

Article

An Efficient Approach to Large-Scale Ab Initio Conformational Energy Profiles of Small Molecules

Yanxing Wang , Brandon Duane Walker , Chengwen Liu  and Pengyu Ren * 

Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712, USA

* Correspondence: pren@utexas.edu

Abstract: Accurate conformational energetics of molecules are of great significance to understand many chemical properties. They are also fundamental for high-quality parameterization of force fields. Traditionally, accurate conformational profiles are obtained with density functional theory (DFT) methods. However, obtaining a reliable energy profile can be time-consuming when the molecular sizes are relatively large or when there are many molecules of interest. Furthermore, incorporation of data-driven deep learning methods into force field development has great requirements for high-quality geometry and energy data. To this end, we compared several possible alternatives to the traditional DFT methods for conformational scans, including the semi-empirical method GFN2-xTB and the neural network potential ANI-2x. It was found that a sequential protocol of geometry optimization with the semi-empirical method and single-point energy calculation with high-level DFT methods can provide satisfactory conformational energy profiles hundreds of times faster in terms of optimization.

Keywords: conformational energy profile; computational efficiency; semi-empirical method; neural network potential; AMOEBA force field



Citation: Wang, Y.; Walker, B.D.; Liu, C.; Ren, P. An Efficient Approach to Large-Scale Ab Initio Conformational Energy Profiles of Small Molecules. *Molecules* **2022**, *27*, 8567. <https://doi.org/10.3390/molecules27238567>

Academic Editor: M. Natália D.S. Cordeiro

Received: 29 October 2022

Accepted: 27 November 2022

Published: 5 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Conformational energy was first developed to describe the ring strain energy of hydrocarbon rings, and has been generalized to include energy coming from bond distortions, angle strain, torsional strain, etc., enabling it to describe overall deviations of molecular geometry from the ideal [1,2]. If limited to systems of biological interest, e.g., ligand-receptor binding, the conformational energy is mostly attributed to the torsional strain from bond rotations. Attaining accurate conformational energetics of molecules is of great significance to calculate and understand many important molecular properties, such as molecular dipole moment, binding affinity, etc. [3,4]. Traditionally, quantum mechanical (QM) methods are applied to obtain the conformational energy landscape of a molecule, but expensive computational cost makes direct usage only suitable for small systems.

Molecular mechanics (MM) can be applied to large biologically interesting systems by using classical potential energy functions instead of solving the Schrödinger equation [5–7]. The core of MM methods is utilizing a set of equations and parameters, also known as force fields (FFs), to model molecular interactions and potential energy surfaces. Typically, the parameters of FFs are derived by fitting to experimental and/or QM data, based on predefined rules of transferability [8–13]. This process needs a medium to large amount of QM calculations to obtain the conformational energetics of small model compounds. Nevertheless, some parameters, e.g., torsion parameters, are sensitive to the local chemical environment and, hence, are not as transferable [14,15]. As a consequence, to attain simulations of quality, researchers have to parameterize novel molecules individually with ab initio QM data, which is time-consuming, but can be made more efficient with automation tools [16–19], such as Poltype [18,19] for the atomic multipole optimized energetics for

biomolecular simulation (AMOEBA) force field [20], the force field builder in the commercial Schrödinger software suite [21] for the optimized potentials for liquid simulations (OPLS) force field [22,23], and so on. With so much QM data to compute, researchers tend to use “poor” transferability rules for torsion to allow less QM. Thus, it becomes valuable to find more efficient alternatives to increase the scalability of computing QM for conformational energy surfaces.

As the data-driven deep learning methods are being incorporated more with FF development and molecular simulations [24–30], there appears to be a huge demand for giant ab initio QM datasets for training [24,31–34]. Neural network potentials (NNPs) could learn the conformational energy landscape accurately, but require vast amounts of data to achieve the desired performance. To deal with expensive dataset generation, for example, the ANI potential utilizes active-learning to intelligently let the model choose what data it wants for improvements. The dataset contains millions of DFT energy points that have been reduced, and ANI potential’s performance was further refined [25,35]. The difficulty of generating such a large dataset impedes the evaluation and application of deep learning models. Researchers are unable to generate their own data set to tailor deep learning models based on their individual needs, such as specific levels of accuracy of QM energy, hybrid models of NNP and FF, etc.

To the best of our knowledge, there are few published articles on the comparison of alternatives to speed up conformational scans without significant loss of accuracy. There has been some work on using NNPs to accelerate conformational scans and refine the torsion parameters for the GAFF2 FF [27]. However, one might not want to use the data generated by another NNP to train their own NNPs. Furthermore, we anticipated that semi-empirical methods [36,37] could be another alternative with accuracy in between DFT and MM methods. Thereby, in this work, we benchmarked the quality of the geometry and energy obtained with AMOEBA, an FF method, GFN2-xTB, a semi-empirical method, ANI-2x, an NNP method, and ω B97XD [38,39], a DFT method in ab initio conformational energy scan settings with a fragment dataset compiled by us, with both good drug-likeness and chemistry coverage considered.

2. Results and Discussion

2.1. Our Dataset Has a Broad Coverage of Chemical Space

The dataset contains a total of 233 fragments with a broad range of sizes (Figure 1A), with the number of total atoms varying from ~ 5 to ~ 40 covering chemical environments of different functional groups. We also analyzed the composing elements of the center atoms (i.e., atoms b,c in torsion $a-b-c-d$) and the neighbors (i.e., atoms a,d). Here, we only considered the elements of biochemical interests, namely C, N, O, P, S, F, Cl, Br, I, and H, of which only the first five are not monovalent; therefore, only they can serve as center atoms in a torsion bond. Our dataset covers the majority of the chemistry defined by this rule (Figure 1B), except those that are unstable or that are rarely found in biology-related compounds, e.g., P-P, P-S, N-P, O-O, N-O, etc. The N-O bond is actually very common, such as in nitro groups, but it appears as terminal bond that cannot serve as a center bond (bond b,c in torsion $a-b-c-d$). Our dataset actually includes nitro groups, as shown in Figure 1C. This figure shows the number of occurrences of the elements of neighbor atoms grouped by the element of the center atom, which illustrates the diversity of the local environments of the torsion bonds. Nitro groups are represented in the point under the center atom N at neighboring atom O. The vacancies with no point suggest bonding environments rarely found in biological systems that were excluded from our dataset, such as hetero atoms neighbored by halogens, etc. The structures of typical fragments randomly sampled from our dataset are presented in Figure 2.

