JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

Article

Atom Types Independent Molecular Mechanics Method for Predicting the Conformational Energy of Small Molecules

Zhaomin Liu, Stephen J. Barigye, Moeed Shahamat, Paul Labute, and Nicolas Moitessier

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/acs.jcim.7b00645 • Publication Date (Web): 18 Dec 2017

ACS Publications

Journal of Chemical Information and Modeling is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Atom Types Independent Molecular Mechanics Method for Predicting the Conformational Energy of Small Molecules.

*Zhaomin Liu,[1] Stephen J. Barigye,[1] Moeed Shahamat,[1‡] Paul Labute,[2] Nicolas Moitessier[1]**

[1]Department of Chemistry, McGill University, 801 Sherbrooke St. W., Montréal, QC, Canada H3A 0B8; [2]Chemical Computing Group Inc., 1010 Sherbrooke St. W., Montréal, QC, Canada H3A 2R7.

ABSTRACT

We previously implemented a well-known qualitative chemical principle into an accurate quantitative model computing relative potential energies of conformers. According to this principle, hyperconjugation strength correlates with electronegativity of donors and acceptors. While this early version of our model applies to σ bonds, lone pairs, disregarded in this early version, also have a major impact on the conformational preferences of molecules. Among the well-established principles used by organic chemists to rationalize some organic chemical behaviors are the anomeric effect, the alpha effect, basicity and nucleophilicity. These effects are directly related to the presence of lone pairs. We report herein our effort to incorporate lone pairs into our model to extend its applicability domain to any saturated small molecules. The developed model H-TEQ 2 has been validated on a wide variety of molecules from polyaromatic molecules to carbohydrates and molecules with high heteroatoms/carbon ratios. Interestingly, this method, in contrast to common force field-based methods, does not rely on atom types and is virtually applicable to any organic molecules.

INTRODUCTION

Molecular Mechanics, Force Fields and Potential Energy Surfaces. One application of Molecular Dynamics (MD) simulations is the study of interactions between ligands and macromolecules of therapeutic interest and/or the investigation of the conformational behavior of these complexes over time. The analysis of these interactions in molecular systems entails the routine exploration of the corresponding potential energy surfaces (PES). While quantum mechanical (QM) methods may offer high quality computations of the PES and thus accurate estimations of the ligand-macromolecule interaction energy,[1] the size of the biomolecular systems considered in drug discovery paradigms renders QM methods computationally intractable for proper exploration of the PES. In this context, the low computational cost molecular mechanics (MM) methods are typically employed.[2] However, the accuracy of MM methods largely depends on the quality of the potentials generated by the underlying empirical force fields (FFs).

Briefly, a FF is a linear combination of potential functions that describe the interactions between bonded atoms (e.g., stretching, bending, and torsions) and non-bonded atoms (e.g., van der Waals (vdW), electrostatics and hydrogen bonding) in a molecular system (Eq. 1) together with a set of parameters. Examples of commonly employed FFs include OPLS,[3-5] AMBER,[6-9] CHARMM[10,11] and MMFF94.[12-16] The accuracy of the potentials obtained with different FFs critically depends on their appropriate parameterization (see below) for the functional groups of interest and the suitability of the employed functional forms.

$$E = E_{bonds} + E_{angles} + E_{torsions} + E_{out\text{-}of\text{-}plane} + E_{van\ der\ waals} + E_{electrostatics} \quad (1)$$

Force Fields and Atom Types. At the core of many FF development protocols is the concept of *atom types*, which define the chemical environment of atoms based on atomic properties such

as the element type and hybridization state. Thus, each atom in a molecule under investigation will be assigned an atom type (e.g., oxygen atom in an ether, nitrogen in an ammonium) and the associated parameters regardless of the other distant functional groups present in the molecule. In this context, FF parameterization consists in the determination of the optimum force constants and equilibrium values (for each function in Eq. 1) for each of the defined atom types. This parameter fitting procedure is performed using experimental and/or high level QM data as references. However, the consensus is that no particular FF is capable of providing accurate predictions of bonding and non-bonding interactions for all structures in the small molecule (e.g., drug) space with an estimated number of 10^{60} possible members, as some uncommon functional groups (e.g., polyfunctionalized heterocycles) may not have any parameters available.[17-20] In an effort to address this limitation, two major approaches have been envisioned. On one hand, automated FF toolkits for systematic derivation of parameters have been developed (e.g., lsfitpar, FFBuilder, GAAMP),[17,21-23] although this limits the throughput of simple MM calculations. On the other hand, efforts have been placed on developing generalized FFs such as GAFF for organic compounds[9], UFF[24] and CGenFF[11]. Conversely, FFs specifically parametrized for particular classes of molecules (e.g., carbohydrates,[25-28] RNA,[29,30] and lipids[31]) have been developed to deal with specific chemical environments (e.g., GLYCAM[32,33], ECEPP[34,35]), thus alleviating the burden of using large training sets, although at the expense of a reduced applicability domain. Notwithstanding these initiatives, recent studies have indicated that current FFs remain unsatisfactorily parameterized for drug-like molecules.[18,36] These studies underscore the poor transferability of atom-type based FFs. Thus, the need to develop alternative MM methods/strategies with greater transferability cannot be overemphasized since it is implausible to develop parameters for all possible molecules.

Encoding Chemical Principles for Conformation Evaluation. Over the years, our research philosophy has been to incorporate classical organic/medicinal chemistry principles, often invoked to rationalize experimental observations, into computational chemistry models. This approach has guided the development of accurate methods for molecular modeling, implemented in our computational programs ACE (e.g., Hammond-Leffler postulate and Curtin-Hammett principles) and FITTED (e.g., pKa and covalent bonding).[37-40] This philosophy has been further applied to develop an MM method to compute small molecules' torsional potential energies based on chemical principles named as H-TEQ (Hyperconjugation for Torsional Energy Quantification).[19] This mechanism-based method, in contrast to atom type based methods, can be applied to virtually any molecules. In a proof-of-principle study, we first sought to comprehend the physical origins of the torsion energy term employed in FFs (Eq. 1), i.e., the influence of the torsion strain on the molecular conformational preferences. Our ultimate aim was to derive quantitative and chemically intuitive models for computing the torsion potential energy. In a previous work, we first identified that hyperconjugation was the major factor that controls torsional rotation in single bonds, in agreement with other reports in the literature.[41-43] We then developed rules to quantify the hyperconjugation energy based on the electronegativity of atoms; organic chemists often relate hyperconjugation strength with the electronegativity of the hyperconjugation acceptor to rationalize the anomeric effect, gauche effect and other experimental observations. As a validation of H-TEQ, the developed rules were employed in predicting the torsion energy profiles of several molecules focusing on rotations about C-C and C-N(+) bonds. The impact of various electron-donating and withdrawing substituents was accurately modeled. These encouraging results led us to extend this mechanism-based approach to all single bonds. We describe herein our initiative to refine and extend our method H-TEQ to

include $n \rightarrow \sigma^*$ hyperconjugation interactions and to investigate the influence of different stereoelectronic effects on the torsional rotation.

FACTORS MODULATING TORSIONAL ENERGY FOR MM IMPLEMENTATION

Hyperconjugation and Conformation. Conformational changes involve an interplay between repulsive and stabilizing interactions (e.g., electrostatic, Pauli's exclusion principle, hyperconjugation, London dispersion force).[41] Although there exists no consensus on which factor plays an overriding role in defining conformational preferences in saturated systems, the hyperconjugation model has been often evoked to rationalize the conformational behavior of some molecules. Examples are the stabilization of the staggered conformation in ethane over the eclipsed one as well as the gauche effect in 1,2-dihaloalkane derivatives,[44] although this has been challenged.[45] Quantitatively, we demonstrated that explicitly assembling the hyperconjugation energy with vdW and electrostatics potentials allowed to accurately reproduce the potential energy profiles obtained with high-level quantum mechanics calculations, indicating that hyperconjugation was a key component of the torsion energy.[19]

Lone Pairs in Organic Chemistry. The gauche effect can now be reproduced by our model H-TEQ 1.4 which evaluates the impact of $\sigma \rightarrow \sigma^*$ on the conformation of small organic molecules. However, other effects such as the anomeric effect, basicity/nucleophilicity and the alpha effect cannot be explained by this type of hyperconjugation. For these three effects, lone pairs must be considered. More specifically, the anomeric effect (preference for the axial position over the equatorial position for the aglycone in carbohydrates) is usually explained by the $n \rightarrow \sigma^*$ hyperconjugation stabilization shown in Figure 1.

The high nucleophilicity observed when a lone-pair containing heteroatom is oriented α to the nucleophile was first investigated in the early 60's when Jencks and Carriuolo observed that hydroxylamine was a lot more nucleophilic than suggested by its pKa.[46] The term "alpha effect", coined by Edwards and Pearson[47] was originally rationalized as the stabilization of the developing positive charge in the transition state by the adjacent lone pair electrons. Ten years later, the frontier molecular orbital (FMO) theory was used to rationalize this phenomenon.[48] According to this theory, the presence of adjacent lone pairs increase the HOMO energy and the alpha effect is highly dependent on the conformation (low for hydrazine most stable conformation shown in Figure 1b).

