Efficient Modeling for Large-Scale Chemical Problems: Force Field Development and Screening Workflows

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

von
Thomas Rose
aus
Haan

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:	11.07.2025
Erscheinungsjahr:	2025

Publications

The following publications are included in this thesis.

1. S. Grimme and T. Rose, *mcGFN-FF: an accurate force field for optimization and energetic screening of molecular crystals*, Z. Naturforsch., B **79** (2024) 191, DOI: 10.1515/znb-2023-0088

2. T. Rose, M. Bursch, J.-M. Mewes, and S. Grimme, *Fast and Robust Modeling of Lanthanide and Actinide Complexes, Biomolecules, and Molecular Crystals with the Extended GFN-FF Model*, Inorg. Chem. **63** (2024) 19364, DOI: 10.1021/acs.inorgchem.4c03215

3. G. A. Lara-Cruz, T. Rose, S. Grimme, and A. Jaramillo-Botero, *Reaction-Free Energies for Complexation of Carbohydrates by Tweezer Diboronic Acids*, J. Phys. Chem. B **128** (2024) 9213, DOI: 10.1021/acs.jpcb.4c04846

4. A. Arissa, T. Rose, N. Leick, S. Grimme, J. C. Johnson, and J. V. Lockard, *Charge Transfer and Recombination Pathways through Fullerene Guests in Porphyrin-Based MOFs*, J. Phys. Chem. C **129** (2025) 8215, DOI: 10.1021/acs.jpcc.5c00161

Additionally, significant contributions have been made to the following publications.

5. A. Katbashev, M. Stahn, T. Rose, V. Alizadeh, M. Friede, C. Plett, P. Steinbach, and S. Ehlert, *Overview on Building Blocks and Applications of Efficient and Robust Extended Tight Binding*, J. Phys. Chem. A **129** (2025) 2667, DOI: 10.1021/acs.jpca.4c08263

6. M. Thiele, T. Rose, M. Lõkov, S. Stadtfeld, S. Tshepelevitsh, E. Parman, K. Opara, C. Wölper, I. Leito, S. Grimme, and J. Niemeyer, *Multifunctional Organocatalysts - Singly-Linked and Macrocyclic Bisphosphoric Acids for Asymmetric Phase-Transfer and Brønsted-Acid Catalysis*, Chem. Eur. J. **29** (2023) e202202953, DOI: 10.1002/chem.202202953

Abstract

The first part of this thesis focuses on the implementation and evaluation of periodic boundary conditions (PBCs) for the GFN-FF force field (FF) method, its optimization for lanthanides, and its parameterization for actinides. The original FF provides a robust and accurate description of geometries, frequencies and non-covalent interactions (GFN) for molecular systems with up to 10,000 atoms, including elements up to radon. To facilitate the FF's application in solid-state chemistry, PBCs were introduced to GFN-FF accounting for interactions with periodic images of the unit cell, up to specified interaction cutoffs. Implemented lattice gradients and periodic optimizer allow for smooth optimizations. The Ewald summation, with an optimal convergence factor, is applied to ensure the convergence of the electrostatic energy. For comprehensive testing of low-level methods, existing molecular crystal benchmark sets were expanded in terms of both atom types and molecular size. The first benchmark set for molecular crystal unit cell volumes (mcVOL22) includes additional elements, such as phosphorus, sulfur, and chlorine, alongside the commonly considered hydrogen, carbon, nitrogen, and oxygen. The second benchmark set includes unit cell volumes and lattice energies of eight peptide crystals (PEP8), with an average molecular size of 96 atoms. A specialized run mode for molecular crystals (mcGFN-FF) was developed, significantly reducing overbinding for these systems by damping non-covalent interactions. The improved method shows a performance comparable to the extended tight-binding method GFN1-xTB on various molecular crystal benchmarks. The FF's applicability is further extended through reparameterization for lanthanides and the inclusion of actinides. Specifically, the assignment of covalent bonds and ligand bond lengths are improved by applying optimized covalent radii. To enable robust modeling of lanthanide and actinide complexes with challenging connectivity, a feature has been introduced that allows the manual adjustment of the automatically assigned covalent bonds. A notable application of the force field was the investigation of host-guest systems involving metal-organic frameworks and C60 molecules, which show promise for use in photocatalysis. In this context, the combination of the force field, tight-binding, and DFT methods provided an efficient approach to determine the C60 positions within the framework and suggested that the formation of a solvent shell around the C60 guest molecules is energetically unfeasible. The robust molecular dynamics simulations and geometry optimizations of metal-organic frameworks, ionic liquids, and lanthanide- or actinide-biocomplexes with mcGFN-FF are valuable assets for multi-level workflows in the corresponding fields of chemistry. Overall, this work significantly facilitates computational investigations of both molecular and periodic systems for almost the entire periodic table.

In addition to the main part of this work, quantum chemical tools are integrated into an efficient workflow for the accurate calculation of reaction-free energies. A particular emphasis is placed on the treatment of solvation effects by inclusion of explicit solvent molecules. The employed microsolvation approach is required for an accurate description of structures with strong solute-solvent interactions, such as the molecular tweezers studied here. Fine-tuned workflow parameters include the number of included solvent molecules, the number of trial calculations, the metadynamic time length, and the combination of density functional approximations. Moreover, the conformational entropy contributions are accounted for – an often overlooked aspect in computational studies but crucial for flexible molecules like those in this study. The developed workflow lays the foundation for large-scale screening of tweezer-like compounds, where mircrosolvation is crucial for accurate reaction-free energy predictions. This facilitates the development of molecular tweezers which are selective for a specific target. Herein, reaction-free energies are calculated as an average of results with different

numbers of explicit solvent molecules. Four final workflow settings were evaluated, showing deviations between 2.1 to 2.4 kcal/mol from experimental data. This accuracy allows screening for candidate ligands with a reasonable threshold. Furthermore, the overall approach and detailed discussion of employed methods and considered contributions can guide future investigations of similar systems.

Contents

CHAPTER 1

Introduction

In modern chemistry, theoretical methods have become indispensable for efficiently addressing chemical challenges. These challenges typically involve identifying a molecule or substance with optimal properties for specific applications. Theoretical calculations can validate experimental results and offer deeper interpretation and understanding. Moreover, significant advancements in computational methods over the last decades have made it possible to predict properties and preselect compounds for specific tasks with reliable accuracy.[7–9] This increase in efficiency accelerates chemical problem-solving and reduces the need for chemicals in product development, promoting environmentally friendly chemistry.[10,11] The first part of this thesis focuses on force field (FF) method development, aiding the initial phases of computational investigations and the study of large systems. To provide context, the following sections will introduce the types of systems studied in this work, outline the investigated properties of interest, provide an overview of computational chemistry methods, and position the GFN-FF model[12] within the broader landscape of existing methods.

Many chemical products are desired in the solid state or in solution for practical reasons, such as storage and application. For computational chemists, this presents the challenge of modeling systems that are essentially infinite at the atomic level. Often, only a portion of the material or key molecules are modeled explicitly, while the rest of the system is approximated.[13–15] Alternatively, periodic boundary conditions (PBC) can be applied, utilizing a repeating unit to model the macroscopic structure.[16] When dealing with solids, computational methods must accommodate their diverse compositions and bonding characteristics, while their properties can vary significantly with small changes in the structure, such as doping.[17] Beyond these challenges, non-covalent interactions (NCIs) play a decisive role in the properties of certain solids, necessitating computational methods to approximate their effects. Molecular crystals and metal-organic frameworks (MOFs) are prime examples of solids where NCIs are particularly important.[18] Molecular crystals, which rely on NCIs such as hydrogen bonding and Van der Waals forces, are widely used in pharmaceuticals, where these forces affect properties like solubility and stability.[19,20] MOFs, composed of organic linkers and inorganic nodes forming porous frameworks, also rely on NCIs, which are crucial for interactions with guest molecules introduced into their pores.[21] These host-guest interactions are central to MOF applications such as gas storage, catalysis, and sensing.[22–24] As such, the accurate computational description of NCIs is indispensable for advancing the understanding and design of these materials.

Modeling the liquid phase presents its own challenges, as solvent effects can significantly influence reaction mechanisms and material properties.[25,26] In many cases, the surrounding solvent

is considered implicitly as a dielectric continuum, effectively capturing bulk solvent effects with reduced computational cost.[14] However, when strong, localized solute-solvent interactions are critical, explicit inclusion of solvent molecules becomes necessary. Microsolvation approaches address this by including explicit solvent molecules in the system, enabling a more accurate representation of conformational preferences.[27]

In recent decades, lanthanides (Ln) and actinides (An) have found important applications, such as therapeutics, imaging, magnets, displays, and detectors.[28–32] The application in biosciences and the search for novel structures poses the need for a robust method that can handle thousands of atoms while modeling the structure and interactions reasonably well. A particular challenge in treating Ln/An complexes with methods relying on element-specific parameterization is that the bond length to a specific element can change significantly depending on the oxidation state of the Ln/An and on the other coordinating ligands.[33–35] Therefore, special attention should be given to the assignment of covalent bonds for Ln and An in such methods.

Computationally investigated properties can be broadly classified into geometry-, energy-, and spectroscopy-related properties. The geometry of the investigated system – specifically the atom coordinates and atom types – forms the foundation of most computational chemistry calculations. Therefore, determining the relevant geometries for a chemical problem is a crucial aspect of this field. For flexible molecules, considering a single conformation is often insufficient to accurately capture their properties and reactivity. Utilizing an ensemble of conformers accounts for the fact that multiple energetically favorable conformers can contribute nearly equally to the molecule's overall energetic properties. Furthermore, interactions with other molecules can shift the relative stability of conformers, making a different conformation more favorable in the complex than in isolation. As a result, thorough conformational searches are essential to accurately model both isolated molecules and their interactions.[36,37] Conformer sampling tools either efficiently screen the degrees of freedom or perform time-dependent simulations of the system.[38] Additionally, geometric properties of interest include how the spatial arrangement and size of molecules can affect their ability to interact or react, as well as the time-dependent motion of molecules within a given environment. Such properties are commonly studied by molecular dynamics (MD) simulations, which determine molecular movements by evaluating atomic forces, and geometry optimizations that minimize the system's energy.[39] One of the most critical observables is the Gibbs free energy, which incorporates solvation effects, thermal contributions, and entropy alongside the electronic energy. Although the total free energies of molecules do not have any direct physical meaning, relative free energies can be used to investigate reaction mechanisms considering their reaction energies, barrier heights, and product yields. Further, the free energy difference between two conformers can help predict the preferred molecular geometry. Depending on the significance of other effects and the required accuracy, relative electronic energies can often provide valuable insights to guide decisions in the early stages of research. Although these contributions are computationally accessible, achieving reliable accuracy and feasible efficiency in their calculation remains an active area of research.[40,41]

Having introduced the relevant fields of study for this thesis, computationally investigated properties, and general approaches to their calculation, this section will provide an overview of the methods commonly applied in computational chemistry. These methods can be categorized into different levels of theory, like FF theory, density functional theory (DFT), or wave function theory (WFT). Figure 1.1 gives an overview of the levels of theory considering the accuracy and computational costs. Since computational costs and accuracy depend on various factors – such as the application, the method, and the software it is implemented in – this illustration should only be understood as a rough orientation.

Depicted applications of the different levels of theory range from modeling membrane proteins with surrounding lipid bilayers, over complex coordination to smaller proteins, to investigating reaction mechanisms and describing intricate electronic structures.

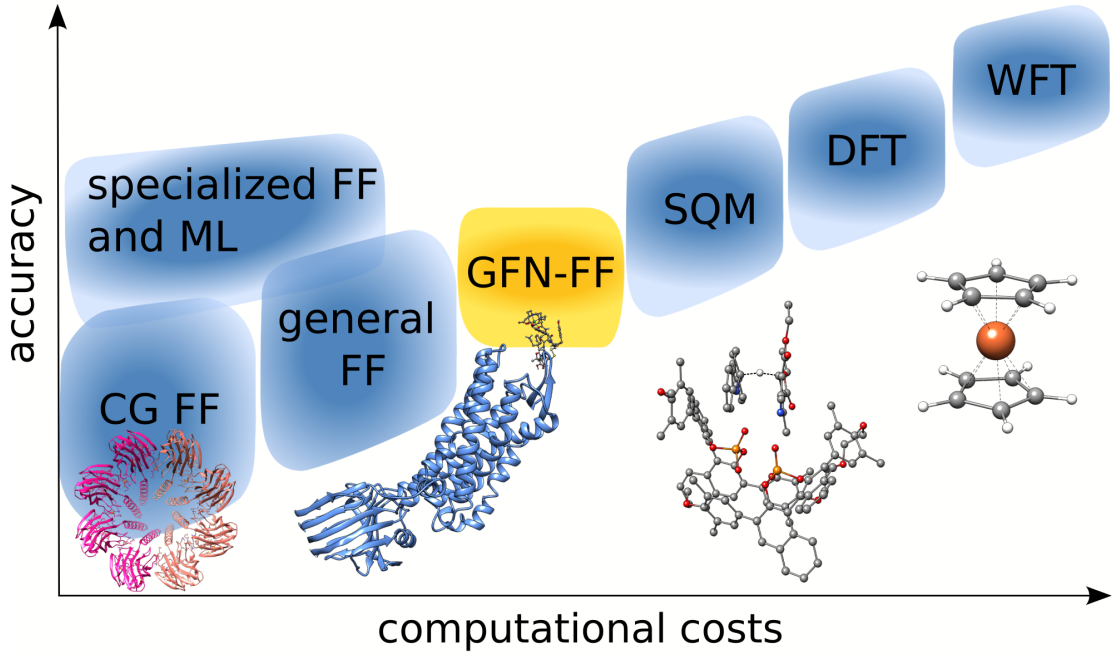


Figure 1.1: Qualitative overview of computational costs and accuracies of commonly applied computational chemistry paradigms. The GFN-FF model is included since it is at the center of this work.

WFT provides some of the most accurate methods commonly used in computational chemistry. These methods can achieve results that rival or surpass experimental accuracy, which is typically assumed to be around 1 kcal/mol for reaction energies.[42] However, this accuracy comes at the cost of exponentially increasing computational costs with system size. Consequently, these WFT methods are frequently employed to calculate reference values for smaller systems, particularly when benchmarking lower-level theories. The coupled-cluster approach with singles, doubles, and perturbative triple excitations in its wave function, CCSD(T), is widely regarded as the "gold standard" of quantum chemistry.[43]

DFT methods are often considered the workhorse of modern computational chemistry.[44] They can achieve experimental-level accuracy while remaining computationally feasible for systems containing hundreds of atoms. In practice, applying higher-level methods has little impact when other factors, such as solvation effects, are represented with lower accuracy. The reduced computational costs allow for the calculation of more expensive properties, such as NMR chemical shifts,[45,46] infrared (IR) spectra,[47] and optical properties.[48,49]

Semiempirical quantum mechanical (SQM) methods enhance efficiency by incorporating empirical parameters fitted to experimental or theoretical reference data. A common approximation in this context is to treat only valence electrons explicitly, assuming that core electrons remain tightly bound to the nuclei. These methods are often employed in multilevel workflows, serving as an efficient intermediate step for generating accurate initial structures or conformer ensembles. Their accuracy is often sufficient for tasks like energetic screening, identifying preliminary transition states, and

performing initial investigations of thermodynamic properties.[50–54]

In atomistic FFs, electronic effects can be emulated using point charges centered on the nuclei. These FFs derive the properties of chemicals from interactions between all atoms, typically using elaborate parameterizations of various interaction potentials. While some FFs aim to be generally applicable,[12,55–57] many other FFs are designed to excel for specific elements or chemical domains.[58–62] Coarse-grained (CG) FFs, which map multiple atoms of a molecule to one building block, are commonly used in biomolecular simulations due to their reduced computational cost.[63] They enable the simulation of large biomolecular systems on the scale of entire membrane proteins including a surrounding lipid bilayer.[64] An issue often omitted in scientific publications is the availability and ease of use of the presented methods. In this regard, the open-source software package, `xtb`,[65] facilitates calculations with minimal user input. This software supports both GFN-FF and extended tight-binding calculations, which can be executed with a straightforward one-line command requiring only the molecular geometry, charge, and multiplicity as input.

In recent years, machine learning (ML) has found its way into theoretical chemistry.[66,67] Key to this development are the availability of structures in large databases[68–73] and the development of efficient software and hardware to generate data sets large enough for training and evaluating ML accurate models.[74–77] In contrast to learning potentials or forces directly, delta learning approaches utilize existing FFs by learning corrections to improve their performance.[78] However, an ML model that offers robust and reasonable energies, geometries, and MD simulations for the entire periodic table has yet to be developed.

At the start of a research project in computational chemistry, the number of candidates for a specific task can lie far beyond thousands of candidates. In such cases, multi-level workflows are efficient means to narrow down promising candidates while increasing the level of theory throughout the workflow stages. Herein, FF or SQM methods can be used to search for conformers or to calculate first estimates of target properties. Finally, DFT and WFT methods can provide accurate properties resulting in a few candidates. If the remaining compounds have not already been investigated experimentally, ML models can be employed to facilitate the development of synthetic routes.[79,80] Thus, integrating computational chemistry into the chemical research process can not only save valuable time but also reduces waste that would otherwise require disposal.

With its publication, the GFN-FF model filled a gap in methods with accuracy and computational costs between universal FFs and SQM levels of theory. While specialized FFs, such as those designed for proteins, excel within their specific domains, GFN-FF is designed to provide reasonable geometries, frequencies, and non-covalent interactions (GFN) for a broad range of elements across the periodic table. The original parametrization is feasible for molecular systems and cutouts with up to 10,000 atoms including elements up to radon. The method's robust performance, delivering accurate geometric properties and reasonable energies, has led to widespread application, encouraging further expansion of its capabilities.

The work included in this thesis focuses on increasing the versatility of the GFN-FF model by making it available for solid-state and actinide chemistry. The periodic implementation of GFN-FF is presented alongside an optimized version for molecular crystals, mcGFN-FF, which improves upon the description of NCIs for these solids. Additionally, molecular crystal benchmark sets are presented to allow a comprehensive analysis of the model.

These improvements in the GFN-FF model significantly facilitate research in solid-state and Ln/An chemistry. With the periodic implementation and extension to actinides, the GFN-FF model provides a versatile and efficient tool for investigating a wide range of systems, from molecular crystals to

solids and actinide chemistry. While primarily suited for obtaining initial, qualitative insights, its ease of use and robust performance allow researchers to model large-scale systems – now including solids with a few thousand atoms in the unit cell. New applications of the extended GFN-FF model include screening for guest positions in host systems such as MOFs.[4] Furthermore, MD simulations can provide average distances in host-guest complexes which are important, *e.g.*, for understanding diffusion, adsorption, or binding processes.[81,82]

In the second part of this thesis, a workflow is presented for calculating accurate reaction-free energies from conformer ensembles that include explicit solvent molecules. Including explicit solvent molecules is critical when the solvent affects reactants, transition states, and products differently.[83] This approach presents an elegant compromise between the insufficient treatment of strong solvent effects in implicit models and the high computational cost of simulating all solvent molecules explicitly. However, microsolvation increases the complexity of screening the conformational space, which is mandatory for flexible molecules. A prominent example of such compounds are molecular tweezers. They are used in applications ranging from molecular recognition to drug delivery and typically feature flexible or switch-like linkers connecting binding sites that can engage with specific substrates. Depending on the flexibility of the linker and the length of the arms, the unoccupied molecular tweezer can experience significant intramolecular interactions at the substrate binding sites. To obtain a complete conformer ensemble with explicit solvent molecules approaches like the quantum cluster growth (QCG)[84] search for possible configurations by running simulations within a repulsive potential. Therefore, the utilized method must provide a relatively accurate potential energy surface, to obtain a feasible number of conformers.

The following chapter provides the theoretical background necessary for understanding the concepts and methods presented in this thesis. Chapter 3 introduces the periodic implementation of the GFN-FF model, as incorporated in the `xtb` package. In Chapter 4, the enhancement of the FF for Ln and its extension to An is presented. An interesting application of the mcGFN-FF model for investigating MOF structures is presented in Chapter 5. Chapter 6 details a quantum chemical workflow designed to calculate accurate reaction-free energies for molecular tweezers. Finally, Chapter 7 summarizes the key findings and outlines potential directions for future research.

CHAPTER 2

Theoretical Background

The content of this chapter summarizes existing knowledge and is included to provide context for the presented studies. For further details on the given topics, the reader is referred to the original literature.[85,86]

2.1 Free Energy Contributions

The Gibbs free energy G is a central quantity in computational chemistry, bridging theoretical calculations with experimental measurements and providing valuable insights into chemical problems. It is typically defined by the thermodynamic relation

$$G = H - TS\,, \tag{2.1}$$

where H is the enthalpy, S is the entropy, and T is the temperature. In practical chemical applications, reactions frequently occur in solution and at temperatures significantly above absolute zero. Under these conditions, the total free energy includes contributions from solvation (solv) and thermostatistical (thermo) effects, in addition to the electronic (el), and zero-point vibrational (ZPV) energy. Consequently, the enthalpy and entropy can be expressed as

$$H = E_{\mathrm{el}} + E_{\mathrm{ZPV}} + H_{\mathrm{thermo}} + H_{\mathrm{solv}} \text{ and} \tag{2.2}$$

$$S = S_{\mathrm{thermo}} + S_{\mathrm{solv}} + S_{\mathrm{conf}} \tag{2.3}$$

respectively. In practice, the free energy is separated into the components that are calculated, meaning that enthalpy and entropy are summed up for solvation and thermodynamic contributions. This leads to the expression

$$G = E_{\mathrm{el}} + \delta G_{\mathrm{solv}} + G_{\mathrm{thermo}} - TS_{\mathrm{conf}}\,, \tag{2.4}$$

where E_{el} is typically calculated with a quantum chemical method such as DFT, δG_{solv} is obtained from a solvation model,[87–93] G_{thermo} can be calculated from rotational, translational and vibrational contributions (including E_{ZPV}),[94] and S_{conf} can be derived from a conformer ensemble.[95–97] With this expression, it is possible to calculate relative free energies which can be compared to experimental

measurements. For example, the binding free energy ΔG_{bind} for the formation of a complex C from two molecules A and B is calculated as:

$$\Delta G_{\text{bind}} = G^{\text{C}} - G^{\text{A}} - G^{\text{B}} \,, \tag{2.5}$$

where the free energy difference between reactants and products quantifies the equilibrium constant. Accurately predicting reaction-free energies for diverse chemical systems has significant implications for molecular design. It facilitates the screening of potential candidates for specific reactivity and thus reduces the experimental resources needed. This motivates the development of methods and workflows that can calculate accurate free energies efficiently. Even though free energy and structural properties are often key quantities in chemical investigations, it should be noted that a variety of properties are computationally accessible.

2.2 Electronic Structure Theory

The electronic structure of a system is fundamental in determining its physical and chemical properties. Practically, the electronic structure can be described with a wave function Ψ and we are interested in finding wave functions that describe chemically relevant electronic states. To this end, a well-known eigenvalue equation of the wave function, the time-independent Schrödinger equation

$$\hat{\text{H}}\Psi = E\Psi, \tag{2.6}$$

is used as a starting point for introducing electronic structure theory. This equation connects specific electronic states, eigenfunctions of the Hamiltonian H, to an energy E. Generally, electronic structure methods differ in speed and accuracy based on how they approximate or substitute either the Hamiltonian, the wave function, or both. Throughout Sections 2.2.1 to 2.2.4 the electronic structure methods relevant to this thesis are introduced.

2.2.1 The Hamiltonian

The total non-relativistic Hamiltonian

$$\hat{\text{H}}_{\text{tot}} = \hat{\text{T}}_{\text{n}} + \hat{\text{T}}_{\text{e}} + \hat{\text{V}}_{\text{ne}} + \hat{\text{V}}_{\text{ee}} + \hat{\text{V}}_{\text{nn}} \tag{2.7}$$

consists of the kinetic (T) and potential (V) energies of the nuclei (n) and electrons (e). One of the most common approximations to the Hamiltonian is the Born-Oppenheimer (BO) approximation. Herein, the motion of the atom nuclei is uncoupled from the electron motions because they take place on different time scales. As a result, the kinetic energy of the nuclei is zero and the potential energy of the nuclei-nuclei repulsion becomes a constant. Thus, Equation 2.7 can be written as

$$\hat{\text{H}}_{\text{tot}} = \hat{\text{T}}_{\text{e}} + \hat{\text{V}}_{\text{ne}} + \hat{\text{V}}_{\text{ee}} + \hat{\text{V}}_{\text{nn}}^{\text{const}}. \tag{2.8}$$

This includes the kinetic energy of the electrons,

$$\hat{\text{T}}_{\text{e}} = -\frac{1}{2}\sum_{i=1}^{N_{\text{e}}} \nabla_i^2, \tag{2.9}$$

the nuclei-electron attraction potential,

$$\hat{\mathrm{V}}_{\mathrm{ne}} = -\sum_{i=1}^{N_n}\sum_{j=1}^{N_e}\frac{Z_i}{\mid \vec{R}_i - \vec{r}_j \mid}, \tag{2.10}$$

the electron-electron repulsion potential,

$$\hat{\mathrm{V}}_{\mathrm{ee}} = \sum_{i=1}^{N_e-1}\sum_{j=i+1}^{N_e}\frac{1}{\mid \vec{r}_i - \vec{r}_j \mid}, \tag{2.11}$$

and the nuclei-nuclei repulsion potential,

$$\hat{\mathrm{V}}_{\mathrm{nn}} = \sum_{i=1}^{N_n-1}\sum_{j=i+1}^{N_n}\frac{Z_i Z_j}{\mid \vec{R}_i - \vec{R}_j \mid}. \tag{2.12}$$

Here, the electron positions are indicated by $\vec{r}$, nuclei positions by $\vec{R}$, the nuclear charge by Z, the number of particles by N, and the nabla operator by ∇. The BO approximation includes the adiabatic approximation, in which coupling terms between different electronic states are neglected. It is assumed, that a system remains in the same electronic state as the nuclei move across the potential energy surface (PES), yielding one electronic surface. These two approximations are typically accurate, but break down when two or more electronic states are energetically degenerate. This often occurs in photochemical systems or during the homolytic dissociation of very polarized diatomic molecules.[98,99]

2.2.2 Hartree-Fock Theory

Hartree-Fock (HF) theory is a fundamental quantum chemistry method for modeling the electronic structure of chemical systems. For systems with a large number of electrons, explicitly calculating the interactions between all electrons quickly becomes computationally unfeasible. To manage this, HF employs the mean-field approximation, where each electron only interacts with the mean-field generated by all the nuclei and all the other electrons. This leads to the definition of the Fock operator

$$\mathbf{F}_i = \mathbf{H}_i^{\mathrm{core}} + \sum_{j}^{N_{\mathrm{e}}}(\mathbf{J}_j - \mathbf{K}_j) \tag{2.13}$$

describing the kinetic energy of an electron i and its attraction to the nuclei with $\mathbf{H}_i^{\mathrm{core}}$, as well as the Coulomb repulsion $\mathbf{J}_j$ and the exchange repulsion $\mathbf{K}_j$ to all other electrons. Before explaining how this operator describes the interactions of electrons with a mean field, it is helpful to introduce the basis set approximation which is used to model the electronic structure. Within this approximation, the atom-centered atomic orbitals (AOs) – *e.g.* 1s, 2s, $2\mathrm{p}_x$, or $3\mathrm{d}_{xy}$ – are each approximated by one basis function χ or a combination of multiple basis functions. The basis set then comprises the parameters defining all required functions to describe the AOs for each included element. Two types of basis functions are commonly used: The Slater-type function

$$\chi_{\mathrm{slater}}[r] = NY_{l,m}r^{n-1}\mathrm{e}^{-\zeta r} \tag{2.14}$$

accurately describes the cusp at the nucleus and exhibits the correct asymptotic behavior. Typically, the normalization constant N is determined by requiring that the overlap $S = \langle\chi|\chi\rangle$ of an AO with itself equals one. However, this function makes solving the electron integrals computationally demanding. Thus, the use of Gaussian-type functions is often preferred. Here a primitive Gaussian function φ is given as:

$$\varphi_{\text{gauss}}[r] = NY_{l,m}r^{2n-2-l}\mathrm{e}^{-\zeta r^2}. \tag{2.15}$$

Typically, multiple primitive Gaussian functions are contracted to a single basis function χ_{gauss} as described below in Section 2.2.3. To obtain the total wave function, molecular orbitals (MOs)

$$\phi_i^{\text{MO}} = \sum_{\alpha}^{M_{\text{basis}}} c_{\alpha i}\chi_{\alpha}^{\text{AO}} \tag{2.16}$$

are built as a linear combination of atomic orbitals (LCAO) introducing coefficients c_α. Then, the total wave function is built with the MOs in the form of a single Slater determinant to fulfill the antisymmetry condition of a fermionic system. Although an AO can be represented by a single basis function or a combination of multiple basis functions, we will denote an AO as χ in the following text. According to the Aufbau principle, the MOs are occupied in ascending order of the corresponding orbital energy. The Pauli principle restricts each MO, to a maximum of two electrons with opposite spin. Now, the Fock matrix from Equation 2.13 can be reformulated concerning interactions between atomic orbitals χ_α and χ_β

$$F_{\alpha\beta} = \langle\chi_\alpha|\mathbf{F}|\chi_\beta\rangle \tag{2.17}$$

$$= \langle\chi_\alpha|\mathbf{h}|\chi_\beta\rangle + \sum_{j}^{\text{occ. MO}} \langle\chi_\alpha|\mathbf{J_j} - \mathbf{K}_j|\chi_\beta\rangle \tag{2.18}$$

$$= \langle\chi_\alpha|\mathbf{h}|\chi_\beta\rangle + \sum_{j}^{\text{occ. MO}}\sum_{\gamma\delta}^{M_{\text{basis}}} c_{\gamma j}c_{\delta j}(\langle\chi_\alpha\chi_\gamma|\mathbf{g}|\chi_\beta\chi_\delta\rangle - \langle\chi_\alpha\chi_\gamma|\mathbf{g}|\chi_\delta\chi_\beta\rangle) \tag{2.19}$$

$$= \langle\chi_\alpha|\mathbf{h}|\chi_\beta\rangle + \sum_{\gamma\delta}^{M_{\text{basis}}} P_{\gamma\delta}(\langle\chi_\alpha\chi_\gamma|\mathbf{g}|\chi_\beta\chi_\delta\rangle - \langle\chi_\alpha\chi_\gamma|\mathbf{g}|\chi_\delta\chi_\beta\rangle) \tag{2.20}$$

which includes a sum over occupied MOs (occ. MOs) for the two-electron integrals in the Coulomb and Exchange interactions. These integrals include the two-electron operator $\mathbf{g} = \frac{1}{|r_i - r_j|}$. This also introduces the density matrix P

$$P_{\gamma\delta} = \sum_{j}^{\text{occ. MO}} c_{\gamma j}c_{\delta j}\,, \tag{2.21}$$

defined by the LCAO coefficients. The mean field that an electron, described by a one-electron basis function, in HF theory interacts with can be separated into two parts. The first part $\langle\chi_\alpha|\mathbf{h}|\chi_\beta\rangle$ covers one-electron contributions to the field. These are the kinetic energy of the electrons and the nuclei-electron attraction potential. The second part covers the averaged field resulting from Coulomb interactions between all unique electron pairs and exchange repulsion between unique electron pairs with the same spin. The product between the density matrix and the two-electron contributions

effectively weights these contributions based on the occupation of the MOs. For further details refer to reference [86] starting at Section 3.4.3.

Instead of Equation 2.6, for a given non-orthonormal basis set we can now solve the Roothaan-Hall equations

$$\mathbf{FC} = \mathbf{SC}\epsilon, \tag{2.22}$$

including the Fock operator, the LCAO coefficients $c_{\alpha i}$ in matrix $\mathbf{C}$, the orbital energies ϵ and the overlap $\mathbf{S}$ between basis functions. After orthogonalizing the basis set, within which the overlap matrix is the unit matrix, the nonlinear eigenvalue equation can be solved for new coefficients by diagonalizing the Fock matrix. The new coefficients can then be inserted in Equation 2.19 to obtain a new Fock matrix. This is repeated until the old MOs, defined by the coefficients, are consistent with the new ones, up to given convergence thresholds. This procedure is called the self-consistent field (SCF) procedure and is described here based on Section 3.4.6 of reference [86]:

1. Define a geometry (a set of atom coordinates, atomic numbers, and number of electrons) and a basis set.

2. Calculate all required molecular integrals, $\mathbf{S}$, $\mathbf{H}$ and $(\mu\nu|\lambda\sigma)$.

3. Obtain an initial guess of the density matrix P.

4. Calculate the Fock matrix according to Equation 2.20.

5. Diagonalize the fock matrix to obtain new coefficients C.

6. Calculate a new density matrix from the coefficients.

7. Determine whether the new density matrix is the same as the old density matrix within the given convergence criteria. Repeat steps 4-7 until the procedure has converged.

Common convergence criteria are the absolute change in total electronic energy $\mid \Delta E_{\mathrm{SCF}} \mid = \mid E^{i}_{\mathrm{SCF}} - E^{i-1}_{\mathrm{SCF}} \mid$, the maximum norm of the change of the density matrix $\mid \Delta P \mid = \mid P^{i}_{\mu\nu} - P^{i-1}_{\mu\nu} \mid_{\infty}$, or the Euclidian norm of the orbital gradient $g_c = \mid \frac{\partial E_{\mathrm{SCF}}}{\partial c_{\mu\nu}} \mid_2$. After the SCF procedure chemical properties can be calculated with the converged density matrix, Fock matrix, and coefficients. The foundation for most properties is the total electronic energy of a system

$$E_{\mathrm{el}} = \frac{1}{2} \sum_{\gamma\delta}^{M_{\mathrm{basis}}} P_{\delta\gamma}(H^{\mathrm{core}}_{\gamma\delta} + F_{\gamma\delta}) + \sum_{i}^{N_{\mathrm{atoms}}-1} \sum_{j=i+1}^{N_{\mathrm{atoms}}} \frac{Z_i Z_j}{R_{ij}}. \tag{2.23}$$

It can be calculated from the density matrix, the core Hamiltonian, the Fock matrix, and the nuclear-nuclear repulsion contribution.

While the exchange interactions between electrons are described exactly in HF theory, a major shortcoming is the neglect of the correlation between electrons. Thus, metal and transition metal complexes are among the most challenging systems for HF, because electron correlation often plays a critical role in describing their electronic structure.[100]

2.2.3 Kohn-Sham Density Functional Theory

As the name implies, density functional theory (DFT) leverages that the electronic energy of a chemical system is a functional of the electron density.[101] A crucial development leading to the success of DFT methods was the re-introduction of orbitals for the description of the electron density as suggested by Kohn and Sham (KS).[102] Both KS-DFT and HF rely on the variational principle,[103] which ensures that the set of KS-orbitals (or the wavefunction for HF) minimizes the total energy. Thereby, the electron density that corresponds to the lowest energy is the electron density of the ground state. In KS-DFT, an approximate electron density is constructed from KS-orbitals, which are MOs built from one-electron functions similar to HF theory. The electron density can thus be expressed in an AO basis as

$$\rho[r] = \sum_{\gamma\delta}^{M_{\text{basis}}} P_{\gamma\delta}\, \chi_\gamma[r]\, \chi_\delta[r], \tag{2.24}$$

where the occupations of the MOs are included in the density matrix **P**. The KS-DFT energy is then calculated as

$$E_{\text{KS-DFT}}[\rho] = T_{\text{S}} + V_{ne}[\rho] + J[\rho] + E_{\text{XC}}[\rho], \tag{2.25}$$

including the kinetic energy T_S, the attraction potential between electrons and nuclei V_{ne}, the Coulomb repulsion potential J, and the exchange-correlation functional E_{XC}. Herein, the kinetic energy is calculated from MOs: $T_{\text{S}} = -\frac{1}{2}\sum_i^{N_{\text{elec}}} \langle\phi_i|\nabla^2|\phi_i\rangle$. While the functional form of $V_{ne}[\rho]$ and $J[\rho]$ are known, the exchange-correlation functional E_{XC} is unknown and many approximations have been presented over the years.[104] Unlike in HF theory, the Coulomb and exchange terms in KS-DFT do not exactly cancel out for the interaction of an electron with itself, leading to the prominent self-interaction error (SIE). This error is especially significant in systems with sparse electron density, such as in hydrogen transfer reactions or, in extreme cases, systems containing only a single electron.[104,105] A second critical problem in KS-DFT is the lack of long-range correlation needed to describe London dispersion. Both issues have been addressed in the literature and are discussed below.

Although DFT methods do not allow for systematic improvement of accuracy as readily as methods based on wavefunctions, the classification of DFT methods by the functional form of their exchange-correlation term helps guide the selection of appropriate methods for specific applications. This ranking, often represented as "Jacob's ladder" by Perdew and collaborators[106,107], is summarized in Table 2.1. Higher-rung density functional approximations (DFAs) tend to include more information about electron interactions, but this does not necessarily mean they will yield better results for all systems.

On the lowest rung are the local density approximation (LDA) methods, which assume that the local density can be described as a uniform electron gas. Thus, the density gradient and higher-order derivatives of the density are neglected in the exchange energy which then only depends on the density. In the more general case, the local spin density approximation (LSDA), depends on the densities for α

Rung	Name	Variables	Examples
1	Local density	ρ	LDA, LSDA, X_α
2	GGA	$\rho, \nabla\rho$	BLYP,[108,109] PBE, B97
3	Meta-GGA	$\rho, \nabla\rho, \nabla^2\rho$	M06-L,[110] TPSS, SCAN
4	Hybrid	$\rho, \nabla\rho, \nabla^2\rho$, HF exchange	PBE0,[111] TPSSh[112]
5	Double-Hybrid	$\rho, \nabla\rho, \nabla^2\rho$, HF exchange, virtual orbitals	PWPB95,[113] DSD-PBEP86[114]

Table 2.1: Jacob's ladder of density functional approximations with considered components (variables) and commonly used DFAs for the respective rungs. Starting with the generalized gradient approximations (GGAs) the methods depend on the density gradient.

spin and β spin electrons separately.

$$E_X^{\mathrm{LDA}}[\rho] = -C_X \int \rho^{\frac{4}{3}}(\mathbf{r}) d\mathbf{r} \tag{2.26}$$

$$E_X^{\mathrm{LSDA}}[\rho] = -2^{\frac{1}{3}} C_X \int \rho_\alpha^{\frac{4}{3}} + \rho_\beta^{\frac{4}{3}} d\mathbf{r} \tag{2.27}$$

$$\text{with} \quad C_X = \frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}}$$

The correlation energy E_C^{LSDA} is obtained through interpolation schemes[115,116] using analytical solutions for low and high density limits.[117,118] LSDA methods can provide accurate results for homogeneous solids where the electron density varies slowly but for molecular systems, too small exchange energies or too large correlation contributions often lead to larger errors.

The second rung comprises generalized gradient approximation (GGA) methods, which include the first derivative of the density $\nabla\rho$ in the exchange part E_{X}. While the specific form varies in different GGAs, the exchange energy of GGAs *e.g.*

$$E_{\mathrm{X}}^{\mathrm{GGA}} = \sum_\sigma \int e_{\mathrm{X}\sigma}^{\mathrm{LSDA}}[\rho_\sigma] g_{\mathrm{X}\sigma}[s_\sigma^2] d^3\mathbf{r} \tag{2.28}$$

includes the LSDA exchange energy $e_{\mathrm{X}\sigma}^{\mathrm{LSDA}}[\rho_\sigma]$ and a gradient factor $g_{\mathrm{X}\sigma}[s_\sigma^2]$. For the given example from the B97 functional[119] the spin-density gradient is given as

$$s_\sigma = \frac{|\nabla\rho_\sigma|}{\rho_\sigma^{\frac{4}{3}}}. \tag{2.29}$$

Typically, GGA methods either follow Becke's approach by including fitted parameters in the exchange part or follow a more conservative approach by Perdew which aims to fulfil physical relations and limits. Overall, GGAs perform significantly better on molecules and inhomogeneous solids with reduced overbinding compared to LSDA methods.[120]

On the third rung are meta-GGAs which include higher-order derivatives of the density ($\nabla^2\rho$) or

the orbital kinetic energy density τ

$$\tau[\mathbf{r}] = \frac{1}{2}\sum_i \mid \nabla\phi_i[\mathbf{r}] \mid^2 \tag{2.30}$$

in the exchange-correlation part. For many chemical systems, meta-GGAs pose an efficient means for obtaining accurate structures, frequencies, conformers, and for describing non-covalent interactions (NCI).[42]

As the name indicates, hybrid functionals include HF exchange for given KS orbitals. The idea is to substitute a part of the unknown or approximated exchange energy with the exact Fock-exchange energy E_X^{HF} obtained from KS-orbitals. The exchange-correlation energy

$$E_{XC}^{\mathrm{hybrid}} = (1 - a_X)E_X^{\mathrm{DFA}} + a_X E_X^{\mathrm{HF}} + E_C^{\mathrm{DFA}} \tag{2.31}$$

then includes a new parameter a_X that scales the portion of E_X^{DFA} (with the DFA being LSDA/GGA/meta-GGA) that is substituted. That HF provides the exact exchange energy can be comprehended by starting at the adiabatic connection formula (ACF)

$$E_{XC} = \int_0^1 \langle\Psi_\lambda|V_{XC}^{\mathrm{hole}}[\lambda]|\Psi_\lambda\rangle\, d\lambda\ , \tag{2.32}$$

which connects the exchange-correlation energy with the corresponding hole potential V_{XC}^{hole}. Here, the interactions between electrons are monitored with the parameter λ. When approximating this integral as the average over the marginal values of the integral

$$E_{XC} \approx \frac{1}{2}(\underbrace{\langle\Psi_0|V_{XC}^{\mathrm{hole}}[0]|\Psi_0\rangle}_{E_X^{\mathrm{HF}}} + \langle\Psi_1|V_{XC}^{\mathrm{hole}}[1]|\Psi_1\rangle)\ , \tag{2.33}$$

the wave function Ψ_0 with non-interacting electrons ($\lambda = 0$) is the single Slater determinant used in HF theory.

While the E_X^{DFA} describes short-range exchange contributions accurately, the long-range contributions are described by E_X^{HF}. While substituting a fixed amount of the exchange energy has a positive effect for many systems, this approach results in a wrong asymptotic behavior. This issue is addressed in locally range-separated (LRS) hybrid functionals,[121] where the exchange contributions are separated into a short-range (SR) and long-range (LR) part:

$$E_{\mathrm{XC}}^{\mathrm{LRS}} = (1 - a_X)(E_{\mathrm{X}}^{\mathrm{DFA,SR}}[f_\omega^{\mathrm{SR}}] + E_{\mathrm{X}}^{\mathrm{HF,LR}}[f_\omega^{\mathrm{LR}}]) + a_X E_X^{\mathrm{HF}} + E_{\mathrm{C}}^{\mathrm{DFA}} \tag{2.34}$$

$$f_\omega^{\mathrm{SR}} + f_\omega^{\mathrm{LR}} \coloneqq \underbrace{\frac{1 - \mathrm{erf}[\omega r]}{r}}_{\text{short range}} + \underbrace{\frac{\mathrm{erf}[\omega r]}{r}}_{\text{long range}} = \frac{1}{r} \tag{2.35}$$

In this manner, the decay of the exchange-correlation energy is described more accurately and LR contributions from static correlation are considered.[122] More importantly, making the portion of substituted exchange dependent on the inter-electronic distance reduces the SIE.[123]

Double-hybrid (DH) functionals[124] on the highest rung build upon hybrid functionals and additionally

substitute a part of the correlation energy E_C^{DFA} with non-local correlation from perturbation theory (PT)[125] or random phase approximation (RPA).[126–128] A common example is the second-order perturbation theory from Møller and Plesset (MP2), which includes virtual excitations represented by doubly excited Slater determinants relative to the KS reference state. The exchange-correlation energy for DH

$$E_{XC}^{\mathrm{DH}} = (1 - a_X)E_X^{\mathrm{DFA}} + a_X E_X^{\mathrm{HF}} + (1 - a_C)E_C^{\mathrm{DFA}} + a_C E_C^{\mathrm{PT}} \tag{2.36}$$

includes a parameter a_C determining the amount of substituted correlation. Due to the inclusion of MP2 correlation, DH functionals can inherit incorrect behavior for systems with a small energetic gap between the highest occupied MO (HOMO) and the lowest unoccupied MO (LUMO).

Basis sets

The accuracy and speed of all methods that use the basis set approximation for describing the electron density or wave function are naturally highly dependent on the number of utilized basis functions. While the introduction of basis sets in Equations 2.14 to 2.16 was concise to keep the focus on HF theory, details on basis sets are given here. For clarity, the following explanation refers to basis sets that utilize contracted Gaussian-type orbitals (GTO) as basis functions. In cartesian coordinates, contracted GTOs are defined as:

$$\chi_{l_x,l_y,l_z}^{\mathrm{gauss}} = N \sum_i^K d_i (x - X)^{l_x} (y - Y)^{l_y} (z - Z)^{l_z} \mathrm{e}^{-\zeta_i |\vec{r} - \vec{R}|^2} \ . \tag{2.37}$$

This includes electron coordinates $\vec{r} = (x, y, z)$, nuclear coordinates $\vec{R} = (X, Y, Z)$, and the variables l_x, l_y, and l_z which are set according to the angular momentum l. For each atom, these basis sets contain the type of the AO that the basis function $\chi_{l_x,l_y,l_z}^{\mathrm{gauss}}$ corresponds to, the exponents ζ and the contraction coefficients d. The type of the AO defines the angular momentum. For example, s-orbitals have $l = 0$, p-orbitals have $l = 1$, and d-orbitals have $l = 2$. Now, the basis functions $\chi_{l_x,l_y,l_z}^{\mathrm{gauss}}$ are defined by requiring $l = l_x + l_y + l_z$. While more permutations exist for orbitals with $l > 1$, the number of AOs corresponding to each orbital type can be reduced to $n_{\mathrm{AO,l}} = 2l + 1$ since the remaining functions correspond to different, lower, orbital types. Several naming conventions have been established for basis sets and the most important ones are mentioned here. The size of the basis set is typically expressed in how many independent basis functions χ are used to describe an AO. Herein, independent basis functions have different exponents "zeta". A minimal basis set or single zeta (SZ) includes one independent basis function for each AO. To be clear, in the case of p-type AOs, this single independent basis function is used to create the p_x, p_y, and p_z AOs by varying l_x, l_y, and l_z respectively. Basis sets with twice the number of independent functions per AO are called double zeta (DZ), with three times as many triple zeta (TZ), and so on. In practice, the corresponding factor is often just applied for AOs in the valence region of the considered element *e.g.* valence double zeta (vDZ). To describe strong polarization effects, it is necessary to include the next higher orbital-type function in the element's basis which is commonly indicated with a P. Again, this is often only applied for the valence region as for the triple zeta valence polarized def2-TZVP basis set.

The most prominent error directly related to the basis set size is the basis set superposition error (BSSE). The error source becomes apparent when looking at relative energies such as complexation

energies

$$\Delta E_{\text{dimer}} = E_{\text{AB}} - E_{\text{A}} - E_{\text{B}} \,, \tag{2.38}$$

calculated from the energy of the complex E_{AB} formed by molecules A and B and the energies of the isolated molecules. In the geometry of the complex, it is in general possible for the basis functions from one molecule to describe the wave function of the other molecule, lowering the total energy of the system. This is not possible in the isolated systems leading to an artificially lowered (more negative) complexation energy. In a complete basis set each fragment has the basis functions needed to model the wave function of the method's ground state exactly. Thereby, additional flexibility from the basis functions of other fragments can not lower the energy any further. On the other extreme, minimal basis sets have a relatively small BSSE since the limited flexibility of the AOs does not allow significant descriptions of other fragments.[129] Of course, the remaining basis set incompleteness error (BSIE) is often increased for minimal basis sets, especially when the description of polarization is important. The counterpoise (CP) correction can be used to approximate the BSSE and correct for it.[130,131] The BSIE and the convergence behavior of many common basis sets have been discussed in the literature.[132]

2.2.4 Semiempirical Tight-Binding Methods

This section is adapted from Reference [133].

In density functional tight-binding (DFTB) methods, the energy E is expanded as a Taylor series around the reference density $\rho_0 = \sum_i^{N_{\text{atoms}}} \rho_0^i$ which is often built as a superposition of spherical neutral atomic reference densities:[134,135]

$$E[\rho] = \sum_{n=0}^{\infty} \frac{E^{(n)}[\rho_0]}{n!} (\rho - \rho_0)^n \,. \tag{2.39}$$

This assumes small fluctuations $\delta\rho = \rho - \rho_0 \approx 0$ of the density from the reference. Typically for DFTB methods, only the valence electrons are considered for fluctuations in density, while the other electrons are assumed to be tightly bound to the core.

The extended tight-binding (xTB) methods GFN1-xTB and GFN2-xTB truncate the Taylor expansion after the third-order term. The overall goal of these methods is to provide accurate geometries, frequencies and non-covalent interactions (GFN). To this end, GFNn-xTB methods rely on large fit sets to determine the optimal parameters for the target properties. This parameterization distinguishes xTB methods from DFTB methods like DFTB3 which use interpolation functions to obtain pairwise parameters from calculations with first principle methods. Hereafter, special attention is given to the energy expressions of the GFN1-xTB and GFN2-xTB methods, since they were crucial for many developments presented in this work. The following zeroth-order, first-order, and second-order terms are formally equivalent in the two methods. The repulsion energy

$$E_{\text{rep}}^{(0)} = \frac{1}{2} \sum_{i,j}^{N_{\text{pair}}} \frac{Z_i^{\text{eff}} Z_j^{\text{eff}}}{R_{ij}} \mathrm{e}^{-\sqrt{\alpha_i \alpha_j} (R_{ij})^{k_f}} \tag{2.40}$$

is calculated for atom pairs with indices i, j and includes the fitted element-specific constants Z^{eff}, the fitted exponential scaling parameter α, global parameter k_f and the distance of the two atoms R_{ij}.

