Machine learning determination of new Hammett's constants for *meta*- and *para*-substituted benzoic acid derivatives employing quantum chemical atomic charge methods †

Gabriel Monteiro-de-Castro, Julio Cesar Duarte, Itamar Borges Jr.*
Instituto Militar de Engenharia, Praça Gen. Tibúrcio, 80, Rio de Janeiro, RJ, 22290-270, Brazil.
* E-mail address: itamar@ime.eb.br

Abstract

Hammett's constants σ quantify the electron donor or electron acceptor power of a chemical group bonded to an aromatic ring. Their experimental values have been successfully used in a large variety of applications, but some of them may have inconsistent values or were not measured. For this reason, developing an accurate and consistent set of Hammett's values is paramount. In this work, we employed the machine learning (ML) regression algorithms *Decision Tree Regressor*, the neural network *Multilayer Perceptron Regressor*, and *Lasso Lars IC* in a cross-validation (CV) approach combined with quantum chemical calculations of atomic charges to estimate theoretically the new Hammett's constants σ_m, σ_p, σ_m^0, σ_p^0, σ_p^+, σ_p^-, σ_R, and σ_I for 90 chemical donor or acceptor groups by employing different types of quantum chemical atomic charges of the groups as input properties. New 219 σ values, including previously unknown ones – 46 σ_p^0, 35 $\sigma_p^{+,-}$, and 11 $\sigma_{I,R}$ – are proposed. The different substituent groups were bonded to benzene and *meta*- and *para*-substituted benzoic acid derivatives. Among the investigated atomic charge methods (Mulliken, Löwdin, Hirshfeld, and ChelpG), Hirshfeld's method showed the best regressions for most of the different kinds of σ values. For each type of Hammett constant, linear expressions depending only on the atomic charges of the group were obtained. Correlation coefficients (R^2) as high as 0.945, mean squared errors (MSE) as low as 0.004, and root mean square errors (RMSE) as low as 0.062 were found. The ML approach, in most cases, showed very close predictions to the original experimental values, with the values from *meta*- and *para*-substituted benzoic acid derivatives showing the most accurate values. A new consistent set of Hammett's constants is presented, as well as simple equations for predicting new values for groups not included in the original set of 90.

Keywords: Hammett's constants; Machine learning; Atomic charges; Substituent effect; Benzoic acid.

† This version of the work has been deposited as a preprint at ChemRxiv (DOI:

1. Introduction

Chemical reactions are highly complex processes whose investigation from a theoretical point of view can be challenging. An essential step in the attempt to model reactions was made by Hammett when he developed an equation from extensive experimental observations.[1,2] Hammett's original idea, and its later extensions, have provided the primary basis for quantitatively determining structure-reactivity relationships in physical organic chemistry.[3]

The substituent effect is among the most critical factors affecting the chemical, physicochemical and biochemical properties of compounds, which are traditionally investigated in the context of Hammett's theory.[4] Initially developed for the ionization of benzoic acid derivatives in water for different substitutions at the *para*, *meta*, and *ortho* positions, the Hammett equation is a linear quantitative model describing differences in free energies of reactions, but which has shown itself, over the course of several decades, to have a much more general character. The Hammett equation assumes that the effects of a substituent X and the reaction conditions can be represented by the expression: [3,5–7]

$$\rho\sigma_{\mathrm{X}} = \log\left(\frac{K_{\mathrm{X}}}{K_{\mathrm{H}}}\right) \qquad (1)$$

where K_{X} is the equilibrium constant for a substituted reactant, K_{H} refers to the unsubstituted reactant where H is the hydrogen atom, ρ is a reaction constant that depends only on the conditions of the reactions under study (temperature, solvent, etc.), and σ_{X}, is a numerical constant depending only on the nature of the substituent X and its position in the molecule, quantifying the magnitude of the electron-donating or accepting nature of the substituent.[7]

Electronic effects, such as the inductive and mesomeric effects, are the determinant phenomena in the substituent effect.[8] Therefore, the *ortho*-position is disregarded due to steric effects. This scenario implies that the original σ values are complex quantities, which led to the formulation of new constants that could, e.g., describe the delocalization of positive and negative charges in the system. In 1957, a new set of parameters named σ^{+} and σ^{-} were suggested based on the solvolysis of substituted phenyldimethylcarbinyl chlorides.[9] These parameters include mesomeric effects when the substituent is in direct conjugation with the molecule's reaction center;[10,11] however, it was soon realized that this new set of σ values would not apply to *meta*-substituted systems due to the impossibility of resonance between a *meta*-substituent and the reaction centers.[9,12–14] As can be seen in Figure 1, there is a lack of substituent-reaction center resonance in *meta*-substituted systems, such as deprotonated benzoic acid molecules containing electron donors or acceptor substituents, since there is no significant partial charge on the carbon bonded to the carboxyl group (reaction center). For this reason, in this case, the parameters σ_p^{+} and σ_p^{-} were defined only for substituents on the *para*-position and labeled as follows: σ_p^{+} describes groups that can stabilize positive charges through resonance whereas σ_p^{-} describes groups that can stabilize negative charges through resonance.

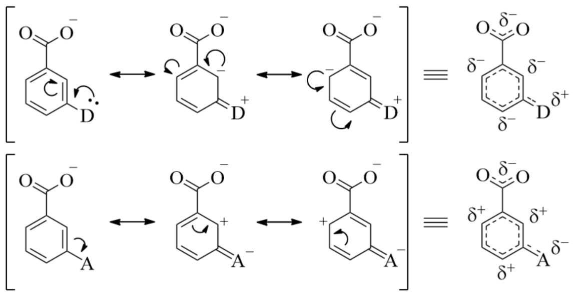

Figure 1. Resonance structures in *meta*-substituted, deprotonated benzoic acid derivatives. "D" is an electron donor group and "A" is an electron acceptor group.

Later, another set of constants was proposed since the original σ set appeared divided into resonance (R) and inductive (I) parts. However, when a series of 4-substituted bicyclo[2,2,2]octane derivatives was investigated in 1953,[15] it became clear that for substituents with a negative resonance effect (-R, negative R) – i.e., for substituents that withdraw electrons from the molecule by delocalization – σ_m and σ_I had (with some exceptions) approximately the same value, thus indicating that σ_m had a significant inductive effect. As a result, the separation between R and I terms was done for the σ_p constant according to the equation

$$\sigma_p = \sigma_I + \sigma_R \tag{2}$$

Therefore, obtaining σ_R values requires the correct determination of the σ_p and σ_I values.[16] Moreover, different σ_R values, such as σ_R^0, σ_R^+, or σ_R^-, can be obtained depending on the kind of σ_p used in Equation (2) – i.e., σ_p^0, σ_p^+, or σ_p^-.[8]

In 1959, another attempt was made to propose parameters that would counteract the resonance effects between the substituent and the reaction center. Therefore, the σ^0 set was created.[17] This set would be defined as unbiased since it would be based on the idea that the resonance between a substituent X and the reaction center (e.g., the carboxyl in the benzoic acid derivatives) should not be considered when analyzing the resonance and inductive effects of the substituent on the molecular electron density.[3,17] Thus, the unbiased set could be evaluated in three ways:[16] I) phenylacetic acid ionization; II) the rates of substituted phenylacetate hydrolysis; or, III) statistical methods. Since, in this case, there is no significant direct substituent-reaction center resonance, the σ_m^0 constants have the intriguing property of being nearly identical to σ_m.[3,16]

Hammett's theory had been recently applied to different chemical problems. Investigations of photophysical properties of a series of 3-amide-6-hydroxy-4-methyl-2-pyridones-5-(4-substituted phenylazo) in the presence of solvents of different polarities;[18] solute-solvent hydrogen bond energies of *para*-substituted benzoic acids in the presence of immiscible solvents;[19] photophysical properties of 14 types of 2-phenylamino-1,10-phenatrolines synthesized with different types of electron-donating and withdrawing substituents are some examples.[20] From a theoretical standpoint, Hammett's constants *per se* were also investigated. A representative work was carried out by Galabov and coworkers[21] who correlated different atomic charge models of carbon atoms in mono-substituted benzene systems with the aforementioned unbiased σ^0 set, which we mentioned, it is based on the notion that the resonance between a substituent X and the reaction center should not be considered.[3,17] The atomic charge methods included Mulliken's,[22] Natural Population analysis (NPA),[23,24] Minimal Basis Set (MBS),[25] Merz-

Kollman (MK),[26] ChelpG,[27] Hirshfeld,[28] and Charge Model 5 (CM5).[29] They showed that the Hirshfeld method provided very good regressions between the substituents' σ^0 values and the *meta*- and *para*-carbon atom (position in relation to the substituent). Batagin-Neto et al.[30] used the resonance and inductive σ values (σ_R and σ_I) of –H, –CCH, –Ph, –Me, –CN, –F, $-NO_2$, $-NH_2$, –OH, and –OMe side groups to study the effects of these substituents on the optoelectronic properties of polypyrrole (PPy). They were able to correlate FMO energies and the maximum absorption wavelength (λ_{max}) with the side groups' parameters and thus propose a set of new PPy derivatives with potential application in optoelectronic devices.

