Substituent Descriptors from the Topology of the Electron Density

Kevin Lefrancois-Gagnon

A thesis presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry and Materials Science

Department of Chemistry
Lakehead University
Thunder Bay, Ontario, Canada

Abstract

This work focuses on determining the suitability of functional group properties from the Quantum Theory of Atoms in Molecules (QTAIM) as descriptors for use in machine learning and linear regression analyses. These were studied as an alternative to the conventionally used proxies to substituent properties. QTAIM integrated group properties and local critical point properties were studied for their suitability of use in these applications.

The sensitivity of these properties to the choice of Density Functional Theory functional and basis set used for determining the density was examined. The properties at other levels of theory were compared to high-level B2PLYPD3-BJ/aug-cc-pV5Z data. Integrated properties are well conserved between the different model chemistries, both in an absolute sense and in linear relationships to the reference values. Local critical point properties are more sensitive to the level of theory, and hybrid functionals with triple-ζ basis sets are recommended for their evaluation.

The integrated properties differ in terms of absolute value between molecular environments. However, they are linearly well-related between substrates, with the exception of a few outliers, all of which are readily interpreted in terms of changing group geometry.

Bond critical point properties between substituent and substrate are sensitive to changing substrate. A way to calculate changing sub-

stituent electronegativity between substrates is described. By accounting for changes in substituent electronegativity between substrates, however, patterns based on atom-type will enable bond critical point properties to be included in machine learning models.

Functional group bond critical point and group property descriptors were examined for describing four currently used proxies to substituents' properties. Principal Components Analysis drastically reduces the dimensionality of the data. Partial Least Squares Regression can simultaneously model the changes to all four proxies using the QTAIM substituent descriptors. This reveals that similar information is contained in the substituent descriptors and the proxies, despite changes in solvation, substrate, and possible experimental errors in the latter.

Dedication

I dedicate this thesis to my wife, Julia, to our upcoming child, and to our pets present (Dobby, Piglet, T’Challa, Rotund and Peggy) and future. You always know how to make me laugh and support me best. That’s how well you know me. You know me in your marrow.

Statement of Contributions

The content of Chapter Three and Appendices A and E was previously published as:

K. M. Lefrancois-Gagnon and R.C. Mawhinney *Toward universal substituent constants: Model chemistry sensitivity of descriptors from the quantum theory of atoms in molecules*. Journal of Computational Chemistry. 2020; 41: 2485– 2503. doi:10.1002/jcc.26404

Author Contributions

The project was proposed by Robert Mawhinney.

KMLG performed the DFT calculations, QTAIM analysis of the data, prepared the figures and wrote the paper with input from Robert Mawhinney

The content of Chapter Four and Appendix B is in review in the Journal of Computational Chemistry

Author Contributions

The project was proposed by Robert Mawhinney.

KMLG performed the DFT calculations, QTAIM analysis of the data, prepared the figures and wrote the paper with input from Robert Mawhinney

A paper based on the content of Chapter 5 and Appendix C is in preparation at the time of writing

Author Contributions

The project was proposed by Robert Mawhinney.

KMLG performed the DFT calculations, QTAIM analysis of the data, prepared the figures and wrote the paper with input from Robert Mawhinney

A paper based on the content of Chapter 6 and Appendix D is in preparation at the time of writing

Author Contributions

The project was proposed by Robert Mawhinney.

KMLG performed the DFT calculations, QTAIM analysis of the data, prepared the figures and wrote the paper with input from Robert Mawhinney

Acknowledgements

I want to thank Dr. Robert Mawhinney, my supervisor for all his support and guidance. Thank you for the experience and knowledge you've given me in the past seven years working in your lab. Thank you for your encouragement, understanding, and kind words. Additional thanks go out to my committee members Dr. Wely Floriano and Dr. Christine Gottardo. Thank you for reviewing this thesis and for all the wonderful classes in which you've taught me over the years.

Thanks to the members of Dr. Mawhinney's research lab over the years: Steven, Niall, Krista and Ashley. It was always great to bounce ideas off of you. Additional thanks to all the staff in the Lakehead University Department of Chemistry, both professors and lab technicians. I've had a chance to work alongside many of you over the years, and I thank you for helping me in my professional development as a student and teaching assistant.

I also want to thank my parents, Dave and Tracey. You've always pushed me to be my best self, and taught me to work hard at the things I want. I appreciate your support, both in the years of my life before my PhD, and during. This would not have been possible without you. To my father and mother-in-law, Eugene and Laura, thank you for your words of encouragement, and also for Eugene's insightful questions.

I'd also like to specifically thank one of my friends and peers who studied alongside me for the majority of my academic life, Courtney

Moore. You made life in the Chemistry department at Lakehead University even more enjoyable than it was. Also, I never did get any of that 10% of your salary that you owed me.

Lastly, to my wife. I would not be where I am today without her unconditional support. Thank you for your patience while I finished my degree, and thank you for always being there for me to lean on when I need to.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The development of machine learning models in chemistry is a current interest.[1,2] These models can discover underlying patterns in data and predict the outcome of chemical reactions.[3,4] Machine learning has also been applied to the prediction of molecular properties.[5–7] The development of physically meaningful descriptors for machine learning models would aid in future applications in predicting properties of reactions.[8]

Machine learning models should incorporate the fact that chemical reactions can be affected by non-participating groups through steric and electronic interactions.[9,10] These groups, termed substituents, are a core part of chemistry. Typically, we understand substituents in terms of their size, electron-donating, and electron-withdrawing abilities.[11,12] The electron-donating and withdrawing abilities are readily interpreted by chemists based on the atomic electronegativity of atoms comprising the substituent, the presence of lone pairs, and the presence of π systems. We know intuitively that an atom with lone pairs of electrons can donate them through resonance, while an electronegative atom will withdraw electron density from the reaction site.[13] We can even intuit if a substituent can even present both, or neither of the mechanisms, that is, donating electrons through resonance, while withdrawing electrons inductively. For the desired outcome of a reaction,

such as increasing the rate or modifying the equilibrium, we can choose substituents whose effect on the reaction site is beneficial. Many examples of this exist. In silosesquioxane-supported Phillips catalysts, the presence of a functional group with lone pairs of electrons near the active site improves activity and molecular weight distribution in ethylene polymerization reactions.[14] Substituents can also affect the Rh(III) catalyzed C-H activation of 2,2'-bipyridine derivatives by weakening the Rh-N bond.[15]

Much work has been done to move beyond intuition by quantifying a substituent's ability to donate or accept electron density. Traditionally, quantification is done using parameters defined in the manner of Hammett.[16] Here, a reference reaction is chosen, and hydrogen's parameter is set to zero. Then, using this reference reaction and substituent, the parameters are derived using a linear relationship. Reference reactions can be chosen to emphasize or neglect certain modes of transmission of the substituent effect.[17]

Due to the way substituent effects are observed, it makes sense that they are typically measured using a probe.[18] We use the term probe to describe a portion of the molecule whose property is easily measured and reflects the desired effect. These probes are held constant along a series of substituted molecules. With the only change in the molecule being the substituent, the change in the probe speaks to the substituent effect.

A straightforward probe to use, commonly used since the introduction of substituent constants, is a carboxylic acid. In 1937, Hammett used the pK_a of *m*- and *p*- substituted benzoic acids to quantify the original substituent parameters, σ_m and σ_p.[16] He proposed Equation 1.1 as a tool to examine the effect and determine these parameters. K is the equilibrium constant for a substituted benzoic acid, and K_0 is the

equilibrium constant of an unsubstituted benzoic acid. The reference reaction for which ρ is unity is the dissociation of substituted benzoic acids, and σ is defined as 0 for unsubstituted benzoic acids. ρ describes the sensitivity of the reaction to the substituent effect, relative to the reference reaction.

$$\log\frac{K}{K_0} = \rho\sigma \tag{1.1}$$

Deviations from Equation 1.1 were observed and thus started a frenzy of development in substituent constants. The first deviations were observed in systems where there was a strong resonance interaction between the substituent and the probe. This could be through a resonance electron-donating substituent and a resonance electron-accepting probe, or a resonance electron-accepting substituent and a resonance electron-donating probe.[19]

Substituent constants were developed for the former case using the solvolysis of *t*-cumyl chlorides by Brown and Okamoto, Figure 1.1.[20] Here, rather than equilibrium constants, the rate of the reaction was used. The carbocation intermediate ensured that strong resonance interactions between the substituent and probe were taken into account. Substituent constants derived in this way were termed σ^+. Other scales of σ^+ were derived.[21–26] Yukawa and Tsuno expanded on this idea, noting that different reactions would have different resonance demands, so a straightforward equation such as Equation 1.2 was insufficient to capture all the information.[27] Noting that all scales of σ^+ were linearly related to each other, Yukawa and Tsuno proposed Equation 1.3, where $\Delta\sigma_R^+$ is defined in Equation 1.4. This reduces to Brown and Okamoto's model since r is 1 for their reference reaction.

$$\log\frac{k}{k_0} = \rho\sigma^+_{\text{Brown}} \tag{1.2}$$

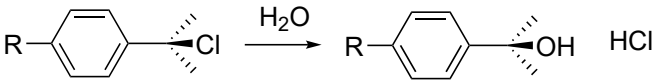

Figure 1.1: The solvolysis reaction of *t*-cumyl chlorides used to define Brown and Okamoto's σ^+

$$\log\frac{k}{k_0} = \rho\left(\sigma + r\Delta\sigma_R^+\right) \tag{1.3}$$

$$\Delta\sigma_R^+ = \sigma_{\text{Brown}}^+ - \sigma \tag{1.4}$$

These models are not sufficient for reactions with a resonance electron-donating reaction site, so further constants were developed. σ^- is defined in the same way as the original Hammett constants, but for the dissociation constant of substituted phenols, anilines, or N,N-dimethylanilines.[19] Another scale of substituent constant is the σ^0 scale, where through resonance is completely omitted. The system of interest here is derived from Hammett's benzoic acid system is taken, but with a methylene group is interposed between the carboxylic acid and the aromatic ring.[28]

Most proposed constants examine the total substituent effect. Intuitively we think of substituents in terms of multiple effects, such as the resonance, and field/inductive effects. Efforts have been made to decompose the substituent constants into these component mechanisms.[17] The field effect is often studied through the dissociation constants of 4-R-[2.2.2]-bicyclooctanecarboxylic acids,[17] or quinuclidines.[29] These constants are often termed σ'. Clearly, this system allows focusing on the field effect from its lack of π-system. Additionally, it is rigid, further allowing focus on the field effect. In these cases where field/inductive effects are easy to quantify, resonance constants are often derived by removing the inductive effect from the total substituent constant.

Swain and Lupton calculated another scale of field and resonance constants.[30] Here it was supposed that the σ' of the 4-R-[2.2.2]-bicyclooctane carboxylic acids are entirely due to field. To compensate for the lack of data points, they use the correlation shown in Equation 1.5 to evaluate a and b, the slopes. For the known data points $\mathscr{F}$ is equal to σ'. For unknown data points (where the 4-R-[2.2.2]-bicyclooctane carboxylic acid data is unavailable), Equation 1.6 is used to calculate $\mathscr{F}$. Writing σ_p as Equation 1.7 and assuming the resonance constant, $\mathscr{R}$, is 0 for NH_3^+ allows the evaluation of f. Other substituent constant scales were able to be described as linear combinations of $\mathscr{F}$ and $\mathscr{R}$. Hansch, Leo, and Taft used similar formulae to derive their F and R for their field and resonance parameters.[31]

$$\sigma' = a\sigma_m + b\sigma_p \tag{1.5}$$

$$\mathscr{F} = a\sigma_m + b\sigma_p \tag{1.6}$$

$$\sigma_p = f\mathscr{F} + r\mathscr{R} \tag{1.7}$$

The advent of computational chemistry has been extremely beneficial for substituent analysis. Substituent constants can be derived in several ways theoretically. Many cases exist of computational methods used to calculate Hammett parameters[32–34] or even define original substituent constant scales that can only be made theoretically. Hammett parameters have been estimated by the bond angles and lengths in the carboxylic acid group in substituted benzoic acids,[34] change in core-electron binding energies in fluorobenzenes,[32,33] and electrophilicity indices.[35] The true power of using computational chemistry in analyzing substituent effects arises from developing new parameters using data only readily available computationally.[36–39] Computational approaches also allow quick screening of many substituent parameters without the

need for synthesis.

Readily available properties from routine computational chemistry calculations are atomic and group charges. The substituent's charge does not correlate to the pK_a of substituted benzoic acids due to the polarity of the R-C bond. However, the inclusion of the charge of C_{ipso} with the charge of the substituent yielded good correlations, as this captured the bond's polarity. The charge of the substituent and C_{ipso} combined are now commonly referred to as the charge of the substituent active region (cSAR).[39] The acronym was formerly qSAR, but the acronym was changed after potential confusion with the acronym QSAR for quantitative structure-activity relationships. This group proposes that Hammett's constants relate to the charge transferred between the active region of the substituent and the probe's active region. Numerous studies have been completed regarding this parameter, looking at its application in physical interpretations of substituent effects.[40–47]

Other work extended the substituent effect model beyond that of just the field and resonance effects, adding in two further parameters: the electronegativity effect and polarizability effect, σ_χ and σ_α.[48] The former σ_χ is associated with the inductive effect, which is often hard to separate from the field effect. Bonds transmit the former and the latter is transmitted through space. Theoretically, the field effect can be measured by measuring an H-H bond's polarization by an H-X molecule separated from each other by 4Å. Equation 1.8 can be used to evaluate the field substituent constant. Inductive constants can be measured by the change in charge of a methyl hydrogen between substituted and unsubstituted methanes, Equation 1.9. Polarizability, quantified by σ_α, measures the sensitivity of the system to the presence of a charge. It is determined by calculated by calculating the polarization potential of

CH_4 and RCH_3 with a point charge 3Åaway and taking the difference, Equation 1.10. Finally, resonance constants, σ_R are calculated using the change in π electron populations, as described in Equation 1.11.[48]

$$\sigma_F = -35.5\Delta q_{(\mathrm{H})\alpha} \tag{1.8}$$

$$\sigma_\chi = 2\left[q_{\mathrm{H(CH4)}} - q_{H(HX)}\right] \tag{1.9}$$

$$\sigma_\alpha = \mathrm{PP}_{RCH3} - \mathrm{PP}_{CH4} \tag{1.10}$$

$$\sigma_R = 0.0047\Delta q_\pi + 0.019 \tag{1.11}$$

Natural Bond Orbital Analysis decomposes atomic populations into different orbitals. Examining σ and π orbital populations also lends insight into the substituent effect. σ and π electron donor-acceptor properties (sEDA and pEDA, respectively) can be calculated.[38] For a simple system like substituted methanes, sEDA is calculated as Equation 1.12, where $\sigma_{\mathrm{CH_4}}$ is the σ electron population of the carbon in methane. A similar descriptor is available in benzene rings, where the populations of all atoms in the ring are instead taken into account. For benzene rings, the σ population is defined as valence s, p_x and p_y orbitals, and the π population by p_z. The sEDA and pEDA are then given by Equations 1.13 and 1.14 respectively. sEDA correlates with group electronegativities per Boyd, Boyd, and Edgecombe, while pEDA correlates with σ_R. Additionally, pEDA descriptors calculated in varying π systems are correlated to each other. The descriptor has been extended to apply to heterocyclic systems as well.[49] This was applied in a variety of cases.[50–53]

$$\mathrm{sEDA(CH_4)} = \sigma_{\mathrm{R\text{-}CH_3}} - \sigma_{\mathrm{CH_4}} \tag{1.12}$$

$$\text{sEDA} = \sum_{j=1}^{6} \sigma^{j}_{\text{R-C}_6\text{H}_5} - \sum_{j=1}^{6} \sigma_{\text{C}_6\text{H}_6} \tag{1.13}$$

$$\text{pEDA} = \sum_{j=1}^{6} \pi^{j}_{\text{R-C}_6\text{H}_5} - \sum_{j=1}^{6} \pi_{\text{C}_6\text{H}_6} \tag{1.14}$$

Substituent effects have also been evaluated using the molecular electrostatic potential, Equation 1.15.[36] Originally, it was common to find critical points in $V(\mathbf{r})$, and use those minima to evaluate substituent effects, but an easier approach was developed.[54] Instead of calculating critical points, $V(\mathbf{r})$ at nuclei in aromatic systems was sufficient to quantify the substituent effect. $V(\mathbf{r})$ should reflect changes in the density at the nuclei, which is why it would be good as a substituent descriptor.

$$V(\mathbf{r}) = \sum_{A} \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} - \int \frac{\rho(\mathbf{r}')}{\mathbf{r} - \mathbf{r}'} d\mathbf{r}' \tag{1.15}$$

Quantum chemical topology, as embodied through the Quantum Theory of Atoms in Molecules (QTAIM), is an excellent tool to describe substituents.[55] The theory's details will be laid out in Chapter 2, but the following summary is sufficient for understanding the literature. QTAIM defines atoms as basins in the electron density, each containing an attractor (the nucleus). Zero-flux surfaces separate them. Critical points, where $\nabla\rho = 0$, are calculated and those along bond paths are termed bond critical points. Atomic properties calculated by integrating a property over the basin and property densities evaluated at the bond critical points can determine substituent properties.

An electronegativity scale derived from QTAIM parameters can define the electronegativity of atoms and groups.[56,57] Determined based on periodic trends, an electronegativity factor is derived, as in Equation 1.16, where r_H is the distance from the hydrogen atom to the BCP of the R-H bond, r_{AH} is the AH bond length, N_A is the valence of atom A

and $\rho(\mathbf{r}_c)$ is the electron density at the bond critical point. The Pauling electronegativity, χ is then determined from this using Equation 1.17, where a and b are determined by setting the electronegativity of fluorine and lithium atoms to 4 and 1, respectively.

$$F_A = \frac{r_H}{r_{AH}\rho(\mathbf{r}_c)N_A} \tag{1.16}$$

$$\chi_A = aF_A^b \tag{1.17}$$

One theoretical analysis of substituent effects also uses BCP properties. Popelier defines a Quantum Topological Molecular Similarity (QTMS) index based on the difference in BCP space.[37,58–61] In an aromatic system, properties at the six aromatic C-C BCPs are evaluated. The difference between them defines the similarity between the molecules, indicating the effect a substituent has on the aromatic system. The difference is defined by Equation 1.18, where $P_{\mathrm{b},j}$ is BCP property k at BCP j. The sum is taken over 9 possible properties, increased from the three of the original paper.[37] The properties are the Laplacian of the electron density, $\nabla^2\rho_\mathrm{b}$; the electron density, ρ_b; eigenvalues of the second derivative of the electron density, $\lambda_{i,\mathrm{b}}$ (i = 1,2,3); potential energy density, V_b; kinetic energy densities, G_b and K_b; ellipticity at the BCP, ε_b; and equilibrium bond length $\mathrm{R_e}$. The distances between individual BCPs are summed to give the similarity descriptor for the molecule, Equation 1.19.

$$d_{ij} = \sqrt{\sum_{i,j}\left(P^k_{\mathrm{b},i} - P^k_{\mathrm{b},j}\right)^2} \tag{1.18}$$

$$d(\mathrm{A,B}) = \sum_{i\in\mathrm{A}}\sum_{j\in\mathrm{B}} d_{ij} \tag{1.19}$$

All of the above substituent parameters have in common that they do not quantify the intrinsic properties of the substituent; they all reference some other part of the molecule to define them. Experimentally, rates and equilibria see the substituent effect and are a useful way to quantify it.[16] Computationally, the effect of the substituent, often on the aromatic ring, is used.[36,37,50] These properties are excellent tools, but it is the effect of the substituents elsewhere being measured in all cases. QTAIM, as mentioned before, can calculate atomic properties that sum to molecular properties.[55] Similarly, by only summing properties in a substituent, intrinsic substituent properties can be measured.

Intrinsic substituent properties from QTAIM have been used to study substituent effects before. Smith used QTAIM substituent dipole moments to describe the field effect on the dissociation of 4-R-[1.1.1]-bicyclopentanoic acids.[62] The dipole in RH molecules was correlated to the dipole in RCH_3 and 4-R-[1.1.1]-bicyclopentane. The aforementioned QTMS descriptors by Popelier are based on QTAIM.[37] The difference between r_H in a substituted and unsubstituted system and the stress tensor polarizability, were found to correlate to experimental Hammett constants.[63] The basicity of substituted anilines was interpreted using BCP properties, Laplacian critical points, N atom population, and delocalization indices.[64] BCP properties were also used to examine substituent effects in ethanes.[65]

This thesis sets out to prove the applicability of integrated and local QTAIM descriptors in small molecules as part of feature sets to QSPR and machine learning studies. To do this, we examine the sensitivity to choice of model chemistry in calculating the descriptors, assess their transferability between different molecular environments, and prove that the information contained in the descriptors is physically relevant. Given an insensitivity in model chemistry, descriptors from different

model chemistries can be included in the same model without compromising accuracy. Transferability between molecular environments limits the need to recalculate the descriptors, allowing one set of descriptors to be used in many analyses. Lastly, the physical relevance ensures that these descriptors are meaningful to include in such a model. This work provides an avenue for intrinsic substituent descriptors to be used in analyses predicting reactivity. These descriptors are easily calculable, physically meaningful, and applicable in numerous situations.

This thesis is presented in chapters to address these key sticking points in calculating substituent properties. Chapter 2 lays out the theoretical framework in which we are working, covering the theoretical background of Density Functional Theory (DFT) and QTAIM. Chapter 3 addresses these descriptors' model chemistry sensitivity, comparing numerous functionals and basis sets to reference values. Chapter 4 addresses the transferability of group descriptors between substrate series common to chemistry. Chapter 5 analyses patterns in bond critical point property changes for inclusion in machine learning models. Chapter 6 compares substituent properties computed using a reference-level DFT method (double-hybrid functional with large basis set) to some of the descriptors we see day to day.

References

[1] Kayala, M. A. and Baldi, P., *J. Chem. Inf. Model.* **2012**, *52*, 2526–2540.

[2] Meuwly, M., *Chem. Rev.* **2021**

[3] Carrera, G., Gupta, S. and Aires-de Sousa, J., *J. Comput. Aided. Mol. Des.* **2009**, *23*, 419–429.

[4] Coley, C. W., Barzilay, R., Jaakkola, T. S., Green, W. H. and Jensen, K. F., *ACS Central Sci.* **2017**, *3*, 434–443.

[5] Browning, N. J., Ramakrishnan, R., von Lilienfeld, O. A. and Roethlisberger, U., *J. Phys. Chem. Lett.* **2017**, *8*, 1351–1359.

[6] Gubaev, K., Podryabinkin, E. V. and Shapeev, A. V., *J. Chem. Phys.* **2018**, *148*, 241727.

[7] Hansen, K., Biegler, F., Ramakrishnan, R., Pronobis, W., von Lilienfeld, O. A., Müller, K.-R. and Tkatchenko, A., *J. Phys. Chem. Lett.* **2015**, *6*, 2326–2331.

[8] Beker, W., Gajewska, E. P., Badowski, T. and Grzybowski, B. A., *Angew. Chem. Int. Ed.* **2019**, *58*, 4515–4519.

[9] Devi, S., Saraswat, M., Grewal, S. and Venkataramani, S., *J. Org. Chem.* **2018**, *83*, 4307–4322.

[10] Golas, P. L., Tsarevsky, N. V. and Matyjaszewski, K., *Macromol. Rapid Commun.* **2008**, *29*, 1167–1171.

[11] Förster, H. and Vögtle, F., *Angew. Chem. Int. Ed.* **1977**, *16*, 429–441.

[12] Miljković, M., Yeagley, D., Deslongchamps, P. and Dory, Y. L., *J. Org. Chem.* **1997**, *62*, 7597–7604.

[13] Domenicano, A. and Murray-Rust, P., *Tetrahedron Lett.* **1979**, *20*, 2283–2286.

[14] Baba, R., Thakur, A., Chammingkwan, P., Terano, M. and Taniike, T., *Dalton Trans.* **2017**, *46*, 12158–12166.

[15] Wu, S., Wang, Z., Bao, Y., Chen, C., Liu, K. and Zhu, B., *Chem. Comm.* **2020**, *56*, 4408–4411.

[16] Hammett, L. P., *J. Am. Chem. Soc.* **1937**, *59*, 96–103.

[17] Holtz, H. D. and Stock, L. M., *J. Am. Chem. Soc.* **1964**, *86*, 5188–5194.

[18] Taft, R. W. and Topsom, R. D., The Nature and Analysis of Substituent Electronic Effects, John Wiley & Sons, Ltd **1987**, pp. 1–83.

[19] Fickling, M. M., Fischer, A., Mann, B. R., Packer, J. and Vaughan, J., *J. Am. Chem. Soc.* **1959**, *81*, 4226–4230.

[20] Brown, H. C. and Okamoto, Y., *J. Am. Chem. Soc.* **1958**, *80*, 4979–4987.

[21] Pearson, D. E., Baxter, J. F. and Martin, J. C., *J. Org. Chem.* **1952**, *17*, 1511–1518.

[22] Kochi, J. K. and Hammond, G. S., *J. Am. Chem. Soc.* **1953**, *75*, 3445–3451.

[23] Deno, N. C. and Schriesheim, A., *J. Am. Chem. Soc.* **1955**, *77*, 3051–3054.

[24] Deno, N. C. and Evans, W. L., *J. Am. Chem. Soc.* **1957**, *79*, 5804–5807.

[25] Miller, J., *Aust. J. Chem.* **1956**, *9*, 61.

[26] Packer, J., Vaughan, J. and Wilson, A., *J. Org. Chem.* **1958**, *23*, 1215.

[27] Yukawa, Y. and Tsuno, Y., *Bull. Chem. Soc. Jpn.* **1959**, *32*, 965–971.

[28] Taft, R. W., *J. Phys. Chem.* **1960**, *64*, 1805–1815.

[29] Grob, C. A. and Schlageter, M. G., *Helv. Chim. Acta* **1976**, *59*, 264–276.

[30] Swain, C. G. and Lupton, E. C., *J. Am. Chem. Soc.* **1968**, *90*, 4328–4337.

[31] Hansch, C., Leo, A., and Taft, R. W.„ *Chem. Rev.* **1991**, *91*, 165–195.

[32] Takahata, Y. and Chong, D. P., *Int. J. Quantum Chem.* **2005**, *103*, 509–515.

[33] Segala, M., Takahata, Y. and Chong, D. P., *J. of Mol. Struct.: THEOCHEM* **2006**, *758*, 61–69.

[34] Gross, K. C., Seybold, P. G., Peralta-Inga, Z., Murray, J. S. and Politzer, P., *J. Org. Chem.* **2001**, *66*, 6919–6925.

[35] Meneses, L., Araya, A., Pilaquinga, F. and Fuentealba, P., *Chem. Phys. Lett.* **2008**, *460*, 27–30.

[36] Gadre, S. R. and Suresh, C. H., *J. Org. Chem.* **1997**, *62*, 2625–2627.

[37] Popelier, P. L. A., *J. Phys. Chem. A* **1999**, *103*, 2883–2890.

[38] Ozimiński, W. P. and Dobrowolski, J. C., *J. Phys. Org. Chem.* **2009**, *22*, 769–778.

[39] Sadlej-Sosnowska, N., *Chem. Phys. Lett.* **2007**, *447*, 192–196.

[40] Krygowski, T. M. and Sadlej-Sosnowska, N., *Struct. Chem.* **2011**, *22*, 17–22.

[41] Siodla, T., Oziminski, W. P., Hoffmann, M., Koroniak, H. and Krygowski, T. M., *J. Org. Chem.* **2014**, *79*, 7321–7331.

[42] Szatylowicz, H., Siodla, T. and Krygowski, T. M., *J. Phys. Org. Chem.* **2017**, *30*, e3694.

[43] Szatylowicz, H., Jezuita, A., Ejsmont, K. and Krygowski, T. M., *Struct. Chem.* **2017**, *28*, 1125–1132.

[44] Zborowski, K. K., Szatylowicz, H., Stasyuk, O. A. and Krygowski, T. M., *Struct. Chem.* **2017**, *28*, 1223–1227.

[45] Varaksin, K. S., Szatylowicz, H. and Krygowski, T. M., *J. Mol. Struct.* **2017**, *1137*, 581–588.

[46] Szatylowicz, H., Jezuita, A., Siodla, T., Varaksin, K. S., Ejsmont, K., Shahamirian, M. and Krygowski, T. M., *Struct. Chem.* **2018**, *29*, 1201–1212.

[47] Szatylowicz, H., Domanski, M. A. and Krygowski, T. M., *ChemistryOpen* **2019**, *8*, 64–73.

[48] Topsom, R. D., Some Theoretical Studies of Electronic Substituent Effects in Organic Chemistry, John Wiley & Sons, Ltd **1987**, pp. 125–191.

[49] Mazurek, A. and Dobrowolski, J. C., *J. Org. Chem.* **2012**, *77*, 2608–2618.

[50] Oziminski, W. P., Krygowski, T. M., Fowler, P. W. and Soncini, A., *Org. Lett.* **2010**, *12*, 4880–4883.

[51] Rode, M. F. and Sobolewski, A. L., *J. Phys. Chem. A* **2010**, *114*, 11879–11889.

[52] Oziminski, W. and Krygowski, T., *Tetrahedron* **2011**, *67*, 6316–6321.

[53] Krygowski, T. M., Oziminski, W. P. and Cyrański, M. K., *J. Mol. Model.* **2012**, *18*, 2453–2460.

[54] Galabov, B., Ilieva, S. and Schaefer, H. F., *J. Org. Chem.* **2006**, *71*, 6382–6387.

[55] Matta, C F; Boyd, R. J., The Quantum Theory of Atoms in Molecules, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim **2007**.

[56] Boyd, R. J. and Edgecombe, K. E., *J. Am. Chem. Soc.* **1988**, *110*, 4182–4186.

[57] Boyd, R. J. and Boyd, S. L., *J. Am. Chem. Soc.* **1992**, *114*, 1652–1655.

[58] O'Brien, S. E. and Popelier, P. L., *Can. J. Chem.* **1999**, *77*, 28–36.

[59] O'Brien, S. E. and Popelier, P. L. A., *J. Chem. Inform. Comput. Sci.* **2001**, *41*, 764–775.

[60] O'Brien, S. E. and Popelier, P. L. A., *J. Chem. Soc., Perkin Trans. 2* **2002**, *2*, 478–483.

[61] Popelier, P. L. A., Chaudry, U. A. and Smith, P. J., *J. Chem. Soc., Perkin Trans. 2* **2002**, *7*, 1231–1237.

[62] Smith, A. P., McKercher, A. E. and Mawhinney, R. C., *J. Phys. Chem.. A* **2011**, *115*, 12544–12554.

[63] Jiajun, D., Maza, J. R., Xu, Y., Xu, T., Momen, R., Kirk, S. R. and Jenkins, S., *J. Comput. Chem.* **2016**, *37*, 2508–2517.

[64] Graña, A. M., Hermida-Ramón, J. M. and Mosquera, R. A., *Chem. Phys. Lett.* **2005**, *412*, 106–109.

[65] Grabowski, S. J., Krygowski, T. M. and Leszczynski, J., *J. Phys. Chem. A* **2009**, *113*, 1105–1110.

[66] Charton, M., *J. Org. Chem.* **1961**, *26*, 735–738.

[67] Charton, M., *J. Org. Chem.* **1965**, *30*, 969–973.

Chapter 2

Theoretical Framework

2.1 Introduction

This thesis uses electronic structure computational procedures to calculate molecular geometries and resulting densities. It is from these determined densities that the substituent properties are calculated. The primary electronic structure method used is Density Functional Theory (DFT). This chapter will commence with a background of electronic structure method development before DFT. An explanation of DFT will follow this. The post-analysis method of choice, QTAIM, is then described in detail and is followed by an overview of the necessary mathematics for the analysis.

2.2 The Schrodinger Equation

The time-independent eigenvalue equation proposed by Schrodinger, Equation 2.1 is the starting point of computational chemical analyses, where $\hat{H}$ is the Hamiltonian operator, $\psi(\mathbf{r}, \mathbf{R})$ is the wavefunction of the system, $\mathbf{r}$ is the coordinates of the electrons, and $\mathbf{R}$ is the coordinates of the nuclei. The Hamiltonian is the sum of the kinetic energy and po-

tential energy operators of the system and can be written for a system of N_{elec} electrons and M nuclei as in Equation 2.2.[1]

$$\hat{H}\psi(\mathbf{r},\mathbf{R}) = E\psi(\mathbf{r},\mathbf{R}) \tag{2.1}$$

$$\hat{H} = \sum_{A=1}^{M} -\frac{1}{2}\nabla_A^2 + \sum_{A=1}^{M}\sum_{B>A}\frac{Z_A Z_B}{r_{AB}} + \sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N_{elec}}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N_{elec}}\sum_{j=1}^{N_{elec}}\frac{1}{r_{ij}} \tag{2.2}$$

This is an incredibly useful equation, as it provides a route to the wavefunction, ψ, containing all information about a given system. For ease of solving this equation, we approximate that the nuclei are stationary relative to the movement of the electrons; this applies to most chemical systems of interest. In terms of Equation 2.2, this is manifested as, $\nabla_A^2 = 0$. Therefore, for a given permutation of the nuclei, the nuclear-nuclear repulsion is a constant that can be added at the last step of a calculation. This leads to the electronic Schrodinger equation, Equation 2.3.[1]

$$\left(\sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N_{elec}}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N_{elec}}\sum_{j=1}^{N_{elec}}\frac{1}{r_{ij}}\right)\psi(\mathbf{r}) = \hat{H}_{\text{elec}}\psi(\mathbf{r}) = E_{\text{elec}}\psi(\mathbf{r}) \tag{2.3}$$

Equation 2.3 is now solvable. Restrictions need to be placed on the solution. ψ must be normalized so that $\int \psi(\mathbf{r})^*\psi(\mathbf{r})d\mathbf{r} = N$, the number of electrons in the system. Additionally, ψ must be antisymmetric for the exchange of electrons. On top of these restrictions, Equation 2.3 can only be solved exactly for one-electron systems. Due to the electron-electron repulsion term, systems with two or more electrons can not be exactly solved. The integrals that would be needed to do this are impossible due to the three-body problem. Approximations

can be made to enable a solution, however.[1]

2.3 Hartree-Fock Theory

A useful first approximation is the Hartree-Fock approximation. This approximation assumes the form of the wavefunction to be that of a single Slater determinant of one-electron orbitals, Equation 2.4. This form ensures the wavefunction is antisymmetric and normalized. Physically, this approximation relates to taking the electron-electron interactions into account in an average sense. Equation 2.4 has energy as written in Equation 2.5, where h_i is a one-electron term, J_{ij} is a two-electron Coulomb term and K_{ij} is an exchange term, defined in Equations 2.6, 2.7 and 2.8, respectively. These equations also defined their respective operators, $\hat{h}$, $\hat{J}$, and $\hat{K}$.[1]

$$\Psi(\mathbf{r}) = \frac{1}{\sqrt{N!}} \left| \begin{bmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) & \dots & \phi_N(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) & \dots & \phi_N(\mathbf{r}_2) \\ \vdots & \vdots & & \vdots \\ \phi_1(\mathbf{r}_N) & \phi_2(\mathbf{r}_N) & \dots & \phi_N(\mathbf{r}_N) \end{bmatrix} \right| \tag{2.4}$$

$$\begin{aligned} E_{\mathrm{HF}} &= \langle \Psi | \mathbf{H}_{\mathrm{elec}} | \Psi \rangle \\ &= \sum_{i=1}^{N} h_i + \sum_{i} \sum_{j>i} (J_{ij} - K_{ij}) + V_{nn} \end{aligned} \tag{2.5}$$

$$h_i = \int \phi_i^*(\mathbf{r}_1) \left(-\frac{1}{2}\nabla_1^2 - \sum_A^M \frac{Z_A}{r_{1A}} \right) \phi_i(\mathbf{r}_1) d\mathbf{r}_1 = \int \phi_i^*(\mathbf{r}_1) \hat{h}_1 \phi_i(\mathbf{r}_1) d\mathbf{r}_1 \tag{2.6}$$

$$J_{ij} = \int \int \frac{\phi_i^*(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2)\phi_j(\mathbf{r}_2)\phi_i(\mathbf{r}_1)}{r_{12}} d\mathbf{r}_2 d\mathbf{r}_1 = \int \phi_i^*(\mathbf{r}_1)\hat{J}_j\phi_i(\mathbf{r}_1)d\mathbf{r}_1 \quad (2.7)$$

$$K_{ij} = \int \int \frac{\phi_i^*(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2)\phi_i(\mathbf{r}_2)\phi_j(\mathbf{r}_1)}{r_{12}} d\mathbf{r}_2 d\mathbf{r}_1 = \int \phi_i^*(\mathbf{r}_1)\hat{K}_j\phi_i(\mathbf{r}_1)d\mathbf{r}_1 \quad (2.8)$$

For Slater determinant wavefunctions, minimizing the energy using a set of Lagrangian multipliers leads to the Hartree-Fock equation, Equation 2.9. $\hat{F}$ is the Fock operator, defined in Equation 2.10 as being composed of the one-electron, Coulomb, and exchange operators. The final Equation 2.9 is for a diagonalized ε.[1]

$$\hat{F}_i\phi_i' = \varepsilon_i\phi_i' \quad (2.9)$$

$$\hat{F}_i = \hat{h}_i + \sum_j^N (\hat{J}_j - \hat{K}_j) \quad (2.10)$$

These Hartree-Fock equations can rewrite the energy in terms of the orbital energy. First, note the orbital energy by multiplying Equation 2.9 by ϕ_i^* and integrating, Equation 2.11. Using these definitions, Equation 2.5 can be rewritten, Equation 2.12.[1]

$$\varepsilon_i = \int \phi_i^* \hat{F}_i \phi_i d\mathbf{r} \quad (2.11)$$

$$E_{\text{HF}} = \sum_i^N \varepsilon_i - \frac{1}{2}\sum_{ij}^N (J_{ij} - K_{ij}) \quad (2.12)$$

A final note is that for practical solutions, the orbitals must be written in terms of a basis set, Equation 2.13. Substituting this into Equation 2.9 gives the Roothan-Hall equations, Equation 2.14, where **F** is the Fock matrix, **S** is the overlap matrix, **C** is the basis set coefficient matrix, and $\boldsymbol{\varepsilon}$ is the orbital energy matrix. Basis sets will be described

later in this chapter.[1]

$$\phi_i' = \sum_{\alpha}^{M_{\text{basis}}} c_{\alpha_i} \chi_\alpha \tag{2.13}$$

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \tag{2.14}$$

In taking electron-electron interactions into account in an average way, Hartree-Fock makes a large approximation. The difference between the true energy and the Hartree-Fock energy is termed the correlation energy. Other wavefunction methods systematically improve on Hartree-Fock and approach the exact solution, but they are computationally expensive. As these are not applied here, they will not be covered in detail.[1]

2.4 Density Functional Theory

It is desirable to improve on the Hartree-Fock method while not incurring a great increase in computational cost. In search of this, another family of electronic structure methods was proposed, termed Density Functional Theory (DFT). In this, it is said that the energy is a functional of the density. Given a density, ρ, the energy functional would determine the energy of the molecular system. Originally, DFT was proposed as an orbital-free method, requiring only $3N$ spatial coordinates. However, orbital-free DFT proved problematic to apply, and an orbital-based method was later introduced.[1]

$$E = E[\rho] \tag{2.15}$$

When Hohenberg and Kohn proposed DFT, they proposed and proved two theorems. The first theorem states a one-to-one mapping between the external potential and the electron density exists, within an additive

constant. The second theorem states that the energy determined from DFT is variational. The exact density will produce lower energy than an approximate density. The energy in DFT can be stated as Equation 2.16. $T[\rho]$ is the kinetic energy, $V_{ee}[\rho]$ is the electron-electron repulsion energy, and $V_{ne}[\rho]$ is electron-nucleus interaction energy, which is known given an electron density ρ and an external potential arising from nuclear point charges, $v(\mathbf{r})$.[2]

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{ne}[\rho] = F_{HK}[\rho] + \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} \tag{2.16}$$

No orbitals are invoked in this formulation, and there are only $3N$ spatial coordinates. Practically, this formulation has been difficult due to expensive expansions of kinetic energy and electron-electron energy. Kohn and Sham proposed an alternate DFT formulation that invokes orbitals, now termed Kohn-Sham DFT (KS-DFT). They utilize a hypothetical system with non-interacting electrons with the same density as the actual system with interacting electrons. This non-interacting system has a wavefunction in the form of a Slater determinant, with kinetic energy exactly defined in Equation 2.17. Approximating the kinetic energy in the real system with this form, and approximating the electron-electron repulsion as the Coulomb repulsion $J[\rho]$, the energy is written as Equation 2.18. $E_{xc}[\rho]$ is the exchange-correlation energy defined in Equation 2.19, gathering the approximations to the kinetic energy and electron-electron energy into one term.[3]

$$T_S = \sum_{i=1}^{N} \langle \phi_i | -\frac{1}{2}\nabla_i^2 | \phi_i \rangle \tag{2.17}$$

$$E[\rho] = T_S[\rho]] + J[\rho] + E_{xc}[\rho] + V_{ne}[\rho]] \tag{2.18}$$

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (V_{ee}[\rho] - J[\rho]) \tag{2.19}$$

The goal of DFT method development is finding good approximations to $E_{xc}[\rho]$. Theoretically, the exact $E_{xc}[\rho]$ exists, and if it could be found, DFT would be an exact method. This is the crux of DFT method development: defining new approximations to $E_{xc}[\rho]$. Often, $E_{xc}[\rho]$ is represented as an additive sum of the exchange and correlation energies, Equation 2.20. ε_x and ε_c are the exchange and correlation energies per particle, respectively. Functionals are grouped into families of similar functional types, based on the approximations used in their definitions. These families are now reviewed on an individual basis.

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho] = \int d\mathbf{r}\rho(\mathbf{r})\varepsilon_x(\rho) + \int d\mathbf{r}\rho(\mathbf{r})\varepsilon_c(\rho) \tag{2.20}$$

2.4.1 Local Spin-Density Approximation

The simplest approximation is treating the density locally as a uniform electron gas, that is, assuming that the density is slowly varying. For this type of system, the exchange energy-per-particle is well known, Equation 2.21. The correlation energy however, is not simply expressed and are determined from a fit of Monte Carlo simulations of the uniform electron gas.[4]

$$\varepsilon_x(\rho) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho(\mathbf{r})^{1/3} \tag{2.21}$$

Numerous issues plague this type of functional in molecular systems due to the erroneous assumption upon which it is based (that of a slowly varying density). The exchange energy is consistently underestimated, the electron correlation is overestimated, and bond strengths are over-

estimated. As such, to properly describe chemical systems, some improvements need to be made.[1]

2.4.2 Generalized Gradient Approxmation

A first improvement to the uniform electron gas approximation is to allow the density to vary more quickly. This is first done by incorporating the first derivative of the density, $\nabla\rho(\mathbf{r})$, into the formulae for exchange and correlation energies. The Fermi and Coulomb holes are further required to integrate to their correct values (-1 and 0, respectively). Functionals that include these criteria are termed Generalized Gradient Approximation (GGA) functionals. These further additions to the uniform electron gas allow the study of molecular systems with more accuracy.[1]

Often these functionals are defined by completely separate exchange and correlation parts. That is, they are derived separately and combined. An example of this is the BLYP functional, which combines B88 exchange[5] with LYP correlation.[6] Taken together, these define the BLYP functional. However, these exchange and correlation parts are not only defined in the context of the other. One might observe B88 exchange being used in combination with different correlation functionals, or LYP correlation being used with different exchange functionals. A different example is the PBE functional, where the exchange and correlation parts were derived in the same work and are typically only applied together.[7]

2.4.3 *meta*-Generalized Gradient Approximation

The next logical step after including the first derivative is to include the second derivative of the density, $\nabla^2\rho(\mathbf{r})$. The kinetic energy density,

τ, related to $\nabla^2\rho$, contains similar information and can also be incorporated here in its place. An example here is the TPSS functional.[8]

2.4.4 Hybrid Methods

It was proposed that the exact exchange from Hartree-Fock theory plays an important part in accurate density functionals. This was proposed because the $\lambda = 0$ limit includes exact exchange in the adiabatic connection formula. Therefore, mixing in some part of exact exchange even after we "turn on" electron-electron interactions makes sense to improve accuracy.[9] Becke introduced a three-parameter model, generally given in Equation 2.22. a is the amount of exact exchange to incorporate (and the amount of LSDA exchange to discount), b is the amount of exchange from the GGA to incorporate, and c is how much GGA correlation to include (and how much LSDA correlation to discount). A common example is B3LYP, where the GGA exchange is from the B88 functional, and the GGA correlation is from the LYP functional. For B3LYP, the fitting parameters a, b, and c were determined by fitting to experimental data.[10] The PBE GGA functional has also been hybridised into PBE0 following Equation 2.23. Here 25% exact exchange is taken, and 25% of PBE exchange is discounted.[11]

$$E_{\mathrm{XC}} = (1-a)E_{\mathrm{X}}^{\mathrm{LSDA}} + aE_x^{\mathrm{exact}} + b\Delta E_x^{\mathrm{GGA}} + (1-c)E_{\mathrm{C}}^{\mathrm{LSDA}} + cE_c^{\mathrm{GGA}} \quad (2.22)$$

$$E_{\mathrm{xc}}^{\mathrm{PBE0}} = E_{\mathrm{xc}}^{\mathrm{PBE}} + \frac{1}{4}\left(E_x^{\mathrm{HF}} - E_x^{\mathrm{PBE}}\right) \quad (2.23)$$

2.5 Double-Hybrid Methods

A further increase in accuracy is possible by taking into account correlation using the Kohn-Sham orbitals in a similar manner to second-order perturbation theory, Equation 2.25. The original example is the B2PLYP functional of Grimme. This incorporates a mixture of GGA and HF exchange, as well as GGA and second-order perturbation theory correlation, Equation 2.25. As indicated by the name, the exchange GGA functional is the B88 functional, and the correlation functional is the LYP functional, both previously described. This has been extensively benchmarked and shown to be an accurate functional.[12]

$$E_c^{\mathrm{PT2}} = \frac{1}{4}\sum_{ia}\sum_{jb}\frac{[(ia|jb)-(ib|ja)]^2}{\varepsilon_i+\varepsilon_j-\varepsilon_a-\varepsilon_b} \tag{2.24}$$

$$E_{xc} = (1-a)E_x^{\mathrm{GGA}} + aE_x^{\mathrm{HF}} + bE_c^{\mathrm{GGA}} + cE_c^{\mathrm{PT2}} \tag{2.25}$$

2.6 Basis

The KS-DFT mentioned above and HF methods both utilize orbitals in their computational procedures. A specific tool is needed for how integrals involving these orbitals are calculated and described. That tool is the basis set. We write the orbital as a linear sum of basis functions, χ. χ can be either Slater, Equation 2.26 or Gaussian, Equation 2.27 basis functions. The exponential terms of Slater functions are to the order of $-\zeta r$, while Gaussian functions are to the order of -ζr^2. Slater basis functions more correctly model the electron density near the nucleus as a cusp, as opposed to the Gaussian basis functions, whose slope is zero at the nucleus. The integrals are difficult to calculate for Slater orbitals, but Gaussian orbitals circumvent that difficulty. Often, the inaccurate

behaviour of Gaussian orbitals is resolved by writing Slater orbitals as a combination of multiple Gaussian orbitals.[1]

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{n-1}e^{-\zeta r} \tag{2.26}$$

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{2n-2-l}e^{-\zeta r^2} \tag{2.27}$$

Basis sets are classified based on the number and type of functions included in the basis set. The smallest basis sets are minimum basis sets with only just enough functions for the electrons in the neutral atoms. The size increases from there to double zet (DZ), triple zet (TZ), and quadruple zet (QZ) basis sets which contain twice, three times, and four times the number of functions needed for the electrons in a neutral atom. The size can further increase to pentuple and beyond. As the most important electrons in chemistry are in the valence electrons, the basis set is often split into the core electrons and the valence electrons. The core electrons are represented by one contracted Gaussian type orbital and the valence by multiple functions.[1]

Aside from size, there are two other standard ways to modify a basis: including diffuse functions or polarization functions. Diffuse functions have a longer tail, allowing the electrons to be found further from the nucleus. These are often essential for the description of anions. Polarization functions are functions of higher polarization than currently included. For example, d-orbitals are not included for carbon atoms in an unpolarized basis set for carbon, but they may often be added as polarization functions.[1]

Various types of basis set are used in this work. First explained are the Pople style basis sets. These can be abbreviated *a*-*bc*G, or *a*-*bcd*G, where a, b, c, and d are positive integers. These are split valence basis

sets and these designations mean that the core electron orbitals are represented by contracted functions comprised of *a* Gaussian type orbitals, and the valence valence orbitals are written with two contracted functions composed of *b* and *c* Gaussian functions. The *a*-*bcd*G type functions instead have the valence split into three contracted functions, composed of *a*, *b*, and *c* contracted orbitals. Diffuse functions are designated as one or two '+' before the G; one plus means diffuse functions on heavy atoms, and two means diffuse functions on heavy atoms and hydrogen. Polarization functions are specified using terms in brackets. (*n*d*m*f,*k*p*l*d) means *n* additional *d* functions and *m* additional *f* functions on heavy atoms, and *k* additional *p* functions and *l* additional d functions on hydrogen.[13–24]

Other types of basis include the correlation consistent basis sets, designated cc-pVXZ, where X is a letter or number signifying the size of the basis sets. These basis sets were designed so that each function introduced at a given stage contributes a similar amount of energy. For example, the 2s and 1p orbitals are included simultaneously and would have similar energy lowering. These basis sets include polarization by default, and diffuse functions are included by specifying the aug- prefix.[25–31]

Jensen defined polarization-consistent basis sets which are optimized for DFT type calculations. Aside from specifically being optimized for DFT, these are similar in concept to correlation consistent. They are named pc-n, where n is the amount of polarization in the basis set. Diffuse functions are also specified here by the aug-prefix.[32–39]

A final basis set type discussed here are the def2 basis sets. These were developed to calculate accurate response properties - specifically the polarizability - for large systems at a minimal computational cost. This was done by minimizing the number of diffuse functions in the

basis set. These are typically referred to as def2-ζZVPP, where ζ is a letter indicating basis set size. The VPP stands for "Valence Plus Polarization".[40]

2.7 QTAIM

While DFT itself provides a way to calculate the density and corresponding energy of a system, chemistry is absent from its output. Chemistry lies in atoms and bonding interactions that make up the molecule which are undefined in raw DFT output. DFT methods result in molecular energy and electron density. Post-analysis methods determine the properties, atoms, and bonds of the system. One such post-analysis method is The Quantum Theory of Atoms in Molecule (QTAIM), which provides a way to define atoms and bonding interactions in a sensible manner, derived from first principles.[41,42]

QTAIM defines structural features in chemistry in a way that is easily seen qualitatively in the first derivative of the electron density, $\nabla\rho$. Atoms are easy to discern in the gradient paths $\nabla\rho$. There is a surface between the atoms over which the gradient lines do not pass. This surface partitions the density into different basins. Atoms are then defined as the union of the electron density and nucleus contained in a basin. An example is in the gradient vector paths of N_2, shown in Figure 2.1. Here, two basins are seen, with a surface separating the two atoms.[41,42]

QTAIM goes beyond this qualitative picture of the atoms in molecules and shows that they are open quantum subsystems. They have an atomic virial theorem and subsystem variation principle. The quantum mechanics of the system is rigorously defined. Therefore the atoms are not just qualitatively observed but are rigorously defined theoretically.