The extended Hückel energy

$$E_{\mathrm{EHT}}^{(1)} = \sum_{\gamma\delta}^{M_{\mathrm{basis}}} P_{\gamma\delta}\hat{H}_{\delta\gamma}^{\mathrm{EHT}} \tag{2.41}$$

describes covalent bonding using the valence electron density matrix $P_{\gamma\delta}$ and the EHT matrix $H_{\delta\gamma}^{\mathrm{EHT}}$. The EHT matrix includes over half of the method's parameters to be able to describe various covalent bonds while avoiding an element pair-specific parameterization. The second-order isotropic electrostatic (IES) and XC energy

$$E_\gamma = E_{\mathrm{IES+IXC}}^{(2)} = \frac{1}{2}\sum_{i,j}^{N_{\mathrm{pair}}}\sum_{k}^{N_{\mathrm{shell},i}}\sum_{l}^{N_{\mathrm{shell},j}} \frac{q_k q_l}{\sqrt{R_{ij}^2 + \eta_{ij,kl}^{-2}}} \tag{2.42}$$

employs partial Mulliken shell charges q and the short-range damping term η. The damping term for GFN1-xTB

$$\eta_{ij,kl}^{\mathrm{GFN1\text{-}xTB}} = 2\left(\frac{1}{\eta_i(1+\kappa_{i,k})} + \frac{1}{\eta_j(1+\kappa_{j,l})}\right)^{-1} \tag{2.43}$$

applies the harmonic mean of the effective shell hardness and the damping term for GFN2-xTB

$$\eta_{ij,kl}^{\mathrm{GFN2\text{-}xTB}} = \frac{1}{2}((1+\kappa_{i,k})\eta_i + (1+\kappa_{j,l})\eta_j) \tag{2.44}$$

the average of the effective shell hardness. The third-order IES and XC energy are also formally equivalent for both xTB methods but differ in the treatment of the partial charges and a global shell-specific parameter K_l^Γ that is only present in GFN2-xTB.

$$E_\Gamma^{\mathrm{GFN1\text{-}xTB}} = E_{\mathrm{IES+IXC}}^{(3),\mathrm{GFN1}} = \frac{1}{3}\sum_{i}^{N_{\mathrm{atoms}}} q_i^3\Gamma_i \tag{2.45}$$

$$E_\Gamma^{\mathrm{GFN2\text{-}xTB}} = E_{\mathrm{IES+IXC}}^{(3),\mathrm{GFN2}} = \frac{1}{3}\sum_{i}^{N_{\mathrm{atoms}}}\sum_{l}^{N_{\mathrm{shell},i}} q_l^3 K_l\Gamma_i \tag{2.46}$$

In contrast to the atomic Mulliken partial charges q_i used in GFN1-xTB, partial shell charges q_l are employed in GFN2-xTB. In GFN1-xTB, an entirely geometry-dependent halogen bond (XB) correction is added for XBs involving a halogen atom X, its covalently bonded neighbor B, and the XB acceptor A.

$$E_{\mathrm{XB}}^{\mathrm{GFN1}} = \sum^{N_{\mathrm{XB}}} f_{AXB}^{\mathrm{damp}} k_X \left(\left(\frac{k_{XR}R_{\mathrm{cov},AX}}{R_{AX}}\right)^{12} - k_{X2}\left(\frac{k_{XR}R_{\mathrm{cov},AX}}{R_{AX}}\right)^{6}\right)\left(\left(\frac{k_{XR}R_{\mathrm{cov},AX}}{R_{AX}}\right)^{12} + 1\right)^{-1} \tag{2.47}$$

Here, f_{AXB}^{damp} is a damping function depending on the angle between X, B, and A. The parameters $k_{XR} = 1.3$ and $k_{X2} = 0.44$ are global parameters and k_X is halogen-specific. Including the D3

dispersion energy $E_{\text{disp}}^{\text{D3}}$, the total energy expression for GFN1-xTB is given as

$$E_{\text{GFN1-xTB}} = E_{\text{rep}}^{(0)} + E_{\text{disp}}^{(0)} + E_{\text{XB}}^{(0)} + E_{\text{EHT}}^{(1)} + E_{\text{IES + IXC}}^{(2)} + E_{\text{IES+IXC}}^{(3)} \tag{2.48}$$

$$= E_{\text{rep}} + E_{\text{disp}}^{\text{D3}} + E_{\text{XB}} + E_{\text{EHT}} + E_{\gamma} + E_{\Gamma} \tag{2.49}$$

The GFN2-xTB energy is given as

$$E_{\text{GFN2-xTB}} = E_{\text{rep}}^{(0)} + E_{\text{disp}}^{(0,1,2)} + E_{\text{EHT}}^{(1)} + E_{\text{IES + IXC}}^{(2)} + E_{\text{AES+AXC}}^{(2)} + E_{\text{IES+IXC}}^{(3)} \tag{2.50}$$

$$= E_{\text{rep}} + E_{\text{disp}}^{\text{D4}'} + E_{\text{EHT}} + E_{\gamma} + E_{\Gamma} + E_{\text{AES}} + E_{\text{AXC}} \tag{2.51}$$

where the anisotropic XC, multipole electrostatics, and charge-dependent dispersion interactions set it apart from preceding tight-binding methods. Dispersion contributions $E_{\text{disp}}^{\text{D4}'}$ are included as a self-consistent variant of the D4 model.[136,137] The anisotropic electrostatic energy

$$E_{\text{AES}} = E_{q\mu} + E_{q\Theta} + E_{\mu\mu} \tag{2.52}$$

$$= \frac{1}{2}\sum_{ij}^{N_{\text{pair}}} \Big(f_3[R_{ij}](q_i\vec{\mu}_j^{\intercal}\vec{R}_{ji} + q_j\vec{\mu}_i^{\intercal}\vec{R}_{ij}) \tag{2.53}$$

$$+ f_5[R_{ij}]\big(q_i\vec{R}_{ij}^{\intercal}\mathbf{\Theta}_j\vec{R}_{ij} + q_j\vec{R}_{ij}^{\intercal}\mathbf{\Theta}_i\vec{R}_{ij}$$

$$- 3(\vec{\mu}_i^{\intercal}\vec{R}_{ij})(\vec{\mu}_j^{\intercal}\vec{R}_{ij}) + \vec{\mu}_i^{\intercal}\vec{\mu}_j R_{ij}^2)\big)\Big)$$

includes the cumulative atomic dipole moment $\vec{\mu}$, the difference vector $\vec{R}_{ij}$ between atoms i and j, the distance R_{ij} between atoms i and j, and the quadrupole moment $\mathbf{\Theta}$. In three-dimensional space, $\vec{R}_{ij}$ and $\vec{\mu}_i$ each have three components and $\mathbf{\Theta}_i$ is a three by three tensor. Thus, the cartesian components (α, β) of $\mathbf{\Theta}_i$ and $\vec{\mu}_i$ are calculated as

$$\Theta_i^{\alpha\beta} = \frac{3}{2}\theta_i^{\alpha\beta} - \frac{\partial_{\alpha\beta}}{2}(\theta_i^{xx} + \theta_i^{yy} + \theta_i^{zz}) \tag{2.54}$$

$$\theta_i^{\alpha\beta} = \sum_{\delta}\sum_{\gamma} P_{\delta\gamma}\big(\alpha_i D_{\gamma\delta}^{\beta} + \beta_i D_{\gamma\delta}^{\alpha} - \alpha_i\beta_i\,\langle\phi_\gamma|\phi_\delta\rangle - \langle\phi_\gamma|\alpha_i\beta_i|\phi_\delta\rangle\big) \tag{2.55}$$

$$\mu_i^{\alpha} = \sum_{\delta}\sum_{\gamma} P_{\delta\gamma}(\alpha_i S_{\gamma\delta} - \langle\phi_\gamma|\alpha_i|\phi_\delta\rangle) \tag{2.56}$$

including the overlap $S_{\gamma\delta} = \langle\phi_\gamma|\phi_\delta\rangle$ of the AOs ϕ, the electric dipole $D_{\gamma\delta}^{\alpha} = \langle\phi_\gamma|\alpha_i|\phi_\delta\rangle$, and the quadrupole moment $Q_{\gamma\delta}^{\alpha\beta} = \langle\phi_\gamma|\alpha_i\beta_i|\phi_\delta\rangle$. Notably, the quadrupole moment is traceless (2.54), and the dipole and quadrupole moment are atom-centred (2.55 and 2.56). The charge

$$q_i = Z_i - \sum\sum P_{\delta\gamma}S_{\gamma\delta} \tag{2.57}$$

is calculated from the charge of the nucleus Z, the density matrix P, and the overlap $S_{\gamma\delta}$.

The anisotropic XC energy

$$E_{\mathrm{AXC}} = \sum_{i}^{N_{\mathrm{atoms}}} (f_{\mathrm{XC}}^{\mu_i} \mid \mu_i \mid + f_{\mathrm{XC}}^{\Theta_i} \mid\mid \Theta_i \mid\mid^2) \tag{2.58}$$

uses the cumulative atomic multipole moments defined above and fitted element-specific parameters f_{XC}. This contribution describes anisotropic deformations of the electron density of an atom. The anisotropic ES and XC contributions are one of the most important improvements of GFN2-xTB compared to GFN1-xTB. These improve the description of directional non-covalent bonding such as pnicogen, hydrogen, or halogen bonding.[133,138] Furthermore, the additional s-AO for H atoms is not necessary anymore in GFN2-xTB making the method slightly faster than its predecessor.

2.2.5 London Dispersion Correction

The accurate treatment of long-range correlation effects, such as London dispersion, remains a challenge for KS-DFT methods.[139–141] London dispersion arises from instantaneous electron correlations in fluctuating charge densities. These fluctuations in electron density, often described as virtual excitations, lead to transient dipole moments that induce corresponding dipoles on other fragments.[142] Therefore, the description of this effect requires the inclusion of virtual orbitals as done in MP2 for example. While double-hybrid functionals do include correlation energy from PT methods, this fixed fraction is often not sufficient or optimal for describing London dispersion accurately.[124,143,144] Addressing this challenge, dispersion correction schemes have proven effective for including London dispersion effects in DFT calculations. From the commonly applied correction schemes – D3[145], D4[136], VV10[146], vdW-DF[147,148], TS[149], or MBD[150,151] – the DFT-D schemes from Grimme are used in this thesis and therefore the latest model (D4) is described here. In the DFT-D4 model, the total energy

$$E_{\text{DFT-D4}} = E_{\text{KS-DFT}} + E_{\text{D4}} \tag{2.59}$$

is given by the sum of the electronic energy from a KS-DFT method $E_{\text{KS-DFT}}$ and the D4-dispersion energy E_{D4}. The latter

$$E_{\text{D4}} = E_{\text{D4}}^{(6,8)} + E_{\text{D4}}^{(9),ATM} \tag{2.60}$$

includes a two-body term $E_{\text{D4}}^{(6,8)}$ and a three-body term $E_{\text{D4}}^{(9),\text{ATM}}$ which are given as

$$E_{\text{D4}}^{(6,8)} = -\sum_{i,j}^{N_{\text{pairs}}} \sum_{n=6,8} s_n \frac{C_{n,ij}}{R_{ij}^n} f_n^{\text{damp}} \tag{2.61}$$

$$E_{\text{D4}}^{(9),ATM} = \sum_{\text{ijk}}^{N_{\text{ATM}}} \frac{C_{9,ijk}(3\cos\theta_i \cos\theta_j \cos\theta_k + 1)}{(R_{ij}R_{jk}R_{ik})^3} f_9^{\text{damp}} \tag{2.62}$$

where s_n are scaling factors of the corresponding multipolar contributions, C_n are dispersion coefficients, f^{damp} are damping functions, and θ are internal angles formed by the three atoms. For the two-body term, the dipole-dipole dispersion coefficients

$$C_{6,ij} = \frac{3}{\pi} \int_0^\infty \alpha_i[\mathrm{i}\omega]\alpha_j[\mathrm{i}\omega]d\omega \tag{2.63}$$

are calculated as the integral over the isotropically averaged atomic dynamic polarizabilities α concerning the frequency. A fundamental characteristic of DFT-D approaches is that they rely on precomputed dynamic polarizabilities, while several other approaches depend on the electron density directly. The higher-order dispersion coefficients C_8 and C_9 are calculated recursively from the C_6 coefficients.[136,145] The Becke-Johnson (BJ) damping function and the zero-damping scheme by Chai et al.[152]

$$f_{n,\mathrm{BJ}}^{\mathrm{damp}} = \frac{R_{ij}^n}{R_{ij}^n + (a_1 R_{0,ij} + a_2)^n} \tag{2.64}$$

$$f_9^{\mathrm{damp}} = \frac{1}{1 + 6(\bar{R}_{ijk})^{-16}} \tag{2.65}$$

are applied as indicated in Equations 2.61 and 2.62. Apart from the parameterization, D4 improves upon D3 by including a new charge-dependent scaling function ζ_q in the calculation of the polarizabilities α and considers three-body interactions $E_{\mathrm{D4}}^{(9),ATM}$ by default. While the three-body contributions are long-ranged compared to the two-body contributions, they mostly only account for 5%-10% of the total dispersion energy.[136]

2.3 Force Field Methods

In contrast to the previous methods, FFs directly relate the geometry of a system to its energy and forces without explicitly considering the electron density or wave function. In most FFs, electrons are not treated explicitly, and their effects are implicitly captured through parameterized potentials. Herein the focus lies on atomistic FFs, which consider all atoms as point charges. Faster FFs, *e.g.* for proteins, consider functional groups or amino acids as the smallest unit. In the following, common FF terms are explained, leading up to the total energy of the GFN-FF model. Figure 2.1 illustrates some of the fundamental interactions and variables in FF methods. This includes the torsion energy E_{tors} with dihedral angle φ, the bond stretching energy E_{bond} with atom distance R, the bending energy E_{bend} based on angle Θ, and the electrostatic energy E_{es} between point charges q. Additionally, repulsion energy E_{rep} and van der Waals energy E_{vdW} are central to modeling non-bonded interactions. Many FFs also incorporate cross potentials E_{cross}, which account for coupling between different interaction terms.

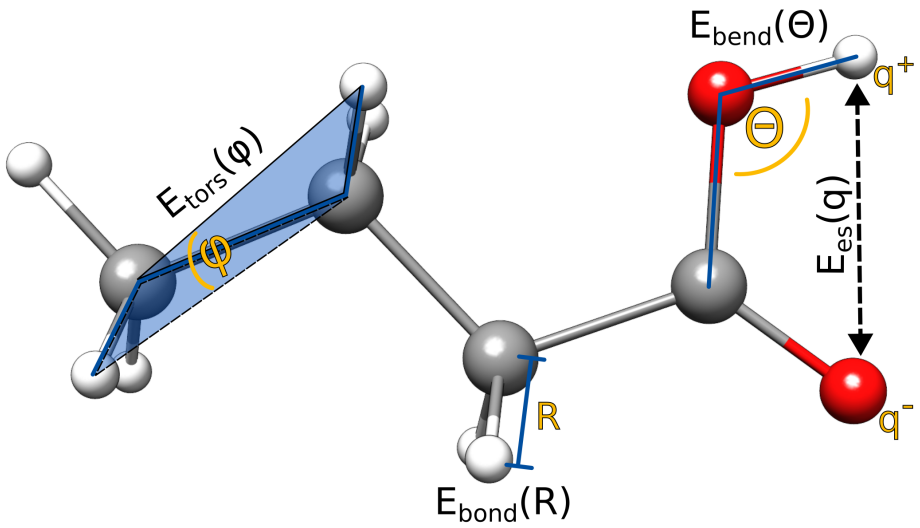


Figure 2.1: Illustration of fundamental FF potentials and corresponding variables on the example of butyric acid.

The bond potential describes the impact of stretching or shortening a bond between two atoms i and j from a reference bond length R^0. This can be expressed via a Tailor expansion

$$E_{\text{bond, ij}} = E(0) + \frac{dE}{dR}(R_{ij} - R_{ij}^0) + \frac{d^2E}{2dR^2}(R_{ij} - R_{ij}^0)^2 \tag{2.66}$$

that depends on the distance R_{ij} of the two atoms. The energy $E(0)$ is arbitrary and therefore set to zero in most cases. The first derivative evaluated at the reference distance should be zero by definition, since the reference distance should be the one of the equilibrium geometry, thus necessitating the first derivative of the energy to be zero. Applying these simplifications and summing over all covalent bonds N_{bond}, the energy expression becomes

$$E_{\text{bond}} = \sum_{ij}^{N_{\text{bond}}} k_{ij}^{\text{bond}}(R_{ij} - R_{ij}^0)^2, \tag{2.67}$$

as expected from Hooke's law. The factor k can vary greatly in complexity depending on the FF method, *e.g.* introducing dependencies on the electronegativity difference of the two atoms like in the GFN-FF model. In this model, the covalent bond energy

$$E_{\text{bond}}^{\text{GFN-FF}} = \sum_{ij}^{N_{\text{bond}}} -k_1 \mathrm{e}^{-k_2(R_{ij} - R_{ij}^0)^2} \tag{2.68}$$

is calculated by a Gaussian-type function. Additionally, the GFN-FF calculates the R_{ij}^0 based on precomputed partial charges of the two atoms. Since bonds can not be deduced from the electron density, many FF require additional input providing the bonding situation of the system. The fully automated setup of the GFN-FF includes the assignment of covalent bonds between atoms, greatly increasing the ease of use of the method. Herein, the R_{ij}^0 is used the determine whether or not there

should be a covalent bond between the two atoms.

The potential for the bending of an angle formed by three atoms Θ_{ijk} from a reference angle Θ^0_{ijk},

$$E_{\text{bend}} = \sum_{ijk}^{N_{\text{bend}}} k^{\text{bend}}_{ijk} (\Theta_{ijk} - \Theta^0_{ijk})^2 \tag{2.69}$$

is derived analogously to the bond potential. In the GFN-FF model Equation 2.69 is only applied for linear angles ($\Theta^0_{ijk} \approx \pi$) and for all other cases, the angles are replaced with their cosine

$$E_{\text{bend}} = \sum_{ijk}^{N_{\text{bend}}} k^{\text{bend}}_{ijk} \left(\cos[\Theta_{ijk}] - \cos[\Theta^0_{ijk}]\right)^2 \tag{2.70}$$

This function accounts for the symmetry of bond angle distortions by incorporating two minima, symmetrically positioned around a maximum at $\Theta_{ijk} = \pi$, within the interval $[0, 2\pi]$.

The torsion potential captures the energy change associated with a rotation around a central bond in a four-atom sequence of covalently bonded atoms. The dihedral angle φ between the two planes defined by atoms i, j, k and j, k, l determines the potential. To reflect the periodicity of this rotation, the potential is written as a Fourier series:

$$E_{\text{tors}} = \sum_{ijkl}^{N_{\text{tors}}} \sum_{n=1} k^{(n)}_{ijkl} \cos[n\varphi_{ijkl}]. \tag{2.71}$$

Here, the number n determines the torsion potential's periodicity and the prefactor's magnitude $k^{(n)}$. For example, a rotation around a carbon-carbon double bond should be periodic by 180° and have a larger barrier than a rotation around a single bond. The GFN-FF model includes damping factors for long distances in the bending and torsion potential to allow bond dissociation.

The electrostatic interactions are derived from the Coulomb potential

$$E_{\text{coul}} = \sum_{ij}^{N_{\text{atoms}}} \frac{q_i q_j}{4\pi\epsilon_{\text{el}} R_{ij}} \tag{2.72}$$

between two point charges q_i and q_j at distance R_{ij} using the electric constant ϵ_{el}. These interactions have a large range with $\frac{1}{R}$ scaling, which requires special treatment in periodic systems as described in Section 2.4. The GFN-FF model uses the electronegativity equilibration (EEQ) model to describe the electrostatic interactions. Applying the Einstein summation convention the electrostatic energy according to the EEQ model can be written as

$$E_{\text{es}} = q_i (\frac{1}{2} \mathbb{A}_{ij} q_j - X_i) \tag{2.73}$$

including the partial charges vector q, Coulomb matrix $\mathbb{A}$, and the coordination number dependend X vector. This model allows the calculation of partial charges while maintaining the total charge of the system.

Van der Waals interactions play a crucial role in many chemically relevant systems and encompass

a range of interactions, including dipole-dipole interactions, dipole-induced dipole interactions, and London dispersion forces. The Lennard-Jones (LJ) 12-6 potential

$$E_{\mathrm{LJ}} = 4E_{\mathrm{min}}\left(\left(\frac{R_{ij}^{\mathrm{zero}}}{R_{ij}}\right)^{12} - \left(\frac{R_{ij}^{\mathrm{zero}}}{R_i j}\right)^{6}\right) \tag{2.74}$$

is often used to describe the London dispersion force. Here, E_{min} is the minimum energy at the equilibrium distance, and R_{ij}^{zero} is the distance at which E_{LJ} is zero. Other potentials applicable in this context are the Morse potential,[153] the Buckingham potential,[154], or the Steele potential used for surfaces.[155] The GFN-FF uses a modified version of the D4 dispersion model for the dispersion potential:

$$E_{\mathrm{disp,GFN\text{-}FF}} = -\sum_{ij}^{N_{\mathrm{atoms}}}\left(s_6\frac{C_{6,ij}}{R_{ij}^6}f_6^{\mathrm{damp}}\right) + \left(s_8\frac{C_{8,ij}}{R_{ij}^8}f_8^{\mathrm{damp}}\right). \tag{2.75}$$

This expression includes the rational Becke-Johnson (BJ) damping function f_n^{damp}, the dispersion coefficients C_n, and scaling factors s_n, where n denotes the order of the dispersion term and takes values of either six or eight. Due to the high computational costs, the GFN-FF model does not include three-body dispersion interactions.

In the GFN-FF, the Pauli repulsion is separated into a covalent and a non-covalent part. Both potentials have the general form

$$E_{\mathrm{rep}} = \sum_{ij} k_{\mathrm{rep},1}\frac{Z_i^{\mathrm{eff}}Z_j^{\mathrm{eff}}}{R_{ij}}\exp\left(-\sqrt{k_{\mathrm{rep},2}R_{ij}^3}\right), \tag{2.76}$$

similar to the one proposed by Grimme et al. [156]

The last common potential is a cross-term, typically describing couplings between the bond, bending, and torsion interactions. For example, the GFN-FF model includes a three-body bond (3bb) correction

$$E^{\mathrm{3bb}} = \sum_{ijk}^{N_{ijk}} C_{ijk}\frac{3\cos[\theta_i]\cos[\theta_j]\cos[\theta_k]+1}{(R_{ij}R_{jk}R_{ik})^3} \tag{2.77}$$

coupling the bond stretching and bending interactions.

In addition to the above-mentioned potentials, the GFN-FF model also describes hydrogen and halogen bonds. For hydrogen bonds, the potential

$$E_{\mathrm{HB}} = -\sum_{DHA}^{N_{\mathrm{DH\,A}}} k_{\mathrm{HB}}\left(\frac{\omega_{\mathrm{DA}}}{R_{\mathrm{DA}}^3} + \frac{\omega_{\mathrm{AH}}}{R_{\mathrm{AH}}^3}\right) \tag{2.78}$$

is calculated for each triplet of donor D, hydrogen H, and acceptor A. The prefactor k_{HB} is rather involved and includes a case separation between hydrogen bonded to the donor $D-H\cdots A$ and hydrogen not bound to the donor $D\cdots H\cdots A$. Halogen bonds are assigned a potential similar to hydrogen bonds.

$$E_{\mathrm{XB}} = -\sum_{D} XA^{N_{\mathrm{XB}}}k_{\mathrm{XB}}\frac{\chi_{XA}}{R_{XA}^3} \tag{2.79}$$

Putting all these interactions together, the GFN-FF potential is given by

$$\begin{aligned} E_{\text{GFN-FF}} &= E_{\text{cov}} + E_{\text{NCI}} \\ &= E_{\text{rep}}^{\text{bond}} + E_{\text{bond}} + E_{\text{bend}} + E_{\text{tors}} + E_{\text{abc}}^{\text{bond}} \\ &+ E_{\text{ES}} + E_{\text{disp}} + E_{\text{HB}} + E_{\text{XB}} + E_{\text{rep}}^{\text{NCI}} . \end{aligned} \tag{2.80}$$

Apart from the elaborate interaction potential, the parameterization plays a key role in the accuracy and robustness of the presented model. In contrast to element-pair-specific parameters, the use of element-specific parameters in the GFN-FF model reduces the number of parameters, is less prone to overfitting, and ensures greater robustness. The fitting procedure minimizes the RMSD between GFN-FF computed data and reference data from PBEh-3c[157] and B97-3c,[158] including equilibrium and distorted geometries, harmonic frequencies, and NCI energies from the GMTKN55 database.[104] Given that the accuracy and robustness of empirical methods depend heavily on the parameterization, careful benchmarking is essential to assess their reliability and ensure they perform well across diverse systems.

2.4 Periodic Boundary Conditions

So far, the chemical systems have been considered to be in the gas phase. Solvation effects in liquids can be addressed through implicit models or by adding explicit solvent molecules, and various approaches exist for the efficient modeling of crystalline solids. These models can be categorized into nonperiodic and periodic, and encompass free clusters, saturated clusters, embedded clusters, cyclic clusters, and the unit cell approach.[13] Here, a cluster refers to a cutout from a periodic structure and the preceding adjectives describe how the omitted surrounding is approximated. In this work, the unit cell approach has been implemented for the GFN-FF model and is therefore focused on hereafter.

When transitioning from molecular to periodic systems, the geometry is defined not only by the atom coordinates and atom types but also by the lattice parameters, which determine the shape of the repeating unit, or unit cell. Figure 2.2 illustrates a unit cell (highlighted in blue) containing four uracil molecules and eight repeated images of the unit cell. For FF methods, it is common that periodic images of the unit cell are generated only up to the cutoff radii specified for NCIs. A useful approach for handling periodic images is the minimum image convention,[16,159] which requires the unit cell to be at least half the size of the largest interaction cutoff. This ensures that atoms do not interact with their own periodic image, effectively minimizing finite-size effects. However, the importance of these effects depends on the investigated problem. For example, in a geometry optimization of the system shown in Figure 2.2, the result should remain unchanged even if the unit cell encompassed all nine depicted cells. This is because each atom in a larger unit cell should behave the same as its counterpart in the original cell. This is not the case for MD simulations, where the periodic images of atoms have to move the same as the original. Since each atom in the unit cell is initialized with a random starting velocity, the size of the unit cell makes a difference in the simulation. Looking at Figure 2.2, it is obvious that the given arrangement of the molecules could not occur in a simulation with only one uracil in the unit cell. Investigations of defects in solids are another example where the unit cell size has a crucial impact. Note that the implementation in `xtb` connected to this thesis, does not enforce the minimum image convention.

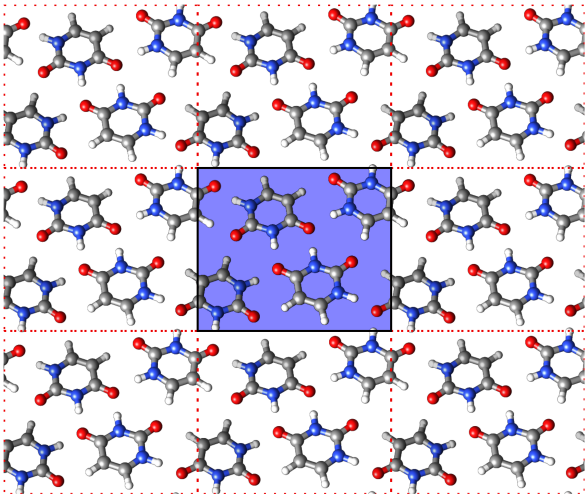


Figure 2.2: Illustration of periodic boundary conditions. The unit cell containing four uracil molecules is highlighted in blue and dotted red lines mark its repeated images.

To perform geometry optimizations in periodic systems, the stress tensor or more precisely the gradient of the lattice parameters has to be calculated. It is defined as the derivative of the energy E by the infinitesimal strain deformations ϵ.

$$\sigma_{ij} = \frac{1}{V} \frac{\partial E}{\partial \epsilon_{ij}} \mid_{\epsilon=0} \tag{2.81}$$

Knuth et al.[160] provide helpful transformations of Equation 2.81 for practical implementations. By convention, the first index defines the face of the unit cell that the stress acts on and the second index defines its direction. Thereby, the stress tensor in three dimensions contains three normal stresses on the diagonal and six shear stresses.

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \tag{2.82}$$

The conservation of the angular momentum leads to a symmetric stress tensor: $\sigma_{ij} = \sigma_{ji} \; \forall \; i, j \in \{x, y, z\}$.[161]

The presented extensions and improvements of the GFN-FF model are intended to model molecular interactions within a solid such as a molecular crystal or a guest host system with a MOF.

Since periodic systems are infinitely large, the convergence behavior of applied potentials needs to be considered. The electrostatic interactions are long-ranged interactions scaling with the inverse distance $\frac{1}{R}$. The sum of these interactions does not converge in real space and different approaches exist to calculate the electrostatic energy efficiently.[162,163] In the Ewald summation, the energy is split into two parts

$$E_{\text{es}} = \frac{q_1 q_2}{4\pi\epsilon_0} \frac{1}{R} \propto \frac{1}{R} = \frac{\text{erf}[\xi R]}{R} + \frac{(1 - \text{erf}[\xi R])}{R} \tag{2.83}$$

by adding and subtracting the error function $\text{erf}[\xi R]$ with Ewald splitting parameter ξ to the numerator.

Now, the first part converges in reciprocal space and the second in real space. This means that both parts can be truncated at finite (inverse) distances, which allows any desired precision. The error function is typically chosen because it allows for smooth handling of the real space and reciprocal space contributions via the Ewald splitting parameter.

2.5 Methods for the Exploration of Conformational Space

In real-world reactions, molecules can partake in the form of different conformers, where the relative portions of present conformers depend on the temperature among other factors. Conformers are defined as local minima on the potential energy surface and therefore depend on the applied method. For many molecules, it is not sufficient for property calculations to include only the structure corresponding to the global energy minimum, but other (nearly degenerate) local minima. An ensemble of conformers can be generated by sampling configurations during an MD simulation. In these simulations, the positions and velocities of atoms are propagated over time based on the force acting on them, governed by Newton's second law:

$$F = ma \tag{2.84}$$

where the force is expressed as the negative gradient of the potential energy:

$$-\frac{dE}{d\vec{r}} = m\frac{d^2\vec{r}}{dt^2} \tag{2.85}$$

The atom positions $\vec{R}$ are updated each time step Δt for example with the Verlet algorithm:

$$\vec{R}_{t+\Delta t} = (2\vec{R}_t - \vec{R}_{t-\Delta t}) + \vec{a}_t(\Delta t)^2 \tag{2.86}$$

where the acceleration a is derived from the force as:

$$\vec{a}_t = -\frac{1}{m_t}\frac{dE}{d\vec{R}_t} \tag{2.87}$$

The Taylor expansion of the atom positions at time t and the transformation to equation 2.86 is given in Section 14.2.1 of Reference [85].

In the `xtb` program NVT simulations are performed by default, which means that the number of particles (N), the unit cell volume (V), and the temperature are fixed. While fixing the number of particles and the unit cell volume is trivial, the temperature – which is given by the kinetic energy of the particles – needs to be regulated. This is often done by coupling the system to a heat bath which effectively scales the velocities of the atoms by a factor

$$f_{\text{velo}} = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_{\text{target}}}{T_{\text{current}}} - 1\right)} \tag{2.88}$$

and depends on the coupling parameter τ. A problem with using MD simulations for sampling the conformational space is that the barriers of the potential energy surface have to be overcome, to sample a new conformer. In consequence, a significant amount of computational resources are spent exploring

the energy wells of the found conformers. This problem has been addressed in meta-dynamics (MTD) simulations by including coarse-grained dynamics and adaptive bias potentials.[164] For example, the conformer rotamer ensemble search tool (CREST) applies a bias potential based on atomic RMSDs.[38] With a complete conformer ensemble, the time dependence can be included indirectly by calculating chemical properties or reaction properties as a weighted average. For example, the free energy of a system G_{avg} can be calculated as the Boltzmann average[165]

$$G_{\mathrm{avg}} = \sum_{i}^{\mathrm{CRE}} G_i \frac{\mathrm{e}^{-G\beta}}{\sum_j^{\mathrm{CRE}} \mathrm{e}^{-G\beta}} \tag{2.89}$$

over a conformer-rotamer ensemble (CRE). Then, reaction properties can be calculated as relative free energies as described in Equation 2.5. It should be noted that Monte Carlo (MC) simulations[166,167] are a common alternative to MD simulations, employing stochastic sampling techniques to explore the potential energy surface.

CHAPTER 3

mcGFN-FF: an accurate force field for optimization and energetic screening of molecular crystals

Stefan Grimme[1] and Thomas Rose[1]

Received: October 5, 2023
Published online: April 5, 2024

Reprinted in Appendix B with permission from S. Grimme and T. Rose, *mcGFN-FF: an accurate force field for optimization and energetic screening of molecular crystals* Z. Naturforsch., B **79** (2024) 191–200 –
DOI: https://doi.org/10.1515/znb-2023-0088

Own contributions

- Implementation of periodic boundary conditions for GFN-FF in the `xtb` code
- Parameterization and implementation of GFN-FF run mode for molecular crystals
- Compilation of new benchmark sets
- Performing all of the calculations
- Interpretation of the results
- Writing the manuscript

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

In this work, PBCs have been implemented for the GFN-FF model. In contrast to most FFs and ML potentials, the presented model applies to almost all elements in the periodic table. This makes the method valuable for candidate screening, crystal structure prediction (CSP), and investigation of interaction sites in large complexes such as MOFs. For extended systems, the convergence of the electrostatic contributions needs to be ensured, which is done with the use of the Ewald summation. Herein, the optimal Ewald splitting parameter is determined with the golden section search.[168] Since the description of NCIs is a focus of the GFN methods, initial tests of the FF were focused on molecular crystal benchmark sets such as the X23.[169,170] These tests revealed a systematic overbinding reflected in lattice energies approximately 5 kcal/mol too negative and unit cell volumes approximately 10 % too small. This encouraged the focus of the `xtb` version of the periodic GFN-FF on these solids, by introducing a specialized run mode named mcGFN-FF for molecular crystals. In this optimized version, the non-covalent parts of the dispersion, electrostatic, hydrogen-bond, and repulsion interactions are scaled according to revised lattice energies on the X23 benchmark. The mcGFN-FF has been benchmarked against diverse test sets varying in composition and system size. Overall, it performs comparable to GFN1-xTB but with a larger standard deviation in the errors in energy and unit cell volume. Previously, Gale et al. have implemented PBC for the FF, named pGFN-FF, overcoming instabilities of the model for ionic crystals.[171] Their implementation features the Wolf sum for handling electrostatic interactions and includes a damping factor for the three-body bond correction (dATM). The normal run mode of GFN-FF, without the NCI scaling, performs similarly to the pGFN-FF without the dATM damping factor. Excluding a small benchmark for ice polymorphs (ICE10)[172], the mean absolute relative deviations (MARD) of mcGFN-FF improve from 33.5 % to 14.1 % for lattice energies and from 11.7 % to 6.4 % for unit cell volumes. The versatility of the periodic FF is showcased on three crystals with intricate electronic structures. The optimized systems comprise metallasilylidyne, osmium, and siladodecahedrane. Even for these intricate systems, the RMSD of the optimized structure from the experimental reference is below 1.0 Å. An off-target evaluation has been performed for twelve MOFs, showing performance comparable to GFN1-xTB.

A requirement for evaluating any method is the availability of reliable reference data. Although crystal structures are abundantly published, thermal effects restrict comparisons to a qualitative level. Thermal expansion typically increases the unit cell volume, for instance up to 10 % for certain molecular crystals[173], but it can also decrease the unit cell volume such as for water near its melting point.[174] Consequently, experimental data must either be back-corrected or reference values need to be calculated with accurate computational methods. Existing benchmarks for molecular crystals primarily focus on elements such as hydrogen (H), carbon (C), nitrogen (N), oxygen (O), and chlorine (Cl).[170,175,176] The Benchmark Data Set of Crystalline Organic Semiconductors (BMCOS) includes many systems with sulfur (S), and a limited number of systems with Cl, fluorine (F), boron (B), phosphorus (P), or silicon (Si). To extend the compositions in atom types of available benchmarks, the presented mcVOL22 provides reference geometries and unit cell volumes for 22 molecular crystals including P and S atoms in addition to H, C, N, O, and Cl. Herein, r^{2}SCAN-D4 is used for optimizing the geometries.[136,177,178] Additionally, a second benchmark set includes lattice energies and unit cell volumes for eight peptide structures (PEP8). This offers a benchmark with significantly larger molecules, which is ideal for testing fast computational methods concerning biological applications.

Overall, implementing PBCs for GFN-FF represents a significant step toward enabling the routine investigation of geometries and energies for larger and more diverse molecular crystals. The development of mcGFN-FF optimizes the description of NCIs, improving the accuracy of lattice energies and unit cell volumes while maintaining computational efficiency. Additionally, introducing

the mcVOL22 and PEP8 benchmarks broadens the range of test systems, incorporating a wider variety of atom types and larger molecular structures, which are particularly relevant for biological and organic materials. Future directions include exploring the model's capabilities for MD simulations of MOFs with guest molecules, further integrating the method into CSP workflows to improve candidate screening, and extending its application to surfaces and localized effects in solids. These areas hold considerable promise for advancing theoretical chemistry and materials science.

CHAPTER 4

Fast and Robust Modeling of Lanthanide and Actinide Complexes, Biomolecules, and Molecular Crystals with the extended GFN-FF Model

Thomas Rose[1], Markus Bursch[2], Jan-Michael Mewes[3], and Stefan Grimme[1]

Received: July 30, 2024 Published online: September 28, 2024

Reprinted in Appendix C with permission from[4] T. Rose, M. Bursch, J.-M. Mewes, and S. Grimme, *Fast and Robust Modeling of Lanthanide and Actinide Complexes, Biomolecules, and Molecular Crystals with the extended GFN-FF model*. Inorg. Chem. **63** (2024) 19364–19374

DOI: https://doi.org/10.1021/acs.inorgchem.4c03215

Own contributions

- Parameterization of the force field for actinides
- Improvement of the force field parameters for lanthanides
- Performing all of the calculations
- Drafting the manuscript
- Major contributions to editing the manuscript
- Major contributions to the interpretation of the results

[1] Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Beringstraße 4, Bonn 53115, Germany
[2] FACCTs GmbH, Rolandstraße 67, 50677, Koeln, Germany
[3] beeOLED GmbH, Niedersedlitzer Str. 75c, 01257 Dresden, Germany
[4] Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

The steady expansion of the synthetically accessible space necessitates computational methods to keep up with this development. With applications in imaging, tumor treatment, organic light-emitting diodes, and treatment of radioactive waste, Ln/An chemistry has become of interest in recent years.[179–181] While DFT and tight-binding methods are available to investigate rare-earth elements, there remains a need for fast and robust methods that can routinely handle molecules and crystals comprising thousands of atoms. The GFN-FF model is applicable to systems of this size and has been frequently used since its publication. Applications include conformer generation and screening in the early stages of different quantum chemistry workflows.[182–184] This makes it sensible to parameterize the FF for An and enhance its performance for Ln. As a starting point, the initial parameters for the An were adopted from the Ln. The extended version of the GFN-FF includes four improvements that optimize the model's performance for 4f and 5f elements. First, the electronegativity equilibration (EEQ) charge model was revised based on reference Hirshfeld charges. The covalent radii used for determining covalent bonds between atoms were optimized to improve geometry optimizations. It should be noted, that there is no distinction between e.g. ionic and covalent bonds, rather all bonds are treated as covalent bonds defined by an automatically determined neighbor list. Furthermore, a systematic shift of the bond length to Ln/An atoms was observed for each included halogen in the LnQM[185] and the fit set compiled from the AcQM benchmark set.[186] Thus, the element-pair-specific parameter for the ideal bond length was scaled based on the mean signed deviation of the halogen f-element bond lengths in the two data sets. Additionally, an assignment of covalent bonds to hydrogen atoms of ligands coordinated to the Ln/An atom is avoided by introducing a partial charge cutoff. With this cutoff, the bond assignment for hydrogen atoms is correct for the entire LnQM. The extended model was tested on the LnQM concerning bond distances and partial charges. The mean absolute deviation of bond length between the central Ln atom and its neighbors is reduced from 0.18 Å to 0.14 Å. However, the root mean squared deviation of the partial charges increased from 0.17 e to 0.20 e. This increase stems from incorporating hydrides and halogen complexes in the fit set, to enhance the robustness of the EEQ model. Geometric properties for An were tested on four 12-coordinated An(IV)-ADC[187] crystals with Th, U, Np, or Pu as the An atom. Bond distances were reproduced accurately with absolute deviations below 0.12 Å and the absolute deviations in pore diameters were below 0.30 Å. Bond angles deviated less than 4.2° from the experimental values and the dihedral angle between the anthracene and the carboxylate group deviated less than 1.4°. The suitability of the method for exploring dynamical properties and conformational spaces of Ln/An complexes is tested by performing MD simulations. The radial distribution functions (RDFs) of the neighboring atoms of the central Ln/An atoms are a good indicator for correct bond assignment and depict the flexibility of the assigned bonds. While the extended model performed better than the original method for all four test complexes, the novel feature that allows the manual assignment of bonds by the user ensured correct bond topologies for the most difficult complexes. Importantly, the MD simulations remained robust even without corrections to the neighbor assignments. A particularly relevant application of the method is the structural investigation of Ln-complexes interacting with biomolecules. This application was showcased in the example of the radiopharmaceutical Lu-dotatate coordinating with the somatostatin receptor (SSTR). The optimization of the entire guest host system comprising 7566 atoms was performed on four CPUs with 2.0 GHz in under 20 hours, including relevant solvation effects via the implicit solvation model ALPB. A second complex with 1671 atoms was optimized in 11 min yielding an RMSD of 1.2 Å. Though geometric properties are the focus of the GFN-FF, a set of relative isomer energies for Ln/An complexes was compiled to test the accuracy of the energies. From the 16 energies, 13 deviate from the r^{2}SCAN-3c reference by less than 4.5 kcal/mol and the

largest deviation is 10.8 kcal/mol. Notably, the FF optimizations each took less than a second, while the reference calculations took 51 min on average. The timings and the robustness of the method were tested with a versatile test set comprising 29 different atom types, of which four are the Ln/An Eu, Lu, Ac, and Th. Optimizations failed for some, but not all systems containing ionic clusters, highlighting an area for future improvement. Similar to the molecular version, the periodic optimizations scale quadratic with the system size where the prefactor depends on the number of optimization cycles needed.

GFN-FF has become a highly versatile computational method that is now capable of handling molecular and periodic systems with remarkable efficiency. With the inclusion of PBCs and parameters for elements up to lawrencium, the extended model significantly broadens the scope of systems that can be studied, from biomolecular complexes to Ln/An-containing materials. This advancement marks an important step toward the routine investigation of large or periodic systems containing Ln/An elements, providing a powerful tool for addressing challenges in modern computational chemistry.

CHAPTER 5

Charge Transfer and Recombination Pathways through Fullerene Guests in Porphyrin-based MOFs

Alison Arissa[1], Thomas Rose[2], Noémi Leick[3], Stefan Grimme[2], Justin C. Johnson[2], Jenny V. Lockard[1]

Received: January 8, 2025
Published online:

Reprinted in Appendix D with permission [4] from Alison Arissa, Thomas Rose, Noémi Leick, Stefan Grimme, Justin C. Johnson, Jenny V. Lockard, *Charge Transfer and Recombination Pathways through Fullerene Guests in Porphyrin-based MOFs*. J. Phys. Chem. C **129** (2025) 8215 –
DOI: https://doi.org/10.1021/acs.jpcc.5c00161

Own contributions

- Performing all calculations for the computational investigation
- Interpretation of the computational results
- Writing the computational part of the manuscript

[1] Department of Chemistry, Rutgers University-Newark, Newark, New Jersey 07102, USA
[2] Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Beringstraße 4, Bonn 53115, Germany
[3] National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, Colorado 80401, USA
[4] Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

Porphyrin-based MOFs are a promising platform for constructing host-guest systems with tunable conductive properties and extended charge separation lifetimes, depending on the electron acceptor guest molecules.[188,189] In this study, two MOFs were synthesized, each featuring Zr_6 clusters connected by porphyrin-based linkers: PCN-222(H_2), containing metal-free porphyrin linkers, and PCN-222(Zn), with zinc porphyrin linkers. Both structures include large hexagonal pores and smaller triangular pores, the latter being of particular interest for fullerene guests due to its matching size. These MOFs were infiltrated with fullerene derivatives C_{60} and C_{61}BM in 1,4-dioxane and in dimethylformamide (DMF) to investigate the effects of varying the linker, guest species, and polarity of the solvent on the properties of the host-guest system. Notably, the infiltration process required approximately 7 days, indicating substantial energetic barriers for fullerene entry into the triangular pores.

Experimental characterization revealed key differences between the two MOFs. N_2 physisorption measurements showed significant reductions in the triangular pore volume upon fullerene infiltration, with PCN-222(H_2) exhibiting a more pronounced decrease (59%) compared to PCN-222(Zn) (24%), suggesting stronger binding interactions in the metal-free system. Photoluminescence quenching experiments further supported this conclusion, as fluorescence was quenched stronger in PCN-222(H_2). Ultrafast transient absorption spectroscopy provided additional insight into the photochemical properties, revealing sub-picosecond electron transfer and charge recombination lifetimes on the order of one nanosecond. Moreover, the measured charge transfer state formation and decay lifetimes were similar for polar DMF solvent and unpolar 1,4-dioxane solvent.

Computational chemistry provided additional context for these experimental findings, with mcGFN-FF allowing investigations directly on the periodic MOF structures. For higher-level methods, a large cutout comprising 1002 atoms was created to model the fullerene infiltration of the triangular pore. A preliminary FF screening of the periodic MOF identified two key positions for C_{60} within the triangular pores: P1, located between porphyrin linkers, and P2, positioned between Zr_6 clusters. These positions were confirmed with nudged elastic band[190–192] calculations utilizing GFN2-xTB. Finally, both positions were evaluated with PBEh-3c including solvent effects with the continuum solvation model SMD.[91] The calculated hybrid DFT interaction energies suggest that both positions can be occupied simultaneously without significant energetic penalty, but do not leave enough space for solvent molecules directly in between.

Molecular dynamics simulations with mcGFN-FF provided further insight into the stability and dynamics of the guest molecules and solvents. Simulations showed that C_{60} molecules did not diffuse between P1 and P2 within a 550 ps timeframe, consistent with the long infiltration time observed experimentally. Further, the mobility of the C_{60} at both positions was quantified by observing the distance of the center of mass from the optimized starting point of the MD. The maximum distance was 1.0 Å starting at P1 and 2.5 Å starting at P2, showing significantly higher mobility at the second position. In additional simulations, solvent molecules were found to be able to move from the trigonal pore to the hexagonal pore in contrast to C60 molecules. This indicates that the solvent molecules will likely move to the hexagonal pore while the triangular pore is predominantly occupied by the fullerene guests. This aligns well with the experimental findings that the photochemical properties are essentially independent of the solvent polarity.

This study demonstrates how experimental and computational methods complement each other to provide a comprehensive understanding of host-guest interactions in porphyrin-based MOFs. In particular, being able to model the host-guest system directly with mcGFN-FF without needing to build a cutout proved helpful for gathering first insights.

CHAPTER 6

Reaction-Free Energies for Complexation of Carbohydrates by Tweezer Diboronic Acids

Gustavo Adolfo Lara-Cruz[1] , Thomas Rose[2], Stefan Grimme[2], Andres Jaramillo-Botero[3]

Received: July 18, 2024
Published online: September 16, 2024

Reprinted in Appendix E with permission from[4] G. A. Lara-Cruz, T. Rose, S. Grimme, and A. Jaramillo-Botero, *Reaction-Free Energies for Complexation of Carbohydrates by Tweezer Diboronic Acids*. J. Phys. Chem. B **128** (2024) 9213–9223

DOI: https://doi.org/10.1021/acs.jpcb.4c04846

Own contributions

- Contributions to the workflow development
- Discussion of the calculation setup
- Contributions to the interpretation of the results
- Writing and revising parts of the manuscript

[1] iOMICAS Research Institute, Pontificia Universidad Javeriana, Calle 17 # 121B-155, Santiago de Cali, Valle del Cauca 760031, Colombia
[2] Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Beringstraße 4, Bonn 53115, Germany
[3] Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States; iOMICAS Research Institute, Pontificia Universidad Javeriana, Calle 17 # 121B-155, Santiago de Cali, Valle del Cauca 760031, Colombia
[4] Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

Accurately calculating reaction-free energies for reactions involving the covalent bonding of a target molecule to a tweezer-like molecule is a challenging task. The significant difference in flexibility between the reactants and the products makes conformational entropy contributions non-negligible, requiring extensive sampling of the conformational space. Additionally, tweezer-like molecules can exhibit intramolecular NCIs, such as hydrogen bonding between the tweezer's ends. The solvent used in the reaction can significantly affect these interactions, making accurate modeling of solvation effects essential. This work presents a workflow designed to address these challenges, achieving accuracy and precision sufficient for screening candidate ligands. The workflow focuses primarily on two ligands containing diboronic acids (DBA-1 and DBA-2), for which reliable experimental data is available. It relies on a conformational search with the Conformer-Rotamer Ensemble Sampling Tool (CREST) followed by an enhancement of the ensemble with the Commandline Energetic SOrting (CENSO) tool. The Quantum Cluster Growth (QCG) method included in CREST was used to generate ensembles with explicit water molecules for microsolvation. In contrast to routine investigations, the most critical parameters of the QCG and CENSO tools have been identified and fine-tuned for efficiency. A key factor investigated was the choice of methods applied in different workflow stages. Though the GFN-FF model could be used in principle to sample the conformational space, the conformer ensemble became unreasonably large due to the larger thresholds required to include all relevant conformers. Given the challenges and the relatively small system size, semiempirical tight-binding methods were found to be more efficient for generating the conformer ensemble. An increase in RMSD concerning the original structure was used to evaluate whether the meta-dynamics sampled the conformational space sufficiently. Based on this approach, an MTD time of 80 ps was found to be effective. The introduction of explicit water molecules, alongside the implicit solvation model, provided the largest improvement in reaction-free energies. However, no clear convergence behavior for the number of explicit water molecules was observed. Therefore, averaging reaction-free energies over multiple simulations with different numbers of explicit solvation molecules is recommended to obtain a final reaction-free energy. Since CREST calculations are not deterministic, three independent calculations were performed for each reactant and product system to obtain a single reaction-free energy. In most cases, a single calculation would have sufficed, but performing multiple calculations helped identify a few outliers. When calculating the reaction-free energies as an average of all possible permutations of the reactant and product energies, rather than merging the ensembles, outliers resulted in a multimodal distribution of the reaction-free energies. The CENSO workflow consists of four parts: part 0 – prescreening, where the FF/SQM ranking from CREST is improved using lightweight DFT and solvation models; part 1 – screening, which refines the ranking with better DFT methods and may include thermal contributions; part 2 – optimization, where conformer geometries are optimized with DFT gradients, and solvation effects and thermal contributions are incorporated; and part 3 – refinement, where high-level DFT single-point calculations are performed to achieve highly accurate reaction-free energies. For the refinement with CENSO, the energy thresholds for sorting out conformers have been investigated for parts zero through two. Respectively, thresholds of 10 kcal/mol, 6 kcal/mol, and 3 kcal/mol were found to be efficient. Different combinations of DFAs were tested. Using the HF-3c method for all CENSO parts was found unreliable. However, using it for the optimization step followed by part 3 with r^{2}SCAN-3c (settings F2) was one of the most effective settings together with B97-D3 followed by r^{2}SCAN-3c for parts 1 and 2 (settings B). With setting B, the final reaction-free energies including conformational entropy contributions deviated between 2.2 kcal/mol and −3.1 kcal/mol from experimental values. With settings F2, the deviations were between 2.0 kcal/mol and −3.1 kcal/mol. The contributions of conformational entropy $T\Delta S$ range

from −3.3 kcal/mol to −4.8 kcal/mol. This work provides valuable insights into the optimization of the CREST and CENSO workflow, offering guidance on the selection and tuning of critical parameters. By thoroughly investigating the effects of different settings, this study enhances the reliability of these tools for generating accurate reaction-free energies, particularly when handling flexible molecules and solvation effects. The findings emphasize the importance of incorporating conformational entropy and a strategy for handling microsolvation that will be valuable to future applications in computational reaction modeling. These advances contribute to improving the precision and reliability of computational methods in ligand screening and reaction prediction.

CHAPTER 7

Summary

Computational chemistry plays a crucial role in the vast and diverse challenges posed by modern chemical research. The scope of chemical space extends from small organic molecules to large biomolecular systems, metal-containing macromolecules, and intricate solid-state materials. An ideal computational tool must handle this diversity while providing different types of functionalities with high efficiency. These range from geometry optimization and MD simulations to accurately determining chemical properties. In the early stages of compound design, it is often sufficient to examine molecular geometries, interatomic distances, or approximate relative energies to make informed decisions about how to proceed. At this stage, the ease of use, robustness, and speed of a computational method often outweigh its absolute accuracy. Even if the results are only qualitatively correct, they can provide critical insights guiding subsequent work. For this reason, having a single computational approach that can be applied broadly and without elaborate preparation is invaluable. Such a method streamlines the process, reducing the time and effort required to learn and set up new methods, especially when diverse chemical systems are under investigation.

The GFN-FF model, with its extensions to PBCs and Ln/An chemistry, effectively addresses these challenges by offering a broadly applicable, user-friendly tool for diverse chemical systems. Its implementation in the open-source code `xtb`, with well-defined defaults and automated setup, allows the users to run most calculations with a simple command, requiring only a few keywords. Moreover, `xtb` can read detailed input files, enabling organized modification of calculation settings. As introduced in this work, the FF now allows the user to manually introduce or remove covalent bonds between selected atoms, after the automated setup of these bonds. This feature empowers expert users to explore systems with unusual or challenging bonding scenarios. The PBC extension incorporates Ewald summation for efficient treatment of electrostatic contributions, a periodic L-BFGS optimizer for stable optimizations, and a specialized mcGFN-FF mode that enhances the modeling of NCIs in molecular crystals. The robustness and accuracy of optimizations with mcGFN-FF are evident in several key metrics: unit cell volumes of molecular crystals show a MARD of about 7%, lattice energies for molecular crystals have a MARD of 14.1% on comprehensive benchmark sets, and geometries of MOFs (bond distances, bite angles, and pore sizes) are mostly within 7% of the reference values. While the method's energetic ranking of polymorphs or isomers may not always be accurate, it provides a useful preliminary screening tool before resorting to more computationally expensive methods. In practice, mcGFN-FF is particularly valuable for periodic systems where NCIs play a significant role. A notable application involves simulating C60 guest molecules within a MOF

composed of zirconium-oxide clusters connected by porphyrin linkers.[4] Initial locations of the guest molecules within the MOF's porous channels were explored using mcGFN-FF and later confirmed with GFN2-xTB calculations. The two energetically favorable positions were then investigated further with DFT methods. Additionally, the MD simulations with the FF elucidated the mobility of the C60 molecules at the two positions and revealed solvent diffusion between porous channels of the MOF. Further analysis of MD trajectories could offer valuable information on interactions, such as average distances between the MOF and C60 or between neighboring guest molecules. The model has been enhanced to better handle the challenging 4f and 5f elements through updates to the electronegativity equilibration (EEQ) charge model, optimized covalent radii, and element-pair-specific bond length corrections. Example applications demonstrate the model's effectiveness in MD simulations of lanthanide-containing compounds. Radial distribution functions obtained from the MD simulations show a good trade-off between robustness and flexibility for these challenging complexes. The presented work demonstrates the model's utility for molecular and periodic systems, including challenging applications such as metal-organic frameworks (MOFs), Ln/An complexes, molecular crystals, and biomolecular systems.

When situating GFN-FF within the landscape of computational methods, it offers an intermediate position between FFs and SQM methods. Its quadratic scaling with the number of atoms makes it more computationally demanding than conventional FFs. However, this is offset by its general applicability across most elements of the periodic table, contrasting with the domain-specific specialization of other FFs. Furthermore, the model performs similarly to the SQM method GFN1-xTB for many tested systems, concerning geometries and lattice energies. Special purpose FFs such as CHARMM36[193], ff14SB[194] and OPLS4[60] excel in biomolecular simulations due to their finely tuned parameters for proteins, nucleic acids, and small organic molecules. These FFs offer simulations of proteins with sizes beyond the capabilities of the GFN-FF model but are mostly limited to the biomolecular field. In contrast, universal FFs like MMFF,[57] UFF,[56] and DREIDING[55] aim for generality across the periodic table but lack the precision required for complex electronic systems or challenging geometries. Tight-binding methods, including GFN1-xTB, GFN2-xTB, and DFTB3[195], offer a compelling middle ground by combining efficiency with good accuracy for medium-sized molecular systems. ML potentials such as ANI[74,196] and packages like DeepMD-kit[197,198] bring high accuracy by leveraging large datasets, often at the DFT level. Despite their promise, these methods require system-specific training data and are often limited in general applicability, making them less versatile for early-stage exploratory research. Unlike specialized or highly parameterized methods, GFN-FF balances general applicability, accuracy, and robustness across a broad spectrum of chemical systems. While not ideal for high-precision tasks or large proteins, GFN-FF excels in early-stage exploration, particularly for systems combining diverse atom types and electronic complexities.