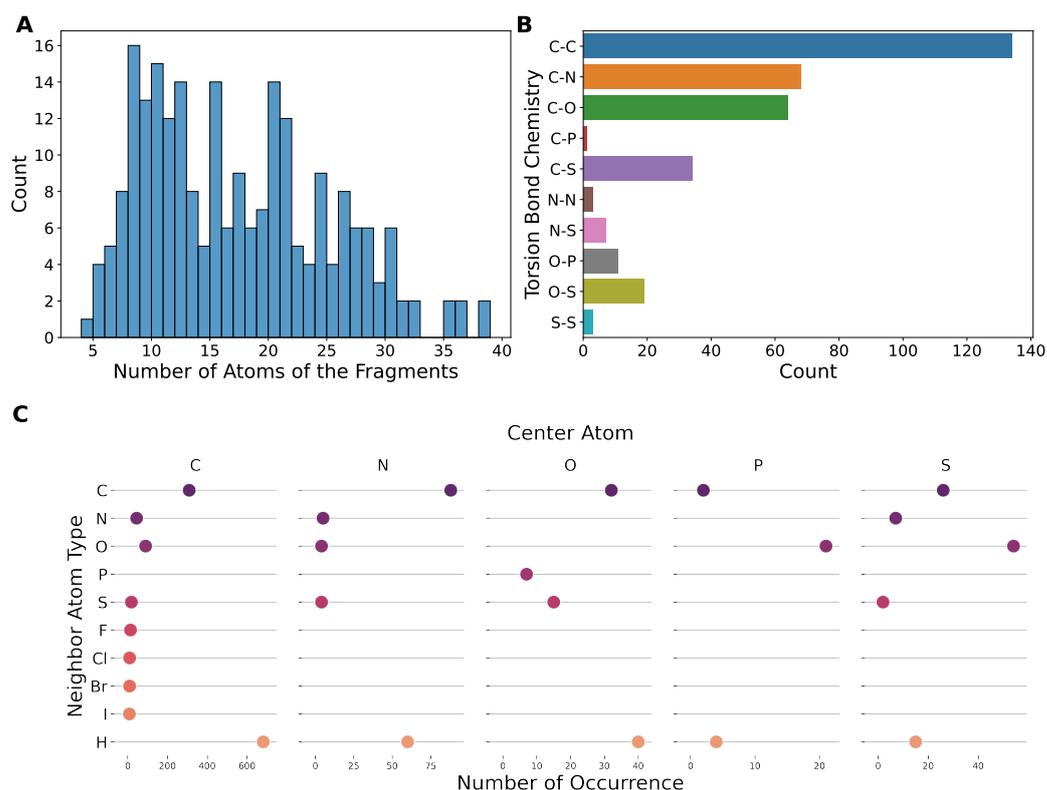


Figure 1. The torsion chemistry coverage of the dataset compiled in this work. **(A)** The histogram of the number of atoms of the fragments in the dataset. **(B)** The chemistry coverage measured by the elements of center atoms. **(C)** The chemistry coverage measured by the elements of neighbor atoms. The categories without a point mean that no such data exists in the dataset.

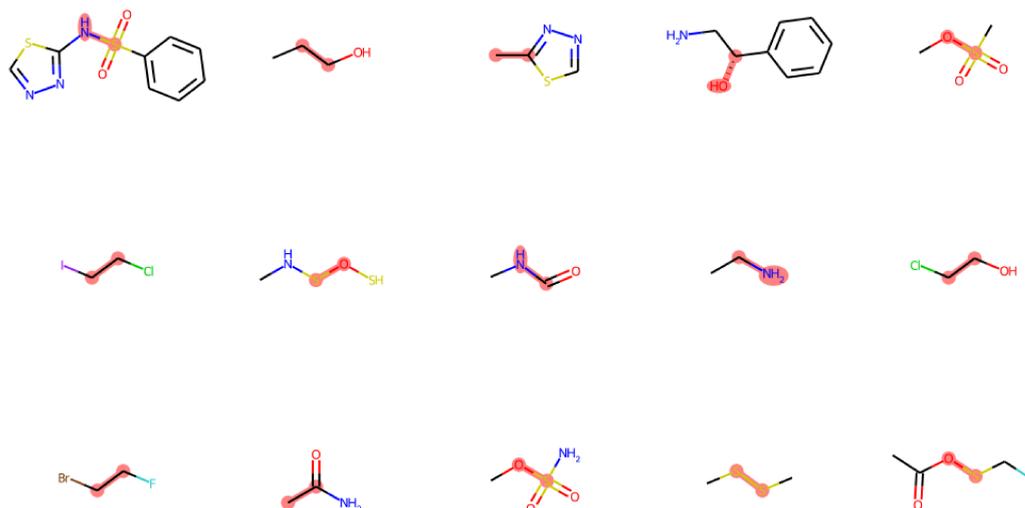


Figure 2. Typical fragments randomly selected from the dataset. The torsion bond and the center atoms are highlighted in red.

2.2. Cheaper Optimization Methods Can Also Provide Acceptable Geometry

We then benchmarked the three methods on this dataset with the scenario of conformational energy scans, assuming that the DFT results are the ground truth. First, we assessed the geometry obtained from the constrained optimization with AMOEBA, ANI, xtb, and ω B97XD/6-311G. Unlike DFT methods that remove the corresponding degrees of freedom for the constraints, MM methods usually restrain the degrees of freedom by adding extra harmonic potentials. Therefore, the constrained torsion dihedrals remain al-

most the same as the targets after ω B97XD optimization, but could change slightly after the AMOEBA, ANI, and xtb optimizations (Figure 3A). The reason is that the energy increase due to the unsatisfied restraints could be compensated for by the other interactions in the molecule. However, acceptable results where the overall deviation is within 1 degree can be achieved by setting sufficient force constants that control the steepness of the added potential wells on the constrained dihedrals. The optimized geometries were subsequently compared with the ω B97XD-optimized one (Figure 3B) by all-atom RMSD. The majority of structures showed an RMSD of less than 0.2 Å, with a few extreme RMSDs of up to ~ 1 Å. In particular, xtb yielded the best agreement with ω B97XD results, suggested by the lowest mean RMSD (the bar inside the violin), although the torsion angle deviation was not as good as other methods. Generally speaking, all of the three MM methods seemed able to provide acceptable geometry for conformational scans.