As the anomeric effect and alpha effect result from the modulation and interaction of the lone pairs FMOs, these effects should be incorporated into our model for the accurate modeling of lone-pair containing molecules.

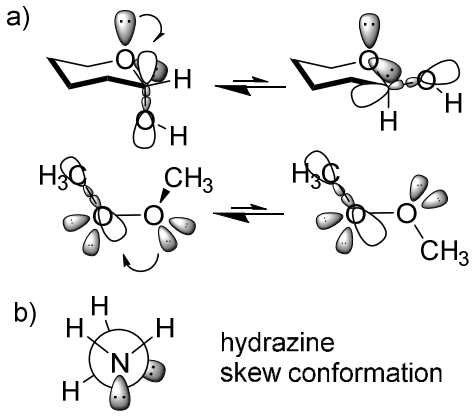


Figure 1. Conformational preference in small molecules resulting from the a) anomeric effect and b) alpha effect

Hyperconjugation involves the transfer of electrons from filled (donor) to empty (acceptor) molecular orbitals (MOs). Thus, the magnitude of the hyperconjugation energy should be related to the energy difference between the donor and acceptor MOs. Computational studies have

demonstrated that the $n \rightarrow \sigma^*$ interactions yield stronger hyperconjugation than the $\sigma \rightarrow \sigma^*$ interactions, given the smaller MO energy gap of the former.[49,50] Therefore, hyperconjugation should play an even more dominant role in dictating the conformational preference of structures with atoms containing lone pairs.[51] It is also important to note that hyperconjugation is influenced by the magnitude of orbital overlap between filled and empty orbitals and thus, geometric orientations that allow for greater spatial proximity would lead to stronger hyperconjugation stabilization.

Deriving Torsional Parameters in Force Fields. Parameterization of torsional energy terms is one of the most challenging phases in FF development, due to the large number of possible torsion types. Notwithstanding the concerted efforts to develop parameters to encompass representative torsion types in drug-like molecules, it has been acknowledged that torsion parameters for new molecules will continue to be missing, underscoring the poor transferability of current torsion parameters.[18] Deriving torsion parameters for a given atom type is commonly performed by fitting the truncated Fourier series (Eq. 2) to QM torsional profiles.

$$E_{torsion} = \sum_{1}^{4} \frac{V_n}{2}(1 + \cos(n\theta + \varphi)) \qquad (2)$$

As previously demonstrated in our study on torsions made of carbons and ammonium nitrogens (sp3, no lone-pairs), the hyperconjugation energy is a major component of the torsional potential.[19] However, when computed for torsions featuring lone pairs, hyperconjugation varies in amplitude, nature (i.e., $n \rightarrow \sigma^*$ or $\sigma \rightarrow \sigma^*$) and/or number (e.g., $n \rightarrow \sigma^*$ is unidirectional since there are no acceptor lone pair MOs). Additionally, the influence of electron withdrawing substituents on the $n \rightarrow \sigma^*$ hyperconjugation energy is significantly stronger than the subtle electronic effect on $\sigma \rightarrow \sigma^*$.[52] Therefore, for lone pair containing molecules, we thought that

evaluating $n \rightarrow \sigma^*$ and $\sigma \rightarrow \sigma^*$ interactions individually and then combining their contributions into a global hyperconjugation energy would yield more accurate torsional profile predictions. Common empirical force fields are typically trained on sets of molecules and implicitly include all the effects. However, classical torsional parameterization does not consider hyperconjugation explicitly, hence the common FF parameters derived for torsion types containing lone pairs do not explicitly distinguish contributions from $\sigma \rightarrow \sigma^*$ and $n \rightarrow \sigma^*$ interactions, despite their different amplitudes.

Hyperconjugation Involving Lone Pairs and Impact on Molecular Conformational Preferences. In order to fully examine the contribution of lone pairs to the hyperconjugation stabilization, we assembled a set of molecules (Figure 2) to specifically investigate $n \rightarrow \sigma^*$ hyperconjugation and its impact on the conformational preferences of molecules. The molecules in this set cover various functional groups such as amines, alcohols, phosphines and thiols, and the bonds in the center of the considered torsions (referred to as central bonds) include a lone-pair containing atom. Additionally, we included molecules in which both atoms forming the central bond contain lone pairs and thus exhibit the alpha effect (e.g., hydrazines, peroxides, disulfides and hydroxylamines). It is important to highlight that the aim of the designed set is not to cover the entire chemical space of fragments associated with lone pairs but rather to include the basic fragments with specific variations in electron donor and acceptor abilities. Consistent with our goal of developing a transferable MM method, rules developed based on the $n \rightarrow \sigma^*$ hyperconjugation profiles of these representative molecular systems should be applicable to the modeling of $n \rightarrow \sigma^*$ hyperconjugation in any molecular system. As we aimed to develop an

integrated method covering both the *n*→σ* and σ→σ* hyperconjugations, methylsilanes as well as the previously investigated ethane derivatives were aggregated to this set.

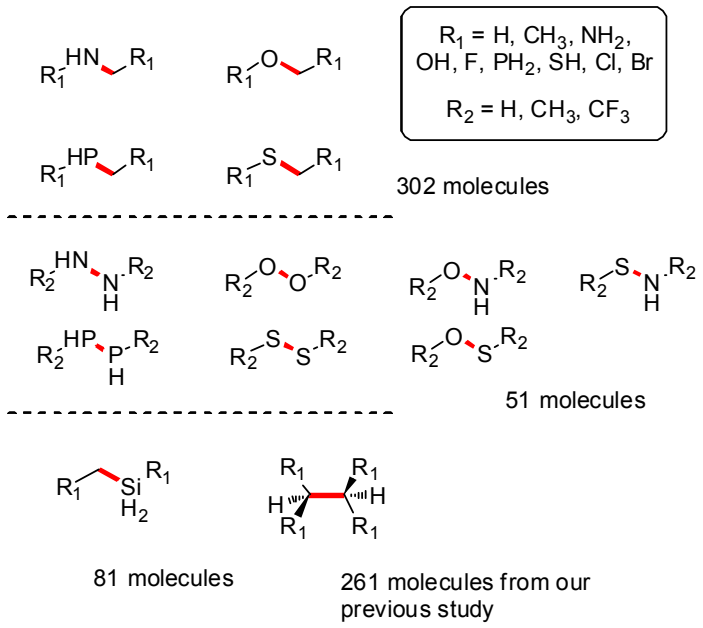


Figure 2. Sets of small molecules studied for hyperconjugation

Conformational Preferences in Molecules Featuring Lone Pairs. In molecules with lone pair containing heteroatoms, the energy of a given conformation depends on a combination of vdW, electrostatics and σ→σ* and n→σ* hyperconjugation contributions. A torsion scan for the H-C-N-H torsion in methylamine revealed an expected symmetrical conformational preference with a period of 3 and minima for the staggered geometry. This outcome results from an interplay between steric and hyperconjugation contributions. However, the preferred geometry changes dramatically when a strong acceptor, such as the C-F bond in fluoromethylamine is introduced (Figure 3): the lowest energy conformation has the lone pair and fluorine in the antiperiplanar conformation. Moreover, the local energy minimum is associated with the dihedral angle Θ (n-N-C-F) set to 0°. As we previously reported, hyperconjugation favors the *anti* geometry over the *syn* geometry. This preference is due to the favorable orbital alignment which

allows for strong donor-acceptor stabilization.[19] Thus, although the eclipsed conformation is sterically disfavored, the strong $n \rightarrow \sigma^*$ hyperconjugation interaction in fluoromethylamine compensates for the steric repulsion. Similar changes in preferred conformations are observed in alcohols, phosphanes and thiols. A discussion on the similar behavior of other small molecules can be found as supporting information (Section S1).

