

ULYSSES: An Efficient and Easy to Use Semiempirical Library for C++

Filipe Menezes* and Grzegorz M. Popowicz*



Cite This: *J. Chem. Inf. Model.* 2022, 62, 3685–3694



Read Online

ACCESS |



Metrics & More

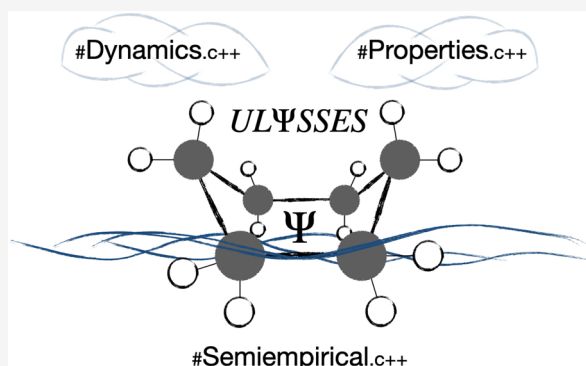


Article Recommendations



Supporting Information

ABSTRACT: We introduce *ULYSSES*, a user-friendly and robust C++ library for semiempirical quantum chemical calculations. In its current version, *ULYSSES* is equipped with a large set of different semiempirical models, most of which are based on the Neglect of Diatomic Differential Overlap (NDDO) approximation. Empirical corrections for dispersion and hydrogen bonding are available for most methods, so that higher quality is achieved in the calculation of energies of nonbonded complexes. The library is furthermore equipped with geometry optimization, as well as modules for calculating molecular properties of general interest. Ideal gas thermodynamics is available and allows single structure as well as conformer (multistructure) averaged properties to be calculated. We offer the possibility to use several vibrational partition functions according to the nature of interactions being studied: for covalent systems, the traditional harmonic oscillator approximation is available; for nonbonded complexes, we systematically extended the partition function proposed by Grimme for all thermodynamic functions. The library is also capable of running Born–Oppenheimer molecular dynamics.



INTRODUCTION

With the advent of local correlation methods, highly efficient variants of several *ab initio* and Density Functional Theory (DFT) methods^{1–6} were made available. These advances permitted for the first time the single-point energy calculation of a small protein of 644 atoms at the Singles Doubles Coupled Cluster with Perturbative Treatment of Triples (CCSD(T)) level of theory.³

One of the most interesting features of local correlation methods is that the basis set superposition error is significantly reduced.^{1,7,8} However, the need for large basis sets is not eliminated, so that mostly (linear scaling) DFT methods find their way into the simulation of large systems, often using mixed QM/MM schemes. Due to the complexity of some of the integrals involved in the calculations, DFT is however still overly demanding in terms of computational power for most interesting problems in applied sciences.⁹ Semiempirical methods, i.e., parametrized low-cost variants of the Hartree–Fock (HF) and DFT methods, aim at introducing important quantum effects^{10–12} at a reduced computational cost. These are of particular importance for large systems, namely the ones of biological interest and on applications related to material sciences.^{13,14} A very interesting achievement of semiempirical quantum chemistry is the realization that the surfaces of proteins transfer a considerable amount of charge to the neighboring solvent molecules.¹⁵ This observation, which was later corroborated by further *ab initio* calculations, would not have been possible using force fields.

Modern semiempirical theory is based mostly on two classes of methods. On one hand, there is the class of Neglect Diatomic Differential Overlap (NDDO) methods, introduced by Pople, Santry, and Segal.¹⁶ The most prominent methods were developed by the school of Dewar, and they include the Modified Neglect Diatomic Differential Overlap (MNDO) and Austin Model 1 (AM1) methods.^{17–45} Other popular methods from the NDDO family are the Parametrized Models (PMs) of Stewart^{46–49} and the Recife Model 1 (RM1),⁵⁰ the latter designed with the particular focus of describing biological systems. There were some attempts to make non- or less-parametrized variants of the NDDO approximation.^{51–53} These never enjoyed the success of the parametrized ones due to the severity of the approximations used.⁵⁴ Another important development was the introduction of the Orthogonalization-Corrected Methods (OMx) by Thiel and co-workers.^{55–57} The introduction of (overlap) orthogonalization effects improves several of the problems related to semiempirical quantum chemistry. For the reader interested in details on all these methods, we recommend a review article.⁵⁸

Density Functional Tight Binding (DFTB) builds the other family of semiempirical methods, and instead of simplifying the

Received: June 15, 2022

Published: August 5, 2022



HF equations, its starting point is DFT. In very broad terms, one takes an expansion of the electronic density in terms of a reference density (which corresponds to the density of isolated atoms) plus charge fluctuations. When this approximation to the total density is used to calculate the energy, it leads to a series expansion of the latter as a function of powers of the charge fluctuations. Different truncations of this energy lead to different DFTB approximations.^{59–67}

Although much has been put forth in semiempirical quantum chemistry, there are still many problems to tackle and much improvement that can be achieved. Here, we introduce the *ULYSSES* library, a carefully designed and header-only C++ library for semiempirical quantum chemistry and property calculation. The current capabilities include several dispersion-corrected semiempirical methods and algorithms based on molecular and metadynamics. Due to its simple architecture, users can easily build their own programs, either from scratch or by using template files, without worrying about complex installation procedures or spending time installing several external dependencies: the only requirement for using *ULYSSES* is a C++ compiler. The primary advantage of *ULYSSES* over other packages is its simplicity and ease of use without sacrificing potential functionality nor efficiency. Furthermore, we are convinced that future developments we will bring into the library will make it attractive for a larger community.

COMPUTATIONAL DETAILS

All calculations were performed using our newly developed library. Unless otherwise stated, geometries were optimized using the Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) algorithm with the Moré-Thuente line search and the default convergence criteria. The Hamiltonian of choice for optimizing geometries is always consistent with the method chosen for energy evaluation. Initial geometries are obtained via standard Universal Force Field (UFF) optimization using Avogadro 1.2.^{68,69} Graphics were generated using ChimeraX.^{70,71} Plots were generated using matplotlib.⁷²

FUNCTIONALITIES AND GENERAL STRUCTURE

ULYSSES is written in C++ as a header-only library. We require three external libraries to compile the code and run calculations: Eigen⁷³ is used for matrix operations; rapidJSON⁷⁴ is used for parsing JSON files; xsum^{75,76} is necessary in some modules requiring accurate summations. All these modules are included within *ULYSSES*'s structure, or they can be allocated separately. This ensures that the main library remains available as header-only. Of all modules and libraries, only xsum requires C++11 compilers. The rest of the code is not compiler specific. Further details on the program philosophy and argumentation for our choices are given in the [Supporting Information](#).

The current list of functionalities of *ULYSSES* includes

- (Empirically corrected) NDDO methods and one DFTB method.
- Geometry optimization.
- Mulliken population analysis.
- Reactivity indices.
- Harmonic frequency calculation.
- Molecular dynamics (and metadynamics) calculations based on semiempirical methods.
- Ideal gas properties.

The program is subdivided into four main modules:

- The quantum chemical calculation, which is coupled to the calculation of molecular properties. These include for instance population analysis and the calculation of harmonic normal modes. The whole module is centralized in the *QCbasis* class which includes the general implementation of the methods.
- The geometry optimization module is a function that takes as input a particular quantum chemical method (or function object) and a solver object.
- Molecular Dynamics is based on the Dynamics class. This takes a quantum chemical object in and solves the Newtonian equations of motion to let the system evolve in time. Note that we have currently no force field available in *ULYSSES*.
- Ideal gas properties are currently available from several classes, all of them possibly (but not necessarily) connected to a particular quantum chemical calculation or a JSON database of structural information. This database allows users to create their own libraries of molecular data for thermodynamical property calculation or also to easily include conformational anharmonicity in the calculation of ideal gas thermodynamic properties.

