Overview on Building Blocks and Applications of Efficient and Robust Extended Tight Binding

Abylay Katbashev,[†] Marcel Stahn,[†,§] Thomas Rose,[†] Vahideh Alizadeh,[†] Marvin Friede,[†] Christoph Plett,[†] Pit Steinbach,[‡] and Sebastian Ehlert*,[¶]

[†]Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany
[‡]Institute of Physical Chemistry, RWTH Aachen University, Melatener Str. 20, 52074 Aachen, Germany
[¶]AI for Science, Microsoft Research, Evert van de Beekstraat 354, 1118 CZ Schiphol, The Netherlands
[§]now at OpenEye, Cadence Molecular Sciences, Ebertplatz 1, 50668 Cologne, Germany

E-mail: sehlert@microsoft.com

Abstract

The extended tight binding (xTB) family of methods opened many new possibilities in the field of computational chemistry. Within just five years, the GFN2-xTB parametrization for all elements up to Z=86 enabled more than a thousand applications, which were previously not feasible with other electronic structure methods. The xTB methods provide a robust and efficient way to apply quantum mechanics based approaches for obtaining molecular geometries, computing free energy corrections or describing non-covalent interactions and found applicability for many more targets. A crucial contribution to the success of the xTB methods is the availability within many simulation packages and frameworks, supported by the open source development of its program library and packages. We present a comprehensive summary of the applications and capabilities of xTB methods in different fields of chemistry. Moreover, we consider the main software packages for xTB calculations, covering their current ecosystem, novel features, and usage by the scientific community.

1 Introduction

For the computational modeling of complex chemical reactions, Kohn–Sham density functional theory (KS-DFT)[1] is one of the most widely used electronic structure methods.[2–4] However, due to its high computational cost, KS-DFT is practically out of reach for larger systems and molecular dynamics on long timescales. In some cases, atomic potentials or force fields offer a low-cost alternative, but the lack of explicit consideration of electronic structure makes them unsuitable for studying chemical reactions or different electronic states.

Simplified electronic structure methods like tight binding (TB) theory[5,6] offer a great reduction of computational cost compared to KS-DFT, while keeping a minimal representation of the electrons which can be conveniently combined or augmented with atomic potentials or force field related models. However, the reduction in computational complexity compared to KS-DFT is offset by introducing (semi-)empirical model parameters. The challenge of obtaining parameters for many different element combinations for a wide range of applications in a chemically diverse space is addressed in the extended tight binding (xTB) Hamiltonian[5] with the corresponding geometry, frequency, and noncovalent interaction (GFN) parameterizations.[7,8] Since the xTB Hamiltonian is based on KS-DFT, it has the flexibility to be adapted and improved for describing properties of interest like, e.g., spin-splittings,[9] magnetic properties,[10] and excited states.[11]

Furthermore, the xTB methods and their parametrizations offer the possibility to tailor the model specifically for a system of interest, thereby enhancing method performance. In this regard, reparametrization techniques, whether through classical frameworks[12] or automated differentiation[13] facilitate the fine-tuning of parameters based on existing parameterizations.[14–16] It is worth noting that xTB methods are also utilized in machine learning (ML), for instance, as dataset generation tools that can be used to train more simplified models. For this, either the xTB method can be used to label data sets,[17–19] or intermediate and output quantities of xTB can be used as feature inputs to models,[20–22] targeting more accurate labels obtained from KS-DFT, higher levels of theory, or experiment.

In general, xTB methods are available for a wide range of applications similar to those of KS-DFT, but at a reduced computational cost and a compromise on the accuracy. Geometry optimizations present one of the primary applications, either for obtaining initial structure guesses, for instance, preoptimizing solvent–solute clusters,[23] screening numerous geometries to, for example, identify a global minimum from thousands of isomer structures,[24] or tackling systems where KS-DFT is no longer feasible, such as in describing large oligomer structures.[25]

Furthermore, the accessibility of analytical energy derivatives (forces) enables the calculation of Hessians to characterize the potential energy surface (PES) and to compute thermo-

statistical corrections to free energy. The latter was employed, for instance, to calculate the free energy of compounds relevant to photochemical reactions.[26,27] Additionally, the impact of solvation, crucial for accurate free energy calculations, can be readily integrated implicitly across all previously mentioned applications,[28,29] enabling the computation of free energies in solution.

However, implicit solvation models might not be able to capture all relevant solvent-solute interactions, and break down, e.g., for participation of a solvent molecule in a reaction. Then, cluster approaches explicitly accounting for the solvent molecules are required. As these additional molecules strongly increase the computational effort, especially in the solute-solvent cluster generation step, KS-DFT is often not applicable, leaving efficient xTB methods a reasonable choice for this task. While they allow for the calculation of larger systems, the accuracy of KS-DFT is sometimes necessary for describing certain types of chemistry, where the parametrizations are insufficient, the model in general is not flexible enough, or the approximations are too crude. In such cases, combined approaches like embedding or QM:QM (KS-DFT:TB) approaches can be used.[30–33]

Investigating the dynamics of the system of interest directly is another application for xTB methods where they offer a computationally feasible alternative to KS-DFT-based ab initio molecular dynamics, with the possibility to reach systems with millions of atoms.[34] Moreover, the combination of xTB methods with enhanced sampling techniques, such as metadynamics with collective variables (e.g. the root-mean-square displacement),[35,36] enables a quite general procedure for the exploration of potential energy surfaces (PES) as realized in the conformer-rotamer ensemble sampling tool (CREST) program package.[37,38] Finally, the application to materials and extended systems, like crystal structure prediction or simulation of liquids, is an emerging opportunity for the application of xTB methods.[39]

In this work, we focus on the current state of the main software packages for xTB methods: the *xtb* program suite and the *tblite* library. We begin with a brief theoretical excursion behind the xTB framework, explaining the basics of the Tight Binding theory and its GFN-specific parametrization. Following that, we describe the technical side of the xTB-based packages, including their code ecosystem, licensing, and their integration and interaction with the broader scientific computational environment. In the application section, we summarize the most common uses of xTB methods, namely conformer sampling, structure optimizations, AIMD simulations, and data generation. To complete the picture, we review several new features available in the *xtb* package: aISS docking tool, QM/MM module, mcGFNFF force field for molecular crystal calculations, and a post-SCF implicit solvation model CPCM-X. Finally, through citation dynamics, we demonstrate the significant influence of the xTB methods, and GFN2-xTB in particular, in the field of computational chemistry.

2 Theory and parametrization

In the following, a brief outline of the key fundamentals of the xTB framework is given. A starting point for the xTB Hamiltonian is a non-local KS-DFT expression:

$$E[\rho] = E_{nn} + T_s[\rho] + V_{ext}[\rho] + J[\rho] + E_{xc}[\rho] \quad (1)$$

where E_{nn} is the nuclear–nuclear repulsion, $T_s[\rho]$ the KS kinetic energy, $V_{ext}[\rho]$ the external potential of the nuclei, $J[\rho]$ the Coulomb repulsion, $E_{xc}[\rho]$ the exchange correlation energy, and ρ is the ground state electron density. The density ρ can be partitioned into two terms: the superposition of the atomic densities ρ_0 and the density fluctuations $\delta\rho$

$$\rho = \rho_0 + \delta\rho = \sum_{A}^{N_{at}} \rho_{0,A} + \delta\rho \quad (2)$$

which is also shown in a schematic way in Fig. 1 by depicting the density fluctuation for a molecule as the difference between total and
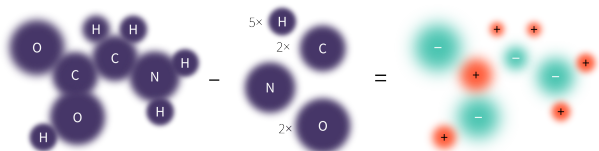
Figure 1: Schematic representation of partitioning of the total density ρ into atomic densities ρ_0 and density fluctuations $\delta\rho$. In this example, the difference in chemical environments is visible for the carbon atoms due to the qualitative difference in the sign of density fluctuations.

atomic densities.

Due to the non-linear appearance of the density in the exchange-correlation functional, for example, a series expansion of the energy in terms of density fluctuations is used as a central expression for all tight binding methods,

$$E[\rho] = E_0[\rho_0] + E_1[\rho_0, \delta\rho] + E_2[\rho_0, (\delta\rho)^2] + \dots \tag{3}$$

For the xTB methods, the energy functional is expanded up to third order in the density fluctuations. The main contributions to the electronic energy are given by the effective one-electron contribution, also known as extended Hückel theory (EHT), computed from the density matrix P with

$$E_1^{\mathrm{EHT}} = \sum_{\kappa\lambda} H_{\kappa\lambda} P_{\lambda\kappa} \tag{4}$$

where κ/λ are the indices of the atomic orbitals and the Hamiltonian H is given as

$$H_{\kappa\lambda} = \frac{h_\kappa + h_\lambda}{2} \cdot S_{\kappa\lambda} \cdot K(R_{\kappa\lambda}; Z_\kappa, Z_\lambda, \ell_\kappa, \ell_\lambda) \tag{5}$$

where S is the overlap, $h_{\kappa/\lambda}$ are atomic energy levels, and K is a non-linear function based on the interatomic distance parameterized for each element and angular momentum. The levels are made dependent on the local environment by a smooth counting function for coordinating atoms around each atomic site. The expression in Eq. 5 is central for the xTB methods, as it contains about half of the model parameters. Those are contained in the definition of the basis functions for computing the overlap matrix, the energy levels and their coordination number dependence, and the parametrization of the scaling function K.

For the second-order contribution to the energy, the density fluctuations are used to express the Coulomb energy $J[\rho]$. This is done by expanding the density fluctuations further in a series of atom-centered multipole moments $\xi_{\mathrm{A}}^{(\ell)}$

$$\delta\rho = \sum_{\mathrm{A}}^{N_{\mathrm{nuc}}} \sum_{\ell=0}^{L} \xi_{\mathrm{A}}^{(\ell)} \tag{6}$$

where the multipole moments are obtained by Mulliken population analysis from the respective moment integrals (overlap, dipole, quadrupole). The second-order energy expression for the Coulomb energy can then be written as

$$E_2^{\mathrm{coul}} = \sum_{\mathrm{AB}}^{N_{\mathrm{nuc}}} \sum_{\ell,\ell'=0}^{L} \xi_{\mathrm{A}}^{(\ell)} \cdot T^{(\ell\ell')}(R_{\mathrm{AB}}; Z_{\mathrm{A}}, Z_{\mathrm{B}}) \cdot \xi_{\mathrm{B}}^{(\ell')} \tag{7}$$

where the elements $T^{(\ell\ell')}$ are the interaction tensors between the multipole moments, with the correct long-range dependency and a parameterized short-range damping in the overlapping regime. For the GFN1-xTB method, the expansion is truncated already for the monopole moments, while in GFN2-xTB, contributions up to the second order in the energy (dipole-dipole and charge-quadruple) are included. Note that the short-range damping of this electrostatic energy implicitly accounts for correlation effects.

For electronically complicated small gap systems, an important contribution in terms of robustness is the electronic entropy $-T_{\mathrm{elec}}S$. It arises from allowing fractional occupation based on a Fermi distribution at finite electronic temperature and enables the handling of systems with static correlation in a reasonable, automatic way.

Since there is no explicit contribution accounting for spin-polarization, all xTB methods energetically favor low-spin solutions, even if the physical ground state would have a higher multiplicity. To overcome this limitation, spin-

4

polarization can be explicitly introduced in the Hamiltonian, allowing xTB to recover high-spin ground states correctly.[9]

Additional energy contributions are the effective nuclear repulsion E_0^{rep} and London dispersion E_n^{disp}. The short-ranged nuclear repulsion function parameterizes the combination of nuclear–nuclear repulsion E_{nn}, the nuclear–electron attraction of the atomic densities $V_{\text{ext}}[\rho_0]$, and Coulomb repulsion of the atomic densities $J[\rho_0]$. The final GFN2-xTB energy is given as,

$$E^{\text{GFN2-xTB}} = E_0^{\text{rep}} + E_1^{\text{EHT}} + E_2^{\text{coul}} \\ + E_3^{\text{coul}} + E^{\text{disp}} - T_{\text{elec}}S \quad (8)$$

where E_n^{disp} is the self-consistent D4 dispersion energy[40] and E_3^{coul} the third-order on-site Coulomb energy. The GFN1-xTB energy expression has a similar structure but differs in the details of some energy contributions, for example, D3(BJ) dispersion[41,42] is used instead of self-consistent D4 dispersion. For brevity, we will omit further details and refer the reader to the full theory and derivation of the xTB methods in Ref. 5.

In general, the xTB methods are parameterized on a chemically diverse set of molecular data. The training sets consist of relative (reactions) energies, molecular gradients, and molecular Hessians. Electronic properties, such as density matrices, partial charges, or dipole moments, were not included as reference. The energy expression of an xTB method is evaluated with a given trial parameterization to obtain energies, gradients, and Hessians in fully self-consistent computations. The parameters are optimized by minimizing the least squares (L2) loss between the xTB computed values and accurate reference data. In the optimization, the Levenberg–Marquardt algorithm[43,44] with line search is used by computing the numerical derivatives of the loss function with respect to the parameters. The parameters are initialized by hand for the core elements HCNO, or for other elements, from the row above or by interpolation between other elements.

