Accurate, Automated Density Functional Theory for Complex Molecules Using On-the-fly Error Correction

Samuel M. Blau[†*], Evan Walter Clark Spotte-Smith[‡§*], Brandon M. Wood[¶], Shyam Dwaraknath[†], and Kristin A. Persson[‡#§]

[†] *Energy Storage and Distributed Resources, Lawrence Berkeley National Laboratory*
[‡] *Materials Science Division, Lawrence Berkeley National Laborory*
[§] *Department of Materials Science and Engineering, University of California Berkeley*
[¶] *National Energy Research Supercomputing Center, Lawrence Berkeley National Laboratory*
[#] *Molecular Foundry, Lawrence Berkeley National Laboratory*
[*] *These authors contributed equally to this work*

Abstract

High-throughput density functional theory (DFT) has been widely utilized to study a variety of materials and molecular properties. However, its application to complex molecular systems, including those relevant to electrochemical reactivity and decomposition, has been limited by insufficient automation. Here, we report a broadly applicable, automated framework for the accurate and robust DFT calculation of molecules, capable of addressing species relevant to electrochemistry. This framework is specifically designed to study molecules with different charge states, open-shell electronic structure, metal coordination, and implicit solvation. We first identify appropriate levels of theory that avoid calculation failures and accurately predict molecular redox potentials. We then describe our framework, including methods to automatically detect and correct errors and to optimize structures from saddle points to potential energy surface minima. To demonstrate the efficacy of this framework, we examine a case study including over 12,000 calculations of reactive molecular fragments. This framework is able to reduce the rate of failure for DFT calculations from 25.1% to 1.2%, significantly improving the degree of automation possible for high-throughput molecular DFT.

Keywords: Density functional theory, high-throughput, error correction, electrochemistry, computational thermochemistry

Introduction

With advancements in available computing power, *ab initio* calculations based on density functional theory (DFT) have become potent tools for the prediction of materials and molecular properties. When used in a high-throughput manner, DFT can generate massive data sets[1,2] and serve to populate databases.[3–10] By leveraging these repositories or other high-throughput calculations, researchers have been able to screen compounds for a broad range of applications, including photovoltaics,[11] metal-ion battery electrolytes[12–14] and electrodes,[15–18] redox-flow batteries,[19] catalysts,[20,21] thermoelectrics,[22] pharmaceuticals,[23–26] and many more. High-throughput DFT has also recently been used to construct computational reaction networks that describe the thermodynamics and kinetics of chemical reactions between large numbers of chemical species.[27,28]

In order to facilitate high-throughput computational chemistry, a number of tools[3,29–34] have been developed to automate DFT calculations, generating input files, interfacing with quantum chemistry software such as the Vienna Ab Initio Simulation Program (VASP),[35] Quantum Espresso,[36] or Q-Chem,[37] and both parsing and analyzing data with little to no human intervention. A key feature of such packages is the ability to autonomously detect and correct errors (issues that cause the DFT code to fail) and problems (issues that do not cause the code to fail but might lead to incorrect results) that may be encountered during calculations.

Numerous studies have explored neutral, closed-shell organic molecules with great success using high-throughput methods. However, certain other types of species remain difficult to simulate in an automated and high-throughput fashion due to their inherent chemical complexity. Specifically, charged molecules, open-shell molecules, molecules coordinated with metal ions, and molecules in solvent can be notoriously difficult to analyze using DFT:

- *Charged molecules:* Without being stabilized (for instance, by a counter-ion or solvent environment), charged molecules can be difficult to optimize to stationary points either during self-consistent field (SCF) calculations or during geometry optimization. This difficulty is especially pronounced for anions due to the expanded electron cloud, which necessitates a large basis set for accurate calculation of the electron density.

- *Open-shell molecules:* Radicals are often unstable and reactive. This instability can make geometry optimization especially challenging, as minima may be extremely shallow and the radical may have a tendency to dissociate or significantly change its bonding during optimization.

- *Metal-coordinated molecules:* The introduction of metals into molecular DFT calculations leads to a number of complications. Metals usually require coordination to be stabilized, necessitating the addition of solvent or ligand species. Because metal coordination occurs on a length scale longer than conventional molecular bonds, it cannot be accurately captured without a sufficiently advanced level of theory (LOT). Additionally, metal chemistry is strongly dependent on oxidation state; in cases where several oxidation states are possible, care must be taken to ensure that the SCF converges to a state where the metal has an appropriate charge.

- *Solvated molecules:* The introduction of solvent generally increases computational complexity. Either the solvent must be explicitly considered, increasing the number of electrons and the dimensionality of the potential energy surface (PES), or an implicit solvent model must be used, which adds additional convergence parameters (converging the solvent cavity) and additional terms to the molecular energy and gradient. In either case, optimization can become both more computationally costly and more difficult.

An example molecule for each category is depicted in Figure 1, along with some possible errors that might arise during calculation. The risk of calculation failure may increase for molecules exhibiting multiple of these challenging traits - for instance, a charged, radical, metal-coordinated molecule; errors may also be more likely in a complex workflow involving many calculations, necessitating robust error handling.

It is important to note that the diverse challenges identified above cannot be easily addressed by a single solution. However, these types of molecules are important in many applications; ionic species and coordinated metals are highly relevant in a range of biological applications,[38–40] radicals play a key role in combustion[41,42] and catalysis,[43] and solvation is essential for many synthetic and reactive processes.[44–46] Further, all of these classes of species are critical in electrochemical systems, particularly Li-ion and multivalent-ion electrolytes as well as catholytes and anolytes in redox-flow batteries. Complex electrochemical reaction cascades can trigger electrolyte decomposition, potentially hampering battery performance[47] or alternatively forming a protective passivating layer in the form of a solid electrolyte interphase (SEI).[48] In this work, we focus on electrochemically relevant molecules, though we emphasize that much of what we discuss is broadly applicable.

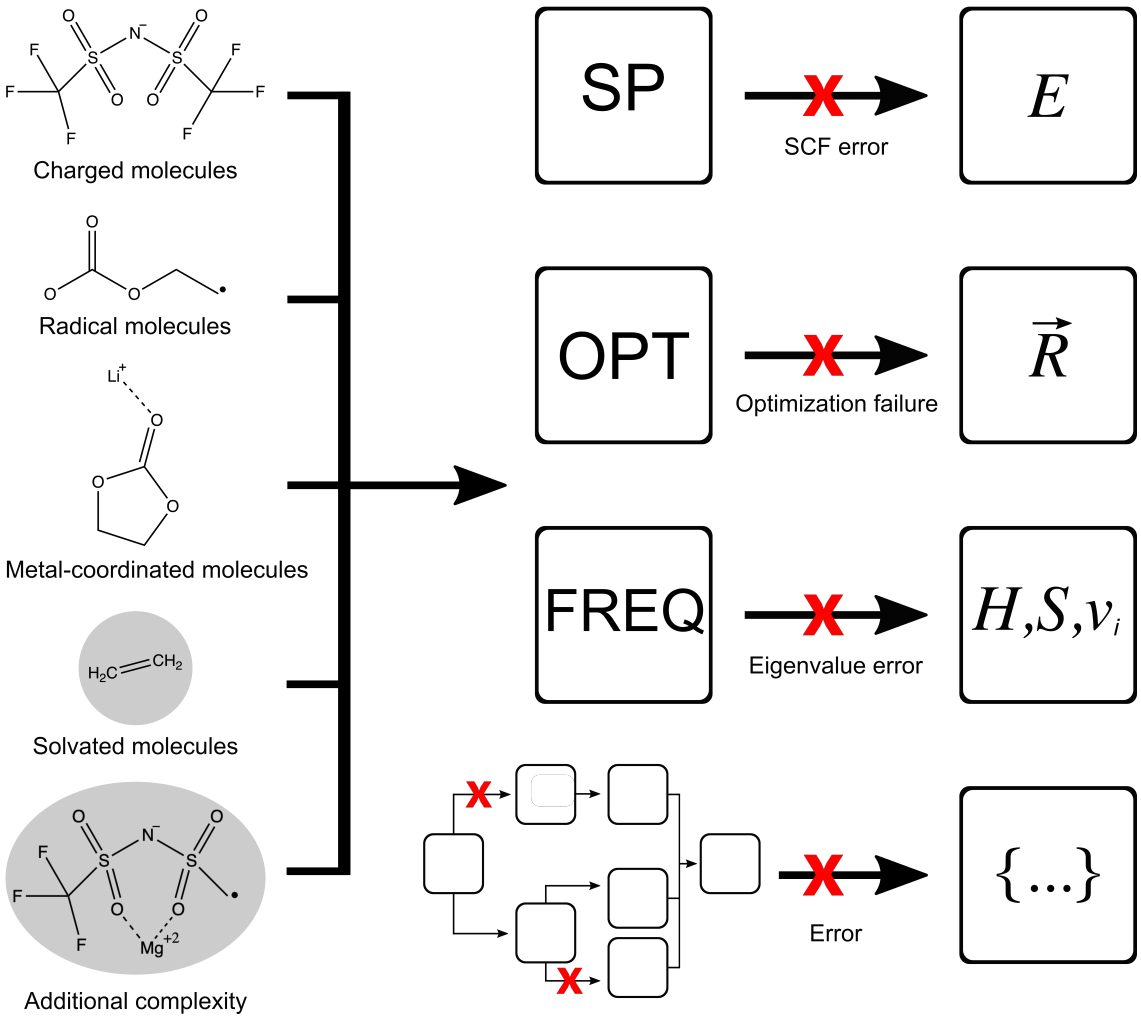


Figure 1: A schematic depicting the importance of error correction. Performing common calculations such as a single-point self-consistent field (SCF) calculation (SP) to obtain the energy E, a geometry optimization (OPT) to obtain the atomic coordinates $\vec{R}$, or a vibrational frequency analysis (FREQ) to obtain the enthalpy H, entropy S, and frequencies ν_i, will frequently result in errors of varying nature. This may be especially true when charged, radical, metal-coordinated, or solvated molecules are involved. In calculations involving several of these challenging features, or in complex workflows involving multiple calculations (bottom), the risk for errors can be compounded.