Details on the main submodules of *ULYSSES* are given in the [Supporting Information](#). The connectivity between modules is presented in [Figure 1](#).

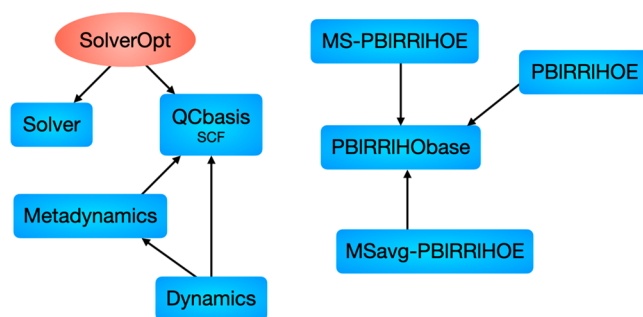


Figure 1. Global program structure and the interconnection between different classes.

AVAILABLE HAMILTONIANS AND CORRECTIONS

We have currently several Hamiltonians available in our code, mostly from the family of NDDO methods. This includes MNDO,^{17,18,22–36} AM1,^{19,37–45,77} Parametrized Model 3 (PM3),^{46–48,77} MNDOd,^{21,78} PM6,⁴⁹ RM1,⁵⁰ PM3-BP,⁷⁹ MNDO-PDDG,^{80,81} and PM3-PDDG.^{80–82} We also have GFN2-xTB available.⁸³ We note that the references given for each method include the method description and the lists of parameters we use in our code. Furthermore, in the case of NDDO Hamiltonians, the wave functions of normal (closed-shell) molecules are obtained using the Restricted Hartree–Fock (RHF) formalism, i.e., by solving the Roothaan–Hall equations. For open-shell systems—radicals and high multiplicity molecules—we implemented the Unrestricted Hartree–Fock (UHF) method that solves instead the Pople–Nesbet equations. This information may be processed automatically, so that the user does not necessarily need to specify any option. The default program output indicates which algorithm is used.

A list of the elements available for each method implemented in *ULYSSES* is given in the [Supporting Information](#). To aid our users, the checkAtoms() function is called automatically

whenever a calculation is started. This function verifies whether parameters are available for all atoms in the chemical system declared. If one element is not parametrized, the calculation is stopped. We wish to stress that this applies to element specific parameters only. For instance, in the PM6 method there are several atom-pair parameters missing in the original method definition. This does not stop the calculation from taking place.

Most semiempirical methods mentioned above are oblivious to or inaccurate for nonbonded interactions, particularly dispersion and hydrogen bonding.^{84–86} We have in our program the possibility to complement NDDO methods using empirical force-field corrections, which include D3H4,^{87,88} D3H4X,⁸⁹ and D3H+.^{90,91} The first correction is available for AM1, PM3, RM1, and PM6, whereas the other two corrections are available for PM6 only.

Despite their simplicity and the use of minimal basis sets, Self-Consistent Field (SCF) convergence problems might arise in NDDO calculations. This is particularly troublesome for methods using *spd* basis sets. Convergence accelerators like Direct Inversion of the Iterative Subspace (DIIS)^{92,93} or Relaxed Constraints Algorithm (RCA)⁹⁴ are embedded in the SCF and used by default. In those situations, the density matrix is obtained from the Molecular Orbital (MO) coefficients calculated by traditional matrix diagonalization. For cases in which the HOMO–LUMO gap is too small, the ordering of orbitals might revert in successive iterations, which keeps the SCF from converging. This abnormal behavior is also reflected in strong oscillations in the electronic energy, from which the SCF may not recover. Level-shifts may be employed to modify predominantly the eigenvalues associated with unoccupied MOs^{95,96} and to efficiently correct the behavior. We have furthermore the possibility to use pseudodiagonalization⁹⁷ and the density damping method of Badziag and Solms⁹⁸ but currently not in combination with DIIS. This wide set of options for aiding SCF convergence makes *ULYSSES* a very robust library for semiempirical quantum chemistry calculation.

Performing a single-point energy calculation is accomplished in four lines of code. For instance, a PM6-D3H+ calculation on a singlet ($2S + 1 = 1$) molecule of charge 0 may be run using the following four lines of code

```
Molecule mol("path/to/geom.xyz",0,1,"C1");
BSet bs(mol,"pm6");
PM6 nddo(bs,mol,"0","D3H+");
nddo.Calculate();
```

Note that by using "0" we let the program decide whether to use the RHF or UHF formalisms. Alternatively, one may use "rhf" or "uhf" to enforce a specific algorithm.

When studying large molecular systems, solvent effects are extremely relevant. Treating (large) solvation shells at quantum chemical levels is resource intensive and brings out other problems. This may be circumvented by means of implicit solvation. *ULYSSES* has currently the ALPB⁹⁹ model available for GFN2-xTB only. This means that we still do not have solvation models to be used with NDDO methods. To use the solvation model, one needs only a call to the `setSolvent()` method. This method takes a string as argument. This must be placed anywhere between method instantiation and the calling to `Calculate()`. The [Supporting Information](#) contains a list of solvents available and the dielectric constants used.

■ AVAILABLE MOLECULAR PROPERTIES

Besides laying down the foundation for the quantum-mechanical methods, *QCbasis* offers the possibility to calculate system specific properties. This includes getting the energies and coefficients for HOMO and LUMO, gradients and Hessian, Mulliken charges and bond orders, Mayer bond orders,¹⁰⁰ Armstrong bond orders,¹⁰¹ and harmonic frequencies. GFN2-xTB also includes the possibility to calculate polarizabilities (atomic and molecular) and molecular dispersion coefficients. For the case of some NDDO methods, we have Charge Model (CM) 1,¹⁰² CM2,¹⁰³ CM3,^{104,105} and CM5¹⁰⁶ charges available (as well as the respective molecular dipole moments). Note though that CM5 can potentially be used with any method available, and the results might be of lower quality because we do not use Hirshfeld charges.¹⁰⁷

QCbasis allows the estimation of Ionization Potentials (IPs) using either Koopman's theorem or the definition

$$\text{IP} = E^+ - E^0$$

In the latter expression, E^0 represents the target system, and E^+ is the target system with one less electron. Electron Affinities (EAs) are available via the definition only. Using these, Mulliken electronegativities and chemical hardnesses can be calculated. Note that these quantities are available, for consistency reasons, from IPs and EAs calculated using the definition. Fukui indices can also be calculated using one of the two modalities available. The most demanding calculation goes via the local reactivity indices as defined by Yang and Mortier.¹⁰⁸ A simpler variant uses the absolute values of the frontier molecular orbitals.¹⁰⁹ The Fukui indices may furthermore be used to calculate local softness parameters.¹¹⁰

■ BENCHMARK

Most semiempirical methods available within *ULYSSES* are well established and sufficiently documented regarding performance and quality. To benchmark the code, we compare the results of calculations against well-established programs, as well as timing.

