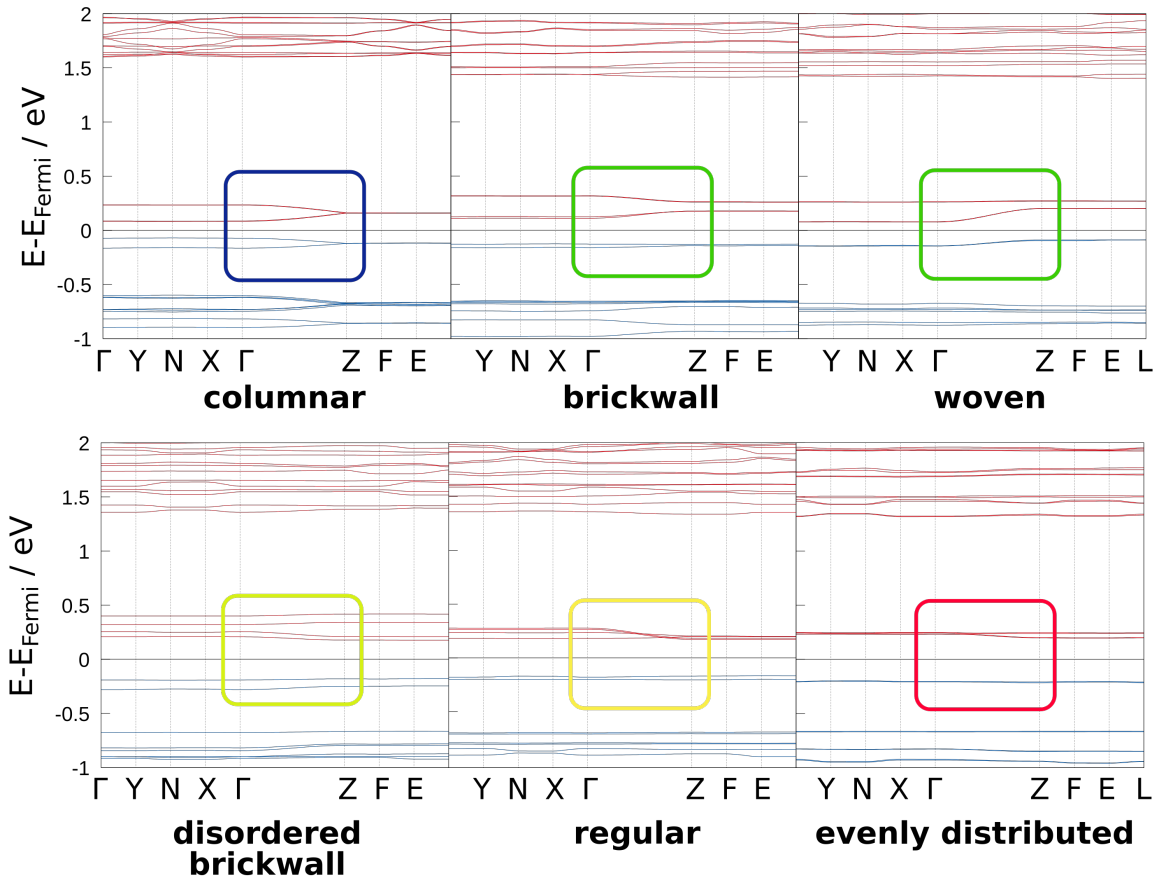

Figure D.11: Band structures for different NDI structural motifs; the boxes denote the two HOCOs and four LUCOs and are color coded according to the band gaps in Figure D.10: The valence bands are in blue and conduction bands, in red. The Fermi level is set to the zeroth of energy.

structure but with a shift close to (30|0), consolidating the findings from the XRD analysis. As 2D

COFs are extremely large systems, supercell calculations to converge long-ranged lattice vibrations were not computationally feasible. As a result, frequencies in the fingerprint region in the near IR up to about 1000 cm^{-1} are not well reproduced *in silico*. Turning to a hybrid functional would be useful to predict the dipole intensities needed for an even more accurate coincidence of calculated and experimental IR spectra, but this is a very costly proposition as well.

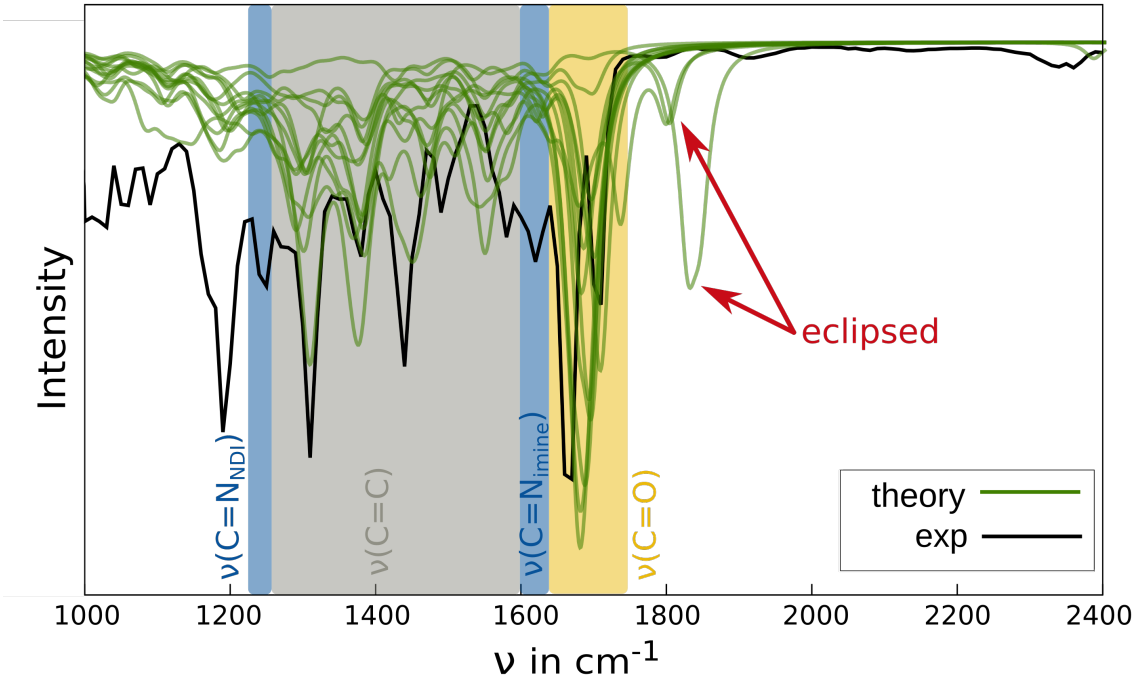

Figure D.12: Computed IR spectra of various structural motifs compared to experimental data from Ref. [259].

D.4 Conclusion & Outlook

In this work, a series of displaced stacking motifs were generated for a representative iminie-based COF, the TAPPy-NDI COF, and evaluated using periodic GFN1-xTB quantum-chemical calculations. All results obtained here with the SQM methodology are in good agreement with previous DFT results for comparable, albeit much smaller systems. GFN1-xTB proved to be numerically very robust with reasonable (short) computer timings.
Importantly, with our results we are able to differentiate the kinetically stabilized porous planar systems, that emerge after specific synthetic procedures, from the thermodynamically stable non-porous undulated systems, that are obtained after annealing. Furthermore, simulated PXRD and IR data indicate that the eclipsed $\mathbf{L_P P}$ (i.e., consisting of large-pored planar structures) class presumably plays a minor role in the experimental mixture, which is more likely composed of serrated $\mathbf{L_P P}$ stacking motifs with a shift of about 30% in one of the in-plane directions and 0-10% in the other direction. In addition to small symmetrical deviations from the eclipsed pattern and the stable corridor at 30% shift, a potential third stable area for experimentally desirable $\mathbf{L_P P}$ systems was discovered to correspond to a high-energy corridor at 50% shift.

It turns out that the interlayer shift is not the all-determining parameter and we find a complex interplay between different structural characteristics and the energetic, electrical, and spectroscopic properties; these interconnections are summarized in Figure D.13. We expect that additional parameters will be needed in further studies to better categorize the structural properties such as the NDI arrangements.
2D COFs pose significant challenges in their computational modeling owing to their vast structural

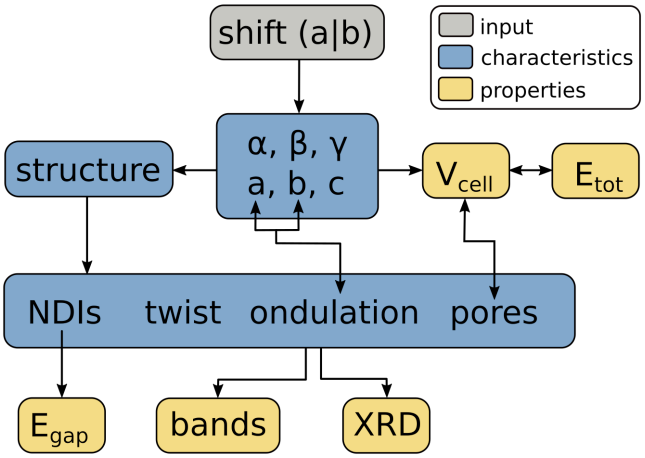

Figure D.13: Cause and effect of different parameters (blue) onto structure and properties (yellow).

and conformational diversity, large unit cells, strong non-covalent interactions, complex stacking patterns, and incomplete experimental reference values. The multi-structural approach we applied here has proven to be closer to reality than single-structure models commonly used to date as it takes into account the inherent structural diversity arising from the isomeric flexibility of the COFs. The determination of the exact composition is challenging due to the extremely large number of degrees of freedom and the unknown contribution of different residues to the ensemble.
To summarize, we have demonstrated the reliable application of periodic GFN1-xTB for a variety of computational investigations of a typical 2D COF providing an overall good agreement with experimental findings. Furthermore, our results point to a need to depart from the conventional view that 2D COFs are largely uniform in structure and essentially consist of eclipsed stacked units. Thus, the use of an ensemble approach is strongly encouraged with more attention paid to intermediately displaced stacking motifs.

D.5 Acknowledgements

The authors acknowledge funding by the DFG (RTG-2591 "TIDE – *Template-designed Organic Electronics*") and the College of Science of The University of Arizona.

D.6 Supporting Information

The electronic supporting information (*ESI*) contains the optimized structures in .cif format, the raw data for band structure visualization in .dat format, the actual visualization of the band structures in .eps format, the raw data for the visualization of PXRD patterns in .dat format, the raw data for the visualization of IR spectra in .dat format, tabulated absolute and relative energies in .ods format, and supporting information and pictures in .pdf format.

APPENDIX E

Efficient Calculation of Electronic Coupling Integrals with the Dimer Projection Method via a Density Matrix Tight-Binding Potential

Julia Kohn,[†] Nora Gildemeister,[‡] Stefan Grimme[†], Daniele Fazzi,[§] Andreas Hansen,[†]

Received: 13 July 2023
Published online: 11 October 2023

Reprinted in Appendix E (adapted) with permission[¶] from
J. T. Kohn, N. Gildemeister, S. Grimme, D. Fazzi, and A. Hansen, *Efficient calculation of electronic coupling integrals with the dimer projection method via a density matrix tight-binding potential*, J. Chem. Phys. **159**.14 (2023), DOI: 10.1063/5.0167484

Own contributions

- coding and implementation
- performing of all calculations except for Gaussian
- interpretation of the results
- writing of the manuscript

[†]Mulliken Center for Theoretical Chemistry, University of Bonn, Bonn, 53115, Germany
[‡]Department of Chemistry, Greinstrasse 4-6, 50939 Köln, Germany.
[§]Dipartimento di Chimica "Giacomo Ciamician", Via Selmi 2, 40126 Bologna, Italy.
[¶]Permission requests to reuse material from this chapter should be directed to AIP Publishing.

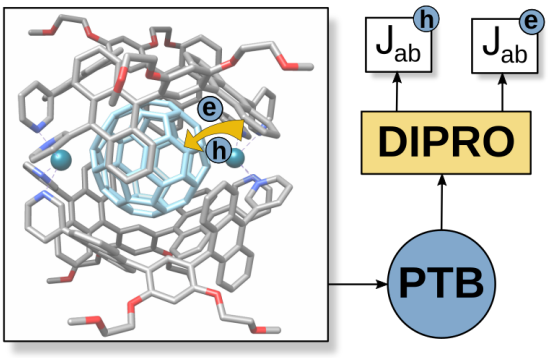

Figure E.1: Table of content graphic (ToC).

Abstract Designing organic semiconductors for practical applications in organic solar cells (OSCs), organic field-effect transistors (OFETs), and organic light-emitting diodes (OLEDs) requires understanding charge transfer mechanisms across different length and time scales. The underlying electron transfer (ET) mechanisms can be efficiently explored using semiempirical quantum mechanical methods (SQM). The dimer projection method (DIPRO) combined with the recently introduced non-selfconsistent density matrix tight-binding potential (PTB [S. Grimme, M. Müller, A. Hansen, *J. Chem. Phys.,* 158, (2023).]) is used in this study to evaluate charge transfer integrals important for understanding charge transport mechanisms. PTB, parameterized for the entire periodic table up to Z=86, incorporates approximate non-local exchange, allowing efficient and accurate calculations for large hetero-organic compounds. Benchmarking against established databases such as Blumberger's HAB sets, or our newly introduced JAB69 set, and comparing with high-level reference data from ωB97X-D4 calculations confirms that DIPRO@PTB consistently performs well among the tested SQM approaches for calculating coupling integrals. DIPRO@PTB yields reasonably accurate results at low computational cost, making it suitable for screening applications and applications to large systems such as metal-organic frameworks (MOFs) and cyanine-based molecular aggregates further discussed in this work.

E.1 Introduction

To develop new functional semiconducting materials for opto-electronic applications, a comprehensive understanding of the electron transfer mechanisms is indispensable. Electron transfer processes are subject to extensive computational studies since the early 1970s and are still a strongly evolving topic.[375] Today, semiempirical methods (SQM) enable the investigation of reasonably large systems to test their suitability for high-performance conductive materials.[376] This is especially important for the design of molecular and polymeric organic semiconductors, with increasing application in organic solar cells (OSCs),[377] organic field effect transistors (OFETs),[378] organic light-emitting diodes (OLEDs),[379] and other organic electronics (OEs).[260–262] In contrast to metallic conductors, that exhibit band transport, organic semiconductors mostly exhibit different variations of hopping transport.[192,380] Researchers have developed various methods to measure and model electron and hole mobilities in recent years[381–383] . The most common approaches utilize Marcus' theory and subsequently employ Monte Carlo simulations to evaluate the charge mobility[189,263,384,385] . In this

approach, the first step involves calculating electron transfer integrals, also known as coupling integrals (H_{ab} or J_{ab}), between different fragments or localized states. Fragmentation of the investigated chemical space can be carried out using a wide range of theories, most of which can be combined with any available orbital localization method[264] . Several methods for calculating coupling integrals exist, including charge constrained density functional theory (CDFT)[265] , fragment orbital DFT (FODFT) or fragment orbital density functional tight binding (FODFTB)[266] , projection-operator diabatization (POD)[267,386,387] , frozen density embedding (FDE)[388] , generalized Mulliken-Hush (GMH)[268] , multistate DFT (MSDFT)[389] , analytic overlap method (AOM)[269] , machine-learning approaches (ML)[390,391] , and the dimer projection method (DIPRO)[270] .
Compounds for OE are often many hundreds of atoms large, have complicated intra- and intermolecular structure, or exhibit a wide conformational variety making the exploration of all three-dimensional coupling possibilities, charge carrier pathways, and their corresponding integrals time-consuming. Additionally, for charge mobility calculations in solids, either crystalline or amorphous, thermal fluctuations have to be accounted for. To achieve that, extensive molecular dynamic simulations (MDs) are required and coupling integrals are evaluated several thousands to tens of thousands of times along a single trajectory. By using SQM, computation times can be significantly reduced while still reasonably accurate results can be obtained, enabling large length- and time-scale investigations such as those involving DNA[277,392] , complex fullerene-based acceptors for organic photovoltaics (OPV) applications[393] , materials design and screening for new OE compounds[271] , impact of thermal disorder effects on charge mobility[394] , charge transport in covalent organic frameworks[272] , pietzo effect[395] , and polymer crystals[273] to become feasible. POD methods[386] can be considered as valuable and effective alternatives to DIPRO, however such methods can be affected by delocalization of the diabatic states over the fragments, thus leading to incomplete diabats. The latter problem was recently mitigated by Oberhofer et al.,[267] showing that modified-POD and DIPRO methods indeed lead to very similar results, even though the mutual polarization effects between fragments are treated differently. In this paper we decided to focus our development and benchmarking on the DIPRO approach, given its consolidated use in literature. The advantages of DIPRO compared to other coupling integral methods is that it is a post-processing method and can in principle be used with any existing code and mean-field theory level. Furthermore, the mathematics behind DIPRO are easily understandable, accessible, and implementable.
In the following, we use the recently introduced semiempirical non-selfconsistent tight-binding potential PTB[160] together with DIPRO to calculate coupling integrals for a wide range of hetero-organic compounds. We transfer our approach to the investigation of challenging systems such as merocyanines, whose resonant electronic structure remains challenging due to their strong intramolecular charge transfer, and large metal-organic frameworks (MOFs), that due to their extended and highly delocalized structure are computationally demanding. We benchmark the results against ωB97X-D4/TZ2P reference DFT data and compare them to other SQM methods using common benchmarks such as Blumberger's HAB7[396] , HAB11[397] , and HAB79[276] sets.

E.2 Theory

The dimer projection method, known as DIPRO[270] and originally introduced by Refs. [274, 275], enables the calculation of the coupling integrals between pairs of molecules (each molecule usually

defined as fragment). Such method is often reported in literature under the name of fragment orbital DFT (FO-DFT),[397] and different "flavors" have been developed over the years. Each "flavor" is mainly related to the choice of the molecular orbitals belonging to each fragment, specifically whether they reflect the neutral or the charged electronic state of the molecule. For a critical review concerning FO-DFT, DIPRO and other diabatization procedures (e.g., POD[387]) we refer the reader to Refs. [260] and [397]. DIPRO necessitates three quantum mechanical single-point calculations from any source (e.g., SQM, HF, DFT): one for each monomer (A,B) and another for the dimer (AB), whereas the optimized monomer structures are used in the unrelaxed dimer geometry. The relevant equations are provided below:

$$\gamma_1^i = \mathbf{C_A}^i \cdot \mathbf{S_{AB}} \cdot \mathbf{C_{AB}} \tag{E.1}$$

$$\gamma_2^j = \mathbf{C_B}^j \cdot \mathbf{S_{AB}} \cdot \mathbf{C_{AB}} \tag{E.2}$$

$$S_{ab}^{ij} = \gamma_1^i \cdot \gamma_2^j \tag{E.3}$$

$$J_{ab}^{ij} = \gamma_1^i \cdot \mathbf{E_{AB}} \cdot \gamma_2^j \tag{E.4}$$

$$J_{ab,eff}^{ij} = \left| \frac{J_{ab}^{ij} - 0.5 \cdot (E_A^i + E_B^j) \cdot S_{ab}^{ij}}{1 - (S_{ab}^{ij})^2} \right|. \tag{E.5}$$

Here, **C** are the orbital coefficients, **S** is the AO overlap matrix, **E** are the orbital energies, i and j denote the molecular orbitals that are considered for the electron or hole transfer. In this work, i and j always correspond to the HOMOs of the coupled molecules.

The DIPRO approach is limited, as it is not valid for very large overlaps (i.e., short intermolecular distances) or for $J_{ab}^{ij} \lessapprox 0.5 \cdot (E_A^i + E_B^j) \cdot S_{ab}^{ij}$ in the case of a local non-trivial non-linear relationship between J_{ab} and $J_{ab,eff}$. One way to address the first limitation is by focusing solely on physically meaningful, i.e., equilibrium distances. The latter issue is commonly resolved by introducing method-specific scaling factors[276,398] . These aspects will be thoroughly discussed in Section E.5.1 .

The variation between different approaches for calculating coupling integrals is relatively small, typically around 1%, as long as the considered transfer states and orbitals are similar[276] . Thus, the accuracy of coupling integrals is not limited by the ET approach but instead heavily relies on the chosen quantum mechanical level of theory. Multi-reference configuration interaction (MRCI) and n-electron valence state perturbation theory (NEVPT2) are generally regarded as the "gold standard" for calculating coupling integrals[399] . Density functional theory (DFT) methods exhibit reduced accuracy along Jacob's ladder, with range-separated hybrids or hybrids with approximately 50% Fock exchange performing the best. Different implementations of the same functional can yield up to a 35 meV deviation in $J_{ab,eff}$, corresponding to a typical relative error of 10%. Moving down the ladder towards generalized gradient approximations (GGAs) results in an accuracy decrease of about 10% compared to RSH methods. Descending further to semiempirical methods leads to a deviation from the reference values in the range of 100 to 120 meV (around 40%)[400] . These trends are also depicted in Fig. S1 in the SI. However, this methodical deviation can be significantly reduced by applying a scaling factor to $J_{ab,eff}$, resulting in improvements of up to one order of magnitude.

E.3 Fit and Test Set

The co-planar dimers, with 3.5Å intermolecular distance, of the HAB79 benchmark by Blumberger *et. al.*[276] serve as a fit set to obtain the empirical scaling factors for the PTB coupling integrals. Here, we introduce a new test set called JAB69, that includes the HAB7[396] and HAB11[397] benchmarks and enhances it by 51 chemically comparable, but larger dimers.
The JAB69 benchmark consists of 69 mostly medium-sized, conjugated, parallel, planar, perfectly eclipsed-stacked, homo-dimers with a distance of 3.5Å between their centers of mass. The set is sorted by element composition, i.e., 20 purely carbon- and hydrogen-containing molecules (CH subset), 27 molecules that additionally contain nitrogen and oxygen (CHNO subset), 16 residues furthermore containing sulfur (CHNOS subset), and 6 residues with other elements (CHNOSE subset). Fig. E.2 depicts the Lewis structures of the monomers (optimized cartesian coordinates are available in the Supporting Information (SI)).

E.4 Computational Details

We computed coupling integrals at the ωB97X-D4/TZ2P,[141,278,279] PBE-D4/TZ2P,[300] PTB,[160] ZINDO[401] and DFTB3(3ob-3-1 parameter set)[402,403] levels of theory. We calculated PTB single points with the `ptb 3.7` standalone program[†] using a customized more verbose output and subsequent post-processing with the development version of our DIPRO in-house code.[‡] ZINDO single points were calculated within `Gaussian16` program version C.01[404] followed by a custom DIPRO post-processing script.[§] ωB97X-D4, PBE-D4, and DFTB3 coupling integrals were calculated using `ADF V.2020.102`.[405,406] We took GMH@NEVPT2 reference data for Figure S1 and geometries of the HAB79 test set from Ref. [276], but used our own DIPRO@ωB97X-D4 references for both test sets, HAB79 and JAB69. The range separated hybrid ωB97X-D4/TZ2P proved to be robust and good performing and is widely used for the calculation of electronic coupling integrals for larger molecules as a kind of silver standard[276]. A comparison of DIPRO@ωB97X-D4 and GMH@NEVPT2 is available in SI.xlsx.
For the statistical evaluation, we used the mean deviation (MD) and relative mean deviation (relMD), the mean absolute deviation (MAD) and relative mean absolute deviation (relMAD), the standard deviation (SD) and relative standard deviation (relSD), the Pearson correlation coefficient (ρ_P), and the Spearman rank coefficient (ρ_S). The respective equations are provided in the SI.

E.5 Results

In general, coupling integrals calculated at the SQM level are smaller than at the DFT level. This is a direct consequence of the minimal basis set used by most SQM methods, thus leading to overlocalization of the coupled states and too fast exponential decay of $|J_{ab,eff}|$ with the distance. As PTB uses a larger vDZP basis set, this trend is less pronounced here compared to, e.g., GFN-xTB or ZINDO. Additionally, most SQM methods underestimate electronic gaps, which also directly affects

[†] available upon request
[‡] available upon request
[§] available upon request

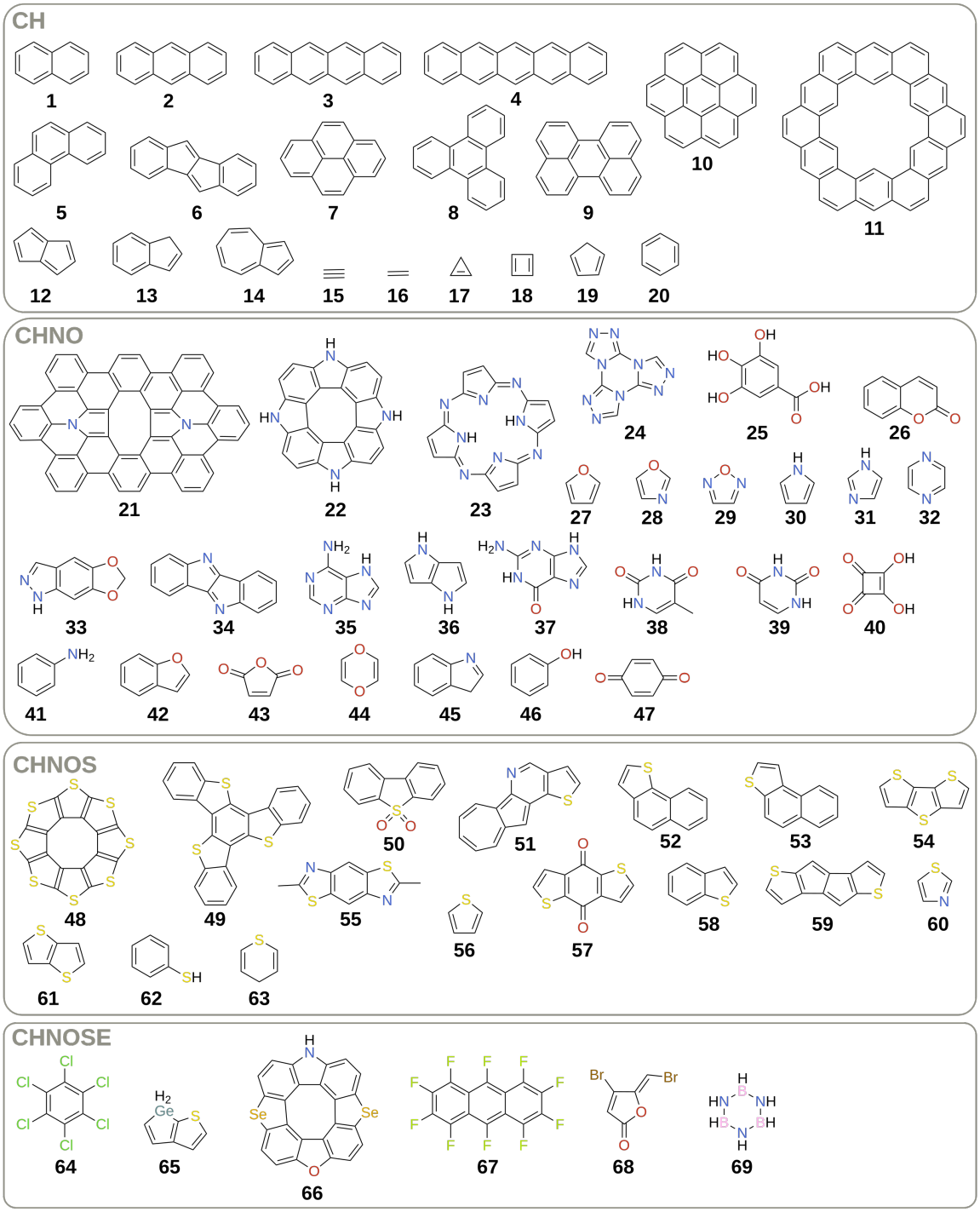

Figure E.2: Monomers of the JAB69 test set sorted by element composition.

the coupling. The DIPRO formula for $|J_{ab,eff}|$ (see eq. 5) exhibits some shortcomings (e.g., it is not valid for very large overlaps, it can exhibit a non-trivial non-linear relationship in certain areas). This can be corrected for by applying a uniform scaling factor f=1.921 to the PTB orbital energies of the dimer, according to the following equation:

$$\gamma_1^i \cdot (\mathbf{E_{AB}} \cdot f) \cdot \gamma_2^j = J_{ab}^{ij} \cdot f \tag{E.6}$$

$$J_{ab,eff}^{ij} = \left| \frac{J_{ab}^{ij} \cdot f - 0.5 \cdot (E_{mon1}^i + E_{mon2}^j) \cdot S_{ab}^{ij}}{1 - (S_{ab}^{ij})^2} \right| \tag{E.7}$$

The scaling factor f is determined by:

$$f = \frac{\sum_{i=1}^{n} |J_{ab}^{ij}(\omega B97X - D4)|/|J_{ab}^{ij}(method)|}{n} \tag{E.8}$$

whereas n denotes the total number of considered systems. We determined the scaling factor on the HAB79 set and then applied it to all other calculations. There are two drawbacks of this procedure. First, the scaling of all values leads to increased SD and RMSD compared to unscaled SQM methods. Second, the scaling can accidentally introduce huge errors and thus biases the statistical evaluation. Although, the calculated scaling factor is not overly sensitive to the underlying test set, it is sensitive to the elemental composition and the dimer distance. The transfer of this scaling factor to less common elements, especially metals and metalloids, is not encouraged. We advise to determine a new scaling factor for these special purposes. No such scaling factors were applied to the other methods, either because it was not necessary (i.e. approximately linear dependency of $|J_{ab,eff}|$ on J_{ab}), or because the chosen program package did not allow user intervention.