Additionally, it should be noted that ANI-2x can only be applied to molecules containing C, H, O, N, S, F, and Cl elements, leading to its incapability of handling many small molecules, e.g., fragments with P, Br, and I in our dataset. We also found that ANI often fails the optimization due to convergence issues. Perhaps the reason for this is due to the optimization engine—atomic simulation environment (ASE)—that ANI employed. This further hinders researchers from utilizing ANI for generating accurate QM datasets. In contrast, the xTB method is capable of handling molecules across the periodic table with up to thousands of atoms [37]. Researchers need not bother with the chemistry coverage and scalability issues.

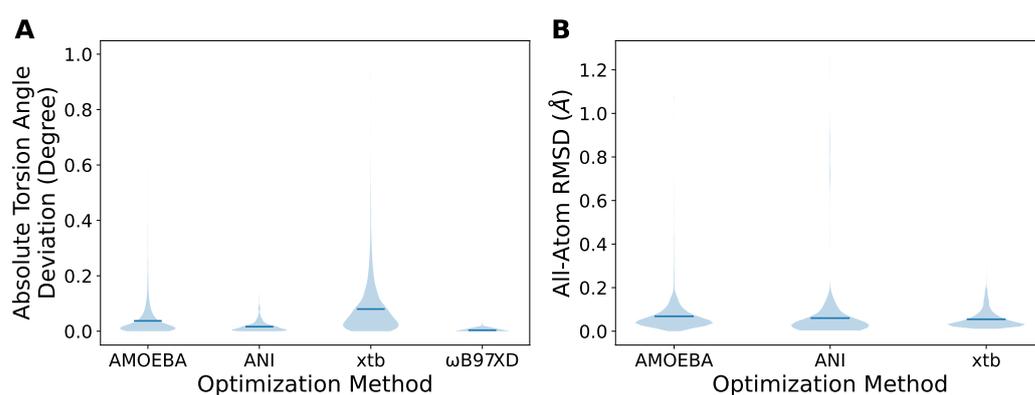


Figure 3. The analysis of the geometry obtained from the constrained optimizations. (A) The deviations of the torsion dihedral angles from the target set in the constraints. (B) The RMSD of the geometry obtained from the indicated optimization methods compared to the ω B97XD-optimized one. The bars in the violins represent the mean values.

2.3. DFT Method Is Still Necessary to Obtain Satisfactory Torsion Energy Profile

We further compared the deviation of the energies obtained through the various combinations of geometry optimization methods and single-point energy methods to the energies from DFT for both, i.e., ω B97XD/6-311G optimization followed by ω B97XD/6-311+G* single-point energy, shortened as ω B97XD- ω B97XD. This notation, *opt-sp*, is also used in the legend of Figure 4. As seen, the energies from xtb-xtb and ANI-ANI were in poor agreement with those from ω B97XD- ω B97XD, as suggested by the overall root mean squared error (RMSE) of ~ 1 kcal/mol (Figure 4A) and the 95% percentile of >2 kcal/mol (Figure 4B). The ANI-2x paper actually reported similar numbers for the comparison between ANI-ANI and DFT-DFT [25]. The overall RMSE of ANI- ω B97XD was similar to ANI-ANI but its 95% percentile, ~ 1 kcal/mol, was better than ANI-ANI, suggesting that the unsatisfactory RMSE might be attributed to the existence of the outliers (Figure 4B). Nevertheless, what excited us most is that xtb- ω B97XD demonstrates an excellent alignment to the reference, ω B97XD- ω B97XD, as indicated by the overall RMSE of 0.41 kcal/mol and the 95% percentile of 0.62 kcal/mol. The AMOEBA- ω B97XD was also included here as the baseline

FF method. These results aligned well with the geometry deviation mentioned before (Figure 3), which gave us an idea of how much difference the subtle deviation in geometry could make to the torsion energy profile.

In MD simulations of ambient conditions, molecular structures fluctuate around the equilibrium conformation. Thus, it is more relevant and important to model the energy profile accurately close to the equilibrium geometry. To this end, we included RMSEs for subsets of data points with different cutoffs on DFT energies (Figure 4A). As expected, all methods performed better with decreasing of DFT energy, i.e., coming closer and closer to the equilibrium. Even for xtb- ω B97XD, the RMSE for the subset with lower DFT energy was slightly smaller than the overall RMSE. The RMSE on data points with DFT energy no greater than 2 kcal/mol is lower than 0.30 kcal/mol. In addition, we also included the Boltzmann-weighted overall RMSEs at temperatures of 300 K and 1000 K (Figure 4A) and the histograms of Boltzmann-weighted absolute deviations at 1000 K (Figure 4C). The Boltzmann-weighted RMSE and 95% percentile xtb- ω B97XD are 0.27 kcal/mol and 0.39 kcal/mol, respectively. These results were consistent with that the xTB methods were parameterized to provide good geometries [37].

