ACS PHYSICAL CHEMISTRY Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

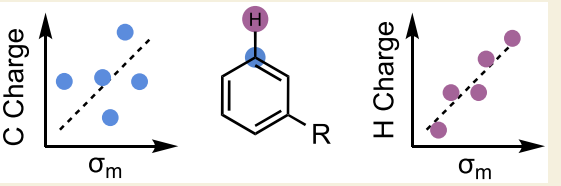
pubs.acs.org/physchemau
Article

Bottom-Up Atomistic Descriptions of Top-Down Macroscopic Measurements: Computational Benchmarks for Hammett Electronic Parameters

Guilian Luchini and Robert S. Paton*

Cite This: ACS Phys. Chem Au 2024, 4, 259–267
Read Online

ABSTRACT: The ability to relate substituent electronic effects to chemical reactivity is a cornerstone of physical organic chemistry and Linear Free Energy Relationships. The computation of electronic parameters is increasingly attractive since they can be obtained rapidly for structures and substituents without available experimental data and can be applied beyond aromatic substituents, for example, in studies of transition metal complexes and aliphatic and radical systems. Nevertheless, the description of "top-down" macroscopic observables, such as Hammett parameters using a "bottom-up" computational approach, poses several challenges for the practitioner. We have examined and benchmarked the performance of various computational charge schemes encompassing quantum mechanical methods that partition charge density, methods that fit charge to physical observables, and methods enhanced by semiempirical adjustments alongside NMR values. We study the locations of the atoms used to obtain these descriptors and their correlation with empirical Hammett parameters and rate differences resulting from electronic effects. These seemingly small choices have a much more significant impact than previously imagined, which outweighs the level of theory or basis set used. We observe a wide range of performance across the different computational protocols and observe stark and surprising differences in the ability of computational parameters to capture para- vs meta-electronic effects. In general, σ_m predictions fare much worse than σ_p. As a result, the choice of where to compute these descriptors—for the ring carbons or the attached H or other substituent atoms—affects their ability to capture experimental electronic differences. Density-based schemes, such as Hirshfeld charges, are more stable toward unphysical charge perturbations that result from nearby functional groups and outperform all other computational descriptors, including several commonly used basis set based schemes such as Natural Population Analysis. Using attached atoms also improves the statistical correlations. We obtained general linear relationships for the global prediction of experimental Hammett parameters from computed descriptors for use in statistical modeling studies.



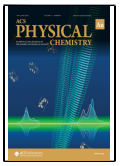
KEYWORDS: *Hammett parameters, molecular descriptors, charge models, population analysis, electronic effects*

INTRODUCTION

Linear Free Energy Relationships (LFERs) quantitatively relate changes in molecular structure to changes in chemical reactivity and represent a well-established and robust approach to deriving mechanistic insight from multiple observations.[1] Historically, LFERs have been carried out using parameters or descriptors obtained from experimental observations for a set of reference processes, such as acid ionization constants (thermodynamic measurements) or relative rate constants (kinetic measurements).[2] These quantitative parameters can, in turn, be related to the outcomes of entirely different reactions. The Hammett relationship is a paradigmatic example in which benzoic acid acidity (an experimentally determined thermodynamic parameter) correlates with the rate for various reactions involving substrates, reagents, or catalysts bearing substituted aromatic rings.[3] A linear correlation between reactivity and the inductive and resonance contributions from meta- or para-substituents on aromatic rings (σ_m and σ_p Hammett parameters, respectively) reveals the sensitivity of a reaction to these electronic influences. Deviations from linearity may be related to changes in the rate-determining step or in the mechanism or transition state of the reaction.[4]

The ability to relate substituent electronic effects to chemical reactivity is a cornerstone of physical organic chemistry.[5] Several variants of the Hammett parameter, σ, have been derived for experimental reference processes, including Brown's σ_p^+ parameters, based on the solvolysis of t-cumyl chlorides,[6] Arnold's σ_α parameters, based on spin delocalization of substituted benzyl radicals,[7] Creary's σ_C parameters, based on methylenecyclopropane rearrangements,[8] and Jiang and Ji's σ_{jj} parameters, based on trifluorostyrene cyclodimerizations.[9] Alternative experimental measures, for exam-

Received: August 19, 2023
Revised: January 14, 2024
Accepted: January 16, 2024
Published: February 6, 2024



ple, electrode polarization effects, have also been related to Hammett parameters.[10] However, as quantum chemical methods have improved, computationally derived parameters, for example, obtained from density functional theory (DFT) calculations, have emerged as an alternative for the quantitative description of molecules and their substituents.[11] Given the broad utility of Hammett relationships in correlating reactivity and selectivity with substituent electronic effects, there have been numerous efforts to obtain parameters/descriptors computationally that avoid the need for measurements of equilibrium constants or which can be employed beyond aromatic substituents.[12] Additional Hammett-like equations have also been derived to capture electronic substituent effects using computational concepts like molecular softness and Fukui functions.[13] Computationally derived electronic parameters are increasingly attractive since they can be obtained for structures and substituents without available experimental data and, due to recent advances in computational workflow automation, are readily accessible for large numbers of functional groups.[14] Another advantage of using computed electronic descriptors vs tabulated parameters is that additivity does not need to be assumed—multiple substituted aromatics can instead be computed directly. Combined steric and electronic effects have been insightful in describing Hammett relationships;[15] however, many studies rely solely on electronic descriptors. Nevertheless, the departure from experimental descriptor acquisition, obtained from macroscopic "top-down" experimental observations to using computational "bottom-up" descriptors presents several challenges for practitioners. For example, the choice of computational protocol(s) and treatment of the conformational ensemble are critical considerations,[16] particularly where descriptor values are obtained as an ensemble average. More fundamentally, since the quantum mechanical wave function/density is inherently delocalized across an entire structure and atomic charges are not themselves experimentally observable, myriad theoretical approaches exist to partition the electronic density, which does not by definition belong to any atom, over the different atoms.[17] Our work is focused on establishing the most appropriate computational protocol for obtaining a correct qualitative and quantitative description of substituent electronic effects at the macroscopic level. To the best of our knowledge, no benchmark comparisons of these protocols against a broad data set of experimental electronic parameters has been previously reported.

