Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters

Alexandre V. Brethomé,[†] Stephen P. Fletcher,*,[†] and Robert S. Paton*,[†,§]

[†]Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, United Kingdom

[§]Department of Chemistry, Colorado State University, Fort Collins, Colorado 80523, United States

S *Supporting Information*

ABSTRACT: Mathematical relationships that relate chemical structure with selectivity have provided quantitative insights underlying catalyst design and informing mechanistic studies. However, flexible compounds can adopt several distinct geometries and can be challenging to describe, using a single structure-based descriptor. How best to quantify the structural characteristics of an ensemble of structure poses both practical and technical difficulties. In this work, we introduce an automated computational workflow that can be used to obtain multidimensional Sterimol parameters for a conformational ensemble of a given substituent from a single command. The Boltzmann-weighted Sterimol parameters obtained from this approach are shown to be useful in multivariate models of enantioselectivity, while the range of values from conformers within 3 kcal/mol of the most stable structure provides a visual way to capture a possible source of uncertainty arising in the resulting models for enantioselectivity. Our approach improves the model performance in cases where particularly flexible substituents have been studied. In all cases, this approach enables the impact of conformational effects on model performance to be quickly diagnosed: in particular, these effects may be more significant than statistical model error such that selectivity prediction should be performed more cautiously. Implementing our approach requires no programming expertise and can be executed from within a graphical user interface using open-source programs.

KEYWORDS: *Sterimol, descriptors, conformation, multivariate modeling, enantioselectivity*
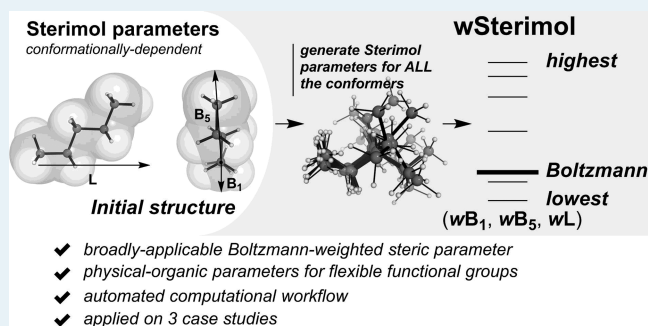
1. INTRODUCTION

Steric effects are key nonbonding interactions that influence molecular conformation and reactivity. These interactions can be harnessed in asymmetric catalysis, where differential steric effects between competing diastereomeric transition structures lead to an energetic difference, resulting in enantioselectivity.[1] Steric effects arise from the fact that each atom in a molecule occupies a certain amount of space. Repulsive interatomic interactions result from the overlap of neighboring electron clouds, leading to an increase in the quantum mechanical exchange energy (Pauli repulsion).[2] Efforts to quantify the relative magnitude of steric effects across different functional groups have traditionally been described using empirical relationships, leading to the establishment of physical-organic parameters such as interference values,[3] A-values,[4] Charton,[5] and Taft parameters.[6] More recent work has also been described.[7,8] However, the steric demands of an anisotropic functional group are not fully captured by a single parameter. Sterimol parameters were developed to be multidimensional, that is, they quantify steric demands along different principal axes, making them better-suited to describe the effects of unsymmetrical substituents.[9] Although they were developed in the 1970s, Sterimol parameters were largely unknown to the organic chemistry community until more recently, and, as a

result, have been underused. A significant breakthrough was made by the Sigman laboratory in 2012, who first showed that Sterimol parameters could be used in the construction of quantitative relationships between structure and enantioselectivity in cases where unidimensional steric parameters failed.[10] This discovery was of importance in igniting a great deal of interest in the use of multivariate relationships in catalysis,[11] and it has also led to the adoption of Sterimol parameters much more widely in describing and predicting functional group effects upon stereoselectivity.

Multidimensional Sterimol parameters were developed by Verloop, and the original Sterimol program by Verloop and Hoogenstraaten, working at Duphar B.V. (The Netherlands)[9] to capture the dimensions (and, hence, steric interactions) of a substituent along different directions. Using Corey−Pauling−Koltun (CPK) molecular models,[12] the principal axes B_1, B_5, and L can be defined (Figure 1) about the point of attachment of a given substituent. In brief, B_1 represents the shortest distance perpendicular from the primary axis of attachment; while B_5 represents the longest distance. These two subparameters can be
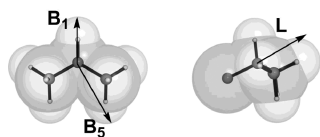
Figure 1. Representation of Sterimol parameters B_1, B_5, and L with front and lateral profiles of an *i*-propyl group using the CPK atomic model.

viewed as the minimum and maximum widths of a substituent and are influenced by the extent of branching. The final subparameter L is the total distance following the primary axis of attachment, which therefore relates the length. Original tabulated values of these parameters were generated more than 40 years ago from three-dimensional (3-D) structures constructed according to CPK definitions of bond lengths and atomic radii, along with idealized (dihedral) angles. Since then, quantum mechanical optimization of molecular structures has become a routine means to obtain 3-D coordinates, from which Sterimol parameters can be computed. However, this still requires an arbitrary description of van der Waals radii to be chosen. CPK values remain faithful to original Sterimol implementation and take factors such as hybridization into account. On the other hand, they are undefined for most elements in the periodic table, which limits their application to organic fragments.

Sterimol parameters were first used in quantitative structure−activity relationship (QSAR) studies applied to medicinal chemistry,[10] in which substituents were typically fairly rigid, and later in asymmetric catalysis to describe diverse substrates and ligands.[13,14] However, the inclusion of conformationally flexible groups can be observed in more recent studies in asymmetric catalysis.[10,15,16]

The construction of a quantitative structure selectivity relationship (QSSR) has developed into a predictive strategy to accelerate ligand design.[17] This approach assumes that structurally (and/or electronically) similar compounds will behave similarly and that modification of substituents will produce predictable changes in selectivity. Thus, if substituent modification does not modify the mechanistic pathway, selectivities between competing pathways ($\Delta\Delta G^{\ddagger}$) will be comparable, and the differences can be rationalized in terms of structural changes, described by steric and electronic parameters. However, an important question arises regarding which set(s) of 3-D coordinates should be used to define these parameters: is it more appropriate to use the most stable conformation, a catalytically relevant conformation, or an ensemble (Boltzmann) average? For instance, *n*-pentyl groups have been considered.[16] With many accessible conformations, the use of just a single conformer may give rise to potentially misleading steric parameters. The all-anti conformer is characterized by large L and small B_S values, but the situation changes dramatically (L decreases, B_S increases) when just one dihedral adopts a *gauche* conformation. One may choose to take the most favorable conformer and assume it describes the main steric interactions, or create different sets (maximum and minimum values for instance) in order to explore a range of Sterimol values.[18] Acknowledging that, outliers in QSAR or QSSR models using Sterimol parameters may emerge more often for flexible groups, even as simple as *n*-butyl. This observation can, of course, be extrapolated to any parameter that is structure-dependent, such as the Tolman's cone angle.[19] The conformation(s) used in tabulated steric parameters found in
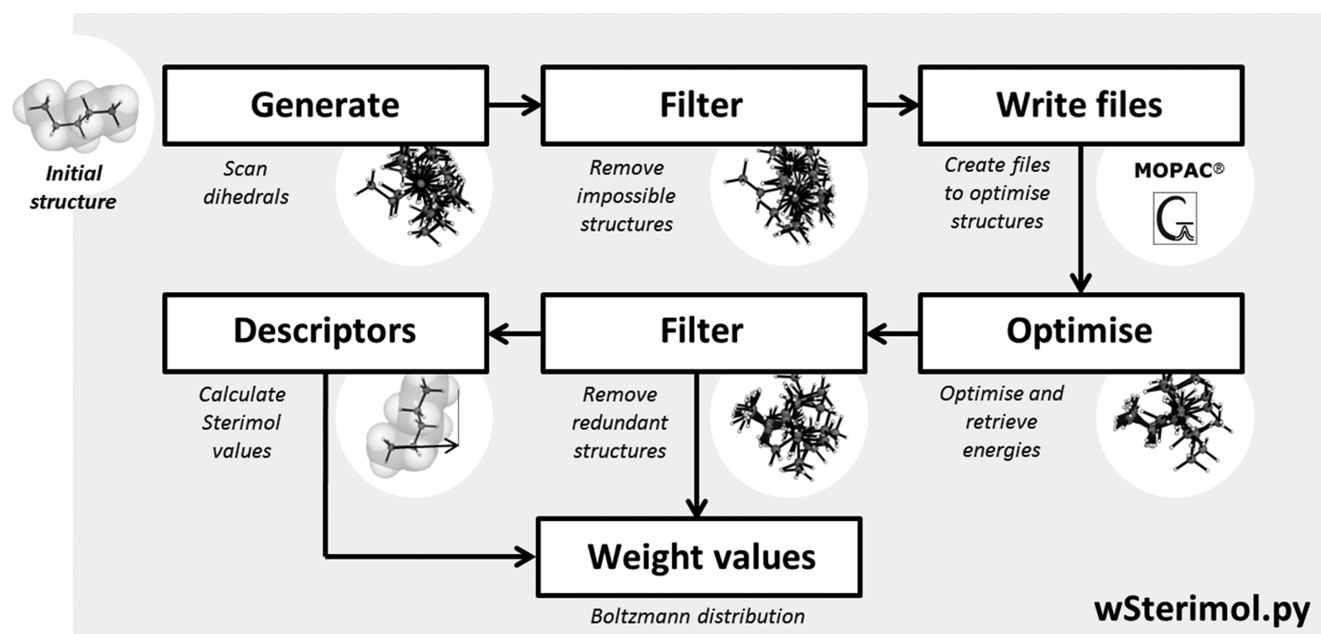
the older literature may not be readily discernible, particularly since high-accuracy quantum chemical methods have only become more available recently, such that those parameters recorded for flexible functional groups should be treated cautiously.

