JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

pubs.acs.org/jcim

Atom Type Independent Modeling of the Conformational Energy of Benzylic, Allylic, and Other Bonds Adjacent to Conjugated Systems

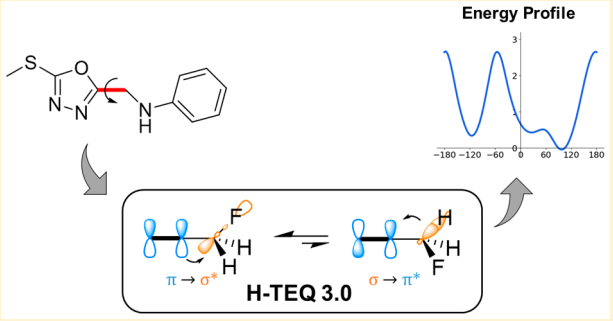
Candide Champion,[†] Stephen J. Barigye,[†,§] Wanlei Wei,[†] Zhaomin Liu,[†,∥] Paul Labute,[‡] and Nicolas Moitessier*[,†]

[†]Department of Chemistry, McGill University, 801 Sherbrooke Street W., Montréal, QC, Canada H3A 0B8

[‡]Chemical Computing Group Inc., 1010 Sherbrooke Street W., Montréal, QC, Canada H3A 2R7

Supporting Information

ABSTRACT: Applications of computational methods to predict binding affinities for protein/drug complexes are routinely used in structure-based drug discovery. Applications of these methods often rely on empirical force fields (FFs) and their associated parameter sets and atom types. However, it is widely accepted that FFs cannot accurately cover the entire chemical space of drug-like molecules, due to the restrictive cost of parametrization and the poor transferability of existing parameters. To address these limitations, initiatives have been carried out to develop more transferable methods, in order to allow for more rigorous descriptions of any drug-like molecule. We have previously reported H-TEQ, a method which does not rely on atom types and incorporates well established chemical principles to assign parameters to organic molecules. The previous implementation of H-TEQ (a torsional barrier prediction method) only covered saturated and lone pair containing molecules; here, we report our efforts to incorporate conjugated systems into our model. The next step was the evaluation of the introduction of unsaturations. The developed model (H-TEQ3.0) has been validated on a wide variety of molecules containing heteroaromatic groups, alkyls, and fused ring systems. Our method performs on par with one of the most commonly used FFs (GAFF2), without relying on atom types or any prior parametrization.



INTRODUCTION

Computational Methods in Drug Discovery and Molecular Mechanics. Computational methods are often quick and cost-effective complements to experiments to identify potential binders to targets of therapeutic interest and/or off-targets. Over the years, computational tools have contributed to many stages of the drug discovery process, from the prediction of drug-likeness[1] following, for example, Lipinski's rule of five or ADME (absorption, distribution, metabolism, and excretion) properties[2] using artificial intelligence (AI) or statistical analysis, to physics-based methods providing insights into the structure and dynamics of molecular systems.[3,4] It is anticipated that structure-based drug design (SBDD) will have even greater relevance in future drug discovery paradigms.[5,6] The accuracy of predicted drug binding affinities depends on several factors such as the scale of the structural model[7] (subatomic, atomic, coarse-grained), the accuracy of the energy potentials computed for molecular conformations,[8,9] as well as the degree to which all energetically accessible conformations are sampled.[10] In this context, quantum mechanical (QM) methods would provide a very accurate depiction of the energetics of molecular systems, allowing the rigorous estimation of ligand−macromolecule binding energies.[11] These methods cannot, however, be carried to high-throughput tasks, to scan large portions of conformational space, or to study large macromolecules, due to their restrictive computational costs. In light of this limitation, molecular mechanics (MM) methods have been developed to evaluate the energetics of molecular systems using simplified potentials with the objective of reproducing experimental data and QM potentials, while reducing computational costs by several orders of magnitude. However, the accuracy of these more empirical MM methods largely depends on the quality of the potentials and parameters of the underlying force fields (FFs).[12,13]

Atom-Type Based FFs. In MM, the potential energy of a molecular system (e.g., small molecules, proteins, nucleic acids, and complexes) is calculated using a FF corresponding to a set of potential energy functions and its associated precomputed parameters (eqs 1−7). The contributions from each term in an FF can be split into two categories, "bonded" interactions (bonds, angles, torsions, out-of-plane angles) which are calculated for atoms within the same molecule and "non-bonded" interactions (e.g., van der Waals and electrostatics) which are calculated for pairs of atoms separated by three or more bonds (intramolecular) or pairs of atoms in different molecules (intermolecular). It should be noted that the

Received: July 15, 2019
Published: October 7, 2019

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

interactions of atoms separated by three bonds are therefore of both types: torsions and nonbonded interactions.

$$E_{\text{total}} = \underbrace{E_{\text{bonds}} + E_{\text{angles}} + E_{\text{torsions}} + E_{\text{out-of-plane}}}_{\textbf{bonded}} + \underbrace{E_{\text{vdW}} + E_{\text{electrostatics}}}_{\textbf{non-bonded}} \quad (1)$$

$$E_{\text{bonds}} = \boldsymbol{K_r}(r - \boldsymbol{r_{\text{eq}}})^2 \quad (2)$$

$$E_{\text{angles}} = \boldsymbol{K_\theta}(\theta - \boldsymbol{\theta_{\text{eq}}})^2 \quad (3)$$

$$E_{\text{torsion}} = \sum_{n=1}^{N} \frac{\boldsymbol{V_n}}{2}[1 + \cos n(\varphi - \boldsymbol{\delta})] \quad (4)$$

$$E_{\text{out-of-plane}} = \boldsymbol{K_\omega}(\omega - \boldsymbol{\omega_{\text{eq}}})^2 \quad (5)$$

$$E_{\text{vdW}} = \sum_{\text{pairs}\,i,j} \boldsymbol{\varepsilon_{ij}}\left[\left(\frac{\boldsymbol{R_{\text{min},i,j}}}{r_{i,j}}\right)^{12} - \left(\frac{\boldsymbol{R_{\text{min},i,j}}}{r_{i,j}}\right)^{6}\right] \quad (6)$$

$$E_{\text{electrostatics}} = \sum_{\text{pairs}\,i,j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{i,j}} \quad (7)$$

Transferability. "Atom types" are central to most widely applied FFs in SBDD, such as the AMBER,[14,15] CHARMM,[16,17] GROMOS,[18,19] and OPLS[20−23] series. In the AMBER protein FF, for example, parameters for aromatic carbons (atom type CA) are different from aliphatic carbons (CT) or carbonyl carbon (C) and other carbon types. However, these definitions are limited to local environments and distant chemical functional groups do not affect the atom type (and set of parameters) assigned to particular moieties, which consequently ignores some electronic effects. For example, electron donating or withdrawing substituents adjacent to aromatic rings are not considered in ring atom types, whereas torsional energy profiles could differ when such substituents are present.

Parametrization of a FF consists in finding the ideal values for all the parameters (shown in bold) associated with each function (eqs 1−7). For example, the bond stretching term (eq 2) describes the ideal bond distance between two atoms and is characterized by an equilibrium bond length ($\boldsymbol{r_{\text{eq}}}$) and a force constant ($\boldsymbol{K_r}$). In order to describe all possible bond stretching events, parameters for the equilibrium value and force constants are required for all combinations of two atom types.[12] This parameter fitting process uses experimental (e.g., H NMR, thermodynamic properties) and/or high-level QM data as reference, which are costly to obtain, ultimately imposing a limit on the size of the training set used to develop parameters. FFs thus rely on the transferability of parameters obtained from molecules in the training set to other similar molecules. The current consensus is that no particular FF could accurately describe the energetics of all possible small drug-like molecules due to the sheer size of the chemical space, and the poor transferability of empirical parameters generated on specific molecules.[24] It is important to keep in mind that not all types of parameters are subject to this lack of transferability; for instance, bonds and angles are generally assumed to be fully covered. However, the authors of OPLS3.0 estimated in 2015 that 33% of drug-like molecules were missing at least one torsion parameter,[21] (a more recent version attempts to solve this limitation[25]) and the treatment of nonbonded interactions has also recently been challenged by the introduction of polarizable FFs (e.g., AMOEBA,[26,27] CHARMM-Drude[28,29]). Current developments in FF methodologies are hence highly focused toward torsional and nonbonded interactions.

To address the liabilities resulting from poor parameter transferability and/or missing parameters, researchers have followed two main approaches. On one hand, automated toolkits such as GAAMP,[30] ffTK,[31] Paramfit,[32] and Parmscan[33] have been developed, allowing to generate accurate parameters for specific molecules of interest from QM data. These user-friendly toolkits are particularly fit for researchers studying the interactions within a ligand/receptor pair using molecular dynamics (MD), since only few parameters need to be generated (usually for the drug-like molecule). However, these tools cannot be carried to high-throughput tasks (e.g., docking libraries of compounds), due to the computational costs associated with the parameter fitting process. While these toolkits allow parameters to be generated for specific studies, they do not solve the problem of parameter transferability. A radically different approach consists in developing MM methods with greater transferabilities without relying on the concept of atom types to determine parameters. To our knowledge, Mobley et al.'s recent attempt with SMIRNOFF,[34,35] a FF which uses direct chemical perception instead of traditional atom types to determine parameters, as well as H-TEQ (developed in our lab),[36,37] to the best of our knowledge the only methods moving away from the atom type paradigm of FFs. Both SMIRNOFF and H-TEQ were shown to perform comparably well to GAFF (one of the most widely used FFs for small organic molecules)[38] in reproducing liquid properties and QM torsional profiles, respectively. The performance of these methods has not yet been extensively tested in the context of SBDD relevant interactions, due to their very recent releases, and we expect further work to allow these methods to cover most (if not all) possible organic molecules with good accuracy. These efforts are highly encouraging in the future ability of atom type free FFs to rival state-of-the-art FFs toward SBDD applications, without requiring any molecule-specific parametrization.