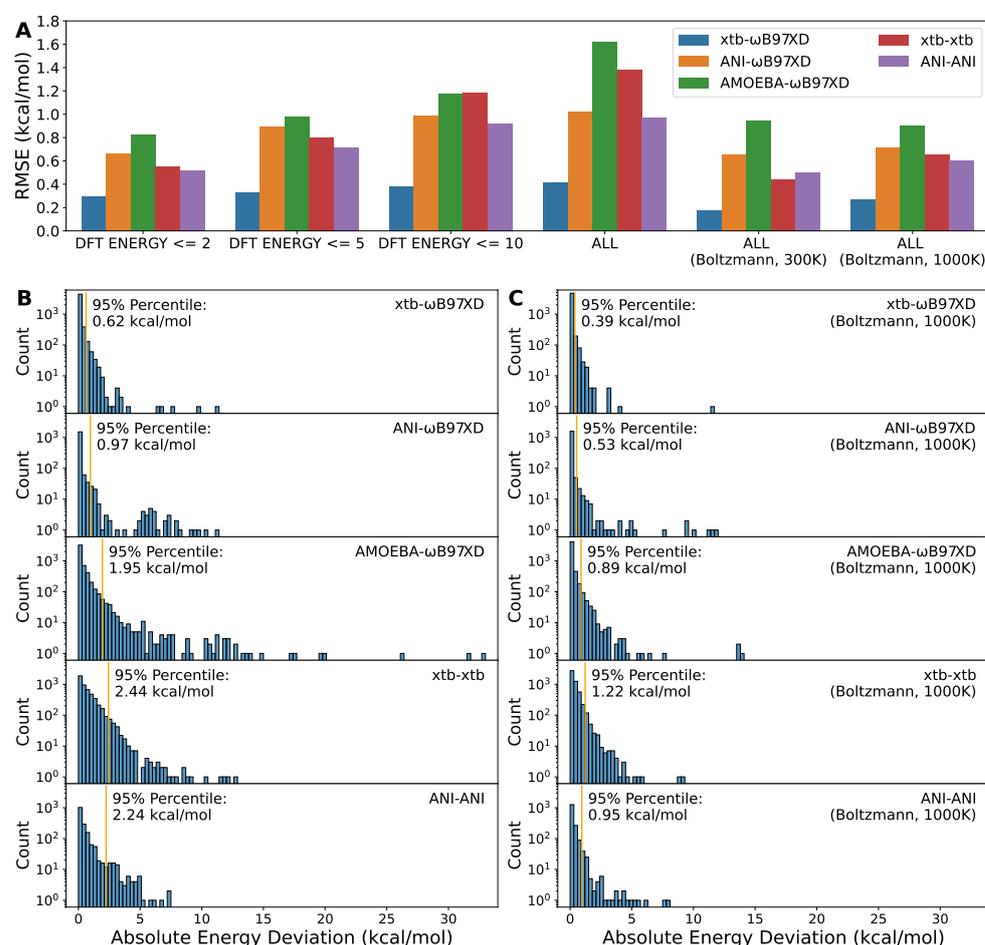


Figure 4. The comparison of energies obtained from various combination of optimization and single-point energy methods. (A) The root mean squared error of the indicated methods to DFT methods, ω B97XD- ω B97XD. This refers to ω B97XD/6-311G optimization followed by ω B97XD/6-311+G* single-point energy. The same notation convention applies to others in the legends. (B) The histograms of absolute errors for the indicated methods. (C) The histograms of Boltzmann-weighted absolute errors at 1000 K for the indicated methods. The orange vertical lines represent the 95% percentiles. The y axes were transformed to log scale to highlight outliers.

2.4. Inexpensive Computational Cost Highlights the Advantage of the Semi-Empirical Method xtb

When it comes to large scale conformational energy scans, time efficiency becomes non-trivial and as important as the accuracy of energy landscapes. Based on what was discussed above, it is tempting to replace the DFT optimization with xtb but for now, the DFT method should be kept for the single-point energy calculations to ensure satisfactory energy profiles. We found that the CPU times consumed for the DFT optimization of each torsion point were mainly distributed under 100 s per iteration (Figure 5A) and under 2000 s per molecule (Figure 5B). However, for the semi-empirical method xtb, the CPU time costs were mostly less than 1 s per iteration and less than 20 s per molecule. We next calculated the ratio and found that for most of fragments in our dataset, xtb optimization is faster than DFT by a median factor of ~ 72 with the majority lying within a factor of 10–1000 (Figure 5C). We also determined the elapsed times for optimization with ANI (Figure S1), which mainly lie in the range of 3.5–6.0 s per molecule. ANI was slightly faster than xtb. Since ANI does not provide an official timer, it was difficult to time each iteration.

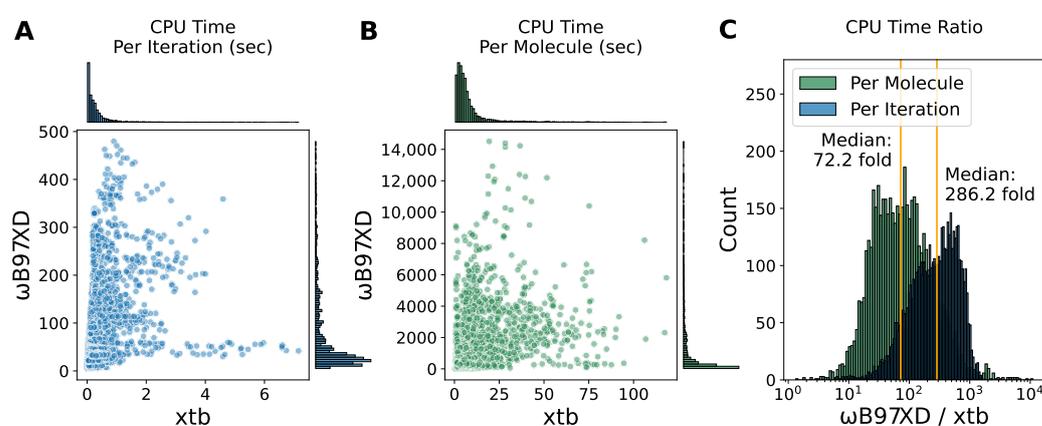


Figure 5. The computational cost comparison between xtb and ω B97XD optimization. (A,B) The scatter plot of the CPU times elapsed per iteration (A) and per molecule (B) for ω B97XD optimization versus xtb optimization. (C) The histogram of the ratio of CPU times for ω B97XD optimization to xtb optimization. The orange vertical line represents the median.