Atomic charges play a fundamental role in qualitative pictures of reactivity and selectivity and are integral ingredients used to construct quantitative relationships between chemical structure and reaction outcomes. The theoretical foundations of different charge models have been described extensively,[18] and significant effort has gone into comparing the various methods for computing atomic charge.[19] Cramer and Truhlar[20] and Martin[18b] have categorized the different available charge models into four separate classes: (I) nonquantum mechanical approaches based on experimentally measured properties, such as deformation densities or dipole moments; (II) those computed by partitioning of the quantum mechanical charge density, with examples including Mulliken,[21] Natural Population Analysis (NPA or Weinhold charges),[22] or Hirshfeld[23] charges; (III) those obtained by fitting to reproduce a physical observable (such as dipole moment or a computed electrostatic potential) seen in the CHarges from ELectrostatic Potentials using a Grid-based method (CHELPG),[24] Merz–Kollman charges using Universal force field radii (MKUFF),[25] Hu, Lu, Yang charges with standard atomic densities (HLYGat),[26] and Atomic Polar Tensor (APT)[27] charges; or (IV) those based on semi-empirical adjustments to Class II or III methods, such as in Charge Model 5 (CM5).[28] With the above categories, another important distinction arises between methods that project the molecular wave function onto atom-centered basis functions, like NPA, and those that are based directly on the electron density as a function of space, such as the Hirshfeld scheme. In this work, we observe a distinct performance difference between these two groupings.

Since partial atomic charge is not an experimental observable, there is no universal objective ranking of these approaches. Each charge method operates on its own scale (although atomic contributions do sum to total molecular charge), and all types have been used for various applications.[29] In this work, we explore the ability of various charge models from Classes II–IV (what we term a bottom-up description) to describe empirically determined (i.e., top-down) Hammett parameters.[30] Benchmark studies have been published comparing smaller subsets of charge values in their predictive ability against Hammett constants, with findings usually naming Hirshfeld charges as the best predictor.[31] Hirshfeld charges have also been shown to be good predictors of electrophilic and nucleophilic trends in contexts outside Hammett constants.[32] The method that Hirshfeld charges are constructed by partitioning electron density into individual atomic contributions to form the molecule as a whole (as in a promolecule)[33] has been shown to be effective at capturing electronic reactivity.[34] Additionally, important general considerations—for example, regarding the more stable performance of Hirshfeld or NPA charges vs Mulliken for different basis set sizes—have been reported previously, and less specific information is available to guide the choice of computational protocol in describing macroscopic electronic parameters. Our quantitative comparisons against Hammett parameters addresses this need.

In addition to atomic charges, the relationship of computed atomic properties such as NMR chemical shifts to Hammett parameter is appealing due to their ability to capture local electronic effects and having a corresponding experimental observable.[35] Perhaps due to more complex nuclear coupling relationships captured in this experimental measurement, correlations to partial atomic charge values may not be as high.[36] Still, successful studies have utilized NMR shifts in their relationship to Hammett parameters and partial atomic charges.[37] A variety of methods also exist to compute NMR shifts, including Gauge-Independent Atomic Orbital (GIAO),[38] Individual Gauges for Atoms in Molecules (IGAIM),[39] and the Continuous Set of Gauge Transformations (CSGT)[40] methods.

Machine learning (ML) models relying on large data sets (over 100,000 data points) to train neural networks have been used to compute partial atomic charges and NMR shifts. When utilized within their domain of applicability, these models give a highly accurate surrogate prediction of QM or experimental properties much faster than traditional QM calculations. These models typically use simple molecular representations, such as an SMILES string, as inputs. The *qmdesc* Python library utilizes a model trained to predict QM Hirshfeld charges (ML-QMDESC Hirshfeld, in this work) and GIAO NMR shifts (ML-QMDESC NMR), among other QM properties.[41] The

https://doi.org/10.1021/acsphyschemau.3c00045
ACS Phys. Chem Au 2024, 4, 259–267

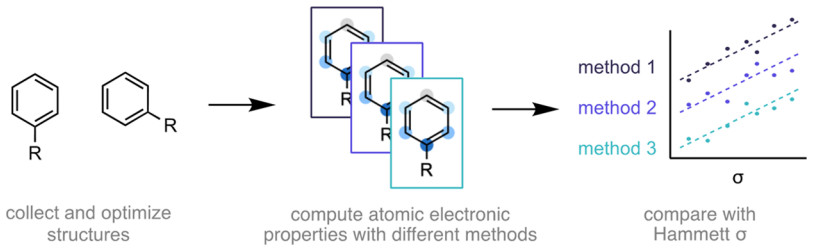


Figure 1. Atomic electronic properties (charges, chemical shifts) obtained from several computational protocols are compared against experimental Hammett (σ_m and σ_p) parameters.

σ_p = -0.83 σ_m = -0.16	σ_p = -0.72 σ_m = -0.23	σ_p = -0.70 σ_m = -0.21	σ_p = -0.66 σ_m = -0.16	σ_p = -0.61 σ_m = -0.24	σ_p = -0.56 σ_m = -0.02	σ_p = -0.45 σ_m = 0.10	σ_p = -0.37 σ_m = 0.12	σ_p = -0.34 σ_m = -0.04	σ_p = -0.27 σ_m = 0.12
σ_p = -0.25 σ_m = 0.10	σ_p = -0.24 σ_m = 0.10	σ_p = -0.21 σ_m = -0.07	σ_p = -0.20 σ_m = -0.10	σ_p = -0.19 σ_m = 0.02	σ_p = -0.17 σ_m = -0.07	σ_p = -0.16 σ_m = -0.08	σ_p = -0.15 σ_m = -0.07	σ_p = -0.15 σ_m = -0.04	σ_p = -0.15 σ_m = -0.08
σ_p = -0.15 σ_m = -0.05	σ_p = -0.14 σ_m = -0.11	σ_p = -0.14 σ_m = -0.05	σ_p = -0.14 σ_m = -0.05	σ_p = -0.13 σ_m = -0.06	σ_p = -0.12 σ_m = -0.07	σ_p = -0.12 σ_m = -0.08	σ_p = -0.12 σ_m = -0.07	σ_p = -0.11 σ_m = -0.03	σ_p = -0.09 σ_m = -0.08
σ_p = -0.08 σ_m = 0.05	σ_p = -0.07 σ_m = 0.08	σ_p = -0.07 σ_m = -0.03	σ_p = -0.05 σ_m = -0.03	σ_p = -0.04 σ_m = 0.06	σ_p = -0.03 σ_m = 0.06	σ_p = -0.03 σ_m = 0.25	σ_p = -0.02 σ_m = 0.03	σ_p = -0.02 σ_m = 0.07	σ_p = -0.01 σ_m = 0.06
σ_p = 0.00 σ_m = 0.00	σ_p = 0.00 σ_m = 0.15	σ_p = 0.00 σ_m = 0.21	σ_p = 0.01 σ_m = 0.00	σ_p = 0.01 σ_m = 0.08	σ_p = 0.01 σ_m = 0.15	σ_p = 0.01 σ_m = 0.07	σ_p = 0.02 σ_m = 0.06	σ_p = 0.03 σ_m = 0.20	σ_p = 0.05 σ_m = 0.09
σ_p = 0.06 σ_m = 0.34	σ_p = 0.06 σ_m = 0.12	σ_p = 0.07 σ_m = 0.06	σ_p = 0.08 σ_m = 0.09	σ_p = 0.09 σ_m = 0.12	σ_p = 0.10 σ_m = 0.15	σ_p = 0.12 σ_m = 0.15	σ_p = 0.12 σ_m = 0.15	σ_p = 0.17 σ_m = 0.33	σ_p = 0.18 σ_m = 0.35
σ_p = 0.18 σ_m = 0.16	σ_p = 0.18 σ_m = 0.31	σ_p = 0.20 σ_m = 0.21	σ_p = 0.23 σ_m = 0.37	σ_p = 0.23 σ_m = 0.39	σ_p = 0.25 σ_m = 0.23	σ_p = 0.26 σ_m = 0.25	σ_p = 0.31 σ_m = 0.39	σ_p = 0.32 σ_m = 0.29	σ_p = 0.35 σ_m = 0.38
σ_p = 0.36 σ_m = 0.28	σ_p = 0.36 σ_m = 0.35	σ_p = 0.41 σ_m = 0.23	σ_p = 0.42 σ_m = 0.35	σ_p = 0.42 σ_m = 0.36	σ_p = 0.43 σ_m = 0.34	σ_p = 0.44 σ_m = 0.27	σ_p = 0.45 σ_m = 0.37	σ_p = 0.45 σ_m = 0.36	σ_p = 0.45 σ_m = 0.37
σ_p = 0.48 σ_m = 0.38	σ_p = 0.50 σ_m = 0.38	σ_p = 0.53 σ_m = 0.23	σ_p = 0.54 σ_m = 0.43	σ_p = 0.60 σ_m = 0.53	σ_p = 0.66 σ_m = 0.56	σ_p = 0.68 σ_m = 0.62	σ_p = 0.72 σ_m = 0.60	σ_p = 0.78 σ_m = 0.71	

