ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Effect of Molecular Structure on the B3LYP-Computed HOMO−LUMO Gap: A Structure −Property Relationship Using Atomic Signatures

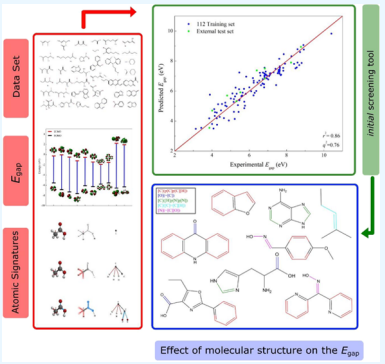
Ahmed Mohamed,* Donald P. Visco, Jr., Karl Breimaier, and David M. Bastidas

Cite This: *ACS Omega* 2025, 10, 2799−2808

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Compounds possessing a small highest occupied molecular orbital−lowest unoccupied molecular orbital (HOMO−LUMO) gap (E_{gap}) are highly desirable due to their instability and reactivity, making them useful for a wide range of applications. However, the search for new organic compounds with a low E_{gap} is an expensive endeavor due to the exponentially increasing pool of virtual compounds. Accordingly, in this study, atomic Signatures were utilized as molecular descriptors to investigate the correlation between the molecular structure and the B3LYP-computed E_{gap}, thus aiding in the development of a quantitative structure−property relationship (QSPR). An easy-to-use robust model was constructed using forward-stepping multilinear regression with leave-one-out cross validation, resulting in a regression coefficient (r^2) of 0.86 and a predictability (q^2) of 0.76. The use of atomic Signatures as molecular descriptors successfully inferred correlations between different structural motifs and E_{gap}. The atomic fragments containing π-bonds in various aromatic compounds were found to be the most significant atomic Signatures, explaining nearly 50% of the variance in the data, with regression coefficients that decreased E_{gap}. This is attributed to π-electron delocalization, making this molecular fragment a reactive site in a molecule. Finally, an external test set was used to further evaluate the model's predictive performance. The developed QSPR can be utilized as a reliable initial screening tool to identify potential candidates possessing low E_{gap} values.



INTRODUCTION

The highest occupied molecular orbital−lowest unoccupied molecular orbital (HOMO−LUMO) gap (E_{gap}) is a quantum chemical property defined as the difference between the energy of the highest occupied and lowest unoccupied molecular orbitals, HOMO and LUMO, respectively. In recent years, this property has been found to be instrumental in different chemical research areas, where, in most cases, researchers are trying to alter the configuration of organic molecules, either experimentally or computationally, in order to minimize this property.[1−4] This is desirable as molecules with low E_{gap} are unstable (i.e., highly reactive), making E_{gap} important in the design and search for new molecules with desired optoelectronic and electron-transfer properties (properties used in the development of organic photovoltaic cells,[5−7] organic film transistors,[8,9] and organic light-emitting diodes).[2,10−14] Photovoltaic cells and organic semiconductors with low E_{gap} are more efficient as they are easily excited by absorbed light, producing electricity by the excitation of electrons from the HOMO to LUMO orbitals.[15,16] Similarly, molecules with lower E_{gap} are highly reactive, increasing their tendency to participate in chemical reactions with their surrounding environment. This phenomenon is critical for the adsorption of chemical compounds onto a substrate, which finds diverse applications in both the fields of catalysis and corrosion inhibition. Specifically, in catalysis, the E_{gap} of reactants can influence the adsorption process and, therefore, the rate and selectivity of the reaction.[17] Additionally, in corrosion inhibition research, it was concluded that E_{gap} is a critical property in designing efficient corrosion inhibitors.[18−21] The reactivity and kinetic instability of inhibitors with low E_{gap} increase their affinity to adsorb and complex with the metal substrate, protecting the surface from the corrosive environment. Furthermore, the E_{gap} has relevant applications in successfully correlating toxicological end points, chemical reactions, spectroscopic data, and biological activities.[22−25] Accordingly, E_{gap} is a critical intrinsic quantum chemical property that is utilized in different industries to make design decisions regarding the development of new organic molecules to optimize a property of interest.

The E_{gap} can be calculated computationally, often using density functional theory (DFT) for the electrical band gap or time-dependent DFT (TD-DFT) for optical properties.[26−29]

Received: September 19, 2024
Revised: December 27, 2024
Accepted: January 3, 2025
Published: January 15, 2025

In this study, the focus is on the HOMO−LUMO gap, which better approximates the electrical band gap.[29] Regardless of the calculation method, the search for new organic compounds with a low E_{gap} is an expensive endeavor, both computationally and experimentally, due to the exponentially increasing pool of virtual compounds. Hence, a computer-assisted method of approximating the E_{gap} can find value as an early stage filter guiding the researcher to a data set of molecules with relevant properties. After that, a more expensive method can be utilized to refine those selections that emerge from this early stage filter.

One way to implement this filtering step is by constructing a quantitative structure−property relationship (QSPR), which creates a statistically optimized regression relation between a property of interest and a set of structurally related descriptors. Different works have been published in the literature to construct an effective and computationally inexpensive QSPR to predict the relationship between the E_{gap} and various descriptors. For example, Xu et al. utilized the smooth overlap atomic position (SOAP) method to provide descriptors, encoding regions of atomic geometries by using local expansion of a Gaussian atomic density based on harmonic and radial basis functions.[30] The data set was comprised of 323 polycyclic aromatic hydrocarbons with E_{gap} values ranging from 0.64 to 6.59 eV.[30] A model was constructed using machine learning (ML) to understand the impact of different functional groups on the E_{gap}; the mean absolute error (MAE) was reported to be 0.19 eV.[30] It was concluded that the inclusion of a ketone group in a structure had the greatest impact on the reduction of the E_{gap}.[30] Another study was conducted by Pereira et al., who utilized different ML methods to precite the HOMO, LUMO, and E_{gap} of 111,725 organic structures (88,537 for training, 9989 for testing, and 13,199 for validation) using a hybrid B3YLP functional with 6-31G* basis set.[31] Different descriptors were tested, and the best results were achieved by using modified distances descriptors with random forest models, achieving an r^2, MAE, and root-mean-square error (RMSE) of 0.88, 0.23, and 0.33 eV, respectively. It was found that the inclusion of the orbital energy calculated using PM7 (a semiempirical method for continuous potential energy surfaces) as an additional descriptor was able to enhance the model's performance.[31]

Furthermore, Montavon et al. constructed multiple QSPRs using a deep learning artificial neural network on thousands of organic molecules to predict various electronic properties including the HOMO and LUMO.[32] The descriptors used were molecular representations derived from stoichiometry and configurational information consisting of nuclear charges and three-dimensional (3D) interatomic distances represented in a matrix.[32] The constructed QSPRs were considered valid as the authors had achieved a MAE of 0.16 and 0.13 eV for the HOMO and LUMO, respectively.[32] Additionally, Mazouin et al. recently used quantum machine learning to accurately predict the E_{gap} using the QM9 and QM7b data sets, which consist of thousands of organic compounds containing up to 7 heavy atoms.[33] It was concluded that earlier classification based on structural features prior to model training enhanced the model accuracy.[33] To achieve this, three organic classifications were introduced: (1) molecules with aromatic rings or carbonyl groups, (2) molecules with single unsaturated bonds, and (3) molecules with saturated bonds.[33] Once this partitioning was applied, a kernel ridge regression was used to construct the model, achieving a 0.1 eV MAE.[33]

As identified above in some previous studies, the classification of organic compounds based on structural features was able to significantly enhance the model accuracy, illustrating the impact of the molecular structure on the E_{gap}. Mozum et al. further illustrated this relationship by demonstrating that three compounds (cyclohexanol, cyclohex-2-ethanol, and phenol) with identical skeletal structures but different bond saturation resulted in a significant change in E_{gap}.[33] This latter phenomenon is hard to quantify since the descriptors used in the literature, specifically the previously mentioned studies, do not directly articulate specific structural features, motifs, and bonding in a simple, easy-to-understand manner.