As described in the pioneering work by Sigman,[10] the Charton parameter is derived from a macroscopic experimental average, and assuming that rotation about the primary bond to the substituent is fast, any spatial anisotropy becomes an irrelevance. Sterimol values outperform the Charton parameter where multifaceted steric interactions must be taken into account, potentially leading to multivariate relationships between selectivity and several substituent parameters.[20] In contrast to Charton parameters, Sterimol parameters are not based on an experimental observation (which inherently describes an ensemble average). Instead, they depend on a particular specification of molecular conformation. Many mechanisms in asymmetric catalysis are not well-understood and the stereo-determining transition states are strongly influenced by conformational effects. As enantioselectivity is a macroscopic observable, one should try to pay attention to conformers when generating descriptor values to keep them as meaningful as possible. The approach taken toward conformational sampling (or even the absence of such considerations) may introduce variations in steric parameters greater than the inherent noise of a statistical model. Given the relatively small error bar that can be tolerated to generate useful predictions of stereoselectivity, due consideration should be given to conformational sampling in the computation of physical-organic parameters for statistical models.

Guidance for the development of statistically significant models has been thoroughly explored, particularly in the QSAR literature.[21−23] Following from this, Sigman has shown how the development of new steric and stereoelectronic parameters aid in mechanistic interrogation so that if a correlation is found, one can formulate mechanistic hypotheses from numerical models.[24] In this work, we turn our attention to the role of parameter uncertainty, brought about through conformational effects. Our workflow provides a visualization of how variation in conformational Sterimol values may add uncertainty to the predictions that result from a multivariate model. Intuitively, one may feel that rigid substituents are more straightforward to model and that more flexible groups are inherently less useful data points which degrade model performance (see the Supporting Information),[15] however, in this work, we aimed to validate quantitatively this hypothesis using an automated approach.

Advances in multivariate modeling may be aided by the discovery of improved statistical algorithms, but undoubtedly also with the development of descriptors that can be chemically interpreted. Automation of descriptor generation will also certainly play an important role in order to simplify routine computational design.[25] Also, we envisage that an automated, easy-to-use and transparent approach to monitor and Boltzmann-weight (if appropriate) conformational effects will be of great use and may help to bring these approaches to a broader section of the organic community.

This paper aims to address the aspect of conformational sampling and the role that this has on the use of steric parameters in physical organic chemistry. We have developed a broadly applicable and robust approach to generate steric parameters for flexible groups, termed *weighted Sterimol* (wSterimol for short). The code is built to be used via a graphical interface on several platforms, and open-source programs can be used. We will

Figure 2. Schematic representation of the weighted Sterimol algorithm triggered by very simple keywords in the *Pymol* console.

explain, in broad terms, the algorithm to tackle this problem as easily as possible for the end-user and the benchmarking realized to get consistent wSterimol values will also be discussed. Finally, this approach has the potential to broaden the community use of steric parameters to flexible substituents. This tool is also hoped to reinforce the awareness of the influence of conformers on model correlation errors as shown in three case studies found in the literature.

2. METHODOLOGY

While it is important to distinguish the different types of steric interactions represented by B_1, B_5, and L, which could have a different role in reactivity or selectivity, the contribution from all conformers can also be described in terms of an ensemble value. Thus, the different conformers can be weighted according to a Boltzmann distribution, using populations derived from computations. This weighting is indeed preferable for linear substituents which may have a flat potential energy surface, for which the different conformers lie very close in energy but may have completely different 3-D shape and, hence, Sterimol values.

Analogously to the state-of-the-art in computational NMR prediction,[26] wSterimol automates conformer generation, geometry optimization, filtering, and the final calculation of Sterimol values (Figure 2). First, all the potential conformers are generated with a conformational sampling method,[27] such as Monte Carlo,[28] molecular dynamics,[29] or SUMM (systemic unbounded multiple minimum).[30] Since systematic torsional exploration of substituents with a small number of rotatable bonds is exhaustive, we chose to implement a SUMM-like approach. The structures are then saved, stored and can be visualized at each step. When exploring the different dihedrals, structures generated with substantial interatomic clashes were

filtered out. Geometry optimization is then performed, using either semiempirical calculations (with Mopac[31]) or density functional theory (DFT, with Gaussian[32]). Following optimization, duplicate conformations are filtered out, based on an RMSD threshold. In each group, only the lowest in energy conformation is kept to be the representative structure of its group, which one would call a conformer. Sterimol parameters are finally computed for each conformer and the Boltzmann distribution outputs three final values: wB_1, wB_5, and wL. For further technical details, see the Supporting Information (SI).

In order to be as easy to handle as possible for the end-user, wSterimol was coded in Python and has to be used directly from within the *Pymol* graphical interface. This software has the advantage of having an open-source version, and a python interface, which does not require the use of a separate terminal window. One can visualize the different conformers at the different stages of the algorithm with the same atomic model used to calculate Sterimol parameters. This visualization can help in understanding the mechanistic explanation behind a potential correlation in multivariate modeling. The underlying computation of Sterimol parameters for a given conformation can be performed from the command line if desired, and this code is freely available.[33] The wSterimol package itself is released under an open-source MIT license and is also freely available.[34]

We implemented an automated, modular workflow that can be executed using a single command to obtain wSterimol values. The first argument is a list of dihedrals to be explored (Figure 2). Those can be easily chosen by looking at them directly in the graphical interface. The last arguments are the atoms to define the primary axis in Sterimol calculation.
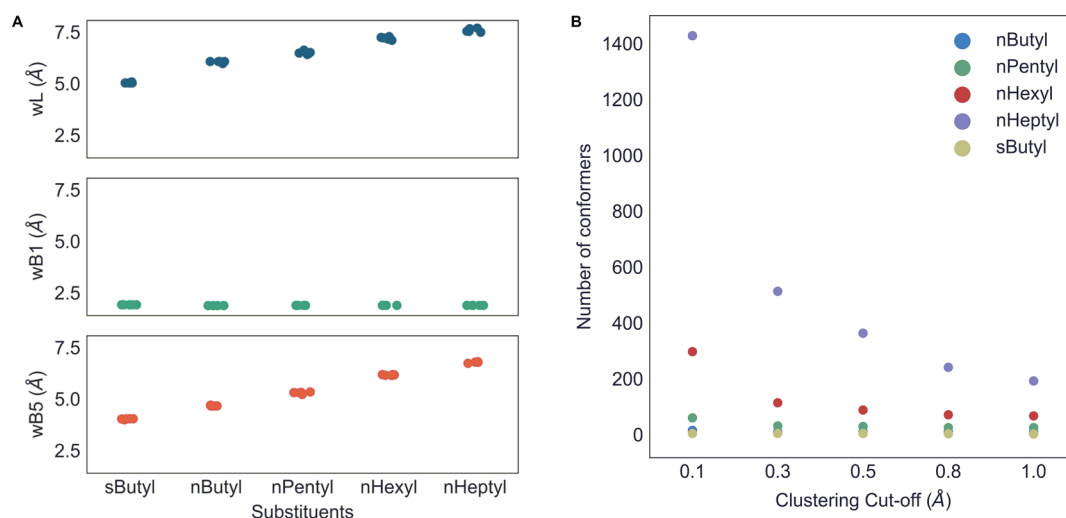
Figure 3. (A) Graphic charts representing the effect of the clustering cutoff on the Sterimol values for different substituents using PM6-DH2 semiempirical level of theory. The cut-offs are indistinguishable. (B) Graphic representing the number of considered "conformers" versus the clustering cutoff used for different substituents.