For, GFN2-xTB the HCNO training data contains approximately 450 gradients, 40 Hessians, and 1300 relative energies. The reference data are usually obtained either at PBEh-3c[45] or B97-3c[46] level, while for relative energies mostly CCSD(T)/CBS values from the literature were used, i.e., subsets from the GMTKN55[47] database (e.g., ACONF, Amino20x4, BHROT27, BUT14DIOL, IDISP, MCONF, PArel, PCONF21, PX14, S22, SCONF, TAUT15, and WATER27). The element training sets for all other elements set contain approximately 100 gradients and 10 Hessians. For the lanthanides (f-block elements), a single training set was used, and the element parameters were linearly interpolated between cerium and lutetium.

Although the GFN parametrizations are designed to be general and robust, specialized applications can benefit from tailored parameters. For example, GFN1-xTB was re-fitted to DFT reference ionization potentials (IPs) and electron affinities (EAs) for the calculation of mass spectra, as IPEA-xTB version.[48] Additionally, GFN1-xTB was also optimized for specific materials, such as organosilicon compounds (GFN1(Si)-xTB) and halide perovskites.[15,16] Even tuning a single parameter can improve the description of exotic interactions.[14] Classical packages like *tblite*[49] or ParAMS[12] provide support for reparametrization and manipulation of parameter files. However, some applications can benefit from tighter integration and more transparent access to the xTB parameters. This is explored with the recently introduced *dxtb* package,[13] which provides an implementation of the GFN1-xTB Hamiltonian in PyTorch.[50] In this approach, the PyTorch framework allows to leverage automatic differentiation of the xTB energy expression, which could be applied for direct optimization of xTB parameters or in combination with machine learning models using xTB outputs. The latter has already proven successful in the context of other semiempirical quantum chemical methods.[51–54]

3 Technical details

Historically, the xTB methods are first made available in the program *xtb*, which has grown from an implementation of the xTB methods with basic functionality like geometry optimization, frequency analysis, and basic molecular dynamics to a full computational chemistry modeling package. As of today, *xtb* provides a wide range of functionalities, such as the possibility to add constraining potentials for fixing distances, angles, and dihedral angles or adding a confinement potential to avoid dissociation of clusters in molecular dynamics.

Next to the *xtb* package, the xTB methods are developed as part of the *tblite* library, which provides a rich library implementation. While *xtb* provides an atomistic simulation package for using xTB methods in many practical computational chemistry workflows, *tblite* is meant to provide the basic building blocks for creating an atomistic simulation package.

Both *xtb* and *tblite* are available under the GNU Lesser General Public License version 3.0 (LGPL-3.0) or later revisions of the license. This choice is made to ensure the package will be available as open source and also be included with proprietary projects. However, the LGPL-3.0 ensures that changes made to the project remain available to the community. Notably, most reimplementations of the xTB methods were done or started before the open source release of the *xtb* package, including the ones available in AMS,[55] CP2K,[56] entos,[57] and TeraChem.[58]

Besides the implementations in *xtb* and *tblite* the xTB methods are available in many other software packages. In the AMS package, the GFN1-xTB method is provided as part of the DFTB module with an MPI parallel implementation. The implementation of GFN1-xTB in CP2K is provided both with MPI and GPU support and has been scaled up for molecular dynamics simulations up to 80 million atoms.[34,56] TeraChem has an implementation of GFN1-xTB and GFN2-xTB (version with D3 dispersion) via a semiempirical integral library (SQMBox),[11] which enabled excited state calculations based on the xTB Hamiltonians.

In the context of machine learning the QCore package of Iambic (formerly Entos) implements GFN1-xTB and the ML method OrbNet which is based on the GFN1-xTB intermediates and outputs.[20,21] An open source framework for tight binding development was created with the TBMaLT package,[59] which has basic support for the xTB core Hamiltonian and is integrated with the *dxtb* library.[13]

Furthermore, many software packages are built with support for *xtb* or *tblite*. This includes tools for molecule generation like Architector,[60] AGOX,[61] cgbind,[62] or stk,[63] reaction network exploration tools like ChemDyME,[64] or Chemoton.[65] Finally, *xtb* is shipped now as a standard component of many established chemistry packages, like Orca,[66,67] Turbomole,[68] Gaussian[69] and OpenMM[70] or integrated via *tblite* directly into the packages like QCxMS,[71] DFTB+[72,73] and NWChem.[74]

4 Applications

In this section, we collect the most popular fields of application for xTB methods within scientific research: conformational analysis, equilibrium geometry searches, ab initio molecular dynamics, and data generation.

4.1 Conformations and structures

One of the prime applications made possible with the introduction of the xTB methods is the large scale screening of molecule structures. This computational efficiency for high throughput calculations in combination with an efficient approach, like metadynamics, to explore the potential energy surface, has been realized with CREST. The CREST program provides the de facto standard for sampling conformer ensembles.[38] This has enabled the exploration of a broader chemical space, facilitated by the extensive parameterization of the xTB methods, leading to publications spanning a wide range of applications. This comprises, for example, drug design[75] and delivery,[76] catalyst design and reaction pathway studies,[77,78] molecular switches,[79,80] and peptide dynam-

6

ics.[81] A more detailed overview of CREST and its applications can be found in the recently published software review.[38]

Aside from CREST, there are several alternative conformer generation tools found in the literature that have been used in combination with *xtb*. Popular choices are: AB-Cluster based on the artificial bee colony algorithm,[82] Molclus program[83] used in combination with the Multiwfn wave-function analyzer,[84] conformational search module of HyperChem,[85] ETKDG method[86] implemented in RDKit library,[87] Confab toolkit,[88] the autodE pipeline[89] and XTBDFT.[90] There are also some projects which prefer to search for conformers manually[91,92] or semi-manually by means of molecular simulation techniques.[93–95]

Even though conformer search is crucial and an integral part of structural investigations, the system size that can be treated is certainly limited. For systems with thousands of atoms, performing such analysis is not computationally feasible on modern CPU architectures. In such cases, *xtb* still serves as a valuable energy minimization tool, as GFN2-xTB was used for geometry optimization in nearly one-fifth of the publications discussed in Section 6. Typical systems include but are not limited to, proteins,[96] polypeptides,[97] metal-organic frameworks,[98] nanotubes,[99,100] large non-covalently bound complexes,[101–103] and porous liquids.[104]

4.2 Dataset generation & data-driven models

High quality datasets are an increasingly important resource in the chemical community as the use of ML models is growing. To ensure the quality and comparability of dataset entries, all structures should be optimized by the same electronic structure method. This step is crucial if the source of an entry uses 1D representations such as SMILES,[105,106] SELFIES,[107] or other, sometimes graph-based, specifications[108] to store chemical information. Semiempirical electronic structure methods are a natural choice for this application since numerous calculations need to be executed when curating a dataset. The robustness and efficiency of

xTB methods made them one of the methods of choice for this purpose.

Using those elaborate datasets, models can be built, that describe a correlation between an input vector, also referred to as "labels", and a target property. During the training phase, a statistical model is optimized to minimize the loss function for predicting the target based on the labels. The complexity of these labels can vary drastically from basic properties like the molecular weight of a molecule to electronic properties such as the Hamiltonian matrix or orbital energies. The labels should represent the necessary chemical information of a molecule to predict the target property. The computation of such labels should be efficient since they are computed not only during the training phase but also during the prediction or inference phase. As discussed in the previous paragraph, xTB methods are commonly used for structure optimizations when generating datasets. The use of labels based on xTB properties is therefore associated with no additional computational cost.

In 2015, Ramakrishnan *et al.*[109] introduced the concept of Δ-ML models to chemical applications. These models are trained to infer the difference in the value of a property computed by two quantum chemical methods. Usually, a lower-quality method is used as the base method upon which the high-level method value is predicted. Regarding the base method, GFN2-xTB is a popular choice due to the factors summarized previously in this work. Data-driven models are also compared against xTB reference values.

Since there is a vast number of examples where xTB methods are used in dataset generation tasks, we will only highlight those that put special emphasis on the use of xTB. One of the largest datasets for small- to medium-sized organic molecules is GEOM.[17] It constitutes a high quality data set collected from various sources with diverse experimental properties as labels. Using CREST and associated workflows, conformer sets for 317,928 molecules were computed at the GFN2-xTB level of theory. For a smaller portion of molecules, the level of theory was extended to DFT quality.
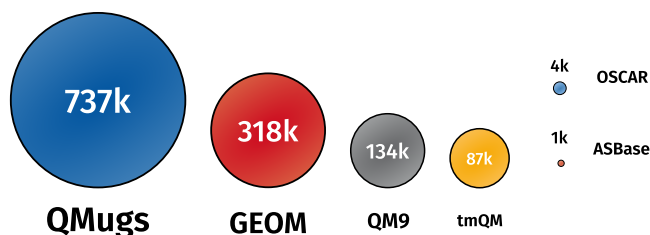
Figure 2: Comparison of the number of systems for the datasets, where xTB methods played a significant role in the generation of structures or labels. The area of the spheres is scaled according to the number of structures. QM9 is used as a size reference.

GEOM thus constitutes a dataset to train models to predict conformer ensembles,[110] or to use conformer ensembles as input as compared to single structures.[111]

Another still growing dataset with a similar scope of molecules and an even greater collection of xTB-based labels is QMugs.[18,19] Here, the conformational space of entries is also explored and included. The recent 1.1 update includes further steric and "dispersion" descriptors based on xTB.

Besides organic molecules, xTB methods are also used in the structure optimization of the tmQM[112] dataset containing 86,665 mononuclear transition metal complexes. For this set, no xTB based labels are published. However, it shows the robustness of these methods beyond organic chemistry. This point is further underlined by smaller databases capturing specific parts of the chemical space: organocatalysts in OSCAR,[113] or aggregate materials in ASBase.[114] In Fig. 2, the size of the datasets featured in this work is compared against the size of QM9[115] as a well-cited reference.

The labels computed from xTB calculations can be sorted into a few categories. First, a distinction between scalar and tensorial descriptors can be made. For the latter type, we can mainly find work by the Miller group. They published a series of neural network models based on quantum chemical tensors and named them OrbNet.[20,21,116] Apart from this family of models, Wang *et al.* published a generative model predicting the 3D structure of drug-like molecules based on experimental and quantum-chemically computed electron densities.[117] In terms of scalar properties, the HOMO-LUMO gap is a popular choice for applications involving electronic excitations.[118,119] In other applications, the models are trained to maximize this gap.[120,121] Besides the HOMO-LUMO gap, HOMO, LUMO and SOMO orbital energies are used as well.[121–123] In contrast to the molecular properties presented to this point, atomic properties are also used by models, most commonly atomic partial charges.[122–126] Usually, only the partial charge of atoms selected by chemical intuition is used, *i.e.*, reaction or coordination centers.

Recognizing the potential of Δ-ML models, the MultiXC-QM9[127] dataset was curated by Nandi *et al.* It encompasses the reaction energies for 162 million reactions with 228 quantum chemical methods, including GFN2-xTB energies. This allows to build elaborate Δ-ML models for small organic molecules. Δ-learning approaches have also been published in combination with ML-potentials. Staub *et al.* combined the SpookyNet[128] framework with GFN2-xTB as a base level.[129] This approach proved to be superior in the prediction of chemical yields for Wilkinson's catalyst, compared to only relying on GFN2-xTB calculations. A similar approach for building potential for molecular dynamics simulations is published by Murakami *et al.*[130] In their work, a polynomial fit is used instead of neural network models.

xTB methods are used in other areas of chemistry to build Δ-models: atmospheric chemistry,[131] medicinal chemistry,[132] and photochemistry.[133] Closely related to Δ-models, there are also studies where the xTB energies are used in neural network models to enrich the data, leading to better prediction performance.[134,135]

4.3 Ab initio molecular dynamics

Ab initio molecular dynamics (AIMD) provides a powerful computational technique to capture effects where approaches based on stationary points are not sufficient anymore. Different levels of electronic structure theory can be applied. In practice, KS-DFT, particularly at the GGA

level, is most commonly used.[136–138] However, the computational cost of running KS-DFT is prohibitive regarding the size of the system of interest, which limits the practical usability of AIMD. The xTB methods have shown, due to their more favorable computational cost, the potential to scale up to large system sizes, with the largest reported AIMD using GFN1-xTB and specialized hardware implementations reaching 80 million atoms.[34]

An example of the usage of AIMD with xTB methods is the investigation of transition state dynamics for transmetalations.[139] Here, GFN2-xTB is used to describe the reaction mechanism in different explicit solvent environments, which crucially participate in the transition state and the post-transition state dynamics. Furthermore, von der Esch et al.[140] compared computed IR spectra extracted from AIMD with experimental values. The benchmark results indicate that the tested KS-DFT functionals perform the best, while GFN2-xTB is an attractive, time-efficient alternative to these KS-DFT methods.

As mentioned above, xTB-driven AIMD in combination with enhanced sampling techniques or metadynamics can be applied for reaction discovery. AIMD simulations with the xTB methods are mainly used in two ways either for direct quantitative analysis of AIMD trajectories,[141–144] or for qualitative and exploratory AIMD runs before applying more accurate and expensive KS-DFT methods.[145] In the approach termed nanoreactor,[36,146] potential reaction products from a given set of starting structures are sampled from an enhanced sampling AIMD. Here, the reactants are confined to a limited space and exposed to a recurrent piston-like force that induces reactions among neighboring molecules.[146] Alternatively, a root-mean-square displacement-based metadynamics run, similar to conformer sampling, was shown to perform well in unimolecular rearrangement or fragmentation reactions.[36] The conformational sampling via metadynamics has already been outlined in section 4.1, which is available in the CREST program.[37,38]

Furthermore, specialized applications for AIMD with xTB are available via the QCxMS program[48,71] for describing molecule fragmentation in high-energy environments like electron ionization mass spectrometry or collision-induced dissociation.[147,148] In QCxMS several parallel high-energy AIMD simulations with xTB can automatically explore the energetically accessible phase space regions, yielding fragmentation products without bias.