E.5.1 HAB79

Fig. E.3 depicts the correlation between different SQM methods and ωB97X-D4/TZ2P reference coupling integrals J_{ab} for the HAB79 set and relates the structure motifs of the outliers.

Except for ZINDO, all examined methods exhibit a significant correlation with the reference values. Among them, PBE-D4 best reproduces the absolute values, followed by PTB. GFN1-xTB and DFTB3 perform equally, but less well than PTB. According to Fig. E.3b we identified three classes of outliers which can be assigned to distinct structure motifs. A forth type, mainly represented by molecules **2** and **3**, is not explained by a certain structure motif but could arise from a non-trivial interplay between the orbitals C and orbital energies E of the dimer. PBE-D4, GFN1-xTB, and DFTB3 exhibit a notable number of structural outliers of class **I.** that result in couplings near zero. On the other hand PTB structural outliers of classes **II.** and **III.** yield systematically under- and overestimated results that are still meaningful, in contrast to the outliers near zero. The two outliers of class **IV.** yield zero couplings for all examined methods except ωB97X-D4. Anyhow, correcting the performance of the identified structure motifs either in PTB or the other methods is very challenging as it necessitates an intrinsic change of the method, i.e. a modified parameterization. Overall, PTB shows the greatest resilience of electronic couplings from the underlying structure motif and can thus be considered as the most robust among the tested methods.

We will delve into the statistical evaluation extensively in Tab. F.3. Generally, PTB and other TB methods adequately describe J_{ab} to derive meaningful $|J_{ab,eff}|$ without sacrificing information,

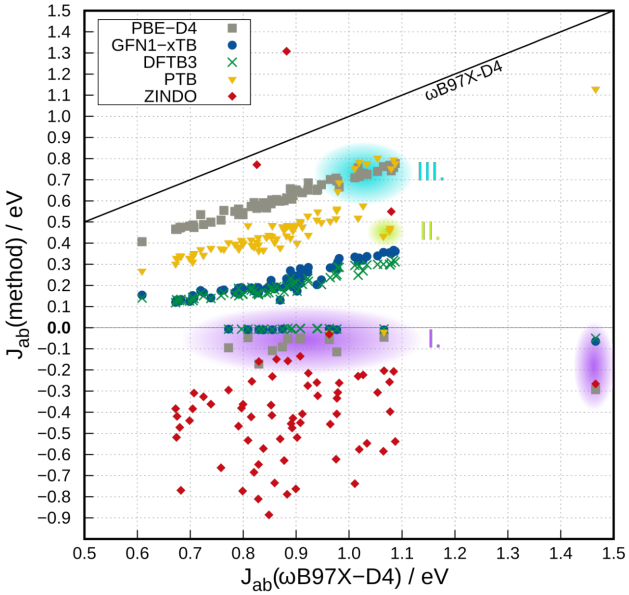

Figure E.3: a) Correlation plot of different tested methods against ωB97X-D4/TZ2P references, for unscaled J_{ab} in eV of the HAB79 benchmark. The black line denotes perfect correlation with the reference. b) Structure motifs of different types of outliers for different methods.

following the aforementioned global scaling with f.
In analogy to Ref. [276] we applied a linear regression to $|J_{ab,eff}|$ for all examined methods. Notably, this statistical scaling differs from the f=1.921 scaling factor introduced above due to physical reasons. The scaling as proposed by Ref. [276] is based on the inverse of the slope of linear fit functions (b denotes the y-axis intersect and m the slope):

$$X_{scaled} = (X_{unscaled} - b) \cdot \frac{1}{m} \tag{E.9}$$

The linear fit functions of all examined methods and test sets are given in the SI. Fig. E.4 and Tab. F.3 depict the resulting statistical measures.
For the full HAB79 set PTB yields the best results of all examined methods, even better than PBE,

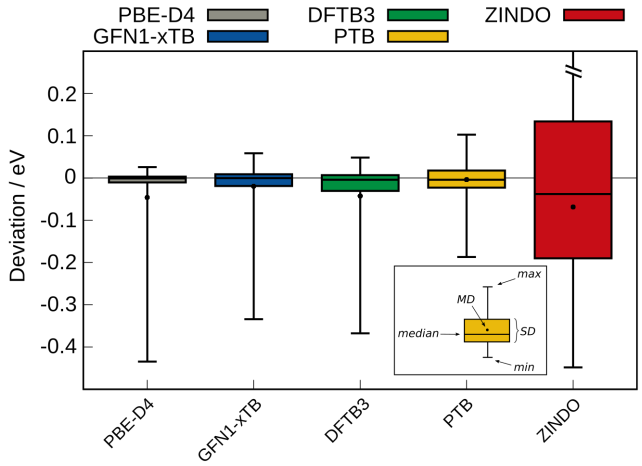

Figure E.4: Statistical measures of different tested methods against ωB97X-D4/TZ2P references, for scaled $|J_{ab,eff}|$ in eV of the HAB79 benchmark with no outliers excluded. The results for ZINDO should be interpreted with care due to the very strong scattering and low correlation with the reference.

Table E.1: Statistical measures, according to equations 6 to 13, of $|J_{ab,eff}|$ values calculated at different levels of theory for the HAB79 benchmark and compared to ωB97X-D4/TZ2P references. Absolute values are given in meV, relative ones in %. The set does not exclude any outliers. ZINDO values could not be scaled due to missing correlation.

	PBE-D4	GFN1-xTB	DFTB3	PTB	ZINDO
MD	-0.046	-0.020	-0.043	-0.004	0.208
MAD	0.051	0.032	0.053	0.033	0.392
SD	0.112	0.060	0.084	0.049	0.928
relMD	-11.9	-4.8	-10.8	-0.9	59.3
relMAD	13.2	8.0	13.4	7.8	102.3
relSD	28.4	13.9	20.7	11.1	247.3
ρ_S	0.808	0.779	0.699	0.794	-0.175
ρ_P	0.451	0.637	0.55	0.762	-0.139

closely followed by GFN1-xTB. As noted above, ZINDO performs worst. Another big advantage of PTB, besides its performance, is its robustness. There are only two 2σ-outliers among 79 systems, whereas GFN1-xTB features 11 outliers, PBE 12 outliers, and DFTB3 even 18 outliers. The underlying structural features that cause these outliers were already discussed at Fig. E.3. Further sources for outliers may be inaccurate orbital energies, an altered orbital order, or a wrong sign of J_{ab} (that may depend on very small contributions of the overlap matrix).
The linear scaling improves the tested methods by up to 65% in relMAD or up to 30% in relSD. In general, SQM methods experience a much stronger improvement than PBE, rendering them competitive to GGA DFT, but at a much lower computational cost. Additional statistics for the unscaled values are given in SI.xlsx.

E.5.2 JAB69

Statistical results for the JAB69 benchmark are presented in Fig. E.5 and Tab. F.2
PBE performs best on the JAB69 benchmark, followed by PTB. ZINDO yields again the worst results,

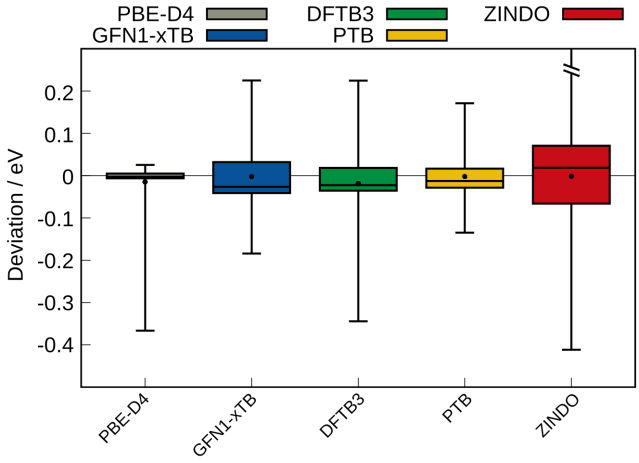

Figure E.5: Statistical measures of different tested methods against ωB97X-D4/TZ2P references, for scaled $|J_{ab,eff}|$ in eV of the JAB69 benchmark with some outliers (e.g. due to convergence issues) or elements excluded due to missing parametrization as mentioned in Tab. F.2.

but in contrast to the HAB79 set, the observed correlation allows for linear regression. This benchmark shows a significantly reduced number of structural outliers close to zero compared to the HAB79 benchmark. There are four outliers for PBE, one each for GFN1-xTB and PTB, seven for DFTB3, and nine for ZINDO. Again, outliers share the commonalities mentioned in Fig. E.3, except for No. **15**, **16**, and **17**, that we will discuss in detail. Excluding the outliers from the statistical evaluation has fewer effects than in the HAB79 benchmark; PBE experiences the strongest improvement with 4% in relMAD and 12% in relSD.
The advantage of employing a uniform scaling factor (f) acting on the PTB dimer orbital energies is a significant improvement in many results with minimal additional effort. However, there is a drawback when it comes to certain molecules that are poorly described by such a simplified approach. This applies specifically to ethyne, ethylene, and cyclopropene (**15**, **16**, and **17** in Figure E.2), respectively. The scaling factor of 1.921, as described in equations 6 to 8, derived primarily from medium-sized

Table E.2: Statistical measures, according to equations 6 to 13, of $|J_{ab,eff}|$ values calculated at different levels of theory for the JAB69 benchmark compared to ωB97X-D4/TZ2P references. Absolute values are given in meV, relative ones in %. No. **15** did not converge with PBE and No. **19** not with ZINDO. There were no parameters available for various residues from the CHNOSE subset in DFTB3 (No. **65**, **66**, **69**) and ZINDO (No. **64**, **65**, **66**, **68**, **69** see Figure E.2). Accordingly, the respective part of the statistics exclude the mentioned molecules.

	PBE-D4	GFN1-xTB	DFTB3	PTB	ZINDO
MD	-0.015	-0.002	-0.018	-0.002	0.031
MAD	0.022	0.059	0.061	0.039	0.109
SD	0.064	0.077	0.095	0.055	0.168
relMD	-2.865	9.531	-0.242	6.338	55.681
relMAD	8.528	25.147	20.558	17.385	74.368
relSD	22.18	55.94	35.47	38.28	273.05
ρ_S	0.972	0.858	0.819	0.904	0.381
ρ_P	0.879	0.821	0.763	0.896	0.315
no. of molecules	68	69	66	69	63

molecules, proves to be too large for these small systems. This scaling factor is size- and distance dependent due to the spatial behavior of molecular orbitals and their overlap.
By increasing the system size of molecules like ethylene and benzene, for instance, through the expansion into homologous rows of polyenes and acenes, the scaling factor converges as the system size grows. Larger systems exhibit optimum scaling factors closer to 2, resulting in smaller final values of $|J_{ab,eff}|$, which arises from the asymptotically decreasing electronic gap. Additional information on this topic can be found in the SI.
The scaling factor is not ideally transferable to non-organic elements such as heavier main group elements or transition metals. Another drawback is the distance dependence of the scaling factor. There are inherent differences in the decrease of $|J_{ab,eff}|$ between SQM and DFT methods due to the minimal basis set of SQM methods lacking long-ranged diffuse functions. As the distance between coupling fragments increases, the scaling factor also increases approximately linear. The scaling factor does not exhibit any angular dependency. Accordingly, the work in progress in our lab includes implementing not only a simple scaling factor but a more accurate and versatile scaling function for an improved description of distance, size, and system dependencies. Further details can be found in the SI.
Presently, our DIPRO implementation only considers single orbitals for the calculation of coupling integrals, specifically the HOMO/LUMO of each monomer if we are interested in hole/electron couplings and transport properties. This may result in significant deviations in $|J_{ab,eff}|$ when dealing with near-degenerate orbitals or cases where the orbital order has been altered. Additionally, in special linear high-symmetry situations like for ethyne, the local coordinate systems assigned to the individual fragments may differ from the local coordinate system of the less symmetric dimer. There are two possible pathways how to account for near-degeneracy in the future. First, the root mean square over the sum of all possible couplings within a certain energy range as it is done in ADF, or second, the transformation of the degenerate orbitals using a symmetry adapted basis to obtain a new orthogonalized Hamiltonian with a unique set of couplings as it is done in Ref. [397].

E.5.3 Challenging Systems

Relevant molecules for organic electronic applications (e.g., OFETs, photodetectors, OPVs) are merocyanines. Those have been studied and named in the late 40s[407] as dyes and photo-agents; and from the early 80s[408] to today[287] they have been utilized and extensively investigated as OSC materials[409,410]. Recently, merocyanines have been studied for applications in the field of bio-imaging ,[411,412] optical sensors for temperature,[413] pH,[414] or chemicals,[415] photosensitizers in nanomedicine and cancer therapy,[416,417] and antimicrobial drugs.[418,419] The computational investigation of merocyanines is challenging due to their electronic structure, namely the resonance between zwitterionic and neutral structures, and thus strong electron correlation effects. Additionally, most often a single-molecule approach, in contrast to a cluster-aggregate or nano-crystalline approach, is not sufficient to describe all properties of interest. Typically, range-separated hybrid functionals (RSH) and high-level wave function methods like CASSCF/NEVPT2 or CCSD(T) are employed for accurate examinations.[420–422] Notably, the PTB method emulates such RSH behavior and additionally is parameterized to yield good hyperpolarizabilities, which are crucial for describing merocyanines.[423–425] Fig. E.6 shows the molecular structures of some merocyanines as well as packing motifs and $|J_{ab,eff}|$ for various SQM methods.

Among the tested SQM methods, ZINDO and PTB show the best agreement with the reference

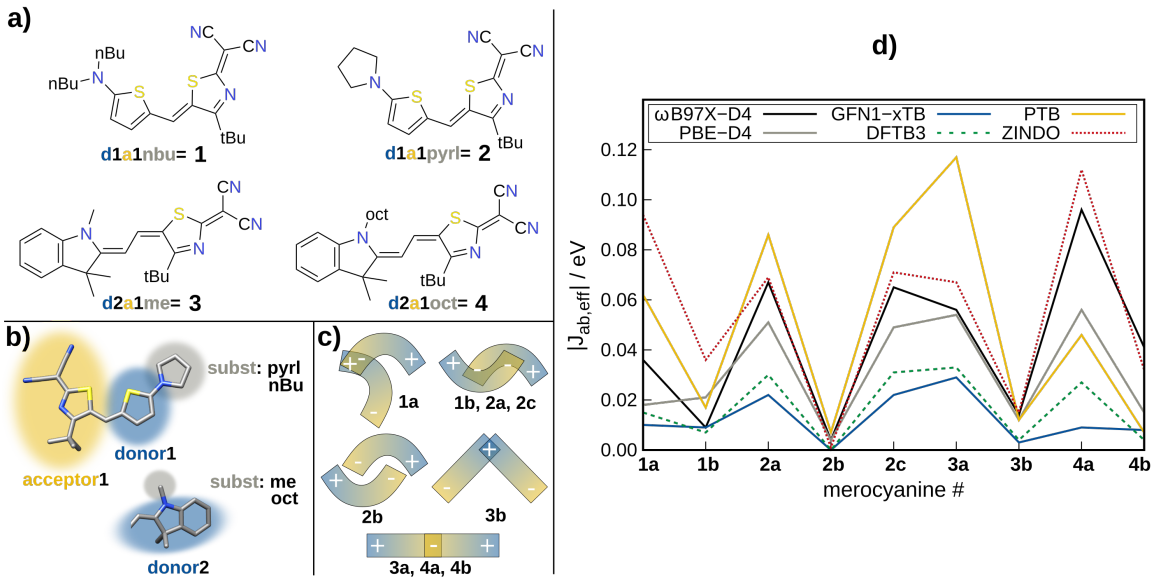

Figure E.6: Panel a): different kinds of merocyanines (1-4) as derived by combination of various donor (d) and acceptor (a) groups. Panel b): molecular structures of donor and acceptor groups, with indication of different substituents, namely: methyl (me), normal-butyl (nBu), octyl (oct) and pyrrolidine (pyrl). Panel c): sketches of possible intermolecular packing motifs. Panel d) coupling integrals $|J_{ab,eff}|$ in eV for different merocyanine dimers at different levels of theory.

ωB97X-D4/TZ2P, reproducing the order of magnitude as well as the relative order of couplings. Noteworthy, the average coupling of the investigated merocyanines is almost by factor 0.5 lower than the estimated target accuracy of the SQM methods as determined for the HAB79 and JAB69 benchmarks. The very good performance of PTB and ZINDO for the merocyanines is rather surprising.

Firstly, merocyanines are dipolar molecules with significant charge delocalization, which is generally challenging for SQM methods to describe accurately. Secondly, the diverse dimer packing motifs, ranging from eclipsed stacking to in-plane coupling and various close dimers (see Fig. E.6c), pose a high challenge for SQM to achieve a uniform treatment. Thirdly, the universal scaling factor used in PTB is not tailored to merocyanines. Lastly, the methods perform reasonably over a coupling range of four orders of magnitude, i.e. spanning from 10^{-3} eV to over 10^{0} eV, for most of which they have not been benchmarked. Considering these factors, the remarkable correlation between PTB and ZINDO is exceptional, particularly because ZINDO has previously shown less favorable results in our JAB69 and HAB79 benchmark study, and PTB has occasionally suffered from overscaling (notably, only merocyanine **3a** is overscaled). So far, we were not able to exploit the exact reasons for the good performance of ZINDO for the merocyanines, which is in contrast to the rather poor performance for the benchmarks sets discussed above. Overall, these findings highlight the favorable transferability of the DIPRO@PTB approach to molecular materials with complex/challenging electronic structure and conformational flexibility.

MOFs are relevant and emerging materials in the development of modern organic electronics.[426–428] They are independent supramolecular building blocks that exhibit a high degree of order and can easily be customized for special applications. Furthermore, MOFs use the advantageous electronic properties of metals while only containing a minimal amount of them, whereas merocyanines are purely organic. In the following, we show the calculation of coupling integrals for a large metal organic cage (MOC) with PTB. MOCs are the one-dimensional variant of MOFs. Our test case is shown in Figure E.7.

The examined MOC is an organic cage constituted of Pd-linked anthracene panels encapsulating

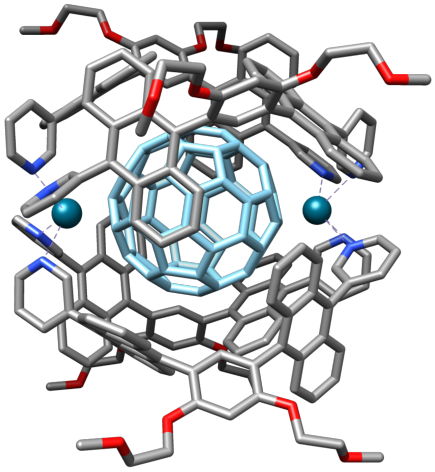

Figure E.7: GFN2-xTB/ALPB($CHCl_3$) optimized structure of the examined MOC[429] . Hydrogens are omitted for clarity, the two displayed metal atoms are Palladium.

a C_{60} fullerene. Such host-guest systems are dominated by non-covalent interactions (NCIs) which are in general difficult to describe as they rely on electron correlation effects. Dispersion corrected DFT and the GFN-xTB methods are able to describe NCI systems reasonably accurate. The $|J_{ab,eff}|$ results for different methods are given in Tab. E.3.

The coupling integrals computed with the low-level methods exhibit good agreement with the ωB97X-D4 reference, both qualitatively and in terms of magnitude. Unfortunately, it was not possible

Table E.3: $|J_{ab,eff}|$ in meV for different methods and different combination of molecular orbitals. h corresponds to the HOMO, h-1 to HOMO-1, and l to LUMO.

orbitals						
cage	C_{60}	ωB97X-D4	PBE-D4	GFN1-xTB	PTB	$PTB_{unscaled}$
h-1	h-1	22	10	13	12	5
h	h-1	60	24	56	79	36
h-1	h	3	8	8	7	3
h	h	23	21	38	57	24
h-1	l	40	155	86	72	33
h	l	434	73	55	639	292
l	l	20	18	68	10	4

to test ZINDO due to the absence of parameters for Pd. As previously mentioned, the scaling factor of 1.921 for PTB may not be suitable for extremely large or small systems. Consequently, we also present the unscaled values in this study. It is important to note that the comparability of coupling integrals between different methods relies on the similarity of the underlying molecular orbitals. The degree of orbital delocalization and degeneracy increases with the extension of the π-system. To address this issue, we conducted a manual inspection of the molecular orbitals near the active space and arranged the couplings to maximize the similarity of transfer orbitals. Orbital visualizations are available in the SI (Fig. S5). In terms of describing charge transfer integrals, PTB outperforms GFN1-xTB and PBE-D4 due to its partial RSH character. Despite the inherent challenges associated with merocyanines as well as MOCs, such as high correlation effects, large system size, intricate electronics, inclusion of metals, orbital ordering, and near-degeneracy, PTB performs well in describing them.
Approximate computation times on a quadruple core computer range from approximately one minute for GFN1-xTB, around six minutes for PTB, and 30 hours for PBE-D4/TZ2P to over five days for the ωB97X-D4/TZ2P reference.

E.6 Conclusion and Outlook

We utilized the recently developed semiempirical PTB method in combination with the dimer projection method DIPRO, to compute intermolecular charge transfer coupling integrals, denoted as $|J_{ab,eff}|$. To enhance the correlation with the reference method ωB97X-D4/TZ2P, we determined a basic scaling factor for PTB, resulting in an improvement of up to 60% for the computed coupling values. Our study involved testing several SQM methods on Blumberger's HAB79 benchmark as well as our newly compiled JAB69 benchmark. The performance of all methods based on tight-binding models was highly satisfactory after scaling, with PTB even surpassing the previously assumed accuracy limit of 100 to 120 meV in MAD for the JAB69 set. Specifically, PTB achieved a MAD of 76 meV, while other tight-binding methods exhibited MAD values around 260 meV and ZINDO showed a MAD of 317 meV. PTB exhibited general robustness against outliers, even slightly outperforming GGA-DFT in this aspect. Additionally, PTB demonstrated the ability to handle heavier main-group elements and transition metals due to its comprehensive parameterization covering all elements. On the other hand,

a major drawback of PTB is its need for a system and distance dependent empirical scaling factor. Furthermore, we demonstrated the transferability of our approach to medium-sized merocyanines and large-sized MOCs, both of which are prominent examples in organic electronics. The computational speedup of PTB achieved in calculating $|J_{ab,eff}|$ for a MOC system with 446 atoms, compared to GGA-DFT, is approximately 300-fold. Yet the PTB results are in good agreement with the hybrid-DFT reference. Further improvements for a better treatment of nearly degenerate and partially occupied orbitals with the DIPRO approach are already in preparation in our lab.

E.7 Supplementary Material

The electronic supplementary information include:
i) the molecular structures of the HAB79 and JAB69 test set and the investigated merocyanines and MOFs as xyz coordinate files in HAB79.zip, JAB69.zip, merocyanines.zip and MOF.xyz
ii) the raw data for the generation of statistics and pictures as SI.xlsx table
iii) SI.pdf with additional figures, tables and investigations useful for understanding and more in-depth knowledge.

E.8 Acknowledgments

The authors thank the *German Research Foundation* (DFG) for funding in the framework of *RTG 2591 - Template-Designed Organic Electronics (TIDE)*.

E.9 Data Availability Statement

The data that support the findings of this study are available within the article and its supplementary material. The customized alpha program versions are available from the authors upon request.

APPENDIX F

A Semi-Automated Quantum-Mechanical Workflow for the Generation of Molecular Monolayers and Aggregates

Julia Kohn,[†] Stefan Grimme[†], Andreas Hansen,[†]

Received: 23 July 2024
Published online: 25 September 2024

Reprinted in Appendix F (adapted) with permission[‡] from
J. T. Kohn, S. Grimme, and A. Hansen, *A Semi-Automated Quantum-Mechanical Workflow for the Generation of Merocyanine Monolayers and Aggregates for Organic Electronics*, J. Chem. Phys. **161**.12 (2024), DOI: 10.1063/5.0230341

Own contributions

- coding and implementation
- performing of all calculations except for coupled cluster
- interpretation of the results
- writing of the manuscript

[†]Mulliken Center for Theoretical Chemistry, University of Bonn, Bonn, 53115, Germany
[‡]Permission requests to reuse material from this Chapter should be directed to AIP Publishing.

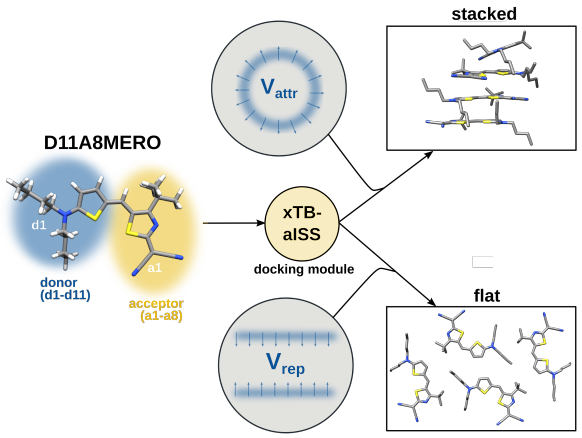

Figure F.1: Table of content graphic (ToC).