3. BENCHMARKING

Although only one simple command (executed in Pymol) is needed to generate the weighted Sterimol values, many default parameters are deployed during the workflow (see the SI for a full list). Some of them influence the output values and so benchmarking studies were therefore performed to identify the limits of wSterimol. Five key parameters are explored in this section.

3.1. Comparison to Sterimol Seminal Code. The weighted Sterimol script uses our Python implementation of Verloop's Sterimol program.[35,36] The main advantages of the updated version are the ability to process Cartesian coordinates; and the inclusion of Bondi atomic van der Waals radii (in addition to CPK), which have wider coverage of the periodic table.[37,38] This work reuses it to compute the different Sterimol parameters for each generated conformers. We compared the parameters obtained from original Fortran and updated Python programs when fed identical structures of 22 common substituents (*n*-alkyl, branched-alkyl, aromatic, benzylic). Most values agree to 2DP, with the largest difference of just 0.01 Å (see the SI).

3.2. Clustering Cutoff. Once several different input structures have been optimized, one must characterize the unique stable conformers. A stable conformer is characterized by a local energy minimum, although a flat potential energy surface (PES) can lead to distinct structures being incorrectly identified as the same conformer. Thus, a cutoff must be chosen to decide if two structures are duplicates and members of the same conformational cluster. The structures are aligned to each other, and the RMSD is retrieved for each pair alignment. If the RMSD is below the decided cutoff, two structures are tagged as duplicates and placed in the same conformational cluster. In each cluster, the structure possessing the lowest energy is defined as being the representative structure of its cluster (in other words, a conformer). Therefore, this is a critical variable that will control the number of conformers considered by the wSterimol parameters. Five different substituents were chosen to benchmark this variable (namely, *n*-butyl, *n*-pentyl, *n*-hexyl, *n*-heptyl, and *s*-butyl) due to their potentially flat PES which would complicate the optimization and make the cutoff very sensitive

(Figure 3A). The PM6-DH2 level of theory was used, because of its rapidity and its good description of aliphatic chains.

To our surprise, however, wSterimol values were not impacted by a clustering cutoff spanning a wide region from 0.1 to 1.0 Å, although the number of conformers involved in the Boltzmann distribution was drastically different (Figure 3B). The cutoff does not impact the weighted values but does modify the calculation time. More conformers imply more Sterimol calculations and the computational cost increases accordingly. A higher cutoff is advised, in terms of cost, but should be limited to avoid merging theoretically different conformers. A very large clustering cutoff would remove any dynamism from the system, leading ultimately to the "classic" Sterimol values. In practice looking at the different structures, 1.0 Å clustering cutoff already started to merge structures which should be classified in different clusters. We advise using a cutoff between 0.3 and 0.5 Å, default in the script being 0.3 Å (see the SI to change it manually).

3.3. Effect of the Atomic Model. Verloop initially used a CPK atomic model, for which connectivity information is used to categorize different "types" of the same atom. For example, the atomic radius of carbon ranges from 1.5 Å to 1.7 Å, depending on formal hybridization, aromaticity, and ring size. Either the atom types can be manually typed in the submitted files, or they can be calculated as shown in our script. However, these CPK values were limited to the elements H, C, N, O, F, P, S, Cl, Br, and I. We considered that Bondi radii could provide an alternative definition of van der Waals radii, which are both simpler (a single value for each element) and with wider coverage of the periodic table, based on crystallographic data (Figure 4A). The effect on the computed weighted Sterimol was explored on five linear substituents and four alkenes to have different hybridization patterns (Figure 4B).

We found a systematic difference between the two atomic models, but the actual variation is far less significant than the effect of other variables such as the level of theory. In terms of computational cost, CPK model is slightly more expansive due to the computing time for connectivity calculation which is not needed for Bondi. As Bondi has the advantage of being defined across the periodic table and provide almost identical results, we preconize its use over the CPK model (see the SI to change it manually).
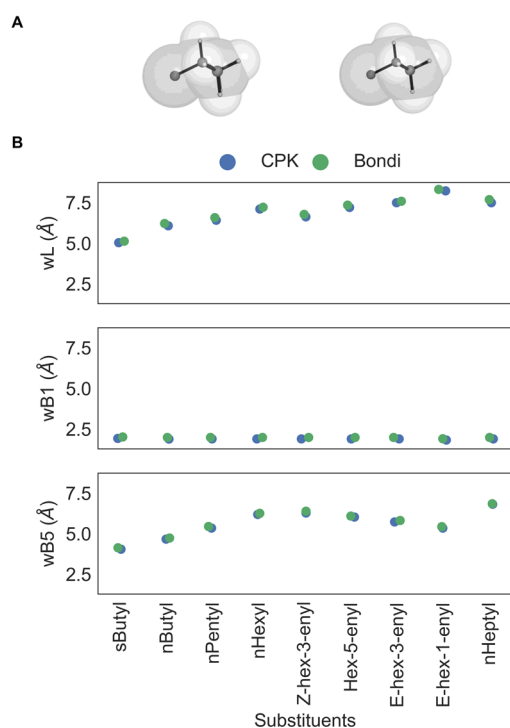
Figure 4. (A) Graphical representation of *i*-Pr group using CPK (left) and Bondi (right) atomic model. (B) Charts representing the effect of CPK and Bondi atomic models on wSterimol values for different substituents at the PM6-DH2 level.

3.4. Convergence Criteria: The Dihedral Division Number.

When generating conformers, each dihedral is explored by systematically rotating through a number of increments. The dihedral division number describes how finely the torsion is searched (Figure 5A). A greater number of divisions leads to greater coverage and a higher chance of locating all conformations on the potential energy surface (PES). For n tending to ∞, a dihedral has been explored in a continuum, maximizing the likelihood of finding all conformers. The computational cost for k dihedral angles, however, also increases with the dihedral division number as n^k. In practice, fewer than ten calculations should be enough to reach the curvature of a local minimum that the optimization will follow. It implies an optimum value is reachable between cost and accuracy. The graphs below present the impact of the dihedral division number on weighted Sterimol values for three linear substituents. The generated structures by dihedral rotation are highly dependent on the initial structure. Thus, nine different initial conformations were taken to get a statistically significant representation.

We demonstrate that the angle division number gives consistent wSterimol parameters for values of five and above (Figure 5B). Below this threshold, the weighted Sterimol values tend to oscillate presumably due to the lack of complete coverage of the PES. Above 5, wSterimol values converge nicely but unfortunately with a higher computational cost. Therefore, we advise the use of a dihedral division number of 5 for a balance between cost and consistency, which is the default value (see the SI to modify it manually).

3.5. Effect of the Level of Theory.

Optimization criteria and computed energies are highly dependent on the level of theory used. As the Boltzmann-weights in the conformational ensemble uses these energies, and clustering uses the coordinates of the optimized structure to define conformers, the level of theory has the most critical influence on the final wSterimol values.