Figure 2. Data set of aryl substituents studied with experimentally determined Hammett σ_p and σ_m values.

CASCADE Python library also can be utilized to predict experimental ^{1}H and ^{13}C NMR shifts (ML-CASCADE NMR).[42]

In this work, we perform a quantitative comparison of a wide range of computed parameters obtained from different protocols against experimental Hammett parameters. We examine the performance of various computational methods, including density functional theory (DFT), semiempirical QM,[43] and ML models (Figure 1). Due to their broad generality and widespread availability of empirical values, our study focuses on Hammett's original definition of σ_m and σ_p values. We assess the suitability of different charge models and computational strategies, such as analyzing the meta/para carbon atoms, the attached meta/para hydrogen atoms, and

https://doi.org/10.1021/acsphyschemau.3c00045
ACS Phys. Chem Au 2024, 4, 259−267

additional ring substituents in deriving computed parameters that are correlated with Hammett parameters. In doing so, we demonstrate a wide range of performance across the different computational protocols and observe stark and surprising differences in the ability of computational parameters to capture *para* vs *meta*-electronic effects. Additionally, we apply these methods to parametrize multiply substituted aromatics, compare against experimental rate data, and develop suggested best practices for "bottom-up" computational models of inductive and resonance effects, as encapsulated by Hammett parameters. We propose linear regression formulas to derive Hammett parameters for substituents without experimental data and for multiply substituted aromatic rings from computed descriptors.

METHODS

Our studies use an experimental data set curated by Ertl,[12b] also utilized in a recent benchmark study by Borges,[31b] based on the 200 most common aryl substituents in the ChEMBL database,[44,31] of which 89 have experimentally determined Hammett parameters available (Figure 2). While Borges and co-workers also evaluated various predictive models that utilize multiple charge values from the aromatic ring to predict Hammett constants, we have chosen to instead investigate a wider range of charge and NMR values to compare, evaluating individual atoms in their univariate correlations with experimental values for this data set. Molecules (in which each substituent is appended to a phenyl group) were first converted from SMILES format to 3D structures with added hydrogens using RDKit.[45] Conformer ensembles for each structure were then generated with CREST,[46] from which all structures were fully optimized with DFT. We used the B3LYP functional[47] with a Becke-Johnson damped Grimme D3-dispersion correction[48] and def2-TZVP basis set[49] in chloroform, a common solvent used in organic reactions and NMR analyses, using SMD implicit solvation for geometry optimizations.[50] All structures were verified as minima by the analysis of vibrational frequencies. The DFT-computed charges and NMR shielding tensors were obtained at the same level of theory. Single-point analyses with other density functionals (M06-2X, wB97XD, mPW1PW91) and solvent models yield parameters that are highly correlated with those discussed in the manuscript; the statistical comparisons made in the main text are independent of the level of theory used (see Supporting Information, section S2 for detailed comparison).

For the DFT optimized structures, Class II, III, and IV atomic charges along with NMR isotropic shielding tensors (proportional to computed NMR chemical shifts) were obtained at the same level of theory, all of which are widely available in modern electronic structure packages. Variations of these charge models were also computed, including Minimal-Basis Mulliken charges (MBS-Mulliken),[51] iterative Hirshfeld, and iterative CM5 methods.[52] ML predicted Hirshfeld charges were calculated by the qmdesc ML package, and NMR chemical shifts were predicted using both *qmdesc* and *CASCADE* models. Additionally, semiempirical xTB Mulliken and CM5 charges were also computed. For species with multiple conformers, the atomic charges and chemical shifts were Boltzmann-averaged using the computed Gibbs free energy values. Computed charges and chemical shift values were parsed from output files (see the repository in the Supporting Information for all Python workflows) for carbon and hydrogen atoms meta and para to the substituent of interest. For unsymmetrical conformations, computed charge and chemical shift values at the two meta positions subtly differ. Since these positions are equivalent at the macroscopic level (due to rapid single bond rotation relative to the NMR time scale), computed meta values were averaged into a single predictor. The values were then compared on a univariate basis to the corresponding experimental σ_p and σ_m values from Figure 2 using Pearson's R^2.