Abstract Organic electronics (OE) such as organic light-emitting diodes (OLEDs) or organic solar cells (OSCs) represent an important and innovative research area to achieve global goals like environmentally friendly energy production. To accelerate OE material discovery, various computational methods are employed. For the initial generation of structures, a molecular cluster approach is employed. Here, we present a semi-automated workflow for the generation of monolayers and aggregates using the GFN*n*-xTB methods and composite density functional theory (DFT-3c). Furthermore, we present the novel D11A8MERO dye interaction energy benchmark with high-level coupled cluster reference interaction energies for the assessment of efficient quantum chemical and force-field methods. GFN2-xTB performs similar to low-cost DFT, reaching DFT/mGGA accuracy at two orders of magnitude lower computational cost. As an example application, we investigate the influence of the dye aggregate size on the optical and electrical properties and show that at least four molecules in a cluster model are needed for a qualitatively reasonable description.

F.1 Introduction

Organic electronics (OE)[14–17] represent a rapidly evolving research area. Since their inception in the 1980s,[26,27] the field has diversified into various applications including organic light-emitting diodes (OLEDs),[22,23] organic field effect transistors (OFETs),[24,25] organic solar cells (OSCs),[18–21] organic circuits,[430–432] and molecular computers.[433–436] This led to famous inventions like OLED displays,[437] smart windows,[37] electronic paper,[38] printed electronics,[35] or conductive ink,[36] spanning the whole process from their initial discovery in the lab to industrial serial production and end customers. These advances often approach or even surpass the quality and efficiency of pre-existing non-organic technologies at usually reduced costs and environmental impact.[438–441] OE have gained renewed attention as a promising field to promote sustainable energy production,[442–444] biomimetics,[445,446] environmental sensing devices and health monitoring,[447,448] and secure digitalization efforts like the Internet of Things (IoT)[449,450] all while being less dependent on rare earth metals.
Given the wide and diverse field of potentially promising materials in OE, quantum chemical modeling and machine learning approaches have significantly accelerated material discovery.[451–456] The overarching goal is to understand the structure-property relationships to predict better candidates

for further research and to understand their function. Key opto-electronic properties of interest include absorption/emission spectra, oscillator strength, charge carrier mobility, band gap and singlet-triplet gap, and quantum yield.[39–41]
In general, there are two ways to approach OE materials computationally, 1) in the bulk using periodic boundary conditions[280,457] and 2) as single molecules or molecular aggregates and clusters.[281,282] Both approaches have their advantages and downsides. The major drawback of 1) is the limited description of the conformational space and polymorphism, while 2) could be a rather crude approximation of the bulk largely neglecting long-range and many-body effects. Both approaches require initial structure guesses, which can be difficult as high quality crystallographic data is not always available. Computational crystal structure prediction (CSP)[458–460] is already extremely challenging for small molecules but becomes almost infeasible for large and flexible organic compounds with many polymorphs. Alternatively, a common way to generate aggregate structures for computational studies is to cut them out from an experimental crystal structure (if available).[288,409] However, this approach is also not always feasible and can limit the exploration of novel structures. Therefor, a reliable, reproducible, unbiased, and efficient workflow for the generation of aggregates, which is able to approximate the bulk by a bottom-up approach instead, is highly desirable. Such workflows have successfully been applied to, e.g., explicit solvation,[461,462] and bound inorganic nanoclusters.[463] In this work, we introduce an adapted workflow and apply it for the generation of various merocyanine tetramers.
Merocyanines are an important molecule class for organic electronics, e.g., for dye-sensitized solar cells (DSSCs),[30,31] photo-detectors,[464,465] and OFETs.[283,466] Discovered in the late 1940s and initially used as dyes and photo-agents,[407] merocyanines attracted significant attention about 40 years later as OPV materials.[408,467] Today, their applications extend to optical sensors for various triggers like chemicals,[415] pH,[414] or temperature,[413] bio-imaging,[411,412] antimicrobial drugs,[418,419] and photo-sensitizers for cancer therapy and nanomedicine.[416,417]
Kohn-Sham density functional theory (DFT) is usually computationally too expensive for large structure generation and ensemble approaches, thus necessitating the use of cheaper but still reliable methods. Semi-empirical quantum mechanical (SQM) methods offer a good trade-off between computation time and accuracy and have already been successfully applied to related problems.[4,133] Additionally, several efficient composite DFT methods have been developed over the last years specifically designed for large molecular systems.[123–125,468] Here, we evaluate the performance of various low-cost DFT and SQM methods for the here newly introduced D11A8MERO dye interaction energy benchmark set. It highlights the differences between flat and stacked aggregates and provides high-level coupled cluster and double-hybrid DFT reference energies for nearly 600 tetrameric systems comprising up to 260 atoms and generated with the proposed workflow. As a practical application we investigate the influence of the aggregate size on the opto-electronic properties and their convergence towards a sufficient approximation for the bulk.

F.2 Workflow

Our workflow is based on the automated interaction site screening algorithm (aISS) as implemented in the `xtb` program package.[138,212,469] This docking module identifies the global energy minimum (most negative interaction energy) between two fragments by docking one molecule at a time producing an energetically sorted ensemble of aggregates. For multiple docked molecules the procedure is

iteratively repeated. The aISS performs an interaction energy pre-screening by moving a charged Kr dummy atom around the target molecule. At favorable positions the docking molecule is initially placed and a grid-based search of further docking motifs is conducted, representing the molecule by its center of mass and principal axes of inertia. This involves a stack search, an angular search, and a directed enzyme-pocket-like search. This step is followed by an energy evaluation and subsequently a genetic optimization with random crossovers and mutations in the intermolecular positions. The last two steps are repeated until convergence and final geometry optimizations and energy evaluations on GFN1-xTB level are performed. As the GFN methods are parameterized up to Z=86, the procedure is not limited to bio-molecules and protein-ligand binding but can also treat metal organic frameworks (MOFs), zeolithes, or transition metal complexes. For more in-depth information about the aISS docking module, see the original publication Ref. [138].

The proposed workflow has two distinct run-types, 1) generation of flat layers and 2) generation of stacks, as schematically shown in Figure F.2. Both run-types combined allow for controlled three-dimensional cluster growth. For the growth of monolayers, we employ a constraining repulsive potential (V_{rep}) in the form of a sandwich as implemented in the `xtb` program package[469] to keep the clusters artificially flat. This approach counteracts the natural tendency of molecules to prefer face-to-face stacking or spherical aggregation to maximize dispersion interactions. It is also possible to apply just one half of the sandwich potential, i.e., a one-sided repulsive surface potential to ensure docking of molecules to only one side of a molecular surface model. This procedure also works for docking onto a surface like h-BN or graphene. Furthermore, rearrangement of the sub-monolayers upon docking of additional molecules can be optionally simulated by geometry optimization.

For the growth of columnar stacks, a directing attractive potential (V_{attr}) is applied to the flat core of the molecule to ensure growth perpendicular to the molecule plane. This step is crucial since clusters, unlike crystals, do not grow in an ordered fashion, i.e., the interaction depends only weakly on whether the next molecule docks to the existing stack or to its side. Consequently, the workflow currently is not able to simulate strongly slipped-stacked or porous growth.

We investigated the necessity of a final optimization of the generated flat tetrameric ensembles with loose convergence criteria at the GFN2-xTB level, which we had initially neglected. As a result, we advise to apply such a final optimization step by default because otherwise significant deviations from the equilibrium geometry due to insufficient relaxation of the intermolecular contact orientations can lead to artificially high-lying structures in subsequent DFTtreatments. For instance, for the system 4_s_d11a7_6 the difference between ωB97X-3c energies and coupled cluster references was reduced from 19.1 kcal/mol down to 7.5 kcal/mol upon relaxation. This shows that relative errors of usually well-behaved methods can artificially rise significantly when the final structure relaxation step is omitted because of non-parallel potential energy surface especially in repulsive regions. (see Supporting Information (SI) Chapter 2 for more details).

Additional options available in the `xtb` program package can also be applied, e.g., implicit solvation, electric field, different GFN methods, or geometrical constraints to enhance flexible usage of our workflow. Looking forward, we plan to develop an extended algorithm that considers factors beyond minimization of the interaction energy such as surface/volume coverage or dipole orientation. Our workflow is not intended for crystal structure prediction (CSP) but aims to identify divergent structure motifs to reduce the chemical space in which CSP has to be performed. This can provide educated guesses and subsidiary conditions for an accelerated CSP or machine learning (ML) approaches.

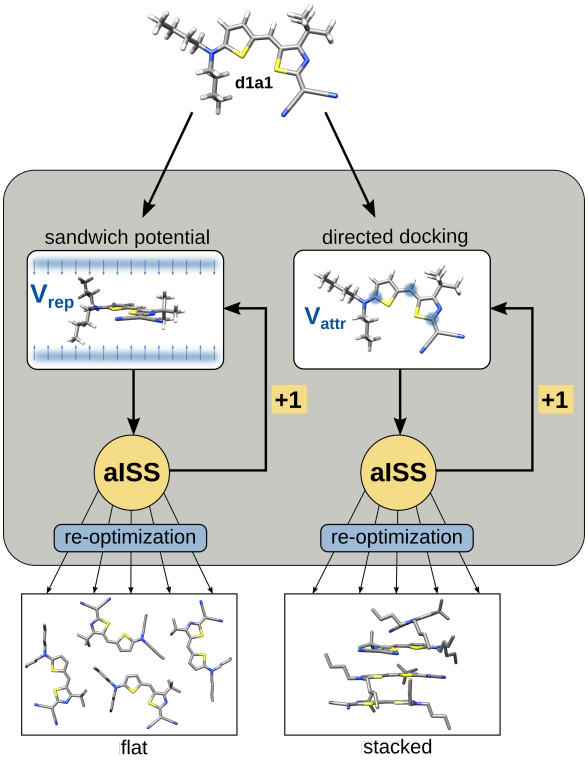

Figure F.2: Schematic workflow for the generation of monolayers and aggregates at the example of the D11A8MERO benchmark (see below, Figure F.3).

F.3 D11A8MERO Benchmark and Structures

Recent studies elucidated the complex structure of merocyanines in the bulk and thin films and revealed how the obtained polymorphs and employed reaction conditions influence the measured opto-electronic properties.[294,470] The computational description of merocyanines is demanding due to their zwitterionic electronic nature with strong electron correlation effects. For the investigation of thin films, a single-structure approach is not sufficient[4,288] to describe all key optical properties adequately, an ensemble or cluster/aggregate approach, such as the one provided by the presented workflow, is necessary. With our workflow, we provide an easy-to-apply tool to support such investigations computationally and merocyanines represent our test cases, also because of their challenging structural diversity and properties.

Inspired by the MEROCY9 set,[5] a set of four different merocyanine dimers in several configurations, we compiled a test set for various monomers comprising eleven different donor units (**d**) and eight different acceptor units (**a**) following synthesized merocyanines (for nomenclature see Figure F.3).[470–472] In principle, donor and acceptor units can be combined arbitrarily and further units with slightly varied alkyl- and aryl-substituents are conceivable, e.g., *n*-butyl instead of ethyl, as long as they are planar and fully conjugated. **d1-6** and **a1-3** were extracted from the MEROCY9 set and extended by ten chemically diverse **d**-**a** combinations from **d7-11** with **a4-8**, with each donor and each acceptor occurring twice, to cover a broader part of the chemical space of merocyanines. Furthermore, we paid attention to include differently flexible cores and side chains to account for broad and narrow

conformational spaces. For the donors **d1-3** which share the same core but different side chains, the influence of the side chains onto the structures and properties was elaborated in greater detail as this core belongs to one of the most investigated merocyanines and many experimental structures are available.[189,287,465,473–475] We used 28 different merocyanines for our investigations. That comprises each of the 18 combination of **d1-6** with **a1-3**, and ten combinations of **d7-11** with **a4-8** as depicted in Fig. F.3 (namely **d7a4**, **d7a6**, **d8a5**, **d8a6**, **d9a7**, **d9a8**, **d10a4**, **d10a8**, **d11a5**, and **d11a7**). All corresponding structures can be found in the SI. The obtained monomers range between a size of 31 to 65 atoms. As described in section F.2, we generated flat and stacked tetramers of each of the 28 molecules, resulting in systems sizes between 124 and 260 atoms. We decided to investigate tetramers and not larger clusters because prior investigations of acenes[288] showed that already tetramers exhibit a right qualitative behavior of the properties of interest compared to the bulk. We wanted to verify that this number also holds true for more complicated systems like merocyanines, which will be elaborated in section F.5.3. Furthermore, structures up to around 200 atoms are still feasible for high level coupled cluster and double hybrid DFT calculations as reference. In the course of our investigations, we reduced the number of generated conformers in the **d7-11**x**a4-8** part from twelve to eight to keep the computational effort feasible.

As shown for the example of **d1a1** in Figures F.4 and F.5, our ensemble approach yields relevant

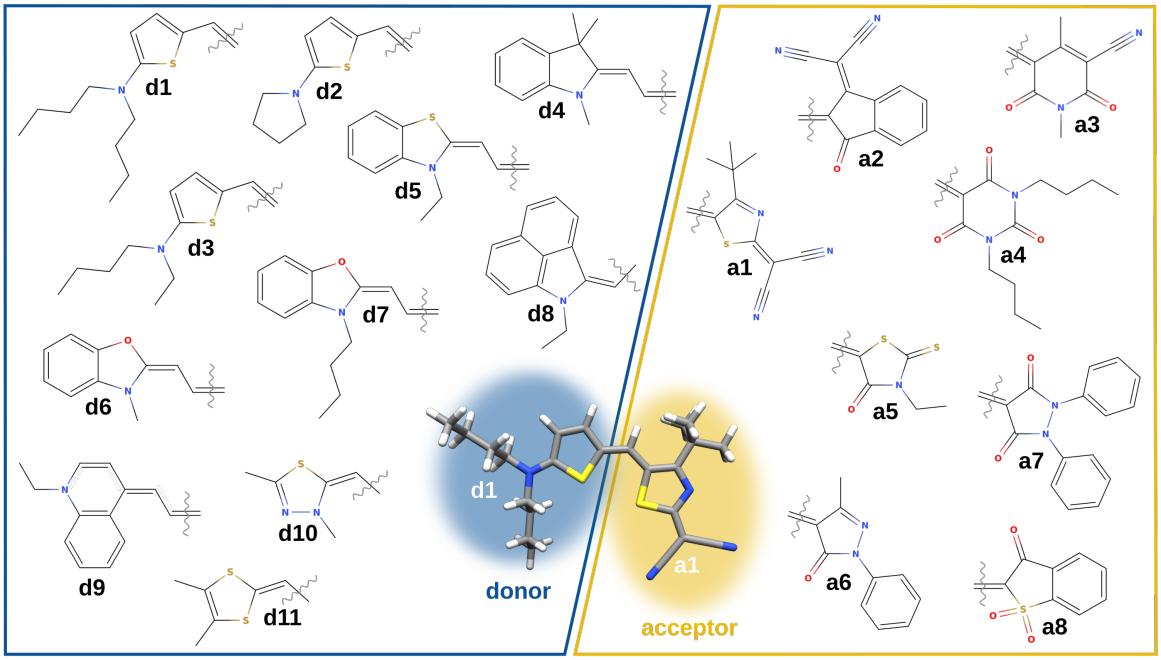

Figure F.3: Representation of the D11A8MERO benchmark set comprising eleven different donor units and eight different acceptor units as found in synthesized merocyanines.

structure motifs that occur in the experimental crystal structures. We also compared our approach to experimentally available structures of **d2a1**, **d3a1**, **d4a1**, and **d1a2** as found in the Cambridge Structural Database (CSD). The general structural and energetic results of this ensemble approach show the potential of our aggregate generation workflow to contribute to joint experimental-theoretical crystal structure determination via supportive computational modeling.

Merocyanines prefer dimer motifs due to their zwitterionic nature. The most prominent motifs are

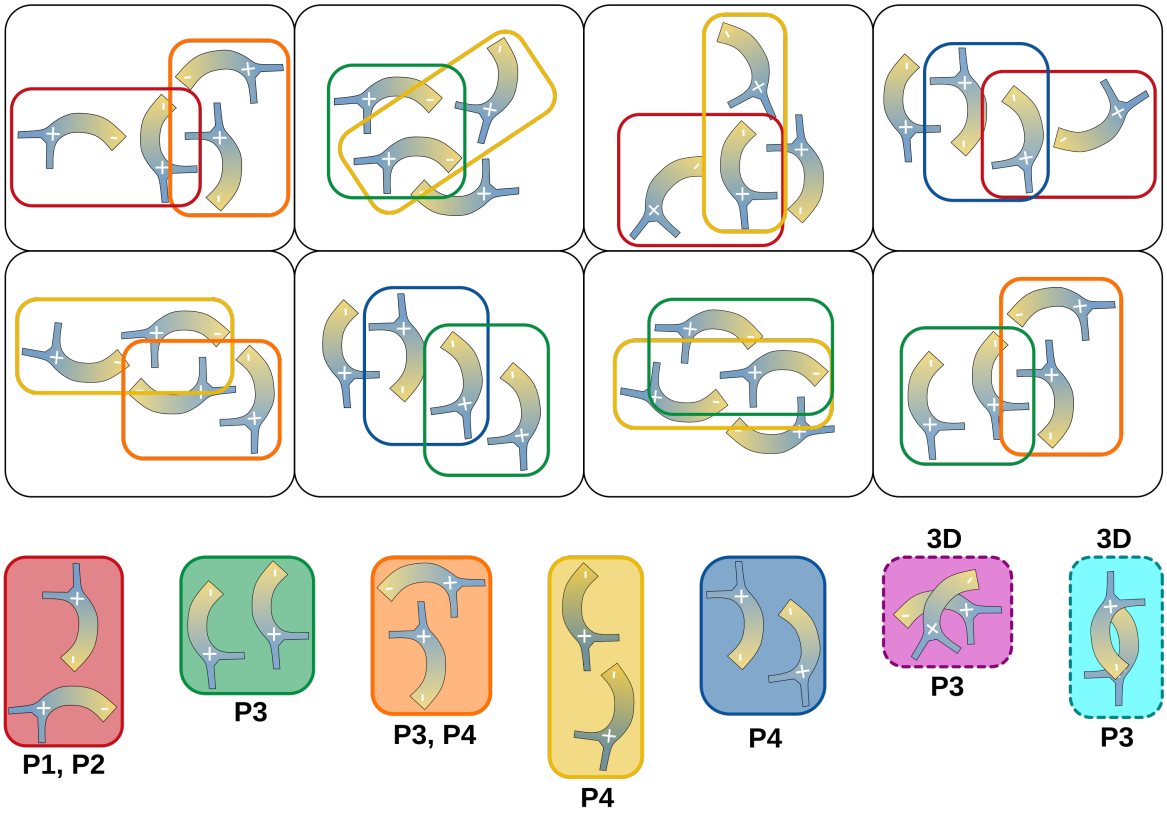

Figure F.4: Schematic representation of flat aggregates of **d1a1** generated with the introduced workflow. Structural motifs that also occur in the experimental crystal structures are marked in different colors. **P1-P4** denote the different available polymorphs of **d1a1**, whereas **P3** is the most likely structural motif found in the respective thin film structure.

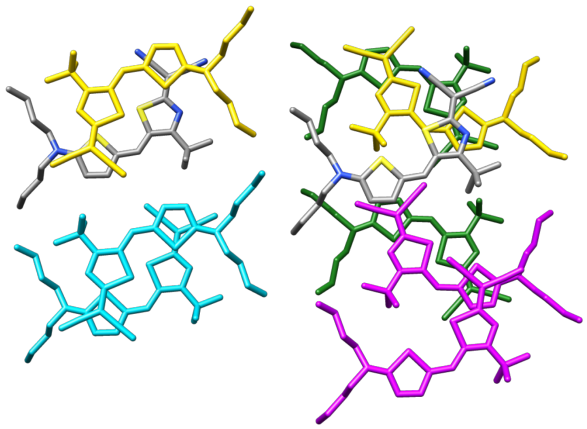

Figure F.5: Cutout of the crystal structure of **P3** with highlighted structural motifs predicted by the presented workflow using the color code introduced in Figure F.4. Hydrogen atoms omitted for clarity.

parallel/antiparallel next to each other (Fig. F.4 green, blue), perpendicular next to each other (red, orange), daisy-chained (yellow), perpendicular stacked (pink), and antiparallel stacked (turquoise). These motifs occur throughout all merocyanines in the D11A8MERO set and all investigated experimental polymorphs. As mentioned previously, we miss out on the strongly slipped stacked motifs as we are still in the molecular regime. Hence, the sterical constraints of a crystal to force especially the side chains into certain less interacting conformations are missing. An energetically lowest structure approach alone is not sufficient to capture all structural motifs possibly occurring in the crystal. The importance of ensembles is already widely known for molecular problems[476,477] and it is equally important for cluster and bulk approaches,[478,479] especially when there are multiple experimental low-energy polymorphs of a compound instead of just one single crystal structure.

F.4 Computational Details

The monomer structures were generated with the built-in molecular editor in `Avogadro`,[368] and were subsequently optimized at the GFN2-xTB[143] level of theory using `xtb` V6.6.1.[469] Tetramers were generated by applying our workflow as described in section F.2 at the GFN1-xTB level[142] with a subsequent geometry refinement with loose optimization criteria at the GFN2-xTB level of theory. For the monomers from the combinations of **d1-6** with **a1-3**, twelve tetramers, each flat and stacked, were generated, whereas for the monomers **d7a4**, **d7a6**, **d8a5**, **d8a6**, **d9a7**, **d9a8**, **d10a4**, **d10a8**, **d11a5**, and **d11a7**, eight were generated for each, the stacked and flat run-type.
For a representative subset of 16 tetramers (D11A8MERO16), high-level coupled cluster single point energies were calculated using PNO-LCCSD(T)-F12b[285,286,480]/aug-cc-pVTZ+D[481] (cc-pVTZ for H; without CABS singles correction for the Hartree-Fock energy) as implemented in Molpro[482,483] (2024.1 release) and compared to various hybrid, range-separated hybrid, and double-hybrid functionals with a def2-TZVPP and def2-QZVP basis,[484] such as B3LYP-D4,[231] PBE0-D4,[301] ωB97M-V,[485] PWPB95-D4,[284] and revDSD-PBEP86-D4,[486] to identify an appropriate theoretical reference level, which is computationally feasible for all structures of the D11A8MERO set. Furthermore, a basis set convergence study at the PBE-D4 level was conducted employing the def2-SVP, def2-SVPD, def2-TZVP, def2-TZVPD, def2-TZVPP, def2-TZVPPD, def2-QZVP, def2-QZVPD, def2-QZVPP, and def2-QZVPPD basis sets using `Turbomole` V7.6.
From all generated 592 tetramer structures, ten outliers from the flat subset and two from the stacked subset were excluded, because they either rearranged to stacked motifs upon geometry refinement or formed fused sidechains. Accordingly, 294 stacked tetramers and 286 flat tetramers are included into the statistics and the final set. Single point energies were conducted at the PWPB95-D4/def2-TZVPP,[284] revDSD-PBEP86-D4/def2-TZVPP,[486] ωB97X-3c,[123] B97-3c,[125] PBEh-3c,[126] r^2SCAN-3c,[124] PBE-D4/def2-TZVP,[300] GFN1-xTB, GFN2-xTB, PM6-D3H4X,[487–489] GFN-FF,[120] MMFF94[490] levels of theory using `xtb`, `dftd3`[146] and `dftd4`[141] standalone programs, `MOPAC2016` V19.179L with `xtb` as driver,[491] `Turbomole` V7.6,[298,492] and `Open Babel` V3.1.1 programs.[493] All double-hybrid calculations employ the RIJCOSX approximation and all hybrid and GGA calculations the RIJ approximation.[494,495]
For the case study of HB238 as presented in section F.5.3, different bulk structure cutouts of varying size (dimer, tetramer, octamer) and orientation (X,Y,Z,XY,XZ,YZ) were taken from the experimental crystal structures of four different polymorphs of **d1a1**, CSD identifiers 2073437, 2073438, 2073461,

and 1856819, consecutively named **P1** to **P4**, see also Ref. [189]. UV/vis spectra were calculated at the time-dependent optimally tuned (OT) ωB97X-D4/def2-SVP(SMD(THF)) level of theory with ω=0.157 as optimized on the monomer and 50 calculated excitations using `ORCA` V5.0.4.[335,496] Charge transfer integrals were calculated at the ωB97X-D4/TZ2P level of theory employing the DIPRO method as implemented in `AMS` V2023.104.[405,406] Furthermore, electronic couplings were also calculated at the PTB level,[160] employing the DIPRO program as described in Ref. [5]. Data and molecule visualization and figure generation were performed with `gnuplot`, `molden`, `inkscape`, and `chimera`.[371,372,497,498]

F.5 Results and Discussion

The interaction energy of the tetramers was calculated according to

$$\Delta E = E_{tetr} - 4 \cdot E_{mon}$$

at various different levels of theory as given in section F.4. Statistical evaluations according to Equations 1-8 in Chapter 1 of the SI were performed on the energy differences between the reference method and the tested methods, whereas the flat and stacked subsets were consequently evaluated separately. In the following, we will first discuss the results of the D11A8MERO-16 subset for which PNO-LCCSD(T)-F12b/aug-cc-pVTZ+D,H=VTZ references were calculated to determine an appropriate DFT reference level. Afterwards, the results for the whole D11A8MERO benchmark will be presented with respect to PWPB95-D4/def2-TZVPP interaction energies. Finally, we show a short case study employing the introduced workflow for the experimentally important merocyanine **d1a1**.