The proper benchmarking of *ULYSSES* in terms of accuracy is reasonably extensive since several aspects must be properly analyzed. We relegate the full details to the [Supporting Information](#) (zip file "Module Benchmark"). Here, we present only a brief summary of the analysis. GFN2-xTB is benchmarked against xTB.¹¹¹ Deviations are always minor and result from differences in the physical constants used and unit conversion. By replacing ours with the constants available in xTB, the discrepancy between programs is reduced to numerical noise. This is discussed for the system with larger deviations and tested also for some other cases. NDDO methods were compared against pDynamo.¹¹² Because we have a common implementation for all NDDO methods, we believe it suffices to test only PM6, which requires *spd* basis of AOs. The discrepancy between implementations is in this case larger than in the case of GFN2-xTB. We show however that this is a result of different overlap integral codes. While pDynamo uses Gaussian orbitals, our implementation uses Slater orbitals (for NDDO methods). Dispersion and hydrogen bond corrections are also tested for meaningful cases, and we show that our code yields the same as reference programs.

ULYSSES's performance is tested by running single-point energy calculations on structure databases as well as some large molecules. All calculations were run on an *imac* 3.6 GHz quad-core Intel Core i7 with 16 GB of RAM. Whenever applicable,

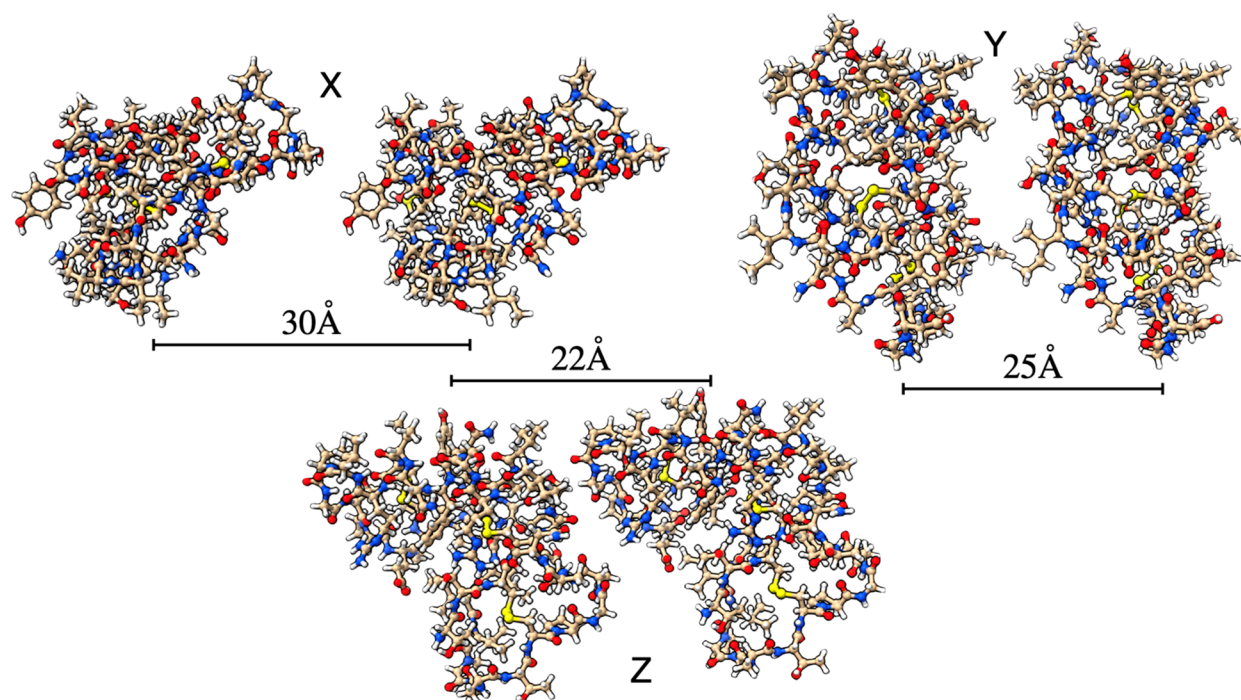


Figure 2. Structure of the 1crn dimers used for benchmarking. The structures are the X, Y, and Z dimers. X is the dimer with the largest internuclear distance, Y is the structure with dispersion dominated interactions, and Z is the H-bond dominated dimer.

protons were added using Avogadro^{68,69} in order to obtain electroneutral systems. Note that we are not interested in reproducing the physics and chemistries of real systems, instead we wish to benchmark performance. Calculations were all run in serial on the CPU with the default options, except when noted otherwise. This means traditional matrix diagonalization with DIIS. We note that *ULYSSES* is currently not yet parallelized.

The first test set is the QM9 database,^{113,114} which contains a total of 133885 structures with up to 9 heavy atoms from the second period (C, N, O, F). The database has on average molecules with 18 atoms; the largest system in the database contains 29 atoms; the smallest molecule has only three atoms. PM6-D3H4X required a total of 5 h and 45 min to run all single-point energy calculations. This makes an average performance of 0.155 s per molecule. The GFN2-xTB calculations required slightly longer, namely 6 h, 8 min. This is a performance of 0.165 s per molecule. The difference in performance is attributed to the fact that GFN2-xTB required more iterations to converge than PM6.

In order to also benchmark the algorithms on larger systems, we performed single-point energy calculations on the PDB structures 1crn¹¹⁵ (640 atoms after adding protons) and 1ceb (1210 atoms after adding protons) available from the PDBbind database.^{116–120} For the calculation on 1crn, GFN2-xTB required 139 s, whereas the PM6-D3H4X calculation finished within 113 s. When used with pseudodiagonalization, the PM6 calculation finishes in 69 s. For 1ceb, the methods required respectively 2864 and 772 s (506 s with pseudodiagonalization). GFN2-xTB's poor performance is related to convergence issues (56 iterations with GFN2-xTB vs 14 iterations for PM6-D3H4X).

We also built three dimers of the protein 1crn (see Figure 2), in which the centers of mass are displaced by 30 Å over the *x* axis, 22 Å over the *y* axis, and 25 Å over the *z* axis. These distances were chosen so that there is no atomic overlap, and the

molecules are still close enough to interact. The orientations were chosen so that different types of interactions are exposed in the calculations. Other displacements/orientations could be easily built using internal functionalities of the library.

The performance of PM6-D3H4X was quite uniform, since the calculations finished within 790 and 794 s. Pseudodiagonalization further reduces the calculation time to 446–468 s, though we typically observe an increase in the number of iterations. On the other hand, GFN2-xTB calculations required about 1600 s.

■ GEOMETRY OPTIMIZATION

The geometry optimization module is centralized in the SolverOpt function, which takes as arguments a function like object and a solver. In our case, the former corresponds typically to a quantum chemical method, though other functions could be straightforwardly used.

Though in principle only one good solver is required for general use, experience shows us that there are always situations in which having more solvers available is advantageous. We have therefore a large library of unconstrained optimization methods available within *ULYSSES*. These range from gradient-based methods, steepest descent and conjugate gradient,^{121–123} to quasi-Newton. Our focus was mostly on the second class of methods, and we include Newton–Raphson, several variants of the BFGS method, Murtagh–Sargent,¹²⁴ Fletcher–Powell,¹²⁵ Shanno,¹²⁶ Barnes–Rosen,^{127,128} Bofill,¹²⁹ Bakken–Helgaker,¹³⁰ and the BFGS update itself,¹³¹ and the Baker¹³² and the Wales variants of Rational Function Optimization (RFO).^{133,134} By default, our implementation of Baker's method uses RFO for the optimization of minima if the Hessian shows the correct spectrum. PRFO is automatically switched on in saddle point optimization or if the Hessian's spectrum is incorrect. When suitable, several line searches are available, namely Davidson's method,¹³⁵ Barzilai–Borwein, Armijo, Moré–Thuente,^{136–139}

Fletcher,¹³¹ and the dogleg trust-region method.¹⁴⁰ Though the dogleg is a trust-region method, its use is made like a line search due to lack of specific implementation details.