Among the qualitative applications of AIMD is its use in a more educative setting: The efficiency of the xTB methods has enabled its use in interactive MD simulations, in which forces can be exerted by the user during the MD run.[149] With this, users can "manually" break bonds during xTB-driven MD runs and observe the reaction of the system afterward.

4.4 Implicit Solvation

In a multitude of the aforementioned applications, *e.g.* the creation of the GEOM[17] dataset, the inclusion of solvation effects plays a crucial role to be able to reproduce experimental conditions. In the xTB methods, this is achieved via the inclusion of tailored implicit solvation models, namely a generalized Born/Surface Area (GBSA) model, and the analytical linearized Poisson Boltzmann (ALPB)[28] model. Both of these models were specifically parametrized for the respective methods and can be invoked by simple command line keywords.

Due to its ease of use, the ALPB model has been used in numerous computational studies since its publication in 2021.[28] However, the GFN methods and, therefore also, the ALPB model were not primarily designed to accurately describe free energies. Still, there have been a limited number of studies that directly assess their performance for high-throughput virtual screening applications, like dissociation constants or enzyme-ligand binding.[150–152] While these studies generally provided promising results, they also revealed the shortcomings of the xTB methods, which could be addressed by re-evaluating the energies at higher levels of theory. Therefore, the usage of a multi-level workflow, where energies are calculated with high-level methods on lower-level geometries, is the common way for calculating solvation-

dependent properties[153] with the help of the xTB methods.

Additionally, the high computational efficiency of the xTB methods, in combination with the ALPB model, gave rise to the creation of physics-informed machine learning models utilizing features or training data obtained with the *xtb* program.[22,154] One example of an approach that uses readily available features for the calculation of solvation free energies is the creation of the implicitly perturbed Hamiltonian.[22] This approach defines energy attributes of the molecular Hamiltonian that are implicitly perturbed by solvents, creating a new class of molecular representations for deep learning approaches.

5 Novel features

This section highlights the recently introduced *xtb* features available in version 6.7.0 or higher. These features include the aISS docking algorithm, CPCM-X solvation model, mcGFNFF periodic force field, and ONIOM multi-layered framework.

5.1 aISS docking algorithm

As molecular phenomena and chemical reactions typically involve more than one particle, non-covalent interactions (NCIs) of different molecules are often crucial when modeling such processes. With the increasing molecular complexity of the interacting systems, possible interaction motifs also become more diverse and harder to predict. Hence, an automated computational prediction of intermolecular geometries is desirable, ranging from two molecules up to large clusters consisting of many particles. For this, the efficiency and reliability of the xTB methods when simulating NCIs are a perfect match and allow an automated computational prediction of intermolecular geometries. This was realized with one of the novel features included in the *xtb* program package, namely the automated interaction site screening (aISS) algorithm, which generates dimers up to large clusters of interacting molecules

fully automated.[155] It combines the intermolecular force field xTB-IFF[156] for a fast and accurate screening of possible interaction sites with the GFN methods for reliable geometry optimizations. The aISS yields structures of similar quality as CREST by being 1–3 orders of magnitude faster. Combined with the broad parametrization of the underlying xTB methods, this allows to also treat systems with 1,000s of atoms almost regardless of the atomic composition. An example of this is shown in Fig. 3.
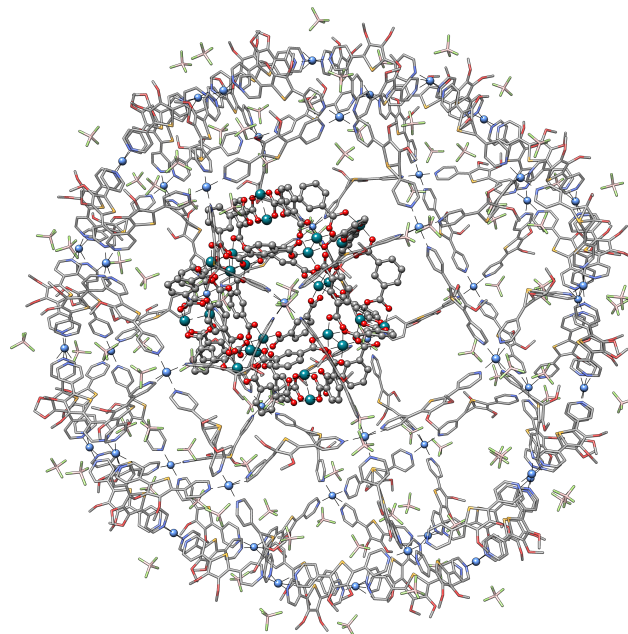

Figure 3: With the aISS predicted structure of the rhodium-organic cuboctahedra inside the $Pd_{48}L_{96}(BF_4)_{96}$ Goldberg polyhedron. Hydrogen atoms are omitted for clarity. Pd is depicted in light blue, Se in orange, B in pinkish, and Rh in light sea green. Reproduced and adapted from Ref. 155. Available under a CC-BY 4.0 license. Copyright 2023 C. Plett, S. Grimme.

The xTB-based docking was already applied to a variety of problems and molecular systems like the investigation of NCIs of supramolecular complexes,[157] ionic liquids,[158,159] metal-organic cages,[160] and nanobelts.[161] In addition, xTB-based docking is employed in the context of explicit and microsolvation, *e.g.*, with the QCG algorithm.[162] This hybrid cluster-continuum approach generates clusters of solutes solvated by solvent molecules that are em-

bedded in implicit solvation models like GBSA or ALPB. It was, for example, successfully employed to model solvation in the context of reaction mechanisms,[163] conformational analysis,[164,165] and spectroscopic properties.[166]

5.2 Extended Conductor-like Polarizable Continuum Model

Even though xTB was already combined with tailored ALPB and GBSA implicit solvation models, which produced good results given their high efficiency, these models are not as accurate as the more advanced implicit solvation models used routinely in combination with DFT methods. However, these advanced solvation models, typically based on polarizable continuum models (PCMs), are not readily available for usage with the xTB methods.

The recent integration of the domain decomposition conductor-like screening model (dd-COSMO)[167] into the *xtb* program package marked a significant leap in the solvation description of the xTB methods. This enhancement, especially when combined with established literature approaches for post-processing results, has greatly improved the performance of xTB methods for describing solvation-related properties. This combined approach is called the extended conductor-like polarizable continuum model (CPCM-X).[29] It leverages a combination of methods first introduced by Klamt in the conductor-like screening model for real solvents (COSMO-RS)[168] and Marenich *et al.* in the universal solvation model based on solute electron density (SMD)[169] and wraps it into a convenient open-source framework, that can be easily interfaced from various software packages and methods. The native implementation of the method in the *xtb* software package using the GFN2-xTB method improves the solvation description for large supermolecular complexes by up to 40 % in comparison to ALPB while still being more than two orders of magnitude more efficient than DFT-based methods. It has been successfully used to develop efficient workflows for calculating solvation-based properties, such as acid dissociation constants.[170]

5.3 Crystal structures

While the xTB methods were parametrized mainly for molecular systems, the application to extended systems under periodic boundary conditions like molecular crystals is an area where fast semiempirical methods can be applied. A full periodic implementation of the xTB methods including k-mesh integration has been realized only in the DFTB+ package and the Atomic Simulation Environment (ASE[171]) via the *tblite* library.[72] The only other known implementation of GFN1-xTB with k-mesh integration is available in the AMS package.[55] Starting with *xtb* version 6.7.0, the generic force field GFN-FF supports periodic boundary conditions, enabling molecular dynamics simulations, pre-optimization, and screening of large unit cells.[172] This implementation features a specialized run mode optimized for molecular crystals (mcGFN-FF) which reduces non-covalent interactions. Despite being less robust, mcGFN-FF achieves accuracy comparable to the GFN1-xTB regarding lattice energies and unit cell volume in many cases. Exploring the initial stages of crystal structure prediction is a promising application that could greatly benefit from this force field. For non-molecular crystals, Gale et al.[173] have developed the periodic GFN-FF (pGFN-FF) which utilizes the Wolf summation and damping of the bonded three-body term to stabilize electrostatic interactions, particularly for ionic crystals. The method has been implemented in the General Utility Lattice Program (GULP) program.

The xTB methods, GFN1-xTB and GFN2-xTB, enable efficient screening of the potential energy surface (PES), facilitating the analysis of crystal polymorphs, adsorption behaviors, and solid-solid transition states.[39,174–176] Molecular dynamics simulations are an invaluable tool for elucidating crystal structures. In particular, GFN2-xTB was used to investigate ligand interactions in MOFs, adsorption behavior, and transitioning between polymorphs.[177–179]

Though GFN has positioned itself as an efficient and robust TB parametrization for materials research, it has its limitations. The direct application of the xTB methods could be chal-

lenging for certain types of solids, such as metals,[180] radicals or charged molecules with double or triple bonds.[181,182] Additionally, a benchmark study on crystalline organic semiconductors has highlighted difficulties in applying GFN2-xTB for periodic systems due to unphysical geometries.[183] However, the same study also demonstrates reasonable performance of GFN1-xTB on similar systems.

For cases where tight-binding methods are expected to perform poorly, multi-level workflows that incorporate subsequent DFT calculations present a promising alternative. Another alternative is the integration of xTB methods with ML potentials,[184–186] which can then be utilized in the form of ΔML corrections to enhance accuracy while retaining computational efficiency.

5.4 Multiscale modelling

The efficiency and robustness of the xTB methods are particularly appealing facets in the integrated approaches, where the chemical problem is addressed through the Divide and Conquer strategy. The core idea is to divide the molecular system into several regions or layers, each calculated at a certain level of theory and merged into a multiscale model. Such models usually target cases where conventional ab initio theories are not feasible. The one prominent example of the integrated model is the ONIOM (Our Own N-layered molecular Orbital and molecular Mechanics) method.[187]

The simple variant of the ONIOM method is a two-layered scheme: the system is partitioned into inner and outer regions. The inner region, usually of higher interest, is treated with a more accurate theory. The environment, or outer region, is accounted for by calculating the inner region and the entire molecule at a lower level of theory and then subtracting the former from the latter. It is crucial that, while the inner region can be treated with ab initio methods (assuming a reasonable cutoff size), the lower-level theory must also sufficiently describe the entire system to ensure sufficient accuracy. In this context, the use of xTB methods becomes particularly advantageous due to their broad parametrization.
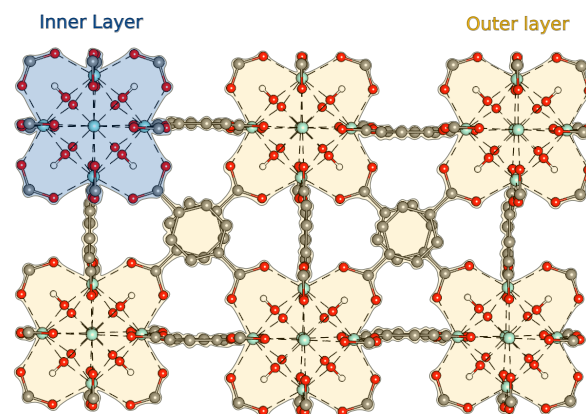

Figure 4: The 2-layered QM/MM scheme is demonstrated using an example of a Zr-functionalized UiO-66 metal-organic framework. Reproduced and adapted from Ref. 31. Available under a CC-BY 3.0 license. Copyright 2023 C. Plett, A. Katbashev, S. Ehlert, S. Grimme, M. Bursch.

The integration of the xTB family into the ONIOM framework is demonstrated using the example of UiO-66.[91] In recent work,[31] this Zr-based metal-organic framework (MOF) was modeled as an octahedral cage, consisting of six metal nodes and a total of 484 atoms, as shown in Fig 4. The cage was optimized by varying the size of the inner region, considering one, three, and six nodes, and using pure as well as inter-mixed methods (DFT, DFT//GFN, GFN).

Although pure DFT methods yielded the smallest deviations from the crystal structure, their steep computational scaling limits routine application for large-scale system size. On the other hand, the ONIOM(DFT//xTB) combined approach can be orders of magnitude faster, depending on the size of the inner region, and offers better accuracy than the underlying xTB low-level theories alone. As a result, the ONIOM framework can be expected to achieve a more balanced cost-efficiency ratio compared to standalone methods.

In the literature, the xTB methods have already been applied in multilayered frameworks.[30,32,33] These frameworks can vary depending on the approach – whether using the subtractive ONIOM or additive QM/MM,[188] the role of GFNn-xTB as a high-level or low-

12

level method, and the treatment of solvation as a continuum or explicitly. The available research spans a diverse chemical space, from water clusters,[189,190] DNA–strands[191,192]-lanthanoid complexes,[193] organoflourine compounds,[194] to (metallo)proteins.[195–203] In addition, there is some interesting work done on benchmarking the size of the inner region,[204] probing QM/MM interface handling,[205] and reparameterizing Lennard-Jones potentials based on the QM/MM results.[206,207]

The ONIOM routine is implemented in the *xtb* package starting from version 6.6.0. In addition to the methods available in *xtb* itself, it interfaces the *ORCA*[66,67] and *Turbomole*[208] program packages. As alternatives, one can use also QM/MM modules in the *ORCA*,[66] *Gaussian*,[209] and *PySCF*,[210] where xTB methods are accessible through external calls to the *xtb* or *tblite* API. A multi-center ONIOM implementation that also utilizes the xTB methods is now available in CREST as well.[211]

6 Citation overview

Over the years, *xtb* has developed into a multifunctional ecosystem incorporating GFN0-xTB,[212,213] GFN1-xTB,[7] GFN2-xTB,[8] and GFN-FF[214] methods. The publications citing these methods were combined into the collective database with the aim to analyze the usage of *xtb* software package in the community.[215] In addition to the aforementioned four "main" publications, the database contains the review on the extended tight binding[5] and does not contain duplicates. The database was processed to assess the popularity of *xtb* over the years, which is shown on Fig. 5a.