F.5.1 Results for D11A8MERO16

16 systems from the D11A8MERO benchmark were selected for a representative subset, namely 4_f_d1a1_7, 4_f_d7a4_2, 4_f_d8a5_1, 4_f_d10a4_7, 4_f_d11a7_1, 4_f_d6a3_3, 4_f_d2a2_3, 4_f_d4a1_7, 4_s_d1a1_9, 4_s_d3a1_10, 4_s_d7a4_6, 4_s_d9a7_4, 4_s_d10a4_6, 4_s_d11a7_6, 4_s_d5a3_5, and 4_s_d9a8_1. The subset was chosen such that most donors and acceptors are covered. To check for residual basis set errors, we tested def2 basis sets from def2-SVP to def2-QZVPPD (see SI Chapter 3 for details). The interaction energies calculated with GGA functionals are essentially converged with a def2-TZVP basis. The difference in computation time between def2-TZVPP and def2-QZVP for double hybrid (DH) DFT amounts to a factor of about four, while the relative improvements in accuracy are relatively small compared to the coupled cluster reference (i.e., relative MAD improved by 1.3% and relative SD by 1.1%; see also SI Chapter 3). Thus, we decided for def2-TZVPP as basis set for the DH-DFT references, which is a computationally feasible setting also for the complete benchmark set.
Tab. F.1 lists error statistics for the performance of selected tested methods against PNO-LCCSD(T)-F12b/aug-cc-pVTZ+D,H=VTZ references in order to identify a reference method, which is still computational feasible for the complete D11A8MERO set (the complete statistics can be found in the SI Chapter 3). Due to the considerable size of the tetramers (up to 260 atoms) it was not possible to perform all coupled-cluster calculations with `tight` domain settings as recommended for NCI

systems[499] and also a CABS singles correction of the Hartree-Fock energies was computationally unfeasible due to the large memory requirements of the algorithm. Therefore, the coupled cluster references presented here have a slightly larger error than usual for systems with less than 150 atoms,[500] especially for the stacked tetramers, where these approximations are most noticeable. Nevertheless, to estimate the corresponding uncertainties, we still performed tight calculations including the CABS singles correction for one flat (conformer 3 of d6a3, ΔE=−21.8 kcal/mol) and one stacked (conformer 6 of d11a7, ΔE=−54.3 kcal/mol) structure (see SI) and calculated the respective CABS singles corrections. For the flat tetramer, the difference in interaction energy with `tight` and `default` settings is only −0.4 kcal/mol and the CABS singles correction contribution to ΔE is −0.7 kcal/mol. Taking into account the remaining basis set errors, we estimate the uncertainty of the calculated coupled cluster references for the flat systems to be ±1.5 kcal/mol. For the stacked system, the estimated uncertainty is about twice as large (ΔE(`tight`)-ΔE(`normal`): −2.7 kcal/mol, CABS singles correction contribution to ΔE: +3.5 kcal/mol). Additionally, we computed a counterpoise (CP) correction for the stacked tetramer to estimate the intermolecular basis set superposition error (BSSE) (see SI). The CP correction is quite small with 0.34 kcal/mol an hence totally insignificantly contributing to the errors in the reference values. Given the somewhat larger uncertainties of the coupled cluster references for the subset, several D4-corrected density functionals evaluated with the def2-TZVPP basis set would be suited. Finally, we decided for PWPB95-D4 as reference for the full benchmark, given its generally good performance for NCI systems.[139,501–504]

Interestingly, although the range-separated hybrid ωB97M-V is known for its overall good performance,[123,139,505] we observed poor results, especially for the stacked complexes. However, with ωB97M-D4 the results are excellent. This can be attributed to the VV10 dispersion correction that misses the ATM three-body dispersion and thus overstabilizes especially the stacked NCI complexes where this contribution is typically large. Future research will elaborate whether a scaled ATM term in addition to the VV10 contribution can be employed to cure this problem systematically.

Table F.1: Statistical measures (cf. Equations 1-8 in Chapter 1 of the SI) of interaction energies for the D11A8MERO16 subset calculated with different levels of theory using a def2-TZVPP basis. Reference values are obtained at the PNO-LCCSD(T)-F12b/aug-cc-pVTZ+D,H=VTZ level of theory. Except for ρ_P, all values are given in kcal/mol. The average interaction energy for the CC reference are -28.7 kcal/mol for the flat subset and -67.2 kcal/mol for the stacked subset.

	MD	MAD	SD	ρ_P
PWPB96-D4	−0.6	1.5	2.0	0.997
revDSD-PBEP86-D4	−3.5	3.5	2.3	0.998
ωB97M-V	−8.0	8.2	6.2	0.991
ωB97M-D4	−1.8	2.0	2.2	0.998
CAMB3LYP-D4	−4.5	4.5	1.8	0.998
PBE0-D4	−2.7	3.1	2.9	0.994
B3LYP-D4	−7.1	7.1	4.5	0.994

F.5.2 Results for D11A8MERO

Interaction energies for the complete D11A8MERO benchmark were calculated with various methods (see F.4), focusing on low-cost DFT and SQM methods, and evaluated against PWPB95-D4/def2-TZVPP reference values. The results for the stacked subset are shown in Figure F.6 bottom and Table F.3, the results for the flat subset are shown in Figure F.6 top and Table F.2. Statistics with alternative revDSD-PBEP86-D4/def2-TZVPP as reference method are given in the SI.

The tested low-cost DFT methods exhibit underbinding behavior compared to the reference, which can

Table F.2: Statistical measures (cf. Equations 1-8 in Chapter 1 of the SI) of interaction energy values calculated at different levels of theory for the flat half of the D11A8MERO benchmark and compared to PWPB95-D4/def2-TZVPP references. Absolute values are given in kcal/mol, relative ones in %. The average interaction energy for the DH-DFT reference is -23.5 kcal/mol for the flat subset.

flat	ωB97X-3c	PBEh-3c	r^{2}SCAN-3c	B97-3c	PBE-D4/TZVP
MD	2.5	2.5	2.5	2.1	−1.0
MAD	2.7	2.7	2.6	2.2	1.5
SD	2.1	2.3	2.2	2.1	1.6
relMD	−10.5	−9.4	−11.1	−7.2	5.0
relMAD	25.6	22.0	26.4	21.9	10.7
relSD	92.4	74.1	99.1	101.1	33.2
ρ_P	0.990	0.988	0.990	0.991	0.995
ρ_S	0.986	0.984	0.986	0.988	0.993
min $\delta\Delta E$	−4.9	−4.5	−1.7	−2.6	−7.6
max $\delta\Delta E$	10.0	8.9	12.6	13.4	5.9
$\delta\Delta E$ range	14.9	13.4	14.3	16.0	13.5

flat	GFN1-xTB	GFN2-xTB	PM6D3H4X	GFN-FF	MMFF94
MD	−1.4	−2.2	2.1	0.4	25.9
MAD	3.1	3.1	4.8	7.3	27.4
SD	4.3	3.9	6.8	9.1	21.8
relMD	6.2	5.3	−8.3	−6.5	−84.6
relMAD	28.1	25.0	26.2	52.8	127.7
relSD	100.2	99.1	59.6	128.6	270.6
ρ_P	0.959	0.966	0.897	0.825	0.075
ρ_S	0.971	0.971	0.937	0.814	−0.054
min $\delta\Delta E$	−17.4	−18.6	−12.7	−19.6	−30.7
max $\delta\Delta E$	16.9	11.5	33.4	24.3	105.9
$\delta\Delta E$ range	34.3	30.1	46.1	43.9	136.6

most likely be attributed to the smaller basis sets, especially for the stacked subset. The difference in accuracy (mean absolute deviation, MAD) between the stacked and flat subset amounts to 3-9 kcal/mol although the error scattering is very similar for both sets (about 2-4 kcal/mol in standard deviation, SD). The error span ($\delta\Delta E$ span) is about twice as large for the GFN methods (about 30 kcal/mol) and three times as large for PM6D3H4X and GFN-FF (about 45 kcal/mol) for the flat subset compared

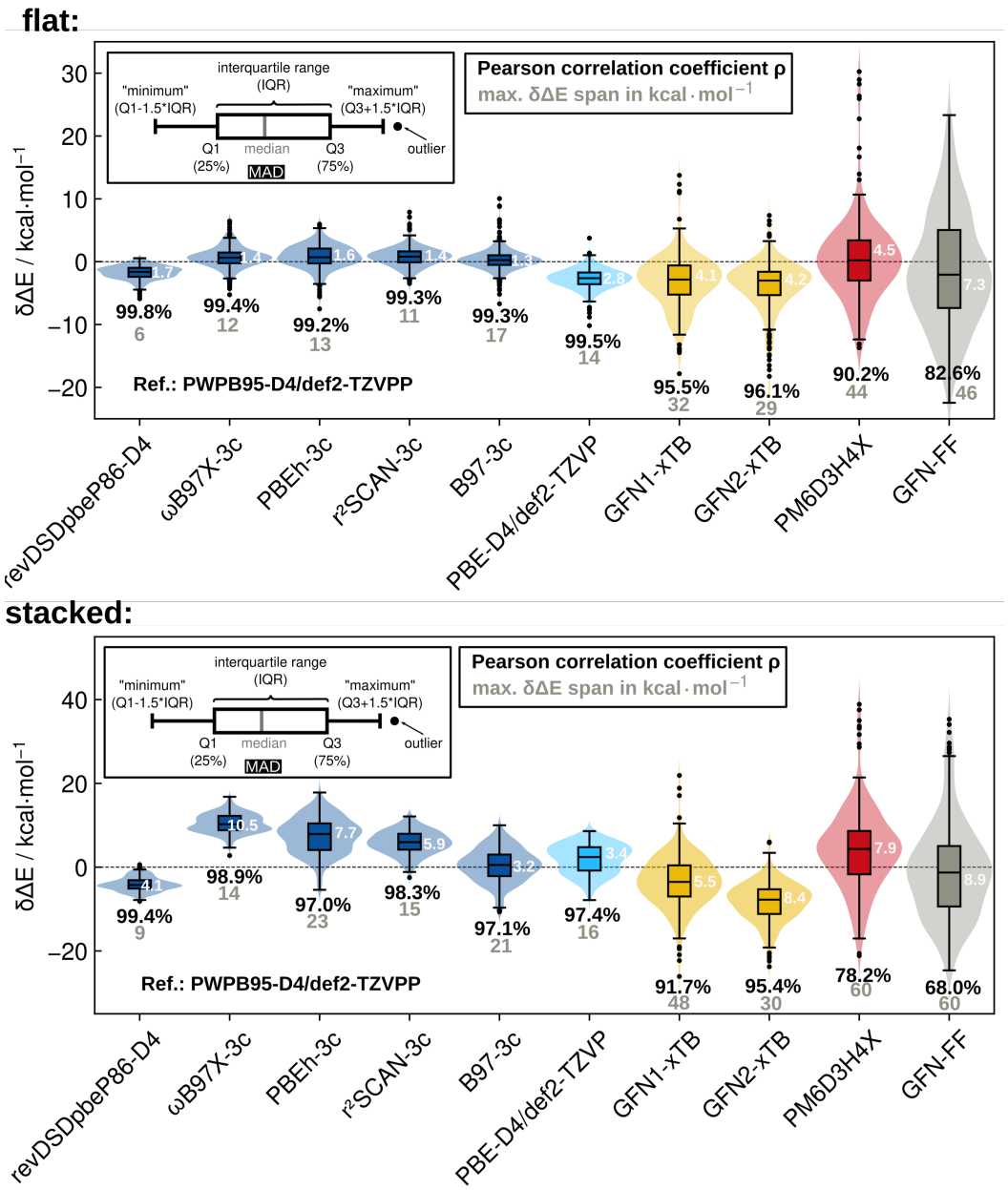

Figure F.6: Violin plot of the error statistics for the interaction energies of the flat (top) and stacked (bottom) subsets of the D11A8MERO benchmark set w.r.t. PWPB95-D4/def2-TZVPP reference values. MMFF94 results are not shown as they do not fit into the given y-axis span. Methods are roughly ordered according to their level of theory, from left to right: (low-cost) DFT methods, SQM methods, and force fields. Negative values of $\delta\Delta E$ indicate overbinding compared to the reference, positive values underbinding. Grey values indicate the $\delta\Delta E$ span. The average interaction energy for the DH-DFT reference is -23.5 kcal/mol for the flat subset and -70.4 kcal/mol for the stacked subset, respectively.

Table F.3: Statistical measures (cf. Equations 1-8 in Chapter 1 of the SI) of interaction energy values calculated at different levels of theory for the stacked half of the D11A8MERO benchmark and compared to PWPB95-D4/def2-TZVPP references. Absolute values are given in kcal/mol, relative ones in %. The average interaction energy for the DH-DFT reference is -70.4 kcal/mol for the stacked subset.

stacked	ωB97X-3c	PBEh-3c	r^2SCAN-3c	B97-3c	PBE-D4/TZVP
MD	14.6	11.6	10.0	4.6	6.1
MAD	14.6	11.6	10.0	4.9	6.1
SD	2.3	3.9	3.1	3.6	3.3
relMD	−20.2	−16.5	−13.7	−6.6	−8.2
relMAD	20.2	16.5	13.7	6.9	8.2
relSD	4.5	6.6	4.6	5.1	4.3
ρ_P	0.988	0.975	0.979	0.977	0.976
ρ_S	0.974	0.951	0.954	0.951	0.954
min $\delta\Delta E$	9.0	0.7	2.1	−4.4	−1.2
max $\delta\Delta E$	20.9	22.7	16.3	12.5	13.7
$\delta\Delta E$ range	11.9	22.0	14.2	16.9	14.9

stacked	GFN1-xTB	GFN2-xTB	PM6D3H4X	GFN-FF	MMFF94
MD	0.6	−4.2	8.1	3.1	76.6
MAD	4.7	4.9	10.2	8.8	76.6
SD	6.3	4.8	9.9	11.2	17.4
relMD	0.0	6.2	−9.7	−2.6	−105.6
relMAD	6.5	7.1	13.2	11.5	105.6
relSD	8.8	7.2	12.3	14.5	27.2
ρ_P	0.914	0.95	0.762	0.692	0.541
ρ_S	0.873	0.917	0.679	0.605	0.484
min $\delta\Delta E$	−22.4	−22.5	−17.9	−18.6	3.4
max $\delta\Delta E$	26.7	10.9	43.9	40.2	124.9
$\delta\Delta E$ range	49.1	33.4	61.8	58.8	121.5

to the DFT methods (about 15 kcal/mol). For the stacked subset, these values are slightly larger but the absolute interaction energies are also larger. The best performing method in terms of SD is ωB97X-3c with only 2.3 kcal/mol and 1.8 kcal/mol, respectively. Unexpectedly, the cheapest among all tested composite methods B97-3c performs significantly better that the tested hybrid and RSH composite methods with MADs of only 1.3 and 3.2 kcal/mol. Likely, this is due to some beneficial error compensation, but this can not be traced back to a specific part of B97-3c as composite methods are designed in such a way that the introduced corrections are optimized together for overall best performance. As shown in Fig. F.6, the tested GGA methods perform better for the relatively uniform part **d1-6**x**a1-3** of the stacked subset, whereas the interaction energies for the more complicated part of **d7-11**x**a4-8** are mostly underestimated. The tested GFN methods do not exhibit such a behavior but rather yield a uniform description of all stacked tetramers. The GFN methods are known to overbind large π-systems dominated by van der Waals interactions,[212] however that is merely pronounced for the D11A8MERO benchmark, even for the stacked subset. The slightly different performance of the two GFN methods may be due to their different partial charge models which are essential for the description of an inherently dipolar system or also maybe due to the different treatment of dispersion in both methods. GFN2-xTB is a reasonable choice for the considered molecules, with a *Pearson* correlation coefficient ρ_P of 0.961 respective 0.954 and a SD of 4.2 and 4.6 kcal, which is comparable with the obtained SDs for the tested composite DFT. PM6D3H4X performs worst among the tested SQM methods and cannot be recommended for such systems. For a force field, GFN-FF performs relatively good with a ρ_P of 0.680 for the stacked set and 0.826 for the flat set and SDs of 11.4 and 9.0 kcal/mol, and may be a suitable choice for the evaluation of very large systems. MMFF94, however, is not applicable at all, since it yields a poor and uncorrelated description of the interactions, especially in the flat systems. Most of the depicted outliers can be attributed to certain molecules and structure motifs. All tested SQM methods struggle to correctly describe **d11a5** in the stacked as well as in the flat subset. This molecule is always strongly underbound compared to the reference. **d11a5** is the most sulfur-rich compound in the benchmark. In previous studies[270] we already discovered that such motifs may be challenging for SQM methods. Additionally, **d10a4** is the second major outlier for the SQM methods, exhibiting strong overbinding in the stacked subset. Interestingly, this does not hold true for GFN2-xTB, which struggles with **d10a8** instead. As there are no difficulties with the description of **d7a4** and **d9a8**, the donor moiety **d10** will likely be the problematic structure motif. It comprises a five-ring heterocycle with three hetero-atoms, the maximum in our benchmark. Despite **d11a5** and **d10** there is a second type of outliers not defined by their chemical composition but their spatial conformational arrangement. These outliers occur only in the flat subset but for both the DFT and SQM methods and exhibit strong underbinding. They are characterized by a shamrock-like arrangement of tetramers with a rather low dipole moment and high symmetry. However, we were not able to identify a detailed reason for these outliers (for more details see SI, Chapter 5).

F.5.3 Case Study for HB238

For all example applications, i.e., charge transport, optical properties, cluster convergence, we focused on **d1a1**, also known under the name HB238.[472] HB238 shows good performance as a transistor, yields smooth and well processable thin films, and exhibits a two-dimensional charge-transport network in the crystal, which are all beneficial for device fabrication. This molecule has widely been used for all kinds of model investigations, computational[189,473] and experimental ones.[471,474,475]

One of the most important properties for OE materials is charge transport. The majority of merocyanine materials exhibit hopping type transport and one suitable description for that is the electron transfer theory by Marcus.[185] The important charge transport parameters within Marcus' theory are electronic coupling, site energies, and reorganization energy. Electronic couplings can for example be calculated with the dimer projection method (DIPRO)[270] via a projected fragment molecular overlap. Site energies and reorganization energy are usually calculated with other approaches. To estimate charge transport for the investigated aggregates, we must determine the appropriate fragment size for converged results. Potentially, we could use it as a secondary condition for aggregate generation (besides interaction energy) to explore other parts of the conformational landscape more relevant to conducting materials. To test the convergence behavior with number of stacked monomers, we generated highly symmetrical model systems up to octamer size and calculated their charge transfer integrals at the PTB@DIPRO and ωB97X-D4/TZ2P@DIPRO levels of theory which were shown to yield reliable results for merocyanines in previous studies.[270] Flat aggregates were not investigated further, neither for charge transport nor optical spectra, as their electronic coupling is negligible.

Figure F.7 (left) shows that only vertical nearest neighbor couplings are relevant for charge transport

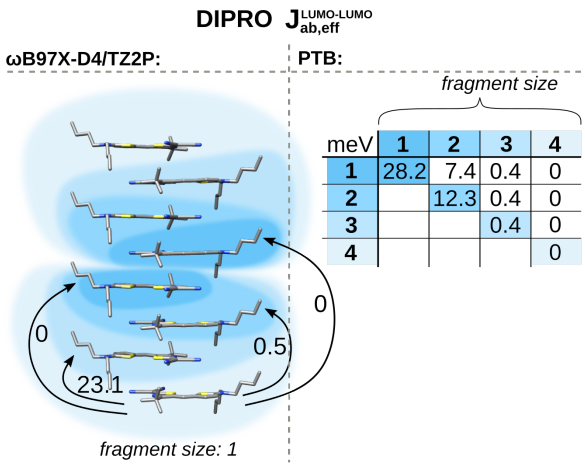

Figure F.7: Electron transfer integrals (LUMO-LUMO) of fragments with sizes varying from monomer to tetramer of a highly symmetrical **d1a1** octamer with up to 472 atoms (right) calculated at the PTB@DIPRO level of theory in meV, and electron transfer integrals of increasingly distant monomeric fragments of dimer cutouts (118 atoms) of the same **d1a1** octamer (left) at the ωB97X-D4/TZ2P level of theory in meV. PTB and ωB97X-D4/TZ2P@DIPRO results differ only slightly from each other (up to 5 meV).

in **d1a1**. Figure F.7 (right) shows that fragments up to a size of dimers each (i.e., in total a tetramer) should be considered in the calculation. Already the coupling between a monomer and a trimer fragment is negligible, as well as trimer with trimer and all larger fragments. Concluding, to obtain reliable qualitative results for electronic properties of merocyanines in the bulk, such as charge transport, the consideration of tetrameric aggregates is sufficient. This is in accordance with Refs.[189] and[506].

Next, we investigate the convergence of optical properties with respect to the number of stacked **d1a1** monomers. When using clusters to approximate bulk optical properties, the choice of the cluster morphology is highly important. According to the good practice of deriving reasonable cutouts,[288,507] there are two options for valid cutouts from **P3**, here named **A** (antiparallel angled stacked dimer)

and **B** (antiparallel cofacially stacked dimer) shown in Chapter 4 of the SI. We investigated for both motifs how the corresponding UV/vis spectra change with increasing cutout size in all three directions. Exemplary results are shown in Figure F.8 for motif **A** and more extensively in Chapter 4 of the SI.
Peak splitting only occurs upon existence of motif **A**. Formally, one of these peaks is red-shifted

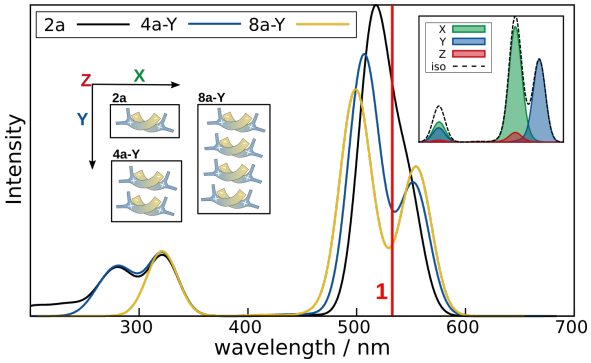

Figure F.8: (Angle resolved) UV/vis spectra of differently sized type **A** cutouts from the **P3** polymorph of **d1a1** at the OT-ωB97X-D4/def2-SVP(SMD(THF)) level of theory. The red **1** denotes the position of the monomer excitation.

and the other one blue-shifted compared to the monomer. In OE material research, this behavior is attributed to J- and H-aggregates, slipped-stacked or parallel arrangements of transition dipoles.[508] H- and J-transitions belong to the same type of excitation differing in their polarizations and energies. E.g., motif **B** is an H-aggregate, underlined by its more pronounced blue-shift upon stacking in y-direction, whereas motif **A** is a J-aggregate. The insert in Fig. F.8 clearly shows the different polarizations of the H- and J-peak. The intensity of the y-peak is proportional to the amount of **A**-motifs that exhibit a ten times stronger transition dipole moment for this excitation than the **B**-motif. In general, the intensity of the angle-resolved peaks (Fig. F.8 inset) increases, if the cluster is elongated in the respective direction.
Although we can observe qualitative spectral trends from dimer up to octamer size (472 atoms), we are not yet able to determine the cluster size for converged optical properties. Computation of spectra for such large systems at hybrid DFT level is computationally very demanding, and larger systems than the investigated ones are currently hardly feasible at this level of theory. A lower theoretical level is no alternative as Fock exchange is absolutely needed to describe charge-transfer states correctly. However, we found that at least tetramers are needed to obtain the correct qualitative description of the basic spectral features. This also allows to assign one of the investigated orientations, i.e., **8a-YZ** (see SI Chapter 4) with the viewing direction along the molecular y-axis, to the experimentally measured thin film structure.[294]

F.6 Conclusion and Outlook

We presented a quantum-chemistry-based workflow for the generation of monolayers and aggregates, using the automated interaction site screening algorithm aISS supplemented by additional constraining and directing potentials to ensure layer-wise and columnar growth. This workflow was then exemplarily applied to generate the novel D11A8MERO benchmark set, composed out of different merocyanine

backbones for the donor and acceptor moieties. 28 conformational ensembles of flat and stacked tetramers were generated, in total 592 molecules with system sizes between 124 and 260 atoms. The respective ensembles contain almost all structure motifs from real crystals, only missing strongly slipped stacked arrangements due to the restrictions of the docking algorithm.
Interaction energies for this test set were calculated at various theoretical levels with a focus on low-cost DFT and SQM methods, evaluated against PWPB95-D4/def2-TZVPP reference energies for forming tetramers, verified against high-level PNO-LCCSD(T)-F12b/aug-cc-pVTZ+D(H=VTZ) values. In general among the SQM methods, GFN2-xTB performs best and robustly across the whole test set, reaching mGGA/TZ accuracy at two orders of magnitude less computational costs. Thus, we recommend GFN2-xTB for such systems, as it presents a good balance between a high quality theoretical description, robustness, and computational efficiency. The successor method GP3-xTB, which is currently developed in our group, also shows very promising preliminary results on this test set. Among all tested low-cost DFT methods, the best performing one in terms of accuracy is B97-3c. ωB97X-3c exhibits the lowest SD but yielded a rather large MAD for the stacked subset. We observe a systematic overbinding for the stacked aggregates for most of the tested DFT methods and small tendencies to underbind for most of the tested SQM methods.
Furthermore, we investigated how optical and electrical properties of the example merocyanine **d1a1** converge with the cluster size. DIPRO results for the hole coupling integrals showed that not more than two molecules per fragment in a two-fragment-evaluation must be considered, i.e., tetramers. This finding was corroborated by angle resolved UV/vis spectra at the OT-ωB97X-D4/def2-SVP(SMD(THF)) level of theory. Although no final convergence with the cluster size was reached yet, we were nevertheless able to identify important and fundamental trends that occur upon crystal growth in different directions, in accordance with previous literature.[189,288,506,507]
As the incomplete generation of J-aggregates only became apparent in the course of the evaluation, future efforts will include a modified runtype for the presented workflow also enabling the generation of slipped-stacked J-aggregates. This is especially challenging as we try to approximate bulk properties, governed by long-range interactions, with a molecular bottom-up approach. Additionally, an alternative secondary condition for the docking algorithm is anticipated, e.g., surface coverage, or dipole moment instead of only the interaction energy, to explore more areas of the conformational space. In general, the D11A8MERO benchmark set could be extended. So far, only 28 out of 88 possible merocyanine backbones have been used and the variation of side chains has not even been fully accounted for yet. With the insights gained from this benchmark study and even better SQM methods, like GP3-xTB which is currently developed in our group, it will be possible to generate even more reliable data for merocyanine aggregates. That would enable the use of machine-learning based approaches to map not only the structure-property relationships across aggregates but also to reduce the complexity for an aggregate-to-bulk learning of e.g., optical properties and crystal structure prediction.
In conclusion, our work provides a robust and efficient workflow for the systematic and reliable generation of monolayers and aggregates. The D11A8MERO benchmark offers valuable insights into challenging NCI systems and can serve as an advanced test case for further method development.

F.7 Supplementary Material

The electronic supplementary information include:
i) the molecular structures of the D11A8MERO benchmark set and the D11A8MERO16 subset as .xyz files and .res files with the references
ii) all used .cif structure files and additional .xyz files for the case study of **d1a1**
iii) the raw data for the generation of statistics and pictures as .ods tables
iv) PDF with additional figures, tables and investigations useful for understanding and more in-depth knowledge.

F.8 Acknowledgments

The authors thank the *German Research Foundation* (DFG) for funding in the framework of *RTG 2591 - Template-Designed Organic Electronics (TIDE)*. The authors gratefully acknowledge the granted access to the MARVIN cluster hosted by the University of Bonn. Finally, we thank Prof. Dr. Daniele Fazzi from the University of Bologna and Prof. Dr. Klaus Meerholz from the University of Cologne for fruitful discussions and support.

F.9 Author Declarations

F.9.1 Conflict of Interest

The authors have no conflicts to disclose.

F.10 Author Contributions

J. K. performed the computational analyses and wrote the manuscript. A. H. conceptualized and supervised the work and performed all coupled cluster calculations. S. G. and A. H. acquired funding. All authors contributed to rationalize the data and correct the manuscript.

F.11 Data Availability Statement

The data that support the findings of this study are available within the article and its supplementary material.

Bibliography

[1] J. T. Kohn, S. Spicher, M. Bursch, and S. Grimme, *Quickstart guide to model structures and interactions of artificial molecular muscles with efficient computational methods*, Chem. Comm. **58** (2022) 258, DOI: 10.1039/d1cc05759f.

[2] J. S. Wössner, J. T. Kohn, D. Wassy, M. Hermann, S. Grimme, and B. Esser, *Increased Antiaromaticity through Pentalene Connection in [n] Cyclo-1, 5-dibenzopentalenes*, Org. Lett. **24** (2022) 983, DOI: 10.1021/acs.orglett.1c03900.