Future applications could include testing the method for modeling surfaces. Furthermore, integrating GFN-FF into automated workflows and screening pipelines will amplify its impact across computational chemistry, from drug design to advanced material discovery.

In the second part of this thesis, a workflow is developed for screening molecular tweezers in terms of their affinity for binding to specific sugars. Molecular tweezers are a crucial class of compounds for molecular recognition and delivery. These tweezers typically feature flexible structures that allow the binding sites to adjust their conformations to interact with the guest optimally. Depending on the flexibility of the tweezer, the binding sites may be able to interact with each other. In such cases, solvent interactions can have a significant impact on reaction-free energies by modulating the free energy of the

unbound tweezer molecule. This workflow is specifically tailored to calculate reaction-free energies for systems involving flexible ligands, where conformational entropy, intramolecular interactions, and solvation effects play a crucial role. In particular, the binding of two diboronic acid (DBA) ligands to glucose or galactose serves as a case study for this methodology due to the availability of detailed experimental data for these reactions.

The workflow comprises two key components: The QCG algorithm in CREST is utilized to create conformer ensembles with a predefined number of explicit solvent molecules added to the binding sites of the molecular tweezers. Together, this allows for largely automated calculation of reaction-free energies. However, both tools' accuracy and efficiency largely depend on the settings of their parameters, making an in-depth study not only valuable for this kind of screening but for other workflows as well that utilize these tools. Sampling efficiency largely depends on the length of the MTDs. In our case, 80 ps was effective for exploring conformational space, as determined by the convergence of the root-mean-square deviation (RMSD) with respect to the starting structure. The completeness of the CREST runs is evaluated by performing multiple conformer searches for each reaction-free energy calculation. Statistical outliers are identified using multimodal Gaussian distributions derived from multiple calculations of reaction-free energies. The addition of explicit water molecules to the conformer ensemble was found to be critical for obtaining accurate free energies. The exact number of solvent molecules to add could not be linked to a geometric indicator, but averaging over a finite number of calculations with different numbers of solvent molecules was effective. The CENSO tool refines the generated conformer ensemble by incrementally increasing the level of theory and incorporating additional contributions. The combination of functionals and the applied sorting thresholds were the main investigated parameters in the ensemble refinement with the CENSO tool. Adjusting the energy thresholds for sorting out conformers has a significant impact on computational efficiency, with optimized settings reducing computational costs by a factor of 22 compared to less stringent thresholds, while still maintaining accuracy. Conformational entropy emerges as a critical factor in this workflow. The formation of a macrocyclic complex from flexible molecules results in substantial entropy penalties, which must be considered to accurately predict reaction-free energies.

This study highlights the challenges posed by microsolvation in the calculation of reaction-free energies, where CREST and CENSO procedure, often effective in other investigations, fails to deliver accurate or efficient results with default settings. The presented work provides a practical guideline for identifying and adjusting key parameters to improve the accuracy and precision of similar investigations. The optimized workflow achieves reaction-free energy predictions with errors as low as 2 kcal/mol, making it a powerful tool for designing DBA ligands with high selectivity for specific carbohydrates.

APPENDIX A

Abbreviations

The abbreviations used throughout this thesis are listed below.

Abbreviation	Meaning
AO	atomic orbital
ACF	adiabatic connection formula
An	actinides
BJ	Becke-Johnson
BO	Born-Oppenheimer
BSIE	basis set incompleteness error
BSSE	basis set superposition error
CG	coarse-grained
CP	counterpoise
CSP	crystal structure prediction
DBA	diboronic acid
DFA	density functional approximation
DFT	density functional theory
DFTB	density functional tight-binding
DH	double hybrid
FF	force field
GTO	Gaussian-type orbitals
GGA	generalized gradient approximation
HF	Hartree-Fock
HOMO	highest occupied molecular orbital
IR	infrared
KS	Kohn-Sham

Abbreviation	Meaning
LCAO	linear combination of atomic orbitals
LDA	local density approximation
LJ	Lennard-Jones
Ln	lanthanides
LSDA	local spin density approximation
LUMO	lowest unoccupied molecular orbital
MARD	mean absolute relative deviation
MC	monte carlo
mcGFN-FF	molecular crystals GFN-FF
MD	molecular dynamics
ML	machine learning
MO	molecular orbital
MOF	metal-organic framework
MSD	mean signed deviation
MTD	meta-dynamics
NCI	non-covalent interaction
PBC	periodic boundary condition
PT	perturbation theory
QSAR	quantitative structure-activity relationship
RDF	radial distribution function
RMSD	root mean squared deviation
RPA	random phase approximation
SCF	self-consistent field
SIE	self-interaction error
SQM	semiempirical quantum mechanical
vdW	van der Waals
WFT	wave function theory
ZPV	zero-point vibrational

APPENDIX B

mcGFN-FF: an accurate force field for optimization and energetic screening of molecular crystals

Stefan Grimme[1] and Thomas Rose[1]

Received: October 5, 2023
Published online: April 5, 2024

Reprinted in Appendix B with permission from S. Grimme and T. Rose, Z. Naturforsch., B **79** (2024) 191–200 –
DOI: https://doi.org/10.1515/znb-2023-0088

Own contributions

- Implementation of periodic boundary conditions for GFN-FF in the `xtb` code
- Parameterization and implementation of GFN-FF run mode for molecular crystals
- Compilation of new benchmark sets
- Performing all of the calculations
- Interpretation of the results
- Writing the manuscript

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

Stefan Grimme* and Thomas Rose

mcGFN-FF: an accurate force field for optimization and energetic screening of molecular crystals

https://doi.org/10.1515/znb-2023-0088
Received October 5, 2023; accepted November 15, 2023

Abstract: This work presents a periodic extension of the GFN-FF force field for molecular crystals named mcGFN-FF. Non-covalent interactions in the force field are adjusted to reduce the systematic overbinding of the original, molecular version for molecular crystals. A diverse set of molecular crystal benchmarks for lattice energies and unit cell volumes is studied. The modified force field shows good results with a mean absolute relative deviation (MARD) of 19.9 % for lattice energies and 10.0 % for unit cell volumes. In many cases, mcGFN-FF approaches the accuracy of the GFN1-xTB quantum chemistry method which has an MARD of 18.7 % for lattice energies and 6.2 % for unit cell volumes. Further, the newly compiled mcVOL22 benchmark set is presented which features r^2SCAN-D4/900 eV DFT reference volumes for molecular crystals with phosphorus-, sulfur-, and chlorine-containing compounds of various sizes. Overall, the mcGFN-FF poses an efficient tool for the optimization and energetic screening of molecular crystals containing elements up to radon.

Keywords: benchmark; periodic; molecular solid

Dedicated to Professor Thomas Bredow of the University of Bonn on the occasion of his 60th birthday.

1 Introduction

Due to the substantial growth of the synthetically accessible chemical space in recent decades [1–3], experimental exploration has become more time-consuming and expensive. Computational chemistry assists these efforts by identifying and ranking energetically favorable candidates. A prominent application in drug development is the mandatory screening for polymorphs. Crystal structure prediction software commonly uses atomistic force fields (FFs) to generate an initial ensemble of polymorphs. Therefore, accurate FF methods are essential to ensure that all relevant structures are included in the ensemble [4]. This work evaluates the newly developed periodic extension of the molecular GFN-FF force field [5] for molecular crystals named mcGFN-FF on lattice energy and unit cell volume benchmarks.

Early FFs were limited to mostly organic molecules or specific applications. Examples are DREIDING [6] for organic and a few inorganic main group elements, and AMBER [7] or CHARMM [8] for proteins. The first universal force field, parameterized for almost the entire periodic table, was introduced with UFF [9]. Being applicable to systems including elements up to radon, the GFN-FF [5] found a good balance between generality and accuracy in addition to its ease of use. These features motivated the implementation of periodic boundary conditions for the GFN-FF published by Gale et al. leading to the pGFN-FF [10]. In their publication, the authors present various adjustments needed to accurately describe various forms of solids. To stay with the original focus of GFN-FF on non-covalent interactions, the mcGFN-FF is specifically designed for molecular crystals.

The next chapter explains the theory of the changes applied in the periodic implementation. Subsequently, different benchmark studies are presented and discussed to demonstrate the performance of mcGFN-FF. The method's versatility is showcased through optimizations of three systems with intricate electronic structures. Finally, limitations are discussed for the energetic screening of polymorphs and off-target calculations on covalently bound metal-organic frameworks.

2 Theory

To apply the GFN-FF method to molecular crystals, periodic boundary conditions are introduced using the supercell method. In this approach, copies of the unit cell are generated around it until all interactions within given cutoffs are considered.

***Corresponding author: Stefan Grimme**, Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Beringstraße 4, Bonn 53115, Germany, E-mail: grimme@thch.uni-bonn.de
Thomas Rose, Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Beringstraße 4, Bonn 53115, Germany, E-mail: thomas.rose@thch.uni-bonn.de

The mcGFN-FF energy $E_{\text{mcGFN-FF}}$ can be split into covalent (cov) and non-covalent interactions (NCI) with a total of ten additive contributions according to

$$E_{\mathbf{mc}\text{GFN-FF}} = E_{\text{cov}} + E_{\mathbf{NCI}} \tag{1}$$

$$E_{\text{cov}} = E_{\text{rep}}^{\text{bond}} + E_{\text{bond}} + E_{\text{bend}} + E_{\text{tors}} + E_{\text{abc}}^{\text{bond}} \tag{2}$$

$$E_{\text{NCI}} = \mathbf{c_{ES}} E_{\text{ES}}^{\mathbf{Ewald}} + E_{\text{disp}}^{\mathbf{s_8}} + \mathbf{c_{HB}} E_{\text{HB}} + E_{\text{XB}} + \mathbf{c_{rep}^{NCI}} E_{\text{rep}}^{\text{NCI}}. \tag{3}$$

The contributions are the bonded repulsion energy $E_{\text{rep}}^{\text{bond}}$, the bond energy E_{bond}, the bending energy E_{bend}, the torsion energy E_{tors}, the three-body bond correction energy $E_{\text{abc}}^{\text{bond}}$, the electrostatic energy E_{ES}, the London dispersion energy E_{disp}, the hydrogen bond energy E_{HB}, the halogen bond energy E_{XB}, and the non-bonded repulsion energy $E_{\text{rep}}^{\text{NCI}}$. Adjustments and scaling factors c_i for the energies of the periodic implementation are printed in bold.

Moving from single molecules to periodic boundary conditions mainly changes the effect of non-covalent interactions. Therefore, changes to the electrostatic energy, London dispersion, hydrogen-bond energy, and non-bonded repulsion have been applied to optimize the performance of the method for molecular crystals.

The electrostatic energy

$$E_{\text{es}} = \frac{1}{2}\vec{q}\left(\mathbf{A}\cdot\vec{q} - \vec{\chi}\right) \tag{4}$$

is calculated from the partial charges $\vec{q}$, the Coulomb matrix $\mathbf{A}$ and the method specific electronegativity $\vec{\chi}$. When introducing periodic boundary conditions, it becomes necessary to ensure convergence of the electrostatic energy, which is done via Ewald summation [11]. This method splits the energy into two parts

$$E_{\text{es}}\left(\frac{1}{R}\right) = E_{\text{es}}\left(\frac{1-\operatorname{erf}(\xi R)}{R} + \frac{\operatorname{erf}(\xi R)}{R}\right),$$

where the first term converges in real space and the second in reciprocal space. This introduces the Ewald splitting parameter ξ that determines the ratio of the two terms. The golden section search (GSS) [12] is applied to obtain an optimal Ewald splitting parameter. This algorithm finds a local minimum of a strictly unimodal function within a given interval. Here, the optimal parameter is the argument of the minimum of

$$\Delta f = |f_{\text{rec}} - f_{\text{real}}| \quad \text{with} \tag{5}$$

$$f_{\text{rec}} = \frac{4\pi}{V r_{\text{rec}}^2}\exp\left(-\frac{r_{\text{rec}}^2}{4\xi^2}\right) \quad \text{and} \tag{6}$$

$$f_{\text{real}} = -\frac{\operatorname{erf}(\xi r_{\text{real}})}{r_{\text{real}}} + \frac{\operatorname{erf}(\gamma r_{\text{real}})}{r_{\text{real}}}, \tag{7}$$

where Δf is the absolute value of the difference between estimates of the largest contribution in reciprocal space f_{rec} and the largest contribution in real space f_{real} to the electrostatic energy. The norms of the smallest reciprocal r_{rec} and real space r_{real} lattice vector are used to obtain the largest contributions. The local minimum is searched in the interval from 10^{-8} to 2.

The dispersion energy

$$E_{\text{disp}}^{(6,8)} = -\sum_{\text{AB}}\sum_{n=6,8} s_n \frac{C_{(n)}^{\text{AB}}}{R_{\text{AB}}} f_{\text{damp}}^{(n)}(R_{\text{AB}}) \tag{8}$$

is calculated from the dispersion coefficients $C_{(n)}^{\text{AB}}$, the atomic distance R_{AB}, the rational Becke-Johnson (BJ) damping function [13] $f_{\text{damp}}^{(n)}$ and the dispersion scaling s_n.

The periodic gradient and stress tensor are implemented to obtain equilibrium geometries using the L-BFGS algorithm [14].

3 Computational details

The optimizations for the mcVOL22 benchmark set were performed with the Vienna Ab Initio Simulation Package (VASP) version 6.3.2 [15–18] utilizing the r^{2}SCAN *meta*-GGA density functional [19] with added D4 dispersion correction [20, 21]. The projector-augmented-wave (PAW) method [22] is applied with POTCAR files taken from the VASP library as listed in Table S6 in the Supporting Information. The calculations include plane waves up to 900 eV and use an automatic mesh of *k*-points as listed in Table S7. Gaussian smearing with a width of 0.01 eV is applied to aid the convergence of the electronic self-consistent field equations. Structures are considered as converged with an energy difference smaller than 1.2×10^{-4} eV, gradient norm smaller than 1.3×10^{-2} eV/Å, and stress tensor norm smaller than 5.7×10^{-2} eV/Å^3. Converged values are also listed in Table S7. Final reference volumes were obtained by fitting the Birch-Murnaghan equation of state [23] to energy-volume data points. The data points were calculated with single-point calculations on scaled unit cells with scaling factors ranging from 0.98 to 1.04 with a step size of 0.02.

All calculations for the GFN-FF and mcGFN-FF were performed with a development version of xtb [24]. The final implementation will be freely available on GitHub.

Calculations with the pGFN-FF were performed with the General Utility Lattice Program (GULP) version 6.1.0 [25, 26]. In their publication, the authors present two extensions to the force field. First, the use of the Wolf sum (W) to handle Coulomb interactions and calculate robust charges,

and second the damping of the three-body bond correction (dATM). Results for pGFN-FF W + dATM are taken from the original publication, while results for pGFN-FF W were calculated with the GULP program. Calculations with GFN1-xTB [27] and UFF were performed with the Amsterdam Modeling Suite (AMS) [28]. To ensure accurate results for GFN1-xTB, the k-space sampling quality was set to "very good" as implemented in the utilized DFTB engine [29].

All energies and unit cell volumes presented in the results section are calculated through free optimizations. In the boxplots discussed below, we exclude outliers defined as deviations that are either larger than the third quartile plus 1.5 times the interquartile range or smaller than the first quartile minus 1.5 times the interquartile range.

4 Results and discussion

4.1 Studied benchmarks

Herein, the various test sets for evaluating the mcGFN-FF are introduced.

The **X23** benchmark [30] comprises a diverse set of 23 small organic molecules with different binding motives. In the revision of the benchmark [31], lattice energies are calculated from experimental sublimation enthalpies by correcting for harmonic vibrational contributions obtained as an average over four density-functional approximations. Reference unit cell volumes were calculated by considering thermal and zero-point energy effects.

The **G60** benchmark [32] primarily consists of rigid molecular crystals and includes multiple compounds containing chlorine or nitro groups. The original publication mentions partially low accuracy of the experimental sublimation enthalpies, with an estimated standard deviation of 10 %. A revision of the benchmark [33] provides a constant correction of 2RT for the experimental values. Here, the revised values are taken as reference values. Since theoretical lattice energies allow a direct comparison of methods, results with sHF-3c [34] reference lattice energies and unit cell volumes as calculated in the revision are presented in the Supporting Information.

To gauge the performance of the force field on larger biological systems, lattice energies and unit cell volumes are evaluated for a set of eight peptide structures (**PEP8**) each consisting of six or seven amino acids. The initial geometries were taken from Schmitz et al. [35] and optimized with r^{2}SCAN-3c [36].

The **ICE10** [37] benchmark provides reference lattice energies and unit cell volumes for ten ice polymorphs.

Ionic liquids (ILs) have become popular in the chemical industry, among other things, due to being good reaction solvents with a low vapor pressure [38]. To test the accuracy of the mcGFN-FF on these difficult systems, back-corrected unit cell volumes of five IL crystals (**IL5**) were extracted from Červinka [39]. The reference values were calculated from the experimental unit cell volumes by approximating the thermal expansion with the quasi-harmonic approximation using B3LYP-D3 [40, 41]/pob-TZVP-rev2 [42].

4.2 Results for the fit set

To optimize the performance of mcGFN-FF for molecular crystals, the s_8 dispersion scaling factor for interactions between atoms belonging to different molecules (intermolecular) is adjusted as well as the newly introduced scaling factors c_i for the electrostatic energy, the hydrogen-bond energy, and the non-bonded repulsion energy. These four scaling factors were fitted simultaneously on the revised energies of the X23 benchmark [31]. In the fitting procedure, the lattice energies are obtained from single-point calculations on structures optimized with the original parameterized periodic implementation of the GFN-FF. The final optimum scaling factors are $c_{\mathrm{ES}} = 0.800$, $s_8^{\mathrm{frag}} = 2.859$, $c_{\mathrm{HB}} = 0.727$, and $c_{\mathrm{rep}}^{\mathrm{NCI}} = 1.343$ for electrostatic, London dispersion, hydrogen-bond, and non-covalent repulsion interactions respectively. Figure 1 depicts deviations from the reference lattice energies for the tested FFs and GFN1-xTB. Overall the methods perform best on the non-polar Van der Waals (VdW) subset and worst on the mixed or hydrogen-bonded subset. The mean absolute deviations (MAD) and mean deviations (MD) for each subset can be found in the Supporting Information in Table S1. The effect of the adjusted parameterization of GFN-FF is seen as a shift towards less stable crystals, with an exception for systems dominated by non-polar Van der Waals (VdW) interactions. This effectively reduces the MAD from 4.8 kcal/mol to 2.0 kcal/mol and reduces the systematic overbinding, reflected in the MD of −4.7 kcal/mol, to zero. The pGFN-FF W + dATM shows a similar progression as GFN-FF with an also similar MAD of 5.0 kcal/mol. After the fit, mcGFN-FF performs even better than GFN1-xTB which has an MAD of 2.4 kcal/mol and an MD of 0.7 kcal/mol. The timings for this benchmark are provided in the Supporting Information.

As a first validation of the fit, the deviations from back-corrected experimental unit cell volumes are depicted in Figure 2 as percentage values. Even though the volumes were not used in the reparameterization of the periodic GFN-FF, the MARD is substantially reduced from 10.2 % to 4.7 %. The underestimation of unit cell volumes is reduced

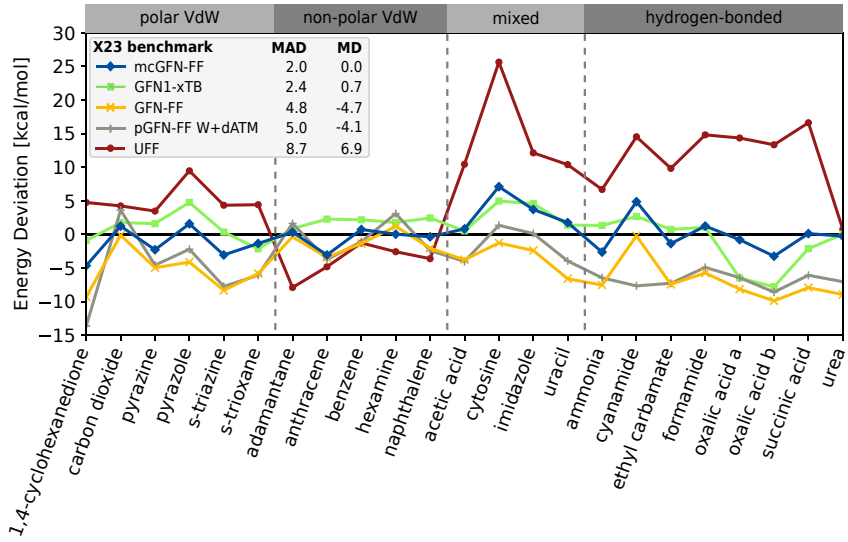


Figure 1: Deviation from back-corrected experimental sublimation enthalpies for compounds of the revised X23 benchmark. The key includes mean absolute deviations and mean deviations for the tested methods in kcal/mol.

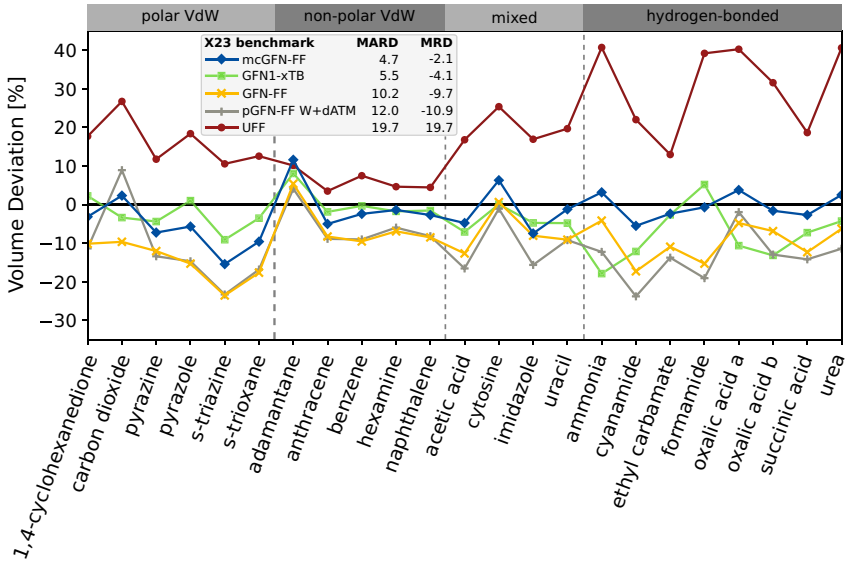


Figure 2: Deviation from back-corrected experimental unit cell volumes of the revised X23 benchmark. The deviations are given as percentage values relative to the reference.

from −9.7 % to −2.1 %. With an MARD of 12.0 % pGFN-FF W + dATM provides slightly worse volumes than GFN-FF. Similar to the lattice energies, GFN1-xTB performs slightly worse than mcGFN-FF with an MARD of 5.5%.

4.3 Molecular crystal volumes benchmark set

In order to cover a wider variety of atom types and structural properties we present the mcVOL22 benchmark set. The set provides r^{2}SCAN-D4 reference volumes for 22 molecular crystals that contain phosphorus, sulfur, and chlorine atoms. Illustrations of representative systems are shown in Figure 3 and the whole set is depicted in the Supporting Information. The deviations from the reference volumes are depicted in Figure 4 for the tested methods. The slight overbinding of the crystals with a mean relative deviation (MRD) of −4.4 % for GFN1-xTB is consistent with the results on the X23 benchmark. The higher-level method only shows one larger deviation for **mcv14**. With an MRD of −7.7 % too small cell volumes, mcGFN-FF is the best-performing force field on this benchmark, followed by UFF with 8.0 %. All GFN-FF type methods show the largest deviation for **mcv02**, a small rigid system including phosphorus. The universal force field has the largest deviation of 22 % for **mcv01** which contains Cl^- counter ions.

4.4 Statistical evaluation

The results for the presented benchmarks are depicted in Figure 5. The MARD for the energies does not include the X23

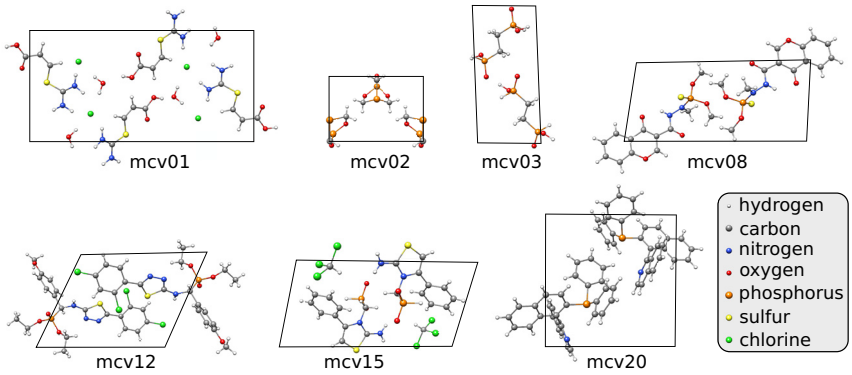


Figure 3: Selected structures from the mcVOL22 benchmark for molecular crystal volumes.

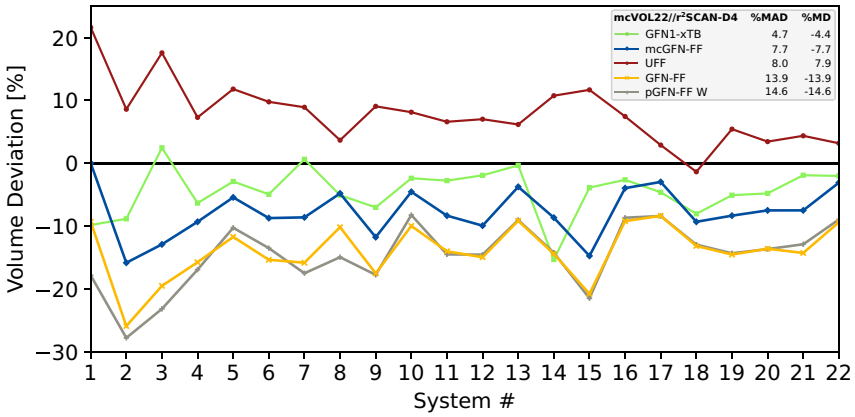


Figure 4: Deviation from r²SCAN-D4 unit cell volumes. The key includes mean absolute relative deviations and mean relative deviations for the tested methods.

benchmark since it was used to fit the mcGFN-FF. Due to small values for lattice energies and unit cell volumes, the ICE10 benchmark has a high impact on the MARD and therefore values without this benchmark are provided as indicated. The boxplots show that the adjustments in the mcGFN-FF consistently reduce the systematic overbinding present in an unmodified periodic implementation of GFN-FF. Overall, our reparameterization improves the results significantly, with an exception for the ICE10 benchmark. In this case, the original method already provides accurate results, leading to larger errors in the adjusted variant due to the reduced stabilization of the crystals. As indicated by the MARD, mcGFN-FF performs comparable to GFN1-xTB, while GFN-FF shows similar results as pGFN-FF. Though the UFF provides accurate cell volumes, it shows the largest standard deviation overall.

4.5 Showcases

In order to demonstrate the versatility of the mcGFN-FF we present three systems with intricate electronic structures, including a system with a molybdenum-silicon triple bond [43], an osmium system with BF_4^- counter ion [44, 45], and a siladodecahedrane system [46]. Optimized structures are illustrated in Figure 6 with an overlay of the experimental structure to show the changes in geometry. In order to ensure a robust optimization of these systems, the applied coordinate displacement in the L-BFGS optimizer is damped with a factor

$$f(i)_{\text{damp}} = \frac{1}{1 + 3000i^{-3}} \tag{9}$$

that scales with the optimization step i.

The structures were downloaded from the Cambridge Structural Database (CSD) with identifiers LIDWED, DIXRIN, and AXOBAT respectively. The metallasylidyne system (a) shows the smallest structural deviation with a root mean squared deviation (RMSD) of 0.53 Å from the experimental crystal structure and a −8.2 % smaller unit cell. After optimization, the silicon-molybdenum triple bond is 0.18 Å longer with a length of 2.47 Å and the silicon-molybdenum single bond becomes 0.15 Å longer with a length of 2.6 Å. The osmium system (b) has slightly larger deviations with an RMSD of 0.55 Å and a −12.2 % smaller unit cell. Finally, the siladodecahedrane system (c) has an RMSD of 0.91 Å and a −20.3 % smaller unit cell. For systems **b** and **c** the inclusion of chloroform molecules in the

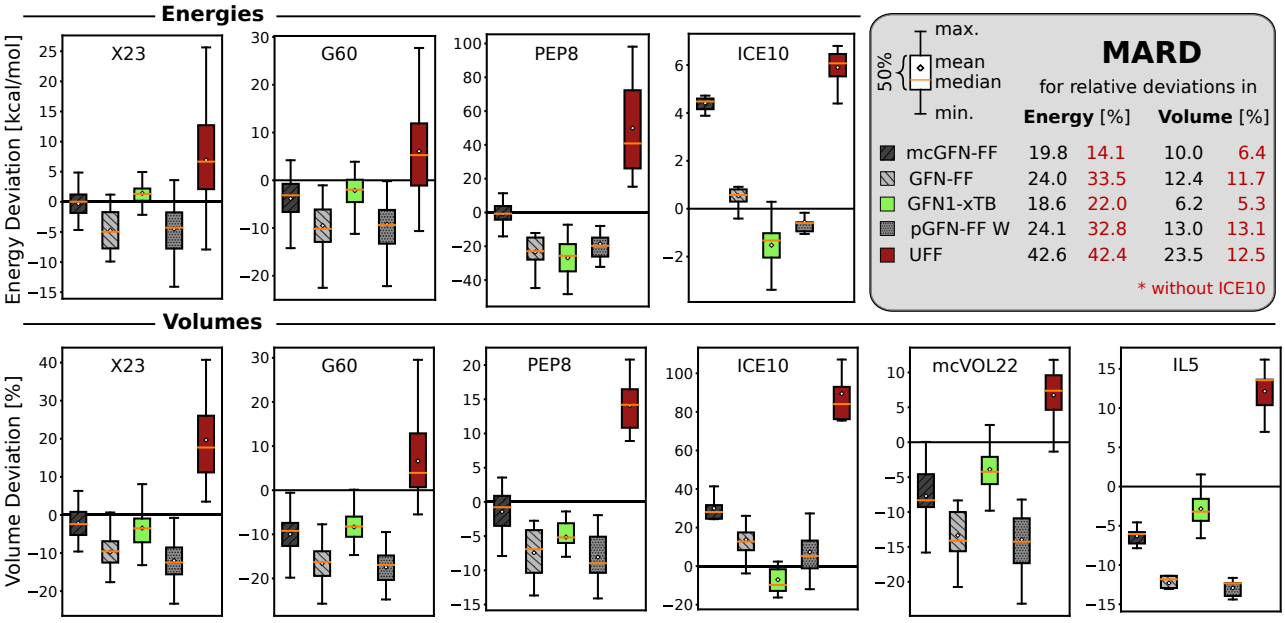


Figure 5: Boxplots for deviations in lattice energies and relative deviations in unit cell volumes for various test sets. The provided MARDs are calculated as averages over the given benchmarks. Since the X23 energies are used to fit the mcGFN-FF, they are excluded from the corresponding MARD for all methods. Maximum and minimum deviations do not include outliers.

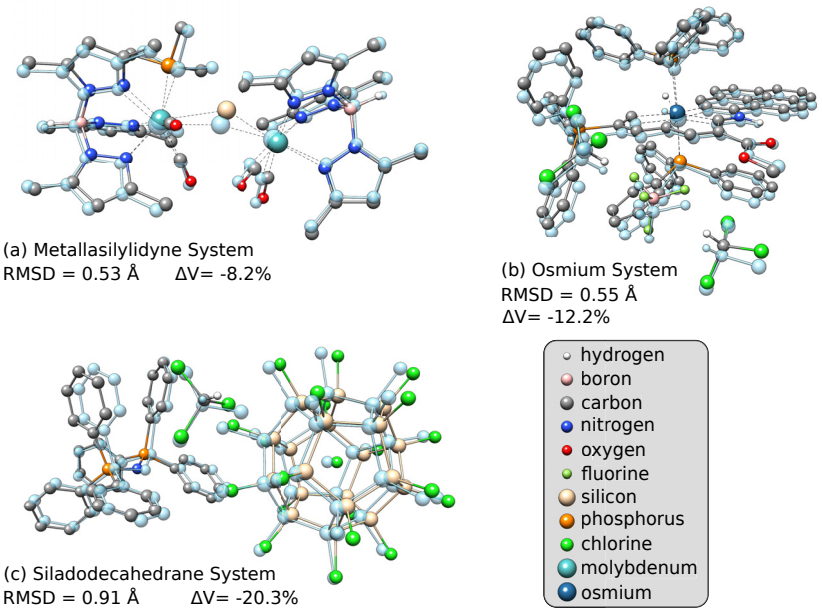


Figure 6: Illustrations of systems with intricate electronic structures that were optimized with mcGFN-FF under periodic boundary conditions. The experimental structure is depicted as a blue overlay. To make the structures easier to recognize, hydrogen atoms that are bound to carbon are hidden and not the entire unit cell is shown. The RMSD and volume deviation are calculated between the entire unit cell of the experiment and the optimized structure from mcGFN-FF.

crystal structures is an additional challenge. System **a** was optimized in 43.7 min and 608 optimizations steps, system **b** was optimized in 12.5 min and 313 steps, and system **c** was optimized in 23.7 min and 318 steps. Even though there are noticeable changes in geometry, it is remarkable that the force field can optimize these systems at all without major structural distortions compared to experiment.

4.6 Polymorph ranking on ROY

Six polymorphs of the well-studied 5-methyl-2-[(2-nitrophenyl) amino]-3-thiophenecarbonitrile crystal (dubbed ROY) [47–53] are ranked energetically with FF methods and GFN1-xTB. The experimental reference is taken from Ref [54]. The polymorphs are named yellow prisms (Y), red prisms (R), orange plates (OP), orange needles (ON), yellow needles (YN), and orange-red plates (ORP). As all polymorphs lie within an energy range of 1.0 kcal/mol, the energetic ranking becomes a tough challenge for all tested methods. The results for this "real-life" application case are depicted in Figure 7. The enthalpic contributions are negligible for comparison to FF methods, as shown by high-level calculations from Beran et al [47]. All force fields assign a rank of two to the most stable form (Y), while GFN1-xTB assigns a rank of four. The widening of the energy range $\Delta E_{\text{range}} = E_{\text{max}} - E_{\text{min}}$ is similar for the GFN-FF related methods with approximately 10 kcal/mol. For GFN1-xTB the widening is smaller with 5.7 kcal/mol and significantly higher for UFF with 19.1 kcal/mol. With a nitro group and a cyano group, the molecule itself is a challenging system let alone energetically ranking its polymorphs with low-cost methods. Therefore, the results should only be considered relative to each other.

4.7 POLY59

The POLY59 benchmark [55] comprises five sets of eleven to fifteen polymorph structures, numbered from 22 through 26 and each set has one (or five for set 23) reference structure that should yield the lowest energy. To evaluate this benchmark, all sets of structures have to be sorted energetically and the rank of the reference structure is summed up over the five sets. The ranking is performed via single-point calculations on given TPSS-D3 structures. Since set 23 has five references the best score is 19 and the worst score is 109. In order to provide an additional reference point, assigning the rank of the reference structure with an equal distribution yields a summed score of 64. Results for the tested FFs as well as results from the benchmark publication are listed in Table 1. With a score ($\sum$) of 42 and 43 respectively, pGFN-FF W and GFN-FF are the best-performing FFs. Though mcGFN-FF is just slightly worse than HF-3c, with a score of 67 compared to 65, the ranking is worse than in the unmodified version although we note a slightly improved energy range. The UFF is not able to work reliably on DFT structures, as reflected in the score of 84 and the immense average energy range between the polymorph with the lowest and the highest energy $\Delta E_{\text{range}}^{\text{avg}}$ of 549.6 kcal/mol. Since the accuracy of force field energies on DFT structures is practically not very relevant, the benchmark has been evaluated with free optimizations of the structures by the FFs. The score and energy range from the optimization is given in brackets for the FFs. While GFN-FF maintains almost the same score, mcGFN-FF and UFF improve substantially with scores of 57 and 48.

Though the accurate energetic ranking of polymorphs with FF methods is not possible yet, testing the methods for the generation of crystal structure ensembles could be worthwhile. In addition, mcGFN-FF may be useful for initial screening steps in multi-level computational workflows.

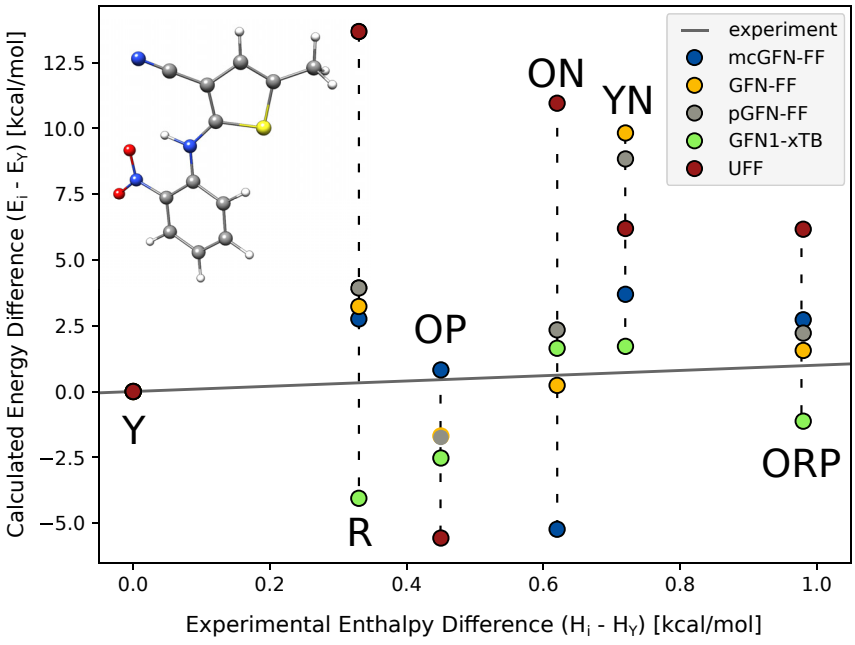


Figure 7: Calculated energy difference between the most stable form (Y) and other ROY polymorphs plotted against experimental enthalpy difference. The molecular structure of ROY is illustrated in the figure.

Table 1: Rank of the energetically lowest polymorph for the five sets according to different methods. The score ($\sum$) for each method is calculated as the sum over the different set ranks. For set 23 there are 5 "lowest" polymorphs $\alpha - \epsilon$. The average energy range from the energetically lowest to the highest polymorph ΔE^{avg}_{range} is given to measure how similar the potential energy surface is to TPSS-D3. The results in brackets are obtained from free optimizations. Values for non-force field methods are from the POLY59 publication, where the value in the 23α column is the sum of the entire set.

Rank	22	23α	23β	23γ	23δ	23ϵ	24	25	26	Score $\sum$		ΔE^{avg}_{range} [kcal/mol]	
mcGFN-FF	10	10	1	14	8	13	4	2	5	67	(57)	21.8	(20.4)
GFN-FF	5	5	2	7	4	8	3	3	6	43	(44)	22.4	(27.9)
pGFN-FF W	4	5	1	7	4	8	4	3	6	42	(43)	24.6	(28.3)
UFF	6	6	11	12	10	15	9	4	11	84	(48)	549.6	(24.6)
HF-3c	5	49	–	–	–	–	1	5	8	65	–	4.9	–
DFTB3-D3	7	52	–	–	–	–	1	6	8	74	–	7.0	–
M06L	1	41	–	–	–	–	1	1	7	51	–	3.5	–
PBE-D3	1	21	–	–	–	–	1	1	1	25	–	2.5	–

4.8 Off-target evaluation on 12 metal-organic frameworks (MOF12)

Though mcGFN-FF was specifically designed for molecular crystals it is evaluated for 12 covalently bound metal-organic frameworks (MOFs) using PBE-D3 reference cell volumes [56]. Note that some optimizations with GFN-FF had to be restarted after deleting the topology file or needed the construction of supercells to ensure convergence. The frameworks each include one of the following elements: silver, cadmium, cobalt, copper, dysprosium, iron, lithium, zinc, or samarium. Figure 8 shows boxplots of the relative deviations from PBE-D3 unit cell volumes for the tested methods. Even though mcGFN-FF is not developed for these systems, it performs comparable to GFN1-xTB.

5 Conclusions

A periodic extension of the GFN-FF for molecular crystals, named mcGFN-FF has been presented. To introduce periodic boundary conditions into the existing implementation, the Ewald summation is applied to converge the electrostatic energy. Herein, the Ewald splitting parameter is determined via the golden section search algorithm. In mcGFN-FF the intermolecular non-covalent energy contributions are adjusted, to reduce the systematic overbinding of molecular crystals observed in the original method.

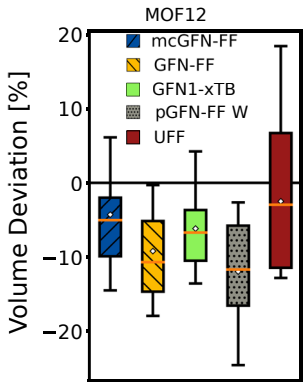


Figure 8: Boxplots of the relative deviations from PBE-D3 unit cell volumes for 12 MOFs.

While the original periodic implementation already shows good and robust results across the diversified benchmarks, mcGFN-FF provides significantly more accurate lattice energies and unit cell volumes in almost all cases. The only exception is the ICE10 benchmark. Here, the unmodified implementation already yields accurate results and the reduced non-covalent binding in mcGFN-FF leads to larger errors.

In principle, the mcGFN-FF can be applied to any periodic system, but the focus on molecular crystals mandates treating other systems with caution. Showing comparable results as low-cost QM methods for polymorph ranking, we suggest the use of mcGFN-FF in the first steps of crystal structure prediction workflows. With promising results on the transition metal showcases, the method poses a useful tool for gathering first insights on intriguing chemical problems.

To this end, precompiled binaries as well as the source code will be freely available on the corresponding Github website (https://github.com/grimme-lab/xtb).

6 Supporting Information

Detailed Benchmark results, illustrations of the mcVOL22 systems, and used POTCAR files for VASP calculations are given as supplementary material and are available online (https://doi.org/10.1515/znb-2023-0088). Furthermore, geometry files, used fit data, and reference values for the mcVOL22 and PEP8 benchmarks are available as a ZIP file.

Acknowledgment: The authors thank S. Ehlert for providing the code for the L-BFGS algorithm, U. Huniar (3DS) and J. M. Mewes for the development version of TURBOMOLE including the periodic r^{2}SCAN-3c implementation, and T. Bredow for assistance with the VASP calculations. Furthermore, the authors thank S. Spicher, S. Ehlert, T. Gasevic, and M. Bursch for many helpful discussions.

Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: The authors state no conflict of interest.

Research funding: None declared.

Data availability: The raw data can be obtained on request from the corresponding author.

References

1. Medina-Franco J. L., Chávez-Hernández A. L., López-López E., Saldívar-González F. I. Chemical multiverse: an expanded view of chemical space. *Mol. Inf.* 2022, *41*, 12, 2200116.
2. Campos K. R., Coleman P. J., Alvarez J. C., Dreher S. D., Garbaccio R. M., Terrett N. K., Tillyer R. D., Truppo M. D., Parmee E. R. The importance of synthetic chemistry in the pharmaceutical industry. *Science* 2019, *363*, 9, eaat0805.
3. Theerthagiri J., Karuppasamy K., Lee S. J., Shwetharani R., Kim H.-S., Pasha S. K. K., Ashokkumar M., Choi M. Y. Fundamentals and comprehensive insights on pulsed laser synthesis of advanced materials for diverse photo- and electrocatalytic applications. *Light: Sci. Appl.* 2022, *11*, 47, 250.
4. Bowskill D. H., Sugden I. J., Konstantinopoulos S., Adjiman C. S., Pantelides C. C. Crystal structure prediction methods for organic molecules: state of the art. *Annu. Rev. Chem. Biomol. Eng.* 2021, *12*, 593–623.
5. Spicher S., Grimme S. Robust atomistic modeling of materials, organometallic, and biochemical systems. *Angew. Chem. Int. Ed.* 2020, *59*, 15665–15673.
6. Mayo S. L., Goddard W. A., Olafson B. D. DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* 1990, *94*, 8897–8909.
7. Weiner S. J., Kollman P. A., Nguyen D. T., Case D. A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* 1986, *7*, 230–252.
8. Brooks B. R., Bruccoleri R. E., Olafson B. D., States D. J., Swaminathan S., Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 1983, *4*, 187–217.
9. Rappe A. K., Casewit C. J., Colwell K. S., Goddard W. A. I., Skiff W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* 1992, *114*, 10024–10035.
10. Gale J. D., LeBlanc L. M., Spackman P. R., Silvestri A., Raiteri P. A universal force field for materials, periodic GFN-FF: implementation and examination. *J. Chem. Theory Comput.* 2021, *17*, 7827–7849.
11. Ewald P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* 1921, *369*, 253–287.
12. Kiefer J. Sequential minimax search for a maximum. *Proc. Am. Math. Soc.* 1953, *4*, 502–506.
13. Becke A. D., Johnson E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* 2005, *123*, 9, 154101.
14. Liu D. C., Nocedal J. On the limited memory BFGS method for large scale optimization. *Math. Program.* 1989, *45*, 503–528.
15. Kresse G., Hafner J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* 1993, *47*, 558–561.
16. Kresse G., Furthmüller J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* 1996, *6*, 15–50.
17. Kresse G., Furthmüller J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* 1996, *54*, 11169–11186.
18. Kresse G., Joubert D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* 1999, *59*, 1758–1775.
19. Furness J. W., Kaplan A. D., Ning J., Perdew J. P., Sun J. Accurate and numerically efficient r2SCAN meta-generalized gradient approximation. *J. Phys. Chem. Lett.* 2020, *11*, 8208–8215.
20. Caldeweyher E., Ehlert S., Hansen A., Neugebauer H., Spicher S., Bannwarth C., Grimme S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* 2019, *150*, 19, 154122.
21. Caldeweyher E., Mewes J.-M., Ehlert S., Grimme S. Extension and evaluation of the D4 London-dispersion model for periodic systems. *Phys. Chem. Chem. Phys.* 2020, *22*, 8499–8512.
22. Blöchl P. E. Projector augmented-wave method. *Phys. Rev. B* 1994, *50*, 17953–17979.
23. Birch F. Finite elastic strain of cubic crystals. *Phys. Rev.* 1947, *71*, 809–824.
24. Bannwarth C., Caldeweyher E., Ehlert S., Hansen A., Pracht P., Seibert J., Spicher S., Grimme S. Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* 2021, *11*, 49, e1493.
25. Gale J. D. GULP: a computer program for the symmetry-adapted simulation of solids. *J. Chem. Soc., Faraday Trans.* 1997, *93*, 629–637.
26. Gale J. D., Rohl A. L. The general utility lattice program (GULP). *Mol. Simul.* 2003, *29*, 291–341.
27. Grimme S., Bannwarth C., Shushkov P. 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). *J. Chem. Theory Comput.* 2017, *13*, 1989–2009.
28. Rüger R., Franchini M., Trnka T., Yakovlev A., van Lenthe E., van Vuren T., Klumpers B., Soini T. *Theoretical Chemistry*; AMS 2022; Vrije Universiteit: Amsterdam, The Netherlands. http://www.scm.com.
29. Rüger R., Yakovlev A., Philipsen P., Borini S., Melix P., Oliveira A., Franchini M., van Vuren T., Soini T., de Reus M., Asl M. G., Teodoro T. Q., McCormack D., Patchkovskii S., Heine T. *Theoretical Chemistry*; AMS DFTB 2023; SCM; Vrije Universiteit: Amsterdam, The Netherlands. http://www.scm.com.
30. Reilly A. M., Tkatchenko A. Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals. *J. Chem. Phys.* 2013, *139*, 12, 024705.
31. Dolgonos G. A., Hoja J., Boese A. D. Revised values for the X23 benchmark set of molecular crystals. *Phys. Chem. Chem. Phys.* 2019, *21*, 24333–24344.
32. Maschio L., Civalleri B., Ugliengo P., Gavezzotti A. Intermolecular interaction energies in molecular crystals: comparison and agreement of localized møller–Plesset 2, dispersion-corrected density functional,

and classical empirical two-body calculations. *J. Phys. Chem. A* 2011, *115*, 11179–11186.
33. Cutini M., Civalleri B., Corno M., Orlando R., Brandenburg J. G., Maschio L., Ugliengo P. Assessment of different quantum mechanical methods for the prediction of structure and cohesive energy of molecular crystals. *J. Chem. Theory Comput.* 2016, *12*, 3340–3352.
34. Caldeweyher E., Brandenburg J. G. Simplified DFT methods for consistent structures and energies of large systems. *J. Phys.: Condens. Matter* 2018, *30*, 40, 213001.
35. Schmitz S., Seibert J., Ostermeir K., Hansen A., Göller A. H., Grimme S. Quantum chemical calculation of molecular and periodic peptide and protein structures. *J. Phys. Chem. B* 2020, *124*, 3636–3646.
36. Grimme S., Hansen A., Ehlert S., Mewes J.-M. r2SCAN-3c: a "Swiss army knife" composite electronic-structure method. *J. Chem. Phys.* 2021, *154*, 18, 064103.
37. Brandenburg J. G., Maas T., Grimme S. Benchmarking DFT and semiempirical methods on structures and lattice energies for ten ice polymorphs. *J. Chem. Phys.* 2015, *142*, 11, 124104.
38. Singh S. K., Savoy A. W. Ionic liquids synthesis and applications: an overview. *J. Mol. Liq.* 2020, *297*, 23, 112038.
39. Červinka C. Tuning the quasi-harmonic treatment of crystalline ionic liquids within the density functional theory. *J. Comput. Chem.* 2022, *43*, 448–456.
40. Lee C., Yang W., Parr R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 1988, *37*, 785–789.
41. Becke A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 1993, *98*, 5648–5652.
42. Laun J., Bredow T. BSSE-corrected consistent Gaussian basis sets of triple-zeta valence with polarization quality of the fifth period for solid-state calculations. *J. Comput. Chem.* 2022, *43*, 839–846.
43. Ghana P., Arz M. I., Chakraborty U., Schnakenburg G., Filippou A. C. Linearly two-coordinated silicon: transition metal complexes with the functional groups M ≡Si-M and M=Si=M. *J. Am. Chem. Soc.* 2018, *140*, 7187–7198.
44. Zhu Q., Zhu C., Deng Z., He G., Chen J., Zhu J., Xia H. Synthesis and characterization of osmium polycyclic aromatic complexes via nucleophilic reactions of osmapentalyne. *Chin. J. Chem.* 2017, *35*, 628–634.
45. Zhu C., Zhu Q., Fan J., Zhu J., He X., Cao X.-Y., Xia H. A metal-bridged tricyclic aromatic system: synthesis of osmium polycyclic aromatic complexes. *Angew. Chem. Int. Ed.* 2014, *53*, 6232–6236.
46. Bamberg M., Bursch M., Hansen A., Brandl M., Sentis G., Kunze L., Bolte M., Lerner H.-W., Grimme S., Wagner M. [Cl@Si20H20]-: parent siladodecahedrane with endohedral chloride ion. *J. Am. Chem. Soc.* 2021, *143*, 10865–10871.
47. Beran G. J. O., Sugden I. J., Greenwell C., Bowskill D. H., Pantelides C. C., Adjiman C. S. How many more polymorphs of ROY remain undiscovered. *Chem. Sci.* 2022, *13*, 1288–1297.
48. Yu L., Stephenson G. A., Mitchell C. A., Bunnell C. A., Snorek S. V., Bowyer J. J., Borchardt T. B., Stowell J. G., Byrn S. R. Thermochemistry and conformational polymorphism of a hexamorphic crystal system. *J. Am. Chem. Soc.* 2000, *122*, 585–591.
49. Lévesque A., Maris T., Wuest J. D. ROY reclaims its crown: new ways to increase polymorphic diversity. *J. Am. Chem. Soc.* 2020, *142*, 11873–11883.
50. Chen S., Xi H., Yu L. Cross-nucleation between ROY polymorphs. *J. Am. Chem. Soc.* 2005, *127*, 17439–17444.
51. Tan M., Shtukenberg A. G., Zhu S., Xu W., Dooryhee E., Nichols S., Ward M. D., Kahr B., Zhu Q. ROY revisited, again: the eighth solved structure. *Faraday Discuss*. 2018, *211*, 477–491.
52. Gushurst K. S., Nyman J., Boerrigter S. X. M. The PO13 crystal structure of ROY. *CrystEngComm* 2019, *21*, 1363–1368.
53. Chen S., Guzei I. A., Yu L. New polymorphs of ROY and new record for coexisting polymorphs of solved structures. *J. Am. Chem. Soc.* 2005, *127*, 9881–9885.
54. Yu L. Polymorphism in molecular solids: an extraordinary system of red, orange, and yellow crystals. *Acc. Chem. Res.* 2010, *43*, 1257–1266.
55. Brandenburg J. G., Grimme S. Organic crystal polymorphism: a benchmark for dispersion-corrected mean-field electronic structure methods. *Acta Crystallogr.* 2016, *B72*, 502–513.
56. Nazarian D., Ganesh P., Sholl D. S. Benchmarking density functional theory predictions of framework structures and properties in a chemically diverse test set of metal–organic frameworks. *J. Mater. Chem. A* 2015, *3*, 22432–22440.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/znb-2023-0088).