Since not all substituents have measurements for all values of all types of σ, efforts have been made to find ways to calculate these parameters. Recently, Ertl[31] developed an approach based on semiempirical atomic charge calculations and the Leave-One-Out (LOO) cross-validation method.[32] The author used the GFN2-xTB semiempirical method, the first broadly parametrized tight-binding method to include electrostatic and exchange-correlation Hamiltonian terms beyond the monopole approximation,[33] and concluded that three carbon atoms in the benzene would be necessary to correlate with the atomic charges. These are the carbons on the *meta*- and *para*-positions in relation to the substituent and the carbon directly attached to the substituent.[31]

Another group employed regressive machine learning (ML) models with Hammett's constants as inputs to predict energies of Frontier Molecular Orbitals (FMO) of tungsten-banzylidyne complexes with different ligands.[34] The authors started by calculating, using Density Functional Theory (DFT)//B3LYP/Def2-SVP, the redox potentials as well as the corresponding experimental values for their complexes. The DFT results and the σ_m and σ_p values of the ligands were then used to train different regressive models. They have shown that their method could accurately predict FMO energies. The recent work of von Lilienfeld et al.[35] is another interesting application: they applied Machine Learning (ML) to predict activation energies of S_N2 reactions in non-aromatic molecular scaffolds based on Hammett's original approach.

Inspired on previous investigations by Galabov et al.[21] and Ertl[31,36] in this work, we use ML methods and different types of quantum chemical atomic charges to accurately determine the values of different Hammett's σ constants for a set of 90 donor- and acceptor-substituents. ML is a computational approach that has been very useful in unveiling several new insights into a plethora of different chemical problems, including predictions of major products of Diels-Alder reactions,[37] reaction yields using perturbation theory combined with ML,[38] activation barrier energies of homogeneous reactions,[39] and even in the atomistic simulations of energetic materials.[40] Therefore, other ML applications to chemistry abound in different and related areas of chemistry, such as materials science.[41–43] A recent review[44] discusses the recent progress of ML models applied to organic photovoltaic (OPV) devices. The authors argue that the ability to predict the performance of OPVs from basic structural data has been demonstrated to be quite promising. However, for reaching a good level of prediction accuracy, a sufficiently large dataset with size, variety, and homogeneity is usually important,[44] although for smaller datasets, accurate data (e.g., from quantum chemical calculations) a careful selection input of features is crucial for obtaining good results, as recently has been recently showed.[45] In this vein, we recently used a ML approach combined with carefully selected quantum chemical input features to successfully investigate molecular properties affecting the sensitivity, thus the safety, of explosives.[46]

2. Computational methods

2.1. Atomic charge methods

Atomic charges are not directly measured because they are not quantum mechanical observables, but are helpful chemical constructs computable with different quantum chemical approaches. They can be used to describe the electron density of a molecule obtained with any electronic structure method (e.g., DFT), thus, to interpret chemical reactivity, non-covalent interactions, rates of reactions, and other properties in many chemical systems. Among the different methods employed to calculate atomic charges, the most popular include Mulliken,[22] Löwdin,[47,48] Hirshfeld,[28] and ChelpG[27] methods, used in this work.

Due to its simplicity and availability as the default output in many computational chemistry software packages, the Mulliken atomic charge method[22] is probably the most used. In this method, the expansion coefficients are taken from the Hartree-Fock (HF) variational method, and the electronic population is separated according to the atomic orbital (AO) contributions. Despite its low computational cost, Mulliken charges suffer from two major problems: (i) the electronic density between two atoms is equally divided between them, regardless of the electronegativity; and (ii) the method employs a set of non-orthogonal basis set that can lead to undesired results.[49]

Löwdin atomic charges[47,48] are an improvement over Mulliken charges because it forms orthogonal basis sets by applying symmetric transformations on all orbitals, hence eliminating the overlap partitions which solves the aforementioned problem (ii). Löwdin charges also have the advantage over Mulliken charges of being less dependent on the size of the basis set.

The Hirshfeld method [28] starts with a "*promolecule*" with neutral spherically symmetric atoms at the same coordinates of the atoms of the real molecule. In this technique, the molecular electron density at a given point in space is shared by the surrounding atoms according to the distance between that point and each atomic nucleus, which is considered when calculating the atomic partial charges in this approach. The electron density of the isolated atom at a certain distance from the nucleus, which corresponds to the distance between that atom's point on the molecule and its nucleus, is included by weighting each atom's contribution. Although this method is an improvement over Mulliken's, there is a general agreement in the literature that the charges calculated using the Hirshfeld method typically are very small, close to zero.[50–52] This is true, according to Ayers, since the weighting factor is evaluated in a way that the molecule's atom is considered to be very similar to the isolated atom. Despite this particularity of the Hirshfeld method, recent studies have concluded that it accurately predicts regioselectivity in electrophilic aromatic substitution reactions, the energy of hydrogen bonds between methane molecules, and the charges of atoms in covalent bonds, among other relevant chemical properties.[49] Our results here indicate also the same behavior concerning Hammett's constants.

Charges employing the Electrostatic Potential on a Grid (ChelpG) methods are based on the original Chelp program[53] in which Lagrange multiplier methods are used to fit atom-centered points to the molecule's electrostatic potential field, keeping the molecular total charge constant as a constraint in the fitting process. The major difference between the Chelp and ChelpG methods is that the ChelpG process uses an algorithm based on regularly spaced points to make point selections.[49] This algorithm differs from the original Chelp program since it was found that it was not invariant to coordinate rotations and internal bond rotations, among other properties.[54]

2.2. Quantum chemical approach

All quantum chemical calculations employed the Orca package version 5.0.3.[55] The first step was to optimize the gas phase geometry of different substituted benzene (B_Z) and *meta*- and *para*-substituted benzoic acid (m-B_AX and p-B_AX, respectively) molecules. Figure 2 shows the investigated B_A and B_Z systems, Table S1 lists their cartesian coordinates, and Table S2 collected the different experimental σ values available in the literature[1,2,7,10,11,21,56–58] for 90 different substituents groups – 89 were the same groups chosen by the author in Ref.31 and, additionally, we included hydrogen as the 90th substituent since all σ values of this atom are taken as the reference value equal zero as suggested by Galabov.[21] The chemical structures of the substituents can be seen in Figure S2 of the Supporting Information (SI). Here, we investigated the following Hammett's parameters: σ_m, σ_p, σ_m^0, σ_p^0, σ_p^+, σ_p^-, σ_R, and σ_I. There are also other types of Hammett's constants, which are based on different reactions or reactions series; however, we chose these ones because they were obtained from the original set (σ_m and σ_p).[59] Since some of the Hammett parameters for several substituents of the dataset are, to the best of our knowledge, still unknown, some approximations were made. For the unknown σ values, the approximations made were: $\sigma_m^0 = \sigma_m$[3,16] for those substituents in which the σ_m^0 values were not found in the literature. Moreover, for each unknown value of σ_R or σ_I, Equation (2) was employed to determine the other constant.

The functional B3LYP[60] combined with the Def2-TZVP[61] basis set was used for the optimization of the molecular geometries. Due to its low computational cost and good accuracy,[62] the B3LYP functional is widely used in the geometrical optimization of organic molecular structures. Vibrational frequency calculations confirmed the nature of the converged structures. When imaginary frequencies were found, the TightOPT keyword in Orca was employed, thereby eliminating them. The converged systems were named B_ZX, m-B_AX, and p-B_AX, where X is the substituent and *m* and *p* refer to the position of the substituent X (*meta* or *para*), as shown in Figure *2*.

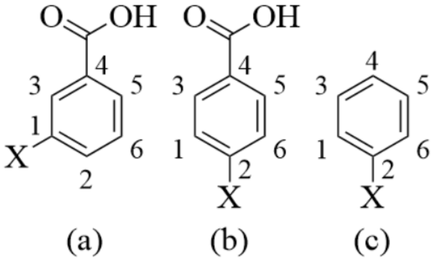

Figure 2. (a) *meta*- (m-B_AX) and (b) *para*-Benzoic Acid (p-B_AX) and (c) Benzene (B_Z) systems.

Single point calculations on the converged geometries were carried out using the Coulomb-attenuating method with B3LYP (CAM-B3LYP)[63] exchange-correlation functional and the same Def2-TZVP[61] basis-set. Mulliken (M), Löwdin (L), Hirshfeld (H), and ChelpG (CG) atomic charges were computed from these single point electronic densities. The values of the atomic charges $q_{i,x}$ (where x = M, L, H, or CG and i = 1 to 6 is the number of the ring carbon atom defined in Figure 2) of all *meta*/*para*-substituted benzoic acid and substituted benzene systems were then subtracted by the corresponding charge $q_{i,x}$ of the unsubstituted B_A and B_Z, respectively (i.e., with H atoms) as suggested by Galabov.[21] Therefore, atomic charge variations ($\Delta q_{i,x}$) are given by the following equations

$$\Delta q_{i,x}^{B_A} = q_{i,x}\left(\mathrm{B_A X}_{m,p}\right) - q_{i,x}(\mathrm{B_A H}) \tag{3a}$$

$$\Delta q_{i,x}^{B_Z} = q_{i,x}(\mathrm{B_Z X}) - q_{i,x}(\mathrm{B_Z H}) \tag{3b}$$

were obtained. In Equations 3, $q_{i,x}(B_A\mathrm{X}_{m,p})$ and $q_{i,x}(B_Z\mathrm{X})$ represent the atomic charge x of carbon i of the X-substituted *m*- or *p*-B_AX and B_ZX systems. In contrast, $q_{i,x}(\mathrm{B_A H})$ and $q_{i,x}(\mathrm{B_Z H})$ are the atomic charge x of carbon i of the unsubstituted B_A and B_Z, respectively.

2.2. Machine learning (ML) algorithms

In the ML approach we used the experimental σ values as the target feature and calculated the atomic charge differences defined by Equations (3) as the input features for training the algorithms. For a preliminary selection of the algorithms, the *Lazy Predict* tool from the Python Scikit-Learn[64] was employed. Many different regression algorithms were evaluated employing their default hyperparameters. We then selected three of these algorithms, namely, the *Decision Tree Regressor*, *Lasso Lars IC*, and *Multilayer Perceptron Regressor (MLPR),* because these regressors provided satisfactory initial results. A brief description of these algorithms is given below.