Four different compounds were taken in order to estimate these differences: (1) a functional group with possible hydrogen
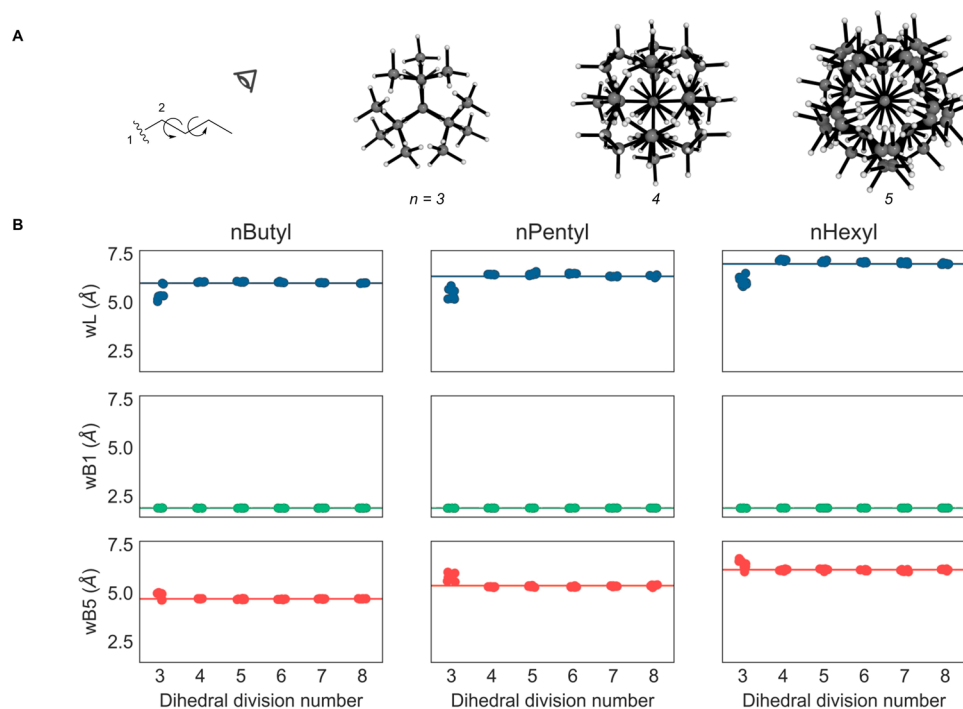

Figure 5. Graphic representing (A) the generated conformers of *n*-butyl, using different dihedral division number from 3 to 5 and (B) the wSterimol values for *n*-butyl, *n*-pentyl, and *n*-hexyl at different dihedral division number starting from 9 different initial structures for statistical significance. The asymptotes are represented by a horizontal line in each graph.
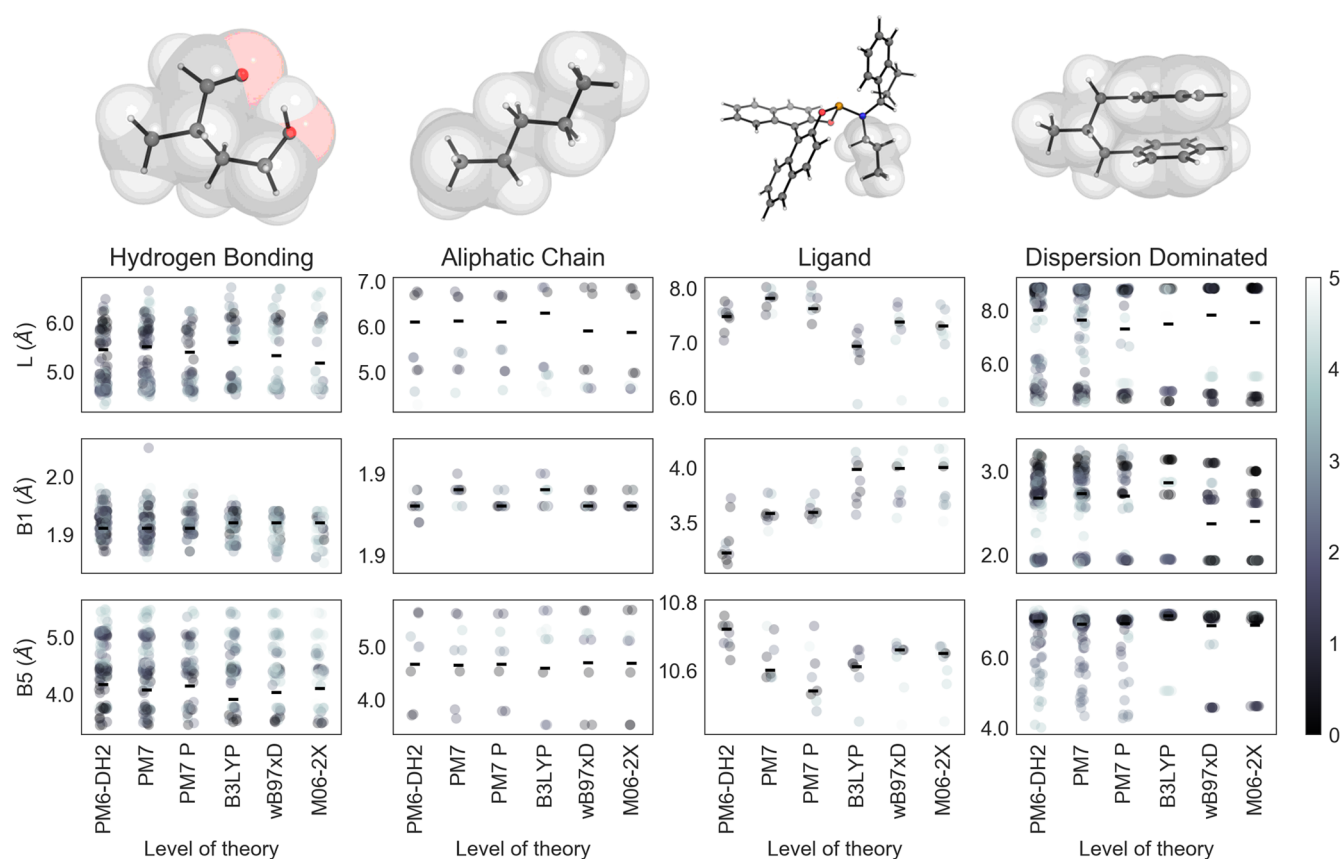
Figure 6. Charts representing the effect of the level of theory on Sterimol values (L, B1, B5) of all the conformers (colored dots) used to calculate wSterimol values (black tick marks) for a moiety that contains hydrogen bonding, for an aliphatic chain, for a ligand containing a complex functional group, and finally for a dispersion-dominated moiety. Sterimol values of all the conformers are colored according to their relative energies, going from 0.0 to 5.0 kcal/mol. A cutoff of 5.0 kcal/mol was used to remove energetically disfavored conformers.

Table 1. Selected Common Functional Groups at the M06-2X/6-311+G(d,p)//M06-2X/6-31G(d) Level of Theory[a]

				L		B1		B5	
group	wL	wB1	wB5	min	max	min	max	min	max
1-Nap-CH$_2$−	5.93	1.89	6.97	5.22	8.91	1.88	1.89	5.88	7.23
4-MePhCH$_2$−	5.96	1.90	7.13	5.62	6.71	1.88	1.90	7.04	7.19
1-Ad−	6.61	3.15	3.49	6.61	6.61	3.15	3.15	3.49	3.49
c-hexyl−	6.65	1.92	3.48	6.65	6.65	1.92	1.92	3.48	3.48
Et−	4.59	1.88	3.16	4.59	4.59	1.88	1.88	3.16	3.16
Et$_2$CH−	5.18	1.94	4.44	4.58	5.55	1.92	2.06	3.79	4.52
i-Pr−	4.56	1.91	3.17	4.56	4.56	1.91	1.91	3.17	3.17
i-Pr$_2$CH−	5.45	2.17	4.46	4.62	5.63	1.97	2.22	4.45	4.52
i-PrCH$_2$−	5.44	1.87	4.42	4.67	5.58	1.87	1.90	4.39	4.42
Me−	3.43	1.88	2.02	3.43	3.43	1.88	1.88	2.02	2.02
n-Bu−	5.80	1.88	4.71	4.63	6.84	1.88	1.89	3.52	5.69
n-pent−	6.14	1.88	5.36	4.60	7.83	1.87	1.91	3.53	6.93
Ph−	6.80	1.70	3.15	6.80	6.80	1.70	1.70	3.15	3.15
Ph$_2$CH−	6.64	1.90	6.15	6.64	6.64	1.90	1.91	6.15	6.15
PhCH$_2$−	5.32	1.90	6.04	5.12	6.71	1.88	1.90	5.82	6.07
s-Bu-	4.97	1.91	4.08	4.60	5.48	1.90	1.92	3.59	4.46
t-Bu−	4.54	2.76	3.17	4.54	4.54	2.76	2.76	3.17	3.17
t-BuCH$_2$−	5.65	1.90	4.39	5.65	5.65	1.90	1.90	4.39	4.39