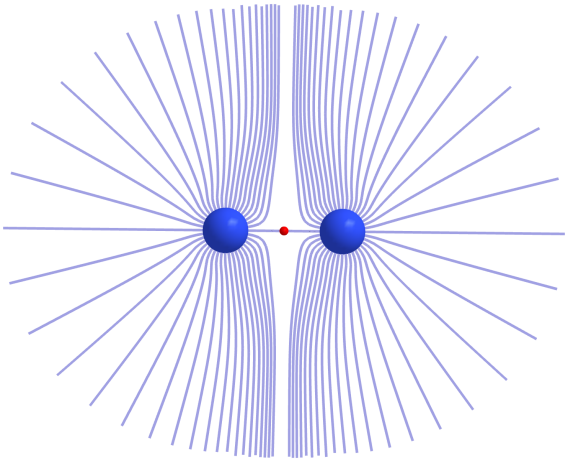

Figure 2.1: Basin paths showing the forms of atoms in N_2

The surface mentioned above is termed the zero-flux surface, $S(\Omega)$, since there is no flux in the density over the surfac (no paths in $\nabla\rho$ cross the surface. It is mathematically defined as Equation 2.28.[41,42]

$$\nabla\rho(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) = 0; \mathbf{r} \in S(\Omega) \tag{2.28}$$

Being proper quantum subsystems, the properties of these atoms are easy to define. Integrating a property density over the basin gives the atomic property corresponding to that operator. For a general property, $O(\Omega)$, with operator $\hat{O}$, the atomic property is defined in Equation 2.29.[41,42]

$$O(\Omega_i) = N\int_{\Omega_i}\left[\int \frac{1}{2}\left[\Psi^*\hat{O}\Psi + (\hat{O}\Psi)^*\Psi\right]d\tau'\right]d\mathbf{r} \tag{2.29}$$

Properties that can be defined include multipoles and energies. The monopole (charge) is defined on the basis of the electron population, whose operator $\hat{O} = 1$, Equations 2.30 and 2.31. The Schrodinger kinetic energy, $K(\Omega)$, and gradient kinetic energy, $G(\Omega)$ can be defined as in Equations 2.32 and 2.33, respectively. As stated, these atoms are proper open systems, so $K(\Omega) = G(\Omega)$, and the kinetic energy is well defined. The kinetic energies, T referring to both, relate to the total

energy through the atomic virial theorem, Equation 2.34 and simply summing the kinetic and potential energy, Equation 2.35.[41,42]

$$N(\Omega) = \int_{\Omega} \rho(\mathbf{r}) d\mathbf{r} \tag{2.30}$$

$$q(\Omega) = Z_{\Omega} - N(\Omega) \tag{2.31}$$

$$K(\Omega) = -\frac{\hbar^2}{4m} N \int_{\Omega} d\mathbf{r} \int d\tau' \left[\psi \nabla^2 \psi^* + \psi^* \nabla^2 \psi \right] \tag{2.32}$$

$$G(\Omega) = -\frac{\hbar^2}{4m} N \int_{\Omega} d\mathbf{r} \int d\tau' \nabla_i \psi^* \cdot \nabla_i \psi \tag{2.33}$$

$$-2T(\Omega) = V(\Omega) \tag{2.34}$$

$$E(\Omega) = T(\Omega) + V(\Omega) = -T(\Omega) = \frac{1}{2} V(\Omega) \tag{2.35}$$

Dipole moments and quadrupole moments are other fundamental electrostatic properties that can be calculated for QTAIM's atoms. The dipole, $\boldsymbol{\mu}(\Omega)$, Equation 2.36 is composed of a charge transfer term, Equation 2.37 and an intraatomic polarization term, Equation 2.38. Here, $Q(\Omega|\Lambda)$ is the charge of groups Λ connected to Ω. In the case of a cyclic system, it is found by solving a set of linear equations. The atomic quadrupole is expressed as a 3x3 symmetric matrix, Equation 2.39.[41,42]

$$\boldsymbol{\mu}(\Omega) = \boldsymbol{\mu}^p(\Omega) + \boldsymbol{\mu}^c(\Omega) \tag{2.36}$$

$$\boldsymbol{\mu}^c(\Omega) = \sum_{\Lambda=1}^{N_b(\Omega)} [\mathbf{R}_{\Omega} - \mathbf{R}_b(\Omega|\Lambda)] Q(\Omega|\Lambda) \tag{2.37}$$

$$\boldsymbol{\mu}^p(\Omega) = \begin{bmatrix} -e\int_\Omega x\rho(\mathbf{r})d\mathbf{r} \\ -e\int_\Omega y\rho(\mathbf{r})d\mathbf{r} \\ -e\int_\Omega z\rho(\mathbf{r})d\mathbf{r} \end{bmatrix} \tag{2.38}$$

$$\begin{aligned} \mathbf{Q}(\Omega) &= -\frac{e}{2}\begin{bmatrix} Q_{xx} & Q_{xy} & Q_{xz} \\ Q_{yx} & Q_{yy} & Q_{yz} \\ Q_{zx} & Q_{zy} & Q_{zz} \end{bmatrix} \\ &= -\frac{e}{2}\begin{bmatrix} \int_\Omega (3x_\Omega^2 - r_\Omega)\rho(\mathbf{r})d\mathbf{r} & 3\int_\Omega x_\Omega y_\Omega \rho(\mathbf{r})d\mathbf{r} & 3\int_\Omega x_\Omega z_\Omega \rho(\mathbf{r})d\mathbf{r} \\ 3\int_\Omega y_\Omega x_\Omega \rho(\mathbf{r})d\mathbf{r} & \int_\Omega (3y_\Omega^2 - r_\Omega)\rho(\mathbf{r})d\mathbf{r} & 3\int_\Omega y_\Omega z_\Omega \rho(\mathbf{r})d\mathbf{r} \\ 3\int_\Omega z_\Omega x_\Omega \rho(\mathbf{r})d\mathbf{r} & 3\int_\Omega z_\Omega y_\Omega \rho(\mathbf{r})d\mathbf{r} & \int_\Omega (3z_\Omega^2 - r_\Omega)\rho(\mathbf{r})d\mathbf{r} \end{bmatrix} \end{aligned} \tag{2.39}$$

The electron density topology also reveals critical points in the electron density, where $\nabla\rho$ equals zero. These are found along bond paths between atoms, in rings, or at the centre of cages. Additionally, the electron density is a cusp at nuclei, which topologically behaves as a critical point. The classifications of critical points are outlined in Figure 2.2. The critical points are classified by rank and signature, referring to the second derivative of the electron density. For molecular structures at equilibrium, the number of non-zero eigenvalues is three. The signature is the sum of the signs of eigenvalues. For nuclear attractor critical points, where the electron density is concentrated, the signs of all eigenvalues are negative, so the signature is -3. For bond critical points, the curvature along the direction corresponding to the bond path is positive, while the others remain negative for a signature of -1. In ring critical points, the eigenvalues for the ring plane are positive, and that for the direction perpendicular to the ring is negative,

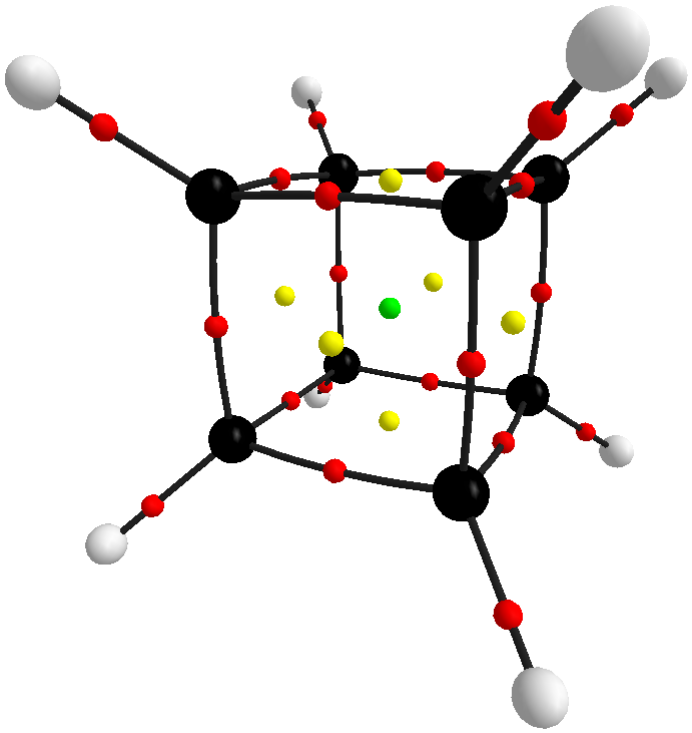

Figure 2.2: Classification of critical points of the electron density for cubane. Nuclear Attractor critical points are shown by nuclei, bond critical points by red dots, ring critical points by yellow dots and the cage critical point by the green dot

for a signature of 1. Lastly, for cage critical points, the eigenvalues are positive in all directions, for a signature of 3.[41,42]

Like for atoms, there are a variety of properties that can be calculated at critical points. When examined at bond critical points, these are used to determine the nature of the bonding interaction. The simplest property to be examined is the electron density at the critical point ρ_c. For what we term covalent interactions, ρ_c is greater than 0.20 au, and for ionic, it is less than 0.10 au. Covalency is also related to the curvature of the density, the second derivative of the density. The complete second derivative, **A**, is a 3x3 matrix, Equation 2.40. However, this can be diagonalized into **Λ**, revealing the eigenvalues of the matrix, λ_i (i = 1, 2, 3, Equation 2.41. These reveal information on the concentration or depletion of electron density along the principal axes corresponding

to the eigenvalues. λ_3 is the eigenvalue along the bond path, and λ_1 and λ_2 are perpendicular to it. We can sum these eigenvalues to give the Laplacian of the electron density, $\nabla^2\rho$, and relate it to the overall accumulation or depletion of electron density at the critical point, Equation 2.42.[41,42]

$$\mathbf{A}(\mathbf{r}_c) = \begin{bmatrix} \frac{\partial^2\rho}{\partial x^2} & \frac{\partial^2\rho}{\partial x\partial y} & \frac{\partial^2\rho}{\partial x\partial z} \\ \frac{\partial^2\rho}{\partial y\partial x} & \frac{\partial^2\rho}{\partial y^2} & \frac{\partial^2\rho}{\partial y\partial z} \\ \frac{\partial^2\rho}{\partial z\partial x} & \frac{\partial^2\rho}{\partial z\partial y} & \frac{\partial^2\rho}{\partial z^2} \end{bmatrix}_{\mathbf{r}=\mathbf{r}_c} \tag{2.40}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \frac{\partial^2\rho}{\partial x'^2} & 0 & 0 \\ 0 & \frac{\partial^2\rho}{\partial y'^2} & 0 \\ 0 & 0 & \frac{\partial^2\rho}{\partial z'^2} \end{bmatrix}_{\mathbf{r}=\mathbf{r}_c} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}_{\mathbf{r}=\mathbf{r}_c} \tag{2.41}$$

$$\nabla^2\rho = \lambda_1 + \lambda_2 + \lambda_3 \tag{2.42}$$

The bond ellipticity, ε_c (Equation 2.43), quantifies the accumulation of the electron density in one plane as compared to another. If the bond is cylindrically symmetric, like a single bond, λ_1 is equal to λ_2, and ε_c is 0. It increases as the bond takes on more multiple bond characteristics, up to a full double bond. As the bond then moves from a double bond to a triple bond, it starts decreasing again, as the triple bond also has cylindrical symmetry.[41,42]

$$\varepsilon_c = \frac{\lambda_1}{\lambda_2} - 1 \tag{2.43}$$

Energy densities at critical points are also used to quantify the interactions. Both the kinetic and potential energy densities provide valuable information. The kinetic energy G is similar to Equation 2.33, but not integrated over the basin, just evaluated locally, Equation 2.44. The

total energy density, H is the sum of G and V, Equation 2.45. V is the average potential field on an electron at $\mathbf{r}$. Information on the meaning of the kinetic and potential energy densities are relayed by the local virial theorem, Equation 2.46. So, for interactions which have more accumulation of density and $\nabla^2\rho < 0$, V dominates, while for interactions with depletion of the electron density, G dominates. Therefore, H can also be related to significance by these statements. H is negative when V dominates for bonds with dominating sharing of electrons.[41,42]

$$G(\mathbf{r}) = -\frac{\hbar^2}{2m} N \int d\tau' \nabla_i \Psi^* \cdot \nabla_i \Psi \tag{2.44}$$

$$H = G + V \tag{2.45}$$

$$\left(\frac{\hbar^2}{4m}\right)\nabla^2\rho = 2G + V \tag{2.46}$$

Finally, QTAIM can calculate the average number of electrons shared between atoms through the Delocalization Index, DI. There is also a corresponding Localization Index, LI, for the number of electrons localized on an atom. The DI is defined in Equation 2.47, where $F^\sigma(A,B)$ is the Fermi correlation, Equation 2.48. The LI is defined from integration the Fermi correlation over the same basin (A=B), Equation 2.49.[41,42]

$$\mathrm{DI}(A,B) = 2|F^\alpha(A,B)| + 2|F^\beta(A,B)| \tag{2.47}$$

$$F^\sigma(A,B) = -\sum_i \sum_j \int_A d\mathbf{r}_1 \int_B d\mathbf{r}_2 \phi_i^*(\mathbf{r}_1)\phi_j(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2)\phi_i(\mathbf{r}_2) \tag{2.48}$$

$$\mathrm{LI}(A,A) = |F^\alpha(A,A)| + |F^\beta(A,A)| \tag{2.49}$$

$$\nabla\rho = \mathbf{i}\frac{\partial\rho}{\partial x} + \mathbf{j}\frac{\partial\rho}{\partial y} + \mathbf{k}\frac{\partial\rho}{\partial z} \tag{2.50}$$

$$n - b + r - c = 1 \tag{2.51}$$

2.8 Multivariate Data Analysis

2.8.1 Principal Components Analysis

Given a large number of properties in a data matrix, **X**, it is desirable to maximize the information in the fewest number of variables. One technique to do this is Principal Components Analysis. Principal Components Analysis seeks to find a scores matrix **F** through a transformation of **X**. The columns of **F** should be orthogonal to each other - the covariance matrix $cov(\mathbf{F}, \mathbf{F})$ should be diagonal. We write the scores **F** as Equation 2.52. We desire to find **Q** that achieves this.[43]

$$\mathbf{F} = \mathbf{XQ} \tag{2.52}$$

We use the formulae for covariance matrix, substitute Equation 2.52, and take **S** as the covariance matrix of **X**, Equation 2.53 and 2.54. We know $\mathbf{X}^\mathsf{T}\mathbf{X}$ is symmetric and therefore diagonalizable. **P** contains the eigenvectors of **S**, and **D** the corresponding eigenvalues, ordered from high to low. Remembering that we need to diagonalize $cov(\mathbf{T}, \mathbf{R})$, setting **Q**=**P** accomplishes this, since $\mathbf{P}^\mathsf{T}\mathbf{P}$ and is the identity matrix due to the transpose **P** being equal to the inverse of **P** for orthogonal matrices.[43,44]

$$cov(\mathbf{F}, \mathbf{F}) = \frac{1}{n-1}\mathbf{F}^\mathsf{T}\mathbf{F} = \frac{1}{n-1}\mathbf{Q}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{XQ} \tag{2.53}$$

$$\mathbf{S} = \frac{1}{n-1}\mathbf{X}^{\mathsf{T}}\mathbf{X} = \mathbf{P}\mathbf{D}\mathbf{P}^{\mathsf{T}} \tag{2.54}$$

$$cov(\mathbf{T}, \mathbf{T}) = \mathbf{P}^{\mathsf{T}}\mathbf{S}\mathbf{P} = \mathbf{Q}^{\mathsf{T}}\mathbf{P}\mathbf{D}\mathbf{P}^{\mathsf{T}}\mathbf{Q} \tag{2.55}$$

This can also be written as a singular value decomposition, Equation 2.56. Here $\mathbf{\Delta}$ is the singular values - the square root of the eigenvalues of $\mathbf{X}\mathbf{X}^{\mathsf{T}}$. This is one way to easily calculate a PCA decomposition, but another involves an iterative method. Start by setting a columnn of $\mathbf{F}$ equal to an arbitrary column in $\mathbf{X}$. Then you iterate between Equations 2.57 and 2.58 until $\mathbf{f}$ stops changing. Then you remove the information from this newly found component from $\mathbf{X}$ and repeat for the next column of $\mathbf{X}$.[43,44]

$$\mathbf{F} = \mathbf{P}\mathbf{\Delta}\mathbf{Q}^{\mathsf{T}} \tag{2.56}$$

$$\mathbf{q} = \frac{1}{\lambda_p}\mathbf{X}^{\mathsf{T}}\mathbf{f} \tag{2.57}$$

$$\mathbf{f} = \mathbf{X}\mathbf{q} \tag{2.58}$$

2.8.2 Partial Least Squares

Partial Least Square (PLS) has a similar philosophy to PCA, but it invokes a response matrix, $\mathbf{Y}$ as well. That is, you are interested in regressing $\mathbf{Y}$ on the data matrix $\mathbf{X}$. Both $\mathbf{Y}$ and $\mathbf{X}$ are decomposed, with each decomposition taking into account information from the other matrix. The decompositions are of form Equation 2.59 and 2.60 This scheme is detailed roughly in the Equation set Equation 2.61. After determining the decompositions of $\mathbf{X}$ and $\mathbf{Y}$ in such a way, regressions of $\mathbf{Y}$ on $\mathbf{X}$ can be

performed.[45]

$$\mathbf{X} = \mathbf{F}\mathbf{Q}^{\mathsf{T}} \tag{2.59}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{P}^{\mathsf{T}} \tag{2.60}$$

$\mathbf{t} := \mathbf{y}_j$

Loop

$$\begin{aligned} \mathbf{q} &:= \frac{\mathbf{X}^{\mathsf{T}}\mathbf{t}}{||\mathbf{X}^{\mathsf{T}}\mathbf{t}||} \\ \mathbf{f} &:= \mathbf{X}\mathbf{q} \\ \mathbf{q} &:= \frac{\mathbf{Y}^{\mathsf{T}}\mathbf{t}}{||\mathbf{Y}^{\mathsf{T}}\mathbf{t}||} \\ \mathbf{t} &:= \mathbf{Y}\mathbf{q} \end{aligned} \tag{2.61}$$

Until **f**stop changing

References

[1] Jensen, F., Introduction to Computational Chemistry, John Wiley & Sons, Inc., Hoboken, NJ, United States, 2nd edition **2006**.

[2] Hohenberg, P. and Kohn, B., *Phys. Rev. B* **1964**, *136*, 864.

[3] Kohn, W. and Sham, L. J., *Phys. Rev. B* **1965**, *140*.

[4] Vosko, S. H., Wilk, L. and Nusair, M., *Can. J. Phys.* **1980**, *58*, 1200–1211.

[5] Becke, A. D., *Phys. Rev. A* **1988**, *38*, 3098–3100.

[6] Lee, C., Yang, W. and Parr, R. G., *Phys. Rev. B* **1988**, *37*, 785–789.

[7] Perdew, J. P., Burke, K., Ernzerhof, M., of Physics, D., *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[8] Tao, J., Perdew, J. P., Staroverov, V. N. and Scuseria, G. E., *Phys. Rev. Lett.* **2003**, *91*, 146401.

[9] Becke, A. D., *J. Chem. Phys* **1993**, *98*, 5648–5652.

[10] Devlin, F. J., Finley, J. W., Stephens, P. J. and Frisch, M. J., *J. Phys. Chem.* **1995**, *99*, 16883–16902.

[11] Adamo, C. and Barone, V., *J. Chem. Phys.* **1999**, *110*, 6158.

[12] Grimme, S., *J. Chem. Phys.* **2006**, *124*, 034108.

[13] Ditchfield, R., Hehre, W. J. and Pople, J. a., *J. Chem. Phys.* **1971**, *54*, 724.

[14] Dill, J. D. and Pople, J. A., *J. Chem. Phys.* **1975**, *62*, 2921.

[15] Clark, T., Chandrasekhar, J., Spitznagel, G. W. and Schleyer, P. V. R., *J. Comput. Chem.* **1983**, *4*, 294–301.

[16] Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., DeFrees, D. J. and Pople, J. A., *J. Chem. Phys.* **1982**, *77*, 3654.

[17] Hariharan, P. C. and Pople, J. A., *Theor. Chim. Acta* **1973**, *28*, 213–222.

[18] Frisch, M. J., Pople, J. A. and Binkley, J. S., *J. Chem. Phys.* **1984**, *80*, 3265–3269.

[19] Rassolov, V. A., Pople, J. A., Ratner, M. A. and Windus, T. L., *J. Chem. Phys.* **1998**, *109*, 1223–1229.

[20] Krishnan, R., Binkley, J. S., Seeger, R. and Pople, J. A., *J. Chem. Phys.* **1980**, *72*, 650–654.

[21] McLean, A. D. and Chandler, G. S., *J. Chem. Phys.* **1980**, *72*, 5639–5648.

[22] Blaudeau, J.-P., McGrath, M. P., Curtiss, L. A. and Radom, L., *J. Chem. Phys.* **1997**, *107*, 5016–5021.

[23] Curtiss, L. A., McGrath, M. P., Blaudeau, J., Davis, N. E., Binning, R. C. and Radom, L., *J. Chem. Phys.* **1995**, *103*, 6104–6113.

[24] Glukhovtsev, M. N., Pross, A., McGrath, M. P. and Radom, L., *J. Chem. Phys.* **1995**, *103*, 1878–1885.

[25] Dunning Jr, T. H., *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[26] Kendall, R. A., Dunning, T. H. and Harrison, R. J., *J. Chem. Phys.* **1992**, *96*, 6796–6806.

[27] Woon, D. E. and Dunning, T. H., *J. Chem. Phys.* **1993**, *98*, 1358–1371.

[28] Peterson, K. A., Woon, D. E. and Dunning, T. H., *J. Chem. Phys.* **1994**, *100*, 7410–7415.

[29] Woon, D. E. and Dunning Jr., T. H., *J. Chem. Phys.* **1995**, *100*, 2975–2988.

[30] Dunning T.H., J., Peterson, K. A. and Wilson, A. K., *J. Chem. Phys.* **1999**, *110*, 7667–7676.

[31] Wilson, A. K., van Mourik, T. and Dunning, T. H., *J. Mol. Struct. THEOCHEM* **1996**, *388*, 339–349.

[32] Jensen, F., *J. Chem. Phys.* **2001**, *115*, 9113–9125.

[33] Jensen, F., *J. Chem. Phys.* **2002**, *116*, 7372–7379.

[34] Jensen, F., *J. Chem. Phys.* **2002**, *117*, 9234–9240.

[35] Jensen, F., *J. Chem. Phys.* **2003**, *118*, 2459–2463.

[36] Jensen, F. and Helgaker, T., *J. Chem. Phys.* **2004**, *121*, 3463–70.

[37] Jensen, F., *J. Phys. Chem. A* **2007**, *111*, 11198–11204.

[38] Jensen, F., *J. Chem. Phys.* **2012**, *136*, 114107.

[39] Jensen, F., *J. Chem. Phys.* **2013**, *138*, 014107.

[40] Rappoport, D. and Furche, F., *J. Chem. Phys.* **2010**, *133*, 143105.

[41] Bader, R., Atoms in Molecules: A Quantum Theory, Clarendon Press, Oxford, U.K. **1994**.

[42] Matta, C F; Boyd, R. J., The Quantum Theory of Atoms in Molecules, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim **2007**.

[43] Abdi, H. and Williams, L. J., *Cumpotational Statistics* **2010**, *2*, 433–470.

[44] Francis, P. J. and Wills, B. J., *Pm&R* **1999**, *6*, 10.

[45] Wold, H., *Journal of Applied Probability* **1975**, *12*, 117–142.

Chapter 3

Model Chemistry Sensitivity of Descriptors from the Quantum Theory of Atoms in Molecules

3.1 Abstract

The Quantum Theory of Atoms in Molecules (QTAIM) provides a theoretical foundation to determine the properties of functional groups through additive atomic contributions. Many studies have used QTAIM in their analyses with a variety of electronic structure methods, but it is unknown if the properties measured using one model chemistry, the combination of the electronic structure method and basis set, can be compared to those measured by another. Here, we evaluate the sensitivity of QTAIM functional group and bond critical point properties using six functionals and seven basis sets. High-level B2PLYPD3-BJ/aug-cc-pV5Z reference values are provided for 116 functional groups and the property sensitivity with respect to these values are evaluated based on

absolute deviations and by assessing linear relationships. Functional group properties, including charges, dipoles, quadrupoles and volumes, were found to be mostly insensitive to choice of computational model chemistry. However, due to structural and topological inconsistencies, the 6-31G(d) basis set is not recommended for use. Bond critical point properties varied with choice of model chemistry, but models incorporating hybrid functionals and triple-ζ basis sets provided values suitable for use in regression studies.

3.2 Introduction

The choice of substituent in a molecule can drastically impact the kinetics and thermodynamics of reactions.[1] To describe this effect, empirical parameters such as Hammett's σ,[2] Taft's σ^*,[3] Charton's σ_I,[4] and Yasuhide's Dual Substituent Parameters[5] were developed based on substituents' influence on a probe. Empirical parameters are inferential parameters, more correctly described as proxies, assessing the effect a substituent has on a measured property, not quantifying the intrinsic properties of a substituent. Regardless, Hammett-type parameters remain the most common tool for describing observed reaction trends and analysing substituent effects in terms of quantitative structure property relationship (QSPR).[6]

Quantum chemical methods can also be used to obtain such descriptors for QSPR analyses.[7] Such new parameters, based on various analyses and tools, include molecular electrostatic potentials,[8] Natural Population Analysis charges of the substituent active region,[9,10] and Quantum Theory of Atoms in Molecules (QTAIM) properties.[11–15] Like empirical parameters, the prior two give values that are not intrinsic substituent properties, because they measure a substituent's

effect elsewhere in a molecule. In contrast, QTAIM group properties reflect the intrinsic properties of substituents. For example, substituent dipole moments have been shown to reflect the field effect in substituted bicyclo[1.1.1]pentane-1-carboxylic acids.[11]

QTAIM divides the electron density into atomic basins, each containing one attractor, which are separated from each other by zero flux surfaces (ZFSs) in the electron density, which no gradient paths cross ($\nabla\rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = 0,\ r \in S$). QTAIM's definition of an atom is the union of an attractor and the electron density in its associated basin.[16] An atomic property can be calculated by integrating that property's associated density over the atomic basin, and the sum over all atoms in the molecule equals the molecular value. As such, the sum of an atomic property over the atoms in a functional group gives the functional group's property.[17,18] Bonding interactions can be evaluated through bond critical point (BCP) properties, which occur at ZFSs and act as the interface between a functional group and the remainder of the molecule, and may also help quantify a functional group's properties, as has previously been applied in describing substituent effects.[14,19–22]

To be used generally, QTAIM substituent properties should be transferable between environments and not be affected by choice of model chemistry (combination of an electronic structure method and basis set).[23–25] The first condition ensures that properties are independent of the substrate and the second allows for values from model chemistries to be compared. This latter requirement has indirectly been the subject of several studies.[26–29] In studying atomic properties computed using HF, MP2 and B3LYP methods, Matta and Popelier independently proved that Density Functional Theory (DFT) is able to reproduce the atomic properties determined by MP2.[26,27] HF methods are able to produce an accurate molecular geometry, but the electon density determined

is not sufficient to give accurate values. Popelier also found that BCP properties were sensitive to choice of electronic structure method, with $\nabla^2\rho$ affected the most.[27] Another study on BCP properties by Tognetti and Joubert analysed the effect of using different DFT functionals. Each property was affected to a different degree, with margins of error of 5%, 25%, and 7% for the electron density, Laplacian, and reduced density gradients, respectively.[28] Jablonski and Palusiak studied the sensitivity of BCP properties to choice of basis set using Pople and Dunning's correlation consistent basis sets. They recommend using the aug-cc-pVTZ or 6-311++G(2df,pd) basis sets in studying bonding interactions and advise against using 6-31G or DZ quality Dunning basis sets.[29] Many of the above papers focus on choice of functional or basis set separately, or only test a very limited number of combinations, and rarely focus on group properties.

In this work, the sensitivity of QTAIM functional group properties to both chosen functional and basis set (42 different model chemistries) is investigated. Group properties are determined for a large set of 116 substituents, and their sensitivity to choice of model chemistry is evaluated. Sensitivity is judged in the sense of magnitude - the ability to exactly replicate reference B2PLYPD3-BJ/aug-cc-pV5Z values, and trends - the quality of the linear relationship between different model chemistries. The two points of comparison allow insight on whether the properties determined by a certain model chemistry are suitable for use in QSPR-type correlative studies where they are applied as regressors. It is shown that local scalar properties are sensitive to the level of theory chosen, while integrated properties are not.

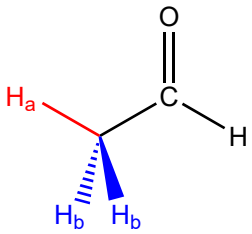

Figure 3.1: Example of the designation of the in-plane (ip) or out-of-plane (op) hydrogen atom in a substituent. H_a is the ip hydrogen, and H_b is the op hydrogen. This substituent is labeled CH_2CHO-ip or CH_2CHO-op for choice of H_a or H_b as the substrate hydrogen, respectively.

3.3 Methodology

The 116 substituents, denoted R, comprise a variety of functional group types (e.g. alkyl, amino, carbonyl, etc., full set shown in Supplementary Figures A1 and A2) and span a wide range of property values. Twelve groups are included in the data set twice because of their ability to exist in two different conformations. In these cases, each inequivalent hydrogen atom is treated as the substrate and are identified as "R-ip" or "R-op" for in-plane (ip) or out-of-plane (op) hydrogens. Figure 3.1 shows an example, where H_a and H_b are the ip and op hydrogen substrate, respectively. B2PLYPD3-BJ/aug-cc-pV5Z properties were chosen as reference values. It is noted that the D3 dispersion is unnecessary in the reference model as it does not directly affect the density; it was included to be consistent with other double-hybrid benchmarking studies. The D3-BJ dispersion can affect the determined optimized geometry, which in turn indirectly affects the density analyzed by QTAIM.[30] It can be neglected without issue. This functional, with the smaller aug-cc-pvTZ basis set, has been shown to accurately reproduce experimental dipole moments and polarizabilities with similar accuracy as CCSD and MP2 methods.[31]

Forty two model chemistries, combining six functionals and seven basis sets, were selected to investigate the sensitivity of QTAIM prop-

erties. The chosen functionals are commonly used and from different rungs of Perdew's Jacob's Ladder.[32] They include one Local Spin Density Approximation (LSDA) Functional: SVWN;[33,34] two Generalized Gradient Approximation (GGA) functionals: BLYP[35,36] and PBE;[37] one meta-GGA functional: TPSS;[38] and two hybrid functionals: B3LYP[39] and PBE0.[40] The popular Minnesota functionals were not selected for study because the densities determined by those functionals have been shown to be inaccurate.[41] The seven chosen basis sets differ in family, size and inclusion of diffuse functions; they are: 6-31G(d),[42–47] pc-1,[48–53] pc-2,[48–53] aug-pc-1,[48–54] aug-pc-2,[48–54] def2-TZVPP,[55] and def2-TZVPPD.[55,56]

All electronic structure calculations were performed using Gaussian09.[57] Geometry optimizations were performed on R-H molecules using each of the selected model chemistries, with harmonic frequencies evaluated to ensure minima were found. H_2CCCO was predicted to have a bent structure (C_s symmetry). Geometry optimization with SVWN/6-31G(d), however, resulted in a C_{2v} symmetric structure, whose coordinate system was adjusted to be close to the C_s symmetry by fixing the angle between one carbon atom, the origin, and the x-axis and using Givens rotation matrices to rotate the coordinates.[58] Extended wavefunction files were generated and QTAIM properties assessed using AIMAll software.[59] Non-nuclear attractors within substituents (present between the carbon atoms in HCCH with PBE0/6-31G(d)) were treated as an additional atom in the substituent and incorporated as such.

Substituent properties of interest include charge, q(R); total dipole, $\boldsymbol{\mu}$(R); charge transfer dipole moment, $\boldsymbol{\mu}_c$(R); polarization dipole moment, $\boldsymbol{\mu}_p$(R); volumes taken at the 0.001 a.u. isosurface, Vol(R); and contributions to molecular quadrupole, $\mathbf{Q}$(R). $\boldsymbol{\mu}$(R), Equation 3.1, is the

sum of $\boldsymbol{\mu}_c$(R) and $\boldsymbol{\mu}_p$(R), defined by Equations 3.2 and 3.3, respectively. **Q**(R) was determined by summing the atomic contributions according to Laidig using Equation 3.4.[60] Here, $\mathbf{R}_\Omega$ is the position of the nuclear attractor in Ω, $\mathbf{R}_b(\Omega|\Lambda)$ is the position of the bond critical point between Ω and attached group Λ, $Q(\Omega|\Lambda)$ is the charge of group Λ, q(Ω) are the atomic charges of basin Ω, χ_Ω^α and χ_Ω^β ($\alpha, \beta = x, y$, or z) are the α or β geometric coordinate of the nucleus of basin Ω, $Q_{\alpha\beta}(\Omega)$ is the atomic quadrupole moment of basin Ω, $\delta_{\alpha\beta}$ is the Kronecker delta, $\langle r^2 \rangle_\Omega$ is the atomic second radial moment of basin Ω, $\mu_\alpha^p(\Omega)$ and $\mu_\beta^p(\Omega)$ are the α or β component of $\boldsymbol{\mu}_p(\Omega)$, and $Q_{\gamma\gamma}$ is the trace of the contribution to the quadrupole matrix. Lastly, R-H BCP properties are assessed, including the potential energy density, V_c; kinetic energy density, G_c; total energy density, H_c, electron density, ρ_c; distance from the substrate hydrogen to the BCP, r(H-BCP); and the Laplacian of the electron density, $\nabla^2\rho_c$, and its associated eigenvalues, λ_i (i=1-3).

$$\boldsymbol{\mu}(\mathrm{R}) = \boldsymbol{\mu}^{\mathrm{c}}(\mathrm{R}) + \boldsymbol{\mu}^{\mathrm{p}}(\mathrm{R}) \tag{3.1}$$

$$\boldsymbol{\mu}^{\mathrm{c}}(\mathrm{R}) = \sum_{\Omega \in \mathrm{R}} \sum_{\Lambda=1}^{N_b(\Omega)} [\mathbf{R}_\Omega - \mathbf{R}_b(\Omega|\Lambda)]\, Q(\Omega|\Lambda) \tag{3.2}$$

$$\boldsymbol{\mu}^{\mathrm{p}}(\mathrm{R}) = \sum_{\Omega \in \mathrm{R}} -\int_\Omega \mathbf{r}\rho(\mathbf{r})d\mathbf{r} \tag{3.3}$$

$$\begin{aligned}\mathbf{Q}(R) = \frac{1}{2}\sum_{\Omega \in R}[3(q(\Omega)\chi_\Omega^\alpha\chi_\Omega^\beta + \frac{Q_{\alpha\beta}(\Omega) + \delta_{\alpha\beta}\{\langle r^2\rangle_\Omega\}}{3} + \chi_\Omega^\beta\mu_\alpha^p(\Omega) \\ + \chi_\Omega^\alpha\mu_\beta^p(\Omega)) - \delta_{\alpha\beta}(Q_{\gamma\gamma})]\end{aligned} \tag{3.4}$$

The chosen model chemistries were assessed through calculation of error scores and analysis of linear plots. The former includes relative

mean absolute deviations, %MAD, Equation 3.5, and relative mean deviations, %MD, Equation 3.6, where, P_i^X represents the value of property P for substituent i using method X, N_{sub} is the number of substituents, and %MAD$(P)^X$ is the percent mean absolute deviation of property P measured using the specified method, X. Many group dipole, charges, and quadrupole values are near zero - either due to symmetry or just inherent low values - causing near zero denominators to unreasonably inflate %MAD. For these properties, errors are evaluated on an absolute scale using MAD and MD statistics. For linear assessments, we examine the degree to which a model chemistry reproduces the reference values by its deviation from the ideal $y = x$ linear model. R and RStudio were used to perform data analysis and prepare graphs.[61,62] The ggplot2 package was used to generate figures,.[63]

$$\%\mathrm{MAD}(P)^X = \frac{100}{N}\sum_{i=1}^{N}\frac{|P_i^X - P_i^{\mathrm{B2PLYP}}|}{|P_i^{\mathrm{B2PLYP}}|} \tag{3.5}$$

$$\%\mathrm{MD}(P)^X = \frac{100}{N}\sum_{i=1}^{N}\frac{P_i^X - P_i^{\mathrm{B2PLYP}}}{|P_i^{\mathrm{B2PLYP}}|} \tag{3.6}$$

3.4 Results and Discussion

Substituents were chosen to represent a variety of different chemical attributes, with QTAIM properties (Reference values provided as Appendix E) varying over a large range. Table 3.1 and 3.2 show the breadth of ranges for each property under consideration, indicating that the test set includes groups representing many different properties of compounds. It is noted that ionic compounds are underrepresented in the test set, but studies on substituent effects often focus on compounds with mainly covalent interactions. As such, the conclusions

from this paper apply mainly to covalent compounds, and due to the limited number of groups, caution is advised in extending our conclusions to ionic groups.

Table 3.1: B2PLYPD3-BJ/aug-cc-pV5Z reference group property ranges

Property	Min	Max	Property	Min	Max
q(R)	-1.43	0.84	μ_x(R)	-1.61	1.19
$\|\|\boldsymbol{\mu}$(R)$\|\|$	0.02	2.91	μ_y(R)	-2.90	0.830
$\|\|\boldsymbol{\mu}_p$(R)$\|\|$	0.12	3.45	μ_z(R)	-1.61	1.65
$\|\|\boldsymbol{\mu}_c$(R)$\|\|$	0.00	4.13	μ_x^p(R)	-1.95	1.58
$\|\|\mathbf{Q}$(R)$\|\|$	0.00	8.20	μ_y^p(R)	-2.70	3.43
Vol(R)	150.9	742.3	μ_z^p(R)	-2.61	2.89
Q_{xx}(R)	-8.16	4.68	μ_x^c(R)	-2.76	2.76
Q_{yy}(R)	-6.57	5.36	μ_y^c(R)	-4.13	3.28
Q_{zz}(R)	-6.09	7.14	μ_z^c(R)	-3.56	3.39

Table 3.2: B2PLYPD3-BJ/aug-cc-pV5Z reference local BCP property ranges

Property	Min	Max
ρ_c	0.057	0.387
r(H-BCP)	0.313	1.56
$\nabla^2\rho_c$	-3.330	0.217
$\lambda_{1,c}$	-2.156	-0.078
$\lambda_{2,c}$	-2.120	-0.078
$\lambda_{3,c}$	0.142	1.212
G_c	0.015	0.138
V_c	-0.942	-0.069
H_c	-0.887	-0.009

3.5 Integrated Properties

3.5.1 Group Charge

Charges are one of the most often applied concepts in chemistry. Using QTAIM, group charges were evaluated and compared with reference values. Table A1 in Appendix A lists comparative statistics including MAD, MD, SD, and maximum positive and negative deviations. Figure 3.2(a)

and (b) displays MD(q(R)) and MAD(q(R)) for all 42 model chemistries, respectively. Generally, MAD are small, with a maximum MAD of 0.065 au for SVWN/6-31G(d), and most others significantly smaller. Regardless of the magnitude, trends in MAD are seen with different functionals and basis sets.

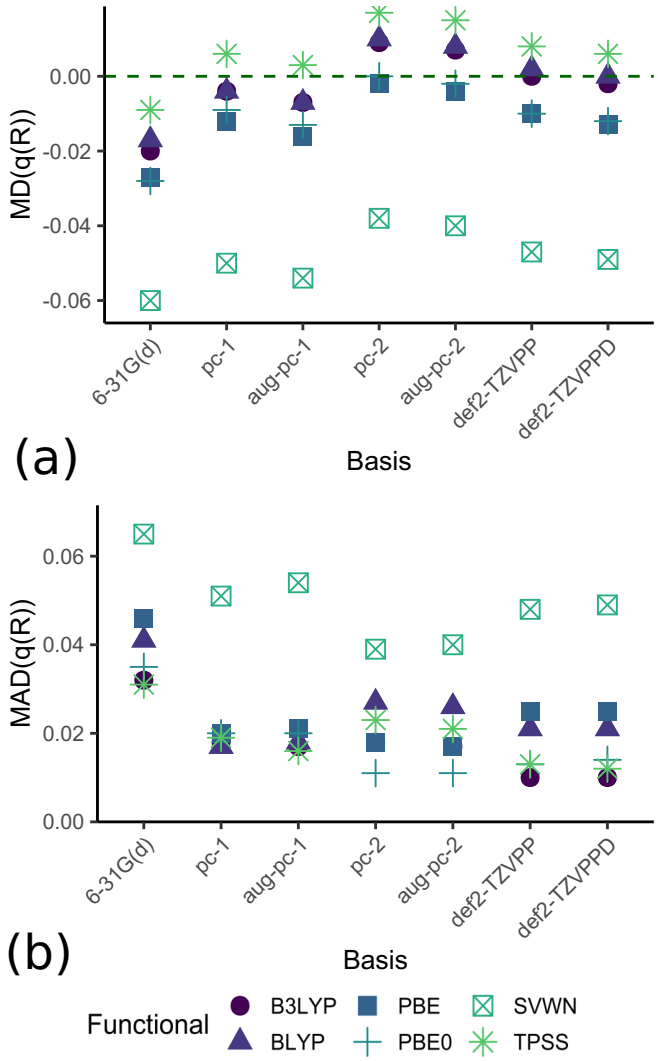

Figure 3.2: (a) MD(q(R)) and (b) MAD(q(R)) of tested model chemistries compared to reference B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green line represents the ideal MD value.

For model chemistries with basis sets of different sizes MAD(q(R)) aligns with expectations. For any given functional, the small Pople style 6-31G(d) basis set is found to have the highest MAD(q(R)), and using other basis sets size decreases MAD. The double-ζ polarization consis-

tent basis sets perform well (MAD less than 0.02 au), and triple-ζ basis sets almost always have low MAD (SVWN being the exception in both cases). Addition of diffuse functions does not change MAD(q(R)) significantly, probably because diffuse functions allow density to be found further from the nucleus but do not remove it from the associated atomic basin. The trend of MAD and basis set size is not followed for the BLYP functional with the polarization consistent basis sets; MAD(q(R)) for the double-ζ basis sets is lower than for triple-ζ basis sets, and so complete basis set extrapolation could be problematic. Here we see that the choice of functional also has an important impact on q(R).

Generally, q(R) becomes more accurate in accordance with the rungs of Perdew's Jacob's ladder. For any given basis set, SVWN has the largest MAD(q(R)), GGA functionals have lower values, and meta-GGA and hybrid functionals have similarly small values. The same trend is echoed in the observed maximum negative deviations (Table A1). A few exceptions exist. For pc-1 and aug-pc-1, the MAD for all functionals except SVWN are quite close to each other. Further, for pc-2 and aug-pc-2 basis, the meta-GGA functional performs worse than the hybrid functionals, and is between BLYP and PBE. PBE, PBE0, and SVWN underestimate q(R). For SVWN, MD(q(R)) and MAD(q(R)) are similar in magnitude, meaning q(R) are consistently underestimated by this functional. TPSS slightly overestimates q(R), but MD is low in magnitude and differs from MAD, so the overestimation is not large.

All above evidence indicates the minimal effect choice of model chemistry has on q(R) values. Even though the magnitude of variation is small, the accuracy of q(R) values varies among basis sets, reinforcing Volkov's finding that QTAIM charges are sensitive to choice of basis set.[64] However, since q(R) differ from reference values by only a small fraction of an electron in most cases, the effect is insignificant from a

practical perspective since reference values range from -1.43 to 0.84 au. Errors for different functionals are also small, with meta-GGA and hybrid functionals having low maximum errors (0.01-0.09 au excluding 6-31G(d) basis sets). Therefore these functionals determine charges accurately, even in worst case scenarios.

Since all model chemistries quantitatively reproduced reference values, it is expected that reference value trends are also reproduced in a linear relationship. The slopes, intercepts and r^2 of all linear relationships with the reference B2PLYPD3-BJ/aug-cc-pV5Z are plotted in Figure 3.3 (values shown in Table A2). The overall quality of the trend is judged by r^2, and the slope and intercept give insight into the magnitude of agreement. All methods exhibit r^2 of 0.99 or above, so values obtained from any of the studied model chemistries, even SVWN/6-31G(d), could be applied as regressors.

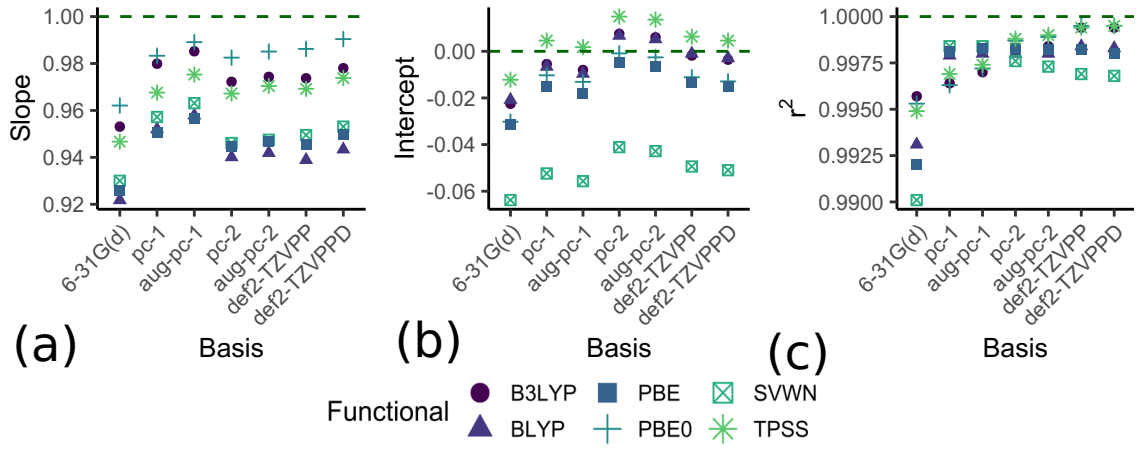

Figure 3.3: q(R) regression parameters (a) slope (b) intercept and (c) Correlation coefficient comparing tested model chemistries to B2PLYPD3-BJ/aug-cc-pV5Z reference. Dashed green lines represent regression parameters in the ideal $y = x$ relationship

For most methods, the slope and intercept are close to the ideal line. Correspondingly, methods with higher MAD, typically those using the SVWN functional and/or the 6-31G(d) basis set, show more deviation. For all basis sets, SVWN exhibits large negative intercepts, demonstrat-

ing its systematic negative deviations. For all functionals, the 6-31G(d) basis set shows the slope most deviated from 1, and the largest magnitude intercept consistent with the observed high absolute deviation. Other methods, which have low MAD, show slopes and intercepts close to 1 and 0, respectively. Based on all evidence presented, for studies where highly accurate group charges are required, hybrid functionals with triple-ζ basis sets are recommended, but group charges from any method can be used when only trends are important. These near ideal slopes and intercepts also support the interpretation that all MAD are still low in a practical sense.

3.5.2 Group Dipole

The total group dipole and its directional components were determined using the standard coordinate system from Gaussian calculations, and errors (see Tables A3- A6) evaluated as described earlier. Figure 3.4 displays MAD and MD statistics graphically. MADs range from 0.005 to 0.041 au for μ_x(R), 0.008 to 0.052 au for μ_y(R), and 0.004 to 0.039 au for μ_z(R). These deviations are similar in magnitude to those observed by Matta, and no individual directional component is the primary source of the total error in $\boldsymbol{\mu}$(R).[26] MDs are small relative to MAD, revealing that values are not consistently over- or underestimated. Worst case errors for μ_z(R) are typically around 0.1 au, but occasionally rise to 0.2 au. μ_x(R) and μ_y(R) have higher maximum errors, only getting as low as 0.1 au, but the relative MAD, MD and max deviations are low compared to spans of 2.8, 3.7, and 3.2 au, for μ_x(R), μ_y(R) and μ_z(R), respectively.

Aside from 6-31G(d), which has the highest MAD, choice of basis set has a low impact on performance. In general, triple-ζ basis sets have lower MAD scores than double-ζ basis sets, and augmenting them with diffuse functions lowers MAD even more. We observe an irregular trend

in MAD with respect to basis set for μ_y(R), μ_z(R), and $||\boldsymbol{\mu}(R)||$. The SVWN functional used with triple-ζ and augmented triple-ζ basis sets has increased error over aug-pc-1. However, SVWN exhibits the expected behaviour for μ_x. Low MAD scores for augmented basis sets could be due to the inclusion of diffuse functions in the reference basis set, but previous work showed that for polarization consistent basis sets, diffuse functions are required to reach the complete basis set limit in a consistent manner.[54] Diffuse functions probably affect MAD($\boldsymbol{\mu}$) values by changing the distribution of electron density within the basins, allowing electron density to be found farther from nuclei. For most accurate dipole values, it is therefore recommended to include diffuse functions.

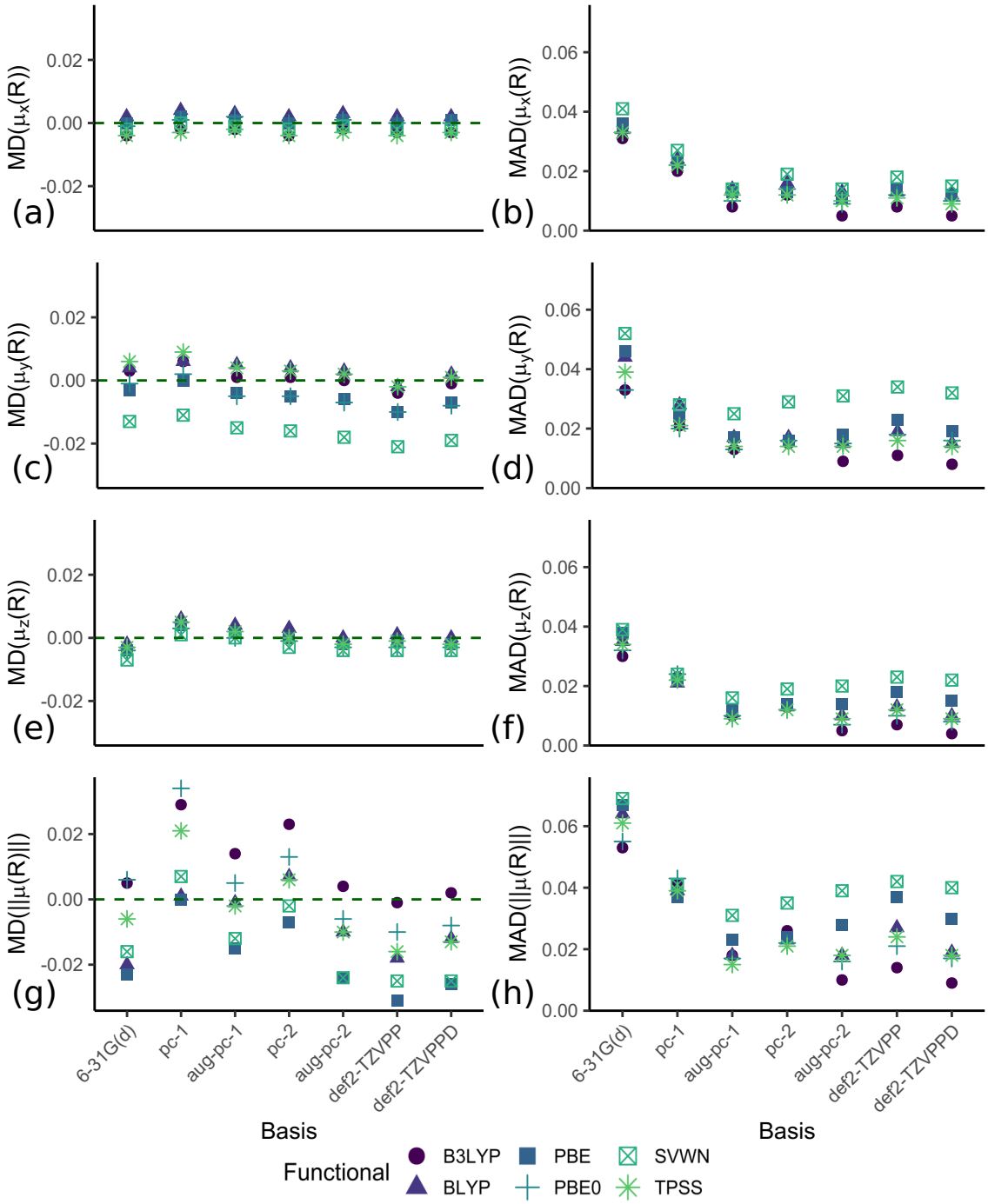

Figure 3.4: (a) MD(μ_x(R)) (b) MAD(μ_x(R)) (c) MD(μ_y(R)) (d) MAD(μ_y(R)) (e) MD(μ_z(R)) (f) MAD(μ_z(R)) (g) MD(||μ(R)||) (h) MAD(||μ(R)||) of tested model chemistries compared to reference B2PLYPD3-BJ/aug-cc-pV5Z values

Trends in MAD based on choice of functional mirror those for MAD(q(R)): the error decreases in a manner consistent with the rungs of Perdew's Jacob's ladder, until the meta-GGA rung. Generally, PBE0 has similar MAD as TPSS, but B3LYP outperforms both. For the GGA functionals, BLYP shows a lower error than PBE, occassionally displaying similar or lower MAD than TPSS or PBE0. These trends are generally followed for all directional components and may be an artifact of functional family (BLYP, B3LYP, B2PLYPD3-BJ), but some exceptions exist.With the pc-1 basis set, all functionals have similar MAD values for μ_z(R) and $||\boldsymbol{\mu}(R)||$, and cases exist where TPSS gives the lowest MAD for a given basis set. Perhaps the greater flexibility provided by a triple-ζ basis set allows MAD for a set of functionals to vary over a larger range.

$\boldsymbol{\mu}$, and therefore MAD($\boldsymbol{\mu}$), depends on two components, as defined in Equation 3.1. MAD($\boldsymbol{\mu}$) is lower than both MAD($\boldsymbol{\mu}^{c}(R)$) and MAD($\boldsymbol{\mu}^{p}(R)$). This can be explained through additional summary statistics for the individual contributions, shown in Tables A7- A12, and displayed in Figures A3 and A4. $\boldsymbol{\mu}^{c}(R)$ have negative mean deviations, while $\boldsymbol{\mu}^{p}(R)$ exhibits positive mean deviations, so the error in $\boldsymbol{\mu}$(R) is a fortuitous cancellation of errors. $\boldsymbol{\mu}^{c}$(R), given by Equation 3.2, depends on the charge of the group Λ connected to Ω. As discussed earlier, not all model chemistries overpredict or underpredict q(R) consistently, but the mean deviation is more often negative. Assuming atoms and other groups have similar deviations, this explains the underestimation of $\boldsymbol{\mu}^{c}(R)$. $\boldsymbol{\mu}^{p}$, Equation 3.3, is an integral so an analysis of the error by its components is not straightforward.

Since all directional components of the dipole have low error scores, the direction of the dipoles is reproduced adequately. This is further confirmed by plotting the directional components of the tested model chemistries and the reference values together and evaluating the lin-

ear relationship. Tables A13- A15 give regressions results for the x, y, and z components of $\boldsymbol{\mu}$(R) and values of the slopes, intercepts, and r^2 are illustrated in Figure 3.5. The lowest r^2 is 0.97 (SVWN/6-31G(d)), but most methods have r^2 of 0.99 or higher, proving that these relationships are highly linear. The observed trends in MAD of the dipole and its contributions can be rationalized by analysis of the regression slopes. Intercepts are near-zero and do not significantly affect the MD values. Increased deviation of the slope from 1 coincides with an increase in MAD for each μ_x, μ_y and μ_z. The slope is still close to 1 in all cases, confirming the adequate quantitative reproduction of the values.

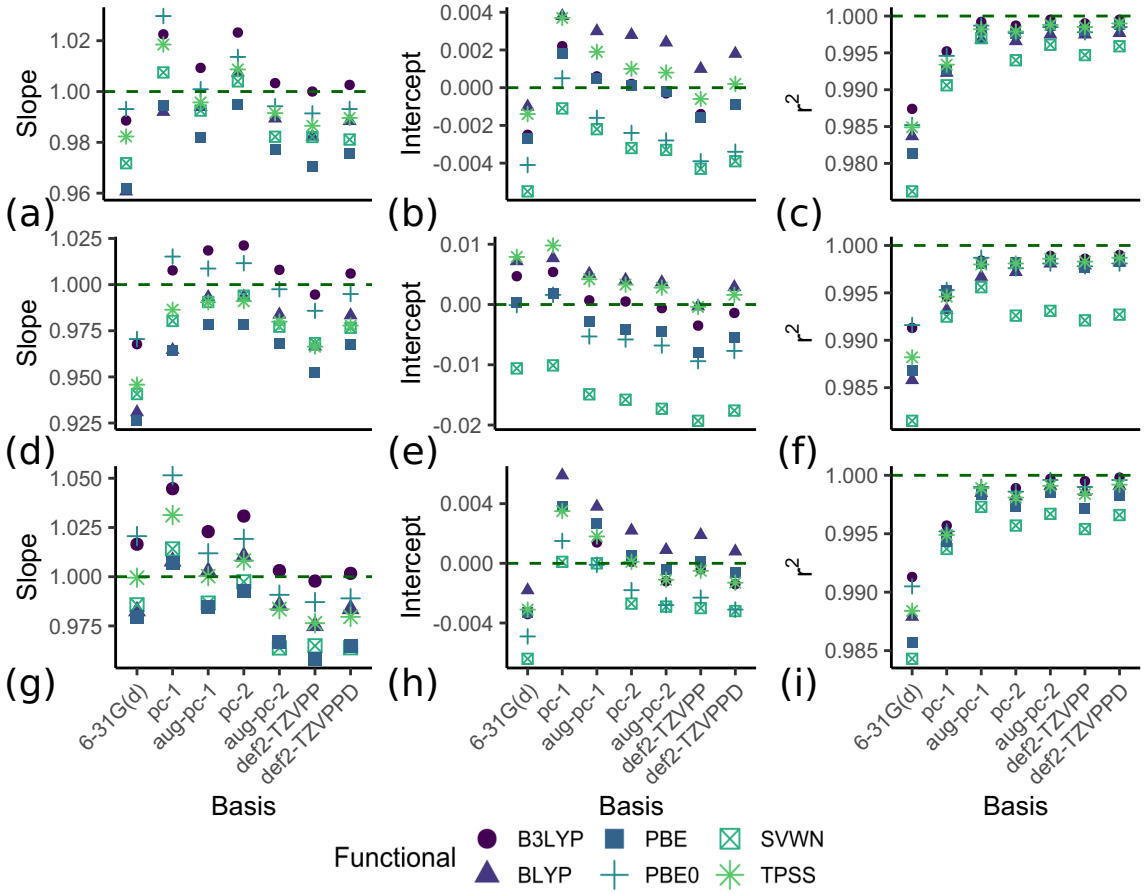

Figure 3.5: Regression parameters for: μ_x(R) (a) slope, (b) intercept, (c) correlation coefficient; μ_y(R): (d) slope, (e) intercept, and (f) correlation coefficient; μ_z(R) (g) slope (h) intercept and (i) correlation coefficient comparing tested model chemistries to B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green lines represent the ideal regression parameter in the relationship $y = x$

Error scores and linear plots both yielded the same information, that the group dipole is relatively unaffected by choice of model chemistry. All methods adequately reproduce the same trend as the reference values. We conclude that all model chemistries determine the same orientation of the dipole, within reasonable error. Additionally, any model chemistry could be used in regression studies. The most accurate values are from methods using meta-GGA or hybrid functionals and augmented-triple-ζ basis sets, so those methods are recommended for most accurate dipole values.

3.5.3 Group Quadrupole

The substituent contribution to the molecular quadrupole moment is a symmetric 3x3 matrix, and error statistics have been calculated for each of the six unique components and the overall magnitude of the quadrupole moment (provided in Tables A16- A22). Because many off-diagonal elements are approximately zero, we will focus only on the diagonal elements, and the overall quadrupole moment, $||\boldsymbol{Q}(\mathrm{R})||$, defined in Equation 3.7 below. Each of the three diagonal elements have similar MAD values, varying with model chemistry from 0.034 to 0.332 au. Each one exhibits similar trends associated with changes in functional and basis set, permitting a discussion of all of their trends simultaneously. MAD($||Q(\mathrm{R})||$) are slightly larger than those of the individual elements, expected based on its formulation (Equation 3.7).

$$||Q(\mathrm{R})|| = \sqrt{\frac{2}{3}\left(Q_{xx}^2 + Q_{yy}^2 + Q_{zz}^2\right)} \tag{3.7}$$

The quadrupole moment elements are composite properties depending on atomic positions, charges, $\boldsymbol{\mu}^p$, atomic quadrupoles and radial second moments as shown in Equation 3.4. As discussed previ-

ously, the charges and dipoles are mostly insensitive to chosen level of theory. Atomic positions are well reproduced.[26] The MAD, MD and other summary statistics of diagonal atomic quadrupoles, $Q_{ii}(\Omega)$ and second radial moments $\langle r^2 \rangle_\Omega$ were calculated and are shown in Tables A23-A26. The atomic quadrupoles are well reproduced, with MAD less than 0.1 au. $\langle r^2 \rangle_\Omega$ have higher MAD however, as high as 0.44 au. These improve as expected with the rungs of Perdew's ladder, and basis set size. The errors in these values inform the error in $Q_{ij}(R)$, and since they are generally accurate, we can expect $Q_{ij}(R)$ to be accurate as well.

For all methods, MAD($Q_{ij}(R)$) and MD($Q_{ij}(R)$) values are low. Using double-ζ basis sets, non-SVWN functionals have MAD close to or less than 0.2 au for all components, while triple-ζ basis sets see MAD close to or less than 0.1 au. A slight outlier for this is PBE, whose MAD(Q_{ii}(R)) range up to 0.16 au. As the quadrupole values range over 13 au, a 0.1 au error is small(0.8 %). MD(Q_{xx}(R)), and MD(Q_{yy}(R)) tend to have positive values, and MD(Q_{zz}(R)) is typically negative. Regardless of the consistent sign, the MD of these components is low in magnitude compared to MAD, showing that there is no large bias above or below the reference values. The MD of ||Q(R)|| does not reflect this, with large negative values for SVWN model chemistries, and mostly negative MD values for all other functionals (B3LYP is an exception).