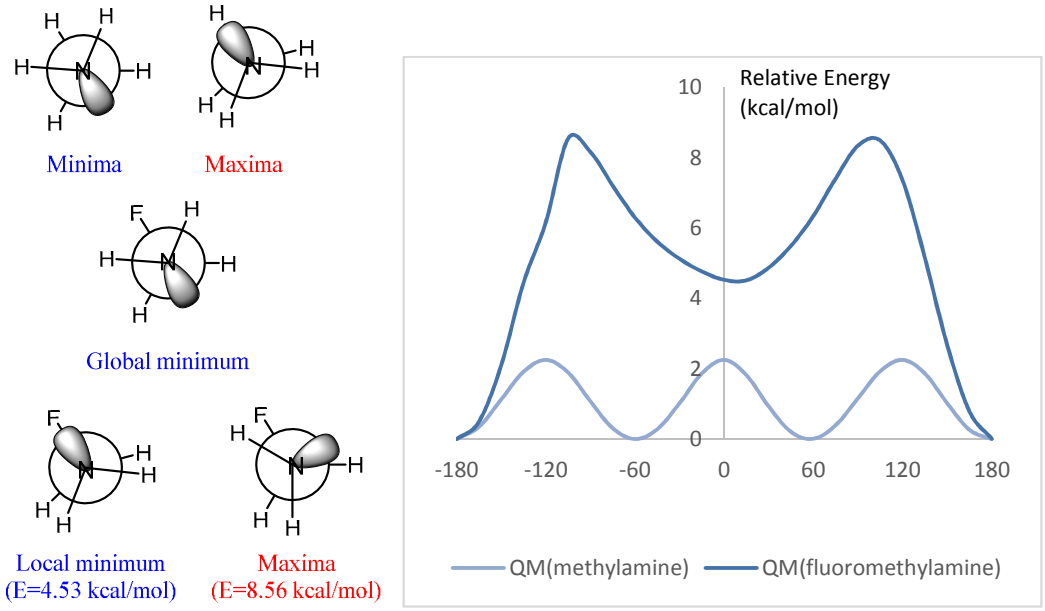


Figure 3. Comparison of the preferred geometries of amines. QM: Quantum Mechanics MP2/6-311+G**.

Extracting Hyperconjugation. In order to complement this description of the impact of $n \rightarrow \sigma^*$ hyperconjugation on the conformation preferences of saturated heteroatomic molecules, we decided to evaluate the contribution of hyperconjugation energy in modeling the overall torsion potential. To this end, we combined the hyperconjugation energy (for $\sigma \rightarrow \sigma^*$ and $n \rightarrow \sigma^*$ interactions) computed by Natural Bond Orbital (NBO) analysis[53] with vdW and electrostatic potentials (obtained from GAFF), and compared the resulting torsion profile with that obtained with high level QM calculations (MP2/6-311+G**). With the substitution of the empirical

torsion term used in GAFF with the NBO-derived hyperconjugation energy, higher accuracy (low deviations relative to the QM profiles) for molecules with strong or weak hyperconjugation acceptors was obtained (Figure 4**Figure 4.**). It is worth mentioning that current FFs, more specifically - GAFF, yielded low deviations for molecules with no substituents (for example methanol in Figure 4a). However, with the introduction of additional functional groups to these molecules (such as bromine in Figure 4b), inaccurate predictions for the energy minima and maxima positions were observed (additional case studies can be found as supporting information, Figure S2**Figure 4.**). This result suggests that our MM method for computing the torsional energy based on the hyperconjugation principle, may exhibit greater transferability.

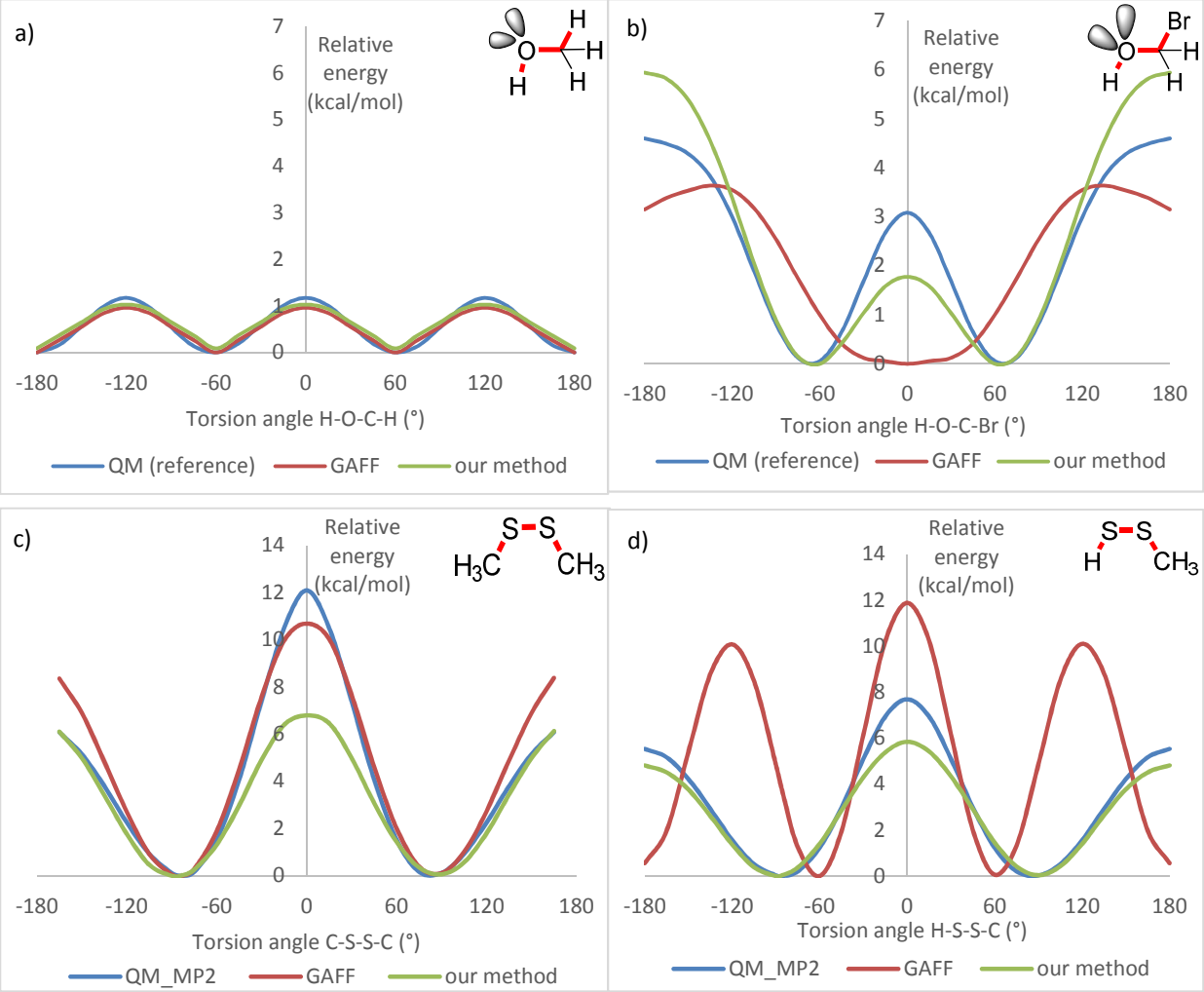


Figure 4. Comparison of the accuracy between our method (GAFF torsion term substituted by NBO-derived hyperconjugation energy) with GAFF on representative molecules - a) methanol, b) bromomethanol, c) dimethyldisulfide, d) methyldisulfide. Our method: NBO-derived hyperconjugation replacing the torsion energy term in GAFF.

We also analyzed the conformational preferences of molecules with both atoms of the central bond featuring lone pairs such as hydrazine, peroxides and disulfides (Figure 2, Figure 4-c,d). Similar to the previous cases (a single atom featuring lone pairs), we observed higher transferability of our method compared to the atom-type based GAFF. As shown with both dimethyl disulfide and methyldisulfide (Figure 4-c,d), our method correctly predicts the positions of the energy minima and maxima with the overall root mean square deviations (RMSDs) of 1.79 kcal/mol and 0.68 kcal/mol, respectively. Although GAFF predicts more accurately the energy profile of dimethyl disulfide (RMSD = 0.97 kcal/mol), it does not predict the correct positions of the local minima and maxima for methyl disulfide, resulting in a much higher RMSD value (4.49 kcal/mol). Once more the poor transferability of the atom-type method is demonstrated. While GAFF includes a specifically trained torsion parameter (c3-ss-ss-c3) for C-S-S-C used for dimethylsulfide, it does not possess a parameter for C-S-S-H, required for the methyl disulfide torsional energy prediction. As a result, a generic GAFF torsion parameter (X-ss-ss-X) was selected for modeling methyldisulfide and clearly turned out to be inappropriate.

Overall, our method underestimated the energy barriers of the molecules we've studied, for example the disulfide molecules illustrated in Figure 4 (additional examples can be found as supporting information, Figure S2**Figure 4.**). After visual inspections of the corresponding conformations, we believed that these lower barriers were resulting from the missing lone pair – lone pair (lp-lp) repulsion. In the most MM methods, electrostatic interaction is computed using

atomic charges, with no explicit treatment of the lone pairs. As a result, the repulsion between two lone pairs could not be captured by the computed electrostatic interaction. Thus, an additional energy component for lp-lp repulsion is required.

Factors Controlling Hyperconjugation: Energy Gap between Donor and Acceptor Molecular Orbitals. Up to now, hyperconjugation was computed using QM methods. With the demonstration that hyperconjugation energy with lone pairs can substitute the empirical torsion term in common FFs, we then needed to derive a method to predict the hyperconjugation stabilization produced by lone pairs in a high-throughput manner. We started by investigating the factors influencing $n \rightarrow \sigma^*$ hyperconjugation and the relations between these factors and atomic and/or bond properties.