As a way to address this shortcoming, we propose a new QSPR study for E_{gap} that will be constructed using a computationally inexpensive fragmental structural descriptor called Signature, a descriptor not utilized in this area of research before.[34] By utilizing Signatures, atomic features will be encoded, helping to illuminate the impact of specific atomic fragments on the E_{gap}.

The Signature molecular descriptor was first introduced by Faulon for structural elucidation studies[35] but eventually found utility in computer-aided molecular design,[36] which continues to today.[37] A Signature is defined as a systematic codification system over an alphabet of atom types, describing the extended valence (i.e., neighborhood) of the atoms in a molecule.[36] In other words, Signatures describe the connectivity of an atom in a molecule to a predefined extent of branching, called height. A height-1 atomic Signature would describe the "root" atom's connectivity to its nearest neighbor(s), while height-2 describes the atom's connectivity to its second-nearest neighbor(s) in a molecule, etc. Summation of atomic Signatures in a structure at a specified height yields the molecular Signature of that structure at the specified height. The occurrences of atomic Signatures (typically the independent variables) in a data set can be correlated to a property of interest using a QSPR, helping to elucidate the impact of structural features.[38]

In previous work, atomic Signatures were used to identify the most significant molecular fragments affecting the adsorption of corrosion inhibitors on carbon steel in chloride-contaminated simulated concrete pore solution.[39] A data set of organic corrosion inhibitors consisting of amines, alkanolamines, and mono- and polycarboxylates were electrochemically tested to study the effect of molecular structure on corrosion inhibition.[39] It was concluded, by utilizing atomic Signatures, that the presence of π-bonds and nitrogen atoms (with lone pair of electrons) in the structure enhanced the adsorption process, positively impacting the dependent variable (i.e., pitting potential) and corroborating electrochemical tests.[39] Additionally, a QSPR model was constructed to predict the pitting potential using stepwise multilinear regression (MLR), achieving an r^2 of 0.75 and 0.87 at heights 1 and 2, respectively.[39] Similarly, Kayello et al. used atomic Signatures as descriptors to infer the importance of certain structural features on the performance of surface tension-reducing agents.[38] From the constructed QSPR (r^2 = 0.97), it was concluded that the addition of alkyl groups reduced the surface tension while an amine group increased it.[38] Accordingly, a computer-aided molecular design (CAMD) was initiated using atomic Signatures as molecular descriptors. This model then helped to identify new, novel structures with optimum surface tension-reducing capabilities.[38] Other studies

https://doi.org/10.1021/acsomega.4c08626
ACS Omega 2025, 10, 2799−2808

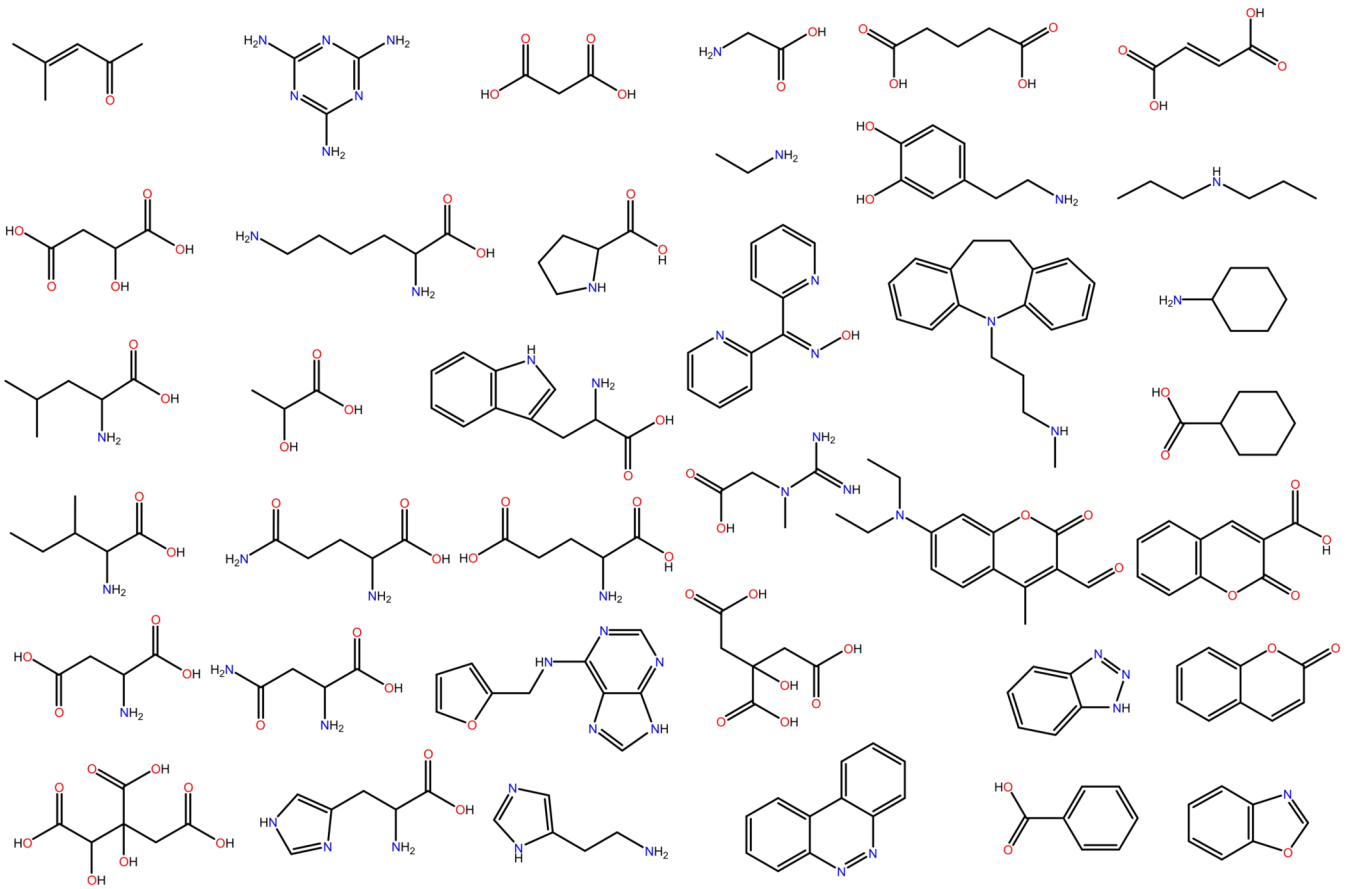


Figure 1. Subset of different organic molecules used as a training set.

exist using Signatures for CAMD, including work in the area of drug delivery,[40] pharmaceuticals,[41] solvent selection,[42] admixture design,[38] ionic liquids,[43] bio-oil additives,[37] organic photovoltaic materials,[44] and foam blowing agents.[45]

Considering the utility of the Signature molecular descriptor and the importance that not only the structure but also the nature of the bonding plays in the calculation of E_{gap}, it would seem reasonable to select Signature as a good candidate for use in designing a QSPR for E_{gap}. Accordingly, in this work, we attempt to develop an accurate, predictive, and computationally inexpensive QSPR to correlate the E_{gap} using atomic Signatures as molecular descriptors. To that end, a total of 112 organic compounds were used as a training data set composed of amines, alkanolamines, carboxylic acids, amino acids, cyclic/aromatic compounds, and hybrid structures. DFT was utilized to identify the E_{gap} of the structurally optimized organic compounds using Becke's three parameter hybrid functional and Lee−Yang− Parr correlation (B3LYP)/6-31G* basis set. Details of the method and results are provided in the next sections. We note that the influence of certain structural fragments is also discussed, which can aid in the design of organic compounds for different applications and environments.

METHODS

Data Set and DFT Calculations. A total of 112 organic compounds were chosen to construct a structurally balanced yet diverse data set consisting of amines, alkanolamines, carboxylic acids, amino acids, cyclic compounds, and "hybrid" compounds (a structure containing two or more functional groups). These functional groups were chosen for the data set as they include a range of unique molecular structures consisting of H, C, O, and N. Elements such as these are critical as they are abundant, aiding in the creation of a more inclusive and general model to be applied to a wide range of structures. A small subset of the database is shown in Figure 1; the entire data set along with the calculated E_{gap} values are available as the Supporting Information.