Decision Tree Regressor (DTR) – Regression and classification models can be constructed using Decision Trees (DT) algorithms.[65,66] The DTs are made from a “root node” containing the input data that then the node splits into sub-nodes until it reaches the “leaves nodes”, i.e., the final nodes which are not split. Splitting is usually based on binary decisions which separates one or more classes from the remaining ones.[67] DTR is able to find complex nonlinear relationships and make very precise predictions when trained by high-quality datasets, such as the present quantum chemical data. An example of this power of DTR algorithms is the 2017 proposition that by simulating the branching structure of DTs with deep neural nets enables one to explain neural network models, thus inheriting the advantages of parametric (the algorithm fits the dataset into a known model) and non-parametric models (any fitting model can be chosen depending on the pattern observed in the input data).[68,69]

Lasso Lars IC (LLIC) – The Lasso Least Angle Regression (LARS) Irrepresentable Condition (IC) algorithm combines: (i) the Lasso method which can simultaneously carry out the estimation of the parameters and model selection in linear regression models;[70,71] (ii) the LARS high-dimensional algorithm which finds the aspect of input features (in our case, the carbon atoms indicated in Figure 2) which provide better correspondence with the target feature (in case, the Hammett’s constants), until a set of the best features is obtained;[72,73] and (iii) the condition which states that Lasso can consistently select the models if the features not included in the model are “irrepresentable” by the features that are in the true model.[74]

MLP Regressor (MLPR) – The Multilayer Perceptron (MLP) is a class of feed-forward Artificial Neural Network (ANN). MLP architecture connects the input, hidden, and output layers in a feed-forward way.[75] As a type of supervised learning, MLP uses backpropagation[76] (a type of gradient-descent algorithm in which predetermined error-function values are calculated) to train the network. The values, which were reintroduced into the network following the computation, are used to adjust the weights of each layer's neurons.[75]

The metrics for all calculations were based on the coefficient of determination (R^2) and the Root Mean Square Error (RMSE), the former is given by

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{O}_i - \bar{O}_i)}{\sum_{i=1}^{n}(O_i - \bar{O}_i)} \tag{4}$$

while the latter can be obtained by

$$\mathrm{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(O_i - \hat{O}_i\right)\right]^{0.5} \tag{4}$$

where n is the total number of samples, O is the observed value, $\hat{O}$ is the value estimated by the algorithm, and $\bar{O}$ is the mean of all values. The Mean Squared Error (MSE) is given by the square of the RMSE and was also used in this work.

To evaluate the performance of the ML procedures, Cross-validation (CV) was carried out by employing the Leave-One-Out (LOO) method. LOO is a K-fold CV method[77] where K is equal to the number of elements in the data, thus, the training set gets K-1 elements while the testing set gets only one element.[32] The R^2, MSE, and RMSE values related to the CV step of the methodology were named R^2_{LOO}, MSE$_{LOO}$, and RMSE$_{LOO}$, respectively. Regressions were carried out using the *cross_val_predict* function of Scikit-learn.[64]

Finally, based on the ML results and the benzoic acid derivative systems, new σ values are proposed. The ML algorithms correspond the atomic charge variations to different available Hammett's constants in a linear fashion, thus generating equations that are able to determine new σ values of different substituent groups not included in the original dataset.

3. Results and discussion

The following discussion is divided into three parts. Firstly, the results obtained from the *GridSearchCV* algorithm that defines the best ML hyperparameters for all regressors and the atomic charges are presented. Secondly, the regressions between the best atomic charge model and the substituted benzene (B$_Z$X) systems are discussed. Finally, we discuss the regressions for both the *meta*- and *para*-substituted benzoic acid (*m*,*p*-B$_A$X) systems employing the best model of atomic charges and, according to the results, we propose a new set of consistent values for the Hammett's constants.

3.1. Machine learning regressors and atomic charges

The charge of all six carbon atoms, comprising the benzene ring in all B_ZX (similarly to Galabov's work[21]) and *m*,*p*-B_AX systems, were included in the dataset of the ML algorithms. The difference between these charges and the unsubstituted B_AH and B_ZH systems are listed in the Supporting Information (SI) in Tables S3-S14. Table S15 in the SI shows the best regressors obtained by the *Lazy Prediction* selection algorithm. As mentioned before, these are the *Decision Tree Regressor*, *MLP Regressor*, and *Lasso Lars IC* (Table S15). The best hyperparameter set for each regressor was found using the *GridSearchCV* function and are shown in Table S16.

Once the best hyperparameters were found, regressions followed by CV – employing a LOO approach – were carried out. Table 1 below shows the errors found for each regression using the atomic charge models and the best hyper-parametrized regressors obtained from the *Lazy Prediction* step. Since not all Hammett's constants are known for all the 90 substituents, the base number n (i.e., the number of molecules used in the ML procedure) varies for each σ. For the 90 substituents used in this work, we were able to use $n = 44$ (σ_p^0), 55 ($\sigma_p^{+,-}$), 79 ($\sigma_{I,R}$), and 90 ($\sigma_{m,p}$ and σ_m^0).

Table 1. R^2_{LOO} values of the CV step involving the DTR, LLIC, and MLPR algorithms for the regressions between Mulliken, Löwdin, Hirshfeld, and ChelpG atomic charges and Hammett's constants. The number of molecules used in the input data for each case is given by n. The color scheme represents greater (green), medium (yellow), and lower (red) values of R^2_{LOO}.

σ	**Regressor**	**Mulliken**	**Löwdin**	**Hirshfeld**	**ChelpG**	***n***
σ_p	DTR	0.560	0.882	0.827	-0.137	90
	LLIC	0.783	0.918	0.931	0.714	
	MLPR	0.798	0.926	0.930	0.714	
σ_m	DTR	0.275	0.644	0.844	-0.018	90
	LLIC	0.541	0.856	0.878	0.758	
	MLPR	0.541	0.862	0.875	0.758	
σ_p^0	DTR	0.545	0.846	0.870	-0.028	44
	LLIC	0.783	0.931	0.945	0.795	
	MLPR	0.782	0.930	0.933	0.795	
σ_m^0	DTR	-0.060	0.639	0.785	0.123	90
	LLIC	0.505	0.864	0.885	0.734	
	MLPR	0.505	0.861	0.883	0.734	
σ_p^+	DTR	0.709	0.849	0.902	-0.133	55
	LLIC	0.875	0.909	0.921	0.635	
	MLPR	0.868	0.907	0.931	0.635	
σ_p^-	DTR	0.499	0.745	0.673	-0.093	55
	LLIC	0.562	0.743	0.742	0.635	
	MLPR	0.541	0.752	0.741	0.636	
σ_R	DTR	0.668	0.768	0.822	-0.150	79
	LLIC	0.788	0.861	0.893	0.513	
	MLPR	0.779	0.869	0.895	0.539	
σ_I	DTR	-0.060	0.790	0.801	0.004	79
	LLIC	0.294	0.799	0.845	0.532	
	MLPR	0.288	0.798	0.850	0.532	

From all the computed regressions using the atomic charge differences as input features defined by Eqs. 3, Hirshfeld's and (to a slightly lower degree) Löwdin's methods gave the best overall good agreement with the different experimental σ values of the substituents shown in Table S1. In Ref. [21], a similar accuracy was obtained by relating σ^0 values with the same variations of Hirshfeld atomic charges of 20 different substituted benzene systems (19 of these substituents were also used in this work). The authors concluded that Hirshfeld charges describe well the properties of aromatic systems,[21] and our work, in a certain sense, extends theirs. From now on, we will discuss results based on the Hirshfeld charges.

Concerning the ML algorithms, our results indicated that the LLIC and MLPR give better linear correlations compared to DTR. However, LLIC was found to be slightly better than MLPR. We were able to achieve a coefficient of determination for the CV step with the LOO algorithm (R^2_{LOO}) up to 0.945 (Table S17, LLIC), 0.895 (Table S18, LLIC), and 0.946 (Table S19, MLPR). The main results, discussed below, were obtained by regressions employing the LLIC algorithm.

3.2. Benzene derivatives

As mentioned before, previous works on the substituted benzene systems inspired this investigation.[21,31,36] Therefore, we first discuss the results involving the benzene derivatives. When comparing Tables S17 and S18 (*p*- and *m*-B_AX) with Table S19 (B_ZX), the regression error trends favor the results using the B_AX derivatives in contrast with the B_ZX systems. This behavior agrees nicely with Hammett's original work on substituted benzoic acids, although in this work, all the systems were modeled in gas phase, whereas Hammett originally performed his experiments in water at 25 ºC.[1,2] The effects that a given substituent X would promote on either B_A and B_Z systems are distinct because the former possesses a carboxyl group which provides additional influence on the electron density in the benzene core.

The aforementioned work by Ertl[31] confirmed, however, that the benzene can still be used as a molecular core system to compute consistent sets of Hammett's constants that can be compared with the available benzoic acid experimental results. Using 89 different substituents, Ertl showed, by means of a CV approach employing the LOO algorithm, that the atomic charge of the carbon directly attached to the substituent and the carbon atoms on the *meta*- and *para*-positions in relation to the substituent (i.e., carbons 2, 3, and 4 in Figure 2c, respectively) resulted in equations with the good metrics of $R^2_{LOO} = 0.873$, $R^2_{regression} = 0.889$, and MAE $= 0.053$ (σ_m) and $R^2_{LOO} = 0.915$, $R^2_{regression} = 0.926$, and MAE $= 0.068$ (σ_p), where MAE is the Mean Absolute Error. Note that our results are more accurate.