In this paper, we describe a set of computational methods that allow for the robust and accurate calculation of molecular geometries and thermochemistry using DFT in a fully automated fashion. Through a benchmark study, we identify LOTs appropriate for use in high-throughput studies of electrochemical systems, including density functional, basis set, and implicit solvent model. These appropriate LOTs are found to accurately predict reduction and oxidation potentials without failure. We present automatic schemes to correct common errors and problems encountered during molecular DFT calculations and describe an open source implementation of these methods. We further demonstrate the efficacy of our implementation through analysis of a set of over 12,000 DFT calculations.

Computational Details

All calculations discussed in this article were performed using version 5.2.2 of the Q-Chem electronic structure code.[37] A large quadrature grid (SG-3) was used for all calculations.[49] Unless otherwise noted, with this exception, all Q-Chem default values (as of the 5.2.2 version) were used for initial calculations, though during

error-correction these default values might be changed. The modification of DFT parameters during error correction will be discussed below.

Level of theory

The LOT for a quantum chemistry calculation defines the methods used to approximate the solution to the many-body Schrödinger equation. For Kohn-Sham DFT calculations, the most important components of the LOT are the density functional and the basis set. The density functional, which depends generally on the spatial electron density and its derivatives, defines the energy of a system of interacting electrons. The basis set is a collection of spatial functions which are used to model the exact many-body electronic wave function. Here, we consider only contracted Gaussian basis sets,[50] which are most commonly used in molecular quantum chemistry.

In addition to the density functional and the basis set, the choice of implicit solvent model is important for approximating the effect of bulk solvent. Implicit solvent models typically approximate bulk solvent as a continuous dielectric medium. The interaction between this medium and the electron density of the molecule of interest approximates the electrostatic interaction between the solvent and the solute, allowing for the calculation of properties including the solvation free energy. While self-consistent iterative procedures involving solvent cavities are often referred to as self-consistent reaction field (SCRF) calculations, we will here use the term SCF to describe such calculations both with and without implicit solvent. More extensive descriptions of the components of the LOT are provided in the Supporting Information.

DFT calculations rely on numerous parameters that can be modified by users. These include convergence parameters for SCF calculations and geometry optimizations, numerical parameters (including the quadrature grid, already mentioned), algorithmic parameters, and others. While these parameters can have significant impacts on the results of DFT calculations - including the accuracy of their results, their computational cost, and their likelihood to fail - they do not affect the underlying physical theory being used to determine the electronic structure. As such, we do not here consider them in our evaluation of LOTs. We do, however, vary some of these parameters in our error correction procedures, described below.

The incredible diversity of density functionals, basis sets, and implicit solvent models available to researchers makes the task of selecting a LOT for a particular application challenging. Here we provide a small combinatorial benchmark that compares a select set of LOTs on three metrics particularly important for high-throughput studies: tendency to fail due to errors or other problems; accuracy (comparing to experimental oxidation and reduction potentials); and computational cost. We use this benchmark to identify LOTs that are generally appropriate for calculations involving complex molecular species in electrochemical applications. Note that we make no claims that the LOTs identified in this benchmark are the best for this application or for any other. There may be functionals, basis sets, and implicit solvent models, or combinations thereof, that are superior in certain dimensions to the top performers of this benchmark. We simply sought to identify a set of LOTs that are adequate for this particular use case. We note that this method of evaluating based on cost and tendency to fail in addition to accuracy is especially useful for benchmarks aiming to identify LOTs for high-throughput applications, where both the computational resources to conduct the calculations and the human resources to analyze failed calculations can be strained.

Table 1: Density functionals, basis sets, and implicit solvent models for benchmark study

Density functionals	**Basis sets**	**Implicit solvent models**	**Solvents**
B3LYP	def2-SVP	COSMO	Tetrahydrofuran (THF)
B3LYP-D3	def2-SVPD	SMD	Water
CAM-B3LYP	def2-TZVPD	Vacuum	
ωB97X-V	def2-TZVPPD	Vacuum + COSMO	
M06-2X-D3		Vacuum + SMD	

The candidate functionals, basis sets, implicit solvent models, and solvents for this benchmark are given in Table 1. These candidates were chosen to represent broader classes of possible choices available to researchers.

Each of the functionals chosen are so-called hybrid functionals, using the exact exchange term from Hartree-Fock theory to correct for electron self-interaction error;[51] among these five functionals are a global

hybrid generalized gradient approximation (GGA) functional (B3LYP),[52,53] a global hybrid GGA functional with dispersion correction (B3LYP-D3), a range-separated hybrid GGA functional (CAM-B3LYP),[54] a range-separated hybrid GGA functional with dispersion correction (ωB97X-V),[55] and a global hybrid meta-globalized gradient approximation (meta-GGA) functional with dispersion correction (M06-2X-D3).[56] Both B3LYP-D3 and M06-2X-D3 use Grimme's D3 dispersion correction,[57] while ωB97X-V uses the VV10 nonlocal correlation functional[58] to account for dispersion.

All basis sets are from the "def2" family developed by Alrichs and coworkers[59] and augmented with diffuse functions by Rappoport and Furche.[60] This family was chosen because its members have similar accuracy across most of the periodic table,[61] making them appropriate for use in a wide range of chemical systems. The specific basis sets chosen (def2-SVP, def2-SVPD, def2-TZVPD, def2-TZVPPD) represent a gradual increase in size through addition of core and valence functions, polarization functions, and diffuse functions.

Two implicit solvent models were selected: COSMO,[62] a member of the PCM family of methods,[63] and SMD,[64] a member of the SMx family of methods and a representative method that includes non-electrostatic energy contributions; all calculations were also conducted in vacuum. In molecular DFT studies, solvent effects are often treated through corrections, wherein geometry optimization and vibrational analysis will take place in vacuum, and the energy will then be corrected by a single-point calculation in solvent. To determine if this method would lead to accurate prediction of redox potentials, we performed single-point calculations on all structures optimized in vacuum, using both COSMO and SMD.

Because the efficacy of implicit solvent models can depend on the solvent of interest, we selected two solvents with very different electrostatic behavior: tetrahydrofuran ($\varepsilon = 7.43$) and water ($\varepsilon = 78.36$). Parameters for these solvents were taken from the Minnesota Solvent Descriptor Database.[65] It should be noted that redox potentials can change based on the solvent used, and so this introduces a possible source of error.

Six molecules relevant to various electrochemical applications (metal-ion battery electrolyte solvents, redox shuttles, and redox-flow catholytes) were selected to form the benchmark set. Three of these molecules - anisole, (2,2,6,6-tetramethylpiperidin-1-yl)oxidanyl (TEMPO), and 10-methylphenothiazine - were chosen to test the ability of LOTs to calculate oxidation potentials, while the other three - lithium ethylene carbonate (LiEC), lithium fluoroethylene carbonate (LiFEC), and lithium ethylene sulfite (LiES) - were chosen to test the ability of LOTs to calculate reduction potentials. These molecules, along with their relevant redox potentials (either oxidation potentials or reduction potentials) can be seen in Table 2.

For this benchmark, redox potentials are calculated with DFT as

$$E(Li/Li^+) = -(G_{reduced} - G_{non-reduced})/F - 1.4\ V \tag{1}$$

for reduction and

$$E(Li/Li^+) = (G_{oxidized} - G_{non-oxidized})/F - 1.4\ V \tag{2}$$

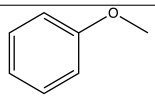
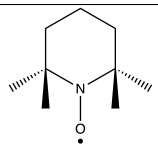
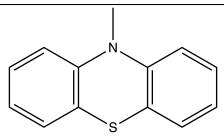
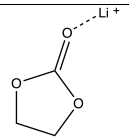
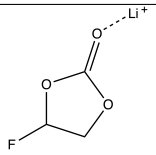
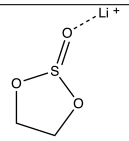
for oxidation, where F is the Faraday constant. The scaling factor of $\sim -1.4\ V$, derived from the difference between the energy of an electron in vacuum ($-4.44\ eV$) and the absolute potential of a Li electrode ($-3.05\ eV$),[66] converts the calculated redox potential to be relative to a Li/Li^+ reference electrode, as is commonly done in studies concerning lithium-ion batteries and related electrochemistry.

The computed redox potential values were compared against experimentally measured values for the test set molecules (Table 2). Each of the 100 possible combinations of the LOTs - including density functional, basis set, and implicit solvent model - were considered. For each LOT (and, where applicable, each combination of LOT and solvent), geometry optimization and vibrational frequency calculations were performed for each molecule in the test set in two charge states; note that this procedure was only performed once in vacuum, although three of the five solvent models use these vacuum-optimized geometries. For the LOTs including "Vacuum + COSMO" or "Vacuum + SMD" implicit solvent methods, a single-point calculation was performed on the final optimized vacuum geometry in solvent in order to correct the final energy. All calculations used the Frequency-Flattening Optimization procedure, which we describe below. We note that the experimental redox potentials for the test set molecules were measured in a variety of solvents. In an attempt to minimize the effect of the solvent on our results, when reporting errors, we report the average of the error for the calculations in water and THF.

In general, for each molecule, a single initial guess geometry was provided and used for both charge states. This was not possible in the case of LiEC and LiFEC. For these two molecules, the carbonate carbon

is planar in the non-reduced (+1) state and nonplanar in the reduced (0) state. In almost all LOTs tested, when a planar geometry was provided for the reduced calculations involving LiEC and LiFEC, the resulting optimized geometry incorrectly remained planar. In some cases, this occurred because of an extra electron residing on the lithium atom, as determined by the Mulliken population analysis.[67] These planar geometries generally resulted in poor estimations of the reduction potential. For these two molecules, to remedy this idiosyncratic issue, slightly non-planar guess geometries were provided for the reduced calculations, which resulted in lower energy non-planar configurations with appropriate Mulliken charges on the lithium atom and better estimates of the reduction potential.

Table 2: Molecules used in the level of theory benchmark, along with their redox potentials vs. Li/Li^+.

Number	**Molecule name**	**Structure**	**Oxidation potential (vs. Li/Li$^+$)**	**Oxidation states**
1	Anisole		4.20[68]	0, +1
2	TEMPO		3.5[68]	0, +1
3	10-methylphenothiazine		3.47[68]	0, +1
Number	**Molecule name**	**Structure**	**Reduction potential (vs. Li/Li$^+$)**	**Oxidation states**
4	LiEC		0.75[69]	0, +1
5	LiFEC		1.15[69]	0, +1
6	LiES		2.1[70]	0, +1

During the course of the benchmark study, eight calculations failed due to various errors, and two failed to optimize to a stationary point. 36 calculations successfully optimized to a stationary point but could not optimize to a PES minimum (this issue is discussed more below). All LOTs involved in these 46 failed calculations, 18 in total, have been excluded from further analysis (see Table S2.1 in the Supporting Information for a list of these LOTs). Given that one of our main goals is to identify a LOT that is not likely to fail given complex and potentially reactive molecules, no LOT that fails during a relatively small and simple benchmark can be considered appropriate for our use case.