2.5. Case Discussion

Along with the overall comparison, herein, we show some of the poor cases we visually inspected. Often, the fragments containing sulfur atoms could be challenges for the methods benchmarked in this work. Figure 6A demonstrates a fragment where ANI and xtb could provide good geometry, but not good energy. Figure 6B shows an example where ANI was outperformed by xtb since ANI could not generate even a reasonable geometry, as indicated by the up and down of the ANI- ω B97XD curve, but the ANI-ANI curve suggests that ANI energy could somehow compensate for the bad geometry. In Figure 6C,D, we present two examples where xtb- ω B97XD exhibited relatively bad RMSEs (1.17 and 0.59 kcal/mol, respectively), but still was able to keep the overall shape of the torsion energy surfaces. Usually, the performance of NNPs is highly dependent on the training data. Since the dataset of ANI-2x was not publicized, we could not do much analysis. However, according to the description of the dataset creation in the ANI-2x paper, the molecules containing sulfur atoms were generated by simply substituting oxygen atoms. We assume that some functional groups containing sulfur atoms were probably not covered, such as sulfonic groups. This could be the reason why ANI-2x behaved poorly on these molecules.

We also found that fragments containing more than one phosphoric acid groups were non-trivial with respect to optimization. On one hand, the P-O-P bond in the triphosphoric acid was flattened during optimization at the level of ω B97XD/6-311G (Figure 7A). On the other hand, a non-negligible number of points failed to converge at that level of theory

with Psi4. Finally, we managed to optimize with Gaussian09 at the level of PBE/6-311G to obtain geometry of reasonable quality. However, the torsion profiles obtained by xtb diverged from those from DFT (Figure 7B). This problem might arise from the strong interactions between the phosphoric acid groups, making it hard to reach the minima without a polarizable continuum model in vacuum.

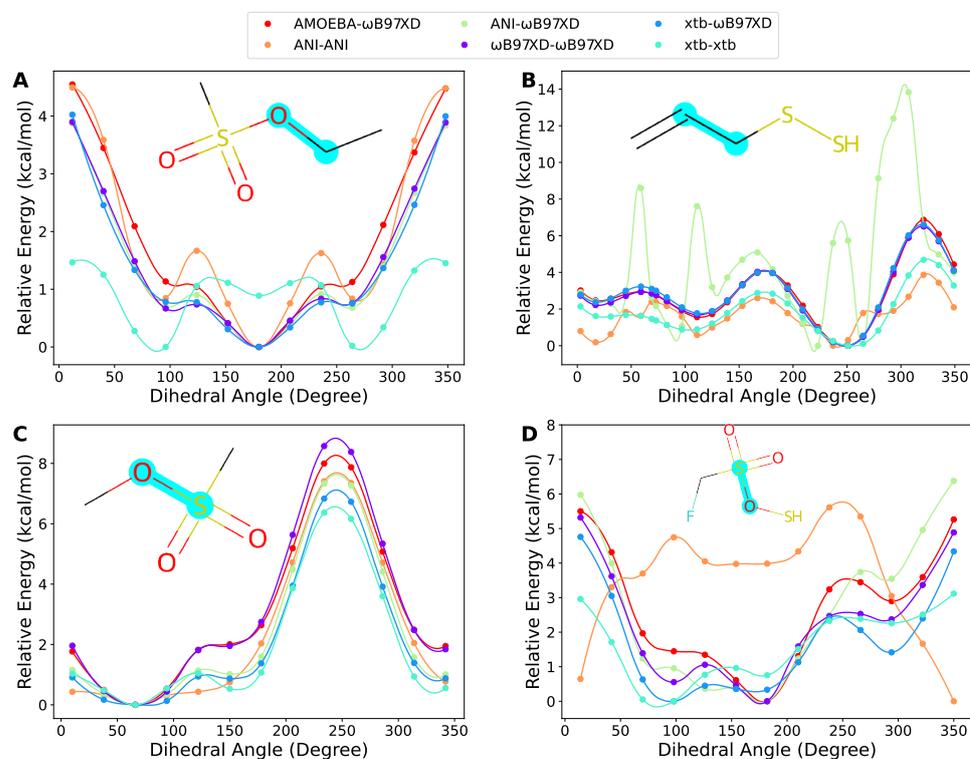


Figure 6. Four tricky cases containing sulfur atoms. (A) A fragment where ANI and xtb could provide good geometry, but not good energy. (B) An example where ANI was outperformed by xtb since ANI could not generate reasonable geometry. (C,D) Two examples where xtb- ω B97XD exhibited relatively bad RMSEs, but still was able to keep the overall shape of the torsion energy surfaces. The scanned torsion bonds are highlighted in cyan. The same notation as in Figure 4 is used in the legends.

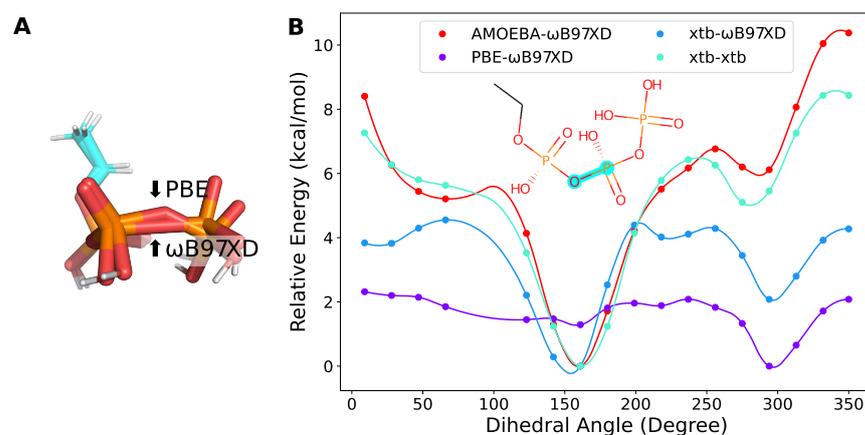


Figure 7. The torsion profile of the triphosphate fragment. (A) The optimized structures at the level of PBE/6-311G or ω B97XD/6-311G. (B) The torsion profile of the triphosphate fragment obtained with different methods. The scanned torsion bonds are highlighted in cyan. The same notation as in Figure 4 is used in the legends.