While our previous versions of H-TEQ focused on saturated compounds, we report here our efforts to incorporate unsaturated compounds.

IMPACT OF UNSATURATIONS ON TORSIONAL ENERGY

Organic Chemistry Principles and Drug Conformational Energy. In organic chemistry, several models have been employed to rationalize the conformational preferences of molecules and stereoselectivity in chemical reactions. For example, the hyperconjugation model is often evoked to rationalize the preference of the staggered conformation in ethane and the anomeric effect in carbohydrate molecules.[39−41] Briefly, hyperconjugation is a stabilizing interaction involving the donation of electrons from a bonding (e.g., σ) to an antibonding (e.g., σ^*) orbital, leading to the formation of a new orbital lying lower in energy (Figure 1).[42] Two main factors influence the strength of hyperconjugation interactions.[43,44] First, a greater physical overlap between interacting orbitals leads to a stronger interaction. This overlap is maximized when σ and σ^* are in *anti* relationship (Figure 1a) rather than *syn* (Figure 1b) and is minimal when orbitals

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

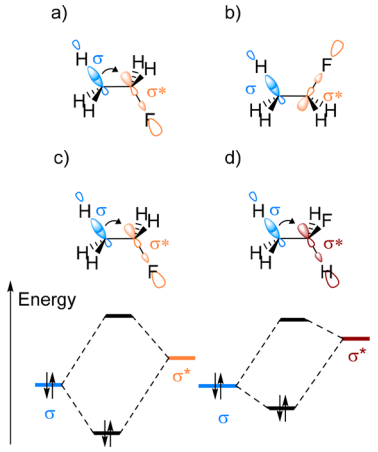


Figure 1. Factors influencing the strength of hyperconjugation interactions in the fluoroethane molecule. The orbital overlap is greater when $\sigma_{(C-H)}$ and $\sigma^*_{(C-F)}$ are anti (a) rather than syn (b). The energy gap between σ and σ^* is smaller between $\sigma_{(C-H)}$ and $\sigma^*_{(C-F)}$ (c) than $\sigma_{(C-H)}$ and $\sigma^*_{(C-H)}$ (d).

are perpendicular to one another. Second, a smaller gap between donating and accepting orbitals energy levels leads to a stronger interaction. For example, the electronegative fluorine in fluoroethane results in a lower lying $\sigma^*_{(C-F)}$ (Figure 1c) than $\sigma^*_{(C-H)}$ (Figure 1d) and reducing the energy gap ultimately favoring the conformation in which $\sigma_{(C-H)}$ and $\sigma^*_{(C-F)}$ (the best acceptor in this molecule) are *anti*.

Although qualitative in nature, these chemistry principles are highly transferable since they follow general principles such as the degree of electron donating or electron withdrawing character, presence of lone pairs and the degree of overlap of molecular orbitals. Indeed, we have demonstrated that if these principles are quantified (using simple atomic properties), universal models for computing the torsional energy of molecules could be developed.[36,37] While our previous studies were focused toward $\sigma \rightarrow \sigma^*$ and n $\rightarrow \sigma^*$ hyperconjugation modes, a large number of drug like molecules contain unsaturations and aromatic ring systems,[45] which exhibit other hyperconjugation modes: $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$, which we will refer to as **π-hyperconjugation**. These additional hyperconjugation interactions must play an important role in determining the conformational preferences of such moieties. Therefore, the goal of the present manuscript is to describe our progress in integrating π interaction modes into our H-TEQ method, to guarantee its applicability to torsions in any drug-like molecule, improving the accuracy of FF-based methods for SBDD applications. In this present work, conjugated systems are not considered, and we focus on the prediction of torsional parameters for $C(sp^2)-C(sp^3)$ bonds only.

Asymmetric Induction and π-Hyperconjugation. As predictive yet qualitative chemical principles, multiple chemical models have been developed to predict diastereoselectivity in nucleophilic addition reactions involving carbonyl groups such as the Cram and Felkin–Anh models.[46,47] The early Cram model states that the ideal path of attack of a nucleophile toward a carbonyl group is essentially that minimizing steric hindrance. The more reliable Felkin–Anh model invokes additional electronic effects which control the diastereoselectivity; a strong electron withdrawing group (R^L in Figure 2) at the vicinal position oriented antiperiplanar to the incoming nucleophile leads to a favorable $\sigma \rightarrow \sigma^*$ interaction stabilizing

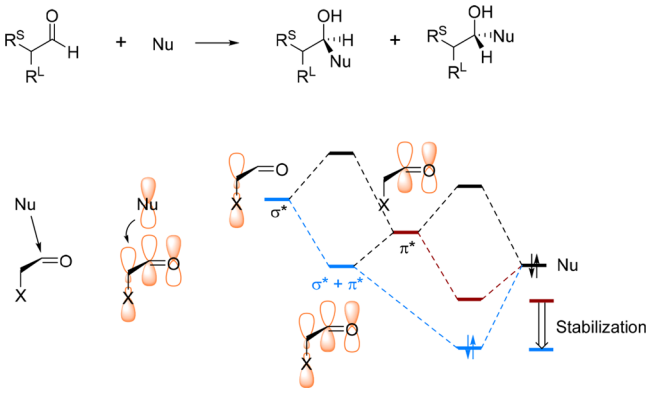


Figure 2. Electronic interactions evoked by the Felkin–Anh model to predict conformational preference.

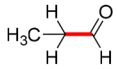
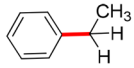
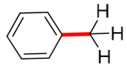
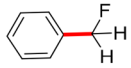
the transition state. Furthermore, the angle of attack of the nucleophile is not 90° but ~107° (Burgi–Dunitz angle)[48] which maximizes the alignment of the nucleophile σ orbital with the carbonyl π^* orbital, ultimately leading to the bond formation.[4−6] Although these different models disagree as to which of these interactions predominate, and do not predict the same stereochemical outcomes, it remains clear that both steric and electronic effects govern nucleophilic addition reactions on carbonyl centers.

While the Felkin–Anh model is in principle intended to predict the orientation of nucleophilic attacks to the C-α, it can also provide an understanding of the $\pi \rightarrow \sigma^*$ or $\sigma \rightarrow \pi^*$ hyperconjugation propensity. From the Felkin–Anh model, strongly electron-withdrawing (EWD) groups play a role similar to large substituents favoring the alignment of the σ^* with π and π^* orbital and, thus, favoring the $\pi \rightarrow \sigma^*$ hyperconjugation. Indeed, as can be observed in Table 1, our calculations with the natural bond orbitals method (NBO) (see Computational Methods) allowed us to quantify the strength of hyperconjugation interactions and showed that strong EWD groups (e.g., fluorine) favor $\pi \rightarrow \sigma^*$ relative to $\sigma \rightarrow \pi^*$. The favorable nucleophilic attack at the carbonyl group may therefore be attributed to the interaction between the $\sigma \rightarrow \pi^*$ resulting in a lower energy LUMO and thus more susceptible to nucleophilic attack. On the other hand, electron donating groups (e.g., CH_3) result in weak σ^* receptors and thus $\sigma \rightarrow \pi^*$ hyperconjugation predominates. In Table 1, we present values for the energy gaps and Fock matrix elements obtained from NBO calculations. In short, the Fock matrix elements are related to the overlap between the interacting orbitals (a larger value/overlap corresponds to stronger hyperconjugation), and energy gaps are self-explanatory (smaller energy gap leads to a stronger interaction as discussed). We would recommend ref 43 to any reader interested in the details of how these values are calculated from NBO.

The qualitative models routinely employed by organic chemists are often transferable, as they tend to isolate the predominant factors to simplify the picture. However, in order to translate these qualitative theories into robust quantitative predictions, an inclusion of other weaker interactions might be necessary. The underlying interrogative is for a given torsion involving $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$, should the predominant hyperconjugation mode be exclusively considered or should contributions from both modes be incorporated in our model; this is particularly interesting for cases where comparable

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Energy Gap and Fock Matrix Elements for $\pi \rightarrow \sigma^*$ and $\sigma \rightarrow \pi^*$ Hyperconjugation

		E_{hyp}(kcal/mol)	ΔE [BD – BD*]/(a.u)	F[BD, BD*]/(a.u)
	π → σ*	1.77	1.10	0.039
	σ → π*	5.96	1.03	0.070
	π → σ*	2.28	1.10	0.045
	σ → π*	9.32	0.91	0.082
	π → σ*	4.5	1.01	0.060
	σ → π*	2.3	1.40	0.051
	π → σ*	4.42	0.90	0.056
	σ → π*	4.7	0.98	0.061
	π → σ*	5.22	0.90	0.061
	σ → π*	7	0.86	0.069
	π → σ*	11.68	0.79	0.086
	σ → π*	2.04	1.34	0.047

magnitudes are observed (e.g., toluene in Table 1). Furthermore, should $\sigma \rightarrow \sigma^*$ hyperconjugation interactions be neglected as they are smaller in magnitude than π-hyperconjugation interactions? It is important to note that $\sigma \rightarrow \sigma^*$ are maximal when the σ and σ^* orbitals are *anti* (in plane with the π-system), whereas $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$ are maximal when the σ and σ^* orbitals are perpendicular to the π-system (Figure 3). Hyperconjugation and π-hyperconjugation hence favor different conformations, thus neglecting weaker competing interactions could hinder the predictive ability of our model.