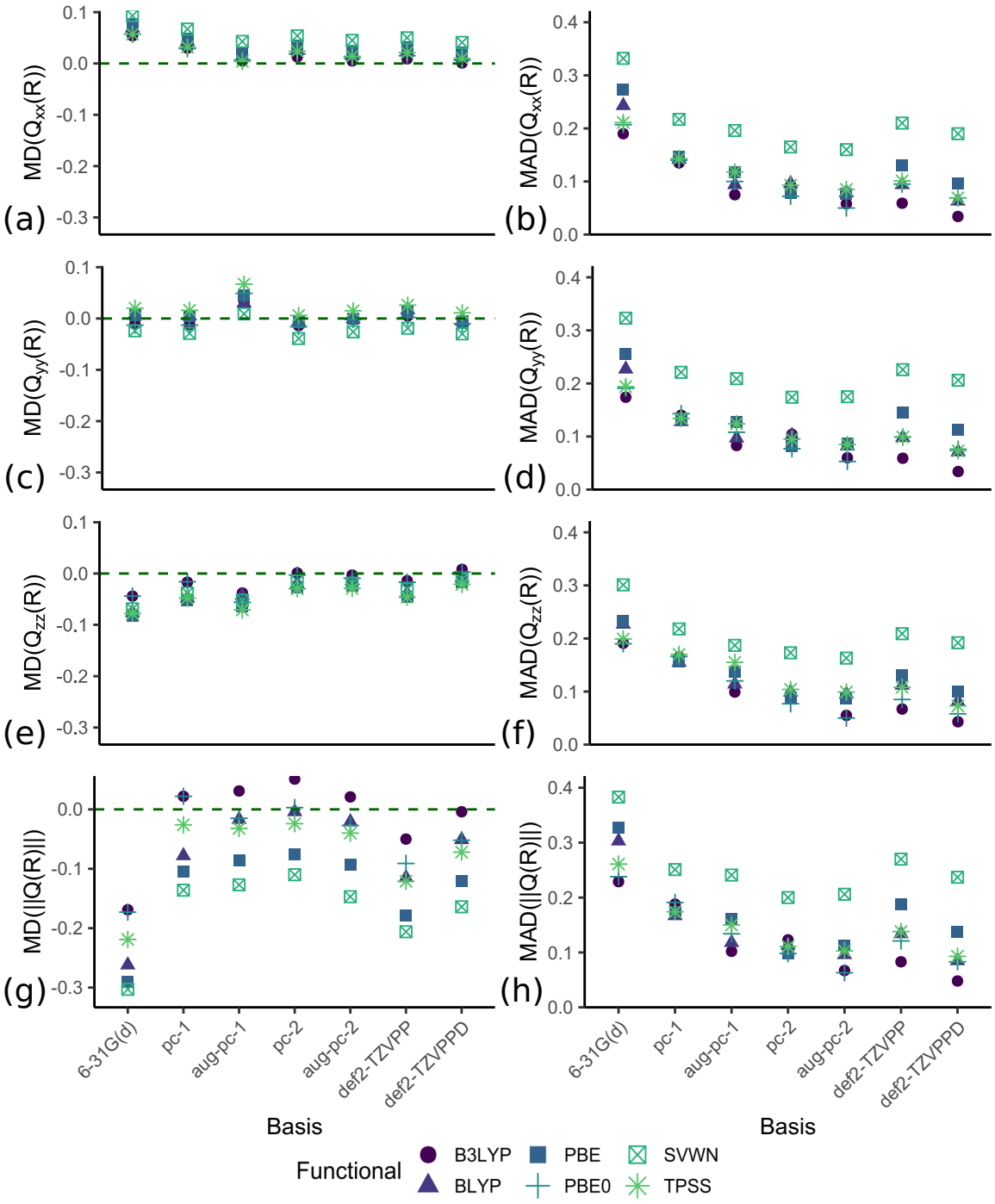

Figure 3.6: (a) MD(Q_{xx}(R)) (b) MAD(Q_{xx}(R)) (c) MD(Q_{yy}(R)) (d) MAD(Q_{yy}(R)) (e) MD(Q_{zz}(R)) (f) MAD(Q_{zz}(R)) (g) MD(||Q(R)||) (h) MAD(||Q(R)||) of tested model chemistries compared to reference B2PLYPD3-BJ/aug-cc-pV5Z values

Figure 3.6 shows the MAD and MD for the different model chemistries. SVWN/6-31G(d) is the least accurate. For SVWN and PBE with the pc-1 or aug-pc-1 basis sets, the MAD is similar to that for the def2-TZVPP and def2-TZVPPD basis sets, respectively. For other functionals, employing triple-ζ basis sets increases the accuracy of the quadrupole values as compared to double-ζ basis sets. Inclusion of diffuse functions generally lowers the MAD of a basis with some exceptions. Mimicking the charges and dipoles, the meta-GGA TPSS and the hybrid PBE0 typically have similar MAD for a given basis, and B3LYP often has a lower MAD than either of those. The least accurate functional is SVWN, with GGA functionals being more accurate, but still less accurate than the meta-GGA and hybrid functionals. The low MAD of the B3LYP functional agrees with previous computational studies on the accuracy with which the B3LYP functional calculates the molecular quadrupole.[65]

The diagonal elements of **Q**(R) were compared linearly to B2PLYPD3-BJ/aug-cc-pV5Z values. r^2, slopes and intercepts of the regressions are shown in Figure 3.7, and data is given in Tables A27- A29. A good linear relationship is observed for all components, with r^2 of 0.97 or greater. For GGA, meta-GGA or hybrid functionals with an aug-pc-1 or triple-ζ basis set, r^2 of 0.99 or greater are observed, showing a great conservation of linearity. Slopes and intercepts are withing 0.05 of the ideal values (unit-less, and au, respectively). As these all compare well to the $y = x$ ideal line, this agrees with the observed low MAD and MD values of the components. This confirms that model chemistry choice has little effect on quadrupole values.

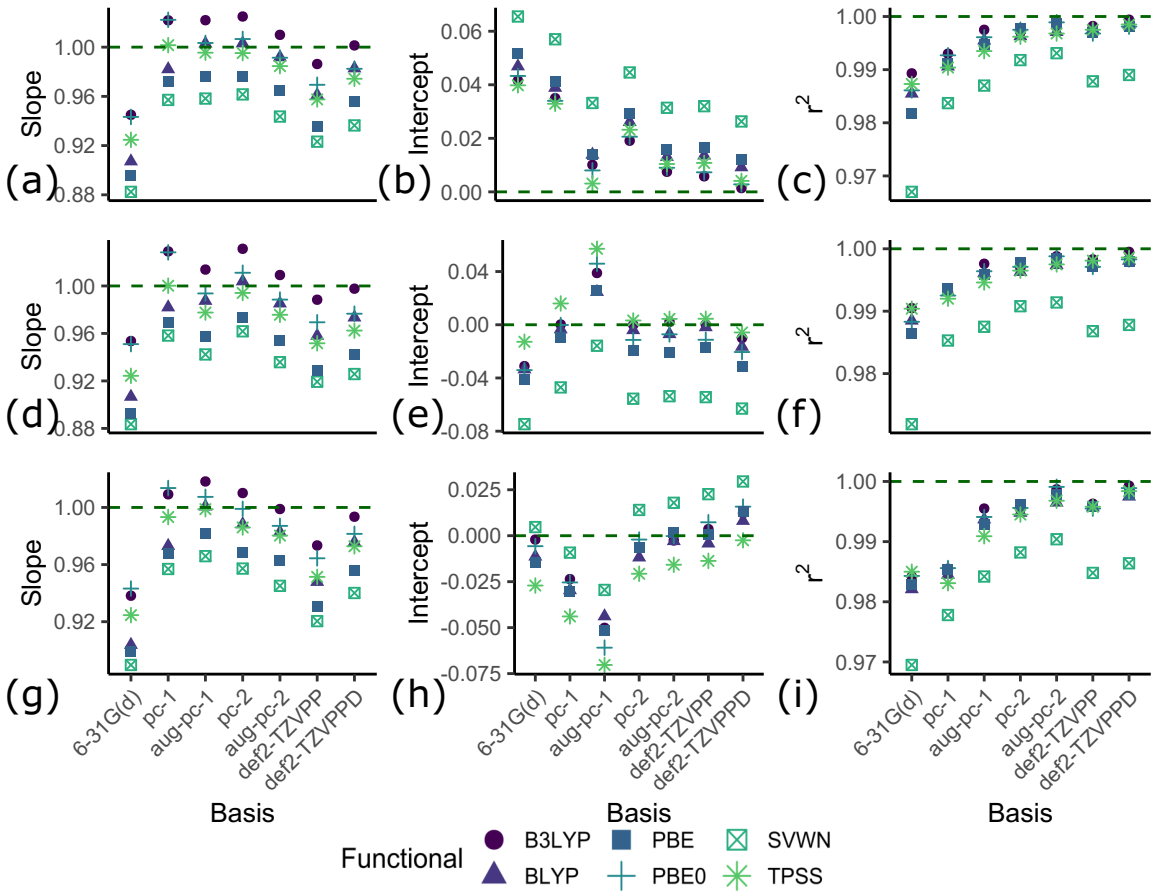

Figure 3.7: Regression parameters for: Q_{xx}(R) (a) slope, (b) intercept, (c) correlation coefficient; Q_{yy}(R): (d) slope, (e) intercept, and (f) correlation coefficient; Q_{zz}(R) (g) slope (h) intercept and (i) correlation coefficient comparing tested model chemistries to B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green lines represent the ideal regression parameter in the relationship $y = x$

3.5.4 Group Volume

The size of a functional group can be measured by the group volume. Functional group size does not change much between conformers (the largest percent difference is 0.7% for R=CH_2NH_2). Since these values are not near-zero, %MAD and %MD statistics are used. The %MAD and %MD values of Vol(R) are shown in Figure 3.8, with data given in Table A30. The choice of basis set has a large effect on %MAD for a given model chemistry. The error for model chemistries incorporating the 6-31G(d) basis set is high, but for other basis sets it is much lower. Including diffuse functions significantly affects group volume because

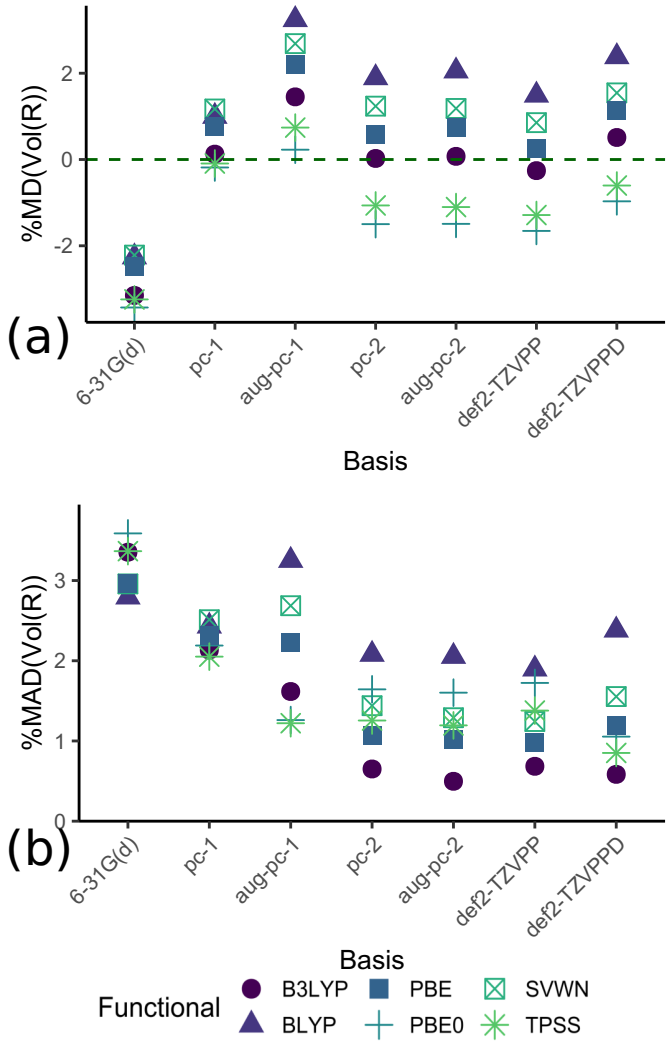

Figure 3.8: (a) MD(Vol(R)) and (b) MAD(Vol(R)) of tested model chemistries compared to reference B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green line represents the ideal MD value.

they allow electron density farther from the nuclei, generally increasing the volume of a group by moving the isodensity surface farther out towards infinity. Including diffuse functions therefore makes %MD more positive, and if %MD is already positive, then diffuse functions further raise %MD.

Choice of functional has a lower effect on observed %MAD than choice of basis set. Overall, %MAD near or less than 2% are observed for PBE, PBE0, TPSS, and B3LYP using any basis set other than 6-31G(d) and pc-1. The uniformity of these errors shows the relative insensitivity of Vol(R) to choice of functional. This makes sense as Vol(R) does not

directly depend on the density, only the volume of space in which the density is found. In spite of this relative insensitivity, the BLYP functional still has %MAD values nearly twice that of the other functionals with the same triple -ζ basis set.

Maximum negative errors are large for methods without diffuse functions since, as previously discussed, those methods underestimate Vol(R). In these cases, anions are typically the maximum error case. Specifically, O^- is the substituent that most often has the largest deviation. Anions are the maximum error case for unaugmented basis sets because diffuse functions are necessary to correctly describe anions. Methods incorporating diffuse functions see other functional groups having the maximum error, but these maximum errors are small compared to the range of values shown in Table 3.1.

In terms of linear plots, Vol(R) values are accurately reproduced. Table A31 lists the slope and intercepts of all regressions, and they are shown in Figure 3.9. The source of error in Vol(R) values is the intercept: all regressions have proper slopes near 1, but the intercepts deviate from 0, indicating a systematic error. The 6-31G(d) and pc-1 basis sets consistently have the most negative intercept, while all other intercepts are smaller than those values. Methods with those two basis sets are the only ones with r^2 values less than 0.99. Both these factors are likely due to these basis sets being the smallest studied. Other intercepts are less than 5 au in magnitude, fairly small for the size of the Vol(R) values. For a given basis set, upon inclusion of diffuse functions, the intercept becomes more positive, expected from the dependence of Vol(R) values on diffuse functions noted above. All of the above information taken into account leads to the conclusion that Vol(R) values linearly relate well to the reference values for all basis sets studied larger than 6-31G(d) and pc-1.

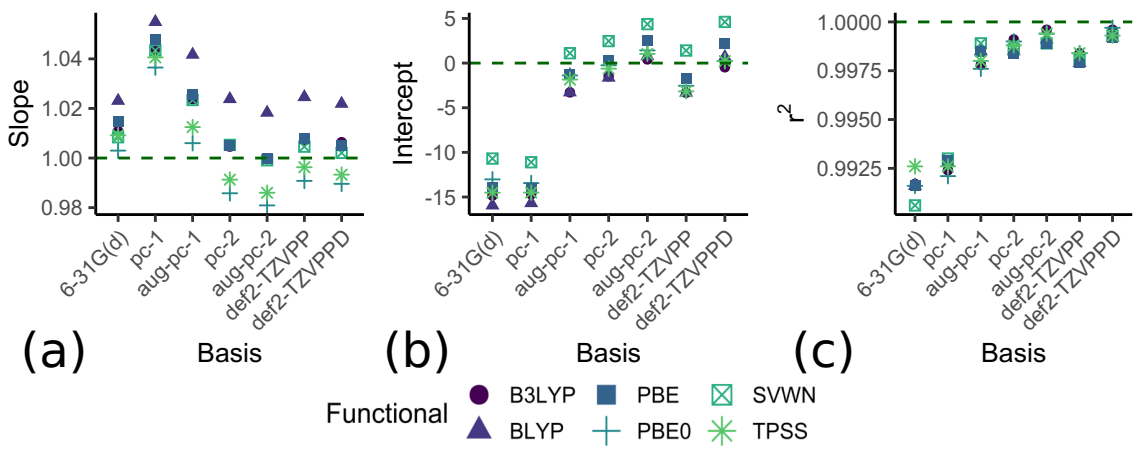

Figure 3.9: Vol(R) regression parameters (a) slope (b) intercept and (c) Correlation coefficient comparing tested model chemistries to B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green lines represent the ideal regression parameter in the relationship $y = x$

While the methods all relate linearly, they have different error statistics. While non-augmented basis sets could have performed better if the reference was not augmented, anions require diffuse functions to be properly described. That fact informs our recommendation that basis sets for calculating Vol(R) include diffuse functions. All diffuse function including basis sets yielded accurate values. Since functionals had a lower impact on Vol(R) any functional can be used, even SVWN. Trends are more important in comparing volumes, so if anions are not in the set of interest, any method can be used in calculating Vol(R).

3.6 Local Properties at Critical Points

3.6.1 Electron Density at the BCP

Figure 3.10(b) plots %MAD(ρ_c) values for the different model chemistries and Table A32 provides other comparative statistics. The %MADs span a broad range from 1.3 to 11.0%, showing that the choice of model chemistry has a notable effect on ρ_c values. Figure 3.10(a), showing %MD(ρ_c), reveals the consistent underestima-

tion of ρ_c; %MD(ρ_c) and %MAD(ρ_c) values are similar in magnitude. Nearly every ρ_c value is underestimated - only 5 model chemistries predict a single positive deviation for any substituent at all: B3LYP/pc-2, B3LYP/aug-pc-2, PBE0/pc-2, PBE0/aug-pc-2 and PBE0/def2-TZVPPD. SVWN/6-31G(d) exhibits a high maximum error of 0.054 au(14% error). Maximum error for non-SVWN functional and triple-ζ basis sets are closer to 6% or better, reasonable maximum errors.

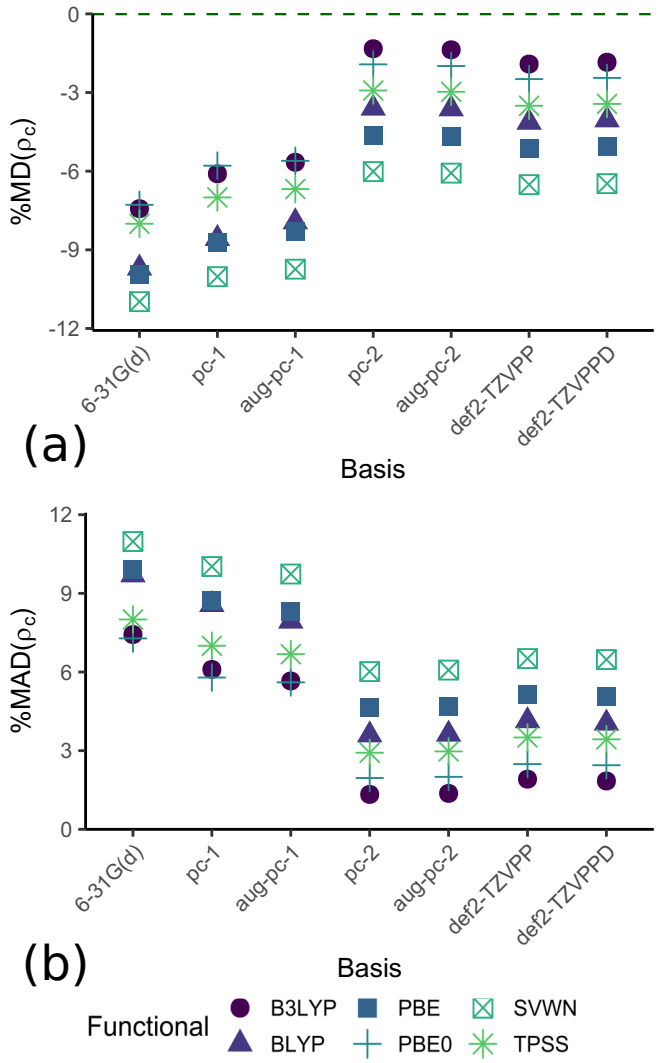

Figure 3.10: (a) MD(ρ_c) and (b) MAD(ρ_c) of tested model chemistries compared to reference B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green line represents the ideal MD value.

%MAD changes consistently with functional type and basis set. Double-ζ basis sets have much larger errors than the triple-ζ basis sets. The pc-2 and aug-pc-2 basis sets consistently perform slightly

better than the def2 basis sets in reproducing reference ρ_c values. Diffuse functions are not observed to have a significant impact because the density at the bond critical point is not significantly affected by diffuse functions. In terms of functional, %MAD decreases as functionals higher in Perdew's Jacob's ladder are used (for %MAD SVWN > GGA > meta-GGA > hybrid). The trends differ from the group properties, where TPSS and the hybrid functionals performed similarly. Specifically, the SVWN/6-31G(d) model performs the worst, judged by a %MAD of almost 11%. Such a large quantitative error limits their use in the applications concerned with magnitude of ρ_c. However, other methods (hybrid functionals with triple-ζ basis sets) show small %MAD(ρ_c), as low as 1.3% for B3LYP/pc-2, and these are suitable for use.

While the large %MAD(ρ_c) for SVWN and double-ζ basis sets limits the direct application of ρ_c values calculated using those methods, they may be sufficiently accurate for linear relationships. Determined ρ_c values were correlated to the reference; slopes, intercepts, and r^2 for all 42 correlations are given in Table A33 and Figure 3.11 shows these value pictorially. The worst performing method, SVWN/6-31G(d), still reproduces the trend in the reference values, having an r^2 greater than 0.99. Increased linearity is seen in other methods, with r^2 generally following the same trends as reduction in %MAD. All methods for the ρ_c values can be used in linear correlations, as shown by the r^2 values above 0.99 for all methods tested. While having a good r^2 value, methods differ in terms of %MAD(ρ_c) due to the deviation of the slope from 1. The values of the slope get closer to 1 as basis set size increases, and with climbing the rungs of Perdew's Jacob's ladder. For example, the B3LYP/aug-pc-2 values, exhibit a slope of 1.000, and therefore reproduce the magnitude of ρ_c well. The intercept is small on the scale of ρ_c values, and shouldn't affect the error much.

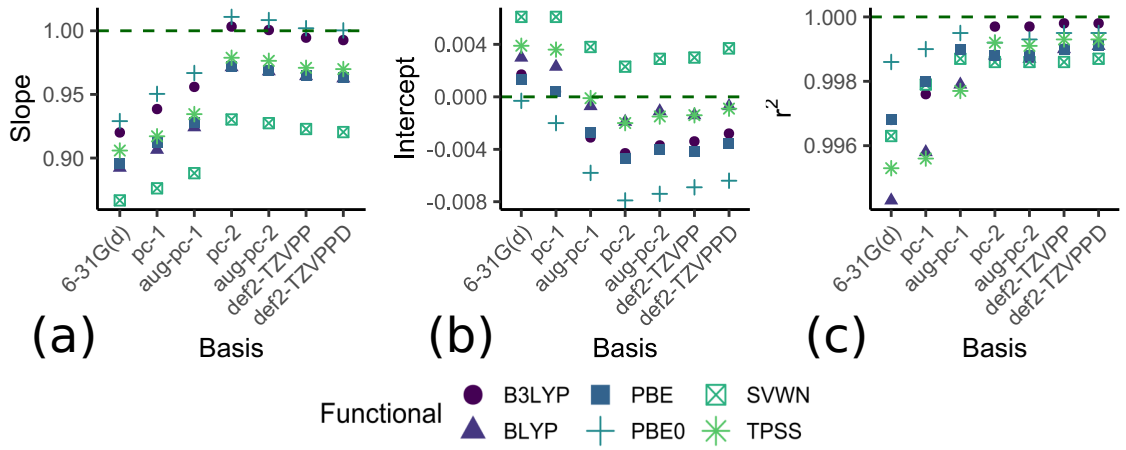

Figure 3.11: ρ_c regression parameters (a) slope (b) intercept and (c) Correlation coefficient comparing tested model chemistries to B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green lines represent the ideal regression parameter in the relationship $y = x$

3.6.2 Distance to the BCP

The distance from the hydrogen atom to the BCP, r(H-BCP), has been employed in electronegativity descriptors in the past.[12,13] In this case, the observation of monotonically decreasing r(H-BCP) with increasing Z_A in a row of the periodic table was employed. Thus, r(H-BCP) is another interesting descriptor that can be related to the electronic properties of a substituent. r(H-BCP) may also provide insight into the source of the error in ρ_c. Table A34 shows the error statistics in r(H-BCP) for the model chemistries under study. %MD(r(H-BCP)) and %MAD(r(H-BCP)) are shown in Figure 3.12(a) and (b), respectively. The relative error in these values is lower than that for ρ_c. r(H-BCP) is neither consistently underestimated nor overestimated, judging from the difference in magnitudes of %MAD and %MD, and the tendency to over- or underestimate is not consistent between functionals.

%MAD exhibits trends with the choice of functional, but with additional dependence on chosen basis set. Hybrid functionals consistently have low %MAD. GGA and meta-GGA functionals however perform differently with choice of basis set. With the pc-2 basis set, these function-

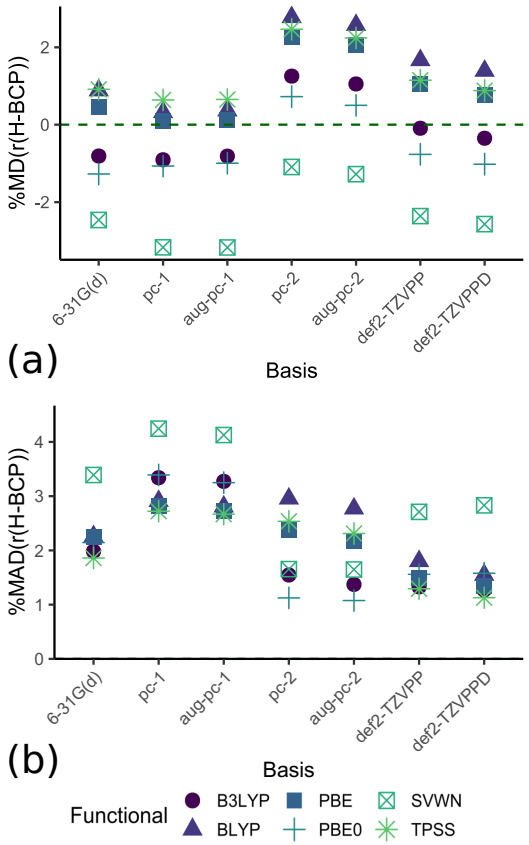

Figure 3.12: (a) %MD(r(H-BCP)) and (b) %MAD(r(H-BCP)) of tested model chemistries compared to reference B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green line represents the ideal MD value.

als perform worse than hybrid functionals, but are similar to the hybrid functionals with the def2-basis sets. SVWN usally performs the worst of the functionals, except for with the pc-2 basis sets, where it performs similar to the hybrid functionals.

Trends in basis set are observed in %MAD(r(H-BCP)) values. The pc-1 basis set performs worse than the 6-31G(d) basis set. Using GGA and meta-GGA functionals, pc-2 basis sets exhibit similar %MAD to the pc-1 basis sets, and increased error from the 6-31G(d) basis set. The def2 basis sets exhibit lower %MAD from the double-ζ basis sets. Additionally, inclusion of diffuse functions does not significantly alter %MAD of a method.

The r(H-BCP) of tested model chemistries are linearly related to the reference values. The regression parameters are given in Table A35.

With exception of the meta-GGA and hybrid functionals using the pc-1 basis sets, all r^2 values are larger than 0.99. The trends in slopes relate to the trends in %MAD. Triple-ζ basis set slopes are closer to 1 than the pc-1 basis sets, but similar or worse than the 6-31G(d) basis sets. Due to the similar slopes between the 6-31G(d) and def2 the lower %MAD of def2 is probably due to the slightly better linearity of def2 (0.9975-0.9993 vs. 0.9944-0.9975). Overall, the error in r(H-BCP) values is low, and any method could be used to obtain accurate enough values of this parameter.

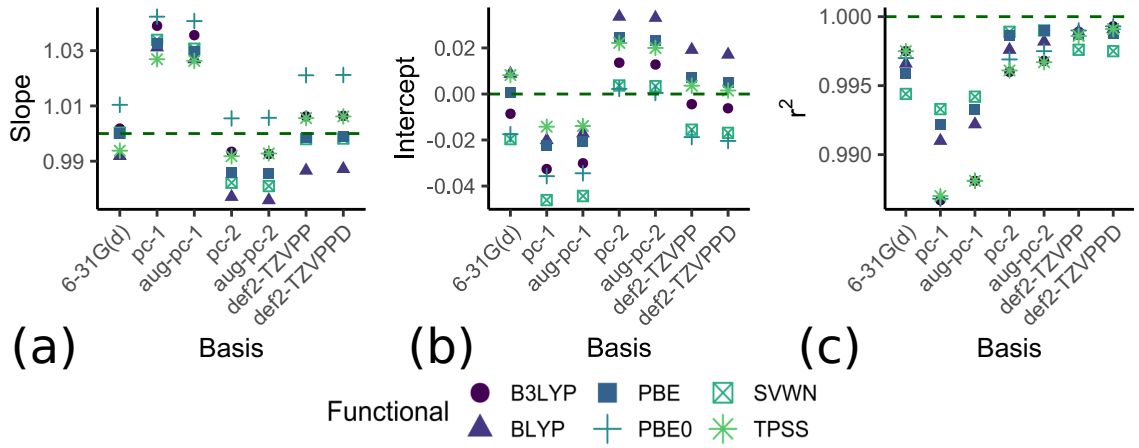

Figure 3.13: r(H-BCP) regression parameters (a) slope (b) intercept and (c) Correlation coefficient comparing tested model chemistries to B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green lines represent the ideal regression parameter in the relationship $y = x$

Figure 3.14 compares the error in ρ_c to that in r(H-BCP) using the worst case scenario (SVWN/6-31G(d)). A clear relationship is seen for substituents bound to hydrogen though C, N or O. As the BCP is located farther from the hydrogen nucleus (closer to the heavy atom), $\Delta\rho_c$ approaches zero. Substituents with second row elements attached to the substrate hydrogen (S, Si, and P), do not follow the relationship. This relationship becomes obfuscated using model chemistries with lower Δr(H-BCP) since, for these methods, the error in r(H-BCP) takes place in a very small range. The error in ρ_c appears similar to a vertical line in

these plots. The vertical appearance is due to the variance of $\Delta\rho$ while Δr(H-BCP) occurs over a very small range (smaller than the error in the otherwise noted relationships between the two values).

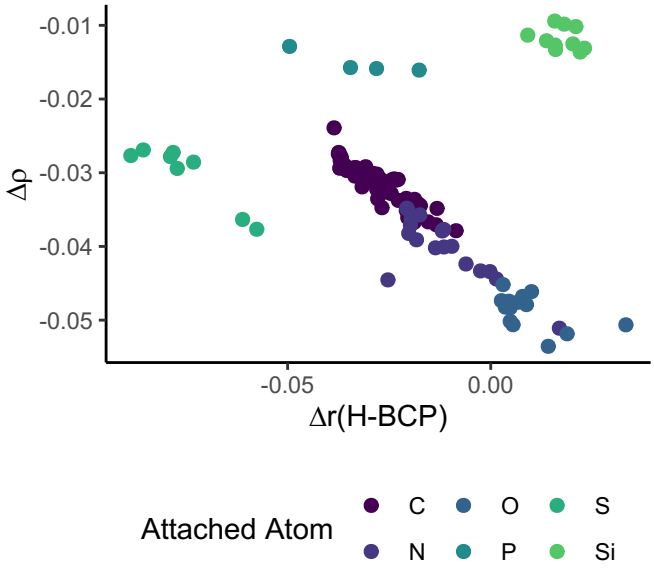

Figure 3.14: Comparison of the SVWN/6-31G(d) error in ρ_c, $\Delta\rho$, and the error in r(H-BCP), Δr(H-BCP).

3.6.3 Laplacian of the Electron Density at the BCP

The eigenvalues, λ_i, of the second derivative matrix, and their sum, $\nabla^2\rho_c$, reveal information on the accumulation or depletion of electron density at the BCP, which can be interpreted in terms of the covalence of the bond. Figure 3.15 shows %MD and %MAD of $\nabla^2\rho_c$ and λ_i for the 42 model chemistries (values in Tables A36-A39). These are very sensitive to the choice of model chemistry, with %MAD as high as 30%. Obviously, %MAD of 30% are unusable from a quantitative standpoint, and even the minimum value, while usable, is not ideal. $\nabla^2\rho_c$, $\lambda_{1,c}$ and $\lambda_{2,c}$ are typically overestimated, but negative deviations occur regularly, unlike for ρ_c values.

Even with larger %MAD($\nabla^2\rho_c$), the trends in %MAD($\nabla^2\rho_c$) values mimic those previously seen for ρ_c. Double-ζ basis sets do not give ac-

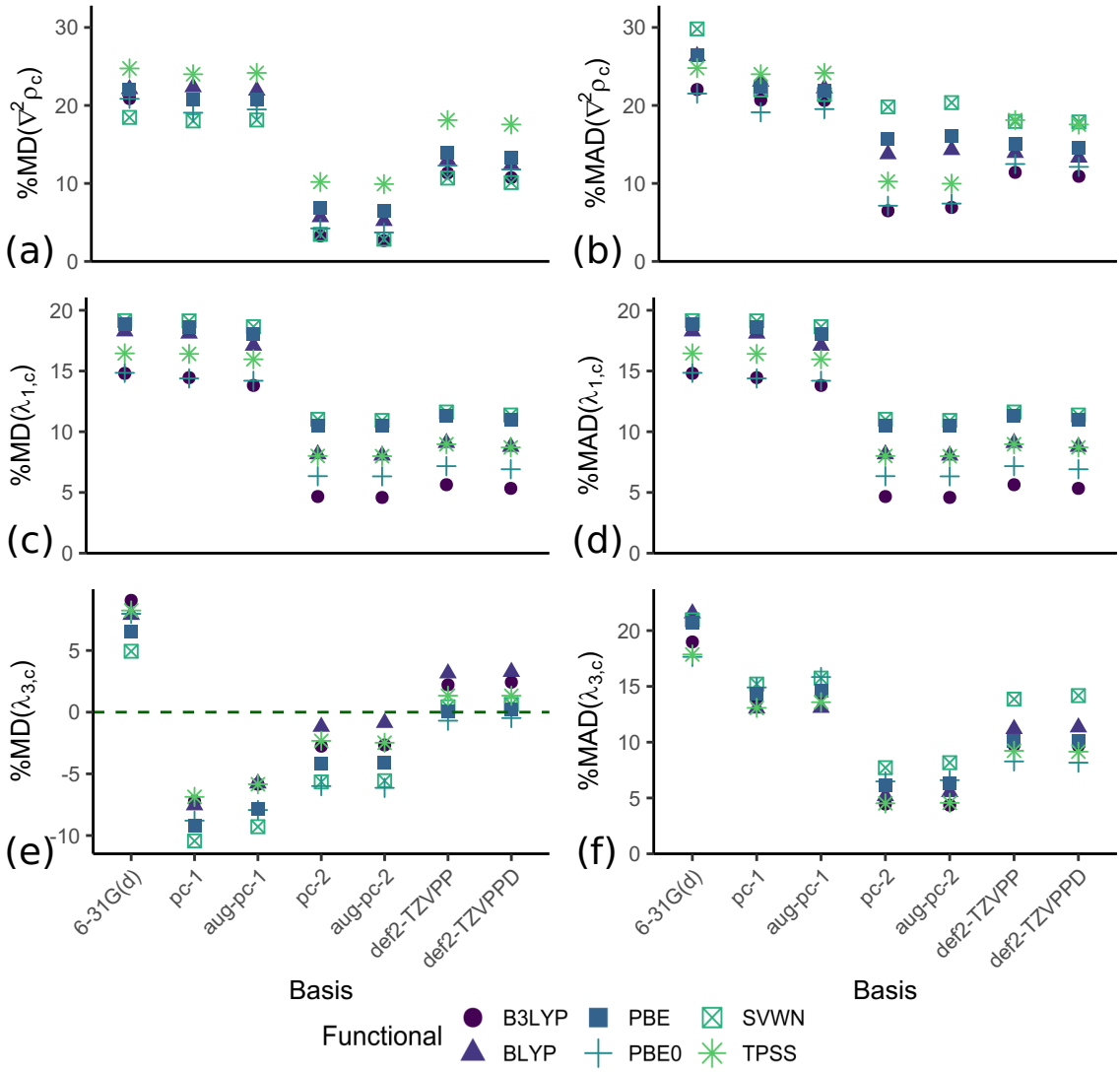

Figure 3.15: (a) %MD($\nabla^2\rho_c$) (b)%MAD($\nabla^2\rho_c$) (c) %MD($\lambda_{1,c}$) (d) %MAD($\lambda_{1,c}$) (e) %MD($\lambda_{3,c}$) (f) %MAD($\lambda_{3,c}$) of tested model chemistries compared to reference B2PLYPD3-BJ/aug-cc-pV5Z values

curate %MAD, regardless of functional, and triple-ζ basis sets improve performance with respect to double-ζ. The pc-2 and aug-pc-2 basis sets most accurately reproduce reference values for a given functional. Diffuse functions do not affect %MAD significantly, with their inclusion accompanied by a minor, but unimportant, change in %MAD($\nabla^2\rho_c$). For some basis sets, the %MAD values do not follow the expected trend with respect to functional. Using the def2 basis sets, TPSS has similar error to SVWN, and with both pc-1 basis sets, SVWN performs better than GGA and TPSS functionals. However, in all cases, hybrid function-

als perform the best, and are therefore recommended for use.

%MAD($\nabla^2\rho_c$) values are higher than those for the individual eigenvalues (Tables A37- A39) due to compounded errors. Because of the symmetry of the R-H bond, λ_1 and λ_2 are nearly identical in magnitude, as shown in Figure A5. Both λ_1 and λ_2 are always overestimated, but λ_3, is not consistently over- or underestimated (observed from its low %MD relative to %MAD). The two similar positive %MD of λ_1 and λ_2 therefore dominate the error for $\nabla^2\rho$, causing it to be overestimated. Because it is generally overestimated, the maximum negative deviations for $\nabla^2\rho_c$ are quite small. For double-ζ basis sets and the SVWN functional, maximum percent errors close to 40% are typical, which are extremely large. For triple-ζ basis sets and hybrid functionals, the maximum percent error decreases to about 16%, which, while large is reasonable for a worst case error. Caution should be used in applying $\nabla^2\rho_c$ values.

The error in ρ_c can explain the error in the eigenvalues. Figure 3.16(a) and 3.17(a) relates the error in ρ_c to the error in $\lambda_{1,c}$ and $\lambda_{3,c}$, respectively, for SVWN/6-31G(d) values. Figure 3.16(b) and 3.17(b) also show the relationship between $\lambda_{i,c}$ and ρ_c, and a clear relationship is present. For λ_1, a general relationship is observed for each different atom through which the substituent attaches to hydrogen. In these relationships, as $|\Delta\rho|$ decreases, $|\Delta\lambda_{1,c}|$ decreases. $|\Delta\rho|$ decreasing corresponds to the density increasing, which leads to λ_1, the negative curvature, becoming more negative due to a sharper peak. No such relationships are present for $\lambda_{3,c}$, where the substituents with the same attached atom have nearly constant $\Delta\lambda_{3,c}$ with varying $\Delta\rho_c$, as a result of the curvature along the bond path being dominated by the atomic densities and minor changes in ρ_c have no effect. Similar relationships are seen for other model chemistries, but for methods with

low %MAD, the relationships are not as clear due to scattering in the group, likely from the small ranges of error in the values. In summary, $\Delta\lambda_{1,c}$ depends on both the attached atom and $\Delta\rho_c$, while $\Delta\lambda_{3,c}$ depends solely on the attached atom.

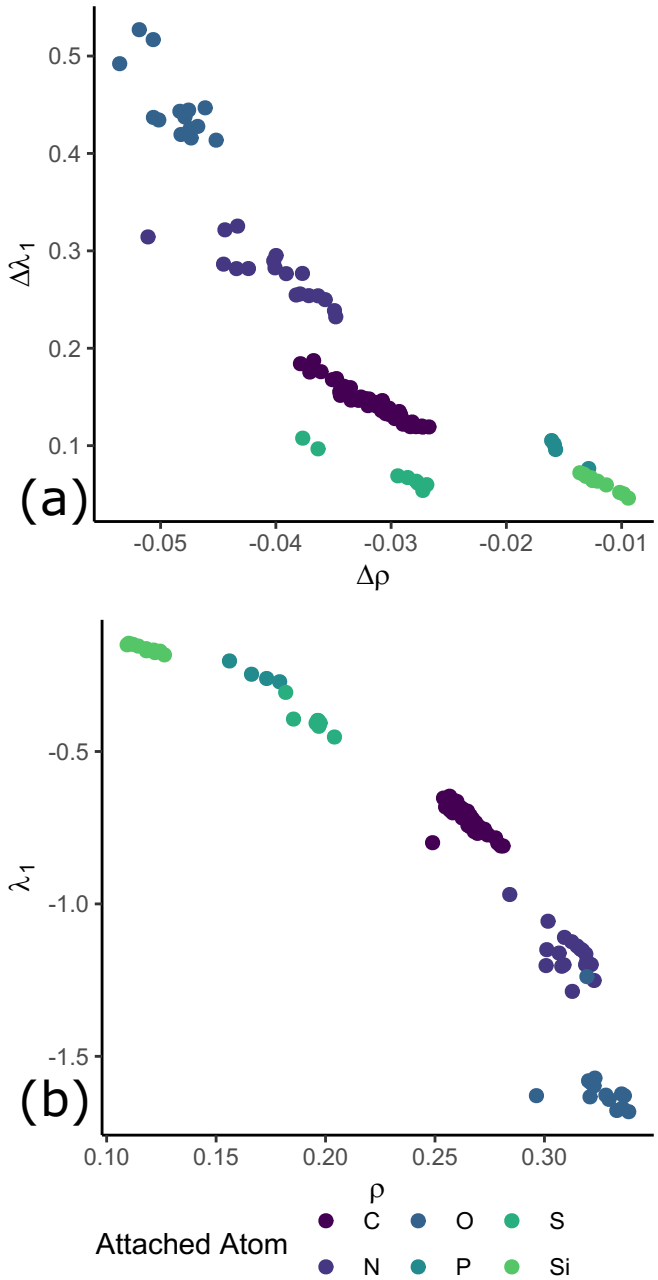

Figure 3.16: Comparison of (a) the error in ρ_c, $\Delta\rho$, to the error in $\lambda_{1,c}$, $\Delta\lambda_{1,c}$ (b) ρ_c to $\lambda_{1,c}$. Values calculated at SVWN/6-31G(d).

$\nabla^2\rho_c$ values are not quantitatively well produced, but they could be applied in correlation analyses. To test if the $\nabla^2\rho_c$ for any model chemistry can be used for this purpose, linear plots were constructed. Ta-

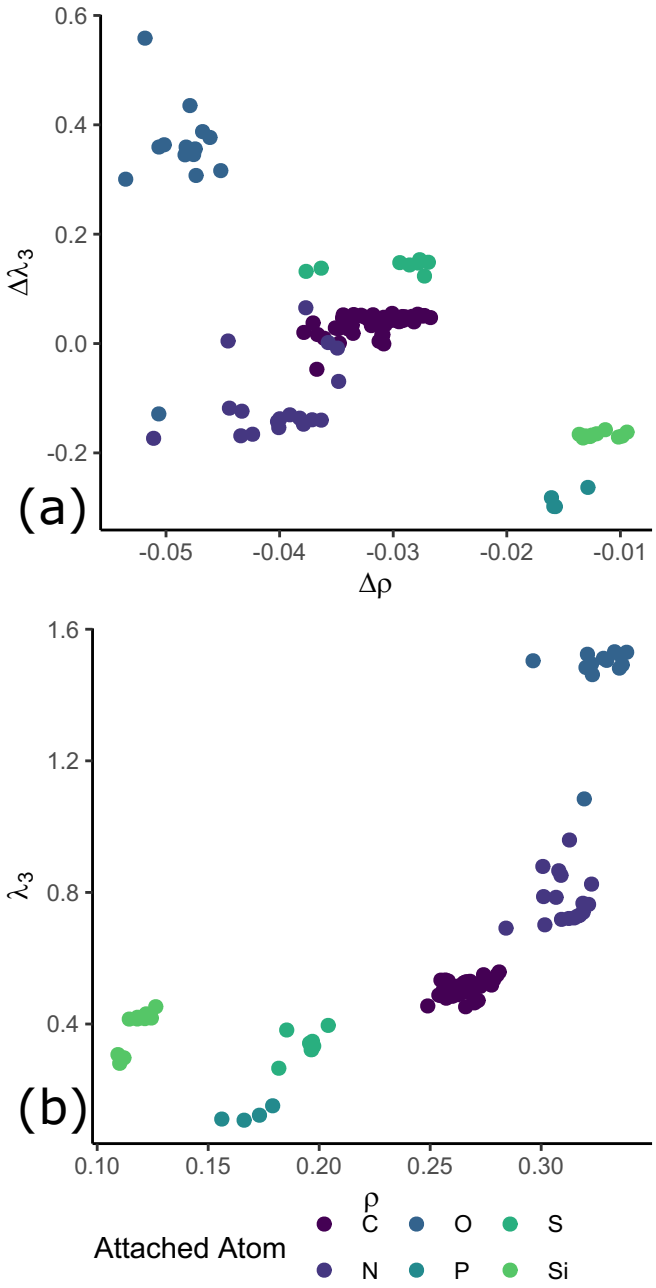

Figure 3.17: Comparison of (a) the error in ρ_c, $\Delta\rho$, to the error in $\lambda_{3,c}$, $\Delta\lambda_{3,c}$ (b) ρ_c to $\lambda_{3,c}$. Values calculated at SVWN/6-31G(d).

ble A40 gives a table containing all slopes, intercepts and r^2 values for the linear plots, shown graphically in Figure 3.18. Unlike ρ_c, the methods with higher %MAD do not linearly relate well to the reference values. The 6-31G(d) basis set has low r^2 values, 0.94 or less, revealing scatter around the linear relationship for those methods. Other double-ζ basis sets have moderate r^2 values (0.95-0.98), which are not sufficiently accurate for use in correlation analyses. All methods employing

a triple-ζ basis set display good linearity, with high r^2 values. Therefore, even for QSAR/QSPR relationships, triple-ζ basis sets are recommended. The r^2 do not change as significantly when varying functional. For triple-ζ basis sets, not much change is seen in r^2 for these, so any functional with the triple-ζ basis set can be used.

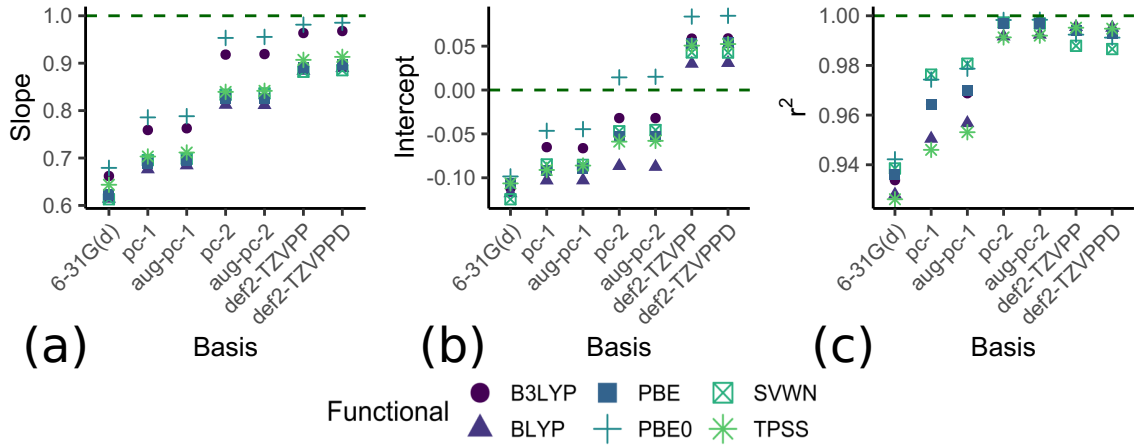

Figure 3.18: $\nabla^2\rho_c$ regression parameters (a) slope (b) intercept and (c) Correlation coefficient comparing tested model chemistries to B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green lines represent the ideal regression parameter in the relationship $y = x$

3.6.4 Energy Densities at the BCP

Kinetic energy density, G_c, potential energy density, V_c, and their sum H_c calculated at BCPs aid in characterizing bonds. Tables A41- A43 give summary statistics characterizing the performance of the 42 model chemistries in reproducing the reference values, with %MD and %MAD displayed in Figure 3.19. Like $\nabla^2\rho_c$, high %MAD are observed for G_c, with all functionals, aside from SVWN, performing similarly. The basis set effect is simlar to that seen in other properties: 6-31G(d) has the highest error, followed by the other double-ζ basis sets, and triple-ζ basis sets have the lowest error. Figure 3.19(a) shows G_c values are nearly always underestimated; the %MD(G_c) are similar in magnitude to %MAD(G_c). Maximum errors close to 30% for double-ζ basis sets and

decrease to around 16% for triple-ζ basis sets mimic the values seen for $\nabla^2\rho_c$, with percent errors.

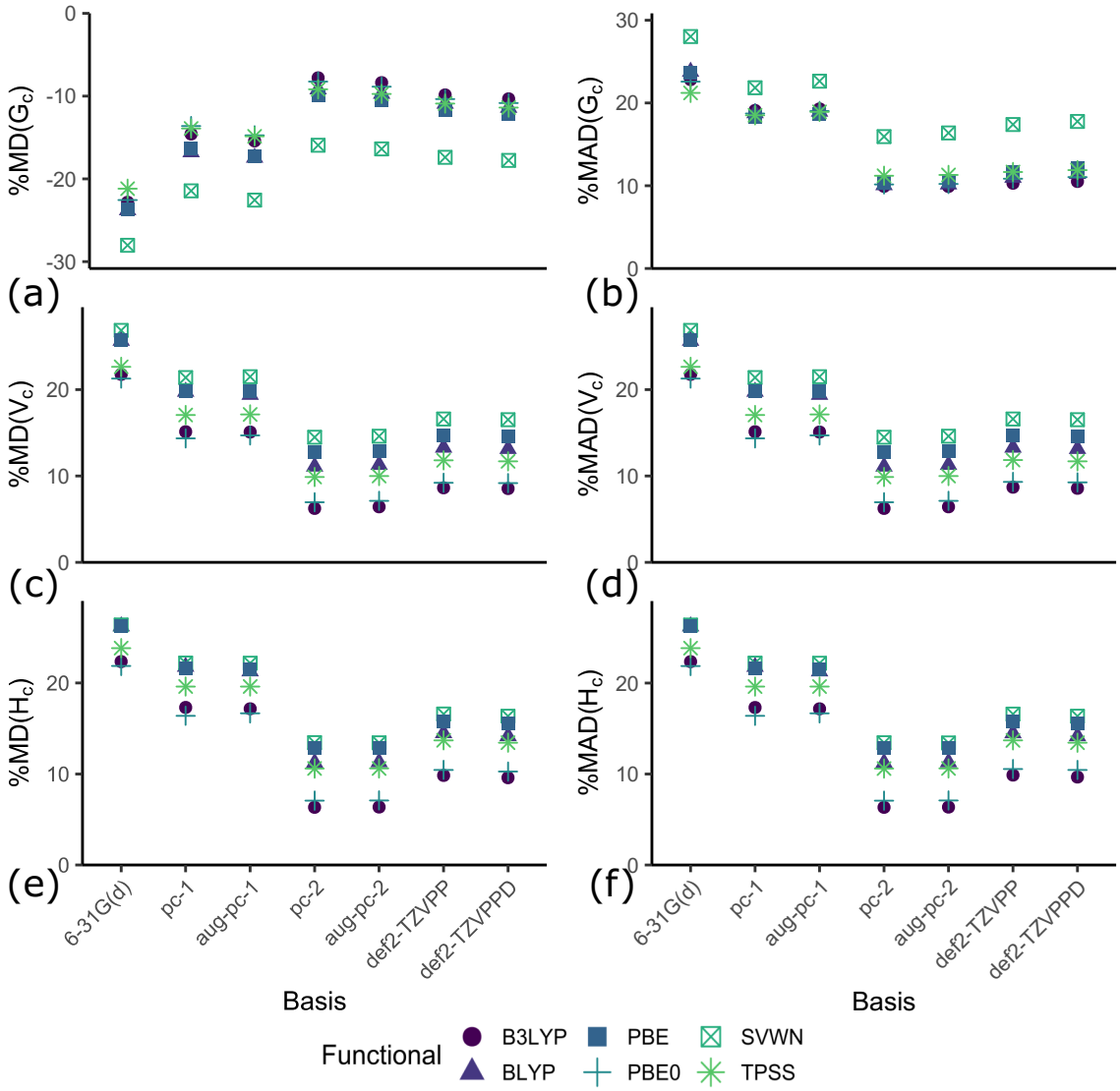

Figure 3.19: (a) %MD(G_c) (b) %MAD(G_c) (c) %MD(V_c) (d) %MAD(V_c) (e) %MD(H_c) (f) %MAD(H_c) of tested model chemistries compared to reference B2PLYPD3-BJ/aug-cc-pV5Z values

Figure 3.19(c) and (d) shows %MD(V_c) and %MAD(V_c), respectively. %MAD(V_c) repeats the same trends with basis set size as %MAD(G_c), but %MAD(V_c) has different functional dependence. A slight improvement is found going from SVWN to GGA, meta-GGA functionals and GGA are roughly equivalent in accuracy, and hybrid functionals perform better than the meta-GGA functional. Unfortunately, even the best performing methods have a slightly higher %MAD than ρ_c and group prop-

erties. V_c are overestimated, as judged by positive MD illustrated in Figure 3.19(c). As the %MD(V_c) and %MAD(V_c) are equal in magnitude, nearly all deviations are positive. Only six cases exhibit negative deviations, shown in Table A42. The same magnitude maximum deviation behaviour for V_c is seen as for G_c. Similar magnitudes of %MAD are observed for both V_c and G_c, even though G_c values are smaller in magnitude. This highlights the benefit of using a percent scale - even though the absolute deviations differ in magnitude, the relative deviations show that, on the scale of each property, the deviations are similar.

Since V_c is overestimated and G_c is underestimated, %MAD(H_c), shown in Figure 3.19(f), are lower than the %MAD for either of its components due to cancellation of errors. Otherwise, the same trends seen in V_c are observed here because V_c dominates H_c. As H_c relates to the covalency of a bond, through its relation to $\nabla^2\rho$ shown in Equation 3.8, all bonds in this study can be said to be dominated by covalent character. Maximum errors are similar in magnitude to maximum errors of V_c and G_c. All maximum error cases are for anions or cations, likely due to a change in the bonding character from typical covalent bonds. For all energy densities, even the best case scenario errors are not small. Hybrid functionals with triple-ζ basis sets are recommended.

$$\frac{\hbar^2}{4m}\nabla^2\rho = V + 2G \tag{3.8}$$

Regardless of the accuracy of the magnitude of the values, they can still be used in linear relationships. Results are shown in Tables A44-A46, and displayed in Figures 3.20- 3.22. G_c do not follow an exact linear relationship with the reference values; many model chemistries exhibit r^2 close to 0.94 or 0.97. Additionally, slopes deviate greatly from 1, and are likely expected to be the main source of error in G_c values.

The same deviation is not seen in similar plots for V_c and H_c, as shown in Figure 3.21 and Figure 3.22. For these properties, large scatter is observed for 6-31G(d) values, but the plot improves to linearity using any other basis set. Since the hybrid functional and triple-ζ basis set show a good linear relationship, V_c and H_c can be used in linear correlations. The effect of the inaccurate G_c values is not observed in H_c because of its small contribution. Note that the slopes of all correlations of the tested model chemistries to the reference values are significantly less than one. This highlights that the tested model chemistries overestimate the magnitude of those energy densities as compared to the reference method. The overestimation leads to the positive MDs observed in Tables A42 and A43. The intercepts of those same correlations are relatively small, but are statistically different from 0. The intercept is larger for correlations involving model chemistries with double-ζ basis sets than those using triple-ζ, highlighting the increased accuracy of the latter methods in determining the magnitude and trend of the properties.

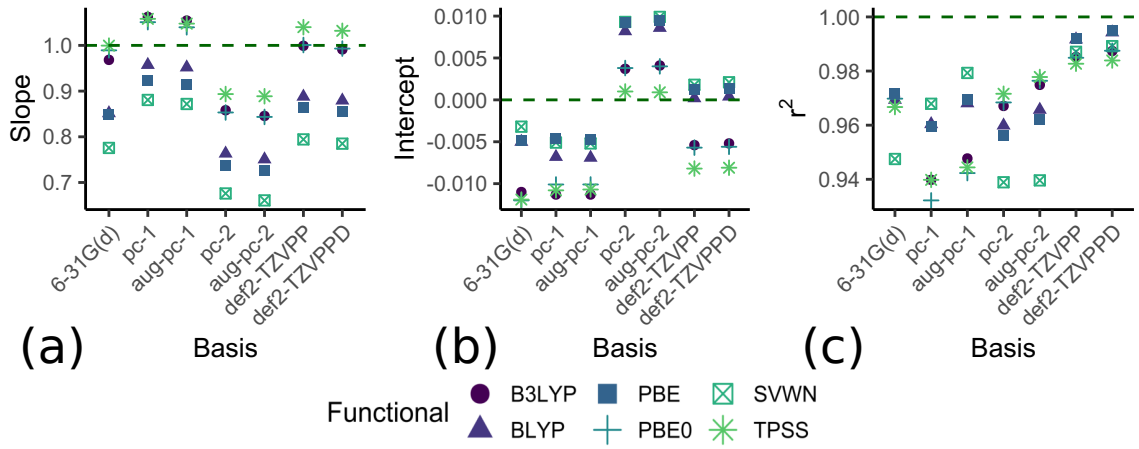

Figure 3.20: G_c regression parameters (a) slope (b) intercept and (c) Correlation coefficient comparing tested model chemistries to B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green lines represent the ideal regression parameter in the relationship $y = x$

3.7 Conclusions

QTAIM group properties show great promise as inherent substituent constants that can be widely applicable to a variety of chemical problems. Due to the large variety of computational model chemistries available, the ability of various model chemistries to be used in different applications was investigated. Integrated properties quantitatively compare to the reference B2PLYPD3-BJ/aug-cc-pV5Z values for 116 substituents, but the SVWN functional and 6-31G(d) basis set showed significantly higher error than other functionals and basis sets. Two cases occured where 6-31G(d) was qualitatively unable to describe the molecule. All group properties displayed high linearity in relation to the reference values. For BCP properties, double-ζ basis sets, and the GGA and SVWN functionals did not show quantitative agreement or a linear relationship with the reference values, while hybrid functionals with triple-ζ basis sets did.