All quantum chemical properties presented in this study were calculated using DFT, which was performed using Gaussian 16 software. The molecule geometry was fully optimized employing Becke's three parameter hybrid functional and the Lee−Yang−Parr correlation (B3LYP) using a 6-31G* basis set. B3LYP is one of the most popularly used functionals to find different quantum properties including E_{gap}, making it a standard benchmark for comparing different and newly constructed models.[23,30,31,33,46,47] All optimizations were performed in the aqueous phase using water as a solvent. The E_{HOMO} and E_{LUMO} were then extracted from the output file using an in-house script before calculating the E_{gap} ($E_{LUMO} - E_{HOMO}$). A histogram presenting the E_{gap} distribution is illustrated in Figure 2. It should be noted that the E_{gap} distribution range from 3.13 to 10.40 eV is larger than what has been typically reported in the literature on E_{gap} studies.[28,30]

Model Training and Validation. Forward-stepping multilinear regression (FSMLR) with leave-one-out cross validation (LOOCV) was used to construct a QSPR using atomic Signatures as molecular descriptors. FSMLR is a simple yet powerful regression approach that was used successfully in previous studies to infer the relation between molecular

https://doi.org/10.1021/acsomega.4c08626
ACS Omega 2025, 10, 2799−2808

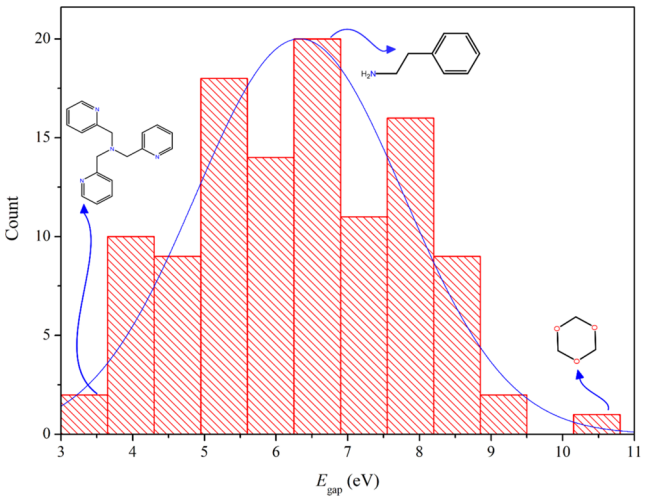


Figure 2. Histogram showing the range and frequency of E_{gap} values from the 112 data set.

fragments and a variety of properties of interest.[38,39] One of the main advantages of FSMLR is its interpretability, easily identifying significant atomic Signatures affecting, either positively or negatively, the E_{gap} through the values of the regression coefficients. The most common way to measure the model's correlative strength is through r^2, which is the ability to correlate the training set's property of interest. Although a high r^2 indicates a correlative model, this does not necessarily guarantee a predictive one. However, a QSPR capable of accurately predicting the property of interest for compounds outside of the training set is critical to the model's usefulness, especially for design. Accordingly, cross validation can be applied to gauge the predictive power of the constructed model, quantifying it through a parameter called q^2. Ultimately, the model must be able to correlate experimental data in the training set and achieve an acceptable level of prediction outside the training set.

Cross validation (CV) is a statistical technique used to evaluate a model's predictive ability through partitioning available dependent variables into different folds or subsets.[48] From this, the model will use one subset as the test set while the other subsets will be used for model training purposes. This process will be systematically repeated until all subsets have been used to train/test the model, and accordingly, the average performance of the model is recorded, and an optimum predictive model is created. LOOCV is a type of CV training technique that treats every data point as a test set, while the remaining ones are used as a training set.[49] It should be noted that this method is typically used on relatively small data sets.

To avoid overfitting, a LOOCV was utilized to assess the model's predictability in this study. All of the atomic Signatures

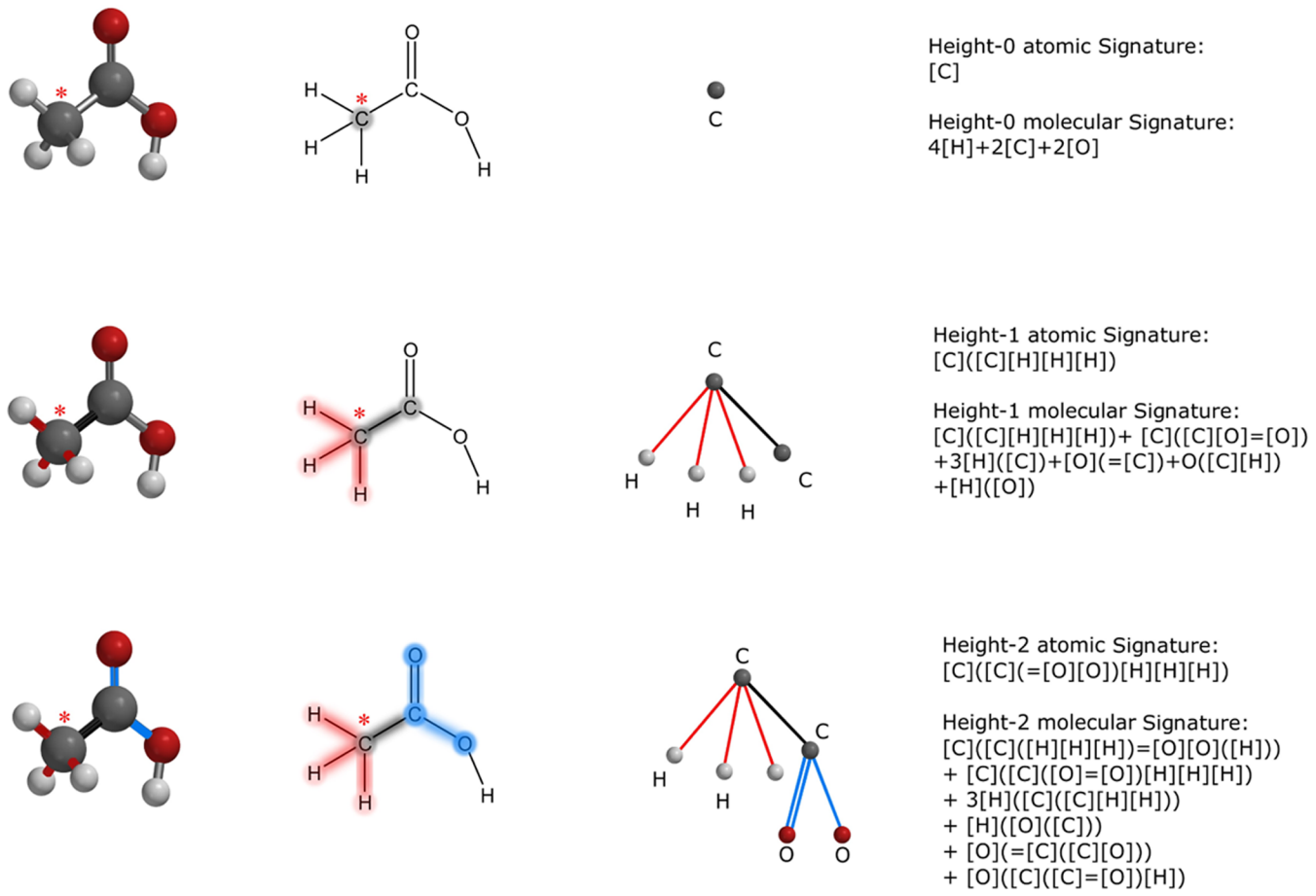


Figure 3. Construction of the molecular Signature of acetic acid at heights 0, 1, and 2. The selected root atom is the carbon indicated by a "*". Acetic acid is presented in 3D and 2D to show the branching of the root atom at the aforementioned heights. The atomic Signature at heights 0, 1, and 2 for this starred root atom are [C], [C]([C][H][H][H]), and [C]([C](=[O][O])[H][H][H]), respectively.

https://doi.org/10.1021/acsomega.4c08626
ACS Omega 2025, 10, 2799−2808

Table 1. List of 67 Height-to-One Atomic Signatures for the 112 Molecular Structures