We ranked the remaining 82 LOTs in terms of their MAE compared to experimental values. The 10 LOTs with lowest mean absolute error (MAE) are listed in Table 3; a complete list is provided in Table S2.2 in the Supporting Information. Table 3 also includes relative cost with respect to the least expensive LOT in the table (B3LYP-D3/def2-TZVPD/COSMO).

Table 3: Benchmark study errors, mean absolute error (MAE), and average costs.

Functional	Basis Set	Solvent Method	1	2	3	4	5	6	MAE	Rel. Cost
B3LYP	def2-TZVPD	SMD	0.413	0.211	0.080	0.152	-0.007	0.306	0.219	7.65
B3LYP	def2-TZVPPD	SMD	0.412	0.221	0.078	0.152	-0.005	0.303	0.220	12.37
B3LYP	def2-TZVPD	Vacuum + SMD	0.402	0.200	0.070	0.051	-0.002	-0.505	0.255	1.09
B3LYP-D3	def2-TZVPD	Vacuum + SMD	0.405	0.198	0.075	0.049	0.001	-0.520	0.259	1.18
B3LYP	def2-SVPD	SMD	0.446	0.304	0.137	0.200	0.099	0.371	0.263	2.16
ωB97X-V	def2-TZVPPD	SMD	0.465	0.241	0.300	0.023	-0.131	0.384	0.274	14.34
ωB97X-V	def2-TZVPD	SMD	0.466	0.241	0.302	0.024	-0.131	0.385	0.275	8.48
B3LYP-D3	def2-TZVPD	COSMO	0.580	0.425	0.182	-0.092	-0.195	0.132	0.278	1.00
B3LYP-D3	def2-TZVPPD	COSMO	0.597	0.428	0.213	-0.090	-0.195	0.130	0.285	1.62
M06-2X-D3	def2-SVPD	SMD	0.643	0.301	0.373	0.159	-0.041	-0.045	0.288	2.57

Table 3: Redox potential errors and overall mean absolute error (MAE) in V for the ten selected LOTs in the benchmark study, along with the average cost relative to the least computationally expensive LOT listed. LOTs are ordered in terms of their MAE.

Even given the limited size of the benchmark, definite trends emerge regarding the suitability of different components of the LOT. It is clear that basis sets augmented by diffuse functions are essential for accurate prediction of redox potentials (10 of the top 10 LOTs use basis sets augmented with diffuse functions), and that triple-zeta basis sets lead to better results than split-valence basis sets (8 of the top 10 use basis sets of triple-zeta quality). It also appears that the SMD model often outperforms COSMO in this domain, with 8 of the top 10 LOTs using SMD either for all calculations or as a single-point correction. The use of an some implicit solvent model is critical for calculations involving molecules in solution; the 13 worst-performing LOTs all performed calculations in vacuum with no implicit solvent correction (see the Supporting Information for details). The effect of the density functional is less clear. Only 6 of the 10 top-performing LOTs use dispersion corrections, and none of the top 3 do. Very few (2 out of the top 10) use density functionals that are range-separated; the majority use global hybrid functionals. This is not to say that attributes like dispersion correction and range separation are not beneficial, but that they are perhaps less essential for the accurate prediction of redox potentials, provided that an appropriate basis set and solvent method are used.

Based on our benchmark, any of the LOTs listed in Table 3 should be considered adequate for the calculation of redox potentials of electrochemically relevant molecules. In this work, we have chosen to use ωB97X-V/def2-TZVPPD/SMD; while several LOTs using B3LYP and its derivatives outperform ωB97X-V/def2-TZVPPD/SMD somewhat in our relatively small benchmark, a previous study with a broader benchmark found that the ωB97X-V functional is better able to predict a wide range of thermochemical as well as kinetic properties when compared to B3LYP.[71] In our experience, this functional also tends to avoid saddle points comparatively well, particularly for complex reactive fragments, which can potentially reduce the likelihood of calculation failure and reduce computational time by avoiding additional calculations to eliminate imaginary frequencies. Generally, we make the following recommendations:

- For applications where computational resources are not significantly limited or where an extremely high degree of accuracy is needed, use a LOT with a hybrid functional, a diffuse-augmented basis set of triple-zeta quality, and an implicit solvent model accounting for non-electrostatic energy contributions. Ex: ωB97X-V/def2-TZVPPD/SMD, B3LYP/def2-TZVPD/SMD.

- If computational resources are limited, or for use in coarse high-throughput screening involving extremely large numbers of species, consider using a diffuse-augmented split-valence basis set. Alternatively, one could consider using a triple-zeta basis set with an implicit solvent model used only as a correction to the electronic energy. A high-quality implicit solvent model should be considered especially important in these cases. Ex: B3LYP/def2-SVPD/SMD, B3LYP/def2-TZVPD/Vacuum + SMD.

Fully Automating Molecular DFT

Implementation Overview

We have implemented a computational infrastructure for high-throughput and automated DFT calculations using the Q-Chem electronic structure code in existing open-source Python packages developed by the Materials Project,[9,10] namely `pymatgen`,[72] `custodian`, and `atomate`.[73] The modules in these codes used specifically for Q-Chem, along with their purposes, are described in Figure 2a.

Some of the work reported here is based on previous methods developed for the Electrolyte Genome project.[74] The workflows designed for the Electrolyte Genome project facilitated analysis of electrolyte molecules, with a particular emphasis on calculation of redox potentials and ion pair dissociation energies. While this effort required many of the features that we describe below, including automated error handling and elimination of imaginary frequencies, its narrow focus on generally stable electrolyte molecules and their reduced or oxidized counterparts meant that the workflows developed for the Electrolyte Genome were not well suited to more complex analysis, for instance the consideration of reactive molecular fragments that we describe below. We have improved upon the work of the Electrolyte Genome to be able to handle more challenging electrochemically relevant molecules and a wider range of molecular properties - including bond dissociation energies - in a high-throughput fashion.

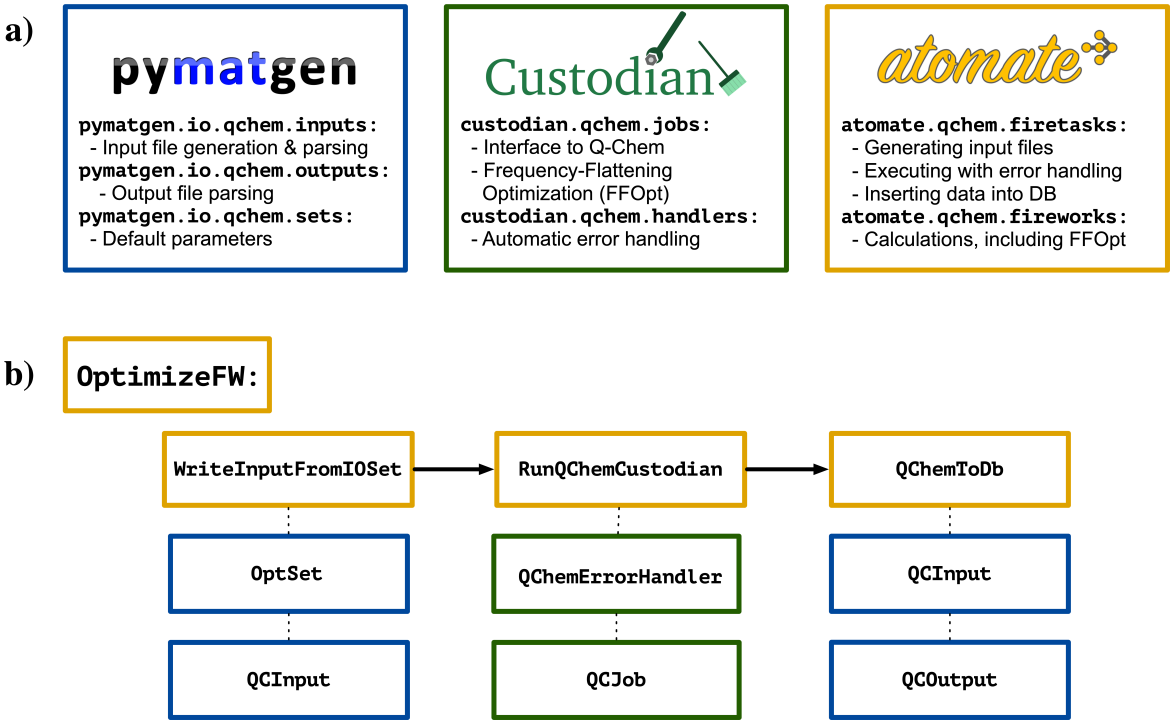


Figure 2: An overview of our automated high-throughput framework, as implemented in `pymatgen` (blue), `custodian` (green), and `atomate` (yellow) (a); an example calculation (Firework) for geometry optimization (b), indicating the different steps and the ways in which `pymatgen`, `custodian`, and `atomate` interact. First, the input file is written using default parameters defined in `pymatgen`. Then, the geometry optimization calculation is performed using the Q-Chem interface in `custodian` and an automated error handler. Finally, once the calculation is finished, the input and output files are parsed using `pymatgen`, and the results from the calculation are added to a database.

The basic functionality to generate, process, analyze, and manipulate molecules is included in `pymatgen`. We have added functionality to read and write Q-Chem input files and to parse Q-Chem output files. In addition, we have developed a number of "Sets", pre-defined collections of input parameters appropriate for common types of calculations. These Sets, by default, use one of the combinations of density functional and basis set identified above as appropriate for electrochemical applications (ωB97X-V/def2-TZVPPD); no

implicit solvent method is used by default, as users may be interested in vacuum calculations. Both PCM and SMD are supported, though as noted above, for electrochemical applications, SMD is strongly preferred.

The `custodian` Q-Chem module defines the interface between Q-Chem and our automation framework in `atomate`. It can execute arbitrary Q-Chem jobs and can automatically check for, detect, and correct errors in Q-Chem calculations. One specific type of calculation, which we call a Frequency-Flattening Optimization, or FFOpt, is also included in `custodian`. This calculation uses the exact Hessian calculated during a vibrational frequency analysis to confirm that an optimized structure is a minimum on the PES, rather than some other stationary point (transition state or higher-order saddle point); if the optimization has not led to a minimum structure, the exact Hessian informs further optimization, leading to imaginary frequencies being successfully "flattened".