The presented evidence shows that for correlation studies with group

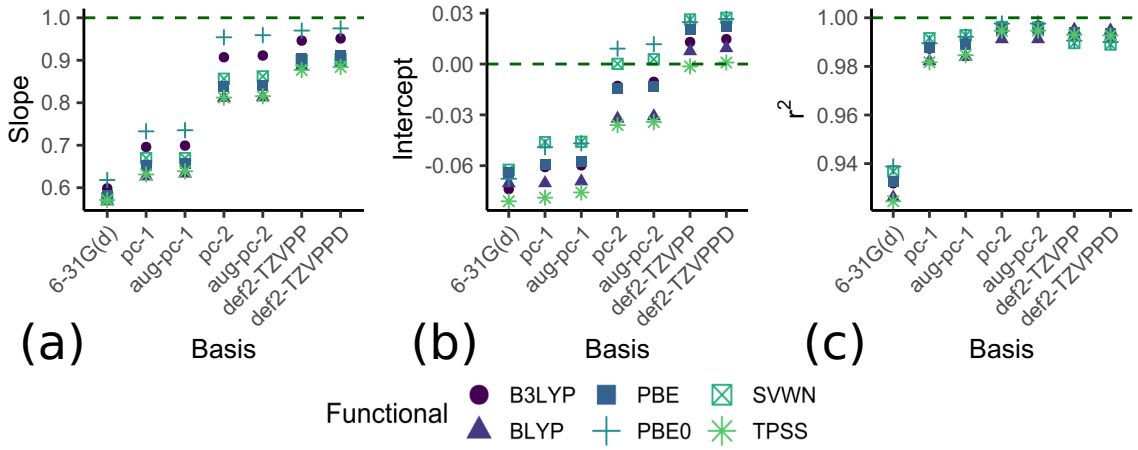

Figure 3.21: V_c regression parameters (a) slope (b) intercept and (c) Correlation coefficient comparing tested model chemistries to B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green lines represent the ideal regression parameter in the relationship $y = x$

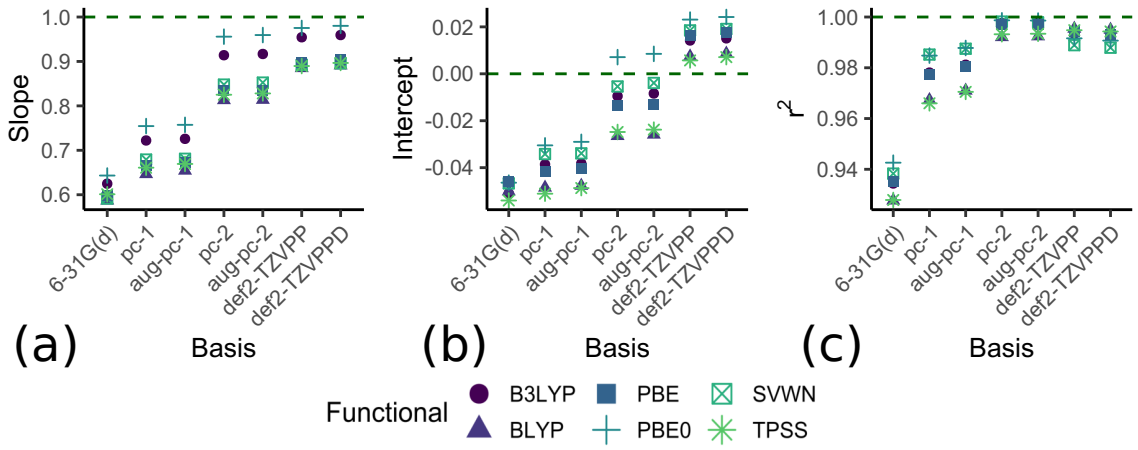

Figure 3.22: H_c regression parameters (a) slope (b) intercept and (c) Correlation coefficient comparing tested model chemistries to B2PLYPD3-BJ/aug-cc-pV5Z values. Dashed green lines represent the ideal regression parameter in the relationship $y = x$

properties as regressors, any model chemistry can be used, aside from the 6-31G(d) basis, which does not consistently describe molecules correctly. Quantitatively, any functional and basis compare fairly well to the reference values. However, for increased accuracy, meta-GGA or hybrid functionals with triple-ζ basis sets are recommended. If quadrupoles or dipoles are among properties of interest, augmenting the basis set with diffuse functions is recommended. For BCP properties, regardless of application, only hybrid functionals with triple-ζ basis sets are recommended for use. Diffuse functions have minimal impact on BCP properties however and are not required for these studies.

This paper presents a first step in determining *ab initio* substituent constants. An extensive study on the extent of transferabilty of the properties between different systems is under way. In implementing these type of descriptors, constants for new substituents can be evaluated easily computationally prior to experiment, enabling computational screening of potential molecules or substituents prior to any syntheses.

3.7.1 Acknowledgements

This research was funded by the National Sciences and Engineering Research Council of Canada, Government of Ontario through the Ontario Graduate Scholarship, and the Lakehead University Faculty of Graduate Studies. This research was enabled in part by support provided by the Shared Hierarchical Academic Research Computing Network (www.sharcnet.ca) and Compute Canada (www.computecanada.ca).

References

[1] Moss, R. A. and Krogh-Jespersen, K., *Tetrahedron Lett.* **2016**, *57*, 611–614.

[2] Hammett, L. P., *J. Am. Chem. Soc.* **1937**, *59*, 96–103.

[3] Taft Jr, R. W., *J. Am. Chem. Soc.* **1953**, *75*, 4231–4238.

[4] Charton, M., *J. Org. Chem.* **1964**, *29*, 1222–1227.

[5] Yasuhide, Y., Yuho, T. and Masami, S., *Bull. Chem. Soc. Jpn.* **1972**, *45*, 1198–1205.

[6] Ponec, R., Amat, L. and Carbó-Dorca, R., *J. Comput. Aided. Mol. Des.* **1999**, *13*, 259–270.

[7] Matta, C. F. and Arabi, A. A., *Futur. Med. Chem.* **2011**, *3*, 969–994.

[8] Gadre, S. R. and Suresh, C. H., *J. Org. Chem.* **1997**, *62*, 2625–2627.

[9] Szatylowicz, H., Jezuita, A., Siodła, T., Varaksin, K. S., Domanski, M. A., Ejsmont, K. and Krygowski, T. M., *ACS Omega* **2017**, *2*, 7163–7171.

[10] Siodla, T., Oziminski, W. P., Hoffmann, M., Koroniak, H. and Krygowski, T. M., *J. Org. Chem.* **2014**, *79*, 7321–7331.

[11] Smith, A. P., McKercher, A. E. and Mawhinney, R. C., *J. Phys. Chem. A* **2011**, *115*, 12544–12554.

[12] Boyd, R. J. and Edgecombe, K. E., *J. Am. Chem. Soc.* **1988**, *110*, 4182–4186.

[13] Boyd, R. J. and Boyd, S. L., *J. Am. Chem. Soc.* **1992**, *114*, 1652–1655.

[14] Popelier, P. L. A., *J. Phys. Chem. A* **1999**, *103*, 2883–2890.

[15] Cortés-Guzmán, F. and Bader R., *J. Phys. Org. Chem.* **2004**, *17*, 95-99

[16] Bader, R. and Matta, C., *Found. Chem.* **2013**, *15*, 253-276

[17] Bader, R., Atoms in Molecules: A Quantum Theory, Clarendon Press, Oxford, U.K. **1994**.

[18] Bader, R. and Bayles, D., *J. Phys. Chem. A* **2000**, *104*, 5579-5589

[19] O'Brien, S. E. and Popelier, P. L., *Can. J. Chem.* **1999**, *77*, 28–36.

[20] O'Brien, S. E. and Popelier, P. L. A., *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 764–775.

[21] O'Brien, S. E. and Popelier, P. L. A., *J. Chem. Soc. Perkin Trans. 2* **2002**, *2*, 478–483.

[22] Popelier, P. L. A., Chaudry, U. A. and Smith, P. J., *J. Chem. Soc. Perkin Trans. 2* **2002**, *7*, 1231–1237.

[23] Bader, R., and Becker, P., *Chem. Phys. Lett.* **1988**, *148*, 452-458

[24] Bader, R., Keith, T., Gough, K., and Laidig, K., *Mol. Phys.* **1992**, *75*, 1167-1189

[25] Cortés-Guzmán, F., and Bader, R., *Chem. Phys. Lett.* **2003**, *379*, 183-192

[26] Matta, C. F., *J. Comput. Chem.* **2010**, *31*, 1297–1311.

[27] Devereux, M., Popelier, P. L. and McLay, I. M., *J. Comput. Chem.* **2009**, *30*, 1300–1318.

[28] Tognetti, V. and Joubert, L., *J. Phys. Chem. A* **2011**, *115*, 5505–5515.

[29] Jabłoński, M. and Palusiak, M., *J. Phys. Chem. A* **2010**, *114*, 2240–2244.

[30] Goerigk, L., in Non-Covalent Interactions in Quantum Chemistry and Physics; de la Roza, A., and DiLabio, G, Eds. Elsevier: Amsterdam, **2017**, Chapter 6, 195-219

[31] Hickey, A. L. and Rowley, C. N., *J. Phys. Chem. A* **2014**, *118*, 3678–3687.

[32] Perdew, J. P. and Schmidt, K., *AIP Conference Proceedings* **2001**, *577*, 1–20.

[33] Kohn, W. and Sham, L. J., *Phys. Rev.* **1965**, *140*, A1133–A1138.

[34] Vosko, S. H., Wilk, L. and Nusair, M., *Can. J. Phys.* **1980**, *58*, 1200–1211.

[35] Becke, A. D., *Phys. Rev. A* **1988**, *38*, 3098–3100.

[36] Lee, C., Yang, W. and Parr, R. G., *Phys. Rev. B* **1988**, *37*, 785–789.

[37] Perdew, J. P., Burke, K. and Ernzerhof, M., *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[38] Tao, J., Perdew, J. P., Staroverov, V. N. and Scuseria, G. E., *Phys. Rev. Lett.* **2003**, *91*, 146401.

[39] Becke, A. D., *J. Chem. Phys* **1993**, *98*, 5648–5652.

[40] Adamo, C. and Barone, V., *J. Chem. Phys.* **1999**, *110*, 6158–6170.

[41] Medvedev, M. G., Bushmarinov, I. S., Sun, J., Perdew, J. P. and Lyssenko, K. A., *Science (80-.).* **2017**, *355*, 49–52.

[42] Hariharan, P. C. and Pople, J. A., *Theor. Chim. Acta* **1973**, *28*, 213–222.

[43] Hehre, W. J., Ditchfield, R. and Pople, J. A., *J. Chem. Phys.* **1972**, *56*, 2257–2261.

[44] Dill, J. D. and Pople, J. A., *J. Chem. Phys.* **1975**, *62*, 2921–2923.

[45] Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., DeFrees, D. J. and Pople, J. A., *J. Chem. Phys.* **1982**, *77*, 3654–3665.

[46] Rassolov, V. A., Pople, J. A., Ratner, M. A. and Windus, T. L., *J. Chem. Phys.* **1998**, *109*, 1223–1229.

[47] Rassolov, V. A., *J. Comput. Chem.* **2001**, *22*, 976–984.

[48] Jensen, F., *J. Chem. Phys.* **2001**, *115*, 9113–9125.

[49] Jensen, F., *J. Chem. Phys.* **2002**, *116*, 7372–7379.

[50] Jensen, F., *J. Phys. Chem. A* **2007**, *111*, 11198–11204.

[51] Jensen, F. and Helgaker, T., *J. Chem. Phys.* **2004**, *121*, 3463–70.

[52] Jensen, F., *J. Chem. Phys.* **2012**, *136*, 114107.

[53] Jensen, F., *J. Chem. Phys.* **2013**, *138*, 014107.

[54] Jensen, F., *J. Chem. Phys.* **2002**, *117*, 9234–9240.

[55] Weigend, F. and Ahlrichs, R., *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–305.

[56] Rappoport, D. and Furche, F., *J. Chem. Phys.* **2010**, *133*, 143105.

[57] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A. J., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Keith, T., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J. and Fox, J. D., Gaussian 09 **2010**.

[58] Golub, G. and Van Loan, C., Matrix Computations, John Hopkins University Press, Baltimore, Md., 3rd edition **1996**.

[59] Keith, T. A., AIMAll **2017**.
URL `aim.tkgristmill.com` Accessed Sept. 1, 2019

[60] Laidig, K., *Chem. Phys. Lett.* **1991**, *185*, 483–489.

[61] R Core Team, R: A Language and Environment for Statistical Computing **2016**.
URL `https://www.r-project.org/` Accessed Sept. 30, 2019

[62] Team, R., RStudio: Integrated Development Environment for R **2016**.
URL `http://www.rstudio.com/` Accessed Sept. 30, 2019

[63] Wickham, H., ggplot2: Elegant Graphics for Data Analysis **2009**. URL `http://ggplot2.org`

[64] Volkov, A., Koritsanszky, T., Chodkiewicz, M. and King, H. F., *J. Comput. Chem.* **2009**, *30*, 1379–1391.

[65] De Proft, F., Tielens, F. and Geerlings, P., *J. Mol. Struct. THEOCHEM* **2000**, *506*, 1–8.

Chapter 4

Transferability of Group Descriptors from the Quantum Theory of Atoms in Molecules

4.1 Abstract

Machine learning in chemistry is a topic of current interest, developing unforeseen relationships between specific molecular properties and large ensembles of feature set descriptors. Computational atomic and functional group descriptors are such a feature set. Ideally, these descriptors would be invariant to the local environment, and hence be considered transferable. Whether this is true is an open question. Here, we evaluated the transferability of Quantum Theory of Atoms in Molecule functional group descriptors for one-hundred and seventeen functional groups in a series of sixteen substrates to determine whether descriptors obtained using hydrogen as substrate are transferable. The functional group volume has a strong, consistent, linear relationship throughout. All other hydrogen-based group descriptors exhibit a relatively strong linear relationship with those in carbon-based substrates

and a reasonable linear relationship with those in non-carbon-based substrates. Outliers are readily interpretable in terms of substrate induced functional group geometry changes. As expected, directional descriptors lying along the substituent-substrate axis are the least conserved.

4.2 Introduction

Machine learning in chemistry, a topic of current interest , attempts to use molecular descriptors to predict outcomes of reactions.[1–4] Descriptors related to substituents on a system are physically meaningful and useful to include in such machine learning models.[5]

Substituent descriptors are traditionally developed by measuring the effect a substituent has elsewhere in a molecule and not based on the substituent's intrinsic properties. Substituent parameters from such indirect measurements are best described as proxies. Hammett developed the first set of proxies, the σ_m and σ_p substituent parameters.[6] Other groups, based on both experimental and theoretical approaches, have defined additional substituent constants, although still using a proxy scheme.[7,8]

Substituent descriptors are often developed within a given molecular framework. For example, Hammett used benzene as a backbone and the pK_a of a carboxylic acid as proxy. However, regardless of their origin, these have been used as parameters in a variety of other systems.[9,10] This is due to the concept of transferability - that a functional group in one molecule has similar properties to those in another. Numerous attempts have been made to exactly quantify transferability, and provide transferable structural units. The applications of this idea includes experimental X-ray diffraction parameter databases,[11,12]

computational pseudoatoms databases,[13–17] force fields,,[18–22] linear regression analysis of molecular properties,[23] and transferable atom equivalents.[24,25] Therefore, evaluating the properties of atoms and functional groups, and examining their transferability is a topic of continuing interest.

There are two types of transferability to consider: perfect and compensatory.[26] The former is when the property of an atom in two different molecules is identical, while the latter relates to the properties of two fragments mutually changing in opposite manners when brought in contact. Sadly, due to the Hohenberg-Kohn theorems and their extensions to subsystems, it is proven that perfect transferability is an unattainable limit, but it can be approached.[27,28] The observed linear relationships between the molecule's energy and the number of methylene groups (in a homologous series) is an example of compensatory transferability, with the destabilization of methylenes near the chain ends compensated for by the stabilization of methyl and other methylene groups.[29]

The Quantum Theory of Atoms in Molecules (QTAIM) is an ideal tool to assess transferability, since it defines atoms and functional groups based on the topology of the electron density. Atoms are based on basins in the density of the molecule, separated by zero flux surfaces, on which $\nabla\rho(\mathbf{r}) \cdot (n)(\mathbf{r}) = 0$ for all points on the surface. These surfaces partion the density completely into unique basins, each containing a single attractor, usually a nucleus, or in some cases a non-nuclear attractor. The union of the attractor and the density in the surrounding basin defines an atom.[30,31] By integrating property operators over the basin, unique atomic properties are obtainable. Since atomic properties sum to the molecular value, group properties are readily obtainaed by summing over the atoms comprising the group of interest.[26]

Much work uses QTAIM to determine functional group properties and examine their transferability. When using QTAIM atomic energies to examine transferability, one should omit the scaling factor for the energy, or use self-consistant virial scaling.[32] The transferability of atomic and group properties in alkanes,[29] alkanols,[33,34] fluoroalkanes,[35] cyanoalkanes,,[36,37] ethers,[38,39] aldehydes, and ketones,[40,41] was considered by Mosquera's group. These studies examined transferability by looking for any numerical deviation between two atoms. In this manner, they define pseudo-transferable atom types. Devereux and Popelier alternatively define a "significant" difference as a considerable variation relative to the total observed range of properties.[42]Similarly, Rykounov assessed transferabilty in hydropyrimidines by comparing the change in properties to their uncertainties.[43] The above studies look for any numerical deviation in QTAIM properties, but the potential of practical application of transferability remains unstudied.

Here, we analyse a large data set of 117 substituents and 17 substrates and assess the transferability of QTAIM-derived functional group properties by addressing which properties are transferable, and investigating the limits of the transferability. Through linear regressions and the analysis of outliers, we interpret the linearity, or lack thereof, of the studied properties. This study enables insight into the ability of these properties to be used in linear regression techniques between substrates and into their eventual use in machine learning as feature sets in regression algorithms.

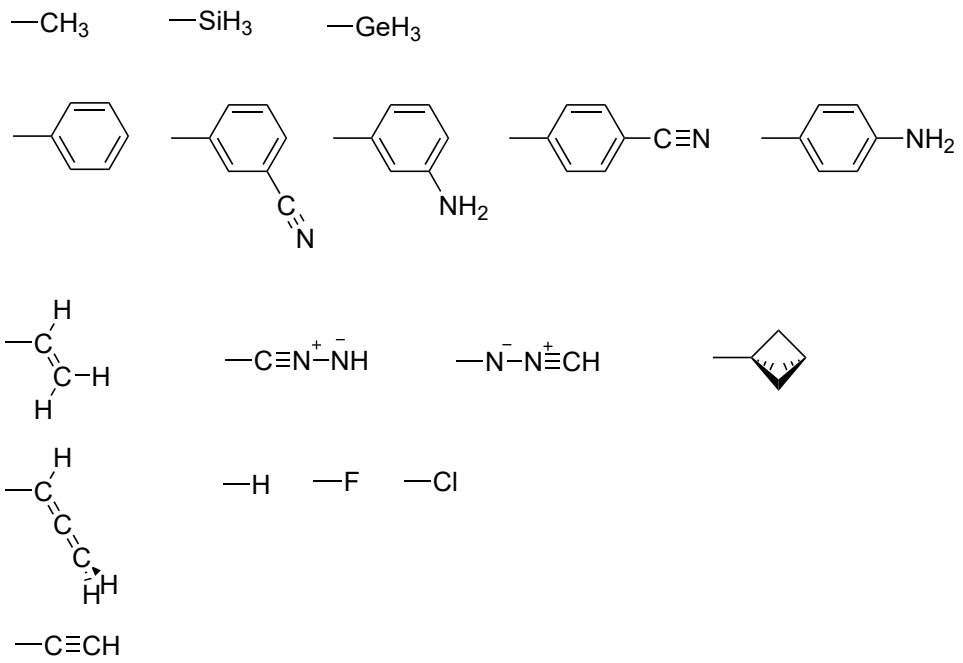

Figure 4.1: Substrates chosen for the study of transferability of QTAIM group properties

4.3 Methodology

A series of 117 substituents, R, were combined with 17 substrates, G, with all R-G structures optimized in Gaussian09.[44] Frequencies were evaluated at the optimized structures to determine if located stationary points are minima. The B3LYP/def2-TZVPPD model chemistry was used on the basis of previous work.[45–48] Substrates, shown in Figure 4.1, were chosen to represent a variety of chemical situations: carbon-based substrates of differing hybridisation (CH_3, $CHCH_2$, $CHCCH_2$, CCH, C_6H_5), aromatic substrates with varying substitution ($m{-}C_6H_4{-}NH_2$, $m{-}C_6H_4CN$, $p{-}C_6H_4{-}NH_2$, $p{-}C_6H_4CN$), substrates of varying electronegativity (F, Cl, GeH_3, SiH_3), flexible substrates (CNNH, NNCH), and rigid substrates ([1.1.1]-bicyclopentane, abbreviated as BP here on). Electronegativity of the substrates in GH molecules was calculated using Boyd and Edgecombe's definition of electronegativity, Equation 4.1 and 4.2.[49] Here, r_H is the distance from the hydrogen atom to the G-H bond critical point, r_{AH} is the bond length between the

hydrogen and the bonded atom in G, N_A is the valence of the atom in G bonded to H, and ρ_c is the density at the G-H bond critical point. The parameters a and b are determined here by calculating F_A for fluorine and lithium in FH and LiH molecules, then setting χ_F to 4, and χ_{Li} to 1.

$$F_A = \frac{r_H}{\rho_c r_{AH} N_A} \tag{4.1}$$

$$\chi_A = a F_A^b \tag{4.2}$$

The 117 substituents, shown in Figures B1 and B2 in Appendix B, comprise a variety of classes (amino, alkyl, carbonyl, etc.). In this study, we use an axis system derived from previous work.[50] Specifically, the negative x-axis lies along R-G bond (in the direction of G). The atom of R bonded to G is positioned at the origin and we further choose a specific atom for each substituent that is guaranteed to be on the positive xy plane for each substituent. This chosen atom is kept constant between all substrates, ensuring that the x, y, and z axes are defined consistently for any given substituent, regardless of substrate. This is illustrated in Figure 4.2

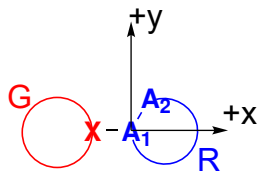

Figure 4.2: Illustration of coordinate system used in study. Atom A_1 of substituent R is positioned at the origin. Atom X of substrate G bonded to R is positioned along the negative x-axis. Arbitrarily chosen atom A_2 of substituent R is positioned on $+y$ axis. Atom A_2 is the same for a given substituent in all substrates G.

Extended wavefunction files were generated, and QTAIM properties assessed using AIMAll 17.11.14.[51] Properties of interest include the group charge, q(R); total dipole, $\boldsymbol{\mu}$(R); charge transfer dipole moment,

$\boldsymbol{\mu}^c$(R); polarization dipole moment, $\boldsymbol{\mu}^p$(R); and contributions to molecular quadrupole, **Q**(R). $\boldsymbol{\mu}$(R), Equation 4.3, is the sum of $\boldsymbol{\mu}^c$(R) and $\boldsymbol{\mu}^p$(R), defined by Equations 4.4 and 4.5, respectively. **Q**(R) was determined by summing the atomic contributions according to Laidig using Equation 4.6.[52] Here, $\mathbf{R}_\Omega$ is the position of the nuclear attractor in Ω, $\mathbf{R}_b(\Omega|\Lambda)$ is the position of the bond critical point between Ω and attached group Λ, $Q(\Omega|\Lambda)$ is the charge of group Λ, q(Ω) are the atomic charges of basin Ω, χ_Ω^α and χ_Ω^β ($\alpha, \beta = x, y$, or z) are the α or β geometric coordinate of the nucleus of basin Ω, $Q_{\alpha\beta}(\Omega)$ is the atomic quadrupole moment of basin Ω, $\delta_{\alpha\beta}$ is the Kronecker delta, $\langle r^2 \rangle_\Omega$ is the atomic second radial moment of basin Ω, $\mu^{p,\alpha}(\Omega)$ and $\mu^{p,\beta}(\Omega)$ are the α or β component of $\boldsymbol{\mu}^p(\Omega)$, and $Q_{\gamma\gamma}$ is the trace of the contribution to the quadrupole matrix. Generally, functional group properties will be refered to as $P(R_{substrate})$, where substrate is one or more specific substrates. $P(R_G)$ however will refer to the property in all substrates except our chosen reference (G=H).

$$\boldsymbol{\mu}(\mathrm{R}) = \boldsymbol{\mu}^{\mathrm{c}}(\mathrm{R}) + \boldsymbol{\mu}^{\mathrm{p}}(\mathrm{R}) \tag{4.3}$$

$$\boldsymbol{\mu}^{\mathrm{c}}(\mathrm{R}) = \sum_{\Omega \in \mathrm{R}} \sum_{\Lambda=1}^{N_b(\Omega)} [\mathbf{R}_\Omega - \mathbf{R}_b(\Omega|\Lambda)]\, Q(\Omega|\Lambda) \tag{4.4}$$

$$\boldsymbol{\mu}^{\mathrm{p}}(\mathrm{R}) = \sum_{\Omega \in \mathrm{R}} -\int_\Omega \mathbf{r}\rho(\mathbf{r})d\mathbf{r} \tag{4.5}$$

$$\begin{aligned}\mathbf{Q}(R) = \frac{1}{2}\sum_{\Omega \in R}[3(q(\Omega)\chi_\Omega^\alpha\chi_\Omega^\beta + \frac{Q_{\alpha\beta}(\Omega) + \delta_{\alpha\beta}\{\langle r^2\rangle_\Omega\}}{3} + \chi_\Omega^\beta\mu^{p,\alpha}(\Omega) \\ + \chi_\Omega^\alpha\mu^{p,\beta}(\Omega)) - \delta_{\alpha\beta}(Q_{\gamma\gamma})]\end{aligned} \tag{4.6}$$

Data analysis was performed using R and RStudio,[53,54] and graphs generated using the ggplot2 package.[55] Assesment of linearity is criti-

cal for this study, as many machine learning models utilise linear modeling. We will judge linearity using a combination of r^2, deviations from the ideal slope of 1, and deviation from the ideal intercept of 0. Transferability for practical application is assigned for r^2 greater than 0.9, slope within 0.1 units of 1.0, and intercept less than 5% of the range. The deviation from these ranges is explained using an analysis of outlying points in the linear regressions. Additionally, for each substituent, we calculate the average, standard deviation, minimum and maximum of a property over all other substrates, and compare to G=H. There are two cases in which these descriptors can be used in future modeling with machine learning. Firstly, all metrics lie within the ranges delineated above, which would indicate a good quantitative transferability sufficient for application. Secondly, if errors are observed, but interpretable and predictable based on some physical observable, the descriptors may still be applied.

4.4 Results and Discussion

In our dataset certain substituent-substrate combinations optimize to structures that differ from the general optimized substituent geometry in R-G molecules. Often, these are conformation dependent. For example, certain substituent-substrate combinations will only optimize to one conformer (ip or op). CH_2NH_2−ip with G=F and Cl optimize to CH_2NH_2−op. For other substrates, $NHNH_2$−op and CH_2OH−ip conformers are not found and are therefore omitted from the dataset. For F alone, CH_2SH−ip is not found, likely due to the stronger effect of F than Cl. CH_2COOH−ip is not found for G=GeH_3. In other cases, a different conformer altogether is located. For R=CH_2COX, strongly electronegative substrates (F and NNCH) see a whole new minimum struc-

ture not seen in other substrates. In these cases, $F-CH_2COX-op$ and $HCNN-CH_2COX-op$ are omitted from the data set. Cases where the substituent-substrate bond is unreasonably elongated, or where the substituent-substrate bond induces a drastic change in the substituent structure are also excluded. They are as follows: Cl-CO_2^-, HCNN-CO, HNNC-OCl, HNNC-OF, HCNN-OF, HCNN-$OO^{\cdot}$. All these cases are rare, corresponding to a total of 22 molecules in the full dataset of 1989 molecules. They typically involve fluorine, or other elements which have strong interactions. Four inclusions in the dataset correspond to imaginary frequencies. These molecules are still included in the dataset, as their geometries can still be compared. The molecules are $H_3GeCH_2COF-ip$, $m-C_6H_4NH_2-CH(CH_3)_2$,$m-C_6H_4NH_2-CH_2COF-ip$, $p-C_6H_4CN-CH_2COOH-ip$

$P(R_H)$ are compared pairwise to the other sixteen $P(R_G)$. If, for arbitrary substrates G_1 and G_2, $P(R_{G_1}) \approx P(R_H)$ and $P(R_{G_2}) \approx P(R_H)$, then by induction $P(R_{G_1}) \approx P(R_{G_2})$. In this way we thoroughly assess transferability in 16 cases between each other, using a single point of comparison, G=H. Transferability will be addressed one property at a time to help identify those whose $P(R_H)$ are suitable for use in other substrates. Of vital importance to the analysis is the range over which the values occur. Table 4.1 clearly depicts that the values span a wide range in the properties of interest. Appendix B shows histograms of the distribution of property values. Other tables containing the summary statistics (mean deviation, mean absolute deviation, max and min) and linear regression parameters (slopes, intercepts, r^2) are given in Appendix B. For electrostatic properties (charge/dipole/quadrupole), universally low deviations for all substrates are not observed, ruling out perfect transferability of electrostatic properties between substrates. This then leads to the investigation into the linearity of properties between dif-

Table 4.1: Maximum and minimum values for studied properties, taking the whole data set into account (i.e. not just for one substrate)

Property	Min	Max	Property	Min	Max
q(R)	-1.63	1.53	$\mu_{p,x}$(R)	-3.36	3.65
μ^x(R)	-2.62	1.02	$\mu_{c,x}$(R)	-4.83	2.82
μ^y(R)	-1.63	1.33	$\mu_{p,y}$(R)	-2.76	2.80
μ^z(R)	-1.33	1.23	$\mu_{c,y}$(R)	-4.21	3.43
Q_{xx}(R)	-3.83	6.07	$\mu_{p,z}$(R)	-2.37	1.57
Q_{yy}(R)	-3.06	6.94	$\mu_{c,z}$(R)	-2.84	3.25
Q_{zz}(R)	-7.79	2.39	Vol(R)	137	788

ferent substrates. Also shown in Appendix B are the electronegativities of the substrates. Based on these values, substrates F, NNCH, and Cl are refered to as electronegativite substrates while SiH_3 and GeH_3 are refered to as the electropositive substrates.

Figure 4.3 show the average slope, intercept, and r^2 for the regressions comparing $P(R_G)$ to $P(R_H)$. The standard deviations are imposed, and limits on optimal values for transfrability are shown in green. The average slopes, plus or minus their standard deviation, for μ_x(R), μ_z(R), q(R), $Q_x x$(R), and Q_{yy}(R) exceed the limits for keeping the slope close to 1. The average $|\frac{\text{intercept}}{\text{range}}| \pm \text{sd}$ is greater than 0.05 for μ_x(R), q(R), and Q_{xx}(R). The r^2 see significant deviations from the ideal for all properties, barring Q_{zz}(R) and Vol(R). The r^2 is the most critical parameter in assessing the linearity for these applications. That is, even if the slope and intercept differed from their ideal values, as long as the r^2 was acceptable, the $P(R_H)$ values could be used in building linear models for other substrates. Of course, the slope and intercept must be the ideal values for transferability in the strictest sense, but linearity observed through r^2 would be sufficient. However, this is not seen and these metrics show strictly that all properties except Q_{zz}(R) and Vol(R) are not quantitatively transferable.

Due to the lack of strict conservation, each property will be investigated thoroughly in more detail. The sources of the non-ideal lin-

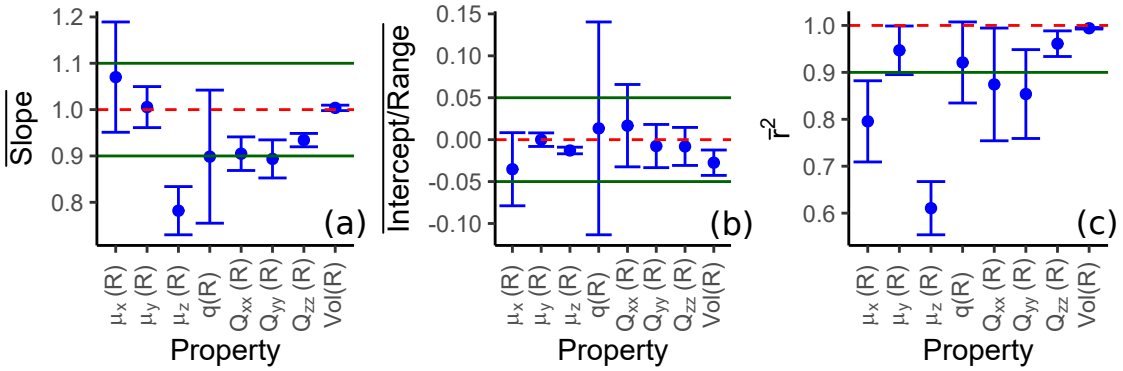

Figure 4.3: Average linear regression parameters, (a) slope, (b) intercept/range, (c) r^2 with their standard deviations for integrated properties under study. The red lines illustrate the ideal values for the slope, intercept, and r^2. To illustrate transferability, the average linear regression parameters and their standard deviation would be found between the green lines in all cases.

ear regression parameters will be investigated on a case-by-case basis. If the errors are interpretable and predictable, these properties can still be used in machine learning models. One mode of analysis in the forthcoming sections are analysing outlying points in the regressions. All outliers detected in the linear relationships are summarized in Tables B30 and B31 in Appendix B.

4.4.1 Group Charge

One of the most straightforward group properties to calculate is the group charge, obtained by subtracting the integrated atomic population from the nuclear charge. Summary statistics show that the group charge intuitively changes with the substrate, consistent with the changing electronegativities from hydrogen. We address the observed non-transferability of the group charge in detail, determining how this limits $q(R_H)$ in other situations. Figure 4.4 shows the linear plot comparing $q(R_G)$ to $q(R_H)$. For ease of analysis, 3-5 substrates are plotted an each graph.

For carbon-based substrates, Figure 4.4A, C and D, q(R) relates lin-

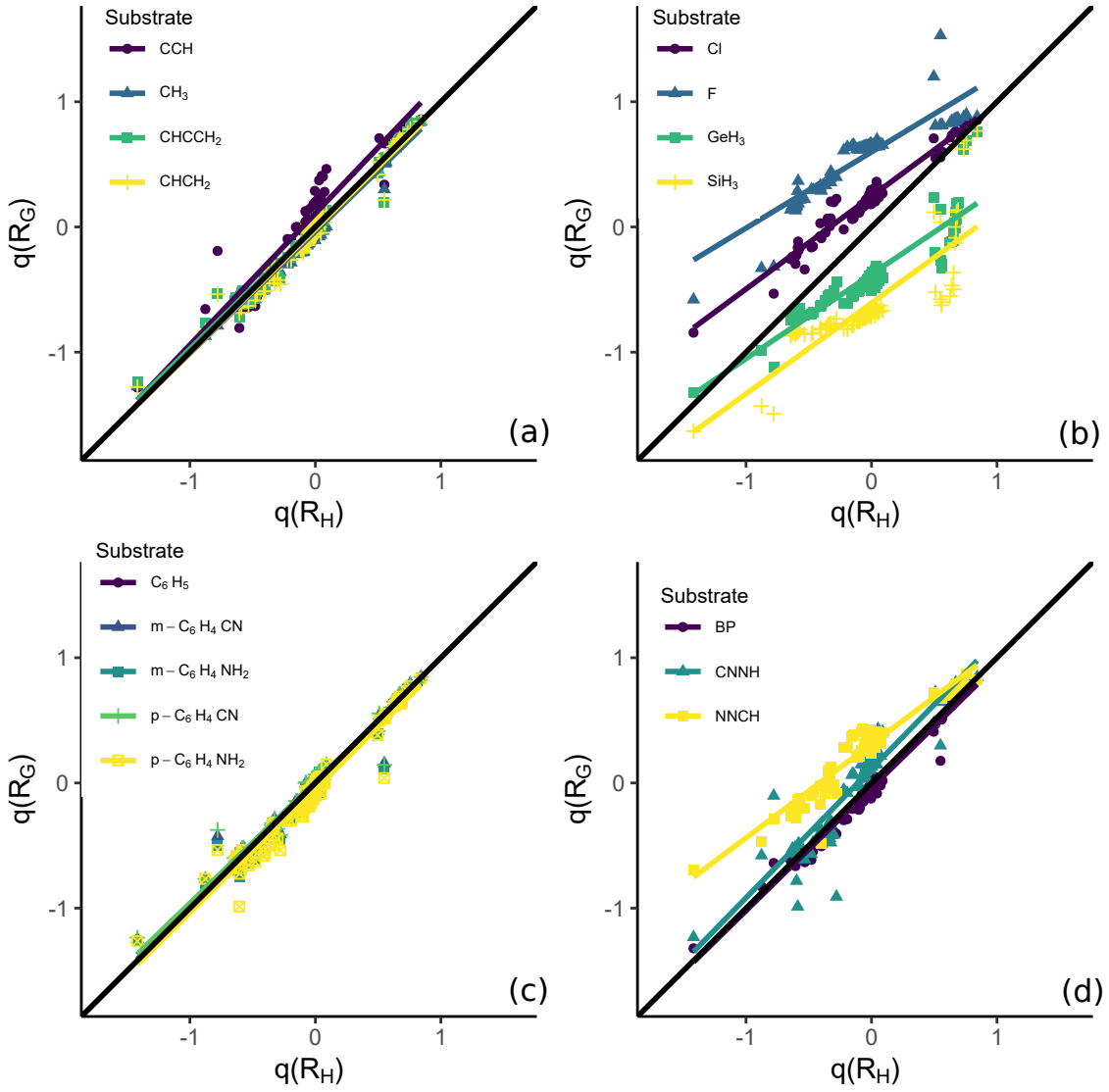

Figure 4.4: Linear plots of q(R) for R in various R-G systems against $q(R_H)$. (a) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (b) G: Cl, F, SiH_3. GeH_3; (c) C_6H_5, m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; (d): BP, CNNH, NNCH

early to $q(R_H)$, with little deviation from the ideal $y = x$ line, except for G=CCH and CNNH. G=CCH and CNNH are related as one of the resonance structures of nitrilimines involves an sp hybridised carbon.[56] $q(R_{CCH})$ differs most from the $y = x$ line when $q(R_H)$ is near zero or above, indicating an interaction between positively charged substituents and CCH. All other carbon based substrates relate well ($r^2 > 0.95$), with only a few outliers. These outliers are charged substituents, S^- and CO^+. Therefore group charges in hydrogen can lin-

early relate to group charges in other carbon-based substrates, though alkyne substrates should be employed more cautiously.

For substrates with varying electronegativity we see less conservation of linearity, with r^2 less than 0.83. The most striking change is the intercepts of the linear lines - they are greater than zero for more electronegative substrates (F, Cl, NNCH) and less than zero for more electropositive substrates (GeH_3 and SiH_3). The general change in intercept is attributed to the electronegativity of the substrates. For electropositive substrates, increased deviations occur for highly charged substituents. The remainder of the data set is quite linear, though the slopes and intercepts differ from the ideal $y = x$, with slopes less than 0.75, indicating that the change in substituent has less effect on charge in non-hydrogen substrates.

Additional information can be intuited when looking at the data set as a whole, taking into account all substrates. For each substituent, the average $q(R_G)$ can be compared with $q(R_H)$. This average, of course has a standard deviation. Figure 4.5 shows $q(R_H)$ and the average value over all other substrates, with standard deviations superimposed as error bars, and maximum and minimum values shown in green. The average follows $q(R_H)$ closely. The standard deviation is near-constant over all substituents, only lowering for highly charged substituents. The minimum line is the most electropositive substrate, SiH_3, and the most electronegative substrate, F, is the maximum line. Spikes in maximum and minimum values are typically due to omissions in the data set, as previously discussed for the fluorine substrate.

While the average is subject to the choice of data set, our choice of substrates was not chosen with the average replicating $q(R_H)$ in mind, it was merely chosen to include substrates of varying properties. Carbon-based substrates have the most weight in the average, as they are

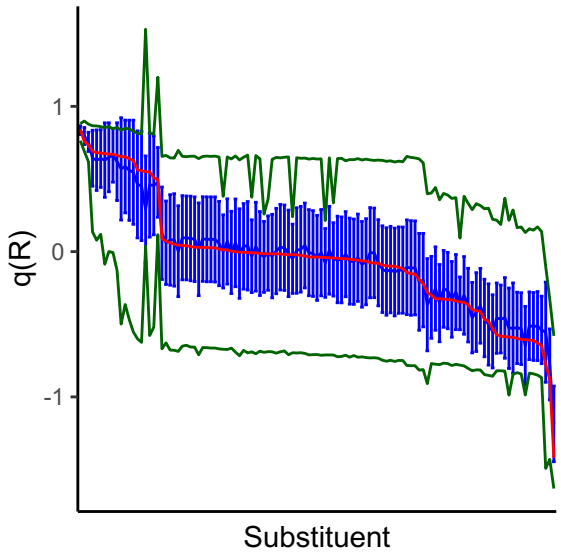

Figure 4.5: Plot displaying q(R) values for $q(R_H)$, the average q(R) over all other substrates. Error bars are the standard deviation of q(R) over all other substrates. Green lines show the range of q(R) values over all other substrates

included most in the dataset. This is justified by the fact that carbon-based systems are the most common systems to which these descriptors would likely be applied.

Figure 4.5 also clearly shows the distribution of charges. Jumps occur in the dataset at intuitive places. q(R) is high for positively charged or electropositive substituents. It then sharply dips close to zero for carbon and sulphur based substituents, with further dips for nitogen-based and oxygen based substituents. More negatively charged substituents are included than positively charged substituents, but negatively charged substituents are also more common in practice - amino, hydroxyl, and so on.

4.4.2 Group Dipole

Another property that could be of great use is the dipole moment, $\boldsymbol{\mu}(R)$, describing the substituent's polarity. In our original paper,[57] the dipole's observed error was 0.008 au for the total dipole moment and

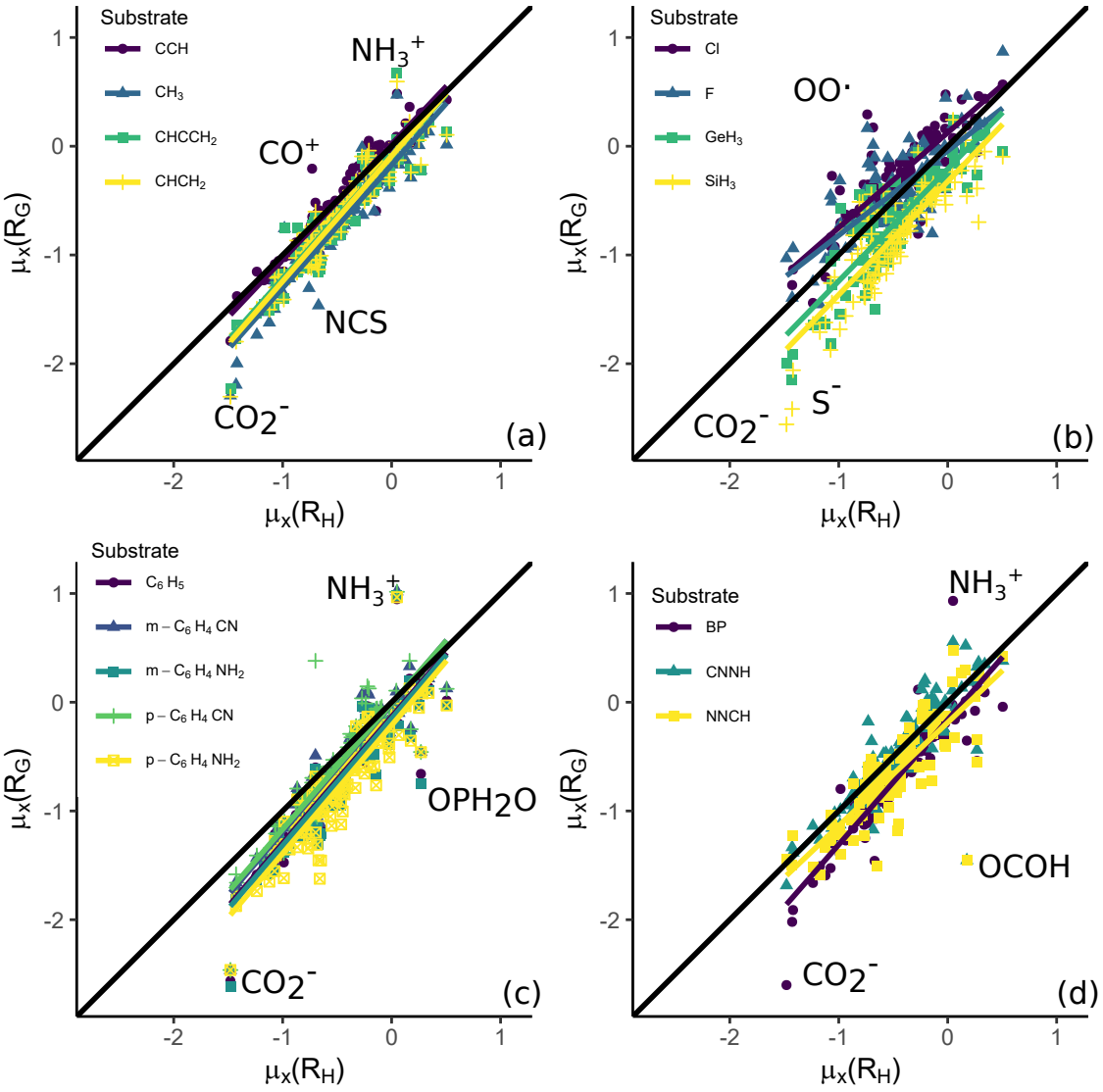

Figure 4.6: Linear plots of μ_x(R) for R in various R-G systems against μ_x(R). (a) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (b) G: Cl, F, SiH_3. GeH_3; (c) C_6H_5, m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; D: BP, CNNH, NNCH

up to 0.03 for the charge-transfer and polarization terms. Here, we examine the linear regression parameter's non-transferability in detail. Figures 4.6- 4.8 show linear plots relating $\mu_i(R_G)$ values to $\mu_i(R_H)$ ($i = x, y,$ or$z$). Of the three directions, μ_x(R) has the most scatter around the regression. Increased scatter in these lines is attributed to the definition of the axis systems. From the choice of coordinate system, the substrate-substituent bond is along the x axis, where most variation is seen. This is not a rebuttal of the axis system, but rather, noting a

logical consequence of the choice of axis system. In most cases good linearity is still qualitatively observed. The main deviation from linearity is due to the presence of outliers. The common outliers are as follows: charged substituents, NCS, CH_2OOCH_3-op, OPH_2O, and OCHO. Non-charged outliers are due to an observed change in geometry - a change in the G-N-C angle for NCS and changed dihedrals for others. Differing electronegativity leads to more observed scatter as observed for Cl, F, NNCH, GeH_3 and SiH_3 substrates. Carbon-based substrates generally see little scatter, only the few outliers. As these outliers are interpretable based on either being charge substituents, or their geometry changing, μ_x(R) is still a promising descriptor for inclusion in machine learning models.

Figure 4.7 illustrates that μ_y(R) has much less observed scatter than μ_x(R) (r^2 typically > 0.9), but a large cluster at (0,0) should be noted; many substituents have near-0 μ_y. Most substrates relate well linearly, but outliers are observed in CCH, CNNH, and F. For CCH, the largest outlier is ONO due a change in bonding character of the N-O bonds in the xy plane; one bond is elongated to 1.6 Å, compared to the other N-O bond of 1.12 Å. Another outlier for many substrates is OPH_2O, whose G-O-P-O dihedral angle is different from that in substrate hydrogen for most substrates. This dihedral change induces a change in μ_y(R).

The largest outlier for the fluorine substrate is ONH_2, which also undergoes a geometry change, due to inclusion of sp carbon resonance structures. In $HONH_2$, the amino hydrogens bend away from the substrate hydrogen, and in $FONH_2$ they bend toward it - this change in geometry leads to the deviation from $\mu_y(R_H)$. The outliers for CNNH are ONO, which sees a similar geomtry change to substrate CCH, and OOH, which optimizes to a planar geometry, therefore also changing the observed μ_y. Non-carbon attached substrates see a slight increase

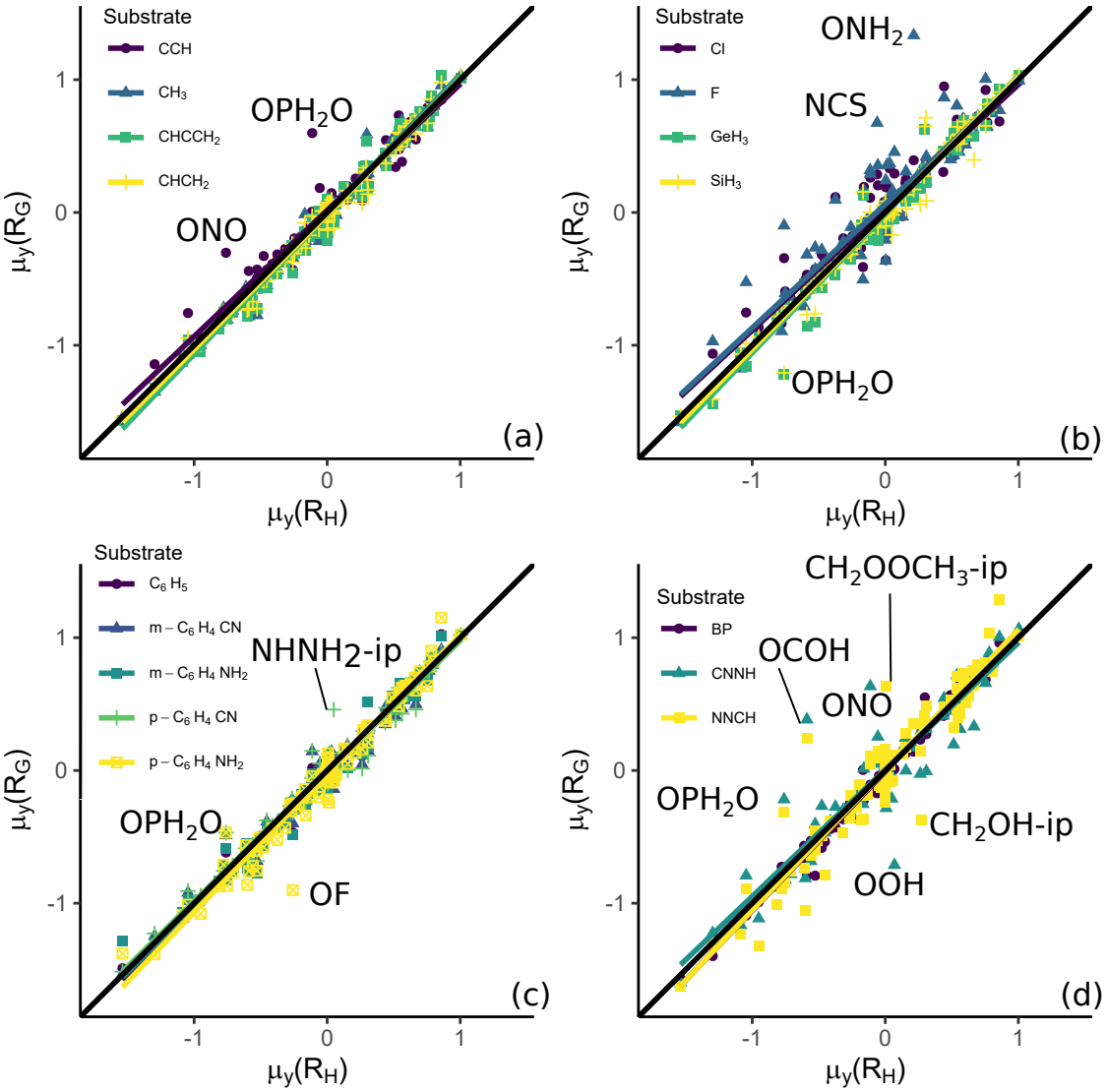

Figure 4.7: Linear plots of μ_y(R) for R in various R-G systems against μ_y(R). (a) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (b) G: Cl, F, SiH_3. GeH_3; (c) C_6H_5, m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; D: BP, CNNH, NNCH

in deviation from the $y = x$ line, but generally, this property is well transfered across substrates.

Evidently from Figure 4.8, μ_z does not behave like the other directions, though OPH_2O is an outlier for many substrates, as in μ_y(R). There is some variation at $\mu_z(R_H)$=0 and many overlapping points near (0,0). This is clearest in Figure 4.8(c). One of the few substituents for which the deviation at $\mu_z(R_H)$=0 occurs is $CONH_2$. We observe that the nitrogen in $CONH_2$ is less planar when attached to a conjugated system. This

is understood by examining the available resonance structures, a few of which are shown in Figure 4.9. Only conjugated systems are capable of contributions from resonance structure C, highlighting why this is an outlier in systems with conjugated substrates. Aside from the outliers at 0, linearity is generally good for all carbon-attached substrates. A few outliers exist, typically corresponding to amino substituents, since they optimize to a more planar geometry for certain substrates. This indicates an error occurs when geometry is significanty changed due to resonance, or other, interactions. A common amino outlier is $N(CH_3)_2$ - the electron donating methyl groups make the nitrogen more able to donate electron density through resonance, increasing the planar structure and deviation from the property when attached to hydrogen substrate. Substrates of varying electronegativity have additional deviation from $y = x$, but this is consistent with observed variant geometries discussed in other directions.

The summary statistics for each substituent are shown in Figures 4.10- 4.12. As expected from the linear plots, μ_x has the highest standard deviations and largest range between minimum and maximum values. It's also evident that on average, $\mu_x(R_G)$ is lower than $\mu_x(R_H)$. $\mu_y(R_G)$ and $\mu_z(R_G)$ have less difference between their values in hydrogen and the average, and low standard deviations. The main error is seen in spikes in the maximum and minimum errors here.

The distributions of the dipoles in Figures 4.10- 4.12 have clear patterns. For $\mu_x(R)$, the values are positive for substituents bonded through an oxygen which also contain carbon. Substituents with $\mu_x(R)$ around -0.1 are typically carbon based substituent (e.g. $CHCCH_2$) As μ_x decreases more, more electronegative atoms are incorporated into the substituents. $\mu_x(R)$ jumps to the lowest values for negatively charged substituents. $\mu_y(R)$ has a steady decrease, barring the plateau of many

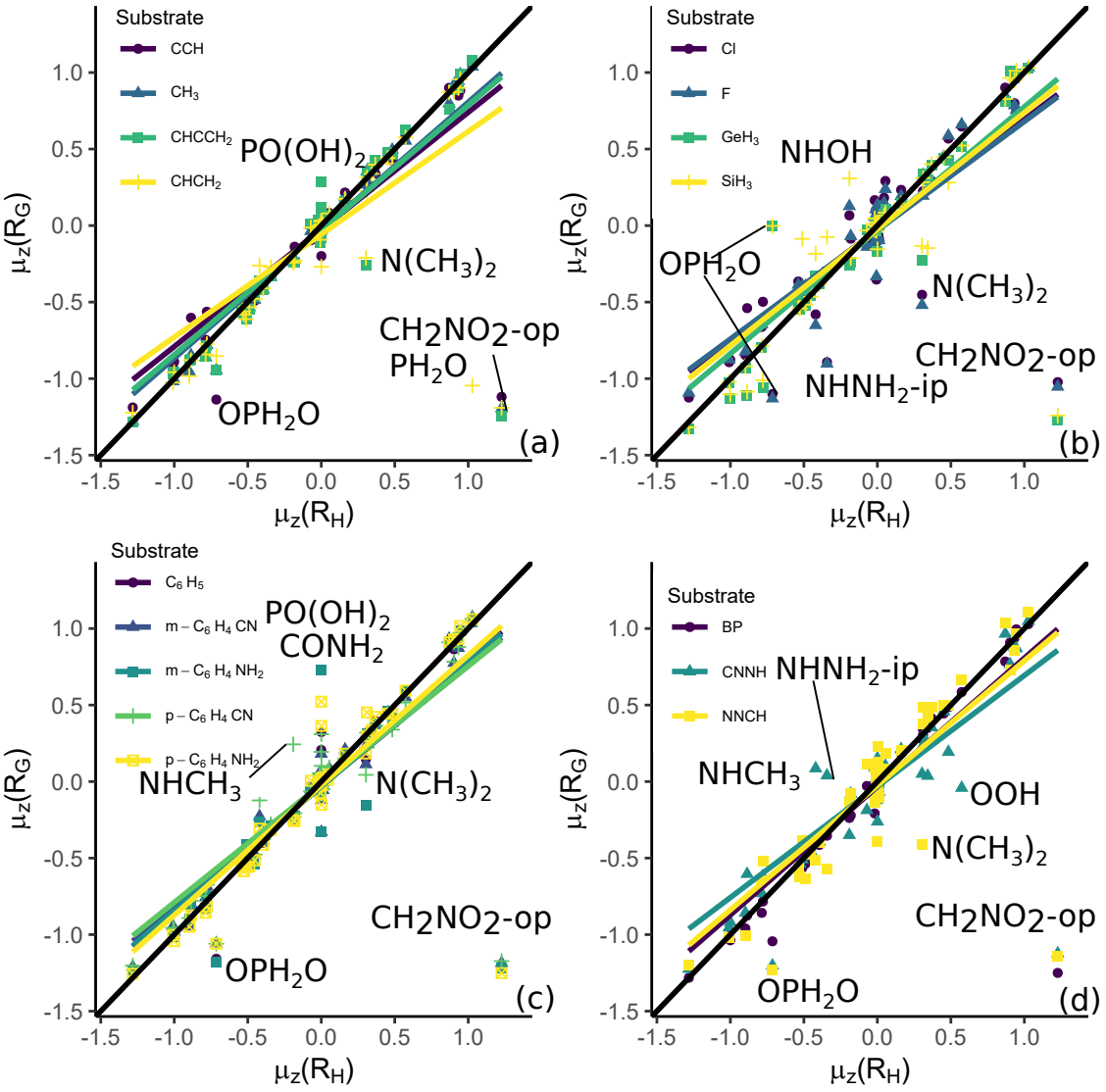

Figure 4.8: Linear plots of μ_z(R) for R in various R-G systems against $\mu_z(R_H)$. (a) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (b) G: Cl, F, SiH_3. GeH_3; (c) C_6H_5, m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; D: BP, CNNH, NNCH

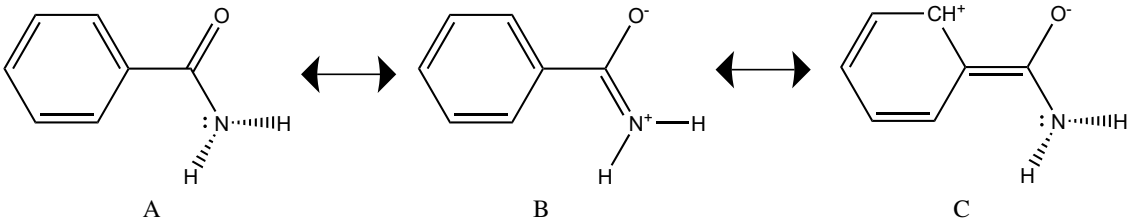

Figure 4.9: Most relevant resonance structures for $C_6H_5-CONH_2$. A and C contribute to a pyramidal nitrogen, while B contributes to a planar nitrogen

substituents with near-zero μ_y(R). μ_z also has a steady change except for the plateau, which is larger than in μ_y(R) due to the many substituents symmetrical around the xy plane.