The hyperconjugation energy E_{hyp} for the $n \rightarrow \sigma^*$ interaction is qualitatively described as a function of (1) the overlap between the donor and an empty anti-bonding orbital and (2) the energy gap (ΔE) between these two orbitals (the smaller, the better).[42,50] Consistent with FMO principles, the lone pair MOs are higher in energy than the σ MOs. Thus, ΔE between n and σ^* is anticipated to be smaller than that between σ and σ^* while the hyperconjugation energy is expected to be larger. Indeed, our computations, summarized in Table S1 (supporting information), confirmed that ΔE (n - σ^*) is smaller than ΔE (σ - σ^*) even with a much stronger acceptor such as $\sigma^*_{(C\text{-}F)}$, and the $n \rightarrow \sigma^*$ interactions are associated with stronger hyperconjugation energy relative to $\sigma \rightarrow \sigma^*$ (refer also to Figure S3 in supporting information). Note that in a given row of the periodic table, the introduction of more electronegative atoms to the C-X σ bonds results in lower energies for both the bonding and antibonding orbitals, and in lower donor ability and increased acceptor ability.[19] Therefore, the general acceptor ability trend follows the (decreasing) order: $\sigma^*_{(C\text{-}F)} > \sigma^*_{(C\text{-}O)} > \sigma^*_{(C\text{-}N)} > \sigma^*_{(C\text{-}C)}$.

There are two lone pairs involved in the n→σ* hyperconjugation for oxygen and sulfur. In the "valence shell electron pair repulsion" (VSEPR) theory, a rabbit-ear model has been used to describe the two lone pairs, considering them as two equivalent sp^3 orbitals. This model is used to rationalize the different angles in alkanes, amines and alcohols. Another model considers a combination of p- and sp- orbitals for the two lone pairs,[54] which has been used in NBO analysis.[55,56] We opted for the model with two equivalent lone pairs since the p+sp model would require two separate rules to model the energy profiles for the p- and sp-type lone pairs. A discussion on these two models can be found as supporting information, S3.

Role of Orbital Overlap in Hyperconjugation Stabilization. Next, we analyzed the contribution of the orbital overlap in the hyperconjugation stabilization of molecules with center atoms in the same column of the period table (e.g., amines and phosphines). While the energy gap for $\Delta E_{n(N) - \sigma^*(C\text{-}H)}$ (1.08 a.u.) and $\Delta E_{n(P) - \sigma^*(C\text{-}H)}$ (1.15 a.u.) are comparable (Table S1), the hyperconjugation stabilization energy for $n_N \rightarrow \sigma^*_{C\text{-}H}$ (10.66 kcal/mol) is much greater than $n_P \rightarrow \sigma^*_{C\text{-}H}$ (3.81 kcal/mol, Table S1). The overriding factor in this case is the orbital overlap which we found greater between n_N and $\sigma^*_{C\text{-}H}$ orbitals ($\Delta S_{ij} = 0.203$, Table S1) than between n_P and $\sigma^*_{C\text{-}H}$ ($\Delta S_{ij} = 0.089$). This difference results from a shorter distance between the n and σ* orbitals in the case of the former, which favors stronger donor-acceptor interactions.[52,57]

Given the influence of the orbital overlap on the hyperconjugation stabilization energy, we sought to investigate the interdependence between the orbital overlap and properties associated with the donor-acceptor bridge, such as the bond length (l) and electronegativity. To this end, we evaluated the relationship between the n and σ^* orbital overlap (ΔS_{ij}) and the bond lengths and electronegativities for C-N and C-P central bonds in a set of functionalized amines (NH_2CH_2X) and phosphines (PH_2CH_2X). Interestingly, we observed that the overlap is related to both the

electronegativity of the X atom (direct correlation[42]) and l^3 (inverse correlation, Figure 5). The linear correlation between overlap and the electronegativity of X is consistent with a previous study which demonstrated that an increase in the electronegativity of X in a C-X bond results in higher polarization of $\sigma*$ towards C, which favors greater n and $\sigma*$ orbital overlap. Likewise, it is plausible that the orbital overlap exhibits an inverse relationship with l^3 since smaller distances between the orbitals favor greater overlap and vice-versa. Therefore, it is anticipated that incorporating bond lengths to H-TEQ method should allow for greater refinement and more accurate predictions of torsion energy profiles.

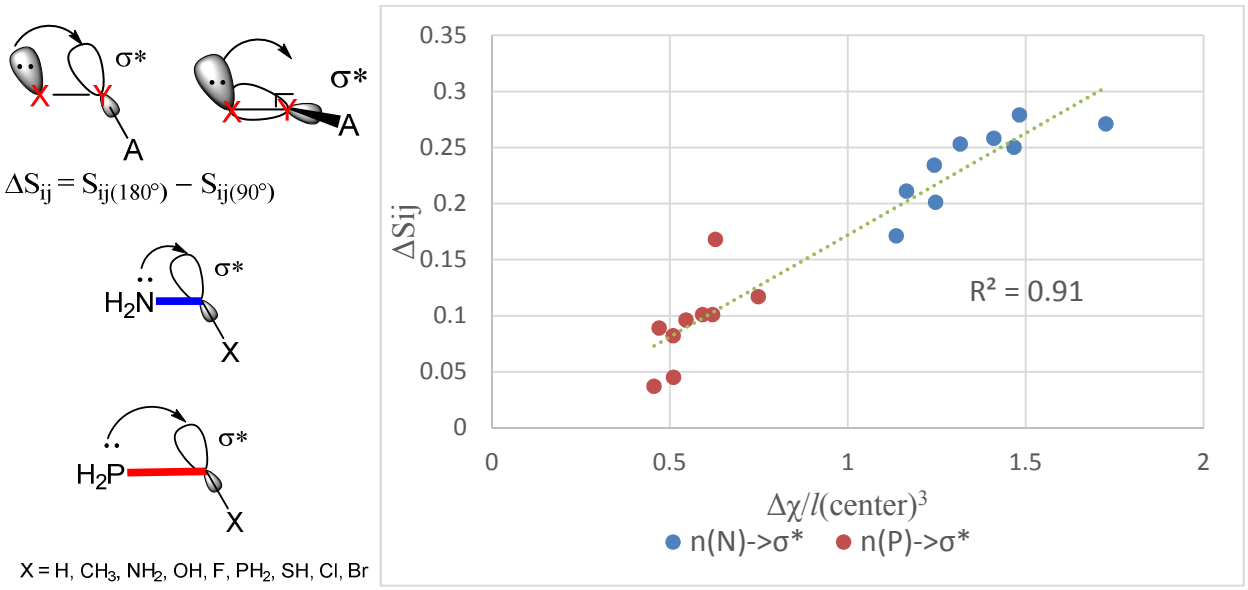


Figure 5. Correlation between orbital overlap and electronegativity as well as central bond length. ΔS_{ij}: orbital overlap index computed by NBO analysis.

The bond length at the donor site and acceptor site also modulates the orbital overlap. Using different acceptors as examples, we first compared the $\sigma^*_{C\text{-}X}$ acceptor ability of the atoms down the group of the periodic table (i.e. F, Cl and Br). Based on the electronegativity differences, we would expect $\sigma^*_{C\text{-}F}$ to be the strongest acceptor and $\sigma^*_{C\text{-}Br}$ the weakest in the halogen group

(iodine has not been investigated). However, our results indicated an inverse order: C-Br > C-Cl > C-F (Table S1), consistent with previous studies reported in the literature.[42,52] This is attributed to a decrease in the $\sigma^*_{C\text{-}X}$ orbital energy with longer C-X bonds, resulting in smaller energy gaps down the group of the periodic table. This bond length effect in a way modulates the *effective* electronegativity on predicting the hyperconjugation. The detailed development of rules to include the bond length and electronegativity factors for predicting hyperconjugation will be discussed in the following sections.

Lone Pair-Lone Pair Interaction and Torsional Energy. The impact of lone pair-lone pair interactions on the torsional energy profiles of hydrazine and hydroxylamine have been previously investigated by Mo and co-workers.[58] In their report, the rotational energy barriers were attributed to a combination of hyperconjugation and steric effects suggesting that the H-TEQ approach should be appropriate for this class of molecules. In order to analyze these rotations in detail, hydrazine and hydroxylamine were investigated more in depth (Table S1, entries 13 and 4). When comparing methylamine with hydrazine and hydroxylamine, our computations suggested that the *n* and *σ** orbital overlap is significantly reduced for the latter two. The lone pair repulsion likely distorts the two lone pairs hence reducing the proximity with the acceptor σ*. This phenomenon can be viewed as the *shielding* of σ* by lone pairs which is more pronounced with hydroxylamine (two lone pairs on the acceptor side, entry 14 in Table S1) than with hydrazine (one lone pair, entry 13 in Table S1).