3. Materials and Methods

3.1. Dataset

We compiled the dataset used in this work by emphasizing both drug-likeness and chemistry coverage. We started with ~ 30 drug molecules randomly selected from the DrugBank database [40], which were cut into fragments with Poltype 2. The fragmenter in Poltype 2 ensures that the electron density of the fragment close to the innermost rotatable bond is similar to the electron density of the parent molecule for the same bond, which reduces the DFT computation cost without sacrificing the accuracy of torsion energy scans of that bond. These drug molecules provided us ~ 100 unique fragments, of which each has one or more torsions that were scanned. We further expanded our dataset by the addition of ~ 100 small organic molecules retrieved from PubChem [41] based on the torsion chemistry uncovered by the drug fragments. Our final dataset was composed of 233 fragments with 344 torsions, which covers 124 unique torsions defined by the elements of the four atoms of the torsion angle $a-b-c-d$. The dataset can be found in Table S1.

3.2. Computational Details

In this work, we focused on the comparison between DFT, semi-empirical, and NNP methods. To this end, we chose a representative method for each of the three categories, namely, GFN2-xTB, ω B97XD, and ANI-2x. GFN2-xTB was published in 2019 as a variant to the tight-binding DFTB3 [42,43] scheme, which includes anisotropic second-order density fluctuation effects via short-range damped interactions of cumulative atomic multipole moments. GFN2-xTB was designed to account for properties around the energetic minimum, e.g., geometries. There have been several studies [36,44] demonstrating that the theoretically sophisticated GFN2-xTB performs better than or comparable with other semi-empirical methods such as AM1 [45], PM3 [46], DFTB3, etc. The ω B97XD functional showed overall better performance than most of the 200 density functionals, including PBE0-D3(BJ) [47,48], PBE-D3(BJ) [48,49], and M06-2X [50], according to the benchmark study published by Head-Gordon and coworkers in 2017 [38]. There have been other publications benchmarking [51] or discussing [52] the performance of ω B97XD. Overall, ω B97XD was considered to offer a good balance between computational cost and accuracy. ANI-2x is the most current version of ANI potentials, with the best performance and broadest chemistry coverage (including seven elements: C, H, O, N, S, F, and Cl).

The overall workflow is as follows. (i) Minimize the input structure at the MP2 [53]/6-31G* [54–59] level of theory to obtain a good initial geometry for the subsequent conformational energy scan. (ii) With all the other torsion dihedrals fixed, rotate the scanned torsion bond to generate structures along the dihedral angle at uniform intervals. The increment of the degree is dependent on the number of cosine terms used in torsion fitting (i.e., three terms in this work) and the number of other torsions around the scanned torsion, due to needing more points sampled than number of parameters being fit for AMOEBA. (iii) Minimize the structures with constraints on all torsions using AMOEBA, and then use the indicated optimization methods, i.e., xtb, ANI, or ω B97XD/6-311G [60–63]. The force constants for xtb were set as 5 kcal/mol/deg². The convergence criteria for Psi4 are default, at {Delta E: 1.00×10^{-4} au, MAX Force: 2.50×10^{-3} au, RMS Force: 1.70×10^{-3} au, MAX Disp: 1.00×10^{-2} au, RMS Disp: 6.70×10^{-3} au}. The convergence criteria for xtb are default, at “normal”, {Econv: 5×10^{-6} au, Gconv: 1×10^{-3} au/ α }. (iv) Calculate the single-point energies with the indicated methods, including xtb, ANI, and ω B97XD/6-311+G* [55,63–66].

GFN2-xTB was readily exploited with xtb software. All DFT calculations involved in this work were performed with Psi4 [67], except for those explicitly mentioned as using Gaussian09 [68]. ANI-2x was accessed by the TorchANI [69] Python module. For simplicity, xtb and ANI will be used to refer to GFN2-xTB and ANI-2x through the article, respectively.

3.3. Metrics

We mainly used RMSE and absolute error (AE) to analyze the results obtained from different methods. In typical torsion parameterization scenarios, the Boltzmann-weighted error is widely adopted as well, which emphasizes the energy deviation close to the equilibrium structures, since this conformational space is of great significance in equilibrium MD simulations. The AE and RMSE are given by

$$AE(i) = |E_i - E_i^{DFT}| \quad (1)$$

$$RMSE = \sqrt{\sum_{i=0}^N \frac{1}{N} AE(i)^2} \quad (2)$$

The Boltzmann-weighted versions, denoted as a superscript of B, are given by

$$AE^B(i) = \frac{N \exp(-\frac{E_i^{DFT}}{k_B T})}{\sum_{i=0}^N \exp(-\frac{E_i^{DFT}}{k_B T})} |E_i - E_i^{DFT}| \quad (3)$$

$$RMSE^B = \sqrt{\sum_{i=0}^N \frac{\exp(-\frac{E_i^{DFT}}{k_B T})}{\sum_{i=0}^N \exp(-\frac{E_i^{DFT}}{k_B T})} AE(i)^2} \quad (4)$$

4. Conclusions

Based on what was reported in this work, we would like to recommend the sequential protocol of optimization with GFN2-xTB and single-point energy at ω B97XD/6-311+G* as a promising alternative for large-scale conformational scans. With running time reduced by factors of hundreds, this protocol can provide geometry in excellent agreement with that obtained by ω B97XD/6-311G, and accurate DFT single-point energy. Other combinations of comparable semi-empirical methods and DFT methods may be able to provide similar results, but relevant benchmark studies are necessary to draw a general conclusion. It should also be noted that researchers need pay attention to the chemistry of their dataset, since certain functional groups can be tricky for all methods. We hope that this work can benefit researchers in need of large-scale conformational energy scans.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules27238567/s1>, Figure S1: The distribution of the elapsed times of optimization with ANI-2x and ASE Python module; Table S1: The SMILES of the fragments composing the dataset in this work.

Author Contributions: Conceptualization, P.R.; methodology, P.R.; software, Y.W., B.D.W., C.L.; validation, Y.W.; formal analysis, Y.W.; investigation, Y.W.; resources, P.R.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W., B.D.W., C.L., P.R.; visualization, Y.W.; supervision, P.R.; project administration, P.R.; funding acquisition, P.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Institutes of Health (R01GM106137 and R01GM114237) and the Welch Foundation (F-2120).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data not reported can be requested from the corresponding author.

Conflicts of Interest: Pengyu Ren is a co-founder of Qubit Pharmaceuticals.