#	height-1 signatures	#	height-1 signatures	#	height-1 signatures	#	height-1 signatures
1	[C](=[C][H][H])	21	[C]([C][O]=[O])	41	[C](p[C]p[C][N])	61	[N](p[N]p[N])
2	[C](=[C][H][N])	22	[C]([C]p[C]p[C])	42	[C](p[C]p[C][O])	62	[O](=[C])
3	[C]([C]=[C][H])	23	[C]([C]p[C]p[N])	43	[C](p[C]p[C]p[C])	63	[O]([C][C])
4	[C]([C][C]=[C])	24	[C]([C]p[C]p[O])	44	[C](p[C]p[C]p[N])	64	[O]([C][H])
5	[C]([C][C]=[N])	25	[C]([C]p[N]p[N])	45	[C](p[C]p[C]p[O])	65	[O]([C][N])
6	[C]([C][C]=[O])	26	[C]([C]p[N]p[O])	46	[C](p[C]p[N]p[N])	66	[O]([H][N])
7	[C]([C][C][C][C])	27	[C]([H][H][H][N])	47	[H]([C])	67	[O](p[C]p[C])
8	[C]([C][C][C][H])	28	[C]([H][H][H][O])	48	[H]([N])		
9	[C]([C][C][C][N])	29	[C]([H][H][O][O])	49	[H]([O])		
10	[C]([C][C][C][O])	30	[C]([H][N]=[O])	50	[N](=[C][H])		
11	[C]([C][C][H][H])	31	[C]([H]p[N]p[N])	51	[N](=[C][O])		
12	[C]([C][C][H][N])	32	[C]([H]p[N]p[O])	52	[N]([C][C][C])		
13	[C]([C][C][H][O])	33	[C]([N][N]=[N])	53	[N]([C][C][H])		
14	[C]([C][H]=[N])	34	[C]([N]p[N]p[N])	54	[N]([C][H][H])		
15	[C]([C][H]=[O])	35	[C](p[C]=[O]p[O])	55	[N]([C][H][O])		
16	[C]([C][H][H][H])	36	[C](p[C][H]p[N])	56	[N]([C]p[C]p[N])		
17	[C]([C][H][H][N])	37	[C](p[C][H]p[O])	57	[N](p[C][H]p[N])		
18	[C]([C][H][H][O])	38	[C](p[C][N]p[N])	58	[N](p[C]p[C])		
19	[C]([C][H][O][O])	39	[C](p[C]p[C]=[O])	59	[N](p[C]p[C][H])		
20	[C]([C][N]=[O])	40	[C](p[C]p[C][H])	60	[N](p[C]p[N])		

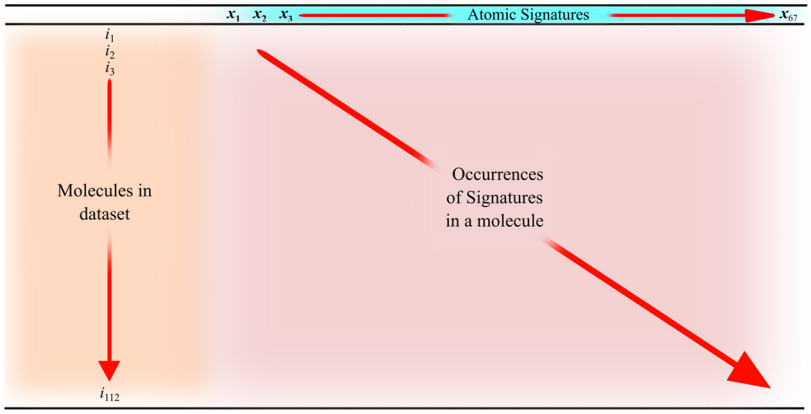


Figure 4. Schematic of the descriptor matrix used. The rows represent the 112 molecular structures, while the columns are the 67 generated atomic Signature. The occurrences of each Signature in a molecule are the values of the matrix.

used in the QSPR had an α-to-enter/remove significance level (α) of 0.1 and a p-value of less than 0.05, indicating the rejection of the null hypothesis. Furthermore, the MAE and RMSE were calculated and used as statistical metrics to assess the model's performance, as seen in eqs 1 and 2

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}\left|E\mathrm{gap}^{i}_{\mathrm{pred}} - E\mathrm{gap}^{i}_{\mathrm{DFT}}\right| \quad (1)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(E\mathrm{gap}^{i}_{\mathrm{pred}} - E\mathrm{gap}^{i}_{\mathrm{DFT}}\right)^2} \quad (2)$$

where m is the number of molecules in the data set, $E\mathrm{gap}^{i}_{\mathrm{pred}}$ is the model-predicted E_{gap} of compound i, and $E\mathrm{gap}^{i}_{\mathrm{DFT}}$ is the DFT-calculated E_{gap} of compound i.

RESULTS AND DISCUSSION

Atomic Molecular Signature. A Signature is a systematic codification system over an alphabet of atom types, describing the extended valence (i.e., neighborhood) of the atoms of a molecule.[34,36] In other words, Signature describes the connectivity of atoms in a molecule to an extent of branching up to a predefined distance called the height. The summation of atomic Signatures at a certain height in a molecule yields the molecular Signature at the predetermined height.

As an example, Figure 3 represents the construction of the molecular Signature (heights 0, 1, and 2) of acetic acid, and the root atom is indicated by a star, which is C in this specific example. As seen from Figure 3, at height = 0, the atomic Signature is simply the atom itself [C]. In contrast, at height = 1, the atomic Signature describes all atoms bonded to the root atom without backtracking. The atomic Signature for this C atom at height = 1 is [C]([C][H][H][H]), meaning that the carbon atom in acetic acid is single-bonded to a carbon and three hydrogens. The parentheses indicate how far an atom is from the root, and the bonds between atoms are assumed to be single unless otherwise specified ("=" for a double bond, "t" for a triple bond, and "p" for the aromatic bond). At height = 2, the atomic Signature becomes more specific and detailed as it describes the connectivity of the root atom to neighboring atoms two bonds away without backtracking. As a result, the atomic Signature at height = 2 is [C]([C]([=O][O])[H]-[H][H]), illustrating that there are two oxygen atoms bonded

https://doi.org/10.1021/acsomega.4c08626
ACS Omega 2025, 10, 2799−2808

(one is single-bonded, while the other is double-bonded) to the carbon atom that is attached to the root carbon atom. Summing up all of the atomic Signatures at a given height will result in the molecular Signature of the molecule, as seen in Figure 3. The number preceding each atomic Signature represents the occurrences of this atomic Signature in a structure.

QSPR Construction. An in-house script was used to deconstruct the molecular data files (MOL files) of each molecular structure in the data set into height-1 atomic Signatures. The script generated 67 unique height-1 atomic Signatures, which are presented in Table 1. These 67 atomic Signatures were used as the independent variables to correlate with the E_{gap} (dependent variable) in a QSPR. As mentioned, an FSMLR with LOOCV was employed to construct the model by selecting the most significant atomic Signatures affecting the property of interest based on the previously defined selection criteria. To illustrate the occurrence of each atomic Signature in a specific molecular structure, a descriptor matrix was constructed where the rows represent the molecules, columns represent the atomic Signatures, and the value of the matrix is the occurrence of each height-1 atomic Signature within a molecule. Figure 4 illustrates a schematic of the descriptor matrix used to construct the QSPR.