RESULTS AND DISCUSSION

A correlation heatmap of Pearson R^2 values is shown in Figure 3 for a complete set of charge and NMR shift models against

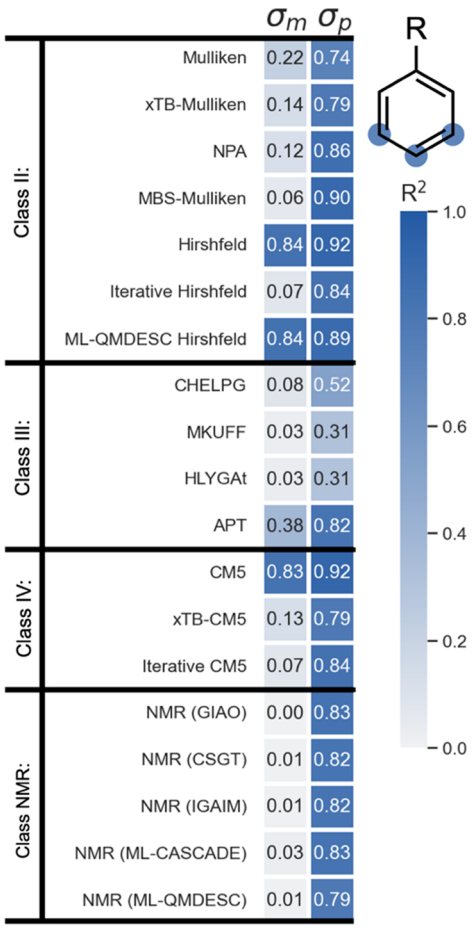


Figure 3. Correlation of computed charges and NMR shifts at carbon with σ_p and σ_m values.

experimental Hammett parameters. For simplicity in analyzing these relationships, we focused on univariate correlations between Hammett values and the computed charge or NMR values on single atoms. The computed values were evaluated at the para and meta carbon atoms of the phenyl ring, an approach commonly employed in literature studies.[12b] Most strikingly, however, we observe very poor correlations between most computed parameter sets and experimental σ_m values. Apart from Hirshfeld, ML-computed Hirshfeld and CM5 charges (which give reasonable R^2 values of 0.84, 0.84 and 0.83), all other methods yield much worse correlations ranging from 0.00 to 0.38. In contrast, the computed parametric parameters compare well, for the most part, with experimental σ_p values producing correlations of R^2 = 0.74 and above. Class III charge methods are the exception, and CHELPG, MKUFF, and HLYGAt charges are all poorly correlated with the experimental σ_p values. These three methods are all based on the molecular electrostatic potential, as opposed to the other Class III method APT based on the dipole moment. Interestingly, NPA charges, used routinely in the literature as electronic parameters, exhibit a large degradation in their correlation with para (R^2 = 0.86) vs meta (R^2 = 0.12) Hammett parameters. Indeed, NPA charges are outperformed

https://doi.org/10.1021/acsphyschemau.3c00045
ACS Phys. Chem Au 2024, 4, 259−267

by several QM approaches, such as MBS-Mulliken (R^2 = 0.90), Hirshfeld (R^2 = 0.92), and CM5 (R^2 = 0.92) charge methods, and even by ML-predicted Hirshfeld charges (R^2 = 0.89). Overall, Hirshfeld and CM5 charges at carbon show the most even performance across the meta- and para-positions in comparison to experimental σ-parameters.

Initially, we attributed the poor performance of several protocols in describing electronic effects at the meta-position to the difference in experimental scales between σ_p (−0.83 to 0.78) and σ_m (−0.24 to 0.71), making the latter intrinsically harder to predict computationally. However, upon further investigation using the computed charge and NMR shift values at the hydrogen (i.e., rather than carbon) atoms, we found this explanation to be incorrect. Indeed, for the meta-H atoms, we obtained an overall improvement in correlation for σ_m across nearly all methods tested, except for Class III charges and NMR values (Figure 4). Correlations for σ_p values remained

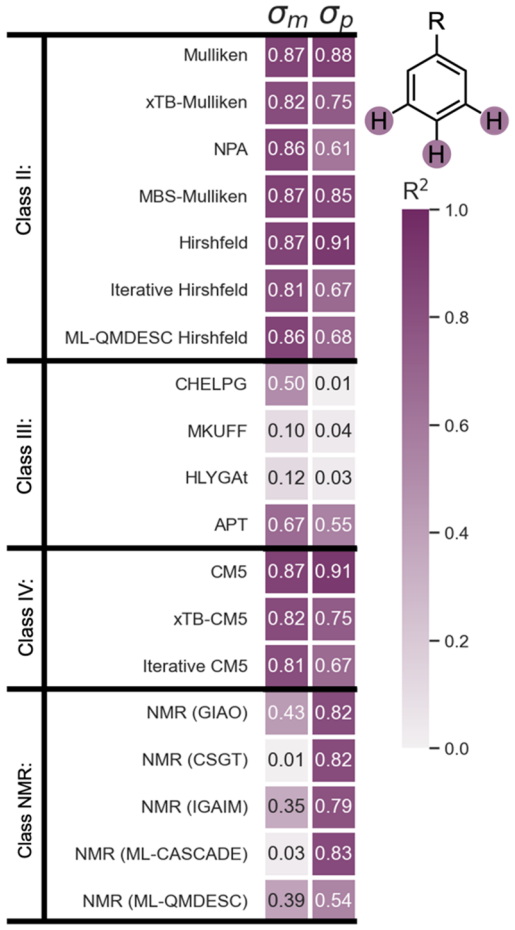


Figure 4. Correlation of computed charges and NMR shifts at hydrogen with σ_p and σ_m values.

relatively consistent relative to those obtained from the analysis of the C atoms. These results are perhaps surprising since the absolute differences between computed charges/chemical shifts are smaller for H than for C, being one bond further away from the substituent of interest, nor are these atoms directly influenced by resonance effects. We wondered if perhaps an indirect influence of the substituent on the C atom, particularly in the meta case, was being captured and in hydrogen, which has less overall electron density and could be more susceptible to changes in electronic environment. Nevertheless, computed descriptors for H atoms statistically outperform C atoms, and we propose that it is the closer proximity of the meta-C atom to the substituent of interest that results in numerical differences between the different methods.

Wave function-based projection schemes, such as Mulliken charges, are well-known to suffer basis set sensitivity due to the problem of attributing spatially diffuse basis functions to a single atom and the partitioning of large overlap populations.[53] While NPA charges remove the problem of orbital overlap through orthogonalization, nevertheless, we find these values are significantly perturbed by the spatial proximity of meta-substituents, yielding unrealistic results that correlate poorly with experimental Hammett parameters. In contrast, integration of the electron density over atomic domains, as in the Hirshfeld scheme, avoids problems inherent to basis set based approaches yielding atomic charges that perform well across para- and meta-substitution patterns and which correlate best to observable Hammett parameters. Perhaps

Hirshfeld or other density-based schemes (CM5, Voronoi) are therefore strongly recommended as electronic parameters/descriptors for future physical-organic or statistical studies over the widely used NPA charges. Additionally, where the practitioner has a choice of atom(s) for which to collect electronic parameters, we observe that locations closer to the site of change (which would be expected to show the largest variations) may, in fact, produce descriptors less correlated to macroscopic observations.