Another work relating atomic charges and specific carbon atoms in substituted-benzene systems was performed by Galabov et al.,[21] who concluded that the Hirshfeld atomic charges of the carbons in the *meta*- and *para*-positions in relation to the substituent's position showed good correlation for the σ^0 set of constants (R^2 up to 0.959), the only ones investigated. This is quite interesting since the suggestion of the σ^0 set constants did not consider the resonance between the substituent and the reaction center when analyzing the resonance and inductive effects of the substituent on the electron density of the molecule.[3,17] Thus, this suggests that the benzene core is a good starting point to determine theoretically Hammett's constants.

Our ML regressions using LLIC together with the CV step for the B_Z derivatives indicated that certain carbon atoms in the benzene ring could be used as the most convenient references to calculate Hammett's parameters. This, however depend on the type of σ value. Overall, the best regressions obtained with the B_Z derivatives were those only for the constants σ_p^0 and σ_p, probably due to possible resonance effects, which are more predominant in *para*-substituted derivatives (see Figure 1).

Consequently, we decided to employ only the benzoic acid systems for obtaining a consistent set of Hammett's parameters for the 90 substituents. The results are discussed in the next section.

3.3. Benzoic acid derivatives

Table 2 below lists the predicted values of Hammett's constants using the *Lasso Lars IC* (LLIC) algorithm and the Hirshfeld charges because they gave the lowest errors. The major outliers only for the σ_p^- constants, namely the –NH_2, –NMe_2, –NEt_2, and –NHC(=O)Me groups (as will be explained shortly), were removed. Equation (2) was used for calculating σ_R the constants, and we employed only the *meta*- and *para*-substituted benzoic acid derivatives. The equations derived from the regressions used to obtain each type the Hammett's constants as function of the charge differences, listed in Table 2, are:

$$\sigma_p = 0.06 + 1.54\Delta q_2 + 7.83\Delta q_3 + 35.51\Delta q_4 + 6.46\Delta q_5 \tag{5a}$$
$$\sigma_p^0 = 0.06 + 19.45\Delta q_4 + 39.53\Delta q_5 + 1.78\Delta q_6 \tag{5b}$$
$$\sigma_p^+ = -0.09 + 2.80\Delta q_2 + 74.78\Delta q_4 - 17.37\Delta q_5 - 6.47\Delta q_6 \tag{5c}$$
$$\sigma_p^- = 0.25 + 38.58\Delta q_4 + 17.57\Delta q_5 + 2.79\Delta q_6 \tag{5d}$$
$$\sigma_m = 0.03 + 1.77\Delta q_1 + 4.40\Delta q_2 + 4.55\Delta q_3 + 29.88\Delta q_4 + 3.25\Delta q_5 + 8.09\Delta q_6 \tag{5e}$$
$$\sigma_m^0 = 0.05 + 1.27\Delta q_1 + 6.34\Delta q_2 + 3.39\Delta q_3 + 35.69\Delta q_4 + 0.55\Delta q_5 + 8.94\Delta q_6 \tag{5f}$$
$$\sigma_I = 0.07 + 1.13\Delta q_1 + 5.89\Delta q_2 + 44.84\Delta q_4 - 6.08\Delta q_5 + 7.70\Delta q_6 \tag{5g}$$
$$\sigma_R = \sigma_p - \sigma_I \tag{5h}$$

Table 2. Machine-Learning-based predictions of different σ values for different substituent groups bonded to the benzoic acid. Values obtained using Eqs. 5 from the LLIC algorithm and the Hirshfeld atomic charge model.

–X	σ_m	σ_p	σ_R	σ_I [a]	σ_p^+	σ_p^- [b]	σ_m^0	σ_p^0
–H	0.03	0.06	-0.02	0.07	-0.09	0.25	0.05	0.06
–Br	0.38	0.28	-0.19	0.47	0.01	0.42	0.41	0.37
–2-pyrimidinyl	0.13	0.21	0.13	0.08	0.24	0.39	0.12	0.09
–2-furyl	0.10	0.00	-0.14	0.15	-0.24	0.13	0.10	0.00
–3-thienyl	0.07	0.00	-0.13	0.13	-0.23	0.12	0.07	0.01
–2-thienyl	0.12	0.03	-0.17	0.20	-0.23	0.16	0.14	0.05
–3-pyridyl	0.15	0.12	-0.07	0.19	-0.04	0.25	0.15	0.09
–2-pyridyl	0.13	0.10	-0.03	0.13	0.06	0.22	0.12	-0.02
–4-pyridyl	0.19	0.19	-0.02	0.21	0.07	0.35	0.19	0.17
–*c*-C_5H_9	-0.03	-0.11	-0.14	0.03	-0.35	-0.02	-0.04	-0.12
–*c*-C_6H_{11}	-0.03	-0.12	-0.15	0.03	-0.36	-0.03	-0.04	-0.13
–C_6H_4-3-Br	0.14	0.11	-0.06	0.17	-0.06	0.26	0.14	0.12
–C_6H_4-4-Br	0.12	0.09	-0.08	0.16	-0.09	0.22	0.12	0.08
–C_6H_4-4-*t*-Bu	0.04	-0.03	-0.12	0.09	-0.25	0.09	0.04	-0.04
–C_6H_4-4-Et	0.04	-0.03	-0.12	0.09	-0.25	0.09	0.04	-0.04

–C_6H_4-4-Me	0.04	-0.03	-0.12	0.09	-0.25	0.09	0.04	-0.04
–C_6H_4-3-Cl	0.14	0.11	-0.07	0.18	-0.06	0.24	0.14	0.09
–C_6H_4-4-Cl	0.11	0.08	-0.08	0.16	-0.10	0.21	0.12	0.07
–C_6H_4-3-F	0.13	0.10	-0.06	0.16	-0.07	0.24	0.13	0.09
–C_6H_4-4-F	0.09	0.04	-0.10	0.14	-0.16	0.17	0.09	0.04
–C_6H_4-3-NO_2	0.22	0.22	-0.04	0.26	0.10	0.36	0.23	0.19
–C_6H_4-4-NO_2	0.23	0.25	-0.01	0.26	0.15	0.42	0.24	0.24
–C_6H_4-4-OMe	0.01	-0.08	-0.14	0.07	-0.33	0.04	0.01	-0.07
–Ph	0.07	0.01	-0.10	0.11	-0.19	0.14	0.07	0.00
–CCH	0.33	0.29	-0.01	0.31	0.20	0.47	0.34	0.26
–*t*-Bu	-0.05	-0.12	-0.12	0.00	-0.35	-0.03	-0.07	-0.14
–CF_3	0.42	0.53	0.10	0.42	0.52	0.79	0.45	0.54
–CH_2Ph	-0.01	-0.05	-0.10	0.05	-0.30	0.07	-0.01	-0.03
–$(CH_2)_2Ph$	0.00	-0.11	-0.17	0.07	-0.36	-0.02	0.00	-0.12
–*n*-C_5H_{11}	-0.03	-0.11	-0.15	0.04	-0.36	-0.02	-0.03	-0.11
–*n*-Bu	-0.03	-0.14	-0.18	0.04	-0.39	-0.06	-0.03	-0.16
–*n*-Pr	-0.02	-0.10	-0.14	0.05	-0.34	-0.01	-0.03	-0.10
–$(CH_2)_2COOH$	0.07	0.03	-0.10	0.13	-0.18	0.15	0.07	0.03
–Et	-0.02	-0.09	-0.14	0.05	-0.34	0.00	-0.02	-0.10
–$CH_2CH=CH_2$	-0.01	-0.04	-0.09	0.05	-0.24	0.05	-0.01	-0.07
–*i*-Bu	0.00	-0.10	-0.15	0.05	-0.34	-0.01	-0.01	-0.12
–CH_2CN	0.22	0.23	-0.05	0.28	0.08	0.33	0.23	0.18
–$CH_2C(=O)NH_2$	0.08	0.06	-0.07	0.13	-0.13	0.19	0.08	0.06
–CH_2NMe_2	-0.01	-0.06	-0.09	0.03	-0.29	0.07	-0.01	-0.06
–CH_2NH_2	0.00	-0.02	-0.09	0.06	-0.22	0.10	0.00	-0.03
–CH_2OMe	-0.02	-0.08	-0.12	0.04	-0.29	-0.01	-0.02	-0.14
–CH_2OH	0.03	0.01	-0.05	0.06	-0.19	0.15	0.04	0.03
–Me	-0.02	-0.09	-0.15	0.06	-0.35	-0.01	-0.02	-0.09
–$CHPh_2$	0.04	-0.04	-0.14	0.10	-0.27	0.06	0.04	-0.07
–$CH=CH_2$	0.10	0.06	-0.06	0.12	-0.11	0.21	0.10	0.05
–*c*-C_3H_5	-0.02	-0.16	-0.24	0.08	-0.50	-0.08	-0.01	-0.13
–*c*-C_4H_7	-0.05	-0.14	-0.16	0.03	-0.41	-0.05	-0.05	-0.14
–*s*-Bu	-0.02	-0.11	-0.15	0.04	-0.35	-0.02	-0.03	-0.12
–*i*-Pr	-0.02	-0.10	-0.14	0.04	-0.34	-0.01	-0.03	-0.11
–CH(Me)OH	-0.02	-0.06	-0.09	0.03	-0.33	0.09	-0.02	0.01
–CHF_2	0.32	0.33	0.02	0.31	0.26	0.54	0.34	0.34
–CHO	0.39	0.60	0.31	0.28	0.73	0.91	0.39	0.55
–Cl	0.38	0.25	-0.22	0.48	-0.04	0.36	0.40	0.33
–CN	0.67	0.73	0.13	0.60	0.79	0.98	0.69	0.68
–C(=O)Ph	0.31	0.38	0.13	0.25	0.41	0.64	0.30	0.34
–C(=O)Et	0.32	0.42	0.17	0.24	0.49	0.70	0.31	0.38
–C(=O)Me	0.32	0.44	0.20	0.24	0.52	0.73	0.31	0.41
–$C(=O)NH_2$	0.32	0.37	0.09	0.28	0.32	0.65	0.31	0.42
–C(=O)NHPh	0.35	0.39	0.06	0.33	0.31	0.67	0.35	0.46
–C(=O)NHMe	0.28	0.32	0.06	0.25	0.23	0.59	0.28	0.37
–C(=O)OEt	0.30	0.42	0.21	0.21	0.51	0.68	0.29	0.35
–C(=O)OMe	0.32	0.45	0.21	0.23	0.54	0.71	0.31	0.37
–C(=O)OH	0.38	0.54	0.25	0.28	0.67	0.81	0.38	0.45
–F	0.40	0.17	-0.40	0.56	-0.24	0.14	0.41	0.26
–I	0.37	0.28	-0.15	0.44	0.04	0.45	0.39	0.36