The Q-Chem module in `atomate` combines the Q-Chem input and output modules in `pymatgen` and the Q-Chem interface and error handlers in `custodian` to perform Q-Chem jobs and analyze their data in a high-throughput fashion. An example calculation, the `OptimizeFW`, is schematically shown in Figure 2b. First, based on some input parameters, a Q-Chem input file for a geometry optimization calculation is written. Then, the optimization job is run, with `custodian` waiting for completion and, upon completion, checking for errors. If the job completes without errors, then the output is parsed and stored in a database. Individual Q-Chem calculations, represented in `atomate` by Fireworks like `OptimizeFW`, can be combined to form more complex workflows.

We note that, while our code is presently designed to work only with Q-Chem, the methods that we describe here could be applied in general to most molecular DFT codes.

Error Correction

Once a calculation, deployed through `custodian` and `atomate`, has terminated, its output file is parsed for errors. If any errors are detected, then a recipe-based error correction process is conducted. Through extensive experience with DFT calculations, we have developed a series of rules governing which remedies should be applied to eliminiate particular errors. If possible, the error handler applies the appropriate remedy, generally by altering the input parameters for the Q-Chem calculation, generating a new input file, and restarting the calculation. If an error is encountered in the re-started calculation, the same recipe-based error correction procedure is applied. If the error handler is unable to interpret the error, or if no remedy has been implemented for a particular error type, the calculation is allowed to fail.

Even if there are several possible remedies, only one remedy is applied at a time. The appropriate remedy for a given error may be sensitive to the parameters with which the calculation was run. Those parameters, in turn, may depend on the type and number of errors that the calculation has encountered previously.

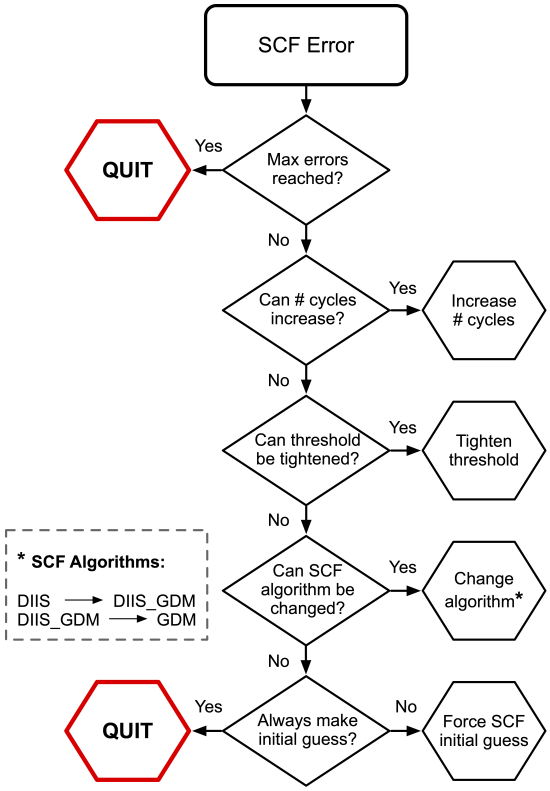


Figure 3: A flowchart for correcting an SCF convergence error. When the error is encountered only a single remedy will be applied. If there is no possible remedy, or if too many errors have already been encountered, then the error handler will quit, and the calculation will be allowed to fail.

To illustrate the error-correction process, Figure 3 depicts the logic dictating how a convergence error for an SCF calculation should be remedied. The first possible remedy involves increasing the number of SCF cycles allowed; if the number of SCF cycles is lower than a maximum defined by `custodian`, then the number of cycles are increased to that maximum. If that remedy cannot be applied, either because it has already been applied or because the user specified a large number of SCF cycles initially, then the error handler will attempt to tighten the threshold for neglect of two-electron integrals. Tightening this threshold means that a greater number of two-electron integrals must be calculated, thereby increasing the computational cost. However, a tight threshold increases precision, which in our experience can improve SCF convergence behavior. If the threshold cannot be tightened, then the next remedy is to alter the SCF algorithm. The geometric direct minimization (GDM) method[75] tends to be highly robust at converging SCF calculations even for challenging molecules. However, because of its higher cost, we generally begin either using the more rapid Direct Inversion of the Iterative Subspace method (DIIS)[76,77] or a combination of the two methods (DIIS_GDM in Q-Chem), using GDM only as a method of last resort. Finally, the SCF settings are altered such that an initial guess electron density is generated for each SCF calculation, with no knowledge of prior calculations. Using the previous solution as a starting point for an SCF calculation can improve efficiency, but it can also fail to capture electronic state reordering in a newly visited region the PES, occasionally resulting in SCF convergence problems. If none of these remedies can be applied, if all of them have been applied already, or if the number of errors encountered in total has exceeded a user-defined limit, then the calculation will fail without further attempt to remedy the error.

In addition to SCF convergence errors, we have implemented remedies for a range of errors that might arise during a calculation (failing to optimize the molecular geometry, failing to transform from internal to Cartesian coordinates, failing to calculate the Hessian eigenvalues for a vibrational frequency calculation, etc.) or while preparing a calculation (failing to parse the input file, failing to access the DFT code executable file, failure to access a license file, etc.). We continue to actively develop these error handlers, adding new types of errors that can be corrected and improving the remedies used to correct those errors.

Frequency Flattening to PES Minima

The goal of geometry optimization is to minimize the energy and to determine the stable molecular geometry. Generally, a optimizer will seek to reduce the gradient to zero, indicating that a stationary point has been found. However, convergence to a stationary point does not guarantee convergence to a minimum of the PES; it is also possible to converge to an nth-order saddle point, where n is the number of imaginary frequencies. It is important to know when a calculation has converged to a saddle point and how to remedy it. Saddle points may provide poor approximations to the minimum-energy structure and may present significantly higher energy than the nearest minimum energy structure. Furthermore, saddle points can exhibit different bonding behavior or even different connectivity from the minimum, as in the example shown in Figure 4.

Most often, geometry optimization in DFT is conducted using a quasi-Newton-Raphson method; at each step, the energy and gradient are calculated, and the gradient is used to generate an approximation of the second derivative (Hessian) matrix.[78] While, in some methods, the exact Hessian is calculated at each step, this is prohibitively expensive in most cases and is therefore inappropriate for high-throughput applications. Because the Hessian used in quasi-Newton-Raphson optimization is not exact, the optimizer's knowledge of the curvature of the PES is limited. This makes it relatively common for geometry optimization algorithms to converge to saddle points instead of minima, especially for complex reactive fragments.

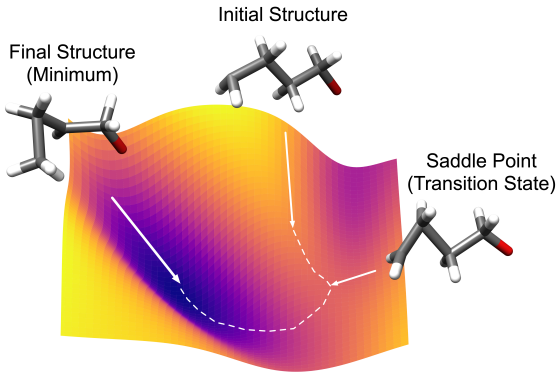


Figure 4: A frequency-flattening optimization trajectory. From an initial structure, the molecule is optimized first to a saddle point (in this case, a first-order saddle point, or transition state) as identified by a frequency calculation at the stationary point. From there, the optimization is restarted with knowledge of the exact Hessian at the saddle point, leading to a true solution, e.g. a converged minimum of the PES. Along this trajectory, the bonding and geometry of the molecule change considerably. Here, an artificial PES is used to illustrate the different features of the potential energy landscape.

During an FFOpt calculation, as illustrated in Figure 5, we perform successive optimization calculations until the structure has converged to a true PES minimum. In order to determine if a converged structure lies on a PES minimum or a saddle point, we perform a vibrational frequency calculation following each completed optimization calculation. Frequency calculations serve a dual purpose, simultaneously providing information about the curvature of the PES (the exact Hessian) and the nature of the converged stationary point while also providing some thermodynamic information, including the molecular enthalpy and entropy. If there are no imaginary frequencies, then the structure is a PES minimum, and no further calculations are needed. If there are imaginary frequencies, then the structure is a PES saddle point. In the subsequent optimization calculation, we use the exact Hessian reported by the frequency calculation to provide a better description of the local PES, allowing the optimizer to move away from the saddle point and towards a true minimum. This procedure can be repeated as many times as needed until a minimum is found.

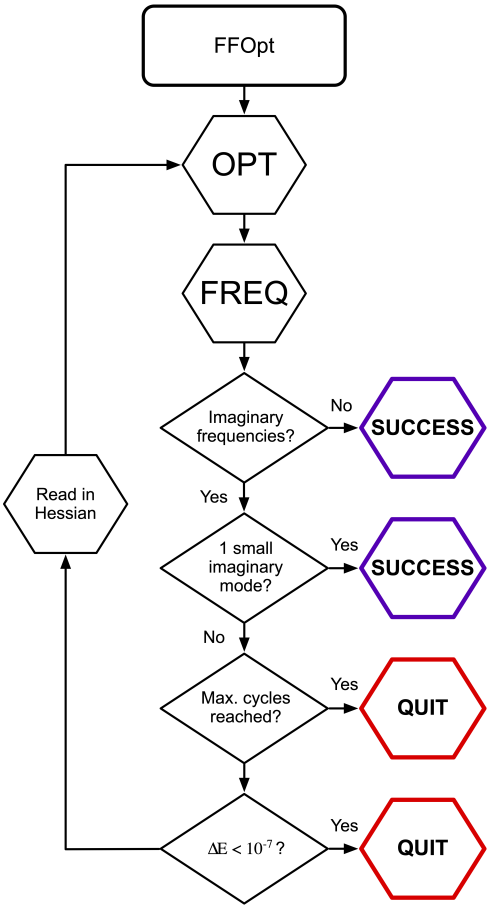


Figure 5: The frequency-flattening optimization (FFOpt) procedure. In the initial step, the geometry is optimized and a vibrational frequency calculation is performed. If there are no imaginary frequencies, or if there is a single imaginary frequency with very small magnitude, the calculation completes successfully. Otherwise, the Hessian from the vibrational frequency calculation will be used to inform the next cycle of optimization.