On top of the shielding effect which reduced the hyperconjugation of lp-X-Y-σ*, lp-lp repulsion has also not been considered in the current MM. Thus, we implemented both effects explicitly in our method.

IMPLEMENTATION OF n→σ* HYPERCONJUGATION

Previous Implementation. With a comprehensive analysis of the factors that influence hyperconjugation interactions, we set out to extend the formerly developed rules (H-TEQ series1) to *n*→σ* hyperconjugation interactions. We previously focused exclusively on σ→σ* and employed the electronegativity to derive rules to predict the torsional energy of molecules.[19] With the aim of developing a chemical principles-based approach, we accounted for the energy gap by considering the electronegativity differences (Δχ) between the acceptor and donor. Additionally, in order to consider the polarization of the central bond, the electronegativity (χ) of the atoms constituting the central bond were included (Δχ). Note that the defined rules in the first version of H-TEQ followed period-specific trends (*i.e.,* equations developed for donor and/or acceptor elements from a particular row could not be extrapolated to other rows). Therefore, different coefficients were employed in the H-TEQ equations depending on the element periodicity of the donor and acceptor.

Development of H-TEQ 2. Bearing in mind that element periodicity implicitly includes the role of the element size and polarizability, we sought to incorporate other observables to the H-TEQ approach to explicitly model these effects and reduce H-TEQ 2 to a single equation. To this end, we compiled the sets of molecules employed in the analysis of the *n*→σ* and σ→σ* interactions in previous sections into a single dataset.

The hyperconjugation torsion profiles computed using NBO were expressed as a Fourier series (Eq. 5) and then fitted to derive the $V_{1\text{-}3}$ parameters for each torsion of the set. Equation 5 has been used to compute the torsion potential in many FFs. As discussed previously,[19] from a quantitative perspective, the V_1 parameter accounts for the *syn*/*anti* conformational preference while V_2 characterizes the amplitude of the hyperconjugation energy; the latter may vary from 2-

30 kcal/mol depending on the donor and acceptor. V_3 is considered as a correcting factor in the scale of 0.3-1.2 kcal/mol for all types of hyperconjugation. Compared to V_1 and V_2, the variation of V_3 is minimal and we thus decided to reduce the number of terms to develop and develop equations only for V_1 and V_2, while keeping V_3 as a constant as proposed in H-TEQ 1.4 (Table S2).[19]

$$E_{hyp} = \frac{V_1}{2}(1+\cos\theta) + \frac{V_2}{2}(1+\cos(2\theta)) + \frac{V_3}{2}(1+\cos(3\theta)) \quad (5)$$

Developing Equations for Deriving V_1 and V_2. In the section above, $V_{1\text{-}3}$ parameters have been derived from NBO data. However, the NBO analysis is computationally involved, as it requires the computation of MOs using high-level QM methods. As achieved previously with the σ→σ* interactions, we intend to derive these parameters on-the-fly with no recourse to expensive computations. From the NBO based study, we found that the bond length, a factor not considered in H-TEQ 1, played an important role in hyperconjugation. So, in the revised version of H-TEQ, the expression for Δχ has been redefined to include the lengths of the donor (l_D), acceptor (l_A) and central (l_C) bonds (Eq. (6-9)) to take into account element periodicity. The central bond length influence was first integrated as a factor of $1/l^3$. l_A and l_D, have also been introduced into effective electronegativities $\chi_{A\text{-effective}}$ and $\chi_{D\text{-effective}}$.

As in our previous report, group electronegativity (derived from element electronegativities) is used for the donors, acceptors and neighbors while Pauling electronegativities were used for the atom of the central bond.[19] The effect of neighboring groups (R_1, R_2, R_3 and R_4 in Figure 6) has also been introduced into this new model. Based on our previous study, high electronegative groups on either donor or acceptor side reduce the hyperconjugation strength.[19] Hence the

electronegativities of donors and acceptors (χ_A and χ_D) were refined using the electronegativities of the neighboring groups ($\chi_{neighbor}$).

The acceptor and donor sites are assigned based on their electronegativity differences (i.e., A is more electronegative than D). If D and A have the same electronegativity values, the site with higher electronegative neighbors is considered as acceptor. It should be noted that in order to extend the same method to n→σ*, χ and *l* values for lone pairs were required and these were empirically defined. Qualitatively, lone pairs occupy higher energy MOs relative to the σ orbitals and are therefore in principle better donors. In this sense, the χ values for lone pairs should be small (Table S3).

Figure 6. Schematic representation of the all the factors considered to model hyperconjugation. Donor (D) and acceptor (A).

As discussed above, lone pair-lone pair interaction in R′-X-Y-R (X/Y=N, O, P, S) results a shielding effect over the σ* orbitals, which yields a reduced $n \rightarrow \sigma^*$ hyperconjugation amplitude. Therefore, this effect required special treatment. Since in H-TEQ 2 lone pairs have been considered as neighboring groups, an additional refinement (S_n) was applied to the acceptor neighboring effect (eq. 7) to take into account this lone pair shielding effect (S_n values for different heteroatoms are provided as supporting information, S3).

$$\Delta\chi = \frac{abs\left((\chi_{\text{centerA}} + \chi_{\text{A-effective}}) - (\chi_{\text{centerD}} + \chi_{\text{D-effective}})\right) - abs\left(\chi_{\text{centerA}} - \chi_{\text{centerD}}\right)}{l_{center}^3} \quad (6)$$

$$\chi_{A-effective} = (\chi_{\text{A}} - \sum \chi_{\text{neighbor-A}} * 0.12 - S_n) * l_A \quad (7)$$

$$\chi_{D-effective} = \frac{(\chi_{\text{D}} + \sum \chi_{\text{neighbor-D}} * 0.12)}{l_D} \quad (8)$$

$$V_2 = -3.86 * \Delta\chi^2 + 1.16\Delta\chi - 2.34 \quad (9)$$

The performance of the H-TEQ 2 in deriving accurate V_2 values has been assessed. The V_2 values derived from H-TEQ 2 has been compared with the V_2 generated by fitting the NBO-derived energy profile. A correlation coefficient (R^2) of 0.84 was obtained with our molecule set (Figure S5). A close look at the outliers revealed that the few NBO-derived V_2 greater than 25 kcal/mol are consistently underestimated by H-TEQ 2. These V_2 values belong to the $n_{(N)} \rightarrow \sigma^*$ with strong acceptors such as F, Cl, Br and will be discussed below.

Computing *Syn/anti*- Preference through Atomic Properties. V_1 models the *syn*-/*anti*-conformational preference induced by hyperconjugation. We observed higher V_1 values with longer acceptor bonds for atoms in different periods (i.e., F vs. Cl, Br; O vs. S and N vs. P as acceptor in Table S4. This could be related to the orbital overlap: for greater bond lengths, rotations from the *anti* to the *syn* conformation results in σ* being geometrically further away from the lone pair, significantly reducing the orbital overlap. Although higher V_1 values with longer acceptor bonds were also observed for the molecules with longer center bonds (i.e. phosphine), this net change is smaller resulting in an overall smaller orbital overlap. The same trends were observed when we focused on the donor site by comparing the donor bond lengths in ethane and methylsilane derivatives and the V_1 values. Using this knowledge, we proposed to model the *syn*/*anti* preference (V_1) using an equation expressed as a function of the bond lengths

exclusively with a direct relation to l_D and l_A and an inverse relation to l_C (Eqn. 10). For $n\rightarrow\sigma^*$, the lone pairs "bond length values" were those defined in the V_2 development (l_n in Table S3). The V_1 parameters for $n\rightarrow\sigma^*$ and $\sigma\rightarrow\sigma^*$ showed two distinct correlations with $l_{effective}$ for $n\rightarrow\sigma^*$ and $\sigma\rightarrow\sigma^*$. Thus, two equations (Eqn. 11, 12) with different coefficients will be used to compute V_1 ($n\rightarrow\sigma^*$) and V_1 ($\sigma\rightarrow\sigma^*$). For a comparison of predicted V_1 with NBO-derived V_1 refer to supporting information Figure S6.

$$l_{effective} = \frac{(l_D + l_A)}{l_C} \qquad (10)$$

$$V_{1(\sigma\rightarrow\sigma^*)} = 4.72 * l_{effective} - 6.58 \qquad (11)$$

$$V_{1(n\rightarrow\sigma^*)} = 10.13 * l_{effective} - 7.22 \qquad (12)$$

As discussed above, the lp-lp repulsion is not included in the electrostatic potentials computed by most MM methods. This repulsion is strong when the two lone pairs face each other (with the θ(n-X-Y-n) = 0°) and weak when they are in the *anti*-conformation. Thus, we could use the cosine function ($E = V_1/2 \cos(\theta)$) to model the lp-lp repulsion ($V_{1\text{lp-lp repulsion}}$ is in Table S3).