Convergence is achieved whenever user-defined criteria/thresholds are satisfied for one of the following sets of options:

- Gradient and change in function/energy.
- Gradient and step.
- Change in function and step.
- All the above.

Our default is the first option. The convergence thresholds are set by default to 10^{-7} arbitrary units for all parameters. For quantum chemistry applications, we recommend using instead 10^{-5} E_h for energies and 10^{-3} E_h/Å for gradients. Tighter convergence criteria might be needed for more flexible systems. Details on modifications we introduced on some of the line-search algorithms and benchmark studies will be provided in another publication with the focus on structure generation and optimization.

Geometry optimization is invoked with five lines of code:

```
BakerRFO solve(4,0,0); //number 4 means BFGS Hessian update
//zeroes imply function minimization without following eigenvalues

Molecule mol("path/to/geom.xyz",-1,1,"C1");
BSet bs(mol,"am1");
AM1 nddo(bs,mol);
SolverOpt(nddo,solve,0,0,1e-6,1e-4);
```

In the code-snippet above, we use Baker's RFO with the BFGS Hessian update to optimize the equilibrium structure of a molecule of charge -1 using the AM1 method. Convergence criteria is set to 10^{-6} E_h for energy and 10^{-4} E_h/Å for gradients. The optimized geometry is going to be available from the quantum chemical method object, which may be used in further studies, printed or written to file.

IDEAL GAS PROPERTIES

The calculation of ideal gas properties takes place over one of three classes. The simplest, *PBIRRIHOE*, implements all the machinery required for calculating the ideal gas properties of a single structure. The other two classes are adequate for multistructure calculations, i.e., they provide different means for accounting conformational anharmonicity. All classes are built on top of *PBIRRIHOEbase*, which simply includes the required machinery for calculations. For convenience, the data required by the models is made available via two different modalities. On one side, one may provide the data structures directly and then perform calculations. Alternatively, all the required information may be stored and read from a JSON database. An example of one of such entries is given in the [Supporting Information](#). The strength of the *PBIRRIHOEbase* objects lies in the support from the JSON database. This has a double advantage for the user. On one hand, temperature-dependent conformational effects are easily applied when calculating thermodynamic properties. On the other hand, particular data may be stored in JSON-format for future (re)utilization. This may include calculated or even experimental information, which allows the user, e.g., to benchmark or validate the models or calculated data.

The calculation of thermodynamic properties of a single component from the database requires a minimum of 6–7 lines of code. Below we show the example of ideal gas properties for adenosine based on GFN2-xTB data. The properties are calculated at 300 K, and we assume a mol of adenosine is in standard conditions.

```
std::vector<std::string> method(3);
method[0] = "adenosine";
method[1] = "GFN2-xTB";
method[2] = "GFN2-xTB";
PBIRRIHOE Mol(300,0,false,NA,0.0224,0.0); //NA → 1 mol of molecules; 0.0224 → volume (m³)
Mol.ReadDB(method,0,1,"main");
double sMol = Mol.S();
```

In the example above, we calculate the entropy of the system. Other thermodynamic functions would require calls to the functions *H()*, *G()*, *A()*, *U()*, *CP()*, or *CV()*. The results come always in SI units. By means of loops, temperature profiles may be easily calculated, and allocation of several species allows the calculation of reaction-related properties.

PARTITION FUNCTIONS INCLUDED

The main model used for thermodynamic property calculation is the Particle-in-a-Box in the classical limit, Rigid Rotor in the classical limit, Harmonic Oscillator and Electron (*PBIRRIHOE*). The electronic partition function is just a simple summation of Boltzmann factors over declared electronic levels and their degeneracies.

It is well-known that the harmonic oscillator behaves poorly for internal motions of low-vibrational frequency. In such cases, it is possible to interpolate between the partition functions for the harmonic oscillator and the classical free-rotor, as suggested by Grimme.¹⁴¹ We apply the free-rotor correction consistently over all thermodynamic functions, not just entropies and Gibbs free energies. Though the correction barely affects enthalpies, we find it more appropriate to have all the thermodynamic properties consistently calculated from the same partition functions. This is for instance important for fine enthalpy–entropy compensation effects. To control the degree of mixing, a threshold frequency is used, ω_0 , which by default takes the value of 75 cm⁻¹ (cf. the original publication for more details).

The other classes aim at introducing conformational anharmonicity. The first one is *MSPBIRRIHOE*, which implements a multistructure extension of the *PBIRRIHOE* model. To some extent this is equivalent to the MS–HO method of the MS–Tor family¹⁴² but with the generalized vibrational partition function. The other class, *MSavgPBIRRIHOE*, applies a Boltzmann weighting to the thermodynamic properties. Conformational anharmonicity is then included via the Gibbs–Shannon terms.¹⁴³ Though there are small differences in heat capacities, entropies by the two methods are numerically equivalent.

Finally, if the systems of interest have one imaginary vibrational frequency, Transition State Theory (TST)^{144,145} is used to calculate the partition function. For convenience of the users, a set of functions is included to calculate rate constants based on TST, which might include simple tunneling corrections.¹⁴⁶

DYNAMICAL SIMULATIONS

ULYSSES is equipped with a dynamics class that allows Born–Oppenheimer Molecular Dynamics (BOMD)^{111,147–151} using the leapfrog or the velocity Verlet algorithms.¹⁵² In this variant of the molecular dynamics technique, electrons are treated quantum mechanically, whereas the nuclei are allowed to move classically along the Born–Oppenheimer potential surface.¹⁴⁹ Two types of ensembles are currently available, NVE for which no thermostating is required or NVT simulations using either the Berendsen thermostat¹⁵³ or the Maxwell–Boltzmann scaling technique of Andrea et al.¹⁵⁴

The SHAKE¹⁵⁵ and RATTLE algorithms¹⁵⁶ are available to constrain sets of bonds. These algorithms are typically very useful to keep reactions from happening during the equilibration stage. Metadynamics¹⁵⁷ is also available from the molecular dynamics class. Several molecular properties mentioned above (Available Molecular Properties) may also be followed along the trajectory. We currently do not have periodic boundary conditions available in the code.

Running a molecular dynamics simulation to follow a particular molecular property in time requires six lines of code. In the code snippet below, we set up and run the dynamics of a cation using GFN2-xTB, following the ionization potential. The simulation runs for 3 ps with steps of 1 fs. Every 10 fs the geometry is written to file. The output has a standard definition, which may be modified by the user according to their needs. The temperature of the system is set to 500 K.

```
Molecule mol("path/to/geom.xyz",1,1,"C1");
BSet bs(mol,"gfn2");
GFN2 tb(bs,mol);
Dynamics MDobj(mol,3,true,true,0.001,0.01);
MDobj.setPropertyToMeasure("IP");
MDobj.runMD(electron,500,0.2,0e-3,200);
Mol.ReadDB(method,0,1,"main");
```

CONCLUSIONS

We introduced the ULYSSES library, a highly efficient and robust header-only C++ library for semiempirical quantum chemical simulations of large systems. With this feature, ULYSSES may be easily coupled with any other C++ code, or one may also use its header-files to build template C++ programs with which quantum chemical calculations are available. The only requirement to use ULYSSES is a C++ compiler. The library is equipped with a vast choice of dispersion and hydrogen-bond corrected semiempirical methods, as well as many simple functions essential for quantum chemical simulation, namely geometry optimization of minima and transition states, ideal gas property calculation, and molecular dynamics. The good performance of the library is shown for several examples, either running on large libraries of compounds or on small proteins.