APPENDIX C

Fast and Robust Modeling of Lanthanide and Actinide Complexes, Biomolecules, and Molecular Crystals with the extended GFN-FF Model

Thomas Rose[1], Markus Bursch[2], Jan-Michael Mewes[3], and Stefan Grimme[1]

Received: July 30, 2024 Published online: September 28, 2024

Reprinted in Appendix C with permission from[4] T. Rose, M. Bursch, J.-M. Mewes, and S. Grimme, *Fast and Robust Modeling of Lanthanide and Actinide Complexes, Biomolecules, and Molecular Crystals with the extended GFN-FF Model*. Inorg. Chem. **63** (2024) 19364–19374

DOI: https://doi.org/10.1021/acs.inorgchem.4c03215

Own contributions

- Parameterization of the force field for actinides
- Improvement of the force field parameters for lanthanides
- Performing all of the calculations
- Drafting the manuscript
- Major contributions to editing the manuscript
- Major contributions to the interpretation of the results

[1] Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Beringstraße 4, Bonn 53115, Germany
[2] FACCTs GmbH, Rolandstraße 67, 50677, Koeln, Germany
[3] beeOLED GmbH, Niedersedlitzer Str. 75c, 01257 Dresden, Germany
[4] Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

Inorganic Chemistry

pubs.acs.org/IC Article

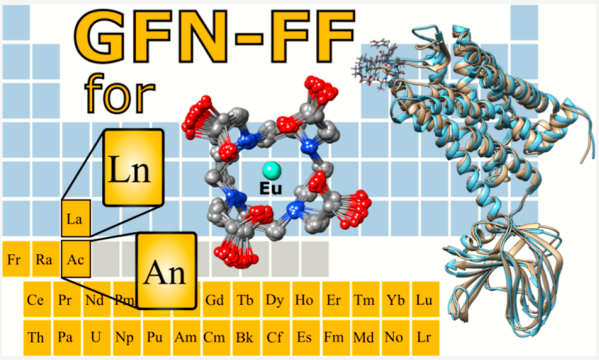
Fast and Robust Modeling of Lanthanide and Actinide Complexes, Biomolecules, and Molecular Crystals with the Extended GFN-FF Model

Thomas Rose, Markus Bursch, Jan-Michael Mewes, and Stefan Grimme*

Cite This: *Inorg. Chem.* 2024, 63, 19364−19374 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Lanthanides (Ln) and actinides (An) have recently become important tools in biomedical and materials science. However, the development of computational methods able to describe such elements in various environments has not kept up with the pace of the field. Addressing this challenge, this work introduces and showcases an extension of the GFN-FF to An alongside a reparameterization for Ln. This development fills a gap for fast computational methods that are out-of-the-box applicable to large f-element-containing systems with thousands of atoms. We discuss the reparameterization of the charge model and the covalent topology setup and showcase the model through various applications: Molecular dynamics simulations, optimization of Ln-containing biomolecules, and optimizations of several periodic structures. With the presented improvements, GFN-FF is a powerful method that routinely delivers robust and accurate geometries for large Ln/An systems with thousands of atoms.

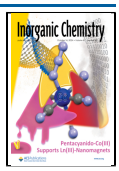


INTRODUCTION

Lanthanides (Ln) and actinides (An) are of significant interest in biomedical research, having applications, e.g., in imaging and targeted therapy using radiopharmaceuticals.[1,2] Important tools for researching Ln- and An-containing molecules are methods that can optimize and simulate large biomolecules at an atomistic level. Although numerous force fields (FFs) exist that accurately describe proteins,[3−10] to the best of our knowledge, only the MMFF[11] and UFF[12] include a rudimentary parametrization for all Ln/An elements. Additionally, we found isolated parameters for certain elements (e.g., Th(IV), Lu(III), and La(III)).[13] Other FF are specially designed for molecular dynamics (MD) simulations of Ln/An-cations or systems in an aqueous environment and cover a wider number of actinides or lanthanides.[14−19] Although quantum chemical tight binding methods are available and valuable tools for theoretical investigations of Ln complexes,[20,21] their application is limited to systems with hundreds of atoms, excluding many biologically relevant molecules. For MD simulations, a density functional tight-binding method has been developed for uranium.[22] Ln/An also play crucial roles in other fields, such as in developing organic light-emitting diodes,[23] and in processing radioactive waste.[24,25] Fast force field methods are valuable tools for screening conformational space in these areas. Providing a method that can elucidate the structures of large molecules or crystals that include Ln/An is essential for advancing both medical and technological applications involving these elements.

However, defining the nature of the chemical bonding with Ln/An atoms is challenging even with DFT methods[26] due to complex electronic configurations and significant relativistic effects.[27] Thus, modeling every aspect of this chemistry with a FF is not feasible. While it is possible to address uncommon bonding motifs as special cases within the method, doing so requires a substantial amount of reference data and can compromise the robustness of the method, e.g., if numerous geometrical constraints are introduced to fix the structure of an Ln/An complex. To ensure maintainability and robustness, new parameters or interaction potentials should be introduced thoughtfully. Consequently, modeling such systems effectively with a limited number of interaction types and parameters necessitates a focus on relevant target properties. Since the focus of the GFN-FF is on geometries, frequencies, and noncovalent interactions (GFN), a reasonable aim is to provide good structural properties. Addressing this challenge, this work presents an extension of the GFN-FF[28] method to An and a reparameterization for Ln that applies to molecular and periodic structures. The method is generic or generally applicable in the sense that it can be applied for systems

Received: July 30, 2024
Revised: September 11, 2024
Accepted: September 17, 2024
Published: September 28, 2024



https://doi.org/10.1021/acs.inorgchem.4c03215
Inorg. Chem. 2024, 63, 19364−19374

including any element up to lawrencium, and enables robust optimizations, energy calculations, and MD simulations of molecular and periodic systems. The implementation of periodic boundary conditions was recently published with the introduction of a specialized run mode for molecular crystals called mcGFN-FF.[29] The widespread application of the original FF, reflected in over 200 citations within about four years, underscores its robustness across a wide array of applications, thereby highlighting the promising potential of this extension. Selected literature discussing the application and performance of the original method is given in refs 28−33.

This paper is structured as follows: The Theory section covers the automatic FF setup and the interaction potentials that have been reparameterized. In the Methods section, we detail the applied algorithms, density functionals as references, and procedures used to enhance and extend the GFN-FF. The Results and Discussion section presents the reparameterization alongside notable test sets that showcase the method's effectiveness for optimizations, energy calculations, and MD simulations. Finally, the conclusion summarizes the key findings that make the method appealing for use in Ln/An computational chemistry.

THEORY

The GFN-FF energy $E_{\text{GFN-FF}}$ consists of covalent (cov) and noncovalent interactions (NCI) with a total of ten additive contributions according to

$$E_{\text{GFN-FF}} = E_{\text{cov}} + E_{\text{NCI}} \tag{1}$$

$$E_{\text{cov}} = E_{\text{rep}}^{\text{bond}} + E_{\text{bond}} + E_{\text{bend}} + E_{\text{tors}} + E_{\text{abc}}^{\text{bond}} \tag{2}$$

$$E_{\text{NCI}} = E_{\text{ES}} + E_{\text{disp}} + E_{\text{HB}} + E_{\text{XB}} + E_{\text{rep}}^{\text{NCI}} \tag{3}$$

These interaction potentials are the bonded repulsion energy $E_{\text{rep}}^{\text{bond}}$, the bond energy E_{bond}, the bending energy E_{bend}, the torsion energy E_{tors}, the three-body bond energy $E_{\text{abc}}^{\text{bond}}$, the electrostatic energy E_{ES}, the London dispersion energy E_{disp}, the hydrogen bond energy E_{HB}, the halogen bond energy E_{XB}, and the nonbonded repulsion energy $E_{\text{rep}}^{\text{NCI}}$. The mainly element-specific parametrization enables selective adjustments for lanthanides and the introduction of new parameters for actinides, without affecting the potentials that do not include these elements. Many of the interactions in the FF are scaled with the partial charges of the atoms, making the underlying electronegativity equilibration (EEQ) model[34] one of the most vital parts. Herein, the partial charges $\vec{q}$ are determined by solving a set of linear equations

$$\begin{pmatrix} \mathbf{A} & \vec{1} \\ \vec{1}^{T} & 0 \end{pmatrix} \begin{pmatrix} \vec{q} \\ \lambda \end{pmatrix} = \begin{pmatrix} \vec{X} \\ q_{\text{tot}} \end{pmatrix} \tag{4}$$

which includes the Coulomb matrix $\mathbf{A}$, the Lagrange multiplier λ, the $\vec{X}$ vector, and the total charge of the system q_{tot}. The Coulomb matrix

$$A_{ij} = \begin{cases} \Gamma_{ii} + \dfrac{2\gamma_{ii}}{\sqrt{\pi}} & \text{for } i = j \\ \dfrac{\text{erf}(\gamma_{ij} R_{ij})}{R_{ij}} & \text{otherwise} \end{cases} \tag{5}$$

includes the two empirical parameters: Gamma Γ and the atomic radii α in $\gamma_{ij} = \frac{1}{\sqrt{\alpha_i^2 + \alpha_j^2}}$. Another two empirical parameters are included in the X vector

$$X_i = \kappa_i \sqrt{\log\left(\frac{1 + \exp(\text{CN}_{\text{max}})}{1 + \exp(\text{CN}_{\text{max}} - \text{mCN}_i)}\right)} - \chi_i \tag{6}$$

calculated from the modified D3 coordination number mCN and the maximum coordination number CN_{max}. These parameters are the coordination number scaling factor κ and the fitted atomic electronegativities χ. Thus, the four empirical parameters adjusted for Ln/An in this work are chi χ, gamma Γ, kappa κ, and alpha α. For the setup of the FF parameters, a second set of topological EEQ partial charges is calculated. Herein, the distances are determined by the shortest covalent path between two atoms. To obtain the distances, the covalent radii of all atom pairs on this path are summed up.

The topology setup, specifically the assignment of covalent bonds is essential to most FF. Since the setup is completely automated, it demands additional attention when handling highly coordinated elements such as Ln/An. In GFN-FF, a covalent bond is assigned between two atoms i and j if the distance between the two atoms R_{ij} (in the initial structure) is smaller than the atom-pair-specific threshold R_{ij}^0 that depends on the empirical element-specific covalent radii r_0^{NB} and is scaled with a charge-dependent scaling factor $f_R(q)$. The inequality for this assignment can be written as

$$R_{ij} < f_R(q_i, q_j) R_{ij}^0 (r_{0,i}^{\text{NB}}, r_{0,j}^{\text{NB}}) \tag{7}$$

and can be solved for the r_0^{NB} of one of the two atoms. Note that in GFN-FF the neighbor list only contains bonded neighbors (NB).

For the interactions between Ln/An atoms and the coordinated ligands, the orbital overlap, Pauli repulsion, and electrostatic interactions vary in importance for different elements.[35,36] Modeling these interactions accurately within the framework of a FF is not a straightforward task, and thus demands a trade-off between computational cost, accuracy, and robustness. A crucial requirement in this regard is that Ln- or An-complexes should not dissociate any ligands during optimizations or MD. Therefore, we choose to assign covalent bonds even for those Ln- or An-bonds dominated by electrostatic interactions as an efficient means to achieve this goal. Specifically, the bond potential

$$E_{\text{bond},ij} = -k_{\text{bond},1} \exp(-k_{\text{bond},2}(R_{ij} - R_{ij}^0)^2) \tag{8}$$

is applied for these bonds. It includes prefactors $k_{\text{bond},1}$ and $k_{\text{bond},2}$ and depends on the distance of the two atoms R_{ij} which is compared to the reference bond distance R_{ij}^0. Other attractive potentials still apply as before. A detailed explanation of the automated setup of the FF parameters and further details on the contents of the energies are given in the Supporting Information of the original publication.[28]

METHODS

The reparameterization and parametrization of the EEQ parameters for Ln/An elements are carried out with the NLOPT python package utilizing the local derivative-free BOBYQA optimizer.[37,38] The fit uses the root-mean-square deviation (rmsd) of partial charges of all atoms regarding DFT reference values and is considered converged if the change in fit parameters is smaller than 1.0×10^{-7}. The fit set for the actinides is derived from the actinides quantum mechanics (AcQM)

https://doi.org/10.1021/acs.inorgchem.4c03215
Inorg. Chem. 2024, 63, 19364−19374

data set presented by Wittmann et al.[39] This data set comprises 2537 actinide complexes with diverse ligands and coordination numbers ranging from four to 10. The data set features PBE0-NL optimized geometries and ωB97M-V/ma-def-TZVP Hirshfeld partial charges. To obtain new EEQ parameters for the lanthanides, all actinide atoms in the AcQM set were replaced by the corresponding lanthanide, and the systems were reoptimized using ωB97M-V/def2-TZVPP[40,41] with def2-ECP effective core potentials[42] and ma-def2-SVP basis set[43] for the lanthanides. The reference Hirshfeld charges are evaluated on the electron densities of the optimized systems. The EEQ parameters are fitted using a quadratic interpolation between La, Gd, and Lu for the lanthanides and between Ac, Cm, and Lr for the actinides. For both fit sets, only reference charges between −1 and 2 are included and systems are excluded as outliers if the deviation of the Ln/An partial charges obtained with the initial GFN-FF from the reference is larger than 1.5*e*. To ensure some robustness of the model, it has been important to include hydrides and halogens in the fit sets in addition to the already diverse ligands. Both fit sets, the input geometries for the MDs, the optimized biomolecular structures, and the isomer structures are available on our GitHub in the repository dedicated to this work.[44]

The covalent radii for the neighbor setup r_0^{NB} are optimized for Ln/An using a semiautomated procedure: First, all r_0^{NB} for Ln/An atoms are calculated using all surrounding atoms within a given distance *d*. Then the atom corresponding to the maximum r_0^{NB} is inspected manually and, if the atom should be a neighbor to the Ln/An atom, the r_0^{NB} is adopted in the implementation. If this is not the case, the distance *d* is reduced by 0.01 ua, and the procedure is repeated until an optimal r_0^{NB} is found for each Ln/An. This procedure utilized the PBE0-D4/def2-SVP geometries from the LnQM data set[45] for the lanthanides and the PBE0-NL geometries in the fit set derived from the AcQM for the actinides. The obtained covalent radii are always rounded up to the second decimal place to ensure that the corresponding neighbor is identified as such. Since the threshold for determining neighbor assignment depends on the atomic charges, the optimal r_0^{NB} are calculated using the reparameterized EEQ model.

The initial parameters for the actinides as well as for francium and radium are taken from the corresponding element of the sixth period. Three exceptions are the global hardness parameter *c*, Pauling electronegativities *en*, and covalent radii r_{topo} used to determine distances for the calculation of topological partial charges. The values for the actinides are taken from the same source as the other elements' parameters but from a later edition.[46−49]

Bond lengths in Ln/An complexes can vary greatly depending on the oxidation state, charges, and surrounding ligands. To tackle even the most challenging systems, we allow the user to manually define the covalent bond neighbors of atoms via a detailed input file that is read by `xtb`. Further information on how to use this feature is given in the Supporting Information and the xTB documentation. The oxidation state itself is not recognized by the method, and it only affects the FF indirectly through the change in total system charge and the input geometry.

In many complexes, hydrogen atoms will be close to the Ln/An center while being covalently bonded only to the ligand atoms. In such cases, the partial charge of the hydrogen atom is slightly positive, based on the electronegativity of common ligand atoms (C, N, and O). In hydride complexes, where the hydrogen is directly bound to an Ln/An atom, the hydrogen has a slightly negative partial charge for the same reason.[50] This allows a distinction between hydridic hydrogen atoms and hydrogen atoms coordinated only to ligand atoms based on the partial charge of the hydrogen. Therefore, to avoid wrong topologies concerning Ln/An-H bonds, the FF removes hydrogen atoms as neighbors to Ln/An, if the topology-based charge of the hydrogen is greater than −0.0281*e*.

The isomers for relative isomer energies have been generated manually, starting from systems of the LnQM data set. All systems have been optimized with r^2SCAN-3c[51] and subsequent single-point calculations with ωB97X-V/def2-mTZVPP[51,52] are performed for validation. The r^2SCAN-3c energies are used as reference values because they proved to be more robust. The relative isomer energies and further discussion are provided in the Supporting Information.

Throughout the manuscript, the FF version including an EEQ reparameterization, optimized covalent radii, shifted bond length parameters for halogens, and a charge-based neighbor threshold for hydrogen atoms is called the extended version.

RESULTS AND DISCUSSION

EEQ Parametrization. The performance with the original lanthanide EEQ parameters is shown in Figure 1a and the

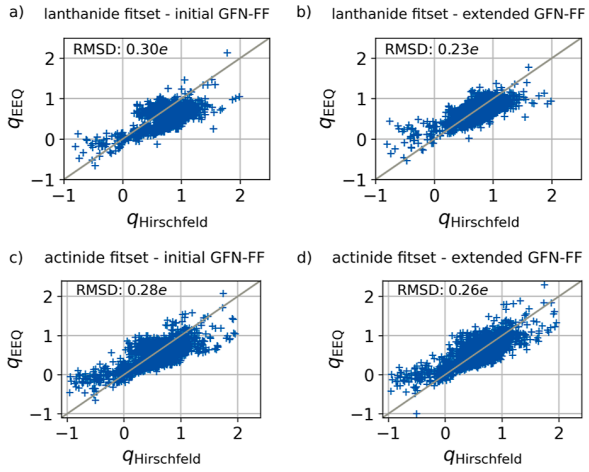


Figure 1. Comparison of Ln/An partial charges from Hirshfeld analysis with EEQ partial charges from (a) the initial parametrization of the GFN-FF on the Ln fit set, (b) the extended GFN-FF on the Ln fit set, (c) the initial parametrization of the GFN-FF on the An fit set, and (d) the extended GFN-FF on the An fit set. The rmsd between the FF charges and the Hirshfeld charges is given for each panel. All partial charges are given in elementary charges.

results with the extended FF are shown in Figure 1b. Even though the new parameters reduce the rmsd of the partial charges by 23%, the large standard deviation of the Ln-partial charges remains, especially for those below 0.0*e* and above 1.5*e*. The results for the actinides are depicted in Figure 1c, d. The fit improves the rmsd by 7% and shows a similar progression in the standard deviation as observed for the lanthanides. The fitted parameters for Ln/An are listed in the Supporting Information in Tables S1 and S2 respectively.

Optimal Covalent Radii for Lanthanides and Actinides. A central challenge in modeling Ln/An complexes using FF with a (mostly) fixed covalent topology is the elusive nature of the bonds to heteroatoms. These bonds are often predominantly electrostatic with minimal covalent contributions. However, any FF requires a topology and preliminary work shows that purely electrostatic models for these bonds have major shortcomings. Hence, we decided to model Ln/An-heteroatom bonds as having covalent character.

The optimal covalent radii are listed in Table 1 for each Ln/An. The parameters for the lanthanides decrease from lanthanum to lutetium in agreement with the lanthanide contraction. This is not the case for the actinides, where the covalent radii decrease from actinium to lawrencium but show a local maximum for berkelium. In this regard, the surroundings of the An-atom have a different effect on actinides than they have on lanthanides. This has also been observed by Serezhkin et al.[53] for AnX crystals with Th, U, Np, Pu, Am, Cm for An and S, Se, and Te for X.

https://doi.org/10.1021/acs.inorgchem.4c03215
Inorg. Chem. 2024, 63, 19364−19374

Table 1. Optimal Covalent Radii Parameter r_0^{NB} for Ln/An Given in Å

Ln	r_0^{NB}	An	r_0^{NB}
La	1.79	Ac	1.82
Ce	1.70	Th	1.65
Pr	1.69	Pa	1.46
Nd	1.68	U	1.42
Pm	1.70	Np	1.57
Sm	1.61	Pu	1.51
Eu	1.63	Am	1.51
Gd	1.63	Cm	1.59
Tb	1.62	Bk	1.68
Dy	1.55	Cf	1.65
Ho	1.58	Es	1.39
Er	1.59	Fm	1.41
Tm	1.54	Md	1.41
Yb	1.55	No	1.46
Lu	1.46	Lr	1.39

LnQM. The LnQM is a large data set including PBE0-D4/def2-SVP optimized structures and Hirshfeld charges on ωB97M-V/def2-SVPD densities for a diverse set of Ln-complexes. With hundreds of accurate geometries for each lanthanide, this set is ideal for testing and improving the performance of GFN-FF.

The Ln-neighbor bond lengths d(Ln-NB) are analyzed separately for each atom type of the neighboring atom as listed in Table 2. This evaluation showed an overall improvement in the extended version and a systematic shift for halogen atoms toward shorter bond distances, excluding Fluorine which is shifted to larger bond distances, with the initial version. In this evaluation, neighbors are defined by the applied FF version. Therefore, the number of neighbors evaluated N_{NB} differs due to the improved covalent radii in the extended version. Since halogen atoms are only included as direct neighbors to the lanthanides in the LnQM, the mean signed deviation (MSD) being close to the mean absolute deviation (MAD) allows for improving their bond length parameters by shifting them according to the MSD. This reduces the bond length MSD and MAD substantially for the extended version. The scaling introduces a pairwise factor exclusively for halogen-actinide bonds and halogen-lanthanide bonds. For the actinides, the halogen bond lengths have been improved based on the actinide fit set, where the total MAD in d(Ln-NB) is reduced from 0.14 to 0.12 Å and the MSD from 0.05 to 0.03 Å. Unlike the initial version, the extended model does not assign any covalent bonds between the central lanthanide atom and hydrogen atoms, which is expected since the LnQM does not include any hydride complexes. The substantial increase in neighbors for carbon atoms is linked to ligands such as carboxylate ligands where the carbon atom is next to the atom that coordinates with the lanthanide. Since the overall bond length varies greatly between different ligands, it can not be avoided that such atoms are bound to the lanthanide in some cases. Excluding specific Ln/An-C bonds is not as simple as the treatment of Ln/An-H bonds and might cause more harm than it does good. This issue was less pronounced for nitrogen atoms in for example nitrate ligands with the initial version. In this case, the problem has been fixed by the new covalent radii in the extended version leading to a reduced number of neighbors, MSD, and MAD. The MADs for the Ln-neighbor bond lengths are listed in Table 3. With the extended FF, the MAD of the bond lengths is reduced from 0.18 to 0.14 Å averaged over all lanthanides.

Finally, the MADs of partial charges of the lanthanide atoms in the LnQM have been evaluated. The rmsd increases from 0.17e to 0.20e since the fit set also considers halogens and hydrides in addition to the already diverse ligands in the LnQM. The MADs are listed in Table 3 for each lanthanide. The charges are plotted against the reference charges in Figure S1 in the Supporting Information. Even though the MAD worsens with the extended version, there is no clear correlation between EEQ and Hirshfeld charges in either version. Hence, the further use of the FF charges for Ln/An complexes is not recommended and resulting dipole moments should be treated with caution.

12-Coordinate Actinide MOFs. Metal−organic frameworks (MOFs) are promising candidates in radioactive waste management.[24] To test the applicability of the FF in this area, porous MOFs (An(IV)-ADC) with 12-coordinated actinides[54] are optimized with the periodic mcGFN-FF and compared to experimental results. The deviations in structural measures for the MOFs containing thorium, uranium, neptunium, or plutonium are listed in Table 4. Absolute deviations in actinide−oxygen bond distances are below 0.12 Å. The rectangular pores of the MOFs are preserved with absolute deviations in pore diameters below 0.31 Å. The bite angles obtained with the FF are accurate with deviations up to 4.2°. The dihedral angles between the anthracene and the carbonyl

Table 2. MSD and MAD of Ln−NB Bond Distances for All Ln Neighbors in the LnQM Sorted by Atom Type of the Neighbor[a]

	initial version				extended version			
	N_{NB}	ref dist.	MSD	MAD	N_{NB}	ref dist.	MSD	MAD
H	93	3.08	−0.77	0.77	0			
C	6767	2.64	−0.27	0.28	8018	2.68	−0.26	0.27
N	49,487	2.54	−0.12	0.20	48,922	2.53	−0.02	0.16
O	48,079	2.37	−0.10	0.15	48,425	2.36	−0.02	0.12
F	2851	2.07	0.07	0.09	2850	2.07	−0.01	0.07
P	3918	3.03	−0.24	0.24	4404	3.05	−0.15	0.15
S	1409	2.96	−0.35	0.35	1485	2.97	−0.26	0.26
Cl	2575	2.63	−0.14	0.15	2575	2.63	0.01	0.07
Br	2524	2.82	−0.19	0.20	2529	2.82	0.01	0.08
I	2489	3.07	−0.23	0.23	2521	3.08	−0.00	0.10

[a]Only atoms identified as neighbors by the used GFN-FF were evaluated and the number of neighbors is given in column N_{NB}. Distances are given in Å.

https://doi.org/10.1021/acs.inorgchem.4c03215
Inorg. Chem. 2024, 63, 19364−19374

Table 3. MAD of Bond Distances for All Ln-Neighbor Bonds d(Ln-NB) in the LnQM for the Initial Version and the Extended Version of the FF, as Well as MADs from Hirshfeld Charges[a]

	d(Ln-NB)		q	
MADs	initial	extended	initial	extended
La	0.19	0.16	0.18	0.23
Ce	0.23	0.16	0.17	0.31
Pr	0.22	0.16	0.16	0.29
Nd	0.21	0.15	0.17	0.25
Pm	0.20	0.14	0.16	0.23
Sm	0.19	0.15	0.16	0.19
Eu	0.19	0.14	0.16	0.17
Gd	0.18	0.14	0.17	0.20
Tb	0.18	0.15	0.18	0.15
Dy	0.18	0.15	0.16	0.14
Ho	0.16	0.13	0.15	0.15
Er	0.16	0.13	0.15	0.15
Tm	0.15	0.14	0.14	0.13
Yb	0.15	0.13	0.15	0.12
Lu	0.14	0.14	0.27	0.18

[a]Here, neighbors are defined by the topology of the FF. The deviations in bond distances are given in Å and the MADs for the partial charges q are given in e.

Table 4. Evaluation of the mcGFN-FF on Four An(IV)-ADC Crystals[a]

An(IV)-ADC	Th	U	Np	Pu
Δd_{An-O} [Å]	−0.09	−0.12	−0.06	−0.07
Δd_{vert} [Å]	0.28	0.17	0.24	0.19
Δd_{hor} [Å]	−0.29	−0.31	−0.21	−0.21
$\Delta \alpha_{bite}$ [deg]	3.42	4.17	3.12	3.17
$\Delta \Theta_{twist,min}$ [deg]	−1.08	−1.16	−1.30	−1.37
$\Delta \Theta_{twist,max}$ [deg]	−0.70	−0.74	−1.14	−1.18

[a]Deviations between the FF and experimental values are given for actinide oxygen distance d_{An-O}, vertical pore diameter d_{vert}, horizontal pore diameter d_{hor}, carboxylate bite angle α_{bite}, and anthracene plane twisting dihedral angle Θ_{twist}. For the dihedral angle, the minimum and maximum values observed in the optimized structure are compared to the experiment.

deviate by up to 1.4° from the experiment. Overall, the FF could be an interesting method to test for prescreening of MOFs in this research area. Since the FF does not apply any symmetry, the distances and angles could differ between the eight actinide centers in the unit cell. However, apart from the anthracene-plane twisting dihedral angle Θ_{twist}, the standard deviation for the reported measures is zero for the given number of decimal points. To capture the range of twisting dihedral angles in the FF geometries, the deviations are reported for the minimum and maximum angles. The calculated values and the unit cell are illustrated in Figure 2. The experimental values, calculated structural properties, timings, and optimization cycles for the optimizations are reported in Table S3 in the Supporting Information.

MD Simulations. A common use case of GFN-FF would be the simulation of a Ln-containing protein or MOF structure. To illustrate the performance and applicability of our improved model, MD simulations (MDs) were conducted for a hierarchy of systems, ranging from small prototypical Ln ligands DOTA and MACROPA to entire Ln-coordinating proteins. The neighbor list is especially important for MDs since missing covalent bonds lead to undesired flexibility. Therefore, the input structure that the topology is based on should be chosen carefully. The Ln-complexes were preoptimized with GFN2-xTB and the Ac−macropa complex is optimized with TPSSh/SCPP as published by Kovács.[55]

Figure 3 depicts the radial distribution function (RDF) of neighbors to central Ln/An atoms in prototypical complexes. The RDFs are obtained from 50 ps MDs with different GFN-FF versions. For each system, the distances from all neighboring atoms to the central Ln- or An-atoms are evaluated indifferent to the neighbor assignment by the FF.

For example, the original parametrization of the FF (gray) only assigns two neighbors to the europium atom in the sterically demanding NMe_6−Eu(II) complex of Allen and co-workers,[56] leading to a dissociation of the other atoms during the MD. The improved parametrization (yellow) assigns six out of eight neighbors correctly, preventing dissociation, while the hill in the RDF at 4.8 Å corresponds to the unbound neighbors. Finally, assigning all neighbors correctly via user input (blue) ensures a stable structure during the entire MD with distances between 2.4 and 3.0 Å. Note that the two unbound iodide counterions have been omitted for clarity.

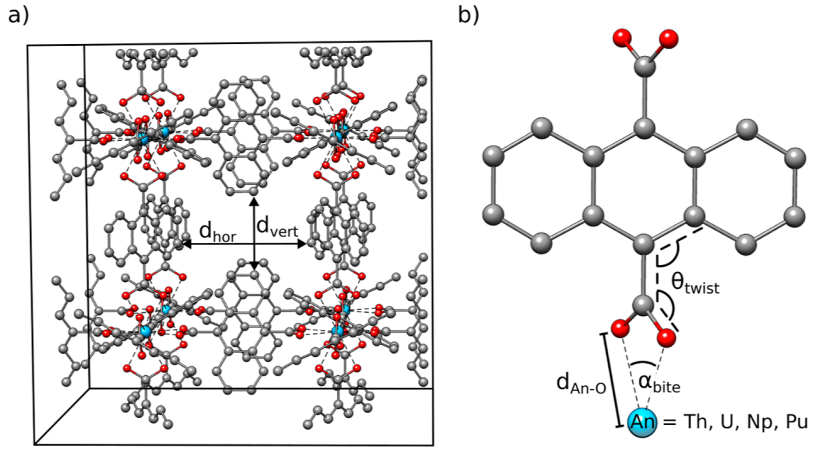


Figure 2. Illustration of the An-ADC crystal. Hydrogen atoms are omitted for clarity. (a) Unit cell with horizontal pore diameter d_{hor} and vertical pore diameter d_{vert}. (b) Illustration of bite angle α_{bite}, twist angle Θ_{twist}, and actinide-oxygen distance d_{An-O}.

https://doi.org/10.1021/acs.inorgchem.4c03215
Inorg. Chem. 2024, 63, 19364−19374

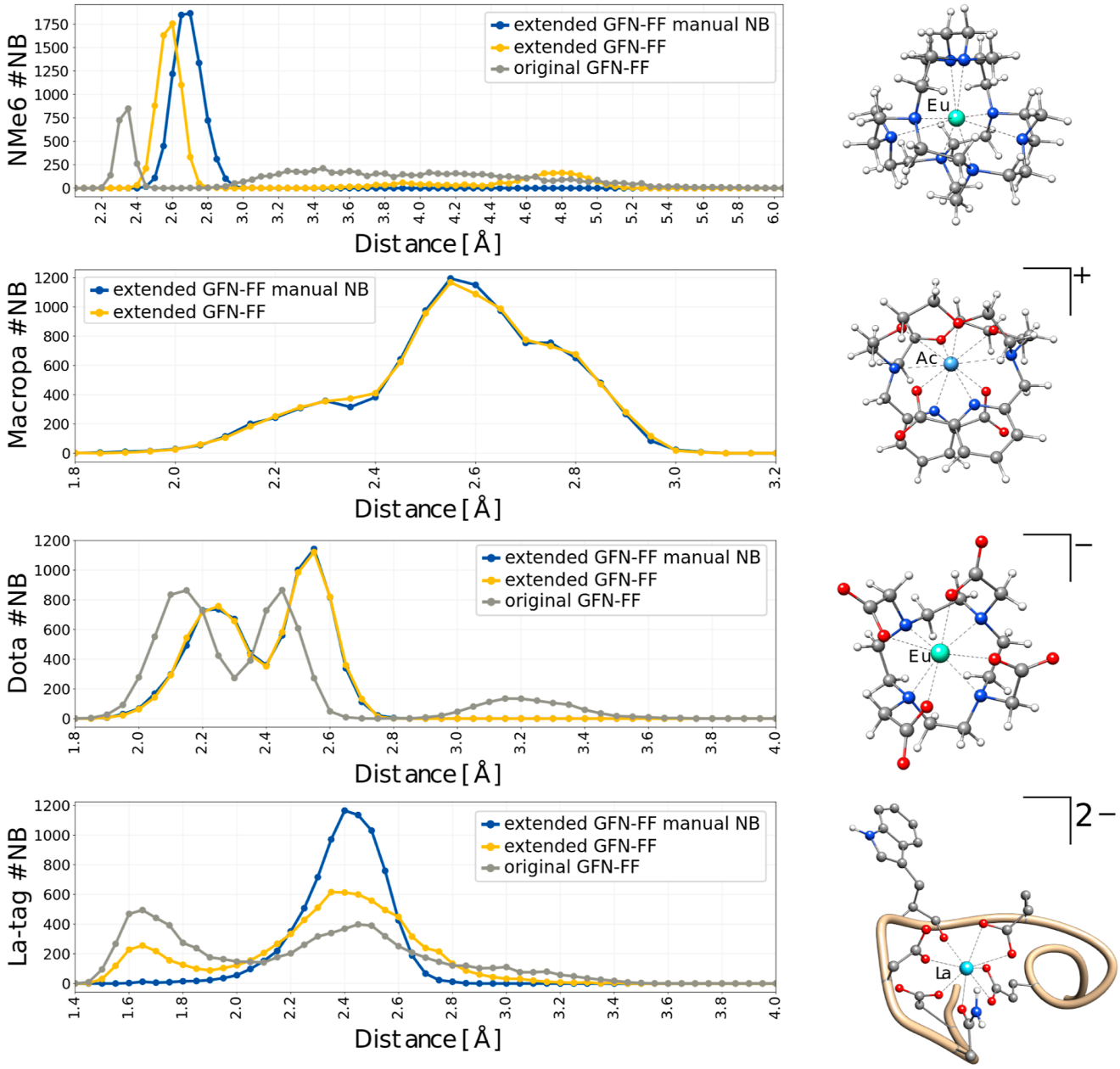


Figure 3. Radial distribution of neighbors to central Ln/An atoms summed up over a 50 ps MD. The systems are depicted on the right side of the plots. In each system, the central atom should be coordinated with eight neighboring atoms.

Since the original parametrization of the GFN-FF does not include actinides, the Ac−macropa complex is only evaluated for the extended version and the extended version with manual neighbor setup. The extended version already assigns all neighbors correctly, giving a similar RDF compared to the FF with a manual neighbor setup.

For the Eu−dota complex, the original parametrization only assigns seven out of eight neighbors correctly leading to a small hill around 3.2 Å. The new parametrization (yellow) and the manual neighbor setup (blue) qualitatively provide the same results with two peaks around 2.2 Å for oxygen neighbors and around 2.6 Å for nitrogen neighbors.

With 256 atoms, the La-tag is by far the largest and most flexible complex simulated. The input structure was generated from RCSB PDB[57] structure with ID 7CCO by adding hydrogen atoms with chimera[58] at physiological pH. Furthermore, solvation effects are included with the implicit solvation model ALPB for water.[59] For this realistic use case, the original and extended version assign the carbon atoms of carboxylate groups as additional neighbors leading to very short distances for some of the oxygen atoms. This effect is less pronounced in the extended version. Again, the correct neighbor assignment leads to a robust MD with neighbor distances between 2.0 and 2.7 Å.

The entire MD calculation with the manual neighbor list setup took 32.6 s for the Eu−NMe_6 complex, 54.3 s for the Ac−macropa complex, 13.9 s for the Eu−dota complex, and 7.1 min for the La-tag. All calculations were performed on four CPUs at 4.2 GHz. Overall, these examples demonstrate that the extended GFN-FF together with the possibility to manually define the neighbor list enables routine investigation of the conformational space of Ln/An systems via MD simulations.

https://doi.org/10.1021/acs.inorgchem.4c03215
Inorg. Chem. 2024, 63, 19364−19374

Especially for systems where the oxidation state has a large impact on the structure, such as NMe_6−Eu(II) and dota−Eu(III), the detailed input proves to be vital for obtaining good results.

Optimization of Lanthanide-Containing Biomolecules. The somatostatin receptor (SSTR) is a target for the treatment of multiple tumors,[60] e.g., with Radiopharmaceuticals such as Lu−dotatate.[61] To showcase an important application of the FF, the Lu−dotatate complex was manually added to the crystal structure of somatostatin receptor 2 (SSTR2) after removing the peptide antagonist. The original protein structure was downloaded from RCSB PDB[57] with Entry ID 7XNA. The 2D structure of Lu−dotatate was downloaded from DrugBank[62] and converted into a 3D structure with GFN-FF and manual adjustments. The optimization was performed on four CPUs with 2.0 GHz. In the start structure, the Lu−dotatate complex had a minimum distance of 3.4 Å to the protein. Figure 4 shows an overlay of

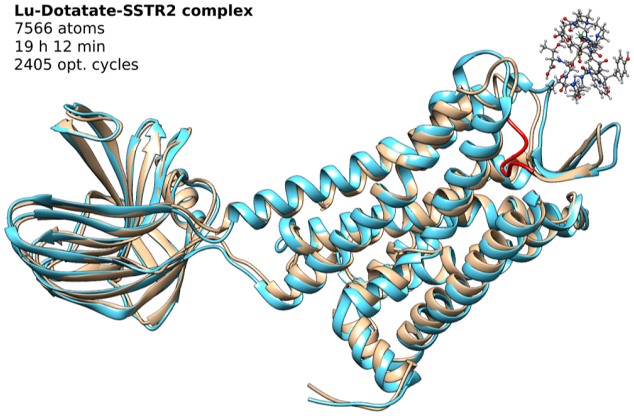


Figure 4. Overlay between the experimental SSTR2 structure (PDB code 7XNA) in blue and red and the optimized structures of the protein (brown) and the Lu−dotatate complex in ball and stick representation. The peptide inhibitor colored in red was removed before the optimization. The FF optimization closes the vacancy and partially encloses the Lu−dotatate molecule in the protein.

the original protein (brown) and removed peptide (red) with the optimized protein (blue) and Lu-complex (colored by atom type). To approximate important solvation effects, the implicit solvation model ALPB was included in the FF optimization with GFN-FF. The optimization closes the cavity left by the peptide and partially encloses the Lu-complex in the protein. On the unaltered side of the structure, the deviations from the experimental structure are reasonable, with the largest deviations for random coils.

A second Ln−biomolecule complex, proposed as a sustainable tool for extracting rare-earth elements, was optimized using the extended FF to validate the accuracy of the optimized structures. An overlay of the optimized structure with the experimental structure (transparent blue) is shown in Figure 5. The integrity of the molecular structure is maintained, which is reflected in the relatively small rmsd of 1.1 Å. The entire calculation was completed in 10 min and 48 s on four CPUs with 4.2 GHz. Of this time, the 541 optimization cycles required 10 min and 27 s, while the setup of the FF parameters took 21 s. For both biomolecules, the Ln-complex remains intact during the optimization and no topological changes occur, making the GFN-FF a vital tool for

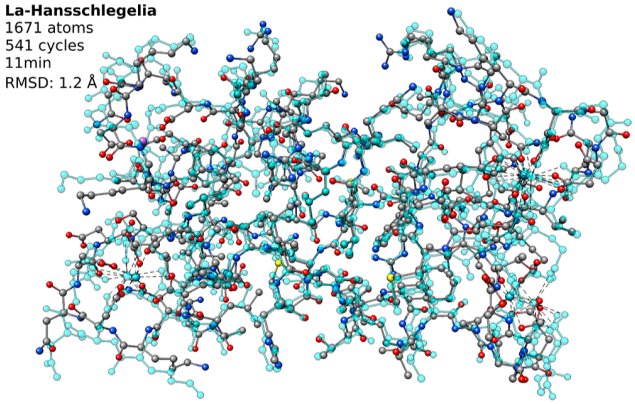


Figure 5. Overlay between the experimental structure of La−Hansschlegelia with La(III) and the structure optimized with the extended GFN-FF and ALPB solvation model. The structure contains three lanthanum atoms and one iodine atom. Hydrogen atoms are omitted for clarity.

elucidating large complex structures including elements up to lawrencium.

Isomerization Energies. Although the primary focus of GFN-FF and the presented extension is on structural properties, it is pertinent to briefly discuss chemical energies. For this purpose, isomerization energies of Ln/An complexes provide an application-oriented test case. To calculate meaningful relative isomer energies, it is crucial to ensure that the GFN-FF neighbor lists of all the isomers are identical. This can be checked with the `--wrtopo nb` keyword, which causes the neighbor list to be printed into a JSON file. If necessary, the neighbor list should be modified through the detailed input file as mentioned previously. Users should be aware that the topology is highly dependent on the input structure, and therefore the FF can only provide reasonable results if the input is reasonable. Figure 6 presents the relative isomer energies calculated with r^2SCAN-3c and GFN-FF. The system names start with the central atom, followed by a unique identifier for the surrounding ligands. The first three systems differ only in the central atom, displaying similar results with r^2SCAN-3c for the same isomers. Here, GFN-FF shows the largest deviations for the square isomer, with up to 10.8 kcal/mol difference for the Lu_ff6372 system. Considering the relative ranking of all the isomers, the FF is only off by 2 kcal/mol or less, which aligns closely with the reference value's accuracy. For the other systems with only two isomers, GFN-FF predicts the relative energies well, with absolute deviations between 0.2 and 3.8 kcal/mol. Only the Nd_c5f44a system is qualitatively wrong. Across the studied systems, except for Eu_ff6372, the FF predicts consistently too low relative isomer energies. However, GFN-FF shows a very reasonable error range and good robustness, with only two deviations larger than 6 kcal/mol. Further testing of the method for screening of isomers seems promising but requires a detailed study of the threshold for sorting out unfavorable isomers.

It should be noted, that even DFT calculations can give larger deviations for such systems. The accuracy of the reference values is discussed in the Supporting Information alongside the calculated energies given in Table S4.

Timings and Limitations for Periodic Systems. To evaluate the timings and limitations of the FF, optimizations are performed on crystal structures including Eu, Lu, Ac, or Th as well as alkali metals, alkaline earth metals, transition metals,

https://doi.org/10.1021/acs.inorgchem.4c03215
Inorg. Chem. 2024, 63, 19364−19374

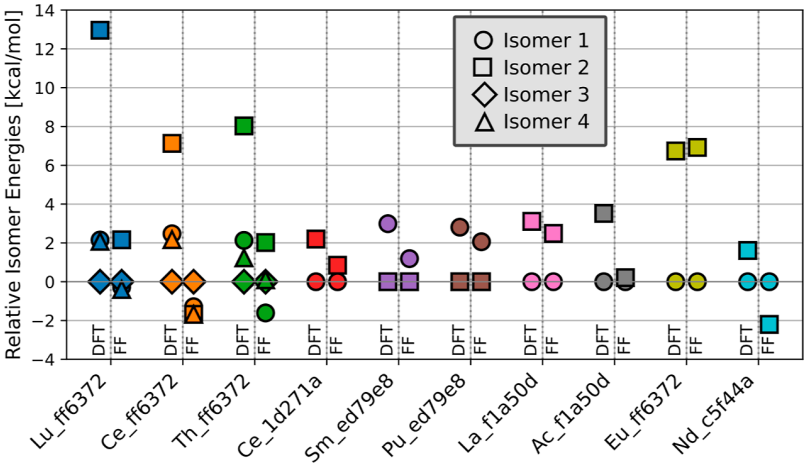


Figure 6. Relative isomer energies of Ln/An complexes concerning the lowest lying isomer according to the r^2SCAN-3c reference energies. GFN-FF results are depicted next to the DFT reference. For each system, corresponding isomers are identified by the shape of the marker. All systems are optimized with the given method.

post-transition metals, halogens, and metalloids. A total of 29 different atom types are included in the 53 optimized systems. The specific elements and the Crystallography Open Database (COD)[63−65] IDs are listed in the Supporting Information alongside further optimization details in Tables S5 and S6. Timings plotted against the initial unit cell volumes are given in Figure 7 together with the deviation of the optimized cell

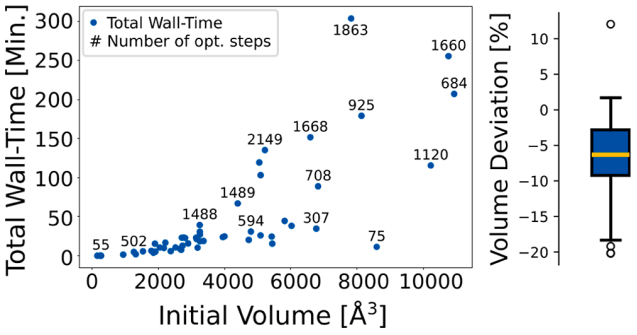


Figure 7. Total wall-time in minutes for the entire optimization of periodic structures including Eu, Lu, Ac, or Th plotted against the initial volume of the unit cell. All calculations were performed on one CPU at 4.2 GHz. The depicted boxplot shows the statistics for the deviation of the unit cell volume before and after the optimization. The box depicts the interquartile range and the median, while the fliers show the minimum and maximum values excluding the outliers which are given as circles.

volume from the experimental cell volume. The number of optimization cycles is given next to the data points for selected systems. The fastest optimization with 17 cycles in 0.17 s is for a cubic closest packed actinium crystal with four atoms in the unit cell (ID:9008458). The longest optimization with 1863 cycles in 304.0 min is for a system with 889 atoms in the unit cell, including tetraphenylborate anions and a Th-complex (ID:1548031). Overall, the FF optimizations of the periodic systems show quadratic scaling with a prefactor depending on how quickly the optimization converges.

During the optimizations, the following limitations of the FF were encountered. Optimizations on systems including silver−iodine or copper−iodine clusters, similar to the system with ID 7704645, became unstable and led to a disintegration of the crystal structure. Furthermore, when two or three cyclopentadienyl derivatives (Cp′) coordinate to a central atom, as in the system with ID 7701390, bond breaking in the Cp-ligand occurs in some cases. If calculations on strongly ionic crystals fail, we recommend testing the pGFN-FF by Gale et al.[66] which now also provides a library for generating parameters via GitHub. Further timings on two systems including francium or radium are included in the Supporting Information.

CONCLUSION

Addressing a lack of fast methods specifically parametrized for lanthanides and actinides, we presented an extension of the GFN-FF to include parameters for these 4f and 5f elements. The extended version comprises a reparameterized EEQ model, optimized neighbor assignment, improved bond length parameters for halogens, and a charge-based neighbor threshold for hydrogen atoms. These modifications are selectively applied only if the atom or its neighboring atom is a Ln/An, ensuring that the performance of systems without these elements is not affected. Parameters for francium and radium are included for consistency, and a proof of principle was demonstrated for these elements. Additionally, the extended FF features a new manual setup for the neighbor list, facilitating robust calculations for otherwise challenging systems.

We validated our extension through MD simulations and geometry optimizations on a diverse set of systems varying in size and composition. Our results show that Ln-ligand bond lengths are accurately described with an MAD of 0.14 Å compared to the LnQM (DFT) geometries. Other structural parameters, such as bite angles or pore sizes in MOFs, exhibit errors below 7% when compared to experimental values. MD simulations on four pivotal ligands used in radiochemistry further underline the robustness of our model. With an MAD of 3.4 kcal/mol for relative isomer energies compared to r^2SCAN-3c, the screening for isomers is an additional application of the method. Another interesting application is the calculation of binding free energies for anion sensors.[67] Timings on highly diverse systems varying in composition and system size show quadratic scaling of the computational costs with the system size. Throughout this study, we tested highly

https://doi.org/10.1021/acs.inorgchem.4c03215
Inorg. Chem. 2024, 63, 19364−19374

charged species (+2 or +3 system charges), including the La-tag, Ce_1d271a, Nd_c5f44a, and various systems in the LnQM. No specific issues were observed with these systems.

These findings demonstrate that the FF can be used effectively to model large systems and to explore the conformational space in Ln/An chemistry. This advancement opens new possibilities for computational studies in radioactive waste treatment, radiopharmaceuticals, and photochemistry, which are challenging yet crucial areas of research.

■ ASSOCIATED CONTENT

Data Availability Statement

Both fit sets, the input geometries for the MDs, the optimized biomolecular structures, and the isomer structures can be found in a GitHub repository at https://github.com/grimme-lab/GFN-FF_Ln_An.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.inorgchem.4c03215.

Tabulated values for discussed results, proof of principle for francium and radium (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Stefan Grimme − *Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Bonn 53115, Germany;* orcid.org/0000-0002-5844-4371; Email: grimme@thch.uni-bonn.de

Authors

Thomas Rose − *Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Bonn 53115, Germany*

Markus Bursch − *FACCTs GmbH, Köln 50677, Germany*

Jan-Michael Mewes − *beeOLED GmbH, Dresden 01257, Germany;* orcid.org/0000-0002-4669-8091

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.inorgchem.4c03215

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge the use of the MARVIN cluster at the University of Bonn for providing computational resources that have contributed to the research results reported within this paper. We thank the HRZ HPC Team and the HPC/A Lab for their support and maintenance of the MARVIN cluster. The authors thank Christian Hölzer and Lukas Wittmann for helpful discussions regarding the LnQM and AcQM data sets, respectively. We acknowledge the use of ChatGPT by OpenAI for assisting in drafting and revising the language and structure of the manuscript.

■ REFERENCES

(1) Lange, R.; Schreuder, N.; Hendrikse, H. In *Practical Pharmaceutics*; Le Brun, P., Crauste-Manciet, S., Krämer, I., Smith, J., Woerdenbag, H., Eds.; Springer International Publishing, 2023; pp 531−550.

(2) Vallabhajosula, S. *Molecular Imaging and Targeted Therapy: Radiopharmaceuticals and Clinical Applications*; Springer International Publishing, 2023.

(3) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36: An improved force field for folded and intrinsically disordered proteins. *Biophys. J.* **2017**, *112*, 175−176.

(4) Tian, C.; Kasavajhala, K.; Belfon, K. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **2020**, *16*, 528−552.

(5) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved peptide and protein torsional energetics with the OPLS-AA force field. *J. Chem. Theory Comput.* **2015**, *11*, 3499−3509.

(6) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; Van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843−856.

(7) De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **2013**, *9*, 687−697.

(8) Kawamoto, S.; Liu, H.; Miyazaki, Y.; Seo, S.; Dixit, M.; DeVane, R.; MacDermaid, C.; Fiorin, G.; Klein, M. L.; Shinoda, W. SPICA force field for proteins and peptides. *J. Chem. Theory Comput.* **2022**, *18*, 3204−3217.

(9) Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagieła, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A. others A unified coarse-grained model of biological macromolecules based on mean-field multipole−multipole interactions. *J. Mol. Model.* **2014**, *20*, 1−15.

(10) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 force field: altius, fortius, citius. *J. Chem. Theory Comput.* **2019**, *15*, 2719−2733.

(11) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(12) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.

(13) Marjolin, A.; Gourlaouen, C.; Clavaguéra, C.; Ren, P. Y.; Wu, J. C.; Gresh, N.; Dognon, J.-P.; Piquemal, J.-P. Toward accurate solvation dynamics of lanthanides and actinides in water using polarizable force fields: from gas-phase energetics to hydration free energies. *Theor. Chem. Acc.* **2012**, *131*, 1198.

(14) Duvail, M.; Souaille, M.; Spezia, R.; Cartailler, T.; Vitorge, P. Pair interaction potentials with explicit polarization for molecular dynamics simulations of La3+ in bulk water. *J. Chem. Phys.* **2007**, *127*, 034503.

(15) Duvail, M.; Vitorge, P.; Spezia, R. Building a polarizable pair interaction potential for lanthanoids(III) in liquid water: A molecular dynamics study of structure and dynamics of the whole series. *J. Chem. Phys.* **2009**, *130*, 104501.

(16) Duvail, M.; Martelli, F.; Vitorge, P.; Spezia, R. Polarizable interaction potential for molecular dynamics simulations of actinoids-(III) in liquid water. *J. Chem. Phys.* **2011**, *135*, 044503.

(17) Acher, E.; Masella, M.; Vallet, V.; Réal, F. Properties of the tetravalent actinide series in aqueous phase from a microscopic simulation self-consistent engine. *Phys. Chem. Chem. Phys.* **2020**, *22*, 2343−2350.