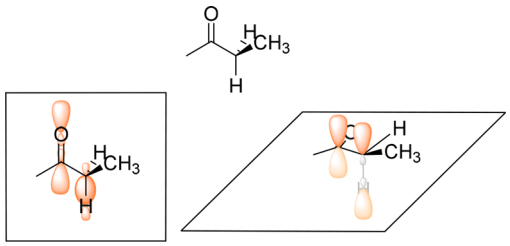


Figure 3. Different conformations favor different hyperconjugation modes. $\sigma \rightarrow \sigma^*$ is favored when C−H and C═O are coplanar (left), and $\pi \rightarrow \sigma^*$ is favored when C−H and C═O are perpendicular (right).

Hyperconjugation and/or Sterics as Major Torsional Energy Contributors. The conformational flexibility of small drug-like molecules essentially stems from rotation around bonds (i.e., dihedral angles). Hence, an accurate prediction of torsional energy profiles is critical for applications in SBDD. Nonbonded interactions (vdW, electrostatics) cannot be neglected, however, as when a bond is rotated, the molecule can reorganize other degrees of freedom in order to minimize steric clash, allow favorable H-bonding or vdW interactions to occur etc.; these additional effects become weaker however as molecules get smaller. Typically, empirical torsional parameters are parametrized last and are the only terms in the FF equation which do not explicitly describe a specific underlying physical interaction. While this empirical nature can make up for errors in nonbonded parameters and improve the accuracy of molecules in the development set, it may be at the root of the poor transferability of torsional parameters.[13] We hence hypothesize that replacing these poorly transferable empirical parameters, by contributions from different hyperconjugation modes ($\sigma \rightarrow \sigma^*$, π-hyperconjugation) will improve the transferability of torsional energies for drug-like molecules. It should be noted that for the purpose of our comparison, the remaining terms of the FF energy will be calculated with the current implementation of GAFF2; hence, residual error from the other parts of the FF are expected to be present.

The first step in our approach was to confirm our hypothesis that hyperconjugation interactions will play an important role in the determination of conformational preference. Rotational energy profiles were computed using QM at the MP2/6-311+G** level of theory which is consistent with previous studies.[36,37] As shown in Figure 4, different π-system reveal varying conformational preferences and radically different rotational profiles (amplitude and periodicity). On one hand, the thiophene and benzene profiles contain two minima (±90°) and two maxima (0°, 180°), whereas the ketone and furan show 3 minima (180°, ±60°) and 3 maxima (0°, ±120°). Inspecting the optimal conformations of each molecule, we notice that for both the benzene and thiophene derivatives, the $-CH_3$ substituent is positioned such as to maximize the overlap between the $\sigma_{(C-C)}$ and π/π^* orbitals in the aromatic ring, at the expense of possible interactions between the $\sigma_{(C-H)}$ and π-orbitals (Figure 4). On the other hand, the furan derivative in its preferred conformation shows reduced orbital overlap between the $\sigma_{(C-C)}$ and the π-orbitals, allowing one of the $\sigma_{(C-H)}$ to partially overlap with the π-orbitals. For the ketone, $\sigma_{(C-C)}$ does not interact at all with the π-orbitals, and both $\sigma_{(C-H)}$ partially overlap with the π-orbitals. In the benzene example, the preference for ±90° can be attributed to unfavorable steric clash between hydrogens at the ortho position when the methyl group is in plane of the π-system. The thiophene derivative reveals a very similar profile, although it is not expected to be subject to a steric clash of the same magnitude, the hydrogen atom being further away (5-membered rings having inherently different geometries than 6-membered rings).

A notable difference between the benzene and thiophene profiles is the broadness of the low energy region. Clearly, the

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

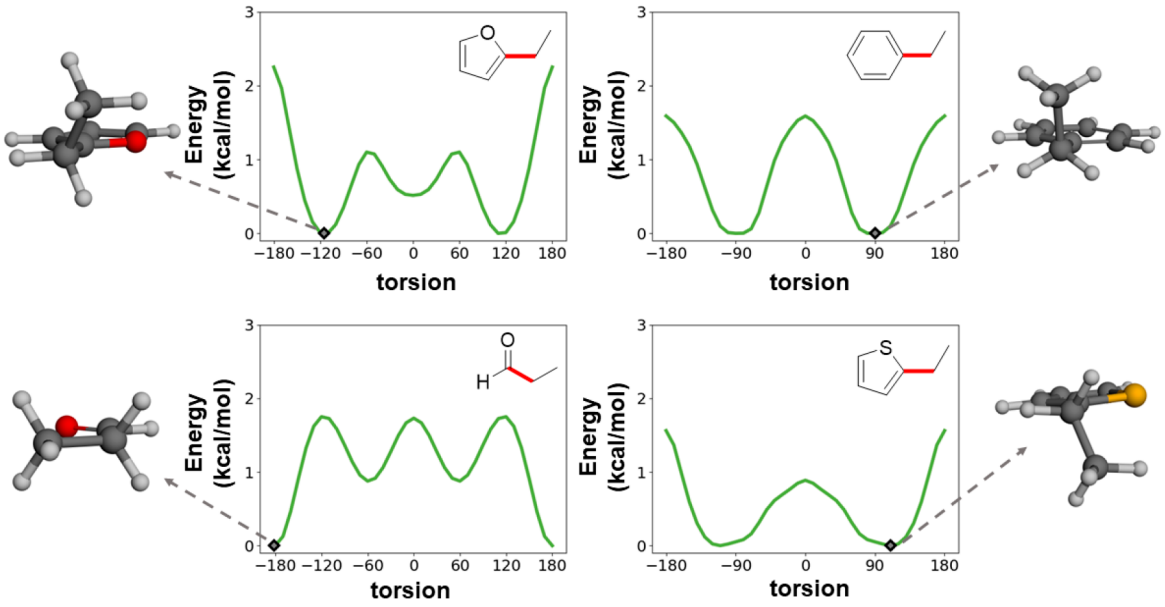


Figure 4. Variety of QM torsional profiles linked to the underlying interactions. Rotated bonds are shown in red.

nature of the π-system is closely linked to which interaction will predominate and, ultimately, to which conformation will be preferred. Three hyperconjugation modes are competing in these systems ($\sigma \rightarrow \sigma^*$, $\sigma \rightarrow \pi^*$, and $\pi \rightarrow \sigma^*$; Figure 3) and their strength depends on two major factors, the energy level difference of the interacting bonding/antibonding orbital pair and the spatial orbital overlap.[42,43,49,50] The nature of the π-system is directly related to the energy levels of π and π^* orbitals, as well as their polarization.[51] While it remains unclear at this stage as to which interaction predominates in each case, we can assess that the same set of interactions (with different electronic effects from various functional groups) may lead to very different profiles (Figure 4).

Understanding Interactions. While quantum chemistry can provide an accurate depiction of molecules, the information which can be extracted remains limited, and it is sometimes impossible to directly translate results obtained from these theoretical calculations into well understood chemical or physical principles. This discrepancy has thus led to a wide range of QM based methods that decompose the quantum energy into more chemically relevant parts. There are currently three main approaches allowing to dissect delocalization interactions (hyperconjugation in this work): natural bond orbital (NBO) analysis,[52] energy decomposition analysis (EDA),[53] and the block localized wave function methods (BLW).[54] While these methods are built around similar concepts (a full wave function or electron density is compared to a localized construct, and the energy difference between both is assumed to be related to delocalization interactions), they operate quite differently, specifically in the way they generate the localized orbitals. While EDA methods and BLW use nonorthogonal orbitals (which increases the role of steric effects), NBO uses orthogonal orbitals to describe the localized reference.[55] Such decomposition schemes were initially developed to study intermolecular interactions[56,57] but have more recently been used to study intramolecular hyperconjugation type interactions.[58] The degree to which the decomposition is performed also varies between methods, and to this day, NBO is the only method which can output an energy value for every bonding/antibonding orbital pair in a molecule. In other methods, hyperconjugation and conjugation energies are agglomerated into a single energy term, hence not giving a chemically relevant picture with the same level of resolution.

Considering the two major interactions present in our systems are $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$, we expect that factors increasing the amplitude of one would decrease the amplitude of the other, as a good σ-donor is usually a poor σ-acceptor, and *vice versa* (see Table 1).[43] Hence, a full decomposition of the interactions resolving both $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$ seems more valuable, our end goal being to understand and develop rules to explain the factors controlling these interactions. NBO has notably been applied to understand the conformational preference of ethane, by invoking the predominant role of hyperconjugation,[39] to discern how different elements within pnictogens (N, P, As)[59] and chalcogens (O, S, Se, Te)[60] impact n $\rightarrow \sigma^*$ hyperconjugation and ultimately the magnitude of the anomeric effect. NBO has also already been used to study the torsional energy profiles of conjugated systems.[61] Overall, NBO has been employed to understand the conformational preference of a wide range of molecules, explaining these preferences with different hyperconjugation modes, as well as to explore how different elements within a group can impact such interactions; it is therefore particularly fit for our purposes.

COMPUTATIONAL METHODS

Construction of the Development Set. In order to complement our previously developed H-TEQ method, our objective was to replace the empirical torsional energy found in GAFF2 by a new function which would describe π-hyperconjugation. This function would be based on atomic properties (e.g., treating all carbons atoms in the same way) and from the topology of the molecule without any prior atom typing of the molecules. It is worth mentioning that atom types are associated with predefined parameters while our approach computes parameters on-the-fly. As a result, each atom of any given molecule will be assigned unique parameters. To that end, we have first constructed a development set (Figure 5), covering a large variety of π-systems and saturated groups

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

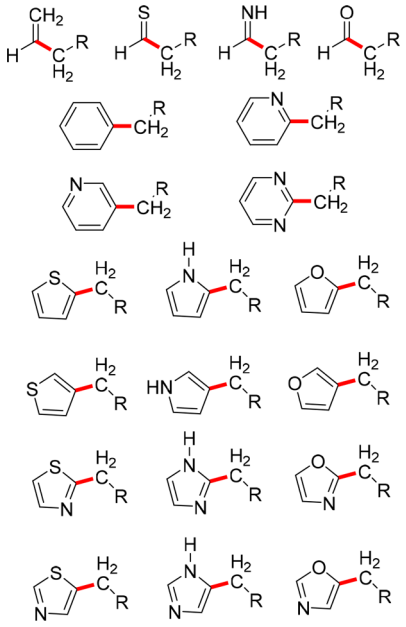


Figure 5. Development set of molecules used to study conformational preference of organic molecules containing π-systems, rotated bonds are shown in red (R = H, F, Cl, CH_3, OH).

positioned at a vicinal position, in order to understand the effects of different moieties on torsional profiles and the underlying interactions giving rise to such profiles. It should be noted that our data set is not built to resemble drug-like molecules, but rather to cover as wide a range of chemical space as possible (−R groups going from very electron withdrawing—e.g. fluorine—to electron donating—e.g. hydroxyl), in order to understand the factors governing the various hyperconjugation modes.