[a]The weighted Sterimol parameters are listed alongside the minimum and maximum values in a 3.0 kcal/mol energy window. Further examples are tabulated in the SI.

bonding between an alcohol and a carbonyl, (2) a linear organic molecule with van der Waals interactions only, (3) a nontrivial functional group contained in a ligand, and (4) a dispersion-dominated interaction between two adjacent aromatic rings

were chosen. Three commonly used DFT (B3LYP, M06-2X, and ωB97xD) and two semiempirical functionals (PM6-DH2 and PM7) were chosen due to their spread use in the literature.[39]

The M06-2X/6-31G(d) level of theory is used as a reference in this benchmark.

As expected, the influence of the level of theory is highly dependent on the functional group (Figure 6). Thus, aliphatic chains are barely affected and show consistent values. This starts to change in the presence of hydrogen bonding or dispersion interactions. Semiempirical methods led to a different distribution of the conformers, compared to DFT functionals, even if the apparent wSterimol values are sometimes almost identical. On average, twice as many conformers are generated with PM6-DH2 and PM7 (semiempirical methods), compared to the DFT functionals, and they are continuously scattered along the *y*-axis. This is attributable to differences in the underlying shape of the PES and the optimization convergence criteria between different packages. Therefore, the conformers were optimized with PM7 at 100 times more precise convergence criteria (called PM7 P in Figure 6; see the SI for details). PM7 indeed shows results that are more similar to B3LYP for the hydrogen-bonding moiety. Semiempirical results are expected to be superior to small-molecule force field methods.[40] However, we still found cases for more-complex functional groups, where they give qualitatively different predicted values than DFT. One can also notice that ωB97xD and M06-2X are very similar, in terms of their conformer distribution, and represent what is probably the best in terms of cost and accuracy. In other words, we advise the use of the most accurate level of theory as it is reasonable in order to generate descriptors. As shown by Grimme and Schreiner, the methodological state-of-the-art, along with computational power, is continually improving.[41] Therefore, descriptors should remain easily computable but as accurate as possible.[20] Using a single machine (e.g., a laptop), wSterimol values are best generated with PM7 (precise optimization). With access to greater computational resources, we recommend using DFT calculations, such as M06-2X 6-31g(d) (Gaussian is currently supported by the code), including single-point energy corrections if possible, e.g., with M06-2X 6-311+G(d,p). Such calculations are still feasible on a single CPU with a few hours calculation time for flexible systems containing 5−6 heavy atoms.

4. TABULATION OF COMMON FUNCTIONAL GROUPS

Herein, we show the tabulated values of common functional groups already found in the literature (Table 1). A longer list can be found in the SI. This table aims to serve as a standard source of information for flexible common substituents, while rigid moieties are also listed for the sake of consistency in modeling. A 3.0 kcal/mol energy window is used because it captures 99% of the population at 298 K in all the substituents shown below at the considered level of theory. Using a substituent as flexible as *n*-heptyl, the captured population remains above 99% at 298 K. Our benchmarking studies illustrate that DFT calculations should be used if practically possible. The wSterimol values in this table are Boltzmann-weighted at the M06-2X/6-311+G-(d,p)//M06-2X/6-31G(d) level of theory.

5. APPLICATIONS

In this section, we demonstrate conformational effects in the application of multivariate models of enantioselective catalysis. We chose case studies for which it has previously been shown that Sterimol parameters *alone* provide a useful structure−selectivity relationship. Models for which steric interactions do

not provide the main contribution were not considered. Our emphasis is not on the predictive and interpretative qualities of these models since these aspects have already been explained, but on the additional information that comes from the application of wSterimol parameters. Three case studies are shown. In the first case, this approach gives equivalent performance to conventional Sterimol parameters and confirms that substituent flexibility has only a minor impact on model performance and interpretation. The second example provides improved performance over static parameters, while also showing that the conformational dependence of parameter values at the higher end of the predicted values can be far greater than the statistical model error. In the final case study, wSterimol helps to identify regions of the model where conformational sampling could have an impact on statistical performance and predictivity.

Miller and Sigman discovered that substrate steric parameters play an essential role in the organocatalysed desymmetrization of bisphenols (Figure 7).[10] A strong relationship exists between the enantioselectivity and the two parameters B_1 and L, such that the optimal substrate for this reaction has a large minimum
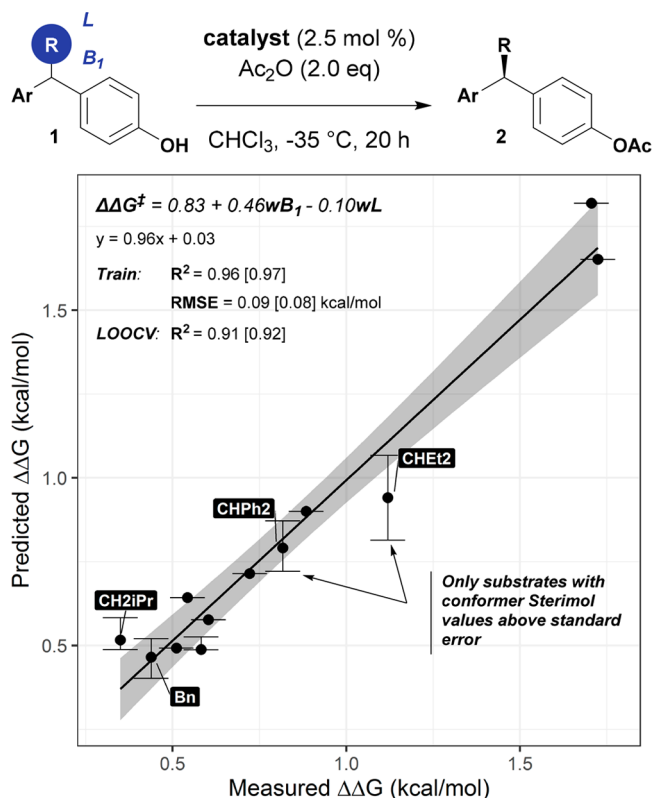

Figure 7. QSSR analysis of the desymmetrization of bisphenol using wSterimol values. Error bars represent the range of Sterimol values of the conformers within 3.0 kcal/mol window. Gray area represents the standard error of the model at the 95% confidence interval. Only the substituents with a range of Sterimol values leading to a range greater than 0.05 kcal/mol were labeled. Original model values have been placed in brackets for comparison. wSterimol model (this work): $\Delta\Delta G^{\ddagger} = 0.83 + 0.46wB_1 - 0.10wL$, Training [$R^2 = 0.96$, RMSE = 0.09 kcal/mol], LOOCV [$R^2 = 0.91$, RMSE = 0.14 kcal/mol], ANOVA [$wB_1 = 2.1 \times 10^{-7}$, $wL = 1.4 \times 10^{-2}$]. Sterimol model (original work): $\Delta\Delta G^{\ddagger} = 0.83 + 0.47B_1 - 0.10L$, Training [$R^2 = 0.97$, RMSE = 0.08 kcal/mol], LOOCV [$R^2 = 0.92$, RMSE = 0.13 kcal/mol], ANOVA [$B_1 = 6.9 \times 10^{-8}$, $L = 7.3 \times 10^{-3}$].

width (B_1) and a short length (L). This correlation implies a propeller-like twist of the aryl rings in the reaction mechanism. As was noted, the use of Sterimol was consistent with the substituent exhibiting a strong conformational preference and that the substituents must be treated as static and not dynamic. Therefore, we decided to test wSterimol on a system that is static and where Sterimol parameters perform very well, to ensure that this did not lead to any degradation of the linear model. Reassuringly, the use of wSterimol showed negligible differences in terms of correlation, error, and significance of the model. We illustrate this by depicting both the standard error of the model (gray bands) and the range of wSterimol values within 3 kcal/mol from the most stable conformer (error bars) in Figure 7. The gray bands around the regression line show the range in which the true regression line lies with 95% confidence. The bands visualize all intervals for every possible x-values and are tightest where the data are grouped more densely. For a given prediction, the experimental value will lie within the upper and lower bounds 95% of the time. One can see that the model standard error is more significant than possible noise attributable to different conformations: this is particularly true for isotropic functional groups (Me, *t*Bu, Ad). One can conclude that conformational effects do not play a significant role in model construction and performance and that the model obtained with wSterimol is consistent with the hypothesized mechanistic explanation observed previously.