–NEt$_2$	-0.27	-0.75	-0.77	0.02	-1.59	-0.86	-0.27	-0.53
–NMe$_2$	-0.26	-0.75	-0.79	0.05	-1.58	-0.86	-0.25	-0.52
–NH$_2$	-0.11	-0.58	-0.76	0.18	-1.39	-0.66	-0.09	-0.34
–NHPh	-0.04	-0.48	-0.69	0.21	-1.20	-0.57	-0.03	-0.32
–NHEt	-0.20	-0.70	-0.80	0.10	-1.55	-0.81	-0.19	-0.49
–NHMe	-0.19	-0.69	-0.80	0.11	-1.55	-0.80	-0.18	-0.48
–NHC(=O)Ph	0.17	-0.12	-0.46	0.34	-0.50	-0.19	0.17	-0.15
–NHC(=O)Me	0.07	-0.02	-0.25	0.22	-0.46	0.00	0.08	0.08
–NHOH	0.00	-0.34	-0.56	0.22	-0.98	-0.39	0.01	-0.21
–NHSO$_2$Me	0.21	-0.07	-0.52	0.44	-0.58	-0.10	0.23	0.03
–NO$_2$	0.72	0.87	0.23	0.64	1.05	1.09	0.73	0.78
–OPh	0.18	-0.20	-0.54	0.34	-0.72	-0.28	0.17	-0.08
–OCF$_3$	0.42	0.23	-0.32	0.55	-0.17	0.24	0.42	0.32
–OPr	-0.01	-0.32	-0.56	0.24	-0.95	-0.40	0.00	-0.19
–OEt	0.04	-0.32	-0.57	0.25	-0.94	-0.39	0.02	-0.19
–OMe	0.06	-0.29	-0.56	0.27	-0.91	-0.36	0.05	-0.15
–O-*i*-Pr	0.03	-0.35	-0.60	0.25	-0.93	-0.47	0.01	-0.25
–OCHF$_2$	0.38	0.13	-0.39	0.51	-0.24	0.09	0.37	0.22
–OC(=O)Me	0.32	0.18	-0.22	0.40	-0.13	0.22	0.32	0.25
–OH	0.14	-0.23	-0.60	0.37	-0.90	-0.27	0.14	-0.02
–P(=O)(OH)$_2$	0.34	0.51	0.21	0.29	0.54	0.82	0.38	0.49
–SMe	0.07	-0.16	-0.41	0.24	-0.64	-0.07	0.10	-0.01
–SO$_2$Ph	0.51	0.60	0.13	0.47	0.63	0.88	0.54	0.60
–SO$_2$Me	0.52	0.64	0.16	0.48	0.69	0.94	0.55	0.65
–SO$_2$NH$_2$	0.51	0.59	0.10	0.49	0.61	0.85	0.55	0.56

[a] Obtained from the regressions of the *m*-B$_A$X derivatives.

[b] The values of σ_p^- were calculated by disregarding, during the ML procedure, the –NH$_2$, –NMe$_2$, –NEt$_2$, and –NHC(=O)Me substituents.

To perform an additional test of the predictability of Eqs. 5, we examined three additional substituent groups, namely –CCl$_3$, –NHCHO, and –NHCONH$_2$. The same optimization and single-point approach used for the benzoic acid systems were done for these substituents, as well as computing the Hirshfeld atomic charge differences. Values are shown in Table below. The larger difference is for the value of σ_m of the –NHCHO group ($\Delta\sigma_p = 0.22$). The other results show that our ML-based equations predict quite accurately the values of different Hammett's parameters for these three chemical groups, not present in the original set of 90.

Table 3. Machine-Learning-based predictions of different σ values for the new *m*-B$_A$X test systems. Calculations with the *Lasso Lars IC* algorithm. Values in parenthesis are the values from the literature.

–X	$\boldsymbol{\sigma_m}$	$\boldsymbol{\sigma_p}$	$\boldsymbol{\sigma_R}$	$\boldsymbol{\sigma_I}$	$\boldsymbol{\sigma_p^+}$	$\boldsymbol{\sigma_p^-}$	$\boldsymbol{\sigma_m^0}$	$\boldsymbol{\sigma_p^0}$
–CCl$_3$	0.32 (0.40)[a]	0.44 (0.46)[a]	0.10 (0.00)[b]	0.34 (0.36)[b]	0.41	0.66	0.35 (0.31)[a]	0.45 (0.33)[a]
–NHCHO	0.14 (0.19)[a]	-0.22 (0.00)[d]	-0.33 (-0.40)[b]	0.29 (0.33)[b]	-0.55	0.04	0.15 (0.19)[a]	0.19 (0.09)[a]
–NHCONH$_2$	0.13 (-0.03)[a]	-0.22 (-0.24)[c]	-0.55 (-0.47)[b]	0.32 (0.23)[b]	-0.84	-0.17	0.13	0.04

[a] Ref. 11; [b] Ref. 78; [c] Ref. 16; [d] Ref. 7.

The values of σ_R were calculated with Equation (2) because we observed a dependence between this constant and both the B_AX systems. This dependence is expected, once it was proposed that σ_p has both a resonant (σ_R) and inductive (σ_I) term, the latter being in almost all situations equal to σ_m.[15,16] Combining then Equations (5a) and (5g) with Equation (2), the errors of σ_R were reduced by using only the m- or p-B_AX systems. These metrics can be seen in Table 4.

Table 4. Regression and cross-validation errors and coefficient of determinations for the Equation (5h). The number of molecules used is represented by n.

σ	**Regressor**	**MSE**	**RMSE**	R^2	**MSE**$_{LOO}$	**RMSE**$_{LOO}$	R^2_{LOO}	n
σ_R	LLIC	0.006	0.074	0.921	0.006	0.077	0.916	79

The values calculated by our ML-based method (Table 2) were then compared with the experimental values of σ collected in Table S1. The correlation plots comparing the predicted and theoretical σ values for each ML algorithm are in Figure 3. The predictions of σ constants are in very good agreement with the experimental constants as can be seen by the almost linear plots in Figure 3, especially for the LLIC algorithm. The outliers in the plots are probably due to experimental errors in the measurement of Hammett's constants. They can propagate in the calculations or measurements of σ_R constant, for instance, since it depends on the accuracy of the determinations of both σ_p and σ_I (likewise σ_m) for computing $\sigma_R = \sigma_p - \sigma_I$)[8,15,16]. However, considering that CAM-B3LYP functional using in the DFT computation of the atomic charges describe accurately charge-transfer (CT) effects in a molecule,[63] (hyper)polarizabilities,[79,80] and spatial molecular orbitals overlaps,[81] we expect that this approach is quite suitable for accurately describing resonance and inductive effects in benzoic acid systems. Therefore, this accurate dataset provides a sound physical basis for using the ML approach for predicting Hammett's parameters, as our results show.