References

1. Wiberg, K.B. The concept of strain in organic chemistry. *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 312–322. [[CrossRef](#)]
2. Gonthier, J.F.; Steinmann, S.N.; Wodrich, M.D.; Corminboeuf, C. Quantification of “fuzzy” chemical concepts: A computational perspective. *Chem. Soc. Rev.* **2012**, *41*, 4671–4687. [[CrossRef](#)] [[PubMed](#)]
3. Peach, M.L.; Cachau, R.E.; Nicklaus, M.C. Conformational energy range of ligands in protein crystal structures: The difficult quest for accurate understanding. *J. Mol. Recognit.* **2017**, *30*, e2618. [[CrossRef](#)] [[PubMed](#)]
4. Mazzanti, A.; Casarini, D. Recent trends in conformational analysis. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 613–641. [[CrossRef](#)]
5. Boyd, D.B.; Lipkowitz, K.B. Molecular mechanics: The method and its underlying philosophy. *J. Chem. Educ.* **1982**, *59*, 269. [[CrossRef](#)]
6. Engler, E.M.; Andose, J.D.; Schleyer, P.V. Critical evaluation of molecular mechanics. *J. Am. Chem. Soc.* **1973**, *95*, 8005–8025. [[CrossRef](#)]
7. Hehre, W.J. *A Guide to Molecular Mechanics and Quantum Chemical Calculations*; Wavefunction: Irvine, CA, USA, 2003; Volume 2.
8. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [[CrossRef](#)]
9. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690. [[CrossRef](#)]
10. Klauda, J.B.; Venable, R.M.; Freites, J.A.; O'Connor, J.W.; Tobias, D.J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A.D., Jr.; Pastor, R.W. Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843. [[CrossRef](#)]
11. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197. [[CrossRef](#)]
12. Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.P.; Ponder, J.W.; Ren, P. AMOEBA polarizable atomic multipole force field for nucleic acids. *J. Chem. Theory Comput.* **2018**, *14*, 2084–2108. [[CrossRef](#)] [[PubMed](#)]
13. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J.W.; Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063. [[CrossRef](#)] [[PubMed](#)]
14. Kania, A.; Sarapata, K.; Gucwa, M.; Wójcik-Augustyn, A. Optimal solution to the torsional coefficient fitting problem in force field parametrization. *J. Phys. Chem. A* **2021**, *125*, 2673–2681. [[CrossRef](#)]
15. Zgarbová, M.; Rosnik, A.M.; Luque, F.J.; Curutchet, C.; Jurečka, P. Transferability and additivity of dihedral parameters in polarizable and nonpolarizable empirical force fields. *J. Comput. Chem.* **2015**, *36*, 1874–1884. [[CrossRef](#)] [[PubMed](#)]
16. Huang, L.; Roux, B. Automated force field parameterization for nonpolarizable and polarizable atomic models based on ab initio target data. *J. Chem. Theory Comput.* **2013**, *9*, 3543–3556. [[CrossRef](#)] [[PubMed](#)]
17. Morado, J.; Mortenson, P.N.; Verdonk, M.L.; Ward, R.A.; Essex, J.W.; Skylaris, C.K. Paramol: A package for automatic parameterization of molecular mechanics force fields. *J. Chem. Inf. Model.* **2021**, *61*, 2026–2047. [[CrossRef](#)]
18. Wu, J.C.; Chatterjee, G.; Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131*, 1–11. [[CrossRef](#)]
19. Walker, B.; Liu, C.; Wait, E.; Ren, P. Automation of AMOEBA polarizable force field for small molecules: Poltype 2. *J. Comput. Chem.* **2022**, *43*, 1530–1542. [[CrossRef](#)]
20. Ponder, J.W.; Wu, C.; Ren, P.; Pande, V.S.; Chodera, J.D.; Schnieders, M.J.; Haque, I.; Mobley, D.L.; Lambrecht, D.S.; DiStasio, R.A., Jr.; et al. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564. [[CrossRef](#)]
21. *Schrödinger Release 2022-3: FEP+*; Schrödinger, LLC: New York, NY, USA, 2021.
22. Jorgensen, W.L.; Maxwell, D.S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
23. Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G.A.; Dahlgren, M.K.; Russell, E.; Von Bargen, C.D.; et al. OPLS4: Improving force field accuracy on challenging regimes of chemical space. *J. Chem. Theory Comput.* **2021**, *17*, 4291–4300. [[CrossRef](#)] [[PubMed](#)]
24. Smith, J.S.; Isayev, O.; Roitberg, A.E. ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192–3203. [[CrossRef](#)] [[PubMed](#)]
25. Devereux, C.; Smith, J.S.; Huddleston, K.K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A.E. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192–4202. [[CrossRef](#)] [[PubMed](#)]
26. Glick, Z.L.; Metcalf, D.P.; Koutsoukas, A.; Spronk, S.A.; Cheney, D.L.; Sherrill, C.D. AP-Net: An atomic-pairwise neural network for smooth and transferable interaction potentials. *J. Chem. Phys.* **2020**, *153*, 044112. [[CrossRef](#)]
27. Wang, X.; Xu, Y.; Zheng, H.; Yu, K. A Scalable Graph Neural Network Method for Developing an Accurate Force Field of Large Flexible Organic Molecules. *J. Phys. Chem. Lett.* **2021**, *12*, 7982–7987. [[CrossRef](#)]
28. Glick, Z.L.; Koutsoukas, A.; Cheney, D.L.; Sherrill, C.D. Cartesian message passing neural networks for directional properties: Fast and transferable atomic multipoles. *J. Chem. Phys.* **2021**, *154*, 224103. [[CrossRef](#)]