Details of the Calculations. First, we have obtained torsional energy profiles for every molecule in our development set using QM at the MP2/6-311+G** level of theory using the software GAMESS-US,[62,63] which is consistent with our prior studies.[36,37] In more detail, the torsional profiles were obtained by freezing the desired torsion from −180° to 180° with 10° increments while allowing all other degrees of freedom to optimize. The resulting optimized conformations were used to perform the MM and NBO calculations.

MM calculations were performed using the AMBER16 package; GAFF2 atom types were automatically assigned with antechamber, partial charges were assigned using the AM1-BCC method on the global minimum structures and were carried to all other conformations of the same molecule. Finally, the GAFF2 energy was calculated by following the sander routine.

NBO calculations were performed with the NBO 6.0[52] program using the same level of theory and basis set as our QM calculations. To verify whether NBO energies could be used, we first replaced the torsional energy (related to the central rotated bond only) in GAFF2, by hyperconjugation energies (from NBO) related to all of the hyperconjugation modes around the central bond of interest. We then resorted to a scaling of these NBO energies down by factors of 0.25 for $\sigma \rightarrow \sigma^*$ and 0.4 for π-hyperconjugation to minimize the root-mean squared error (RMSE) between QM and scaled NBO profiles (see Table S1). Note that the NBO profiles were obtained by replacing the torsional energy in GAFF2 by scaled NBO hyperconjugation energies. Only the torsional energy related to the central rotated bond were replaced by NBO energies. A Fourier regression of the hyperconjugation profiles of every molecule in the development set was performed, such that each $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$ would be associated with a set of V_{1-3} parameters.

The RMSE calculations were performed with eq 8, in which every point in the torsional profile is compared to the QM reference. Prior to the RMSE calculation, profiles were rescaled such that the point with lowest energy was set to 0 kcal/mol. RMSEs being an imperfect measure, cut-offs or Boltzmann weights are sometimes used to discard or scale down the contribution from points of the potential energy surface that are higher in energy.[64] We did not notice any significant difference in RMSEs when using such schemes (Table S2) and have kept the simpler RMSE scheme shown in eq 8.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_n (E_{\mathrm{QM}} - E_{\mathrm{MM}})^2}{n}} \tag{8}$$

Finally, to compare the performance of our method H-TEQ 3.0, we have replaced the torsional energy in GAFF2 by the equations we have developed (eqs 9 and 10) which are meant to reproduce NBO calculated π-hyperconjugation energies. $\sigma \rightarrow \sigma^*$ hyperconjugation energies were also included by using our previously developed set of equations.[37] It is important to note that only the torsional energies related to the central rotated bond were replaced by H-TEQ 3.0 values. Indeed, our method cannot yet cover all possible chemical groups (notably torsions involving $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ interactions), and some of the molecules in our validation set could not be supported by H-TEQ 3.0. Hence to treat every molecule with the exact same methodology, we kept all of the other (noncentral) GAFF2 torsional energies. Generally, rotation around these other torsions is minimal as the central bond is rotated, particularly when the molecule is small. A few exceptions in which the other parts of the molecule reorganized considerably were observed though (Figure S1).

Construction of the Validation Set. To evaluate the performance of H-TEQ 3.0, we have developed a validation set of 50 diverse drug molecules from a larger set previously developed in our lab.[65] In order to reduce the computational cost associated with the obtention of QM torsional profiles, molecules were fragmented, keeping only the most relevant parts. For example, molecules were fragmented at sp^3–sp^3 bonds, replacing parts of the molecule by hydrogens atoms. Furthermore, torsional profiles were obtained by using 15° (rather than 10°) increments, thereby reducing the number of conformations in a profile from 36 to 24. Overall, our validation set does not contain any molecule used to train the model, and a variety of novel π-systems were used (e.g., extended conjugated systems, fused rings). The full set could be found in the Supporting Information (Figure S2).

RESULTS AND DISCUSSION

Quantifying Hyperconjugation from NBO. The first step to our approach consisted in verifying that NBO hyperconjugation values could be used to replace the torsional energy within GAFF2. When replacing GAFF2 torsional energies by scaled NBO hyperconjugation values, we noted a significant decrease in RMSE from 0.84 to 0.55 kcal/mol with respect to the QM reference (full results in Table S1). A more detailed account of NBO's performance can be obtained by

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

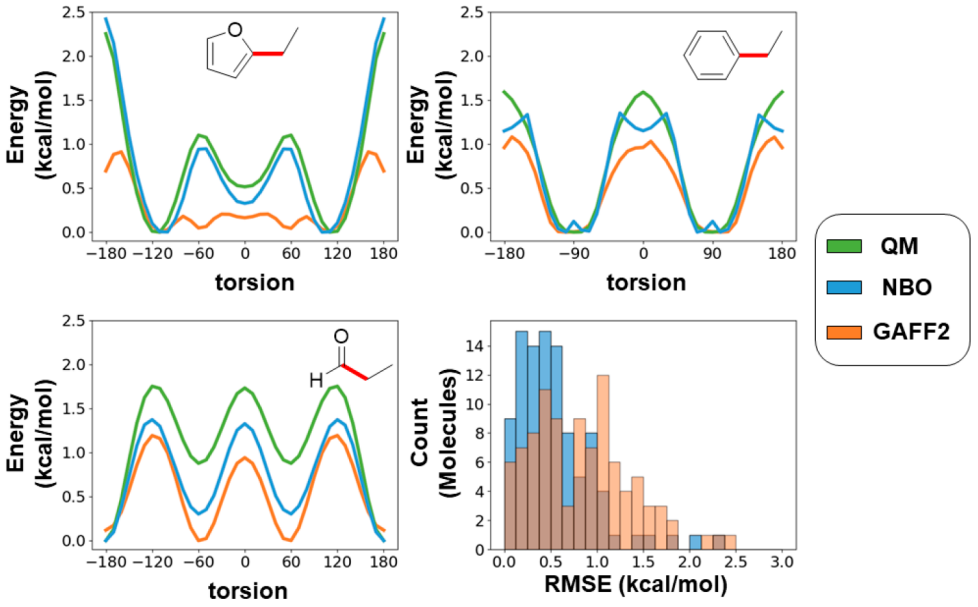


Figure 6. Replacing the torsional energy term in GAFF2, by hyperconjugation obtained from NBO (with scaling factors of 0.4 and 0.25). Histogram distribution of the RMSEs of the 98 molecules in our development set between NBO/QM (blue) and GAFF2/QM (orange).

inspecting the histogram shown in Figure 6. As expected from the average values shown in Table S1, RMSEs obtained by NBO are lower than those obtained using GAFF2. More interestingly, GAFF2 reveals the presence of two populations, which we can interpret as molecules having been explicitly parametrized (low RMSE) and those for which parameters were transferred from similar molecules but which suffer from poor transferability (larger RMSEs). The development set used herein consists of very small molecules only (<20 atoms), and we expect the lack of transferability to be further exacerbated in larger, more diverse drug-like molecules, as the probability that more parameters will be suboptimal is larger, and smaller errors will accumulate. In Figure 6, the torsional profiles for three molecules are shown using QM, GAFF2, and NBO (replacing the torsion energy of GAFF2). Although for some molecules (e.g., ethylbenzene) the impact of replacing torsional energies with hyperconjugation was low, the QM torsional profile was reproduced much more accurately in the furan and ketone examples. While the energy barriers remained underestimated in the ketone profile, NBO correctly predicted that the three energy minima were not equivalent. Furthermore, the furan profile was modeled with far greater accuracy by NBO, which can be explained by the fact that 5-membered rings are poorly parametrized (some not parametrized at all) in GAFF2; hence, calculations often rely on "generic" parameters.

While NBO energies were better at reproducing QM profiles, these calculations cannot be run to generate torsional parameters every time parameters are required. Hence, following an approach used successfully in the development of earlier versions of H-TEQ, our first objective was to understand factors governing the strength of the different interactions based on NBO generated data, and to develop a set of rules based on atomic properties (electronegativity, bond length, aromaticity etc.) which would reproduce NBO interaction energies and could be calculated on-the-fly. Such a method would allow chemists to generate parameters for any drug-like molecule containing π-systems for use in MD simulations or docking studies.

Electronegativity, Aromaticity, and π-Hyperconjugation. As previously mentioned, two major factors impact the strength of the $\sigma \rightarrow \pi^*/\pi \rightarrow \sigma^*$ interactions. First, the energy gap between bonding and antibonding interacting orbitals is directly related to electronegativity. More electronegative elements yield lower-in-energy orbitals; for example, the π and π^* orbital energy levels are higher in benzene, than in pyridine. Thus, introduction of heteroatoms into an aromatic ring, or nonaromatic conjugated systems leads to a lowering of the energy levels. The energy levels of π and π^* orbitals being quite disparate, we can expect that a lowering of the energy levels would favor interactions unidirectionally ($\sigma \rightarrow \pi^*$ or $\pi \rightarrow \sigma^*$), as when the energy gap decreases for one interaction, it is expected to increase for the other.