DATA AND SOFTWARE AVAILABILITY

ULYSSES 1.0 will be made available, free of charge, under a gitlab repository upon publication of this manuscript. The official link is <https://gitlab.com/siriuz/ulysses.git>. The software is available under the AGPL 3 license.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.2c00757>.

Choice of programming language and philosophy; description of central objects; elements available and respective references; and detailed descriptions of ideal gas property calculation (PDF)

Benchmark of implemented methods against well-established software (ZIP)

AUTHOR INFORMATION

Corresponding Authors

Filipe Menezes — Institute of Structural Biology, Helmholtz Zentrum Muenchen, 85764 Neuherberg, Germany;
orcid.org/0000-0002-7630-5447;

Email: filipe.menezes@helmholtz-muenchen.de

Grzegorz M. Popowicz — Institute of Structural Biology, Helmholtz Zentrum Muenchen, 85764 Neuherberg, Germany;
Email: grzegorz.popowicz@helmholtz-muenchen.de

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jcim.2c00757>

Funding

This work was funded by the Bundesministerium fuer Wirtschaft und Klimaschutz (BMWi) via ZIM, grant number KK 5197901TS0, as well as the Bundesministerium fuer Bildung und Forschung (BMBF), project SUPREME, number 031L0268.

Notes

The authors declare no competing financial interest.

REFERENCES

- (1) Pulay, P.; Saebo, S. Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory. *Theor. Chim. Acta* **1986**, *69*, 357.
- (2) Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, No. 034106.
- (3) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.
- (4) Ma, Q.; Schwilk, M.; Köppl, C.; Werner, H.-J. Scalable Electron Correlation Methods. 4. Parallel Explicitly Correlated Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD-F12). *J. Chem. Theory Comput.* **2017**, *13*, 4871–4896.
- (5) Menezes, F.; Kats, D.; Werner, H.-J. Local complete active space second-order perturbation theory using pair natural orbitals (PNO-CASPT2). *J. Chem. Phys.* **2016**, *145*, 124115.
- (6) Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, No. 1085.
- (7) Saebo, S.; Pulay, P. Fourth-order Møller–Plesset perturbation theory in the local correlation treatment. I. Method. *J. Chem. Phys.* **1987**, *86*, 914–922.
- (8) Saebo, S.; Tong, W.; Pulay, P. Efficient elimination of basis set superposition errors by the local correlation method: Accurate ab initio studies of the water dimer. *J. Chem. Phys.* **1993**, *98*, 2170–2175.
- (9) Elstner, M.; Frauenheim, Th.; Kaxiras, E.; Seifert, G.; Suhai, S. A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules. *Phys. Stat. Sol. (b)* **2000**, *217*, 357–376.
- (10) Giese, T. J.; York, D. M. Improvement of semiempirical response properties with charge-dependent response density. *J. Chem. Phys.* **2005**, *123*, 164108.
- (11) Bredow, T.; Jug, K. Theory and range of modern semiempirical molecular orbital methods. *Theor. Chem. Acc.* **2005**, *113*, 1–14.
- (12) Clark, T. J. *Mol. Struct. (Theorchem)* **2000**, *530*, 1–10.
- (13) Yilmazer, N. D.; Korth, M. Enhanced semiempirical QM methods for biomolecular interactions. *Comput. Struct. Biotechn. J.* **2015**, *13*, 169.
- (14) Stewart, J. J. P. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int. J. Quantum Chem.* **1996**, *58*, 133–146.
- (15) Merz, K. M. Using quantum mechanical approaches to study biological systems. *Acc. Chem. Res.* **2014**, *47*, 2804–2811.