(18) Moreno Martinez, D.; Guillaumont, D.; Guilbaud, P. Force Field Parameterization of Actinyl Molecular Cations Using the 12−6-4 Model. *J. Chem. Inf. Model.* **2022**, *62*, 2432−2445.

(19) Duvail, M.; Moreno Martinez, D.; Žiberna, L.; Guillam, E.; Dufrêche, J. F.; Guilbaud, P. Modeling Lanthanide Ions in Solution: A Versatile Force Field in Aqueous and Organic Solvents. *J. Chem. Theory Comput.* **2024**, *20*, 1282−1292.

(20) Bursch, M.; Hansen, A.; Grimme, S. Fast and Reasonable Geometry Optimization of Lanthanoid Complexes with an Extended Tight Binding Quantum Chemical Method. *Inorg. Chem.* **2017**, *56*, 12485−12491.
(21) Liu, C.; Aguirre, N. F.; Cawkwell, M. J.; Batista, E. R.; Yang, P. Efficient Parameterization of Density Functional Tight-Binding for 5f-Elements: A Th-O Case Study. *J. Chem. Theory Comput.* **2024**, *20*, 5923−5936.
(22) Carlson, R. K.; Cawkwell, M. J.; Batista, E. R.; Yang, P. Tight-Binding Modeling of Uranium in an Aqueous Environment. *J. Chem. Theory Comput.* **2020**, *16*, 3073−3083.
(23) Wang, L.; Fang, P.; Zhao, Z.; Huang, Y.; Liu, Z.; Bian, Z. Rare Earth Complexes with 5d−4f Transition: New Emitters in Organic Light-Emitting Diodes. *J. Phys. Chem. Lett.* **2022**, *13*, 2686−2694.
(24) Patra, K.; Ansari, S. A.; Mohapatra, P. K. Metal-organic frameworks as superior porous adsorbents for radionuclide sequestration: Current status and perspectives. *J. Chromatogr. A* **2021**, *1655*, 462491.
(25) Niu, K.; Yang, F.; Gaudin, T.; Ma, H.; Fang, W. Theoretical Study of Effects of Solvents, Ligands, and Anions on Separation of Trivalent Lanthanides and Actinides. *Inorg. Chem.* **2021**, *60*, 9552−9562.
(26) Dognon, J.-P. Theoretical insights into the chemical bonding in actinide complexes. *Coord. Chem. Rev.* **2014**, *266*−*267*, 110−122.
(27) Das, A.; Das, U.; Das, A. K. Relativistic effects on the chemical bonding properties of the heavier elements and their compounds. *Coord. Chem. Rev.* **2023**, *479*, 215000.
(28) Spicher, S.; Grimme, S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angew. Chem., Int. Ed.* **2020**, *59*, 15665−15673.
(29) Grimme, S.; Rose, T. mcGFN-FF: an accurate force field for optimization and energetic screening of molecular crystals. *Z. Naturforsch., B: J. Chem. Sci.* **2024**, *79*, 191−200.
(30) Plett, C.; Grimme, S.; Hansen, A. Conformational energies of biomolecules in solution: Extending the MPCONF196 benchmark with explicit water molecules. *J. Comput. Chem.* **2024**, *45*, 419−429.
(31) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.; Spicher, S.; Steinbach, P.; Wesołowski, P. A.; Zeller, F. CREST-A Program for the Exploration of Low-Energy Molecular Chemical Space. *J. Chem. Phys.* **2024**, *160*, 114110.
(32) Chen, Y.-q.; Sheng, Y.-j.; Ma, Y.-q.; Ding, H.-m. Efficient Calculation of Protein−Ligand Binding Free Energy Using GFN Methods: The Power of the Cluster Model. *Phys. Chem. Chem. Phys.* **2022**, *24*, 14339−14347.
(33) Gorges, J.; Bädorf, B.; Grimme, S.; Hansen, A. Efficient Computation of the Interaction Energies of Very Large Non-covalently Bound Complexes. *Synlett* **2023**, *34*, 1135−1146.
(34) Mortier, W. J.; Ghosh, S. K.; Shankar, S. Electronegativity-Equalization Method for the Calculation of Atomic Charges in Molecules. *J. Am. Chem. Soc.* **1986**, *108*, 4315−4320.
(35) Chandrasekar, A.; Ghanty, T. K. Uncovering Heavy Actinide Covalency: Implications for Minor Actinide Partitioning. *Inorg. Chem.* **2019**, *58*, 3744−3753.
(36) Varathan, E.; Gao, Y.; Schreckenbach, G. Computational Study of Actinyl Ion Complexation with Dipyriamethyrin Macrocyclic Ligands. *J. Phys. Chem. A* **2021**, *125*, 920−932.
(37) Johnson, S. G. The NLopt nonlinear-optimization package, version 2.7.1. https://github.com/stevengj/nlopt, 2007 (accessed Jan 25, 2024).
(38) Powell, M. J. The BOBYQA algorithm for bound constrained optimization without derivatives. In *Cambridge NA Report NA2009/06*; University of Cambridge: Cambridge, 2009; Vol. *26*, pp 26−46.
(39) Wittmann, L.; Gordiy, I.; Friede, M.; Helmich-Paris, B.; Grimme, S.; Hansen, A.; Bursch, M. Extension of the D3 and D4 London Dispersion Corrections to the full Actinides Series. *Phys. Chem. Chem. Phys.* **2024**, *26*, 21379−21394.
(40) Mardirossian, N.; Head-Gordon, M. ωB97M-V. A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.
(41) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(42) Dolg, M.; Stoll, H.; Preuss, H. Energy-adjusted *ab initio* pseudopotentials for the rare earth elements. *J. Chem. Phys.* **1989**, *90*, 1730−1734.
(43) Zheng, J.; Xu, X.; Truhlar, D. G. Minimally augmented Karlsruhe basis sets. *Theor. Chem. Acc.* **2011**, *128*, 295−305.
(44) Rose, T. *GFN-FF for Lanthanides and Actinides*; GitHub Repository, 2024. https://github.com/grimme-lab/GFN-FF_Ln_An.
(45) Hölzer, C.; Gordiy, I.; Grimme, S.; Bursch, M. Hybrid DFT Geometries and Properties for 17k Lanthanoid Complexes - The LnQM Data Set. *J. Chem. Inf. Model.* **2024**, *64*, 825−836.
(46) Allred, A. Electronegativity values from thermochemical data. *J. Inorg. Nucl. Chem.* **1961**, *17*, 215−221.
(47) Pauling, L. The nature of the chemical bond. IV. The energy of single bonds and the relative electronegativity of atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570−3582.
(48) Haynes, W. M. *CRC Handbook of Chemistry and Physics*, 95th ed.; CRC Press, OCLC, 2014; p 908078665.
(49) Ghosh, D. C.; Islam, N. Semiempirical evaluation of the global hardness of the atoms of 103 elements of the periodic table using the most probable radii as their size descriptors. *Int. J. Quantum Chem.* **2010**, *110*, 1206−1213.
(50) Jabł oński, M. a. Charge-inverted hydrogen bond vs. other interactions possessing a hydridic hydrogen atom. *Chem. Phys.* **2014**, *433*, 76−84.
(51) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r^2SCAN-3c: A "Swiss army knife" composite electronic-structure method. *J. Chem. Phys.* **2021**, *154*, 064103.
(52) Mardirossian, N.; Head-Gordon, M. ωB97X-V. A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904−9924.
(53) Serezhkin, V. N.; Savchenkov, A. V.; Serezhkina, L. B. Features of actinide contraction in AnX crystals (X = S, Se or Te). *Solid State Sci.* **2021**, *121*, 106734.
(54) Lv, K.; Urbank, C.; Patzschke, M.; März, J.; Kaden, P.; Weiss, S.; Schmidt, M. MOFs with 12-Coordinate 5f-Block Metal Centers. *J. Am. Chem. Soc.* **2022**, *144*, 2879−2884.
(55) Kovács, A. Theoretical Study of Actinide Complexes with Macropa. *ACS Omega* **2020**, *5*, 26431−26440.
(56) Jenks, T. C.; Bailey, M. D.; Corbin, B. A.; Kuda-Wedagedara, A. N.; Martin, P. D.; Schlegel, H. B.; Rabuffetti, F. A.; Allen, M. J. Photophysical characterization of a highly luminescent divalent-europium-containing azacryptate. *Chem. Commun.* **2018**, *54*, 4545−4548.
(57) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(58) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(59) Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. *J. Chem. Theory Comput.* **2021**, *17*, 4250−4261.
(60) Zhao, W.; Han, S.; Qiu, N.; Feng, W.; Lu, M.; Zhang, W.; Wang, M.; Zhou, Q.; Chen, S.; Xu, W.; et al. Structural insights into ligand recognition and selectivity of somatostatin receptors. *Cell Res.* **2022**, *32*, 761−772.
(61) Kurz, S. C.; Zan, E.; Cordova, C.; Troxel, A. B.; Barbaro, M.; Silverman, J. S.; Snuderl, M.; Zagzag, D.; Kondziolka, D.; Golfinos, J. G.; Chi, A. S.; Sulman, E. P. Evaluation of the SSTR2-targeted Radiopharmaceutical ^{177}Lu-DOTATATE and SSTR2-specific ^{68}Ga-

https://doi.org/10.1021/acs.inorgchem.4c03215
Inorg. Chem. 2024, 63, 19364−19374

DOTATATE PET as Imaging Biomarker in Patients with Intracranial Meningioma. *Clin. Cancer Res.* **2024**, *30* (4), 680−686.
(62) Wishart, D. S. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, 668−672.
(63) Vaitkus, A.; Merkys, A.; Sander, T.; Quirós, M.; Thiessen, P. A.; Bolton, E. E.; Gražulis, S. A workflow for deriving chemical entities from crystallographic data and its application to the Crystallography Open Database. *J. Cheminf.* **2023**, *15*, 123.
(64) Downs, R. T.; Hall-Wallace, M. The American Mineralogist Crystal Structure Database. *Am. Mineral.* **2003**, *88*, 247−250.
(65) Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database—an open-access collection of crystal structures. *J. Appl. Crystallogr.* **2009**, *42*, 726−729.
(66) Gale, J. D.; LeBlanc, L. M.; Spackman, P. R.; Silvestri, A.; Raiteri, P. A Universal Force Field for Materials, Periodic GFN-FF: Implementation and Examination. *J. Chem. Theory Comput.* **2021**, *17*, 7827−7849.
(67) Bodman, S. E.; Breen, C.; Kirkland, S.; Wheeler, S.; Robertson, E.; Plasser, F.; Butler, S. J. Sterically demanding macrocyclic Eu(III) complexes for selective recognition of phosphate and real-time monitoring of enzymatically generated adenosine monophosphate. *Chem. Sci.* **2022**, *13*, 3386−3394.

https://doi.org/10.1021/acs.inorgchem.4c03215
Inorg. Chem. 2024, 63, 19364−19374

APPENDIX D

Charge Transfer and Recombination Pathways through Fullerene Guests in Porphyrin-based MOFs

Alison Arissa[1], Thomas Rose[2], Noémi Leick[3], Stefan Grimme[2], Justin C. Johnson[2], Jenny V. Lockard[1]

Received: January 8, 2025
Published online:

Reprinted in Appendix D with permission [4] from Alison Arissa, Thomas Rose, Noémi Leick, Stefan Grimme, Justin C. Johnson, Jenny V. Lockard, *Charge Transfer and Recombination Pathways through Fullerene Guests in Porphyrin-based MOFs*. J. Phys. Chem. C **129** (2025) 8215 –
DOI: https://doi.org/10.1021/acs.jpcc.5c00161

Own contributions

- Performing all calculations for the computational investigation
- Interpretation of the results
- Writing the computational part of the manuscript

[1] Department of Chemistry, Rutgers University-Newark, Newark, New Jersey 07102, USA
[2] Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Beringstraße 4, Bonn 53115, Germany
[3] National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, Colorado 80401, USA
[4] Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

THE JOURNAL OF PHYSICAL CHEMISTRY C
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCC

Article

Charge Transfer and Recombination Pathways through Fullerene Guests in Porphyrin-Based MOFs

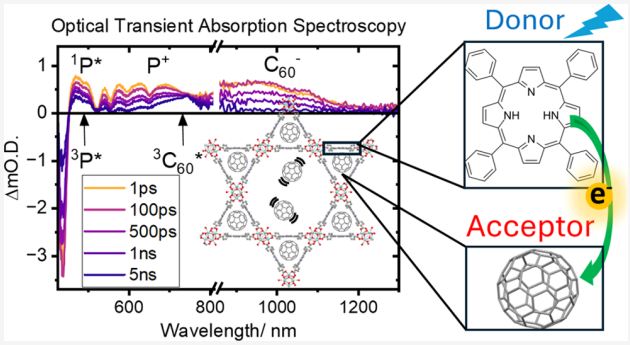
Alison Arissa, Thomas Rose, Noémi Leick, Stefan Grimme, Justin C. Johnson,* and Jenny V. Lockard*

Cite This: https://doi.org/10.1021/acs.jpcc.5c00161

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Porphyrin-based metal−organic frameworks (MOFs) offer a unique platform for building porous donor−acceptor networks that exhibit long-lived charge separation and transport upon incorporation of electron acceptor guest species. Here, porphyrin-based MOFs, PCN-222(H_2) and PCN-222(Zn), synthesized as nanoparticle suspensions, are successfully infiltrated with fullerene acceptor molecules, C_{60} or $PC_{61}BM$, in both polar and nonpolar solvent environments. The location and relative binding strength of these guest species are evaluated through a combination of N_2 physisorption measurements, photoluminescence quenching, and UV−vis absorption titration experiments. Semiempirical tight binding calculations are used to screen potential locations of the fullerene guest within the MOF pores, and hybrid density functional theory (DFT)-computed interaction energies confirm the energetically favorable positions. The fundamental photophysics of these donor−acceptor host−guest combinations are probed using ultrafast transient absorption spectroscopy. Sub-picosecond electron transfer involving initial exciplex population is observed, with slow charge recombination lifetimes on the order of $\tau \sim 1$ ns for all systems in both dimethylformamide and 1,4-dioxane. Charge recombination occurs through population of fullerene and/or framework porphyrin triplet states depending on the porphyrin metalation status. The photophysics of the fullerene-loaded MOFs are discussed in the context of relevant porphyrin−fullerene donor−acceptor molecules to highlight the unique role of the framework environment in dictating photoinduced electron transfer and decay pathways.



INTRODUCTION

Metal−organic frameworks (MOFs) present both technological promise in a wide array of applications and a fruitful platform for fundamental studies of host−guest chemistry, electron transfer, and photophysics in three-dimensional (3D) coordination space.[1−6] Composed of metal ions or clusters connected through coordination bonds with organic or organometallic linkers, MOFs self-assemble under controlled solvothermal conditions to yield a variety of microporous environments and precise chemical tunability not found in other solid-state hybrid materials.[7,8] The retention of crystallinity upon solvent removal or exchange renders the porous frameworks accessible to other potential guest species during subsequent postsynthetic modification (PSM) treatment. This provides another handle to further alter the framework composition and electronic structure. The resulting diversity in chemical makeup and tunable, permanent porosity make MOFs attractive candidates for potential adsorption-based applications. While initial targets focused on gas separation and storage,[9−12] and catalysis,[13−15] more recent efforts to impart and study electron transfer and transport properties in these materials pave the way for applications that rely on MOF conductivity and/or long-lived charge separation such as resistive sensors, electrochromic devices, and electro- or photocatalysis.[16−26] In one approach, introducing redox-active guest species via PSM can yield MOF host−guest donor−acceptor (D−A) systems with the desired charge transfer (CT) and transport properties. Leveraging established molecular D−A combinations, focusing solely here on archetypal electron-rich porphyrins and electron-deficient fullerenes, organized D−A arrangements within the porous structures can be achieved through host−guest interaction. Supramolecular porphyrin−fullerene D−A complexes, which have been studied extensively,[27−30] typically rely on self-assembly methods to build extended arrays to reach longer length scales needed for device fabrication.[31,32] In MOF-based systems, the introduction of fullerene acceptor guest molecules via postsynthetic modification into preformed MOF structures containing porphyrin donor linkers can promote these D−A interactions upon confinement within the porous structure

Received: January 8, 2025
Revised: April 11, 2025
Accepted: April 14, 2025

while avoiding additional self-assembly steps. The ideal framework for most applications, and in particular photo-(electro)catalysis, would promote selective infiltration of these fullerene guests to foster the D−A host−guest interactions yet possess a pore structure that remains accessible to other guest molecules. PCN-222 is a MOF structure predicted to exhibit such behavior.

The PCN-222 framework contains nodes of Zr_6 clusters connected by tetracarboxyphenylporphyrin linkers to form a 3D architecture with two types of one-dimensional (1D) channel pores that are considerably different in size and shape.[33] The smaller of the two channels is ideally suited for housing electron-deficient molecules such as C_{60} and other fullerene derivatives. Preferential confinement within these triangular pores would likely facilitate their D−A interactions with the host framework linkers, triggering directional charge transport. The much larger hexagonal porous channels, however, would not likely have this confinement effect on fullerene, leaving them effectively available for other guest species. The isostructural framework, NU-1000 (with tetracarboxyphenyl pyrene in place of the porphyrin linkers)[34] provides precedent for this host−guest arrangement and D−A behavior. Introducing metal(IV) bis(dicarbollide) complexes as electron-deficient guests to this framework leads to preferential confinement within the smaller channels, and the composite host−guest material shows increased electrical conductivity.[35] Furthermore, photoconductive behavior is predicted for this guest molecule as well as C_{60} incorporated in metal-free or zinc porphyrin versions of PCN-222.[36] The photophysics and electrochemistry of similar porphyrin units, along with the precedent for noncovalent D−A interaction with electron acceptor molecules in cage complexes and other frameworks,[29,30,37−39] suggests their likely participation as electron donors upon visible light irradiation in these PCN-222 host−guest systems. An additional advantage of MOFs with multiple pore types is the unique chemical environments that they might exhibit due to steric effects, mass transport limitations, or pore surface chemistry that is distinct in different pores.

In this investigation, we examine a series of PCN-222 frameworks loaded with fullerene guest molecules to experimentally confirm the predicted confinement within the porous structure and accompanying D−A interaction with the porphyrin linkers. While charge transport is expected to be an important property of these materials for some applications, this study focuses on understanding the initial photoinduced CT process and associated decay pathway of these systems. C_{60} or fullerene derivative, [6,6]phenyl-C_{61}-butyric acid methyl ester ($PC_{61}BM$), guest species are incorporated in either the metal-free or zinc porphyrin version of this MOF, PCN-222(H_2,Zn), to probe the steric effects of guest functionalization and influence of porphyrin metalation status on the host−guest cofacial binding strength and resulting D−A excited state decay pathways. Both *N*′,*N*-dimethylformamide (DMF) and 1,4-dioxane solvent environments are employed to determine the influence of the solvent polarity on these excited state dynamics. Steady state electronic spectroscopy combined with N_2 gas physisorption measurements are used to establish the location and relative binding strength of the fullerene guests within the frameworks, while ultrafast optical transient absorption (TA) spectroscopy studies reveal their excited state landscapes and associated dynamics. Optical spectroscopic measurements of MOF materials, in general, can be challenging because of light scattering interference. Here, we use a modified MOF synthesis method that generates extremely stable nanoparticle suspensions of PCN-222 to allow for in-depth steady state and time-resolved optical spectroscopy characterization. The photophysics of the porphyrin and fullerene components in other D−A contexts are well established,[40−44] and therefore provide a useful benchmark to help interpret the TA results of the MOF host−guest systems and the observed trends. By tracking the spectral signatures of the charge transfer and triplet excited states following porphyrin photoexcitation, we observe subtle differences in both the rate of electron transfer and the recombination pathways depending on the fullerene guest/porphyrin linker combination. Furthermore, global fitting and target analysis of these TA data reveal that exciplex formation likely precedes the CT state population along the decay path. These results are discussed in the context of related D−A molecular systems and their established photoinduced electron transfer pathways. The comparisons highlight how the pore structure of the MOF uniquely promotes confinement of the acceptor species along the rigid framework channels, which in turn influences the charge transfer and recombination behavior toward specific triplet state populations.

EXPERIMENTAL METHODS

Materials. *meso*-Tetracarboxyphenylporphyrin (TCPP), tetraphenylporphyrin (TPP), and zinc tetraphenylporphyrin (ZnTPP) were synthesized using literature precedent.[45−47] Zirconyl chloride octahydrate (99% purity) was purchased from Sigma-Aldrich. *N*′,*N*-Dimethylformamide (DMF), toluene, and 1,2-orthodichlorobenzene (ODCB) were purchased from Millipore Sigma. Dichloroacetic acid (DCA) (97%) was purchased from TCI Chemicals. Buckminsterfullerene (C_{60}) (99.9% purity) was obtained from Thermo Fisher Scientific, and [6,6]-phenyl-C_{61}-butyric acid methyl ester ($PC_{61}BM$) was obtained from Nano-C.

Synthesis. PCN-222 nanoparticles were synthesized following literature precedent[48] with some procedure modifications. 24 mg of TCPP was added to a 100 mL pressure vessel containing 60 mL of DMF and ultrasonicated until dissolved. 180 mg of $ZrOCl_2{\cdot}8H_2O$ and 1 mL of DCA were subsequently added, and the reaction flask was ultrasonicated for another 10 min. Finally, the reaction flask was placed in a 135 °C oven for 24 h. After cooling to room temperature, solutions were centrifuged in 40 mL of DMF at 10,000 rpm for 10 min. Subsequently, the supernatant was decanted, 40 mL of fresh DMF was added, and the nanoparticles were resuspended through ultrasonication for 30 min. This process was repeated two times. To ensure the removal of $ZrOCl_2{\cdot}8H_2O$ and TCPP starting material, the nanoparticles were resuspended in 20 mL of a 1 M HCl/DMF mixture using ultrasonication and left in a 120 °C oven for 18 h. Next, 20 mL of DMF was added to the suspension, which was then centrifuged for 10 min at 10,000 rpm. Finally, the supernatant was decanted and 40 mL of DMF was added to resuspend the nanoparticles. PCN-222(Zn) nanoparticles were synthesized by postmetalation of PCN-222(H_2) using a previously reported procedure[49] with slight modification. 100 mg of $ZnCl_2$ was added to 40 mg of PCN-222(H_2) suspended in 20 mL DMF. The mixture was heated to 90 °C for 24 h. After cooling to room temperature, the PCN-222(Zn) suspension was centrifuged at 10,000 rpm for 10 min, decanted, and resuspended in 20 mL 0.5 M HCl/DMF. Subsequently, the 0.5 M HCl/DMF suspension was

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

centrifuged at 10,000 rpm for 10 min, decanted, and resuspended in 40 mL fresh DMF. After the centrifugation step in fresh DMF was repeated two more times, the PCN-222(Zn) nanoparticles were resuspended a final time in DMF. To generate these MOF suspensions in 1,4-dioxane, solvent exchange was performed by centrifuging the DMF nanoparticle suspension at 10,000 rpm for 10 min followed by decantation and resuspension of the solid in 40 mL of 1,4-dioxane. This process was repeated five times.

C_{60}⊂PCN-222(H_2,Zn) and $PC_{61}BM$⊂PCN-222(H_2,Zn). For powder X-ray diffraction and N_2 physisorption measurements, 40 mg of washed PCN-222(H_2) or PCN-222(Zn) nanoparticles suspended in 10 mL of DMF were combined with 100 mg of C_{60} or $PC_{61}BM$ dissolved in 10 mL of ODCB and allowed to soak for 7 days. This 8:1 fullerene/porphyrin ratio PCN-222(H_2,Zn) nanoparticle suspension was then centrifuged at 10,000 rpm for 10 min. After the supernatant was decanted, the nanoparticles were subsequently washed and centrifuged in 40 mL of toluene two times to achieve solvent exchange and remove unbound C_{60}. Following the final decantation step, the sample was allowed to air-dry overnight. For transient absorption spectroscopy, PCN-222(H_2) or PCN-222(Zn) nanoparticle suspensions in ∼10 mL of DMF (or 1,4-dioxane) were concentration adjusted to achieve an optical density of ∼0.2 at 400 nm absorption wavelength (resulting porphyrin concentration is $\sim 6.7 \times 10^{-6}$ M). Subsequently, a 200 μL aliquot of C_{60} in toluene was added to achieve a final molar ratio in a solution of 8:1 C_{60} guest to porphyrin linker. The samples were allowed to soak for 7 days.

Characterization. The zinc content in PCN-222(Zn) was evaluated using atomic absorption spectroscopy (AAS) (Thermo Scientific iCE 3500 with a Zn hollow cathode lamp). 10 mg of PCN-222(Zn) was activated in a vacuum oven at 130 °C for 24 h. Then, 7 mg was added to 10 mL of saturated NaOH and allowed to digest overnight prior to the AAS measurement. PCN-222(H_2,Zn) nanoparticles were characterized by scanning electron microscopy (SEM) and dynamic light scattering (DLS). SEM images were captured on a Hitachi S-4800 field emission scanning electron microscope using a secondary electron detector with an accelerating voltage of 15 kV and a probe current of 10 mA. SEM samples were prepared on a fixed aluminum stub, carbon conductive tape, and 1 mm of iridium sputter treatment. DLS measurements of the PCN-222(H_2,Zn) nanoparticle suspensions in DMF were obtained using a Malvern Zetasizer Nano-ZS instrument equipped with a 4 mW, 633 nm He−Ne laser and an Avalanche photodiode detector at an angle of 173°. Powder X-ray diffraction (PXRD) measurements were completed on a Rigaku MiniFlex 6G from 2 to 60° 2θ running at 40 kV and 15 mA (600 W). UV−visible absorption spectra were collected on a Cary 5000 UV−vis spectrophotometer in a quartz cell with a 1 cm optical path. The fluorescence data were collected on a Horiba Fluorolog-3 spectrofluorometer using λ = 421 or 430 nm excitation wavelength for the metal-free porphyrin- and Zn porphyrin-containing frameworks, respectively. Emission lifetime measurements were collected using a Light Conversion Harpia ultrafast spectroscopy system equipped with time-correlated single-photon counting (TCSPC) capability. λ = 421 or 430 nm excitation pump pulses were generated using a Yb:KGW pumped femtosecond laser source run at 10 kHz with noncollinear optical parametric amplification.

N_2 Physisorption. The samples were degassed against vacuum (1×10^{-5} Torr) at room temperature for 15 h, heated to 120 °C in 30 min, and held at 120 °C for 2.5 h prior to surface area and pore size distribution measurements. This procedure was performed in a home-built temperature-programmed desorption setup equipped with a residual gas analyzer (RGA) able to monitor mass-to-charge ratios up from 1 to 200 amu. Using RGA, this degassing procedure was determined to be sufficient to remove remaining solvents and water potentially obstructing N_2 binding sites. N_2 physisorption isotherms at 77 K performed in a Micromeritics ASAP 2020 were collected with 45 s equilibration time in the p/p_0 range of 0−0.001 decreased to 10 s for $p/p_0 > 0.001$. From these isotherms, the specific surface area was extracted through the Brunauer−Emmett−Teller (BET) model in the range of p/p_0 from 10^{-5} to 0.1 and respecting the Rouquerol criterion.[50] The pore size distribution and pore volume of the samples were modeled in the range of $5 \times 10^{-5} < p/p_0 < 0.8$ using the commercially available DFT model for cylindrical geometries called "N_2@77 K—oxide cylindrical pores Tarazona" available in the Micromeritics software.

Optical Titration Measurements. For UV−vis titration measurements and photoluminescence (PL) quenching studies, C_{60}⊂PCN-222(H_2,Zn) and $PC_{61}BM$⊂PCN-222(H_2,Zn) suspension samples with a range of C_{60} or $PC_{61}BM$ concentration were prepared by adding 200 mL total volume aliquots of toluene containing 0.1−8 mol equiv of C_{60} or 0.1−16 mol equiv $PC_{61}BM$ per porphyrin linker to 10 mL PCN-222(H_2,Zn) suspensions in DMF. The final porphyrin linker concentration in each sample is 1.67×10^{-6} M. Samples were allowed to soak for 7 days to ensure diffusion of guest into pores prior to UV−vis and photoluminescence spectroscopy measurements.

Transient Absorption Spectroscopy. MOF suspension samples were measured in a 2 mm optical path quartz cuvette. Transient absorption data sets were acquired using a Coherent Libra Ti:sapphire laser, with an output of 800 nm at 1 kHz. A Light Conversion TOPAS-C OPA was used to generate the ∼150 fs pump pulse tuned to 400 nm for these studies to excite near the peak of the porphyrin Soret band. The pump pulse energy was typically 100 nJ, and the pump spot size was found to be ∼300 μm through the use of a beam profiler (ThorLabs). In an Ultrafast Systems Helios Spectrometer, a small amount of 800 nm light was used to pump a 1 mm sapphire crystal to generate 450−800 nm probe light for UV−vis TA. NIR supercontinuum was generated in a 10-mm-thick sapphire crystal. A delay up to 5 ns can be achieved with the Helios. Two-dimensional (2D) maps were processed averaging three spectra for background and scattering light subtraction, chirp correction, and single-wavelength kinetic slices using Surface Xplorer. Data were then imported into MATLAB software to extract decay and species-associated spectra (SAS) using a custom global fitting routine.

■ COMPUTATIONAL METHODS

Nudged elastic band (NEB)[51−53] calculations are performed using the GFN2-xTB[54] method in the ORCA[55,56] program package. The previously reported C_{60} position[36] centered within the triangularly arranged porphyrin linkers (P1) is taken as the start structure, and the position in between adjacent porphyrin trimers along the *c*-axis of the framework is taken as the final position (P2) of the NEB search. Due to the symmetry of the MOF along the *c*-axis, this search for the energetically lowest path from P1 to P2 covers all possible positions of the fullerene guest species along this pore. The

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

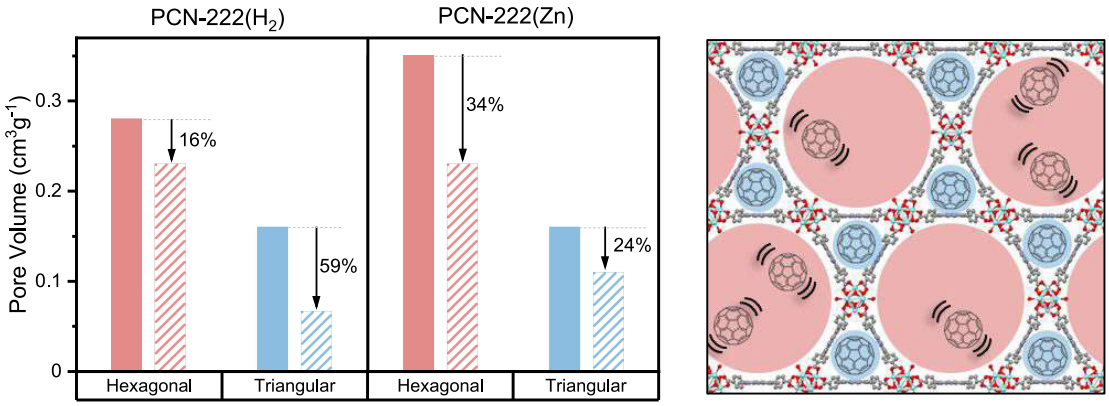


Figure 1. (Left) Pore volume of PCN-222(H_2,Zn) (solid) and C_{60}⊂PCN-222(H_2,Zn) (patterned) in the hexagonal and triangular pores determined through fitting N_2 physisorption isotherms. Arrow labels indicate the percent pore volume reduction upon introduction of C_{60} in PCN-222(H_2,Zn). (Right) Hexagonal (red) and triangular (blue) pores of the PCN-222 structure are illustrated, where the proposed location of C_{60} is graphically depicted and the weak interaction of C_{60} in the hexagonal pores is highlighted by using parentheses.

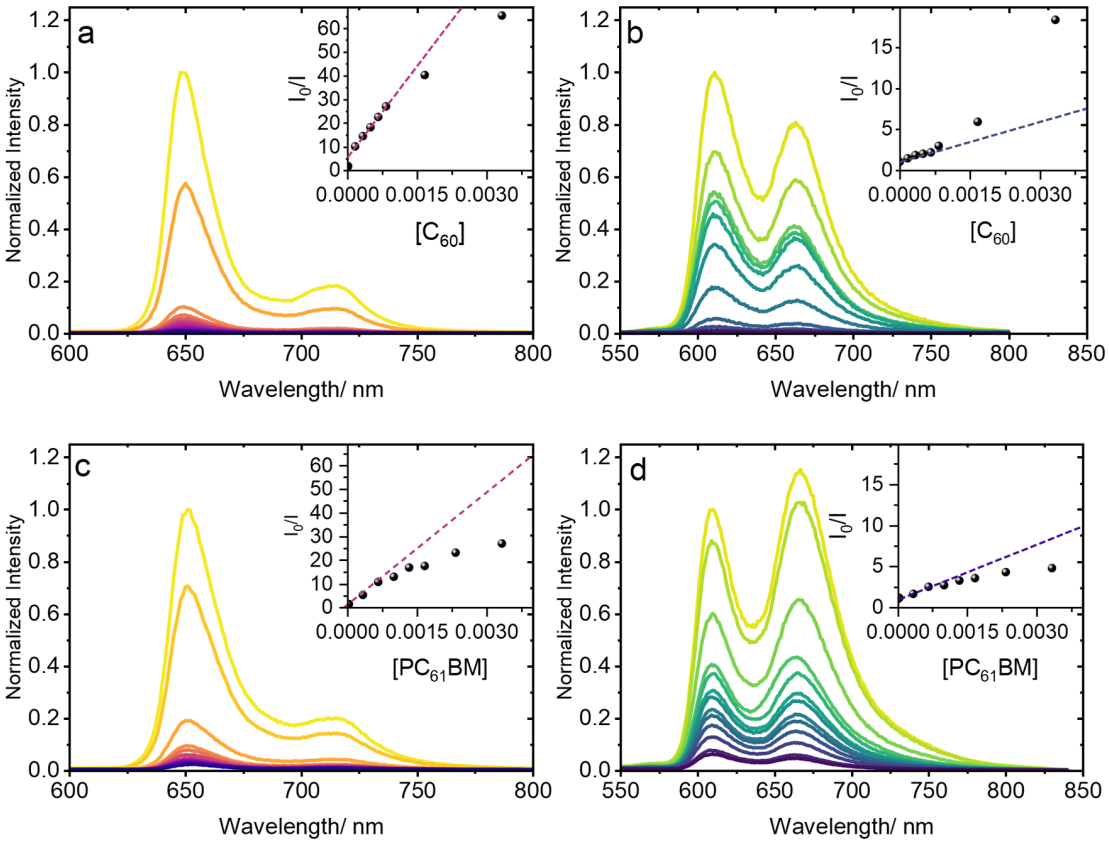


Figure 2. Fluorescence titration measurements of (a) C_{60}⊂PCN-222(H_2), (b) C_{60}⊂PCN-222(Zn), (c) $PC_{61}BM$⊂PCN-22$_2$(H_2), and (d) $PC_{61}BM$⊂PCN-222(Zn) in DMF. Insets: Stern−Volmer plots of I_0/I as a function of the C_{60} or $PC_{61}BM$ concentration (mM).

calculation is performed with a cutout of the MOF as illustrated in Figures 4 and S9. The resulting energy curve (Figure S10) does not show any local minima between P1 and P2, allowing further investigations to focus on these two positions. Interaction energies between the C_{60} guest and the framework are calculated with PBEh-3c[57] on GFN2-xTB geometries. For structures in which both P1 and P2 are occupied, interaction energies are calculated with and without a single explicit solvent molecule (DMF or 1,4-dioxane) located between the two C_{60} molecules, along with an implicit solvation model using SMD[58] for the respective solvent. Molecular dynamics (MD) simulations with mcGFN-FF[59] and NEB calculations with GFN2-xTB are performed to model the diffusion of these solvent or C_{60} guests from the small to large pores of the framework.

RESULTS AND DISCUSSION

Synthesis and Characterization. PCN-222 nanoparticles were characterized by PXRD, SEM, and DLS (Figures S1 and S2). While diffraction measurements confirmed the crystallinity and phase of the framework, SEM revealed particle sizes of ∼120 nm. DLS characterization showed similar results,

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

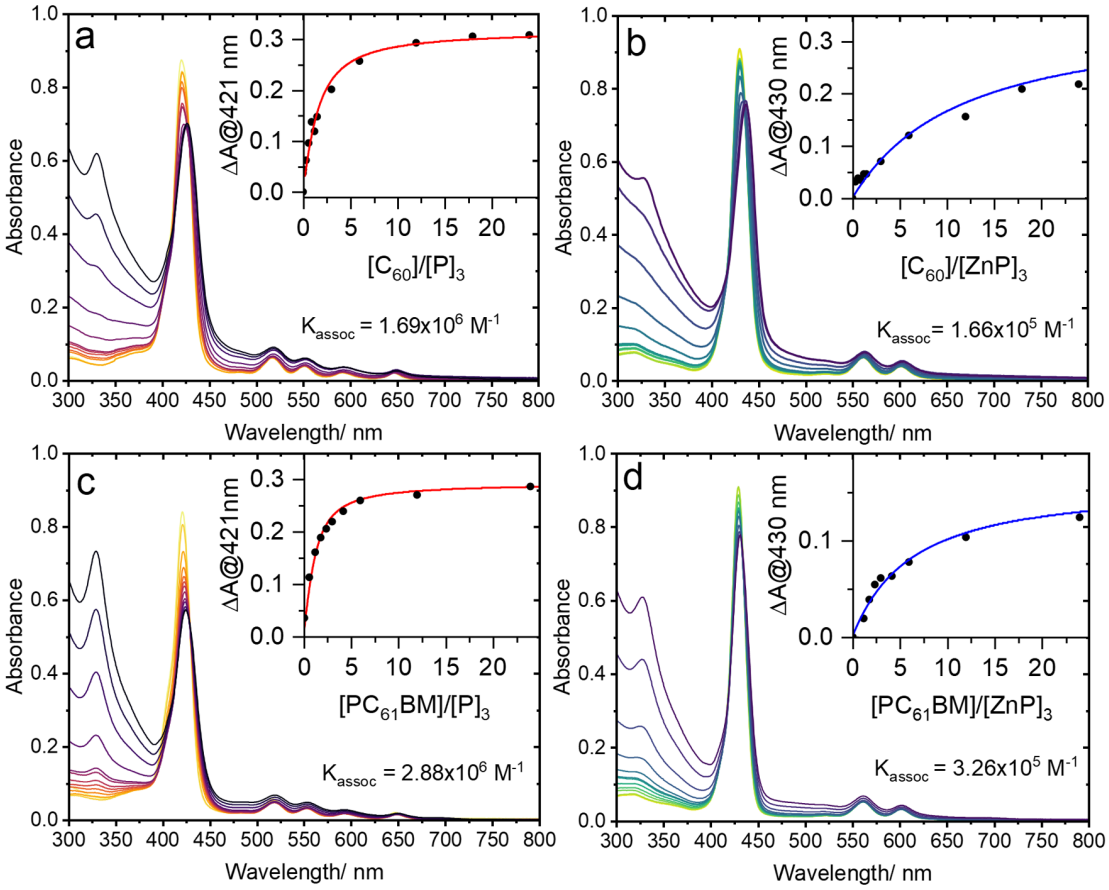


Figure 3. UV−vis titration measurements for (a) C_{60}⊂PCN-222(H_2), (b) C_{60}⊂PCN-222(Zn), (c) $PC_{61}BM$⊂PCN-222(H_2), and (d) $PC_{61}BM$⊂PCN-222(Zn) in DMF. Inset: binding isotherm generated by monitoring absorbance at λ = 421 nm (metal-free versions) and λ = 430 nm (Zn versions) as a function of relative fullerene/porphyrin concentration, expressed as fullerene to porphyrin trimer ratio, (e.g., $[C_{60}]/[P]_3$) in line with a triangular pore location.

yielding an average particle size of ∼140 nm. The PCN-222 DMF suspensions display UV−vis absorption spectra typical of porphyrins like tetraphenylporphyrin (TPP) in solution (Figure S3). Notably, the Soret band for PCN-222 is red-shifted compared to that of analogous complex TPP in DMF. Postsynthetic metalation of the free base porphyrin linker sites of PCN-222 with Zn was first indicated by the distinct color change of the MOF nanoparticle suspension from magenta to violet and then confirmed by UV−vis absorption spectroscopy as revealed by the characteristic change in the Q-band region (Figure S3). Elemental analysis by AAS characterization is consistent with complete metalation of the porphyrin linkers (5.7 wt % Zn measured compared to the theoretical 5.2 wt %). Upon introduction of the fullerene guests (C_{60} or $PC_{61}BM$), PXRD characterization confirmed the retention of crystallinity and phase of the PCN-222 hosts in each case (Figure S1).

Characterization of Pore Volume Changes by N_2 Physisorption. N_2 physisorption measurements provide some insight into the location and distribution of the fullerene guest species in the MOFs. N_2 physisorption isotherms and the results of pore size distribution analyses for PCN-222(H_2,Zn), C_{60}⊂PCN-222(H_2,Zn), and $PC_{61}BM$⊂PCN-222(H_2,Zn) are shown in Figure S4 and summarized in Table S1. Figure 1 summarizes the pore volume changes derived from these adsorption data when C_{60} is infiltrated in either the metal-free or Zn porphyrin versions of PCN-222. Similar results are found for $PC_{61}BM$ guests. The surface area and pore volume of the C_{60}- and $PC_{61}BM$-infiltrated MOFs are lower than those of the parent PCN-222 material, confirming the incorporation of these guest species within the porous frameworks. The pore size distribution analysis of these data reveals two different pore sizes consistent with the ∼1.2 nm-wide triangular porous channels and 3.7 nm-wide hexagonal porous channels of the PCN-222 framework. Notably, in the metal-free MOFs, the triangular pore volume decreased by nearly 59% while the hexagonal pore volume decreased by only 16% upon incorporation of C_{60}. For PCN-222(Zn), both the small and large pore volumes decreased by similar amounts, 24% for the triangular pores and 34% for the hexagonal pores, upon introduction of C_{60}. The same volume reduction trends are observed for the $PC_{61}BM$-loaded MOFs. The significantly larger reduction in small pore volume in PCN-222(H_2) upon incorporation of fullerene guests indicates higher loading of C_{60} or $PC_{61}BM$ in the triangular channels of this MOF compared to PCN-222(Zn). This preferential adsorption may be a consequence of stronger binding affinity of the metal-free porphyrin linker sites for fullerene guests, as will be explored below.

Optical Electronic Spectroscopy Evidence of Fullerene−Porphyrin Linker Interaction. *Photoluminescence Quenching.* While the N_2 physisorption characterization revealed the relative loading of the fullerene within the small triangular and large hexagonal pores of the framework, it does not offer direct evidence of interaction between the framework

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

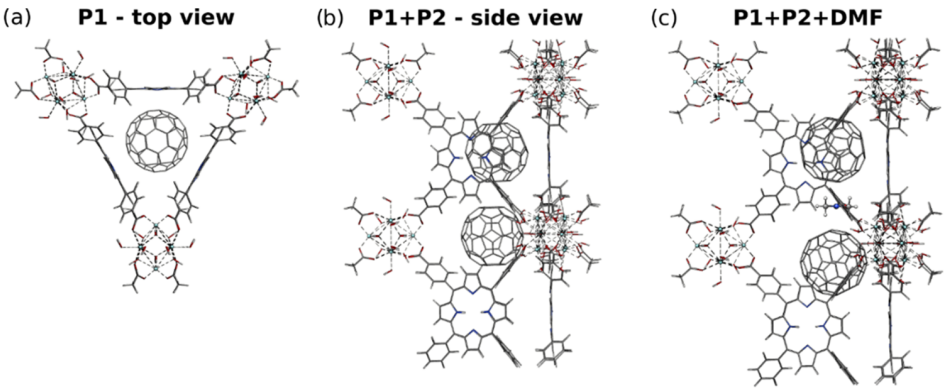


Figure 4. Illustration of C_{60} positions in the PCN-222 model system. (a) Top view along the triangular porous channel including a C_{60} molecule at position 1 in between the porphyrin linkers. (b) Side view of the model system with two C_{60} molecules located at positions 1 and 2. (c) Side view of the model system with a single DMF molecule between the two C_{60} at positions 1 and 2.

porphyrin linkers and fullerene guests. Photoluminescence quenching titration measurements provide initial confirmation of these binding interactions and insight into their relative strengths for the two MOFs and two fullerenes. Figure 2 illustrates the characteristic porphyrin linker emission of PCN-222(H_2) and PCN-222(Zn), along with the quenching observed upon the introduction of C_{60} or $PC_{61}BM$ in each case. Compared to the analogous fluorescence titration measurements of TPP or ZnTPP in solution, which show negligible quenching upon fullerene titration (Figure S5), the fluorescence quenching behavior observed for the MOF systems indicates linker−guest interaction and implicates the unique confinement effect of the framework as its primary driving force. The PCN-222(H_2) framework exhibits more pronounced fluorescence quenching compared to PCN-222(Zn) upon introduction of either C_{60} or $PC_{61}BM$, indicating a greater association of the fullerene guests with the metal-free porphyrin linker sites compared with the Zn porphyrin analogues.

The inset graphs of Figure 2 show Stern−Volmer plots derived from the titration measurements. The plots further illustrate the marked difference in fluorescence quenching between the metal-free and Zn porphyrin-based MOFs upon introduction of fullerene guests. Notable deviations from linearity are observed in the higher fullerene concentration range in each case. Interpreting these deviations, however, is complicated by the multiple factors likely influencing fluorescence quenching behavior in these materials. While combined static and dynamic quenching processes as well as sphere-of-action quenching are marked by upward curving Stern−Volmer plots, fractional accessibility of fluorophores yields the opposite trend, producing downward curvature of the plot with higher quencher concentration.[60] In fluorescent MOFs such as PCN-222, all of these factors may be influencing the observed trends. Pore confinement, coupled with different fullerene−porphyrin ground state binding strengths that depend on the porphyrin metalation status, would likely influence the relative contribution of static, dynamic, and sphere-of-action fluorescence quenching by the fullerene guests, whereas incomplete diffusion of the fullerene guest molecules within the porous structure would render some framework porphyrin fluorophore linker sites effectively inaccessible to the fullerene guests. The net downward trend for the PCN-222(H_2) Stern−Volmer plot suggests that the latter factor dominates at higher C_{60} and $PC_{61}BM$ concentrations. However, the near-complete fluorescence quenching observed even at modest fullerene concentrations suggests that this population of inaccessible fluorophores is quite small. The overall upward curvature of the PCN-222(Zn) Stern−Volmer plot at higher C_{60} concentration indicates the dominant effects of increased contribution of collision and/or sphere-of-action quenching between excited porphyrin linkers and weakly/nonbound, yet confined, fullerene guests. Evidence for this behavior can be found in a changing fluorescence lifetime with increased C_{60} loading (Figures S6 and S7). Interestingly, this upward curvature trend is not observed for $PC_{61}BM$ in PCN-222(Zn). Its downward curving Stern−Volmer plot suggests that the added steric bulk of this fullerene guest may inhibit its diffusion, outweighing any collision or sphere-of-action quenching contributions.

UV−Vis Titration Measurements. UV−vis absorption spectroscopy characterization provides evidence for a ground state interaction between the fullerene guest and porphyrin linkers in PCN-222(H_2,Zn) through the observed red shift and absorbance decrease of the Soret band (Figure 3). These spectral changes occur when electron density is withdrawn from the porphyrin units by cofacial van der Waals interactions with the fullerene molecules within the porous framework.[61] Note, no porphyrin-based Soret or Q-band spectral changes are observed upon addition of $PC_{61}BM$ or C_{60} to TPP or ZnTPP in DMF solutions (Figure S8), confirming the unique confinement effect of the porphyrin-based frameworks on the fullerene guests.

UV−vis titration experiments (Figure 3) are used to establish association constants K_{assoc} for each fullerene-loaded MOF system. Titration isotherms are generated from the Soret band changes measured for the MOF suspensions upon introduction of the fullerene guests. On the basis of this analysis,[62,63] the K_{assoc} values were evaluated to be 1.69×10^6 M^{-1} for C_{60}⊂PCN-222(H_2), 2.88×10^6 M^{-1} for $PC_{61}BM$⊂PCN-222(H_2), 1.66×10^5 M^{-1} for C_{60}⊂PCN-222(Zn), and 3.26×10^5 M^{-1} for $PC_{61}BM$⊂PCN-222(Zn). These results show that C_{60} and $PC_{61}BM$ have similar binding strengths within PCN-222. However, the metal-free porphyrin framework exhibits higher binding affinities for these fullerene species compared with the zinc analogue. While this trend contradicts the reported computational studies of C_{60} binding strength on this system,[36] analogous titration studies of related metal-free and zinc porphyrin complexes with cofacial C_{60} interaction also showed stronger binding for the metal-free

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

porphyrin versions.[64] Notably, those studies similarly predicted a higher affinity for the Zn porphyrin version as well. The single inflection point in these titration data suggests that only one axial coordination site involves significant binding interaction. Based on the appropriately sized small triangular pores of the PCN-222 structure, we hypothesize that this is the location of the bound fullerene guest species. This binding location is both theoretically predicted[36] and in line with reported cyclic porphyrin trimer supramolecular cage systems with similar dimensions that form analogous complexes with fullerenes.[65] UV−vis titration measurements of that supramolecular system yielded association constants on the order of 10^4−10^6 M^{-1} depending on the solvent.

Computational Investigation of Fullerene Locations and Solvation Environment. Computational modeling further elucidates the likely locations of the fullerene guest species within the triangular pore volume of the framework. In previous computational studies,[36] only the position of the guest species that are cofacially associated with the porphyrin linkers in the small triangular pore has been considered. However, other positions along this channel might be occupied and could affect the properties of the MOF. To this extent, possible locations of the C_{60} molecule along the triangular porous channel of the PCN-222 structure were investigated by using the NEB reaction path exploration algorithm. Two minima were identified: one associated with the C_{60} interacting with the triangularly arranged porphyrin linkers (P1) and the other for the C_{60} located at the position between adjacent porphyrin sites along the triangular channel (P2) (see Figure 4).

Calculated interaction energies, listed in Table 1, show that both positions can be occupied simultaneously without an energetic penalty. Specifically, the interaction energy with both positions occupied is 1.7−1.8 kcal/mol lower than the sum of the interaction energies of only P1 or P2 occupied depending on the implicit solvent. To investigate whether a solvation shell around the C_{60} molecules at P1 and P2 is feasible, we additionally calculated interaction energies with a single explicit solvent molecule between the two C_{60} molecules along with implicit solvation contributions. The energies reported in Table 1 show that with the addition of a DMF molecule, the interaction energy is reduced by 82%, and for 1,4-dioxane it is reduced by 108%. This indicates that occupying both positions in the small triangular pore with C_{60} molecules does not leave enough space for solvent molecules. Furthermore, molecular dynamics (MD) simulations showed that both solvent molecules can move from the small pore to the large pore through the gap between the two linkers in PCN-222, while the C_{60} molecules are unable to move through this gap. This suggests that upon fullerene loading, the small triangular pores are mostly occupied by the C_{60} molecules and the solvent molecules are displaced into the larger voids of the hexagonal pores. The results of the MD simulation are presented in Figures S11−S13 of the Supporting Information. The geometries for all calculations and the reaction equations for the interaction energies are given in the Supporting Information.

Table 1. Calculated Interaction Energies for C_{60} Molecules at the PBEh-3c Composite DFT Level on Optimized GFN2-xTB Geometries Located at Local Minima P1 and P2 within the One-Dimensional Triangular Pore of PCN-222 Model[a]

solvent	C_{60} location	energy (kcal/mol) (implicit solvent)	energy (kcal/mol) (implicit + explicit solvent)[b]
DMF	P1	−22.5	
	P2	−25.2	
	P1 + P2	−45.9	−8.5
1,4-dioxane	P1	−25.3	
	P2	−24.9	
	P1 + P2	−48.5	+3.7

[a]The energies include solvation contributions from the SMD model as indicated. [b]A single solvent molecule located between the two C_{60} molecules as described in the text.

Characterization of Charge Transfer and Recombination by Ultrafast Transient Absorption Spectroscopy. To analyze the excited state dynamics of these fullerene-loaded MOF systems, we turned to femtosecond transient absorption (TA) spectroscopy. TA data collected for C_{60}⊂PCN-222(H_2) and C_{60}⊂PCN-222(Zn) are presented in Figure 5. The top panels (Figure 5a,b) contain spectral slices extracted at different time delays over the 5 ns window and kinetic traces (inset graphs) collected at diagnostic probe wavelengths associated with different transient species. The bottom panels (Figure 5c,d) show the species-associated spectra (SAS) of these two systems derived through the TA data global fitting analyses (vide infra). The TA results for the $PC_{61}BM$-loaded MOFs are found in Figure S14, and those of the MOF nanoparticle suspensions without fullerene guests are shown in Figure S15. Upon photoexcitation, each MOF system displays characteristic porphyrin excited state absorption (ESA) at ∼480 nm along with Soret and Q-band ground state bleach (GSB) features, reflecting the immediate population of the S_1 excited state of the porphyrin moiety (following fast internal conversion from S_2 that occurs within the instrument response time). While PCN-222(H_2) and PCN-222(Zn) in the absence of fullerene exhibit a decay of these features that indicates direct return to the ground state, the presence of the acceptor guest species in these MOFs yields TA spectral signatures that reflect CT excited state population along their decay path in each case. Following well-established literature precedent, the broad absorption band between 600 and 800 nm arises from the porphyrin radical cation,[42] and the 1080 and 1020 nm ESA bands are due to the C_{60} and $PC_{61}BM$ radical anion[66] species, respectively. By 5 ns, the broad visible ESA band converges into a feature centered around ∼740 nm, while the broad ESA in the NIR region beyond 900 nm decays to zero. This spectral evolution at longer delay times suggests charge recombination to the fullerene triplet excited state[67] based on similar observations of electron transfer pathways in molecular porphyrin-C_{60} systems.[42] At this long delay time, TA spectra for the metal-free porphyrin frameworks with fullerene guests also display residual absorption at ∼475 nm, accompanied by the GSB features, which is the hallmark of porphyrin locally excited triplet state population[68] and has been documented in molecular porphyrin-C_{60} analogues.[37,42] The persistence of both ESA features along with the GSB at long time delays in this case indicates that charge recombination occurs via both porphyrin and fullerene triplet excited state decay paths. In contrast, TA spectra for the Zn-metalated porphyrins in PCN-222(Zn) with fullerene guests only show absorption around 740 nm by 5 ns, with all other features decaying to zero, indicating that the reverse charge transfer (RCT) event decays primarily to the C_{60} triplet state.