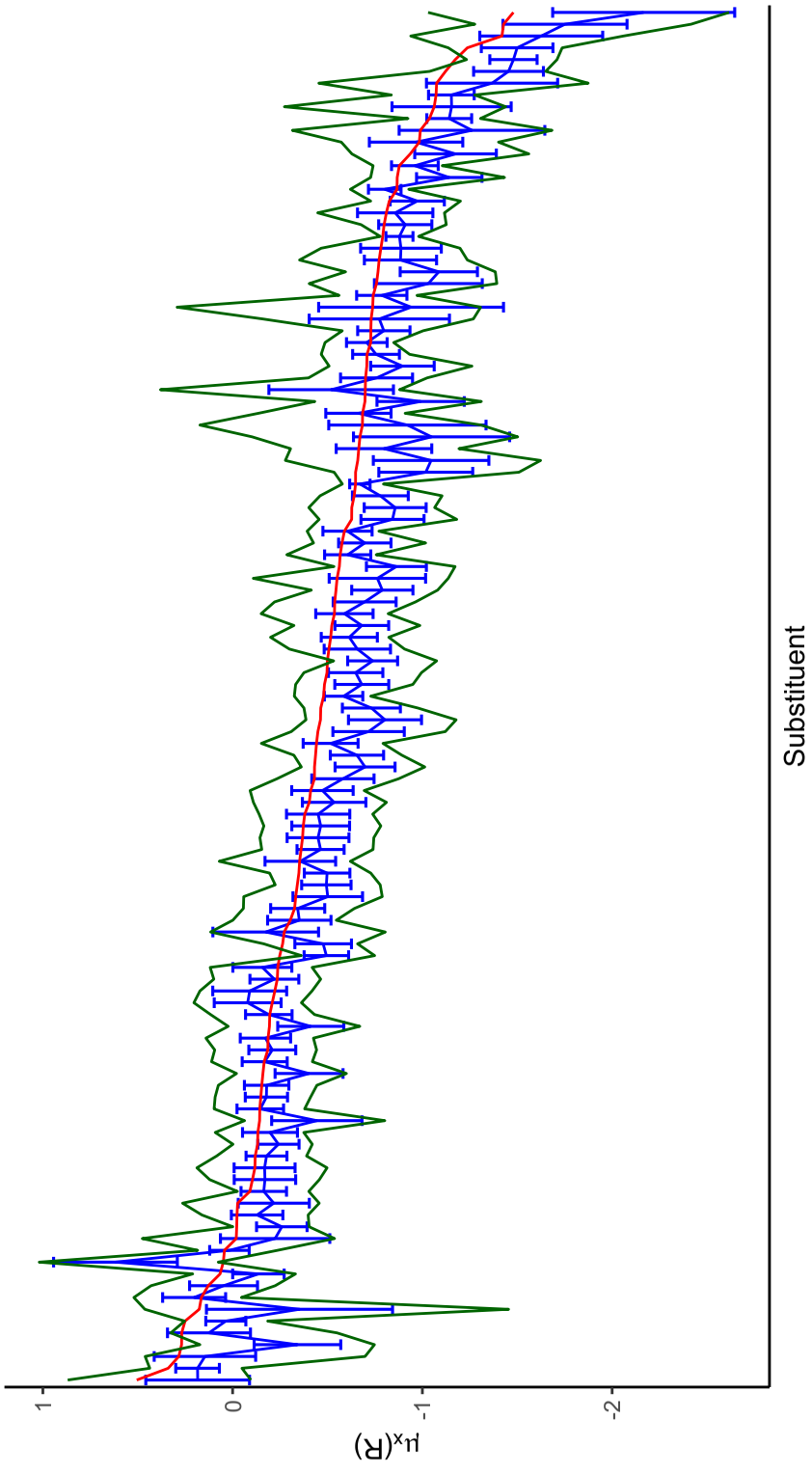

Figure 4.10: Plot displaying $\mu_x(R)$ values for $\mu_x(R_H)$, the average $\mu_x(R)$ over all other substrates. Error bars are the standard deviation of $\mu_x(R)$ over all other substrates. Green lines show the range of $\mu_x(R)$ values over all other substrates

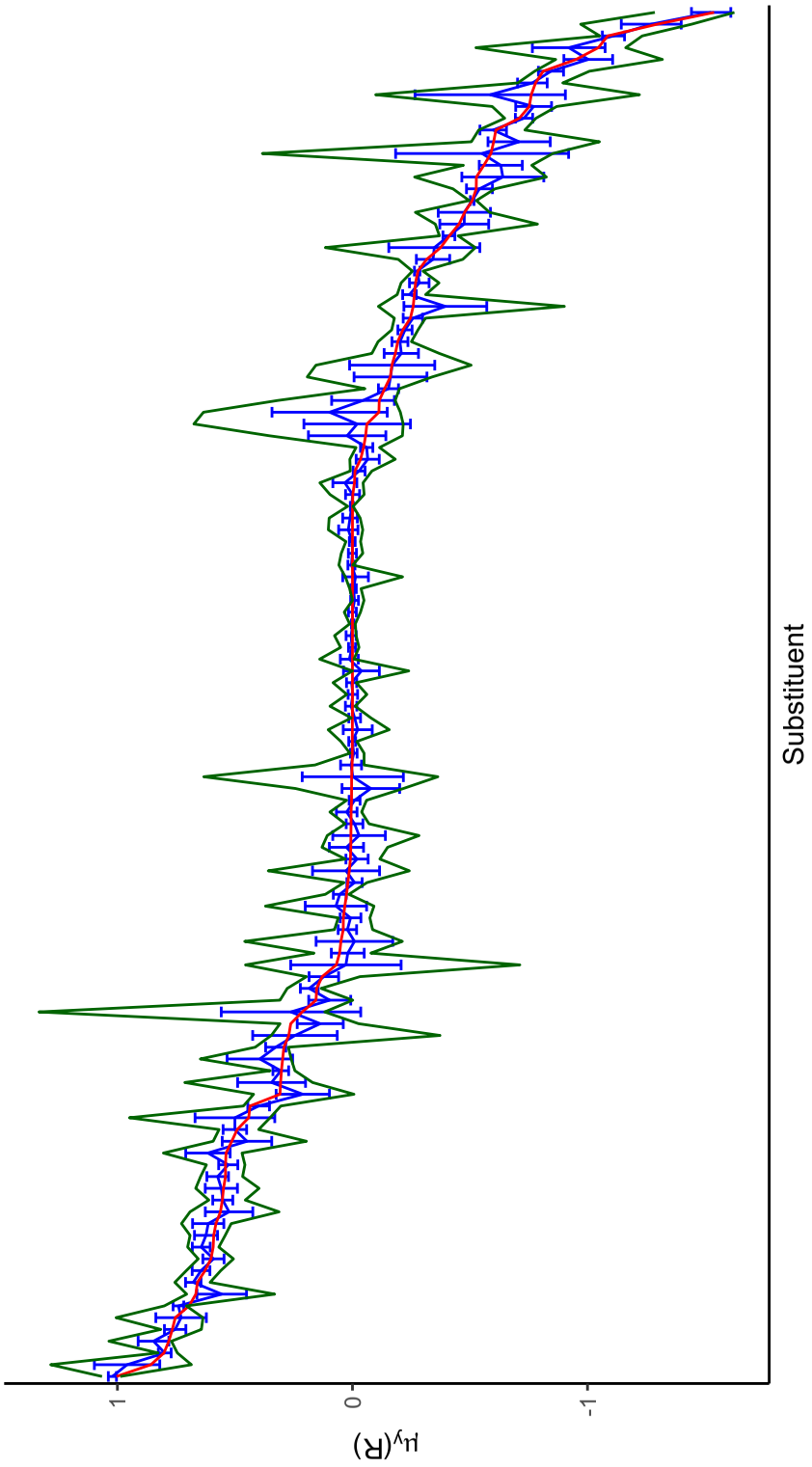

Figure 4.11: Plot displaying μ_y(R) values for $\mu_y(R_H)$, the average μ_y(R) over all other substrates. Error bars are the standard deviation of μ_y(R) over all other substrates. Green lines show the range of μ_y(R) values over all other substrates

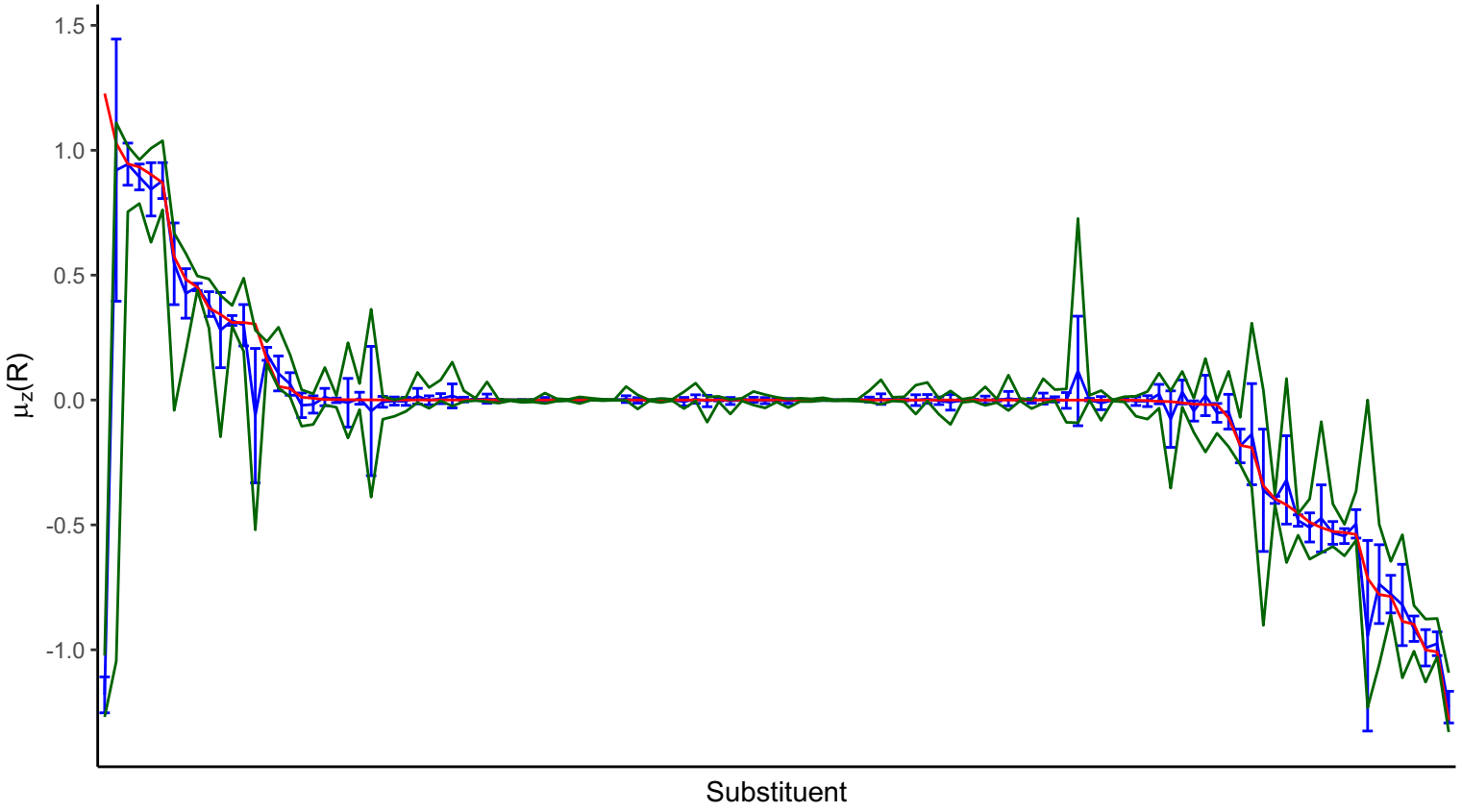

Figure 4.12: Plot displaying μ_z(R) values for $\mu_z(R_H)$, the average μ_z(R) over all other substrates. Error bars are the standard deviation of μ_z(R) over all other substrates. Green lines show the range of μ_z(R) values over all other substrates

The group dipole is a composite property, the sum of a through-bond dipole, $\boldsymbol{\mu}^c$, and a polarization dipole, $\boldsymbol{\mu}^p$. Linear plots are shown in Appendix B, but much information is readily gleaned from the substituent summary statistics (Figures 4.13- 4.15 and Figures 4.16- 4.18 for $\boldsymbol{\mu}^c$ and $\boldsymbol{\mu}^p$, respectively). Here, we observed that μ_x^c and μ_x^p each have higher standard deviation than μ_x. The lower standard deviation in μ_x is a result of the cancellation of errors between μ_x^c and μ_x^p. This is confirmed in the linear plots in Appendix B, Figure B17 and B22, the clearest cases (GeH_3, SiH_3, Cl, and F) given in Figure 4.19. In Figure 4.19(a), the regression lines for $\mu_x^c(R_{F,\,Cl})$ lie above the $y = x$ line and in Figure 4.19(b) $\mu_x^p(R_{F,\,Cl})$ lie below the $y = x$ line. The opposite is true for $\mu_x^c(R_{GeH_3,\,SiH_3})$, illustrating the cancellation of errors leading to the lower standard deviation in $\mu_x(R)$. Figures 4.13- 4.15 and Figures 4.16- 4.18 show that $\boldsymbol{\mu}^c$ and $\boldsymbol{\mu}^p$ have low standard deviations and ranges for the y and z direction - the most variance is seen in the x direction, mirroring the observations in the total dipole.

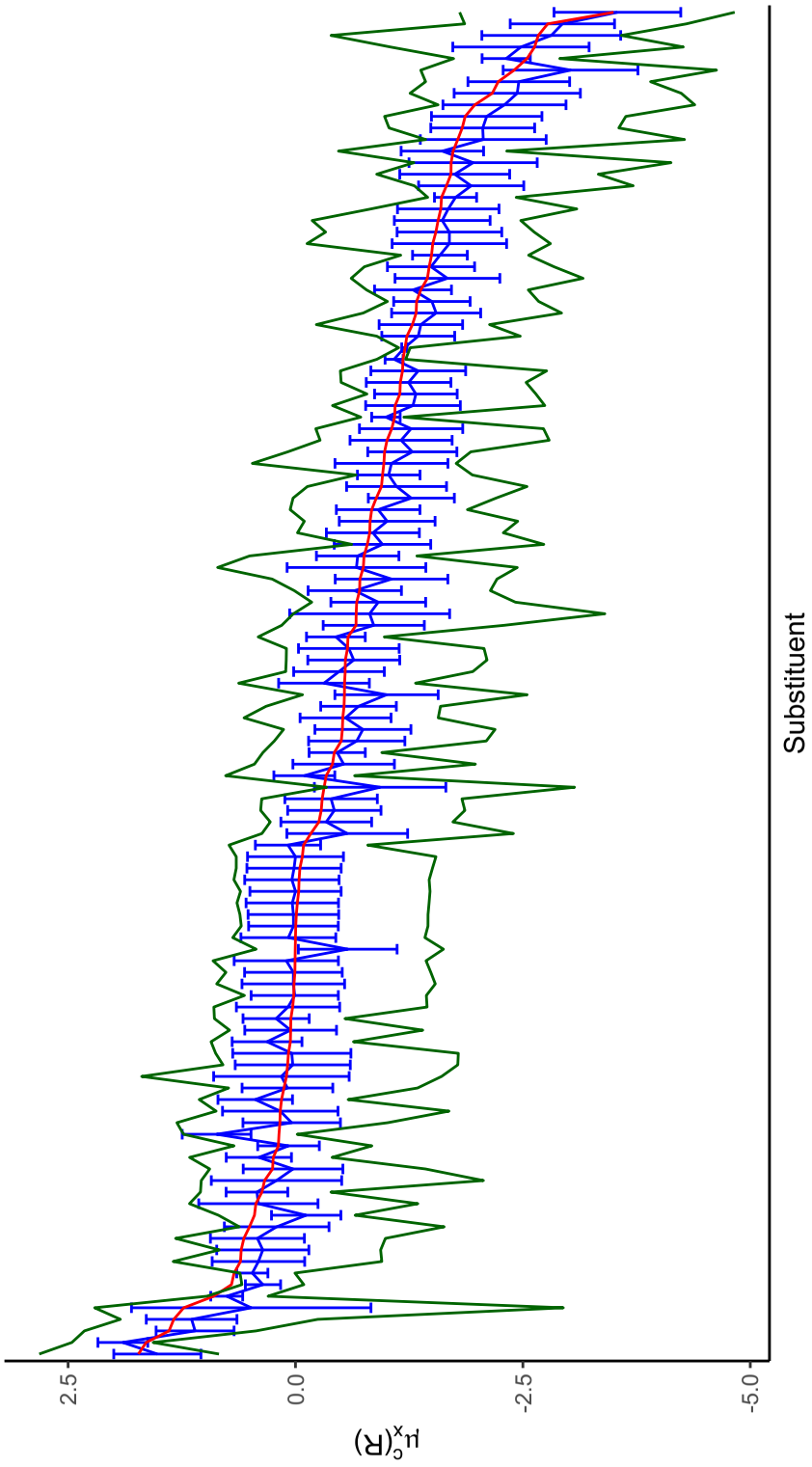

Figure 4.13: Plot displaying μ_x^c(R) values for $\mu_x^c(R_H)$, the average μ_x^c(R) over all other substrates. Error bars are the standard deviation of μ_x^c(R) over all other substrates. Green lines show the range of μ_x^c(R) values over all other substrates

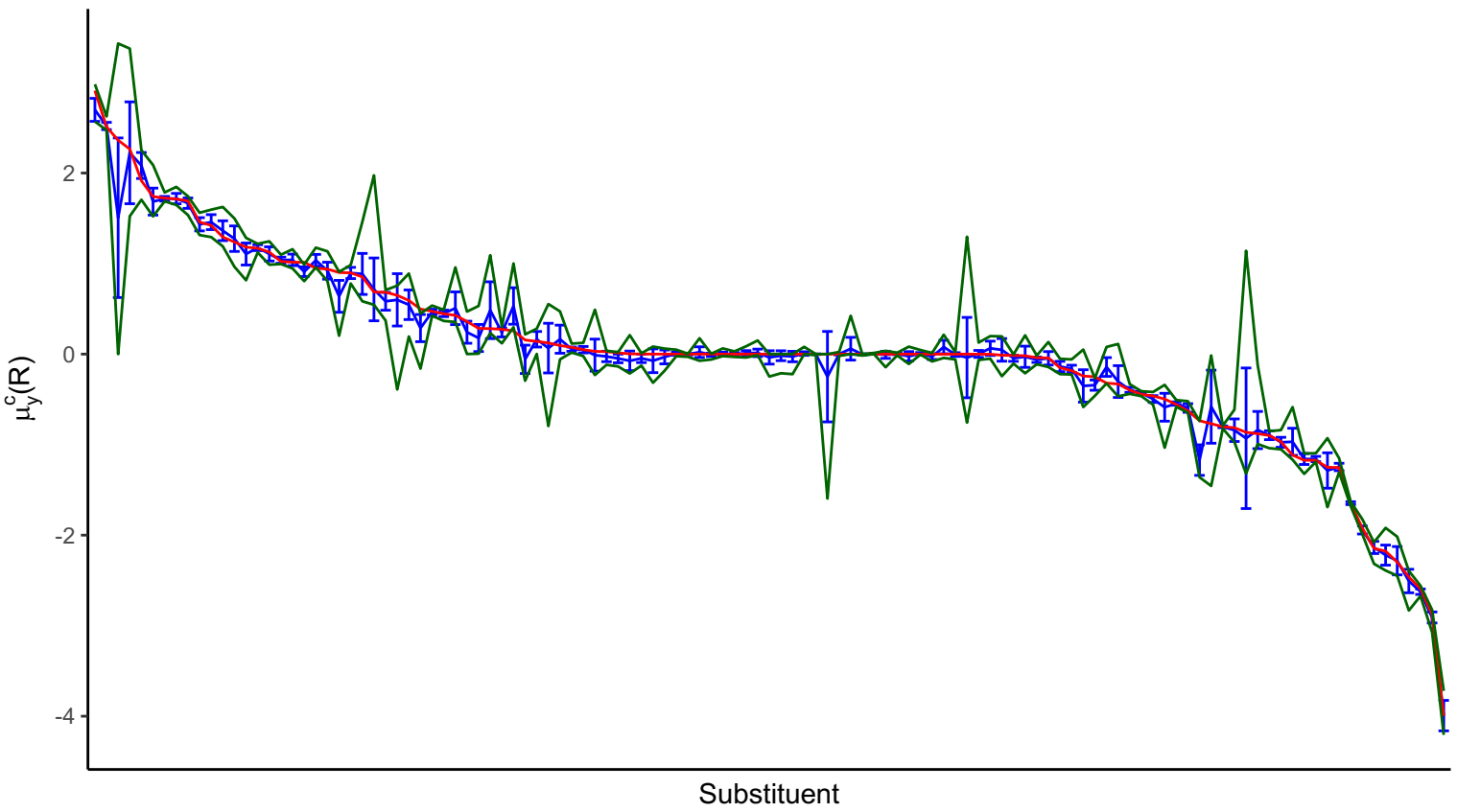

Figure 4.14: Plot displaying μ_y^c(R) values for $\mu_y^c(R_H)$, the average μ_y^c(R) over all other substrates. Error bars are the standard deviation of μ_y^c(R) over all other substrates. Green lines show the range of μ_y^c(R) values over all other substrates

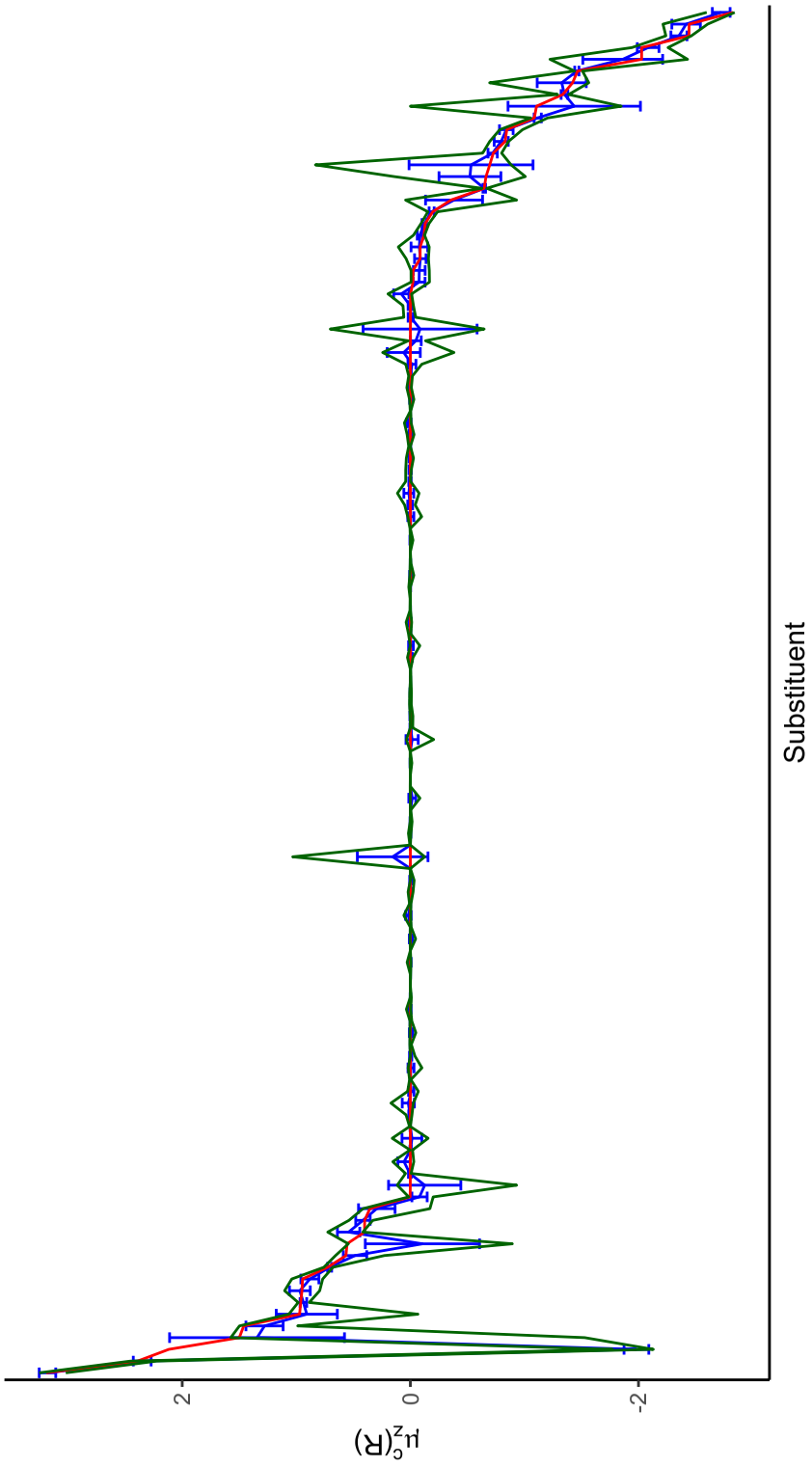

Figure 4.15: Plot displaying $\mu_z^c(R)$ values for $\mu_z^c(R_H)$, the average $\mu_z^c(R)$ over all other substrates. Error bars are the standard deviation of $\mu_z^c(R)$ over all other substrates. Green lines show the range of $\mu_z^c(R)$ values over all other substrates

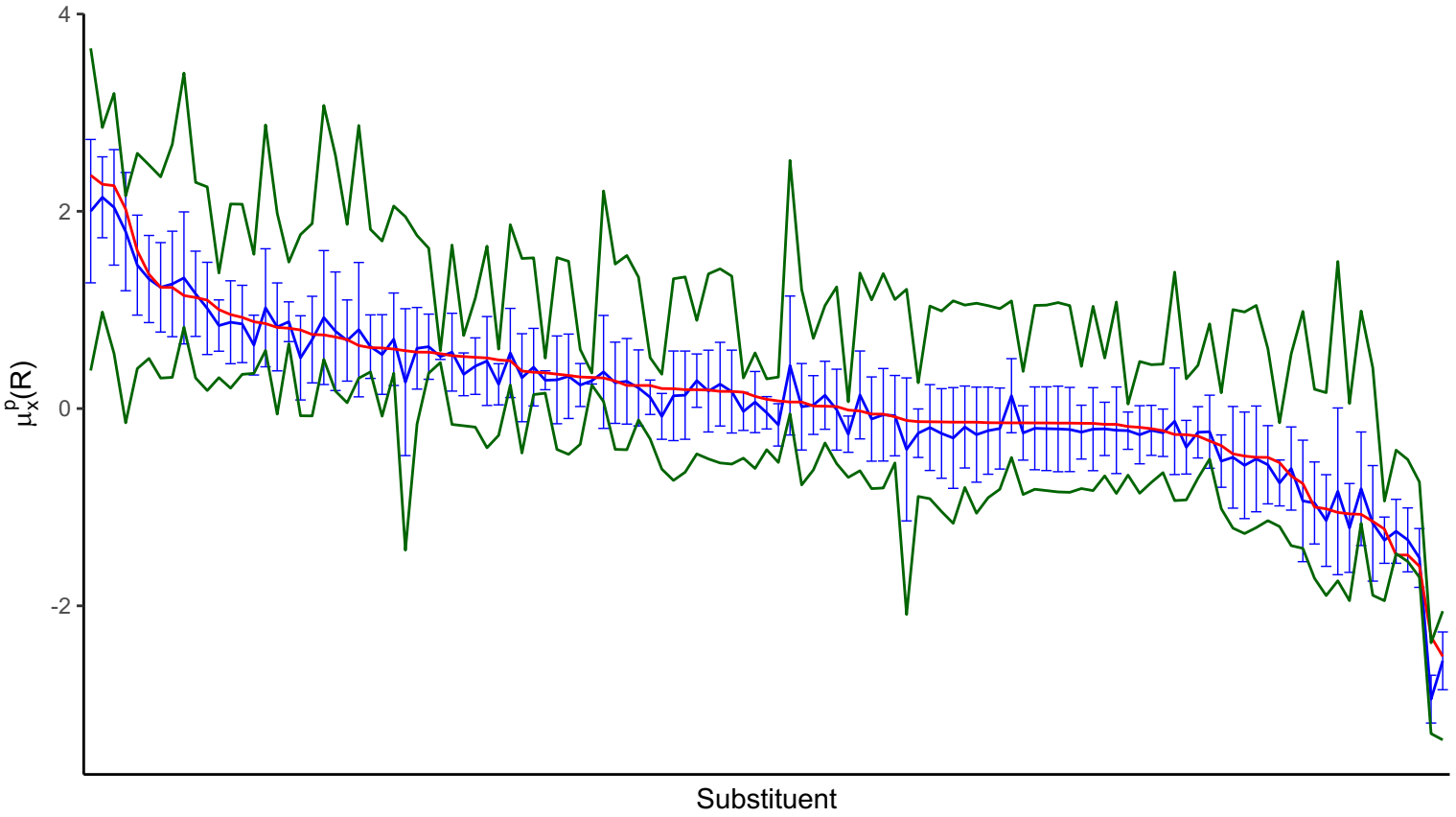

Figure 4.16: Plot displaying μ_x^p(R) values for $\mu_x^p(R_H)$, the average μ_x^p(R) over all other substrates. Error bars are the standard deviation of μ_x^p(R) over all other substrates. Green lines show the range of μ_x^p(R) values over all other substrates

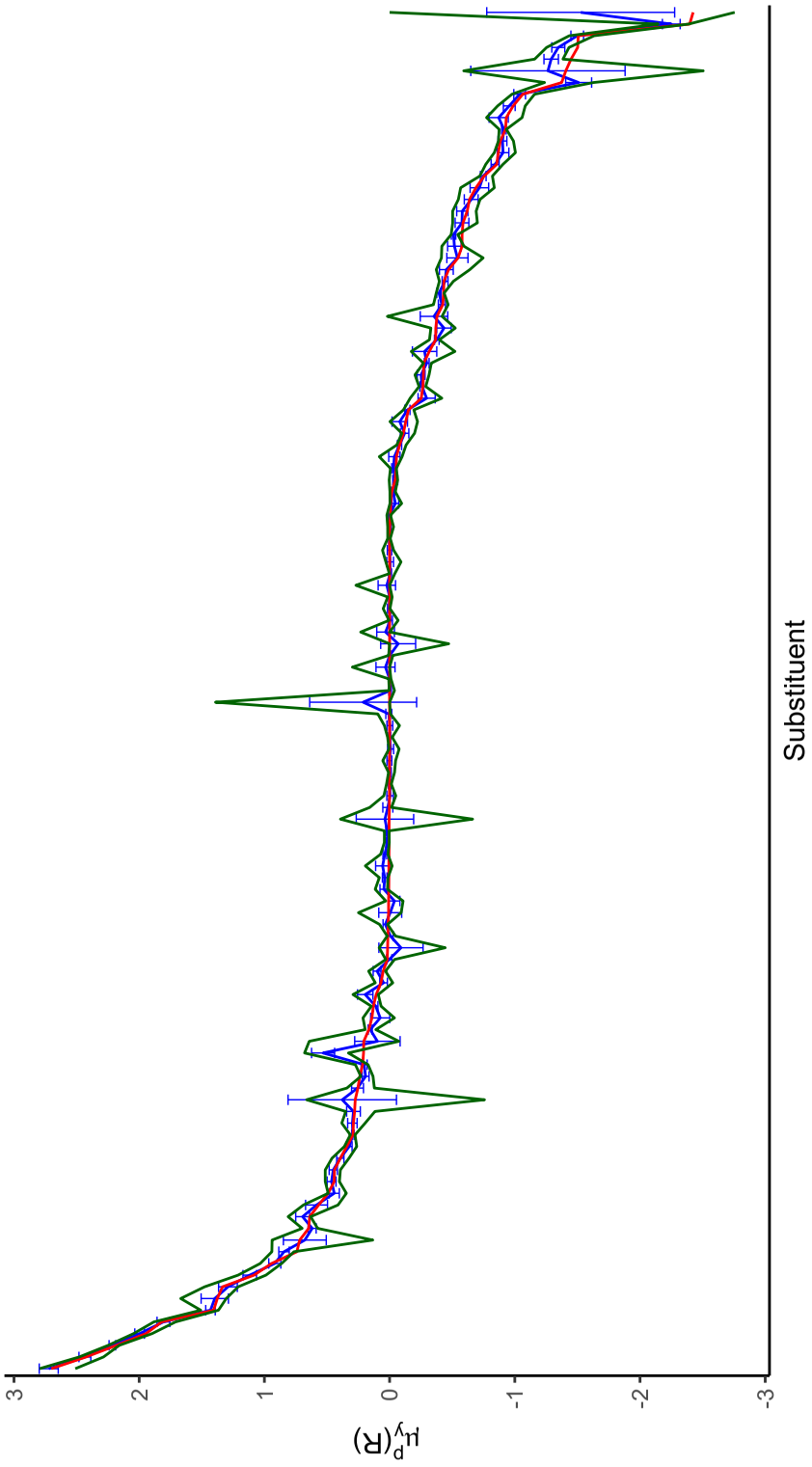

Figure 4.17: Plot displaying $\mu_y^p(R)$ values for $\mu_y^p(R_H)$, the average $\mu_y^p(R)$ over all other substrates. Error bars are the standard deviation of $\mu_y^p(R)$ over all other substrates. Green lines show the range of $\mu_y^p(R)$ values over all other substrates

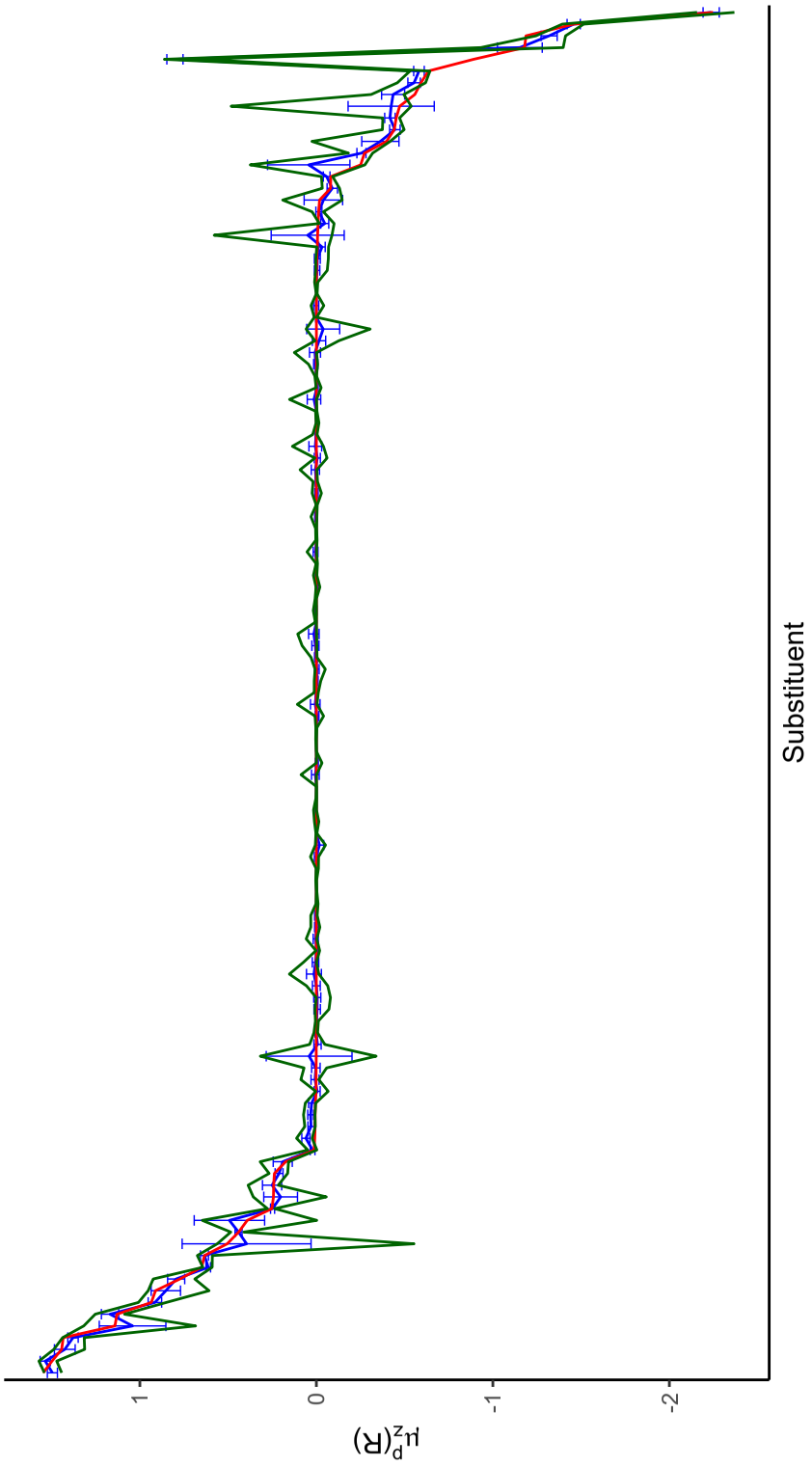

Figure 4.18: Plot displaying μ_z^p(R) values for $\mu_z^p(R_H)$, the average μ_z^p(R) over all other substrates. Error bars are the standard deviation of μ_z^p(R) over all other substrates. Green lines show the range of μ_z^p(R) values over all other substrates

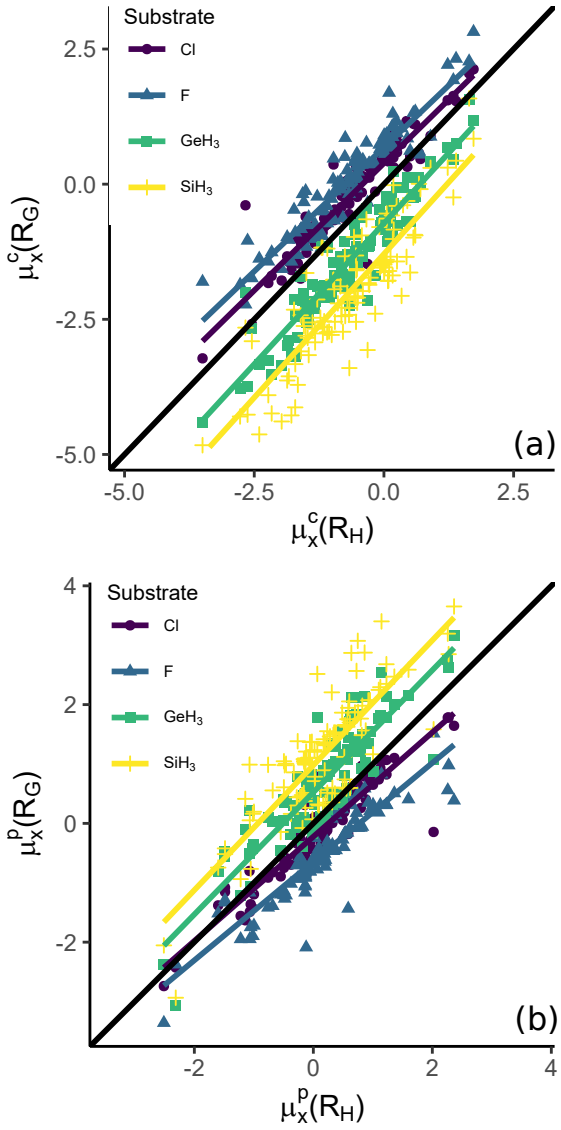

Figure 4.19: Linear plots of (a) $\mu_x^c(R_G)$ and (b) $\mu_x^p(R_G)$ against $\mu_x^c(R_H)$ and $\mu_x^p(R_H)$, respectively

Overall, functional group dipole moment vectors are transferable between carbon-based substrates and substrate hydrogen. For specific interactions, there is some outliers in μ_y and μ_z. These may be applied with only small lack of accuracy. Due to the extra outliers in substrates of differing electronegativity, more caution should be used in using $\boldsymbol{\mu}(R_H)$ values in those substrates, but generally these outliers are minimal. Additionally, the outliers can be intuited in terms of changing geometry, which can be accounted for in machine learning models.

4.4.3 Group Quadrupole

The next order multipole can also be quite useful in explaining electrostatic properties of substituents. The quadrupoles have found use in explaining substituent effects before.[58] Figures 4.20- 4.22 show the linear relationships between $Q_{ii}(R_G)$ and $Q_{ii}(R_H)$ ($i = x$, y, or z). Off diagonal elements are not examined here due to many near-zero values. For Q_{xx}(R), qualitatively good linear relationships are seen, with a few outliers. In Figure 4.20A one of the outliers has been previously discussed. This is NCO, which exhibits geometry differences between substrates. Another outlier is $OCOCH_3$ for which there is a dihedral change. The outlier in Figure 4.20B for GeH_3 and Cl is CH_2NO_2-op.

Like Q_{xx}(R), Q_{yy}(R) is exceedingly well conserved, except for a few outliers. Many outliers are for the same substituents as previously discussed, as in CH_2OOCH_3-op, NCO, and OPH_2O in a variety of substrates. These were due to geometry changes in the located optimized structures. The large outlier in Figure 4.21B is CHCCO, whose optimized geometry differs from H_2CCCO for those substrates. It is worth noting that some of the observed excellent conservation in Q_{yy}(R) may be due to the definition of the coordinate system. Here, the y axis is fixed for each given substituent, the same between all substrates. These

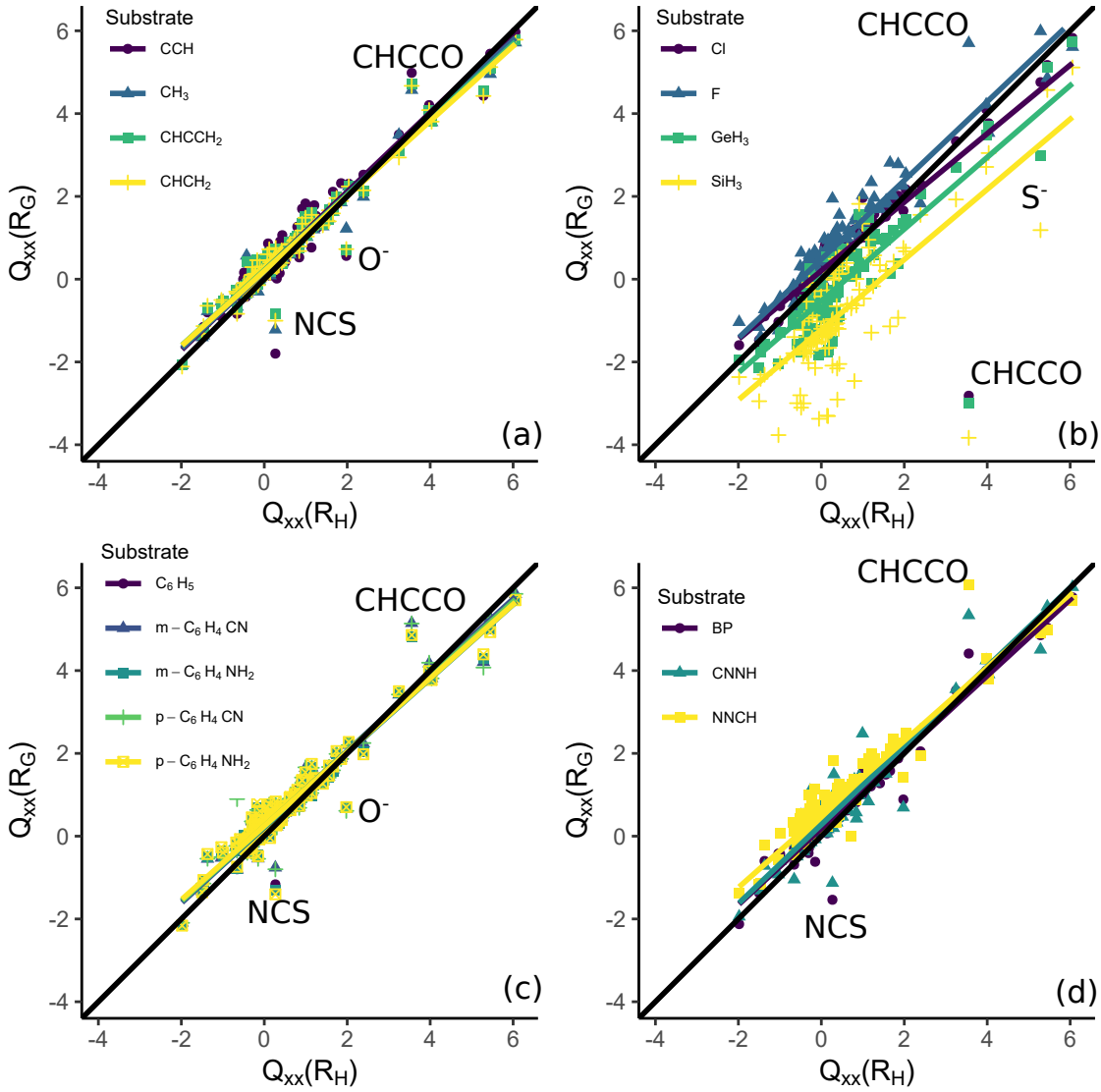

Figure 4.20: Linear plots of Q_{xx}(R) for R in various R-G systems against Q_{xx}(R). (a) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (b) G: Cl, F, SiH_3. GeH_3; (c) C_6H_5, m−C_6H_4CN, m−$C_6H_4NH_2$, p−C_6H_4CN, p−$C_6H_4NH_2$; D: BP, CNNH, NNCH

are still sensitive to geometry changes, but if no significant geometry changes occur, there may be a bias to the conservation of *y*-directional properties.

Visually, Q_{zz}(R) appears the best conserved of all three directions. This aligns with the expectations observed by averaging the linear regression parameters. High linearity is observed in all cases. Of these, the substrates of varying electronegativity in Figure 4.22B exhibit the most non-linearity, but even this is trivial, as each substrate only has

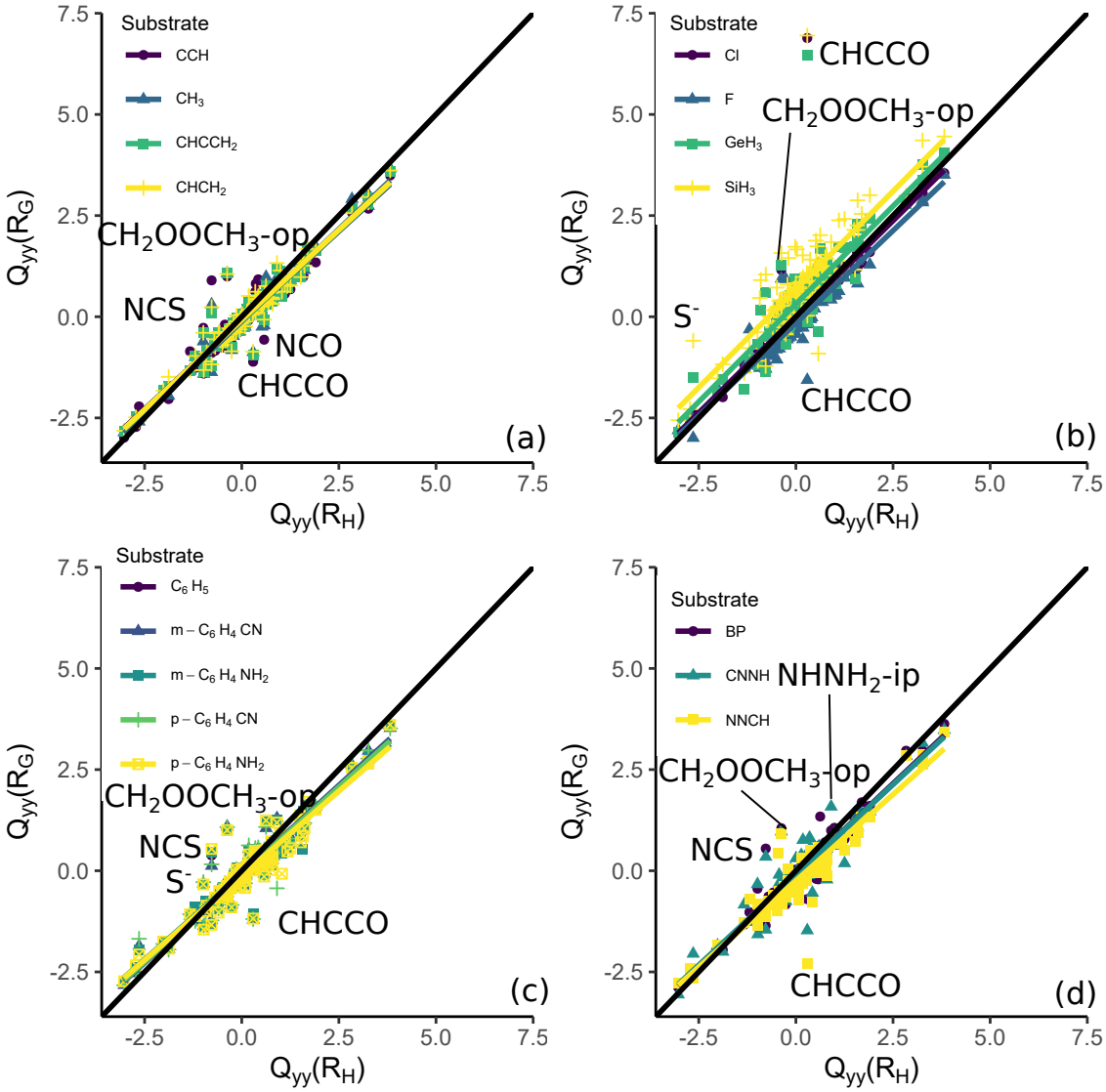

Figure 4.21: Linear plots of Q_{yy}(R) for R in various R-G systems against Q_{yy}(R). (a) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (b) G: Cl, F, SiH_3. GeH_3; (c) C_6H_5, m−C_6H_4CN, m−$C_6H_4NH_2$, p−C_6H_4CN, p−$C_6H_4NH_2$; D: BP, CNNH, NNCH

a few outliers (usually the same substituent across all substrates). A large reason for this is many points are positioned at (0,0). This is due to many planar or symmetric substituents around the *xy* plane. The few small outliers observed are mainly OPH_2O, and CH_2OOCH_3-op, who exhibit geometry discrepancies in their dihedral angles relative to substrate hydrogen.

Similar to the dipoles, there is additional insight to be gained in plotting the average property value, property value attached to hydro-

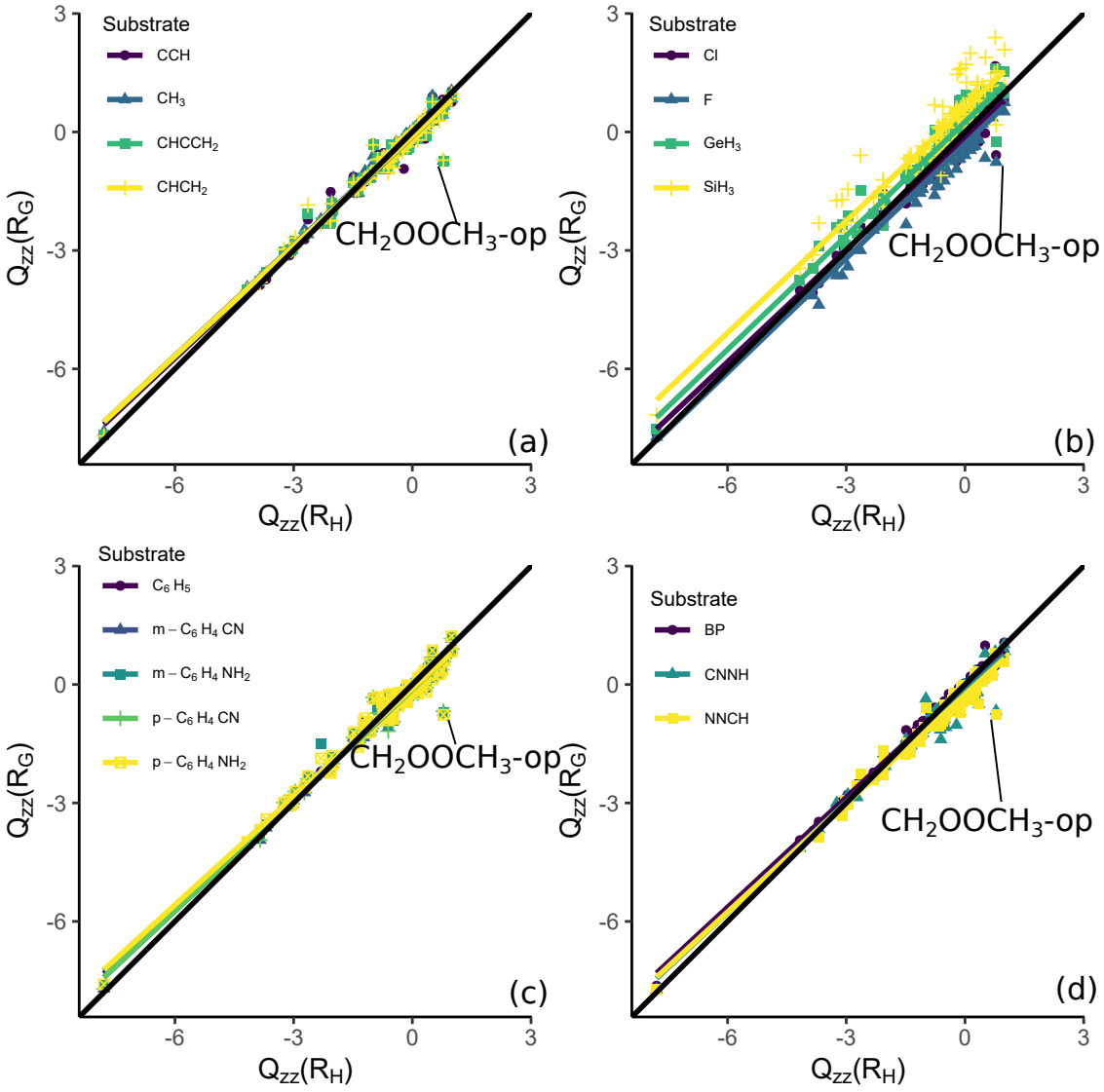

Figure 4.22: Linear plots of Q_{zz}(R) for R in various R-G systems against Q_{zz}(R). (a) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (b) G: Cl, F, SiH_3. GeH_3; (c) C_6H_5, m−C_6H_4CN, m−$C_6H_4NH_2$, p−C_6H_4CN, p−$C_6H_4NH_2$; D: BP, CNNH, NNCH

gen, and the maximum and minimum values. These are shown in Figure 4.23-Figure 4.25. For each substituent, Q_{xx}(R) has the largest range and standard deviation difference of the three direcitonal quadrupoles, as the substrate-substituent bond lies in this direction. Figure 4.25 illustrates more clearly the large number of substituents with near-zero Q_{zz}(R), comprising most of the data set. Not as many substituents have near-zero Q_{xx}(R) and Q_{yy}(R) as compared to Q_{zz}(R). For Q_{yy}(R) and Q_{zz}(R), the maximum and minimum error values are fairly close to-

gether for most substituents. The large spike in Figure 4.24 for Q_{yy}(R) is the substituent CHCCO, which observes a geometry difference between certain substrates and hydrogen, as previously discussed.

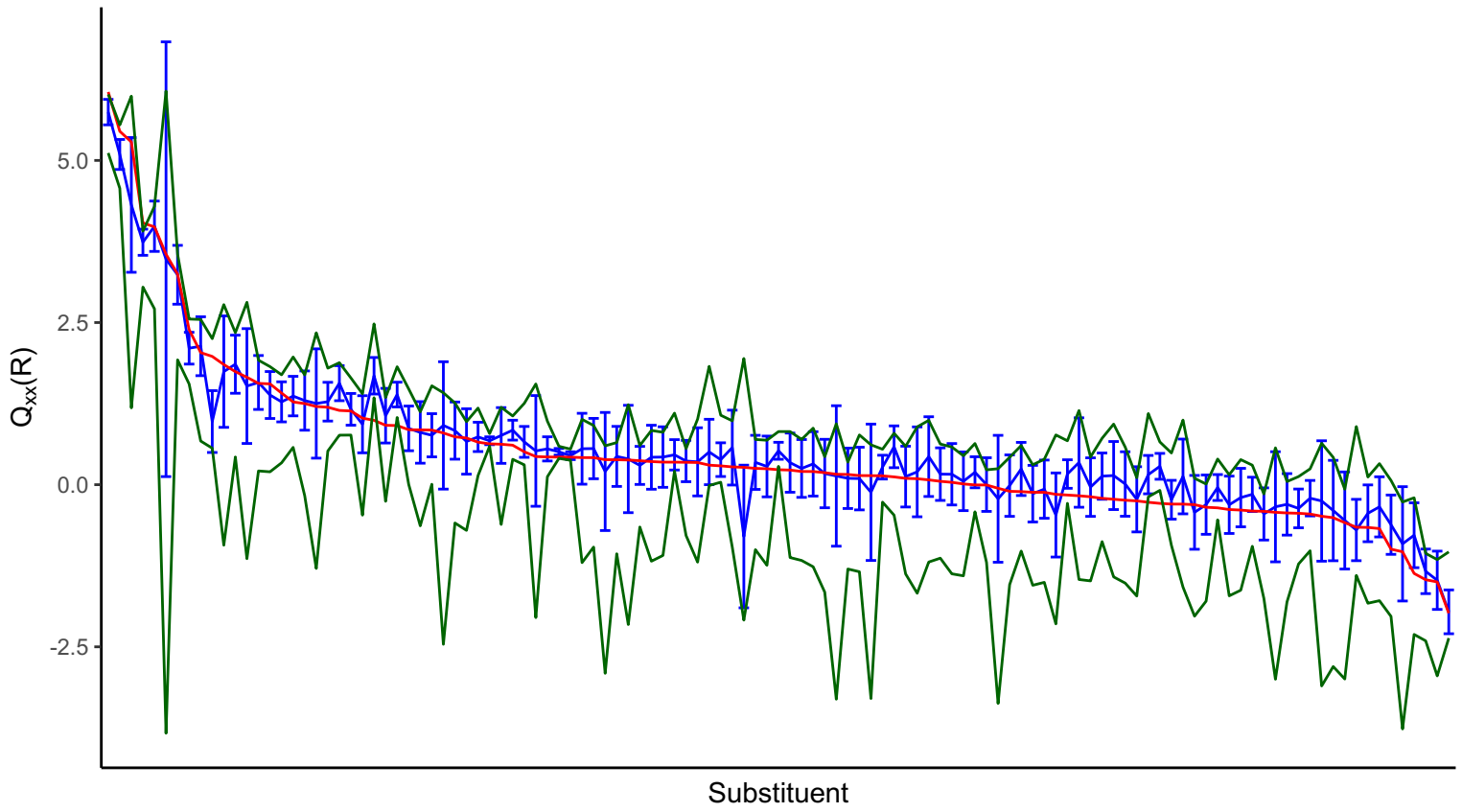

Figure 4.23: Plot displaying Q_{xx}(R) values for $Q_{xx}(R_H)$, the average q(R) over all other substrates. Error bars are the standard deviation of Q_{xx}(R) over all other substrates. Green lines show the range of Q_{xx}(R) values over all other substrates.