- (16) Pople, J. A.; Santry, D. P.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *J. Chem. Phys.* **1965**, *43*, S129.
- (17) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899.
- (18) Dewar, M. J. S.; Thiel, W. A semiempirical model for the two-center repulsion integrals in the NDDO approximation. *Theor. Chim. Acta (Berl.)* **1977**, *46*, 89.
- (19) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
- (20) Thiel, W.; Voityuk, A. A. Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group. *J. Phys. Chem.* **1996**, *100*, 616–626.
- (21) Thiel, W.; Voityuk, A. A. Extension of the MNDO formalism to d orbitals: Integral approximations and preliminary numerical results. *Theor. Chim. Acta* **1992**, *81*, 391.
- (22) Dewar, M. J. S.; Rzepa, H. S. Ground states of molecules. 40. MNDO results for molecules containing fluorine. *J. Am. Chem. Soc.* **1978**, *100*, 58–67.
- (23) Dewar, M. J. S.; Reynolds, C. H. An improved set of MNDO parameters for sulfur. *J. Comput. Chem.* **1986**, *7*, 140.
- (24) Dewar, M. J. S.; McKee, M. L. Ground states of molecules. 41. MNDO results for molecules containing boron. *J. Am. Chem. Soc.* **1977**, *99*, 5231.
- (25) Dewar, M. J. S.; Rzepa, H. S. Ground states of molecules. 45. MNDO results for molecules containing beryllium. *J. Am. Chem. Soc.* **1978**, *100*, 777.
- (26) Dewar, M. J. S.; Rzepa, H. S. Ground states of molecules. 53. MNDO calculations for molecules containing chlorine. *J. Comput. Chem.* **1983**, *4*, 158.
- (27) Dewar, M. J. S.; Healy, E. Ground states of molecules. 64. MNDO Calculations for compounds containing bromine. *J. Comput. Chem.* **1983**, *4*, 542.
- (28) Dewar, M. J. S.; Healy, E.; Stewart, J. J. P. Ground States of Molecules. 67. MNDO Calculations for Compounds Containing Iodine. *J. Comput. Chem.* **1984**, *5*, 358.
- (29) Davis, L. P.; Guidry, R. M.; Williams, J. R.; Dewar, M. J. S.; Rzepa, H. S. MNDO calculations for compounds containing aluminum and boron. *J. Comput. Chem.* **1981**, *2*, 433.
- (30) Dewar, M. J. S.; Grady, G. L.; Healy, E. F. Ground states of molecules. 73. MNDO calculations for compounds containing germanium. *Organometallics* **1987**, *6*, 186.
- (31) Dewar, M. J. S.; Holloway, M. K.; Grady, G. L.; Stewart, J. J. P. Ground states of molecules. Part 79. MNDO calculations for compounds containing lead. *Organometallics* **1985**, *4*, 1973.
- (32) Dewar, M. J. S.; Grady, G. L.; Merz, K. M.; Stewart, J. J. P. MNDO calculations for compounds containing mercury. *Organometallics* **1985**, *4*, 1964.
- (33) Dewar, M. J. S.; Grady, G. L.; Stewart, J. J. P. MNDO Calculations for Compounds Containing Tin. *J. Am. Chem. Soc.* **1984**, *106*, 6771.
- (34) Dewar, M. J. S.; Merz, K. M. Ground states of molecules. Part 84. MNDO calculations for compounds containing zinc. *Organometallics* **1986**, *5*, 1494.
- (35) Dewar, M. J. S.; McKee, M. L.; Rzepa, H. S. MNDO parameters for third period elements. *J. Am. Chem. Soc.* **1978**, *100*, 3607.
- (36) Dewar, M. J. S.; Healy, E. F.; Stewart, J. J. P.; Friedheim, J.; Grady, G. Revised MNDO parameters for silicon. *Organometallics* **1986**, *5*, 375.
- (37) Dewar, M. J. S.; Jie, C. AM1 Calculations for Compounds Containing Germanium. *Organometallics* **1989**, *8*, 1544.
- (38) Dewar, M. J. S.; Jie, C. AM1 calculations for compounds containing mercury. *Organometallics* **1989**, *8*, 1547.
- (39) Dewar, M. J. S.; Jie, C. AM1 Calculations for Compounds Containing Silicon. *Organometallics* **1987**, *6*, 1486.
- (40) Dewar, M. J. S.; Holder, A. J. AM1 Parameters for Aluminum. *Organometallics* **1990**, *9*, 508.
- (41) Dewar, M. J. S.; Jie, C.; Zoebisch, E. G. AM1 calculations for compounds containing boron. *Organometallics* **1988**, *7*, 513.
- (42) Dewar, M. J. S.; Jie, C. AM1 parameters for phosphorus. *J. Mol. Struct. (Theochem)* **1989**, *187*, 1–13.
- (43) Dewar, M. J. S.; Yuan, Y.-C. AM1 parameters for Sulfur. *Inorg. Chem.* **1990**, *29*, 3881.
- (44) Dewar, M. J. S.; Merz, K. M. AM1 parameters for zinc. *Organometallics* **1988**, *7*, 522.
- (45) Dewar, M. J. S.; Zoebisch, E. G. Extension of AM1 to the Halogens. *J. Mol. Struct. (Theochem)* **1988**, *180*, 1.
- (46) Stewart, J. J. P. Optimization of Parameters for Semi-Empirical Methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209.
- (47) Stewart, J. J. P. Optimization of parameters for semiempirical methods. III Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi. *J. Comput. Chem.* **1991**, *12*, 320.
- (48) Anders, E.; Koch, R.; Freunsch, P. Optimization and application of lithium parameters for PM3. *J. Comput. Chem.* **1993**, *14*, 1301.
- (49) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations. *J. Mol. Model.* **2007**, *13*, 1173.
- (50) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1: a reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101.
- (51) Cook, D. B.; Hollis, P. C.; McWeeny, R. Approximate ab initio calculations on polyatomic molecules. *Mol. Phys.* **1967**, *13*, 553.
- (52) Brown, R. D.; Roby, K. R. Approximate Molecular Orbital Theory for Inorganic Molecules I. Analysis of Possible Integral Approximations. *Theor. Chim. Acta (Berl.)* **1970**, *16*, 175.
- (53) Chandrasekhar, J.; Mehrotra, P. R.; Subramanian, S.; Manoharan, P. T. NDDO MO calculations I. Analysis of the method. *Theor. Chim. Acta (Berl.)* **1976**, *41*, 243.
- (54) Tu, Y.; Laaksonen, A. Towards Fast and Reliable Quantum Chemical Modelling of Macromolecules. In *New Algorithms for Macromolecular Simulation*; Springer Berlin Heidelberg: 2006; pp 315–341, DOI: 10.1007/3-540-31618-3_17.
- (55) Kolb, M.; Thiel, W. Beyond the MNDO model: Methodical considerations and numerical results. *J. Comput. Chem.* **1993**, *14*, 775.
- (56) Weber, W.; Thiel, W. Orthogonalization corrections for semiempirical methods. *Theor. Chem. Acc.* **2000**, *103*, 495.
- (57) Dral, P. O.; Wu, X.; Spoerkel, L.; Koslowski, A.; Weber, W.; Steiger, R.; Scholten, M.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters. *J. Chem. Theory Comput.* **2016**, *12*, 1082.
- (58) Husch, T.; Vaucher, A. C.; Reiher, M. Semiempirical molecular orbital models based on the neglect of diatomic differential overlap approximation. *Int. J. Quantum Chem.* **2018**, *118*, No. e25799.
- (59) Seifert, G. Tight-Binding Density Functional Theory: An Approximate Kohn–Sham DFT Scheme. *J. Phys. Chem. A* **2007**, *111*, 5609.
- (60) Frauenheim, Th.; Seifert, G.; Elstner, M.; Hajnal, M. Z.; Jungnickel, G.; Porezag, D.; Suhai, S.; Scholz, R. A Self-Consistent Charge Density-Functional Based Tight-Binding Method for Predictive Materials Simulations in Physics, Chemistry and Biology. *Phys. Stat. Sol. (b)* **2000**, *217*, 41.
- (61) Elstner, M. SCC-DFTB: What Is the Proper Degree of Self-Consistency? *J. Phys. Chem. A* **2007**, *111*, 5614.
- (62) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931.
- (63) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method: Third-Order Expansion of the Density Functional Theory Total Energy and Introduction of a Modified Effective Coulomb Interaction. *J. Phys. Chem. A* **2007**, *111*, 10861.
- (64) Cui, Q.; Elstner, M. Density functional tight binding: values of semi-empirical methods in an ab initio era. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14368.