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

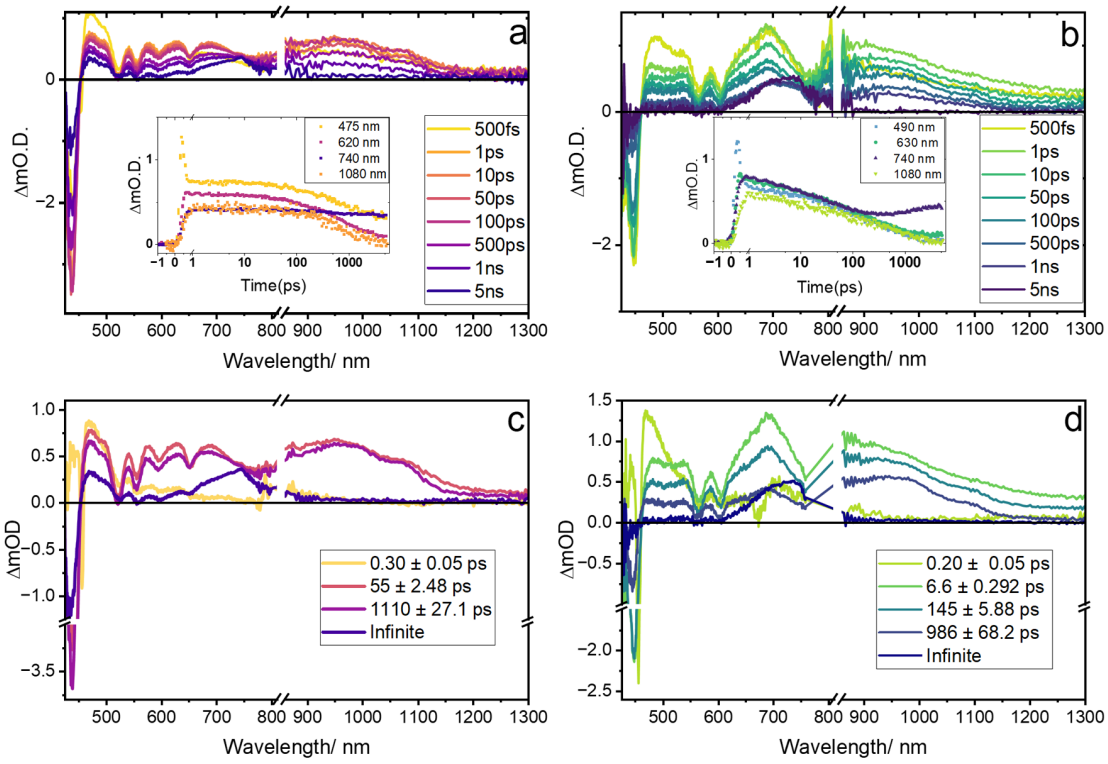


Figure 5. TA spectral overlays of (a) C_{60}⊂PCN-222(H_2) and (b) C_{60}⊂PCN-222(Zn); the insets show kinetic slices obtained from the corresponding spectra at 475, 740, and 1080 nm. Species-associated spectra of (c) C_{60}⊂PCN-22$_2$(H_2) and (d) C_{60}⊂PCN-222(Zn). Samples were measured as nanoparticle suspensions in DMF with λ = 400 nm excitation.

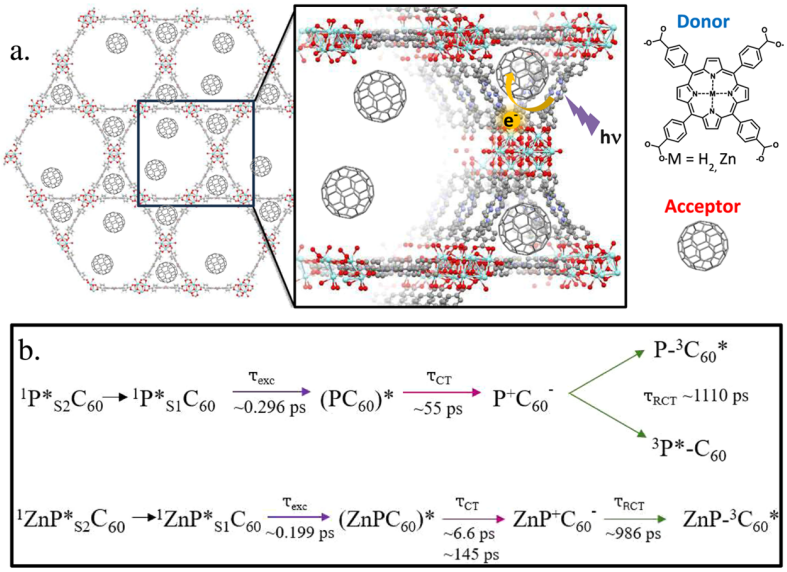


Figure 6. (a) Illustration of photoinduced electron transfer pathway as it relates to the proposed encapsulated location of the fullerene guests within the small triangular pores of the MOF. (b) Proposed excited state decay pathway of C_{60}⊂PCN-222(H_2) (top) and C_{60}⊂PCN-222(Zn) (bottom).

The global fits of the TA data provide further insight into the excited state decay pathways and kinetics for these fullerene guest-containing MOFs. Target analyses (see SI for more details) reveal four intermediate state components for both C_{60}⊂PCN-222(H_2) and $PC_{61}BM$⊂PCN-222(H_2) and five components for C_{60}⊂PCN-222(Zn) and $PC_{61}BM$⊂PCN-222(Zn). The SAS with the fastest lifetime component for both C_{60}⊂PCN-222(H_2) and C_{60}⊂PCN-222(Zn) is assigned to the singlet excited state of the porphyrin donor linkers, ^{1}P*-C_{60} or ^{1}ZnP*-C_{60} with lifetimes τ = ∼0.3 and ∼0.2 ps, respectively. In C_{60}⊂PCN-222(H_2), this initial porphyrin excited state decays to subsequent intermediate species, one with τ = ∼55 ps and another with τ = ∼1110 ps. These SAS components each display absorption features at ∼625 and ∼1080 nm (i.e., porphyrin cation and C_{60} anion spectroscopic signatures, respectively), indicating population of excited states with CT character. Notably, the τ = ∼55 ps SAS component displays broader spectral features that become slightly more

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

resolved by the third component. These global fit results align with those reported for related porphyrin−fullerene molecular dyads that involve through-space electron transfer.[69−71] In these systems, exciplex formation is invoked to explain the shorter lifetime component, indicating that its population precedes that of the formal charge transfer excited states along the decay pathway. The exciplex species in this molecular dyad, characterized by broader spectral features compared to the pure CT state, is a geometrical rearrangement of the donor and acceptor in the excited state that delocalizes the initial excitation of the donor with the possibility of some CT character. Thus, we assign the shorter and longer lifetime intermediate SAS components in C_{60}⊂PCN-222(H_2) to the exciplex (PC_{60})* and $P^+C_{60}^-$ CT state decay pathways, respectively. The C_{60}⊂PCN-222(Zn) excited state deactivation pathway also involves exciplex ($ZnPC_{60}$)* and CT excited state, $ZnP^+C_{60}^-$ populations; however, the exciplex involves two decay components (with $\tau_{exciplex}$ = 6.6 and 145 ps). These observed kinetics may be attributed to weaker fullerene binding affinities of the Zn porphyrin linkers (Figure 3), which allow multiple relaxation steps toward the optimal exciplex geometry. The formal CT excited state, with population similarly confirmed by the sharpened spectral features, has a lifetime of τ_{RCT} = 986 ps. The final "infinite" SAS component for C_{60}⊂PCN-222(H_2) that persists beyond the 5 ns time window of the TA experiment, displays absorption features at ∼475 and ∼740 nm, corresponding to the triplet states of the porphyrin[68] and C_{60},[67] respectively. The final SAS component for C_{60}⊂PCN-222(Zn) only displays the ∼740 nm feature confirming exclusive decay via the C_{60} triplet excited state as previously discussed. A summary of the proposed excited state deactivation pathways derived from these TA analyses for the two MOF systems is depicted in Figure 6b.

Aside from the differences in exciplex and triplet decay pathways, both metal-free and zinc porphyrin frameworks loaded with C_{60} acceptor guests exhibit similar excited state dynamics with respect to charge transfer behavior and reverse charge transfer lifetimes (∼1 ns). The deviation in triplet decay pathways between C_{60}⊂PCN-222(H_2) and C_{60}⊂PCN-222(Zn) can be rationalized based on the established triplet state energy levels of relevant porphyrin and zinc porphyrin complexes relative to that of C_{60}. While the triplet energies reported for metal-free and zinc tetraphenylporphyrin complexes are 1.43 and 1.61 eV, respectively,[72] that of the fullerene triplet state $^3C_{60}$* is 1.55 eV.[73] Assuming similar triplet state energies of these moieties in the MOF-C_{60} donor−acceptor systems, dual population of the lower-energy metal-free porphyrin triplet state along with $^3C_{60}$* upon charge recombination is plausible, while the analogous population of the Zn porphyrin triplet state may be fleeting or energetically inaccessible. Furthermore, the significantly enhanced binding interaction and greater pore volume reduction for the fullerene-loaded PCN-222(H_2) framework compared to those of PCN-222(Zn) are a potential source of divergent behavior. If the triplet energy migrates to more weakly bound fullerene in the large pore of PCN-222(Zn), the proposed equilibrium between ^{3}P* and $^3C_{60}$* could be interrupted by subsequent fullerene diffusion that renders triplet energy transfer irreversible to ^{3}P*. The global fit analyses of the PC_{61}BM⊂PCN-222(H_2) and PC_{61}BM⊂PCN-222(Zn) TA results (Figures S19 and S20) reveal similar decay pathways and kinetics as the C_{60}-loaded MOFs, including exciplex state formation, RCT lifetimes, and triplet state populations, indicating minimal impact of the butyric-methyl ester appendage on the electron transfer photophysics in this host−guest arrangement. This similar excited state behavior for the two fullerene intercalants, despite the apparent absence of a dynamic quenching channel for free or weakly bound PC_{61}BM, suggests that energy offsets in the triplet manifolds may be the primary factor that dictates the product of reverse charge transfer.

The TA results presented thus far were obtained using DMF as the solvent environment for the fullerene-loaded framework suspensions. TA measurements were also collected for these MOF systems upon 1,4-dioxane solvent exchange to probe the influence of solvent polarity on their photophysical responses following the C_{60} introduction. Charge-separated excited state energies in most D−A systems are heavily influenced by solvent polarity, which can lead to drastic changes in observed photoinduced electron transfer and recombination pathways.[44,69,74−79] While polar or moderately polar solvents such as DMF can facilitate charge separation by stabilizing the CT state, nonpolar media can hinder photoinduced charge separation by rendering the CT state less energetically favorable, or even inaccessible. The TA data collected in 1,4-dioxane are shown in Figure S16 with the results of the global fitting analysis depicted in Figures S21 and S22. Using the same sequence models applied to these systems measured in DMF, global fitting yielded analogous intermediate SAS components associated with singlet porphyrin-localized excited state, exciplex, and CT excited state populations. Furthermore, the TA analyses reveal comparable charge recombination pathways dominated by C_{60} triplet excited state population for C_{60}⊂PCN-222(Zn) or a dual population of both porphyrin and C_{60} triplet excited states for C_{60}⊂PCN-222(H_2). Notably, despite the significant difference in solvent polarity between DMF (ε = 36.7) and 1,4-dioxane (ε = 2.20) environments, the TA results revealed similar CT state formation and decay lifetimes for these fullerene-loaded MOF materials in both solvents. This suggests that the C_{60} acceptor species are sufficiently shielded when confined within the small triangular channels of the framework, such that the resulting D−A CT excited state energies are essentially impervious to the surrounding solvent environment. As a result, only the electrostatic interaction between the framework porphyrin linkers and fullerene guest species influences their electron transfer behavior and kinetics. The expulsion of solvent from the small triangular channels upon introduction of the fullerene guests is further supported by the computational results presented above, which reveal that the C_{60}-loaded MOF in the absence of solvent is more thermodynamically favorable than this system in the presence of interstitial DMF or 1,4-dioxane solvent molecules.

Comparison with Molecular Porphyrin−Fullerene D−A Systems. Comparing electronic structure and excited state dynamics derived from the observed photophysics of these fullerene-loaded MOF materials with those of relevant molecular porphyrin−fullerene D−A systems highlights the unique role of the framework in dictating photoinduced electron transfer and decay pathways. It is well established that covalently linked C_{60}-porphyrin(Zn,H_2) dyads undergo D−A electron transfer following local porphyrin donor excitation, with charge separation and recombination pathways and kinetics determined by the linker distance between the porphyrin donor and fullerene acceptor moieties as well as

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

the solvation environment.[44,69,74,75] In these molecular examples, however, the fullerene and porphyrin moieties are connected via a single linker, which leads to substantial conformational flexibility and negligible direct $\pi-\pi$ interaction between the porphyrin plane and the convex π surface of the fullerene. Cyclophane-like C_{60}-porphyrin dyads, with two separate linkers symmetrically connecting the porphyrin and fullerene components, yield face-to-face π-stacked structures,[61,70] while cage complexes such as the covalent organic polyhedron, COP-5, possess a fixed cofacial orientation of two porphyrin sites for encapsulation of fullerene (C_{70} in the reported system).[37] These molecular analogues display through-space $\pi-\pi$ interactions that in some ways resemble those of the fullerene-loaded porphyrin-based MOFs. In each molecular case, direct interchromophore interaction, with similar fixed edge-to-edge donor−acceptor distances, R_{EE} ~2.8 Å, is evinced by absorption spectra perturbations like those observed in $C_{60}\subset$ and $PC_{61}BM\subset$PCN-222(H_2,Zn). Unlike the MOF systems, however, the molecular analogues exhibit additional NIR absorption and emission features (at least in nonpolar solvents). The appearance of the new absorption band indicates a direct excitation to the CT excited state, while the emission band is attributed to its radiative decay.[37,61] For the cyclophane-like dyads, despite the lack of this CT emission in polar solvents, exciplex formation is assigned as a precursor to the fully charge-separated state, as mentioned above. In the C_{70}@COP-5 system, however, the charge transfer state forms from the COP-5 porphyrin singlet excited state without evidence of an intervening exciplex, regardless of solvent polarity. While the PCN-222 framework rigidity and cage-like encapsulation of the fullerene guest molecules might predict similar behavior, exciplex intermediate is nonetheless detected in this system in both polar and nonpolar solvent environments. The different pathways may be attributed to the fullerene triangular pore location and subsequent electronic coupling with three porphyrin moieties in PCN-222 (rather than two in the case of COP-5) as well as the different fullerene species encapsulated in each case (C_{60} vs C_{70}). Regardless of the participation of exciplex states, the electron transfer process occurs on the ultrafast time scale in each case.

The reverse charge transfer lifetime trends observed for the $C_{60}\subset$ and $PC_{61}BM\subset$PCN-222(H_2,Zn) series resemble the behavior reported for the fullerene-loaded COP-5 system. Namely, electron transfer occurs for the cage complex in both polar and nonpolar solvents with similar charge recombination lifetimes, despite a large difference in solvent dielectric constant. The prolonged CR lifetimes are attributed to the rigidity of the porphyrin cage in COP-5 and very small reorganization energy upon photoinduced electron transfer. Framework rigidity in the PCN-222 series likely plays a similar role along with the solvent-excluded pore location of the acceptor species, which essentially eliminates the solvent reorganization effects on the energetics of the CT state and, therefore, the rate of charge recombination. Notably, the lifetime is even longer than the 300−600 ps range reported for the COP-5 system, and more impervious to solvent polarity with τ_{RCT} ~1 ns for all MOF systems in both polar and nonpolar solvents. The deviation in charge recombination pathways further highlights how the unique arrangement, rigidity, and solvent exposure of the fullerene and porphyrin moieties in each case influence the relative energy of the CT excited state. The cofacial C_{60}-porphyrin molecular systems that form charge-separated excited states in polar solvents decay via direct recombination to the ground state, bypassing any triplet excited state population. Only in the C_{70}@COP-5 system with a nonpolar solvent environment is a porphyrin triplet state invoked as the dominant charge recombination pathway. The dual triplet decay pathway observed upon charge recombination within the fullerene-loaded PCN-222(H_2) framework suggests near-energetic resonance between the CT state and both porphyrin and fullerene triplets.

CONCLUSIONS

The framework-imposed arrangement of porphyrin linkers in PCN-222 is responsible for the unique donor−acceptor interactions with encapsulated fullerene guests. The distinct occupation, binding strength, and solvent accessibility of acceptors in different pore types of these tailored MOFs lead to electronic structure perturbations that are unmatched in molecular analogues. Although in all cases fast charge transfer occurs from porphyrin to fullerene, the subsequent flow of energy and charge is redirected by the tunable energy landscape—most notably, through varying excimer relaxation pathways and an eventual exclusive formation of fullerene triplet in the Zn-metalated MOF. Employing this energy flow strategically in a photochemical or catalytic reaction is a future goal that may require additional structural optimization and the inclusion of catalytic metal species.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.5c00161.

Powder XRD, SEM, DLS, UV−vis and fluorescence titration control experiments, N_2 physisorption isotherms, pore size distributions, further computational results, additional TA results, details of global fitting procedure, and global fitting analysis results (PDF)

Geometries for interaction energies, MD trajectories, and NEB trajectories (ZIP)

AUTHOR INFORMATION

Corresponding Authors

Justin C. Johnson − *National Renewable Energy Laboratory, Golden, Colorado 80401, United States;* orcid.org/0000-0002-8874-6637; Email: justin.johnson@nrel.gov

Jenny V. Lockard − *Department of Chemistry, Rutgers University-Newark, Newark, New Jersey 07102, United States;* orcid.org/0000-0001-7758-169X; Email: jlockard@newark.rutgers.edu

Authors

Alison Arissa − *Department of Chemistry, Rutgers University-Newark, Newark, New Jersey 07102, United States*

Thomas Rose − *Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Bonn 53115, Germany*

Noémi Leick − *National Renewable Energy Laboratory, Golden, Colorado 80401, United States;* orcid.org/0000-0002-2014-6264

Stefan Grimme − *Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie,*

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

Rheinische Friedrich-Wilhelms Universität Bonn, Bonn 53115, Germany; orcid.org/0000-0002-5844-4371

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.5c00161

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The development of synthesis, structural, and steady state characterization and their analysis were supported by the National Science Foundation under Grant No. DMR-2423113. Transient absorption spectroscopy, N_2 physisorption experiments, and their analysis were supported by the Solar Photochemistry Program, Division of Chemical Sciences, Geosciences, and Biosciences (CSGB) Division, within the Office of Basic Energy Sciences, Office of Science, DOE. This work was authored in part by the National Renewable Energy Laboratory, operated for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. A.A. would like to acknowledge partial support provided by an NIGMS Graduate Research Training Initiative for Student Enhancement (G-RISE) Grant (5T32GM140951-03) and National GEM Consortium fellowship. The views expressed in the article do not necessarily represent the views of the Department of Energy or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

■ REFERENCES

(1) Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M. The Chemistry and Applications of Metal-Organic Frameworks. *Science* **2013**, *341*, No. 1230444.

(2) Kole, G. K.; Vittal, J. J. Solid-state reactivity and structural transformations involving coordination polymers. *Chem. Soc. Rev.* **2013**, *42*, 1755−1775.

(3) Foo, M. L.; Ryotaro, M.; Susumu, K. Functional Hybrid Porous Coordination Polymers. *Chem. Mater.* **2014**, *26*, 310−322.

(4) Wang, C.; Demin, L.; Wenbin, L. Metal−Organic Frameworks as A Tunable Platform for Designing Functional Molecular Materials. *J. Am. Chem. Soc.* **2013**, *135*, 13222−13234.

(5) Yuan, S.; Feng, L.; Wang, K.; Pang, J.; Bosch, M.; Lollar, C.; Sun, Y.; Qin, J.; Yang, X.; Zhang, P.; et al. Stable Metal−Organic Frameworks: Design, Synthesis, and Applications. *Adv. Mater.* **2018**, *30*, No. 1704303.

(6) Zhao, Y.; Wu, L.; Wu, K.; Wei, R.-J.; Zeng, H.; Pang, H.; Lu, W.; Li, D. Host-guest interactions in the confined spaces of metal−organic frameworks: Design principles, characterizations, and applications. *Coord. Chem. Rev.* **2025**, *524*, No. 216302.

(7) Ji, Z.; Wang, H.; Canossa, S.; Wuttke, S.; Yaghi, O. M. Pore Chemistry of Metal−Organic Frameworks. *Adv. Funct. Mater.* **2020**, *30*, No. 2000238.

(8) Gao, W.-Y.; Chrzanowski, M.; Ma, S. Metal-Metalloporphyrin Frameworks: A Resurging Class of Functional Materials. *Chem. Soc. Rev.* **2014**, *43*, 5841−5866.

(9) Liu, J.; Thallapally, P. K.; McGrail, B. P.; Brown, D. R.; Liu, J. Progress in adsorption-based CO2 capture by metal-organic frameworks. *Chem. Soc. Rev.* **2012**, *41*, 2308−2322.

(10) Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R. Carbon Dioxide Capture in Metal−Organic Frameworks. *Chem. Rev.* **2012**, *112*, 724−781.

(11) Yang, X.; Xu, Q. Bimetallic Metal−Organic Frameworks for Gas Storage and Separation. *Cryst. Growth Des.* **2017**, *17*, 1450−1455.

(12) Wang, H.; Li, J. General strategies for effective capture and separation of noble gases by metal−organic frameworks. *Dalton Trans.* **2018**, *47*, 4027−4031.

(13) Pascanu, V.; Miera, G. G.; Inge, A. K.; Martín-Matute, B. Metal−Organic Frameworks as Catalysts for Organic Synthesis: A Critical Perspective. *J. Am. Chem. Soc.* **2019**, *141*, 7223−7234.

(14) Rogge, S. M. J.; Bavykina, A.; Hajek, J.; Garcia, H.; Olivos-Suarez, A. I.; Sepulveda-Escribano, A.; Vimont, A.; Clet, G.; Bazin, P.; Kapteijn, F.; et al. Metal-organic and covalent organic frameworks as single-site catalysts. *Chem. Soc. Rev.* **2017**, *46*, 3134−3184.

(15) Zhao, M.; Ou, S.; Wu, C.-D. Porous Metal−Organic Frameworks for Heterogeneous Biomimetic Catalysis. *Acc. Chem. Res.* **2014**, *47*, 1199−1207.

(16) D'Alessandro, D. M. Exploiting redox activity in metal−organic frameworks: concepts, trends and perspectives. *Chem. Commun.* **2016**, *52*, 8957−8971.

(17) Calbo, J.; Golomb, M. J.; Walsh, A. Redox-active metal−organic frameworks for energy conversion and storage. *J. Mater. Chem. A* **2019**, *7*, 16571−16597.

(18) Cong, C.; Ma, H. Photoelectroactive metal−organic frameworks. *J. Mater. Chem. A* **2023**, *11*, 13065−13088.

(19) Fumanal, M.; Ortega-Guerrero, A.; Jablonka, K. M.; Smit, B.; Tavernelli, I. Charge Separation and Charge Carrier Mobility in Photocatalytic Metal-Organic Frameworks. *Adv. Funct. Mater.* **2020**, *30*, No. 2003792.

(20) Li, X.; Yu, J.; Lu, Z.; Duan, J.; Fry, H. C.; Gosztola, D. J.; Maindan, K.; Rajasree, S. S.; Deria, P. Photoinduced Charge Transfer with a Small Driving Force Facilitated by Exciplex-like Complex Formation in Metal−Organic Frameworks. *J. Am. Chem. Soc.* **2021**, *143*, 15286−15297.

(21) Rajasree, S. S.; Fry, H. C.; Gosztola, D. J.; Saha, B.; Krishnan, R.; Deria, P. Symmetry-Breaking Charge Transfer in Metal−Organic Frameworks. *J. Am. Chem. Soc.* **2024**, *146*, 5543−5549.

(22) Wang, Q.; Gao, Q.; Al-Enizi, A. M.; Nafady, A.; Ma, S. Recent advances in MOF-based photocatalysis: environmental remediation under visible light. *Inorg. Chem. Front.* **2020**, *7*, 300−339.

(23) Yan, M.; Johnson, E. M.; Morris, A. J. Redox Hopping in Metal−Organic Frameworks through the Lens of the Scholz Model. *J. Phys. Chem. Lett.* **2023**, *14*, 10700−10709.

(24) Ahrenholtz, S. R.; Epley, C. C.; Morris, A. J. Solvothermal Preparation of an Electrocatalytic Metalloporphyrin MOF Thin Film and its Redox Hopping Charge-Transfer Mechanism. *J. Am. Chem. Soc.* **2014**, *136*, 2464−2472.

(25) Yan, M.; Bowman, Z.; Knepp, Z. J.; Peterson, A.; Fredin, L. A.; Morris, A. J. Reaction-Type-Dependent Behavior of Redox-Hopping in MOFs—Does Charge Transport Have a Preferred Direction? *J. Phys. Chem. Lett.* **2024**, *15*, 11919−11926.

(26) Sun, M.-L.; Wang, Y.-R.; He, W.-W.; Zhong, R.-L.; Liu, Q.-Z.; Xu, S.; Xu, J.-M.; Han, X.-L.; Ge, X.; Li, S.-L.; et al. Efficient Electron Transfer from Electron-Sponge Polyoxometalate to Single-Metal Site Metal−Organic Frameworks for Highly Selective Electroreduction of Carbon Dioxide. *Small* **2021**, *17*, No. 2100762.

(27) Chang, X.; Xu, Y.; von Delius, M. Recent advances in supramolecular fullerene chemistry. *Chem. Soc. Rev.* **2024**, *53*, 47−83.

(28) Wang, C.; Bo, W.; Chunru, W. Rational Construction and Efficient Regulation of Stable and Long-Lived Charge-Separation State in Fullerene Materials. *Acc. Mater. Res.* **2024**, *5*, 426−437.

(29) Liao, M.-S.; Watts, J. D.; Huang, M.-J. Interaction of Metal Porphyrins with Fullerene C60: A New Insight. *J. Phys. Chem. B* **2007**, *111*, 4374−4382.

(30) Boyd, P. D. W.; Reed, C. A. Fullerene−Porphyrin Constructs. *Acc. Chem. Res.* **2005**, *38*, 235−242.

(31) Sakaguchi, K.-i.; Kamimura, T.; Uno, H.; Mori, S.; Ozako, S.; Nobukuni, H.; Ishida, M.; Tani, F. Phenothiazine-Bridged Cyclic Porphyrin Dimers as High-Affinity Hosts for Fullerenes and Linear Array of C60 in Self-Assembled Porphyrin Nanotube. *J. Org. Chem.* **2014**, *79*, 2980−2992.

(32) Kaur, R.; Sen, S.; Larsen, M. C.; Tavares, L.; Kjelstrup-Hansen, J.; Ishida, M.; Zieleniewska, A.; Lynch, V. M.; Bähring, S.; Guldi, D. M.; et al. Semiconducting Supramolecular Organic Frameworks Assembled from a Near-Infrared Fluorescent Macrocyclic Probe and Fullerenes. *J. Am. Chem. Soc.* **2020**, *142*, 11497−11505.
(33) Feng, D.; Gu, Z.-Y.; Li, J.-R.; Jiang, H.-L.; Wei, Z.; Zhou, H.-C. Zirconium-Metalloporphyrin PCN-222: Mesoporous Metal−Organic Frameworks with Ultrahigh Stability as Biomimetic Catalysts. *Angew. Chem., Int. Ed.* **2012**, *51*, 10307−10310.
(34) Mondloch, J. E.; Bury, W.; Fairen-Jimenez, D.; Kwon, S.; DeMarco, E. J.; Weston, M. H.; Sarjeant, A. A.; Nguyen, S. T.; Stair, P. C.; Snurr, R. Q.; et al. Vapor-Phase Metalation by Atomic Layer Deposition in a Metal−Organic Framework. *J. Am. Chem. Soc.* **2013**, *135*, 10294−10297.
(35) Kung, C.-W.; Otake, K.; Buru, C. T.; Goswami, S.; Cui, Y.; Hupp, J. T.; Spokoyny, A. M.; Farha, O. K. Increased Electrical Conductivity in a Mesoporous Metal−Organic Framework Featuring Metallacarboranes Guests. *J. Am. Chem. Soc.* **2018**, *140*, 3871−3875.
(36) Pratik, S. M.; Gagliardi, L.; Cramer, C. J. Engineering Electrical Conductivity in Stable Zirconium-Based PCN-222 MOFs with Permanent Mesoporosity. *Chem. Mater.* **2020**, *32*, 6137−6149.
(37) Ortiz, M.; Cho, S.; Niklas, J.; Kim, S.; Poluektov, O. G.; Zhang, W.; Rumbles, G.; Park, J. Through-Space Ultrafast Photoinduced Electron Transfer Dynamics of a C70-Encapsulated Bisporphyrin Covalent Organic Polyhedron in a Low-Dielectric Medium. *J. Am. Chem. Soc.* **2017**, *139*, 4286−4289.
(38) Takai, A.; Chkounda, M.; Eggenspiller, A.; Gros, C. P.; Lachkar, M.; Barbe, J.-M.; Fukuzumi, S. Efficient Photoinduced Electron Transfer in a Porphyrin Tripod−Fullerene Supramolecular Complex via $\pi-\pi$ Interactions in Nonpolar Media. *J. Am. Chem. Soc.* **2010**, *132*, 4477−4489.
(39) Kieran, A. L.; Pascu, S. I.; Jarrosson, T.; Sanders, J. K. M. Inclusion of C60 into an adjustable porphyrin dimer generated by dynamic disulfide chemistry. *Chem. Commun.* **2005**, 1276−1278.
(40) Sension, R. J.; Szarka, A. Z.; Smith, G. R.; Hochstrasser, R. M. Ultrafast photoinduced electron transfer to C60. *Chem. Phys. Lett.* **1991**, *185*, 179−183.
(41) Guldi, D. M.; Hartmut, H.; Eberhard, J.; Klaus Dieter, A. Redox processes and alkylation reactions of fullerene C60 as studied by pulse radiolysis. *J. Phys. Chem. A* **1993**, *97*, 11258−11264.
(42) D'Souza, F.; Suresh, G.; Zandler, M. E.; Arkady, K.; El-Khouly, M. E.; Fujitsuka, M.; Ito, O. Studies on Covalently Linked Porphyrin−C60 Dyads: Stabilization of Charge-Separated States by Axial Coordination. *J. Phys. Chem. A* **2002**, *106*, 12393−12404.
(43) Karolczak, J.; Kowalska, D.; Lukaszewicz, A.; Maciejewski, A.; Steer, R. P. Photophysical Studies of Porphyrins and Metalloporphyrins: Accurate Measurements of Fluorescence Spectra and Fluorescence Quantum Yields for Soret Band Excitation of Zinc Tetraphenylporphyrin. *J. Phys. Chem. A* **2004**, *108*, 4570−4575.
(44) Schuster, D. I.; Cheng, P.; Jarowski, P. D.; Guldi, D. M.; Luo, C.; Echegoyen, L.; Pyo, S.; Holzwarth, A. R.; Braslavsky, S. E.; Williams, R. M.; Klihm, G. Design, Synthesis, and Photophysical Studies of a Porphyrin-Fullerene Dyad with Parachute Topology; Charge Recombination in the Marcus Inverted Region. *J. Am. Chem. Soc.* **2004**, *126*, 7257−7270.
(45) Asano, N.; Uemura, S.; Kinugawa, T.; Akasaka, H.; Mizutani, T. Synthesis of Biladienone and Bilatrienone by Coupled Oxidation of Tetraarylporphyrins. *J. Org. Chem.* **2007**, *72*, 5320−5326.
(46) Adler, A. D.; Longo, F. R.; Finarelli, J. D.; Goldmacher, J.; Assour, J.; Korsakoff, L. A simplified synthesis for meso-tetraphenylporphine. *J. Org. Chem.* **1967**, *32*, 476.
(47) Adler, A. D.; Longo, F. R.; Kampas, F.; Kim, J. On the preparation of metalloporphyrins. *J. Inorg. Nucl. Chem.* **1970**, *32*, 2443−2445.
(48) Bonnett, B. L.; Smith, E. D.; De La Garza, M.; Cai, M.; Haag, J. V.; Serrano, J. M.; Cornell, H. D.; Gibbons, B.; Martin, S. M.; Morris, A. J. PCN-222 Metal−Organic Framework Nanoparticles with Tunable Pore Size for Nanocomposite Reverse Osmosis Membranes. *ACS Appl. Mater. Interfaces* **2020**, *12*, 15765−15773.
(49) Zhang, L.; Shi, X.; Zhang, Z.; Kuchel, R. P.; Namivandi-Zangeneh, R.; Corrigan, N.; Jung, K.; Liang, K.; Boyer, C. Porphyrinic Zirconium Metal−Organic Frameworks (MOFs) as Heterogeneous Photocatalysts for PET-RAFT Polymerization and Stereolithography. *Angew. Chem., Int. Ed.* **2021**, *60*, 5489−5496.
(50) Rouquerol, J.; Llewellyn, P.; Rouquerol, F. Is the BET Equation Applicable to Microporous Adsorbents?. In *Studies in Surface Science and Catalysis*; Llewellyn, P. L.; Rodriquez-Reinoso, F.; Rouqerol, J.; Seaton, N., Eds.; Elsevier, 2007; Vol. *160*, pp 49−56.
(51) Ásgeirsson, V.; Benedikt Orri, B.; Ragnar, B.; Ute, B.; Frank, N.; Christoph, R.; Hannes, J. Nudged Elastic Band Method for Molecular Reactions Using Energy-Weighted Springs Combined with Eigenvector Following. *J. Chem. Theory Comput.* **2021**, *17*, 4929−4945.
(52) Mills, G.; Jónsson, H.; Schenter, G. K. Reversible work transition state theory: application to dissociative adsorption of hydrogen. *Surf. Sci.* **1995**, *324*, 305−337.
(53) Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific, 1998; pp 385−404.
(54) Bannwarth, C.; Sebastian, E.; Stefan, G. 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.* **2019**, *15*, 1652−1671.
(55) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73−78.
(56) Neese, F. Software update: The ORCA program system—Version 5.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2022**, *12*, No. e1606, DOI: 10.1002/wcms.1606.
(57) Stefan, G.; Jan Gerit, B.; Christoph, B.; Andreas, H. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.* **2015**, *143*, No. 054107.
(58) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. 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* **2009**, *113*, 6378−6396.
(59) Grimme, S.; Rose, T. mcGFN-FF: an accurate force field for optimization and energetic screening of molecular crystals. *Z. Naturforsch., B: J. Chem. Sci.* **2024**, *79*, 191−200.
(60) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 2nd ed.; Springer: New York, NY, 2004.
(61) Chukharev, V.; Tkachenko, N. V.; Efimov, A.; Guldi, D. M.; Hirsch, A.; Scheloske, M.; Lemmetyinen, H. Tuning the Ground-State and Excited-State Interchromophore Interactions in Porphyrin−Fullerene π-Stacks. *J. Phys. Chem. B* **2004**, *108*, 16377−16385.
(62) Wang, A.; Barcus, K.; Cohen, S. M. Quantifying Ligand Binding to the Surface of Metal−Organic Frameworks. *J. Am. Chem. Soc.* **2023**, *145*, 16821−16827.
(63) Lahanas, N.; Kucheryavy, P.; Lalancette, R. A.; Lockard, J. V. Crystallographic identification of a series of manganese porphyrin complexes with nitrogenous bases. *Acta Crystallogr., Sect. C: Struct. Chem.* **2019**, *75*, 304−312.
(64) Sun, D.; Tham, F. S.; Reed, C. A.; Chaker, L.; Boyd, P. D. W. Supramolecular Fullerene-Porphyrin Chemistry. Fullerene Complexation by Metalated "Jaws Porphyrin" Hosts. *J. Am. Chem. Soc.* **2002**, *124*, 6604−6612.
(65) Gil-Ramírez, G.; Karlen, S. D.; Atsuomi, S.; Kyriakos, P.; Yasuhiro, I.; Briggs, G. A. D.; Morton, J. J. L.; Anderson, H. L. A Cyclic Porphyrin Trimer as a Receptor for Fullerenes. *Org. Lett.* **2010**, *12*, 3544−3547.
(66) Kamimura, T.; Ohkubo, K.; Kawashima, Y.; Ozako, S.; Sakaguchi, K.-i.; Fukuzumi, S.; Tani, F. Long-Lived Photoinduced Charge Separation in Inclusion Complexes Composed of a Phenothiazine-Bridged Cyclic Porphyrin Dimer and Fullerenes. *J. Phys. Chem. C* **2015**, *119*, 25634−25650.
(67) Guldi, D. M.; Maurizio, P. Excited-State Properties of C60 Fullerene Derivatives. *Acc. Chem. Res.* **2000**, *33*, 695−703.

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

(68) Rogers, J. E.; Nguyen, K. A.; Hufnagle, D. C.; McLean, D. G.; Su, W.; Gossett, K. M.; Burke, A. R.; Vinogradov, S. A.; Pachter, R.; Fleitz, P. A. Observation and Interpretation of Annulated Porphyrins: Studies on the Photophysical Properties of meso-Tetraphenylmetalloporphyrins. *J. Phys. Chem. A* **2003**, *107*, 11331−11339.
(69) Kesti, T. J.; Tkachenko, N. V.; Visa, V.; Hiroko, Y.; Hiroshi, I.; Shunichi, F.; Helge, L. Exciplex Intermediates in Photoinduced Electron Transfer of Porphyrin−Fullerene Dyads. *J. Am. Chem. Soc.* **2002**, *124*, 8067−8077.
(70) Al-Subi, A. H.; Niemi, M.; Tkachenko, N. V.; Lemmetyinen, H. Effect of Anion Ligation on Electron Transfer of Double-Linked Zinc Porphyrin−Fullerene Dyad. *J. Phys. Chem. A* **2011**, *115*, 3263−3271.
(71) Tkachenko, N. V.; Helge, L.; Junko, S.; Kei, O.; Tomoo, S.; Hiroshi, I.; Shunichi, F. Ultrafast Photodynamics of Exciplex Formation and Photoinduced Electron Transfer in Porphyrin−Fullerene Dyads Linked at Close Proximity. *J. Phys. Chem. A* **2003**, *107*, 8834−8844.
(72) Hou, L.; Zhang, X.; Pijper, T. C.; Browne, W. R.; Feringa, B. L. Reversible Photochemical Control of Singlet Oxygen Generation Using Diarylethene Photochromic Switches. *J. Am. Chem. Soc.* **2014**, *136*, 910−913.
(73) Goldoni, A.; Cepek, C.; Modesti, S. The surface triplet exciton of C60(111). *Synth. Met.* **1996**, *77*, 189−194.
(74) Imahori, H.; El-Khouly, M. E.; Mamoru, F.; Osamu, I.; Yoshiteru, S.; Shunichi, F. Solvent Dependence of Charge Separation and Charge Recombination Rates in Porphyrin−Fullerene Dyad. *J. Phys. Chem. A* **2001**, *105*, 325−332.
(75) Imahori, H.; Kiyoshi, H.; Masanori, A.; Tsuyoshi, A.; Seiji, T.; Tadashi, O.; Masahiro, S.; Yoshiteru, S. Linkage and Solvent Dependence of Photoinduced Electron Transfer in Zincporphyrin-C60 Dyads. *J. Am. Chem. Soc.* **1996**, *118*, 11771−11782.
(76) Armaroli, N.; Marconi, G.; Echegoyen, L.; Bourgeois, J.-P.; Diederich, F. Charge-Transfer Interactions in Face-to-Face Porphyrin-Fullerene Systems: Solvent-Dependent Luminescence in the Infrared Spectral Region. *Chem. - Eur. J.* **2000**, *6*, 1629−1645.
(77) El-Khouly, M. E.; Ito, O.; Smith, P. M.; D'Souza, F. Intermolecular and supramolecular photoinduced electron transfer processes of fullerene−porphyrin/phthalocyanine systems. *J. Photochem. Photobiol., C* **2004**, *5*, 79−104.
(78) Nakamura, T.; Ikemoto, J.-y.; Fujitsuka, M.; Araki, Y.; Ito, O.; Takimiya, K.; Aso, Y.; Otsubo, T. Control of Photoinduced Energy- and Electron-Transfer Steps in Zinc Porphyrin−Oligothiophene−Fullerene Linked Triads with Solvent Polarity. *J. Phys. Chem. B* **2005**, *109*, 14365−14374.
(79) Fujitsuka, M.; Ito, O.; Yamashiro, T.; Aso, Y.; Otsubo, T. Solvent Polarity Dependence of Photoinduced Charge Separation in a Tetrathiophene-C60 Dyad Studied by Pico- and Nanosecond Laser Flash Photolysis in the Near-IR Region. *J. Phys. Chem. A* **2000**, *104*, 4876−4881.

https://doi.org/10.1021/acs.jpcc.5c00161
J. Phys. Chem. C XXXX, XXX, XXX−XXX

APPENDIX E

Reaction-Free Energies for Complexation of Carbohydrates by Tweezer Diboronic Acids

Gustavo Adolfo Lara-Cruz[1] , Thomas Rose[2], Stefan Grimme[2], Andres Jaramillo-Botero[3]

Received: July 18, 2024
Published online: September 16, 2024

Reprinted in Appendix E with permission from[4] G. A. Lara-Cruz, T. Rose, S. Grimme, and A. Jaramillo-Botero, *Reaction-Free Energies for Complexation of Carbohydrates by Tweezer Diboronic Acids*. J. Phys. Chem. B **128** (2024) 9213–9223

DOI: https://doi.org/10.1021/acs.jpcb.4c04846

Own contributions

- Contributions to workflow development
- Discussion of calculation setup
- Interpretation of the results
- Writing and revising parts of the manuscript

[1] iOMICAS Research Institute, Pontificia Universidad Javeriana, Calle 17 # 121B-155, Santiago de Cali, Valle del Cauca 760031, Colombia
[2] Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Beringstraße 4, Bonn 53115, Germany
[3] Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States; iOMICAS Research Institute, Pontificia Universidad Javeriana, Calle 17 # 121B-155, Santiago de Cali, Valle del Cauca 760031, Colombia
[4] Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Reaction-Free Energies for Complexation of Carbohydrates by Tweezer Diboronic Acids

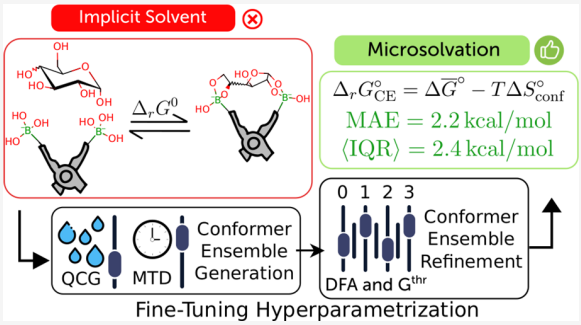
Gustavo Adolfo Lara-Cruz, Thomas Rose, Stefan Grimme, and Andres Jaramillo-Botero*

Cite This: J. Phys. Chem. B 2024, 128, 9213−9223

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The accurate calculation of reaction-free energies (Δ_rG°) for diboronic acids and carbohydrates is challenging due to reactant flexibility and strong solute−solvent interactions. In this study, these challenges are addressed with a semiautomatic workflow based on quantum chemistry methods to calculate conformational free energies, generate microsolvated solute structural ensembles, and compute Δ_rG°. Workflow parameters were optimized for accuracy and precision while controlling computational costs. We assessed the accuracy by studying three reactions of diboronic acids with glucose and galactose, finding that the conformational entropy contributes significantly (by 3−5 kcal/mol at room temperature). Explicit solvent molecules improve the computed Δ_rG° accuracy by about 4 kcal/mol compared to experimental data, though using 13 or more water molecules reduced precision and increased computational overhead. After fine-tuning, the workflow demonstrated remarkable accuracy, with an absolute error of about 2 kcal/mol compared to experimental Δ_rG° and an average interquartile range of 2.4 kcal/mol. These results highlight the workflow's potential for designing and screening tweezer-like ligands with tailored selectivity for various carbohydrates.

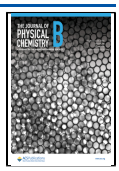


INTRODUCTION

In plants, carbohydrates play key roles, serving as the primary energy source for metabolic processes,[1] acting as signaling molecules that initiate cascades of response reactions,[1,2] and providing carbon atoms essential for synthesizing polysaccharides needed for tissue regeneration.[1] For these reasons, there is a need for the development of sensors with the capability to selectively detect and quantify carbohydrates under in vivo conditions. These sensors could play a crucial role in breeding new agricultural plant varieties more resilient to the increasing stress conditions due to climate change and global warming. In this context, the reactions involving phenylboronic acids (PBA) and carbohydrates to produce boron diester compounds align well with several prerequisites for creating sensors to quantify carbohydrates in living systems.[2−5] These reactions are reversible under normal physiological conditions, PBA compounds are biocompatible,[4,5] and the binding process is detectable through a wide diversity of signal transduction methods,[6] making PBA highly attractive candidates for designing sensors capable of specifically detecting carbohydrates in aqueous environments.

Diboronic acids (DBA), as shown in Figure 1, are examples of PBA-based molecules successfully designed for selective detection and quantification of glucose (Glc) and galactose (Gal). Depending on the structural details of the DBA, these molecules bind to glucose and galactose via four B−O−C bonds with the hydroxyl groups in 1,2 and 1,3 positions, forming cyclic boron diesters of five and six members, respectively.[7−10] In this way, the DBA molecules behave as tweezer-like ligands reacting preferentially with glucose and galactose, despite the higher reactivity of PBA toward fructose.[4,8] The relative distance and orientation of boronic groups ($R-B^-(OH)_3$) and the rigidity of the DBA-tweezer ligand are crucial factors in achieving high selectivity toward glucose.[4,8,11] Although optimal control of these factors is limited by synthetic accessibility and chemical intuition for the design of the DBA ligands, computational chemistry is a valuable tool to alleviate these limitations, as demonstrated by Yang and co-workers,[11] who found the optimal positioning of the two PBA molecules to bind glucose by means of geometry optimizations, designing one of the best performing DBA for glucose.[11] Even so, Yang's design[11] only used a "structure-based" methodology, which does not guarantee the covalent binding between the reactant molecules, if applied to the design of DBAs for other carbohydrates. Therefore, additional descriptors for the covalent binding processes, such as the reaction-free energy (Δ_rG°),[12] are required to complement the

Received: July 18, 2024
Revised: September 2, 2024
Accepted: September 5, 2024
Published: September 16, 2024



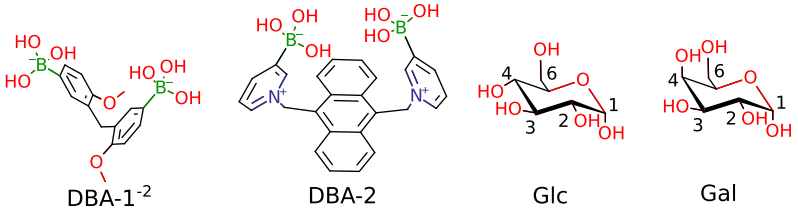


Figure 1. Structures of the diboronic acids (DBAs) **1** and **2** and the monosaccharides glucose and galactose used in this work for the ab initio calculation of reaction-free energies $\Delta_r G^\circ$.

"structure-based" methodologies for the computational design of DBA ligands for carbohydrates.

Carbohydrates are highly solvated and flexible molecules in solution,[13] and DBAs reported in the literature are mostly flexible molecules as well.[4,8] Thus, several conformations, for reactants and products, are needed for accurate calculations of the reaction-free energies $\Delta_r G^\circ$ in these systems.[14,15] Accurate calculation of the solvation contribution becomes difficult when explicit solute molecules lead to a significant change in the investigated property, as commonly used implicit models are no longer viable. Furthermore, modern workflows for reactions where two flexible molecules covalently bond to a single molecule demand the investigation of the change in entropy.[16,17]

In this work, we selected the reactions **DBA-1**$^{-2}$ + Glc, **DBA-1**$^{-2}$ + Gal, and **DBA-2** + Glc as example systems to present a state-of-the-art workflow for the calculation of reaction-free energies. The systems studied in this work show all these aforementioned challenges, and the well-established experimental information about these reactions like the structures of reactants and products, alongside reported equilibrium constants $(K(T))$,[7−10] allow us to perform a rigorous evaluation of the workflow here presented.

This work is organized as follows: The Theory section explains the general approach to calculate free energies of conformer ensembles. In the section Computational Details and Experimental Data, we present the methodology for the construction of the microsolvated systems, the calculation of the reaction-free energies, the sampling of the conformational space, and the generation of the final conformer ensemble. In the Results and Discussion section, reaction-free energies of representative test systems are reported and parameters that affect the results are discussed. Finally, the Conclusion summarize key findings on how to efficiently calculate reaction-free energies of flexible systems that require microsolvation.

THEORY

For a reaction in solution with a standard state "°" of 1 M, the experimental $\Delta_r G^\circ$ is calculated from the observed equilibrium reaction constant K_{obs}

$$\Delta_r G^\circ = -RT\ln(K_{obs}) = -RT\ln\left(\frac{[C][D]}{[A][B]}\right) \quad (1)$$

where R is the gas constant, and T is the absolute temperature of the system at equilibrium. For the chemical reaction $A + B \rightleftharpoons C + D$, the equilibrium constant can be expressed with the molar concentrations $[A]$, $[B]$, $[C]$, and $[D]$. Theoretically, the reaction-free energy can be computed according to

$$\Delta_r G^\circ = G^\circ(C) + G^\circ(D) - G^\circ(A) - G^\circ(B) \quad (2)$$

from the respective absolute free energies G° of the products and reactants. For a given molecule X, the free energy

$$G^\circ(X) = E_{gas}(X) + G^\circ_{trv}(X) + \delta G_{solv}(X) \quad (3)$$

is calculated from the electronic energy in the gas phase $E_{gas}(X)$, the free energy from the translational, rotational, and vibrational motions $G^\circ_{trv}(X)$ (including the zero point energy) accessible for the molecule X at temperature T, and the solvation free energy $\delta G_{solv}(X)$. Hence, accurate computations by this brute force approach require high accuracy for all the individually computed (free) energy components. Moreover, for flexible molecules, the calculation of $G^\circ(X)$ requires considering the thermodynamic average over the (at best) complete conformer ensemble (CE). If the flexibility of any of the reactants is likely to change due to the reaction, the conformational entropy contribution $-TS^\circ_{conf}$ should be included in the molecular Gibbs free energy for the CE[18]

$$G^\circ_{CE}(X) = \sum_i^{N_{conf}} p_i \cdot G^\circ_i(X) - TS^\circ_{conf}(X) = \overline{G}^\circ(X) - TS^\circ_{conf}(X) \quad (4)$$

Here, $\overline{G}^\circ(X)$ is the weighted Boltzmann average over the CE, where the weight for conformation i

$$p_i = \frac{e^{-G^\circ_i(X)/kT}}{\sum_j^{N_{conf}} e^{-G^\circ_j(X)/kT}} \quad (5)$$

is calculated with the corresponding molecular free energy $G^\circ_i(X)$, the Boltzmann constant k and temperature T. j is an index that runs over the conformers (N_{conf}) used to calculate the weights p_i. By substituting eq 3 into eq 4, the molecular free energy for a CE

$$G^\circ_{CE}(X) = \overline{E}_{gas}(X) + \overline{G}^\circ_{trv}(X) + \overline{\delta G}_{solv}(X) - TS^\circ_{conf}(X) \quad (6)$$

is obtained. Finally, the reaction-free energy (eq 2) considering the CE for reactants and products is calculated by the equation

$$\Delta_r G^\circ_{CE} = \Delta\overline{G}^\circ - T\Delta S^\circ_{conf} \quad (7)$$

where ΔS°_{conf} accounts for the conformational entropy changes in the reaction. The $G^\circ_{CE}(X)$ equations used to calculate $\Delta_r G^\circ_{CE}$ for the case of microsolvated reactions are presented in eqs S2−S6 in the Supporting Information.