[3] P. Krämer, J. Kohn, D. A. Hofmeister, M. Kersten, C. Sterzenbach, A. Gres, A. Hansen, S.-S. Jester, S. Grimme, and S. Höger, *Size-Increased All-Phenylene Molecular Spoked Wheels–A Combined Theoretical and Experimental Approach*, Angew. Chem. Int. Ed. (2024) e202411092, DOI: 10.1002/anie.202411092.

[4] J. T. Kohn, H. Li, A. M. Evans, J.-L. Bredas, and S. Grimme, *Quantum Chemistry Insight into the Multifaceted Structural Properties of Two-Dimen-sional Covalent Organic Frameworks*, Chem. Mater. **35** (2023) 2820, DOI: 10.1021/acs.chemmater.2c03555.

[5] J. T. Kohn, N. Gildemeister, S. Grimme, D. Fazzi, and A. Hansen, *Efficient calculation of electronic coupling integrals with the dimer projection method via a density matrix tight-binding potential*, J. Chem. Phys. **159** (2023), DOI: 10.1063/5.0167484.

[6] J. T. Kohn, S. Grimme, and A. Hansen, *A Semi-Automated Quantum-Mechanical Workflow for the Generation of Merocyanine Monolayers and Aggregates for Organic Electronics*, J. Chem. Phys. **161** (2024), DOI: 10.1063/5.0230341.

[7] T. A. Schaub, E. A. Prantl, J. T. Kohn, M. Bursch, C. R. Marshall, E. J. Leonhardt, T. C. Lovell, L. N. Zakharov, C. K. Brozek, S. R. Waldvogel, et al., *Exploration of the solid-state sorption properties of shape-persistent macrocyclic nanocarbons as bulk materials and small aggregates*, J. Am. Chem. Soc. **142** (2020) 8763, DOI: 10.1021/jacs.0c01117.

[8] M. Bursch, A. Hansen, P. Pracht, J. T. Kohn, and S. Grimme, *Theoretical study on conformational energies of transition metal complexes*, Phys. Chem. Chem. Phys. **23** (2021) 287, DOI: 10.1039/d0cp04696e.

[9] M. Hermann, D. Wassy, J. T. Kohn, P. Seitz, M. U. Betschart, S. Grimme, and B. Esser, *Chiral Dibenzopentalene-Based Conjugated Nanohoops through Stereoselective Synthesis*, Angew. Chem. Int. Ed. **60** (2021) 10680, DOI: 10.1002/anie.202016968.

[10] T. J. Keller, C. Sterzenbach, J. Bahr, T. L. Schneiders, M. Bursch, J. Kohn, T. Eder, J. M. Lupton, S. Grimme, S. Höger, et al., *Nanopatterns of molecular spoked wheels as giant homologues of benzene tricarboxylic acids*, Chem. Sci. **12** (2021) 9352.

[11] J. T. Kohn, M. Bursch, A. Hansen, and S. Grimme, *Computational study of ground-state properties of μ 2-bridged group 14 porphyrinic sandwich complexes*, J. Comp. Chem. **44** (2023) 229, DOI: 10.1002/jcc.26870.

[12] S. Rickert, S.-X. L. Luo, J. Bahr, J. T. Kohn, M. Xue, A. Hansen, S. Grimme, S.-S. Jester, T. M. Swager, and S. Höger, *Flexible Phenanthracene Nanotubes for Explosive Detection*, J. Am. Chem. Soc. **146** (2024) 2986, DOI: 10.1021/jacs.3c08131.

[13] K. Gratzfeld, A. J. Kny, J. T. Kohn, A. Hansen, S. Grimme, K. Wandelt, and M. Sokolowski, *Tuning the Adsorption Geometry of Structures of Perylene Tetracarboxylic Acid on the Au(111) Electrode by the Potential* (*in preparation*).

[14] S. Ogawa, *Organic Electronics Materials and Devices*, Springer, 2015.

[15] H. Klauk, *Organic Electronics: Materials, Manufacturing and Applications*, vol. 2, 2006 1325.

[16] S. R. Forrest, *Organic electronics: foundations to applications*, Oxford University Press, USA, 2020.

[17] M. C. Petty, *Organic and molecular electronics: from principles to practice*, John Wiley & Sons, 2019.

[18] O. Inganäs, *Organic photovoltaics over three decades*, Adv. Mater. **30** (2018) 1800388.

[19] K. A. Mazzio and C. K. Luscombe, *The future of organic photovoltaics*, Chem. Soc. Rev. **44** (2015) 78.

[20] Y.-W. Su, S.-C. Lan, and K.-H. Wei, *Organic photovoltaics*, Mater. Today **15** (2012) 554.

[21] G. Zhang, F. R. Lin, F. Qi, T. Heumüller, A. Distler, H.-J. Egelhaaf, N. Li, P. C. Chow, C. J. Brabec, A. K.-Y. Jen, et al., *Renewed prospects for organic photovoltaics*, Chem. Rev. **122** (2022) 14180.

[22] G. Hong, X. Gan, C. Leonhardt, Z. Zhang, J. Seibert, J. M. Busch, and S. Bräse, *A brief history of OLEDs—emitter development and industry milestones*, Adv. Mater. **33** (2021) 2005630.

[23] B. Geffroy, P. Le Roy, and C. Prat, *Organic light-emitting diode (OLED) technology: materials, devices and display technologies*, Polym. Int. **55** (2006) 572.

[24] X. Ren, F. Yang, X. Gao, S. Cheng, X. Zhang, H. Dong, and W. Hu, *Organic field-effect transistor for energy-related applications: low-power-consumption devices, near-infrared phototransistors, and organic thermoelectric devices*, Adv. Energy Mater. **8** (2018) 1801003.

[25] G. Horowitz, *Organic field-effect transistors*, Adv. Mater. **10** (1998) 365.

[26] C. W. Tang and S. A. VanSlyke, *Organic electroluminescent diodes*, Appl. Phys. Lett. **51** (1987) 913.

[27] H. Koezuka, A. Tsumura, and T. Ando, *Field-effect transistor with polythiophene thin film*, Synth. Met. **18** (1987) 699.

[28] J. E. Carlé and F. C. Krebs, *Technological status of organic photovoltaics (OPV)*, Sol. Energy Mater. Sol. Cells **119** (2013) 309.

[29] M. Moser, A. Wadsworth, N. Gasparini, and I. McCulloch, *Challenges to the success of commercial organic photovoltaic products*, Adv. Energy Mater. **11** (2021) 2100056.

[30] H. Tian and F. Meng, "Solar cells based on cyanine and polymethine dyes," *Organic Photovoltaics*, CRC Press, 2017 313.

[31] Z. Chen, F. Li, and C. Huang, *Organic D-π-A dyes for dye-sensitized solar cell*, Curr. Org. Chem. **11** (2007) 1241.

[32] G. Li, R. Zhu, and Y. Yang, *Polymer solar cells*, Nat. Photonics **6** (2012) 153.

[33] T. Liu, T. Yang, R. Ma, L. Zhan, Z. Luo, G. Zhang, Y. Li, K. Gao, Y. Xiao, J. Yu, et al., *16% efficiency all-polymer organic solar cells enabled by a finely tuned morphology via the design of ternary blend*, Joule **5** (2021) 914.

[34] M. Wang, R. Dong, and X. Feng, *Two-dimensional conjugated metal–organic frameworks (2D c-MOFs): chemistry and function for MOFtronics*, Chem. Soc. Rev. **50** (2021) 2764.

[35] G. L. Goh, H. Zhang, T. H. Chong, and W. Y. Yeong, *3D printing of multilayered and multimaterial electronics: a review*, Adv. Electron. Mater. **7** (2021) 2100445.

[36] E. Dimitriou and N. Michailidis, *Printable conductive inks used for the fabrication of electronics: an overview*, Nanotechnol. **32** (2021) 502009.

[37] Y. Ke, J. Chen, G. Lin, S. Wang, Y. Zhou, J. Yin, P. S. Lee, and Y. Long, *Smart windows: electro-, thermo-, mechano-, photochromics, and beyond*, Adv. Energy Mater. **9** (2019) 1902066.

[38] J. Heikenfeld, P. Drzaic, J.-S. Yeo, and T. Koch, *A critical review of the present and future prospects for electronic paper*, J. Soc. Inf. Disp. **19** (2011) 129.

[39] Y. Chen and C. Wang, *Designing high performance organic batteries*, Acc. Chem. Res. **53** (2020) 2636.

[40] J. C. Costa, R. J. Taveira, C. F. Lima, A. Mendes, and L. M. Santos, *Optical band gaps of organic semiconductor materials*, Opt. Mater. **58** (2016) 51.

[41] J. Song, H. Lee, E. G. Jeong, K. C. Choi, and S. Yoo, *Organic light-emitting diodes: pushing toward the limits and beyond*, Adv. Mater. **32** (2020) 1907539.

[42] T.-P. Nguyen, *Defects in organic electronic devices*, Phys. Status Solidi **205** (2008) 162.

[43] M. O'Neill and S. M. Kelly, *Ordered materials for organic electronics and photonics*, Adv. Mater. **23** (2011) 566.

[44] S. S. Lee and Y.-L. Loo, *Structural complexities in the active layers of organic electronics*, Annu. Rev. Chem. Biomol. Eng. **1** (2010) 59.

[45] M. Fahlman, S. Fabiano, V. Gueskine, D. Simon, M. Berggren, and X. Crispin, *Interfaces in organic electronics*, Nat. Rev. Mater. **4** (2019) 627.

[46] A. Troisi, G. Orlandi, and J. E. Anthony, *Electronic interactions and thermal disorder in molecular crystals containing cofacial pentacene units*, Chem. Mater. **17** (2005) 5024.

[47] W. Zhang, J. H. Bombile, A. R. Weisen, R. Xie, R. H. Colby, M. J. Janik, S. T. Milner, and E. D. Gomez, *Thermal fluctuations lead to cumulative disorder and enhance charge transport in conjugated polymers*, Macromol. Rapid Commun. **40** (2019) 1900134.

[48] J. W. Steed and J. L. Atwood, *Supramolecular chemistry*, John Wiley & Sons, 2022.

[49] J.-M. Lehn, *Towards complex matter: supramolecular chemistry and self-organization*, Eur. Rev. **17** (2009) 263.

[50] E. Mattia and S. Otto, *Supramolecular systems chemistry*, Nat. Nanotechnol. **10** (2015) 111.

[51] I. V. Kolesnichenko and E. V. Anslyn, *Practical applications of supramolecular chemistry*, Chem. Soc. Rev. **46** (2017) 2385.

[52] M. J. Webber and R. Langer, *Drug delivery by supramolecular design*, Chem. Soc. Rev. **46** (2017) 6600.

[53] J.-M. Lehn, "Supramolecular chemistry—molecular information and the design of supramolecular materials," *Makromolekulare Chemie. Macromolecular Symposia*, vol. 69, 1, Wiley Online Library, 1993 1.

[54] X. Y. Ling, D. N. Reinhoudt, and J. Huskens, *From supramolecular chemistry to nanotechnology: assembly of 3D nanostructures*, Pure Appl. Chem. **81** (2009) 2225.

[55] C. J. Pedersen, *Cyclic polyethers and their complexes with metal salts*, J. Am. Chem. Soc. **89** (1967) 7017.

[56] J. Lehn, "Cryptates: inclusion complexes of macropolycyclic receptor molecules," *Chemistry for the Welfare of Mankind*, Elsevier, 1979 871.

[57] J. Li, *Chemistry of Buckybowl from Closed-Shell to Open-Shell*, Illinois Institute of Technology, 2019.

[58] R. Semino, J. C. Moreton, N. A. Ramsahye, S. M. Cohen, and G. Maurin, *Understanding the origins of metal–organic framework/polymer compatibility*, Chem. Sci. **9** (2018) 315.

[59] Feynman, R. P.; et. al., There's plenty of room at the bottom, *CIT Engineering and Science Magazine* **1960**.

[60] A. Cnossen, J. C. Kistemaker, T. Kojima, and B. L. Feringa, *Structural dynamics of overcrowded alkene-based molecular motors during thermal isomerization*, J. Org. Chem. **79** (2014) 927.

[61] M. von Delius, E. M. Geertsema, D. A. Leigh, and D.-T. D. Tang, *Design, synthesis, and operation of small molecules that walk along tracks*, J. Am. Chem. Soc. **132** (2010) 16134.

[62] C.-X. Zhao, Q. Zhang, G. London, and D.-H. Qu,
Functional Rotaxanes: From Synthetic Methodology to Functional Molecular Materials,
Handbook of Macrocyclic Supramolecular Assembly (2020) 277.

[63] J. F. Stoddart,
Mechanically interlocked molecules (MIMs)-molecular shuttles, switches, and machines,
Angew. Chem. Int. Ed **56** (2017) 11094.

[64] G. Gil-Ramírez, D. A. Leigh, and A. J. Stephens,
Catenane: fünfzig Jahre molekulare Verschlingungen, Angew. Chem. **127** (2015) 6208.

[65] A. Ol'ga, D. G. Samsonenko, and V. P. Fedin, *Supramolecular chemistry of cucurbiturils*,
Russ. Chem. Rev. **71** (2002) 741.

[66] R. S. Forgan, J.-P. Sauvage, and J. F. Stoddart,
Chemical topology: complex molecular knots, links, and entanglements,
Chem. Rev. **111** (2011) 5434.

[67] S. D. Fielden, D. A. Leigh, and S. L. Woltering, *Molecular knots*,
Angew. Chem. Int. Ed. **56** (2017) 11166.

[68] S. A. Meißner, T. Eder, T. J. Keller, D. A. Hofmeister, S. Spicher, S.-S. Jester, J. Vogelsang,
S. Grimme, J. M. Lupton, and S. Höger, *Nanoscale π-conjugated ladders*,
Nat. Commun. **12** (2021) 6614.

[69] C. M. Gangemi, R. Puglisi, A. Pappalardo, and G. T. Sfrazzetto,
Supramolecular complexes for nanomedicine, Bioorg. Med. Chem. Lett. **28** (2018) 3290.

[70] K. Ariga, *Mechano-Nanoarchitectonics: Design and Function*,
Small Methods **6** (2022) 2101577.

[71] R. Kumar, C. Seth, V. Kaliginedi, and R. Venkatramani,
A conceptual framework for designing and analyzing complex molecular circuits,
J. Mater. Chem. C **11** (2023) 14680.

[72] M. N. Stojanovic, D. Stefanovic, and S. Rudchenko, *Exercises in molecular computing*,
Acc. Chem. Res. **47** (2014) 1845.

[73] J. Singh, H. Kim, and K.-W. Chi, *Non-Covalent Interaction-Directed Coordination-Driven
Self-Assembly of Non-Trivial Supramolecular Topologies*, Chem. Rec. **21** (2021) 574.

[74] T. Saito, *Computational materials design*, vol. 34, Springer Science & Business Media, 2013.

[75] A. R. Oganov, G. Saleh, and A. G. Kvashnin, *Computational materials discovery*,
Royal Society of Chemistry, 2018.

[76] G. Bozzolo, R. D. Noebe, and P. B. Abel,
Applied computational materials modeling: theory, simulation and experiment,
Springer Science & Business Media, 2007.

[77] S. Datta and J. P. Davim, *Computational Approaches to Materials Design: Theoretical and
Practical Aspects: Theoretical and Practical Aspects*, IGI Global, 2016.

[78] J. H. Panchal, S. R. Kalidindi, and D. L. McDowell,
Key computational modeling issues in integrated computational materials engineering,
Comput. Aided Des. **45** (2013) 4.

[79] L. A. Curtiss and M. S. Gordon,
Computational materials chemistry: methods and applications,
Springer Science & Business Media, 2004.

[80] J. Hafner, *Atomic-scale computational materials science*, Acta Mater. **48** (2000) 71.

[81] K. T. Butler, J. M. Frost, J. M. Skelton, K. L. Svane, and A. Walsh,
Computational materials design of crystalline solids, Chem. Soc. Rev. **45** (2016) 6138.

[82] J. Hoja, A. M. Reilly, and A. Tkatchenko, *First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure*, Comput. Mol. Sci. **7** (2017) e1294.

[83] P. Biswas, D. Tafen, F. Inam, B. Cai, and D. Drabold,
Materials modeling by design: applications to amorphous solids,
J. Phys. Condens. Matter **21** (2009) 084207.

[84] T. M. Boland and A. K. Singh,
Computational synthesis of 2D materials: A high-throughput approach to materials design,
Comput. Mater. Sci. **207** (2022) 111238.

[85] A. Rege, *A perspective on methods to computationally design the morphology of aerogels*,
Adv. Eng. Mater. **25** (2023) 2201097.

[86] M. G. Corradini and M. A. Rogers,
Molecular gels: improving selection and design through computational methods,
Curr. Opin. Food Sci. **9** (2016) 84.

[87] A. Gooneie, S. Schuschnigg, and C. Holzer,
A review of multiscale computational methods in polymeric materials, Polymers **9** (2017) 16.

[88] M. Bauchy, *Deciphering the atomic genome of glasses by topological constraint theory and molecular dynamics: a review*, Comput. Mater. Sci. **159** (2019) 95.

[89] C. Cavallotti, M. Di Stanislao, D. Moscatelli, and A. Veneroni, *Materials computation towards technological impact: the multiscale approach to thin films deposition*,
Electrochim. Acta **50** (2005) 4566.

[90] C. Altintas, G. Avci, H. Daglar, A. Nemati Vesali Azar, S. Velioglu, I. Erucar, and S. Keskin,
Database for CO2 separation performances of MOFs based on computational materials screening, Appl. Mater. Interfaces **10** (2018) 17257.

[91] Z.-W. Zhao, O. H. Omar, D. Padula, Y. Geng, and A. Troisi, *Computational identification of novel families of nonfullerene acceptors by modification of known compounds*,
J. Phys. Chem. Lett. **12** (2021) 5009.

[92] Y. Wang, K. J. Stebe, C. de la Fuente-Nunez, and R. Radhakrishnan,
Computational design of peptides for biomaterials applications,
Appl. Bio Mater. **7** (2023) 617.

[93] Z.-C. Mu, Y.-L. Tan, J. Liu, B.-G. Zhang, and Y.-Z. Shi,
Computational modeling of DNA 3D structures: From dynamics and mechanics to folding,
Molecules **28** (2023) 4833.

[94] T. Le, V. C. Epa, F. R. Burden, and D. A. Winkler,
Quantitative structure–property relationship modeling of diverse materials properties,
Chem. Rev. **112** (2012) 2889.

[95] A. M. Mroz, V. Posligua, A. Tarzia, E. H. Wolpert, and K. E. Jelfs, *Into the unknown: how computation can help explore uncharted material space*, J. Am. Chem. Soc. **144** (2022) 18730.

[96] C. Kuehmann and G. Olson, *Computational materials design and engineering*, Mater. Sci. Technol. **25** (2009) 472.

[97] A. R. Oganov, C. J. Pickard, Q. Zhu, and R. J. Needs, *Structure prediction drives materials discovery*, Nat. Rev. Mater. **4** (2019) 331.

[98] J. Hafner, C. Wolverton, and G. Ceder, *Toward computational materials design: the impact of density functional theory on materials research*, MRS Bull. **31** (2006) 659.

[99] A. M. Teale, T. Helgaker, A. Savin, C. Adamo, B. Aradi, A. V. Arbuznikov, P. W. Ayers, E. J. Baerends, V. Barone, P. Calaminici, et al., *DFT exchange: sharing perspectives on the workhorse of quantum chemistry and materials science*, Phys. Chem. Chem. Phys. **24** (2022) 28700.

[100] K. Masuda-Jindo, *Tight-binding theory in the computational materials science*, Mater. Transact. **42** (2001) 979.

[101] W. Thiel, *Semiempirical quantum–chemical methods*, Comput. Mol. Sci. **4** (2014) 145.

[102] J. A. Harrison, J. D. Schall, S. Maskey, P. T. Mikulski, M. T. Knippenberg, and B. H. Morrow, *Review of force fields and intermolecular potentials used in atomistic computational materials research*, Appl. Phys. Rev. **5** (2018).

[103] N. E. Jackson, *Coarse-graining organic semiconductors: the path to multiscale design*, J. Phys. Chem. B **125** (2020) 485.

[104] T. M. Dieb, S. Ju, J. Shiomi, and K. Tsuda, *Monte Carlo tree search for materials design and discovery*, MRS Commun. **9** (2019) 532.

[105] G. Pilania, *Machine learning in materials science: From explainable predictions to autonomous design*, Comput. Mater. Sci. **193** (2021) 110360.

[106] T. Gates, G. Odegard, S. Frankland, and T. Clancy, *Computational materials: multi-scale modeling and simulation of nanostructured materials*, Compos. Sci. Technol. **65** (2005) 2416.

[107] M. J. Buehler, J. Dodson, A. C. Van Duin, and W. A. Goddard III, *The Computational Materials Design Facility (CMDF): A powerful framework for multi-paradigm multi-scale simulations*, MRS Online Proceedings Library **894** (2005) 0894.

[108] G. Wang, L. Peng, K. Li, L. Zhu, J. Zhou, N. Miao, and Z. Sun, *ALKEMIE: An intelligent computational platform for accelerating materials discovery and design*, Comput. Mater. Sci. **186** (2021) 110064.

[109] M. Vaz Jr., E. A. de Souza Neto, and P. A. Muñoz-Rojas, "Materials Modeling – Challenges and Perspectives," *Advanced Computational Materials Modeling*, John Wiley & Sons, Ltd, 2010, chap. 1 1.

[110] E. Wimmer, *Computational materials design: A perspective for atomistic approaches*, J. Comp. Aided Mater. Des. **1** (1994) 215.

[111] R. De Borst, *Challenges in computational materials science: Multiple scales, multi-physics and evolving discontinuities*, Comput. Mater. Sci. **43** (2008) 1.

[112] F. Jensen, *Introduction to computational chemistry*, John wiley & sons, 2017.

[113] B. Chen, J. Xing, and J. I. Siepmann,
Development of polarizable water force fields for phase equilibrium calculations,
J. Phys. Chem. B **104** (2000) 2391.

[114] F. Dommert, K. Wendler, R. Berger, L. Delle Site, and C. Holm, *Force fields for studying the structure and dynamics of ionic liquids: a critical review of recent developments*,
Chem. Phys. Chem. **13** (2012) 1625.

[115] T. E. Cheatham III, P. Cieplak, and P. A. Kollman, *A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat*,
J. Biomol. Struc. Dyn. **16** (1999) 845.

[116] J. Rey, S. Murail, S. de Vries, P. Derreumaux, and P. Tuffery,
PEP-FOLD4: a pH-dependent force field for peptide structure prediction in aqueous solution,
Nucleic Acids Res. **51** (2023) W432.

[117] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, et al., *Parmbsc1: a refined force field for DNA simulations*,
Nat. Methods **13** (2016) 55.

[118] A. P. Lyubartsev and A. L. Rabinovich,
Force field development for lipid membrane simulations,
Biochim. Biophys. Acta **1858** (2016) 2483.

[119] K. H. DuBay, M. L. Hall, T. F. Hughes, C. Wu, D. R. Reichman, and R. A. Friesner,
Accurate force field development for modeling conjugated polymers,
J. Chem. Theory Comput. **8** (2012) 4556.

[120] S. Spicher and S. Grimme,
Robust atomistic modeling of materials, organometallic, and biochemical systems,
Angew. Chem. Int. Ed. **59** (2020) 15665.

[121] A. K. Rappé, C. J. Casewit, K. Colwell, W. A. Goddard III, and W. M. Skiff, *UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations*,
J. Am. Chem. Soc. **114** (1992) 10024.

[122] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg,
A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J. Chem. Phys. **132** (2010).

[123] M. Müller, A. Hansen, and S. Grimme, *ωB97X-3c: A composite range-separated hybrid DFT method with a molecule-optimized polarized valence double-ζ basis set*,
J. Chem. Phys. **158** (2023).

[124] S. Grimme, A. Hansen, S. Ehlert, and J.-M. Mewes,
r2SCAN-3c: A "Swiss army knife" composite electronic-structure method,
J. Chem. Phys. **154** (2021).

[125] J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme,
B97-3c: A revised low-cost variant of the B97-D density functional method,
J. Chem. Phys. **148** (2018).

[126] S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen, *Consistent structures and interactions by density functional theory with small atomic orbital basis sets*,
J. Chem. Phys. **143** (2015) 054107.

[127] R. Sure and S. Grimme, *Corrected small basis set Hartree-Fock method for large systems*,
J. Comp. Chem. **34** (2013) 1672.

[128] L. W. Chung, W. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, et al., *The ONIOM method and its applications*,
Chem. Rev. **115** (2015) 5678.

[129] H. Lin and D. G. Truhlar,
QM/MM: what have we learned, where are we, and where do we go from here?
Theor. Chem. Acc. **117** (2007) 185.

[130] J. W. Ponder and D. A. Case, *Force fields for protein simulations*,
Adv. Prot. Chem. **66** (2003) 27.

[131] H. Kruse and S. Grimme,
A geometrical correction for the inter-and intra-molecular basis set superposition error in Hartree-Fock and density functional theory calculations for large systems,
J. Chem. Phys. **136** (2012).

[132] A. Dittmer, R. Izsak, F. Neese, and D. Maganas,
Accurate band gap predictions of semiconductors in the framework of the similarity transformed equation of motion coupled cluster theory, Inorg. Chem. **58** (2019) 9303.

[133] P. Pracht, F. Bohle, and S. Grimme,
Automated exploration of the low-energy chemical space with fast quantum chemical methods,
Phys. Chem. Chem. Phys. **22** (2020) 7169.

[134] S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher, and M. Stahn, *Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules*,
J. Phys. Chem. A **125** (2021) 4039.

[135] S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy,
The high-throughput highway to computational materials design, Nat. Mater. **12** (2013) 191.

[136] Y. Yan, L. Zhang, S. Li, H. Liang, and Z. Qiao,
Adsorption behavior of metal-organic frameworks: From single simulation, high-throughput computational screening to machine learning, Comput. Mater. Sci. **193** (2021) 110383.

[137] E. Martínez-Núñez,
An automated method to find transition states using chemical dynamics simulations,
J. Comp. Chem. **36** (2015) 222.

[138] C. Plett and S. Grimme,
Automated and efficient generation of general molecular aggregate structures,
Angew. Chem. Int. Ed. **62** (2023) e202214477.

[139] M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme,
Best-practice DFT protocols for basic molecular computational chemistry,
Angew. Chem. Int. Ed. **61** (2022) e202205735.

[140] A. Szabo and N. S. Ostlund,
Modern quantum chemistry: introduction to advanced electronic structure theory,
Courier Corporation, 2012.

[141] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, *A generally applicable atomic-charge dependent London dispersion correction*,
J. Chem. Phys. **150** (2019) 154122.

[142] S. Grimme, C. Bannwarth, and P. Shushkov, *A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z= 1–86)*,
Journal of chemical theory and computation **13** (2017) 1989.

[143] C. Bannwarth, S. Ehlert, and S. Grimme,
GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions,
J. Chem. Theory Comput. **15** (2019) 1652.

[144] S. Grimme, *Semiempirical GGA-type density functional constructed with a long-range dispersion correction*, J. Comp. Chem. **27** (2006) 1787.