Typically, in order to limit the computational cost of an individual calculation, the user will define some maximum number of optimization-frequency cycles. Additional cycles will not be pursued if there is only one imaginary mode with a very small frequency magnitude ($|\nu| \leq 15\ cm^{-1}$) or if the energy has changed by less than 10^{-7} Hartree from the previous cycle to the current cycle. Very small, singular imaginary frequencies are allowed because they may not correspond to true transition states; rather, they could be artifacts of numerical noise in the frequency calculation. If there is a single imaginary mode with a small frequency magnitude, the calculation is still considered a success; otherwise, a calculation which terminates with at least one imaginary frequency is considered a failure.

Case study - solid-electrolyte interphase formation in Li-ion batteries

To demonstrate both the necessity of error correcting and frequency-flattening procedures and the efficacy of these procedures as we have implemented them, we attempted 12,218 FFOpt calculations on molecules - primarily reactive fragments - that may be involved in electrolyte decomposition and SEI formation in Li-ion batteries. The methods used to generate these molecular fragments are outside the scope of this work and will be discussed in a future publication. All calculations were conducted at the ωB97X-V/def2-TZVPPD/SMD level of theory, identified above to be appropriate for electrochemical applications. We employed the measured dielectric constant of a 3:7 ethylene carbonate (EC)/ethyl methyl carbonate (EMC)

mixture ($\varepsilon = 18.5$),[79] a commonly used electrolyte solvent blend, with all other SMD solvent parameters based on EC.[74] The threshold for neglect of two-electron integrals was set to 10^{-14}, the most stringent threshold allowed in Q-Chem. We note that both the use of this advanced LOT and the elevated threshold should reduce the number of errors encountered during these calculations.

Data set composition

Some basic information about the calculations included in the case study is provided in Figure 6. The molecules involved in these calculation include 9 different atomic species (C, H, O, N, F, P, S, Li, and Mg), three molecular charge states (-1, 0, and +1), and three spin multiplicities (singlet, doublet, and triplet). While this data set does include some neutral, closed-shell organic molecules, 66.1% (8,082) are charged, 53.8% (6,573) are radical, and 60.7% (7,419) contain at least one coordinated metal. All calculations were conducted in an SMD implicit solvent environment. Thus, on the whole, these calculations pose a significant challenge, even considering the LOT used.

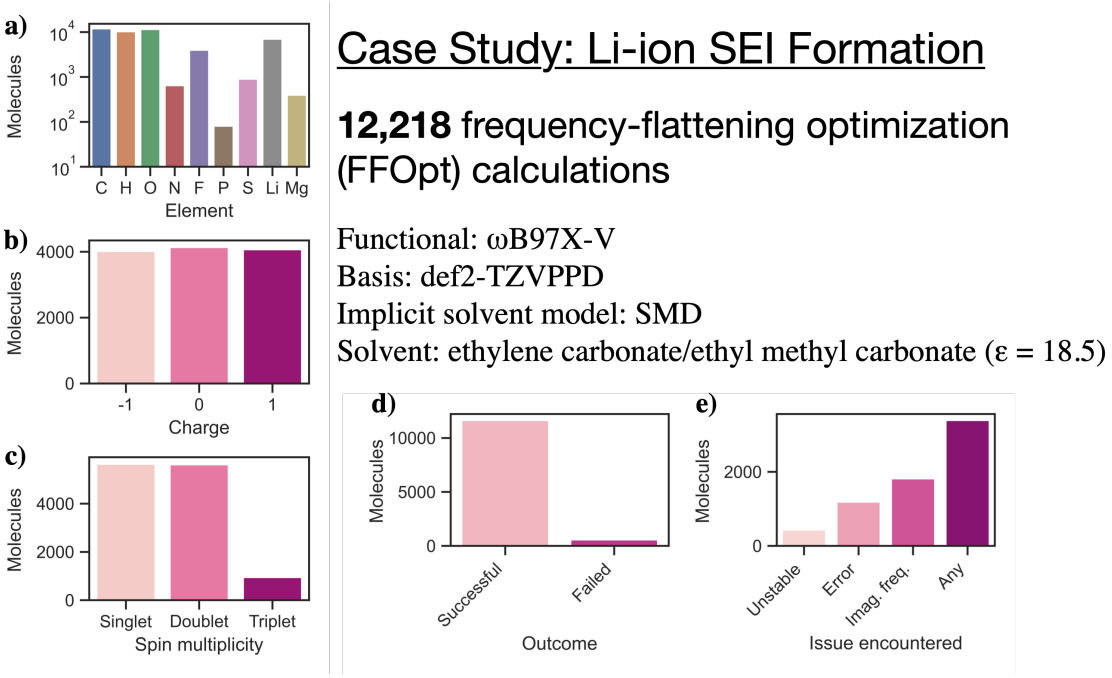


Figure 6: Composition of a data set of 12,218 FFOpt calculations at the ωB97X-V/def2-TZVPPD/SMD level of theory, including the number of calculations that include at least one instance of the elements C, H, O, N, F, P, S, Li, and Mg (a) and with various charges (b) and spin multiplicities (c). Information about calculation outcomes (d) and the occurrence of various issues in the test set (e) is also included.

From Figure 6 we note that error correction and frequency flattening were both used frequently while compiling this data set, signaling their importance for reducing failure rates in high-throughput molecular DFT. Error correction procedures were used in 9.2% (1,187) of calculations, and frequency flattening was used in 14.9% (1,816) of calculations. Molecules were found to be unstable in 3.6% (434) of cases. These molecules dissociated into multiple fragments during geometry optimization, and the fragments continued to move further apart as the optimization proceeded, implying that the most stable arrangement was infinite separation. In total, some problem was identified in 27.8% (3,392) of calculations. For reference, in the dataset generated by the Electrolyte Genome project,[74] problems were encountered in only 9.2% of cases, indicating that the molecules considered here are considerably more challenging to simulate.

Calculation failure

Remarkably, in spite of the large number of calculations that encounter errors or converge to saddle points, only 4.7% (576) of the calculations failed due to any error or other problem. Table 4 lists the various ways in which calculations in this data set failed. Perhaps surprisingly, calculation failure by errors in the DFT calculation are relatively uncommon (8.0% of all failures). Within this category, the most common errors encountered are related to failure to converge SCF calculations, failing to optimize the geometry to a stationary point in a finite number of optimization steps, and failing to converge the rational function optimization step during a geometry optimization. Failure to flatten imaginary frequencies causes 16.7% of all failed calculations. In these cases, the calculation was able to optimize to a saddle point but could not move from that saddle point to a PES minimum in the user-defined number (here chosen as 10) of frequency-flattening cycles. The remainder (75.3%) of calculation failures are caused by molecular instability. These failures imply that the molecule being optimized is inherently unstable at the particular level of theory, though that does not necessarily mean that they would be unstable in reality. For instance, such molecules could be stabilized by explicit solvent molecules. If unstable species are removed from consideration, leaving 11,784 calculations, then from an initial failure rate of 25.1% (2,958), we are able to reduce the error rate through frequency flattening and error correction to just 1.2% (142). We justify this exclusion by noting that a species that is found to be unstable could never be expected to be optimized successfully at the chosen LOT.

Table 4: Causes of failure for the case study data set

Failure type	**Number of failures**
Error in DFT calculation	46
Unstable molecule	434
Frequency flattening failure	96
Total	**576**

Frequency flattening

In this data set, 1,720 calculations used frequency flattening successfully to converge to a PES minimum. In 14.9% (257) of cases, the geometry after the first optimization cycle (at a saddle point) was not isomorphic to the final optimized geometry; in other words, the bonding changed during frequency flattening. On average, when a single frequency was flattened (the geometry converged from a first-order saddle point to a minimum), the electronic energy decreased by 0.271 eV. Even for reasonably small imaginary frequencies ($-50cm^{-1} < \nu < -15cm^{-1}$), the electronic energy decreased by 0.093 eV from the saddle point to the minimum on average. Thus, we can say that frequency flattening can be critically important to obtain accurate bonding and energies.

Ideally, when an imaginary frequency is encountered, a single frequency-flattening cycle would be sufficient to move from the saddle point to a true PES minimum. This is indeed often the case, even for higher order saddle points. However, even when the full Hessian is provided at the beginning of an optimization calculation, after a number of steps, the calculation's internal approximate Hessian can rapidly diverge from reality, causing the optimizer to converge once again to a saddle point (though usually a different one).

Table 5 reports the average number of frequency-flattening cycles required to eliminate all significant imaginary frequencies (with magnitude $|\nu| > 15cm^{-1}$) for calculations that initially converged to saddle points of various orders (1-7). Note that we encountered only one sixth-order saddle point and one seventh-order saddle point that converged to a minimum; these data points are included in Table 5 for completeness but should not be considered in statistical analysis due to the extremely small sample size. Fifth-order saddle points, for which there are only three data points in this set, could be excluded on similar grounds.

Table 5: Average number of frequency-flattening cycles required to converge saddle points to minima

Order	Average cycles	Data points
1	1.08 (0.01)	1154
2	1.18 (0.03)	402
3	1.35 (0.07)	124
4	1.34 (0.14)	35
5	1.67 (0.67)	3
6	7.00 (N/A)	1
7	3.00 (N/A)	1

Table 5: Values in parentheses are standard errors of the mean.

As might be expected, higher-order saddle points are relatively uncommon. First-order saddle points are the most common, making up 67.1% of all successful cases of frequency flattening, though second-order (23.4%) and third-order (7.2%) saddle points also occur with reasonably high frequency. The number of frequency-flattening cycles required to reach a minimum increases as the order of the saddle point increases; however, from first-order saddle points to fifth-order saddle points, this average increases by less than one frequency-flattening cycle. Therefore, even when higher-order saddle points are encountered, the computational cost of flattening the imaginary frequencies does not increase dramatically.

Discussion

Methods of Reducing Cost

Often in high-throughput molecular DFT studies, a low-cost pre-optimization step will be used to determine a reasonable guess geometry in order to reduce the overall computational cost of the geometry optimization. Common tactics for pre-optimization include using a force-field or semi-empirical method or using a lower LOT within DFT (most commonly using a smaller basis set). For calculations of solvated molecules, many studies begin by optimizing in vacuum before optimizing in an implicit solvent environment.