COMPUTATIONAL DETAILS AND EXPERIMENTAL DATA

Methods and Workflow. The conformational space of reactants and products for the reactions are sampled with the Conformer-Rotamer Ensemble Sampling Tool (CREST)[19] (version 2.11.2) while using the included Quantum Cluster

https://doi.org/10.1021/acs.jpcb.4c04846
J. Phys. Chem. B 2024, 128, 9213−9223

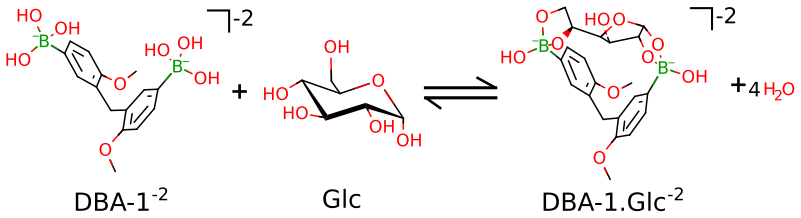


Figure 2. Reaction of **DBA-1** with glucose in aqueous alkaline solution at pH = 11.3.[27]

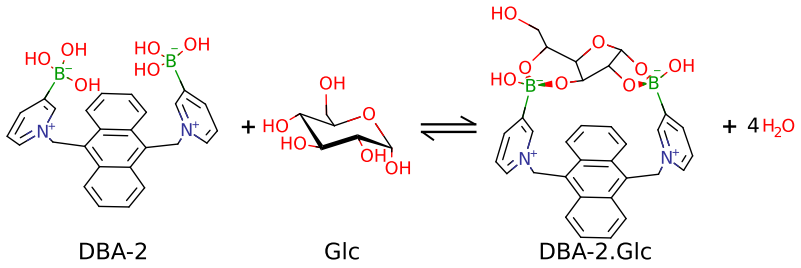


Figure 3. Reaction of **DBA-2** with glucose in aqueous solution at pH = 7.4.[9]

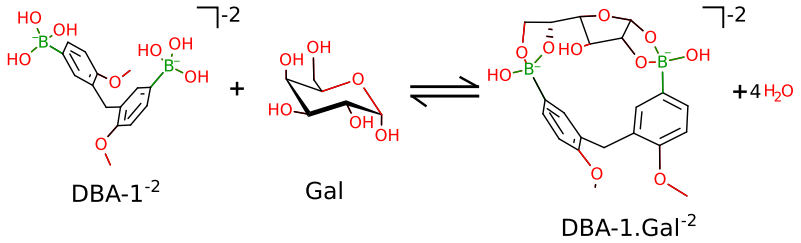


Figure 4. Proposed structures for the reaction of **DBA-1**$^{-2}$ with galactose at the conditions reported in reference in 27 (aqueous solution at pH = 11.3).

Growing method[20] (QCG) for microsolvation with water. The meta-dynamics length was set to 80 ps for all calculations to ensure a well-converged sampling of the conformational space for these complicated systems. Subsequently, a refined energetic sorting of the obtained ensemble is calculated with the Commandline Energetic SOrting (CENSO) tool[18] (version 1.2.0). Here, the refinement can be divided into parts 0−3, where all parts allow the individual setup of the used density-functional approximation (DFA) for the electronic energy and the threshold for sorting out the high-lying conformers. In part 0, a cheap prescreening is performed based on the electronic energy from a single-point (SP) calculation. Further prescreening in part 1 uses free energies by including thermostatistical contributions[21] with the GFN2-*x*TB[ALPB] method, which is an extended semiempirical tight binding approach that includes anisotropic second-order density fluctuation effects via short-range damped interactions of cumulative atomic multipole moments and an analytical linearized Poisson−Boltzmann (ALPB) model to account for solvation effects. In part 2, the remaining structures are optimized at the chosen level of theory. Finally, part 3 performs single-point calculations on the previously optimized structures. The choice of functional and sorting thresholds for the different parts will be discussed in the results section. While implicit solvation is included in part 0 via the generalized Born model augmented with the hydrophobic solvent accessible surface area (GBSA),[22−24] the succeeding parts use the SMD solvation model.[25] Both, explicit and implicit water models are used in the conformer ensemble generation with QCG.

Since the calculation of molecular free energies for flexible systems based on meta-dynamics involves randomness, multiple trials (N) are needed to adequately sample the value distribution of $G^{\circ}_{CE}(X)$. With $G^{\circ}_{CE,i}(X)$ being one of the sampled molecular free energies, for a particular permutation of the group $\{i, j, k, l\}$ (corresponding to the index $x \in \{1, 2, 3, ..., N^4\}$) among the N-sampled molecular free energies for the reactants and products, the eq 2 turns into

$$\Delta_r G^{\circ}_{CE,x} = G^{\circ}_{CE,k}(C) + G^{\circ}_{CE,l}(D) - G^{\circ}_{CE,i}(A) - G^{\circ}_{CE,j}(B) \quad (8)$$

with $i, j, k, l \in \{1, 2, 3, ..., N\}$. As the choice of $\{i, j, k, l\}$ indices is arbitrary, we defined the reaction-free energy

$$\Delta_r G^{\circ}_{CE} = \mathrm{med}(\{\Delta_r G^{\circ}_{CE,x}\}) \quad (9)$$

as the median of the distribution of the set of values $\Delta_r G^{\circ}_{CE,x}$, ordered from smallest to greatest and calculated for all possible combinations of $\{i, j, k, l\}$ indices. Now, the precision of the workflow can be defined as the interquartile range (IQR) of this distribution.

The thermostatistical contributions $G^{\circ}_{trv}(X)$ are calculated with semiempirical GFNn-xTB methods, and the implicit solvation method ALPB[26] is used for calculation of solvation free energy $\delta G_{solv}(X)$. While $\overline{E}_{gas}$, $\overline{G}^{\circ}_{trv}$, and $\delta\overline{G}_{solv}$ from eq 6 are automatically calculated by CENSO with the given settings. The conformational entropy S°_{conf} is determined by the ensemble entropy calculation from CREST based on GFN2-xTB generated ensembles with the GBSA implicit solvent model for water.

https://doi.org/10.1021/acs.jpcb.4c04846
J. Phys. Chem. B 2024, 128, 9213−9223

Table 1. Acidity Constants (pK_a) and the Corresponding Solution's pH Reported for the Calculation of the Observed Equilibrium Constants K_{obs} (L mol^{-1}) for the Reactions of DBA-1^{-2} + Glc, DBA-2 + Glc, and DBA-1^{-2} + Gal[a]

	pK_{a1}	pK_{a2}	reaction	pH	α_{Sugar}	α_{DBA}	K_{obs}	$\Delta_rG^\circ_{Exp}$
DBA-1^{-2}	8.5[b]	9.5[b]	**DBA-1**$^{-1}$ + Glc	11.3	0.89	0.98	19,000[c]	−5.8
DBA-2	3.7[9]	4.7[9]	**DBA-2** + Glc	7.4	1.0	1.0	2510 ± 1.2[d]	− 4.6 ± 0.1
Glc[28,29]	12.1	13.9					10,000[e]	−5.5
Gal	∼12.0		**DBA-1**$^{-1}$ + Gal	11.3	∼0.90	0.98	2200[c]	−4.2

[a]α is the speciation fraction of the corresponding reactant at the given reaction pH's. The experimental reaction-free energies $\Delta_rG^\circ_{Exp}$ (kcal/mol) were calculated with eq 1. [b]See the Supporting Information for the calculation of these values. [c]K_{obs} calculated with circular dichroism spectroscopy using the Benesi−Hildebrand method, and assuming a reaction yield of 100% for the products **DBA-1.Glc**$^{-2}$ and **DBA-1.Gal**$^{-2}$ in references 30 and 31. [d]K_{obs} from potentiometric and fluorescence titrations.[9] [e]K_{obs} from NMR spectroscopy.[9]

Table 2. Settings for CENSO Calculations with Assigned Letters for Referencing[a]

set.	part 0	part 1	part 2	part 3	G_0^{thr} [kcal/mol]	G_1^{thr} [kcal/mol]	G_2^{thr} [kcal/mol]
A	B97-D3	r^{2}SCAN-3c	r^{2}SCAN-3c		6	4	2.5
B	B97-D3	r^{2}SCAN-3c	r^{2}SCAN-3c		10	6	3
C	B97-D3	r^{2}SCAN-3c	r^{2}SCAN-3c		15	10	6
D	HF-3c	HF-3c	HF-3c		15	10	6
E	HF-3c	B97-D3	B97-D3		15	10	6
F	B97-D3	HF-3c	HF-3c	r^{2}SCAN-3c	15	10	6
G	B97-D3	B97-D3	B97-D3		15	10	6

[a]Columns part 0−3 show the functional used in the cheap prescreening, prescreening, optimization, and refinement parts of the CENSO program. The energy windows used for screening out are given by G_i^{thr} with the index corresponding to the different parts.

Experimental Reference Data. To assess the accuracy of our workflow, we selected the reactions **DBA-1**$^{-2}$ + Glc (Figure 2), **DBA-2** + Glc (Figure 3), and **DBA-1**$^{-2}$ + Gal (Figure 4) because the structure of reactants and products has been correctly characterized. Furthermore, the structures of the products in the reactions **DBA-1**$^{-2}$ + Glc and **DBA-2** + Glc have been elucidated by analyzing the NMR spectra of proton (^{1}H) and carbon (^{13}C) nuclei, and by determining dipolar couplings (J_{C-C} and J_{H-H}) to accurately identify the hydroxyl groups of glucose bonded to the boron atoms in the DBAs.[9,27]

In all our calculations we used the α-glucopyranose and α-galactopyranose isomers because the 1,2-hydroxyls in cis position (see Figure 1) make these isomers the most reactive toward **DBA-1**$^{-2}$ and **DBA-2**. Meanwhile, in the β-isomer, the 1,2-diols are in trans positions which do not form a stable five-member diboron ester ring.[8]

The reactions **DBA-1**$^{-2}$ + Glc, **DBA-2** + Glc, and **DBA-1**$^{-2}$ + Gal were performed at conditions where the solution's pH > pK_as of the DBAs (see Table 1). In the Supporting Information, we show that the equilibrium side reactions for reactants and products are unimportant at the corresponding reaction's pH. Then, the equilibrium reaction constants K_{obs} and the experimental reaction-free energies $\Delta_rG^\circ_{Exp}$ can be safely assigned to the chemical processes shown in Figures 2−4. Consequently, the reference reaction-free energies $\Delta_rG^\circ_{Exp}$ are calculated directly from the corresponding experimental K_{obs} using eq 1 and are presented in Table 1. Hereafter, we will also use the symbols $\Delta_rG^\circ_{CD}$, $\Delta_rG^\circ_{PT}$, and $\Delta_rG^\circ_{NMR}$ to refer to the experimental reaction-free energies ($\Delta_rG^\circ_{Exp}$) calculated from the equilibrium constants (K_{obs}) from circular dichroism (CD), potentiometric titrations (PT), and nuclear magnetic resonance (NMR) experiments, respectively.

Despite the reported value[30,31] of the equilibrium constant K_{obs} for the reaction of **DBA-1**$^{-2}$ + Gal (Figure 4), the structure of the product **DBA-1.Gal**$^{-2}$ has not yet been characterized. Nevertheless, the structural and chemical similarities between glucose and galactose allow us to propose a reasonable structure for **DBA-1.Gal**$^{-2}$, as shown in Figure 4, for our calculations of the reaction-free energy. Additionally, the reaction conditions for **DBA-1**$^{-2}$ + Gal (pH = 11.3) allow us to discard the side equilibrium reactions and use similar speciation fractions α as in the case of **DBA-1**$^{-2}$ + Glc reaction, as shown in Table 1.

Finally, the K_{obs} were calculated with diluted solutions. For the **DBA-1**$^{-2}$ + Glc and **DBA-1**$^{-2}$ + Gal reactions,[30,31] the initial molar fractions of the reactants range from 4.5×10^{-5} to 7.2×10^{-5}. In the case of **DBA-2** + Glc reaction,[9] the initial molar fractions range from 4.5×10^{-7} to 4.5×10^{-3}. At equilibrium, these molar fractions decrease further due to the production of four water molecules in the reaction. Thus, the systems studied in this work (Figures 2−4) can be closely represented by our computational workflow. Therefore, solute−solute interactions are expected to have minimal weight in the discrepancies between our calculated reaction-free energies ($\Delta_rG^\circ_{CE}$) and experimental results ($\Delta_rG^\circ_{Exp}$).

Systematic Screening of Computational Parameters. The technical parameters of the methods used in this work (QCG, CREST, and CENSO) form a vast space, hence identifying the optimal combination for calculating the reaction-free energies $\Delta_rG^\circ_{CE}$ is a complex endeavor. Additionally, the determination of the number of water molecules required to refine the representation of solute−solvent interactions poses another parameter to optimize. Nevertheless, the computational cost increases and the conformational sampling becomes more complicated with the addition of explicit water molecules for microsolvation. Hence, we performed a parameter optimization step for the calculation of the reaction-free energies $\Delta_rG^\circ_{CE}$ for the reactions of **DBA-1**$^{-2}$ + Glc (Figure 2), **DBA-2** + Glc (Figure 3), and **DBA-1**$^{-2}$ + Gal (Figure 4).

The parameter search comprised two phases. The first phase involved optimizing the CENSO parameters to compute the energy thresholds needed to sort out energetically unfavorable conformers. Table 2 tabulates the combinations of theory

https://doi.org/10.1021/acs.jpcb.4c04846
J. Phys. Chem. B 2024, 128, 9213−9223

levels and the corresponding energy thresholds (G_i^{thr}) used in the first phase of the parameter search. Each combination was selected taking into account that the computational cost increases when moving forward in the CE refinement with CENSO. That excludes settings D where we were interested in the performance of the HF-3c method for optimization. For the first phase of the parameter search, we only use the reaction of the smaller **DBA-1**$^{-2}$ with glucose (Figure 2) to reduce computational costs during the parameter search.

The 3c methods were chosen due to their good cost-performance ratio, needed for screening a large amount of highly flexible systems[32−35] while using a noncomposite functional like B97-D3 allows increasing the basis set size with the CENSO parts. The B97-D3 functional is used with the AO basis set def2-SV(P) in part 0, and the def2-TZVP(-f) AO basis in parts 1 and 2. Concerning this, the 3c methods have been designed to accurately calculate noncovalent interactions using a tailor-made basis set. The accurate description of noncovalent interactions is especially important when microsolvation is considered. For these reasons, different combinations of the methods HF-3c, B97-D3, and r^{2}SCAN-3c are evaluated.

The second phase of the parameter search screens the number of water molecules necessary to achieve accurate free energies $\Delta_r G^\circ_{CE}$ for the reactions of **DBA-1**$^{-2}$ and **DBA-2** with glucose, using the best-performing combination of DFA methods from the first phase. The number of water molecules for microsolvation was gradually increased in all the reactants and products, to assess the precision, accuracy, and robustness of the DFA methods in CENSO for calculating $\Delta_r G^\circ_{CE}$. Although most of the QCG and CREST parameters were left as default, all the meta-dynamics simulations for conformational sampling were conducted for 80 ps. To avoid any bias in the conformational sampling, a new CE was generated for all the reactants and products for each calculation of the reaction-free energy.

RESULTS AND DISCUSSION

Conformational Entropy Contribution. Given the substantial reduction in molecular flexibility from the reactants (**DBA-1**$^{-2}$, **DBA-2**, Glc, and Gal) to the bound complexes (**DBA-1.Glc**$^{-2}$, **DBA-2.Glc**, and **DBA-1.Gal**$^{-2}$), the contribution of conformational entropy to the reaction-free energy is likely to be significant and should be taken into account. To investigate the impact of entropic penalties in our specific set of systems, the entropic contribution $-T\Delta S^\circ_{conf}$ is calculated from the change in entropy ΔS°_{conf} from reactants to products at 298.15 K. The statistical noise in the ensemble generation with CREST is taken into account by averaging over three calculations for each system.

The entropic contribution to the free energy for the **DBA-1**$^{-2}$ + Glc, **DBA-2** + Glc, and for **DBA-1**$^{-2}$ + Gal reaction are presented in Table 3. The negative $T\Delta S^\circ_{conf}$ values qualitatively agree with the closure of two rings via the formation of four boronate ester (B−O−C) bonds, depicted in the Figures 2−4, and the locks of the dihedral rotations of the PBA groups. Additionally, the reduction in conformational entropy $T\Delta S^\circ_{conf}$ is comparatively lower for **DBA-2** + Glc than for **DBA-1**$^{-2}$ + Glc and **DBA-1**$^{-2}$ + Gal reactions. These differences arise because, in **DBA-2**, the rotations of the PBA groups are dampened by the anthracene ring, lowering the S°_{conf} for **DBA-2** compared with **DBA-1**$^{-2}$, in which the PBA groups are freer to rotate.

Table 3. Calculated Conformational Entropy Contribution for the Reaction Leading to the Given Complex[a]

reaction	$T\Delta S^\circ_{conf}$ [kcal/mol]
DBA-1^{-2} + Glc ⇌ **DBA-1.Glc**$^{-2}$	−4.82
DBA-2 + Glc ⇌ **DBA-2.Glc**	−3.31
DBA-1^{-2} + Gal ⇌ **DBA-1.Gal**$^{-2}$	−4.11

[a]The four water molecules produced after the reaction were removed from the reaction equations because they did not contribute to the conformational entropy change.

These $T\Delta S^\circ_{conf}$ values introduce a shift of the reaction-free energies $\Delta_r G^\circ_{CE}$, significantly improving the accuracy (see below). Generally, it is advisable to analyze how the number of relevant conformers changes between the reactants and the products. Practically, the number of conformers needed to recover 90% of the Boltzmann weight can be used as a measure of flexibility and can be obtained from a CREST calculation. If this measure changes significantly during the reaction, the conformational entropy should be considered in the free energy calculation, as is exemplified by the reactions reported in this work. Comparing the number of relevant conformers for the Boltzmann average after the CENSO refinement, the product conformers decrease by approximately a factor of ∼12 compared with the reactant conformers, no matter the level of theory used for the final refinement with CENSO. These results are reported in the first rows ($(H_2O)_0$) of Tables S4 and S6 in the Supporting Information. Hereafter, the reported reaction-free energies $\Delta_r G^\circ_{CE}$ (including those in the Supporting Information) incorporate the corresponding conformational entropy changes $T\Delta S^\circ_{conf}$ from Table 3, as given by eq 7.

Reaction-Free Energies ($\Delta_r G°$) with the Implicit Solvation Model. Implicit solvation models offer an efficient means of capturing the interactions between solute and solvent molecules, making them a preferred choice for calculating the physicochemical properties of molecules and reactions in solvent environments. In Table 4, we present the free energies

Table 4. Reaction-Free Energies (kcal/mol) for the Reactions DBA-1^{-2} **+ Glc and DBA-2 + Glc Calculated with the SMD Model and Different Levels of Theory**[a]

level of theory	**DBA-1**$^{-2}$ + Glc	**DBA-2** + Glc
B97-D3	− 12.6 ± 0.1	− 17.0 ± 0.2
r^{2}SCAN-3c	− 9.2 ± 0.1	− 13.3 ± 0.2
PW6B95-D4/def2-TZVPD[b]	− 12.5 ± 0.4	− 16.0 ± 0.1
ωB97X-D4/def2-TZVPD[b]	− 11.0 ± 0.1	− 15.0 ± 0.1
experiment	−5.8[30,31]	− 4.6 ± 0.1[9] −5.5[9]

[a]The energies are calculated from the median (eq 9) of a distribution of reaction-free energies $\{\Delta_r G^\circ_{CE,x}\}$ generated with three trials ($N = 3$) values of molecular free energies $G^\circ_{CE,i}(X)$. [b]The molecular free energies $G^\circ_{CE}(X)$ were calculated over one of the CE optimized at level of theory r^{2}SCAN-3c.

$\Delta_r G^\circ_{CE}$ for the reactions of **DBA-1**$^{-2}$ + Glc and **DBA-2** + Glc employing the SMD implicit solvent model and different levels of theory. Irrespective of the level of theory, the computed $\Delta_r G^\circ_{CE}$ values for both reactions significantly overestimate the corresponding experimental references, by approximately a factor of 2 for **DBA-1** + Glc and a factor of 3 for **DBA-2** + Glc. This discrepancy suggests a systematic error source in the computation of the reaction-free energies $\Delta_r G^\circ_{CE}$.

https://doi.org/10.1021/acs.jpcb.4c04846
J. Phys. Chem. B 2024, 128, 9213−9223

Figure 5 depicts representative structures of the energetically most favorable conformers of **DBA-1**$^{-2}$ and **DBA-2**, as

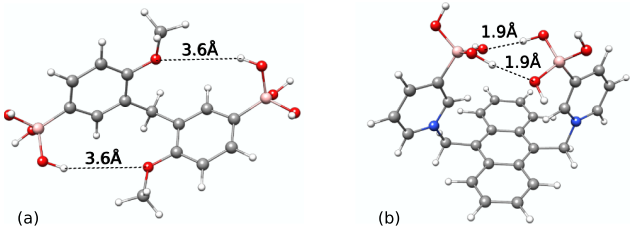


Figure 5. Stable conformers of **DBA-1**$^{-2}$ (a) and **DBA-2** (b) molecules in implicit solvent. The dashed lines are a visual aid for the intramolecular hydrogen bond interactions.

obtained from CENSO refinement of the CE with setting B, and the implicit solvation SMD. Both molecules exhibit intramolecular hydrogen bonds. In **DBA-1**$^{-2}$, these form between the B−O−H and methoxy groups (see Figure 5a) and are present in an energy range of, approximately, 0.5 kcal/mol relative to the lowest conformer. In **DBA-2**, the intramolecular hydrogen bonds are located between the B−O−H groups, as shown in Figure 5b, and are present in an energy range of 2.0 kcal/mol relative to the lowest conformer. We attribute the overstabilization of the given conformers to the inability of implicit solvent models to accurately represent the strong interactions between the solute and solvent molecules,[20,36−39] particularly when molecules feature highly polarized groups such as the boronic groups in the DBA reactants. Forming these hydrogen bonds may also induce molecular stress in the geometry of the reactants **DBA-1**$^{-2}$ and **DBA-2**, increasing their molecular energies $G^{\circ}_{CE}(X)$. Additionally, implicit solvation models struggle to calculate the solvation free energy for anionic molecules[40,41] because estimating the solvent accessible surface (SAS) is difficult due to the potentially incorrect radii used for the cavity construction. The poor radii are discussed as a reason for the poor performance of implicit solvation models for negatively charged molecules.[40,42] This problem has recently been addressed by the dynamic radii adjustment for continuum solvation (DRACO).[41]

A visual inspection of the CE for glucose does not reveal any of the above-mentioned problems for **DBA-1**$^{-2}$ and **DBA-2**. Moreover, Feng et al.[43] showed that HF/6-31+G(d) level of theory with the implicit solvation model CPCM is already accurate enough for the calculation of acidity constants pK_a and proton affinities (PA) for glucose.

To improve the representation of the aqueous media in our calculations, we decided to add explicit water molecules around the solute molecules in a cluster-continuum approach. The results of the first phase of the parameter search are presented in the next section.

Screening of CENSO Parameters for Microsolvation. In the previous section, the average computation time per conformer for the calculation of CENSO's part 2 of a CE of **DBA-1**$^{-2}$**.Glc** is half of the time compared to **DBA-2**.Glc. Therefore, we have selected the reaction **DBA-1**$^{-2}$ + Glc for the calculations in this section to save computation time. The number of water molecules for microsolvation was set arbitrarily to seven water molecules for the free ligand **DBA-1**$^{-2}$ and glucose respectively, to provide a first insight into the impact of microsolvation. The complex **DBA-1.Glc**$^{-2}$ is calculated with six water molecules even though four hydroxyl groups are part of the boron ester bonds, to avoid big differences in the number of explicit solvent molecules compared to the reactants. As the distributions sampled in the three trials might not follow a normal distribution, we report median values and interquartile ranges (IQR) hereafter. Histograms of the sampled energy distributions $\{\Delta_r G^{\circ}_{CE,x}\}$ are reported in Figure S6 in the Supporting Information.

The results in Table 5 demonstrate that incorporating explicit water molecules for all the reactants and products, in a

Table 5. Reaction-Free Energies $\Delta_r G^{\circ}_{CE}$, Interquartile Range (IQR), and Deviation of from Experiment $\Delta\Delta_r G^{\circ}_{CE\text{-}CD}$ for the Reaction DBA-1^{-2} + Glc with the CENSO Settings Listed in Table 2[a,b]

set.	$\Delta_r G^{\circ}_{CE}$	IQR	$\Delta\Delta_r G^{\circ}_{CE\text{-}CD}$
E	−6.7	2.6	−0.9
F	−2.9	1.6	2.9
C	−2.7	4.1	3.1
G	−2.2	2.3	3.6
B	−2.0	1.7	3.8
A	0.1	7.0	5.9
D	−19.3	2.0	−13.5
experiment	−5.8		

[a]All values are reported in kcal/mol. [b]The sets of $\{\Delta_r G^{\circ}_{CE,x}\}$ and the IQRs for the **DBA-1**$^{-2}$ + Glc reaction were calculated with the eqs 9, S5, and S6, using three trials for both $\{G^{\circ}_{CE,\alpha}(X(H_2O)_n)\}$ and $\{G^{\circ}_{CE,\beta}((H_2O)_n)\}$.

microsolvation approach with the QCG workflow, enables more accurate calculations of free energies for the reaction **DBA-1**$^{-2}$ + Glc $\overset{K}{\rightleftharpoons}$ **DBA-1.Glc**$^{-2}$ + 4 H_2O. The systematic error in $\Delta_r G^{\circ}_{CE}$ is reduced from around 6 kcal/mol (with only the implicit solvation model) to approximately 3 kcal/mol. The only exceptions are settings A and D, which still show significant deviations from the experimental reference.

Calculations with settings A, B, and C were performed to evaluate the influence of energy thresholds G^{thr}_i used to sort out the high-lying energy conformers with CENSO on the free energies G°. The Kruskal−Wallis test[44] and Dunn's test indicate significant differences between settings A, B, and C concerning the calculated reaction-free energies. Further details are provided in the Supporting Information in Table S1. These results imply an important influence of the CENSO energy thresholds G^{thr}_i in the distribution of the values $\{\Delta_r G^{\circ}_{CE,x}\}$. Nevertheless, the difference in median values between settings B and C is less than 1.0 kcal/mol meaning that, with settings B, the differences in the PES between GFN2-xTB (used for the CE sampling) and the DFA methods used in the CENSO parts are already taken into account to an acceptable amount. However, setting B is, on average, 22 times faster than setting C for refining a CE of **DBA-1**$^{-2}$.Glc, highlighting the significant impact of proper tuning of G^{thr}_i on the computational efficiency of the workflow. Meanwhile, the difference in median values of settings A and B highlights the importance of CENSO energy thresholds G^{thr}_i as crucial parameters for the accuracy of $\Delta_r G^{\circ}_{CE}$. The current default thresholds of G^{thr}_0 = 4.0 kcal/mol, G^{thr}_1 = 3.5 kcal/mol, and G^{thr}_2 = 2.5 kcal/mol recommend a careful evaluation of the optimal values for these parameters before performing a large-scale calculation for molecular free energies $G^{\circ}_{CE}(X)$ of similar

https://doi.org/10.1021/acs.jpcb.4c04846
J. Phys. Chem. B 2024, 128, 9213−9223

systems that require microsolvation due to strong solute−solvent interactions.

Settings C, D, E, F, and G from Table 2 were selected to test the accuracy of different combinations of DFA methods for calculating $\Delta_r G^\circ_{CE}$. While setting F gives similar results compared with settings C and G ($|\Delta\Delta_r G^\circ_{CE\text{-}CD}| \sim 3$ kcal/mol), setting D gives the worst results among the settings tested ($|\Delta\Delta_r G^\circ_{CE\text{-}CD}| > 10$ kcal/mol). The lack of accuracy in setting D likely stems from the electronic energy $\bar{E}_{gas}(X)$ and solvation free energy $\delta\bar{G}_{solv}(X)$ with the HF-3c/MINIX method, which, due to the minimal basis set, fails to accurately capture polarization effects,[45] crucial for representing solute−solvent interactions and B−O bond polarization in DBA molecules. Therefore, setting D was excluded from further analysis.

For settings C, E, F, and G, Dunn's test showed significant differences ($p < 0.05$) between the medians values for $\Delta_r G^\circ_{CE}$. However, the median values for settings C, F, and G are in close agreement within a range of less than 1 kcal/mol, with an absolute error of around 3 kcal/mol to the experimental reference $\Delta_r G^\circ_{CD}$ for the reaction of reaction **DBA-1**$^{-2}$ + Glc. Moreover, setting E gives the most accurate result with an absolute error of 0.9 kcal/mol compared with the experimental value. For all these settings, the final molecular free energy G°_{CE} is calculated with a TZ basis set, which, along with microsolvation, explains the agreement with the experimental reaction-free energy. Nevertheless, in the case of setting E, we do not discard a fortuitous (passive) error cancellation as the main reason for the improved accuracy because the only difference between settings E and G is the level of theory used in part 0 (cheap prescreening) which, according to the results with the other settings, should not considerably improve the calculated $\Delta_r G^\circ_{CE}$.

When the given workflows are applied for screening purposes, a small IQR becomes more important, since the ranking of the target property is ideally not influenced by the accuracy of the method but by its precision. In this regard, settings B and F are the most favorable with an IQR of 1.7 and 1.6 kcal/mol respectively. Furthermore, a robust or precise workflow, with the optimal tuning of energy threshold G_i^{thr}, allows the use of fewer trials, thus saving valuable computational costs. When comparing IQRs to define the precision in this way, it is important to ensure that the underlying distributions are unimodal and free of outliers. Considering the accuracy, precision, and computational costs of the workflows discussed in this section, the best-performers are settings B, E, F, and G.

Screening for the Optimal Number of Explicit Water Molecules. Since microsolvation is imperative to obtain accurate reaction-free energies for the DBAs, the number of explicit water molecules $(H_2O)_n$ is screened using the best-performing settings from the previous section (B, E, F, and G). To save computational time, the energy thresholds G_i^{thr} are reduced for all workflows to those used in setting B. All reactants and products were microsolvated with the same number of water molecules $(H_2O)_n$, and the energy distributions $\{\Delta_r G^\circ_{CE,x}\}$ were calculated using the molecular free energies $G^\circ_{CE,i}(X)$ from three trials. To avoid confusion with the previous configurations, in this section we renamed the settings E, F, and G as E2, F2, and G2, respectively.

Figures 6 and 7 show the energy distributions as boxplots for the reactions **DBA-1**$^{-2}$ + Glc and **DBA-2** + Glc, respectively. Tables S3 and S5 show the $\Delta_r G^\circ_{CE}$, IQR, and signed error

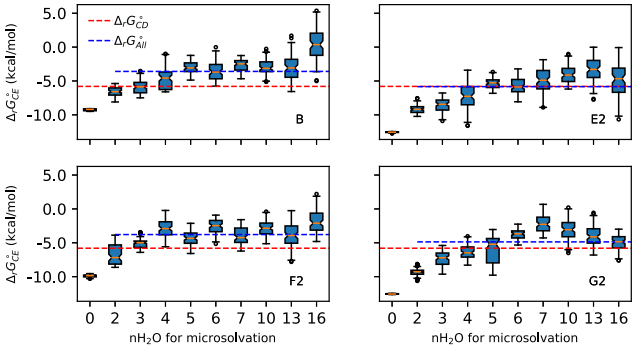


Figure 6. BoxPlot diagrams for the reaction-free energies ($\Delta_r G^\circ_{CE}$) of **DBA-1**$^{-2}$ and glucose, with different numbers of water molecules for microsolvation. The dashed red line serves as a visual guide for the experimental reaction-free energy ($\Delta_r G^\circ_{CD}$) calculated from circular dichroism (CD) experiments.[30,31]

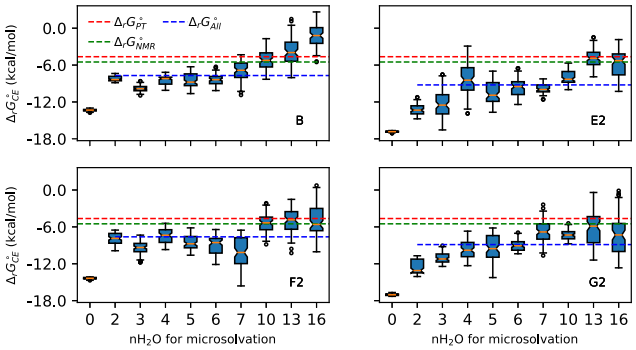


Figure 7. BoxPlot diagrams for the reaction-free energies ($\Delta_r G^\circ_{CE}$) of **DBA-2** and glucose, with different numbers of water molecules for microsolvation. The dashed red and green lines serves as a visual guide for the experimental reaction-free energies ($\Delta_r G^\circ_{PT}$ and $\Delta_r G^\circ_{NMR}$) calculated from potentiometric tritiations (PT) and nuclear magnetic resonance (NMR) experiments, respectively.[9]

$\Delta\Delta_r G^\circ_{CE\text{-}Exp}$ for Figures 6 and 7, respectively. Introducing explicit water molecules improves the representation of solute−solvent interactions, but also increases the IQR of $\Delta_r G^\circ_{CE}$ from less than 0.5 kcal/mol to over 3 kcal/mol, highly reducing the precision in some cases. Furthermore, using CENSO's part 2 of the DBA-1^{-2}.Glc product as a case example, the computational time per conformer for setting B increases proportionally with the number of water molecules used in microsolvation compared to the implicit solvent model ($n = 0$). The decrease in precision and the rise in computational cost are both expected, as the system's degrees of freedom increase with the number of explicit solvent molecules, necessitating more extensive conformational space sampling. Overall, the fluctuations of the reaction-free energies concerning the number of explicit water molecules lie around 8 kcal/mol irrespective of the DBA or applied method. Although there is a number of solvent molecules for each setting that reproduces the experimental reaction-free energies with high accuracy, there are no clear criteria in the literature to define when a system is correctly microsolvated. Matching the calculated $\Delta_r G^\circ_{CE}$ and experimental $\Delta_r G^\circ_{Exp}$ is not feasible for predictive workflows where experimental data is unavailable. Furthermore, different numbers of explicit solvation molecules can lead to matches in reaction-free energies, as observed for **DBA-1**$^{-2}$ with setting G2.

https://doi.org/10.1021/acs.jpcb.4c04846
J. Phys. Chem. B 2024, 128, 9213−9223

In principle, the double zwitterionic nature of **DBA-2** and **DBA-2.Glc** at pH 7.4 makes these molecules more polar, necessitating a greater number of explicit water molecules to accurately represent the solute−solvent interactions.

Other criteria for considering a solute to be correctly microsolvated may be related to structural features. Ideally, the optimal number of solvent molecules would correlate with a relaxation of a bond angle or dihedral angle in the reactants. However, based on the investigated angles given in Tables S4 and S6 it is not possible to decide on one specific number of solvent molecules in our case. The main problem is that the fluctuations in the bond angle when looking at different conformers, changing the number of solvent molecules, or changing the applied method are as large as the observed relaxation. In the absence of a universal criterion, we propose to use the median of all reaction-free energies that include microsolvation $\Delta_r G^\circ_{All}$ for each of the settings, to provide a single value for each setting. For **DBA-1**$^{-2}$ + Glc the final energy values deviate from the experimental values $\Delta_r G^\circ_{CD}$ by 2.2, -3.5×10^{-2}, 2.0, and 0.9 kcal/mol for settings B, E2, F2, and G2, respectively. For **DBA-2** + Glc, the calculated energies for settings B, E2, F2, and G2 deviate by −3.1, −4.6, −3.1, and −4.3 kcal/mol from the experimental value $\Delta_r G^\circ_{PT}$, and by −2.2, −3.7, −2.2, and −3.4 kcal/mol from the experimental value $\Delta_r G^\circ_{NMR}$. With the given precision and accuracy, our workflow can be used for screening candidate DBAs when using a reasonably large threshold.

In the following, further findings on the application of microsolvation are discussed. The Root Mean Square Deviation (RMSD) from the initial configuration (Figures S2, S4, S8, S10, and S13) decreases when increasing the number of water molecules for microsolvation, while the time constants (T_x) for the RMSD's autocorrelation functions (Figures S3, S5, S9, S11, and S14) become greater. This indicates that a strong network of hydrogen bonds formed by the water molecules[38] reduces the conformational changes of the solutes during the 80 ps of meta-dynamics sampling with CREST. As a consequence, the number of relevant conformers in the Boltzmann average for $G^\circ_{CE}(X)$ also decreases when adding water molecules to the solute, as reported in Tables S4, S6, and S7. Simm et al. also reported a decrease in relevant conformers as the number of water molecules increased in the microsolvation of acetonitrile with water.[38] Therefore, this trend is not related to any numerical artifact in our workflow. Additionally, it has been shown that random error propagation increases when few microstates (or conformers) are used in the calculation of statistical-derived thermodynamic properties[46] (like the molecular free energies $G^\circ_{CE}(X)$ and reaction-free energies $\Delta_r G^\circ_{CE}$), which overall agrees with the trend of the IQR trends in Figures 6 and 7, and Tables S3 and S5. Considering the above-mentioned arguments, we can not discard the matching between the calculated $\Delta_r G^\circ_{CE}$ and experimental $\Delta_r G^\circ_{Exp}$, as a consequence of poor sampling of the configuration space or incorrect selection of the DFA methods for the CE refinement, being an example of "get the correct answer by the wrong reasons". We conducted pairwise Dunn's tests to assess the convergence of calculated reaction-free energies depending on the number of water molecules used for microsolvation. Heatmaps for the pairwise Dunn's tests are reported in Figures S7 and S12 for reactions the reactions **DBA-1**$^{-2}$+ Glc and **DBA-2** + Glc, respectively. In general, for **DBA-1**$^{-2}$ + Glc reaction, the optimal number of water molecules for microsolvation is in a range between 5 and 10 molecules because Dunn's test does not show significant differences between the $\Delta_r G^\circ_{CE}$ in this range. In the case of **DBA-2** + Glc reaction, Dunn's test does not show significant differences for $\Delta_r G^\circ_{CE}$ in two ranges, a first range of 4−6 water molecules where the calculated amount of energy released by the reaction is overestimated compared to the experimental values, and a second range from 10 to 16 water molecules where the reaction-free energy for **DBA-2** + Glc match with experimental references $\Delta_r G^\circ_{PT}$ and $\Delta_r G^\circ_{NMR}$ in most of the settings. However, pairwise Dunn's tests may require more trials N for $G^\circ_{CE}(X)$ to achieve conclusive results regarding the convergence of $\Delta_r G^\circ_{CE}$, with the number of water molecules for microsolvation. Due to the computational cost, it may be worth exploring this criterion in other systems with smaller conformational space for the solute.

The reliability of the results presented in Figures 6 and 7, and the optimal parameters for microsolvation, are validated in the next section, by using more trials for the calculation of reaction-free energies.

Workflow Accuracy. All the previous calculations of the reaction-free energy $\Delta_r G^\circ_{CE}$ were performed with three trials. To test the accuracy of our workflow, we increased the number of samples for $G^\circ_{CE,i}(X)$ up to ten trials (N = 10) for the calculation of the energy distributions $\{\Delta_r G^\circ_{CE,x}\}$. In this section, we also calculated the free energy for the **DBA-1**$^{-2}$ + Gal reaction. Increasing the number of trials N reduces the statistical noise for sampling the free energies $G^\circ_{CE}(X)$.

Tables 6−8 show the results of $\Delta_r G^\circ_{CE}$, IQR, and the signed error $\Delta\Delta_r G^\circ_{CE\text{-}Exp}$ with respect to the experimental reference

Table 6. Free Energies for the Reaction DBA-1^{-2} + Glc $\overset{K}{\rightleftharpoons}$ DBA-1.Glc^{-2} + 4 H_2O, Calculated Using the Setting B for CE Refinement, Seven Water Molecules for Microsolvation $(H_2O)_7$, and Increasing Number of Trials N for the Molecular Free Energies $G^\circ_{CE,i}(X)$[a]

N	$\Delta_r G^\circ_{CE}$	IQR	$\Delta\Delta_r G^\circ_{CE\text{-}CD}$
3	−2.5	1.4	3.3
5	−2.1	1.6	3.7
7	−2.6	1.9	3.2
10	−2.4	1.9	3.4

[a] $\Delta_r G^\circ_{CE}$, IQR, and $\Delta\Delta_r G^\circ_{CE\text{-}CD}$ reported in kcal/mol.

Table 7. Free Energies for the Reaction DBA-2 + Glc $\overset{K}{\rightleftharpoons}$ DBA-2.Glc + 4 H_2O, Calculated Using the Setting B for CE Refinement, Seven Water Molecules for Microsolvation $(H_2O)_7$, and Increasing Number of Trials N for the Molecular Free Energies $G^\circ_{CE,i}(X)$[a]

N	$\Delta_r G^\circ_{CE}$	IQR	$\Delta\Delta_r G^\circ_{CE\text{-}PT}$	$\Delta\Delta_r G^\circ_{CE\text{-}NMR}$
3	−6.6	2.3	−2.0	−1.1
5	−6.2	2.3	−1.6	−0.7
7	−6.4	2.0	−1.8	−0.9
10	−6.4	2.0	−1.8	−0.9

[a] $\Delta_r G^\circ_{CE}$, IQR, and $\Delta\Delta_r G^\circ_{CE\text{-}CD}$ reported in kcal/mol.

(CD or PT). For the reactions **DBA-1**$^{-2}$ + Glc and **DBA-2** + Glc, the statistical values fluctuate in an energy range less than 1 kcal/mol with the number of trials N, coinciding with a shape convergence to a monomodal distribution, as shown in the Figure S15a,b. For the **DBA-1**$^{-2}$ + Gal reaction, Figure S15c shows an initial bimodal shape for $\{\Delta_r G^\circ_{CE,x}\}$, converging

https://doi.org/10.1021/acs.jpcb.4c04846
J. Phys. Chem. B 2024, 128, 9213−9223

Table 8. Free Energies for the Reaction DBA-1^{-2} + Gal $\overset{K}{\rightleftharpoons}$ DBA-1.Gal^{-2} + 4 H_2O, Calculated Using the Setting B for CE Refinement, Seven Water Molecules for Microsolvation $(H_2O)_7$, and Increasing Number of Trials N for the Molecular Free Energies $G^{\circ}_{CE,i}(X)$[a]

N	$\Delta_r G^{\circ}_{CE}$	IQR	$\Delta\Delta_r G^{\circ}_{CE\text{-}CD}$
3	−9.3	4.3	−5.1
5	−6.5	4.0	−2.3
7	−6.7	3.9	−2.5
10	−6.1	3.5	−1.9

[a] $\Delta_r G^{\circ}_{CE}$, IQR, and $\Delta\Delta_r G^{\circ}_{CE\text{-}CD}$ reported in kcal/mol.

to a monomodal distribution with increasing the samples N, and showing variations of less than 1 kcal/mol for the statistical values when $N \geq 5$. The initial bimodal distribution of $\{\Delta_r G^{\circ}_{CE,x}\}$ is due to outliers in the sampled $G^{\circ}_{CE,i}(X)$. Microsolvated configurations with outlier energy values have also been reported for the SN2 reaction $CH_3Cl + NH_3$ in aqueous media, where some configurations may even eliminate the reaction energy barrier of 23.5 kcal/mol.[47] However, in our approach, the impact of energy outliers diminishes as the number of samples (N) increases, leading to an improvement in accuracy, as shown in Table 8.

Finally, for the set of three reactions in this section, we report for our workflow an average absolute error $\langle|\Delta\Delta_r G^{\circ}_{CE\text{-}Exp}|\rangle$ = 2.2 kcal/mol respect the experimental references, and average $\langle IQR\rangle$ = 2.4 kcal/mol, when using ten trials (N = 10) of molecular free energies $G^{\circ}_{CE}(X)$, setting B, and seven water molecules for microsolvation of all the solutes. With the above-mentioned settings, the performance of our workflow is remarkable considering that, in general, experimental setups hardly achieve measurement precision of 1−2 kcal/mol.[17] Most importantly, the number of trials can significantly impact the accuracy and it is critical to ensure that the distribution of reaction-free energies is approximately monomodal, especially if experimental data is unavailable.

CONCLUSIONS

In this work, we performed accurate calculations of reaction-free energies $\Delta_r G^{\circ}_{CE}$ for reactions of diboronic acids (DBA) with the monosaccharides glucose (Glc) and galactose (Gal) in aqueous solvent. Specifically, the reactions studied are **DBA-1**$^{-2}$ + Glc, **DBA-2** + Glc, and **DBA-1**$^{-2}$ + Gal. To account for the high molecular flexibility of reactants and products, significant structural changes after the reaction, strong solute−solvent interactions, and the complicated electronic structure around the R-B$(OH)_3^{-1}$ groups, the molecular free energies of the reactants and products are calculated ($G^{\circ}_{CE}(X)$) in a workflow composed of the CREST[19] and CENSO[18] programs.

The conformational entropy change plays a crucial role in accurately calculating the reaction-free energies accounting for an energy penalty of approximately 3−5 kcal/mol. In general, conformational entropy should be considered when the flexibility of the reactants changes significantly due to the reaction, for example by the formation of rings in the products. To this end, the CREST program can be used to identify the number of relevant conformers for the reactants and products and to calculate the conformational entropy.

The reaction-free energy is systematically overestimated with the SMD implicit solvent, compared with experimental reference data, indicating the need for microsolvation to improve the description of the solute−solvent interactions. We conducted a systematic screening to choose the most significant parameters that affect the accuracy, precision, and computational cost of the workflow. These parameters include DFA combinations and energy thresholds (G_i^{thr}) for the sorting of conformers with CENSO, the number of water molecules for microsolvation, and the number of trials (N) of calculating molecular free energies $G^{\circ}_{CE,i}(X)$. After conducting a parameter screening, we found that using tight energy windows G_i^{thr} decreases the accuracy of the calculations. We recommend carefully tuning these windows before large-scale free energy calculations to avoid computational overhead. In general, the use of HF-3c, B97-D3, and r^{2}SCAN-3c DFAs are suitable for CENSO parts in the CE refinement. When sorting the conformers for $G^{\circ}_{CE}(X)$, it is important to use a basis set of TZ quality. The methods B97-D3 and r^{2}SCAN-3c are both suitable for this step, but r^{2}SCAN-3c provided more consistent results for $\Delta_r G^{\circ}_{CE}$ with an increasing number of water molecules for microsolvation.

The number of water molecules required for microsolvation depends on the system being studied. Instead of a definitive number, we find a range that improves the accuracy of $\Delta_r G^{\circ}_{CE}$. However, increasing the number of water molecules decreases the accuracy due to the increased complexity of sampling the conformational space. Moreover, a strong network of hydrogen bonds formed by the water molecules decreases the number of conformers in the CE for $G^{\circ}_{CE}(X)$, also increasing the aleatory error. Therefore, the selection of the optimal number of water molecules for the calculation of $\Delta_r G^{\circ}_{CE}$ is a trade-off between accuracy, precision, and computational cost.

Finally, the workflow accuracy is tested by calculating the $\Delta_r G_{CE}$ for the reactions **DBA-1**$^{-2}$ + Glc, **DBA-2** + Glc, and **DBA-1**$^{-2}$ + Gal, increasing the number of trials up to ten samples (N = 10) of $G^{\circ}_{CE,i}(X)$, using seven water molecules for microsolvation, and the setting B for sorting the CE of reactants and products. The median and IQR values for reaction-free energy $\{\Delta_r G^{\circ}_{CE,x}\}$ with monomodal distribution fluctuate in an energy range of less than 1 kcal/mol. In the case of bimodal distribution, a small fluctuation of the statistical values is achieved with five or more trials ($N \geq 5$). The high accuracy and precision of our workflow are demonstrated by the average absolute error $\langle|\Delta\Delta_r G^{\circ}_{CE\text{-}Exp}|\rangle$ = 2.2 kcal/mol with respect to the experimental references, and average $\langle IQR\rangle$ = 2.4 kcal/mol for the reaction-free energies of the three set of reactions **DBA-1**$^{-2}$ + Glc, **DBA-2** + Glc, and **DBA-1**$^{-2}$ + Gal.

With properly tuned parameters and an increased number of trials, our workflow achieves an accuracy of approximately 1−2 kcal/mol for the reaction-free energies, as shown in Tables 7 and 8. The precision, given through the IQR of 2−3.5 kcal/mol, is comparable to the uncertainty in experimental values. This makes the workflow appealing for the computational design of new diboronic acid ligands tailored for selective reactions with carbohydrates. For large-scale screening applications, however, simplified schemes with reduced computational cost have to be developed in the future.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.4c04846.

https://doi.org/10.1021/acs.jpcb.4c04846
J. Phys. Chem. B 2024, 128, 9213−9223

Equations for calculation of reaction-free energies with cluster-continuum solvation; equilibrium reactions for all solutes; comparison test results between CENSO's setups; detailed information for the reactions **DBA-1**$^{-2}$ + Glc and **DBA-2** + Glc; distribution of reaction-free energies relative to the number of trials $G^{\circ}_{CE,i}(X)$ (PDF)

XYZ files for conformational ensembles of microsolvated reactants and products (ZIP)

AUTHOR INFORMATION

Corresponding Author

Andres Jaramillo-Botero − *Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States; iOMICAS Research Institute, Pontificia Universidad Javeriana, Santiago de Cali, Valle del Cauca 760031, Colombia;* orcid.org/0000-0003-2844-0756; Email: ajaramil@caltech.edu

Authors

Gustavo Adolfo Lara-Cruz − *iOMICAS Research Institute, Pontificia Universidad Javeriana, Santiago de Cali, Valle del Cauca 760031, Colombia;* orcid.org/0000-0003-4253-9824

Thomas Rose − *Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Bonn 53115, Germany*

Stefan Grimme − *Mulliken Center for Theoretical Chemistry, Clausius-Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Bonn 53115, Germany;* orcid.org/0000-0002-5844-4371

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.4c04846

Author Contributions

G.A.L.C. and T.R. contributed equally to the planning, writing, and reviewing of this manuscript. G.A.L.C. performed all calculations except for the conformational entropy contributions of the DBA-1−2 + Glc and DBA-2 + Glc reactions, which were conducted by T.R. A.J-B. conceived the project and co-supervised all modeling and simulations with S.G. All authors made substantial intellectual contributions to the research and manuscript preparation.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was partially funded by the "OMICAS Alliance: Optimización Multiescala In-silico de Cultivos Agrícolas Sostenibles" Scientific Ecosystem, part of the Colombia Científica Program, sponsored by The World Bank, The Ministry of Science, Technology and Innovation (MINCIENCIAS), ICETEX, the Colombian Ministry of Education and the Colombian Ministry of Commerce, Industry and Tourism, under ID: FP44842-217-2018(Award ID: 792-61187). G.A.L.C. wants to thank Professor Carlos Navarro at iOMICAS Research Institute for helpful discussions about the statistical analysis of the results. MarvinSketch was used for drawing and displaying the chemical structures in Figures 1−4, MarvinSketch 23.3.0, Chemaxon (https://www.chemaxon.com).

REFERENCES

(1) Halford, N.; Curtis, T.; Muttucumaru, N.; Postles, J.; Mottram, D. Sugars in crop plants. *Ann. Appl. Biol.* **2011**, *158*, 1−25.

(2) Striegler, S. Selective Carbohydrate Recognition by Synthetic Receptors in Aqueous Solution. *Curr. Org. Chem.* **2003**, *7*, 81−102.

(3) Arnaud, J.; Audfray, A.; Imberty, A. Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. *Chem. Soc. Rev.* **2013**, *42*, 4798−4813.

(4) Bian, Z.; Liu, A.; Li, Y.; Fang, G.; Yao, Q.; Zhang, G.; Wu, Z. Boronic acid sensors with double recognition sites: a review. *Analyst* **2020**, *145*, 719−744.

(5) Rowan, A. E.; Rowan, S. J.; Aida, T.; James, T. D.; Phillips, M. D.; Shinkai, S. *Boronic Acids in Saccharide Recognition*; Stoddart, J. F., Ed.; Monographs in Supramolecular Chemistry; The Royal Society of Chemistry, 2006; pp. 1−174.

(6) Lara-Cruz, G. A.; Jaramillo-Botero, A. Molecular Level Sucrose Quantification: A Critical Review. *Sensors* **2022**, *22*, 9511.

(7) Sandanayake, K. R. A. S.; Nakashima, K.; Shinkai, S. Specific recognition of disaccharides by trans-3,3′-stilbenediboronic acid: rigidification and fluoresecence enhancement of the stilbene skeleton upon formation of a sugar−stilbene macrocycle. *J. Chem. Soc., Chem. Commun.* **1994**, 1621−1622.

(8) Wu, X.; Li, Z.; Chen, X.-X.; Fossey, J. S.; James, T. D.; Jiang, Y.-B. Selective sensing of saccharides using simple boronic acids and their aggregates. *Chem. Soc. Rev.* **2013**, *42*, 8032−8048.

(9) Eggert, H.; Frederiksen, J.; Morin, C.; Norrild, J. C. A New Glucose-Selective Fluorescent Bisboronic Acid. First Report of Strong α-Furanose Complexation in Aqueous Solution at Physiological pH1. *J. Org. Chem.* **1999**, *64*, 3846−3852.

(10) Bielecki, M.; Eggert, H.; Chr Norrild, J. A fluorescent glucose sensor binding covalently to all five hydroxy groups of α-D-glucofuranose. A reinvestigation. *J. Chem. Soc., Perkin Trans.* **1999**, *2*, 449−456.