The effect of the number of atomic Signatures added to the model on explaining the variance in the data (i.e., r^2) is illustrated in Figure 5. The first atomic Signature (see Figure 5) was able to explain 48.6% of the variance, indicating it to be the most significant. With every subsequent addition, the effect decreased until a plateau was reached, beyond which adding more descriptors would lead to overfitting. Out of the 67 atomic Signatures, 18 were identified as significant according to a significance level (α) of 0.1 included in the QSPR. The selected atomic Signatures are listed in Table 2 in the order they were added (most significant to least significant), along with their corresponding regression coefficient and p-value. The p-values of the selected atomic Signatures are lower than 0.05, rejecting the null hypotheses, thus establishing their statistical significance.[50] Moreover, the lack of fit p-value was found to be 0.09, making it insignificant and, in turn, indicating that there is no strong evidence to support the lack of fit,

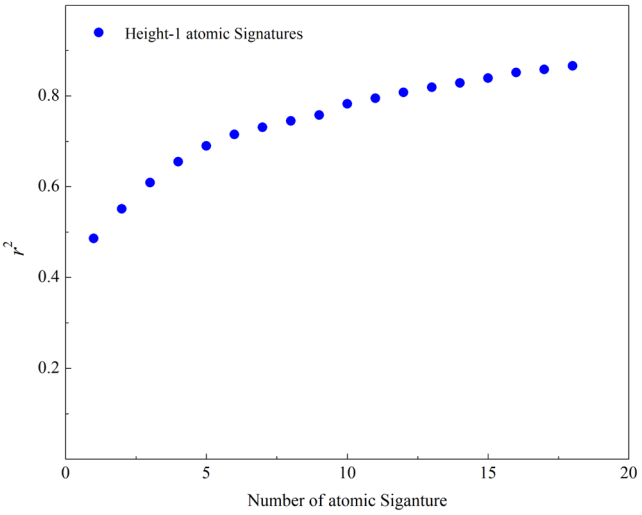


Figure 5. Regression coefficient as a function of the number of atomic Signatures added to the QSPR.

Table 2. QSPR Regression Coefficients of Selected Atomic Signatures (in Order of Decreasing Significance) along with Their Corresponding p-Value

variables	coefficient	p-value
constant	8.475	–
[C](p[C]p[C][H])	−0.2552	0.000
[O](=[C])	−0.7567	0.000
[C]([H]p[N]p[N])	−0.476	0.035
[C]([C]=[C][H])	−0.893	0.000
[N](=[C][O])	−1.345	0.000
[C]([C]p[N]p[N])	1.721	0.000
[H]([N])	−0.558	0.001
[C](p[C]p[C][O])	−1.172	0.000
[O]([C][C])	0.569	0.000
[C]([C][C][H][H])	0.2168	0.000
[N]([C][H][O])	1.326	0.005
[H]([C])	−0.0596	0.002
[O](p[C]p[C])	−0.926	0.000
[C](p[C][H]p[N])	−0.541	0.000
[N]([C][H][H])	0.809	0.020
[C](p[C]p[C]p[N])	−0.418	0.012
[C](=[C][H][H])	−0.937	0.003
[C](p[C]=[O]p[O])	−0.709	0.042

further attesting to the model's strength.[51] The achieved r^2 and q^2 are 0.86 and 0.76, respectively, making the model correlative as well as predictive. To further assess the model's statistical performance, the RMSE and MAE were calculated using eqs 1 and 2 and found to be 0.528 and 0.408 eV, respectively. The MAE and RMSE are within an acceptable range for the proposed QSPR model, which is designed to serve as an efficient initial screening tool and to establish correlations between structural motifs and E_{gap}. As highlighted in the introduction, while more accurate machine learning models, such as those by Mazouin et al. (MAE = 0.1 eV) and Pereira et al. (MAE = 0.23 eV), exist, this study focuses on providing a computationally inexpensive and interpretable model for early stage screening, providing researchers with a practical tool for identifying promising compounds before employing more resource-intensive methods.

As a way to further evaluate the predictive power of the model created, we selected a ten-compound external test set. The ten compounds were selected to be representative of the original data set, having an E_{gap} range spanning from 3.8 to 8.7 eV. This test set, along with their HOMO, LUMO, and E_{gap}, is presented in Figure 6. The constructed QSPR (see Figure 7) was able to successfully predict the E_{gap} values of the test set achieving an r^2 of 0.91, corroborating the model's predictive power. It should be noted that the QSPR is only used to establish prediction and correlation of the E_{gap} for organic compounds composed of C, H, O, and N atoms using the B3LYP functional with 6-31G* basis set in aqueous solution.

Impact of Molecular Structure. The QSPR of the E_{gap} is represented as the summation of the product of the regression coefficients and occurrence of each atomic Signature in a molecule as seen in eq 3. Although the atomic Signatures are slightly correlated, as they are not perfectly orthogonal, one can still infer the effect and impact of each atomic Signature on the E_{gap} through the magnitude and sign of the regression coefficient. For example, the regression coefficient of [C](p-[C]p[C][H]), the most impactful atomic Signature affecting the E_{gap}, is negative, which indicates that the presence of

https://doi.org/10.1021/acsomega.4c08626
ACS Omega 2025, 10, 2799−2808

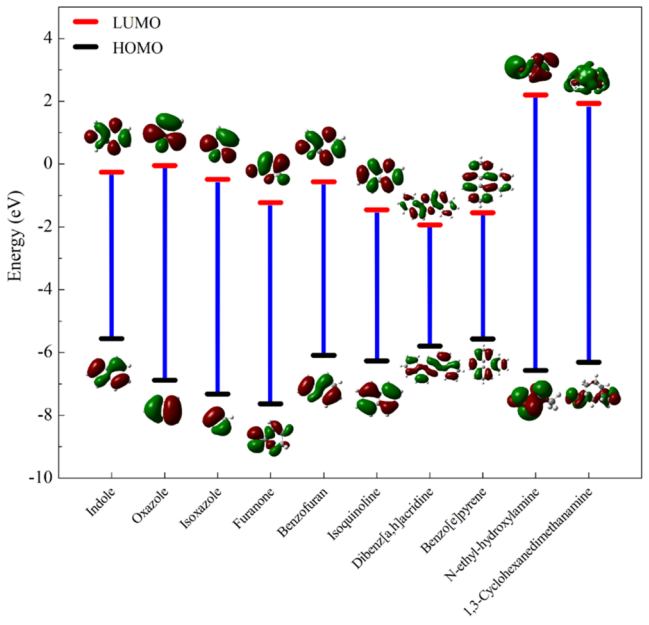


Figure 6. HOMO and LUMO energies of the test set used to validate the constructed QSPR. The difference between both molecular orbitals is the E_{gap} (blue line).

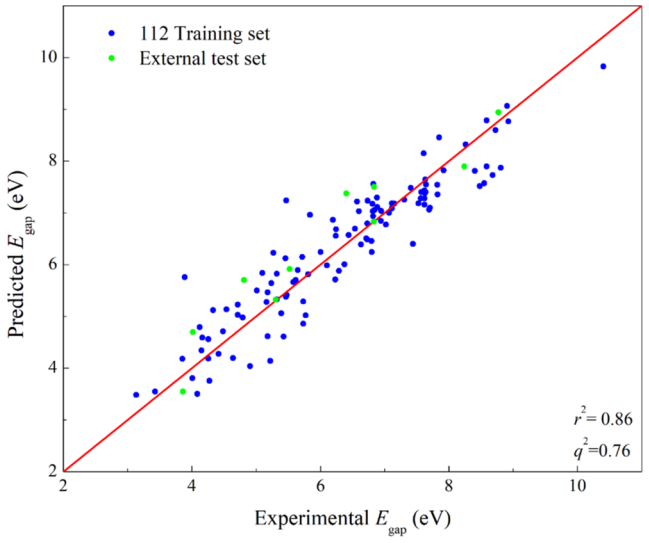


Figure 7. Predicted vs DFT-calculated E_{gap} values for the training and test sets using the height-1 constructed QSPR.

aromatic bonds in a structure can lower the HOMO−LUMO gap. This is analogous to the work done by F. Aihara, as he stated that a reduced HOMO−LUMO gap by definition is closely related to aromaticity.[2] Furthermore, the presence of [O](=[C]), [N](=[C][O]), and [C]([H]p[N]p[N]), which also have negative regression coefficients, decreases the E_{gap} value, which can be attributed to double bonds, aromatics, and heteroatoms being reactive sites in a molecule. Similarly, [C]([C]=[C][H]) had a regression coefficient of −0.893, showing its importance in reducing the E_{gap}, which is attributed to the π-conjugated nature of the atomic Signature, agreeing with the literature.[52] On the other hand, the presence of an internal alkane chain group, [C]([C][C][H][H]), and terminal nitrogen atomic Signature, [N]([C][H][H]), in a molecule increases the E_{gap} value due to a positive regression coefficient. It should be noted that the magnitude of the regression coefficients is also dependent on the number of occurrences of the atomic Signature in the descriptor matrix. For example, the regression coefficient of [H]([C]) is the smallest due to the common presence of this molecular fragment in a molecule; [H]([C]) is present 871 times in this 112-molecule data set.