As we have shown, the choice of the LOT is important in order to reduce the likelihood of errors and improving accuracy. For the challenging, electrochemically relevant molecules that we have focused on here, especially charged, radical, metal-coordinated, or solvated molecules, we have found that pre-optimization methods may be inappropriate or lead to poor results. Semi-empirical codes are often parameterized exclusively or primarily against relatively stable and well-behaved (neutral and closed-shell) molecules,[80,81] making their efficacy for charged and open-shell molecules suspect. It should be noted that there are some semi-empirical codes that are designed to be utilized with metal-containing systems.[82,83]

Even using lower LOTs within DFT can be problematic; in general, there is no guarantee that two levels of theory will produce similar optimized geometries. For calculations of solvated molecules, the elimination of the implicit solvent method from the LOT will generally affect the optimized structure and may be especially damaging for calculations involving ions, as the solvent often critically stabilizes such species. In our benchmark, we did find that in some cases, pre-optimizing in vacuum with a large basis set, when corrected by an advanced implicit solvent model, can lead to adequate results at relatively low cost. However, we note that our benchmark did not include the types of reactive fragment species where spurious bonding changes in vacuum would be more likely and especially detrimental. Further, calculations in vacuum also appear to be more prone to failure. The majority of the LOTs excluded from the benchmark because the calculation failures were conducted in vacuum; this remains true even if solvent-corrected LOTs (Vacuum + SMD, Vacuum + COSMO) are not counted.

To examine the possible utility of low-cost computational methods to pre-optimize, we performed geometry optimization calculations of all 11,642 successful calculations in our case study using the GFN2-xTB method from Grimme's group.[83] We chose GFN2-xTB because it was designed in part to produce reasonably accurate molecular geometries and to model non-covalent interactions such as long-range coordinate bonding. In addition, GFN2-xTB includes an implicit solvation model, allowing for a closer comparison to our chosen LOT (ωB97X-V/def2-TZVPPD/SMD). In all geometry optimizations, the initial structure used for the DFT calculations were also used for the calculations with GFN2-xTB. We used the most stringent optimization criteria ("extreme", requiring a maximum energy change of 5.0×10^{-8} Ha, maximum gradient

of 5.0×10^{-5} Ha $\mathring{A}^{-1}$). The implicit solvation model in GFN2-xTB is only parameterized for a set of common solvents; we used acetone, as its dielectric constant ($\varepsilon = 20.493$) is similar to that used in the case study ($\varepsilon = 18.5$).

Overall, it appears that for this case study, GFN2-xTB did not provide suitable geometries. Of the 11,642 structures, 42 failed to optimize completely, and only 6,286 (54.2%) of the remaining 11,600 calculations optimized to a structure with the same connectivity as the DFT-optimized structure. As Table 6 shows, the results are similar regardless of the molecule's charge state or spin multiplicity, though GFN2-xTB is somewhat more successful on molecules without metals (65.7%) than for molecules with metals (46.8%). If a semi-empirical method designed for molecular geometry optimization only captures the correct bonding behavior in slightly over one-half of molecules, to say nothing of the actual molecular coordinates (bond lengths, angles, dihedrals, etc.), we suggest that semi-empirical methods should be avoided for data sets involving molecules with complex electronic structure, such as those described in this work.

Table 6: Performance of GFN2-xTB geometry optimization on the case study data set

Type of Molecule	**Number of Molecules**	**Successful Optimizations**	**Success Rate (%)**
Neutral	3,948	2,340	59.3
Charged	7,652	3,946	51.6
Closed-shell	5,277	2,579	48.9
Radical	6,323	3,707	58.6
Contains Metal	7,076	3,314	46.8
No metal	4,524	2,973	65.7
Overall	11,600	6,286	54.2
Failed to Converge	42	N/A	N/A

Table 6: Here, a successful optimization is one that produces a structure that is isomorphic (has the same bonding) as the DFT-optimized geometry, while an unsuccessful optimization produces a non-isomorphic structure. Radical molecules here include both doublets and triplets.

Handling Failed Calculations

The error correction procedures that we have described and implemented are based on template responses. These responses have been designed based on human intuition coupled with extensive experience. While not reported here, tens of thousands of calculations were used to test a wide variety of possible error correction schemes, leading to those currently included in `custodian`. These error handlers are capable of reducing calculation failure rate from over 25% to under 5% if unstable molecules are included, or just above 1% if they are excluded. Nonetheless, these error correction procedures do still fail in a non-trivial number of cases.

Calculation failure can be reflective of problems in the original molecular geometry, in the LOT used, or in the error correction scheme (if applied). For example, it is possible that the particular error encountered simply had not been adequately addressed by the error handlers. Without further information about the specific calculation involved, it may be impossible to know which of these possibilities is true.

Regardless of the root cause, failed calculations necessitate more thorough human intervention. In our case study, over 500 calculations failed with some error or problem. Analyzing these calculations individually could take many hours of human time. In some cases, analyzing all failures individually may be truly necessary. However, depending on the scale of the high-throughput effort, this level of effort may be daunting or even impossible.

Assuming that our case study is a representative data set for the study of complex and reactive molecules, unstable geometries provide the most significant challenge for high-throughput DFT. Two options available to researchers aiming to avoid analysis of every error would be either to perturb the structures that failed to converge to a minimum (providing a possibly better guess geometry) or to optimize at a different, perhaps more accurate, level of theory. In either case, the primary goal would be to avoid errors and successfully optimize the structure. However, even if these modified calculations fail as well, this could provide valuable additional information about the nature of the failure. Note that, if the molecule is predicted to be unstable at a high LOT, it is unlikely that any further remedies will be effective.

The Effect of Explicit Solvent

All calculations conducted here, either as part of the level of theory benchmark study or the case study, have exclusively relied on implicit solvent models to approximate the effect of a solvent environment. In some cases, for redox potentials, the effect of explicit solvent is small,[84] and an implicit solvent model may be highly accurate. However, as Cheng et al. pointed out,[14] for applications such as multivalent batteries, explicit solvation may be highly important for accurate redox potential calculations. Aside from redox potentials, it is well known that the presence of explicit solvent can in some cases significantly effect reaction energy barriers.[85–87]

The consideration of explicit solvation for high-throughput calculations is inherently challenging. Not only do such calculations require some idea of the solvent shell around the solute of interest, but they also dramatically increase the computational cost, possibly limiting the amount of calculations which can be conducted or the level of theory that can be used. To the knowledge of the authors, no high-throughput DFT study has included explicit solvent in all calculations. However, such methods may be capable of achieving even higher accuracy than those discussed here.

Conclusion

In this work, we developed a methodology for accurate high-throughput molecular DFT calculations with a low rate of failure. Through a combinatorial benchmark of five hybrid density functionals, four basis sets, and five implicit solvent models, we identified a group of 10 levels of theory that are appropriate for use in high-throughput electrochemical studies. We also presented an open-source framework for automated high-throughput DFT calculations involving robust error correction procedures. Using one of the identified appropriate levels of theory (ωB97X-V/def2-TZVPPD/SMD) and our automation framework, we conducted a case study on over 12,000 calculations of reactive, electrochemically relevant molecules. These calculations fail in only 1.2% of cases, compared to 25.1% without any intervention. While this work has primarily focused on electrochemistry, these methods could be widely applied to a variety of complex chemistries.

Acknowledgments

The software and data infrastructure framework presented here was collaboratively supported. Application to multivalent electrolyte molecules was supported by the Joint Center for Energy Storage Research, an Energy Innovation Hub funded by the US Department of Energy, Office of Science, Basic Energy Sciences. Calculation of electrolyte molecules relevant for Si anode applications was supported the Silicon Electrolyte Interface Stabilization (SEISta) Consortium directed by Brian Cunningham under the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Vehicle Technologies of the U.S. Department of Energy, Contract No. DE-AC02-05CH11231. Additional support for Li metal electrolyte molecules comes from the Battery Materials Research (BMR) program directed by Tien Duong under the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Vehicle Technologies of the U.S. Department of Energy, Contract DE-AC02-05CH11231. Data for this study was produced using computational resources provided by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility under Contract No. DE-AC02-05CH11231, the Eagle HPC system at the National Renewable Energy Laboratory (NREL), and the Lawrencium HPC cluster at Lawrence Berkeley National Laboratory.

Author Contributions

S.M.B led development of the automation and error handling framework for Q-Chem calculations, with assistance from E.W.C.S.-S., B.W., and S.D. E.W.C.S.-S. designed and performed the benchmark study. S.M.B. and E.W.C.S.-S. performed the calculations used in the case study. E.W.C.S.-S. wrote the original manuscript, and all authors assisted in editing the manuscript. K.A.P. conceived of the study and secured funding and computational resources.

Supporting Information

Description of components of the level of theory; complete level of theory benchmark results

References

[1] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data*, 1(1):1–7, August 2014.

[2] Justin S. Smith, Olexandr Isayev, and Adrian E. Roitberg. ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Sci. Data*, 4(1):170193, December 2017.

[3] Scott Kirklin, James E. Saal, Bryce Meredig, Alex Thompson, Jeff W. Doak, Muratahan Aykol, Stephan Rühl, and Chris Wolverton. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput Mater*, 1(1):1–15, December 2015.

[4] James E. Saal, Scott Kirklin, Muratahan Aykol, Bryce Meredig, and C. Wolverton. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM*, 65(11):1501–1509, November 2013.

[5] Johannes Hachmann, Roberto Olivares-Amaya, Sule Atahan-Evrenk, Carlos Amador-Bedolla, Roel S. Sánchez-Carrera, Aryeh Gold-Parker, Leslie Vogt, Anna M. Brockway, and Alán Aspuru-Guzik. The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid. *J. Phys. Chem. Lett.*, 2(17):2241–2251, September 2011.

[6] Stefano Curtarolo, Wahyu Setyawan, Shidong Wang, Junkai Xue, Kesong Yang, Richard H. Taylor, Lance J. Nelson, Gus L.W. Hart, Stefano Sanvito, Marco Buongiorno-Nardelli, Natalio Mingo, and Ohad Levy. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci*, 58:227–235, June 2012.

[7] David D. Landis, Jens S. Hummelshøj, Svetlozar Nestorov, Jeff Greeley, Marcin Dułak, Thomas Bligaard, Jens K. Nørskov, and Karsten W. Jacobsen. The Computational Materials Repository. *Comput. Sci. Eng.*, 14(6):51–57, November 2012.