[145] A. J. Varandas, *Straightening the hierarchical staircase for basis set extrapolations: A low-cost approach to high-accuracy computational chemistry*,
Ann. Rev. Phys. Chem. **69** (2018) 177.

[146] S. Grimme, *Density functional theory with London dispersion corrections*,
Comput. Mol. Sci. **1** (2011) 211.

[147] A. Tkatchenko and M. Scheffler, *Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density<? format?> and Free-Atom Reference Data*,
Phys. Rev. Lett. **102** (2009) 073005.

[148] A. D. Becke and E. R. Johnson, *Exchange-hole dipole moment and the dispersion interaction*,
J. Chem. Phys. **122** (2005).

[149] A. Tkatchenko, R. A. DiStasio Jr, R. Car, and M. Scheffler,
Accurate and efficient method for many-body van der Waals interactions,
Phys. Rev. Lett. **108** (2012) 236402.

[150] B. Axilrod and E. Teller, *Interaction of the van der Waals type between three atoms*,
J. Chem. Phys. **11** (1943) 299.

[151] Y. Muto, *Force between nonpolar molecules*, J. Phys. Math. Soc. Jpn **17** (1943) 629.

[152] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist,
Van der Waals density functional for general geometries, Phys. Rev. Lett. **92** (2004) 246401.

[153] O. A. Vydrov and T. Van Voorhis,
Nonlocal van der Waals density functional: The simpler the better,
J. Chem. Phys. **126** (2014) 1984.

[154] S. Grimme, S. Ehrlich, and L. Goerigk,
Effect of the damping function in dispersion corrected density functional theory,
J. Comp. Chem. **32** (2011) 1456.

[155] S. A. Ghasemi, A. Hofstetter, S. Saha, and S. Goedecker,
Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network, Phys. Rev. B **92** (2015) 045131.

[156] J. Kohn, S. Grimme, and A. Hansen, unpublished ().

[157] H. Neugebauer, B. Bädorf, S. Ehlert, A. Hansen, and S. Grimme,
High-throughput screening of spin states for transition metal complexes with spin-polarized extended tight-binding methods, J. Comp. Chem. **44** (2023) 2120.

[158] M. de Wergifosse and S. Grimme,
Nonlinear-response properties in a simplified time-dependent density functional theory (sTD-DFT) framework: Evaluation of the first hyperpolarizability, J. Chem. Phys. **149** (2018).

[159] M. Friede, C. Hölzer, S. Ehlert, and S. Grimme,
dxtb—An efficient and fully differentiable framework for extended tight-binding,
T. Chem. Phys. **161** (2024).

[160] S. Grimme, M. Müller, and A. Hansen, *A non-self-consistent tight-binding electronic structure potential in a polarized double-ζ basis set for all spd-block elements up to Z= 86*,
The Journal of Chemical Physics **158** (2023).

[161] J. M. Herbert, *Dielectric continuum methods for quantum chemistry*,
Comput. Mol. Sci. **11** (2021) e1519.

[162] C. J. Cramer, D. G. Truhlar, et al.,
Implicit solvation models: equilibria, structure, spectra, and dynamics,
Chem. Rev. **99** (1999) 2161.

[163] Q. Cai, J. Wang, M.-J. Hsieh, X. Ye, and R. Luo,
"Poisson–Boltzmann implicit solvation models," *Annu. Rep. Comput. Chem.* Vol. 8,
Elsevier, 2012 149.

[164] S. Miertuš, E. Scrocco, and J. Tomasi, *Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects*,
Chem. Phys. **55** (1981) 117.

[165] C. Plett, M. Stahn, M. Bursch, J.-M. Mewes, and S. Grimme, *Improving Quantum Chemical Solvation Models by Dynamic Radii Adjustment for Continuum Solvation (DRACO)*,
J. Phys. Chem. Lett. **15** (2024) 2462.

[166] J. Zhu, Y. Shi, and H. Liu, *Parametrization of a generalized Born/solvent-accessible surface area model and applications to the simulation of protein dynamics*,
J. Phys. Chem. B **106** (2002) 4844.

[167] S. Ehlert, M. Stahn, S. Spicher, and S. Grimme,
A robust and efficient implicit solvation model for fast semiempirical methods,
J. Chem. Theory Comput. **17** (2021) 4250.

[168] A. Klamt, *The COSMO and COSMO-RS solvation models*,
Comput. Mol. Science **1** (2011) 699.

[169] A. V. Marenich, C. J. Cramer, and D. G. Truhlar,
Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions,
J. Phys. Chem. B **113** (2009) 6378.

[170] S. A. Hollingsworth and R. O. Dror, *Molecular dynamics simulation for all*,
Neuron **99** (2018) 1129.

[171] D. Bertsimas and J. Tsitsiklis, *Simulated annealing*, Stat. Sci. **8** (1993) 10.

[172] D. Anstine, R. Zubatyuk, and O. Isayev, *AIMNet2: a neural network potential to meet your neutral, charged, organic, and elemental-organic needs* (2024).

[173] S. Grimme, C. Bannwarth, E. Caldeweyher, J. Pisarek, and A. Hansen,
A general intermolecular force field based on tight-binding quantum chemical calculations,
J. Chem. Phys. **147** (2017).

[174] V. M. Anisimov, G. Lamoureux, I. V. Vorobyov, N. Huang, B. Roux, and A. D. MacKerell,
Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator, J. Chem. Theory Comput. **1** (2005) 153.

[175] C. S. De Magalhães, D. M. Almeida, H. J. C. Barbosa, and L. E. Dardenne,
A dynamic niching genetic algorithm strategy for docking highly flexible ligands,
Inf. Sci. **289** (2014) 206.

[176] M. Pardakhti, E. Moharreri, D. Wanik, S. L. Suib, and R. Srivastava,
Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs),
Comb. Sci. **19** (2017) 640.

[177] D. Vemula, P. Jayasurya, V. Sushmitha, Y. N. Kumar, and V. Bhandari,
CADD, AI and ML in drug discovery: A comprehensive review,
Eur. J. Pharm. Sci. **181** (2023) 106324.

[178] D. Weininger, *SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules*, J. Chem. Inf. Comput. Sci. **28** (1988) 31.

[179] S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi, and I. Pletnev,
InChI-the worldwide chemical structure identifier standard, J. Chem. Inf. **5** (2013) 1.

[180] N. van Staalduinen and C. Bannwarth, *MolBar: A Molecular Identifier for Inorganic and Organic Molecules with Full Support of Stereoisomerism* (2024).

[181] `https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html`. *Accessed Aug 13th 2024.*

[182] S. Ash, M. A. Cline, R. W. Homer, T. Hurst, and G. B. Smith,
SYBYL line notation (SLN): A versatile language for chemical structure representation,
J. Chem. Inf. Comput. Sci. **37** (1997) 71.

[183] W. Tress and W. Tress, *Organic solar cells*, Springer, 2014.

[184] G. Zhang, C. Xie, P. You, and S. Li, *Introduction to Organic Electronic Devices*,
Springer, 2022.

[185] R. A. Marcus, *Chemical and electrochemical electron-transfer theory*, Annu. Rev. Phys. Chem. **15** (1964) 155.

[186] E. Buhks, M. Bixon, J. Jortner, and G. Navon, *Quantum effects on the rates of electron-transfer reactions*, J. Phys. Chem. **85** (1981) 3759.

[187] N. R. Kestner, J. Logan, and J. Jortner, *Thermal electron transfer reactions in polar solvents*, J. Phys. Chem. **78** (1974) 2148.

[188] V. Levich, R.-R. Dogonadze, E. German, A. Kuznetsov, and Y. I. Kharkats, *Theory of homogeneous reactions involving proton transfer*, Electrochim. Acta **15** (1970) 353.

[189] N. Gildemeister, G. Ricci, L. Böhner, J. M. Neudörfl, D. Hertel, F. Würthner, F. Negri, K. Meerholz, and D. Fazzi, *Understanding the structural and charge transport property relationships for a variety of merocyanine single-crystals: a bottom up computational investigation*, J. Mater. Chem. C **9** (2021) 10851.

[190] A. Callear, "Basic RRKM Theory," *Comprehensive Chemical Kinetics*, vol. 24, Elsevier, 1983 333.

[191] V. Coropceanu, J. Cornil, D. A. da Silva Filho, Y. Olivier, R. Silbey, and J.-L. Brédas, *Charge transport in organic semiconductors*, Chem. Rev. **107** (2007) 926.

[192] V. Coropceanu, J. Cornil, D. A. da Silva Filho, Y. Olivier, R. Silbey, and J.-L. Brédas, *Charge transport in organic semiconductors*, Chem. Rev. **107** (2007) 926.

[193] S. Giannini, A. Carof, M. Ellis, H. Yang, O. G. Ziogos, S. Ghosh, and J. Blumberger, *Quantum localization and delocalization of charge carriers in organic semiconducting crystals*, Nat. Commun. **10** (2019) 3843.

[194] M. Hayashi and K. Katsuki, *Hydrogen-like absorption spectrum of cuprous oxide*, J. Phys. Soc. Japan **7** (1952) 599.

[195] O. V. Mikhnenko, P. W. Blom, and T.-Q. Nguyen, *Exciton diffusion in organic semiconductors*, Energy Env. Sci. **8** (2015) 1867.

[196] J. Frenkel, *On the transformation of light into heat in solids. I*, Phys. Rev. **37** (1931) 17.

[197] G. H. Wannier, *The structure of electronic excitation levels in insulating crystals*, Phys. Rev. **52** (1937) 191.

[198] T. Förster, *Zwischenmolekulare energiewanderung und fluoreszenz*, Ann. Phys. **437** (1948) 55.

[199] D. L. Dexter, *A theory of sensitized luminescence in solids*, J. Chem. Phys. **21** (1953) 836.

[200] M. Schwarze, W. Tress, B. Beyer, F. Gao, R. Scholz, C. Poelking, K. Ortstein, A. A. Günther, D. Kasemann, D. Andrienko, et al., *Band structure engineering in organic semiconductors*, Science **352** (2016) 1446.

[201] P. Fang, *Analysis of conversion efficiency of organic-semiconductor solar cells*, J. Appl. Phys. **45** (1974) 4672.

[202] W. Shockley and H. Queisser, "Detailed balance limit of efficiency of p–n junction solar cells," *Renewable energy*, Routledge, 2018 Vol2_35.

[203] A. Tsumura, H. Koezuka, and T. Ando,
Macromolecular electronic device: Field-effect transistor with a polythiophene thin film,
Appl. Phys. Lett. **49** (1986) 1210.

[204] R. Van Noorden and D. Castelvecchi, *World's tiniest machines win chemistry Nobel*,
Nature News **538** (2016) 152.

[205] P. Ghosh, G. Federwisch, M. Kogej, C. A. Schalley, D. Haase, W. Saak, A. Lützen, and
R. M. Gschwind, *Controlling the rate of shuttling motions in [2] rotaxanes by electrostatic interactions: a cation as solvent-tunable brake*, Org. Biomol. Chem. **3** (2005) 2691.

[206] S. Ernst and G. Haberhauer, *A Unidirectional Open–Close Mechanism of Metal-Ion-Driven Molecular Hinges with Adjustable Amplitude*, Chem. Eur. J. **15** (2009) 13406.

[207] M. T. Nistor and A. G. Rusu, "Nanorobots with applications in medicine,"
Polymeric Nanomaterials in Nanotherapeutics, Elsevier, 2019 123.

[208] S. E. Lyshevski, *Nano and molecular electronics handbook*, CRC Press, 2018.

[209] M. Suyetin and T. Heine,
A memory nanodevice based on Zn-MOF-74: a molecular dynamics study,
J. Mater. Chem. **8** (2020) 1567.

[210] S. Spicher, E. Caldeweyher, A. Hansen, and S. Grimme, *Benchmarking London dispersion corrected density functional theory for noncovalent ion–π interactions*,
Phys. Chem. Chem. Phys. **23** (2021) 11635.

[211] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme,
A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions,
Phys. Chem. Chem. Phys. **19** (2017) 32184.

[212] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and
S. Grimme, *Extended tight-binding quantum chemistry methods*,
WIREs Comput. Mol. Sci. **11** (2021) e1493.

[213] S. Spicher and S. Grimme,
Robust atomistic modeling of materials, organometallic, and biochemical systems,
Angew. Chem. Int. Ed. **59** (2020) 15665.

[214] C. Romuald, E. Busseron, and F. Coutrot, *Very contracted to extended co-conformations with or without oscillations in two-and three-station [c2] daisy chains*,
J. Org. Chem. **75** (2010) 6516.

[215] R. E. Dawson, S. F. Lincoln, and C. J. Easton,
The foundation of a light driven molecular muscle based on stilbene and α-cyclodextrin,
Chem. Commun. (2008) 3980.

[216] M. C. Jimenez-Molero, C. Dietrich-Buchecker, and J.-P. Sauvage,
Chemically induced contraction and stretching of a linear rotaxane dimer,
Chem. Eur. J. **8** (2002) 1456.

[217] C. J. Bruns, J. Li, M. Frasconi, S. T. Schneebeli, J. Iehl, H.-P. Jacquot de Rouville, S. I. Stupp, G. A. Voth, and J. F. Stoddart,
An electrochemically and thermally switchable donor–acceptor [c2] daisy chain rotaxane, Angew. Chem. Int. Ed. **126** (2014) 1984.

[218] F. Durola, V. Heitz, F. Reviriego, C. Roche, J.-P. Sauvage, A. Sour, and Y. Trolez,
Cyclic [4] rotaxanes containing two parallel porphyrinic plates: Toward switchable molecular receptors and compressors, Acc. Chem. Res. **47** (2014) 633.

[219] C. J. Bruns and J. F. Stoddart, *Rotaxane-based molecular muscles*,
Acc. Chem. Res. **47** (2014) 2186.

[220] E. J. Leonhardt and R. Jasti, *Emerging Applications of Carbon Nanohoops*,
Nat. Rev. Chem. **3** (2019) 672.

[221] D. Wu, W. Cheng, X. Ban, and J. Xia, *Cycloparaphenylenes (CPPs): An Overview of Synthesis, Properties, and Potential Applications*, Asian J. Org. Chem. **7** (2018) 2161.

[222] Y. Xu and M. von Delius, *The Supramolecular Chemistry of Strained Carbon Nanohoops*,
Angew. Chem. Int. Ed. **59** (2020) 559.

[223] K. Yazaki, L. Catti, and M. Yoshizawa,
Polyaromatic Molecular Tubes: From Strategic Synthesis to Host Functions,
Chem. Commun. **54** (2018) 3195.

[224] R. Friederich, M. Nieger, and F. Vögtle, *Auf dem Weg zu makrocyclischen para-Phenylenen*,
Chem. Ber. **126** (1993) 1723.

[225] M. Hermann, D. Wassy, and B. Esser,
Conjugated Nanohoops Incorporating Donor, Acceptor, Hetero- or Polycyclic Aromatics,
Angew. Chem. Int. Ed. **60** (2021) 15743.

[226] M. Schmidt, D. Wassy, M. Hermann, M. T. González, N. Agräit, L. A. Zotti, B. Esser, and E. Leary, *Single-Molecule Conductance of Dibenzopentalenes: Antiaromaticity and Quantum Interference*, Chem. Commun. **57** (2021) 745.

[227] J. Wu, Y. Chen, J. Liu, Z. Pang, G. Li, Z. Lu, Y. Huang, A. Facchetti, and T. J. Marks,
Tuning the Antiaromatic Character and Charge Transport of Pentalene-Based Antiaromatic Compounds by Substitution, J. Mater. Chem. C (2021).

[228] J. S. Wössner, D. Wassy, A. Weber, M. Bovenkerk, M. Hermann, M. Schmidt, and B. Esser,
[N]Cyclodibenzopentalenes as Antiaromatic Curved Nanocarbons with High Strain and Strong Fullerene Binding, J. Am. Chem. Soc. **143** (2021) 12244.

[229] P. v. R. Schleyer, C. Maerker, A. Dransfeld, H. Jiao, and N. J. R. v. E. Hommes,
Nucleus-Independent Chemical Shifts: A Simple and Efficient Aromaticity Probe,
J. Am. Chem. Soc. **118** (1996) 6317.

[230] R. Gershoni-Poranne and A. Stanger,
The NICS-XY-Scan: Identification of Local and Global Ring Currents in Multi-Ring Systems,
Chem. Eur. J. **20** (2014) 5673.

[231] A. D. Becke,
Density-functional thermochemistry. I. The effect of the exchange-only gradient correction,
J. Chem. Phys. **96** (1992) 2155.

[232] E. Van Lenthe and E. J. Baerends, *Optimized Slater-Type Basis Sets for the Elements 1-118*, J. Comput. Chem. **24** (2003) 1142.

[233] B. Zhang, S. Wu, X. Hou, G. Li, Y. Ni, Q. Zhang, J. Zhu, Y. Han, P. Wang, Z. Sun, et al., *A graphyne spoked wheel*, Chem. **8** (2022) 2831.

[234] Y. Liu, A. Narita, J. Teyssandier, M. Wagner, S. De Feyter, X. Feng, and K. Müllen, *A shape-persistent polyphenylene spoked wheel*, J. Am. Chem. Soc. **138** (2016) 15539.

[235] D. Mössinger, J. Hornung, S. Lei, S. De Feyter, and S. Höger, *Molecularly defined shape-persistent 2D oligomers: the covalent-template approach to molecular spoked wheels*, Angew. Chem. **119** (2007) 6926.

[236] D. Mössinger, D. Chaudhuri, T. Kudernac, S. Lei, S. De Feyter, J. M. Lupton, and S. Höger, *Large all-hydrocarbon spoked wheels of high symmetry: modular synthesis, photophysical properties, and surface assembly*, J. Am. Chem. Soc. **132** (2010) 1410.

[237] J. Burdynska, Y. Li, A. V. Aggarwal, S. Höger, S. S. Sheiko, and K. Matyjaszewski, *Synthesis and arm dissociation in molecular stars with a spoked wheel core and bottlebrush arms*, J. Am. Chem. Soc. **136** (2014) 12762.

[238] A. Idelson, C. Sterzenbach, S.-S. Jester, C. Tschierske, U. Baumeister, and S. Höger, *A liquid-crystalline phenylene-based shape-persistent molecular spoked wheel*, J. Am. Chem. Soc. **139** (2017) 4429.

[239] C. Sterzenbach, T. J. Keller, D. Kraus, J. M. Lupton, S.-S. Jester, and S. Höger, *Expanded all-phenylene molecular spoked wheels: cutouts of graphenylene-3*, Org. Chem. Front. **8** (2021) 4980.

[240] K. Sonogashira, Y. Tohda, and N. Hagihara, *A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines*, Tetrahedron Lett. **16** (1975) 4467.

[241] N. Miyaura, K. Yamada, H. Suginome, and A. Suzuki, *Novel and convenient method for the stereo-and regiospecific synthesis of conjugated alkadienes and alkenynes via the palladium-catalyzed cross-coupling reaction of 1-alkenylboranes with bromoalkenes and bromoalkynes*, J. Am. Chem. Soc. **107** (1985) 972.

[242] T. Yamamoto, S. Wakabayashi, and K. Osakada, *Mechanism of C-C coupling reactions of aromatic halides, promoted by Ni (COD) 2 in the presence of 2, 2'-bipyridine and PPh3, to give biaryls*, J. Orgmet. Chem. **428** (1992) 223.

[243] B. Peters, A. Heyden, A. T. Bell, and A. Chakraborty, *A growing string method for determining transition states: Comparison to the nudged elastic band and string methods*, J. Chem. Phys. **120** (2004) 7877.

[244] A. M. Evans, M. J. Strauss, A. R. Corcos, Z. Hirani, W. Ji, L. S. Hamachi, X. Aguilar-Enriquez, A. D. Chavez, B. J. Smith, and W. R. Dichtel, *Two-Dimensional Polymers and Polymerizations*, Chem. Rev. **122** (2021) 442.

[245] F. Haase, K. Gottschling, L. Stegbauer, L. Germann, R. Gutzler, V. Duppel, V. Vyas, K. Kern, R. Dinnebier, and B. Lotsch, *Tuning the stacking behaviour of a 2D covalent organic framework through non-covalent interactions*, Mater. Chem. Front. **1** (2017) 1354.

[246] F. Liu, Y. He, X. Liu, Z. Wang, H.-L. Liu, X. Zhu, C.-C. Hou, Y. Weng, Q. Zhang, and Y. Chen, *Regulating Excitonic Effects in Covalent Organic Frameworks to Promote Free Charge Carrier Generation*, ACS Catal. **12** (2022) 9494.

[247] R. B. Rashid, A. M. Evans, L. A. Hall, R. R. Dasari, E. K. Roesner, S. R. Marder, D. M. D'Allesandro, W. R. Dichtel, and J. Rivnay, *A Semiconducting Two-Dimensional Polymer as an Organic Electrochemical Transistor Active Layer*, Adv. Mater. (2022) 2110703.

[248] F. Haase and B. V. Lotsch, *Solving the COF trilemma: towards crystalline, stable and functional covalent organic frameworks*, Chem. Soc. Rev. **49** (2020) 8469.

[249] X. Wu, X. Han, Y. Liu, Y. Liu, and Y. Cui, *Control interlayer stacking and chemical stability of two-dimensional covalent organic frameworks via steric tuning*, J. Am. Chem. Soc. **140** (2018) 16124.

[250] A. M. Pütz, M. W. Terban, S. Bette, F. Haase, R. E. Dinnebier, and B. V. Lotsch, *Total scattering reveals the hidden stacking disorder in a 2D covalent organic framework*, Chem. Sci. **11** (2020) 12647.

[251] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Deshaye, T. Dumitrică, A. Dominguez, et al., *DFTB+, a software package for efficient approximate density functional theory based atomistic simulations*, J. Chem. Phys. **152** (2020) 124101.

[252] C. Feriante, A. M. Evans, S. Jhulki, I. Castano, M. J. Strauss, S. Barlow, W. R. Dichtel, and S. R. Marder, *New mechanistic insights into the formation of imine-linked two-dimensional covalent organic frameworks*, J. Am. Chem. Soc. **142** (2020) 18637.

[253] S. Xun, H. Li, G. Sini, and J.-L. Bredas, *Impact of Imine Bond Orientations on the Geometric and Electronic Structures of Imine-based Covalent Organic Frameworks*, Chem. Asian J. **16** (2021) 3781.

[254] D. B. Shinde, G. Sheng, X. Li, M. Ostwal, A.-H. Emwas, K.-W. Huang, and Z. Lai, *Crystalline 2D covalent organic framework membranes for high-flux organic solvent nanofiltration*, J. Am. Chem. Soc. **140** (2018) 14342.

[255] H. Lyu, H. Li, N. Hanikel, K. Wang, and O. M. Yaghi, *Covalent organic frameworks for carbon dioxide capture from air*, J. Am. Chem. Soc. **144** (2022) 12989.

[256] X. Zhao, P. Pachfule, and A. Thomas, *Covalent organic frameworks (COFs) for electrochemical applications*, Chem. Soc. Rev. **50** (2021) 6871.

[257] S. Bhunia, K. A. Deo, and A. K. Gaharwar, *2D covalent organic frameworks for biomedical applications*, Adv. Func. Mater. **30** (2020) 2002046.

[258] C. H. Feriante, S. Jhulki, A. M. Evans, R. R. Dasari, K. Slicker, W. R. Dichtel, and S. R. Marder, *Rapid Synthesis of High Surface Area Imine-Linked 2D Covalent Organic Frameworks by Avoiding Pore Collapse During Isolation*, Adv. Mater. **32** (2020) 1905776.

[259] A. M. Evans, K. A. Collins, S. Xun, T. G. Allen, S. Jhulki, I. Castano, H. L. Smith, M. J. Strauss, A. K. Oanta, L. Liu, et al.,
Controlled n-Doping of Naphthalene-Diimide-Based 2D Polymers,
Adv. Mater. **34** (2022) 2101932.

[260] H. Oberhofer, K. Reuter, and J. Blumberger,
Charge transport in molecular materials: An assessment of computational methods,
Chem. Rev. **117** (2017) 10319.

[261] V. Coropceanu, Y. Li, Y. Yi, L. Zhu, and J.-L. Brédas, *Intrinsic charge transport in single crystals of organic molecular semiconductors: A theoretical perspective*,
MRS Bull. **38** (2013) 57.

[262] S. Giannini and J. Blumberger, *Charge transport in organic semiconductors: the perspective from nonadiabatic molecular dynamics*, Acc. Chem. Res. **55** (2022) 819.

[263] A. Baggioli, M. Casalegno, G. Raos, L. Muccioli, S. Orlandi, and C. Zannoni,
Atomistic simulation of phase transitions and charge mobility for the organic semiconductor Ph-BTBT-C10, Chem. Mater. **31** (2019) 7092.

[264] Y. Mao, A. Montoya-Castillo, and T. E. Markland, *Accurate and efficient DFT-based diabatization for hole and electron transfer using absolutely localized molecular orbitals*,
J. Chem. Phys. **151** (2019) 164114.

[265] Q. Wu and T. Van Voorhis,
Extracting electron transfer coupling elements from constrained density functional theory,
J. Chem. Phys. **125** (2006) 164105.

[266] C. Schober, K. Reuter, and H. Oberhofer, *Critical analysis of fragment-orbital DFT schemes for the calculation of electronic coupling values*, J. Chem. Phys. **144** (2016) 054103.

[267] S. Ghan, C. Kunkel, K. Reuter, and H. Oberhofer, *Improved projection-operator diabatization schemes for the calculation of electronic coupling values*,
J. Chem. Theory Comput. **16** (2020) 7431.

[268] R. J. Cave and M. D. Newton, *Generalization of the Mulliken-Hush treatment for the calculation of electron transfer matrix elements*, Chem. Phys. Lett. **249** (1996) 15.

[269] F. Gajdos, S. Valner, F. Hoffmann, J. Spencer, M. Breuer, A. Kubas, M. Dupuis, and J. Blumberger, *Ultrafast estimation of electronic couplings for electron transfer between π-conjugated organic molecules*, J. Chem. Theory Comput. **10** (2014) 4653.

[270] B. Baumeier, J. Kirkpatrick, and D. Andrienko, *Density-functional based determination of intermolecular charge transfer properties for large-scale morphologies*,
Phys. Chem. Chem. Phys. **12** (2010) 11103.

[271] T. Nematiaram, D. Padula, A. Landi, and A. Troisi,
On the largest possible mobility of molecular semiconductors and how to achieve it,
Adv. Func. Mater. **30** (2020) 2001906.

[272] H. Kitoh-Nishioka, K. Welke, Y. Nishimoto, D. G. Fedorov, and S. Irle,
Multiscale simulations on charge transport in covalent organic frameworks including dynamics of transfer integrals from the FMO-DFTB/LCMO approach,
J. Phys. Chem. C **121** (2017) 17712.

[273] Z.-F. Yao, Y.-Q. Zheng, J.-H. Dou, Y. Lu, Y.-F. Ding, L. Ding, J.-Y. Wang, and J. Pei, *Approaching crystal structure and high electron mobility in conjugated polymer crystals*, Adv. Mater. **33** (2021) 2006794.

[274] E. F. Valeev, V. Coropceanu, D. A. da Silva Filho, S. Salman, and J.-L. Brédas, *Effect of electronic polarization on charge-transport parameters in molecular organic semiconductors*, J. Am. Chem. Soc. **128** (2006) 9882.