Second, the spatial overlap between interacting orbitals. From heterocyclic chemistry, it is known that the introduction of heteroatoms into aromatic systems results in differences in atomic charges, as well as a greater shielding effect due to heteroatom substituents.[66] It is also expected that more electronegative heteroatoms lead to a polarization of the π orbitals, which strongly impacts the ability of vicinal (σ or σ^*) orbitals to overlap and interact. π and π^* orbitals have opposite polarization, further reinforcing the fact that, as one interaction becomes stronger, the other weakens, as it is impossible for two orbitals with opposite polarizations to have simultaneous strong overlaps with vicinal orbitals. In nonpolar π-systems such as alkenes, π and π^* orbitals are equally distributed toward both atoms of the double bond. In contrast, introduction of more electronegative heteroatoms (N, O) leads to the polarization of the π orbital toward the heteroatom, which ultimately limits the ability of the π-system to partake in $\pi \rightarrow \sigma^*$ donation (lower orbital overlap, increase in energy gap). On the other hand, the π^* orbital is polarized toward the carbon atom of the double bond leading to stronger overlap for the $\sigma \rightarrow \pi^*$ interaction, resulting in a more prominent acceptor ability. As a rule of thumb, good π acceptors will be poor π donors and vice versa (although in some cases both interactions occur with similar magnitudes, Table 1). In Figure 7, NBO profiles are shown for these

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

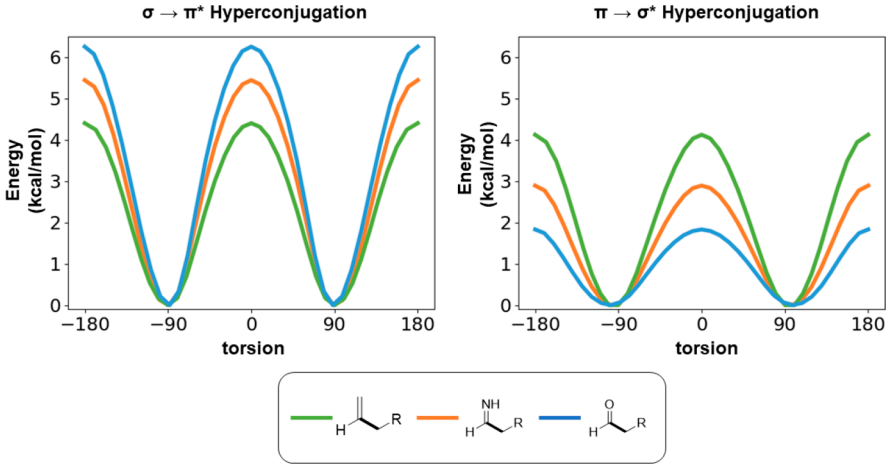


Figure 7. Electronegativity of elements in π systems has an opposing effect for $\pi \rightarrow \sigma^*_{(C-R)}$ and $\sigma_{(C-R)} \rightarrow \pi^*$ interactions. In this example −R is a methyl group, although this trend extends to all −R groups studied (H, Cl, F, OH, CH_3).

specific interactions, indeed we observe that less electronegative elements in the π-system leads to a stronger $\pi \rightarrow \sigma^*$, but weaker $\sigma \rightarrow \pi^*$ interaction. The concept of electronegativity is therefore central in understanding the propensity of a system to contribute to π-hyperconjugation.

Similarly, the elements involved in the σ group with which the π-system is interacting impact the torsional energy profiles. More electronegative elements act as better donors, and weaker acceptors, and less electronegative elements will be better donors and weaker acceptors. The concept of electronegativity will hence be our major descriptor for π-hyperconjugation.

Developing Equations for π-Hyperconjugation. Traditionally, the torsional component of the energy in FFs is calculated using a truncated Fourier series (eq 4); the number of terms (N) included varies depending on the FF but usually does not exceed 6 with 3 being most common. While FFs are empirical in nature, it is essential to understand that each term (V_n) can be interpreted in the context of rotation around a bond and assigned a corresponding chemical meaning. The V_1 terms relates to the *syn* or *anti* preference of two groups, since π-orbitals are somewhat symmetrical (similar density above and below the ring/double bond), the V_1 term should be negligible in our equation. The V_2 term describes the energy cost of rotating around a bond and is related to the strength of the interaction which is maximal at 90° (maximal orbital overlap) and minimal at 0° (no orbital overlap). Finally, the V_3 term can be understood as a correcting factor which can weakly shift the energy barrier, this V_3 term is also related to orbital overlap in the sense that it controls whether it is possible to rotate the bond slightly away from the ideal conformation (±90°) without losing π-hyperconjugation (Figure S3).

For our purposes, V_2 and V_3 components of the torsion energy were sufficient to describe π-hyperconjugation interactions; π-hyperconjugation torsion profiles obtained from NBO were fitted to derive the V_2 and V_3 parameters for each molecule in the set. Since we are treating both interactions independently, each interaction will be assigned its corresponding V_2 and V_3 value, which will then be summed to describe π-hyperconjugation fully. Furthermore, as we noticed that $\sigma \rightarrow \sigma^*$ could not be neglected, we will also add the classical hyperconjugation contribution (weaker), by using our previously developed set of equations.[37] As expected, we found that only the V_2 term was subject to large variations from a molecule to another, hence the V_3 term was assigned a constant value of −0.5 kcal/mol for all molecules. We then concentrated our efforts into the development of an equation to model the V_2 term for both $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$ interactions, based on our understanding of the effects of electronegativity (χ) on energy levels and polarization of the orbitals involved in these interactions. More specifically how the strengths of these interactions are modulated by the electronegativity of relevant parts of the molecule.

$$V_2(\sigma \rightarrow \pi^*) = a\,\frac{\chi^{\pi 1} + \chi^{\pi 2}}{\chi^{\sigma}} + b \tag{9}$$

$$V_2(\pi \rightarrow \sigma^*) = c\,\frac{\chi^{\sigma}}{\chi^{\pi 1} + \chi^{\pi 2}} + d \tag{10}$$

$$\chi_{\text{group}} = \frac{1}{\omega + N}\left(\omega \times \chi_{\text{central}} + \sum_{\text{neighbor}=1}^{N} \chi_{\text{neighbor}}\right) \tag{11}$$

Our group electronegativity assignment scheme not only considered the *central* atom (circled in Figure 8) but also

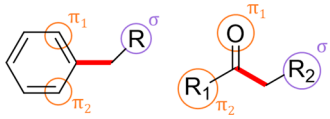


Figure 8. Parts of the molecule considered to predict the strength of multiple interactions. The electronegativities of circled atoms are used to calculate the interactions of $\sigma_{(C-R)} \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*_{(C-R)}$ (benzene analog) and $\sigma_{(C-R2)} \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*_{(C-R2)}$ (ketone).

covalently bound neighboring atoms. The central atoms were accounted for with a greater weight ω (eq 11). The electronegativities of the π-system ($\chi^{\pi 1}$ and $\chi^{\pi 2}$) were calculated by considering both sides of the ring, and the convention was kept for less aromatic molecules (or non-aromatic) such as 5-membered rings or conjugated chains (Figure 8). This could be rationalized by inductive effects of the groups attached to the π-system (without forming a double bond); strongly electron-withdrawing groups (EWG) will increase the propensity of $\sigma \rightarrow \pi^*$, and strongly electron-

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

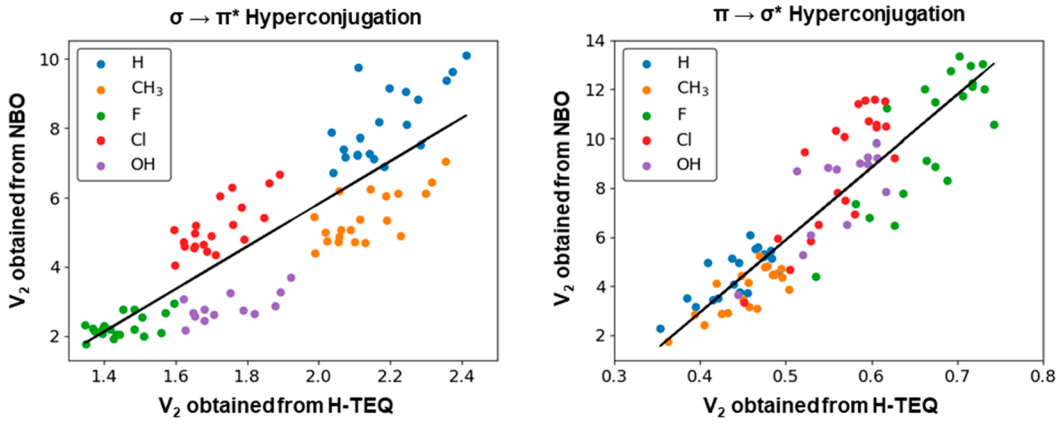


Figure 9. Comparison of rules developed (eqs 9 and 10) to describe both π-hyperconjugation modes ($\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$) with values calculated using NBO analysis. Correlation coefficients obtained are $r^2 = 0.71$ ($\sigma \rightarrow \pi^*$) and $r^2 = 0.81$ ($\pi \rightarrow \sigma^*$).

donating groups (EDG) will increase the propensity of $\pi \rightarrow \sigma^*$.