In addition to considering σ_m and σ_p parameters separately, we investigated the applicability of each computed parameter to the global prediction of Hammett parameters by concatenating meta and para data and computing a Pearson's R^2 against the corresponding Hammett value (Figure 5). As above, higher correlations are obtained using the charges or chemical shift values of H rather than C atoms, and the best correlation with experiment is obtained with Hirshfeld and CM5 charges (R^2 = 0.88), which significantly outperform NPA charges. The linear relationship between computed charge and Hammett parameters (e.g., σ = 79.64 $q(\mathrm{H})_{\mathrm{Hirshfeld}}$ −3.21) can subsequently be used to predict Hammett parameters for functional groups without experimental values: using Hirshfeld charges, we calculate mean absolute errors (MAEs) of 0.07. Inspired by Tantillo's *CHESHIRE* database relating experimental and computed chemical shifts, we have derived and tabulated linear scaling factors for the holistic prediction of Hammett σ values using each of the protocols used in this work (see Supporting Information, section S4).[54]

To further test the hypothesis that computed descriptors for aromatic substituents perform better than the ring atoms themselves in describing electronic effects, we sought a more challenging set of experimental data. We selected the oxidative cleavage of olefins with photoexcited nitroarenes developed by Leonori and co-workers. In this work, the electronic influence of nitroarene substituent(s) on the rate of cycloaddition with a model alkene (adamantylideneadamantane) was explored by an experimental Hammett study using tabulated sigma values.[55] Importantly, several of these substrates have multiple meta- and para- substituents, for which electronic effects were assumed to be additive in the original study (Figure 6). The nitroarenes are shown with the logarithm of the experimental reaction rates relative to (unsubstituted) nitrobenzene (log-(k_X/k_H)).

Another advantage of using computed electronic descriptors vs tabulated parameters is that additivity does not need to be

https://doi.org/10.1021/acsphyschemau.3c00045
ACS Phys. Chem Au 2024, 4, 259−267

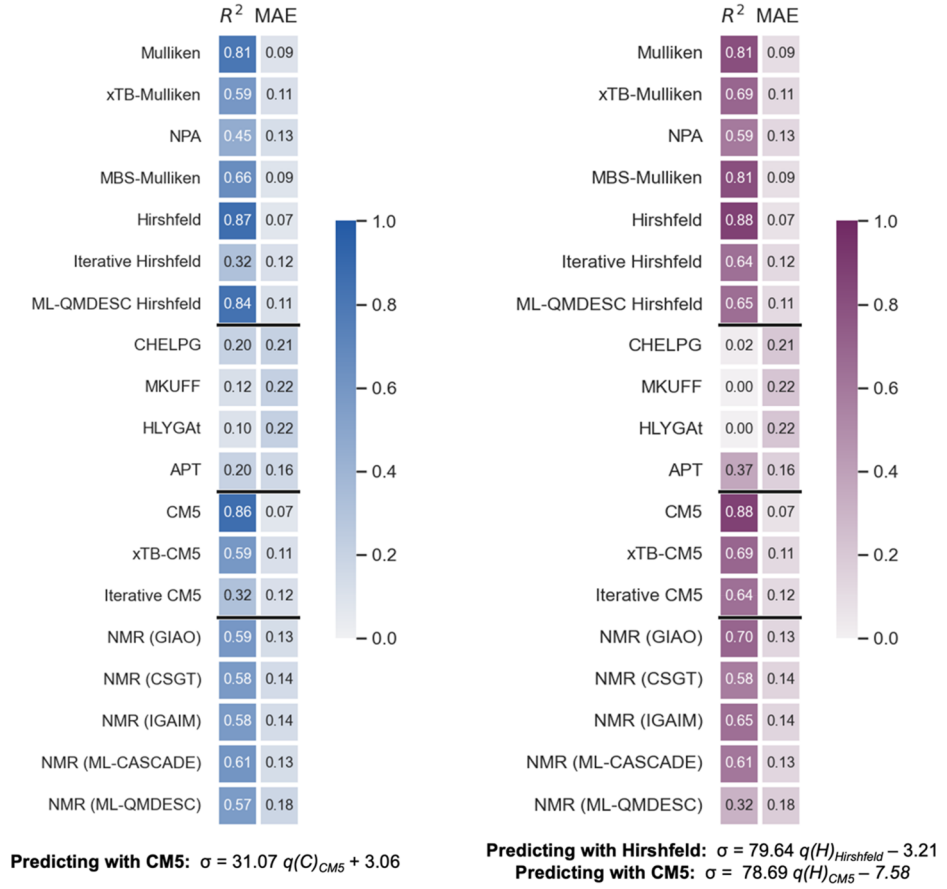
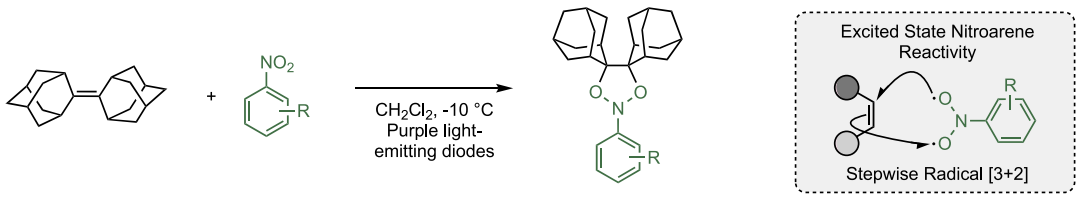


Figure 5. Correlation of descriptors taken from carbon (left, blue) and hydrogen (right, purple) against experimental Hammett parameters (combined σ_p and σ_m values). Error values (MAE) were obtained from linear regression formulas relating each computational scheme with experiment.



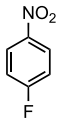
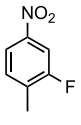
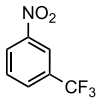
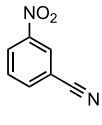
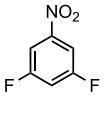
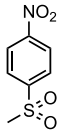
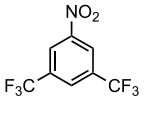
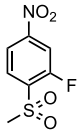
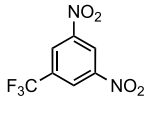
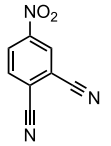
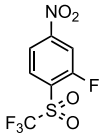
						
-0.0716	0	0.0181	0.0653	0.2513	0.3328	0.4593
						
0.4469	0.5019	0.7078	0.8351	0.8729	1.0272	1.1454

Figure 6. Cycloaddition of multiply substituted nitroarenes with a tetrasubstituted olefin. Electronically differentiated nitroarenes and reported log(k_{rel}) values are shown below.