[275] K. Senthilkumar, F. Grozema, F. Bickelhaupt, and L. Siebbeles, *Charge transport in columnar stacked triphenylenes: Effects of conformational fluctuations on charge transfer integrals and site energies*, J. Chem. Phys. **119** (2003) 9809.

[276] O. G. Ziogos, A. Kubas, Z. Futera, W. Xie, M. Elstner, and J. Blumberger, *HAB79: A new molecular dataset for benchmarking DFT and DFTB electronic couplings against high-level ab initio calculations*, J. Chem. Phys. **155** (2021) 234115.

[277] H. Kitoh-Nishioka and K. Ando, *Charge-transfer matrix elements by FMO-LCMO approach: Hole transfer in DNA with parameter tuned range-separated DFT*, Chem. Phys. Lett. **621** (2015) 96.

[278] J.-D. Chai and M. Head-Gordon, *Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections*, Phys. Chem. Chem. Phys. **10** (2008) 6615.

[279] D. P. Chong, E. Van Lenthe, S. Van Gisbergen, and E. J. Baerends, *Even-tempered slater-type orbitals revisited: From hydrogen to krypton*, J. Comp. Chem. **25** (2004) 1030.

[280] T. Onishi, *Theoretical Chemistry for Advanced Nanomaterials: Functional Analysis by Computation and Experiment*, Springer, 2020.

[281] W. Liu, Q. Liu, C. Xiang, H. Zhou, L. Jiang, and Y. Zou, *Theoretical exploration of optoelectronic performance of PM6: Y6 series-based organic solar cells*, Surf. Interfaces **26** (2021) 101385.

[282] M. Raftani, T. Abram, N. Bennani, and M. Bouachrine, *Theoretical study of new conjugated compounds with a low bandgap for bulk heterojunction solar cells: DFT and TD-DFT study*, Results Chem. **2** (2020) 100040.

[283] A. V. Mumyatov, L. A. Frolova, L. G. Gutsev, E. A. Khakina, N. A. Sanina, S. M. Aldoshin, and P. A. Troshin, *A light-triggered molecular switch for an efficient OFET-based organic memory device*, J. Mater. Chem. **11** (2023) 963.

[284] L. Goerigk and S. Grimme, *Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions*, J. Chem. Theory Comput. **7** (2011) 291.

[285] Q. Ma, M. Schwilk, C. Köppl, and H. J. Werner, *Scalable Electron Correlation Methods. 4. Parallel Explicitly Correlated Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD-F12)*, J. Chem. Theory Comput. **13** (2017) 4871.

[286] Q. Ma and H. J. Werner,
Explicitly correlated local coupled-cluster methods using pair natural orbitals,
Comput. Mol. Sci. **8** (2018) e1371.

[287] H. Bürckstümmer, N. M. Kronenberg, M. Gsänger, M. Stolte, K. Meerholz, and F. Würthner,
Tailored merocyanine dyes for solution-processed BHJ solar cells,
J. Mater. Chem. **20** (2010) 240.

[288] B. Engels and V. Engel, *The dimer-approach to characterize opto-electronic properties of and exciton trapping and diffusion in organic semiconductor aggregates and crystals*,
Phys. Chem. Chem. Phys. **19** (2017) 12604.

[289] G. Du, E. Moulin, N. Jouault, E. Buhler, and N. Giuseppone, *Muscle-like supramolecular polymers: Integrated motion from thousands of molecular machines*,
Angew. Chem. **124** (2012) 12672.

[290] M.-M. Titirici, R. J. White, N. Brun, V. L. Budarin, D. S. Su, F. Del Monte, J. H. Clark, and M. J. MacLachlan, *Sustainable carbon materials*, Chem. Soc. Rev. **44** (2015) 250.

[291] A. Weber, J. Kohn, B. Esser, and S. Grimme, unpublished ().

[292] J. E. Moses and A. D. Moorhouse, *The growing applications of click chemistry*,
Chem. Soc. Rev. **36** (2007) 1249.

[293] R. Du, R. Xin, H. Wang, W. Zhu, R. Li, and W. Liu, *Machine learning: An accelerator for the exploration and application of advanced metal-organic frameworks*,
Chem. Eng. J. (2024) 151828.

[294] L. Böhner, *Optical, Morphological, and Electrical Properties of Merocyanine Molecular Aggregate Thin Films-A Correlated study-Abstract* (2022).

[295] T. Husch, A. C. Vaucher, and M. Reiher, *Semiempirical molecular orbital models based on the neglect of diatomic differential overlap approximation*,
Int. J. Quantum Chem. **118** (2018) e25799.

[296] A. S. Christensen, T. Kubar, Q. Cui, and M. Elstner, *Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications*,
Chem. Rev. **116** (2016) 5301.

[297] "Semiempirical Extended Tight-Binding Program Package `xtb`",
`https://github.com/grimme-lab/xtb`. Accessed: 2020-06-20.

[298] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, *Turbomole*,
Comput. Mol. Science **4** (2014) 91.

[299] R. Ahlrichs, M. Armbruster, M. Bär, et. al., "*TURBOMOLE 7.5* Program Package" **2020**,
Universität Karlsruhe. `http://www.turbomole.com`.

[300] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized gradient approximation made simple*,
Phys. Rev. Lett. **77** (1996) 3865.

[301] C. Adamo and V. Barone,
Toward reliable density functional methods without adjustable parameters: The PBE0 model,
J. Chem. Phys. **110** (1999) 6158.

[302] S. Li, D. Taura, A. Hashidzume, and A. Harada, *Light-switchable janus [2] rotaxanes based on α-cyclodextrin derivatives bearing two recognition sites linked with oligo (ethylene glycol)*, Chem. Asian J. **5** (2010) 2281.

[303] S. Spicher and S. Grimme, *Single-Point Hessian Calculations for Improved Vibrational Frequencies and Rigid-Rotor-Harmonic-Oscillator Thermodynamics*, J. Chem. Theory Comput. **17** (2021) 1701.

[304] Y. Segawa, H. Ito, and K. Itami, *Structurally Uniform and Atomically Precise Carbon Nanostructures*, Nat. Rev. Mater. **1** (2016) 15002.

[305] S. Mirzaei, E. Castro, and R. Hernández Sánchez, *Conjugated Molecular Nanotubes*, Chem. Eur. J. **27** (2021) 8642.

[306] M. A. Majewski and M. Stępień, *Bowls, Hoops, and Saddles: Synthetic Approaches to Curved Aromatic Molecules*, Angew. Chem. Int. Ed. **58** (2019) 86.

[307] B. Esser and M. Hermann, *Buckling up Zigzag Nanobelts*, Nat. Chem. **13** (2021) 209.

[308] R. Gleiter and G. Haberhauer, *Aromaticity and Other Conjugation Effects*, Weinheim, Germany: Wiley-VCH, 2012.

[309] A. Konishi and M. Yasuda, *Breathing New Life into Nonalternant Hydrocarbon Chemistry: Syntheses and Properties of Polycyclic Hydrocarbons Containing Azulene, Pentalene, and Heptalene Frameworks*, Chem. Lett. **50** (2021) 195.

[310] J. Wilbuer, D. C. Grenz, G. Schnakenburg, and B. Esser, *Donor- and Acceptor-Functionalized Dibenzo[a,e]Pentalenes: Modulation of the Electronic Band Gap*, Org. Chem. Front. **4** (2017) 658.

[311] B. Esser, *Theoretical Analysis of [5.5.6]Cyclacenes: Electronic Properties, Strain Energies and Substituent Effects*, Phys. Chem. Chem. Phys. **17** (2015) 7366.

[312] M. Hermann, T. Böttcher, M. Schorpp, S. Richert, D. Wassy, I. Krossing, and B. Esser, *Cations and Anions of Dibenzo[a,e]Pentalene and Reduction of a Dibenzo[a,e]Pentalenophane*, Chem. Eur. J. **27** (2021) 4964.

[313] P. K. Anjalikrishna, S. R. Gadre, and C. H. Suresh, *Antiaromaticity–Aromaticity Interplay in Fused Benzenoid Systems Using Molecular Electrostatic Potential Topology*, J. Phys. Chem. A **125** (2021) 5999.

[314] M. Nakano, I. Osaka, K. Takimiya, and T. Koganezawa, *Novel Dibenzo[a,e]Pentalene-Based Conjugated Polymers*, J. Mater. Chem. C **2** (2014) 64.

[315] D. C. Grenz, M. Schmidt, D. Kratzert, and B. Esser, *Dibenzo[a,e]Pentalenes with Low-Lying LUMO Energy Levels as Potential n-Type Materials*, J. Org. Chem. **83** (2018) 656.

[316] M. Hermann, D. Wassy, D. Kratzert, and B. Esser, *Dibenzo[a,e]Pentalenophanes: Bending a Non-Alternant Hydrocarbon*, Chem. Eur. J. **24** (2018) 7374.

[317] D. Wassy, M. Hermann, J. S. Wössner, L. Frédéric, G. Pieters, and B. Esser, *Enantiopure Nanohoops through Racemic Resolution of Diketo[n]CPPs by Chiral Derivatization as Precursors to DBP[n]CPPs*, Chem. Sci. **12** (2021) 10150.

[318] D. Wassy, M. Pfeifer, and B. Esser, *Synthesis and Properties of Conjugated Nanohoops Incorporating Dibenzo[a,e]Pentalenes: [2]DBP[12]CPPs*, J. Org. Chem. **85** (2020) 34.

[319] S. Yamago, Y. Watanabe, and T. Iwamoto, *Synthesis of [8]Cycloparaphenylene from a Square-Shaped Tetranuclear Platinum Complex*, Angew. Chem. Int. Ed. **49** (2010) 757.

[320] S. Hitosugi, W. Nakanishi, T. Yamasaki, and H. Isobe, *Bottom-up Synthesis of Finite Models of Helical (n,m)-Single-Wall Carbon Nanotubes*, Nat. Commun. **2** (2011) 492.

[321] M. Hermann, R. Wu, D. C. Grenz, D. Kratzert, H. Li, and B. Esser, *Thioether- and Sulfone-Functionalized Dibenzopentalenes as n-Channel Semiconductors for Organic Field-Effect Transistors*, Journal of Materials Chemistry C **6** (2018) 5420.

[322] H. Jia, G. Zhuang, Q. Huang, J. Wang, Y. Wu, S. Cui, S. Yang, and P. Du, *Synthesis of Giant* Π*-Extended Molecular Macrocyclic Rings as Finite Models of Carbon Nanotubes Displaying Enriched Size-Dependent Physical Properties*, Chem. Eur. J. **26** (2020) 2159.

[323] E. R. Darzi and R. Jasti, *The Dynamic, Size-Dependent Properties of [5]–[12]Cycloparaphenylenes*, Chem. Soc. Rev. **44** (2015) 6401.

[324] H. Omachi, S. Matsuura, Y. Segawa, and K. Itami, *A Modular and Size-Selective Synthesis of [n]Cycloparaphenylenes: A Step toward the Selective Synthesis of [n,n] Single-Walled Carbon Nanotubes*, Angew. Chem. Int. Ed. **49** (2010) 10202.

[325] R. Bandari, T. Höche, A. Prager, K. Dirnberger, and M. R. Buchmeiser, *Ring-Opening Metathesis Polymerization Based Pore-Size-Selective Functionalization of Glycidyl Methacrylate Based Monolithic Media: Access to Size-Stable Nanoparticles for Ligand-Free Metal Catalysis*, Chem. Eur. J. **16** (2010) 4650.

[326] Y. Segawa, H. Omachi, and K. Itami, *Theoretical Studies on the Structures and Strain Energies of Cycloparaphenylenes*, Org. Lett. **12** (2010) 2262.

[327] C. E. Colwell, T. W. Price, T. Stauch, and R. Jasti, *Strain Visualization for Strained Macrocycles*, Chem. Sci. **11** (2020) 3923.

[328] J. Shen, D. Yuan, Y. Qiao, X. Shen, Z. Zhang, Y. Zhong, Y. Yi, and X. Zhu, *Diaceno[a,e]Pentalenes from Homoannulations of o-Alkynylaryliodides Utilizing a Unique Pd(OAc)2/n-Bu4NOAc Catalytic Combination*, Org. Lett. **16** (2014) 4924.

[329] A. Meijere, *Carbon Rich Compounds I*, 1998.

[330] M. Haley, J. Pak, S. Brand, and A. de Meijere), *Carbon Rich Compounds II*, Topics, in: Current Chemistry **201** (1999) 81.

[331] M. M. Haley and R. R. Tykwinski, *Carbon-rich compounds: from molecules to materials*, John Wiley & Sons, 2006.

[332] S. Lei, A. Ver Heyen, S. De Feyter, M. Surin, R. Lazzaroni, S. Rosenfeldt, M. Ballauff, P. Lindner, D. Mössinger, and S. Höger, *Two-Dimensional Oligo (phenylene-ethynylene-butadiynylene) s: All-Covalent Nanoscale Spoked Wheels*, Chem. Eur. J. **15** (2009) 2518.

[333] S. Vanhee, R. Rulkens, U. Lehmann, C. Rosenauer, M. Schulze, W. Köhler, and G. Wegner, *Synthesis and characterization of rigid rod poly (p-phenylenes)*, Macromol. **29** (1996) 5136.

[334] R. May, S.-S. Jester, and S. Höger, *A giant molecular spoked wheel*, J. Am. Chem. Soc. **136** (2014) 16732.

[335] F. Neese, *Software update: The ORCA program system—Version 5.0*, Comput. Mol. Sci. **12** (2022) e1606.

[336] G. Ohlendorf, C. W. Mahler, S.-S. Jester, G. Schnakenburg, S. Grimme, and S. Höger, *Highly strained phenylene bicyclophanes.*, Angew. Chem. **125** (2013).

[337] T. Zimmermann and G. W. Fischer, *Pyryliumverbindungen. 37. Arylbenzene aus 2, 4, 6-Triaryl-pyryliumsalzen und Carbonsäureanhydriden*, J. Prak. Chem. **329** (1987) 975.

[338] C. Mahler, U. Müller, W. M. Müller, V. Enkelmann, C. Moon, G. Brunklaus, H. Zimmermann, and S. Höger, *Synthesis of highly phenylene substituted p-phenylene oligomers from pyrylium salts*, Chem. Comm. (2008) 4816.

[339] D. Grabowski, S. Alef, S. Becker, U. Müller, G. Schnakenburg, and S. Höger, *Condensation of pyrylium salts with mixed anhydrides: aryl ethers, aryl amines and sterically congested aromatics*, Org. Chem. Front. **9** (2022) 294.

[340] T. Hama, S. Ge, and J. F. Hartwig, *Palladium-Catalyzed α-Arylation of Zinc Enolates of Esters: Reaction Conditions and Substrate Scope*, J. Org. Chem. **78** (2013) 8250.

[341] S. K. Weber, F. Galbrecht, and U. Scherf, *Preferential oxidative addition in Suzuki cross-coupling reactions across one fluorene unit*, Org. Lett. **8** (2006) 4039.

[342] D. Schmitz and S. Höger, *Preferential oxidative addition in mixed iodo/bromo quinquephenylenes*, Tetrahedron **70** (2014) 3726.

[343] S. Hoger and K. Bonrad, *(3-Cyanopropyl) dimethylsilyl] acetylene, a Polar Analogue of (Trimethylsilyl) acetylene: Synthesis and Applications in the Preparation of Monoprotected Bisacetylenes*, J. Org. Chem. **65** (2000) 2243.

[344] S. Kotha, E. Brahmachary, and K. Lahiri, *Transition metal catalyzed [2+ 2+ 2] cycloaddition and application in organic synthesis*, Eur. J. Org. Chem. **2005** (2005) 4741.

[345] S. Hecht and J. M. Fréchet, *An alternative synthetic approach toward dendritic macromolecules: novel benzene-core dendrimers via alkyne cyclotrimerization*, J. Am. Chem. Soc. **121** (1999) 4084.

[346] A. P. Cote, A. I. Benin, N. W. Ockwig, M. O'Keeffe, A. J. Matzger, and O. M. Yaghi, *Porous, crystalline, covalent organic frameworks*, Science **310** (2005) 1166.

[347] A. P. Cote, H. M. El-Kaderi, H. Furukawa, J. R. Hunt, and O. M. Yaghi, *Reticular synthesis of microporous and mesoporous 2D covalent organic frameworks*, J. Am. Chem. Soc. **129** (2007) 12914.

[348] D. Jiang, *Covalent organic frameworks: an amazing chemistry platform for designing polymers*, Chem. **6** (2020) 2461.

[349] R.-R. Liang, S.-Y. Jiang, A. Ru-Han, and X. Zhao, *Two-dimensional covalent organic frameworks with hierarchical porosity*, Chem. Soc. Rev. **49** (2020) 3920.

[350] H. R. Abuzeid, A. F. EL-Mahdy, and S.-W. Kuo, *Covalent organic frameworks: Design principles, synthetic strategies, and diverse applications*, Giant **6** (2021) 100054.

[351] K. Geng, T. He, R. Liu, S. Dalapati, K. T. Tan, Z. Li, S. Tao, Y. Gong, Q. Jiang, and D. Jiang, *Covalent organic frameworks: design, synthesis, and functions*, Chem. Rev. **120** (2020) 8814.

[352] H. Li and J.-L. Bredas, *Impact of structural defects on the elastic properties of two-dimensional covalent organic frameworks (2D COFs) under tensile stress*, Chem. Mater. **33** (2021) 4529.

[353] Q. Fang, C. Sui, C. Wang, T. Zhai, J. Zhang, J. Liang, H. Guo, E. Sandoz-Rosado, and J. Lou, *Strong and flaw-insensitive two-dimensional covalent organic frameworks*, Matter **4** (2021) 1017.

[354] X. Zhang, K. Geng, D. Jiang, and G. D. Scholes, *Exciton Diffusion and Annihilation in an sp2 Carbon-Conjugated Covalent Organic Framework*, J. Am. Chem. Soc. **144** (2022) 16423.

[355] W. Wang, H. Wang, X. Tang, J. Huo, Y. Su, C. Lu, Y. Zhang, H. Xu, and C. Gu, *Phenothiazine-based covalent organic frameworks with low exciton binding energies for photocatalysis*, Chem. Sci. **13** (2022) 8679.

[356] S. Bi, C. Yang, W. Zhang, J. Xu, L. Liu, D. Wu, X. Wang, Y. Han, Q. Liang, and F. Zhang, *Two-dimensional semiconducting covalent organic frameworks via condensation at arylmethyl carbon atoms*, Nat. Commun. **10** (2019) 1.

[357] T. Li, W.-D. Zhang, Y. Liu, Y. Li, C. Cheng, H. Zhu, X. Yan, Z. Li, and Z.-G. Gu, *A two-dimensional semiconducting covalent organic framework with nickel (II) coordination for high capacitive performance*, J. Mater. Chem. A **7** (2019) 19676.

[358] A. M. Evans, A. Giri, V. K. Sangwan, S. Xun, M. Bartnof, C. G. Torres-Castanedo, H. B. Balch, M. S. Rahn, N. P. Bradshaw, E. Vitaku, et al., *Thermally conductive ultra-low-k dielectric layers based on two-dimensional covalent organic frameworks*, Nat. Mater. **20** (2021) 1142.

[359] A. Giri, A. M. Evans, M. A. Rahman, A. J. McGaughey, and P. E. Hopkins, *Highly Negative Poisson's Ratio in Thermally Conductive Covalent Organic Frameworks*, ACS Nano **16** (2022) 2843.

[360] Z. Jiang, R. Dong, A. M. Evans, N. Biere, M. A. Ebrahim, S. Li, D. Anselmetti, W. R. Dichtel, and A. G. Livingston,
Aligned macrocycle pores in ultrathin films for accurate molecular sieving,
Nature **609** (2022) 58.

[361] N. A. Khan, R. Zhang, X. Wang, L. Cao, C. S. Azad, C. Fan, J. Yuan, M. Long, H. Wu, M. Olson, et al., *Assembling covalent organic framework membranes via phase switching for ultrafast molecular transport*, Nat. Commun. **13** (2022) 1.

[362] R. Shevate and D. L. Shaffer, *Large-Area 2D Covalent Organic Framework Membranes with Tunable Single-Digit Nanopores for Predictable Mass Transport*, ACS Nano **16** (2022) 2407.

[363] B. Lukose, A. Kuc, and T. Heine, *The structure of layered covalent-organic frameworks*, Chem. Eur. J. **17** (2011) 2388.

[364] C. Kessler, R. Schuldt, S. Emmerling, B. V. Lotsch, J. Kästner, J. Gross, and N. Hansen, *Influence of layer slipping on adsorption of light gases in covalent organic frameworks: A combined experimental and computational study*,
Microporous Mesoporous Mater. **336** (2022) 111796.

[365] S. A. Ahmed, Q.-B. Liao, Q. Shen, M. M. F. Ashraf Baig, J. Zhou, C.-F. Shi, P. Muhammad, S. Hanif, K. Xi, X.-H. Xia, et al.,
pH-Dependent Slipping and Exfoliation of Layered Covalent Organic Framework,
Chem. Eur. J. **26** (2020) 12996.

[366] S. Jhulki, C. H. Feriante, R. Mysyk, A. M. Evans, A. Magasinski, A. S. Raman, K. Turcheniuk, S. Barlow, W. R. Dichtel, G. Yushin, et al.,
A Naphthalene Diimide Covalent Organic Framework: Comparison of Cathode Performance in Lithium-Ion Batteries with Amorphous Cross-linked and Linear Analogues, and Its Use in Aqueous Lithium-Ion Batteries, ACS Appl. Energy Mater. **4** (2021) 350.

[367] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, et al.,
The atomic simulation environment—a Python library for working with atoms,
J. Phys. Condens. Matter **29** (2017) 273002.

[368] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, *Avogadro: an advanced semantic chemical editor, visualization, and analysis platform*,
J. Cheminformatics **4** (2012) 17.

[369] `http://www.jp-minerals.org/vesta/en/download.html`.

[370] K. Momma and F. Izumi,
VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data,
J. Appl. Crystallogr. **44** (2011) 1272.

[371] e. a. T. Williams C. Kelley, *Gnuplot 5.2: an interactive plotting program*,
`http://gnuplot.sourceforge.net/`, 2020.

[372] `Inkscape` Version 0.92.5, **2020**, `https://inkscape.org`. Accessed 2022-07-17.

[373] "Density Functional Tight Binding Plus `dftb+`",
`https://github.com/dftbplus/dftbplus`. Accessed: 2022-08-16.

[374] "Semiempirical Extended Tight-Binding Program Package `xtb`", `https://github.com/grimme-lab/xtb`. Accessed: 2022-06-20.

[375] R. R. Nazmutdinov and J. Ulstrup, *Retrospective and prospective views of electrochemical electron transfer processes: theory and computations*, Atomic-Scale Modelling of Electrochemical Systems (2021) 25.

[376] P. Ramos, M. Mankarious, and M. Pavanello, *A critical look at methods for calculating charge transfer couplings fast and accurately*, Practical Aspects of Computational Chemistry IV (2016) 103.

[377] A. Y. Sosorev, *Simple charge transport model for efficient search of high-mobility organic semiconductor crystals*, Mater. Des. **192** (2020) 108730.

[378] H. Dong, X. Fu, J. Liu, Z. Wang, and W. Hu, *25th anniversary article: key points for high-mobility organic field-effect transistors*, Adv. Mater. **25** (2013) 6158.

[379] B. Baumeier, F. May, C. Lennartz, and D. Andrienko, *Challenges for in silico design of organic semiconductors*, J. Mater. Chem. **22** (2012) 10971.

[380] W. Brütting, *Introduction to the physics of organic semiconductors*, Physics of organic semiconductors (2005) 1.

[381] H. Bässler and A. Köhler, *Charge transport in organic semiconductors*, Unimolecular and Supramolecular Electronics I: Chemistry and Physics Meet at Metal-Molecule Interfaces (2012) 1.

[382] A. Nenashev, J. Oelerich, and S. Baranovskii, *Theoretical tools for the description of charge transport in disordered organic semiconductors*, J. Phys. Condens. Matter **27** (2015) 093201.

[383] S. D. Baranovskii, *Mott lecture: Description of charge transport in disordered organic semiconductors: Analytical theories and computer simulations*, Phys. Status Solidi **215** (2018) 1700676.

[384] S. Di Motta, E. Di Donato, F. Negri, G. Orlandi, D. Fazzi, and C. Castiglioni, *Resistive molecular memories: influence of molecular parameters on the electrical bistability*, J. Am. Chem. Soc. **131** (2009) 6591.

[385] S. Di Motta, M. Siracusa, and F. Negri, *Structural and thermal effects on the charge transport of core-twisted chlorinated perylene bisimide semiconductors*, J. Phys. Chem. C **115** (2011) 20754.

[386] I. Kondov, M. Čížek, C. Benesch, H. Wang, and M. Thoss, *Quantum Dynamics of Photoinduced Electron-Transfer Reactions in Dye- Semiconductor Systems: First-Principles Description and Application to Coumarin 343- TiO2*, J. Phys. Chem. C **111** (2007) 11970.

[387] Z. Futera and J. Blumberger, *Electronic couplings for charge transfer across molecule/metal and molecule/semiconductor interfaces: Performance of the projector operator-based diabatization approach*, J. Phys. Chem. C **121** (2017) 19677.

[388] P. Ramos, M. Papadakis, and M. Pavanello,
Performance of frozen density embedding for modeling hole transfer reactions,
J. Phys. Chem. B **119** (2015) 7541.

[389] H. Ren, M. R. Provorse, P. Bao, Z. Qu, and J. Gao,
Multistate density functional theory for effective diabatic electronic coupling,
J. Phys. Chem. Lett. **7** (2016) 2286.

[390] C.-I. Wang, I. Joanito, C.-F. Lan, and C.-P. Hsu,
Artificial neural networks for predicting charge transfer coupling,
J. Chem. Phys. **153** (2020) 214113.

[391] R. Hafizi, J. Elsner, and J. Blumberger,
Ultrafast Electronic Coupling Estimators: Neural Networks versus Physics-Based Approaches,
J. Chem. Theory Comput. (2023).

[392] A. Aggarwal, V. Vinayak, S. Bag, C. Bhattacharyya, U. V. Waghmare, and P. K. Maiti,
Predicting the DNA conductance using a deep feedforward neural network model,
J. Chem. Inf. Model. **61** (2020) 106.

[393] A. Pal, L. K. Wen, C. Y. Jun, I. Jeon, Y. Matsuo, and S. Manzhos,
Comparative density functional theory–density functional tight binding study of fullerene derivatives: effects due to fullerene size, addends, and crystallinity on band structure, charge transport and optical properties, Phys. Chem. Chem. Phys. **19** (2017) 28330.

[394] A. Troisi and G. Orlandi,
Dynamics of the intermolecular transfer integral in crystalline organic semiconductors,
J. Phys. Chem. A **110** (2006) 4065.

[395] A. Landi, A. Peluso, and A. Troisi,
Quantitative Prediction of the Electro-Mechanical Response in Organic Crystals,
Adv. Mater. **33** (2021) 2008049.