- (65) Seifert, G.; Joswig, J.-O. Density-functional tight binding—an approximate density-functional theory method. *WIREs Comput. Mol. Sci.* **2012**, 2, 456.
- (66) Giese, T. J.; York, D. M. Density-functional expansion methods: grand challenges. *Theor. Chem. Acc.* **2012**, 131, 1145.
- (67) Koskinen, P.; Mäkinen, V. Density-functional tight-binding for beginners. *Comput. Mater. Sci.* **2009**, 47, 237.
- (68) Avogadro: an open-source molecular builder and visualization tool, version 1.20. <https://avogadro.cc/> (accessed 2022-08-02).
- (69) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminf.* **2012**, 4, 17.
- (70) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators. *Protein Sci.* **2021**, 30, 70–82.
- (71) Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Meeting modern challenges in visualization. *Protein Sci.* **2018**, 27, 14–25.
- (72) Hunter, J.; Matplotlib, D. A 2D graphics environment. *Comput. Sci. & Eng.* **2007**, 9, 90.
- (73) Guennebaud, G.; Jacob, B.; Avery, P.; Bachrach, A.; Barthélemy, S.; Becker, C.; Benjamin, D.; Berger, C.; Berres, A.; Blanco, J. L.; Borgerding, M.; Bossart, R.; Brix, K.; Brun, G.; Büttgenbach, P.; Capricelli, T.; Carre, N.; Ceccato, J.; Chalupceky, V.; Chrétien, B.; Coles, A.; “complexzeros”, J.; Constantino, E.; Danoczy, M.; Dean, J.; Drenkhahn, G.; Ehrlicher, C.; Fernandes, M.; Ferro, D. G.; Garg, R.; Gautier, M.; Gladky, A.; Glaser, S.; Glisse, M.; Gosselin, F.; Grüninger, C.; Hamelin, P.; Hanwell, M. D.; Harmon, D.; He, C.-P.; Heibel, H.; Hertzberg, C.; Holoborodko, P.; Holy, T.; Irons, T.; de Jong, B.; Kim, K.; Klammler, M.; Köhler, C.; Korepanov, A.; Krivenko, I.; Kruisselbrink, M.; Kundu, A.; Lenz, M.; Larsen, R. M.; Li, B.; Lippner, S.; Lowenberg, D.; Luitz, D. J.; Maks, N.; Mantzaflaris, A.; Marcin, D. J.; Margaritis, K. A.; Martin, R.; Marxer, R.; Di Massa, V.; Mayer, C.; Meier-Dörnberg, F.; Mierle, K.; Montel, L.; Nerbonne, E.; Neundorff, A.; Newton, J.; Niesen, J.; Nuentsu, D.; Oberländer, J.; van den Oever, J.; Olbrich, M.; Pilgrim, S.; Piltz, B.; Piwowarski, B.; Ploskey, Z.; Po, G.; Popov, S.; Rajagopalan, M.; Rajko, S.; Repinc, J.; Riddle, K. F.; Roberts, R.; Rodriguez, A.; Román, P.; Ruepp, O.; Rusu, R. B.; Sanches, A. C.; Saupin, G.; Saut, O.; Schindler, B.; Schmidt, M.; Schridde, D.; Schwendner, J.; Seiler, C.; Senst, M.; Sheorey, S.; Somerville, A.; Stapleton, A.; Steiner, B.; Strothoff, S.; Swirski, L.; Szalkowski, A.; Traversaro, S.; Trojanek, P.; Truchet, A.; Tsouroukssidian, A. R.; Tellenbach, D.; Tyrer, J. R.; Ulerich, R.; de Valence, H.; Vanhassel, I.; Van Dyck, M.; Wheeler, S.; Witherden, F.; Wolfer, U.; Yguel, M.; Zoppitelli, P.; Adler, J. *Eigen v3*; 2010. eigen.tuxfamily.org (accessed 2022-08-02).
- (74) Yip, M. rapidjson.org (accessed 2022-08-02).
- (75) Neal, R. M. Fast exact summation using small and large superaccumulators. 2015, *arXiv*:<https://arxiv.org/abs/1505.05571> (accessed 2022-08-02).
- (76) Afshar, Y. github.com/yafshar/xsum.git (accessed 2022-08-02).
- (77) Brothers, E. N.; Merz, K. M. Sodium Parameters for AM1. *J. Phys. Chem. B* **2002**, 106, 2779–2785.
- (78) Thiel, W.; Voityuk, A. A. Extension of MNDO to d-Orbitals - Parameters and Results for the 2nd Row Elements and for the Zinc Group. *J. Chem. Phys.* **1996**, 100, 616.
- (79) Giese, T. J.; Sherer, E. C.; Cramer, C. J.; York, D. M. A Semiempirical Quantum Model for Hydrogen-Bonded Nucleic Acid Base Pairs. *J. Chem. Theory Comput.* **2005**, 1, 1275.
- (80) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. PDDG/PM3 and PDDG/MNDO: improved semiempirical methods. *J. Comput. Chem.* **2002**, 23, 1601.
- (81) Tubert-Brohman, I.; Guimarães, C. R. W.; Repasky, M. P.; Jorgensen, W. L. Extension of the PDDG/PM3 and PDDG/MNDO semiempirical molecular orbital methods to the halogens. *J. Comput. Chem.* **2004**, 25, 138.
- (82) Tubert-Brohman, I.; Guimarães, C. R. W.; Jorgensen, W. L. Extension of the PDDG/PM3 Semiempirical Molecular Orbital Method to Sulfur, Silicon, and Phosphorus. *J. Chem. Theory Comput.* **2005**, 1, 817.
- (83) 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. *J. Chem. Theory Comput.* **2019**, 15, 1652–1671.
- (84) Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-Cost Quantum Chemical Methods for Noncovalent Interactions. *J. Phys. Chem. Lett.* **2014**, 5, 4275–4284.
- (85) Hostaš, J.; Řezáč, J.; Hobza, P. On the performance of the semiempirical quantum mechanical PM6. *Chem. Phys. Lett.* **2013**, 568–569, 161–166.
- (86) Swart, M.; Solà, M.; Bickelhaupt, F. M. Inter- and intramolecular dispersion interactions. *J. Comput. Chem.* **2011**, 32, 1117–1127.
- (87) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, 132, 154104.
- (88) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding. *J. Chem. Theory Comput.* **2012**, 8, 141–151.
- (89) Řezáč, J.; Hobza, P. A halogen-bonding correction for the semiempirical PM6 method. *Chem. Phys. Lett.* **2011**, 506, 286–289.
- (90) Korth, M. Third-Generation Hydrogen-Bonding Corrections for Semiempirical QM Methods. *J. Chem. Theory Comput.* **2010**, 6, 3808–3816.
- (91) Kromann, J. C.; Christensen, A. S.; Steinmann, C.; Korth, M.; Jensen, J. H. A third-generation dispersion and third-generation hydrogen bonding corrected PM6 method: PM6-D3H+. *Peer J.* **2014**, 2, e449.
- (92) Pulay, P. Convergence acceleration of iterative sequences. The case of scf iteration. *Chem. Phys. Lett.* **1980**, 73, 393.
- (93) Pulay, P. Improved SCF convergence acceleration. *J. Comput. Chem.* **1982**, 3, 556.
- (94) Cancès, E.; Le Bris, C. Can we outperform the DIIS approach for electronic structure calculations? *Int. J. Quantum Chem.* **2000**, 79, 82.
- (95) Dixon, S. L.; Merz, K. M. Fast, accurate semiempirical molecular orbital calculations for macromolecules. *J. Chem. Phys.* **1997**, 107, 879–893.
- (96) Mitin, A. V. The dynamic “level shift” method for improving the convergence of the SCF procedure. *J. Comput. Chem.* **1988**, 9, 107–110.
- (97) Stewart, J. J. P.; Császár, P.; Pulay, P. Fast semiempirical calculations. *J. Comput. Chem.* **1982**, 3, 227–228.
- (98) Badziag, P.; Solms, F. An improved SCF iteration scheme. *Comp. & Chem.* **1988**, 12, 233–236.
- (99) Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. *J. Chem. Theory Comput.* **2021**, 17, 4250.
- (100) Mayer, I. Charge, bond order and valence in the AB initio SCF theory. *Chem. Phys. Lett.* **1983**, 97, 270–274.
- (101) Armstrong, D. R.; Perkins, P. G.; Stewart, J. J. P. Bond indices and valency. *J. Chem. Soc., Dalton Trans.* **1973**, 8, 838–840.
- (102) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class IV charge models: A new semiempirical approach in quantum chemistry. *J. Comp.-Aid. Mol. Des.* **1995**, 9, 87.
- (103) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. New Class IV Charge Model for Extracting Accurate Partial Charges from Wave Functions. *J. Phys. Chem. A* **1998**, 102, 1820–1831.
- (104) Winget, P.; Thompson, J. D.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. Charge Model 3: A Class IV Charge Model Based on Hybrid Density Functional Theory with Variable Exchange. *J. Phys. Chem. A* **2002**, 106, 10707–10717.
- (105) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. Parameterization of charge model 3 for AM1, PM3, BLYP, and B3LYP. *J. Comput. Chem.* **2003**, 24, 1291–1304.
- (106) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the

Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.

(107) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44*, 129–138.

(108) Yang, W.; Mortier, W. J. The use of global and local molecular parameters for the analysis of the gas-phase basicity of amines. *J. Am. Chem. Soc.* **1986**, *108*, 5708–5711.

(109) Jensen, F. *Introduction to Computational Chemistry*; Wiley: 1998.