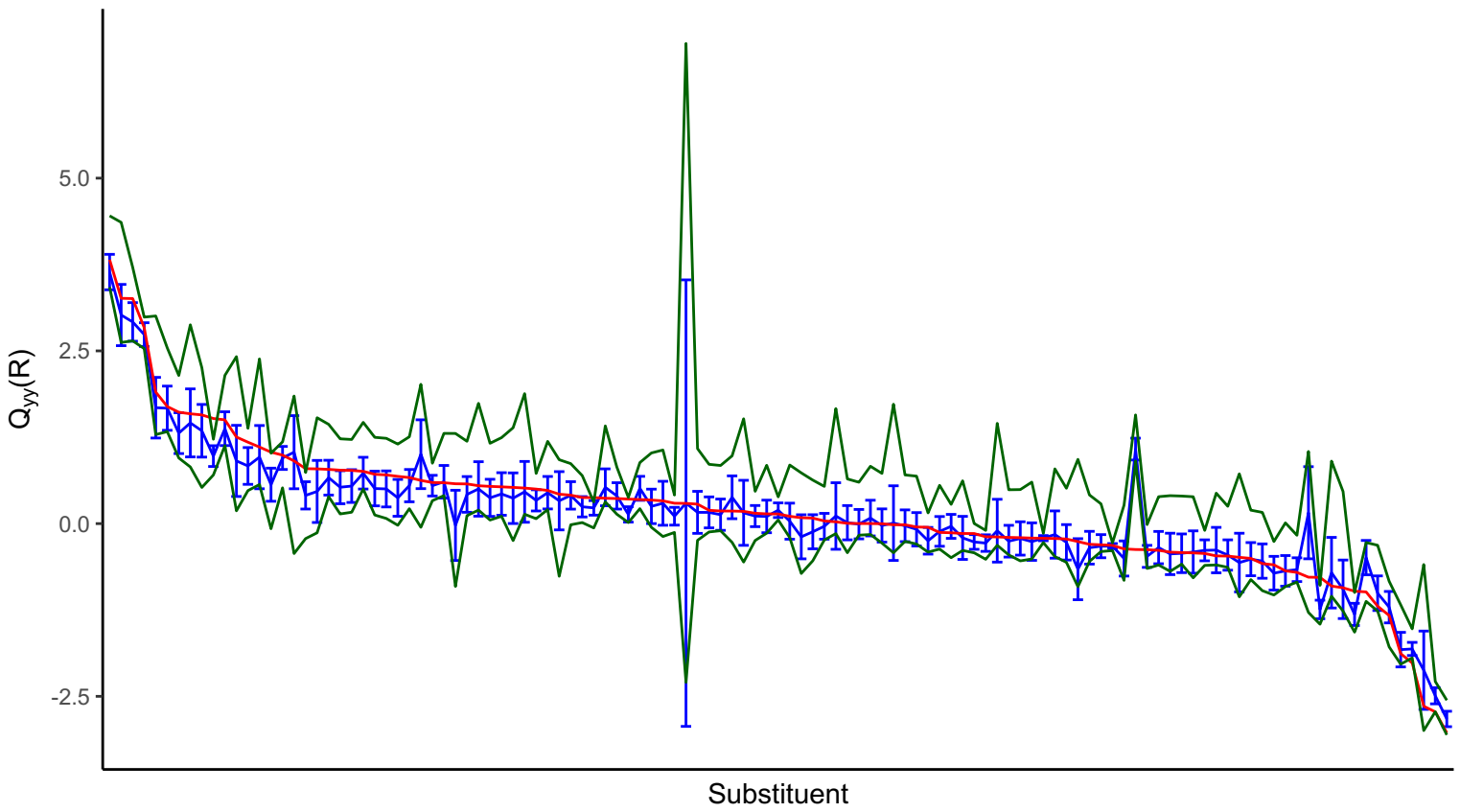

Figure 4.24: Plot displaying Q_{yy}(R) values for $Q_{yy}(R_H)$, the average q(R) over all other substrates. Error bars are the standard deviation of Q_{yy}(R) over all other substrates. Green lines show the range of Q_{yy}(R) values over all other substrates.

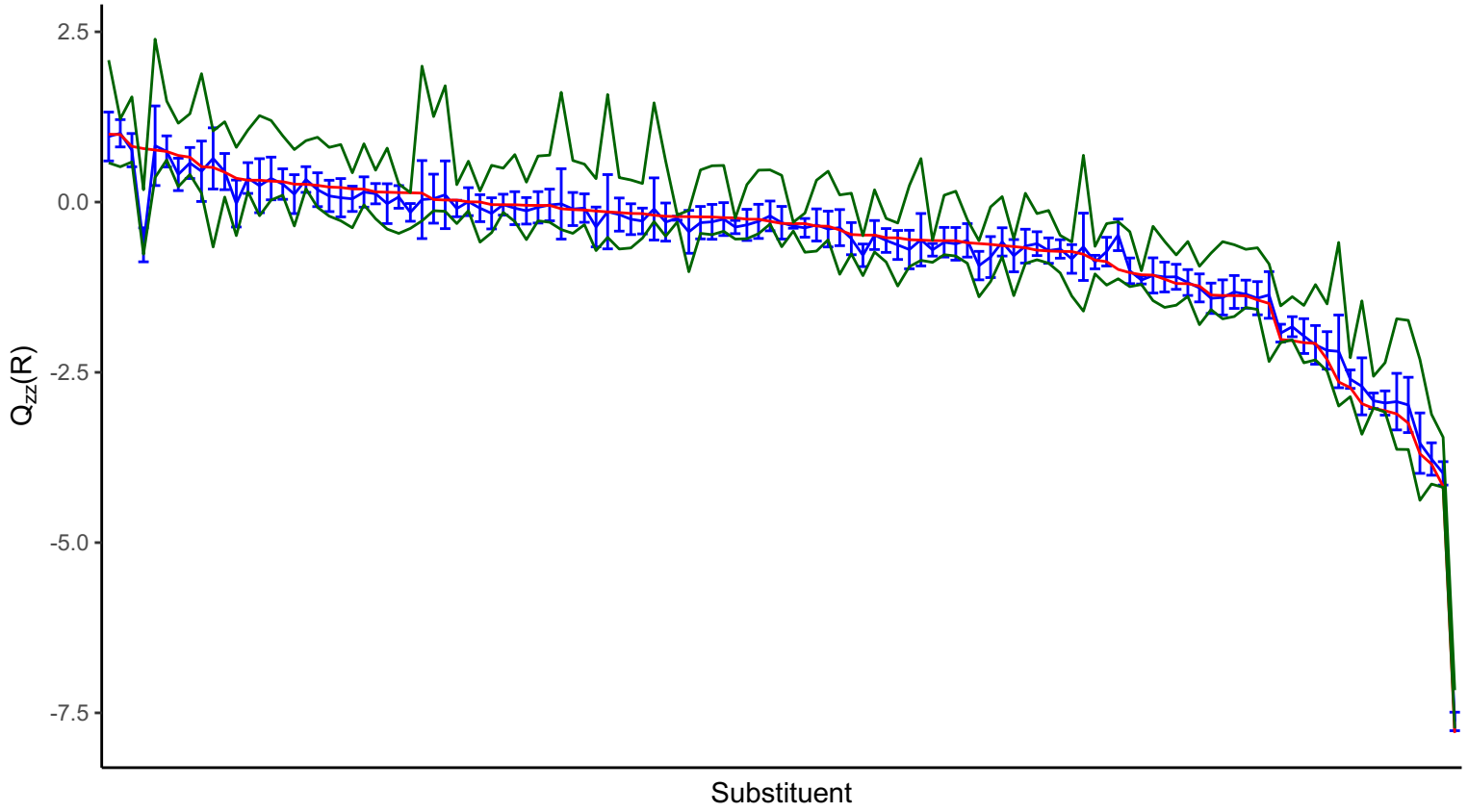

Figure 4.25: Plot displaying Q_{zz}(R) values for $Q_{zz}(R_H)$, the average q(R) over all other substrates. Error bars are the standard deviation of Q_{zz}(R) over all other substrates. Green lines show the range of Q_{zz}(R) values over all other substrates.

In the distribution of quadrupole values seen in Figures 4.23- 4.25, and there are clear patterns. Q_{xx}(R) is large and positive for substituents with sp or sp^2 hybridized carbon atoms. CCCl and CCF have the largest Q_{xx}(R). Carbon and hydrogen based substituents have low positive Q_{xx}(R). Negative values are associated with substituents containing electronegative atoms like chlorine and fluorine (with the exceptions of CCCl and CCF). For Q_{yy}(R), there are a few substituents with large positive values, and many substituents with Q_{yy}(R) ranging from 1 to -1. The substituents with the most negative Q_{yy}(R) are CCCl and CCF. Large negative Q_{zz}(R) are observed for planar substituents with π systems. Substituents with lone pairs have moderate negative Q_{zz}(R).

All the above evidence points to the quadrupoles being suitable for general use in linear regression and machine learning models, as they exhibit an acceptable amount of transferability between substrates. The main outliers are easily observed from geometric descriptors of the molecule. If another substrate, such as a carbon containing one, was chosen as reference, these outliers would disappear. In fact, hydrogen as a substrate is itself an outlier caused by its complete absence of steric bulk. Many substituents that appear as outliers in these graphs are not outliers when the data set as a whole is taken into account without a reference.

4.4.4 Group Volume

The steric bulk of a substituent is obtained by QTAIM through the total volume of the atomic basins comprising the substituent atoms. The volume of the substituent is universally excellently conserved ($r^2 > 0.99$), as revealed in the linear plots shown in Figure 4.26. Deviations from linearity are minimal, if observed at all. Points exhibiting deviance from Vol(R_H) are charged substituents, with the maximum negative error be-

ing attributed to O^- in all substrates. Even this error is minimal and has only a small effect on the observed linear relationships, most of which exhibit a slope of 1. Some deviance in the intercept is observed, but even with this deviance, the lines are still close together, almost overlapping. Overall, these group volumes are excellently transferable, within a small additive factor.

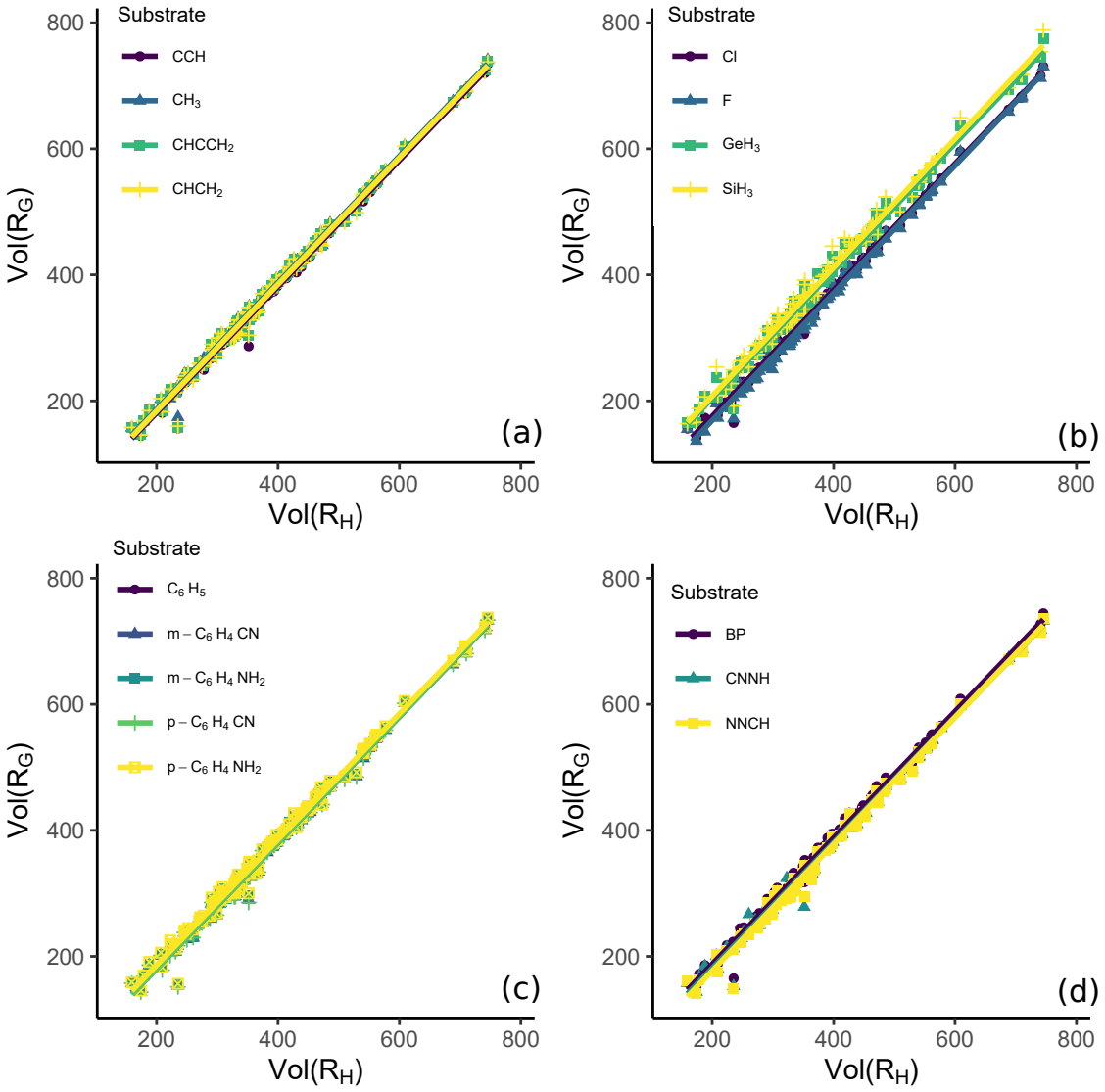

Figure 4.26: Linear plots of Vol(R) for R in various R-G systems against Vol(R). (a) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (b) G: Cl, F, SiH_3. GeH_3; (c) C_6H_5, m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; D: BP, CNNH, NNCH

The plot of the average, maximum, minimum, and Vol(R_H) values in Figure 4.27 confirms the transferability of functional group volumes. The range over which the values occur is low. However, the average is slightly lower than the Vol(R_H) line, representative of the change in intercepts of the linear plots. The takeaway is that the group volumes are highly transferable, and exhibit a good linear trend, only shifted by the intercepts. They could prove useful in regressions and machine learning models as a measure of the substituent steric bulk. Therefore, the volume of a substituent is perfectly transferable.

Figure 4.27 illustrates that a variety of sizes of substituents were chosen, with most having volumes below 600 au^3. The largest substituents have 4 or more carbons, or multiple chlorine atoms. The data set has no overabundance of substituents of a given size, observed from the lack of a plateau in the values. However, there are no middling substituent sizes between about 600 and 700 au^3.

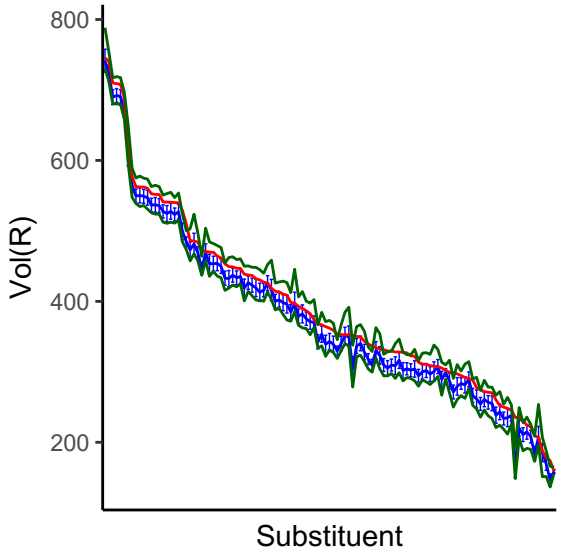

Figure 4.27: Plot displaying Vol(R) values for Vol(R_H), the average Vol(R) over all other substrates. Error bars are the standard deviation of Vol(R) over all other substrates. Green lines show the range of Vol(R) values over all other substrates, Interaction means a large change in the expected substituent-substrate bond is undergone, and planarity means one of the atoms in the substituent adopts a more planar geometry

4.5 Conclusions

With the advent of machine learning in chemistry, generating physically meaningful variables for constructing these models is vital. The observed linear relationships of QTAIM functional group descriptors calculated between different substrates illustrate the potential of these descriptors for machine learning models, as deviations from linearity are interpretable and predictable based on substituent geometry. These changes in geometry can be included in machine learning models. Properties calculated for the functional groups bonded to hydrogen are often sufficient for use in building these models for other systems. Volumes are perfectly transferable and other properties have generally acceptable linearity between $P(R_G)$ and $P(R_H)$ are often observed. The lack of linearity in many cases stems from a small number of interpretable outlying points.

This paper represents another step for applying QTAIM in machine learning models. The non-ideal linear relationships in the direction along the substituent-substrate bond present another task to understand if these descriptors are to be used this way. Work is being undertaken on elucidating the mutual interaction between the substituents and substrates. A clear understanding of the interactions, and the transferability of the descriptors paves the way for this application. Through using these descriptors and machine learning models, a clearer description of computationally predicting and modeling reactivity may someday be achieved.

4.5.1 Acknowledgements

This research was funded by the National Sciences and Engineering Research Council of Canada, Government of Ontario through the Ontario

Graduate Scholarship, and the Lakehead University Faculty of Graduate Studies. This research was enabled in part by support provided by the Shared Hierarchical Academic Research Computing Network (www.sharcnet.ca) and Compute Canada (www.computecanada.ca).

References

[1] Strieth-Kalthoff, F., Sandfort, F., Segler, M. H. S. and Glorius, F., *Chem. Soc. Rev.* **2020**, *49*, 6154–6168.

[2] Pflüger, P. M. and Glorius, F., *Angew. Chem. Int. Ed.* **2020**, *59*, 18860–18865.

[3] Panteleev, J., Gao, H. and Jia, L., *Bioorg. & Med. Chem. Lett.* **2018**, *28*, 2807–2815.

[4] Ramakrishnan, R., Dral, P. O., Rupp, M. and von Lilienfeld, O. A., *J. Chem. Theory and Comput.* **2015**, *11*, 2087–2096.

[5] Beker, W., Gajewska, E. P., Badowski, T. and Grzybowski, B. A., *Angew. Chem. Int. Ed.* **2019**, *58*, 4515–4519.

[6] Hammett, L. P., *J. Am. Chem. Soc.* **1937**, *59*, 96–103.

[7] Krygowski, T. M. and Sadlej-Sosnowska, N., *Struct. Chem.* **2011**, *22*, 17–22.

[8] Galabov, B., Ilieva, S. and Schaefer, H. F., *J. Org. Chem.* **2006**, *71*, 6382–6387.

[9] Charton, M., *J. Org. Chem.* **1961**, *26*, 735–738.

[10] Charton, M., *J. Org. Chem.* **1965**, *30*, 969–973.

[11] Pichon-Pesme, V., Lecomte, C. and Lachekar, H., *J. Phys. Chem.* **1995**, *99*, 6242–6250.

[12] Zarychta, B., Pichon-Pesme, V., Guillot, B., Lecomte, C. and Jelsch, C., *Acta Crystallogr. Sect. A* **2007**, *A63*, 108–125.

[13] Koritsanszky, T., Volkov, A. and Coppens, P., *Acta Crystallogr. Sect. A* **2002**, *A58*, 464–472.

[14] Volkov, A., Li, X., Koritsanszky, T. and Coppens, P., *J. Phys. Chem. A* **2004**, *108*, 4283–4300.

[15] Volkov, A., Messerschmidt, M. and Coppens, P., *Acta Crystallogr. Sect. D* **2007**, *D63*, 160–170.

[16] Domagała, S., Fournier, B., Liebschner, D., Guillot, B. and Jelsch, C., *Acta Crystallogr. Sect. A* **2009**, *A68*, 337–351.

[17] Leduc, T., Aubert, E., Espinosa, E., Jelsch, C., Iordache, C. and Guillot, B., *J. Phys. Chem. A* **2019**, *123*, 7156–7170.

[18] Popelier, P. L. A. and Aicken, F. M., *J. Am. Chem. Soc* **2003**, *125*, 1284–1292.

[19] Popelier, P. L. A. and Aicken, F. M., *ChemPhysChem* **2003**, *4*, 824–829.

[20] Fletcher, T. L. and Popelier, P. L. A., *Theor. Chem. Acc.* **2015**, *134*, 1–16.

[21] Popelier, P. L. A., *Physica Scripta* **2016**, *91*, 033007.

[22] Maxwell, P. and Popelier, P. L. A., *Mol. Phys.* **2016**, *114*, 1304–1316.

[23] Cortés-Guzmán, F. and Bader, R. F. W., *J. Phys. Org. Chem.* **2004**, *17*, 95–99.

[24] Breneman, M., Thompson, R., Rhem, M. and Dung, M., *Computers. Chem.* **1995**, *19*, 161–179.

[25] Whitehead, C., Breneman, C., Sukumar, N. and Ryan, M. D., *J. Comput. Chem.* **2003**, *24*, 512–529.

[26] Bader, R. F. W. and Bayles, D., *J. Phys. Chem. A* **2000**, *104*, 5579–5589.

[27] Riess, J. and Münch, W., *Theoretica chimica acta* **1981**, *58*, 295–300.

[28] Bader, R. F. and Becker, P., *Chem. Phys. Lett.* **1988**, *148*, 452–458.

[29] Lorenzo, L. and Mosquera, R. A., *Chem. Phys.* **2002**, *356*, 305–312.

[30] Bader, R. F. W. and Matta, C. F., *Found. Chem.* **2013**, *15*, 253–276.

[31] Bader, R. F. W., *J. Chem. Phys.* **1972**, *56*, 3320.

[32] Cortés-Guzmán, F. and Bader, R., *Chem. Phys. Lett.* **2003**, *379*, 183–192.

[33] Mandado, M., Graña, A. M. and Mosquera, R. A., *J. Mol. Struct.: THEOCHEM* **2001**, *572*, 223–233.

[34] Mandado, M., Graña, A. M. and Mosquera, R. A., *J. Mol. Struct.: THEOCHEM* **2002**, *584*, 221–234.

[35] Quiñónez, P. B., Vila, A., Graña, A. M. and Mosquera, R. A., *Chem. Phys.* **2003**, *287*, 227–236.

[36] López, J. L., Mandado, M., Graña, A. M. and Mosquera, R. A., *Int. J. Quantum Chem.* **2002**, *86*, 190–198.

[37] López, J. L., Mandado, M., González Moa, M. J. and Mosquera, R. A., *Chem. Phys. Lett.* **2006**, *422*, 558–564.

[38] Vila, A., Carballo, E. and Mosquera, R. A., *Can. J. Chem.* **2000**, *78*, 1535–1543.

[39] Vila, A. and Mosquera, R. A., *Chem. Phys. Lett.* **2000**, *332*, 474–480.

[40] Graña, A. M. and Mosquera, R. A., *J. Chem. Phys.* **1999**, *110*, 6606–6616.

[41] Graña, A. M. and Mosquera, R. A., *J. Chem. Phys.* **2000**, *113*, 1492–1500.

[42] Popelier, P. L. A., Devereux, M. and Rafat, M., *Acta Crystallogr. Sect. A* **2004**, *A60*, 427–433.

[43] Rykounov, A. A. and Tsirelson, V. G., *J. Mol. Struct.: THEOCHEM* **2009**, *906*, 11–24.

[44] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A. J., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Keith, T., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J. and Fox, J. D., Gaussian 09 **2010**.

[45] Becke, A. D., *Phys. Rev. A* **1988**, *38*, 3098–3100.

[46] Lee, C., Yang, W. and Parr, R. G., *Phys. Rev. B* **1988**, *37*, 785–789.

[47] Becke, A. D., *J. Chem. Phys* **1993**, *98*, 5648–5652.

[48] Rappoport, D. and Furche, F., *J. Chem. Phys.* **2010**, *133*, 143105.

[49] Boyd, R. J. and Edgecombe, K. E., *J. Am. Chem. Soc.* **1988**, *110*, 4182–4186.

[50] Smith, A. P., McKercher, A. E. and Mawhinney, R. C., *J. Phys. Chem. A* **2011**, *115*, 12544–12554.

[51] Keith, T. A., AIMAll **2017**.
URL `aim.tkgristmill.com` Accessed Sept. 1, 2019

[52] Laidig, K., *Chem. Phys. Lett.* **1991**, *185*, 483–489.

[53] R Core Team, R: A Language and Environment for Statistical Computing **2016**.
URL `https://www.r-project.org/` Accessed Sept. 30, 2020

[54] Team, R., RStudio: Integrated Development Environment for R **2016**.
URL `http://www.rstudio.com/` Accessed Sept. 30, 2020

[55] Wickham, H., ggplot2: Elegant Graphics for Data Analysis **2009**.
URL `http://ggplot2.org` Accessed Sept. 30, 2020

[56] Mawhinney, R. C., Muchall, H. M. and Peslherbe, G. H., *Chem. Commun.* **2004**, *5*, 1862–1863.

[57] Lefrancois-Gagnon, K. M. and Mawhinney, R. C., *J. Comput. Chem.* **2020**, *41*, 2485–2503.

[58] Gottardo, C., Kraft, T. M., Hossain, M. S., Zawada, P. V. and Muchall, H. M., *Can. J. Chem.* **2008**, *86*, 410–415.

Chapter 5

Parametrizing Bond Critical Point Properties with Electronegativity Descriptors

5.1 Abstract

Development of physically meaningful parameters for inclusion in machine learning is an important field in chemistry. These parameters are ideally transferable between molecules for the machine learning model to be robust and apply in different situations. Through a study on the bond critical point properties of 117 substituents and 16 substrates we find that BCP properties are not transferable. However, a substituent's changing electronegativity between substrates helps quantify the observed variation. The relationship between changes in bond critical point properties and changes in substituent electronegativity between substrates is universal, with only a few exceptions (constant errors and and changes related to resonance). These patterns enable the use of bond critical point properties in machine learning models.

5.2 Introduction

Machine learning is a frontier for discovery in chemistry, assisting in predicting reaction outcomes or molecular design for specific properties.[1,2] Inclusion and changing of functional groups in molecules modify molecular properties.[3] Changing functional groups at remote sites can make acids more or less acidic.[4] The development of functional group properties to model these types of changes is a continuing field of interest.

One way of quantifying functional group properties is through substituent parameters.[5–8] The first, developed by Hammett in 1937, used the substituent's effect on a carboxylic acid probe in substituted benzoic acids to measure the effective property of the substituent. They must be considered proxies since they do not measure the intrinsic property of the substituent. Other experimental and computational substituent parameters have been developed.[9–15] For machine learning, physically meaningful substituent descriptors are desired.[16] Beker et al. use Hammett's parameters as part of their physically meaningful feature set. However, these descriptors would ideally measure intrinsic properties of substituents rather than measuring their effects through a proxy.

One way of measuring these types of substituent descriptors is through the Quantum Theory of Atoms in Molecules (QTAIM).[17] QTAIM defines atoms as regions of electron density containing a nucleus and separated from other atomic basins by zero-flux surfaces.[18,19] These atomic basins have properties, which can be determined by integrating property densities over the basin.[20] Substituent and functional group properties can then be calculated by summing the atomic properties of the atoms comprising the group.[21] QTAIM also analyses the first derivative of the electron density, $\nabla\rho$, for critical points, where $\nabla\rho = 0$. These

critical points are classified by the number of non-zero curvatures of the electron density and their signs. Bond critical points (BCP) have three non-zero curvatures, two of which are negative. BCPs indicate bonding interactions between atoms.

Scalar properties measured at BCPs give direct information on the electronic character of the bond, efficiently giving insight into the character of the substituent.[22] Previously used physically meaningful parameters used in machine learning, like Hammett's parameters,[16] only indirectly measure the substituent's properties. Additionally, BCP descriptors are easy to evaluate for systems for which the BCP property is not known.

Substituent parameters for use in machine learning are ideally insensitive to the level of theory chosen to measure them and transferable between different substrates. We have shown that QTAIM properties are insensitive to the level of theory chosen, though, for best accuracy, a hybrid functional and triple-ζ basis are recommended,[23] and that group properties are linearly transferable between different substrates.

Bond critical point properties, of course, are expected to vary between substrates, as the substituent atom and the substrate atom to which it is connected both contribute to the property. If the substrate is changed, half the contributions to the bond properties change. Patterns can be of assistance in assessing these types of changes. Patterns are what led Boyd and Edgecombe to their model of electronegativity from QTAIM in AH molecules.[24,25] They observed trends in the electron density at the bond critical point and the distance to bond critical points associated with trends in the periodic table and developed a model of electronegativity. They defined an electronegativity factor, F_A, Equation 5.1 based on these trends, and electronegativity, χ, Equation 5.2. Here, r_H is the distance from the hydrogen to the bond critical point,

r_{AH} is the distance between atoms A and H, ρ_c is the electron density at the A-H BCP, and N_A is the valence of atom A. a, and b are calculated by fits to $\chi_L i$ and χ_F of 1 and 4, respectively.

$$F_A = \frac{r_H}{r_{AH}\rho_c N_A} \tag{5.1}$$

$$\chi_A = aF_A^b \tag{5.2}$$

Here we seek to analyze changes in the interactions between a set of 16 substrates, G, and 117 substituents, R. The expected non-transferability of BCP properties is proven. Electronegativities of the substituents in each substrate are calculated. Changes in BCP properties can be accounted for by parameterizing the error with a change in substituent electronegativity, and what atom of the substituent is bonded to the substrate. It is shown that differences in electronegativity can parametrize the non-transference. This allows BCP properties to be included as physically meaningful descriptors in machine learning models.

5.3 Methodology

All possible combinations of 117 substituents, R, and 16 substrates, G, were optimized using Gaussian09.[26] Additionally, Li-G and F-G were optimized to calculate substituent electronegativity. The model chemistry used incorporated the B3LYP functional[27–29] and def2-TZVPPD basis set,[30] by recommendation of Chapter 3.[23] The chosen substituents are shown in Figures C1 and C2 in Appendix C. Figure 5.1 shows the chosen substrates, which illustrate a variety of bonding situations: carbon based substrates of various hybridisations (CH_3, $CHCH_2$, C_6H_5,

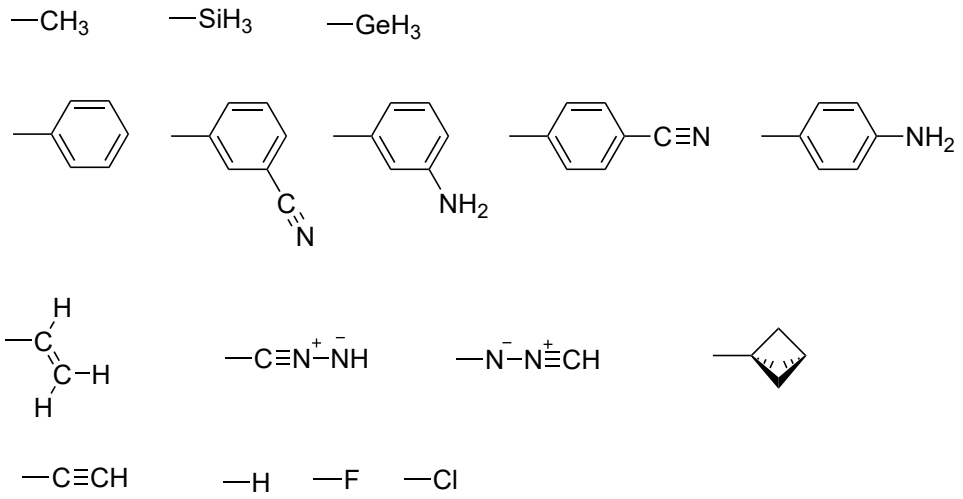

Figure 5.1: Substrates used in the present study

and CCH), non-carbon substrates of varying electronegativity (H, F, Cl, GeH_3, and SiH_3), substrates with varying aromatic substitution (m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$) and substrates of varying flexibility (BP, CNNH, NNCH).

Extended wavefunction files were generated, and QTAIM properties assessed using AIMAll 17.11.14.[31] The R-H BCP properties of interest include the potential energy density, V_c; kinetic energy density, G_c; total energy density, H_c; electron density, ρ_c; Laplacian of the electron density, $\nabla^2\rho_c$; and its associated eigenvalues, λ_i (i=1-3). The generalized electronegativity model for R in an R-G molecule can be written as Equations 5.3 and 5.4. Here r_X is the distance from the atom bonded to R to the R-G BCP, and r_{AX} is the A-X bond length, Figure 5.2. Obtaining χ_R(G) uses the data from G-Li and G-F to fit χ_{Li}(G) and χ_F(G) to 1 and 4, respectively.

$$F_R(G) = \frac{r_X}{r_{AX}\rho_c N_A} \tag{5.3}$$

$$\chi_R(G) = aF_R(G)^b \tag{5.4}$$

Data analysis is performed in R and RStudio.[32,33] Graphics genera-

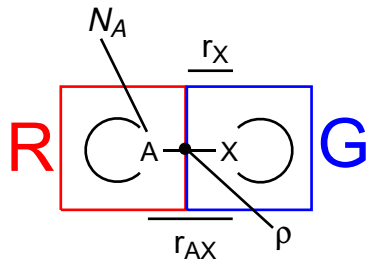

Figure 5.2: Diagram of the system used to calculate χ_R(G)

tion was done using the ggplot2 package.[34] In assessing transferability, MD and MAD summary statistics help summarize changes. Additionally, if properties in RH molecules linearly relate to those in RG molecules, then the properties in RH molecules can be used for those in RG molecules. Slope, intercept, and r^2 of correlations assess linearity. The ideal values are 1, 0 and 1, respectively. Tolerances of those metrics that would indicate transferability are 0.1, 5% of the range over which the property occurs, and 0.1, respectively.

5.4 Results and Discussion

5.4.1 Non-transferability of BCP properties

First, the extent of BCP parameter transferability must be determined. BCP properties studied include ρ, $\nabla^2\rho$, G and V. These values will change between substrates, as one of the two atoms that determine the bond's properties is changing. Table 5.1 shows the ranges over which the various BCP properties of concern occur at the R-H BCP. To compare MD and MAD, it is essential to compare the deviation to the range over which the values occur. The MAD can be evaluated as a percentage over the range the values occur simply by dividing the MAD by the range, Equation 5.5. Here, P represents any given BCP property under consideration. The ranges of $\nabla^2\rho$, G, V, and λ_i (i=1-3) change

drastically between G=H and G≠H, indicating that the BCP properties change between substrates. These changing properties are indicative of non-transferability, which can be further proven using MD and MAD.

$$\%_{\text{rng}}MAD(P) = \frac{\text{MAD}}{\max(P(H)) - \min(\text{P(H)})} \tag{5.5}$$

Table 5.1: Ranges over which the BCP properties are observed

Property	$\text{Min}_{\text{G=H}}$	$\text{Max}_{\text{G=H}}$	$\text{Min}_{\text{G}\neq\text{H}}$	$\text{Max}_{\text{G}\neq\text{H}}$
ρ	0.055	0.380	0.039	0.419
$\nabla^2\rho$	-2.871	0.256	-1.177	1.365
G	0.012	0.141	0.007	0.970
V	-0.862	-0.063	-1.715	-0.032
λ_1	-2.01	-0.073	-1.119	-0.033
λ_2	-1.950	-0.073	-1.119	-0.033
λ_3	0.111	1.123	0.007	2.542
H	-0.790	-0.004	-0.745	0.011

Figure 5.3 shows the MD and MAD of ρ, $\nabla^2\rho$, G and V for all the substrates included in the study. Appendix C contains the raw values of these MD and MAD, and the regression parameters for all linear regressions studied. MD and MAD are high relative to the range for all properties. $\%_{\text{rng}}MAD(P)\rho_c$ is greater than 5% in all cases, typically lying between 10-20%. These large errors indicate the non-transferability of the properties. ρ at the R-G bond is lower than that in the R-H bond, while $\nabla^2\rho$ and G_c are greater. V_c are not consistently over- nor underestimated across all substrates. MD and MAD typically increase in magnitude when the substituent is not attached to a substrate through a carbon atom. That is, for G=Cl, F, NNCH, GeH_3 and SiH_3, increased MAD are often observed. Based these high MAD values, BCP properties are not transferable between substrates.

Regardless of the magnitude of observed deviations, as was previously observed for functional groups in Chapter 4, a linearity might be present between BCP properties evaluated at the R-G BCP and those

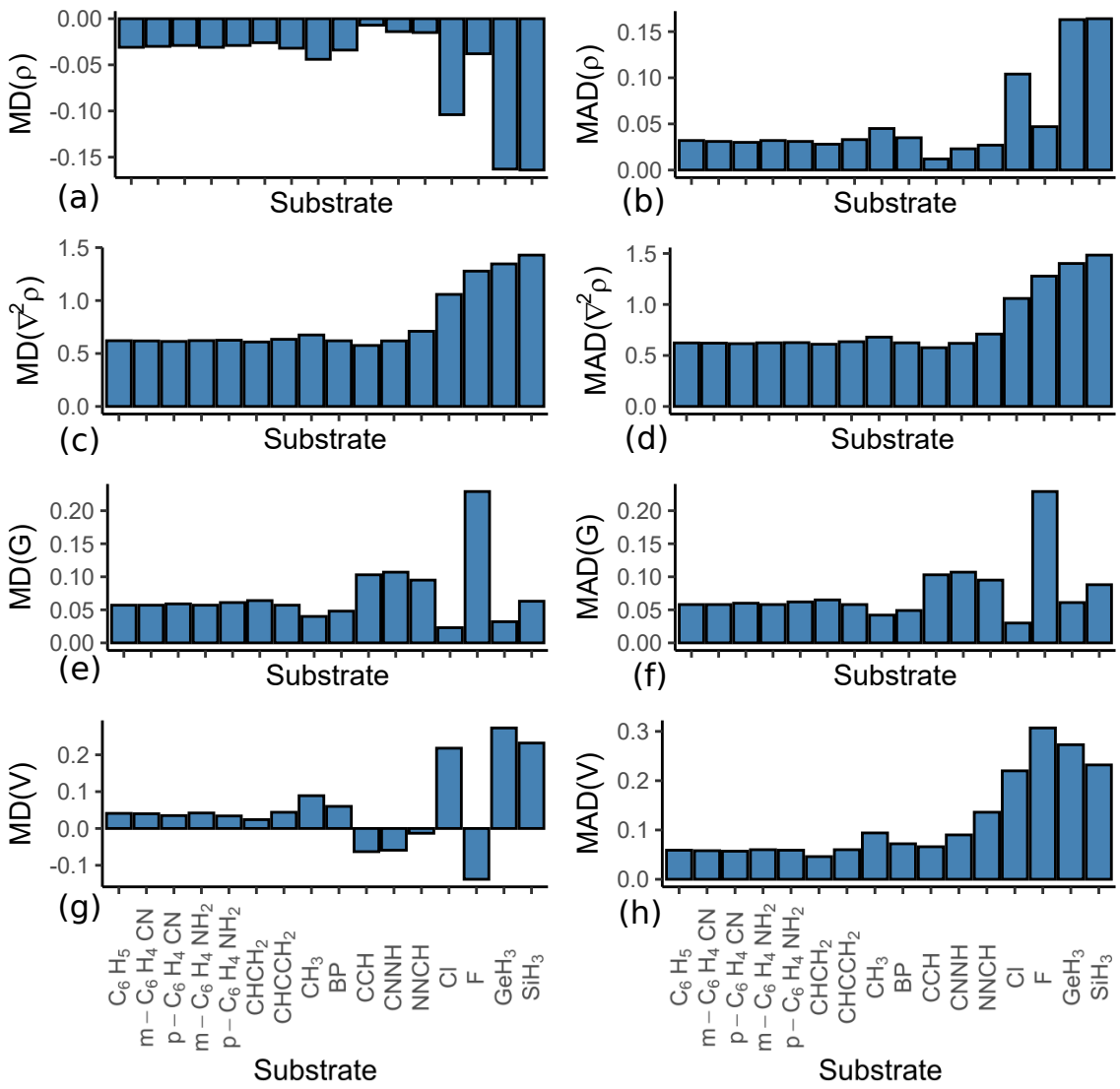

Figure 5.3: Showing the (a) MD(ρ), (b) MAD(ρ), (c) MD($\nabla^2\rho$), (d) MAD($\nabla^2\rho$), (e) MD(G), (f) MAD(G), (g) MD(V), (h) MAD(V) across the set of substrates in the study. Low, constant, MD and MAD of a property would indicate transferability of that property.

evaluated at the R-H BCP. Figure 5.4 shows these plots. The changes in range are clearly presented, with G and V having much larger ranges for other substrates than observed at the R-H BCP. In all cases, slopes deviate from 1. ρ displays the most linear behaviour of any of the properties, but there are still some large deviations present. ρ(G) shifts vertically as G changes, hinting at an underlying pattern in the changes in BCP properties. For machine learning models to use BCP properties, a way to parameterize the observed changes is desired. This could then

be incorporated allowing BCP properties to be used as feature sets in machine learning. In spite of the non-transferability, this is worth accomplishing as BCP properties contain a wealth of information on bonding interactions.

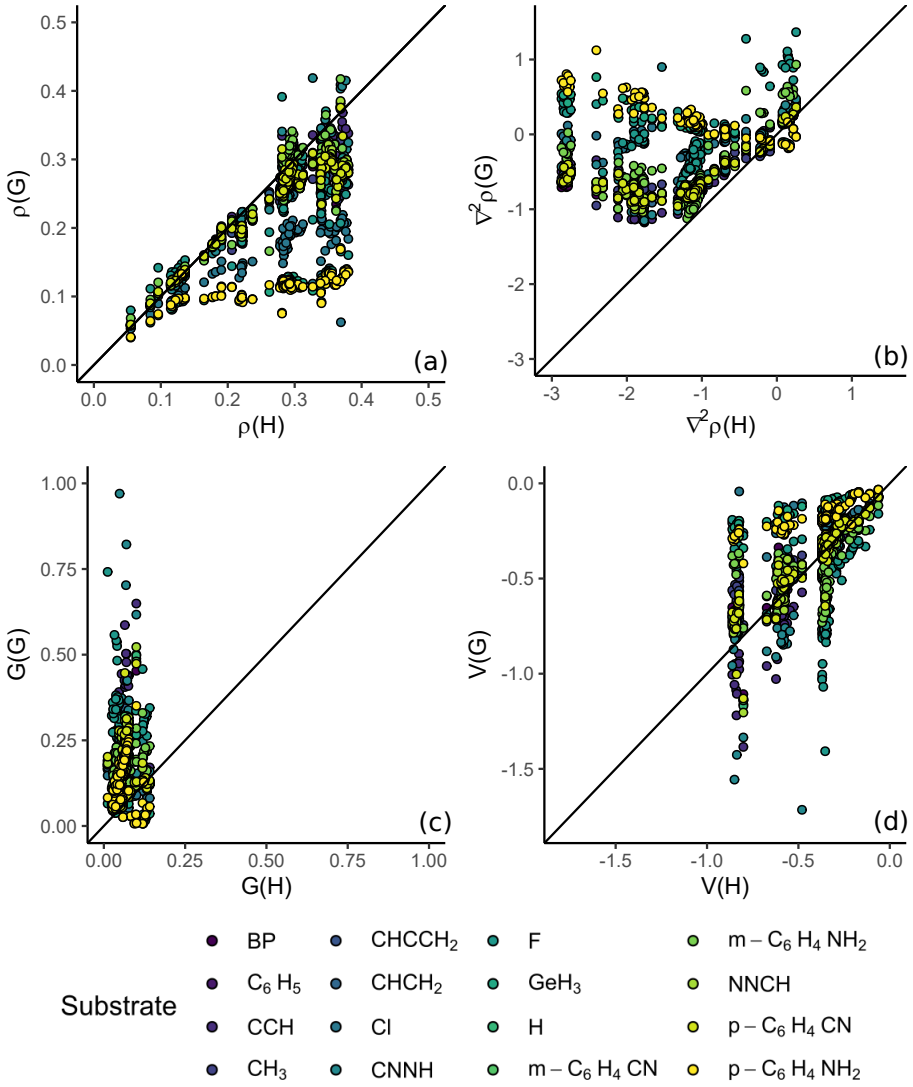

Figure 5.4: BCP properties at the R-G BCP plotted as compared to those in the R-H BCP. (a) ρ, (b) $\nabla^2\rho$, (c) V, and (d) G. The BCP properties would be transferable if the points for each substrate laid exactly on the $y = x$ line

5.4.2 Generalized Electronegativity Scale

One way of parameterizing the changing interactions lies in the concept of electronegativity. Boyd and Edgecombe developed a functional

group electronegativity scale for substituents, R, in RH molecules. As the interaction between a substituent and the group to which it is attached changes from R-H, its effective electronegativity may differ. Here, we recalculated Boyd and Edgecombe's a and b for 15 substrates. This requires only the optimization of G-Li and G-F for all G of interest. The substrate $CHCCH_2$ from our prior studies is omitted due to the irregular nature of the optimized structure, shown in Figure C11. The a and b for each substrate are shown in Appendix C, Figure C12.

Definite changes in the electronegativity are observed. Figure 5.5 shows the MD and MAD for χ_R(G) related to χ_R(H) for all 15 substrates. MD are positive , showing that electronegativity increases in substrates other than hydrogen. The MD and MAD are similar in magnitude, illustrating a consistent increase in values. For G=Cl or F, there is a small difference between MD and MAD, indicating χ_R(Cl) and χ_R(F) are not consistently larger than χ_R(H).

Linear plots also illustrate the change in these values. Figure 5.6 shows linear plots comparing χ_R(G) to χ_R(H). The slopes are greater than 1, consistent with the increase of electronegativity revealed by the MD. $\chi(R_G)$ for carbon attached substrates have strong linear relationships with χ_R(H), revealing that they contain similar information. However, points deviate from the regression line at high χ, indicating an effect on χ that is not only taken into account by a linear relationship. χ_R(CCH) and χ_R(CNNH) see more deviations from linearity than χ_R(CH_3,C_6H_5,$CHCH_2$,BP), due to interactions between the substituents and this substrate. Non-carbon attached substrates do not linearly relate to χ_R(H). Here, the electronegativities change drastically. The points with low χ for substrates with varying electronegativity are relatively linear. Only above $\chi = 2$ do we see most deviations from linearity. G=Cl and F have the highest deviations at high χ values. GeH_3 and SiH_3

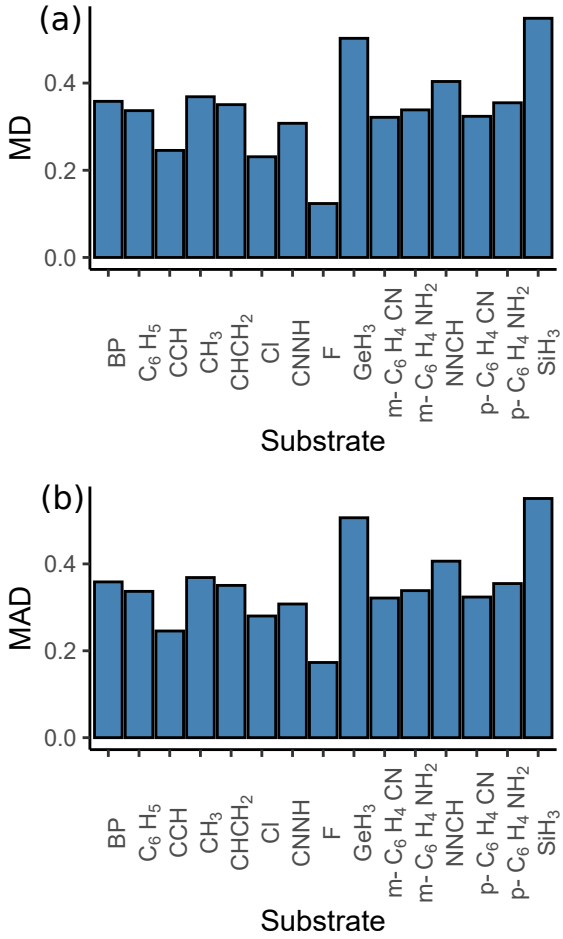

Figure 5.5: Plots showing the (a) MD and (b) MAD of substituent electronegativities in different substrates as compared to G=H

differ from linearity largely above $\chi = 2$. This changing behaviour stems from the character of the substrates, GeH_3 and SiH_3 are electropositive, while F and Cl are electronegative. Deviations are high for Cl and F at high χ because, as those are both electronegative, they compete with the substituent for the electrons.

χ_R(G) descriptors capture a changing interaction between the substituent and the substrate leading to changing substituent electronegativity. Theoretically, it should also be possible to study a substrate's changing electronegativity as the substituent to which it is attached changes. However, this would require optimization of Li-R molecules for all substituents, but the lithium atom interacts with substituents strongly, changing bonding patterns and preventing the determination

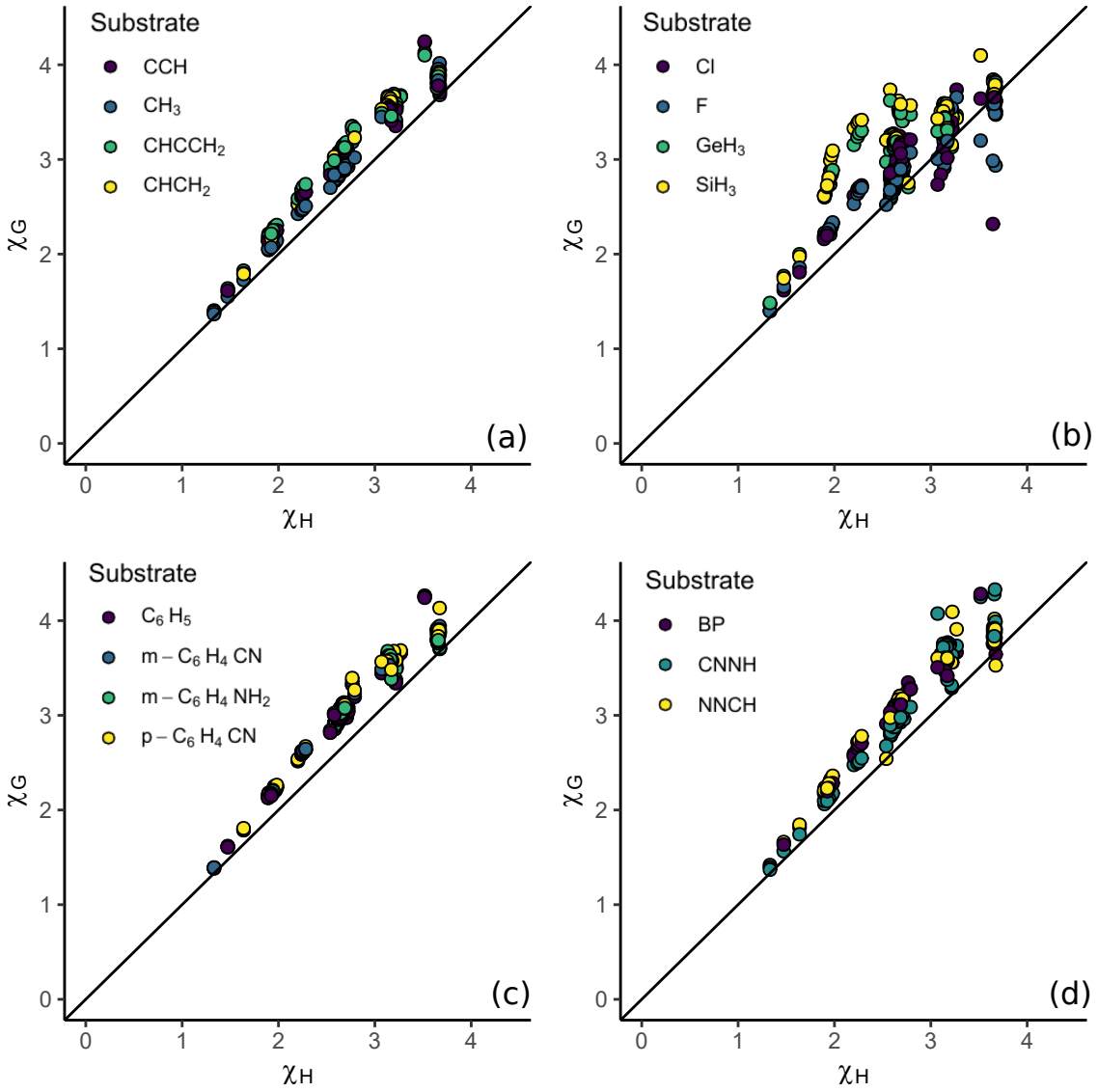

Figure 5.6: Electronegativity of substituents in different substrates (a) CH_3, $CHCH_2$,CCH, andC_6H_5; (b) F, Cl, GeH_3, and SiH_3; (c) m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; and (d) BP, CNNH, and NNCH against the electronegativities in hydrogen. The black line highlights the ideal $y = x$ line.

of optimized Li-R structures needed to construct χ_G(R). The changing interactions captured by the electronegativity are promising for use in analyzing patterns in BCP property changes.

5.4.3 Parameterizing the Error in ρ_c

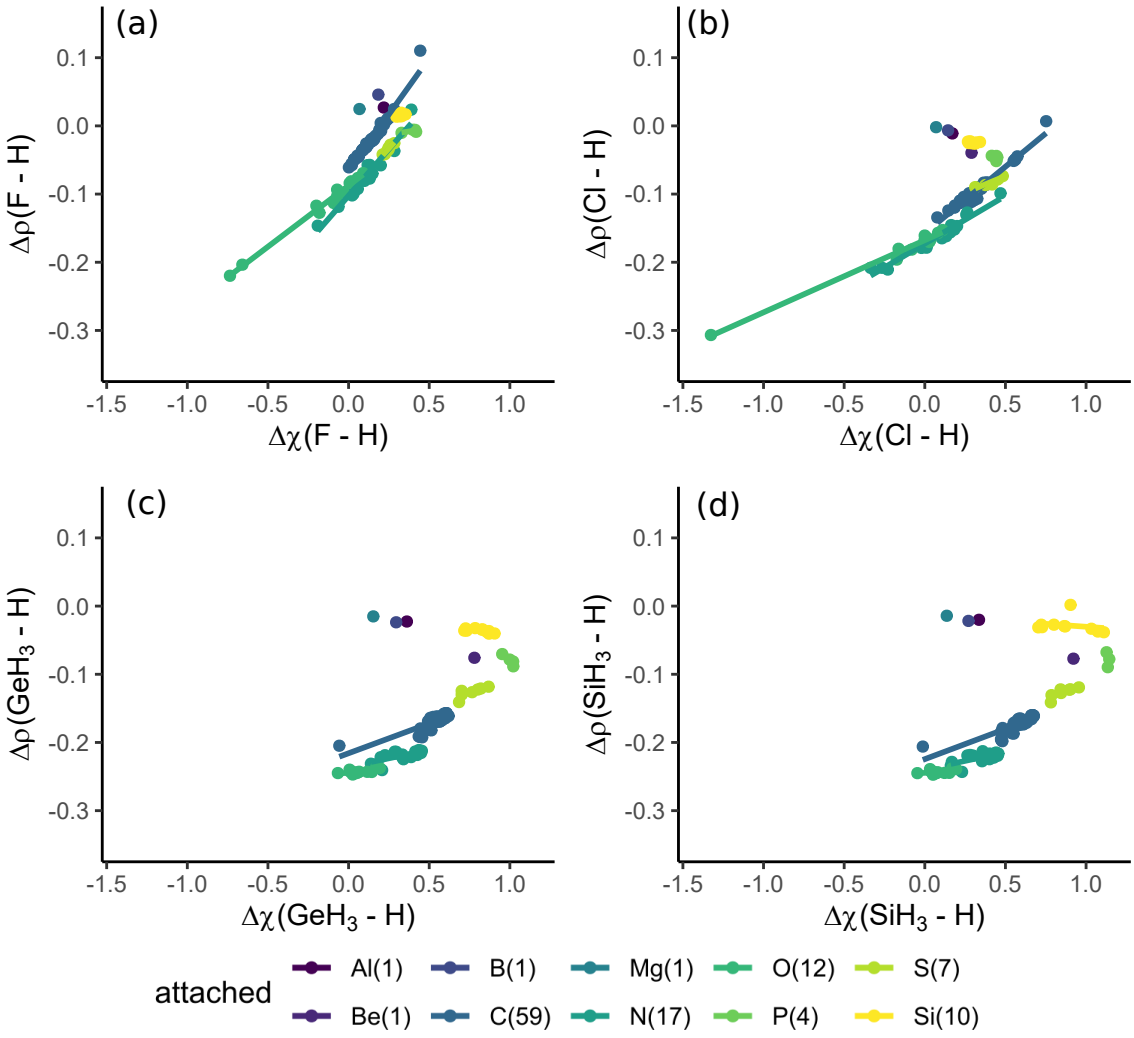

Figure 5.7: Figure showing $\Delta\rho(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = F, (b) G=Cl, (c) G=SiH_3, (d) G = GeH_3

We define a parameter to capture the changing electronegativity, $\Delta\chi$(G-H), Equation 5.6, to provide information on a substituent's changing electronegativity between two substrates. Similarly, we calculate the change in a BCP property, Equation 5.7. Ideally, these would linearly relate to each other. This is not observed over all substituents, but when substituents are separated based on type of substrate connected atom (the definition of $F_{\mathrm{R}}(G)$ contains the N_A term in the denominator, so this separation is intuitive), interesting relations occur.

$$\Delta\chi(G-H) = \chi_{\mathrm{R}}(\mathrm{G}) - \chi_{\mathrm{R}}(\mathrm{H}) \tag{5.6}$$

$$\Delta\mathrm{P}(G-H) = \mathrm{P}_{\mathrm{R}}(\mathrm{G}) - \mathrm{P}_{\mathrm{R}}(\mathrm{H}) \tag{5.7}$$

Figure 5.7 shows the relationship between $\Delta\rho$ and $\Delta\chi$ for one series of substrates, G=Cl, F, SiH_3, and GeH_3. For the F and Cl substrates, strong linear relationships are present for all substituent types. A few singular points correspond to substituents for which there are no others in the set (Al, B, Be, Mg), and thus no linear relationships can be formed. For GeH_3 and SiH_3, graphically, linear patterns are observed. However, these are not accompanied by good r^2 values. For most substituent types for these substrates, r^2 are less than 0.4. N and O attached substituents have nearly constant $\Delta\rho$, however. A constant error based on the atom attached to the substrate is still viable for inclusion in machine learning methods. Since there error is constant however, the $\Delta\chi$ parameter is not needed, and poor r^2 are seen. These substrates all have differing electronegativities from each other and hydrogen, but the patterns are seen in the changes leading to being viable for inclusion in machine learning.