$$\begin{aligned} E_{gap} = {} & -0.2552[C](p[C]p[C][H]) - 0.7567[O](=[C]) \\ & - 0.476[C]([H]p[N]p[N]) - 0.893[C]([C] \\ & = [C][H]) - 1.345[N](=[C][O]) + 1.721[C] \\ & ([C]p[N]p[N]) - 0.558[H]([N]) - 1.172[C] \\ & (p[C]p[C][O]) + 0.569[O]([C][C]) + 0.2168 \\ & [C]([C][C][H][H]) + 1.326[N]([C][H][O]) \\ & - 0.0596[H]([C]) - 0.926[O](p[C]p[C]) \\ & - 0.541[C](p[C][H]p[N]) + 0.809[N]([C][H] \\ & [H]) - 0.418[C](p[C]p[C]p[N]) - 0.937[C](= \\ & [C][H][H]) - 0.709[C](p[C]=[O]p[O]) \\ & + 8.475 \end{aligned} \quad (3)$$

According to Table 2, the five most significant atomic Signatures explain ∼70% of the data variation and negatively impact the E_{gap} value prediction. These atomic Signatures consist mainly of aromatics, heteroatoms, and double bonds, indicating that their presence in a molecule is associated with a decrease in the HOMO−LUMO gap. This is attributed to high electron density, making these structural features reactive sites, specifically due to $\pi-\pi$ interactions, lone pair electrons, and/or delocalized π-bonds. This phenomenon can be clearly seen in the example demonstrated by B. Mazouin et al., where the addition of aromatic bonds decreased the E_{gap}, as aromatic bonds can contribute to a delocalized π-electron system, thus increasing the tendency of a molecule to participate in electron transfer and be more reactive.[33] It should be noted that the effect of aromatics ([C](p[C]p[C][H])) and carbonyl groups ([O](=[C])) on the E_{gap} is analogous with other studies found in the literature.[2,30,53] Figure 8 shows the five most significant atomic Signatures color-coded to fragmentally illustrate them in different molecular structures. This illustration provides evidence of the ability of Signatures to encapsulate different structural motifs, making it an efficient and chemically interpretable descriptor, an advantage missing in other available molecular descriptors.

On the contrary, the presence of internal alkane chain groups, [C]([C][C][H][H]), in a structure causes an increase in the E_{gap} due to their nonconjugated nature. This causes the HOMO and LUMO orbitals to most likely not overlap, resulting in decreased reactivity. Additionally, as seen in eq 3, the presence of [N]([C][H][H]) in a molecular structure is positively correlated to the E_{gap}, which could be due to a steric hindrance effect that can cause a decrease in the rate reaction due to an increased HOMO LUMO gap.[54] Similarly, the increased E_{gap} observed in the presence of the [C]([C]p[N]-p[N]) atomic Signature may be attributed to the carbon atom disrupting the electron delocalization. It should be noted that the effect of atomic Signatures depends on the molecular structure studied and the position of the molecular fragment in the structure itself.

https://doi.org/10.1021/acsomega.4c08626
ACS Omega 2025, 10, 2799−2808

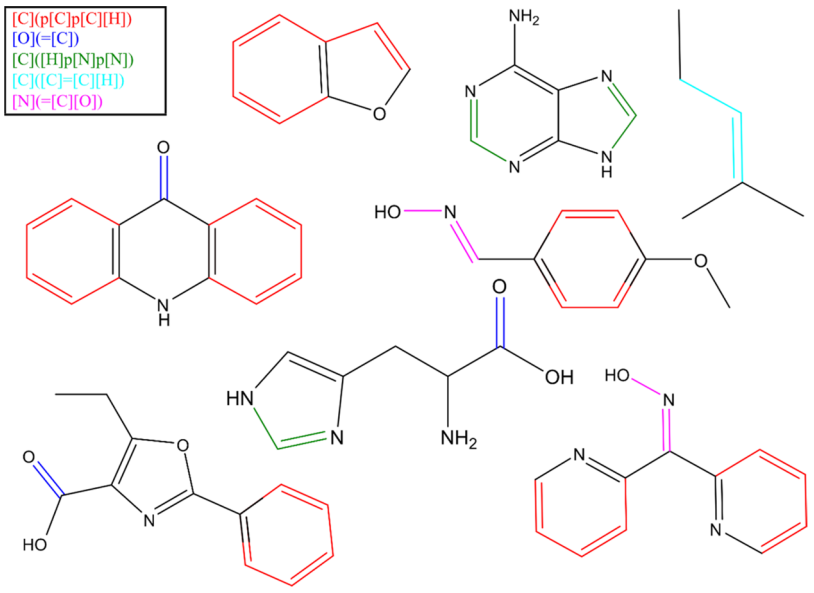


Figure 8. Five of the most significant atomic Signatures, negatively affecting the E_{gap}, are color-coded in different structures, illustrating how atomic Signatures can be precise and easily interpretable.

Benefits of Signatures as Molecular Descriptors. By utilizing atomic Signature, a robust model was constructed using a relatively small data set, where different structural fragments and motifs infer relations to the property of interest (i.e., E_{gap}). Atomic Signatures are one of the very few invertible molecular descriptors, a descriptor that can be used to directly build new chemical structures, helping in optimizing a property of interest through a CAMD. As has been shown, one of the major benefits of atomic Signatures as molecular descriptors is their ability to add physical insights into identifying key structural fragments and motifs contributing to a property. This is due to Signatures intrinsically encapsulating specific atomic bonding, thus making it a highly transferable molecular descriptor that can be applied to structurally diverse systems, facilitating the design and screening of new compounds with optimized properties of interest. Unlike molecular fingerprints that include explicit valence, formal charge, and hybridization type, Signatures provide a clear path from descriptor to structure, making them a more advantageous molecular descriptor to be utilized in correlating, screening, and designing new desired compounds.

CONCLUSIONS

By utilizing atomic Signatures as molecular descriptors, a robust, computationally inexpensive QSPR was constructed to correlate the effect of different structural motifs on the B3LYP-computed E_{gap} of organic compounds. A database was constructed of 112 different amines, alkanolamines, amino acids, carboxylic acids, cyclic compounds, and hybrids of these compounds. Sixty-seven height-1 atomic Signatures were generated and then correlated to the E_{gap} of the compounds using forward-stepping multilinear regression with LOOCV. r^2 and q^2 values of 0.86 and 0.76 were achieved, respectively, indicating a robust model. The model was able to establish that the presence of signatures encapsulating aromatics, heteroatoms, and π-bonds in a molecule was able to significantly reduce the E_{gap}. This is attributed to these structural fragments' high electron density and unpaired electrons, making them reactive sites due to $\pi-\pi$ interactions, lone pair electrons, and/or delocalized π-bonds. In contrast, the presence of terminal nitrogen or internal alkane chain groups causes an increase in the E_{gap}, which can be attributed to steric hindrance and nonconjugated bonding, respectively. To further corroborate the model's performance, the model was able to successfully predict the E_{gap} of an external test set, achieving an r^2 of 0.91.

ASSOCIATED CONTENT

Data Availability Statement

An excel file is available as Supporting Information. The first and second worksheets include the E_{gap} values and descriptor matrix of the training set, while the third and fourth includes the E_{gap} values and descriptor matrix of the external test set. An in-house script was employed to generate the atomic signatures, thereby creating the descriptor matrix. All molecules utilized for both training and testing purposes are provided within the aforementioned excel file.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c08626.