We suggest that the wSterimol approach of including weighted parameter and taking into account possible ranges of values, can be used in the same domain of applicability as conventional Sterimol parameters. If a system is static, wSterimol will take it into account and behave like a conventional Sterimol parameter. However, it is advantageous for describing a dynamic system. The same conclusion can be drawn from the Nozaki–Hiyama–Kishi (NHK) propargylation of acetophenone (see the SI).

Our second case-study considered the asymmetric addition of diethylzinc on benzaldehyde using chiral phosphoramide ligands (Figure 8).[15] A correlation between steric effects and observed enantioselectivity can be obtained using ligand substituent Sterimol parameters (as opposed to the entire ligand). Sterimol descriptions of the entire amine group of the N-substituted ligands originally led to a poor correlation. The involvement of minimal width $(B_1^{R_1})$ of R$_1$ and R$_2$ $(B_1^{R_2})$ parameters led to the mechanistic hypothesis that the carbonyl group is pushed away from the Zn center and that a larger proximal steric bulk will increase the enantioselectivity. Noting that steric descriptions of the more flexible, full system perform worse than those of the more rigid subsystems in multivariate linear regression, we reasoned that a potential source of this difference could be the much greater conformational degrees of freedom accessible to the larger system. By using wSterimol, two outliers were identified using the boxplot visualization technique (see the SI) and the correlation in Figure 8A was obtained. Compared to the previous model, performing the regression using weighted Sterimol values, rather than singular values, results in a better fit and emphasizes the importance of conformer sampling. As shown below, the amplitude in Sterimol values at the extremities of the model is far greater than the statistical model error, which further reinforces that the predictive power of the model in these regions may be weaker. We also performed the regression using the entire amino group. The use of weighted Sterimol parameters on the full amine gives a weaker correlation than the subsystems as previously shown,
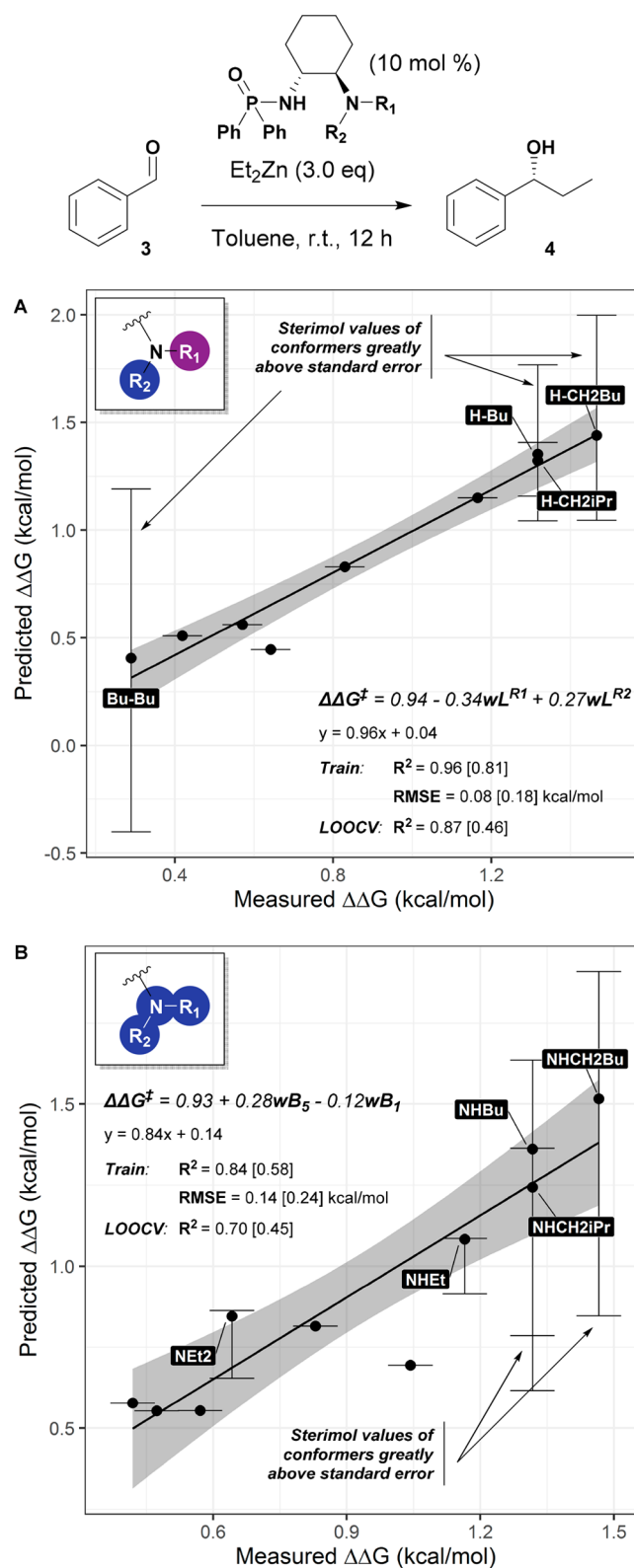

Figure 8. QSSR analysis of the asymmetric addition of diethylzinc on benzaldehyde using chiral phosphoramide ligands. Gray area represents standard error of the model at the 95% confidence interval. Error bars represent the range of Sterimol values of the conformers within a 3.0 kcal/mol window. Only the substituents with a range of Sterimol values leading to a range greater than 0.05 kcal/mol were labeled. Original model values were put in bracket for comparison. (A) Model using wSterimol on R$_1$ and R$_2$ of the ligand, with two outliers removed. wSterimol model (this work): $\Delta\Delta G^\ddagger = 0.94 - 0.34 wL^{R1} + 0.27 wL^{R2}$,

Figure 8. continued

Training [$R^2 = 0.96$, RMSE = 0.08 kcal/mol], LOOCV [$R^2 = 0.87$, RMSE = 0.15 kcal/mol], ANOVA [$wL^{R_1} = 1.4 \times 10^{-4}$, $wL^{R_2} = 1.7 \times 10^{-4}$]. Sterimol model (original work): $\Delta\Delta G^{\ddagger} = 1.04 - 0.36B_1^{R_1} + 0.29B_1^{R_2}$, Training [$R^2 = 0.80$, RMSE = 0.18 kcal/mol], LOOCV [$R^2 = 0.46$, RMSE = 0.32 kcal/mol], ANOVA [$B_1^{R_1} = 5.7 \times 10^{-3}$, $B_1^{R_2} = 3.8 \times 10^{-2}$]. (B) Model using wSterimol on the full amine of the ligand, with two outliers removed. wSterimol model (this work): $\Delta\Delta G^{\ddagger} = 0.93 + 0.28wB_5 - 0.12wB_1$, Training [$R^2 = 0.84$, RMSE = 0.14 kcal/mol], LOOCV [$R^2 = 0.70$, RMSE = 0.20 kcal/mol], ANOVA [$wB_5 = 7.6 \times 10^{-4}$, $wB_1 = 5.1 \times 10^{-2}$]. Sterimol model (original work): $\Delta\Delta G^{\ddagger} = 0.95 + 0.27B_5$, Training [$R^2 = 0.58$, RMSE = 0.24 kcal/mol], LOOCV [$R^2 = 0.45$, RMSE = 0.28 kcal/mol], ANOVA [$B_5 = 1.0 \times 10^{-2}$].

with the same outliers (Figure 8B). The extreme difficulty of regression is emphasized visually in this case by the magnitude of the range of conformer values at the higher end of the predicted values: this would be particularly important to realize if the model is being used to optimize (i.e., enhance) enantioselectivity, since the likely predictive power is degraded in this region in particular. Such an observation may direct additional experimental effort in obtaining data for a more-rigid group in this region of parameter space or urge more caution in generating the descriptor values and selecting the level of theory. The use of weighted Sterimol shows an increase in the performance and the consistency of the models.