In 2017, the first member of the GFN-xTB family, GFN1-xTB (originally abbreviated as GFN-xTB), was published. Two years later, the second-generation GFN2-xTB was released, gathering significant attention in the community and contributing to the overall popularity of *xtb*.

Since 2017, the number of papers citing *xtb* has grown quadratically, with almost 1300 citations in the previous year alone, averaging three
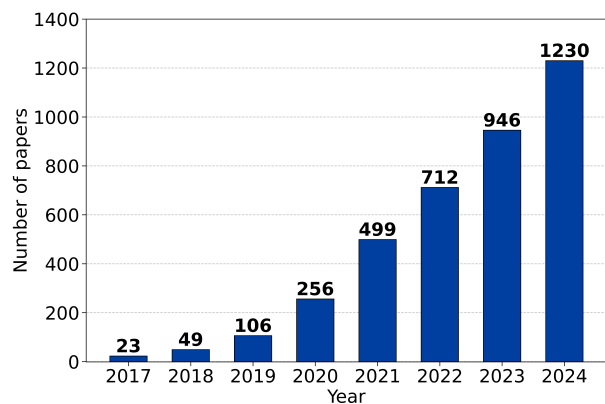
(a) Number of papers citing *xtb* over time.
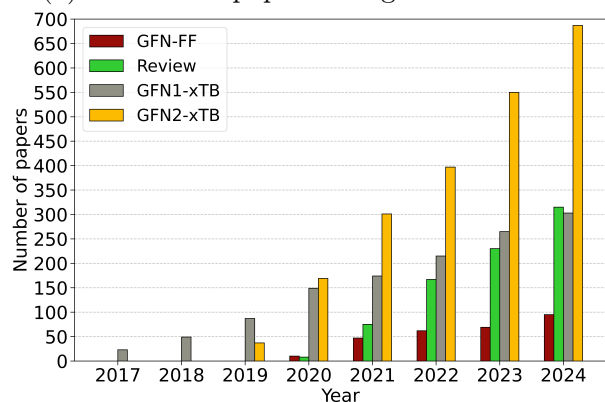
(b) Number of papers citing *xtb*-related publications over time.

Figure 5: The annual citation count for xtb-related methods, both collectively (a) and separately (b). Plot (a) includes only the mutually related GFN-type xTB methods but excludes the earlier presented sTDA-xTB method[216] for excited state calculations.

to four citations per day. The total number of citations has already passed the 3800 mark and is expected to continue growing as *xtb* (and *tblite*) expand their toolbox with a new semi-empirical method in the near future.

A more in-depth analysis is illustrated in Fig. 5b, where the citation dynamics for each method are considered independently. The overall trend is upward, indicating that all listed publications are gaining more and more recognition every year. The steepest curve belongs to GFN2-xTB, exhibiting a polynomial increase, with more than 2000 citations over a span of seven years and around 700 citations last year alone. Similarly, GFN1-xTB shows steady growth in the number of citations
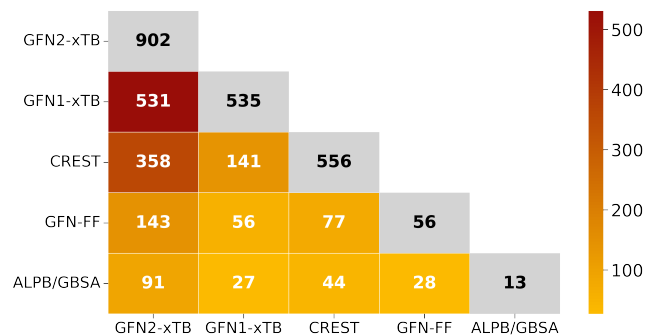
13

Figure 6: Heatmap illustrating the overlap between the citation bases of GFN1-xTB, GFN2-xTB, GFN-FF, CREST, and ALPB/GBSA. The diagonal shows the number of unique citations, while off-diagonals represent the number of cross-referenced citations.

Figure 7: The aim-specific categorization of the publications citing GFN2-xTB. Appl: Application, Dev: Development.

(around 1300), though the growth is more linear compared to GFN2-xTB. The general-purpose force field GFN-FF gains considerable recognition, continuously extending its user base.

In addition to the discussed statistics, it is interesting to demonstrate the interconnection between the *xtb* components. To complete the graph, citation bases of the CREST[37] as well as the implicit solvation models (ALPB/GBSA)[28] were added to the *xtb* publication database, as shown in Fig. 6.

From this analysis, it is evident that GFN2-xTB and GFN1-xTB are the most co-cited approaches. One possible explanation for this could be that a significant portion of these co-citations come from papers related to method development and benchmarking. The same reasoning can be applied to the regular co-citation of GFN-FF with xTB methods, with GFN2-xTB being cited in approximately half of all GFN-FF cited works.

The intersection of the GFN family members with CREST is essential to estimate their applications in the conformer search. In this regard, CREST is most frequently used with the GFN2-xTB method, with its usage twice that of the GFN1-xTB:CREST combination. A similar relationship is observed between GFN1-xTB:CREST and GFN-FF:CREST combinations.

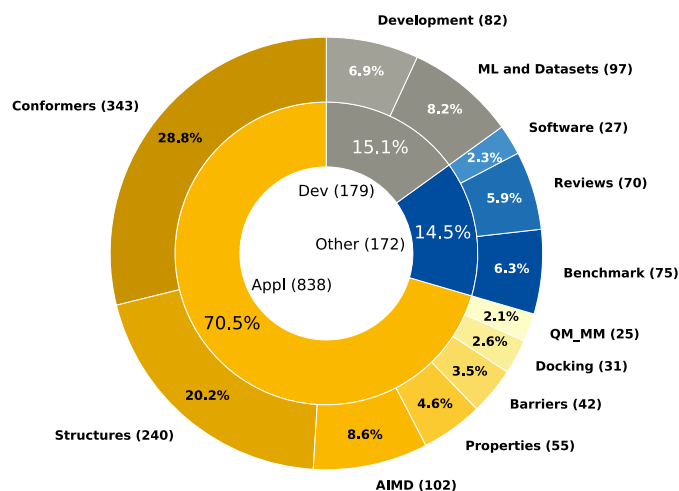The ALPB/GBSA implicit solvation models were used the least in publications, which can

partly be attributed to the fact that this publication is the newest among the topics investigated. Alternatively, it could also indicate that most calculations were either done in the gas phase or used different solvation models.

Finally, GFN2-xTB was chosen to demonstrate the end-purpose-specific partitioning of its citation base. Unfortunately, due to limited access, not all publications are included in the assessment. This assessment is depicted in Fig. 7 in the form of a pie chart. Despite the potential for slight inaccuracies due to the size of the final database (over 1300 papers) and limited human resources, the overall trends remain accurate and can be used to draw meaningful conclusions. During the partitioning, the smallest division to be considered was set to be more than 20 papers.

During the partitioning, the smallest division considered included more than 20 papers. For this reason, some applications are not listed as separate sections and have been added to the most suitable category. In addition, all cross-references, such as GFN2-xTB being used for both conformer search and structure optimization, are addressed by considering only the most relevant usage to avoid multidimensional overlaps and exhaustive statistics.

The dominant portion (70.5%) of users utilized GFN2-xTB for general-purpose applications. Within this category, conformer search,

which also includes manual sampling, represents the largest fraction, followed by structure optimization. The third-largest fraction, ab-initio molecular dynamics (AIMD), is merged with periodic calculations on crystals due to its small size (less than 0.9%).

The remaining segments use *xtb* for various property calculations (e.g., HOMO-LUMO gap,[217] interaction energies,[218] UV and ECD spectra,[219] charges[220]), as well as barriers, molecular docking (with and without aISS), and QM/MM applications.

More than 15% of publications cite GFN2-xTB in the context of method development. This includes the development of new methods (including workflows), and its use for machine learning purposes (primarily dataset generation).

The last piece is more general, and includes reviews, software papers, and benchmarking of the method.

The corresponding databases used in this analysis can be found in a public GitHub repository: https://github.com/grimme-lab/GFN2_citation_base.git

7 Conclusion

We presented the current state of the ecosystem and community around the software packages *xtb* and *tblite* as well as the xTB methods. Within five years, the GFN2-xTB method already enabled thousands of publications, demonstrating its potential to be a new de facto standard method for computational chemistry. A crucial contribution to this success is the wide availability of the xTB methods through open source program simulation packages, mainly provided by interfacing with the *xtb* package, integrating the *tblite* library, or directly using the *xtb* package. The xTB methods make it possible to cover a wide range of tasks in computational chemistry, like geometry optimizations, frequency analysis, free energy computations, solvent modeling, and molecular dynamics to highlight a few. More specialized applications, like reparametrization or multiscale modeling, are supported and readily available. Special

attention should be given to the crucial role of xTB methods in large scale data generation campaigns where xTB methods are present as the initial screening step, intermediate refinement, or even for providing data labels, which makes it crucial in the development of data sets purposed to train large scale ML models.

Going forward, the development will continue to focus on better integration of hardware acceleration, like GPU support (available for Mulliken-approximated Fock exchange in SQM-Box),[11] and parametrization tools, like automatic differentiation via *dxtb*. On the methodological side, the development of improved sampling approaches together with xTB methods for improved conformer sampling, reaction discovery, or chemical space exploration are targets. An outstanding challenge is to bring the xTB methods into the realm of material discovery and crystal structure prediction, by improving support and ease of handling periodic systems and investigating the need for specialized parametrizations. Integration of future xTB methods like the currently developed general purpose parametrization (g-xTB) is a high priority within the framework of the *tblite* library. Overall, the xTB methods will continue providing the basic building blocks for high-throughput and screening applications for molecular properties and be indispensable for a wide range of computational simulations.

Author declarations

Conflict of interest

The authors have no conflicts to disclose.

15

kehrprogramm.

References

(1) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140*, A1133–A1138, Publisher: American Physical Society.

(2) Teale, A. M.; Helgaker, T.; Savin, A.; Adamo, C.; Aradi, B.; Arbuznikov, A. V.; Ayers, P. W.; Baerends, E. J.; Barone, V.; Calaminici, P. et al. DFT exchange: sharing perspectives on the workhorse of quantum chemistry and materials science. *Physical Chemistry Chemical Physics* **2022**, *24*, 28700–28781, Publisher: The Royal Society of Chemistry.

(3) Bursch, M.; Mewes, J.-M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry**. *Angewandte Chemie International Edition* **2022**, *61*, e202205735, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202205735.

(4) Morgante, P.; Peverati, R. The devil in the details: A tutorial review on some undervalued aspects of density functional theory calculations. *International Journal of Quantum Chemistry* **2020**, *120*, e26332, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.26332.

(5) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *WIREs Computational Molecular Science* **2021**, *11*, e1493, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1493.

(6) Elstner, M.; Seifert, G. Density functional tight binding. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2014**, *372*, 20120483, Publisher: Royal Society.

(7) 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). *Journal of Chemical Theory and Computation* **2017**, *13*, 1989–2009, Publisher: American Chemical Society.

(8) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation* **2019**, *15*, 1652–1671, Publisher: American Chemical Society.

(9) Neugebauer, H.; Bädorf, B.; Ehlert, S.; Hansen, A.; Grimme, S. High-throughput screening of spin states for transition metal complexes with spin-polarized extended tight-binding methods. *Journal of Computational Chemistry* **2023**, *44*, 2120–2129, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.27185.

(10) Cheng, C. Y.; Wibowo-Teale, A. M. Semiempirical Methods for Molecular Systems in Strong Magnetic Fields. *Journal of Chemical Theory and Computation* **2023**, *19*, 6226–6241, Publisher: American Chemical Society.

(11) Bannwarth, C.; Martínez, T. J. SQMBox: Interfacing a semiempirical integral library to modular ab initio electronic structure enables new semiempirical methods. *The Journal of Chemical Physics* **2023**, *158*, 074109.

(12) Komissarov, L.; Rüger, R.; Hellström, M.; Verstraelen, T. ParAMS: Parameter Optimization for Atomistic and Molecular Simulations. *Journal of Chemical Information and Modeling* **2021**, *61*, 3737–3743, Publisher: American Chemical Society.

(13) Friede, M.; Hölzer, C.; Ehlert, S.; Grimme, S. dxtb – An Efficient And Fully Differentiable Framework For Extended Tight-Binding. *J. Chem. Phys.* **2024**, *161*, 062501.

(14) Mewes, J.-M.; Hansen, A.; Grimme, S. Comment on "The Nature of Chalcogen-Bonding-Type Tellurium–Nitrogen Interactions": Fixing the Description of Finite-Temperature Effects Restores the Agreement Between Experiment and Theory. *Angewandte Chemie International Edition* **2021**, *60*, 13144–13149, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202102679.

(15) Komissarov, L.; Verstraelen, T. Improving the Silicon Interactions of GFN-xTB. *Journal of Chemical Information and Modeling* **2021**, *61*, 5931–5937, Publisher: American Chemical Society.