(11) Yang, W.; He, H.; Drueckhammer, D. G. Computer-Guided Design in Molecular Recognition: Design and Synthesis of a Glucopyranose Receptor. *Angew. Chem., Int. Ed.* **2001**, *40*, 1714−1718.

(12) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous Free Energy Simulations in Virtual Screening. *J. Chem. Inf. Model.* **2020**, *60*, 4153−4169.

(13) Abb, S.; Tarrat, N.; Cortés, J.; Andriyevsky, B.; Harnau, L.; Schön, J. C.; Rauschenbach, S.; Kern, K. Carbohydrate Self-Assembly at Surfaces: STM Imaging of Sucrose Conformation and Ordering on Cu(100). *Angew. Chem., Int. Ed.* **2019**, *58*, 8336−8340.

(14) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847−2862.

(15) Pracht, P.; Grimme, S. Calculation of absolute molecular entropies and heat capacities made simple. *Chem. Sci.* **2021**, *12*, 6551−6568.

(16) Harvey, J.; Himo, F.; Maseras, F.; Perrin, L. Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis. *ACS Catal.* **2019**, *9*, 6803−6813.

(17) Bursch, M.; Mewes, J.-M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202205735.

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

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

(20) Spicher, S.; Plett, C.; Pracht, P.; Hansen, A.; Grimme, S. Automated Molecular Cluster Growing for Explicit Solvation by Efficient Force Field and Tight Binding Methods. *J. Chem. Theory Comput.* **2022**, *18*, 3174−3189.

(21) Spicher, S.; Grimme, S. Single-Point Hessian Calculations for Improved Vibrational Frequencies and Rigid-Rotor-Harmonic-Oscillator Thermodynamics. *J. Chem. Theory Comput.* **2021**, *17*, 1701−1714.
(22) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383−394.
(23) Sigalov, G.; Fenley, A.; Onufriev, A. Analytical electrostatics for biomolecules: Beyond the generalized Born approximation. *J. Chem. Phys.* **2006**, *124*, No. 124902.
(24) Lange, A. W.; Herbert, J. M. Improving Generalized Born Models by Exploiting Connections to Polarizable Continuum Models. I. An Improved Effective Coulomb Operator. *J. Chem. Theory Comput.* **2012**, *8*, 1999−2011.
(25) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. 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* **2009**, *18*, 6378−6396.
(26) Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. *J. Chem. Theory Comput.* **2021**, *17*, 4250−4261.
(27) Norrild, J. C.; Eggert, H. Evidence for Mono- and Bisdentate Boronate Complexes of Glucose in the Furanose Form. Application of 1JC-C Coupling Constants as a Structural Probe. *J. Am. Chem. Soc.* **1995**, *117*, 1479−1484.
(28) Urban, F.; Shaffer, P. A. THE ACIDIC PROPERTY OF SUGARS. *J. Biol. Chem.* **1932**, *94*, 697−715.
(29) Urban, F.; Williams, R. THE ACIDIC PROPERTY OF SUGARS. II. *J. Biol. Chem.* **1933**, *100*, 237−241.
(30) Shiomi, Y.; Saisho, M.; Tsukagoshi, K.; Shinkai, S. Specific complexation of glucose with a diphenylmethane-3,3′ -diboronic acid derivative: correlation between the absolute configuration of mono- and di-saccharides and the circular dichroic activity of the complex. *J. Chem. Soc., Perkin Trans.* **1993**, *1*, 2111−2117.
(31) Tsukagoshi, K.; Shinkai, S. Specific complexation with mono- and disaccharides that can be detected by circular dichroism. *J. Org. Chem.* **1991**, *56*, 4089−4091.
(32) Brandenburg, J. G.; Maas, T.; Grimme, S. Benchmarking DFT and semiempirical methods on structures and lattice energies for ten ice polymorphs. *J. Chem. Phys.* **2015**, *142*, No. 124104.
(33) Ehlert, S.; Grimme, S.; Hansen, A. Conformational Energy Benchmark for Longer n-Alkane Chains. *J. Phys. Chem. A* **2022**, *126*, 3521−3535.
(34) Müller, M.; Hansen, A.; Grimme, S. ωB97X-3c: A composite range-separated hybrid DFT method with a molecule-optimized polarized valence double-ζ basis set. *J. Chem. Phys.* **2023**, *158*, No. 014103.
(35) Spicher, S.; Caldeweyher, E.; Hansen, A.; Grimme, S. Benchmarking London dispersion corrected density functional theory for noncovalent ion−π interactions. *Phys. Chem. Chem. Phys.* **2021**, *23*, 11635−11648.
(36) Pliego, J. R., Jr; Riveros, J. M. Hybrid discrete-continuum solvation methods. *WIREs Comput. Mol. Sci.* **2020**, *10*, No. e1440.
(37) Sure, R.; el Mahdali, M.; Plajer, A.; Deglmann, P. Towards a converged strategy for including microsolvation in reaction mechanism calculations. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 473−492.
(38) Simm, G. N.; Türtscher, P. L.; Reiher, M. Systematic microsolvation approach with a cluster-continuum scheme and conformational sampling. *J. Comput. Chem.* **2020**, *41*, 1144−1155.
(39) Zhang, J.; Zhang, H.; Wu, T.; Wang, Q.; van der Spoel, D. Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *J. Chem. Theory Comput.* **2017**, *13*, 1034−1043.
(40) Kröger, L. C.; Müller, S.; Smirnova, I.; Leonhard, K. Prediction of Solvation Free Energies of Ionic Solutes in Neutral Solvents. *J. Phys. Chem. A* **2020**, *124*, 4171−4181.
(41) Plett, C.; Stahn, M.; Bursch, M.; Mewes, J.-M.; Grimme, S. Improving Quantum Chemical Solvation Models by Dynamic Radii Adjustment for Continuum Solvation (DRACO). *J. Phys. Chem. Lett.* **2024**, *15*, 2462−2469.
(42) Stahn, M.; Ehlert, S.; Grimme, S. Extended Conductor-like Polarizable Continuum Solvation Model (CPCM-X) for Semiempirical Methods. *J. Phys. Chem. A* **2023**, *127*, 7036−7043.
(43) Feng, S.; Bagia, C.; Mpourmpakis, G. Determination of Proton Affinities and Acidity Constants of Sugars. *J. Phys. Chem. A* **2013**, *117*, 5211−5219.
(44) Kruskal, W. H.; Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583−621.
(45) Müller, M.; Hansen, A.; Grimme, S. An atom-in-molecule adaptive polarized valence single-χ atomic orbital basis for electronic structure calculations. *J. Chem. Phys.* **2023**, *159*, No. 164108.
(46) Faver, J. C.; Yang, W.; Merz, K. M. The Effects of Computational Modeling Errors on the Estimation of Statistical Mechanical Variables. *J. Chem. Theory Comput.* **2012**, *8*, 3769−3776.
(47) Li, Y.; Hartke, B. Assessing Solvation Effects on Chemical Reactions with Globally Optimized Solvent Clusters. *ChemPhysChem* **2013**, *14*, 2678−2686.

https://doi.org/10.1021/acs.jpcb.4c04846
J. Phys. Chem. B 2024, 128, 9213−9223

Bibliography

[1] S. Grimme and T. Rose, *mcGFN-FF: an accurate force field for optimization and energetic screening of molecular crystals*, en, Z. Naturforsch., B **79** (2024) 191, DOI: 10.1515/znb-2023-0088.

[2] T. Rose, M. Bursch, J.-M. Mewes, and S. Grimme, *Fast and Robust Modeling of Lanthanide and Actinide Complexes, Biomolecules, and Molecular Crystals with the Extended GFN-FF Model*, en, Inorg. Chem. **63** (2024) 19364, DOI: 10.1021/acs.inorgchem.4c03215.

[3] G. A. Lara-Cruz, T. Rose, S. Grimme, and A. Jaramillo-Botero, *Reaction-Free Energies for Complexation of Carbohydrates by Tweezer Diboronic Acids*, en, J. Phys. Chem. B **128** (2024) 9213, DOI: 10.1021/acs.jpcb.4c04846.

[4] A. Arissa, T. Rose, N. Leick, S. Grimme, J. C. Johnson, and J. V. Lockard, *Charge Transfer and Recombination Pathways through Fullerene Guests in Porphyrin-Based MOFs*, J. Phys. Chem. C **129** (2025) 8215, DOI: 10.1021/acs.jpcc.5c00161.

[5] A. Katbashev, M. Stahn, T. Rose, V. Alizadeh, M. Friede, C. Plett, P. Steinbach, and S. Ehlert, *Overview on Building Blocks and Applications of Efficient and Robust Extended Tight Binding*, J. Phys. Chem. A **129** (2025), PMID: 40013428 2667, DOI: 10.1021/acs.jpca.4c08263.

[6] M. Thiele et al., *Multifunctional Organocatalysts - Singly-Linked and Macrocyclic Bisphosphoric Acids for Asymmetric Phase-Transfer and Brønsted-Acid Catalysis*, en, Chem. Eur. J. **29** (2023) e202202953, DOI: 10.1002/chem.202202953.

[7] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, *The high-throughput highway to computational materials design*, en, Nat. Mater. **12** (2013) 191, DOI: 10.1038/nmat3568.

[8] N. Mardirossian and M. Head-Gordon, *Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals*, en, Molecular Physics **115** (2017) 2315, DOI: 10.1080/00268976.2017.1333644.

[9] C. D. Parks et al., *D3R grand challenge 4: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies*, en, J. Comput.-Aided Mol. Des. **34** (2020) 99, DOI: 10.1007/s10822-020-00289-y.

[10] P. Deglmann, A. Schäfer, and C. Lennartz, *Application of quantum calculations in the chemical industry—An overview*, en, Int. J. Quantum Chem. **115** (2015) 107, DOI: 10.1002/qua.24811.

[11] K. R. Campos, P. J. Coleman, J. C. Alvarez, S. D. Dreher, R. M. Garbaccio, N. K. Terrett, R. D. Tillyer, M. D. Truppo, and E. R. Parmee,
The importance of synthetic chemistry in the pharmaceutical industry, en,
Science **363** (2019) eaat0805, DOI: 10.1126/science.aat0805.

[12] S. Spicher and S. Grimme,
Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems, en,
Angew. Chem., Int. Ed. **59** (2020) 15665, DOI: 10.1002/anie.202004239.

[13] K. Jug and T. Bredow, *Models for the treatment of crystalline solids and surfaces*, en,
J. Comput. Chem. **25** (2004) 1551, DOI: 10.1002/jcc.20080.

[14] J. Tomasi, B. Mennucci, and R. Cammi,
Quantum Mechanical Continuum Solvation Models, en, Chem. Rev. **105** (2005) 2999,
DOI: 10.1021/cr9904009.

[15] S. Ringe, N. G. Hörmann, H. Oberhofer, and K. Reuter,
Implicit Solvation Methods for Catalysis at Electrified Interfaces, en,
Chem. Rev. **122** (2022) 10777, DOI: 10.1021/acs.chemrev.1c00675.

[16] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller,
Equation of State Calculations by Fast Computing Machines, en,
J. Chem. Phys. **21** (1953) 1087, DOI: 10.1063/1.1699114.

[17] A. Zunger and O. I. Malyi, *Understanding Doping of Quantum Materials*, en,
Chem. Rev. **121** (2021) 3031, DOI: 10.1021/acs.chemrev.0c00608.

[18] M. Ernst and G. Gryn'ova, *Strength and Nature of Host-Guest Interactions in Metal-Organic Frameworks from a Quantum-Chemical Perspective*, en,
ChemPhysChem **23** (2022) e202200098, DOI: 10.1002/cphc.202200098.

[19] A. S. Cannon and J. C. Warner,
Noncovalent Derivatization: Green Chemistry Applications of Crystal Engineering, en,
Cryst. Growth Des. **2** (2002) 255, DOI: 10.1021/cg0255218.

[20] E. Stoler and J. Warner, *Non-Covalent Derivatives: Cocrystals and Eutectics*, en,
Molecules **20** (2015) 14833, DOI: 10.3390/molecules200814833.

[21] H. Furukawa, K. E. Cordova, M. O'Keeffe, and O. M. Yaghi,
The Chemistry and Applications of Metal-Organic Frameworks, en,
Science **341** (2013) 1230444, DOI: 10.1126/science.1230444.

[22] H. Li, K. Wang, Y. Sun, C. T. Lollar, J. Li, and H.-C. Zhou,
Recent advances in gas storage and separation using metal–organic frameworks, en,
Mater. Today **21** (2018) 108, DOI: 10.1016/j.mattod.2017.07.006.

[23] T. A. Goetjen, J. Liu, Y. Wu, J. Sui, X. Zhang, J. T. Hupp, and O. K. Farha, *Metal–organic framework (MOF) materials as polymerization catalysts: a review and recent advances*, en,
Chem. Commun. **56** (2020) 10409, DOI: 10.1039/D0CC03790G.

[24] E. A. Dolgopolova, A. M. Rice, C. R. Martin, and N. B. Shustova, *Photochemistry and photophysics of MOFs: steps towards MOF-based sensing enhancements*, en,
Chem. Soc. Rev. **47** (2018) 4710, DOI: 10.1039/C7CS00861A.

[25] R. E. Skyner, J. L. McDonagh, C. R. Groom, T. Van Mourik, and J. B. O. Mitchell, *A review of methods for the calculation of solution free energies and the modelling of systems in solution*, en, Phys. Chem. Chem. Phys. **17** (2015) 6174, DOI: 10.1039/C5CP00288E.

[26] C. J. Cramer and D. G. Truhlar, *Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics*, en, Chem. Rev. **99** (1999) 2161, DOI: 10.1021/cr960149m.

[27] C. Merten, *Modelling solute–solvent interactions in VCD spectra analysis with the micro-solvation approach*, en, Phys. Chem. Chem. Phys. **25** (2023) 29404, DOI: 10.1039/D3CP03408A.

[28] Q. Zhang, S. O'Brien, and J. Grimm, *Biomedical Applications of Lanthanide Nanomaterials, for Imaging, Sensing and Therapy*, en, Nanotheranostics **6** (2022) 184, DOI: 10.7150/ntno.65530.

[29] K. Lv, S. Fichter, M. Gu, J. März, and M. Schmidt, *An updated status and trends in actinide metal-organic frameworks (An-MOFs): From synthesis to application*, en, Coord. Chem. Rev. **446** (2021) 214011, DOI: 10.1016/j.ccr.2021.214011.

[30] S. Zhang, P. Winter, K. Wu, and A. D. Sherry, *A Novel Europium(III)-Based MRI Contrast Agent*, en, J. Am. Chem. Soc. **123** (2001) 1517, DOI: 10.1021/ja005820q.

[31] H.-K. Kim, G. H. Lee, and Y. Chang, *Gadolinium as an MRI Contrast Agent*, en, Future Med. Chem. **10** (2018) 639, DOI: 10.4155/fmc-2017-0215.

[32] H. Peng, M. I. J. Stich, J. Yu, L.-n. Sun, L. H. Fischer, and O. S. Wolfbeis, *Luminescent Europium(III) Nanoparticles for Sensing and Imaging of Temperature in the Physiological Range*, en, Adv. Mater. **22** (2010) 716, DOI: 10.1002/adma.200901614.

[33] M. N. Bochkarev, *Molecular compounds of "new" divalent lanthanides*, en, Coord. Chem. Rev. **248** (2004) 835, DOI: 10.1016/j.ccr.2004.04.004.

[34] D. Guillaumont, *Actinide(III) and lanthanide(III) complexes with nitrogen ligands: Counterions and ligand substituent effects on the metal–ligand bond*, en, J. Mol. Struct.:THEOCHEM **771** (2006) 105, DOI: 10.1016/j.theochem.2006.03.021.

[35] M. R. MacDonald, J. E. Bates, M. E. Fieser, J. W. Ziller, F. Furche, and W. J. Evans, *Expanding Rare-Earth Oxidation State Chemistry to Molecular Complexes of Holmium(II) and Erbium(II)*, en, J. Am. Chem. Soc. **134** (2012) 8420, DOI: 10.1021/ja303357w.

[36] P. C. D. Hawkins, *Conformation Generation: The State of the Art*, en, J. Chem. Inf. Model. **57** (2017) 1747, DOI: 10.1021/acs.jcim.7b00221.

[37] D. Timmer et al., *Structural Flexibility Slows Down Charge Transfers in Diaminoterephthalate-C_{60} Dyads*, en, J. Phys. Chem. C **128** (2024) 2380, DOI: 10.1021/acs.jpcc.3c08270.

[38] P. Pracht, F. Bohle, and S. Grimme, *Automated exploration of the low-energy chemical space with fast quantum chemical methods*, en, Phys. Chem. Chem. Phys. **22** (2020), Publisher: The Royal Society of Chemistry 7169, DOI: 10.1039/C9CP06869D.

[39] S. A. Hollingsworth and R. O. Dror, *Molecular Dynamics Simulation for All*, Neuron **99** (2018), Publisher: Elsevier 1129, DOI: 10.1016/j.neuron.2018.08.011.

[40] H.-X. Zhou and M. K. Gilson, *Theory of Free Energy and Entropy in Noncovalent Binding*, en, Chem. Rev. **109** (2009) 4092, DOI: 10.1021/cr800551w.

[41] V. Limongelli, *Ligand binding free energy and kinetics calculation in 2020*, en, WIREs Comput. Mol. Sci. **10** (2020) e1455, DOI: 10.1002/wcms.1455.

[42] M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme, *Best-Practice DFT Protocols for Basic Molecular Computational Chemistry***, en, Angew. Chem., Int. Ed. **61** (2022) e202205735, DOI: 10.1002/anie.202205735.

[43] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, *A fifth-order perturbation comparison of electron correlation theories*, en, Chem. Phys. Lett. **157** (1989) 479, DOI: 10.1016/S0009-2614(89)87395-6.

[44] A. M. Teale et al., *DFT exchange: sharing perspectives on the workhorse of quantum chemistry and materials science*, en, Phys. Chem. Chem. Phys. **24** (2022) 28700, DOI: 10.1039/D2CP02827A.

[45] E. Jonas, S. Kuhn, and N. Schlörer, *Prediction of chemical shift in NMR: A review*, en, Magn. Reson. Chem. **60** (2022) 1021, DOI: 10.1002/mrc.5234.

[46] T. Gasevic, J. B. Kleine Büning, S. Grimme, and M. Bursch, *Benchmark Study on the Calculation of ^{207} Pb NMR Chemical Shifts*, en, Inorg. Chem. **63** (2024) 5052, DOI: 10.1021/acs.inorgchem.3c04539.

[47] P. Pracht, D. F. Grant, and S. Grimme, *Comprehensive Assessment of GFN Tight-Binding and Composite Density Functional Theory Methods for Calculating Gas-Phase Infrared Spectra*, en, J. Chem. Theory Comput. **16** (2020) 7044, DOI: 10.1021/acs.jctc.0c00877.

[48] P. Makkar and N. N. Ghosh, *A review on the use of DFT for the prediction of the properties of nanomaterials*, en, RSC Adv. **11** (2021) 27897, DOI: 10.1039/D1RA04876G.

[49] L. Kunze, T. Froitzheim, A. Hansen, S. Grimme, and J.-M. Mewes, *ΔDFT Predicts Inverted Singlet–Triplet Gaps with Chemical Accuracy at a Fraction of the Cost of Wave Function-Based Approaches*, en, J. Phys. Chem. Lett. **15** (2024) 8065, DOI: 10.1021/acs.jpclett.4c01649.

[50] N. Cao, A. C. Castro, D. Balcells, U. Olsbye, and A. Nova, *Copper(II)-Oxyl Formation in a Biomimetic Complex Activated by Hydrogen Peroxide: The Key Role of Trans-Bis(Hydroxo) Species*, en, Inorg. Chem. **63** (2024) 23082, DOI: 10.1021/acs.inorgchem.4c01948.

[51] A. M. Cheplakova, K. A. Kovalenko, D. G. Samsonenko, V. A. Lazarenko, P. V. Dorovatovskii, A. S. Nizovtsev, and V. P. Fedin, *Highly Porous Scandium(III) Tetrafluoroisophthalate Framework for Adsorptive Separation of Light Alkanes*, Inorg. Chem. **63** (2024), Publisher: American Chemical Society, DOI: 10.1021/acs.inorgchem.4c03990.

[52] Y. Stöferle, P. P. Kalapos, P. Willi, and P. Chen,
Activation of Methyltrioxorhenium for Olefin Metathesis by a Frustrated Lewis Pair, en,
J. Am. Chem. Soc. **146** (2024) 33214, DOI: 10.1021/jacs.4c12888.

[53] T. B. O. Costa et al., *Impact of UV-B Photoaging on Chlorpyrifos Adsorption by PET Microplastics: Insights from Experimental and DFT Analysis*, en,
ACS Omega **9** (2024) 46439, DOI: 10.1021/acsomega.4c07521.

[54] É. M. Foyle, R. J. Goodwin, C. J. T. Cox, B. R. Smith, A. L. Colebatch, and N. G. White,
Expedient Decagram-Scale Synthesis of Robust Organic Cages That Bind Sulfate Strongly and Selectively in Water, en, J. Am. Chem. Soc. **146** (2024) 27127,
DOI: 10.1021/jacs.4c09930.

[55] Mayo, Stephen L, Olafson, Barry D, and Goddard, William A,
DREIDING: a generic force field for Molecular Simulations, J. Phys. Chem. **94** (1990) 8897,
DOI: 10.1021/j100389a010.

[56] 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, DOI: 10.1021/ja00051a040.

[57] T. A. Halgren, *Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94*, en, J. Comput. Chem. **17** (1996) 490,
DOI: 10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P.

[58] P. Li and K. M. Merz, *Metal Ion Modeling Using Classical Mechanics*, en,
Chem. Rev. **117** (2017) 1564, DOI: 10.1021/acs.chemrev.6b00440.

[59] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids*, en,
J. Am. Chem. Soc. **118** (1996) 11225, DOI: 10.1021/ja9621760.

[60] C. Lu et al.,
OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space, en,
J. Chem. Theory Comput. **17** (2021) 4291, DOI: 10.1021/acs.jctc.1c00302.

[61] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma,
The missing term in effective pair potentials, en, J. Phys. Chem. **91** (1987) 6269,
DOI: 10.1021/j100308a038.

[62] R. Fuentes-Azcatl, N. Mendoza, and J. Alejandre,
Improved SPC force field of water based on the dielectric constant: SPC/ ϵ, en,
Phys. A (Amsterdam, Neth.) **420** (2015) 116, DOI: 10.1016/j.physa.2014.10.072.

[63] H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. De Jong, S. M. Gopal, X. Periole, and S. J. Marrink, *The power of coarse graining in biomolecular simulations*, en,
WIREs Comput. Mol. Sci. **4** (2014) 225, DOI: 10.1002/wcms.1169.

[64] P. C. T. Souza et al.,
Martini 3: a general purpose force field for coarse-grained molecular dynamics, en,
Nat. Methods **18** (2021) 382, DOI: 10.1038/s41592-021-01098-3.

[65] S. Grimme and contributors, *xtb version 6.7.1*,
GitHub repository, available at github.com/grimme-lab/xtb, 2024.

[66] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, *Machine learning for molecular and materials science*, en, Nature **559** (2018) 547, DOI: 10.1038/s41586-018-0337-2.

[67] M. Meuwly, *Machine Learning for Chemical Reactions*, en, Chem. Rev. **121** (2021) 10218, DOI: 10.1021/acs.chemrev.1c00033.

[68] S. Kim et al., *PubChem 2023 update*, en, Nucleic Acids Res. **51** (2023) D1373, DOI: 10.1093/nar/gkac956.

[69] S. Gražulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. Le Bail, *Crystallography Open Database – an open-access collection of crystal structures*, J. Appl. Crystallogr. **42** (2009) 726, DOI: 10.1107/S0021889809016690.

[70] S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quirós, N. R. Serebryanaya, P. Moeck, R. T. Downs, and A. Le Bail, *Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration*, en, Nucleic Acids Res. **40** (2012) D420, DOI: 10.1093/nar/gkr900.

[71] H. E. Pence and A. Williams, *ChemSpider: An Online Chemical Information Resource*, en, J. Chem. Educ. **87** (2010) 1123, DOI: 10.1021/ed100697w.

[72] C. Knox et al., *DrugBank 6.0: the DrugBank Knowledgebase for 2024*, en, Nucleic Acids Res. **52** (2024) D1265, DOI: 10.1093/nar/gkad976.

[73] H. M. Berman, *The Protein Data Bank*, Nucleic Acids Res. **28** (2000) 235, DOI: 10.1093/nar/28.1.235.

[74] J. S. Smith, R. Zubatyuk, B. Nebgen, N. Lubbers, K. Barros, A. E. Roitberg, O. Isayev, and S. Tretiak, *The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules*, en, Sci. Data **7** (2020) 134, DOI: 10.1038/s41597-020-0473-z.

[75] R. Zubatyuk, J. S. Smith, J. Leszczynski, and O. Isayev, *Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network*, en, Sci. Adv. **5** (2019) eaav6490, DOI: 10.1126/sciadv.aav6490.

[76] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, *Less is more: Sampling chemical space with active learning*, en, J. Chem. Phys. **148** (2018) 241733, DOI: 10.1063/1.5023802.

[77] S. Zhang et al., *Exploring the frontiers of condensed-phase chemistry with a general reactive machine learning potential*, en, Nat. Chem. **16** (2024) 727, DOI: 10.1038/s41557-023-01427-3.

[78] C. Hölzer, R. Oerder, S. Grimme, and J. Hamaekers, *ConfRank: Improving GFN-FF Conformer Ranking with Pairwise Training*, en, J. Chem. Inf. Model. **64** (2024) 8909, DOI: 10.1021/acs.jcim.4c01524.

[79] A. Cho, H. Yun, J. H. Park, S. Y. Lee, and S. Park, *Prediction of novel synthetic pathways for the production of desired chemicals*, en, BMC Syst. Biol. **4** (2010) 35, DOI: 10.1186/1752-0509-4-35.

[80] J. Li and M. D. Eastgate, *Making better decisions during synthetic route design: leveraging prediction to achieve greenness-by-design*, en, React. Chem. Eng. **4** (2019) 1595, DOI: 10.1039/C9RE00019D.

[81] M. Tafipolsky, S. Amirjalayer, and R. Schmid, *Ab initio parametrized MM3 force field for the metal-organic framework MOF-5*, en, J. Comput. Chem. **28** (2007) 1169, DOI: 10.1002/jcc.20648.

[82] Y. Eken, N. M. S. Almeida, C. Wang, and A. K. Wilson, *SAMPL7: Host–guest binding prediction by molecular dynamics and quantum mechanics*, en, J. Comput.-Aided Mol. Des. **35** (2021) 63, DOI: 10.1007/s10822-020-00357-3.

[83] R. Sure, M. El Mahdali, A. Plajer, and P. Deglmann, *Towards a converged strategy for including microsolvation in reaction mechanism calculations*, en, J. Comput.-Aided Mol. Des. **35** (2021) 473, DOI: 10.1007/s10822-020-00366-2.

[84] 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*, en, J. Chem. Theory Comput. **18** (2022) 3174, DOI: 10.1021/acs.jctc.2c00239.

[85] F. Jensen, *Introduction to computational chemistry*, 2nd ed, OCLC: ocm70707839, Chichester, England ; Hoboken, NJ: John Wiley & Sons, 2007.

[86] A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*, eng, Mineola, New York: Dover Publications, Inc, 2012.

[87] A. Klamt and G. Schüürmann, *COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient*, en, J. Chem. Soc., Perkin Trans. 2 (1993) 799, DOI: 10.1039/P29930000799.

[88] A. Klamt, *Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena*, en, J. Phys. Chem. **99** (1995) 2224, DOI: 10.1021/j100007a062.

[89] V. Barone and M. Cossi, *Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model*, en, J. Phys. Chem. A **102** (1998) 1995, DOI: 10.1021/jp9716997.

[90] A. Onufriev, D. Bashford, and D. A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized born model*, en, Proteins:Struct., Funct., Bioinf. **55** (2004) 383, DOI: 10.1002/prot.20033.

[91] 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*, en, J. Phys. Chem. B **113** (2009) 6378, DOI: 10.1021/jp810292n.

[92] K. Mathew, R. Sundararaman, K. Letchworth-Weaver, T. A. Arias, and R. G. Hennig, *Implicit solvation model for density-functional study of nanocrystal surfaces and reaction pathways*, en, J. Chem. Phys. **140** (2014) 084106, DOI: 10.1063/1.4865107.

[93] S. Ehlert, M. Stahn, S. Spicher, and S. Grimme,
Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods, en,
J. Chem. Theory Comput. **17** (2021) 4250, DOI: 10.1021/acs.jctc.1c00471.

[94] S. Grimme, *Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory*, en, Chem. - Eur. J. **18** (2012) 9955, DOI: 10.1002/chem.201200497.

[95] P. Pracht and S. Grimme,
Calculation of absolute molecular entropies and heat capacities made simple, en,
Chem. Sci. **12** (2021) 6551, DOI: 10.1039/D1SC00621E.

[96] 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), Publisher: American Chemical Society 4039,
DOI: 10.1021/acs.jpca.1c00971.

[97] J. H. Jensen, *Predicting accurate absolute binding energies in aqueous solution: thermodynamic considerations for electronic structure methods*, en,
Phys. Chem. Chem. Phys. **17** (2015) 12441, DOI: 10.1039/C5CP00628G.

[98] L. J. Butler, *CHEMICAL REACTION DYNAMICS BEYOND THE BORN-OPPENHEIMER APPROXIMATION*, en, Annu. Rev. Phys. Chem. **49** (1998) 125,
DOI: 10.1146/annurev.physchem.49.1.125.

[99] G. A. Worth and L. S. Cederbaum,
BEYOND BORN-OPPENHEIMER: Molecular Dynamics Through a Conical Intersection, en,
Annu. Rev. Phys. Chem. **55** (2004) 127,
DOI: 10.1146/annurev.physchem.55.091602.094335.

[100] E. R. Davidson, *Computational Transition Metal Chemistry*, en, Chem. Rev. **100** (2000) 351,
DOI: 10.1021/cr980385s.

[101] P. Hohenberg and W. Kohn, *Inhomogeneous Electron Gas*, en, Phys. Rev. **136** (1964) B864,
DOI: 10.1103/PhysRev.136.B864.

[102] W. Kohn and L. J. Sham,
Self-Consistent Equations Including Exchange and Correlation Effects, en,
Phys. Rev. **140** (1965) A1133, DOI: 10.1103/PhysRev.140.A1133.

[103] W. Kohn,
Nobel Lecture: Electronic structure of matter—wave functions and density functionals, en,
Rev. Mod. Phys. **71** (1999) 1253, DOI: 10.1103/RevModPhys.71.1253.

[104] 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, en,
Phys. Chem. Chem. Phys. **19** (2017) 32184, DOI: 10.1039/C7CP04913G.

[105] P. Mishra, Y. Yamamoto, J. K. Johnson, K. A. Jackson, R. R. Zope, and T. Baruah,
Study of self-interaction-errors in barrier heights using locally scaled and Perdew–Zunger self-interaction methods, en, J. Chem. Phys. **156** (2022) 014306, DOI: 10.1063/5.0070893.

[106] Kurth, Stefan, Perdew, John P., and Blaha, Peter, *Molecular and solid-state tests of density functional approximations: LSD, GGAs, and meta-GGAs*, **75** (1999) 889.

[107] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka, *Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits*, en, J. Chem. Phys. **123** (2005) 062201, DOI: 10.1063/1.1904565.

[108] A. D. Becke, *Density-functional exchange-energy approximation with correct asymptotic behavior*, en, Phys. Rev. A **38** (1988) 3098, DOI: 10.1103/PhysRevA.38.3098.

[109] C. Lee, W. Yang, and R. G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*, en, Phys. Rev. B **37** (1988) 785, DOI: 10.1103/PhysRevB.37.785.

[110] Y. Zhao and D. G. Truhlar, *A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions*, en, J. Chem. Phys. **125** (2006) 194101, DOI: 10.1063/1.2370993.

[111] C. Adamo and V. Barone, *Toward reliable density functional methods without adjustable parameters: The PBE0 model*, en, J. Chem. Phys. **110** (1999) 6158, DOI: 10.1063/1.478522.

[112] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, *Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes*, en, J. Chem. Phys. **119** (2003) 12129, DOI: 10.1063/1.1626543.

[113] 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*, en, J. Chem. Theory Comput. **7** (2011) 291, DOI: 10.1021/ct100466k.

[114] S. Kozuch and J. M. L. Martin, *DSD-PBEP86: in search of the best double-hybrid DFT with spin-component scaled MP2 and dispersion corrections*, en, Phys. Chem. Chem. Phys. **13** (2011) 20104, DOI: 10.1039/c1cp22592h.

[115] S. H. Vosko, L. Wilk, and M. Nusair, *Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis*, en, Can. J. Phys. **58** (1980) 1200, DOI: 10.1139/p80-159.

[116] J. P. Perdew and Y. Wang, *Accurate and simple analytic representation of the electron-gas correlation energy*, en, Phys. Rev. B **45** (1992) 13244, DOI: 10.1103/PhysRevB.45.13244.

[117] W. J. Carr, *Energy, Specific Heat, and Magnetic Properties of the Low-Density Electron Gas*, en, Phys. Rev. **122** (1961) 1437, DOI: 10.1103/PhysRev.122.1437.

[118] W. J. Carr and A. A. Maradudin, *Ground-State Energy of a High-Density Electron Gas*, en, Phys. Rev. **133** (1964) A371, DOI: 10.1103/PhysRev.133.A371.

[119] A. D. Becke, *Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals*, en, J. Chem. Phys. **107** (1997) 8554, DOI: 10.1063/1.475007.

[120] S. F. Sousa, P. A. Fernandes, and M. J. Ramos,
General Performance of Density Functionals, en, J. Phys. Chem. A **111** (2007) 10439, DOI: 10.1021/jp0734474.

[121] I. C. Gerber and J. G. Ángyán, *Hybrid functional with separated range*, en, Chem. Phys. Lett. **415** (2005) 100, DOI: 10.1016/j.cplett.2005.08.060.

[122] A. V. Krukau, G. E. Scuseria, J. P. Perdew, and A. Savin,
Hybrid functionals with local range separation, en, J. Chem. Phys. **129** (2008) 124103, DOI: 10.1063/1.2978377.

[123] J. L. Bao, L. Gagliardi, and D. G. Truhlar,
Self-Interaction Error in Density Functional Theory: An Appraisal,
J. Phys. Chem. Lett. **9** (2018) 2353, DOI: 10.1021/acs.jpclett.8b00242.

[124] L. Goerigk and S. Grimme, *Double-hybrid density functionals*, en, WIREs Comput. Mol. Sci. **4** (2014) 576, DOI: 10.1002/wcms.1193.

[125] C. Møller and M. S. Plesset,
Note on an Approximation Treatment for Many-Electron Systems, en,
Phys. Rev. **46** (1934) 618, DOI: 10.1103/PhysRev.46.618.

[126] D. Bohm and D. Pines,
A Collective Description of Electron Interactions. I. Magnetic Interactions, en,
Phys. Rev. **82** (1951) 625, DOI: 10.1103/PhysRev.82.625.

[127] D. Pines and D. Bohm, *A Collective Description of Electron Interactions: II. Collective vs Individual Particle Aspects of the Interactions*, en, Phys. Rev. **85** (1952) 338, DOI: 10.1103/PhysRev.85.338.

[128] D. Bohm and D. Pines, *A Collective Description of Electron Interactions: III. Coulomb Interactions in a Degenerate Electron Gas*, en, Phys. Rev. **92** (1953) 609, DOI: 10.1103/PhysRev.92.609.

[129] R. Sure, J. G. Brandenburg, and S. Grimme,
Small Atomic Orbital Basis Set First-Principles Quantum Chemical Methods for Large Molecular and Periodic Systems: A Critical Analysis of Error Sources, en,
ChemistryOpen **5** (2016) 94, DOI: 10.1002/open.201500192.

[130] S. Boys and F. Bernardi, *The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors*, en,
Molecular Physics **19** (1970) 553, DOI: 10.1080/00268977000101561.

[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, en,
J. Chem. Phys. **136** (2012) 154101, DOI: 10.1063/1.3700154.

[132] J. Witte, J. B. Neaton, and M. Head-Gordon,
Push it to the limit: Characterizing the convergence of common sequences of basis sets for intermolecular interactions as described by density functional theory, en,
J. Chem. Phys. **144** (2016) 194306, DOI: 10.1063/1.4949536.

[133] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and S. Grimme, *Extended tight-binding quantum chemistry methods*, en, WIREs Comput. Mol. Sci. **11** (2021), _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1493 e1493, DOI: 10.1002/wcms.1493.

[134] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, *Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties*, en, Phys. Rev. B **58** (1998) 7260, DOI: 10.1103/PhysRevB.58.7260.

[135] A. S. Christensen, T. Kubař, Q. Cui, and M. Elstner, *Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications*, en, Chem. Rev. **116** (2016) 5301, DOI: 10.1021/acs.chemrev.5b00584.

[136] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, *A generally applicable atomic-charge dependent London dispersion correction*, en, J. Chem. Phys. **150** (2019) 154122, DOI: 10.1063/1.5090222.

[137] 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), Publisher: American Chemical Society 1652, DOI: 10.1021/acs.jctc.8b01176.

[138] D. Setiawan, E. Kraka, and D. Cremer, *Strength of the Pnicogen Bond in Complexes Involving Group Va Elements N, P, and As*, en, J. Phys. Chem. A **119** (2015) 1642, DOI: 10.1021/jp508270g.

[139] J. Pérez-Jordá and A. Becke, *A density-functional study of van der Waals forces: rare gas diatomics*, en, Chem. Phys. Lett. **233** (1995) 134, DOI: 10.1016/0009-2614(94)01402-H.

[140] E. R. Johnson and G. A. DiLabio, *Structure and binding energies in van der Waals dimers: Comparison between density functional theory and correlated ab initio methods*, en, Chem. Phys. Lett. **419** (2006) 333, DOI: 10.1016/j.cplett.2005.11.099.

[141] S. Kristyán and P. Pulay, *Can (semi)local density functional theory account for the London dispersion forces?* en, Chem. Phys. Lett. **229** (1994) 175, DOI: 10.1016/0009-2614(94)01027-7.

[142] S. Grimme, *Density functional theory with London dispersion corrections*, en, WIREs Comput. Mol. Sci. **1** (2011) 211, DOI: 10.1002/wcms.30.

[143] S. Spicher, E. Caldeweyher, A. Hansen, and S. Grimme, *Benchmarking London dispersion corrected density functional theory for noncovalent ion-π interactions*, en, Phys. Chem. Chem. Phys. **23** (2021) 11635, DOI: 10.1039/D1CP01333E.

[144] L. Wittmann, H. Neugebauer, S. Grimme, and M. Bursch, *Dispersion-corrected r2SCAN based double-hybrid functionals*, en, J. Chem. Phys. **159** (2023) 224103, DOI: 10.1063/5.0174988.

[145] 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, en, J. Chem. Phys. **132** (2010) 154104, DOI: 10.1063/1.3382344.

[146] O. A. Vydrov and T. Van Voorhis,
Nonlocal van der Waals density functional: The simpler the better, en,
J. Chem. Phys. **133** (2010) 244103, DOI: 10.1063/1.3521275.

[147] O. A. Vydrov and T. Van Voorhis,
Nonlocal van der Waals Density Functional Made Simple, en,
Phys. Rev. Letters **103** (2009) 063004, DOI: 10.1103/PhysRevLett.103.063004.

[148] J. Klimeš, D. R. Bowler, and A. Michaelides,
Chemical accuracy for the van der Waals density functional,
J. Phys.:Condens. Matter **22** (2010) 022201, DOI: 10.1088/0953-8984/22/2/022201.

[149] A. Tkatchenko and M. Scheffler, *Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data*, en,
Phys. Rev. Letters **102** (2009) 073005, DOI: 10.1103/PhysRevLett.102.073005.

[150] A. Ambrosetti, A. M. Reilly, R. A. DiStasio, and A. Tkatchenko,
Long-range correlation energy calculated from coupled atomic response functions, en,
J. Chem. Phys. **140** (2014) 18A508, DOI: 10.1063/1.4865104.

[151] A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler,
Accurate and Efficient Method for Many-Body van der Waals Interactions, en,
Phys. Rev. Letters **108** (2012) 236402, DOI: 10.1103/PhysRevLett.108.236402.

[152] J.-D. Chai and M. Head-Gordon, *Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections*, en, Phys. Chem. Chem. Phys. **10** (2008) 6615, DOI: 10.1039/b810189b.

[153] P. M. Morse,
Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels, en,
Phys. Rev. **34** (1929) 57, DOI: 10.1103/PhysRev.34.57.

[154] R. A. Buckingham, *The classical equation of state of gaseous helium, neon and argon*, en,
Proc. R. Soc. London, Ser. A **168** (1938) 264, DOI: 10.1098/rspa.1938.0173.

[155] W. A. Steele, *The physical interaction of gases with crystalline solids*, en,
Surf. Sci. **36** (1973) 317, DOI: 10.1016/0039-6028(73)90264-1.

[156] 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)*, en,
J. Chem. Theory Comput. **13** (2017) 1989, DOI: 10.1021/acs.jctc.7b00118.

[157] S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen, *Consistent structures and interactions by density functional theory with small atomic orbital basis sets*, en,
J. Chem. Phys. **143** (2015) 054107, DOI: 10.1063/1.4927476.

[158] J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme,
B97-3c: A revised low-cost variant of the B97-D density functional method, en,
J. Chem. Phys. **148** (2018) 064104, DOI: 10.1063/1.5012601.

[159] U. K. Deiters, *Efficient Coding of the Minimum Image Convention*, en,
Z. Phys. Chem. (Berlin, Ger.) **227** (2013) 345, DOI: 10.1524/zpch.2013.0311.

[160] F. Knuth, C. Carbogno, V. Atalla, V. Blum, and M. Scheffler, *All-electron formalism for total energy strain derivatives and stress tensor components for numeric atom-centered orbitals*, en,
Comput. Phys. Commun. **190** (2015) 33, DOI: 10.1016/j.cpc.2015.01.003.

[161] S. K. Godunov and E. I. Romenskii,
Elements of Continuum Mechanics and Conservation Laws, en,
Boston, MA: Springer US, 2003, DOI: 10.1007/978-1-4757-5117-8.

[162] P. P. Ewald, *Die Berechnung optischer und elektrostatischer Gitterpotentiale*, en,
Ann. Phys. (Berlin, Ger.) **369** (1921) 253, DOI: 10.1002/andp.19213690304.

[163] D. Wolf, P. Keblinski, S. R. Phillpot, and J. Eggebrecht, *Exact method for the simulation of Coulombic systems by spherically truncated, pairwise r-1 summation*, en,
J. Chem. Phys. **110** (1999) 8254, DOI: 10.1063/1.478738.

[164] A. Laio and M. Parrinello, *Escaping free-energy minima*, en,
Proc. Natl. Acad. Sci. U. S. A. **99** (2002) 12562, DOI: 10.1073/pnas.202427399.

[165] P. Pracht and S. Grimme, *Efficient Quantum-Chemical Calculations of Acid Dissociation Constants from Free-Energy Relationships*, en, J. Phys. Chem. A **125** (2021) 5681,
DOI: 10.1021/acs.jpca.1c03463.

[166] N. Metropolis and S. Ulam, *The Monte Carlo Method*, en, J. Am. Stat. Assoc. **44** (1949) 335,
DOI: 10.1080/01621459.1949.10483310.

[167] W. Jorgensen, *Biological reactions are slow in absence of a catalyst*,
Adv. Chem. Phys **70** (1988) 469.

[168] J. Kiefer, *Sequential minimax search for a maximum*, en, Proc. Am. Math. Soc. **4** (1953) 502,
DOI: 10.1090/S0002-9939-1953-0055639-3.

[169] A. M. Reilly and A. Tkatchenko, *Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals*, en,
J. Chem. Phys. **139** (2013) 024705, DOI: 10.1063/1.4812819.

[170] G. A. Dolgonos, J. Hoja, and A. D. Boese,
Revised values for the X23 benchmark set of molecular crystals, en,
Phys. Chem. Chem. Phys. **21** (2019) 24333, DOI: 10.1039/C9CP04488D.

[171] J. D. Gale, L. M. LeBlanc, P. R. Spackman, A. Silvestri, and P. Raiteri, *A Universal Force Field for Materials, Periodic GFN-FF: Implementation and Examination*, en,
J. Chem. Theory Comput. **17** (2021) 7827, DOI: 10.1021/acs.jctc.1c00832.

[172] J. G. Brandenburg, T. Maas, and S. Grimme, *Benchmarking DFT and semiempirical methods on structures and lattice energies for ten ice polymorphs*, en,
J. Chem. Phys. **142** (2015) 124104, DOI: 10.1063/1.4916070.

[173] Y. N. Heit and G. J. O. Beran, *How important is thermal expansion for predicting molecular crystal structures and thermochemistry at finite temperatures?*
Acta Crystallogr., Sect. B:Struct. Sci., Cryst. Eng. Mater. **72** (2016) 514, DOI: 10.1107/S2052520616005382.

[174] R. Mittal, M. Gupta, and S. Chaplot,
Phonons and anomalous thermal expansion behaviour in crystalline solids, en,
Prog. Mater. Sci. **92** (2018) 360, DOI: 10.1016/j.pmatsci.2017.10.002.

[175] L. Maschio, B. Civalleri, P. Ugliengo, and A. Gavezzotti,
Intermolecular Interaction Energies in Molecular Crystals: Comparison and Agreement of Localized Møller–Plesset 2, Dispersion-Corrected Density Functional, and Classical Empirical Two-Body Calculations, en, J. Phys. Chem. A **115** (2011) 11179,
DOI: 10.1021/jp203132k.

[176] J. G. Brandenburg and S. Grimme, *Organic crystal polymorphism: a benchmark for dispersion-corrected mean-field electronic structure methods*,
Acta Crystallogr., Sect. B:Struct. Sci., Cryst. Eng. Mater. **72** (2016) 502,
DOI: 10.1107/S2052520616007885.

[177] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun,
Accurate and Numerically Efficient r ^{2} SCAN Meta-Generalized Gradient Approximation, en,
J. Phys. Chem. Lett. **11** (2020) 8208, DOI: 10.1021/acs.jpclett.0c02405.

[178] E. Caldeweyher, J.-M. Mewes, S. Ehlert, and S. Grimme,
Extension and evaluation of the D4 London-dispersion model for periodic systems, en,
Phys. Chem. Chem. Phys. **22** (2020) 8499, DOI: 10.1039/D0CP00502A.

[179] V. V. Utochnikova, "Lanthanide complexes as OLED emitters," en,
Handbook on the Physics and Chemistry of Rare Earths, vol. 59, Elsevier, 2021 1,
DOI: 10.1016/bs.hpcre.2021.05.001.

[180] S. V. Yudintsev, *Isolation of Separated Waste of Nuclear Industry*, en,
Radiochemistry **63** (2021) 527, DOI: 10.1134/S1066362221050015.

[181] A. P. Bidkar, L. Zerefa, S. Yadav, H. F. VanBrocklin, and R. R. Flavell,
Actinium-225 targeted alpha particle therapy for prostate cancer, en,
Theranostics **14** (2024) 2969, DOI: 10.7150/thno.96403.

[182] N. Ree, A. H. Göller, and J. H. Jensen,
Automated quantum chemistry for estimating nucleophilicity and electrophilicity with applications to retrosynthesis and covalent inhibitors, en, Digital Discovery **3** (2024) 347,
DOI: 10.1039/D3DD00224A.

[183] E. Karunaratne, D. W. Hill, K. Dührkop, S. Böcker, and D. F. Grant,
Combining Experimental with Computational Infrared and Mass Spectra for High-Throughput Nontargeted Chemical Structure Identification, en,
Anal. Chem. **95** (2023) 11901, DOI: 10.1021/acs.analchem.3c00937.

[184] M. Strandgaard, J. Seumer, and J. H. Jensen,
Discovery of molybdenum based nitrogen fixation catalysts with genetic algorithms, en,
Chem. Sci. **15** (2024) 10638, DOI: 10.1039/D4SC02227K.

[185] C. Hölzer, I. Gordiy, S. Grimme, and M. Bursch, *Hybrid DFT Geometries and Properties for 17k Lanthanoid Complexes - The LnQM Data Set*, en, J. Chem. Inf. Model. **64** (2024) 825, DOI: 10.1021/acs.jcim.3c01832.

[186] L. Wittmann, I. Gordiy, M. Friede, B. Helmich-Paris, S. Grimme, A. Hansen, and M. Bursch, *Extension of the D3 and D4 London dispersion corrections to the full actinides series*, en, Phys. Chem. Chem. Phys. **26** (2024) 21379, DOI: 10.1039/D4CP01514B.

[187] K. Lv, C. Urbank, M. Patzschke, J. März, P. Kaden, S. Weiss, and M. Schmidt, *MOFs with 12-Coordinate 5f-Block Metal Centers*, en, J. Am. Chem. Soc. **144** (2022) 2879, DOI: 10.1021/jacs.1c13127.

[188] S. M. Pratik, L. Gagliardi, and C. J. Cramer, *Engineering Electrical Conductivity in Stable Zirconium-Based PCN-222 MOFs with Permanent Mesoporosity*, Chem. Mater. **32** (2020) 6137, DOI: 10.1021/acs.chemmater.0c01847.

[189] Z. Liang, H.-Y. Wang, H. Zheng, W. Zhang, and R. Cao, *Porphyrin-Based Frameworks for Oxygen Electrocatalysis and Catalytic Reduction of Carbon Dioxide*, Chem. Soc. Rev. **50** (2021) 2540, DOI: 10.1039/D0CS01482F.

[190] H. JÓNSSON, G. MILLS, and K. W. JACOBSEN, "Nudged elastic band method for finding minimum energy paths of transitions," *Classical and Quantum Dynamics in Condensed Phase Simulations* 385, DOI: 10.1142/9789812839664_0016.

[191] G. Mills, H. Jónsson, and G. K. Schenter, *Reversible work transition state theory: application to dissociative adsorption of hydrogen*, Surface Science **324** (1995) 305, DOI: https://doi.org/10.1016/0039-6028(94)00731-4.

[192] V. Ásgeirsson, B. O. Birgisson, R. Bjornsson, U. Becker, F. Neese, C. Riplinger, and H. Jónsson, *Nudged Elastic Band Method for Molecular Reactions Using Energy-Weighted Springs Combined with Eigenvector Following*, J. Chem. Theory Comput. **17** (2021) 4929, DOI: 10.1021/acs.jctc.1c00462.

[193] J. Huang and A. D. MacKerell, *CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data*, en, J. Comput. Chem. **34** (2013) 2135, DOI: 10.1002/jcc.23354.

[194] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB*, en, J. Chem. Theory Comput. **11** (2015) 3696, DOI: 10.1021/acs.jctc.5b00255.

[195] M. Gaus, Q. Cui, and M. Elstner, *DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB)*, en, J. Chem. Theory Comput. **7** (2011) 931, DOI: 10.1021/ct100684s.

[196] J. S. Smith, O. Isayev, and A. E. Roitberg, *ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules*, Sci. Data **4** (2017) 170193, DOI: 10.1038/sdata.2017.193.

[197] H. Wang, L. Zhang, J. Han, and W. E, *DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics*, en,
Comput. Phys. Commun. **228** (2018) 178, DOI: 10.1016/j.cpc.2018.03.016.

[198] J. Zeng et al., *DeePMD-kit v2: A software package for deep potential models*, en,
J. Chem. Phys. **159** (2023) 054801, DOI: 10.1063/5.0155600.

Acknowledgements

Finishing this thesis was only possible with the support and encouragement of many amazing people along the way. I'm deeply grateful to everyone who helped make this possible and I want to thank several people. First and foremost, I sincerely thank my supervisor, Stefan Grimme, for the opportunity to work on many exciting projects and for his guidance throughout my studies. I am especially grateful for the chance to attend the MQM conference in Blacksburg in 2020 and the STC conference in Heidelberg in 2022, both of which were valuable and inspiring experiences. I also want to thank Prof. Bredow for being my second reviewer and for taking the time to explain the intricacies of VASP to me. Many thanks to Sebastian Spicher and Sebastian Ehlert, who got me started on my work with the GFN-FF model. I am grateful to my collaborators, Jan-Michael Mewes and Markus Bursch, for sharing their expertise in lanthanide and actinide chemistry, as well as to Gustavo Adolfo Lara-Cruz for a great collaboration on workflow development and for a great time during his stay here. Special thanks to Benedikt Bädorf, Thomas Gasevic, Hagen Neugebauer, and Christoph Plett for proofreading my thesis. Further, I would also like to thank the developers of ChatGPT for creating such an invaluable tool. It has been a great help throughout my research. The friendly and supportive atmosphere in our group has been a constant source of motivation. I want to thank my colleagues Fabian Bohle, Marvin Friede, Christian Hölzer, Julia Kohn, Abylay Katbashev, Lukas Kunze, Sarah Löffelsender, Marcel Müller, Marcel Stahn, and Lukas Wittman. My sincere thanks also go to Claudia Kronz and Jens Mekelburger for their administrative and technical support.

My family's constant encouragement and unwavering support have been invaluable to me, and I am deeply grateful for it. I am also thankful for all my old and new friends who thought it would be a good idea to spend more time with me. Thank you for joining me in bouldering, ultimate frisbee, canoeing, and enjoying game nights together. I am happy to have made many wonderful memories that shifted my perspective toward the brighter side of life, especially when my code was doing what I programmed – and not what I wanted it to do.