29. Kumar, A.; Pandey, P.; Chatterjee, P.; MacKerell, A.D., Jr. Deep Neural Network Model to Predict the Electrostatic Parameters in the Polarizable Classical Drude Oscillator Force Field. *J. Chem. Theory Comput.* **2022**, *18*, 1711–1725. [[CrossRef](#)]
30. Park, C.W.; Kornbluth, M.; Vandermause, J.; Wolverson, C.; Kozinsky, B.; Mailoa, J.P. Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. *NPJ Comput. Mater.* **2021**, *7*, 1–9. [[CrossRef](#)]
31. Smith, J.S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A.E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Sci. Data* **2020**, *7*, 1–10. [[CrossRef](#)]
32. Donchev, A.G.; Taube, A.G.; Decolvenaere, E.; Hargus, C.; McGibbon, R.T.; Law, K.H.; Gregersen, B.A.; Li, J.L.; Palmo, K.; Siva, K.; et al. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Sci. Data* **2021**, *8*, 1–9. [[CrossRef](#)]
33. Eastman, P.; Behara, P.K.; Dotson, D.L.; Galvelis, R.; Herr, J.E.; Horton, J.T.; Mao, Y.; Chodera, J.D.; Pritchard, B.P.; Wang, Y.; et al. SPICE, A Dataset of Drug-like Molecules and Peptides for Training Machine Learning Potentials. *arXiv* **2022**, arXiv:2209.10702.
34. Axelrod, S.; Gomez-Bombarelli, R. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Sci. Data* **2022**, *9*, 1–14. [[CrossRef](#)] [[PubMed](#)]
35. Smith, J.S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A.E. Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*, 241733. [[CrossRef](#)] [[PubMed](#)]
36. 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. [[CrossRef](#)] [[PubMed](#)]
37. Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, e1493. [[CrossRef](#)]
38. Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315–2372. [[CrossRef](#)]
39. Chai, J.D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620. [[CrossRef](#)] [[PubMed](#)]
40. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672. [[CrossRef](#)]
41. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395. [[CrossRef](#)]
42. Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260.
43. 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*, 931–948. [[CrossRef](#)] [[PubMed](#)]
44. Sharapa, D.I.; Genae, A.; Cavallo, L.; Minenkov, Y. A robust and cost-efficient scheme for accurate conformational energies of organic molecules. *ChemPhysChem* **2019**, *20*, 92–102. [[CrossRef](#)] [[PubMed](#)]
45. Dewar, M.J.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909. [[CrossRef](#)]
46. Stewart, J.J. Optimization of parameters for semiempirical methods II. Applications. *J. Comput. Chem.* **1989**, *10*, 221–264. [[CrossRef](#)]
47. Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170. [[CrossRef](#)]
48. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465. [[CrossRef](#)]
49. Perdew, J.P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865. [[CrossRef](#)]
50. Zhao, Y.; Truhlar, D.G. 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. Accounts* **2008**, *120*, 215–241.
51. Kesharwani, M.K.; Karton, A.; Martin, J.M. Benchmark ab initio conformational energies for the proteinogenic amino acids through explicitly correlated methods. Assessment of density functional methods. *J. Chem. Theory Comput.* **2016**, *12*, 444–454. [[CrossRef](#)]
52. Bursch, M.; Mewes, J.M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry. *Angew. Chem.* **2022**, *134*, e202205735. [[CrossRef](#)]
53. Møller, C.; Plesset, M.S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618. [[CrossRef](#)]
54. Ditchfield, R.; Hehre, W.J.; Pople, J.A. Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **1971**, *54*, 724–728. [[CrossRef](#)]
55. Francl, M.M.; Pietro, W.J.; Hehre, W.J.; Binkley, J.S.; Gordon, M.S.; DeFrees, D.J.; Pople, J.A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665. [[CrossRef](#)]
56. Gordon, M.S.; Binkley, J.S.; Pople, J.A.; Pietro, W.J.; Hehre, W.J. Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements. *J. Am. Chem. Soc.* **1982**, *104*, 2797–2803. [[CrossRef](#)]
57. Hariharan, P.C.; Pople, J.A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222. [[CrossRef](#)]

58. Hehre, W.J.; Ditchfield, R.; Pople, J.A. Self—Consistent molecular orbital methods. XII. Further extensions of Gaussian—Type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261. [[CrossRef](#)]
59. Rassolov, V.A.; Ratner, M.A.; Pople, J.A.; Redfern, P.C.; Curtiss, L.A. 6-31G* basis set for third-row atoms. *J. Comput. Chem.* **2001**, *22*, 976–984. [[CrossRef](#)]
60. Curtiss, L.A.; McGrath, M.P.; Blaudeau, J.P.; Davis, N.E.; Binning Jr, R.C.; Radom, L. Extension of Gaussian-2 theory to molecules containing third-row atoms Ga–Kr. *J. Chem. Phys.* **1995**, *103*, 6104–6113. [[CrossRef](#)]
61. Glukhovtsev, M.N.; Pross, A.; McGrath, M.P.; Radom, L. Extension of Gaussian-2 (G2) theory to bromine-and iodine-containing molecules: Use of effective core potentials. *J. Chem. Phys.* **1995**, *103*, 1878–1885. [[CrossRef](#)]
62. Frisch, M.J.; Pople, J.A.; Binkley, J.S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265–3269. [[CrossRef](#)]
63. Krishnan, R.; Binkley, J.S.; Seeger, R.; Pople, J.A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650–654. [[CrossRef](#)]
64. Clark, T.; Chandrasekhar, J.; Spitznagel, G.W.; Schleyer, P.V.R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+ G basis set for first-row elements, Li–F. *J. Comput. Chem.* **1983**, *4*, 294–301. [[CrossRef](#)]
65. McLean, A.; Chandler, G. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z= 11–18. *J. Chem. Phys.* **1980**, *72*, 5639–5648. [[CrossRef](#)]
66. Spitznagel, G.W.; Clark, T.; von Ragué Schleyer, P.; Hehre, W.J. An evaluation of the performance of diffuse function-augmented basis sets for second row elements, Na–Cl. *J. Comput. Chem.* **1987**, *8*, 1109–1116. [[CrossRef](#)]
67. Smith, D.G.; Burns, L.A.; Simmonett, A.C.; Parrish, R.M.; Schieber, M.C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A.; et al. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **2020**, *152*, 184108. [[CrossRef](#)] [[PubMed](#)]
68. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 09, Revision A. 02*; Gaussian, Inc.: Wallingford, CT, USA, 2009.
69. Gao, X.; Ramezanghorbani, F.; Isayev, O.; Smith, J.S.; Roitberg, A.E. TorchANI: A free and open source PyTorch-based deep learning implementation of the ANI neural network potentials. *J. Chem. Inf. Model.* **2020**, *60*, 3408–3415. [[CrossRef](#)]