Finally, we also studied an asymmetric Henry reaction using chiral 1,2-amino phosphinamide ligands (Figure 9).[16] After screening several ligands, a quantitative correlation between the observed enantioselection was obtained using steric parameters to describe each ligand's N-substituents. This shows a correlation using Sterimol values, B_1 of R_1 and B_5 of R_2, underlying the importance of the width of these two subgroups. By using wSterimol values, a slightly smaller correlation was obtained than that observed previously, where singular Sterimol values were used (Figure 9). The amplitude of possible Sterimol values for low-energy conformers within 3.0 kcal/mol noticeably exceeds the model error for two ligands, showcasing the importance of descriptor generation for confident prediction. Again, the wSterimol values (i.e., the Boltzmann-weighted central points) give at least as good correlations as when using a singular value, and the additional information regarding the variability of these values provides context in interpreting the statistical performance and aids to identify regions of the model that may be weaker.

6. LIMITATIONS AND FUTURE DEVELOPMENTS

The current implementation of wSterimol automates the generation, optimization, clustering, and analysis of an acyclic conformational ensemble, generating individual and weighted Sterimol values, which are then easily applied to models such as those above. Ring conformers are currently not sampled; however, this can be implemented using, e.g., distance-geometry methods, molecular dynamics. However, one can generate structures with a third-party program then continue with the different modules in wSterimol. Although rotatable dihedrals are not detected automatically, this can be advantageous, since it allows selected torsions to be excluded from sampling if desired. Given the modular nature of the plugins we have written, expansion to incorporate additional steric parameters for automated conformational weighting is also straightforward.
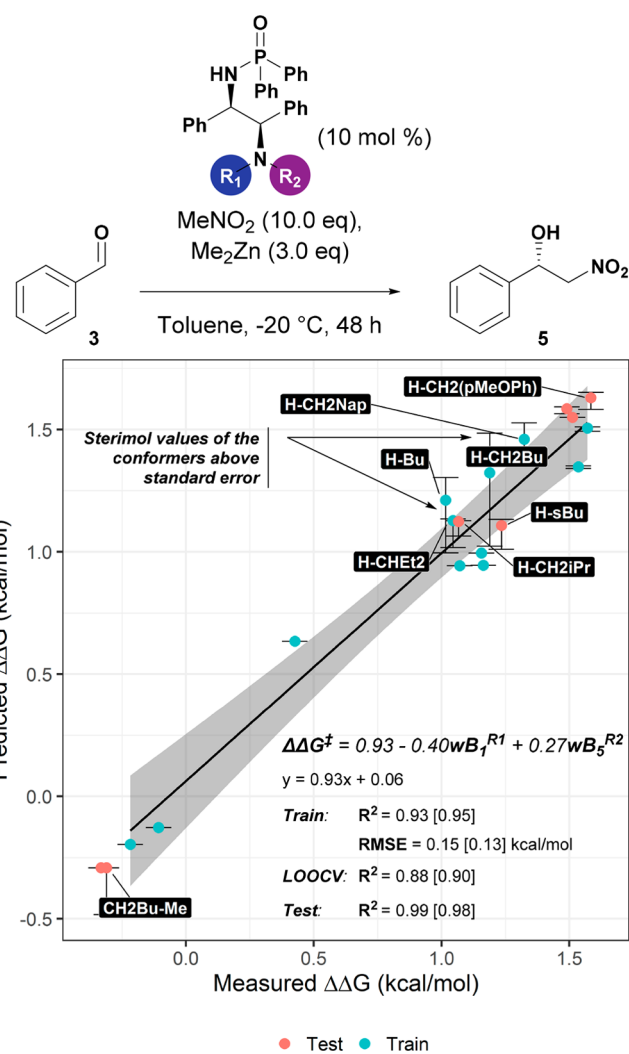

Figure 9. QSSR analysis of asymmetric Henry reaction using wSterimol values from chiral 1,2-amino phosphonamide ligands. Gray area represents the standard error of the model at the 95% confidence interval. Error bars represent the range of Sterimol values of the conformers within 3.0 kcal/mol window. Only the substituents with a range of Sterimol values leading to a range greater than 0.05 kcal/mol were labeled. Original model values are shown in brackets for comparison. wSterimol model (this work): $\Delta\Delta G^{\ddagger} = 0.93 - 0.40wB_1^{R_1} + 0.27wB_5^{R_2}$, Training [$R^2 = 0.93$, RMSE = 0.15 kcal/mol], LOOCV [$R^2 = 0.88$, RMSE = 0.20 kcal/mol], Test [$R^2 = 0.99$, RMSE = 0.07 kcal/mol], ANOVA [$wB_1^{R_1} = 3.5 \times 10^{-6}$, $wB_5^{R_2} = 8.2 \times 10^{-4}$]. Sterimol model (original work): $\Delta\Delta G^{\ddagger} = 0.93 - 0.40B_1^{R_1} + 0.28B_5^{R_2}$, Training [$R^2 = 0.95$, RMSE = 0.13 kcal/mol], LOOCV [$R^2 = 0.90$, RMSE = 0.18 kcal/mol], Test [$R^2 = 0.98$, RMSE = 0.11 kcal/mol], ANOVA [$B_1^{R_1} = 1.2 \times 10^{-6}$, $B_5^{R_2} = 2.9 \times 10^{-4}$].

7. PERSPECTIVES

Although conformational analysis is a familiar topic to organic chemists, coupling practical implementations of this with electronic structure calculations, and using the results to derive ensemble-average properties can be a daunting challenge, and we hope to have alleviated much, if not most, of the pain involved. We have created a computational workflow, wSterimol, to enable the computation of physical organic parameters for a conformational ensemble. The design philosophy has been to build a tool that can be used by nonexperts, using open-source software, with a single command accessed via a graphical interface or from the command line.

This approach provides a capability for automated conformational sampling and avoids complex recipes involving multiple programs, spreadsheets, or scripting. The availability of this tool allows powerful multivariate models and Sterimol parameters to be applied to synthetically relevant functional groups, where tabulated values may either be absent from the literature, or worse, lack conformational context. This package can be used alongside freely available programs, maximizing the reproducible nature of the computed parameters. We have demonstrated the importance of such an approach with three case studies: where flexible substituents were present, this resulted in an improvement in the final model, while for rigid substituents, the model performs equally well. Conformer effects in models found in the literature are sometimes more significant than the confidence interval of the model, meaning conformational sampling is a greater factor of error in the quantitative prediction of selectivity than the inherent model noise, sometimes with 1 kcal/mol of difference for flexible linear substituents. Therefore, we hope that this tool broadens the scope of substituents that can be reliably used in multivariate models of enantioselectivity and a means to quantify the effect of conformational flexibility on statistical models using 3-D physical-organic descriptors such as Sterimol parameters.

ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscatal.8b04043.

> All code has been made openly available for download and reuse at https://github.com/bobbypaton/wSterimol (DOI: 10.5281/zenodo.1434439) and is also provided here (ZIP)

> Full details of wSterimol parameters generation (download, install, and tutorials), tabulation of wSterimol values of common functional groups, details of model construction of the three case studies presented in this paper, two other case studies not discussed and Cartesian coordinates of the tabulated functional groups with their associated wSterimol outputs are disclosed (PDF)

> Full details of the data generated for the benchmark and used to create the corresponding figures are provided (XLSX)

AUTHOR INFORMATION

Corresponding Authors
*E-mail: stephen.fletcher@chem.ox.ac.uk (S.P.F.).
*E-mail: robert.paton@colostate.edu (R.S.P.).