PERFORMANCE AND VALIDATION

As a validation, we applied H-TEQ 2 to a diverse set of molecules including: 1) ammonium derivatives (73 molecules) and the H-TEQ 1.4 validation set (59 molecules); 2) phosphonium derivatives (46 molecules) necessary to examine the ability to predict torsion potentials for dihedrals containing longer central bonds 3) aminoborane derivatives containing extremely polarized center bonds (15 molecules) 4) a subset of most structurally diverse molecules from the

MMFF94 validation dataset (100 molecules, Figure 7a), and 5) monosaccharides (36 molecules) (Figure 7b) for a total of 1022 molecules (including the molecules shown in Figure 2).

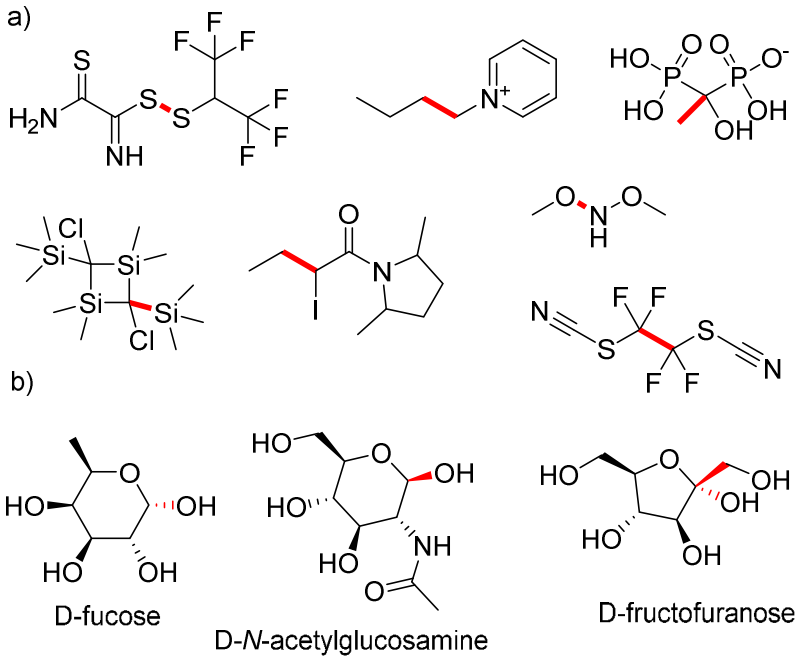


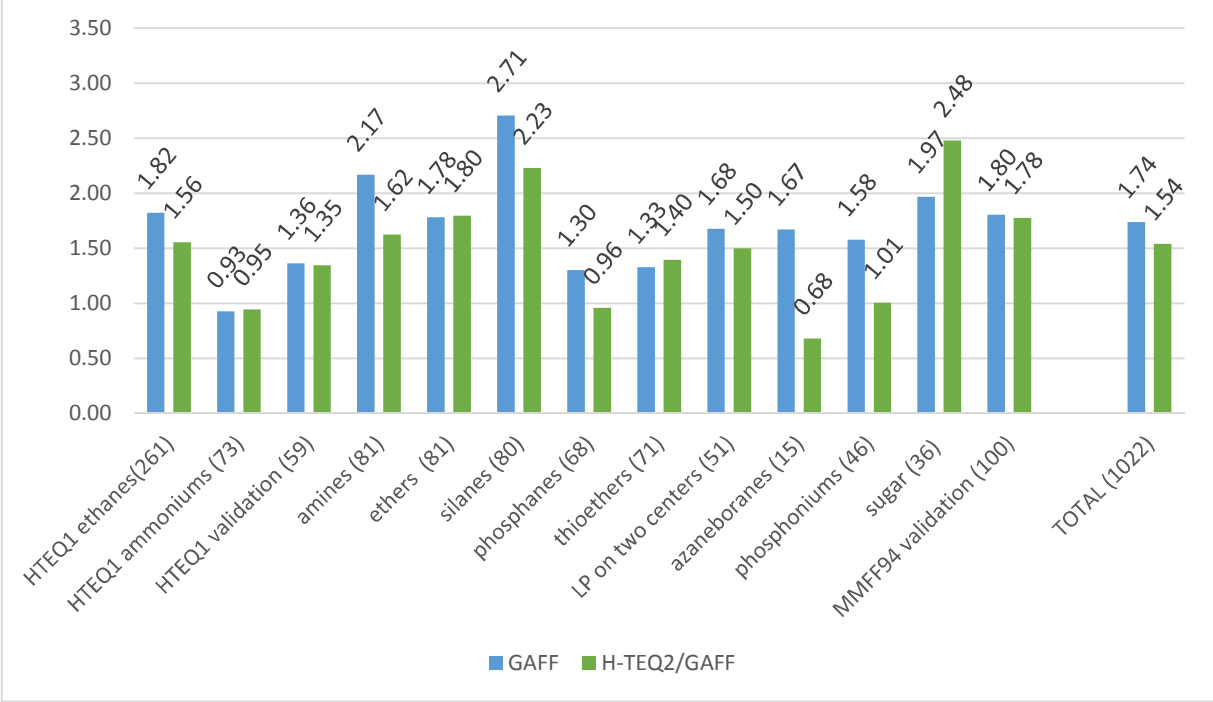
Figure 7. Representative molecules used for validation; bonds in red are the center bonds of the considered torsions.

As the developed method is intended for torsional energy quantification, we tested its accuracy by replacing the existent torsional components of three widely used FFs - GAFF, MMFF94 and Parm@Frosst by H-TEQ 2 while keeping the other energy terms, including bond, angle, vdW and electrostatics intact. We then compared the performance of the three resulting H-TEQ 2/FF and the original FFs in predicting QM torsion rotation profiles. Gratifyingly, H-TEQ 2 improved the accuracy of the three FFs (Table 1).

Table 1. Average RMSD of energy over the torsional profile of FF vs. QM reference (MP2/6-311+G**) (kcal/mol) obtained using a 5000-fold bootstrap validation.

1022 molecules	Current FF	H-TEQ 2
GAFF	1.74 ± 0.04	1.54 ± 0.04
MMFF94s	1.58 ± 0.04	1.44 ± 0.03
Parm@Frosst	1.87 ± 0.05	1.60 ± 0.04

A more detailed analysis showed that the H-TEQ 2 method led to an improvement in accuracy for most of the subsets (Figure 8) compared to the three FFs and to H-TEQ 1.4 (Table S5). This demonstrated that the incorporation of additional molecular properties to the H-TEQ method, in addition to yielding a more generalized framework for computing torsional energies, improved its applicability to different set of molecules.