It is important to note that the major factors in eqs 9 and 10 are inverses of one another, as effects favoring one interaction disfavor the other. Here V_2 is shown as proportional to a function based on the electronegativity (χ) of various parts of the molecule (Figure 8). The values for electronegativities were obtained from the Pauling scale (Table S4)[67] and electronegativity was calculated using the concept of group electronegativity as discussed previously.[36,37] Indeed for $-CF_3$ or $-CH_3$ substituents, we expect the electronegativity of the carbon atom to be much larger in trifluoromethyl than methyl, due to the neighboring chemical environment. Sanderson's electronegativity equilibration principle[68] states that the electronegativity of both atoms in a diatomic system equilibrate to give rise to a new value related to the equilibrium charge distribution in the molecule (this postulate can be extended to all molecules, not only diatomic). Indeed, while electronegativity measures the ability of an atom to attract electrons toward itself, in polar molecules after the electron density has found its ideal distribution, there is no net flux of electrons away from this optimal distribution; in principle, electronegativity needs to be the same for every atom. A large amount of work has been dedicated to understanding the relationship between electron density and electronegativity, from which researchers have developed many schemes to calculate "group electronegativity" for specific parts of a molecule.[69−72] Although this area of research received a lot of attention in the 80s and 90s, no recent contributions were found in the literature. Some of the schemes for group electronegativity rely on calculating the partial charge of every atom, ultimately requiring QM calculations and are, hence, not consistent with our objective to develop a method applicable to high-throughput tasks. We have thus opted to use the simple equation described here by Smith et al.[72] It is interesting to note that electronegativity equalization methods have also been applied to derive partial charges,[73−75] for example the current implementation of CGenFF (CHARMM force field for drug-like molecules) uses a method which draws from these ideas to generate partial charges.[76]

The strengths of the $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$ obtained from NBO correlate well ($r^2 = 0.71$ and 0.81) with the developed rules (Figure 9), linear least-squares regression provided us with the values for the ***a***, ***b***, ***c***, and ***d*** coefficients in eqs 9 and 10. It should be noted that changing the electronegativity scale (e.g., Pauling units vs Mulliken units) or weights in the group electronegativity calculation impacted the correlation of our method with NBO derived values. Indeed, while modifications to the equation could in principle improve the accuracy (stronger correlation to NBO) of one of the interactions (e.g., $\sigma \rightarrow \pi^*$), it usually led to a decrease in the correlation with the other ($\pi \rightarrow \sigma^*$). We thus decided to keep the simplest equation (with a weight of 2 for the central atom) and Pauling units as used previously to limit overfitting. Overall, eqs 9 and 10 both use the same electronegativity scales and weights. It should be noted that minor scaling factors were used to differentiate different kinds of π-systems (6-membered, 5-membered, double bonds) such that the same equation could be used for all molecules (Figure S4).

These equations were implemented into H-TEQ3.0, a program deriving V_{1-3} parameters for MM calculations. The developed Java program also includes the equations from the previous versions of H-TEQ. The parameters (***a***−***d***) used in eqs 9 and 10 are shown in Table 2. Considering the V_2 values

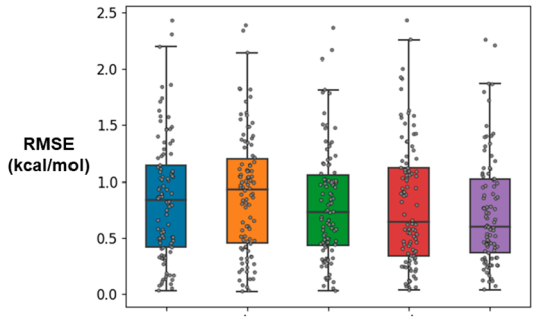
Table 2. Parameters Obtained from the Linear Regression

parameter	value (kcal/mol)
a	6.16
b	−6.50
c	29.52
d	−8.88

generated for $\pi \rightarrow \sigma^*$ and $\sigma \rightarrow \pi^*$ are summed to make an overall V_2 term, parameters ***b*** and ***d*** could be agglomerated into a single parameter. We present them as separate parameters to highlight the fact they were obtained from a linear least-squares regression around NBO data. To test the transferability of these parameters, we have also performed a bootstrapping analysis (Table S5).

Evaluation. The performance of this newly developed method (H-TEQ 3.0) on the development set of molecules was compared alongside GAFF2 against QM energies (Figure 10). The contribution of $\sigma \rightarrow \sigma^*$ was calculated using the previously published version of H-TEQ 2.0.[37] Overall, our method was found to perform better than GAFF2 when a V_3 correction factor of −0.5 kcal/mol was used. While the inclusion of $\sigma \rightarrow \sigma^*$ had a minimal impact on the overall RMSE, the marginally lowest RMSE was found when $\sigma \rightarrow \sigma^*$ hyperconjugation was omitted, which contradicted with the

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX



Method	GAFF2	H-TEQ 3.0			
Inclusion of V_3 correction	n/a	×	×	✓	✓
Inclusion of σ → σ*	n/a	✓	×	✓	×

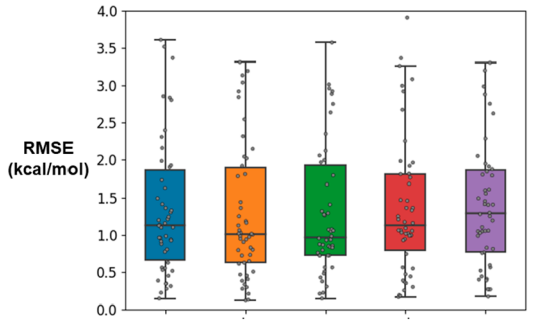
Figure 10. Performance of GAFF2 and H-TEQ3.0 methods over the development set of 98 molecules. Contributions of $\sigma \rightarrow \sigma^*$ hyperconjugation and the V_3 correction factor were also monitored to understand their impact on our method. The black line at the center of each box corresponds to the median value.

results found when replacing raw NBO values (Table S1). This discrepancy likely results from the equation modeling $\sigma \rightarrow \sigma^*$ hyperconjugation being trained only on sp^3 centers, which might not be fully transferable to conjugated (sp^2) centers. In conjugated systems, shielding of the σ orbitals by π orbitals is expected, and different geometries of the substituents (109.5° vs 120°) modify the ability of orbitals to overlap.

Overall, our method was found to be more accurate than GAFF2 in reproducing QM profiles, doing so without requiring the use of atom type description of the molecules, or any prior parametrization.

Performance and Validation. To validate our findings, we have applied the H-TEQ3.0 method to a diverse set of 50 drug molecules (Figure S2) which does not contain any molecule used to train the model, and a variety of novel π-systems (extended conjugated systems, fused rings, Figure 11).

Furthermore, we rotated bonds that were located both in the center and at extremities of the molecules, the former being more important as they lead to the most prominent conformational changes. As for the development set, our method was compared to the widely used GAFF2, and torsional energy was replaced by our equations for π-hyperconjugation and previous equations from H-TEQ 2.0 were employed for $\sigma \rightarrow \sigma^*$ hyperconjugation. Again, the effects of the V_3 correcting factor and $\sigma \rightarrow \sigma^*$ interactions were monitored by switching them on/off (Figure 12).



Method	GAFF2	H-TEQ 3.0			
Inclusion of V_3 correction	n/a	×	×	✓	✓
Inclusion of σ → σ*	n/a	✓	×	✓	×

Figure 12. Performance of GAFF2 and H-TEQ3.0 methods over the validation set of 50 molecules. Contributions of $\sigma \rightarrow \sigma^*$ hyperconjugation and the V_3 correction factor were also monitored to understand their impact on our method. One outlier with a large RMSE (~20 kcal/mol) is not shown (Figure S1). The black line at the center of each box corresponds to the median value.

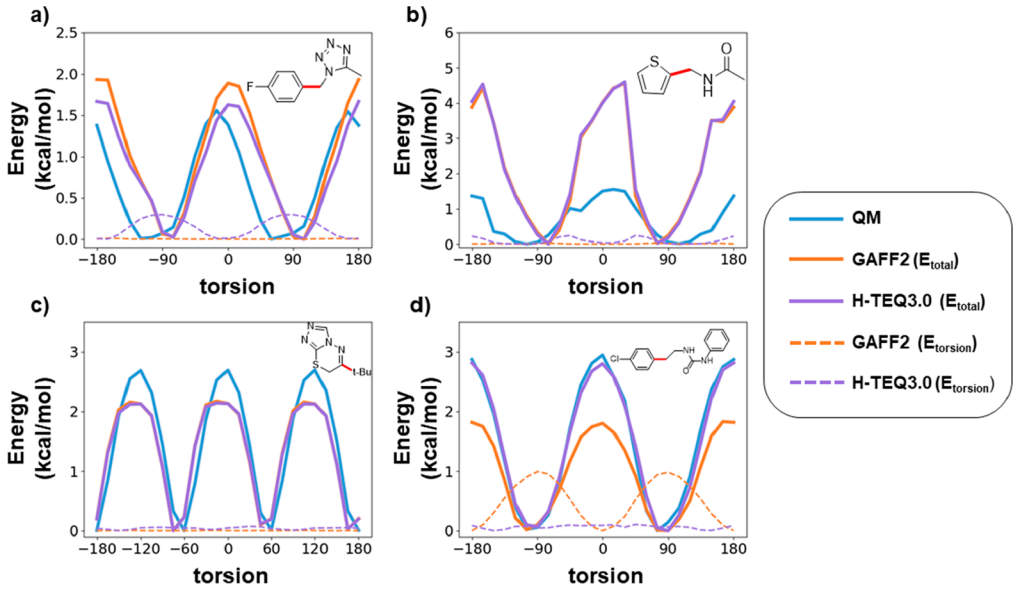


Figure 11. Performance of our method on four drug-like molecules chosen from the validation set. Full lines correspond to the total energy predicted by each method, and dashed lines correspond to the torsional component (of the central bond in red) only. The four molecules shown are (a) 1-(4-fluorobenzyl)-5-methyl-1*H*-tetrazole, (b) *N*-(thiophen-2-ylmethyl)acetamide, (c) 6-(*tert*-butyl)-7*H*-[1,2,4]triazolo[3,4-*b*][1,3,4]-thiadiazine, and (d) 1-(4-chlorophenethyl)-3-phenylurea.