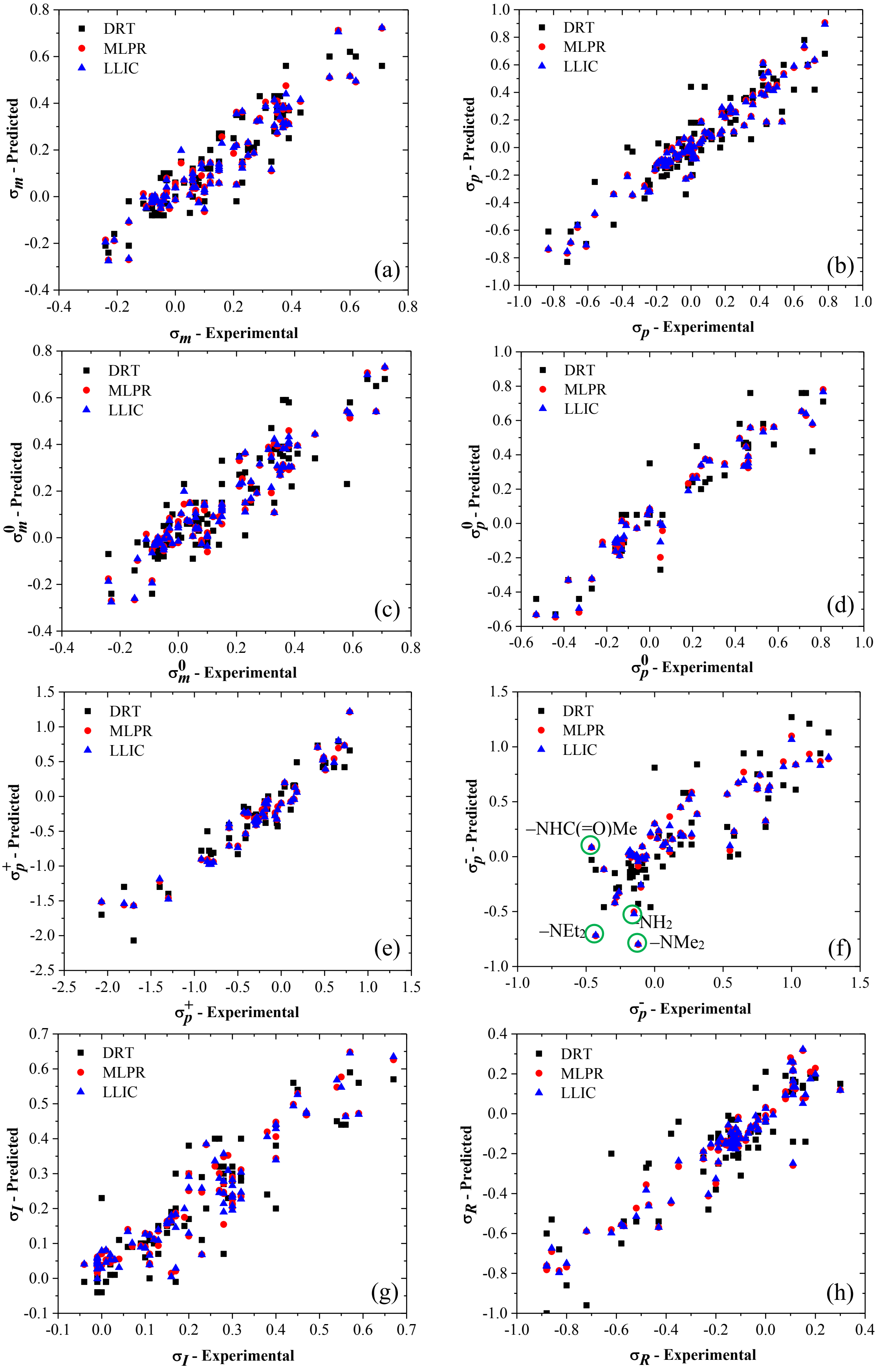

Figure 3. Comparisons between predicted (Table 2) and experimental (Table S1) σ values of all 90 substituent groups calculated from the *meta*- and *para*-substituted benzoic acid

derivatives using the Hirshfeld atomic charge method. Major outliers with the LLIC algorithm are surrounded by a green circle.

The worst result of all sets is of Hammett's constants σ_p^-, with $R_{LOO}^2 = 0.673$ (DTR), 0.742 (LLIC), and 0.741 (MLPR) (see Figure 3f and Table S17). Some of the substituent groups mentioned above (–NH_2, –NMe_2, –NEt_2, and –NHC(=O)Me) were the most significant outliers in this case. These substituents are electron donors, thus, seem not to be well represented by their σ_p^- values (-0.15, -0.12, -0.43, and -0.46, respectively, see Table S1), which are typical of electron acceptors. Given that Hammett's parameters are inferred from the experiment, and thus depend on the reaction conditions, described by the ρ parameter in Equation (1), their values can vary and eventually be inconsistent. Considering that the measured constants have experimental errors, it was suggested that some of Hammett's parameters should be revised, considering that the available σ data are, in some instances, contradictory.[58] Rablen and Yett[58] performed G4 Gaussian model chemistry calculations of the $\sigma_{m,p}$, $\sigma^{+,-}$, and newly defined σ_m^- constants based on a quantum chemical computation of Gibbs free energies of deprotonation of substituted benzoic acid and phenol systems. The G4 procedure is part of the Gn series[82–85] (n = 1, 2, and 3) and combines the high-level coupled-cluster singles and doubles with perturbative triple excitation (CCSD(T))[86] correlation calculation and moderate basis-sets size with lower-level calculations and larger basis-sets. The G4 model was able to reach the outstanding mean absolute error of 0.83 kcal/mol for a data set of 454 experimental energies.[87] Employing this model,[58] the substituents –NH_2, –NMe_2, and –NHC(=O)Me could have values of σ_p^- equal to -1,69, -1.02, and 0.30, respectively, rather than the accepted -0.15, -0.12, and -0.46 (Table S1). This trend was observed in this work where the first two σ_p^- values are more negative than the experimental ones (–NH_2: -0.66 and –NMe_2: -0.86) while the third constant is greater than the accepted one (–NHC(=O)Me: 0.00) – see Table 2. This result is particularly interesting because these three predicted values are close to the corresponding σ_p value reported in the literature, which would be expected for electron donating groups.[15]

Rerunning our ML procedure with σ_p^- values from Rablen and Yett[58] and disregarding the –NEt_2 group, improves the deviations to $\text{MSE}_{LOO}(\text{LLIC}) = 0.056$, $\text{RMSE}_{LOO}(\text{LLIC}) = 0.236$, and $R_{LOO}^2(\text{LLIC}) = 0.807$. Moreover, rerunning the ML algorithms without all four substituents (–NH_2, –NMe_2, –NEt_2, and –NHC(=O)Me) resulted in the following improvements: $\text{MSE}_{LOO}(\text{LLIC}) = 0.036$, $\text{RMSE}_{LOO}(\text{LLIC}) = 0.191$, and $R_{LOO}^2(\text{LLIC}) = 0.821$. These increasingly improvements suggest, like in Ref. [58], that revisions of some accepted σ values must be done. Consequently, we did to not consider the aforementioned substituent groups in our ML investigation. The result was Eq. 5d shown above.

4. Conclusion

The substituent effect is one the most important phenomena in chemistry usually rationalized using Hammett's theory. Hammett's constants (σ) are then a concise way to quantify the electronic effects of a given substituent. Since many of the most common substituents lack experimental σ values, sound theoretical approaches to predict these values are especially important.

We used machine learning (ML)-based regressions and cross-validations (CV), combined with Density Functional Theory (DFT) for computing accurate atomic charges, to obtain different types of σ values (σ_m, σ_p, σ_m^0, σ_p^0, σ_p^+, σ_p^-, σ_R, and σ_I) following an approach based on the original approach of Hammett by using *meta*/*para*-substituted benzoic acid derivatives.[1,2] Using the latter, instead of benzene derivatives, predicted the most accurate σ values.

Among the tested atomic charge methods (Mulliken, Löwdin, Hirshfeld, and ChelpG), Hirshfeld's method, and in a slightly lower degree Löwdin charges, predicted the best set of constants as compared with the experimental σ values. These atomic charge methods are known to describe accurately the molecular electron distribution density, and our results confirm that. Therefore, we found the Hirshfeld method to be the most accurate approach to calculate the different types of Hammett's constants.

From ML regressions, we obtained several equations based on atomic charge values that allowed us compute Hammett's the σ values with small deviations. The results showed a very good agreement between predicted values and the experimental data from literature. Furthermore, these equations can be used for predicting σ constants of other donor or acceptor groups only by computing their charges in benzoic systems.

When using DFT-modeled benzene derivatives instead of benzoic acid systems, σ values could also be calculated, but the results were slightly worse when compared to the use of benzoic acid systems. Therefore, in this work we provide a consistent set of values of different type of Hammett's constants and simple equations to predict them for their chemical groups.

Data availability statement

Data is available upon request.

Supporting Information availability statement

The Supporting Information (SI) contains molecular structures, cartesian coordinates, Hammett's constants, atomic charge differences, and machine learning data.

Authors contributions

Gabriel Monteiro-de-Castro – Data curation; Formal Analysis; Investigation; Methodology; Software; Validation; Visualization; Writing - original draft; Writing & editing.
Julio Cesar Duarte – Conceptualization; Data curation; Project administration; Resources; Supervision; Formal Analysis; Software; Visualization.
Itamar Borges Jr. – Conceptualization; Data curation; Formal Analysis; Funding Acquisition; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing, review & editing.

Conflicts of Interest

There are no conflicts of interest to declare.

Acknowledgments

I. B. thanks the Brazilian agencies CNPq (Grant numbers 304148/2018-0 and 409447/2018-8) and FAPERJ (Grant number E26/201.197/2021) for support. G. M. C thanks CAPES for a Ph.D. fellowship.

References

(1) Hammett, L. P. Some Relations between Reaction Rates and Equilibrium Constants. *Chem Rev* **1935**, *17*, 125–136.

(2) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Temperature and Solvent Influences. *J Chem Phys* **1937**, *4* (9), 613–617. https://doi.org/10.1063/1.1749914.

(3) Johnson, C. D. *The Hammett Equation*; Ebsworth, E. A. V., Elmore, D. T., Padley, P. J., Schofield, K., Eds.; Syndics of the Cambridge University Press, 1973.

(4) Jezuita, A.; Ejsmont, K.; Szatylowicz, H. Substituent Effects of Nitro Group in Cyclic Compounds. *Struct Chem* **2021**, *32* (1), 179–203. https://doi.org/10.1007/s11224-020-01612-x.

(5) Echegoyen, L. Modern Physical Organic Chemistry. *Journal of Physical Organic Chemistry*. Science Books 2011, p 743. https://doi.org/10.1002/poc.1909.

(6) Costa, P. F. V.; Esteves, P.; Vasconcellos, M. *Ácidos E Bases Em Química Orgânica*; Porto Alegre, 2005.

(7) Hansch, C.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem Rev* **1991**, *91* (2), 165–195. https://doi.org/10.1021/cr00002a004.

(8) *Advances in Linear Free Energy Relationships*, 1st ed.; Chapman, N. B., Shorter, J., Eds.; Springer US: Boston, MA, 1972. https://doi.org/10.1007/978-1-4615-8660-9.