ORCID ⓘ
Stephen P. Fletcher: 0000-0001-7629-0997
Robert S. Paton: 0000-0002-0104-4166

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Knowles, R. R.; Jacobsen, E. N. Attractive Noncovalent Interactions in Asymmetric Catalysis: Links between Enzymes and Small Molecule Catalysts. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20678−20685.

(2) Alabugin, I. V. *Stereoelectronic Effects: A Bridge Between Structure and Reactivity*; Wiley: Chichester, U.K., 2016.

(3) Bott, G.; Field, L. D.; Sternhell, S. Steric Effects. A Study of a Rationally Designed System. *J. Am. Chem. Soc.* **1980**, *102*, 5618−5626.

(4) Winstein, S.; Holness, N. J. Neighboring Carbon and Hydrogen. XIX. t-Butylcyclohexyl Derivatives. Quantitative Conformational Analysis. *J. Am. Chem. Soc.* **1955**, *77*, 5562−5578.

(5) Charton, M. Steric Effects. I. Esterification and Acid-Catalyzed Hydrolysis of Esters. *J. Am. Chem. Soc.* **1975**, *97*, 1552−1556.

(6) Taft, R. W. Linear Free Energy Relationships from Rates of Esterification and Hydrolysis of Aliphatic and Ortho-Substituted Benzoate Esters. *J. Am. Chem. Soc.* **1952**, *74*, 2729−2732.

(7) Reid, J. P.; Goodman, J. M. Goldilocks Catalysts: Computational Insights into the Role of the 3,3′ Substituents on the Selectivity of BINOL-Derived Phosphoric Acid Catalysts. *J. Am. Chem. Soc.* **2016**, *138*, 7910−7917.

(8) Falivene, L.; Credendino, R.; Poater, A.; Petta, A.; Serra, L.; Oliva, R.; Scarano, V.; Cavallo, L. SambVca 2. A Web Tool for Analyzing Catalytic Pockets with Topographic Steric Maps. *Organometallics* **2016**, *35*, 2286−2293.

(9) Verloop, A. *Drug Design*, Vol. *III*; Ariens, E. J., Ed.; Academic Press: New York, 1976.

(10) Harper, K. C.; Bess, E. N.; Sigman, M. S. Multidimensional Steric Parameters in the Analysis of Asymmetric Catalytic Reactions. *Nat. Chem.* **2012**, *4*, 366−374.

(11) Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A. The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016**, *49*, 1292−1301.

(12) Pauling, L.; Corey, R. Atomic Coordinates and Structure Factors for Two Helical Configurations of Polypeptide Chains. *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 235−240.

(13) Ardkhean, R.; Roth, P. M. C.; Maksymowicz, R. M.; Curran, A.; Peng, Q.; Paton, R. S.; Fletcher, S. P. Enantioselective Conjugate Addition Catalyzed by a Copper- Phosphoramidite Complex : Computational and Experimental Exploration of Asymmetric Induction. *ACS Catal.* **2017**, *7*, 6729−6737.

(14) Ardkhean, R.; Mortimore, M.; Paton, R. S.; Fletcher, S. P. Formation of Quaternary Centres by Copper Catalysed Asymmetric Conjugate Addition to β-Substituted Cyclopentenones with the Aid of a Quantitative Structure−selectivity Relationship. *Chem. Sci.* **2018**, *9*, 2628−2632.

(15) Huang, H.; Zong, H.; Bian, G.; Song, L. Constructing a Quantitative Correlation between N-Substituent Sizes of Chiral Ligands and Enantioselectivities in Asymmetric Addition Reactions of Diethylzinc with Benzaldehyde. *J. Org. Chem.* **2012**, *77*, 10427−10434.

(16) Huang, H.; Zong, H.; Bian, G.; Yue, H.; Song, L. Correlating the Effects of the N-Substituent Sizes of Chiral 1,2-Amino Phosphinamide Ligands on Enantioselectivities in Catalytic Asymmetric Henry Reaction Using Physical Steric Parameters. *J. Org. Chem.* **2014**, *79*, 9455−9464.

(17) Wiest, O.; Helquist, P. Striking a Balance to Control Stereochemistry. *Science* **2011**, *333*, 1831−1832.

(18) Niemeyer, Z. L.; Milo, A.; Hickey, D. P.; Sigman, M. S. Parameterization of Phosphine Ligands Reveals Mechanistic Pathways and Predicts Reaction Outcomes. *Nat. Chem.* **2016**, *8*, 610−617.

(19) Tolman, C. A. Steric Effects of Phosphorus Ligands in Organometallic Chemistry and Homogeneous Catalysis. *Chem. Rev.* **1977**, *77*, 313−348.

(20) Santiago, C. B.; Guo, J. Y.; Sigman, M. S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**, *9*, 2398−2412.

(21) Tropsha, A.; Gramatica, P.; Gombar, V. K. The Importance of Being Earnest: Validation Is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003**, *22*, 69−77.

(22) Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inf.* **2010**, *29*, 476−488.

(23) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57*, 4977−5010.

(24) Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parametrization of Non-Covalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, *139*, 6803−6806.

(25) Guan, Y.; Wheeler, S. E. Automated Quantum Mechanical Predictions of Enantioselectivity in a Rhodium-Catalyzed Asymmetric Hydrogenation. *Angew. Chem., Int. Ed.* **2017**, *56*, 9101−9105.

(26) Nguyen, Q. N. N.; Tantillo, D. J. Using Quantum Chemical Computations of NMR Chemical Shifts to Assign Relative Configurations of Terpenes from an Engineered Streptomyces Host. *J. Antibiot.* **2016**, *69*, 534−540.

(27) Brueckner, A. C.; Ogba, O. M.; Snyder, K. M.; Richardson, H. C.; Cheong, P. H.-Y. Conformational Searching for Complex, Flexible Molecules. In *Applied Theoretical Organic Chemistry*; World Scientific: Singapore, 2018; pp 147−164.

(28) Chang, G.; Guida, W. C.; Still, W. C. An Internal Coordinate Monte Carlo Method for Searching Conformational Space. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.

(29) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919−11929.

(30) Goodman, J. M.; Still, W. C. An Unbounded Systematic Search of Conformational Space. *J. Comput. Chem.* **1991**, *12*, 1110−1117.

(31) Stewart, J. *MOPAC2016*; Stewart Computational Chemistry; Colorado Springs, CO, 2016.

(32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A. J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision D.01*; Gaussian Inc.: Wallingford, CT, 2009.

(33) Installation on OSX/Linux Using: 'Pip Install Sterimol'. https://Github.Com/Bobbypaton/Sterimol, DOI: 10.5281/Zenodo.1320767.

(34) https://Github.Com/Bobbypaton/WSterimol, DOI: 10.5281/Zenodo.1434439.

(35) Piou, T.; Romanov-Michailidis, F.; Romanova-Michaelides, M.; Jackson, K. E.; Semakul, N.; Taggart, T. D.; Newell, B. S.; Rithner, C. D.; Paton, R. S.; Rovis, T. Correlating Reactivity and Selectivity to Cyclopentadienyl Ligand Properties in Rh(III)-Catalyzed C−H Activation Reactions: An Experimental and Computational Study. *J. Am. Chem. Soc.* **2017**, *139*, 1296−1310.

(36) Piou, T.; Romanov-Michailidis, F.; Ashley, M. A.; Romanova-Michaelides, M.; Rovis, T. Stereodivergent Rhodium(III)-Catalyzed Cis-Cyclopropanation Enabled by Multivariate Optimization. *J. Am. Chem. Soc.* **2018**, *140*, 9587−9593.

(37) Rowland, R. S.; Taylor, R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van Der Waals Radii. *J. Phys. Chem.* **1996**, *100*, 7384−7391.

(38) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441−451.

(39) Goerigk, L.; Grimme, S. A Thorough Benchmark of Density Functional Methods for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670−6688.

(40) Kanal, I. Y.; Keith, J. A.; Hutchison, G. R. A Sobering Assessment of Small-Molecule Force Field Methods for Low Energy Conformer Predictions. *Int. J. Quantum Chem.* **2018**, *118*, e25512.

(41) Grimme, S.; Schreiner, P. R. Computational Chemistry: The Fate of Current Methods and Future Challenges. *Angew. Chem., Int. Ed.* **2018**, *57*, 4170−4176.