[396] A. Kubas, F. Gajdos, A. Heck, H. Oberhofer, M. Elstner, and J. Blumberger,
Electronic couplings for molecular charge transfer: benchmarking CDFT, FODFT and FODFTB against high-level ab initio calculations. II,
Phys. Chem. Chem. Phys. **17** (2015) 14342.

[397] A. Kubas, F. Hoffmann, A. Heck, H. Oberhofer, M. Elstner, and J. Blumberger,
Electronic couplings for molecular charge transfer: Benchmarking CDFT, FODFT, and FODFTB against high-level ab initio calculations, J. Chem. Phys. **140** (2014) 104105.

[398] H. Kitoh-Nishioka and K. Ando, *Calculation of Charge-Transfer Electronic Coupling with Nonempirically Tuned Range-Separated Density Functional*,
J. Phys. Chem. C **123** (2019) 11351.

[399] A. Pershin and P. G. Szalay, *Improving the Accuracy of the Charge Transfer Integrals Obtained by Coupled Cluster Theory, MBPT (2), and TDDFT*,
J. Chem. Theory Comput. **11** (2015) 5705.

[400] M. Pavanello, T. Van Voorhis, L. Visscher, and J. Neugebauer, *An accurate and linear-scaling method for calculating charge-transfer excitation energies and diabatic couplings*,
J. Chem. Phys. **138** (2013) 054101.

[401] M. C. Zerner, *Reviews in computational chemistry*,
p. 313-365, KB Lipkowitz, DB Boyd Eds., VCH Publishers, Inc (2007).

[402] M. Gaus, Q. Cui, and M. Elstner, *DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB)*,
J. Chem. Theory Comput. **7** (2011) 931.

[403] M. Kubillus, T. Kubar, M. Gaus, J. Rezac, and M. Elstner, *Parameterization of the DFTB3 method for Br, Ca, Cl, F, I, K, and Na in organic and biological systems*,
J. Chem. Theory Comput. **11** (2015) 332.

[404] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji, et al., *Gaussian 16 Revision C. 01. 2016; Gaussian Inc*, Wallingford CT **421** (2016).

[405] G. Te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. van Gisbergen, J. G. Snijders, and T. Ziegler, *Chemistry with ADF*, J. Comp. Chem. **22** (2001) 931.

[406] G. Te Velde, F. Bickelhaupt, E. Baerends, et. al., "`ADF 2020.102`, SCM", Theoretical Chemistry - Vrije Universiteit Amsterdam. `http://www.scm.com`.

[407] L. Brooker, G. Keyes, R. Sprague, R. VanDyke, E. VanLare, G. VanZandt, F. White, H. Cressman, and S. Dent Jr, *Color and constitution. X. 1 Absorption of the merocyanines2*, J. Am. Chem. Soc. **73** (1951) 5332.

[408] G. Chamberlain, P. Cooney, and S. Dennison,
Photovoltaic properties of merocyanine solid-state photocells, Nature **289** (1981) 45.

[409] F. Würthner and K. Meerholz, *Systems chemistry approach in organic photovoltaics*, Chem. Eur. J. **16** (2010) 9366.

[410] A. Liess, L. Huang, A. Arjona-Esteban, A. Lv, M. Gsänger, V. Stepanenko, M. Stolte, and F. Würthner,
Organic Thin Film Transistors Based on Highly Dipolar Donor–Acceptor Polymethine Dyes, Adv. Func. Mater. **25** (2015) 44.

[411] C. J. MacNevin, D. Gremyachinskiy, C.-W. Hsu, L. Li, M. Rougie, T. T. Davis, and K. M. Hahn, *Environment-sensing merocyanine dyes for live cell imaging applications*, Bioconjugate Chem. **24** (2013) 215.

[412] E. M. Santos, T. Berbasova, W. Wang, R. E. Salmani, W. Sheng, C. Vasileiou, J. H. Geiger, and B. Borhan,
Engineering of a Red Fluorogenic Protein/Merocyanine Complex for Live-Cell Imaging, Chem. Bio Chem. **21** (2020) 723.

[413] Y. Shiraishi, M. Itoh, and T. Hirai, *Thermal isomerization of spiropyran to merocyanine in aqueous media and its application to colorimetric temperature indication*,
Phys. Chem. Chem. Phys. **12** (2010) 13737.

[414] F. L. Coelho, R. da Costa Duarte, C. de Ávila Braga, J. M. Toldo, P. F. B. Gonçalves, F. da Silveira Santos, and F. S. Rodembusch,
Benzothiazole merocyanine dyes as middle pH optical sensors,
Dyes Pigm. **176** (2020) 108193.

[415] H. Hisamoto, H. Tohma, T. Yamada, K.-i. Yamauchi, D. Siswanta, N. Yoshioka, and K. Suzuki, *Molecular design, characterization, and application of multi-information dyes for multi-dimensional optical chemical sensing. Molecular design concepts of the dyes and their fundamental spectral characteristics*, Anal. Chim. Acta **373** (1998) 271.

[416] H. Shindy, A. Khalafalla, M. Goma, and A. Eed, *Synthesis, photosensitization and antimicrobial activity evaluation of some novel Merocyanine dyes*, Chem. Int. **2** (2016) 114.

[417] M. R. Detty, S. L. Gibson, and S. J. Wagner, *Current clinical and preclinical photosensitizers for use in photodynamic therapy*, J. Med. Chem. **47** (2004) 3897.

[418] F. Sieber, J. M. O'Brien, G. J. Krueger, S. L. Schober, W. H. Burns, S. J. Sharkis, and L. L. Sensenbrenner, *Antiviral activity of merocyanine 540*, Photochem. Photobiol. **46** (1987) 707.

[419] H.-Y. Lin, C.-T. Chen, and C.-T. Huang, *Use of merocyanine 540 for photodynamic inactivation of Staphylococcus aureus planktonic and biofilm cells*, Appl. Environ. Microbiol. **70** (2004) 6453.

[420] B. Tirri, G. Mazzone, A. Ottochian, J. Gomar, U. Raucci, C. Adamo, and I. Ciofini, *A combined Monte Carlo/DFT approach to simulate UV-vis spectra of molecules and aggregates: Merocyanine dyes as a case study*, J. Comp. Chem. **42** (2021) 1054.

[421] C. Brückner, C. Walter, M. Stolte, B. Braida, K. Meerholz, F. Würthner, and B. Engels, *Structure–Property Relationships for Exciton and Charge Reorganization Energies of Dipolar Organic Semiconductors: A Combined Valence Bond Self-Consistent Field and Time-Dependent Hartree-Fock and DFT Study of Merocyanine Dyes*, J. Phys. Chem. C **119** (2015) 17602.

[422] L. Lescos, S. P. Sitkiewicz, P. Beaujean, M. Blanchard-Desce, B. Champagne, E. Matito, and F. Castet, *Performance of DFT functionals for calculating the second-order nonlinear optical properties of dipolar merocyanines*, Phys. Chem. Chem. Phys. **22** (2020) 16579.

[423] Y. Siqueira, M. L. Lyra, T. N. Ramos, B. Champagne, and V. Manzoni, *Unveiling the relationship between structural and polarization effects on the first hyperpolarizability of a merocyanine dye*, J. Chem. Phys. **156** (2022) 014305.

[424] M. E. Reish, A. J. Kay, A. Teshome, I. Asselberghs, K. Clays, and K. C. Gordon, *Testing computational models of hyperpolarizability in a merocyanine dye using spectroscopic and DFT methods*, J. Phys. Chem. A. **116** (2012) 5453.

[425] F. Castet, M. Blanchard-Desce, F. Adamietz, Y. M. Poronik, D. T. Gryko, and V. Rodriguez, *Experimental and Theoretical Investigation of the First-Order Hyperpolarizability of Octupolar Merocyanine Dyes*, Chem. Phys. Chem. **15** (2014) 2575.

[426] M. D. Allendorf, A. Schwartzberg, V. Stavila, and A. A. Talin, *A roadmap to implementing metal–organic frameworks in electronic devices: challenges and critical directions*, Chem. Eur. J. **17** (2011) 11372.

[427] M. Souto, K. Strutyński, M. Melle-Franco, and J. Rocha, *Electroactive organic building blocks for the chemical design of functional porous frameworks (MOFs and COFs) in electronics*, Chem. Eur. J. **26** (2020) 10912.

[428] M. Wang, R. Dong, and X. Feng, *Two-dimensional conjugated metal–organic frameworks (2D c-MOFs): chemistry and function for MOFtronics*, Chem. Soc. Rev. **50** (2021) 2764.

[429] N. Kishi, Z. Li, K. Yoza, M. Akita, and M. Yoshizawa, *An M2L4 molecular capsule with an anthracene shell: encapsulation of large guests up to 1 nm*, J. Am. Chem. Soc. **133** (2011) 11438.

[430] H. Chen and J. Fraser Stoddart, *From molecular to supramolecular electronics*, Nat. Rev. Mater. **6** (2021) 804.

[431] V. Balzani, A. Credi, and M. Venturi, *Molecular logic circuits*, Chem. Phys. Chem. **4** (2003) 49.

[432] D. Vuillaume, *Molecular nanoelectronics*, Proc. Inst. Electr. Electron. Eng. **98** (2010) 2111.

[433] Z. Zhu, Y. Guo, and Y. Liu, *Application of organic field-effect transistors in memory*, Mater. Chem. Front. **4** (2020) 2845.

[434] S. Varghese, J. A. Elemans, A. E. Rowan, and R. J. Nolte, *Molecular computing: paths to chemical Turing machines*, Chem. Sci. **6** (2015) 6050.

[435] W. Chang and A. V. Vasilakos, *Molecular Computing*, Springer, 2014.

[436] X. Liang, W. Zhu, Z. Lv, and Q. Zou, *Molecular computing and bioinformatics*, Molecules **24** (2019) 2358.

[437] Y. Huang, E.-L. Hsiang, M.-Y. Deng, and S.-T. Wu, *Mini-LED, Micro-LED and OLED displays: present status and future perspectives*, Light Sci. Appl. **9** (2020) 105.

[438] M. P. Cenci, T. Scarazzato, D. D. Munchen, P. C. Dartora, H. M. Veit, A. M. Bernardes, and P. R. Dias, *Eco-friendly electronics—a comprehensive review*, Adv. Mater. Technol. **7** (2022) 2001263.

[439] I. McCulloch, M. Chabinyc, C. Brabec, C. B. Nielsen, and S. E. Watkins, *Sustainability considerations for organic electronic products*, Nat. Mater. **22** (2023) 1304.

[440] J. Ajayan, D. Nirmal, P. Mohankumar, M. Saravanan, M. Jagadesh, and L. Arivazhagan, *A review of photovoltaic performance of organic/inorganic solar cells for future renewable and sustainable energy technologies*, Superlattices Microstruct. **143** (2020) 106549.

[441] R. Chawla, P. Singhal, and A. K. Garg, *Photovoltaic review of all generations: environmental impact and its market potential*, Trans. Electr. Electron. Mater. **21** (2020) 456.

[442] N. A. Nordin, M. N. M. Ansari, S. M. Nomanbhay, N. A. Hamid, N. M. Tan, Z. Yahya, and I. Abdullah, *Integrating photovoltaic (PV) solar cells and supercapacitors for sustainable energy devices: a review*, Energies **14** (2021) 7211.

[443] A. Jain, R. Kothari, V. Tyagi, R. K. Rajamony, M. S. Ahmad, H. M. Singh, S. Raina, and A. Pandey, *Advances in organic solar cells: Materials, progress, challenges and amelioration for sustainable future*, Sustain. Energy Technol. Assess. **63** (2024) 103632.

[444] B. Sun and Y. Li, *Ubiquitous clean and sustainable energy-driven self-rechargeable batteries realized by and used in organic electronics*, J. Mater. Chem. **10** (2022) 388.

[445] D. Pan, J. Hu, B. Wang, X. Xia, Y. Cheng, C.-H. Wang, and Y. Lu,
Biomimetic Wearable Sensors: Emerging Combination of Intelligence and Electronics,
Adv. Sci. **11** (2024) 2303264.

[446] K.-N. Kim, M.-J. Sung, H.-L. Park, and T.-W. Lee,
Organic synaptic transistors for bio-hybrid neuromorphic electronics,
Adv. Electron. Mater. **8** (2022) 2100935.

[447] L. Luo and Z. Liu, *Recent progress in organic field-effect transistor-based chem/bio-sensors*,
View **3** (2022) 20200115.

[448] S. G. Surya, H. N. Raval, R. Ahmad, P. Sonar, K. N. Salama, and V. R. Rao, *Organic field effect transistors (OFETs) in environmental sensing and health monitoring: A review*,
Trends Analyt. Chem. **111** (2019) 27.

[449] W. Shi, Y. Guo, and Y. Liu, *When flexible organic field-effect transistors meet biomimetics: a prospective view of the internet of things*, Adv. Mater. **32** (2020) 1901493.

[450] C. S. Buga and J. C. Viana,
A review on materials and technologies for organic large-area electronics,
Adv. Mater. Technol. **6** (2021) 2001016.

[451] P. Friederich, A. Fediai, S. Kaiser, M. Konrad, N. Jung, and W. Wenzel,
Toward design of novel materials for organic electronics, Adv. Mater. **31** (2019) 1808256.

[452] A. Saeki and K. Kranthiraja, *A high throughput molecular screening for organic electronics via machine learning: present status and perspective*, Jpn. J. Appl. Phys. **59** (2020) SD0801.

[453] Ö. H. Omar, M. Del Cueto, T. Nematiaram, and A. Troisi, *High-throughput virtual screening for organic electronics: a comparative study of alternative strategies*,
J. Mater. Chem. **9** (2021) 13557.

[454] A. Zhugayevych and S. Tretiak, *Theoretical description of structural and electronic properties of organic photovoltaic materials*, Annu. Rev. Phys. Chem. **66** (2015) 305.

[455] C. Duan, F. Liu, A. Nandy, and H. J. Kulik, *Putting density functional theory to the test in machine-learning-accelerated materials discovery*, J. Phys. Chem. Lett. **12** (2021) 4628.

[456] C. Duan, A. Nandy, and H. J. Kulik,
Machine learning for the discovery, design, and engineering of materials,
Ann. Rev. Chem. Biomol. Eng. **13** (2022) 405.

[457] C. S. Sarap, Y. Singh, J. M. Lane, and N. Rai,
Electronic properties and optical spectra of donor–acceptor conjugated organic polymers,
Sci. Rep. **13** (2023) 21587.

[458] Q. Zhu and S. Hattori,
Organic crystal structure prediction and its application to materials design,
J. Mater. Res. **38** (2023) 19.

[459] J. E. Arnold and G. M. Day, *Crystal Structure Prediction of Energetic Materials*,
Cryst. Growth Des. **23** (2023) 6149.

[460] D. H. Bowskill, I. J. Sugden, S. Konstantinopoulos, C. S. Adjiman, and C. C. Pantelides,
Crystal structure prediction methods for organic molecules: State of the art,
Annu. Rev. Chem. Biomol. Eng. **12** (2021) 593.

[461] S. Spicher, C. Plett, P. Pracht, A. Hansen, and S. Grimme, *Automated molecular cluster growing for explicit solvation by efficient force field and tight binding methods*, J. Chem. Theory Comput. **18** (2022) 3174.

[462] G. N. Simm, P. L. Türtscher, and M. Reiher, *Systematic microsolvation approach with a cluster-continuum scheme and conformational sampling*, J. Comput. Chem. **41** (2020) 1144.

[463] R. L. Johnston, *Evolving better nanoparticles: Genetic algorithms for optimising cluster geometries*, Dalton Trans. (2003) 4193.

[464] L.-N. Fu, B. Leng, Y.-S. Li, and X.-K. Gao, *Photoresponsive organic field-effect transistors involving photochromic molecules*, Chin. Chem. Lett. **27** (2016) 1319.

[465] A. Liess, A. Arjona-Esteban, A. Kudzus, J. Albert, A.-M. Krause, A. Lv, M. Stolte, K. Meerholz, and F. Würthner, *Ultranarrow bandwidth organic photodiodes by exchange narrowing in merocyanine H-and J-aggregate excitonic systems*, Adv. Func. Mater. **29** (2019) 1805058.

[466] L. Huang, M. Stolte, H. Bürckstümmer, and F. Würthner, *High-performance organic thin-film transistor based on a dipolar organic semiconductor.*, Adv. Mater. **24** (2012) 5750.

[467] N. M. Kronenberg, V. Steinmann, H. Bürckstümmer, J. Hwang, D. Hertel, F. Würthner, and K. Meerholz, *Direct comparison of highly efficient solution-and vacuum-processed organic solar cells based on merocyanine dyes.*, Adv. Mater. **22** (2010) 4193.

[468] J. Witte, J. B. Neaton, and M. Head-Gordon, *Effective empirical corrections for basis set superposition error in the def2-SVPD basis: gCP and DFT-C*, J. Chem. Phys. **146** (2017).

[469] "Semiempirical Extended Tight-Binding Program Package `xtb`", `https://github.com/grimme-lab/xtb`. Accessed: 2023-12-01.

[470] A. Arjona-Esteban, J. Krumrain, A. Liess, M. Stolte, L. Huang, D. Schmidt, V. Stepanenko, M. Gsänger, D. Hertel, K. Meerholz, et al., *Influence of solid-state packing of dipolar merocyanine dyes on transistor and solar cell performances*, J. Am. Chem. Soc. **137** (2015) 13524.

[471] A. Liess, L. Huang, A. Arjona-Esteban, A. Lv, M. Gsänger, V. Stepanenko, M. Stolte, and F. Würthner, *Organic thin film transistors based on highly dipolar donor–acceptor polymethine dyes*, Adv. Func. Mater. **25** (2015) 44.

[472] H. Bürckstümmer, E. V. Tulyakova, M. Deppisch, M. R. Lenze, N. M. Kronenberg, M. Gsänger, M. Stolte, K. Meerholz, and F. Würthner, *Efficient solution-processed bulk heterojunction solar cells by antiparallel supramolecular arrangement of dipolar donor–acceptor dyes*, Angew. Chem. **123** (2011) 11832.

[473] R. Tomar, L. Bernasconi, D. Fazzi, and T. Bredow, *Theoretical Study on the Optoelectronic Properties of Merocyanine-Dyes*, J. Phys. Chem. **127** (2023) 9661.

[474] A. J. Kny, M. Reimer, N. Al-Shamery, R. Tomar, T. Bredow, S. Olthof, D. Hertel, K. Meerholz, and M. Sokolowski, *Chiral self-organized single 2D-layers of tetramers from a functional donor–acceptor molecule by the surface template effect*, Nanoscale **15** (2023) 10319.

[475] R. Schäfer, L. Böhner, M. Schiek, D. Hertel, K. Meerholz, and K. Lindfors, *Strong Light–Matter Interaction of Molecular Aggregates with Two Excitonic Transitions*, ACS Photonics **11** (2023) 111.

[476] B. Chandramouli, S. Del Galdo, M. Fusè, V. Barone, and G. Mancini, *Two-level stochastic search of low-energy conformers for molecular spectroscopy: implementation and validation of MM and QM models*, Phys. Chem. Chem. Phys. **21** (2019) 19921.

[477] N. J. King, I. LeBlanc, and A. Brown, *A Variant on the CREST iMTD Algorithm for Non-Covalent Clusters of Flexible Molecules* (2024).

[478] L. Cheng, W. Cai, and X. Shao, *A conformational analysis method for understanding the energy landscapes of clusters*, Chem. Phys. Chem. **8** (2007) 569.

[479] J. Zhang and V.-A. Glezakou, *Global optimization of chemical cluster structures: Methods, applications, and challenges*, Int. J. Quantum Chem. **121** (2021) e26553.

[480] Q. Ma and H. J. Werner, *Scalable Electron Correlation Methods. 8. Explicitly Correlated Open-Shell Coupled-Cluster with Pair Natural Orbitals PNO-RCCSD(T)-F12 and PNO-UCCSD(T)-F12*, J. Chem. Theory Comput. **17** (2021) 902.

[481] T. H. Dunning, *Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen*, J. Chem. Phys. **90** (1989) 1007.

[482] *The Molpro quantum chemistry package*, J. Chem. Phys. **152** (2020) 144107.

[483] H.-J. Werner et al., *MOLPRO, 2022.3 , a package of ab initio programs*.

[484] F. Weigend and R. Ahlrichs, *Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy*, Phys. Chem. Chem. Phys. **7** (2005) 3297.

[485] N. Mardirossian and M. Head-Gordon, *ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation*, J. Chem. Phys. **144** (2016).

[486] G. Santra, N. Sylvetsky, and J. M. Martin, *Minimally empirical double-hybrid functionals trained against the GMTKN55 database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4*, J. Phys. Chem. A **123** (2019) 5129.

[487] J. J. Stewart, *Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements*, J. Mol. Model. **13** (2007) 1173.

[488] P. S Brahmkshatriya, P. Dobes, J. Fanfrlik, J. Rezac, K. Paruch, A. Bronowska, M. Lepsik, and P. Hobza, *Quantum mechanical scoring: structural and energetic insights into cyclin-dependent kinase 2 inhibition by pyrazolo [1, 5-a] pyrimidines*, Curr. Comput. Aided Drug Des. **9** (2013) 118.

[489] J. Rezac and P. Hobza, *Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods*, J. Chem. Theory Comput. **8** (2012) 141.

[490] T. A. Halgren, *Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94*, J. Comput. Chem. **17** (1996) 490.

[491] J. J. P. Stewart, *MOPAC2016, Version: 19.179L*, http://OpenMOPAC.net, 2018.

[492] R. Ahlrichs, M. Armbruster, M. Bär, et. al., "*TURBOMOLE 7.6* Program Package" **2023**, Universität Karlsruhe. `http://www.turbomole.com`.

[493] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, *Open Babel: An open chemical toolbox*, J. Cheminform. **3** (2011) 1.

[494] F. Neese, *An improvement of the resolution of the identity approximation for the formation of the Coulomb matrix*, J. Comp. Chem. **24** (2003) 1740.

[495] F. Neese, F. Wennmohs, A. Hansen, and U. Becker, *Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree–Fock exchange*, Chem. Phys. **356** (2009) 98.

[496] F. Neese, T. Petrenko, C. Riplinger, et. al., "`ORCA V.5.0.4.`" **2023**, Universität Mühlheim. `https://cec.mpg.de`.

[497] G. Schaftenaar and J. H. Noordik, *Molden: a pre-and post-processing program for molecular and electronic structures*, J. Comp.-Aided Mol. Design **14** (2000) 123.

[498] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, *UCSF Chimera—a visualization system for exploratory research and analysis*, J. Comp. Chem. **25** (2004) 1605.

[499] H.-J. Werner and A. Hansen, *Accurate Calculation of Isomerization and Conformational Energies of Larger Molecules Using Explicitly Correlated Local Coupled Cluster Methods in Molpro and ORCA*, J. Chem. Theory Comput. **19** (2023) 7007.

[500] T. Gasevic, M. Bursch, Q. Ma, S. Grimme, H.-J. Werner, and A. Hansen, *The p-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations*, Phys. Chem. Chem. Phys. **26** (18 2024) 13884.

[501] S. Spicher, E. Caldeweyher, A. Hansen, and S. Grimme, *Benchmarking London dispersion corrected density functional theory for noncovalent ion–π interactions*, Phys. Chem. Chem. Phys. **23** (2021) 11635.

[502] C. Plett, S. Grimme, and A. Hansen, *Conformational energies of biomolecules in solution: Extending the MPCONF196 benchmark with explicit water molecules*, J. Comp. Chem. (2024).

[503] H. Mun, W. Lorpaiboon, and J. Ho,
In Search of the Best Low-Cost Methods for Efficient Screening of Conformers,
J. Phys. Chem. A (2024).

[504] A. Karton, *Catalysis on pristine 2D materials via dispersion and electrostatic interactions*,
J. Phys. Chem. A **124** (2020) 6977.

[505] L. Wittmann, H. Neugebauer, S. Grimme, and M. Bursch,
Dispersion-corrected r2SCAN based double-hybrid functionals, J. Chem. Phys. **159** (2023).

[506] D. Bialas, C. Zhong, F. Wurthner, and F. C. Spano, *Essential states model for merocyanine dye stacks: bridging electronic and optical absorption properties*,
J. Phys. Chem. C **123** (2019) 18654.

[507] L. Craciunescu, M. Asbach, S. Wirsing, S. Hammer, F. Unger, K. Broch, F. Schreiber, G. Witte, A. Dreuw, P. Tegeder, et al., *Cluster-Based Approach Utilizing Optimally Tuned TD-DFT to Calculate Absorption Spectra of Organic Semiconductor Thin Films*,
J. Chem. Theory Comput. **19** (2023) 9369.

[508] N. J. Hestand and F. C. Spano, *Expanded theory of H-and J-molecular aggregates: the effects of vibronic coupling and intermolecular charge transfer*, Chem. Rev. **118** (2018) 7069.

List of Figures

List of Tables

Acknowledgements

During the course of my Ph.D. studies I encountered great support and here I would like to take the opportunity to thank al those involved.
The first and biggest acknowledgment goes to my supervisor Stefan Grimme who hosted me for over six years now in his working group with financial and professional support, including bachelor's and master's thesis, and gave me the opportunity to work in quantum chemistry and introduced me to the doctoral training program TIDE. Great acknowledgments also for Andreas Hansen who took over the duties for project supervision and first contact person.
Big thanks also go to Thomas Bredow for agreeing to be the second reviewer of my thesis, Sigurd Höger as the third reviewer, and Stilla Frede as the forth reviewer.
TIDE - Template-designed Organic Electronics - is a research training group for doctoral students by the universities of Bonn and Cologne funded by the DFG. I spent my whole Ph.D. as part of TIDE. This included not only intensive scientific and soft skill training but also team building activities, funding, and international conference visits. Interdisciplinary exchange, networking, and collaborations were fostered and a strongly bonded community of young scientists in an interesting scientific focus area was created. TIDE gave me the opportunity to spend four month abroad as a visiting researcher in the lab of Jean-Luc Brédas at the University of Arizona, a fascinating and wonderful time of scientific and personal growth. At this point, many thanks to Jean-Luc Brédas and Hong Li who supervised me there and welcomed me in their working group. In the further TIDE context I would like to thank the organizers Klaus Meerholz and Arne Lützen, and the different coordinators who took over most of the paper work, and my supervision committee consisting of Stefan Grimme, Andreas Hansen, and Daniele Fazzi. And of course all members of TIDE for the great time together, especially the students from the focus area E - Theory, Nora Gildemeister and Ritu Tomar.
Great science is most often a collaborative effort and I would like to thank all my cooperation partners, Nora Gildemeister, Simon Rickert, Kristin Gratzfeld, Austin Evans, as well as various members from the working groups of Sigurd Höger, Thomas Bredow, Stefan Grimme, Sven-Stefan Jester, Jean-Luc Brédas, and Birgit Esser.
Studying is the one part but a good working atmosphere is the other part. I thank all members of my working group for the nice and supportive time, helpful discussions, and pronounced group spirit.
A big thank-you also goes to all the proofreaders of my thesis, especially Thomas Froitzheim, Hagen Neugebauer, Nora Gildemeister, Benedikt Bädorf, and Andreas Hansen.
Special thanks go to my office mate Johannes Gorges.
For administrative and technical support, I want to thank Claudia Kronz and Jens Mekelburger.
Outside the scientific community, I want to thank my family and sports mates for their loving support, without which this work would have never been possible.