assumed—multiply substituted aromatics can instead be computed directly. We applied the same computational protocols discussed above to obtain charge and NMR shift values for the 14 nitroarenes. We separately obtained parameters at the ipso-C atom attached to the nitro group and the N atom itself. When compared against experimental relative rates (Figure 7), for the C atom parameters we observe similar trends in the performance as observed above for meta-

https://doi.org/10.1021/acsphyschemau.3c00045
ACS Phys. Chem Au 2024, 4, 259−267

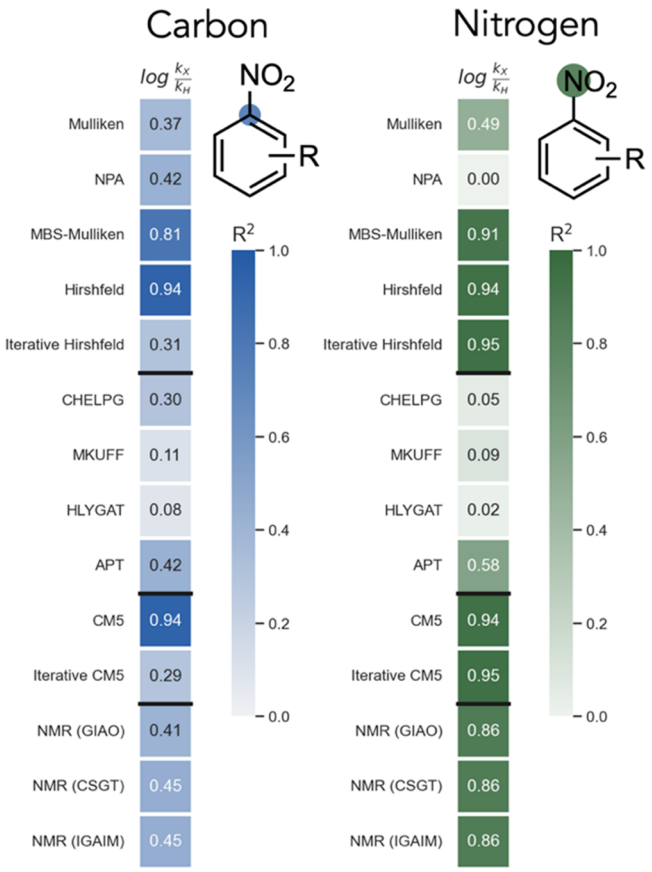


Figure 7. Heatmap correlation plots of computed values from the carbon attached to the nitro group compated with experimental $\log(k_x/k_H)$ values (left) and computed values from the nitro nitrogen compared with experimental $\log(k_x/k_H)$ values (right).

substitution. The best-performing methods are Hirshfeld (R^2 = 0.94), CM5 (R^2 = 0.86), and MBS-Mulliken (R^2 = 0.81), while other methods yield disappointing R^2 values of 0.11−0.42. For parameters obtained at the nitro-group N atom, we obtain an overall improvement in R^2 in most cases, again, except for Class III charges. Mulliken and NPA charges still yield a poor correlation. These results reinforce the idea that the protocol used to describe atomic charge and the location of the atomic position chosen both influence the quantitative ability to describe experimental electronic effects. In both examples studied here, atoms further away from substituents of interest provide improved performance in the statistical modeling of either tabulated Hammett parameters or relative experimental rates.

Our computed parameters and results above were obtained for ground singlet state nitroarenes; however, we also computed parameters for the triplet excited state, which is the reactive intermediate involved in the cycloaddition, observing similar results. Additionally, experimental values were compared against charge and NMR values from the whole NO_2 group, again with similar results (see Supporting Information, section S5).

CONCLUSIONS

We have evaluated the ability of several "bottom-up" computational descriptors to describe experimental "top-down" Hammett parameters and measured relative rates that are influenced by electronic effects. The statistical correlation, as quantified by R^2, varies dramatically according to the type of partial atomic charge scheme (or NMR chemical shift) used. Class III charges such as CHELPG, MKUFF, and HLYGAT correlate poorly, while Hirshfeld and CM5 charges yield the highest correlations with σ_m or σ_p values (R^2 = 0.8−0.9). Higher performance of Hirshfeld charges is in agreement with other literature studies,. Since CM5 charges are based on parametrized corrections to Hirshfeld charges, it is no surprise that they are also similarly high-performing. Another factor influencing the correlation between computed descriptors and experiment is the choice of probe atom: meta- and para-carbon atoms in the aromatic ring produce inferior results to the attached H atoms or other ring substituents. This effect is most pronounced for the prediction of meta-substituent effects, where computed results are the most variable and most methods evaluated at carbon atoms correlate very poorly indeed with σ_m values. These observations suggest that density-based partition schemes such as Hirshfeld should be generally preferred over basis-set-based approaches in obtaining electronic descriptors (such as Hammett parameters) for organic structures, since they are more stable to unphysical perturbations of charge that result from nearby functional groups. From the perspective of predicting unknown Hammett constants or those for multiply substituted systems, we have obtained linear scaling relationships between experimental parameters and all computed protocols studied here. Additionally, we would expect that for more recently developed sigma constants (σ_p^+, σ_α, σ_C, or σ_{jj} parameters), if linear correlations may be formed between the original Hammett constants and these, then these results would propagate to the newer sigma constants. Hirshfeld charges are recommended, yielding the highest R^2 values (0.88 across all σ_m and σ_p values) and smallest errors (0.07) with respect to experiment. The more balanced description across both positions obtained from Hirshfeld charges also implies that this should be the charge model of choice in describing data sets with varied substrates and substitution patterns. While computed NPA charges (used routinely) and NMR chemical shifts produce lower correlations than Hirshfeld charges overall, the performance of much faster ML methods such as *qmdesc* and *CASCADE* are encouraging in specific examples: *qmdesc* Hirshfeld charges evaluated at carbon atoms perform well globally (R^2 = 0.88), while *CASCADE* chemical shifts perform well (R^2 = 0.84) in predicting σ_p values. Overall, these observations lead us to conclude that electronic parametrization strategies should carefully consider the impact of location of probe atoms and how partial atomic charges are derived, since we observe a larger impact than from the choice of density functional and basis set used. Our studies were restricted to the computational description of Hammett parameters; however, as archetypal electronic descriptors that have been applied to LFERs across thousands of studies, we suggest that our findings and recommendations should be considered as part of general computational approaches to electronic parametrization.

ASSOCIATED CONTENT

Data Availability Statement

Cartesian coordinate (xyz format) files and Jupyter Notebooks used for data processing and statistical comparisons are available an accompanying data repository at https://github.com/patonlab/SI-Hammett-Computational-Data.

https://doi.org/10.1021/acsphyschemau.3c00045
ACS Phys. Chem Au 2024, 4, 259−267

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsphyschemau.3c00045.