[8] Maho Nakata and Tomomi Shimazaki. PubChemQC Project: A Large-Scale First-Principles Electronic Structure Database for Data-Driven Chemistry. *J. Chem. Inf. Model.*, 57(6):1300–1308, June 2017.

[9] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.*, 1(1):011002, July 2013.

[10] Anubhav Jain, Joseph Montoya, Shyam Dwaraknath, Nils E. R. Zimmermann, John Dagdelen, Matthew Horton, Patrick Huck, Donny Winston, Shreyas Cholia, Shyue Ping Ong, and Kristin Persson. The Materials Project: Accelerating Materials Design Through Theory-Driven Data and Tools. In Wanda Andreoni and Sidney Yip, editors, *Handbook of Materials Modeling*, pages 1–34. Springer International Publishing, Cham, 2018.

[11] Ilana Y. Kanal, Steven G. Owens, Jonathon S. Bechtel, and Geoffrey R. Hutchison. Efficient Computational Screening of Organic Polymer Photovoltaics. *J. Phys. Chem. Lett.*, 4(10):1613–1623, May 2013.

[12] Martin Korth. Large-scale virtual high-throughput screening for the identification of new battery electrolyte solvents: evaluation of electronic structure theory methods. *Phys. Chem. Chem. Phys.*, 16(17):7919–7926, 2014.

[13] Mathew D. Halls and Ken Tasaki. High-throughput quantum chemistry and virtual screening for lithium ion battery electrolyte additives. *J. Power Sources*, 195(5):1472–1478, March 2010.

[14] Lei Cheng, Rajeev S. Assary, Xiaohui Qu, Anubhav Jain, Shyue Ping Ong, Nav Nidhi Rajput, Kristin Persson, and Larry A. Curtiss. Accelerating Electrolyte Discovery for Energy Storage with High-Throughput Screening. *J. Phys. Chem. Lett.*, 6(2):283–291, January 2015.

[15] Muratahan Aykol, Soo Kim, Vinay I. Hegde, David Snydacker, Zhi Lu, Shiqiang Hao, Scott Kirklin, Dane Morgan, and C. Wolverton. High-throughput computational design of cathode coatings for Li-ion batteries. *Nat. Comm.*, 7(1):13779, December 2016.

[16] Yihan Xiao, Lincoln J. Miara, Yan Wang, and Gerbrand Ceder. Computational Screening of Cathode Coatings for Solid-State Batteries. *Joule*, 3(5):1252–1275, May 2019.

[17] Tim Mueller, Geoffroy Hautier, Anubhav Jain, and Gerbrand Ceder. Evaluation of Tavorite-Structured Cathode Materials for Lithium-Ion Batteries Using High-Throughput Computing. *Chem. Mater.*, 23(17):3854–3862, September 2011.

[18] Miao Liu, Anubhav Jain, Ziqin Rong, Xiaohui Qu, Pieremanuele Canepa, Rahul Malik, Gerbrand Ceder, and Kristin A. Persson. Evaluation of sulfur spinel compounds for multivalent battery cathode applications. *Energy Environ. Sci.*, 9(10):3201–3209, 2016.

[19] Liang Su, Magali Ferrandon, Jeffrey A. Kowalski, John T. Vaughey, and Fikile R. Brushett. Electrolyte Development for Non-Aqueous Redox Flow Batteries Using a High-Throughput Screening Platform. *J. Electrochem. Soc.*, 161(12):A1905, September 2014.

[20] Peter Strasser, Qun Fan, Martin Devenney, W. Henry Weinberg, Ping Liu, and Jens K. Nørskov. High Throughput Experimental and Theoretical Predictive Screening of Materials – A Comparative Study of Search Strategies for New Fuel Cell Anode Catalysts. *J. Phys. Chem. B*, 107(40):11013–11021, October 2003.

[21] Jeff Greeley, Thomas F. Jaramillo, Jacob Bonde, Ib Chorkendorff, and Jens K. Nørskov. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nat. Mater.*, 5(11):909–913, November 2006.

[22] Hong Zhu, Geoffroy Hautier, Umut Aydemir, Zachary M. Gibbs, Guodong Li, Saurabh Bajaj, Jan-Hendrik Pöhls, Danny Broberg, Wei Chen, Anubhav Jain, Mary Anne White, Mark Asta, G. Jeffrey Snyder, Kristin Persson, and Gerbrand Ceder. Computational and experimental investigation of TmAgTe2 and XYZ2 compounds, a new group of thermoelectric materials identified by first-principles high-throughput screening. *J. Mater. Chem. C*, 3(40):10554–10565, October 2015.

[23] Xiaochen Sun, Santiago Vilar, and Nicholas P. Tatonetti. High-Throughput Methods for Combinatorial Drug Discovery. *Sci. Transl. Med.*, 5(205):205rv1–205rv1, October 2013.

[24] Noriaki Okimoto, Noriyuki Futatsugi, Hideyoshi Fuji, Atsushi Suenaga, Gentaro Morimoto, Ryoko Yanai, Yousuke Ohno, Tetsu Narumi, and Makoto Taiji. High-Performance Drug Discovery: Computational Screening by Combining Docking and Molecular Dynamics Simulations. *PLoS Comput. Biol.*, 5(10):e1000528, October 2009.

[25] Jürgen Bajorath. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov.*, 1(11):882–894, November 2002.

[26] Sharangdhar S. Phatak, Clifford C. Stephan, and Claudio N. Cavasotto. High-throughput and in silico screenings in drug discovery. *Expert Opin. Drug Discov.*, 4(9):947–959, September 2009.

[27] Gregor N. Simm and Markus Reiher. Context-Driven Exploration of Complex Chemical Reaction Networks. *J. Chem. Theory Comput.*, 13(12):6108–6119, December 2017.

[28] Laura E. Rush, Paul G. Pringle, and Jeremy N. Harvey. Computational Kinetics of Cobalt-Catalyzed Alkene Hydroformylation. *Angewandte Chemie International Edition*, 53(33):8672–8676, 2014.

[29] Giovanni Pizzi, Andrea Cepellotti, Riccardo Sabatini, Nicola Marzari, and Boris Kozinsky. AiiDA: Automated Interactive Infrastructure and Database for Computational Science. *Comput. Mater. Sci.*, 111:218–230, January 2016.

[30] Stefano Curtarolo, Wahyu Setyawan, Gus L. W. Hart, Michal Jahnatek, Roman V. Chepulskii, Richard H. Taylor, Shidong Wang, Junkai Xue, Kesong Yang, Ohad Levy, Michael J. Mehl, Harold T. Stokes, Denis O. Demchenko, and Dane Morgan. AFLOW: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.*, 58:218–226, June 2012.

[31] Tam Mayeshiba, Henry Wu, Thomas Angsten, Amy Kaczmarowski, Zhewen Song, Glen Jenness, Wei Xie, and Dane Morgan. The MAterials Simulation Toolkit (MAST) for atomistic modeling of defects and diffusion. *Comput. Mater. Sci.*, 126:90–102, January 2017.

[32] Jaron T. Krogel. Nexus: A modular workflow management system for quantum simulation codes. *Comput. Phys. Commun.*, 198:154–168, January 2016.

[33] Kiran Mathew, Joseph H. Montoya, Alireza Faghaninia, Shyam Dwarakanath, Muratahan Aykol, Hanmei Tang, Iek-heng Chu, Tess Smidt, Brandon Bocklund, Matthew Horton, John Dagdelen, Brandon Wood, Zi-Kui Liu, Jeffrey Neaton, Shyue Ping Ong, Kristin Persson, and Anubhav Jain. Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Comput. Mater. Sci.*, 139:140–152, November 2017.

[34] Felipe Zapata, Lars Ridder, Johan Hidding, Christoph R Jacob, Ivan Infante, and Lucas Visscher. Qmflows: A tool kit for interoperable parallel workflows in quantum chemistry. *J. Chem. Info. Model.*, 59(7):3191–3197, 2019.

[35] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54(16):11169–11186, October 1996.

[36] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21(39):395502, September 2009.

[37] Yihan Shao, Zhengting Gan, Evgeny Epifanovsky, Andrew T. B. Gilbert, Michael Wormit, Joerg Kussmann, Adrian W. Lange, Andrew Behn, Jia Deng, Xintian Feng, Debashree Ghosh, Matthew Goldey, Paul R. Horn, Leif D. Jacobson, Ilya Kaliman, Rustam Z. Khaliullin, Tomasz Kuś, Arie Landau, Jie Liu, Emil I. Proynov, Young Min Rhee, Ryan M. Richard, Mary A. Rohrdanz, Ryan P. Steele, Eric J. Sundstrom, H. Lee Woodcock III, Paul M. Zimmerman, Dmitry Zuev, Ben Albrecht, Ethan Alguire, Brian Austin, Gregory J. O. Beran, Yves A. Bernard, Eric Berquist, Kai Brandhorst, Ksenia B. Bravaya, Shawn T. Brown, David Casanova, Chun-Min Chang, Yunqing Chen, Siu Hung Chien, Kristina D. Closser, Deborah L. Crittenden, Michael Diedenhofen, Robert A. DiStasio Jr, Hainam Do, Anthony D. Dutoi, Richard G. Edgar, Shervin Fatehi, Laszlo Fusti-Molnar, An Ghysels, Anna Golubeva-Zadorozhnaya, Joseph Gomes, Magnus W. D. Hanson-Heine, Philipp H. P. Harbach, Andreas W. Hauser, Edward G. Hohenstein, Zachary C. Holden, Thomas-C. Jagau, Hyunjun Ji, Benjamin Kaduk, Kirill Khistyaev, Jaehoon Kim, Jihan Kim, Rollin A. King, Phil Klunzinger, Dmytro Kosenkov, Tim Kowalczyk, Caroline M. Krauter, Ka Un Lao, Adèle D. Laurent, Keith V. Lawler, Sergey V. Levchenko, Ching Yeh Lin, Fenglai Liu, Ester Livshits, Rohini C. Lochan, Arne Luenser, Prashant Manohar, Samuel F. Manzer, Shan-Ping Mao, Narbe Mardirossian, Aleksandr V. Marenich, Simon A. Maurer, Nicholas J. Mayhall, Eric Neuscamman, C. Melania Oana, Roberto Olivares-Amaya, Darragh P. O'Neill, John A. Parkhill, Trilisa M. Perrine, Roberto Peverati, Alexander Prociuk, Dirk R. Rehn, Edina Rosta, Nicholas J. Russ, Shaama M. Sharada, Sandeep Sharma, David W. Small, Alexander Sodt, Tamar Stein, David Stück, Yu-Chuan Su, Alex J. W. Thom, Takashi Tsuchimochi, Vitalii Vanovschi, Leslie Vogt, Oleg Vydrov, Tao Wang, Mark A. Watson,