E_{gap} values for the training set (Table S1); descriptor matrix for the training set (Table S2); E_{gap} values for the test set (Table S3); descriptor matrix for the test set (Table S4) (XLSX)

AUTHOR INFORMATION

Corresponding Author

Ahmed Mohamed − *National Center for Education and Research on Corrosion and Materials Performance, NCERCAMP-UA, Dept. Chemical, Biomolecular, and Corrosion Engineering, The University of Akron, Akron, Ohio 44325-3906, United States;* orcid.org/0000-0002-3878-1346; Email: ame118@1870.uakron.edu

Authors

Donald P. Visco, Jr. − *National Center for Education and Research on Corrosion and Materials Performance, NCERCAMP-UA, Dept. Chemical, Biomolecular, and Corrosion Engineering, The University of Akron, Akron,*

https://doi.org/10.1021/acsomega.4c08626
ACS Omega 2025, 10, 2799−2808

Ohio 44325-3906, United States; orcid.org/0000-0002-8359-5825

Karl Breimaier − *National Center for Education and Research on Corrosion and Materials Performance, NCERCAMP-UA, Dept. Chemical, Biomolecular, and Corrosion Engineering, The University of Akron, Akron, Ohio 44325-3906, United States*

David M. Bastidas − *National Center for Education and Research on Corrosion and Materials Performance, NCERCAMP-UA, Dept. Chemical, Biomolecular, and Corrosion Engineering, The University of Akron, Akron, Ohio 44325-3906, United States;* orcid.org/0000-0002-8720-7500

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c08626

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge funding from Firestone Research Grant 639430 and The University of Akron Fellowship FRC−207160 and FRC-207865. The authors thank the technical support and facilities from The National Center for Education and Research on Corrosion and Materials Performance (NCERCMAP-UA), The College of Engineering and Polymer Science, and The University of Akron.

■ REFERENCES

(1) Perepichka, D. F.; Bryce, M. R. Molecules with Exceptionally Small HOMO−LUMO Gaps. *Angew. Chem., Int. Ed.* **2005**, *44* (34), 5370−5373.

(2) Aihara, J.-i. Reduced HOMO−LUMO Gap as an Index of Kinetic Stability for Polycyclic Aromatic Hydrocarbons. *J. Phys. Chem. A* **1999**, *103* (37), 7487−7495.

(3) Berlin, A.; Zotti, G.; Zecchin, S.; Schiavon, G.; Vercelli, B.; Zanelli, A. New Low-Gap Polymers from 3,4-Ethylenedioxythiophene-Bis-Substituted Electron-Poor Thiophenes. The Roles of Thiophene, Donor−Acceptor Alternation, and Copolymerization in Intrinsic Conductivity. *Chem. Mater.* **2004**, *16* (19), 3667−3676.

(4) Li, H.; Zhang, Z.; Liu, Y.; Cen, W.; Luo, X. Functional Group Effects on the HOMO−LUMO Gap of g-C3N4. *Nanomaterials* **2018**, *8* (8), No. 589.

(5) Hussain, R.; Mehboob, M. Y.; Khan, M. U.; Khalid, M.; Irshad, Z.; Fatima, R.; Anwar, A.; Nawab, S.; Adnan, M. Efficient designing of triphenylamine-based hole transport materials with outstanding photovoltaic characteristics for organic solar cells. *J. Mater. Sci.* **2021**, *56* (8), 5113−5131.

(6) Fu, Z.; Shen, W.; He, R.; Liu, X.; Sun, H.; Yin, W.; Li, M. Theoretical studies on the effect of a bithiophene bridge with different substituent groups (R = H, CH_3, OCH_3 and CN) in donor−π−acceptor copolymers for organic solar cell applications. *Phys. Chem. Chem. Phys.* **2015**, *17* (3), 2043−2053.

(7) Nabil, E.; Hasanein, A. A.; Alnoman, R. B.; Zakaria, M. Optimizing the Cosensitization Effect of SQ02 Dye on BP-2 Dye-Sensitized Solar Cells: A Computational Quantum Chemical Study. *J. Chem. Inf. Model.* **2021**, *61* (10), 5098−5116.

(8) Sui, Y.; Deng, Y.; Du, T.; Shi, Y.; Geng, Y. Design strategies of n-type conjugated polymers for organic thin-film transistors. *Mater. Chem. Front.* **2019**, *3* (10), 1932−1951.

(9) Higashino, T.; Mori, T. Small-molecule ambipolar transistors. *Phys. Chem. Chem. Phys.* **2022**, *24* (17), 9770−9806.

(10) Mucur, S. P.; Canımkurbey, B.; Kavak, P.; Akbaş, H.; Karadağ, A. Charge carrier performance of phosphazene-based ionic liquids doped hole transport layer in organic light-emitting diodes. *Appl. Phys. A* **2020**, *126* (12), No. 923.

(11) Patil, Y.; Jadhav, T.; Dhokale, B.; Misra, R. Design and Synthesis of Low HOMO−LUMO Gap N-Phenylcarbazole-Substituted Diketopyrrolopyrroles. *Asian J. Org. Chem.* **2016**, *5* (8), 1008−1014.

(12) Misra, R.; Gautam, P. Tuning of the HOMO−LUMO gap of donor-substituted symmetrical and unsymmetrical benzothiadiazoles. *Org. Biomol. Chem.* **2014**, *12* (29), 5448−5457.

(13) Sosorev, A. Y.; Nuraliev, M. K.; Feldman, E. V.; Maslennikov, D. R.; Borshchev, O. V.; Skorotetcky, M. S.; Surin, N. M.; Kazantsev, M. S.; Ponomarenko, S. A.; Paraschuk, D. Y. Impact of terminal substituents on the electronic, vibrational and optical properties of thiophene−phenylene co-oligomers. *Phys. Chem. Chem. Phys.* **2019**, *21* (22), 11578−11588.

(14) Lu, S.-Y.; Mukhopadhyay, S.; Froese, R.; Zimmerman, P. M. Virtual Screening of Hole Transport, Electron Transport, and Host Layers for Effective OLED Design. *J. Chem. Inf. Model.* **2018**, *58* (12), 2440−2449.

(15) Colladet, K.; Nicolas, M.; Goris, L.; Lutsen, L.; Vanderzande, D. Low-band gap polymers for photovoltaic applications. *Thin Solid Films* **2004**, *451*−*452*, 7−11.

(16) Zhou, H.; Yang, L.; Price, S. C.; Knight, K. J.; You, W. Enhanced Photovoltaic Performance of Low-Bandgap Polymers with Deep LUMO Levels. *Angew. Chem., Int. Ed.* **2010**, *49* (43), 7992−7995.

(17) Fang, J.; Zheng, W.; Liu, K.; Li, H.; Li, C. Molecular design and experimental study on the synergistic catalysis of cellulose into 5-hydroxymethylfurfural with Brønsted−Lewis acidic ionic liquids. *Chem. Eng. J.* **2020**, *385*, No. 123796.

(18) Obot, I. B.; Macdonald, D. D.; Gasem, Z. M. Density functional theory (DFT) as a powerful tool for designing new organic corrosion inhibitors. Part 1: An overview. *Corros. Sci.* **2015**, *99*, 1−30.

(19) Feiler, C.; Mei, D.; Vaghefinazari, B.; Würger, T.; Meißner, R. H.; Luthringer-Feyerabend, B. J. C.; Winkler, D. A.; Zheludkevich, M. L.; Lamaka, S. V. In silico screening of modulators of magnesium dissolution. *Corros. Sci.* **2020**, *163*, No. 108245.

(20) Mohamed, A.; Visco, D. P.; Bastidas, D. M. Sodium Succinate as a Corrosion Inhibitor for Carbon Steel Rebars in Simulated Concrete Pore Solution. *Molecules* **2022**, *27* (24), No. 8776.

(21) Mohamed, A.; Martin, U.; Bastidas, D. M. Adsorption and Surface Analysis of Sodium Phosphate Corrosion Inhibitor on Carbon Steel in Simulated Concrete Pore Solution. *Materials* **2022**, *15* (21), No. 7429.

(22) Chattaraj, P. K.; Giri, S.; Duley, S. Update 2 of: Electrophilicity Index. *Chem. Rev.* **2011**, *111* (2), PR43−PR75.