(16) Raaijmakers, S.; Pols, M.; Vicent-Luna, J. M.; Tao, S. Refined GFN1-xTB Parameters for Engineering Phase-Stable CsPbX$_3$ Perovskites. *The Journal of Physical Chemistry C* **2022**, *126*, 9587–9596.

(17) Axelrod, S.; Gómez-Bombarelli, R. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data* **2022**, *9*, 185, Number: 1 Publisher: Nature Publishing Group.

(18) Isert, C.; Atz, K.; Jiménez-Luna, J.; Schneider, G. QMugs, quantum mechanical properties of drug-like molecules. *Scientific Data* **2022**, *9*, 273, Number: 1 Publisher: Nature Publishing Group.

(19) Neeser, R. M.; Isert, C.; Stuyver, T.; Schneider, G.; Coley, C. W. QMugs 1.1: Quantum mechanical properties of organic compounds commonly encountered in reactivity datasets. *Chemical Data Collections* **2023**, *46*, 101040.

(20) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F., III OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *The Journal of Chemical Physics* **2020**, *153*, 124111.

(21) Christensen, A. S.; Sirumalla, S. K.; Qiao, Z.; O'Connor, M. B.; Smith, D. G. A.; Ding, F.; Bygrave, P. J.; Anandkumar, A.; Welborn, M.; Manby, F. R. et al. OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy. *The Journal of Chemical Physics* **2021**, *155*, 204103.

(22) Alibakhshi, A.; Hartke, B. Implicitly perturbed Hamiltonian as a class of versatile and general-purpose molecular representations for machine learning. *Nature Communications* **2022**, *13*, 1245, Number: 1 Publisher: Nature Publishing Group.

(23) Vogel, Y. B.; Evans, C. W.; Belotti, M.; Xu, L.; Russell, I. C.; Yu, L.-J.; Fung, A. K. K.; Hill, N. S.; Darwish, N.; Gonçales, V. R. et al. The corona of a surface bubble promotes electrochemical reactions. *Nature Communications* **2020**, *11*, 6323, Number: 1 Publisher: Nature Publishing Group.

(24) Wang, Y.; Liu, T.; Zhang, Y.-Y.; Li, B.; Tan, L.; Li, C.; Shen, X.-C.; Li, J. Cross-catenation between position-isomeric metallacages. *Nature Communications* **2024**, *15*, 1363, Number: 1 Publisher: Nature Publishing Group.

(25) Fa, S.; Shi, T.-h.; Akama, S.; Adachi, K.; Wada, K.; Tanaka, S.; Oyama, N.; Kato, K.; Ohtani, S.; Nagata, Y. et al. Real-time chirality transfer monitoring from statistically random to discrete homochiral nanotubes. *Nature Communications* **2022**, *13*, 7378, Number: 1 Publisher: Nature Publishing Group.

(26) Rabbani, R.; Saeedi, S.; Nazimuddin, M.; Barbero, H.; Kyritsakas, N.; White, T. A.; Masson, E. Enhanced photoreduction of water catalyzed by a cucurbit[8]uril-secured platinum dimer. *Chemical Science* **2021**, *12*, 15347–15352, Publisher: The Royal Society of Chemistry.

(27) Kolarski, D.; Steinbach, P.; Bannwarth, C.; Klaue, K.; Hecht, S. Diaryltriazolium Photoswitch: Reaching a Millisecond Cycloreversion with High Stability and NIR Absorption. *Angewandte Chemie International Edition* **2024**, *63*, e202318015, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202318015.

(28) Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. *Journal of Chemical Theory and Computation* **2021**, *17*, 4250–4261, Publisher: American Chemical Society.

(29) Stahn, M.; Ehlert, S.; Grimme, S. Extended Conductor-like Polarizable Continuum Solvation Model (CPCM-X) for Semiempirical Methods. *The Journal of Physical Chemistry A* **2023**, *127*, 7036–7043, Publisher: American Chemical Society.

(30) Hudson, P. S.; Aviat, F.; Meana-Pañeda, R.; Warrensford, L.; Pollard, B. C.; Prasad, S.; Jones, M. R.; Woodcock, H. L.; Brooks, B. R. Obtaining QM/MM binding free energies in the SAMPL8 drugs of abuse challenge: indirect approaches. *Journal of Computer-Aided Molecular Design* **2022**, *36*, 263–277.

(31) Plett, C.; Katbashev, A.; Ehlert, S.; Grimme, S.; Bursch, M. ONIOM meets xtb: efficient, accurate, and robust multi-layer simulations across the periodic table. *Physical Chemistry Chemical Physics* **2023**, *25*, 17860–17868, Publisher: The Royal Society of Chemistry.

(32) Jiang, Z.; Zhang, M.; Chen, X.; Wang, B.; Fan, W.; Yang, C.; Yang, X.; Zhang, Z.; Yang, X.; Li, C. et al. A Bismuth-Based Zeolitic Organic Framework with Coordination-Linked Metal Cages for Efficient Electrocatalytic CO2 Reduction to HCOOH. *Angewandte Chemie International Edition* **2023**, *62*, e202311223, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202311223.

(33) Castro, M. E.; Rangel-Galván, M.; Dávila, E. M.; Caballero, N. A.; Melendez, F. J. Theoretical NMR and IR spectroscopic analyses of the preferred conformers of the neurotransmitter anandamide. *International Journal of Quantum Chemistry* **2023**, *123*, e27098, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.27098.

(34) Schade, R.; Kenter, T.; Elgabarty, H.; Lass, M.; Schütt, O.; Lazzaro, A.; Pabst, H.; Mohr, S.; Hutter, J.; Kühne, T. D. et al. Towards electronic structure-based *ab-initio* molecular dynamics simulations with hundreds of millions of atoms. *Parallel Computing* **2022**, *111*, 102920.

(35) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.

(36) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *Journal of Chemical Theory and Computation* **2019**, *15*, 2847–2862, Publisher: American Chemical Society.

(37) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical Chemistry Chemical Physics* **2020**, *22*, 7169–7192, Publisher: The Royal Society of Chemistry.

(38) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C. et al. CREST—A program for the exploration of low-energy molecular chemical space. *The Journal of Chemical Physics* **2024**, *160*, 114110.

(39) Pracht, P.; Morgan, J. W. R.; Wales, D. J. Exploring energy landscapes for solid-state systems with variable cells at the extended tight-binding level. *The Journal of Chemical Physics* **2023**, *159*, 064801.

(40) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *The Journal of Chemical Physics* **2019**, *150*, 154122.

(41) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(42) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21759.

(43) Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **1944**, *2*, 164–168, Publisher: American Mathematical Society (AMS).

(44) Marquardt, D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics* **1963**, *11*, 431–441, Publisher: Society for Industrial and Applied Mathematics.

(45) Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *The Journal of Chemical Physics* **2015**, *143*, 054107.

(46) Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S. B97-3c: A revised low-cost variant of the B97-D density functional method. *The Journal of Chemical Physics* **2018**, *148*, 064104.

(47) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Physical Chemistry Chemical Physics* **2017**, *19*, 32184–32215, Publisher: The Royal Society of Chemistry.

(48) Ásgeirsson, V.; Bauer, C. A.; Grimme, S. Quantum chemical calculation of electron ionization mass spectra for general organic and inorganic molecules. *Chemical Science* **2017**, *8*, 4879–4895, Publisher: The Royal Society of Chemistry.

(49) Light-weight tight-binding framework. 2021; `https://github.com/tblite/tblite` (accessed 15th of June 2024).

(50) Ansel, J.; Yang, E.; He, H.; Gimelshein, N.; Jain, A.; Voznesensky, M.; Bao, B.; Bell, P.; Berard, D.; Burovski, E. et al. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24). 2024.

(51) Dral, P. O.; Von Lilienfeld, O. A.; Thiel, W. Machine Learning of Parameters for Accurate Semiempirical Quantum Chemical Calculations. *Journal of Chemical Theory and Computation* **2015**, *11*, 2120–2125.

(52) Zubatiuk, T.; Nebgen, B.; Lubbers, N.; Smith, J. S.; Zubatyuk, R.; Zhou, G.; Koh, C.; Barros, K.; Isayev, O.; Tretiak, S. Machine learned Hückel theory: Interfacing physics and deep neural networks. *The Journal of Chemical Physics* **2021**, *154*, 244108.

(53) Zhou, G.; Lubbers, N.; Barros, K.; Tretiak, S.; Nebgen, B. Deep learning of dynamically responsive chemical Hamiltonians with semiempirical quantum mechanics. *Proceedings of the National Academy of Sciences* **2022**, *119*, e2120333119, Publisher: Proceedings of the National Academy of Sciences.

(54) Vargas–Hernández, R. A.; Jorner, K.; Pollice, R.; Aspuru–Guzik, A. Inverse molecular design and parameter optimization with Hückel theory using automatic differentiation. *The Journal of Chemical Physics* **2023**, *158*, 104801.

(55) Rüger, R.; Franchini, M.; Trnka, T.; Yakovlev, A.; van Lenthe, E.; Philipsen, P.; van Vuren, T.; Klumpers, B.; Soini, T. Amsterdam Modelling Suite. SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, 2024; `http://www.scm.com` (accessed 15th of June 2024).

(56) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F. et al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *The Journal of Chemical Physics* **2020**, *152*, 194103.

(57) Manby, F.; Miller, T.; Bygrave, P.; Ding, F.; Dresselhaus, T.; Batista-Romero, F.; Buccheri, A.; Bungey, C.; Lee, S.; Meli, R. et al. entos: A quantum molecular simulation package. **2019**, `https://doi.org/10.26434/chemrxiv.7762646.v2`.

(58) Seritan, S.; Bannwarth, C.; Fales, B. S.; Hohenstein, E. G.; Isborn, C. M.; Kokkila-Schumacher, S. I. L.; Li, X.; Liu, F.; Luehr, N.; Snyder Jr., J. W. et al. TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics. *WIREs Computational Molecular Science* **2021**, *11*, e1494, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1494.

(59) McSloy, A.; Fan, G.; Sun, W.; Hölzer, C.; Friede, M.; Ehlert, S.; Schütte, N.-E.; Grimme, S.; Frauenheim, T.; Aradi, B. TBMaLT, a flexible toolkit for combining tight-binding and machine learning. *The Journal of Chemical Physics* **2023**, *158*, 034801.

(60) Taylor, M. G.; Burrill, D. J.; Janssen, J.; Batista, E. R.; Perez, D.; Yang, P. Architector for high-throughput cross-periodic table 3D complex building. *Nature Communications* **2023**, *14*, 2786, Number: 1 Publisher: Nature Publishing Group.

(61) Christiansen, M.-P. V.; Rønne, N.; Hammer, B. Atomistic global optimization X: A Python package for optimization of atomistic structures. *The Journal of Chemical Physics* **2022**, *157*, 054701.

(62) Young, T. A.; Gheorghe, R.; Duarte, F. cgbind: A Python Module and Web App for Automated Metallocage Construction and Host–Guest Characterization. *Journal of Chemical Information and Modeling* **2020**, *60*, 3546–3557, Publisher: American Chemical Society.

(63) Turcani, L.; Tarzia, A.; Szczypiński, F. T.; Jelfs, K. E. stk: An extendable Python framework for automated molecular and supramolecular structure assembly and discovery. *The Journal of Chemical Physics* **2021**, *154*, 214102.

(64) Shannon, R. J.; Martínez-Núñez, E.; Shalashilin, D. V.; Glowacki, D. R. ChemDyME: Kinetically Steered, Automated Mechanism Generation through Combined Molecular Dynamics and Master Equation Calculations. *Journal of Chemical Theory and Computation* **2021**, *17*, 4901–4912, Publisher: American Chemical Society.

(65) Unsleber, J. P. Accelerating Reaction Network Explorations with Automated Reaction Template Extraction and Application. *Journal of Chemical Information and Modeling* **2023**, *63*, 3392–3403, Publisher: American Chemical Society.

(66) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *The Journal of Chemical Physics* **2020**, *152*, 224108.

(67) Neese, F. Software update: The ORCA program system—Version 5.0. *WIREs Computational Molecular Science* **2022**, *12*, e1606, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1606.

(68) Balasubramani, S. G.; Chen, G. P.; Coriani, S.; Diedenhofen, M.; Frank, M. S.; Franzke, Y. J.; Furche, F.; Grotjahn, R.; Harding, M. E.; Hättig, C. et al. TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations. *The Journal of Chemical Physics* **2020**, *152*, 184107.

(69) gau_xtb: A Gaussian interface for xtb code. 2023; `http://sobereva.com/soft/gau_xtb/` (accessed 15th of June 2024).

(70) OpenMM plugin to interface with XTB. 2023; `https://github.com/openmm/openmm-xtb` (accessed 15th of June 2024).

(71) Koopman, J.; Grimme, S. Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules. *Journal of the American Society for Mass Spectrometry* **2022**, *33*, 2226–2242, Publisher: American Society for Mass Spectrometry. Published by the American Chemical Society. All rights reserved.

(72) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshaye, M. Y.; Dumitrică, T.; Dominguez, A. et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *The Journal of Chemical Physics* **2020**, *152*, 124101, `https://github.com/dftbplus/dftbplus` (accessed 15th of June 2024).

(73) Hourahine, B.; Berdakin, M.; Bich, J. A.; Bonafé, F. P.; Camacho, C.; Cui, Q.; Díaz Mirón, G.; Ehlert, S.; Elstner, M.; Frauenheim, T. et al. Recent developments in DFTB+, a software package for efficient atomistic quantum mechanical simulations. *J. Phys. Chem. A* **2025**, *0*, 0, manuscript in preparation.