Jan Wenzel, Alec White, Christopher F. Williams, Jun Yang, Sina Yeganeh, Shane R. Yost, Zhi-Qiang You, Igor Ying Zhang, Xing Zhang, Yan Zhao, Bernard R. Brooks, Garnet K. L. Chan, Daniel M. Chipman, Christopher J. Cramer, William A. Goddard III, Mark S. Gordon, Warren J. Hehre, Andreas Klamt, Henry F. Schaefer III, Michael W. Schmidt, C. David Sherrill, Donald G. Truhlar, Arieh Warshel, Xin Xu, Alán Aspuru-Guzik, Roi Baer, Alexis T. Bell, Nicholas A. Besley, Jeng-Da Chai, Andreas Dreuw, Barry D. Dunietz, Thomas R. Furlani, Steven R. Gwaltney, Chao-Ping Hsu, Yousung Jung, Jing Kong, Daniel S. Lambrecht, WanZhen Liang, Christian Ochsenfeld, Vitaly A. Rassolov, Lyudmila V. Slipchenko, Joseph E. Subotnik, Troy Van Voorhis, John M. Herbert, Anna I. Krylov, Peter M. W. Gill, and Martin Head-Gordon. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.*, 113(2):184–215, January 2015.

[38] Dimitris P. Kessissoglou, editor. *Bioinorganic Chemistry*. Springer Netherlands, Dordrecht, 1995.

[39] Nicolas P. E. Barry and Peter J. Sadler. Exploration of the medical periodic table: towards new targets. *ChemComm*, 49(45):5106–5131, 2013.

[40] Kathryn L. Haas and Katherine J. Franz. Application of Metal Coordination Chemistry To Explore and Manipulate Cell Biology. *Chem. Rev.*, 109(10):4921–4960, October 2009.

[41] H Richter and J. B Howard. Formation of polycyclic aromatic hydrocarbons and their growth to soot—a review of chemical reaction pathways. *Prog. Energy Combust. Sci.*, 26(4):565–608, August 2000.

[42] Barry Dellinger, Slawomir Lomnicki, Lavrent Khachatryan, Zofia Maskos, Randall W. Hall, Julien Adounkpe, Cheri McFerrin, and Hieu Truong. Formation and stabilization of persistent free radicals. *P. Combust. Inst.*, 31(1):521–528, January 2007.

[43] Jay Kochi. *Organometallic Mechanisms and Catalysis: The Role of Reactive Intermediates in Organic Processes*. Elsevier, December 2012.

[44] Muhammad Z. Iqbal, Marios S. Katsiotis, Saeed M. Alhassan, Matthew W. Liberatore, and Ahmed A. Abdala. Effect of solvent on the uncatalyzed synthesis of aminosilane-functionalized graphene. *RSC Adv.*, 4(13):6830–6839, 2014.

[45] Virender Kumar, Kulwinder Singh, Akshay Kumar, Manjeet Kumar, Karamjit Singh, Ankush Vij, and Anup Thakur. Effect of solvent on crystallographic, morphological and optical properties of SnO 2 nanoparticles. *Mater. Res. Bull.*, 85:202–208, January 2017.

[46] C. Cativiela, J. I. García, J. A. Mayoral, and L. Salvatella. Modelling of solvent effects on the Diels–Alder reaction. *Chem. Soc. Rev.*, 25(3):209–218, 1996.

[47] Marc-Antoni Goulet, Liuchuan Tong, Daniel A Pollack, Daniel P Tabor, Susan A Odom, Alán Aspuru-Guzik, Eugene E Kwan, Roy G Gordon, and Michael J Aziz. Extending the lifetime of organic flow batteries via redox state management. *J. Am. Chem. Soc.*, 141(20):8014–8019, 2019.

[48] Xin-Bing Cheng, Rui Zhang, Chen-Zi Zhao, Fei Wei, Ji-Guang Zhang, and Qiang Zhang. A review of solid electrolyte interphases on lithium metal anode. *Adv. Sci.*, 3(3):1500213, 2016.

[49] Saswata Dasgupta and John M. Herbert. Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3. *J. Comput. Chem.*, 38(12):869–882, 2017.

[50] Frank Jensen. Atomic orbital basis sets. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 3(3):273–295, 2013.

[51] Thierry Leininger, Hermann Stoll, Hans-Joachim Werner, and Andreas Savin. Combining long-range configuration interaction with short-range density functionals. *Chem. Phys. Lett.*, 275(3-4):151–160, August 1997.

[52] Axel D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98(7):5648–5652, April 1993.

[53] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.*, 98(45):11623–11627, November 1994.

[54] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.*, 393(1-3):51–57, July 2004.

[55] Narbe Mardirossian and Martin Head-Gordon. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.*, 16(21):9904–9924, 2014.

[56] Yan Zhao and Donald G Truhlar. The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theor. Chem. Acc.*, 120(1-3):215–241, 2008.

[57] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.*, 32(7):1456–1465, 2011.

[58] Oleg A. Vydrov and Troy Van Voorhis. Nonlocal van der Waals density functional: The simpler the better. *J. Chem. Phys.*, 133(24):244103, December 2010.

[59] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7(18):3297–3305, August 2005.

[60] Dmitrij Rappoport and Filipp Furche. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.*, 133(13):134105, October 2010.

[61] J. Grant Hill. Gaussian basis sets for molecular applications. *International Journal of Quantum Chemistry*, 113(1):21–34, 2013.

[62] A. Klamt and G. Schüürmann. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. II*, 0(5):799–805, 1993.

[63] Jacopo Tomasi. Selected features of the polarizable continuum model for the representation of solvation. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 1(5):855–867, 2011.

[64] Aleksandr V. Marenich, Christopher J. Cramer, and Donald G. Truhlar. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B*, 113(18):6378–6396, May 2009.

[65] Paul Winget, Derek M Dolney, David J Giesen, Christopher J Cramer, and Donald G Truhlar. Minnesota Solvent Descriptor Database. 1999.

[66] Sergio Trasatti. The absolute electrode potential: an explanatory note. *Pure Appl. Chem*, 58(7):955–966, 1986.

[67] Robert S Mulliken. Electronic population analysis on lcao–mo molecular wave functions. i. *J. Chem. Phys.*, 23(10):1833–1840, 1955.

[68] R L Wang, Claudia Buhrmester, and J R Dahn. Calculations of Oxidation Potentials of Redox Shuttle Additives for Li-Ion Cells. *J. Electrochem. Soc.*, page 6, 2006.

[69] Jian Xia, Remi Petibon, A. Xiao, W. M. Lamanna, and J. R. Dahn. Some Fluorinated Carbonates as Electrolyte Additives for Li(Ni $_{0.4}$ Mn $_{0.4}$ Co $_{0.2}$)O $_2$ /Graphite Pouch Cells. *J. Electrochem. Soc.*, 163(8):A1637–A1645, 2016.

[70] Gerhard H. Wrodnigg. Ethylene Sulfite as Electrolyte Additive for Lithium-Ion Cells with Graphitic Anodes. *J. Electrochem. Soc.*, 146(2):470, 1999.

[71] Narbe Mardirossian and Martin Head-Gordon. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.*, 115(19):2315–2372, October 2017.

[72] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.*, 68:314–319, February 2013.

[73] Anubhav Jain, Yongwoo Shin, and Kristin A. Persson. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.*, 1(1):1–13, January 2016.

[74] Xiaohui Qu, Anubhav Jain, Nav Nidhi Rajput, Lei Cheng, Yong Zhang, Shyue Ping Ong, Miriam Brafman, Edward Maginn, Larry A Curtiss, and Kristin A Persson. The electrolyte genome project: A big data approach in battery materials discovery. *Comput. Mater. Sci.*, 103:56–67, 2015.

[75] Troy Van Voorhis and Martin Head-Gordon. A geometric approach to direct minimization. *Mol. Phys.*, 100(11):1713–1721, June 2002.

[76] Péter Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.*, 73(2):393–398, July 1980.

[77] P. Pulay. Improved SCF convergence acceleration. *J. Comput. Chem.*, 3(4):556–560, 1982.

[78] H. Bernhard Schlegel. Geometry optimization. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 1(5):790–809, 2011.

[79] David S. Hall, Julian Self, and J. R. Dahn. Dielectric Constants for Quantum Chemistry and Li-Ion Batteries: Solvent Blends of Ethylene Carbonate and Ethyl Methyl Carbonate. *J. Phys. Chem. C*, 119(39):22322–22330, October 2015.

[80] James J. P. Stewart. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.*, 13(12):1173–1213, December 2007.

[81] Michael Gaus, Qiang Cui, and Marcus Elstner. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.*, 7(4):931–948, April 2011.

[82] Stefan Grimme, Christoph Bannwarth, and Philip Shushkov. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86). *J. Chem. Theory Comput.*, 13(5):1989–2009, May 2017.

[83] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.*, 15(3):1652–1671, March 2019.

[84] Tingzheng Hou, Guang Yang, Nav Nidhi Rajput, Julian Self, Sang-Won Park, Jagjit Nanda, and Kristin A. Persson. The influence of FEC on the solvation structure and reduction reaction of LiPF6/EC electrolytes and its implication for solid electrolyte interphase formation. *Nano Energy*, 64:103881, October 2019.

[85] Mathew J. Boyer and Gyeong S. Hwang. Theoretical Prediction of the Strong Solvent Effect on Reduced Ethylene Carbonate Ring-Opening and Its Impact on Solid Electrolyte Interphase Evolution. *J. Phys. Chem. C*, 123(29):17695–17702, July 2019.

[86] Kevin Leung. Two-electron reduction of ethylene carbonate: a quantum chemistry re-examination of mechanisms. *Chem. Phys. Lett.*, 568-569:1–8, May 2013.

[87] Tao Song and P. Hu. Insight into the solvent effect: A density functional theory study of cisplatin hydrolysis. *J. Chem. Phys.*, 125(9):091101, September 2006.