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Regardless of the specific method used (inclusion or not of the V_3, etc.), our method performs on par with the current implementation of GAFF2 (full results in Table S6). The V_3 factor was found to slightly negatively impact the accuracy of our method, while the effects from $\sigma \rightarrow \sigma^*$ hyperconjugation were found to be minimal. In Table 3, results are summarized

Table 3. Accuracy of GAFF2 and H-TEQ3.0 to Reproduce the Torsional Profiles over the Development and Validation Sets of Molecules

method compared to MP2/6-311+G**	set of molecules used	average RMSE (kcal/mol)
GAFF	development set	0.84
GAFF (no torsions)[a]	development set	0.93
H-TEQ3.0[b]	development set	0.80
GAFF	validation set	1.69
GAFF (no torsions)[a]	validation set	1.78
H-TEQ3.0[b]	validation set	1.71

[a]The torsional energies related to the central bond were set to 0. [b]Both $\sigma \rightarrow \sigma^*$ and V_3 were included.

for the version of H-TEQ including both V_3 and $\sigma \rightarrow \sigma^*$, the slight increase in accuracy over GAFF2 in the development set was lost in the validation set. This should not be confused as a lack of transferability of H-TEQ however, and the larger RMSEs in the validation set result from the larger prevalence of nonbonded interactions as torsions are rotated, in these larger drug-like molecules. Indeed, torsional barriers in the validation set were larger (average highest point on the profile at 5.10 kcal/mol, compared to 2.60 kcal/mol for the development set). Considering the development set was built to go from minimal to maximal hyperconjugation strength, molecules in the validation should not have stronger hyperconjugation interactions. Therefore, the additional energy in the validation set profiles must come from nonbonded interactions, which were calculated using GAFF2 (our method has no impact on the accuracy of these energy terms). The prevalence of nonbonded interactions is further demonstrated in Figure 11, indeed GAFF2 and H-TEQ3.0 profiles are very similar, and the torsional component of the energy is weak compared to the overall predicted energy barriers (Figure 11a−d).

Figure 11c shows an example of a correct prediction of GAFF2 and H-TEQ3.0 where the torsional component is equal to 0 as a result of the phase cancelation of the torsion energy, and the weak vdW contribution to the energy alone can correctly predict the energy barriers. In Figure 11b, a similar phase cancelation is observed, although the overall profile overestimated the height of the energy barrier by a factor of 2.5. In this case, the flexibility of other parts of the molecule was not modeled well by other energy terms of GAFF2 (vdW, electrostatics). Profiles shown in Figure 11a and d are more interesting. In Figure 11a the weak torsional component to the energy predicted by H-TEQ3.0 (out of phase with the overall profile), replacing the null contribution of GAFF2 led to a slightly more accurate profile, while in Figure 11d, the opposite is seen, an incorrect torsional energy is predicted by GAFF2, which when replaced by a null contribution from H-TEQ led to a much more accurate profile.

An understanding of the magnitude of various π-hyperconjugation modes can explain the origins of the phase cancelation of the torsional terms. Indeed, sp^3 carbons involved in $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$ interactions of interest have three substituents which are separated by dihedrals of 120°. Considering the interactions are essentially modeled by a V_2 term, if the V_2 terms are of similar magnitude they will cancel out. In the development set, the substituent that was modified could be more electronegative (F, Cl, and OH) than the two other H atoms on the sp^3 carbon; hence, phase cancelation was not observed for these molecules. On the other hand, the majority sp^3 carbon atoms bound to π-systems in drug-like molecules will have two H atoms and another larger group as a substituent, the atom directly attached to the sp^3 carbon being in most cases a carbon atom. The propensity of π-hyperconjugation is similar for both −H and $-C(R_1R_2R_3)$ unless R_{1-3} are very electronegative, as predicted by NBO calculations (Figure 9), which explains why phase cancelation is observed for many drug-like molecules in the validation set. As a null hypothesis, we have also calculated the RMSEs where central torsional energies in GAFF2 were set to 0 and have observed larger RMSEs than when using GAFF2 with torsions or H-TEQ 3.0 (Table 3). However, the differences were rather small, which supports the idea that nonbonded interactions play an important role in determining these torsional profiles.

As a result, a major contributor to the energy in bulky drug-like molecules, when a torsion at the center of the molecule is rotated is sterics. Consequently, a correct modeling of nonbonded interactions is of greater importance to correctly predict the conformational energy landscape of such molecules. Polarizable force fields are expected to predict these nonbonded interactions with greater accuracy. Methods such as AMOEBA FF may provide a much more thorough treatment of electrostatic interactions (performed using dipole and quadrupoles moments). However, an automated tool to generate AMOEBA atom types and parameters is yet to be developed, which hindered our ability to perform and include such a comparison in the present study, as atom types and parameters would have to be assigned manually.

CONCLUSION

We have shown that replacing the torsional energy calculated by empirical parameters in FFs with more chemically meaningful potentials describing hyperconjugation interactions in conjugated molecules led to accuracies comparable to the widely used GAFF2, without relying on atom types description or parametrization, avoiding common drawbacks known to be associated with these methods. As opposed to previous work performed on saturated molecules, hyperconjugation is not the predominant factor in determining conformational preference. The self-consistency of FFs (empirical torsion making up for poor nonbonded parameters) thus explains why, for the time being, transferable methods like H-TEQ3.0 do not perform significantly better than empirical methods as long as the other terms (especially nonbonded terms) are not trained concomitantly. The nontransferability of parameters remains a central challenge in FF development, and we expect chemically relevant transferable methods like H-TEQ3.0 to provide more accurate depictions of the energetics of small drug-like molecules, provided that nonbonded interactions are more accurately transcribed. Future research goals include the comparison of our method in MD and docking studies, comparison against a wider range of FFs (including polarizable FFs).

Finally, as the treatment of nonbonded interactions was shown to be problematic in this study, application of the atom type free methodology to describe nonbonded interactions

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

could also be examined, removing entirely the need for atom typing in FFs, ultimately allowing more transferable methods to be applied toward many different SBDD programs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.9b00581.

Additional figures and tables supporting and/or illustrating the conformational preferences of selected molecules (PDF)

Molecule sets (TXT)

Molecule sets (TXT)

■ AUTHOR INFORMATION

Corresponding Author

*Email: nicolas.moitessier@mcgill.ca.

ORCID

Stephen J. Barigye: 0000-0003-3547-8293

Wanlei Wei: 0000-0002-2643-1365

Nicolas Moitessier: 0000-0001-6933-2079

Present Addresses

[§]S.J.B.: Departamento de Química Física Aplicada, Facultad de Ciencias, Universidad Autónoma de Madrid (UAM), 28049 Madrid, Spain.

[‖]Z.L.: Computational ADME, Drug Disposition, Lilly Research Laboratory, Eli Lilly and Company, Indianapolis, IN, 46285 USA.

Notes

The authors declare no competing financial interest.

All of the data (development and validation sets) and the H-TEQ method are freely available on our Web site at http://moitessier-group.mcgill.ca/software.html#hteq.

■ ACKNOWLEDGMENTS

We thank NSERC (CRD program) for financial support. Calcul Québec and Compute Canada are acknowledged for generous CPU allocations. W.W. thanks the FRQNT for a graduate scholarship.

■ REFERENCES

(1) Tian, S.; Wang, J.; Li, Y.; Li, D.; Xu, L.; Hou, T. The application of in silico drug-likeness predictions in pharmaceutical research. *Adv. Drug Delivery Rev.* **2015**, *86*, 2−10.

(2) Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717.

(3) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 9.

(4) Huang, N.; Jacobson, M. P. Physics-based methods for studying protein-ligand interactions. *Curr. Opin. Drug Disc. Devel.* **2007**, *10*, 325−331.

(5) Cheng, T.; Li, Q.; Zhou, Z.; Wang, Y.; Bryant, S. H. Structure-based virtual screening for drug discovery: a problem-centric review. *AAPS J.* **2012**, *14*, 133−41.

(6) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66*, 334−395.

(7) Dans, P. D.; Walther, J.; Gomez, H.; Orozco, M. Multiscale simulation of DNA. *Curr. Opin. Struct. Biol.* **2016**, *37*, 29−45.

(8) Halgren, T. A. Potential-Energy Functions. *Curr. Opin. Struct. Biol.* **1995**, *5*, 205−210.

(9) Lazaridis, T.; Karplus, M. Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* **2000**, *10*, 139−145.

(10) Wang, Z.; Sun, H. Y.; Yao, X. J.; Li, D.; Xu, L.; Li, Y. Y.; Tian, S.; Hou, T. J. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964−12975.

(11) Vivo, M. D. Bridging quantum mechanics and structure-based drug design. *Front. Biosci., Landmark Ed.* **2011**, *16*, 1619−1633.

(12) Vanommeslaeghe, K.; Guvench, O.; MacKerell, A. D. Molecular Mechanics. *Curr. Pharm. Des.* **2014**, *20*, 3281−3292.

(13) Vanommeslaeghe, K.; Yang, M. J.; MacKerell, A. D. Robustness in the Fitting of Molecular Mechanics Parameters. *J. Comput. Chem.* **2015**, *36*, 1083−1101.

(14) 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.* **1996**, *118*, 2309−2309.

(15) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(16) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(17) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(18) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.

(19) Daura, X.; Mark, A. E.; van Gunsteren, W. F. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.* **1998**, *19*, 535−547.

(20) Damm, W.; Frontera, A.; TiradoRives, J.; Jorgensen, W. L. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* **1997**, *18*, 1955−1970.

(21) Harder, E.; Damm, W.; Maple, J.; Wu, C. J.; Reboul, M.; Xiang, J. Y.; Wang, L. L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.