(9) Brown, H. C.; Okamoto, Y. Substituent Constants for Aromatic Substitution 1-3. *J Am Chem Soc* **1957**, *79* (8), 1913–1917. https://doi.org/10.1021/ja01565a039.

(10) Carey, F. A.; Sundberg, R. J. *Advanced Organic Chemistry*, 5th ed.; Advanced Organic Chemistry; Springer US: Boston, MA, 1990. https://doi.org/10.1007/978-1-4613-9795-3.

(11) Chapman, N. B.; Shorter, J. *Correlation Analysis in Chemistry*; Chapman, N. B., Shorter, J., Eds.; Springer US: Boston, MA, 1978. https://doi.org/10.1007/978-1-4615-8831-3.

(12) Brown, H. C.; Okamoto, Y. Electrophilic Substituent Constants. *J Am Chem Soc* **1958**, *80* (18), 4979–4987. https://doi.org/10.1021/ja01551a055.

(13) Yukawa, Y.; Tsuno, Y. Resonance Effect in Hammett Relationship. II. Sigma Constants in Electrophilic Reactions and Their Intercorrelation. *Bull Chem Soc Jpn* **1959**, *32* (9), 965–971. https://doi.org/10.1246/bcsj.32.965.

(14) Yukawa, Y.; Tsuno, Y. Resonance Effect in Hammett Relationship. III. The Modified Hammett Relationship for Electrophilic Reactions. *Bull Chem Soc Jpn* **1959**, *32* (9), 971–981. https://doi.org/10.1246/bcsj.32.971.

(15) Roberts, J. D.; Moreland, W. T. Electrical Effects of Substituent Groups in Saturated Systems. Reactivities of 4-Substituted Bicyclo [2.2.2]Octane-1-

Carboxylic Acids 1. *J Am Chem Soc* **1953**, *75* (9), 2167–2173. https://doi.org/10.1021/ja01105a045.

(16) Hansch, C.; Leo, A. *Exploring QSAR.: Fundamentals and Applications in Chemistry and Biology*; Hansch, C., Leo, A., Eds.; American Chemical Society, 1995.

(17) van Bekkum, H.; Verkade, P. E.; Wepster, B. M. A Simple Re-evaluation of the Hammett Pσ Relation. *Recueil des Travaux Chimiques des Pays-Bas* **1959**, *78* (10), 815–850. https://doi.org/10.1002/recl.19590781009.

(18) Porobić, S. J.; Božić, B. Đ.; Dramićanin, M. D.; Vitnik, V.; Vitnik, Ž.; Marinović-Cincović, M.; Mijin, D. Ž. Absorption and Fluorescence Spectral Properties of Azo Dyes Based on 3-Amido-6-Hydroxy-4-Methyl-2-Pyridone: Solvent and Substituent Effects. *Dyes and Pigments* **2020**, *175*, 108139. https://doi.org/10.1016/j.dyepig.2019.108139.

(19) Rachuru, S.; Skelton, A. A.; Vandanapu, J. Application of Hammett Equation to Hydrogen Bond Interactions of Benzoic Acid in Chloroform/Water System and Explanation for Non-Linear Hammett Relation to Partition Coefficients for the Same System. *Comput Theor Chem* **2020**, *1190*. https://doi.org/10.1016/j.comptc.2020.113024.

(20) Teixeira, R. I.; da Silva, R. B.; Gaspar, C. S.; de Lucas, N. C.; Garden, S. J. Photophysical Properties of Fluorescent 2-(Phenylamino)-1,10-Phenanthroline Derivatives†. *Photochem Photobiol* **2021**, *97* (1), 47–60. https://doi.org/10.1111/php.13303.

(21) Nikolova, V.; Cheshmedzhieva, Di.; Ilieva, S.; Galabov, B. Atomic Charges in Describing Properties of Aromatic Molecules. *J Org Chem* **2019**, *84* (4), 1908–1915. https://doi.org/10.1021/acs.joc.8b02908.

(22) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J Chem Phys* **1955**, *23* (10), 1833–1840. https://doi.org/10.1063/1.1740588.

(23) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *J Chem Phys* **1985**, *83* (2), 735–746. https://doi.org/10.1063/1.449486.

(24) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Intermolecular Interactions from a Natural Bond Orbital, Donor-Acceptor Viewpoint*. https://pubs.acs.org/sharingguidelines.

(25) Montgomery, J. A.; 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* (15), 6532–6542. https://doi.org/10.1063/1.481224.

(26) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J Comput Chem* **1990**, *11* (4), 431–439. https://doi.org/10.1002/jcc.540110404.

(27) 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* (3), 361–373. https://doi.org/10.1002/jcc.540110311.

(28) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor Chim Acta* **1977**, *44* (2), 129–138. https://doi.org/10.1007/BF00549096.

(29) 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* (2), 527–541. https://doi.org/10.1021/ct200866d.

(30) Coleone, A. P.; Lascane, L. G.; Batagin-Neto, A. Polypyrrole Derivatives for Optoelectronic Applications: A DFT Study on the Influence of Side Groups. *Physical Chemistry Chemical Physics* **2019**, *21* (32), 17729–17739. https://doi.org/10.1039/c9cp02638j.

(31) Ertl, P. A Web Tool for Calculating Substituent Descriptors Compatible with Hammett Sigma Constants**. *Chemistry–Methods* **2022**, *202200041*, 1–7. https://doi.org/10.1002/cmtd.202200041.

(32) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer New York: New York, NY, 2009. https://doi.org/10.1007/978-0-387-84858-7.

(33) 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* (3), 1652–1671. https://doi.org/10.1021/acs.jctc.8b01176.

(34) Chang, A. M.; Freeze, J. G.; Batista, V. S. Hammett Neural Networks: Prediction of Frontier Orbital Energies of Tungsten–Benzylidyne Photoredox Complexes. *Chem Sci* **2019**, *10* (28), 6844–6854. https://doi.org/10.1039/C9SC02339A.

(35) Bragato, M.; von Rudorff, G. F.; von Lilienfeld, O. A. Data Enhanced Hammett-Equation: Reaction Barriers in Chemical Space. *Chem Sci* **2020**, *11* (43), 11859–11868. https://doi.org/10.1039/d0sc04235h.

(36) Ertl, P. Simple Quantum Chemical Parameters as an Alternative to the Hammett Sigma Constants in QSAR Studies. *Quantitative Structure-Activity Relationships* **1997**, *16* (5), 377–382. https://doi.org/10.1002/qsar.19970160505.

(37) Beker, W.; Gajewska, E. P.; Badowski, T.; Grzybowski, B. A. Prediction of Major Regio-, Site-, and Diastereoisomers in Diels–Alder Reactions by Using Machine-Learning: The Importance of Physically Meaningful Descriptors. *Angewandte Chemie - International Edition* **2019**, *58* (14), 4515–4519. https://doi.org/10.1002/anie.201806920.

(38) Simón-Vidal, L.; García-Calvo, O.; Oteo, U.; Arrasate, S.; Lete, E.; Sotomayor, N.; González-Díaz, H. Perturbation-Theory and Machine Learning (PTML) Model for High-Throughput Screening of Parham Reactions: Experimental and Theoretical Studies. *J Chem Inf Model* **2018**, *58* (7), 1384–1396. https://doi.org/10.1021/acs.jcim.8b00286.

(39) Lewis-Atwell, T.; Townsend, P. A.; Grayson, M. N. Machine Learning Activation Energies of Chemical Reactions. *WIREs Computational Molecular Science* **2022**, *12* (4). https://doi.org/10.1002/wcms.1593.

(40) Lindsey, R. K.; Huy Pham, C.; Goldman, N.; Bastea, S.; Fried, L. E. Machine-Learning a Solution for Reactive Atomistic Simulations of Energetic Materials. *Propellants, Explosives, Pyrotechnics* **2022**, *47* (8). https://doi.org/10.1002/prep.202200001.

(41) Juan, Y.; Dai, Y.; Yang, Y.; Zhang, J. Accelerating Materials Discovery Using Machine Learning. *Journal of Materials Science and Technology*. Chinese Society of Metals July 20, 2021, pp 178–190. https://doi.org/10.1016/j.jmst.2020.12.010.

(42) Casey, A. D.; Son, S. F.; Bilionis, I.; Barnes, B. C. Prediction of Energetic Material Properties from Electronic Structure Using 3D Convolutional Neural

Networks. *J Chem Inf Model* **2020**, *60* (10), 4457–4473. https://doi.org/10.1021/acs.jcim.0c00259.

(43) Tian, X. lan; Song, S. wei; Chen, F.; Qi, X. juan; Wang, Y.; Zhang, Q. hua. Machine Learning-Guided Property Prediction of Energetic Materials: Recent Advances, Challenges, and Perspectives. *Energetic Materials Frontiers*. KeAi Communications Co. September 1, 2022, pp 177–186. https://doi.org/10.1016/j.enmf.2022.07.005.

(44) Zhao, Z.; Geng, Y.; Troisi, A.; Ma, H. Performance Prediction and Experimental Optimization Assisted by Machine Learning for Organic Photovoltaics. *Advanced Intelligent Systems* **2022**, *4* (6), 2100261. https://doi.org/10.1002/aisy.202100261.

(45) Elton, D. C.; Boukouvalas, Z.; Butrico, M. S.; Fuge, M. D.; Chung, P. W. Applying Machine Learning Techniques to Predict the Properties of Energetic Materials. *Sci Rep* **2018**, *8* (1), 9059. https://doi.org/10.1038/s41598-018-27344-x.