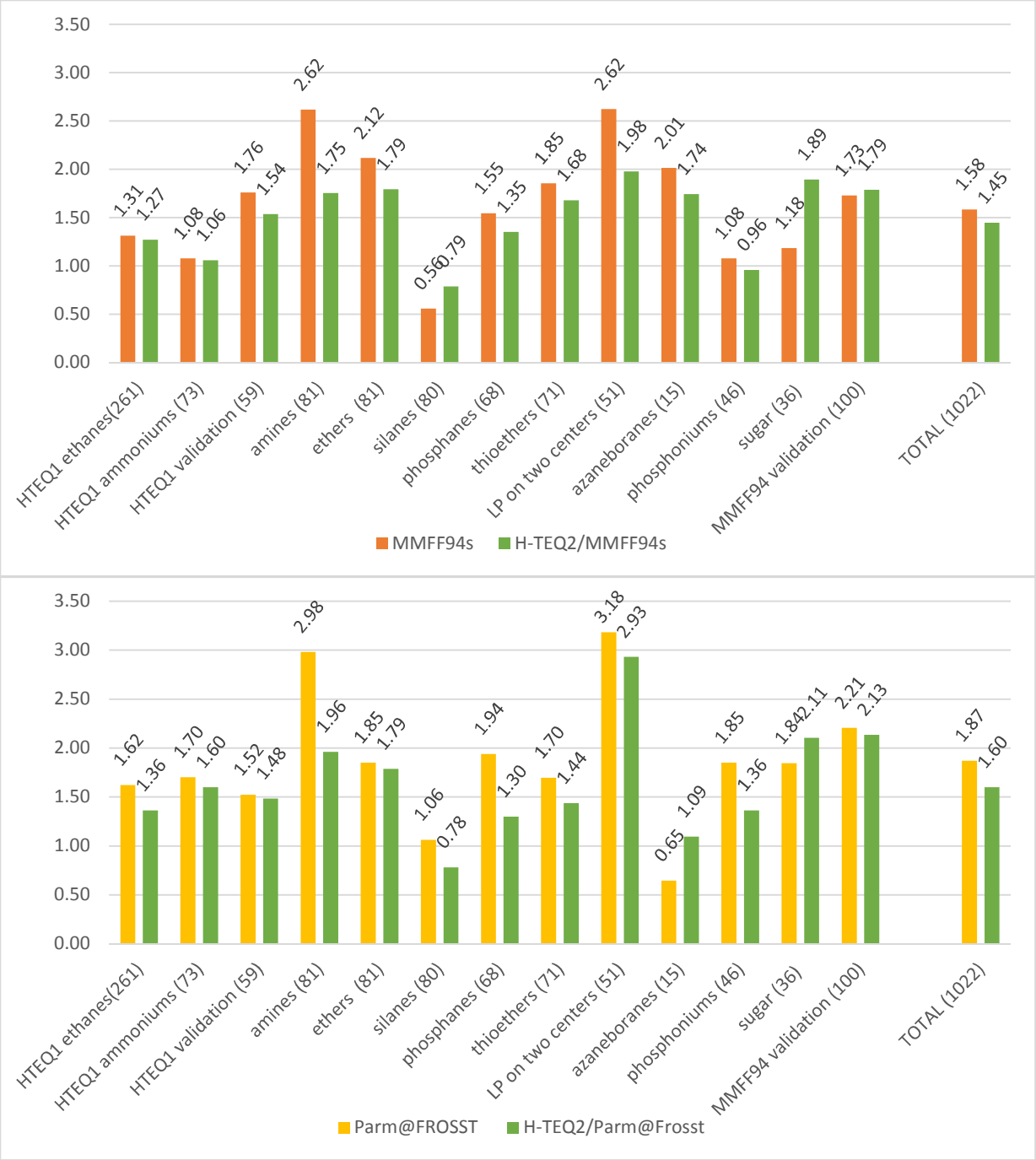


Figure 8. Comparison of the predicted accuracy between H-TEQ 2 and current FFs for each molecular dataset.

The accuracy of the developed method with torsions involving highly polarized bonds has also been assessed. For example torsional energy profiles for amino borane derivatives with the

extremely polarized N(+)-B(-) bond were accurately predicted with RMSD as low as 0.68 kcal/mol RMSD when H-TEQ 2/GAFF was employed. Note that the N(+)-B(-) bond was not considered in the method development, further demonstrating transferability of our method. Similarly, torsion profiles for phosphonium derivatives (featuring long C-P(+) central bonds) are more accurately predicted by H-TEQ 2 (with RMSDs dropping over the three FFs; analysis of the statistical significance of these results can be found as supporting information Table S6).

However, we observed that our approach does not improve the accuracy of any of the three FFs in predicting the torsion energy of the carbohydrate derivatives. A detailed analysis of the torsional profiles for these sugars revealed that this drop in accuracy is mainly attributed to the intramolecular hydrogen bonding and steric effects, which are not contemplated in the hyperconjugation model (examples can be found as supporting information, Figure S7). It is important to highlight that one of the key characteristics of current FFs is their self-consistency, in that constituent parameters are adjusted to accommodate additional effects for adequate fitting. Therefore, bearing in mind that our method doesn't include the retraining of other parameters, it is plausible that a loss of accuracy is observed for molecules where the hyperconjugation model is not amenable to rationalizing observed conformation behavior. Even then, our method seems not to be prone to critical compatibility issues as good accuracy is generally obtained across the different FFs and set of molecules.

Finally, our approach yields comparable accuracy to MMFF94 in predicting the torsional energies for the structurally diverse set of 100 molecules, randomly selected from a repository specifically designed to validate the same FF (Figure 8). H-TEQ 2 yields low accuracy for two molecules in this set, both of which contain a cyclopropane moiety. This is attributed to the fact that H-TEQ 2 computes the torsional energy based on the hyperconjugation model which

considers the orbitals and bonds as aligned (Figure 9). However, this model is inaccurate in the case of the cyclopropane moiety as its orbitals are bent to alleviate the ring strain. The unusual bond angles displace the torsion energy maximum while H-TEQ assumes regular C-C-C angles. As a result, the torsion profiles computed by our method are phase shifted (relative to the QM-derived profile). Distortions induced by small rings are formally excluded in the current version of the method.

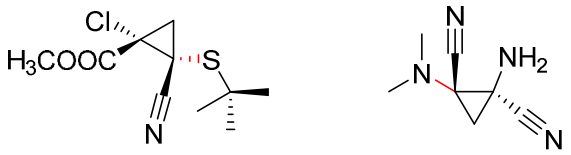


Figure 9. Cyclopropyl derivatives with inaccurate H-TEQ 2 torsion potentials.

CONCLUSION

Lone pairs significantly impact the torsional energy barriers of not only carbon-hetereoatom single bonds but also heteroatom-heteroatom single bonds. We have shown herein that the lone pair effects on the torsional energy can be modeled as a combination of hyperconjugation, vdW and electrostatic effects, allowing us to develop an updated version of H-TEQ with a wider applicability domain.

In contrast to H-TEQ 1.0-1.4, H-TEQ 2 relies on a single unified equation and models hyperconjugation as a function of electronegativity, bond lengths and lone pair-lone pair interactions. Two factors control the hyperconjugation strength: the energy gap between hyperconjugation donors and acceptors and the orbital overlap. In H-TEQ 2, while electronegativity was used to compute the energy gap, bond length was a descriptor used to account for orbital overlap. This new model was integrated into 3 well established FFs and led to

not only an improvement of their accuracies on a variety of molecules but also improved transferability to diverse functional groups.

Supporting Information. Addition Figures and Tables supporting and/or illustrating the conformational preferences of selected molecules are provided as supporting information, molecule sets are available as sdf files. This material is available free of charge via the Internet at http://pubs.acs.org. A Java-based implementation of the H-TEQ 2 method is freely available for download at http://moitessier-group.mcgill.ca/.

AUTHOR INFORMATION

Corresponding Author

* nicolas.moitessier@mcgill.ca.

ORCID

Nicolas Moitessier: 0000-0001-6933-2079

‡ current address: ChemEssen Inc., 812, Ace High Tech City 2, 25 Seonyu-ro, Yeongdeungpo-gu, 150-096 Seoul, South Korea

ACKNOWLEDGMENT

We thank NSERC (CRD program) for financial support. Calcul Québec and Compute Canada are acknowledged for generous CPU allocations.

EXPERIMENTAL SECTION

Hyperconjugation Energy Calculation from NBO. Molecules initially underwent a full optimization followed by freezing the desired torsion at defined degrees (from -180° to 180° with 15° increment) and reoptimizing at MP2/6-311+G** level using software GAMESS-US[59,60]. Basis sets not available in the GAMESS-US package were downloaded from basis set exchange (https://bse.pnl.gov/bse/portal) based on previous studies. Reoptimization provided the optimal starting geometry and orbital alignment with respect to the defined torsion. NBO calculation was then performed with NBO 6.0 using these conformations with the second order perturbation analysis applied to obtain the hyperconjugation energy at the same level of theory and with the same basis set. Donations from the bonding orbital as well as the available lone pairs of the terminal atoms, for example three lone pairs of fluorine in F-C→C-F* were combined together for this hyperconjugation. The energy gaps between the donors and acceptors as well as the orbital overlaps were computed using the NBO method.

MM Calculations. The AMBER11 package was used to perform the calculation with GAFF while the MOE platform[61] was used to compute the energy profiles with MMFF94s and Parm@Frosst. GAFF atom types were assigned using the tLeap routine and the parmchk was applied to automatically write additional required force field parameters (frcmod file). The partial charges were assigned using the AM1-BCC method on the global minimum structures, and these same charges were applied for the other conformations of the same molecules. The GAFF-derived potential energy is computed using the Sander routine. MMFF94s and Parm@Frosst atom types and atom partial charges were assigned in MOE. Partial charges were added prior to energy calculation.

We computed the torsional energy contributions for all the torsions related to the central bond as measured by the FFs. These energies were replaced by H-TEQ 2 and all the other energy contributions were retained to evaluate the FF accuracy.

We selected a set of 100 most structurally diverse molecules from the MMFF94 validation dataset available in MOE, using the diversity subset functionality, available in the MOE platform. This dataset contains a wide variety of chemical functional groups and thus suitable for comparing the performance of small molecule FFs. Rotatable bonds were randomly selected for each of these molecules using an in-house Python-based program, with greater weight for non-terminal rotatable bond (weighted random selection). The corresponding QM torsion scans were performed for each of the molecules.

REFERENCES

(1) De Vivo, M. Bridging Quantum Mechanics and Structure-Based Drug Design. *Front. Biosci., Landmark Ed.* **2011**, *16*, 1619−1633.

(2) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC biology* **2011**, *9*, 71.

(3) 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.

(4) Kaminski, G.; Jorgensen, W. L. Performance of the Amber94, Mmff94, and Opls-Aa Force Fields for Modeling Organic Liquids. *J. Phys. Chem.* **1996**, *100*, 18010-18013.

(5) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the Opls-Aa Force Field for Proteins Via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem.* **2001**, *105*, 6474-6487.

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

(7) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force-Field for Simulations of Proteins and Nucleic-Acids. *J. Comp. Chem.* **1986**, *7*, 230-252

(8) 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.

(9) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comp. Chem.* **2004**, *25*, 1157-1174.

(10) Brooks, B. R.; Brooks III, C. L.; MacKerell Jr, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. Charmm: The Biomolecular Simulation Program. *J. Comp. Chem.* **2009**, *30*, 1545-1614.