Computational details and additional references; supplementary figures; tabulated values for all computed descriptors; tabulated linear scaling factors for the prediction of Hammett parameters (PDF)

AUTHOR INFORMATION

Corresponding Author

Robert S. Paton − *Department of Chemistry, Colorado State University, Ft. Collins, Colorado 80523-1872, United States;* orcid.org/0000-0002-0104-4166; Email: robert.paton@colostate.edu; http://patonlab.colostate.edu/

Author

Guilian Luchini − *Department of Chemistry, Colorado State University, Ft. Collins, Colorado 80523-1872, United States;* orcid.org/0000-0003-0135-9624

Complete contact information is available at: https://pubs.acs.org/10.1021/acsphyschemau.3c00045

Author Contributions

G.L. and R.S.P. conceived and designed the study, performed the study, and cowrote the paper.

Funding

The authors thank the NSF under the CCI Center for Computer-Assisted Synthesis (CHE- 2202693) for funding.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work utilized the RMACC Summit supercomputer, which is supported by the NSF (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University, and the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) through allocation TG-CHE180056. We thank Dr. Graham Haug for helpful comments regarding this work.

REFERENCES

(1) Wells, P. R. Linear Free Energy Relationships. *Chem. Rev.* **1963**, *63*, 171−219.

(2) Williams, W. L.; Zeng, L.; Gensch, T.; Sigman, M. S.; Doyle, A. G.; Anslyn, E. V. The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **2021**, *7*, 1622−1637.

(3) (a) Hammett, L. P. Some Relations between Reaction Rates and Equilibrium Constants. *Chem. Rev.* **1935**, *17*, 125−136. (b) Dippy, J. F. J.; Watson, H. B. 105. Relationships between reaction velocities and ionization constants. *J. Chem. Soc.* **1936**, 436−440. (c) Burkhardt, G. N.; Ford, W. G. K.; Singleton, E. 4. The hydrolysis of arylsulphuric acids. Part I. *J. Chem. Soc.* **1936**, 17−25. (d) Hammett, L. P. The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96−103.

(4) Schreck, J. O. Nonlinear Hammett relationships. *J. Chem. Educ.* **1971**, *48*, 103.

(5) Hansch, C.; Leo, A.; Taft, R. A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* **1991**, *91*, 165−195.

(6) Brown, H. C.; Okamoto, Y. Electrophilic Substituent Constants. *J. Am. Chem. Soc.* **1958**, *80*, 4979−4987.

(7) Dust, J. M.; Arnold, D. R. Substituent effects on benzyl radical ESR hyperfine coupling constants. The.sigma..alpha..cntdot. scale based upon spin delocalization. *J. Am. Chem. Soc.* **1983**, *105*, 1221−1227.

(8) (a) Creary, X.; Mehrsheikh-Mohammadi, M. E.; McDonald, S. Methylenecyclopropane rearrangement as a probe for free radical substituent effects. $\sigma^{\bullet}$ values for commonly encountered conjugating and organometallic groups. *J. Org. Chem.* **1987**, *52*, 3254−3263. (b) Creary, X. Super Radical Stabilizers. *Acc. Chem. Res.* **2006**, *39*, 761−771.

(9) Jiang, X.; Ji, G. A self-consistent and cross-checked scale of spin-delocalization substituent constants, the $\sigma^{\bullet}_{JJ}$ scale. *J. Org. Chem.* **1992**, *57*, 6051−6056.

(10) Sarkar, S.; Patrow, J. G.; Voegtle, M. J.; Pennathur, A. K.; Dawlaty, J. M. Electrodes as Polarizing Functional Groups: Correlation between Hammett Parameters and Electrochemical Polarization. *J. Phys. Chem. C* **2019**, *123*, 4926−4937.

(11) (a) Durand, D. J.; Fey, N. Computational Ligand Descriptors for Catalyst Design. *Chem. Rev.* **2019**, *119*, 6561−6594. (b) 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. (c) Gallegos, L. C.; Luchini, G.; St. John, P. C.; Kim, S.; Paton, R. S. Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties. *Acc. Chem. Res.* **2021**, *54*, 827−836.

(12) (a) Charton, M. The estimation of hammett substituent constants. *J. Org. Chem.* **1963**, *28*, 3121−3124. (b) Ertl, P. A Web Tool for Calculating Substituent Descriptors Compatible with Hammett Sigma Constants. *Chemistry-Methods* **2022**, *2*, No. e202200041. (c) Miranda-Quintana, R. A.; Deswal, N.; Roy, R. K. Hammett constants from density functional calculations: charge transfer and perturbations. *Theor. Chem. Acc.* **2022**, *141*, 4.

(13) Pérez, P.; Simón-Manso, Y.; Aizman, A.; Fuentealba, P.; Contreras, R. Empirical Energy−Density Relationships for the Analysis of Substituent Effects in Chemical Reactivity. *J. Am. Chem. Soc.* **2000**, *122*, 4756−4762.

(14) (a) Alegre-Requena, J. V.; Sowndarya, S. V.; Pérez-Soto, R.; Alturaifi, T. M.; Paton, R. S. AQME: Automated quantum mechanical environments for researchers and educators. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2023**, *13*, No. e1663. (b) Żurański, A. M.; Wang, J. Y.; Shields, B. J.; Doyle, A. G. Auto-QChem: an automated workflow for the generation and storage of DFT calculations for organic molecules. *React. Chem. Eng.* **2022**, *7*, 1276−1284.

(15) Santiago, C. B.; Milo, A.; Sigman, M. S. Developing a Modern Approach To Account for Steric Effects in Hammett-Type Correlations. *J. Am. Chem. Soc.* **2016**, *138*, 13424−13430.

(16) Brethomé, A. V.; Fletcher, S. P.; Paton, R. S. Conformational effects on physical-organic descriptors: The case of sterimol steric parameters. *ACS Catal.* **2019**, *9*, 2313−2323.

(17) Bachrach, S. M. Population Analysis and Electron Densities from Quantum Mechanics. *Rev. Comput. Chem.* **1994**, *5*, 171−228.

(18) (a) Gonthier, J. F.; Steinmann, S. N.; Wodrich, M. D.; Corminboeuf, C. Quantification of "fuzzy" chemical concepts: a computational perspective. *Chem. Soc. Rev.* **2012**, *41*, 4671. (b) Cho, M.; Sylvetsky, N.; Eshafi, S.; Santra, G.; Efremenko, I.; Martin, J. M. The atomic partial charges arboretum: trying to see the forest for the trees. *ChemPhysChem* **2020**, *21*, 688−696.