(74) Aprà, E.; Bylaska, E. J.; de Jong, W. A.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; van Dam, H. J. J.; Alexeev, Y.; Anchell, J. et al. NWChem: Past, present, and future. *The Journal of Chemical Physics* **2020**, *152*, 184102.

(75) da Fonseca, C. A. R.; Prado, V. C.; Paltian, J. J.; Kazmierczak, J. C.; Schumacher, R. F.; Sari, M. H. M.; Cordeiro, L. M.; da Silva, A. F.; Soares, F. A. A.; Oliboni, R. d. S. et al. 4-(Phenylselanyl)-2H-chromen-2-one-Loaded Nanocapsule Suspension—A Promising Breakthrough in Pain Management: Comprehensive Molecular Docking, Formulation Design, and Toxicological and Pharmacological Assessments in Mice. *Pharmaceutics* **2024**, *16*, 269, Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

(76) Hoti, G.; Ferrero, R.; Caldera, F.; Trotta, F.; Corno, M.; Pantaleone, S.; Desoky, M. M. H.; Brunella, V. A Comparison between the Molecularly Imprinted and Non-Molecularly Imprinted Cyclodextrin-Based Nanosponges for the Transdermal Delivery of Melatonin. *Polymers* **2023**, *15*, 1543, Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

(77) Herbstritt, D.; Tomar, P.; Müller, R.; Kaupp, M.; Braun, T. A 2,2-Difluoroimidazolidine Derivative for Deoxyfluorination Reactions: Mechanistic Insights by Experimental and Computational Studies. *Chemistry – A European Journal* **2023**, *29*, e202301556, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/chem.202301556.

(78) Greb, J.; Drennhaus, T.; Klischan, M. K. T.; Schroeder, Z. W.; Frey, W.; Pietruszka, J. A Common C2-Symmetric 2,2'-Biphenol Building Block and its Application in the Synthesis of (+)-di-epi-Gonytolide A. *Chemistry – A European Journal* **2023**, *29*, e202300941, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/chem.202300941.

(79) Opačak, S.; Babić, D.; Perić, B.; Zeljko Marinić; Smrečki, V.; Pem, B.; Vrček, I. V.; Kirin, S. I. A ferrocene-based pseudopeptide chiroptical switch. *Dalton Transactions* **2021**, *50*, 4504–4511, Publisher: The Royal Society of Chemistry.

(80) Rabe von Pappenheim, F.; Wensien, M.; Ye, J.; Uranga, J.; Irisarri, I.; de Vries, J.; Funk, L.-M.; Mata, R. A.; Tittmann, K. Widespread occurrence of covalent lysine–cysteine redox switches in proteins. *Nature Chemical Biology* **2022**, *18*, 368–375, Number: 4 Publisher: Nature Publishing Group.

(81) Osifová, Z.; Kalvoda, T.; Galgonek, J.; Culka, M.; Vondrášek, J.; Bouř, P.; Bednárová, L.; Andrushchenko, V.; Dračínský, M.; Rulíšek, L. What are the minimal folding seeds in proteins? Experimental and theoretical assessment of secondary structure propensities of small peptide fragments. *Chemical Science* **2024**, *15*, 594–608, Publisher: The Royal Society of Chemistry.

(82) Zhang, J.; Dolg, M. ABCluster: the artificial bee colony algorithm for cluster global optimization. *Physical Chemistry Chemical Physics* **2015**, *17*, 24173–24181.

(83) Molclus program. 2016; `http://www.keinsci.com/research/molclus.html` (accessed 15th of June 2024).

(84) Lu, T.; Chen, F. Multiwfn: A multifunctional wavefunction analyzer. *Journal of Computational Chemistry* **2012**, *33*, 580–592.

(85) Laxmi, D.; Priyadarshy, S. HyperChem 6.03. *Biotech Software & Internet Report* **2002**, *3*, 5–9.

(86) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *Journal of Chemical Information and Modeling* **2015**, *55*, 2562–2574.

(87) RDKit. 2019; `https://github.com/rdkit/rdkit` (accessed 15th of June 2024).

(88) O'Boyle, N. M.; Vandermeersch, T.; Flynn, C. J.; Maguire, A. R.; Hutchison, G. R. Confab - Systematic generation of diverse low-energy conformers. *Journal of Cheminformatics* **2011**, *3*, 8.

(89) Young, T. A.; Silcock, J. J.; Sterling, A. J.; Duarte, F. autodE: Automated Calculation of Reaction Energy Profiles— Application to Organic and Organometallic Reactions. *Angewandte Chemie* **2021**, *133*, 4312–4320.

(90) Lin, S.; Elanany, M.; Khawaji, M. XTBDFT: Automated workflow for conformer searching of minima and transition states powered by extended tight binding and density functional theory. *SoftwareX* **2022**, *20*, Publisher: Elsevier.

(91) Tan, T. T. Y.; Li, X.; Otake, K.-i.; Tan, Y. C.; Loh, X. J.; Kitagawa, S.; Lim, J. Y. C. UiO-66 metal organic frameworks with high contents of flexible adipic acid co-linkers. *Chemical Communications* **2022**, *58*, 11402–11405, Publisher: The Royal Society of Chemistry.

(92) Ge, Y.; Ma, Y.; Zhao, M.; Wei, J.; Wu, X.; Zhang, Z.; Yang, H.; Lei, H.; Wu, B. Exploring gabosine and chlorogentisyl alcohol derivatives from a marine-derived fungus as EcGUS inhibitors with informatic assisted approaches. *European Journal of Medicinal Chemistry* **2022**, *242*, 114699.

(93) Demarque, D. P.; Kemper, M.; Merten, C. VCD spectroscopy reveals that a water molecule determines the conformation of azithromycin in solution. *Chemical Communications* **2021**, *57*, 4031–4034, Publisher: The Royal Society of Chemistry.

(94) Yu, L.; Shi, M.; Wang, Z.; Xing, X.; Umair Ali, M.; He, Y.; Meng, H. Tuning the UV/Vis Absorption Spectra of Electrochromic Small Molecular Radicals Through Bridge Modulation. *ChemPhysChem* **2021**, *22*, 1684–1691, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cphc.202100369.

(95) Huang, J.; Huang, C.; Hou, Q.; Wu, M.; Wu, X.; Zhang, Y.; Tian, G.; Zhang, F. A theoretical study on dissociative photoionization and photoionization cross-sections of a typical ketohydroperoxide in n-butane low-temperature oxidation. *Combustion and Flame* **2023**, *257*, 112490.

(96) Rincón, D. A.; Doerr, M.; Daza, M. C. Hydrogen Bonds and n → π* Interactions in the Acetylation of Propranolol Catalyzed by Candida antarctica Lipase B: A QTAIM Study. *ACS Omega* **2021**, *6*, 20992–21004, Publisher: American Chemical Society.

(97) Wang, M.; Zhao, Y.; Zhang, L.; Deng, J.; Qi, K.; Zhou, P.; Ma, X.; Wang, D.; Li, Z.; Wang, J. et al. Unexpected Role of Achiral Glycine in Determining the Suprastructural Handedness of Peptide Nanofibrils. *ACS Nano* **2021**, *15*, 10328–10341, Publisher: American Chemical Society.

(98) Ossinger, S.; Näther, C.; Tuczek, F. Crystal structure of bis[dihydrobis(pyrazol-1-yl)borato-κ2N2,N2´](1,10-phenanthroline-κ2N,N´)zinc(II). *Acta Crystallographica Section E: Crystallographic Communications* **2019**, *75*, 1112–1116, Publisher: International Union of Crystallography.

(99) Wang, Q.; Li, Y.; Meng, T.; Huang, B.; Hu, L.; Su, H.; Meng, C.; Tong, Y. Engineering Heterostructure-Incorporated Metal Silicates Anchored on Carbon Nanotubes for Highly Durable Lithium Storage. *ACS Applied Energy Materials* **2021**, *4*, 1548–1559, Publisher: American Chemical Society.

(100) Khalafalla, M. A. H.; Hadia, N. M. A.; Elsayed, A. M.; Alruqi, M.; El Malti, W.; Shaban, M.; Rabia, M. ATO/Polyaniline/PbS Nanocomposite as Highly Efficient Photoelectrode for Hydrogen Production from Wastewater with Theoretical Study for the Water Splitting. *Adsorption Science & Technology* **2022**, *2022*, e5628032, Publisher: Hindawi.

(101) Liu, Y.-l.; Huang, H.; Peng, Y.-j. Fluorescent probe for simultaneous detection of human serum albumin and sulfite: A theoretical analysis. *Journal of Molecular Structure* **2022**, *1255*, 132441.

(102) Zhu, S.-f.; Gan, Q.; Feng, C. Multimolecular Complexes of CL-20 with Nitropyrazole Derivatives: Geometric, Electronic Structure, and Stability. *ACS Omega* **2019**, *4*, 13408–13417, Publisher: American Chemical Society.

(103) Zheng, K.; Li, D.; Jiang, L.; Li, X.; Xie, C.; Feng, L.; Qin, J.; Qian, S.; Pang, Q. Revisiting stacking interactions in tetrathiafulvalene and selected derivatives using tight-binding quantum chemical calculations and local coupled-cluster method. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2021**, *77*, 311–320, Publisher: International Union of Crystallography.

(104) Wang, Y.; Sun, Y.; Bian, H.; Zhu, L.; Xia, D.; Wang, H. Cyclodextrin Porous Liquid Materials for Efficient Chiral Recognition and Separation of Nucleosides. *ACS Applied Materials & Interfaces* **2020**, *12*, 45916–45928, Publisher: American Chemical Society.

(105) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* **1988**, *28*, 31–36, Publisher: American Chemical Society.

(106) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for generation of unique SMILES notation. *Journal of Chemical Information and Computer Sciences* **1989**, *29*, 97–101, Publisher: American Chemical Society.

(107) Krenn, M.; Ai, Q.; Barthel, S.; Carson, N.; Frei, A.; Frey, N. C.; Friederich, P.; Gaudin, T.; Gayle, A. A.; Jablonka, K. M. et al. SELFIES and the future of molecular string representations. *Patterns* **2022**, *3*, 100588.

(108) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *Journal of chemical information and computer sciences* **1992**, *32*, 244–255, ISBN: 0095-2338 Publisher: ACS Publications.

(109) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *Journal of Chemical Theory and Computation* **2015**, *11*, 2087–2096, Publisher: American Chemical Society.

(110) Alhamoud, K.; Ghunaim, Y.; Alshehri, A. S.; Li, G.; Ghanem, B.; You, F. Leveraging 2D molecular graph pretraining for improved 3D conformer generation with graph neural networks. *Computers & Chemical Engineering* **2024**, *183*, 108622.

(111) Axelrod, S.; Gómez-Bombarelli, R. Molecular machine learning with conformer ensembles. *Machine Learning: Science and Technology* **2023**, *4*, 035025, Publisher: IOP Publishing.

(112) Balcells, D.; Skjelstad, B. B. tmQM Dataset—Quantum Geometries and Properties of 86k Transition Metal Complexes. *Journal of Chemical Information and Modeling* **2020**, *60*, 6135–6146, Publisher: American Chemical Society.

(113) Gallarati, S.; Gerwen, P. v.; Laplaza, R.; Vela, S.; Fabrizio, A.; Corminboeuf, C. OSCAR: an extensive repository of chemically and functionally diverse organocatalysts. *Chemical Science* **2022**, *13*, 13782–13794, Publisher: The Royal Society of Chemistry.

(114) Gong, J.; Gong, W.; Wu, B.; Wang, H.; He, W.; Dai, Z.; Li, Y.; Liu, Y.; Wang, Z.; Tuo, X. et al. ASBase: The universal database for aggregate science. *Aggregate* **2023**, *4*, e263, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/agt2.263.

(115) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data* **2014**, *1*, 140022, Publisher: Nature Publishing Group.

(116) Qiao, Z.; Christensen, A. S.; Welborn, M.; Manby, F. R.; Anandkumar, A.; Miller, T. F. Informing geometric deep learning with electronic interactions to accelerate quantum chemistry. *Proceedings of the National Academy of Sciences* **2022**, *119*, e2205221119, Publisher: Proceedings of the National Academy of Sciences.

(117) Wang, Y.; Xiong, J.; Peng, F.; Li, Q.; Zeng, M.-H. Building a supramolecular gel with an ultra-low-molecular-weight Schiff base gelator and its multiple-stimulus responsive properties. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2022**, *640*, 128445.

(118) Cheng, Z.; Liu, J.; Jiang, T.; Chen, M.; Dai, F.; Gao, Z.; Ke, G.; Zhao, Z.; Ou, Q. Automatic Screen-out of Ir(III) Complex Emitters by Combined Machine Learning and Computational Analysis. *Advanced Optical Materials* **2023**, *11*, 2301093, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/adom.202301093.

(119) Nigam, A.; Pollice, R.; Krenn, M.; Gomes, G. d. P.; Aspuru-Guzik, A. Beyond generative models: superfast traversal, optimization, novelty, exploration and discovery (STONED) algorithm for molecules using SELFIES. *Chemical Science* **2021**, *12*, 7079–7090, Publisher: The Royal Society of Chemistry.

(120) Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. Leveraging large language models for predictive chemistry. *Nature Machine Intelligence* **2024**, 1–9, Publisher: Nature Publishing Group.

(121) Weiss, T.; Mayo Yanes, E.; Chakraborty, S.; Cosmo, L.; Bronstein, A. M.; Gershoni-Poranne, R. Guided diffusion for inverse molecular design. *Nature Computational Science* **2023**, *3*, 873–882, Number: 10 Publisher: Nature Publishing Group.