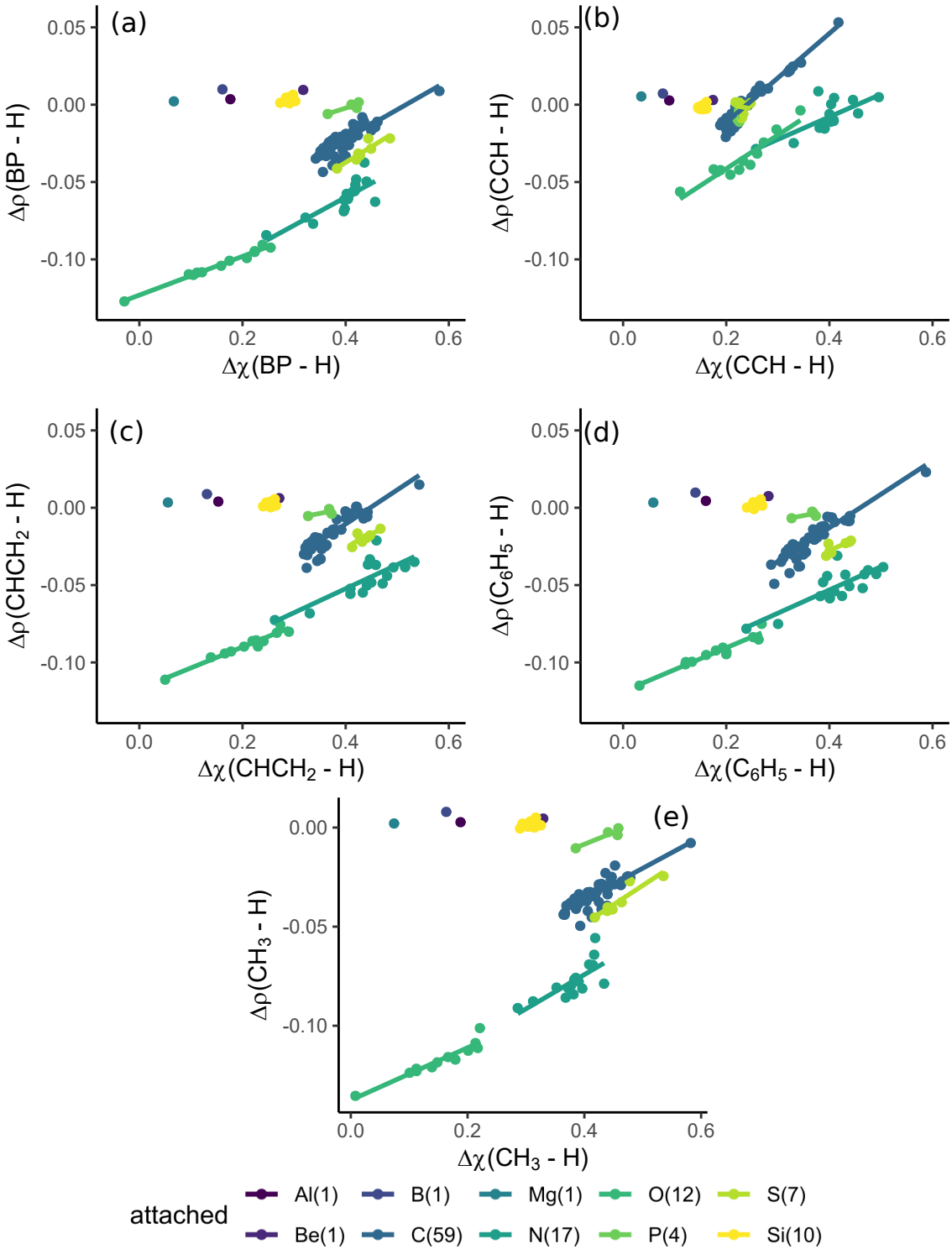

Figure 5.8: Figure showing $\Delta\rho(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = BP, (b) G=CCH, (c) G=$CHCH_2$, (d) G = C_6H_5

Various carbon based substrates also have observable changes in ρ explainable by $\Delta\chi$. Figure 5.8 shows plots for carbon substrates of varying hybridisation. The linear relationship between $\Delta\rho$ and $\Delta\chi$ describes the changes. For the various aromatic substrates, Figure 5.9 shows these changes. The r^2 for these relations is small, despite graphically being well depicted. This is due to the definition of r^2, Equation 5.8. Here, $\hat{y}$ refers to the predicted values of y by the linear relation, and $\bar{y}$ is the mean y value. r^2 depends on the mean y value, and in this data set, the mean y value is skewed. Reduced r^2 are seen. Regardless of quantitative reproduction, the values are qualitatively reproduced well in all cases.

$$r^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{5.8}$$

$\Delta\chi$ even reflects changes in $\Delta\rho$ for substrates who may change electronic character with substitution. Nitrilimine, which can have the substituent attached at either the carbon (CNNH) or nitrogen (NNCH) is one such substrate.[35] Figure 5.10 shows these linear relationships. The relationships exhibit high quality. For NNCH, there is one large outlier skewing the data for the nitrogen substituents, but otherwise all substituents are well explained by linear relationships. With these malleable substrates, the success of this parameter is due to mutual change in the substituent-substrate interaction. The substrate's electronitc structure changes, inducing a change in the substituent-substrate bond, and thus the substituent's effective electronegativity.

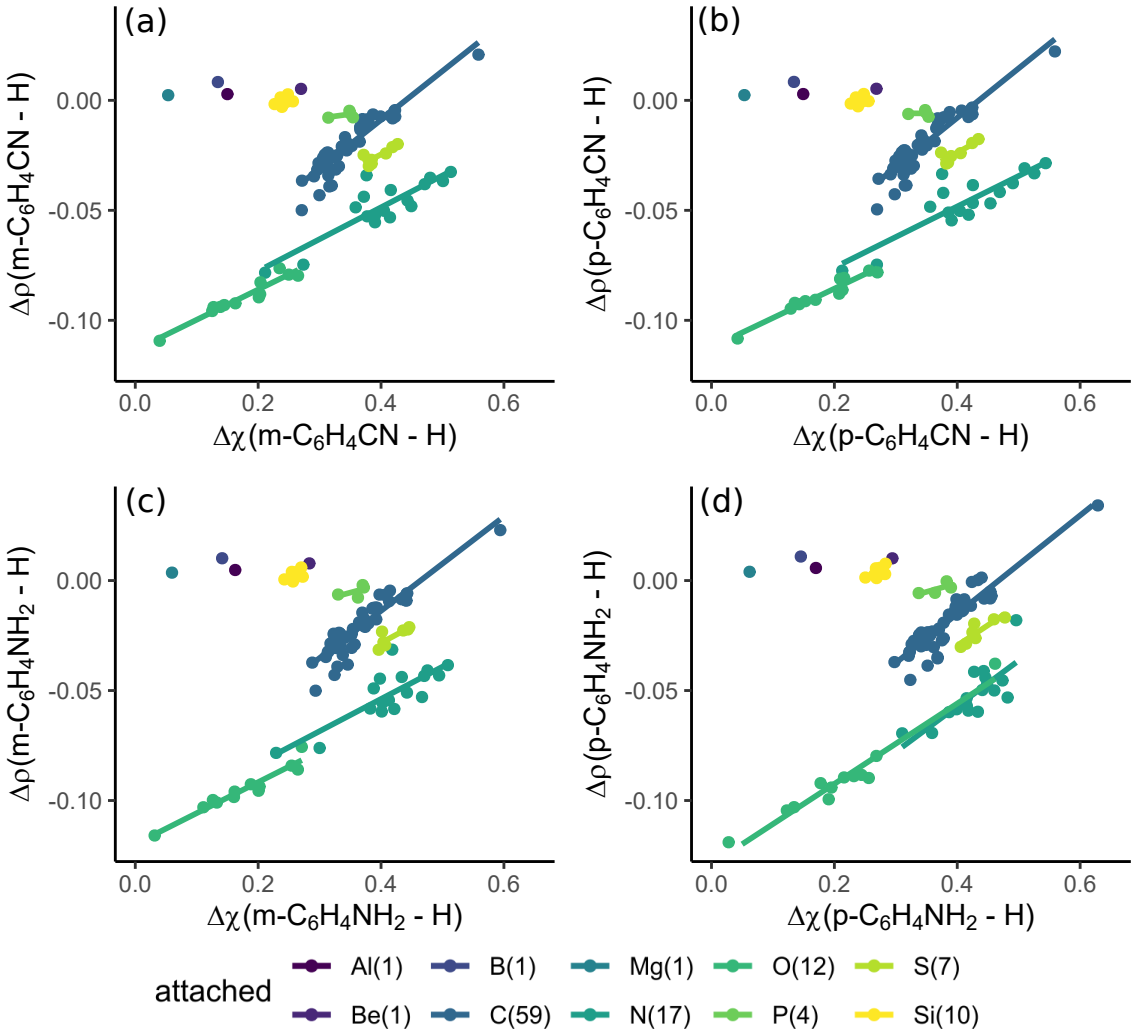

Figure 5.9: Figure showing $\Delta\rho(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = m$-C_6H_4CN$, (b) G=p$-C_6H_4CN$, (c) G=m$-C_6H_4NH_2$, (d) G = p$-C_6H_4NH_2$

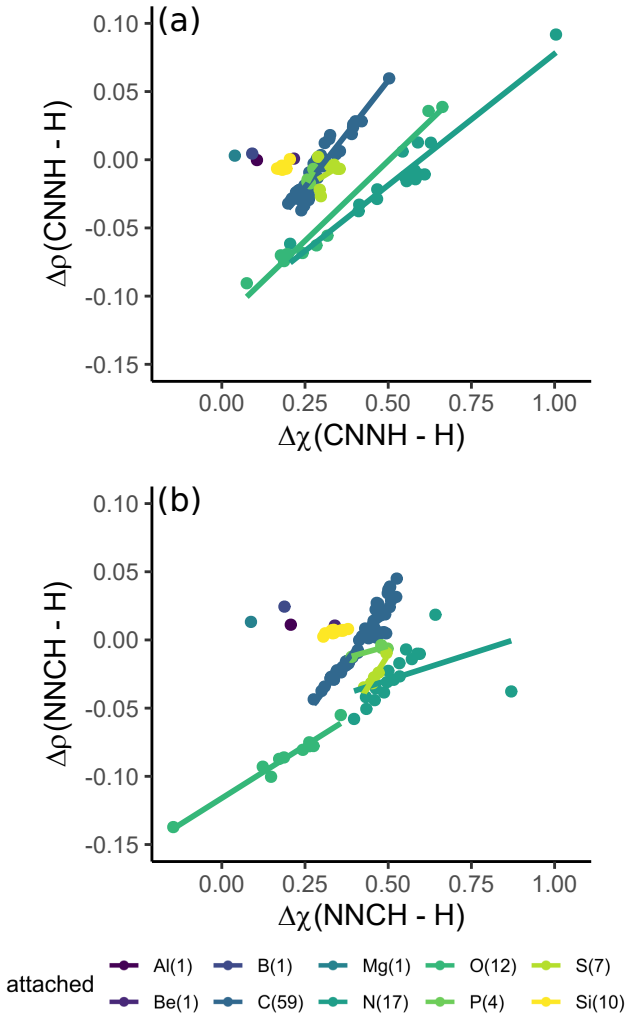

Figure 5.10: Figure showing $\Delta\rho(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = CNNH, (b) G=NNCH

5.4.4 Parameterizing the Error in $\nabla^2\rho_c$

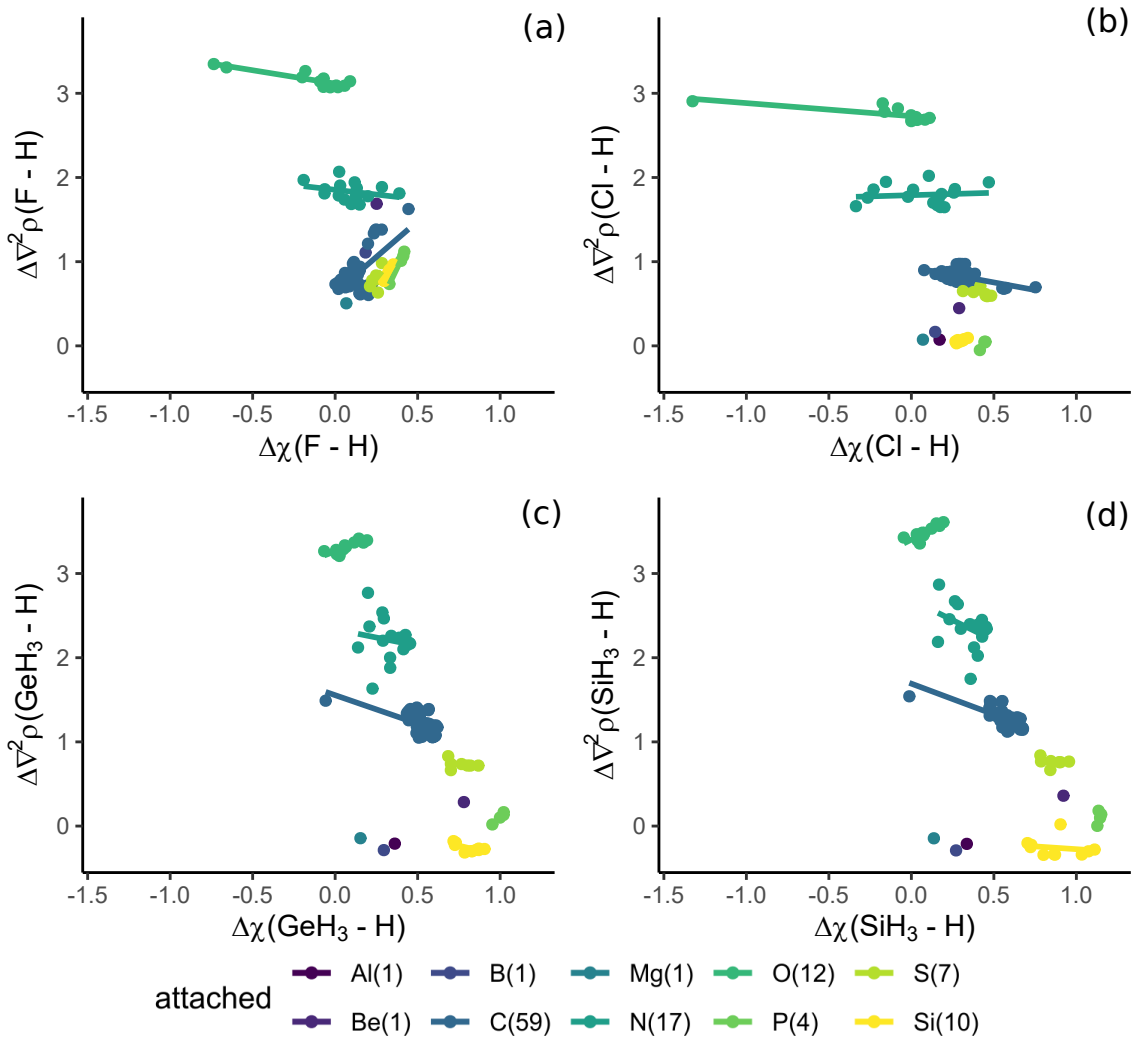

Figure 5.11: Figure showing $\Delta\nabla^2\rho(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = F, (b) G=Cl, (c) G=SiH_3, (d) G = GeH_3

The BCP density itself is only a first test case, as, naturally, χ is calculated based on it. Further properties evaluated at the bond critical point prove the true test for $\Delta\chi$ in describing observed changes in the properties. Here, we examine the Laplacian, $\nabla^2\rho$. Figure 5.11 demonstrates the error, $\Delta\nabla^2\rho$, which is nearly constant for a given atom type. This itself is useful to know in parameterizing a machine learning model, as a constant error can be used as well as a parameterized error. The limits of $\Delta\chi$ are seen in this figure however. When $\Delta\nabla^2\rho$ does change, the regression line doesn't capture it well for substituents attached through

nitrogen atoms. Nitrogen substituents are one of the few functional group types for which $\Delta\nabla^2\rho$ changes.

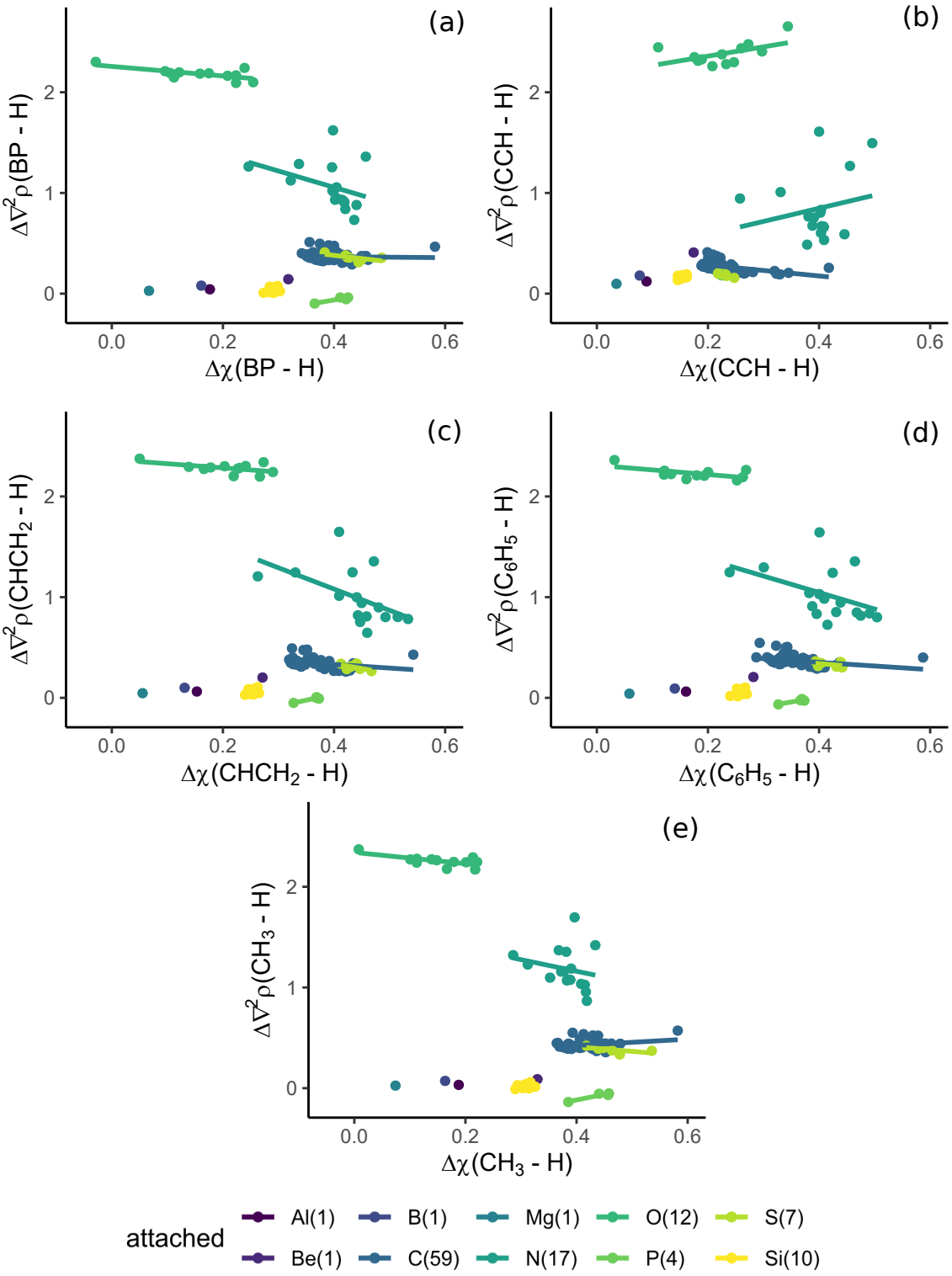

Figure 5.12: Figure showing $\Delta\nabla^2\rho(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = BP, (b) G=CCH, (c) G=CHCH$_2$, (d) G = C$_6$H$_5$

Changing the hybridisation of the carbon reveals more information on the parameterization. Figure 5.12 shows the plot of $\Delta\nabla^2\rho$ for the carbon based substrates. For most atom types, near constant errors are seen. However, nitrogen bonded substituents typically see non-constant error not described by $\Delta\chi$. This could be due to $\Delta\chi$ not capturing resonance effects, as it is only an electronegativity parameter. An alternative is that a wide variety of nitrogen containing substituents (sp^2, sp^3) are included in the data set. It may be hard to simultaneously capture sp^2 nitrogens as well as sp^3. In Chapter 4, changes in resonance effects in substituents were interpretable with geometric changes (for example the planarization of amino nitrogen atoms).

As expected based on the above results, the changes in the R-G bond for varying aromatic substitutions show low change in $\Delta\nabla^2\rho$, Figure 5.13. $\Delta\chi$ does not parameterize low $\Delta\nabla^2\rho$ well due to the constancy. Nitrogen-attached substituents still see scatter. The parameterization is likely unnecessary in this case, so this is not a limiting situation for $\Delta\chi$.

Except for nitrogen bonded substituents in CNNH, $\Delta\chi$ represents $\Delta\nabla^2\rho$ well in the nitrilimine substrates, Figure 5.14. For those outlying cases there is the potential for large interactions with CNNH via resonance, skewing their results. These outliers also highlight the possibility that $\Delta\chi$ does not explain the resonance effect in these systems. Overall, changes in $\nabla^2\rho$ are adequately described by $\Delta\chi$, except for possible resonance interactions.

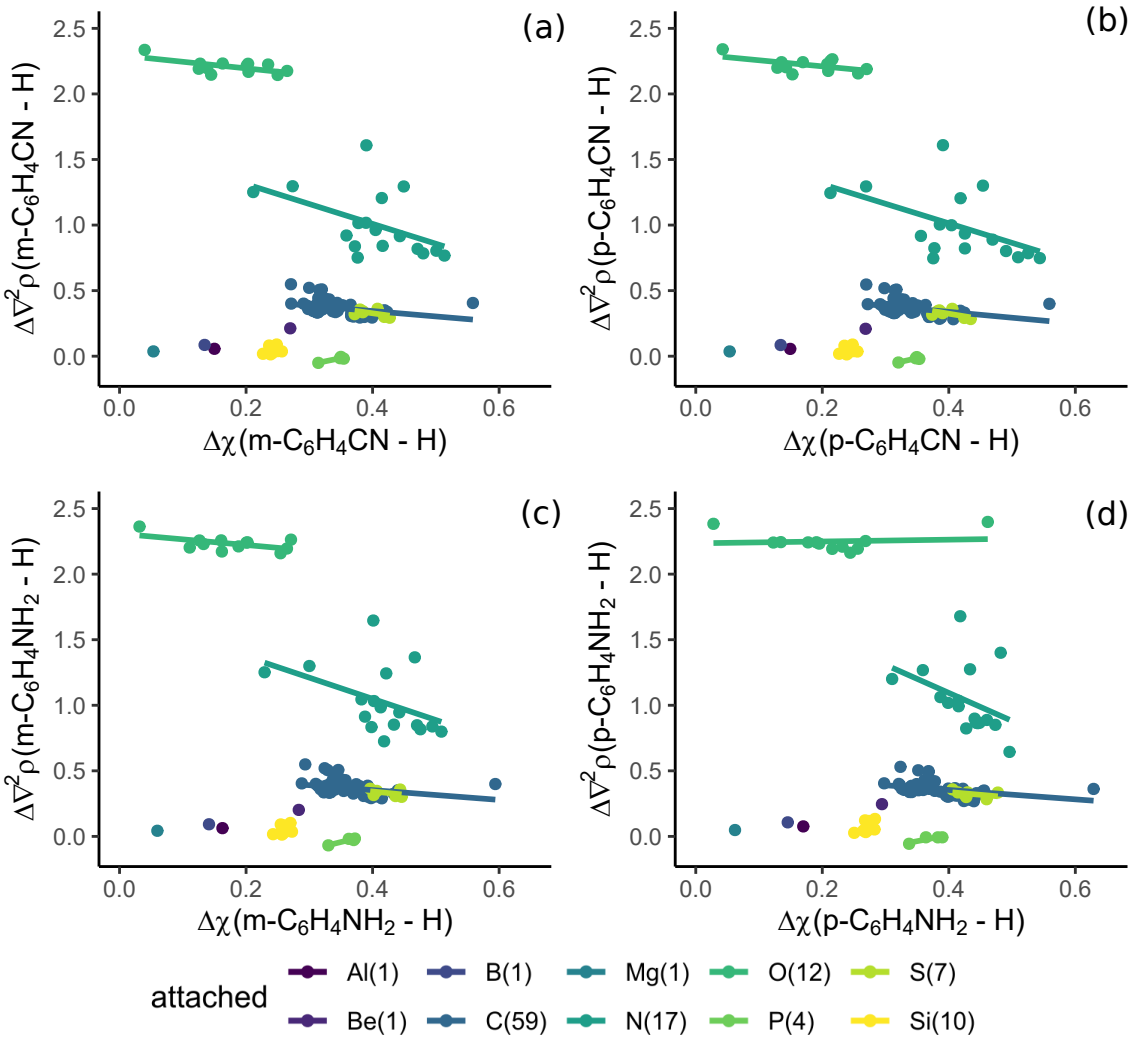

Figure 5.13: Figure showing $\Delta\nabla^2\rho(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = m$-C_6H_4CN$, (b) G=p$-C_6H_4CN$, (c) G=m$-C_6H_4NH_2$, (d) G = p$-C_6H_4NH_2$

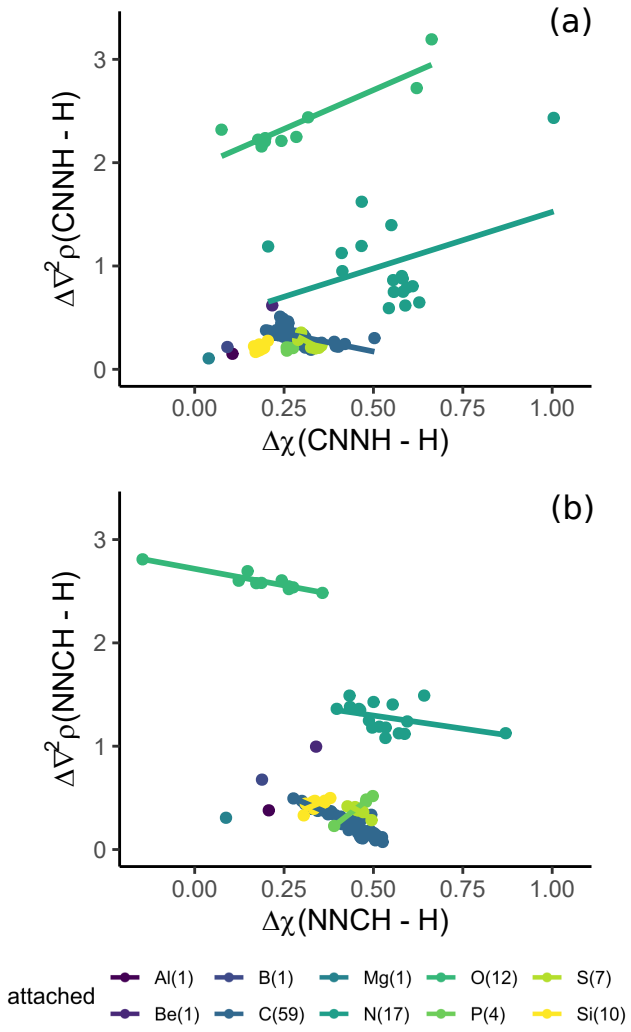

Figure 5.14: Figure showing $\Delta\nabla^2\rho(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = CNNH, (b) G=NNCH

5.4.5 Parameterizing the Error in Energy Density Properties

Another scalar property at bond critical points often used to interpret the covalency and ionicity of the bonding interactions is the energy densities. Figure 5.15 and 5.16 show the parameterization for substrates varying electronegativity. Both ΔV and ΔG are well described by $\Delta\chi$. Carbon attached substituents observe excellent linearity except for in GeH_3 and SiH_3, where they tend to cluster together. All other attached atoms see excellent linearity.

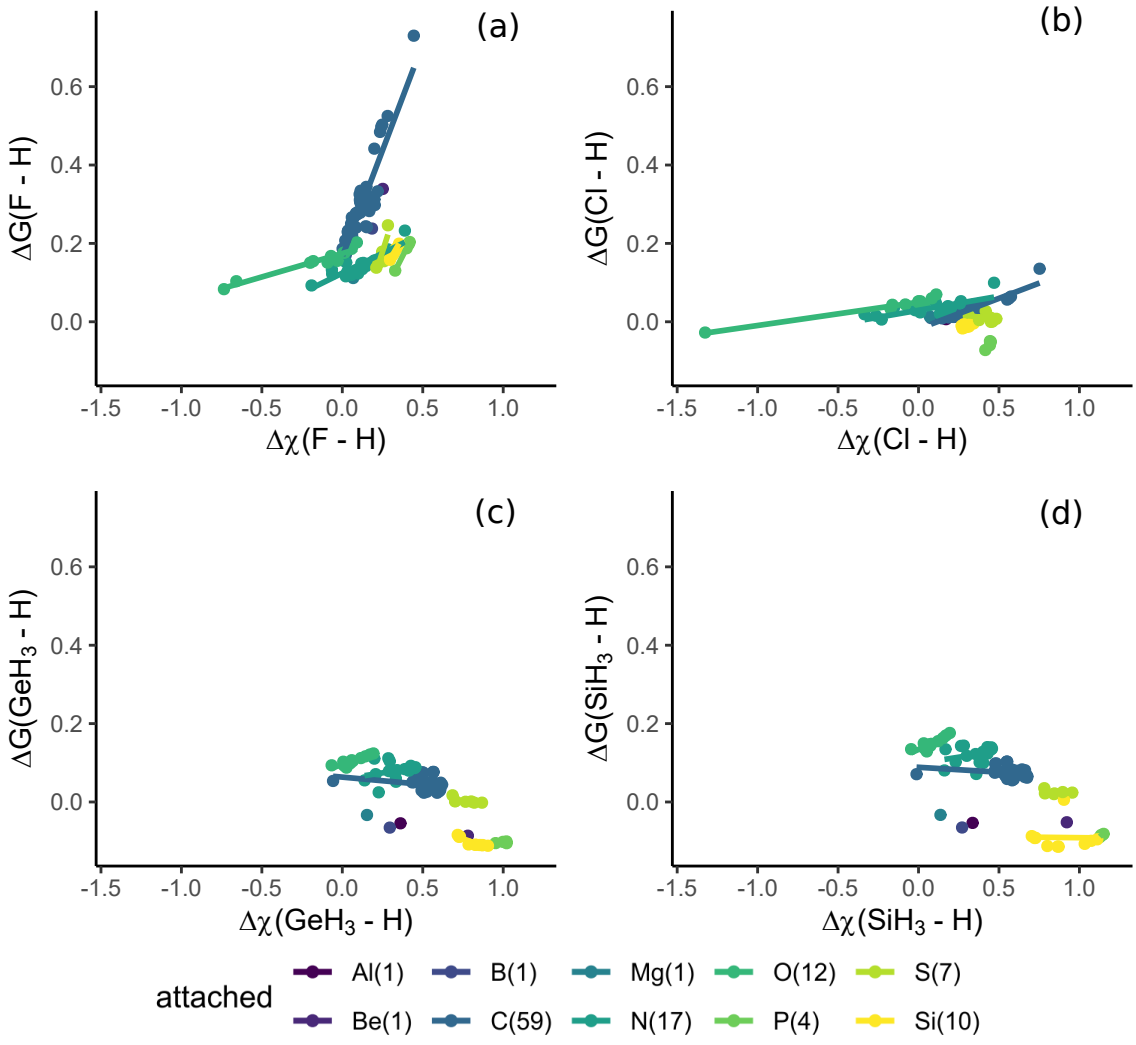

Figure 5.15: Figure showing $\Delta G(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = F, (b) G=Cl, (c) G=SiH_3, (d) G = GeH_3

Figures 5.17 and 5.18 show ΔG while varying the hybridisation of

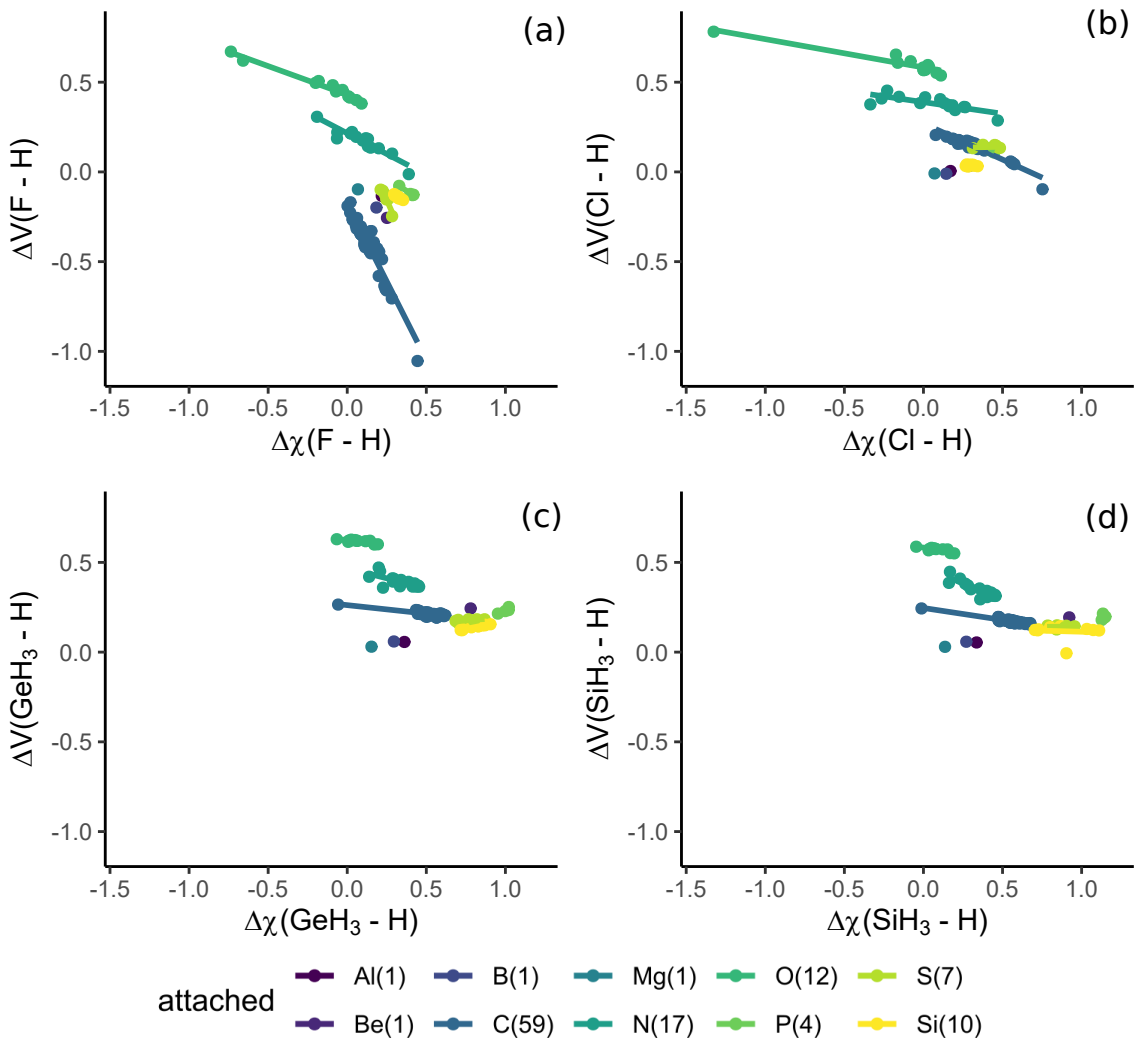

Figure 5.16: Figure showing $\Delta V(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = F, (b) G=Cl, (c) G=SiH_3, (d) G = GeH_3

the carbon atom attached. Oxygen and carbon-attached substituents are generally well described, with exception of a few outliers. Outliers are present in the parameterization of nitrogen atoms. ΔV also linearly relates well in almost all cases, excepting nitrogen outliers for ΔV(CCH - H). r^2 values are good. The changes in V and G for the aromatic series mimic the above noted changes, Figures 5.19 and 5.19. That is, regressions are good, excepting a small number of outliers.

For many atom types, the potential energy differences in nitrilimines are well parameterized. For the kinetic energy density and potential energy density, in G=CNNH, all atom types are described well. A wide

scatter is seen for carbon atoms, but visually, the paramaterization is acceptable in quality. This highlights that the changes in the potential energy densities in NNCH are well described by $\Delta\chi$.

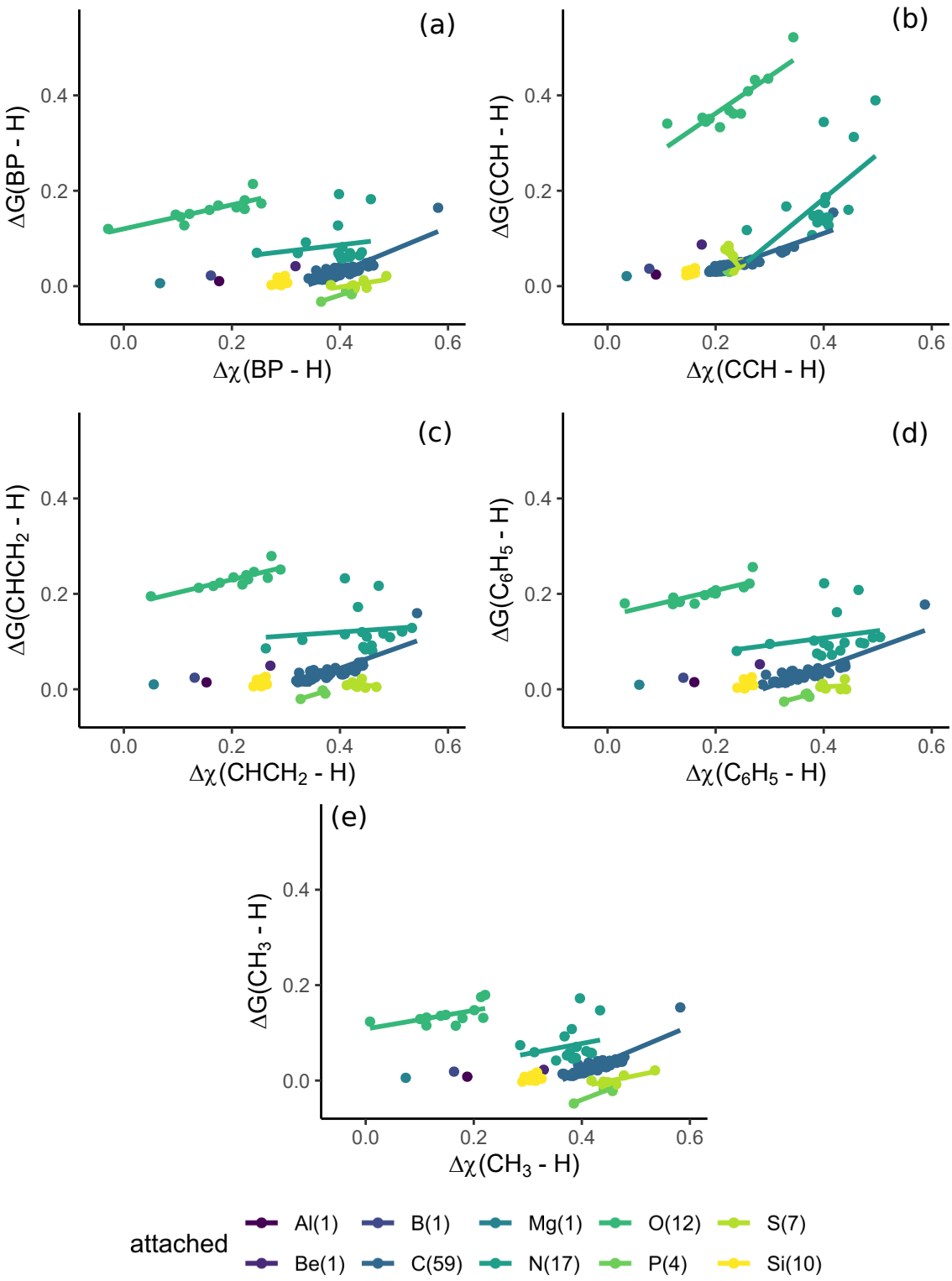

Figure 5.17: Figure showing $\Delta G(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = BP, (b) G=CCH, (c) G=$CHCH_2$, (d) G = C_6H_5

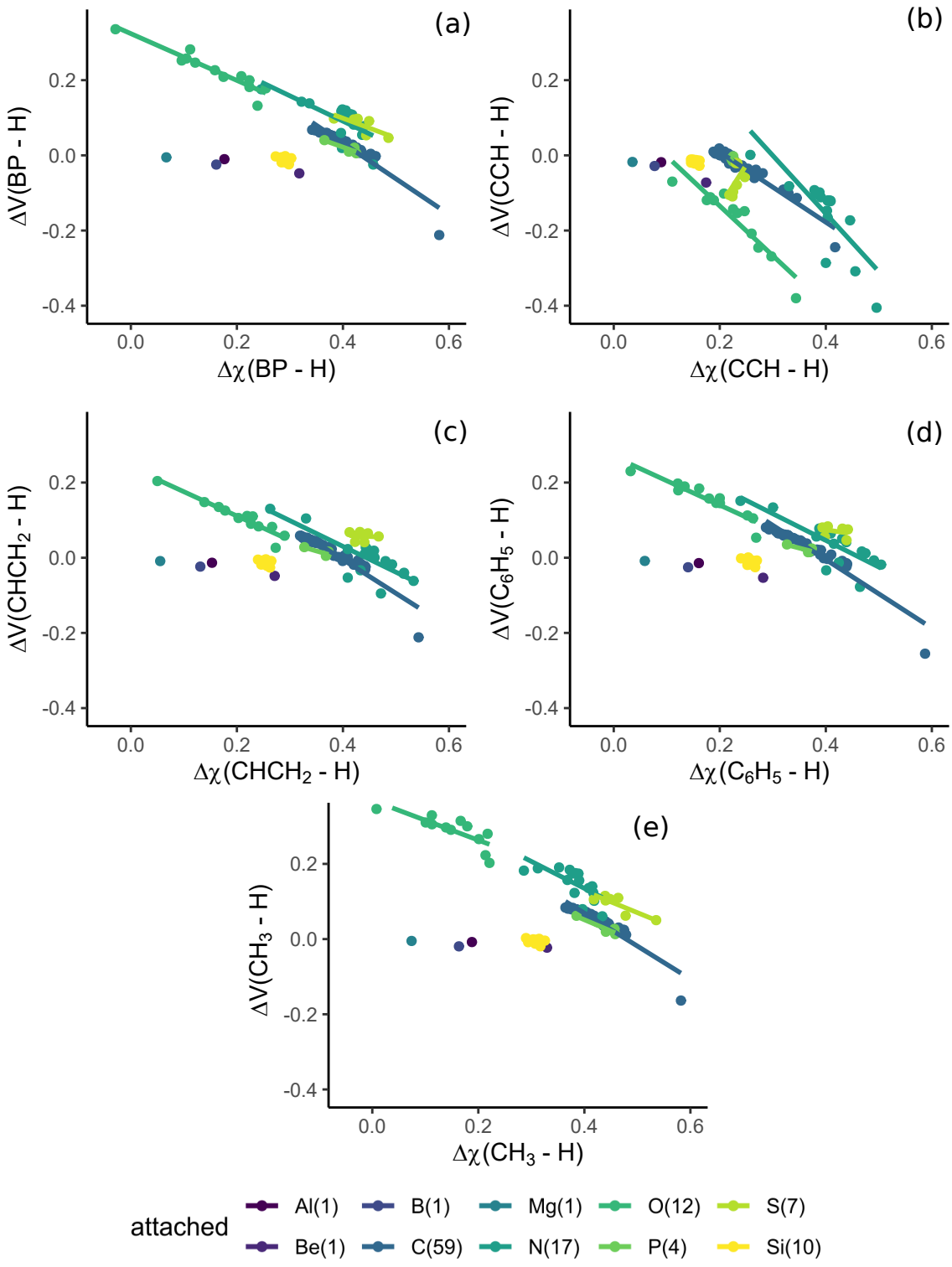

Figure 5.18: Figure showing $\Delta V(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = BP, (b) G=CCH, (c) G=$CHCH_2$, (d) G = C_6H_5

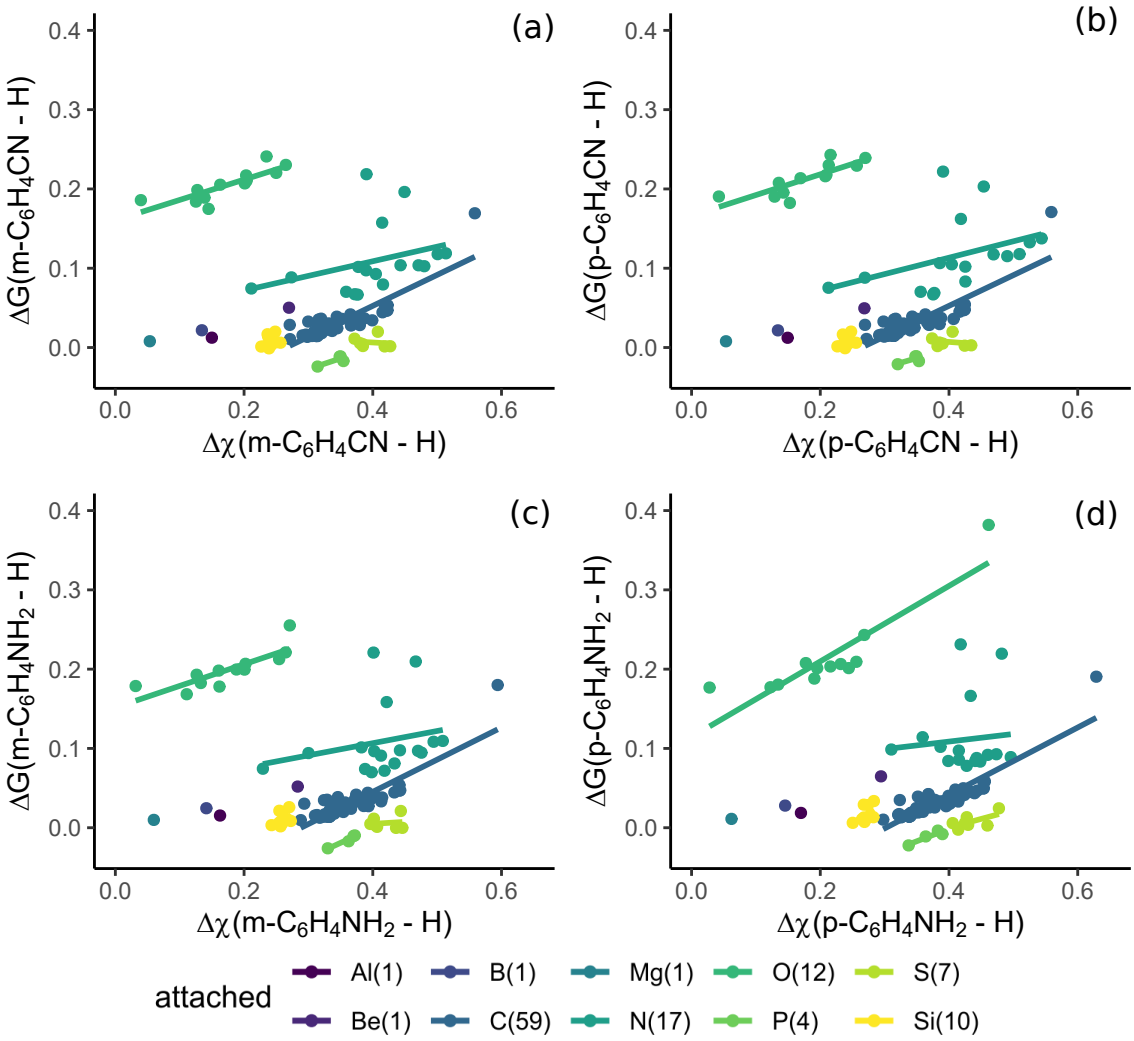

Figure 5.19: Figure showing $\Delta G(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = m$-C_6H_4CN$, (b) G=p$-C_6H_4CN$, (c) G=m$-C_6H_4NH_2$, (d) G = p$-C_6H_4NH_2$

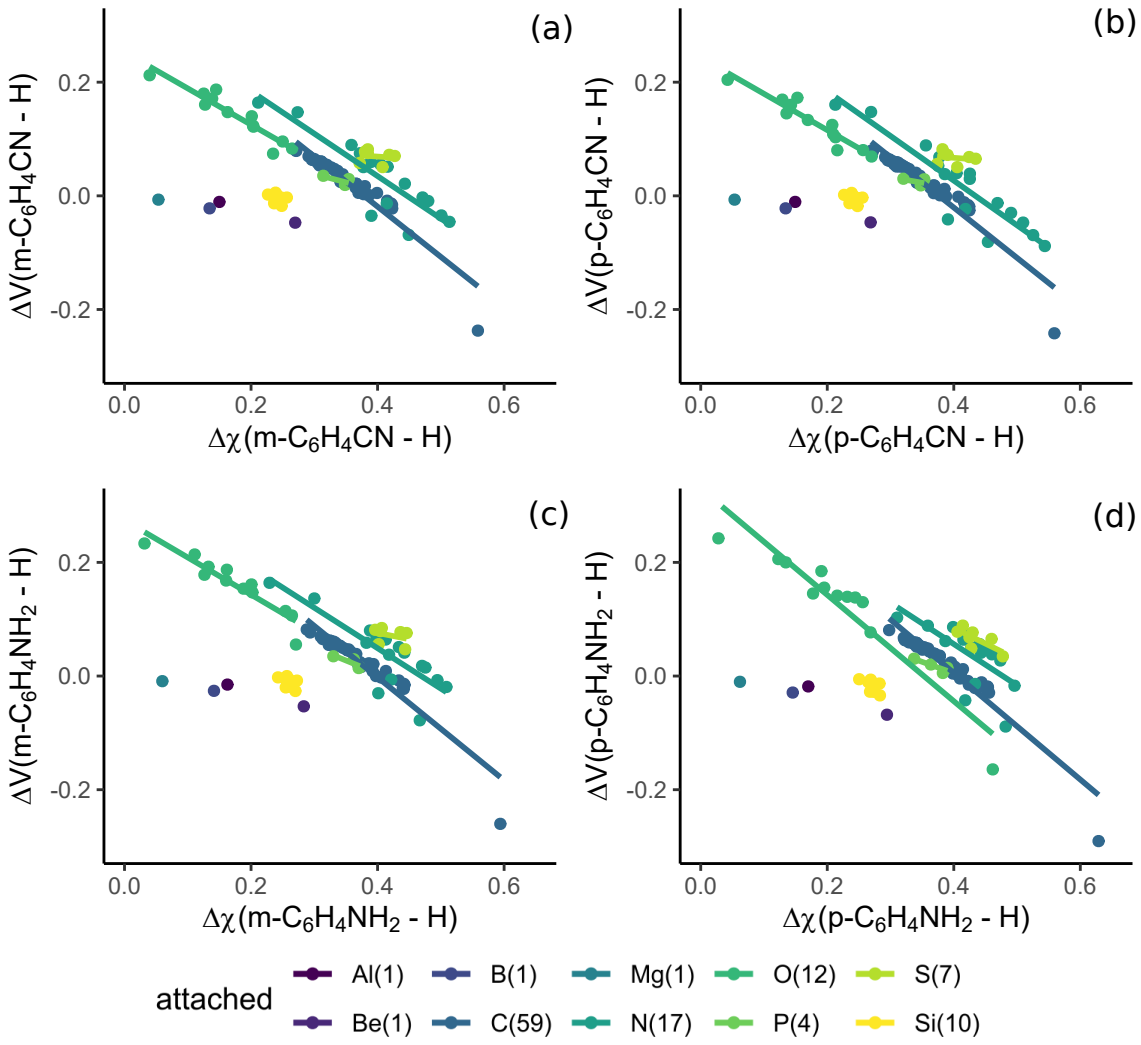

Figure 5.20: Figure showing $\Delta V(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = m$-C_6H_4CN$, (b) G=p$-C_6H_4CN$, (c) G=m$-C_6H_4NH_2$, (d) G = p$-C_6H_4NH_2$

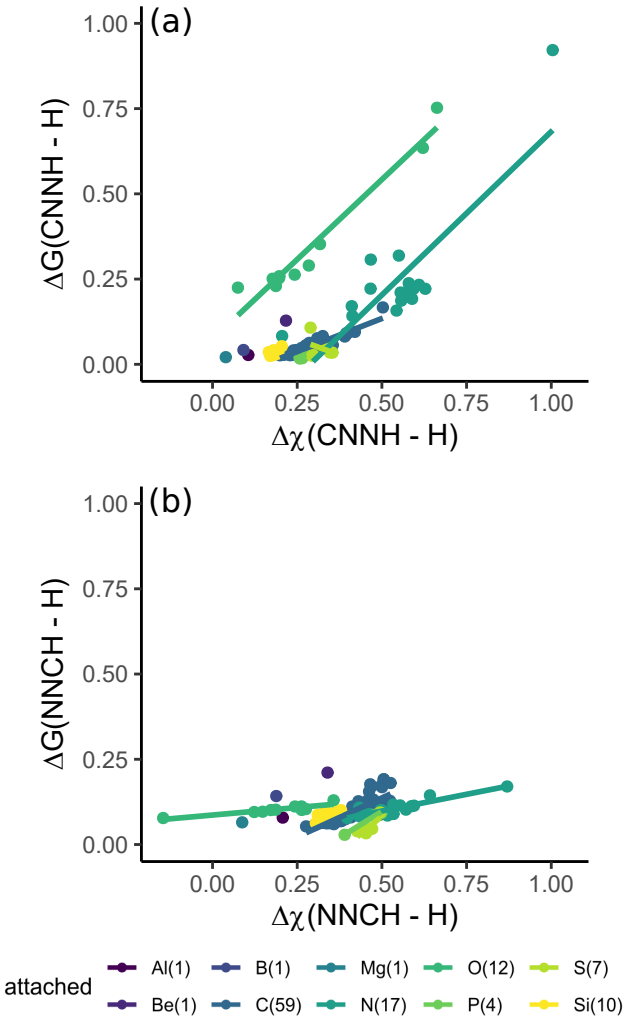

Figure 5.21: Figure showing $\Delta G(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = CNNH, (b) G=NNCH

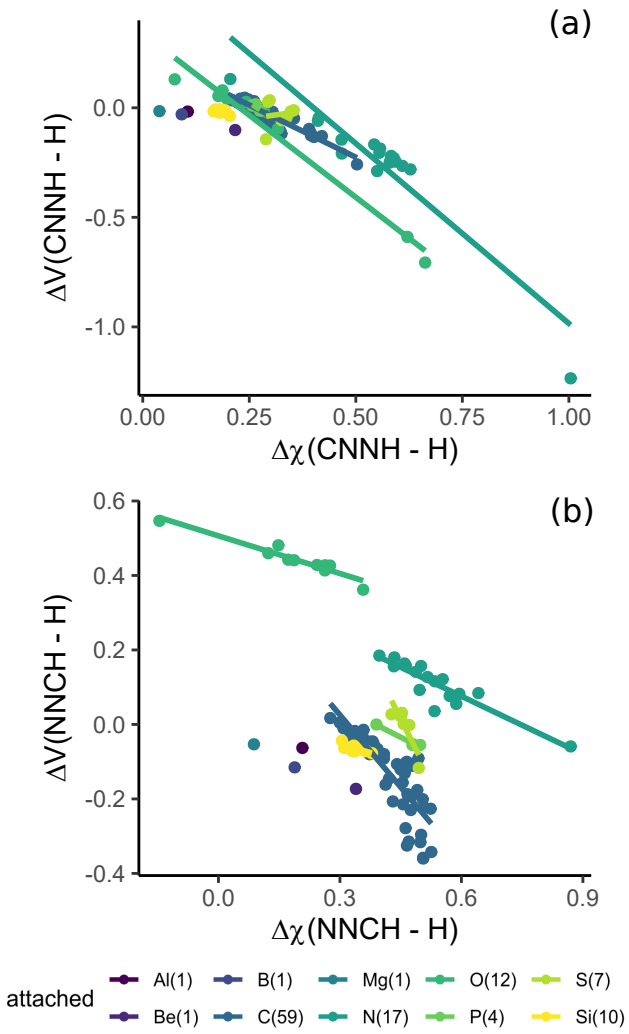

Figure 5.22: Figure showing $\Delta V(G-H)$ compared to $\Delta\chi$(G - H) for the electronegativity series, (a) G = CNNH, (b) G=NNCH

5.5 Conclusions

The contributions of this paper are twofold. First, a generalized group electronegativity model was presented. While it is related to Boyd and Edgecombe's prior electronegativity scale, it captures information regarding the changing interaction of substituents as they are transferred between substrates. An easy method is presented to calculate the substituent electronegativity in any substrate.