(23) Pereira, F.; Latino, D. A. R. S.; Aires-de-Sousa, J. Estimation of Mayr Electrophilicity with a Quantitative Structure−Property Relationship Approach Using Empirical and DFT Descriptors. *J. Org. Chem.* **2011**, *76* (22), 9312−9319.

(24) Wang, Y.; Chen, J.; Ge, L.; Wang, D.; Cai, X.; Huang, L.; Hao, C. Experimental and theoretical studies on the photoinduced acute toxicity of a series of anthraquinone derivatives towards the water flea (Daphnia magna). *Dyes Pigm.* **2009**, *83* (3), 276−280.

(25) Xiong, L.; Tang, J.; Li, Y.; Li, L. Phototoxic risk assessment on benzophenone UV filters: In vitro assessment and a theoretical model. *Toxicol. In Vitro* **2019**, *60*, 180−186.

(26) Makuła, P.; Pacia, M.; Macyk, W. How To Correctly Determine the Band Gap Energy of Modified Semiconductor Photocatalysts Based on UV−Vis Spectra. *J. Phys. Chem. Lett.* **2018**, *9* (23), 6814−6817.

(27) Viñes, F.; Lamiel-García, O.; Ko, K. C.; Lee, J. Y.; Illas, F. Systematic study of the effect of HSE functional internal parameters on the electronic structure and band gap of a representative set of metal oxides. *J. Comput. Chem.* **2017**, *38* (11), 781−789.

(28) Wang, J.; Wang, Y.; Huang, Y.; Peijnenburg, W. J. G. M.; Chen, J.; Li, X. Development of a nano-QSPR model to predict band gaps of spherical metal oxide nanoparticles. *RSC Adv.* **2019**, *9* (15), 8426−8434.

(29) Kimber, P.; Plasser, F. Energy component analysis for electronically excited states of molecules: why the lowest excited

state is not always the HOMO/LUMO transition. *J. Chem. Theory Comput.* **2023**, *19* (8), 2340−2352.
(30) Xu, Y.; Chu, Q.; Chen, D.; Fuentes, A. HOMO−LUMO Gaps and Molecular Structures of Polycyclic Aromatic Hydrocarbons in Soot Formation. *Front. Mech. Eng.* **2021**, *7*, No. 744001, DOI: 10.3389/fmech.2021.744001.
(31) Pereira, F.; Xiao, K.; Latino, D. A. R. S.; Wu, C.; Zhang, Q.; Aires-de-Sousa, J. Machine Learning Methods to Predict Density Functional Theory B3LYP Energies of HOMO and LUMO Orbitals. *J. Chem. Inf. Model.* **2017**, *57* (1), 11−21.
(32) Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.-R.; Anatole von Lilienfeld, O. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **2013**, *15* (9), No. 095003.
(33) Mazouin, B.; Schöpfer, A. A.; von Lilienfeld, O. A. Selected machine learning of HOMO−LUMO gaps with improved data-efficiency. *Mater. Adv.* **2022**, *3*, 8306−8316, DOI: 10.1039/D2MA00742H.
(34) Faulon, J.-L.; Visco, D. P.; Pophale, R. S. The Signature Molecular Descriptor. 1. Using Extended Valence Sequences in QSAR and QSPR Studies. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (3), 707−720.
(35) Faulon, J.-L. Stochastic Generator of Chemical Structure. 1. Application to the Structure Elucidation of Large Molecules. *J. Chem. Inf. Comput. Sci.* **1994**, *34* (5), 1204−1218.
(36) Faulon, J.-L.; Churchwell, C. J.; Visco, D. P. The Signature Molecular Descriptor. 2. Enumerating Molecules from Their Extended Valence Sequences. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (3), 721−734.
(37) Chong, J. W.; Thangalazhy-Gopakumar, S.; Muthoosamy, K.; Chemmangattuvalappil, N. G. Design of bio-oil additives via molecular signature descriptors using a multi-stage computer-aided molecular design framework. *Front. Chem. Sci. Eng.* **2022**, *16* (2), 168−182.
(38) Kayello, H. M.; Tadisina, N. K. R.; Shlonimskaya, N.; Biernacki, J. J.; Visco, D. P., Jr An Application of Computer-Aided Molecular Design (CAMD) Using the Signature Molecular Descriptor—Part 1. Identification of Surface Tension Reducing Agents and the Search for Shrinkage Reducing Admixtures. *J. Am. Ceram. Soc.* **2014**, *97* (2), 365−377.
(39) Mohamed, A.; Visco, D. P.; Bastidas, D. M. Significance of π−Electrons in the Design of Corrosion Inhibitors for Carbon Steel in Simulated Concrete Pore Solution. *Corrosion* **2021**, *77* (9), 976−990.
(40) Chen, J. J. F.; Visco, D. P., Jr Developing an in silico pipeline for faster drug candidate discovery: Virtual high throughput screening with the Signature molecular descriptor using support vector machine models. *Chem. Eng. Sci.* **2017**, *159*, 31−42.
(41) Weis, D. C.; Visco, D. P.; Faulon, J.-L. Data mining PubChem using a support vector machine with the Signature molecular descriptor: Classification of factor XIa inhibitors. *J. Mol. Graphics Modell.* **2008**, *27* (4), 466−475.
(42) Weis, D. C.; Visco, D. P. Computer-aided molecular design using the Signature molecular descriptor: Application to solvent selection. *Comput. Chem. Eng.* **2010**, *34* (7), 1018−1029.
(43) Weis, D. C.; MacFarlane, D. R. Computer-Aided Molecular Design of Ionic Liquids: An Overview. *Aust. J. Chem.* **2012**, *65* (11), 1478−1486.
(44) Meftahi, N.; Klymenko, M.; Christofferson, A. J.; Bach, U.; Winkler, D. A.; Russo, S. P. Machine learning property prediction for organic photovoltaic devices. *npj Comput. Mater.* **2020**, *6* (1), No. 166.
(45) Weis, D. C.; Faulon, J.-L.; LeBorne, R. C.; Visco, D. P. The Signature Molecular Descriptor. 5. The Design of Hydrofluoroether Foam Blowing Agents Using Inverse-QSAR. *Ind. Eng. Chem. Res.* **2005**, *44* (23), 8883−8891.
(46) Camacho-Mendoza, R. L.; Gutiérrez-Moreno, E.; Guzmán-Percástegui, E.; Aquino-Torres, E.; Cruz-Borbolla, J.; Rodríguez-Ávila, J. A.; Alvarado-Rodríguez, J. G.; Olvera-Neria, O.; Thangarasu, P.; Medina-Franco, J. L. Density Functional Theory and Electrochemical Studies: Structure−Efficiency Relationship on Corrosion Inhibition. *J. Chem. Inf. Model.* **2015**, *55* (11), 2391−2402.
(47) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1* (1), No. 140022.
(48) Browne, M. W. Cross-Validation Methods. *J. Math. Psychol.* **2000**, *44* (1), 108−132.
(49) Wong, T.-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* **2015**, *48* (9), 2839−2846.
(50) Altman, D. G.; Bland, J. M. How to obtain the P value from a confidence interval. *BMJ* **2011**, *343*, No. d2304, DOI: 10.1136/bmj.d2304.
(51) Aerts, M.; Claeskens, G.; Hart, J. D. Testing lack of fit in multiple regression. *Biometrika* **2000**, *87* (2), 405−424.
(52) Botelho, A. L.; Shin, Y.; Liu, J.; Lin, X. Structure and Optical Bandgap Relationship of π-Conjugated Systems. *PLoS One* **2014**, *9* (1), No. e86370.
(53) Teunissen, J. L.; De Proft, F.; De Vleeschouwer, F. Tuning the HOMO−LUMO Energy Gap of Small Diamondoids Using Inverse Molecular Design. *J. Chem. Theory Comput.* **2017**, *13* (3), 1351−1365.
(54) Koepnick, B. D.; Lipscomb, J. S.; Taylor, D. K. Effect of Substitution on the Optical Properties and HOMO−LUMO Gap of Oligomeric Paraphenylenes. *J. Phys. Chem. A* **2010**, *114* (50), 13228−13233.

https://doi.org/10.1021/acsomega.4c08626
ACS Omega 2025, 10, 2799−2808