(122) Sinha, V.; Laan, J. J.; Pidko, E. A. Accurate and rapid prediction of pKa of transition metal complexes: semiempirical quantum chemistry with a data-augmented approach. *Physical Chemistry Chemical Physics* **2021**, *23*, 2557–2567, Publisher: The Royal Society of Chemistry.

(123) Shiotake, Y.; Takano, Y.; Saito, T. Semi-empirical and machine learning-based prediction of site of metabolisms mediated by aldehyde oxidase. *Chemical Physics Letters* **2023**, *833*, 140942.

(124) Mohanty, S.; Stevenson, J.; Browning, A. R.; Jacobson, L.; Leswing, K.; Halls, M. D.; Afzal, M. A. F. Development of scalable and generalizable machine learned force field for polymers. *Scientific Reports* **2023**, *13*, 17251, Number: 1 Publisher: Nature Publishing Group.

(125) Ondar, E. E.; Polynski, M. V.; Ananikov, V. P. Predicting 195Pt NMR Chemical Shifts in Water-Soluble Inorganic/Organometallic Complexes with a Fast and Simple Protocol Combining Semiempirical Modeling and Machine Learning. *ChemPhysChem* **2023**, *24*, e202200940, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cphc.202200940.

(126) Ranković, B.; Griffiths, R.-R.; Moss, H. B.; Schwaller, P. Bayesian optimisation for additive screening and yield improvements – beyond one-hot encoding. *Digital Discovery* **2023**, *3*, 654–666.

(127) Nandi, S.; Vegge, T.; Bhowmik, A. MultiXC-QM9: Large dataset of molecular and reaction energies from multi-level quantum chemical methods. *Scientific Data* **2023**, *10*, 783, Number: 1 Publisher: Nature Publishing Group.

(128) Unke, O. T.; Chmiela, S.; Gastegger, M.; Schütt, K. T.; Sauceda, H. E.; Müller, K.-R. SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature Communications* **2021**, *12*, 7273, Number: 1 Publisher: Nature Publishing Group.

(129) Staub, R.; Gantzer, P.; Harabuchi, Y.; Maeda, S.; Varnek, A. Challenges for Kinetics Predictions via Neural Network Potentials: A Wilkinson's Catalyst Case. *Molecules* **2023**, *28*, 4477, Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

(130) Murakami, T.; Ibuki, S.; Hashimoto, Y.; Kikuma, Y.; Takayanagi, T. Dynamics study of the post-transition-state-bifurcation process of the (HCOOH)H+ → CO + H3O+/HCO+ + H2O dissociation: application of machine-learning techniques. *Physical Chemistry Chemical Physics* **2023**, *25*, 14016–14027, Publisher: The Royal Society of Chemistry.

(131) Kubečka, J.; Christensen, A. S.; Rasmussen, F. R.; Elm, J. Quantum Machine Learning Approach for Studying Atmospheric Cluster Formation. *Environmental Science & Technology Letters* **2022**, *9*, 239–244, Publisher: American Chemical Society.

(132) Atz, K.; Isert, C.; Böcker, M. N. A.; Jiménez-Luna, J.; Schneider, G. Δ-Quantum machine-learning for medicinal chemistry. *Physical Chemistry Chemical Physics* **2022**, *24*, 10775–10783, Publisher: The Royal Society of Chemistry.

(133) Yan, J.; Rodríguez-Martínez, X.; Pearce, D.; Douglas, H.; Bili, D.; Azzouzi, M.; Eisner, F.; Virbule, A.; Rezasoltani, E.; Belova, V. et al. Identifying structure–absorption relationships and predicting absorption strength of non-fullerene acceptors for organic photovoltaics. *Energy & Environmental Science* **2022**, *15*, 2958–2973, Publisher: The Royal Society of Chemistry.

(134) Cremer, J.; Medrano Sandonas, L.; Tkatchenko, A.; Clevert, D.-A.; De Fabritiis, G. Equivariant Graph Neural Networks for Toxicity Prediction. *Chemical Research in Toxicology* **2023**, *36*, 1561–1573, Publisher: American Chemical Society.

(135) Iovanac, N. C.; Savoie, B. M. Simpler is Better: How Linear Prediction Tasks Improve Transfer Learning in Chemical Autoencoders. *The Journal of Physical Chemistry A* **2020**, *124*, 3679–3685, Publisher: American Chemical Society.

(136) Marx, D.; Hutter, J. Ab initio molecular dynamics: Theory and implementation. *Modern methods and algorithms of quantum chemistry* **2000**, *1*, 301–449.

(137) Pestana, L. R.; Mardirossian, N.; Head-Gordon, M.; Head-Gordon, T. Ab initio molecular dynamics simulations of liquid water using high quality meta-GGA functionals. *Chemical Science* **2017**, *8*, 3554–3565, Publisher: The Royal Society of Chemistry.

(138) Villard, J.; Bircher, M. P.; Rothlisberger, U. Structure and dynamics of liquid water from ab initio simulations: adding Minnesota density functionals to Jacob's ladder. *Chemical Science* **2024**, *15*, 4434–4451, Publisher: The Royal Society of Chemistry.

(139) Pu, M.; Nielsen, C. D.-T.; Senol, E.; Sperger, T.; Schoenebeck, F. Post-Transition-State Dynamic Effects in the Transmetalation of Pd(II)-F to Pd(II)-CF3. *JACS Au* **2024**, *4*, 263–275, Publisher: American Chemical Society.

(140) Esch, B. v. d.; Peters, L. D. M.; Sauerland, L.; Ochsenfeld, C. Quantitative Comparison of Experimental and Computed IR-Spectra Extracted from Ab Initio Molecular Dynamics. *Journal of Chemical Theory and Computation* **2021**, *17*, 985–995, Publisher: American Chemical Society.

(141) Zou, Y.; Houk, K. N. Mechanisms and Dynamics of Synthetic and Biosynthetic Formation of Delitschiapyrones: Solvent Control of Ambimodal Periselectivity. *Journal of the American Chemical Society* **2021**, *143*, 11734–11740, Publisher: American Chemical Society.

(142) Asif, M.; Sajid, H.; Ullah, F.; Khan, S.; Ayub, K.; Amjad Gilani, M.; Arshad, M.; Salim Akhter, M.; Mahmood, T. Quantum chemical study on sensing of NH3, NF3, NCl3 and NBr3 by using cyclic tetrapyrrole. *Computational and Theoretical Chemistry* **2021**, *1199*, 113221.

(143) Yañez, O.; Rios, R. P. Revisiting hollow all-carbon structures: Aromaticity and kinetic stability. *Diamond and Related Materials* **2023**, *136*, 110057.

(144) Wang, Z.; Leng, Z.; Wang, D.; Si, Z.; Duan, Q.; Han, D. Zn(II) cation-induced solution emission of 4-(4-(1,2,2-triphenylvinyl)phenyl)benzoic acid: Intuitionistic evidence of RIM mechanism. *Journal of Molecular Structure* **2021**, *1234*, 130128.

(145) Yu, C.; Li, D.; Han, W.; Kong, Y.; Zhang, L.; Chen, M.; Hu, K. Improvement of photostability and thermal stability of PVC by carbon quantum dots loaded on TiO2. *Journal of Vinyl and Additive Technology* **2023**, *29*, 864–876, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/vnl.22012.

(146) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nature Chemistry* **2014**, *6*, 1044–1048, Publisher: Nature Publishing Group.

(147) Zhang, B.; Hao, M.; Xiong, J.; Li, X.; Koopman, J. Ab initio molecular dynamics calculations on electron ionization induced fragmentations of C4F7N and C5F10O for understanding their decompositions under discharge conditions. *Physical Chemistry Chemical Physics* **2023**, *25*, 7540–7549, Publisher: The Royal Society of Chemistry.

(148) Fu, D.; Habtegabir, S. G.; Wang, H.; Feng, S.; Han, Y. Understanding of protomers/deprotomers by combining mass spectrometry and computation. *Analytical and Bioanalytical Chemistry* **2023**, *415*, 3847–3862.

(149) Seritan, S.; Wang, Y.; Ford, J. E.; Valentini, A.; Gold, T.; Martínez, T. J. InteraChem: Virtual Reality Visualizer for Reactive Interactive Molecular Dynamics. *Journal of Chemical Education* **2021**, *98*, 3486–3492, Publisher: American Chemical Society.

(150) Pracht, P.; Grimme, S. Efficient Quantum-Chemical Calculations of Acid Dissociation Constants from Free-Energy Relationships. *The Journal of Physical Chemistry A* **2021**, *125*, 5681–5692, Publisher: American Chemical Society.

(151) Pracht, P.; Bannwarth, C. Fast Screening of Minimum Energy Crossing Points with Semiempirical Tight-Binding Methods. *Journal of Chemical Theory and Computation* **2022**, *18*, 6370–6385.

(152) Chan, B.; Dawson, W.; Nakajima, T. Sorting drug conformers in enzyme active sites: the XTB way. *Physical Chemistry Chemical Physics* **2024**, *26*, 12610–12618.

(153) Stahn, M.; Grimme, S.; Salthammer, T.; Hohm, U.; Palm, W.-U. Quantum chemical calculation of the vapor pressure of volatile and semi volatile organic compounds. *Environmental Science: Processes & Impacts* **2022**, *24*, 2153–2166, Publisher: The Royal Society of Chemistry.

(154) Williams, D. C.; Inala, N. Physics-Informed Generative Model for Drug-like Molecule Conformers. *Journal of Chemical Information and Modeling* **2024**, *64*, 2988–3007.

(155) Plett, C.; Grimme, S. Automated and Efficient Generation of General Molecular Aggregate Structures. *Angewandte Chemie International Edition* **2023**, *62*, e202214477, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202214477.

(156) Grimme, S.; Bannwarth, C.; Caldeweyher, E.; Pisarek, J.; Hansen, A. A general intermolecular force field based on tight-binding quantum chemical calculations. *The Journal of Chemical Physics* **2017**, *147*, 161708.

(157) Lambert, H.; Castillo Bonillo, A.; Zhu, Q.; Zhang, Y.-W.; Lee, T.-C. Supramolecular gating of guest release from cucurbit[7]uril using de novo design. *npj Computational Materials* **2022**, *8*, 1–8, Number: 1 Publisher: Nature Publishing Group.

(158) Salman, M.; Lee, J. W.; Lee, S. H.; Lee, M. H.; Pham, V. D.; Kim, M.-S.; Cho, D.; Lee, H. J. A comparative study of ammonia solubility in imidazolium-based ionic liquids with different structural compositions. *Heliyon* **2024**, *10*, Publisher: Elsevier.

(159) Izquierdo, S.; Cintas, P.; J. Durán-Valle, C.; Concepción, J. G. d. l.; M. López-Coca, I. Reinvigorating aza-Michael reactions under ionic liquid catalysis: a greener approach. *Organic & Biomolecular Chemistry* **2024**, *22*, 2423–2434.

(160) Paul, B.; Natarajan, R. Metal–Organic Cage Receptors for Encapsulation and Sensing of Bile Acids. *Inorganic Chemistry* **2024**, acs.inorgchem.4c00934.

(161) Aguiar, C.; Dattani, N.; Camps, I. Möbius boron-nitride nanobelts interacting with heavy metal nanoclusters. *Physica B: Condensed Matter* **2023**, *668*, 415178.

(162) 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. *Journal of Chemical Theory and Computation* **2022**, *18*, 3174–3189.

(163) Spino, C.; Latil, M.; Lessard, R.; Fevre-Renault, Q.; Legault, C. Y. N-Oxides as Control Element for the Direction of a Sigmatropic Rearrangement: Application as a Switch for Fluorescence. *Chemistry – A European Journal* **2023**, *29*, e202301356, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/chem.202301356.

(164) Covington, C. L.; Puente, A. R.; Polavarapu, P. L. Pitfalls in the Optimization of Conformer Populations to Maximize the Similarity between Predicted and Experimental Chiroptical Spectra. *The Journal of Physical Chemistry A* **2024**, *128*, 129–138.

(165) Plett, C.; Grimme, S.; Hansen, A. Conformational energies of biomolecules in solution: Extending the MPCONF196 benchmark with explicit water molecules. *Journal of Computational Chemistry* **2024**, *45*, 419–429, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.27248.

(166) Katsyuba, S. A.; Burganov, T. I. Computational analysis of the vibrational spectra and structure of aqueous cytosine. *Physical Chemistry Chemical Physics* **2023**, *25*, 24121–24128, Publisher: The Royal Society of Chemistry.

(167) Lipparini, F.; Stamm, B.; Cancès, E.; Maday, Y.; Mennucci, B. Fast Domain Decomposition Algorithm for Continuum Solvation Models: Energy and First Derivatives. *Journal of Chemical Theory and Computation* **2013**, *9*, 3637–3648.

(168) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *The Journal of Physical Chemistry* **1995**, *99*, 2224–2235.

(169) 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. *The Journal of Physical Chemistry B* **2009**, *113*, 6378–6396.

(170) Wagen, C.; Wagen, A. Efficient and Accurate pKa Prediction Enabled by Pre-Trained Machine-Learned Interatomic Potentials. *ChemRxiv* **2024**, {\url{10.26434/chemrxiv-2024-8489b } (accessed 15th of June 2024).

(171) Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C. et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

(172) Grimme, S.; Rose, T. mcGFN-FF: an accurate force field for optimization and energetic screening of molecular crystals. *Zeitschrift für Naturforschung B* **2024**, *79*, 191–200.