(46) Duarte, J. C.; Dias Da Rocha, R.; Borges, I. Which Molecular Properties Determine the Impact Sensitivity of an Explosive? A Machine Learning Quantitative Investigation of Nitroaromatic Explosives. *Journal: Physical Chemistry Chemical Physics*. https://doi.org/10.1039/D2CP05339J.

(47) Löwdin, P. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J Chem Phys* **1950**, *18* (3), 365–375. https://doi.org/10.1063/1.1747632.

(48) Löwdin, P. Approximate Formulas for Many-Center Integrals in the Theory of Molecules and Crystals. *J Chem Phys* **1953**, *21* (2), 374–375. https://doi.org/10.1063/1.1698901.

(49) Botelho, F.; Oliveira, R.; Almeida, J.; França, T.; Borges, I. Comparação Entre Métodos Para Determinação de Cargas Atômicas Em Sistemas Moleculares: A Molécula N-{N-(Pterina-7-Il)Carbonilglicil}-L-Tirosina (NNPT). *Quim Nova* **2020**, *44* (2), 161–171. https://doi.org/10.21577/0100-4042.20170683.

(50) Wang, B.; Li, S. L.; Truhlar, D. G. Modeling the Partial Atomic Charges in Inorganometallic Molecules and Solids and Charge Redistribution in Lithium-Ion Cathodes. *J Chem Theory Comput* **2014**, *10* (12), 5640–5650. https://doi.org/10.1021/ct500790p.

(51) Voityuk, A. A.; Stasyuk, A. J.; Vyboishchikov, S. F. A Simple Model for Calculating Atomic Charges in Molecules. *Physical Chemistry Chemical Physics* **2018**, *20* (36), 23328–23337. https://doi.org/10.1039/C8CP03764G.

(52) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical Analysis and Extension of the Hirshfeld Atoms in Molecules. *J Chem Phys* **2007**, *126* (14), 144111. https://doi.org/10.1063/1.2715563.

(53) Chirlian, L. E.; Francl, M. M. Atomic Charges Derived from Electrostatic Potentials: A Detailed Study. *J Comput Chem* **1987**, *8* (6), 894–905. https://doi.org/10.1002/jcc.540080616.

(54) Cox, S. R.; Williams, D. E. Representation of the Molecular Electrostatic Potential by a Net Atomic Charge Model. *J Comput Chem* **1981**, *2* (3), 304–323. https://doi.org/https://doi.org/10.1002/jcc.540020312.

(55) Neese, F. Software Update: The ORCA Program System—Version 5.0. *WIREs Computational Molecular Science* **2022**, *12* (5), 1–15. https://doi.org/10.1002/wcms.1606.

(56) Coleone, A. P.; Lascane, L. G.; Batagin-Neto, A. Polypyrrole Derivatives for Optoelectronic Applications: A DFT Study on the Influence of Side Groups.

Physical Chemistry Chemical Physics **2019**, *21* (32), 17729–17739. https://doi.org/10.1039/C9CP02638J.

(57) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science Books, 2005.

(58) Yett, A.; Rablen, P. R. A G4 Approach to Computing the Hammett Substituent Constants σ_{p} , σ_{m} , σ^{-} , σ^{+} , and $\sigma^{+}{}_{\mathrm{m}}$. *J Phys Org Chem* **2022**. https://doi.org/10.1002/poc.4436.

(59) Gardner Swain, C.; Lupton, E. C. Field and Resonance Components of Substituent Effects. *J Am Chem Soc* **1968**, *90* (16), 4328–4337. https://doi.org/10.1021/ja01018a024.

(60) 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* (45), 11623–11627. https://doi.org/10.1021/j100096a001.

(61) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Physical Chemistry Chemical Physics* **2005**, *7* (18), 3297. https://doi.org/10.1039/b508541a.

(62) Tirado-Rives, J.; Jorgensen, W. L. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *J Chem Theory Comput* **2008**, *4* (2), 297–306. https://doi.org/10.1021/ct700248k.

(63) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem Phys Lett* **2004**, *393* (1–3), 51–57. https://doi.org/10.1016/j.cplett.2004.06.011.

(64) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(65) Géron, A. *Hands-on Machine Learning Whith Scikit-Learing, Keras and Tensorfow*; 2019.

(66) Kingsford, C.; Salzberg, S. L. What Are Decision Trees? *Nat Biotechnol* **2008**, *26* (9), 1011–1013. https://doi.org/10.1038/nbt0908-1011.

(67) Xu, M.; Watanachaturaporn, P.; Varshney, P. K.; Arora, M. K. Decision Tree Regression for Soft Classification of Remote Sensing Data. *Remote Sens Environ* **2005**, *97* (3), 322–336. https://doi.org/10.1016/j.rse.2005.05.008.

(68) Luo, H.; Cheng, F.; Yu, H.; Yi, Y. SDTR: Soft Decision Tree Regressor for Tabular Data. *IEEE Access* **2021**, *9*, 55999–56011. https://doi.org/10.1109/ACCESS.2021.3070575.

(69) Frosst, N.; Hinton, G. Distilling a Neural Network Into a Soft Decision Tree. *CEUR Workshop Proc* **2017**, *2071*. https://doi.org/https://doi.org/10.48550/arXiv.1711.09784.

(70) Wu, L.; Yang, Y.; Liu, H. Nonnegative-Lasso and Application in Index Tracking. *Comput Stat Data Anal* **2014**, *70*, 116–126. https://doi.org/10.1016/j.csda.2013.08.012.

(71) Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **1996**, *58* (1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

(72) Zou, H.; Hastie, T.; Tibshirani, R. On the “Degrees of Freedom” of the Lasso. *The Annals of Statistics* **2007**, *35* (5), 2173–2192. https://doi.org/10.1214/009053607000000127.

(73) Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R. Least Angle Regression. *The Annals of Statistics* **2004**, *32* (2), 407–499. https://doi.org/10.1214/009053604000000067.

(74) Zhao, P.; Yu, B. On Model Selection Consistency of Lasso. *Journal of Machine Learning Research* **2006**, *7*, 2541–2563.

(75) Dutt, M. I.; Saadeh, W. A Multilayer Perceptron (MLP) Regressor Network for Monitoring the Depth of Anesthesia. In *2022 20th IEEE Interregional NEWCAS Conference (NEWCAS)*; IEEE, 2022; pp 251–255. https://doi.org/10.1109/NEWCAS52662.2022.9842242.

(76) Zhixin, S.; Bingqing, L. Research of Improved Back-Propagation Neural Network Algorithm1. In *International Conference on Communication Technology Proceedings, ICCT*; IEEE, 2010; pp 763–766. https://doi.org/10.1109/ICCT.2010.5688628.

(77) Refaeilzadeh, P.; Tang, L.; Liu, H. Cross-Validation. In *Encyclopedia of Database Systems*; Springer US: Boston, MA, 2009; pp 532–538. https://doi.org/10.1007/978-0-387-39940-9_565.

(78) Charton, M. Electrical Effect Substituent Constants for Correlation Analysis. In *Journal of Theoretical Biology*; 1981; Vol. 91, pp 119–251. https://doi.org/10.1002/9780470171929.ch3.

(79) Srivastava, R.; Al-Omary, F. A. M.; El-Emam, A. A.; Pathak, S. K.; Karabacak, M.; Narayan, V.; Chand, S.; Prasad, O.; Sinha, L. A Combined Experimental and Theoretical DFT (B3LYP, CAM-B3LYP and M06-2X) Study on Electronic Structure, Hydrogen Bonding, Solvent Effects and Spectral Features of Methyl 1H-Indol-5-Carboxylate. *J Mol Struct* **2017**, *1137* (2), 725–741. https://doi.org/10.1016/j.molstruc.2017.02.084.

(80) Limacher, P. A.; Mikkelsen, K. V.; Lüthi, H. P. On the Accurate Calculation of Polarizabilities and Second Hyperpolarizabilities of Polyacetylene Oligomer Chains Using the CAM-B3LYP Density Functional. *J Chem Phys* **2009**, *130* (19), 194114. https://doi.org/10.1063/1.3139023.

(81) Komjáti, B.; Urai, Á.; Hosztafi, S.; Kökösi, J.; Kováts, B.; Nagy, J.; Horváth, P. Systematic Study on the TD-DFT Calculated Electronic Circular Dichroism Spectra of Chiral Aromatic Nitro Compounds: A Comparison of B3LYP and CAM-B3LYP. *Spectrochim Acta A Mol Biomol Spectrosc* **2016**, *155*, 95–102. https://doi.org/10.1016/j.saa.2015.11.002.

(82) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 Theory: A General Procedure for Prediction of Molecular Energies. *J Chem Phys* **1989**, *90* (10), 5622–5629. https://doi.org/10.1063/1.456415.

(83) Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. Gaussian-1 Theory of Molecular Energies for Second-row Compounds. *J Chem Phys* **1990**, *93* (4), 2537–2545. https://doi.org/10.1063/1.458892.

(84) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 Theory for Molecular Energies of First- and Second-row Compounds. *J Chem Phys* **1991**, *94* (11), 7221–7230. https://doi.org/10.1063/1.460205.

(85) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) Theory for Molecules Containing First and Second-Row Atoms. *J Chem Phys* **1998**, *109* (18), 7764–7776. https://doi.org/10.1063/1.477422.

(86) Harding, M. E.; Metzroth, T.; Gauss, J.; Auer, A. A. Parallel Calculation of CCSD and CCSD(T) Analytic First and Second Derivatives. *J Chem Theory Comput* **2008**, *4* (1), 64–74. https://doi.org/10.1021/ct700152c.

(87) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory. *J Chem Phys* **2007**, *126* (8), 084108. https://doi.org/10.1063/1.2436888.