(110) Nguyen, L. T.; Le, T. N.; De Proft, F.; Chandra, A. K.; Langenaeker, W.; Nguyen, M. T.; Geerlings, P. Mechanism of [2 + 1] Cycloadditions of Hydrogen Isocyanide to Alkynes: Molecular Orbital. *J. Am. Chem. Soc.* **1999**, *121*, 5992–6001.

(111) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* **2020**, *11*, e01493.

(112) Field, M. J. The pDynamo Program for Molecular Simulations using Hybrid Quantum Chemical and Molecular Mechanical Potentials. *J. Chem. Theory Comput.* **2008**, *4*, 1151.

(113) Ruddigkeit, L.; van Deursen, E.; Blum, L. C.; Raymond, J.-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864.

(114) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 140022.

(115) Teeter, M. M. Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. *Proc. Nat. Acad. Sci.* **1984**, *81*, 6014–6018.

(116) Liu, Z.; Li, Y.; Han, L.; Li, J.; Liu, J.; Zhao, Z.; Nie, W.; Liu, Y.; Wang, R. PDB-wide collection of binding data: current status of the PDBbind database. *J. Bioinform.* **2015**, *31*, 405–412.

(117) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein-Lig. *J. Med. Chem.* **2004**, *47*, 2977–2980.

(118) Li, Y.; Liu, Z.; Li, J.; Han, L.; Liu, J.; Zhao, Z.; Wang, R. Comparative Assessment of Scoring Functions on an Updated Benchmark: 1. Compilation of the Test Set. *J. Chem. Inf. Model.* **2014**, *54*, 1700–1716.

(119) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative Assessment of Scoring Functions on a Diverse Test Set. *J. Chem. Inf. Model.* **2009**, *49*, 1079–1093.

(120) Liu, Z.; Su, M.; Han, L.; Liu, J.; Yang, Q.; Li, Y.; Wang, R. Forging the Basis for Developing Protein-Ligand Interaction Scoring Functions. *Acc. Chem. Res.* **2017**, *50*, 302–309.

(121) Fletcher, R.; Reeves, C. M. Function minimization by conjugate gradients. *Comput. J.* **1964**, *7*, No. 149.

(122) Polak, E.; Ribière, G. Note sur la convergence de méthodes de directions conjuguées. *Rev. Franç. Inf. Rech. Opér.* **1969**, *3*, 35.

(123) Gilbert, J. C.; Nocedal, J. Global Convergence Properties of Conjugate Gradient Methods for Optimization. *SIAM J. Opt.* **1992**, *2*, 21.

(124) Murtagh, B. A.; Sargent, R. W. H. Computational experience with quadratically convergent minimisation methods. *Comp. J.* **1970**, *13*, No. 185.

(125) Fletcher, R.; Powell, M. J. D. A Rapidly Convergent Descent Method for Minimization. *Comp. J.* **1963**, *6*, No. 163.

(126) Shanno, D. F. Conditioning of quasi-Newton methods for function minimization. *Math. Comp.* **1970**, *24*, 647.

(127) Barnes, J. G. P. An Algorithm for Solving Non-Linear Equations Based on the Secant Method. *Comp. J.* **1965**, *8*, No. 66.

(128) Rosen, E. M. A review of quasi-Newton methods in nonlinear equation solving and unconstrained optimization. *Nat. Conf. ACM* **1966**, *37*, 37.

(129) Bofill, J. M. Updated Hessian matrix and the restricted step method for locating transition structures. *J. Comput. Chem.* **1994**, *15*, 1.

(130) Bakken, V.; Helgaker, T. The efficient optimization of molecular geometries using redundant internal coordinates. *J. Chem. Phys.* **2002**, *117*, 9160.

(131) Fletcher, R. *Practical Methods of Optimization: Unconstrained Optimization*; Wiley: 1980; Vol. 1.

(132) Baker, J. An algorithm for the location of transition states. *J. Comput. Chem.* **1986**, *7*, 385–395.

(133) Wales, D. J. Theoretical study of some small van der Waals complexes containing inert gas atoms. *Mol. Phys.* **1991**, *74*, 1.

(134) Wales, D. J. Rearrangements of 55-atom Lennard-Jones and (C60)55 clusters. *J. Chem. Phys.* **1994**, *101*, 3750.

(135) Davidon, W. C. Variable Metric Method for Minimization. *Comp. J.* **1968**, *10*, 406.

(136) Moré, J. J.; Thuente, D. J. Line Search Algorithms with Guaranteed Sufficient Decrease. *Trans. Math. Soft.* **1994**, *20*, 286.

(137) Melville, J. github.com/jlmelville/mize (accessed 2022-08-02).

(138) O'Leary, D. P. www.cs.umd.edu/users/oleary/software/ (accessed 2022-08-02).

(139) Moré, J. J.; Garbow, B. S.; Hillstom, K. E. *User guide for MINPACK-1*; ANL-80-74; 1980; DOI: 10.2172/6997568.

(140) Kelley, C. T. *Iterative Methods for Optimization*; Society for Industrial and Applied Mathematics: 1999; DOI: 10.1137/1.9781611970920.

(141) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem.—Eur. J.* **2012**, *18*, 9955–9964.

(142) Zheng, J.; Mielke, S. L.; Clarkson, K. L.; Truhlar, D. G. MSTor: A program for calculating partition functions, free energies, enthalpies, entropies, and heat capacities of complex molecules including torsional anharmonicity. *Comput. Phys. Commun.* **2012**, *183*, 1803.

(143) Suárez, D.; Díaz, N. Direct methods for computing single-molecule entropies from molecular simulations. *WIREs Comp. Mol. Sci.* **2015**, *5*, 1–26.

(144) Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 107.

(145) Wynne-Jones, W. F. K.; Eyring, H. The Absolute Rate of Reactions in Condensed Phases. *J. Chem. Phys.* **1935**, *3*, 492.

(146) Garrett, B. C.; Truhlar, D. G. Semiclassical tunneling calculations. *J. Phys. Chem.* **1979**, *83*, 2921.

(147) Leforestier, C. Classical trajectories using the full ab initio potential energy surface $H^- + CH_4 \rightarrow CH_3 + H^-$. *J. Chem. Phys.* **1978**, *68*, 4406.

(148) Warshel, A.; Karplus, M. Semiclassical trajectory approach to photoisomerization. *Chem. Phys. Lett.* **1975**, *32*, 11.

(149) Marion, A.; Gokcan, H.; Monard, G. Semi-Empirical Born-Oppenheimer Molecular Dynamics (SEBOMD) within the Amber Biomolecular Package. *J. Chem. Inf. Model.* **2019**, *59*, 206.

(150) Margraf, J. T.; Hennemann, M.; Clark, T. EMPIRE: a highly parallel semiempirical molecular orbital program: 3: Born-Oppenheimer molecular dynamics. *J. Mol. Model.* **2020**, *26*, 43.

(151) Kühne, T. D.; Iannuzzi, M.; Ben, M. D.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borštnik, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Müller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Glöck, A.; Lass, M.; Bethune, I.; Mundy, C. J.; Plessl, C.; Watkins, M.; Vondele, J. V.; Krack, M.; Hutter, J. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.

(152) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: 2017; DOI: 10.1093/oso/9780198803195.001.0001.

(153) Berendsen, H. J. C.; Postma, J. P. M.; van Gusteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684.

(154) Andrea, T.; Swope, W. C.; Andersen, H. C. The role of long ranged forces in determining the structure and properties of liquid water. *J. Chem. Phys.* **1983**, *79*, 4576.

(155) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327.

(156) Andersen, H. C. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.

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