(173) 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. *Journal of Chemical Theory and Computation* **2021**, *17*, 7827–7849, Publisher: American Chemical Society.

(174) Gao, W.; Li, Y.; Zhoujin, Y.; Zeng, Q.; Fang, L.; Parkin, S.; Li, T.; Ruan, S.; Long, S. A Peptoid-like Approach Led to Lactam–Lactam Dimer Formation in 2-Hydroxy- *N* -alkyl- *N* -phenyl-nicotinamides and Their Polymorphism and Solvatomorphism. *ACS Omega* **2023**, *8*, 45447–45456.

(175) Moore, C.; Zhang, D.; Rousseau, R.; Glezakou, V.-A.; McEwen, J.-S. Determining the Adsorption Energetics of 2,3-Butanediol on RuO2(110): Coupling First-Principles Calculations With Global Optimizers. *Frontiers in Energy Research* **2022**, *9*, 781001.

(176) Gouveia, J. D.; Novell-Leruth, G.; Reis, P. M. L. S.; Viñes, F.; Illas, F.; Gomes, J. R. B. First-Principles Calculations on the Adsorption Behavior of Amino Acids on a Titanium Carbide MXene. *ACS Applied Bio Materials* **2020**, *3*, 5913–5921, Publisher: American Chemical Society.

(177) Priandani, L.; Aliefa, A.; Arjasa, O. P.; Pambudi, F. I. An Experimental and Computational Study of Zeolitic Imidazole Framework (ZIF-8) Synthesis Modulated with Sodium Chloride and Its Interaction with CO2. *Bulletin of Chemical Reaction Engineering & Catalysis* **2023**, *18*, 521–538, Publisher: Masyarakat Katalis Indonesia - Indonesian Catalyst Society (MKICS).

(178) Wang, J.; Hou, Y.-C.; Li, L.; Guo, Y.-R.; Ding, S.-D.; Pan, Q.-J. Locating nitrate ions in the extraction and separation of trivalent Am and Eu by NTAamide: Quantum chemistry and molecular dynamics probe. *Separation and Purification Technology* **2024**, *337*, 126442.

(179) Davies, D. W.; Seo, B.; Park, S. K.; Shiring, S. B.; Chung, H.; Kafle, P.; Yuan, D.; Strzalka, J. W.; Weber, R.; Zhu, X. et al. Unraveling two distinct polymorph transition mechanisms in one n-type single crystal for dynamic electronics. *Nature Communications* **2023**, *14*, 1304, Number: 1 Publisher: Nature Publishing Group.

(180) Cui, M.; Reuter, K.; Margraf, J. T. Obtaining Robust Density Functional Tight-Binding Parameters for Solids across the Periodic Table. *Journal of Chemical Theory and Computation* **2024**, *20*, 5276–5290.

(181) Germain, A.; Corno, M.; Ugliengo, P. Computing Binding Energies of Interstellar Molecules by Semiempirical Quantum Methods: Comparison Between DFT and GFN2 on Crystalline Ice. Computational Science and Its Applications – ICCSA 2021. Cham, 2021; pp 632–645.

(182) Vicent-Luna, J. M.; Apergi, S.; Tao, S. Efficient Computation of Structural and Electronic Properties of Halide Perovskites Using Density Functional Tight Binding: GFN1-xTB Method. *Journal of Chemical Information and Modeling* **2021**, *61*, 4415–4424, Publisher: American Chemical Society.

(183) Zhugayevych, A.; Sun, W.; van der Heide, T.; Lien-Medrano, C. R.; Frauenheim, T.; Tretiak, S. Benchmark Data Set of Crystalline Organic Semiconductors. *Journal of Chemical Theory and Computation* **2023**, *19*, 8481–8490, Publisher: American Chemical Society.

(184) Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science* **2022**, *2*, 718–728.

(185) Batatia, I.; Benner, P.; Chiang, Y.; Elena, A. M.; Kovács, D. P.; Riebesell, J.; Advincula, X. R.; Asta, M.; Avaylon, M.; Baldwin, W. J. et al. A foundation model for atomistic materials chemistry. 2024; `http://arxiv.org/abs/2401.00096`, arXiv:2401.00096.

(186) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* **2023**, *5*, 1031–1041.

(187) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. *The Journal of Physical Chemistry* **1996**, *100*, 19357–19363, Publisher: American Chemical Society.

(188) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts* **2007**, *117*, 185–199.

(189) Tinacci, L.; Germain, A.; Pantaleone, S.; Ceccarelli, C.; Balucani, N.; Ugliengo, P. Theoretical Water Binding Energy Distribution and Snowline in Protoplanetary Disks. *The Astrophysical Journal* **2023**, *951*, 32, Publisher: The American Astronomical Society.

(190) Molpeceres, G.; Kästner, J. Adsorption of H2 on amorphous solid water studied with molecular dynamics simulations. *Physical Chemistry Chemical Physics* **2020**, *22*, 7552–7563, Publisher: The Royal Society of Chemistry.

(191) Arsenault-Escobar, S.; Fuentes-Galvez, J. F.; Orellana, C.; Bollo, S.; Sierra–Rosales, P.; Miranda-Rojas, S. Unveiling the tartrazine binding mode with ds–DNA by UV–visible spectroscopy, electrochemical, and QM/MM methods. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2023**, *292*, 122400.

(192) Lucia-Tamudo, J.; Díaz-Tendero, S.; Nogueira, J. J. Intramolecular and intermolecular hole delocalization rules the reducer character of isolated nucleobases and homogeneous single-stranded DNA. *Physical Chemistry Chemical Physics* **2023**, *25*, 14578–14589, Publisher: The Royal Society of Chemistry.

(193) Syakaev, V. V.; Masliy, A. N.; Podyachev, S. N.; Sudakova, S. N.; Shvedova, A. E.; Lentin, I. I.; Gorbunov, A. N.; Vatsouro, I. M.; Lapaev, D. V.; Sh. Mambetova, G. et al. NMR and DFT-study of new luminescent Eu3+ complexes based on calix[4]arenes with 1,2,3-triazole and 1,3-diketone groups. *Inorganica Chimica Acta* **2024**, *561*, 121848.

(194) Cariello, G.; Zeoly, L. A.; Piscelli, B. A.; Lectka, T.; Cormanich, R. A. The pitfalls of using JHF spin–spin coupling constants to infer hydrogen bond formation in organofluorine compounds. *Chemical Communications* **2023**, *59*, 14661–14664, Publisher: The Royal Society of Chemistry.

(195) Machado, T. F. G.; Purg, M.; McMahon, S. A.; Read, B. J.; Oehler, V.; Åqvist, J.; Gloster, T. M.; da Silva, R. G. Dissecting the Mechanism of (R)-3-Hydroxybutyrate Dehydrogenase by Kinetic Isotope Effects, Protein Crystallography, and Computational Chemistry. *ACS Catalysis* **2020**, *10*, 15019–15032, Publisher: American Chemical Society.

(196) Peralta, F. A.; Huidobro-Toro, J. P.; Mera-Adasme, R. Hybrid QM/MM Simulations Confirm Zn(II) Coordination Sphere That Includes Four Cysteines from the P2 × 4R Head Domain. *International Journal of Molecular Sciences* **2021**, *22*, 7288, Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

(197) Nkungli, N. K.; Fouegue, A. D. T.; Tasheh, S. N.; Bine, F. K.; Hassan, A. U.; Ghogomu, J. N. In silico investigation of falcipain-2 inhibition by hybrid benzimidazole-thiosemicarbazone antiplasmodial agents: A molecular docking, molecular dynamics simulation, and kinetics study. *Molecular Diversity* **2023**, 475–496.

(198) Platzer, G.; Ptaszek, A. L.; Böttcher, J.; Fuchs, J. E.; Geist, L.; Braun, D.; McConnell, D. B.; Konrat, R.; Sánchez-Murcia, P. A.; Mayer, M. Ligand 1H NMR Chemical Shifts as Accurate Reporters for Protein-Ligand Binding Interfaces in Solution**. *ChemPhysChem* **2024**, *25*, e202300636, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cphc.202300636.

(199) Morris, R. H. Methane activation by a single copper center in particulate methane monooxygenase: A computational study. *Inorganica Chimica Acta* **2020**, *503*, 119441.

(200) Yan, Z.; Li, X.; Chung, L. W. Multiscale Quantum Refinement Approaches for Metalloproteins. *Journal of Chemical Theory and Computation* **2021**, *17*, 3783–3796, Publisher: American Chemical Society.

(201) Lashkov, A. A.; Eistrich-Geller, P. A.; Samygina, V. R.; Rubinsky, S. V. QM/MM Study of a Nucleophilic Substitution Reaction Catalyzed by Uridine Phosphorylase from Vibrio cholerae. *Crystals* **2023**, *13*, 803, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

(202) Ojha, A. A.; Votapka, L. W.; Amaro, R. E. QMrebind: incorporating quantum mechanical force field reparameterization at the ligand binding site for improved drug-target kinetics through milestoning simulations. *Chemical Science* **2023**, *14*, 13159–13175, Publisher: The Royal Society of Chemistry.

(203) Suenaga, S.; Takano, Y.; Saito, T. Unraveling Binding Mechanism and Stability of Urease Inhibitors: A QM/MM MD Study. *Molecules* **2023**, *28*, 2697, Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

(204) Mera-Adasme, R.; Domínguez, M.; Denis-Alpizar, O. A benchmark for the size of the QM system required for accurate hybrid QM/MM calculations on the metal site of the protein copper, zinc superoxide dismutase. *Journal of Molecular Modeling* **2019**, *25*, 176.

(205) Gallmetzer, H. G.; Hofer, T. S. Probing the range of applicability of structure- and energy-adjusted QM/MM link bonds II: Optimized link bond parameters for density functional tight binding approaches. *Journal of Computational Chemistry* **2022**, *43*, 746–756, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.26830.

(206) Liu, H.; Fu, H.; Shao, X.; Cai, W.; Chipot, C. Accurate Description of Cation$-\pi$ Interactions in Proteins with a Nonpolarizable Force Field at No Additional Cost. *Journal of Chemical Theory and Computation* **2020**, *16*, 6397–6407, Publisher: American Chemical Society.

(207) Liu, H.; Fu, H.; Chipot, C.; Shao, X.; Cai, W. Accuracy of Alternate Nonpolarizable Force Fields for the Determination of Protein–Ligand Binding Affinities Dominated by Cation$-\pi$ Interactions. *Journal of Chemical Theory and Computation* **2021**, *17*, 3908–3915, Publisher: American Chemical Society.

(208) Franzke, Y. J.; Holzer, C.; Andersen, J. H.; Begušić, T.; Bruder, F.; Coriani, S.; Della Sala, F.; Fabiano, E.; Fedotov, D. A.; Fürst, S. et al. TURBOMOLE: Today and Tomorrow. *Journal of Chemical Theory and Computation* **2023**, *19*, 6859–6890, Publisher: American Chemical Society.

(209) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian~16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

(210) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z.-H. et al. Recent developments in the P y SCF program package. *The Journal of Chemical Physics* **2020**, *153*, 024109.

(211) Wesołowski, P. A.; Wales, D. J.; Pracht, P. Multilevel Framework for Analysis of Protein Folding Involving Disulfide Bond Formation. *The Journal of Physical Chemistry B* **2024**, *128*, 3145–3156, Publisher: American Chemical Society.

(212) Pracht, P.; Caldeweyher, E.; Ehlert, S.; Grimme, S. A Robust Non-Self-Consistent Tight-Binding Quantum Chemistry Method for large Molecules. 2019; `https://chemrxiv.org/engage/chemrxiv/article-details/60c742abbdbb890c7ba3851a`.

(213) Pracht, P.; Bannwarth, C. Finding Excited-State Minimum Energy Crossing Points on a Budget: Non-Self-Consistent Tight-Binding Methods. *The Journal of Physical Chemistry Letters* **2023**, *14*, 4440–4448, Publisher: American Chemical Society.

(214) Spicher, S.; Grimme, S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angewandte Chemie International Edition* **2020**, *59*, 15665–15673, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202004239.

(215) Public Zotero library of xTB related citations. `https://www.zotero.org/groups/5374558/xtb`, 2024; (accessed 15th of June 2024).

(216) Grimme, S.; Bannwarth, C. Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB). *The Journal of chemical physics* **2016**, *145*.

(217) Van Khai, T.; Viet Hai, L.; Thi Thu Ha, N.; Thi Thom, N.; Van Trang, N.; Thi Nam, P.; Ngoc Ha, N.; Dai Lam, T. Combined experimental and theoretical studies on enlarged bandgap and improved photoelectrochemical properties of reduced graphene oxide film by hydrogen annealing. *Journal of Electroanalytical Chemistry* **2021**, *900*, 115722.

(218) Alammar, A.; Park, S.-H.; Ibrahim, I.; Arun, D.; Holtzl, T.; Dumée, L. F.; Lim, H. N.; Szekely, G. Architecting neonicotinoid-scavenging nanocomposite hydrogels for environmental remediation. *Applied Materials Today* **2020**, *21*, 100878.

(219) Lin, M.; Adpressa, D. A.; Feng, M.; Lu, Y.; Clark, B. R. Brevianthrones, bianthrones from a Chinese isolate of the endophytic fungus *Colletotrichum brevisporum*. *Phytochemistry* **2021**, *188*, 112792.

(220) Ertl, P.; Altmann, E.; Racine, S.; Decoret, O. Which boronic acids are used most frequently for synthesis of bioactive molecules? *Bioorganic & Medicinal Chemistry* **2023**, *91*, 117405.

TOC Graphic