(19) (a) Wiberg, K. B.; Rablen, P. R. Comparison of atomic charges derived via different procedures. *J. Comput. Chem.* **1993**, *14*, 1504−1518. (b) Heidar-Zadeh, F.; Ayers, P. W.; Verstraelen, T.; Vinogradov, I.; Vöhringer-Martinez, E.; Bultinck, P. Information-theoretic approaches to atoms-in-molecules: Hirshfeld family of partitioning schemes. *J. Phys. Chem. A* **2018**, *122*, 4219−4245.

(20) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class IV charge models: A new semiempirical approach in quantum chemistry. *J. Comput. Aided Mol. Des.* **1995**, *9*, 87−110.

(21) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.

(22) Weinhold, F; Landis, C. R. *Valency and bonding: a natural bond orbital donor-acceptor perspective*; Cambridge University Press, 2005.
(23) (a) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44*, 129−138. (b) Hirshfeld, F., XVII Spatial partitioning of charge density. *Isr. J. Chem.* **1977**, *16*, 198−201.
(24) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361−373.
(25) Besler, B. H.; Merz, K. M., Jr; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431−439.
(26) Hu, H.; Lu, Z.; Yang, W. Fitting molecular electrostatic potentials from quantum mechanical calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1004−1013.
(27) Cioslowski, J. A new population analysis based on atomic polar tensors. *J. Am. Chem. Soc.* **1989**, *111*, 8333−8336.
(28) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.
(29) Meister, J.; Schwarz, W. Principal components of ionicity. *J. Phys. Chem.* **1994**, *98*, 8245−8252.
(30) Frisch, M. J. et al. *Gaussian 16*, Revision C.01; Gaussian Inc., Wallingford, CT, 2016.
(31) (a) Galabov, B.; Ilieva, S.; Schaefer, H. F. An Efficient Computational Approach for the Evaluation of Substituent Constants. *J. Org. Chem.* **2006**, *71*, 6382−6387. (b) Monteiro-de-Castro, G.; Duarte, J. C.; Borges, I., Jr Machine Learning Determination of New Hammett's Constants for meta- and para-Substituted Benzoic Acid Derivatives Employing Quantum Chemical Atomic Charge Methods. *J. Org. Chem.* **2023**, *88*, 9791−9802.
(32) (a) Wang, B.; Rong, C.; Chattaraj, P. K.; Liu, S. A comparative study to predict regioselectivity, electrophilicity and nucleophilicity with Fukui function and Hirshfeld charge. *Theor. Chem. Acc.* **2019**, *138*, 124. (b) Rong, C.; Wang, B.; Zhao, D.; Liu, S. Information-theoretic approach in density functional theory and its recent applications to chemical problems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10*, No. e1461.
(33) Nalewajski, R. F.; Parr, R. G. Information theory, atoms in molecules, and molecular similarity. *Proc. Natl. Acad. Sci.* **2000**, *97*, 8879−8882.
(34) Liu, S.; Rong, C.; Lu, T. Information conservation principle determines electrophilicity, nucleophilicity, and regioselectivity. *J. Phys. Chem. A* **2014**, *118*, 3698−3704.
(35) Ewing, D. F., Correlation of NMR Chemical Shifts with Hammett σ Values and Analogous Parameters. In *Correlation Analysis in Chemistry: Recent Advances*; Chapman, N. B.; Shorter, J., Eds.; Springer US: Boston, MA, 1978; pp 357−396. DOI: 10.1007/978-1-4615-8831-3_8.
(36) Erdmann, P.; Greb, L. What distinguishes the strength and the effect of a Lewis acid: analysis of the Gutmann-Beckett method. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202114550.
(37) (a) Abraham, R. J.; Mobli, M. The prediction of ^{1}H NMR chemical shifts in organic compounds. *Spectrosc. Eur.* **2004**, *16*, 16−22. (b) Schulman, E.; Christensen, K.; Grant, D. M.; Walling, C. Substituent effects on carbon-13 chemical shifts in 4-substituted biphenyls and benzenes. Substituent effect transmitted through eight covalent bonds. *J. Org. Chem.* **1974**, *39*, 2686−2690.
(38) Schreckenbach, G.; Ziegler, T. Calculation of NMR shielding tensors using gauge-including atomic orbitals and modern density functional theory. *J. Phys. Chem.* **1995**, *99*, 606−611.
(39) Keith, T.; Bader, R. Calculation of magnetic response properties using atoms in molecules. *Chem. Phys. Lett.* **1992**, *194*, 1−8.
(40) Keith, T. A.; Bader, R. F. Calculation of magnetic response properties using a continuous set of gage transformations. *Chem. Phys. Lett.* **1993**, *210*, 223−231.
(41) Guan, Y.; Coley, C. W.; Wu, H.; Ranasinghe, D.; Heid, E.; Struble, T. J.; Pattanaik, L.; Green, W. H.; Jensen, K. F. Regio-selectivity prediction with a machine-learned reaction representation and on-the-fly quantum mechanical descriptors. *Chem. Sci.* **2021**, *12*, 2198−2208.
(42) Guan, Y.; Shree Sowndarya, S. V.; Gallegos, L. C.; St. John, P. C.; Paton, R. S. Real-time prediction of ^{1}H and ^{13}C chemical shifts with DFT accuracy using a 3D graph neural network. *Chem. Sci.* **2021**, *12*, 12012−12026.
(43) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.
(44) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magariños, M. P.; Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945−D954.
(45) *RDKit: Open-source cheminformatics*. https://www.rdkit.org/ (accessed Dec 15th 2023).
(46) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169−7192.
(47) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(48) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computer Chemistry* **2011**, *32*, 1456−1465.
(49) Schäfer, A.; Huber, C.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829−5835.
(50) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.
(51) Montgomery, J. A., Jr; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VII. Use of the minimum population localization method. *J. Chem. Phys.* **2000**, *112*, 6532−6542.
(52) Vassetti, D.; Labat, F. Evaluation of the performances of different atomic charges and nonelectrostatic models in the finite-difference Poisson-Boltzmann approach. *Int. J. Quantum Chem.* **2021**, *121*, No. e26560.
(53) Fonseca Guerra, C.; Handgraaf, J.-W.; Baerends, E. J.; Bickelhaupt, F. M. Voronoi deformation density (VDD) charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD methods for charge analysis. *J. Comput. Chem.* **2004**, *25*, 189−210.
(54) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. Computational prediction of 1H and 13C chemical shifts: a useful tool for natural product, mechanistic, and synthetic organic chemistry. *Chem. Rev.* **2012**, *112*, 1839−1862.
(55) Ruffoni, A.; Hampton, C.; Simonetti, M.; Leonori, D. Photoexcited nitroarenes for the oxidative cleavage of alkenes. *Nature* **2022**, *610*, 81−86.

https://doi.org/10.1021/acsphyschemau.3c00045
ACS Phys. Chem Au 2024, 4, 259−267