(11) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. Charmm General Force Field: A Force Field for Drug-Like Molecules Compatible with the Charmm All-Atom Additive Biological Force Fields. *J. Comp. Chem.* **2010**, *31*, 671-690.

(12) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of Mmff94. *J. Comp. Chem.* **1996**, *17*, 490-519.

(13) Halgren, T. A. Merck Molecular Force Field. Ii. Mmff94 Van Der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comp. Chem.* **1996**, *17*, 520-552.

(14) Halgren, T. A. Merck Molecular Force Field. Iii. Molecular Geometries and Vibrational Frequencies for Mmff94. *J. Comp. Chem.* **1996**, *17*, 553-586.

(15) Halgren, T. A.; Nachbar, R. B. Merck Molecular Force Field. Iv. Conformational Energies and Geometries for Mmff94. *J. Comp. Chem.* **1996**, *17*, 587-615.

(16) Halgren, T. A. Merck Molecular Force Field. V. Extension of Mmff94 Using Experimental Data, Additional Computational Data, and Empirical Rules. *J. Comp. Chem.* **1996**, *17*, 616-641.

(17) 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.

(18) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, 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.

(19) Liu, Z.; 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.

(20) Kirkpatrick, P.; Ellis, C. Chemical Space. *Nature* **2004**, *432*, 823-823.

(21) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885-1891.

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

(23) Huang, L. L. L. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput.* **2013**, *9*, 3543-3556.

(24) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard Iii, W. A.; Skiff, W. M. Uff, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024-10035.

(25) Guvench, O.; Greenr, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell Jr, A. D. Additive Empirical Force Field for Hexopyranose Monosaccharides. *J. Comp. Chem.* **2008**, *29*, 2543-2564.

(26) Hatcher, E. R.; Guvench, O.; MacKerell Jr, A. D. Charmm Additive All-Atom Force Field for Acyclic Polyalcohols, Acyclic Carbohydrates, and Inositol. *J. Chem. Theory Comput.* **2009**, *5*, 1315-1327.

(27) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. Charmm Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate-Protein Modeling. *J. Chem. Theory Comput.* **2011**, *7*, 3162-3180.

(28) MacKerell Jr, A. D.; Raman, E. P.; Guvench, O. Charmm Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses. *J. Phys. Chem. B* **2010**, *114*, 12981-12994.

(29) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; MacKerell Jr, A. D. Impact of 2′-Hydroxyl Sampling on the Conformational Properties of Rna: Update of the Charmm All-Atom Additive Force Field for Rna. *J. Comp. Chem.* **2011**, *32*, 1929-1943.

(30) Baker, C. M.; Anisimov, V. M.; MacKerell, A. D. Development of Charmm Polarizable Force Field for Nucleic Acid Bases Based on the Classical Drude Oscillator Model. *J. Phys. Chem. B* **2011**, *115*, 580-596.

(31) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell Jr, A. D.; 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.

(32) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. Molecular Mechanical and Molecular Dynamical Simulations of Glycoproteins and Oligosaccharides. 1. Glycam_93 Parameter Development. *J. Phys. Chem.* **1995**, *99*, 3832-3846.

(33) Kirschner, K. N.; Woods, R. J. Solvent Interactions Determine Carbohydrate Conformation. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10541-10545.

(34) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. Energy Parameters in Polypeptides. Vii. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids. *J. Phys. Chem.* **1975**, *79*, 2361-2381.

(35) Arnautova, Y. A.; Jagielska, A.; Scheraga, H. A. A New Force Field (Ecepp-05) for Peptides, Proteins, and Organic Molecules. *J. Phys. Chem. B* **2006**, *110*, 5025-5044.

(36) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695-2703.

(37) Corbeil, C. R.; Thielges, S.; Schwartzentruber, J. A.; Moitessier, N. Toward a Computational Tool Predicting the Stereochemical Outcome of Asymmetric Reactions: Development and Application of a Rapid and Accurate Program Based on Organic Principles. *Angew. Chem. Int. Ed.* **2008**, *47*, 2635-2638.

(38) Weill, N.; Corbeil, C. R.; De Schutter, J. W.; Moitessier, N. Toward a Computational Tool Predicting the Stereochemical Outcome of Asymmetric Reactions: Development of the Molecular Mechanics-Based Program Ace and Application to Asymmetric Epoxidation Reactions. *J. Comp. Chem.* **2011**, *32*, 2878-2889.

(39) Pottel, J.; Therrien, E.; Gleason, J. L.; Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 6. Development and Application to the Docking of Hdacs and Other Zinc Metalloenzymes Inhibitors. *J. Chem. Inf. Model.* **2014**, *54*, 254-265.

(40) De Cesco, S.; Deslandes, S.; Therrien, E.; Levan, D.; Cueto, M.; Schmidt, R.; Cantin, L. D.; Mittermaier, A.; Juillerat-Jeanneret, L.; Moitessier, N. Virtual Screening and Computational Optimization for the Discovery of Covalent Prolyl Oligopeptidase Inhibitors with Activity in Human Cells. *J. Med. Chem.* **2012**, *55*, 6306-6315.

(41) Pophristic, V.; Goodman, L. Hyperconjugation Not Steric Repulsion Leads to the Staggered Structure of Ethane. *Nature* **2001**, *411*, 565-568.

(42) Alabugin, I. V.; Gilmore, K. M.; Peterson, P. W. Hyperconjugation. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 109-141.

(43) Weinhold, F. Rebuttal to the Bickelhaupt–Baerends Case for Steric Repulsion Causing the Staggered Conformation of Ethane. *Angew. Chem. Int. Ed.* **2003**, *115*, 4320-4326.

(44) Goodman, L.; Gu, H.; Pophristic, V. Gauche Effect in 1,2-Difluoroethane. Hyperconjugation, Bent Bonds, Steric Repulsion. *J. Phys. Chem. A* **2005**, *109*, 1223-1229.

(45) Mo, Y. Computational Evidence That Hyperconjugative Interactions Are Not Responsible for the Anomeric Effect. *Nat. Chem.* **2010**, *2*, 666-671.

(46) Jencks, W. P.; Carriuolo, J. General Base Catalysis of the Aminolysis of Phenyl Acetate. *J. Am. Chem. Soc.* **1960**, *82*, 675-681.

(47) Edwards, J. O.; Pearson, R. G. The Factors Determining Nucleophilic Reactivities. *J. Am. Chem. Soc.* **1962**, *84*, 16-24.

(48) Aubort, J. D.; Hudson, R. F. Enhanced Reactivity of Nucleophiles: Orbital Symmetry and the So-Called "[Small Alpha]-Effect". *Chem. Commun.* **1970**, *15*, 937-938.

(49) Alabugin, I. V.: *Stereoelectronic Effects: A Bridge between Structure and Reactivity*; John Wiley & Sons, 2016.

(50) Weinhold, F.; Landis, C. R.: *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*; Cambridge University Press, 2005.

(51) Cary, F.; Sundberg, R.: *Advanced Organic Chemistry Part A: Structure and Mechanisms*; Springer, New York, 2007.

(52) 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.

(53) Glendening, E.; Badenhoop, J.; Reed, A.; Carpenter, J.; Bohmann, J.; Morales, C.; Landis, C.; Weinhold, F. Nbo 6.0; University of Wisconsin: Madison, Wi. **2013**.

(54) Clauss, A. D.; Nelsen, S. F.; Ayoub, M.; Moore, J. W.; Landis, C. R.; Weinhold, F. Rabbit-Ears Hybrids, Vsepr Sterics, and Other Orbital Anachronisms. *Chem. Educ. Res. Pract.* **2014**, *15*, 417-434.

(55) Glendening, E. D.; Landis, C. R.; Weinhold, F. Natural Bond Orbital Methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 1-42.

(56) Glendening, E. D.; Landis, C. R.; Weinhold, F. Nbo 6.0: Natural Bond Orbital Analysis Program. *J. Comp. Chem.* **2013**, *34*, 1429-1437.

(57) 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.

(58) Song, L.; Liu, M.; Wu, W.; Zhang, Q.; Mo, Y. Origins of Rotational Barriers in Hydrogen Peroxide and Hydrazine. *J. Chem. Theory Comput.* **2005**, *1*, 394-402.

(59) 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.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J. Comp. Chem.* **1993**, *14*, 1347-1363.

(60) Gordon, M. S.; Schmidt, M. W.: Chapter 41 - Advances in Electronic Structure Theory: Gamess a Decade Later. In *Theory and Applications of Computational Chemistry*; Scuseria, C. E. D. F. S. K. E., Ed.; Elsevier: Amsterdam, 2005; pp 1167-1189.

(61) Molecular Operating Environment (MOE). Chemical Computing Group Ulc, 1010 Sherbooke St. West, Suite #910, Montreal, Qc, Canada, H3A 2R7. **2017**.

Table of Contents Graphic