(22) Abel, R.; Harder, E.; Damm, W.; Reboul, M.; Maple, J.; Wu, C. J.; Xiang, J.; Cerutti, D.; Lupyan, D.; Wang, L. L.; Dahlgren, M.; LeBard, D. OPLS3 force field: An improved classical force field for the modeling of drug-like small molecules, proteins, RNA, and DNA. *Abstr. Pap. Am. Chem. Soc.* **2015**, 250.

(23) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, 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.

(24) Kirkpatrick, P.; Ellis, C. Chemical space. *Nature* **2004**, *432*, 823−823.

(25) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1863−1874.

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(26) Shi, Y.; Xia, Z.; Zhang, J. J.; Best, R.; Wu, C. J.; Ponder, J. W.; Ren, P. Y. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046−4063.
(27) Ren, P. Y.; Wu, C. J.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143−3161.
(28) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013**, *9*, 5430−5449.
(29) Huang, J.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model. *J. Phys. Chem. Lett.* **2014**, *5*, 3144−3150.
(30) 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.
(31) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid parameterization of small molecules using the force field toolkit. *J. Comput. Chem.* **2013**, *34*, 2757−2770.
(32) Betz, R. M.; Walker, R. C. Paramfit: Automated optimization of force field parameters for molecular dynamics simulations. *J. Comput. Chem.* **2015**, *36*, 79−87.
(33) Wang, J.; Kollman, P. A. Automatic parameterization of force field by systematic search and genetic algorithms. *J. Comput. Chem.* **2001**, *22*, 1219−1228.
(34) Mobley, D.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Open Force Field Consortium: Escaping atom types using direct chemical perception with SMIRNOFF v0.1. *bioRxiv* **2018**, 286542.
(35) Chodera, J.; Wang, L.-P.; Smith, D.; Stern, C.; Qiu, Y. Torsion fitting and small molecule fragmentation. *Open Force Field Consortium* **2019**, DOI: 10.5281/zenodo.3243674.
(36) Liu, Z. M.; Pottel, J.; Shahamat, M.; Tomberg, A.; Labute, P.; Moitessier, N. Elucidating Hyperconjugation from Electronegativity to Predict Drug Conformational Energy in a High Throughput Manner. *J. Chem. Inf. Model.* **2016**, *56*, 788−801.
(37) Liu, Z. M.; Barigye, S. J.; Shahamat, M.; Labute, P.; Moitessier, N. Atom Types Independent Molecular Mechanics Method for Predicting the Conformational Energy of Small Molecules. *J. Chem. Inf. Model.* **2018**, *58*, 194−205.
(38) Wang, J. M.; 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.
(39) Pophristic, V.; Goodman, L. Hyperconjugation not steric repulsion leads to the staggered structure of ethane. *Nature* **2001**, *411*, 565−568.
(40) Hoffmann, R.; Radom, L.; Pople, J. A.; Schleyer, P. v. R.; Hehre, W. J.; Salem, L. Strong conformational consequences of hyperconjugation. *J. Am. Chem. Soc.* **1972**, *94*, 6221−6223.
(41) Juaristi, E.; Cuevas, G. Recent studies of the anomeric effect. *Tetrahedron* **1992**, *48*, 5019−5087.
(42) Alabugin, I. V.; Gilmore, K. M.; Peterson, P. W. Hyperconjugation. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 109−141.
(43) Alabugin, I. V.; Zeidan, T. A. Stereoelectronic Effects and General Trends in Hyperconjugative Acceptor Ability of σ Bonds. *J. Am. Chem. Soc.* **2002**, *124*, 3175−3185.
(44) Alabugin, I. V. Stereoelectronic Effects with Donor and Acceptor Separated by a Single Bond Bridge. In *Stereoelectronic Effects*; 2016.
(45) Taylor, R. D.; MacCoss, M.; Lawson, A. D. G. Rings in Drugs. *J. Med. Chem.* **2014**, *57*, 5845−5859.
(46) Cram, D. J.; Elhafez, F. A. A. Studies in Stereochemistry. 10. The Rule of Steric Control of Asymmetric Induction in the Syntheses of Acyclic Systems. *J. Am. Chem. Soc.* **1952**, *74*, 5828−5835.
(47) Nguyen, T. A.; Eisenstein, O.; Lefour, J. M.; Tran Huu Dau, M. E. Orbital Factors and Asymmetric Induction. *J. Am. Chem. Soc.* **1973**, *95*, 6146−6147.
(48) Burgi, H. B.; Dunitz, J. D. From Crystal Statics to Chemical-Dynamics. *Acc. Chem. Res.* **1983**, *16*, 153−161.
(49) Al-Masri, I. M.; Mohammad, M. K.; Taha, M. O. Discovery of DPP IV inhibitors by pharmacophore modeling and QSAR analysis followed by in silico screening. *ChemMedChem* **2008**, *3*, 1763−1779.
(50) Alabugin, I. V.; Bresch, S.; dos Passos Gomes, G. Orbital hybridization: a key electronic factor in control of structure and reactivity. *J. Phys. Org. Chem.* **2015**, *28*, 147−162.
(51) Fleming, I. *Molecular Orbitals and Organic Chemical Reactions*, reference ed.; John Wiley & Sons: New York, 2010.
(52) Glendening, E. D.; Landis, C. R.; Weinhold, F. NBO 6.0: Natural bond orbital analysis program. *J. Comput. Chem.* **2013**, *34*, 1429−1437.
(53) Hopffgarten, M. v.; Frenking, G. Energy decomposition analysis. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 43−62.
(54) Mo, Y. R.; Gao, J. L.; Peyerimhoff, S. D. Energy decomposition analysis of intermolecular interactions using a block-localized wave function approach. *J. Chem. Phys.* **2000**, *112*, 5530−5538.
(55) Weinhold, F.; Carpenter, J. E. Some remarks on nonorthogonal orbitals in quantum chemistry. *J. Mol. Struct.: THEOCHEM* **1988**, *165*, 189−202.
(56) Morokuma, K. Molecular Orbital Studies of Hydrogen Bonds. 3. C = O H-O Hydrogen Bond in H2co H2o and H2co 2h2o. *J. Chem. Phys.* **1971**, *55*, 1236.
(57) Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. Hybrid density functional theory for π-stacking interactions: Application to benzenes, pyridines, and DNA bases. *J. Comput. Chem.* **2006**, *27*, 491−504.
(58) Fernández, I.; Frenking, G. Direct Estimate of the Strength of Conjugation and Hyperconjugation by the Energy Decomposition Analysis Method. *Chem. - Eur. J.* **2006**, *12*, 3617−3629.
(59) Carballeira, L.; Pérez-Juste, I. Ab Initio Study and NBO Interpretation of the Anomeric Effect in CH2(XH2)2 (X = N, P, As) Compounds. *J. Phys. Chem. A* **2000**, *104*, 9362−9369.
(60) Salzner, U.; Schleyer, P. v. R. Generalized anomeric effects and hyperconjugation in CH2(OH)2, CH2(SH)2, CH2(SeH)2, and CH2(TeH)2. *J. Am. Chem. Soc.* **1993**, *115*, 10231−10236.
(61) Lu, K. T.; Weinhold, F.; Weisshaar, J. C. Understanding Barriers to Internal-Rotation in Substituted Toluenes and Their Cations. *J. Chem. Phys.* **1995**, *102*, 6787−6805.
(62) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic-Structure System. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(63) Gordon, M. S.; Schmidt, M. W. Advances in electronic structure theory: GAMESS a decade later. *Theory and Applications of Computational Chemistry: The First Forty Years* **2005**, 1167−1189.
(64) Robertson, M. J.; Qian, Y.; Robinson, M. C.; Tirado-Rives, J.; Jorgensen, W. L. Development and Testing of the OPLS-AA/M Force Field for RNA. *J. Chem. Theory Comput.* **2019**, *15*, 2734−2742.
(65) Campagna-Slater, V.; Pottel, J.; Therrien, E.; Cantin, L. D.; Moitessier, N. Development of a Computational Tool to Rival Experts in the Prediction of Sites of Metabolism of Xenobiotics by P450s. *J. Chem. Inf. Model.* **2012**, *52*, 2471−2483.
(66) Katritzky, A. R.; Ramsden, C. A.; Joule, J. A.; Zhdankin, V. V. 2.3−Structure of Five-Membered Rings with One Heteroatom. In *Handbook of Heterocyclic Chemistry*, Third ed.; Elsevier: Amsterdam, 2010; pp 87−138.
(67) Pauling, L. The nature of the chemical bond IV The energy of single bonds and the relative electronegativity of atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570−3582.
(68) Sanderson, R. T. An Interpretation of Bond Lengths and a Classification of Bonds. *Science* **1951**, *114*, 670−672.
(69) Bratsch, S. G. A Group Electronegativity Method with Pauling Units. *J. Chem. Educ.* **1985**, *62*, 101−103.

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(70) Mullay, J. Calculation of Group Electronegativity. *J. Am. Chem. Soc.* **1985**, *107*, 7271−7275.
(71) Yang, Z. Z.; Wang, C. S. Atom-bond electronegativity equalization method. 1. Calculation of the charge distribution in large molecules. *J. Phys. Chem. A* **1997**, *101*, 6315−6321.
(72) Smith, D. W. Group electronegativities from electronegativity equilibration - Applications to organic thermochemistry. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 201−205.
(73) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(74) No, K. T.; Grant, J. A.; Jhon, M. S.; Scheraga, H. A. Determination of Net Atomic Charges Using a Modified Partial Equalization of Orbital Electronegativity Method. 2. Application to Ionic and Aromatic-Molecules as Models for Polypeptides. *J. Phys. Chem.* **1990**, *94*, 4740−4746.
(75) Rappe, A. K.; Goddard, W. A. Charge Equilibration for Molecular-Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358−3363.
(76) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.

DOI: 10.1021/acs.jcim.9b00581
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX