Sooting Tendency of Substituted Aromatic Oxygenates: The Role of Functional Groups and Positional Isomerism in Vanillin Isomers

Hojin Jung[a,‡], Jaeyoung Cho[b,‡,*], Yeonjoon Kim[c], Zhanhong Xiang[d], Sabari Kumar[a], Piper Barnard[d], Charles S. McEnally[d], Lisa D. Pfefferle[d], Seonah Kim[a,*]

[a]*Department of Chemistry, Colorado State University, Fort Collins, CO 80523, USA*
[b]*Department of Aerospace and Mechanical Engineering, The University of Texas at El Paso, El Paso, TX 79968, USA*
[c]*Department of Chemistry, Pukyong National University, Busan 48513, Korea*
[d]*Department of Chemical and Environmental Engineering, Yale University, New Haven, CT 06520, USA*

Abstract

Substituted aromatics are commonly observed in lignin-based biofuel; however, their high sooting tendency prevents direct utilization in commercial combustors. Recent studies have revealed that oxygenated functional group substitution could effectively suppress the soot emission from aromatic biofuels. This study aims to enhance the understanding of sooting tendencies in aromatic oxygenates with mono-, di-, and tri-substitutions, focusing on various functional groups and their positional isomerism. We established a yield sooting index (YSI) database of 42 single-ring aromatic compounds, including 30 new measurements from the present study. The constructed database was utilized to develop a multivariate linear regression (MLR) model to predict the YSI of substituted aromatic oxygenates based on their structural features. The fitted coefficients of the MLR model indicate vastly different impacts of hydroxyl, formyl, and methoxy functional group, as well as the importance of positional isomerism. To understand the role of oxygenated functional groups, we used substituted vanillin isomers containing hydroxyl, methoxy, and formyl groups as a model system. Comparing the sooting tendencies of these compounds revealed a high sensitivity of YSI to positional isomerism. A further mechanistic study using quantum-mechanical calculations showed that subtle interactions between three oxygenated functional groups in vanillin isomers can alter their thermal decomposition pathway, affecting the sooting tendencies of these aromatic fuels. The present study provides a novel statistical and theoretical explanation of how oxygenated substitution and its positional isomerism influence sooting behaviors, facilitating the rational design of lignin-based biofuels with minimal soot emission.

Keywords: Yield sooting index; Multivariate linear regression; Density function theory; Isomer effect; Vanillin

---

*Corresponding author.
‡Equal contribution.

1. Introduction

Biomass has been a dominant source of alternative fuels (biofuels) for decades, decarbonizing the transportation and industrial sector with its low carbon footprint. Most current biofuels come from sugar, starch, and vegetable oils, which can easily be converted into aliphatic oxygenates (alcohol, ether, esters) through thermochemical processes and fermentation [1,2]. These compounds, especially bioethanol and fatty-acid methyl esters, were successfully integrated into mass production and blended into conventional fuels in many countries [3].

Lignocellulosic biomass, meanwhile, has been considered a challenging but promising source of future biofuels. The elevated lignin content in lignocellulosic biomass results in a high concentration of aromatics compounds in the derived biofuel, as the lignin transforms to various oxygenated aromatics such as phenol, cresol, and xylenol, during depolymerization [4,5]. The aromatic content in the fuel is sometimes beneficial for applications requiring low-reactivity fuels (e.g. spark-ignition engine), but in general has been avoided due to its higher sooting tendency [6]. Still, the lignocellulosic biofuel has a huge advantage in its even lower carbon footprint. Spatari et al. [7] showed that the greenhouse gas emission from lignocellulosic (corn stover) biofuel is only ~50 % of those from energy crops (switchgrass), primarily attributed to the fact that it does not require N-fertilizer emitting N_2O.

This low environmental impact of lignocellulosic biofuel has driven numerous research studies to mitigate the soot emissions from aromatic compounds. Kim et al. [8] compared the sooting tendency of 15 different mono- and di-substituted aromatic compounds using a yield sooting index (YSI). Their statistical analysis showed a clear correlation between YSIs and formation enthalpy (ΔH_f) of the fuel radicals from the weakest bond-fission reaction, which may explain the lower sooting tendency of the oxygenated aromatic compounds. The following synergistic experiment-theory analysis revealed that the hydroxyl group effectively eliminates carbon atoms in the aromatic ring during thermal decomposition, reducing the sooting tendency of fuel. This finding also aligns with recent observations of the combustion community [9,10], where oxygenated aromatics showed a systematically lower sooting tendency than aromatic hydrocarbons.

Despite the advantages of the sooting tendency, oxygenated functional groups should be carefully added to the fuel, as they sometimes sacrifice other fuel properties. For example, higher oxygen content in the fuel leads to decreased heating values. The hydroxyl group, especially, is known to lower the melting point and increase the acidity. This clear trade-off between YSI and other fuel properties motivated the present study that aims to find the most effective way to mitigate soot emissions of aromatic compounds with minimal oxygen content.

The previous experiments from Wang et al. [11] showed that the benzyl alcohol emits more soot particles than anisole when blended into *n*-heptane, which may indicate the different impacts of hydroxy and methoxy substitution. Meanwhile, Etz et al. [12] compared the sooting tendency of 1- and 2-phenylethanol, showing dramatic sensitivity of YSI to positional isomerism. They explained that the higher YSI of 2-ethylphenol comes from its lower energy barrier to hydroxyl group elimination. Kim et al. [8] observed that the hydroxyl group's soot mitigation is higher when substituted on an ortho-position than on a meta-position by comparing the differential sooting behavior of 2- and 3-ethylphenol. Their mechanistic study revealed that positional isomerism alters the resonance stabilization of the radicals intermediate, affecting the formation of soot precursors from the flame. However, these studies investigated only mono- and di-substituted aromatics and tri-substituted ones have not been explored.

The present study aims to provide a more systematic and comprehensive understanding of the sooting tendency of aromatic oxygenates with mono-, di-, and tri-substitution, with varying functional groups and their positional isomerism. First, a YSI database for 42 single-ring aromatic compounds was collected, 30 of which were newly measured for the present study. The curated data was then utilized to develop a multivariate linear regression model (MLR) for YSI prediction, targeting bio-derived aromatic compounds. The fitted coefficients of the MLR model indicate the distinct impacts of functional groups, substituent positional isomerism, and the chemical stability of the fuel radicals on the aromatics' sooting tendencies.

These intriguing findings regarding such structural effects led to further mechanistic studies of soot precursor formation pathways of model compounds, vanillin, and its isomers, which contain hydroxyl, methoxy, and formyl groups. Vanillin presents high contents in bio-oils, as it is the direct product of the depolymerization of lignin [13], drawing the attention of combustion researchers for decades [14–16]. Additionally, vanillin and its isomer showed vastly different YSI depending on their substituted site, so they are appropriate for examining the effect of oxygenated functional groups and positional isomerism. A further mechanistic study using quantum-mechanical calculations revealed that three oxygenated functional groups in vanillin closely interact during the thermal decomposition, altering the resultant sooting tendency of aromatic fuels. The present study provides the novel statistical and theoretical explanation regarding the impact of oxygenated substitution and its positional isomerism on the sooting tendency, facilitating the rational design of lignin-based biofuels with low soot emission.

2. Methods

2.1 YSI measurements

Sooting tendencies of aromatic compounds listed in Table 1 were measured using a yield-based approach we developed previously [17]. The specific procedure used in this study is described in [18]; it consists of three steps: i) we separately doped 1000 µmol/mol (1000 ppm) of *n*-heptane (H), toluene (T), and each test sample (TF) into the fuel of a base methane/air flame; ii) we measured the maximum soot concentration in each flame with line-of-sight spectral radiance (LSSR; L); and iii) we rescaled the results into a YSI defined as:

$$\mathrm{YSI_{TF}} = (\mathrm{YSI_T} - \mathrm{YSI_H}) \times \frac{L_{\mathrm{TF}} - L_{\mathrm{H}}}{L_{\mathrm{T}} - L_{\mathrm{H}}} + \mathrm{YSI_H} \qquad (1)$$

This rescaling eliminates sources of systematic uncertainty, such as errors in the methane and air flowrates. Furthermore, it allows the new results to be compared quantitatively with a database containing measured YSIs for hundreds of organic compounds [19]. The parameters $\mathrm{YSI_T}$ and $\mathrm{YSI_H}$ are constants that define the YSI scale; their values – 170.9 and 36.0 – were taken from the database so that the newly measured YSIs would be on the same scale for a direct comparison.

Figure 1a shows the apparatus used in these measurements. For all the flames, the gas-phase flowrate of air is 50,000 $\mathrm{cm^3}$/min and the gas-phase flowrates of dopant, $\mathrm{CH_4}$, and $\mathrm{N_2}$, are 0.395 $\mathrm{cm^3}$/min, 282.425 $\mathrm{cm^3}$/min, and 112.18 $\mathrm{cm^3}$/min, respectively. Mass flow controllers actively governed the flowrates of the gas-phase components ($\mathrm{CH_4}$ and $\mathrm{N_2}$). A syringe pump controlled the injection rates of the liquid dopants into the gaseous $\mathrm{CH_4}/\mathrm{N_2}$ mixture. SI A lists the liquid-phase flowrates calculated for each test compound to produce the target gas-phase dopant flowrate, and the property values used in these calculations. Resistive tapes heated the fuel line to at least 70 °C and above, depending on the test compounds, and the burner to 170 °C. Given this heating, the dopants vaporized rapidly upon injection and were swept to the burner by the other fuel components. A PID controller maintained the burner temperature to within ±1 °C, which was crucial since the soot concentrations in the flame depend sensitively on the initial temperature of the reactants.

The figure 1a also shows the LSSR apparatus: a UV silica window embedded in the chimney transmits light from the flame to the rest of the setup; a fused silica biconvex lens focuses the light onto a 1 mm diameter circular aperture; an interference filter (Thorlabs FB660-10, frequency width at half maximum = 10 nm, center wavelength = 660 ± 2 nm) and an infrared-blocking filter (Schott KG2) isolate the light at 660 nm; and a red-enhanced photomultiplier tube (PMT; Oriel 77,348) detects the light. An A/D converter (LeCroy LT342, 1 MΩ input impedance) samples the PMT output at 5 Hz. Each sample is an average of 50,000 8-bit measures recorded 2 µs apart.

Figure 1b shows details of the burner. All dimensions are given in the original units specified by the suppliers. Reference [20] provides a detailed burner description, including CAD drawings. This burner is being employed by multiple research groups as a standard configuration for experimental and numerical studies of soot formation under the auspices of the International Sooting Flame Workshop [21].

Each test fuel was injected for 600 s, and L was averaged from 300 s to 600 s; SI B shows that the initial 300 s is more than adequate for the test fuel to equilibrate with the walls of the fuel line and burner with this heating. Isooctane was included in each measurement set as an internal standard; its measured YSIs were consistent over time and agreed with previous measurements [19].

Some of the test fuels in this study are solids at room temperature, so they could not be injected into the fuel mixture as a pure compound. Instead, they were dissolved in ethanol (E) with a TF:E molar ratio of 1:4 or 1:6. The endpoints H and T were also dissolved in E at the same molar ratios. Then

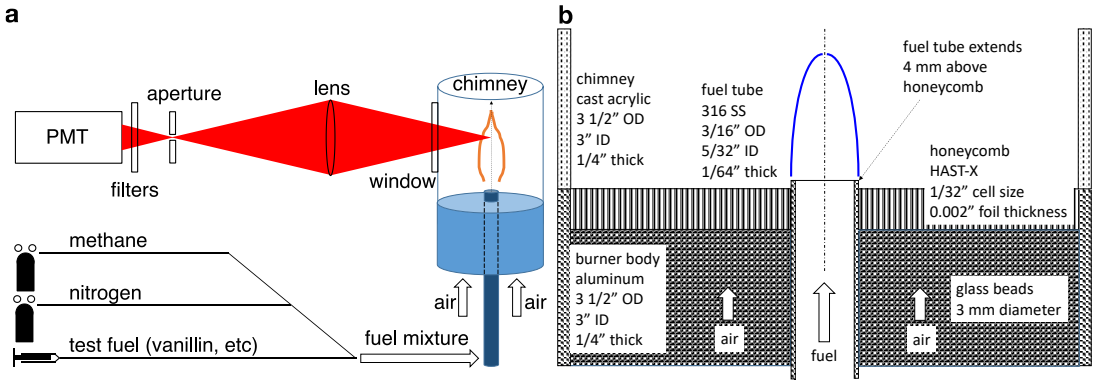

Figure 1. (a) Schematic diagram of the apparatus for measuring L. (b) Details of the specific burner.

5000 μmol/mol of each solution was injected into the flame, values of L_{TF}, L_H, and L_T were measured for these solutions, and YSI_{TF} was calculated from Eq. (1). SI C shows that the L of E and 24BZA mixture increased linearly with the 24BZA mole fraction, which experimentally confirms that the components were not condensing in the fuel delivery system.

Each YSI was measured three times and then averaged. The random uncertainty is estimated to be ±2% based on the standard deviation of the measured values for the internal standard. The systematic uncertainty is estimated to be ±1%, mainly due to uncertainty in the ratio of the mass densities between species. There is an additional uncertainty when comparing the new measurements with the earlier measurements in [19], which we estimate to add ±2% based on the measured values of the internal standard versus its value in [19]. Overall, we estimate that the uncertainty in the measured YSI is ±5% or ±5.0, whichever is larger. The total uncertainty of ±5% is supported by previous results including: (1) agreement between YSIs measured with different soot diagnostics [22], (2) correlation between measured YSIs and sooting tendencies based on smoke point [23], (3) agreement between measured YSIs of mixtures and linear mixing rules [24], and (4) agreement between measured YSIs and YSIs simulated with detailed kinetic mechanisms [25].

2.2 Multivariate linear regression

To gain a systematic understanding of the structural effect on the aromatics' sooting tendencies, an MLR model for YSI prediction was developed with diverse types of descriptors representing i) functional group contribution, ii) substituent positional isomerism, and iii) the chemical stability of fuel radicals.

Chemical descriptors were calculated to parameterize the fuel molecules as inputs into the MLR model. First, the functional group effect was accounted for by counting the number of each functional group in the substituent of the aromatic compounds. The substituents found in the compounds in Table 1 include hydroxyl, methoxy, formyl, methyl, and ethyl, all commonly observed in lignin-based biofuels. Moreover, the effect of positional isomerism was considered by quantifying the instances of ortho-, meta-, and para-substitutions (N_{ortho}, N_{meta}, N_{para}) in di- and tri-substituted aromatic oxygenates.

To account for the chemical stability of the fuel radicals, we utilized the standard enthalpy of formation ($\Delta H_{f,RSR}$) of resonance-stabilized radicals (RSRs) formed from the weakest bond fission reaction: Parent aromatic → •RSR + •R, where the R represents the co-product radicals

from the bond fission reactions, such as H, CH_3, CH_3O, etc.

Kim et al. [8], who previously employed $\Delta H_{f,RSR}$ for YSI prediction, calculated the $\Delta H_{f,RSR}$ of each aromatic compound using an equation derived from the bond dissociation enthalpy (BDE) definition:

$$\Delta H_{f,RSR} = \Delta H_{BDE} + \Delta H_{f,parent} - \Delta H_{f,R} \quad (2)$$

, where ΔH_{BDE} is the weakest BDE, and $\Delta H_{f,parent}$ and $\Delta H_{f,R}$ represent the standard enthalpy of formation of the parent aromatic compound and the co-product radical R, respectively.

Table 1. Measured YSI (YSI_{EXP}) for 42 single-ring aromatic compounds.

Test Fuel	YSI_{EXP}	Test Fuel	YSI_{EXP}	Test Fuel	YSI_{EXP}
toluene	170.9 ± 7.7[a]	4-methylanisole	123.0 ± 6.2[b]	1,2-benzenediol (catechol)	56.8 ± 5.0[b]
ethylbenzene	216.0 ± 10.8[a]	3-methylanisole	126.0 ± 6.3[b]	4-methoxyphenol	54.9 ± 5.0[b]
phenol	81.3 ± 5.0[b]	2-methylanisole	122.0 ± 6.1[b]	3-methoxyphenol	64.1 ± 5.0[b]
Anisole	111.0 ± 5.6[a]	4-methylbenzaldehyde (p-tolualdehyde)	176.7 ± 8.8[b]	2-methoxyphenol	64.0 ± 5.0[b]
benzaldehyde	119.0 ± 6.0[b]	3-methylbenzaldehyde (m-tolualdehyde)	188.0 ± 9.4[b]	2-hydroxybenzaldehyde	114.6 ± 5.7[b]
1,4-dimethylbenzene (p-xylene)	202.0 ± 10.1[a]	2-methylbenzaldehyde (o-tolualdehyde)	186.1 ± 9.3[b]	4-mehtoxybenzaldehyde	116.8 ± 5.8[b]
1,3-dimethylbenzene (m-xylene)	221.6 ± 11.0[a]	1,4-diethylbenzene	270.7 ± 13.5[a]	3-methoxybenzaldehyde	121.8 ± 6.1[b]
1,2-dimethylbenzene (o-xylene)	200.0 ± 10.0[a]	1,3-diethylbenzene	320.9 ± 16.0[a]	1,3,5-trihydroxybenzene	9.7 ± 5.0[b]
1-ethyl-4methylbenzene	257.1 ± 12.9[a]	1,2-diethylbenzene	376.3 ± 18.8[a]	1,2,4-trihydroxybenzene	15.8 ± 5.0[b]
1-ethyl-3methylbenzene	278.0 ± 13.9[a]	4-ethylphenol	129.2 ± 6.5[b]	1,2,3-trihydroxybenzene	34.0 ± 5.0[b]
1-ethyl-2methylbenzene	267.0 ± 13.4[a]	3-ethylphenol	134.8 ± 6.7[b]	4-hydroxy-3-methoxybenzaldehyde (vanillin)	47.7 ± 5.0[b]
4-methylphenol (p-cresol)	104.4 ± 5.2[b]	2-ethylphenol	118.9 ± 5.9[b]	2-hydroxy-3-methoxybenzaldehyde (o-vanillin)	69.7 ± 5.0[b]
3-methylphenol (m-cresol)	107.0 ± 5.4[b]	1,4-benzenediol (hydroquinone)	37.8 ± 5.0[b]	2-hydroxy-4-methoxybenzaldehyde (24BZA)	64.2 ± 5.0[b]
2-methylphenol (o-cresol)	101.9 ± 5.1[b]	1,3-benzenediol (resorcinol)	34.1 ± 5.0[b]	2-hydroxy-5-methoxybenzaldehyde (25BZA)	56.4 ± 5.0[b]

[a] From McEnally et al. [19], [b] Present work

While $\Delta H_{f,R}$ values were sourced from the NIST Chemistry WebBook database [26], not all $\Delta H_{f,parent}$ values were available there. Therefore, we calculated these $\Delta H_{f,parent}$ values using quantum mechanical methods, employing the Gaussian 16 software suite [27] (Section 2.3, *vide infra*). This approach involved quasi-isodesmic reactions with CH_4 and C_6H_6, ensuring the preservation of the aromatic chemical environments of both reactants and products for more accurate heat of formation calculations [28]. A comprehensive list of all collected descriptors for the 42 compounds used in the MLR model is available in the Supplemental Material (Table SI D.)

Among all the descriptors described above, the most relevant descriptors were selected using LASSO (Least Absolute Shrinkage and Selection Operator) regularization [29], effectively reducing prediction variance by adjusting or nullifying coefficients of descriptors in the MLR model. Additionally, the statistical significance of each descriptor was further evaluated using SHAP (Shapley Additive exPlanations) [30] values.

2.3 Quantum mechanics calculation

We adopted the M06-2X density functional along with the def2-TZVP basis set for all geometry optimizations, single-point energy calculations, and Hessian calculations. The M06-2X/def2-TZVP level of theory produces energy results with 1-2 kcal/mol uncertainty, while preventing excessive amounts of spin contamination [31]. Verification of transition states was achieved by detecting a singular imaginary frequency, while intrinsic reaction coordinate analysis was performed to ascertain the associated reactants and products for each identified transition state. The ALFABET (A machine-Learning derived, Fast, Accurate Bond dissociation Enthalpy Tool) [32] was utilized for determining the weakest bond in aromatic compounds and estimating its bond dissociation enthalpy (BDE), whose uncertainty bound is known to be within 1 kcal/mol for most neutral organic compounds consisting of C, H, O, and N atoms. All quantum-mechanical calculations were conducted using the Gaussian 16 program suite [27].

3. Results and discussion

3.1 YSIs of aromatic compounds of interest

We expanded the YSI measurement database by conducting YSI measurements on additional 30 aromatic oxygenates, resulting in an updated database that now encompasses 5 mono-, 30 di-, and 7 tri-substituted compounds, as detailed in Table 1. These compounds feature diverse substituents, including hydroxyl, methoxy, and formyl groups commonly found in biomass and methyl and ethyl groups prevalent in petroleum products positioned at different locations.

As described in Table 1, the measured YSIs of aromatics tend to be lower when there are oxygenated substitutions, which can be seen by comparing the YSI of toluene to that of phenol, anisole, and benzaldehyde. On the other hand, the substitution of alkyl groups (methyl or ethyl) leads to higher soot emission, as illustrated by the >200 YSIs from xylene and ethyl-methyl-benzene isomers. Interestingly, the YSIs of bi-substituted and tri-substituted aromatics indicate a significant impact of positional isomerism on the sooting tendency. For example, catechol (1,2-benzenediol) has ~20 units higher YSI than its positional isomers, resorcinol (1,3-benzenediol) and hydroquinone (1,4-benzenediol).

This large range in YSIs of single ring aromatic compounds, spanning from 9.7 (1,3,5-trihydroxybenzene) to 376.3 (1,2-diethylbenzene), motivated the statistical analysis of the YSI database using an MLR model in the following section.

3.2 Multivariate Linear Regression

The developed MLR model incorporates LASSO for descriptor selection and the fitted coefficients of each descriptor as follows:

$$\frac{\mathrm{YSI_{EXP}}}{\mathrm{MW}} = 0.54N_{\mathrm{methyl}} + 0.98N_{\mathrm{ethyl}} - 1.08N_{\mathrm{hydroxyl}} - 0.34N_{\mathrm{methoxy}} - 0.08N_{\mathrm{para}} + 0.18\Delta H_{\mathrm{f,RSR}} \quad (3)$$

, where MW represents the molecular weight of the tested fuel. Figure 2a presents the parity plot for the developed MLR model. The model performs strongly in predicting sooting tendencies, as evidenced by the

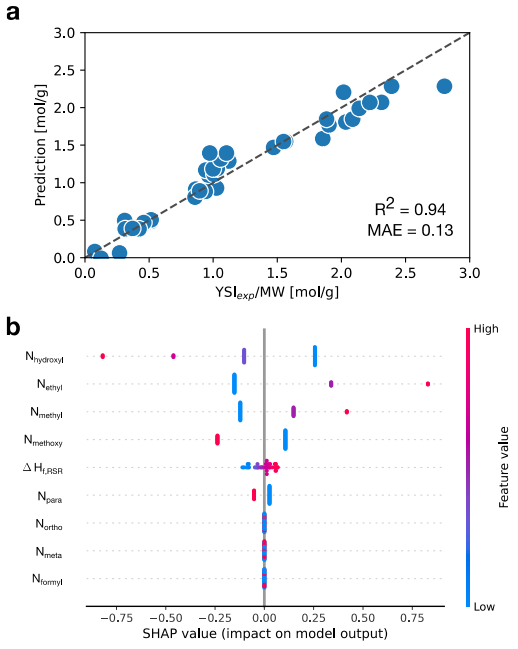

Figure 2. (a) Parity plot for YSI/MW prediction with multivariate linear regression model using LASSO. (b) SHAP values for features used in the model.

R^2 value of 0.94 and a mean absolute error (MAE) of 0.13 across 42 compounds.

Analysis of the model's fitted coefficients in Eq. (3) reveals insights into the impact of various factors on the normalized YSI. The coefficients associated with different functional groups indicate their quantitative contribution to the sooting tendency of aromatics; more methyl and ethyl groups increase the YSI, while hydroxyl and methoxy groups mitigate the soot emission. Meanwhile, the fitted coefficient for N_{para} is negative, implying a lower YSI of di- and tri-substituted aromatics in the para-position, as opposed to those substituted in ortho- and meta-positions. Notably, the positive coefficients of $\Delta H_{f,RSR}$ align with Kim et al.'s research [8], which showed the positive correlation between YSI/MW and $\Delta H_{f,RSR}$ of 15 compounds. This consistency implies that the resonance-stabilized structure of the fuel radical is still one of the important features for understanding the sooting tendency of di- and tri-substituted aromatics.

Figure 2b displays the SHAP values for all the descriptors in our model, shedding light on their relative importance and impact on the model's predictions. It is observed that descriptors like N_{ortho}, N_{meta}, N_{formyl} were excluded through the LASSO process.

Despite the effectiveness of the developed regression model in predicting YSI_{EXP}/MW, there are still outliers presenting larger errors than the MAE. Moreover, the current YSI database includes only a limited number of positional isomers of di- and tri-substituted compounds, which may have led to the statistical insignificance of N_{ortho} and N_{meta} in SHAP analysis (Fig. 1b). This lack of significance should not be misinterpreted as invalidating the potential impact of ortho and meta substitutions on the YSI values of certain compounds. This potential influence is clearly shown by the variations in the YSI values of diethylbenzene isomers, as listed in Table 1. The current model does not possess the capacity to delve into how positional isomerism affects reaction pathways. To address this limitation of the MLR model, we conducted mechanistic studies that elucidate substituent effects on sooting tendencies with quantum-mechanical calculations, particularly for four vanillin isomers having all hydroxyl, methoxy, and formyl groups altogether.

3.3 Mechanistic studies – substituent effects on sooting tendencies of vanillin isomers

Figure 3 illustrates the proposed thermal decomposition pathways of four vanillin isomers and their calculated ground-state energies with zero-point energy correction (*E*+*ZPE*). As noted, the selected vanillin isomers show vastly different YSI depending on the positional isomerism. Vanillin, where the methoxy and formyl groups are attached at ortho- and para-position relative to hydroxyl groups, has the lowest YSI of 47.7 among the tested isomers. Meanwhile, by simply relocating the formyl groups to ortho-position (*o*-vanillin), its YSI increases by 22.0 units. Interestingly, YSIs of vanillin isomers gradually decrease as methoxy groups are away from the hydroxyl group, as can be inferred by comparing *o*-vanillin, 24BZA, and 25BZA.

We posited that the soot formation from vanillin isomers is initiated by unimolecular thermal decomposition rather than bimolecular hydrogen abstraction, especially in the diffusion flame burner used in the present study. This assumption is based on a relatively low dissociation temperature of vanillin. Wu et al. [14] executed a microtubular experiment on the vanillin's thermal decomposition with a residence time of 10 – 100 μs, where they observed the vanillin starts to dissociate around 850 K. This temperature corresponds to ~1 cm height above the burner (HAB)

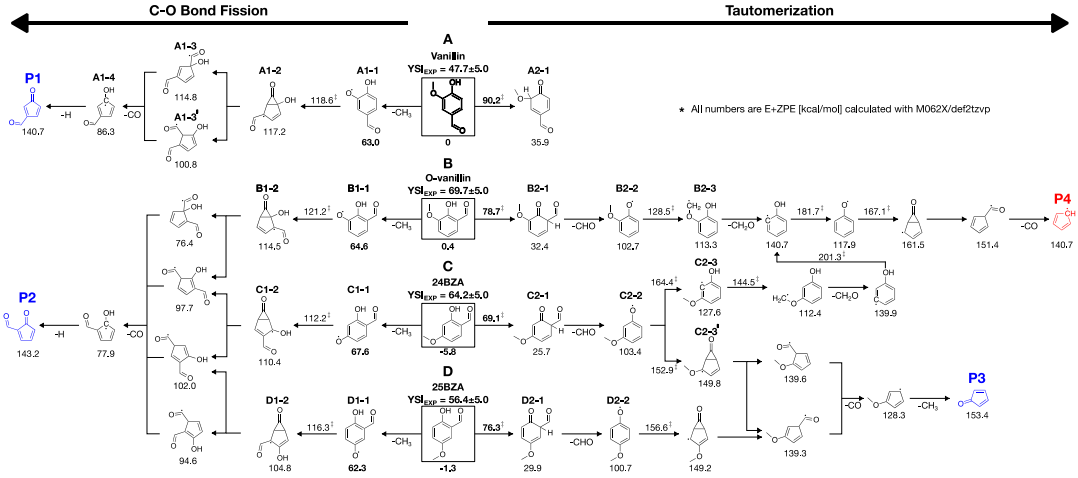

Figure 3. Thermal decomposition pathway of four vanillin isomers, starting from the C-O bond fission or tautomerization.

of the present experiment, while the flame tip is located at ~6 cm HAB [33]. The reactive radicals such as H, OH, and CH_3 usually do not reach this lower part (~1 cm) of diffusion flame; thus, the bimolecular hydrogen abstraction by these reactive radicals hardly competes with the fast-unimolecular reactions of vanillin. Still, it is noteworthy that the bimolecular reaction of "vanillin + OH = product" favors hydrogen abstractions at the formyl groups [34]; thus, its subsequent reactions may resemble those of the tautomerization pathway in Figure 3 (*vide infra* Section 3.3.2).

The unimolecular decomposition of the vanillin isomers can be initiated either through i) the C-O bond fission of the methoxy group, which is the weakest bond, or through ii) tautomerization, where a hydrogen atom in the hydroxyl group migrates to another carbon atom to form a ketone. The details of each pathway are described below.

3.3.1 C-O bond fission

Breaking the weakest C-O bond in the methoxy group of vanillin (YSI=47.7) makes a phenoxy radical that readily eliminates the CO through intermediate radicals with 3-membered rings (A ➔ A1-1 ➔ A1-2 ➔ A1-3/A1-3' ➔ A1-4) [35]. The resultant hydroxyl and formyl substituted cyclopentadienyl radical (A1-4) undergoes subsequent O-H bond fission (A1-4 ➔ P1), culminating in the formation of P1. Similarly, the C-O bond fission of other vanillin isomers leads to the formation of P2, which is a positional isomer of P1. Both P1 and P2 can potentially lose a formyl group, forming a radical that leads to the creation of cyclopentadienone (P3). Cyclopentadienone is known for its propensity to undergo further unimolecular decomposition, resulting in cyclobutadiene and carbon monoxide through ring contraction and CO elimination [36]. Given its 4-membered cyclic and unstable hydrocarbon structure, cyclobutadiene can break down further, generating acetylenes. Consequently, the C-O bond fission of vanillin isomers minimally contributes to soot formation, resulting in relatively smaller species, such as acetylene.

The mechanistic study above indicates that the C-O bond fission alone does not fully account for the differences in YSI_{EXP} across the isomers. In this regard, we postulated that the tautomerization pathway may be where the positional isomerism effect originated.

3.3.2 Tautomerization

Figure 3 shows that the tautomerization has a higher energy barrier than C-O bond fission in all the tested vanillin isomers. Vanillin tautomerization faces an energetic barrier of 90.2 kcal/mol (A ➔ A2-1), while its C-O bond fission energy is only 63.0 kcal/mol (A ➔ A1-1). This drastically different energy barrier indicates that the C-O bond cleavage is vanillin's single dominant dissociation pathway, leading to its lowest YSI among the four isomers.

However, other vanillin isomers – *o*-vanillin, 24BZA, and 25BZA – have a relatively lower energy barrier of 74.9 – 78.3 kcal/mol for the tautomerization pathway. Although this is still higher than the C-O bond fission barrier, the tautomerization channel can compete with C-O bond fission owing to the different nature of the transition state: the C-O bond fission proceeds through a loose transition state; thus, its rate coefficient decreases more drastically at low pressure than the tautomerization that goes through a tight transition state [37]. Considering the present experiment was conducted at atmospheric pressure, the pressure effect may lead to significant reaction flux via tautomerization. Moreover, tunneling effects are more facile through a tight transition state [37], especially when it involves H migration, making the tautomerization channel competitive to C-O fission despite a higher energy barrier.

Following the tautomerization of *o*-vanillin (B ➔ B2-1), its weakest bond shifts to the C-C bond between the formyl group and the α-carbon. Subsequent C-C bond fission (B2-1 ➔ B2-2) leads to the loss of the formyl group and the formation of a phenoxy radical (B2-2). Interestingly, when the methoxy group is adjacent to the radical center of the phenoxy radical, like in *o*-vanillin, the hydrogen in the methoxy group easily migrates to the radical center [38] (B2-2 ➔ B2-3). Once it migrates, the subsequent reactions dispose of all oxygen atoms from the molecules, forming the cyclopentadienyl radical (P4). The P4 radical has been known as a strong soot precursor [39], and it explains the highest YSI (69.7) of *o*-vanillin among the tested isomers.

On the other hand, the phenoxy radical from the tautomerization of 24BZA (C2-2) has a relatively higher energy barrier to forming a P4 radical owing to the longer distance between the phenoxy radical center and the methoxy group. Thus, the P4 formation pathway (C2-2 ➔ C2-3) can compete with the CO disposal pathway (C2-2 ➔ C2-3'), with a similar energy barrier (61 kcal/mol vs. 49.5 kcal/mol). The CO disposal from phenoxy radical results in the formation of P3, which quickly dissociates to CO and two acetylenes [36] and thus minimally contributes to soot formation. Consequently, the 24BZA has a lower YSI (64.2) compared to *o*-vanillin.

Lastly, the phenoxy radical from 25BZA's tautomerization (D2-2) almost exclusively leads to P3 formation because the methoxy group is located at the para-position, making its YSI as low as 56.4. The mechanistic study on 25BZA revealed that its low YSI_{EXP} is due to the comparatively least interaction between the methoxy group and the radical center. This observation aligns with the negative coefficient of N_{para} in the MLR model.

The mechanistic study above clearly suggests that the subtle change in the positional isomerism of formyl, methoxy, and hydroxyl groups can significantly alter the thermal dissociation pathway of

vanillin isomers, thus explaining the varying YSI measurements observed.

4. Conclusion

The present study examined the sooting tendencies of mono-, di-, and tri-substituted aromatic oxygenates, focusing especially on the role of different functional groups and their positional isomerism. A comprehensive yield sooting index (YSI) database for 42 single-ring aromatic compounds, including 30 new entries, was established. Leveraging the curated data, a multivariate linear regression (MLR) model was developed for YSI prediction. The fitted coefficients of the developed MLR model indicated a positive contribution of $\Delta H_{f,RSR}$, $N_{methyl,}$ and N_{ethyl} to the YSI, while the $N_{hydroxy}$, $N_{methoxy}$, and N_{para} coefficients negatively contribute to the YSI. A mechanistic study was performed on the thermal decomposition pathway of vanillin and its isomers to understand the effect of substituent position in aromatic oxygenates. The mechanistic study suggested that the C-O bond fission of vanillin and its isomers led to their complete breakdown, minimally contributing to soot formation. However, *o*-vanillin and 24BZA can form potent soot precursors through tautomerization, explaining their high YSI compared to the other isomers. Interestingly, the reaction pathway following the tautomerization has a strong sensitivity to the distance between the hydroxyl and methoxy group, which aligns with the positive contribution of N_{para} in the MLR model. These insights, combining statistical and theoretical methodologies, contribute to developing lignin-based biofuels with reduced soot emissions, promoting the creation of environmentally friendly fuel alternatives.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This journal article was developed based on funding from the Alliance for Sustainable Energy, LLC, Managing and Operating Contractor for the National Renewable Energy Laboratory for the U.S. Department of Energy. It was conducted as part of the Co-Optimization of Fuels & Engines (Co-Optima) project sponsored by the Bioenergy Technologies and Vehicle Technologies Offices of the U.S. Department of Energy. This work for CSU was supported by the Colorado State University startup funds for PI (Seonah Kim) and the U.S. National Science Foundation award CBET 2210894. Computer time was provided by the NSF Extreme Science and Engineering Discovery Environment (XSEDE, now ACCESS), Grant no. TG-CHE210034. P.B. was supported by a STARS Summer Research Fellowship from Yale College.

Appendix A. Supplementary Information

References

[1] R.K. Srivastava, N.P. Shetti, K.R. Reddy, E.E. Kwon, M.N. Nadagouda, T.M. Aminabhavi, Biomass utilization and production of biofuels from carbon neutral materials, Environ. Pollut. 276 (2021) 116731.

[2] D.P. Ho, H.H. Ngo, W. Guo, A mini review on renewable sources for biofuel, Bioresour. Technol. 169 (2014) 742–749.

[3] OECD, F. and A.O. of the U. Nations, OECD-FAO Agricultural Outlook 2021-2030, (2021).

[4] V. Ashokkumar, R. Venkatkarthick, S. Jayashree, S. Chuetor, S. Dharmaraj, G. Kumar, W.-H. Chen, C. Ngamcharussrivichai, Recent advances in lignocellulosic biomass for biofuels and value-added bioproducts - A critical review, Bioresour. Technol. 344 (2022) 126195.

[5] A.T. Hoang, H.C. Ong, I.M.R. Fattah, C.T. Chong, C.K. Cheng, R. Sakthivel, Y.S. Ok, Progress on the lignocellulosic biomass pyrolysis for biofuel production toward environmental sustainability, Fuel Process. Technol. 223 (2021) 106997.

[6] W.A. A. Demirbas M.A. Balubaid, A.M. Basahel, M.H. Sheikh, Octane Rating of Gasoline and Octane Booster Additives, Pet. Sci. Technol. 33 (2015) 1190–1197.

[7] S. Spatari, H.L. MacLean, Characterizing Model Uncertainties in the Life Cycle of Lignocellulose-Based Ethanol Fuels, Environ. Sci. Technol. 44 (2010) 8773–8780.

[8] Y. Kim, B.D. Etz, G.M. Fioroni, C.K. Hays, P.C. St. John, R.A. Messerly, S. Vyas, B.P. Beekley, F. Guo, C.S. McEnally, L.D. Pfefferle, R.L. McCormick, S. Kim, Investigation of structural effects of aromatic compounds on sooting tendency with mechanistic insight into ethylphenol isomers, Proc. Combust. Inst. 38 (2021) 1143–1151.

[9] Z. Xiang, F. Guo, D. Curtis, C.S. McEnally, L.D. Pfefferle, J. Zhu, Sooting Tendencies of Phenolic Hydrocarbons, (2022).

[10] B.P. Beekley, C.S. McEnally, P.C.S. John, S. Kim, A. Jain, H. Kwon, Y. Xuan, L.D. Pfefferle, Sooting Tendencies of Aromatic Hydrocarbons With Oxygen-Containing Side-Chains, (2020).

[11] W. Wang, Z. Zhang, D. Shi, Y. Huang, L. Zhou, Study of soot formations in co-flow laminar diffusion flames of n-heptane and oxygenated aromatic biofuels from atmospheric condition to 2.3 bar, Fuel. 297 (2021) 120753.

[12] B.D. Etz, G.M. Fioroni, R.A. Messerly, M.J. Rahimi, P.C. St. John, D.J. Robichaud, E.D. Christensen, B.P. Beekley, C.S. McEnally, L.D. Pfefferle, Y. Xuan, S. Vyas, R.S. Paton, R.L. McCormick, S. Kim, Elucidating the chemical pathways responsible for the sooting tendency of 1 and 2-phenylethanol, Proc. Combust. Inst. 38 (2021) 1327–1334.

[13] C. Isola, H.L. Sieverding, A.M. Numan-Al-Mobin, R. Rajappagowda, E.A. Boakye, D.E. Raynie, A.L.

Smirnova, J.J. Stone, Vanillin derived from lignin liquefaction: a sustainability evaluation, Int. J. Life Cycle Assess. 23 (2018) 1761–1772.
[14] X. Wu, Z. Pan, S. Bjelić, P. Hemberger, A. Bodi, Unimolecular thermal decarbonylation of vanillin stifled by the bimolecular reactivity of methyl-loss intermediate, J. Anal. Appl. Pyrolysis. 161 (2022) 105410.
[15] M. Pelucchi, C. Cavallotti, A. Cuoci, T. Faravelli, A. Frassoldati, E. Ranzi, Detailed kinetics of substituted phenolic species in pyrolysis bio-oils, React. Chem. Eng. 4 (2019) 490–506.
[16] E.-J. Shin, M.R. Nimlos, R.J. Evans, A study of the mechanisms of vanillin pyrolysis by mass spectrometry and multivariate analysis, Fuel. 80 (2001) 1689–1696.
[17] C.S. McEnally, L.D. Pfefferle, Improved sooting tendency measurements for aromatic hydrocarbons and their implications for naphthalene formation pathways, Combust. Flame. 148 (2007) 210–222.
[18] J. Zhu, J.V. Alegre-Requena, P. Cherry, D. Curtis, B.G. Harvey, M.A. Jabed, S. Kim, C.S. McEnally, L.D. Pfefferle, J.-D. Woodroffe, Sooting tendencies of terpenes and hydrogenated terpenes as sustainable transportation biofuels, Proc. Combust. Inst. 39 (2023) 877–887.
[19] C.S. McEnally, D.D. Das, L.D. Pfefferle, Yield Sooting Index Database Volume 2: Sooting Tendencies of a Wide Range of Fuel Compounds on a Unified Scale, (2017).
[20] J. Gau, D. Das, C.S. McEnally, D. Giassi, N. Kempema, M. Long, Yale Coflow Burner Information and CAD Drawings, (2017).
[21] B. Franzelli, M. Roussillo, P. Scouflaire, J. Bonnety, R. Jalain, T. Dormieux, S. Candel, G. Legros, Multi-diagnostic soot measurements in a laminar diffusion flame to assess the ISF database consistency, Proc. Combust. Inst. 37 (2019) 1355–1363.
[22] D.D. Das, P.C. St. John, C.S. McEnally, S. Kim, L.D. Pfefferle, Measuring and predicting sooting tendencies of oxygenates, alkanes, alkenes, cycloalkanes, and aromatics on a unified scale, Combustion and Flame 190 (2018) 349–364.
[23] D.D. Das, C.S. McEnally, T.A. Kwan, J.B. Zimmerman, W.J. Cannella, C.J. Mueller, L.D. Pfefferle, Sooting tendencies of diesel fuels, jet fuels, and their surrogates in diffusion flames, Fuel 197 (2017) 445–458.
[24] C.S. McEnally, Y. Xuan, P.C. St. John, D.D. Das, A. Jain, S. Kim, T.A. Kwan, L.K. Tan, J. Zhu, L.D. Pfefferle, Sooting tendencies of co-optima test gasolines and their surrogates, Proceedings of the Combustion Institute 37 (2019) 961–968.
[25] H. Kwon, S. Lapointe, K. Zhang, S.W. Wagnon, W.J. Pitz, J. Zhu, C.S. McEnally, L.D. Pfefferle, Y. Xuan, Sooting tendencies of 20 bio-derived fuels for advanced spark-ignition engines, Fuel 276 (2020) 118059.
[26] P.J. Linstrom, W.G. Mallard, The NIST Chemistry WebBook: A Chemical Data Resource on the Internet, J. Chem. Eng. Data. 46 (2001) 1059–1063.
[27] M. J. Frisch, et al. Gaussian 16, Rev. C.01, (2016) Wallingford, CT.
[28] A. Karton, B. Chan, Accurate Heats of Formation for Polycyclic Aromatic Hydrocarbons: A High-Level Ab Initio Perspective, J. Chem. Eng. Data. 66 (2021) 3453–3462.
[29] R. Tibshirani, Regression Shrinkage and Selection Via the Lasso, J. R. Stat. Soc. Ser. B Methodol. 58 (1996) 267–288.
[30] S.M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, in: Adv. Neural Inf. Process. Syst., Curran Associates, Inc., 2017.
[31] S. Maier, B. Thapa, K. Raghavachari, G4 accuracy at DFT cost: unlocking accurate redox potentials for organic molecules using systematic error cancellation, Phys. Chem. Chem. Phys. 22 (2020) 4439–4452.
[32] P.C. St. John, Y. Guan, Y. Kim, S. Kim, R.S. Paton, Prediction of organic homolytic bond dissociation enthalpies at near chemical accuracy with sub-second computational cost, Nat. Commun. 11 (2020) 2328.
[33] Y. Xuan, G. Blanquart, Numerical modeling of sooting tendencies in a laminar co-flow diffusion flame, Combust. Flame. 160 (2013) 1657–1666.
[34] Y. Sun, L. Liu, M. Li, X. Chen, F. Xu, Theoretical investigation on the mechanisms and kinetics of OH/NO_3-initiated atmospheric oxidation of vanillin and vanillic acid, Chemoshpere, Volume 288, Part 2, 2022, 132544.
[35] L. Pratali Maffei, M. Pelucchi, T. Faravelli, C. Cavallotti, Theoretical study of sensitive reactions in phenol decomposition, React Chem Eng. 5 (2020) 452–472.
[36] A.R. Ghildina, A.D. Oleinikov, V.N. Azyazov, A.M. Mebel, Reaction mechanism, rate constants, and product yields for unimolecular and H-assisted decomposition of 2,4-cyclopentadienone and oxidation of cyclopentadienyl with atomic oxygen, Combust. Flame. 183 (2017) 181–193.
[37] J. Cho, A.W. Jasper, S.J. Klippenstein, R. Sivaramakrishnan, Nonthermal Effects in the Dissociation of HOCO and Other Carbonyl-Centered Free Radicals, 13th U.S. National Combustion Meeting. (2022).
[38] D.J. Robichaud, A.M. Scheer, C. Mukarakate, T.K. Ormond, G.T. Buckingham, G.B. Ellison, M.R. Nimlos, Unimolecular thermal decomposition of dimethoxybenzenes, J. Chem. Phys. 140 (2014) 234302.
[39] J.A. Mulholland, M. Lu, D.-H. Kim, Pyrolytic growth of polycyclic aromatic hydrocarbons by cyclopentadienyl moieties, Proc. Combust. Inst. 28 (2000) 2593–2599.