Secondly, we establish the non-transferability of BCP properties of R-H molecules to other substrates. Despite this non-transferability, a parameter, $\Delta\chi$ based on the electronegativity scale above describes observed changes. Changes in $\Delta\chi$ between substrates linearly relate to changes in BCP properties between substrates. This is ideal for implementation in machine learning, as such a parameter can easily be incorporated alongside the BCP properties. There are two cases requiring caution for using $\Delta\chi$ to describe changes in BCP properties. Firstly, evidence suggests that $\Delta\chi$ does not capture changes in the substituent related to resonance effects. Secondly, in cases where a property has a constant error between G and H for a given substituent type, $\Delta\chi$(G - H) is not necessary to parameterize the error, and numerical errors may be prevalent if parameterization were to be attempted. Future work will focus on incorporating transferable group properties and parameterized changing BCP properties into a machine learning model.

5.5.1 Acknowledgements

This research was funded by the Ontario Graduate Scholarship, and the Lakehead University Faculty of Graduate Studies. This research was enabled in part by support provided by the Shared Hierarchical Academic Research Computing Network (www.sharcnet.ca) and Compute Canada

(www.computecanada.ca).

References

[1] Ramakrishnan, R., Dral, P. O., Rupp, M. and von Lilienfeld, O. A., *J. Chem. Theor. Comput.* **2015**, *11*, 2087–2096.

[2] Dral, P. O., *J. Phys. Chem. Lett.* **2020**, *11*, 2336–2347.

[3] Pelzer, K. M., Cheng, L. and Curtiss, L. A., *J. Phys. Chem. C* **2017**, *121*, 237–245.

[4] Soriano-Correa, C., Barrientos-Salcedo, C., Francisco-Márquez, M. and Sainz-Díaz, C. I., *J. Mol. Graph. Model.* **2018**, *81*, 116–124.

[5] Hammett, L. P., *J. Am. Chem. Soc.* **1937**, *59*, 96–103.

[6] Swain, C. G. and Lupton, E. C., *J. Am. Chem. Soc.* **1968**, *90*, 4328–4337.

[7] Takahata, Y. and Chong, D. P., *Int. J. Quantum Chem.* **2005**, *103*, 509–515.

[8] Krygowski, T. M. and Sadlej-Sosnowska, N., *Struct. Chem.* **2011**, *22*, 17–22.

[9] Grob, C. A. and Schlageter, M. G., *Helv. Chim. Acta* **1976**, *59*, 264–276.

[10] Taft, R. W., *J. Phys. Chem.* **1960**, *64*, 1805–1815.

[11] Oziminski, W. P., Krygowski, T. M., Fowler, P. W. and Soncini, A., *Org. Lett.* **2010**, *12*, 4880–4883.

[12] Rode, M. F. and Sobolewski, A. L., *J. Phys. Chem. A* **2010**, *114*, 11879–11889.

[13] Oziminski, W. and Krygowski, T., *Tetrahedron* **2011**, *67*, 6316–6321.

[14] Krygowski, T. M., Oziminski, W. P. and Cyrański, M. K., *J. Mol. Model.* **2012**, *18*, 2453–2460.

[15] Galabov, B., Ilieva, S. and Schaefer, H. F., *J. Org. Chem.* **2006**, *71*, 6382–6387.

[16] Beker, W., Gajewska, E. P., Badowski, T. and Grzybowski, B. A., *Angew. Chem. Int. Ed.* **2019**, *58*, 4515–4519.

[17] Bader, R., Atoms in Molecules: A Quantum Theory, Clarendon Press, Oxford, U.K. **1994**.

[18] Bader, R. F. W. and Matta, C. F., *Found. Chem.* **2013**, *15*, 253–276.

[19] Bader, R. F. W., *J. Chem. Phys.* **1972**, *56*, 3320.

[20] Matta, C F; Boyd, R. J., The Quantum Theory of Atoms in Molecules, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim **2007**.

[21] Bader, R. F. W. and Bayles, D., *J. Phys. Chem. A* **2000**, *104*, 5579–5589.

[22] Grabowski, S. J., Krygowski, T. M. and Leszczynski, J., *J. Phys. Chem. A* **2009**, *113*, 1105–1110.

[23] Lefrancois-Gagnon, K. M. and Mawhinney, R. C., *J. Comput. Chem.* **2020**, *41*, 2485–2503.

[24] Boyd, R. J. and Edgecombe, K. E., *J. Am. Chem. Soc.* **1988**, *110*, 4182–4186.

[25] Boyd, R. J. and Boyd, S. L., *J. Am. Chem. Soc.* **1992**, *114*, 1652–1655.

[26] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A. J., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Keith, T., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J. and Fox, J. D., Gaussian 09 **2010**.

[27] Becke, A. D., *Phys. Rev. A* **1988**, *38*, 3098–3100.

[28] Lee, C., Yang, W. and Parr, R. G., *Phys. Rev. B* **1988**, *37*, 785–789.

[29] Becke, A. D., *J. Chem. Phys* **1993**, *98*, 5648–5652.

[30] Rappoport, D. and Furche, F., *J. Chem. Phys.* **2010**, *133*, 143105.

[31] Keith, T. A., AIMAll **2017**.
URL `aim.tkgristmill.com` Accessed Sept. 1, 2019

[32] R Core Team, R: A Language and Environment for Statistical Computing **2016**.
URL `https://www.r-project.org/` Accessed Sept. 30, 2019

[33] Team, R., RStudio: Integrated Development Environment for R **2016**.
URL http://www.rstudio.com/ Accessed Sept. 30, 2019

[34] Wickham, H., ggplot2: Elegant Graphics for Data Analysis **2009**.
URL http://ggplot2.org Accessed Sept. 30, 2019

[35] Mawhinney, R. C., Peslherbe, G. H. and Muchall, H. M., *J. Phys. Chem.. B* **2008**, *112*, 650–5.

Chapter 6

Relating QTAIM Functional Group Descriptors to Substituent Effect Proxies

6.1 Abstract

Machine learning to describe chemical reactions is a topic of much interest, and physically meaningful descriptors are valuable for inclusion in these models. Substituents modulate reactions, but are commonly described using proxies to their true properties. Substituent descriptors from the Quantum Theory of Atoms in Molecules are related here to these proxies, which have historically had chemically intuitable effects. Due to the large number of descriptors, multivariate analysis is used to intuit their meaning. Multiple linear regression, Principal Components, and Partial Least Squares Regression analyses highlight that these substituent descriptors can predict the proxies, while being truly substituent properties. Sources of error limiting quantitative reproduction of the proxies data include transferability, experimental accuracy, and solvation issues.

6.2 Introduction

Machine learning for predicting outcomes of chemical reactions is a new field in chemistry.[1,2] Physically meaningful descriptors are desirable inputs to these machine learning models, as it makes them easily interpretable.[3] Since substituents modulate chemical reactions, the development of appropriate substituent descriptors for use in machine learning is vital. Currently, proxies measured by the substituent's effect elsewhere in a molecule quantify substituent properties.[4,5] This dates back to the original parameters developed by Hammett, using the pK_a of substituted benzoic acids.[4] The problem with proxies is that they do not reflect the intrinsic properties of the substituent, but only its effect elsewhere in a molecule. Furthermore, they cannot be determined *a priori*.

The majority of all substituent constants followed this model. Ways of measuring the substituent effect include changing the measured probe or substrate or deconstructing Hammett parameters into components. The former derives a new proxy similar to Hammett's, while the latter separates the substituent effect into different mechanisms of transmission. The two most common mechanisms are resonance and field.[6] Inductive (through bond) and polarization effects may be incorporated as well.[7] Equation 6.1 shows a general expression of the determination of field and resonance separation. Hansch, Leo, and Taft calculated the fitting parameters a and b to develop their scale of F and R from σ_m and σ_p.[8] Their specific values of a and b are given in Equation 6.2.

$$\begin{aligned} \sigma_p &= F + R \\ F &= a\sigma_m + b\sigma_p \\ R &= \sigma_p - F \end{aligned} \tag{6.1}$$

$$F = 1.297(\pm 00.147)\sigma_m - 0.385(\pm 0.089)\sigma_p + 0.033(\pm 0.026) \tag{6.2}$$

The advent of computational chemistry provides another route to substituent parameters. sEDA and pEDA descriptors use the change in π and σ electron populations of the aromatic system to which the substituent is attached to quantify the substituent effect.[9] An alternative method uses the molecular electrostatic potential in an aromatic ring.[5] The potential at the meta or para position in the ring corresponds nicely to the related σ_m and σ_p. None of these, however, are true substituent constants. They, too, are proxies, measured by the substituent's effect elsewhere in a molecule.

The Quantum Theory of Atoms in Molecules provides a route to intrinsic substituent constants.[10] Bader defined atoms as regions of space in the molecule containing a nucleus and separated by zero-flux surfaces.[11,12] These surfaces partition molecular space into open quantum systems, which have definable properties obtained by integrating property operators over the basin.[13] Functional group properties are thus sums of the atomic properties of atoms comprising the group.[14] Additionally, QTAIM classifies critical points in the electron density where $\nabla\rho = 0$. Possible classes include nuclear attractor critical points (NACP), bond critical points (BCP), ring critical points (RCP), and cage critical points (CCP) based on the signs of the curvatures of the electron density at that point. Scalar properties at bond critical

points describe the characteristics of the bonding interaction between two atoms.[15]

QTAIM descriptors have been applied in QSPR in the past.[16] Bader and Cortés-Guzmán related the energies of alkanes to a regression on the energy of CH_2 groups.[17] Popelier's Quantum Topological Molecular Similarity relates a distance in BCP space of an aromatic ring to Hammett's descriptors.[18–22] pK_a in solution have been modeled using the atomic energy of hydrogen atoms.[23] QTAIM BCP properties model the energy of intramolecular hydrogen bonds.[24] Machine learning assisted QTAIM dipoles and charges describe out-of-plane infrared intensities in boron halides.[25] QTAIM charges qualitatively describe substituent effects through the σ_m^0 and σ_p^0 constants.[26] QTAIM electron densities aid in describing dihydrofolate reductase inhibitor interactions accurately.[27] The dipole moment along the substituent-substrate bond directly relates to the field effect.[28]

The above work focuses either on group properties or bond critical point properties. A comprehensive study on the application of QTAIM substituent descriptors in modelling reactivity and the relation of the descriptors to traditional proxies is lacking. Chapters 3, 4, and 5 focused on the descriptors' model chemistry sensitivity, the transferability of integrated properties, and patterns in local properties.[29] Model chemistry sensitivity was low, and perfect transferability was not observed, but the integrated properties in one substrate were still linearly related to the properties in other substrates. The properties evaluated at BCPs were non-transferable, but still exhibited useful patterns that could be of use in applying them in modelling reactivities. Due to the wealth of information on bonding characteristics BCP properties contain, it is still desirable to attempt including them in models.[15]

This chapter examines the interrelation of QTAIM substituent de-

scriptors and their use to model available proxies by using multivariate data analysis. To incorporate the numerous QTAIM descriptors available, we utilize Principal Components Analysis, Partial Least Squares Regression, and Multiple Linear Regression to analyze the structure of the descriptors data matrix and build a model to predict the proxies. The descriptors used include multipoles (charge, dipoles, quadrupoles) which describe the substituents' internal properties and BCP properties (ρ, λ_i (i = 1,2,3), G, V) which describe the substituents' bonding properties. We show that these QTAIM substituent descriptors can model commonly used proxies while reflecting the true nature of the substituent.

6.3 Methodology

A set of 52 common substituents, R, for which proxy data is available were selected from Hansch, Leo, and Taft's paper.[8] The substituents chosen are small substituents able to be calculated with a reference level model chemistry, spanning a range of σ_m and σ_p values. RH molecules were optimized in Gaussian09[30] using the B2PLYPD3/aug-cc-pV5Z[31–34] that was used to generate reference values for our model chemistry assessment in Chapter 3. These molecules were rotated to a coordinate system from prior work.[28] In this coordinate system, the atom directly bonded to the substrate hydrogen is positioned at the origin, and the hydrogen is positioned along the $-x$-axis, Figure 6.1.

Extended wavefunction files were generated in this coordinate system, and QTAIM properties assessed using AIMAll 17.11.14.[35] Properties of interest include the group charge, q(R); total dipole, $\boldsymbol{\mu}$(R); charge transfer dipole moment, $\boldsymbol{\mu}^c$(R); polarization dipole moment, $\boldsymbol{\mu}^p$(R); and contributions to molecular quadrupole, $\mathbf{Q}$(R). $\boldsymbol{\mu}$(R), Equation 6.3, is the sum of $\boldsymbol{\mu}^c$(R) and $\boldsymbol{\mu}^p$(R), defined by Equations 6.4 and 6.5, respectively.

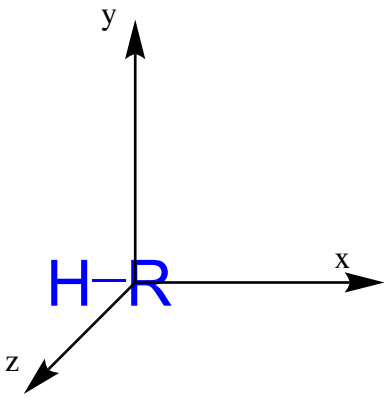

Figure 6.1: Coordinate system of use in the study. The atom of R bonded to H is positioned at the origin. and the R-H bond lies along the $-x$-axis

Q(R) was determined by summing the atomic contributions according to Laidig using Equation 6.6.[36] Here, $\mathbf{R}_\Omega$ is the position of the nuclear attractor in Ω, $\mathbf{R}_b(\Omega|\Lambda)$ is the position of the bond critical point between Ω and attached group Λ, $Q(\Omega|\Lambda)$ is the charge of group Λ, q(Ω) are the atomic charges of basin Ω, χ^α_Ω and χ^β_Ω ($\alpha, \beta = x, y$, or z) are the α or β geometric coordinate of the nucleus of basin Ω, $Q_{\alpha\beta}(\Omega)$ is the atomic quadrupole moment of basin Ω, $\delta_{\alpha\beta}$ is the Kronecker delta, $\langle r^2 \rangle_\Omega$ is the atomic second radial moment of basin Ω, $\mu^{p,\alpha}(\Omega)$ and $\mu^{p,\beta}(\Omega)$ are the α or β component of $\boldsymbol{\mu}^p(\Omega)$, and $Q_{\gamma\gamma}$ is the trace of the contribution to the quadrupole matrix. Lastly, R-H BCP properties are assessed, including the potential energy density, V_c; kinetic energy density, G_c; total energy density, H_c, electron density, ρ_c; distance from the substrate hydrogen to the BCP, r(H-BCP); and the Laplacian of the electron density, $\nabla^2\rho_c$, and its associated eigenvalues, λ_i (i=1-3). The ellipticity of the bond, ε_c is defined as $\lambda_1/\lambda_2 - 1$.

$$\boldsymbol{\mu}(\mathrm{R}) = \boldsymbol{\mu}^{\mathrm{c}}(\mathrm{R}) + \boldsymbol{\mu}^{\mathrm{p}}(\mathrm{R}) \tag{6.3}$$

$$\boldsymbol{\mu}^{\mathrm{c}}(\mathrm{R}) = \sum_{\Omega \in \mathrm{R}} \sum_{\Lambda=1}^{N_b(\Omega)} \left[\mathbf{R}_\Omega - \mathbf{R}_b(\Omega|\Lambda)\right] Q(\Omega|\Lambda) \tag{6.4}$$

$$\boldsymbol{\mu}^{\mathrm{p}}(\mathrm{R}) = \sum_{\Omega \in \mathrm{R}} -\int_{\Omega} \mathbf{r}\rho(\mathbf{r})d\mathbf{r} \tag{6.5}$$

$$\begin{aligned}\mathbf{Q}(R) = \frac{1}{2}\sum_{\Omega \in R}[\,3(q(\Omega)\chi^{\alpha}_{\Omega}\chi^{\beta}_{\Omega} + \frac{Q_{\alpha\beta}(\Omega) + \delta_{\alpha\beta}\{\langle r^2\rangle_{\Omega}\}}{3} + \chi^{\beta}_{\Omega}\mu^{p,\alpha}(\Omega) \\ + \chi^{\alpha}_{\Omega}\mu^{p,\beta}(\Omega)) - \delta_{\alpha\beta}(Q_{\gamma\gamma})]\end{aligned} \tag{6.6}$$

Principal Components Analysis (PCA) and Partial Least Squares Regression (PLSR) were used to analyze patterns in the data. PCA decomposes the data into orthogonal components, $\mathbf{C}_i$, maximizing the variance in the components in decreasing order. The decomposition uses Singular Value Decomposition according to Equation 6.7, where **Q** is the matrix of right singular vectors, **Δ** is the matrix of singular values, and **P** is the matrix of left singular vectors. The first component has the maximum variance, while the second has the maximum variance after the first is removed and is orthogonal to the first. PLSR works similarly to PCA but also takes into account correlation between the descriptors, **X**, and the responses, **Y**. PLSR can analyze multiple related responses simultaneously and are assessed using the root mean square error of cross validation for the training set used to build the model, the root mean square error of prediction for the test set not included in building the model, the r^2 of the predictions of the training data, and Q^2, the r^2 for the predictions of the test set of data.

$$\mathbf{C} = \mathbf{XQ} = \mathbf{P\Delta Q}^{\mathrm{T}} \tag{6.7}$$

Here we take our response matrix, **Y**, as the set of proxies (σ_m, σ_p, F, and R), and our descriptor matrix, **X**, is a matrix of the 19 properties of interest (q(R), μ_x(R), μ^c_x(R), μ^p_x(R), Q_{xx}(R), $||\boldsymbol{\mu}(\mathrm{R})||$, $||\boldsymbol{\mu}^c(\mathrm{R})||$, $||\boldsymbol{\mu}^p(\mathrm{R})||$, $||\mathbf{Q}(\mathrm{R})||$, ρ_c, ε_c, $\nabla^2\rho_c$, $\lambda_{\mathrm{i,c}}$ (i = 1,2,3), G_c, V_c, H_c, and DI(R,H)). The

response matrix is chosen as a set of proxies rather than reactivities because of the non-homogeneity of reactivity data. The equilibrium constants are determined in various solvation environments, and experimental temperatures. They all, however, seek to return values on a uniform scale, that of Hammett's parameters. Therefore, it is most straightforward to compare to the proxies rather than reactivity. The set of descriptors are the determinable QTAIM gas phase descriptors that are well defined for the system. That is, as z and y are undefined. z and y directional properties are not included. It is possible to determine QTAIM descriptors in solvated systems using implicit or explicit solvation models, but to keep the descriptors as easily calculable as possible, it is ideal that they be gas phase.

Three different multivariate data analysis techniques were used. PCA was performed on **X** and **Y** separately to analyze patterns in the data. Multiple Linear Regression (MLR) was used to correlate the raw descriptors to the proxies. Prescaling was performed. PLSR analysis was performed relating **Y** to **X**, done using the kernel algorithm in the pls package,[41] with the data randomly subset into a training and test set, consisting of 80% and 20% of the available data, respectively. Internal validation of the training set was performed using 10-fold cross-validation. The predicted proxy from a MLR or PLSR model is noted with hat notation, for example, $\hat{\sigma}_m$. Data analysis was performed in R using RStudio.[37,38] Graphics generation was done using the ggplot2 package.[39] PCA analysis was performed with the FactoMineR and factoextra packages.[40]

6.4 Results and Discussion

The data to be analyzed includes four response variables (σ_m, σ_p, F, and R) for 52 substituents from Hansch, Leo and Taft's paper.[8] 19 substituent descriptors are calculated for each of those 52 data points. All data used is given in Appendix D. The substituent descriptors included in the study have some intercorrelation between them, which must be taken into account in the analyses. Figure 6.2 highlights these correlations. Not all variables are highly intercorrelated, but intercorrelation is present in blocks. λ_3 and ρ_c strongly intercorrelate with the rest of the BCP properties, as well as with the group charge. The various dipole moment descriptors intercorrelate with each other, and numerous weak correlations are observed. Considering the number of descriptors in our dataset, this clearly illustrates why multivariate analysis is useful for this dataset.

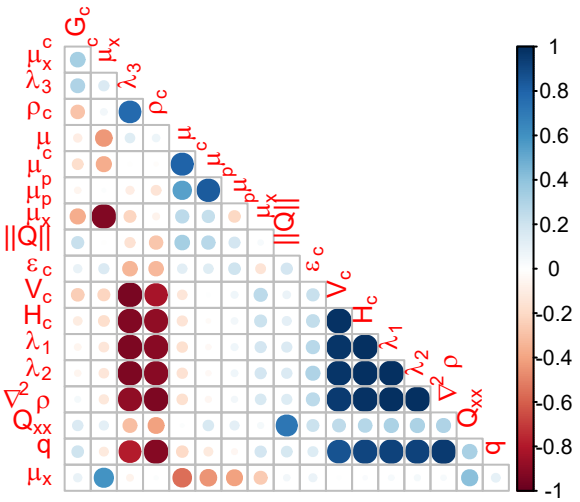

Figure 6.2: Correlation matrix showing intercorrelation of the various QTAIM substituent descriptors. Red indicates inversely correlated values, and blue positively. Circle size and opacity directly relate to the strength of the correlation.

6.4.1 Multiple Linear Regression Analysis

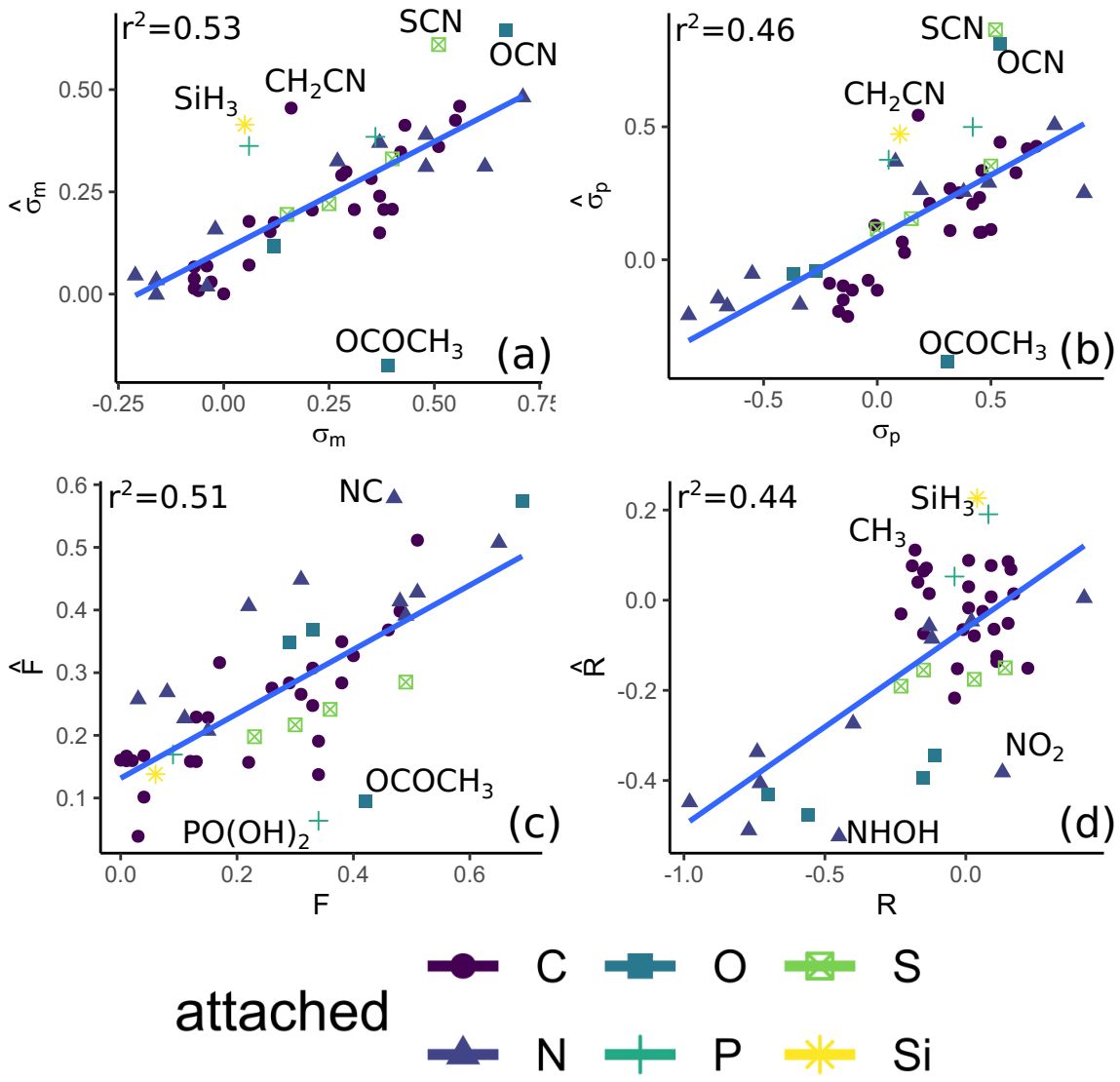

Figure 6.3: Results from bivariate MLR for fitting (A) σ_m, (B) σ_p, (C) F, and (D) R to the data set.

A preliminary Multiple Linear Regression (MLR) study illustrates the potential use of the descriptors, while highlighting the need for more sophisticated multivariate analysis. Bearing the correlation matrix of Figure 6.2 in mind, MLR was attempted. Bivariate MLR modelled the data using all possible combinations of two independent variables. Figure 6.3 shows the best resulting correlations (those with highest r^2) for each proxy. Different proxies were best fit with different substituent

properties. σ_m and σ_p saw the best fit from a combination of μ_x(R) and Q_{xx}(R). These show adequate linear representations, with the presence of a significant number of outliers. F was best fit by combining q(R) and μ_x(R). F represents the field-effect that previously corresponded to μ_x in substituted [1.1.1]-bicyclopentanecarboxylic acids.[28] That μ_x is involved in the best fit for this property confirms these previous findings. The resonance parameter is best fit by completely different properties: λ_1 and λ_2, the curvatures perpendicular to the bond path. As noted in Chapter 3, however, in RH systems, λ_1, and λ_2 are highly linearly dependent on each other, so this is effectively a univariate regression. This fit makes chemical sense, though, as changes in the density perpendicular to the bond path are intuitively related to what are classically viewed as resonance effects. Surprisingly, the properties in R-H molecules relate to the parameters derived from aromatic systems, despite such resonance effects not associated with RH bonding interactions. That BCP properties appear to aid in describing resonance effects further justifies their inclusion in the data set despite lack of transferability.

The r^2 for these plots are low. In the case of σ_m and σ_p, the representation generally looks accurate but has several points not following the linear trend. F and R do not look as well represented, generally. As using two of the multipole properties (dipole, quadrupole and charge) give the highest correlation, one might expect that the inclusion of the third (either q(R) or Q_{xx}(R)) might improve the linear fits. This is not the case, and a trivariate analysis using q(R), μ_x(R), and Q_{xx}(R) to model σ_m, σ_p, and F does not improve the fits. Overall trivariate regression is not attempted due to the overbearing amount of combinations of descriptors present by choosing 3 of 19 variables, and for fear of overfitting. To include information on more variables, PLSR was applied

following a PCA analysis.

6.4.2 PCA decomposition of the data

Prior to performing a PLSR analysis, PCA was used to better understand the data and its structure and validate the use of multi-response PLSR. PCA was used to decompose the proxy data matrix first. To justify using multiresponse PLSR, the responses must be related to each other. We use PCA here to examine σ_m, σ_p, F, and R, and their interrelation. As expected by their formulation (Equation 6.1), the variance in the data set, two components explain all the variance in the proxies, illustrated by the percentage of variance shown in Figure 6.4. The first component explains nearly 80% of the variance in the data set, while the second component explains the remaining 20% of the variation. The third and fourth components can be disregarded in the remaining analysis.

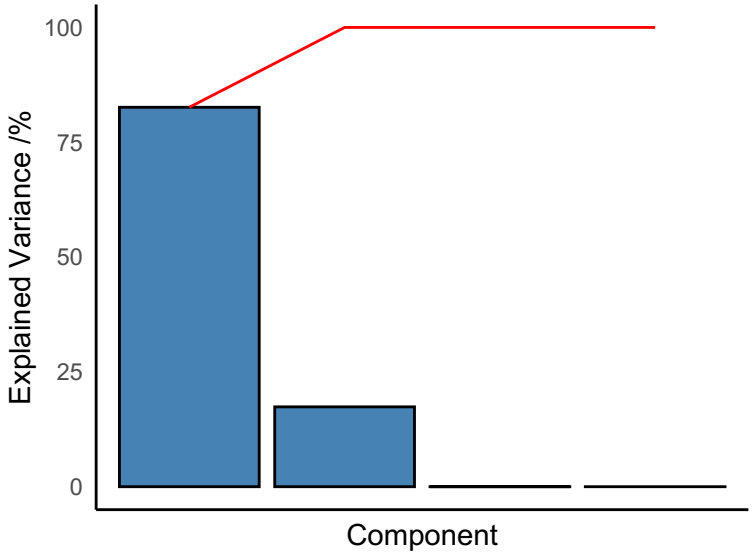

Figure 6.4: Percentage of variance explained by the components in the PCA decomposition of the proxy data

Figure 6.5 shows the contributions of the proxies to each component. All four contribute to the first component, but the primary contributors are σ_m and σ_p, each contributing near 30%. The field and resonance

parameters each contribute roughly 20%. The field and resonance parameters are the main contributors to the second dimension. Figure 6.6 illustrates a biplot of the first two components in this PCA analysis. The biplot shows a plot of the two components against each other, with axes corresponding to the variables in the analysis. Variable axes pointing in similar dimensions are closely correlated to each other, and points far along a variable axis have high values in that variable. We see that σ_p and R point in similar directions, emphasizing that the resonance substituent effect occurs most in the *para* position. σ_m and the field-effect also relate closely. Substituents are roughly separated in this plot based on the atom of the substituent bonded to the substrate. The carbon substituents are all clustered together, and the more electronegative attached atoms also cluster together.

Our substituent descriptors are also decomposed with a PCA, Figure 6.7. Here we include 19 descriptors in the set. As expected, more components are needed to explain the variance in the data. Five components explain 80% of the variance. This is a great reduction in the dimensionality of the data. We now examine what variables contribute to each component, Figure 6.8. Here we examine the first four components, explaining 70% of the variance. Figure 6.8A shows that the first component is composed of many BCP descriptors and the group charge. BCP descriptors not incorporated in this component include $\nabla^2\rho_c$, V_c, and H_c, which are generally associated with a description of the covalency of the bond. The other descriptors incorporated are all correlated together, Figure 6.2. The second component, Figure 6.8B, relates to the dipole moment of the substituent. Component 3, Figure 6.8C, contains information on the dipole, quadrupole, and covalency ($\nabla^2\rho_c$ and V_c). The quadrupole and polarization dipole strongly contribute to component 4, Figure 6.8D. It is intuitive that the polar-

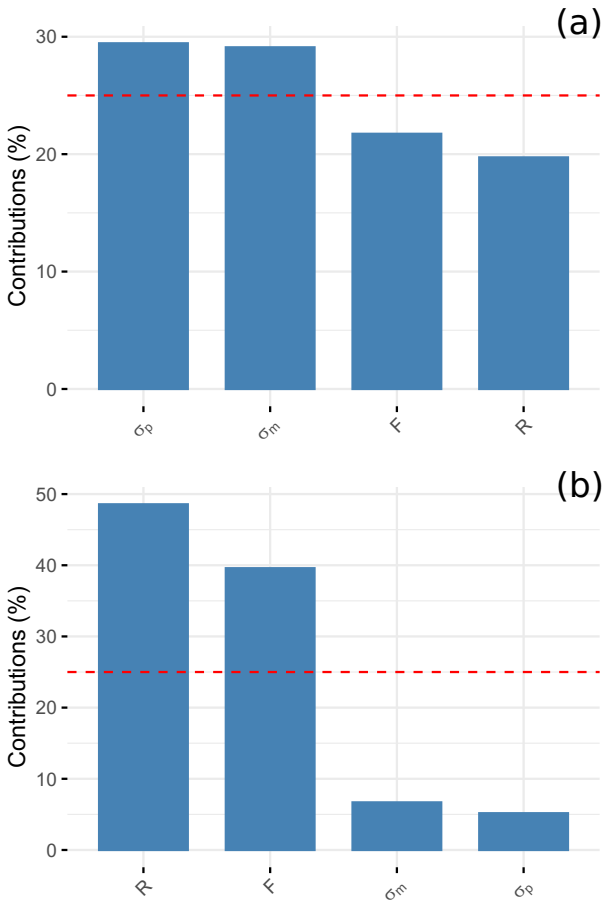

Figure 6.5: Contribution of the proxies, σ_m, σ_p, F, and R to the principal components, (a) Component 1, (b) Component 2. The red dotted line illustrates the average contribution expected for a variable (25% for four variables)

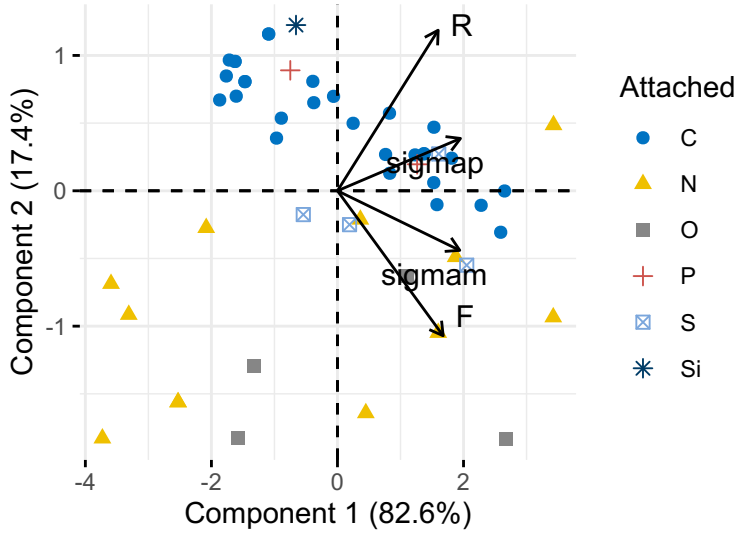

Figure 6.6: PCA biplot showing principal component 1 plotted against principal component 2, with variable axes superimposed over it

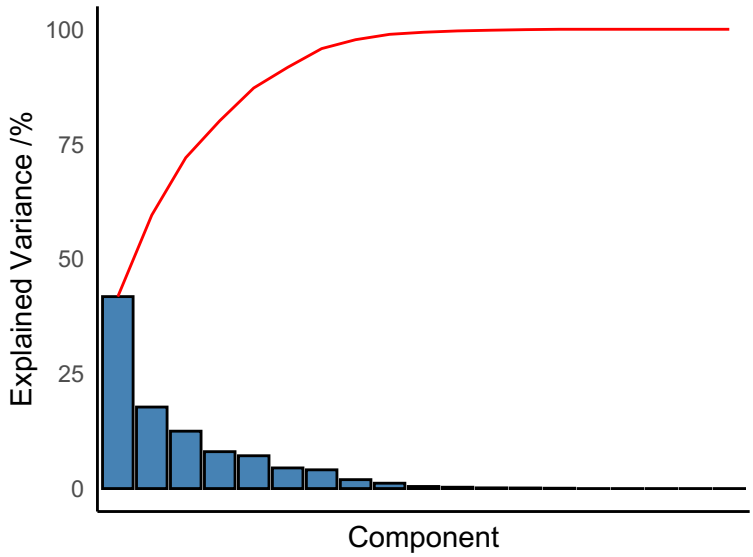

Figure 6.7: Percentage of variance explained by the components in the PCA decomposition of the properties data. Red line is the cumulative percent of variance explained. The red line indicates cumulative percent of explained variance.

ization dipole and quadrupole are grouped together in components, as the polarization dipole is involved in the formulation of the quadrupole, shown in Figure 6.6.

Additionally, there appears to be a separation of the information based on order of multipole - the zeroth order (charge) multipole mainly contributes to the first component, the first order multipole (dipole) mainly contributes to components two and three, and the second order multipole (quadrupole) contributes to component 4. Higher order multipoles therefore contain information that is separate from the lower order multipoles; if they contained the same information they would be in the same component. This could justify the future inclusion of octopoles or hexadecapoles, though that is beyond the scope of this work. Since there is a separation based on order of multipole, the application of Partial Least Squares requires multiple components to describe the proxies if more than one order of multipole is needed, as will be shown.

Understanding the composition of the different components, we can plot the first four components against each other, Figure 6.9. Principal

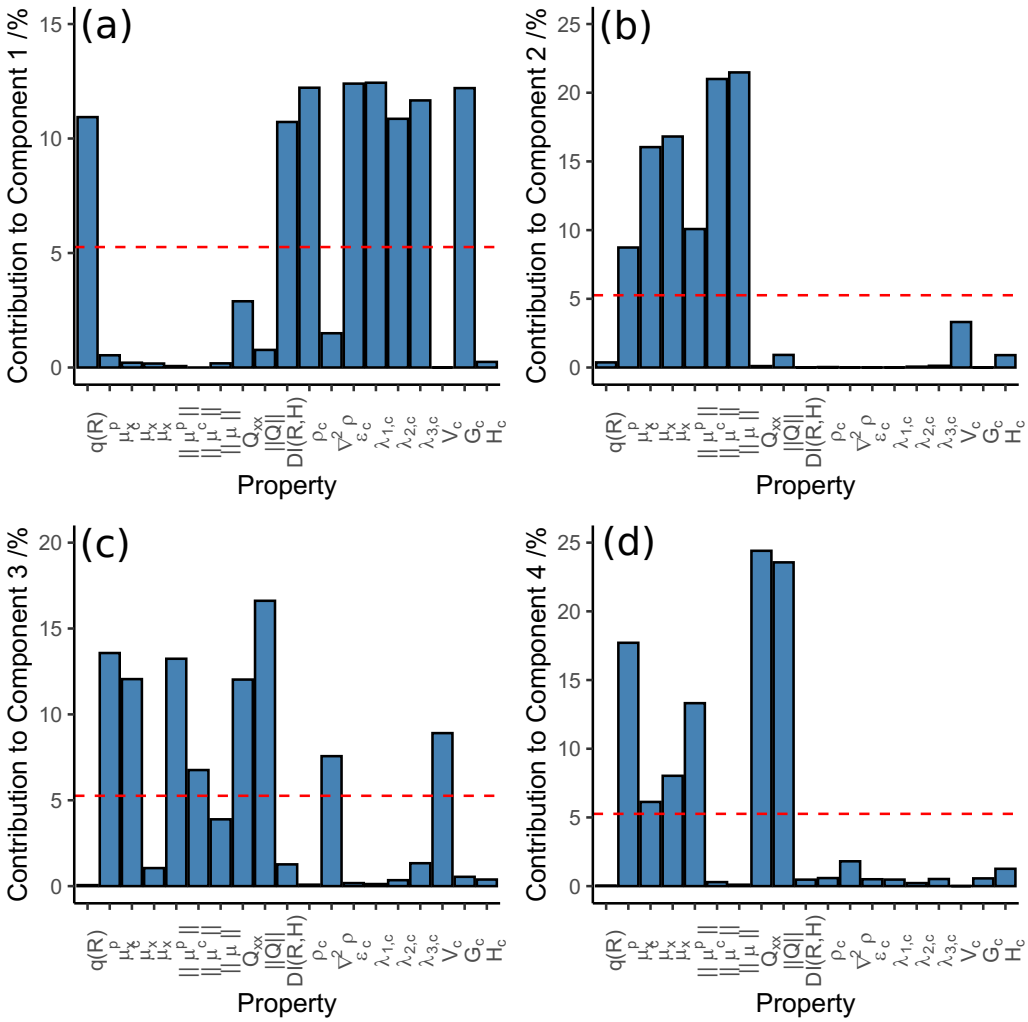

Figure 6.8: Contribution of the variables to the principal components. The red dotted line illustrates the average contribution expected for a variable (5.26% for 19 variables)

Component 1 has a robust physical meaning. It separates the substituents with different types of attached atoms from each other. This relates to the observation in Chapter 5 in which the non-transferability of BCP properties was separated based on the type of attached atom. It further expands on that, revealing that it may not be necessary to include the type of attached atom in a model, but it is instead determinable solely on the BCP properties. The separations are clear and well delineated in all cases, with no erroneous classifications. As the first component reflected information on the R-H critical point, it is clear that it separates the different atom types. Other components are not

separated this way and mostly cluster together. The physical meaning of these other separations is unclear. These components are somewhat interrelated, though, being contributed to by dipoles and quadrupoles, so maybe no clear clustering is due to that.

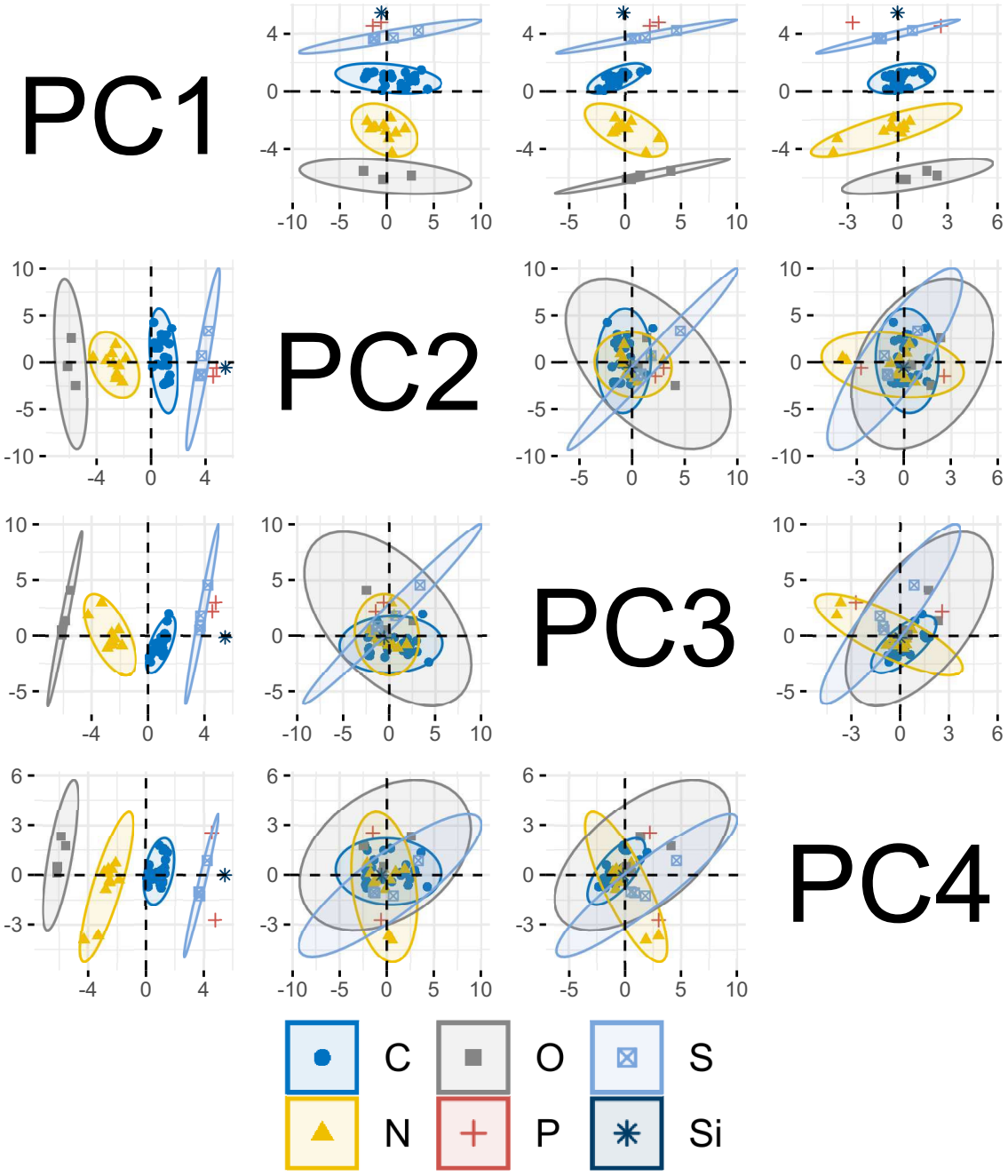

Figure 6.9: Plots of the PCA components against each other

Here, we have understood the structure of the data being studied, and its decomposition. Reducing the dimensionality is feasible, and the

resulting principal components are chemically meaningful. Due to the interrelations between the variables multiresponse PLSR is viable, so we now proceed with modelling the data in this manner.

6.4.3 Partial Least Squares Analysis

The data set was split into a training set and test set for the PLSR model, containing 80% and 20% of the data set, respectively. A multiresponse PLSR model to predict the proxies using the 19 substituent descriptors as the data matrix was constructed. In building the model on the training set, 10-fold cross-validation is used, and RMSEP calculated accordingly. Figure 6.10 shows the RMSEP with the number of components. A high number of components have larger RMSEP due to a lack of ability to predict the points omitted in cross-validation. The number of components with the lowest RMSEP is around four components. After an initial decrease, the model's RMSEP plateaus around four components before increasing again with even higher numbers of components. The RMSEP at five components is roughly the same.

Figure 6.11 and Figure 6.12 show the explained variance of the substituent descriptors and proxies. Four components describe 75% of the variance in the x directional descriptors, and five components explain 81%. In the preliminary PCA analysis, four components explained 70% of the variance and five explained 80%. This decomposition only included a subset (80%) of the data used in the PCA. At four components, where RMSEP reaches its lowest, the proxies only see 58%, 64%, 72% and 77% of the variance explained for σ_m, σ_p, F, and R, respectively. At 5 components, these increase to 75%, 77%, 72% and 78%. This marked increase in the explanation of F suggests that five components are ideal for describing the data since this results in a sufficient amount of variance explained with little change in RMSEP.

The proxies can be predicted based on the model, Figure 6.13. This model acceptably predicts the proxies to an acceptable accuracy. This is remarkable, giving that the proxy data were determined using aromatic data in solvated systems. The r^2 range from 0.72 to 0.78. The

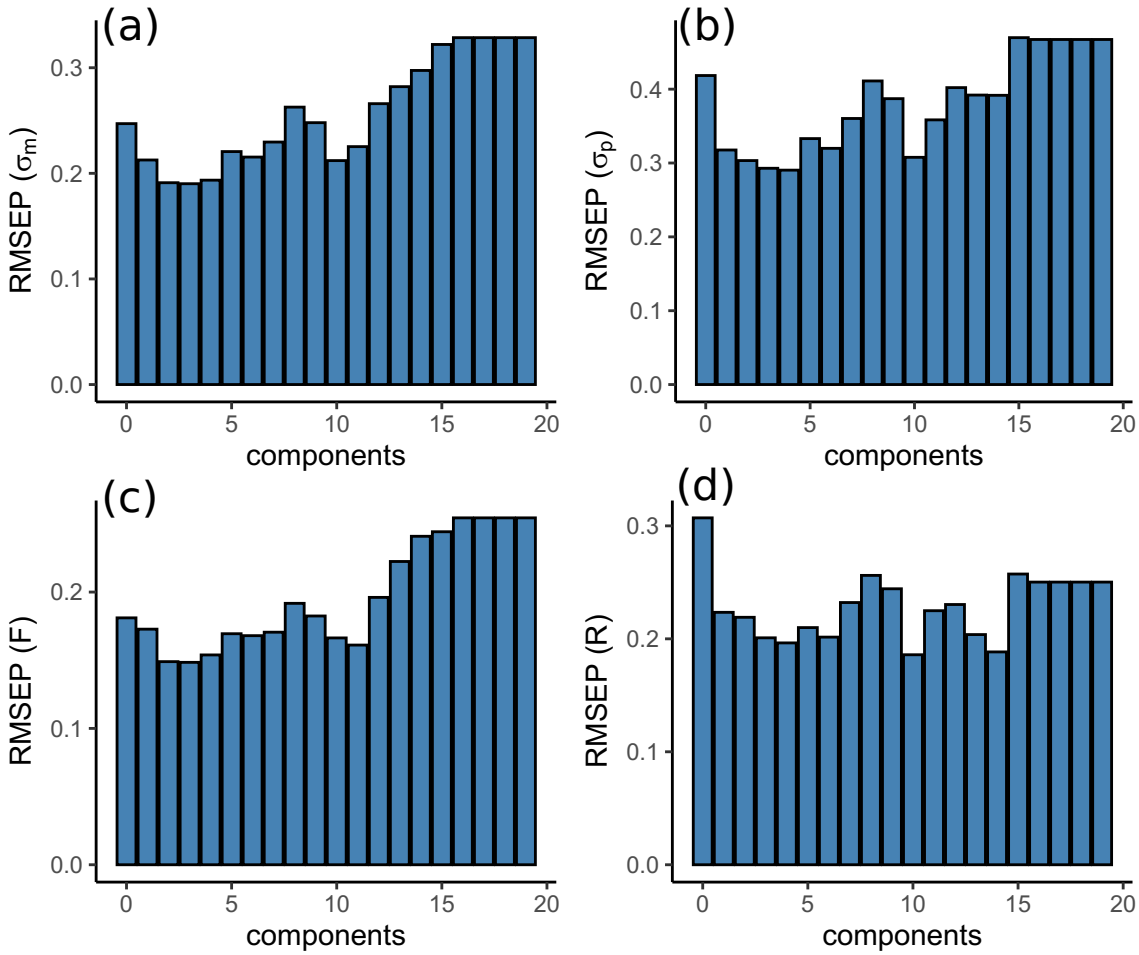

Figure 6.10: Root mean square errors of prediction in the training set, using 5-fold cross validation

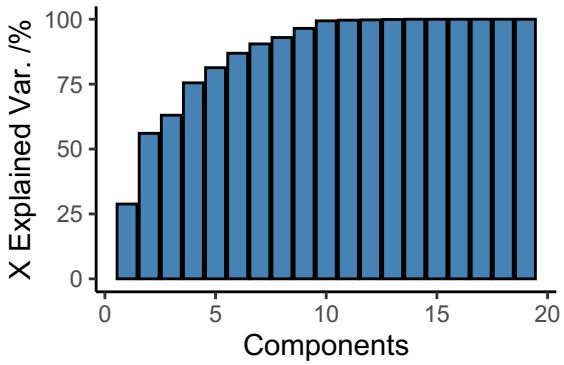

Figure 6.11: Cumulative explained variance of the substituent descriptors

PLSR model can be reduced to the standard linear regression form, $\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E}$, and the coefficients of this regression are shown in Figure 6.14. Most proxies have large coefficients for μ_x, with contributions from Q_{xx}. The coefficients in the model of R are the most starkly differ-

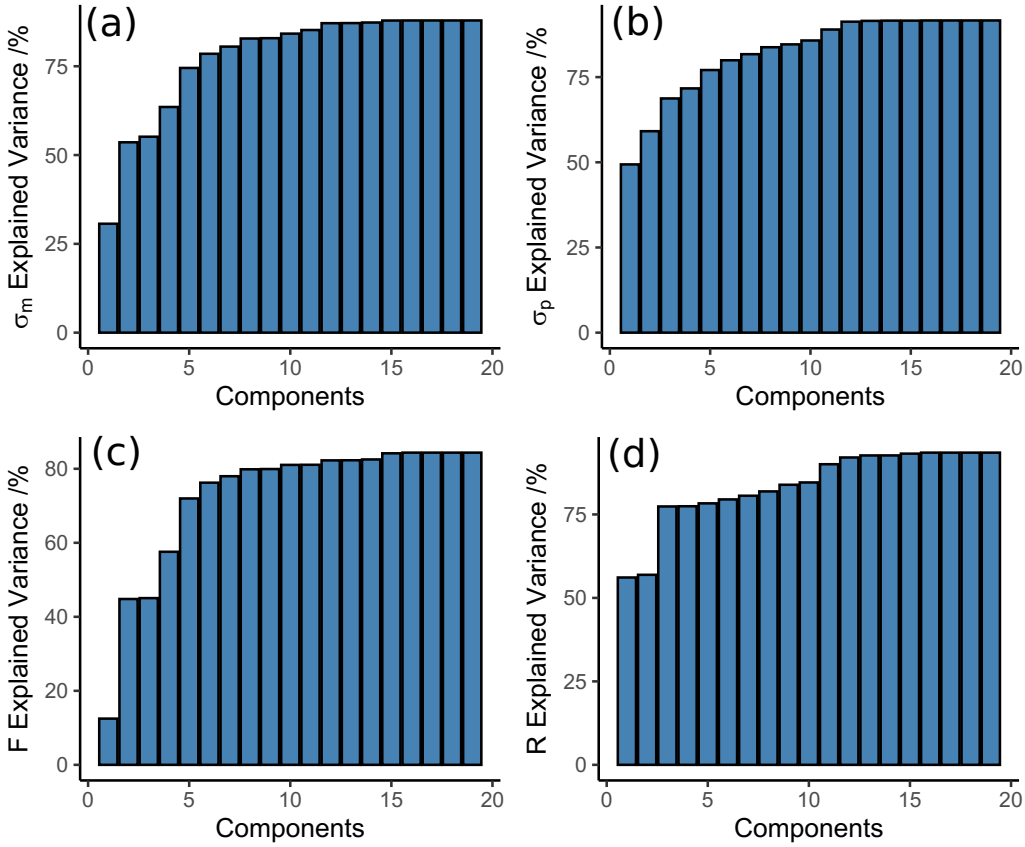

Figure 6.12: Cumulative explained variance of the proxies

ent. Here, G_c, ε_c, and DI(R,H) also play a large role in describing resonance, justifying the inclusion of BCP properties in the models despite issues of non-transferability from Chapter 5. We see these also have an effect, though with smaller coefficients, in σ_p. This makes sense since σ_p is related to the resonance effect by definition of R, but the coefficients are not as large because σ_p also includes information on the field effect.

The model is externally validated using the remaining 20% of the data not included in the training set. The predictions on this external data are shown in Figure 6.15. These predictions show a decent linear model with one outlier. This outlier is the same between the different proxies, $OCOCH_3$. In terms of the resonance effect, the sign of $\hat{R}$ and $\hat{F}$ for $OCOCH_3$ are correct but far too large in magnitude. For σ_m and σ_p, neither the sign, nor the magnitude are correct. The Q^2 for the predic-

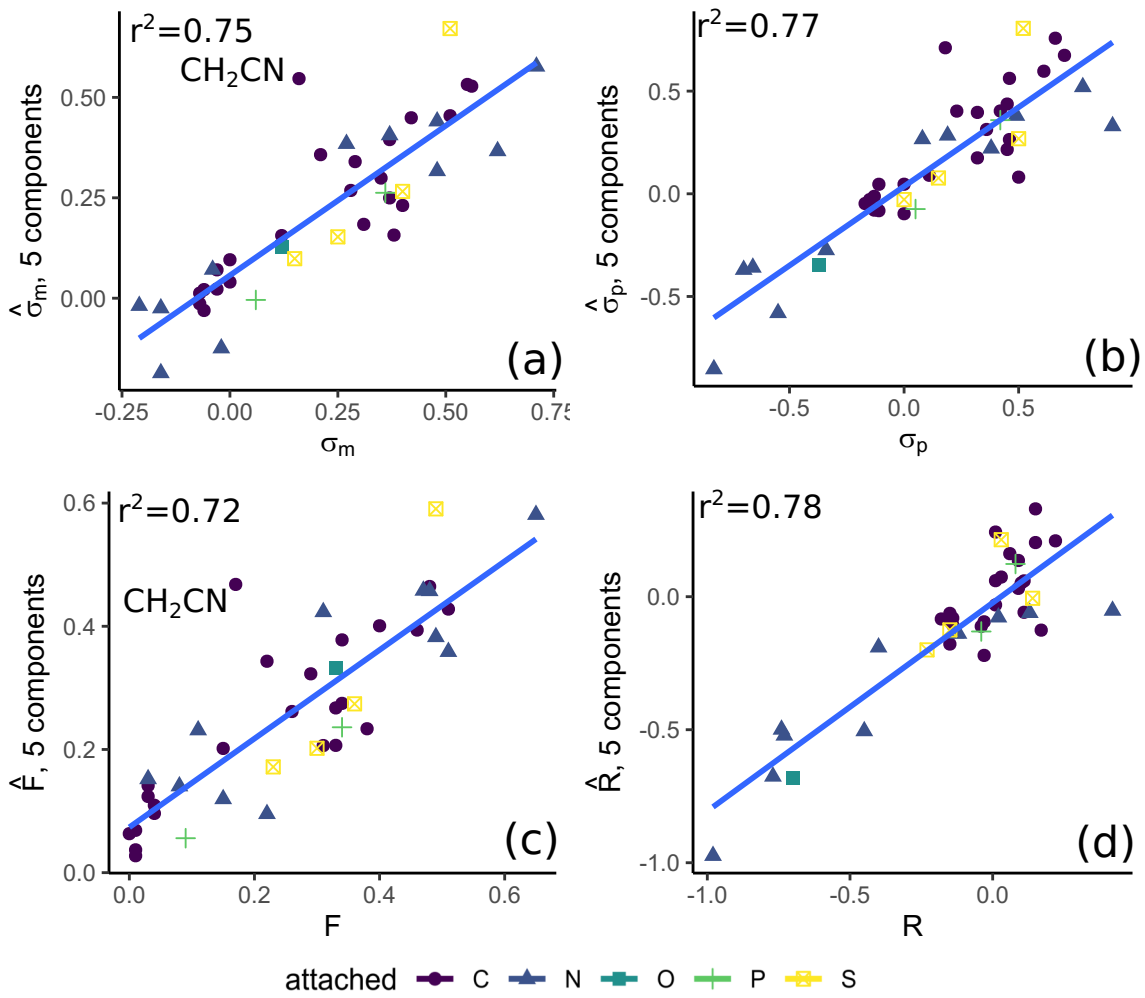

Figure 6.13: Predictions of the proxies in the training set by the PLSR model

tions are lower than ideal due to this outlier. There is no physical reason to exclude the outlier from the set, however. Overall, the model is met with moderate accuracy, indicating that QTAIM substituent descriptors can predict experimental proxies.

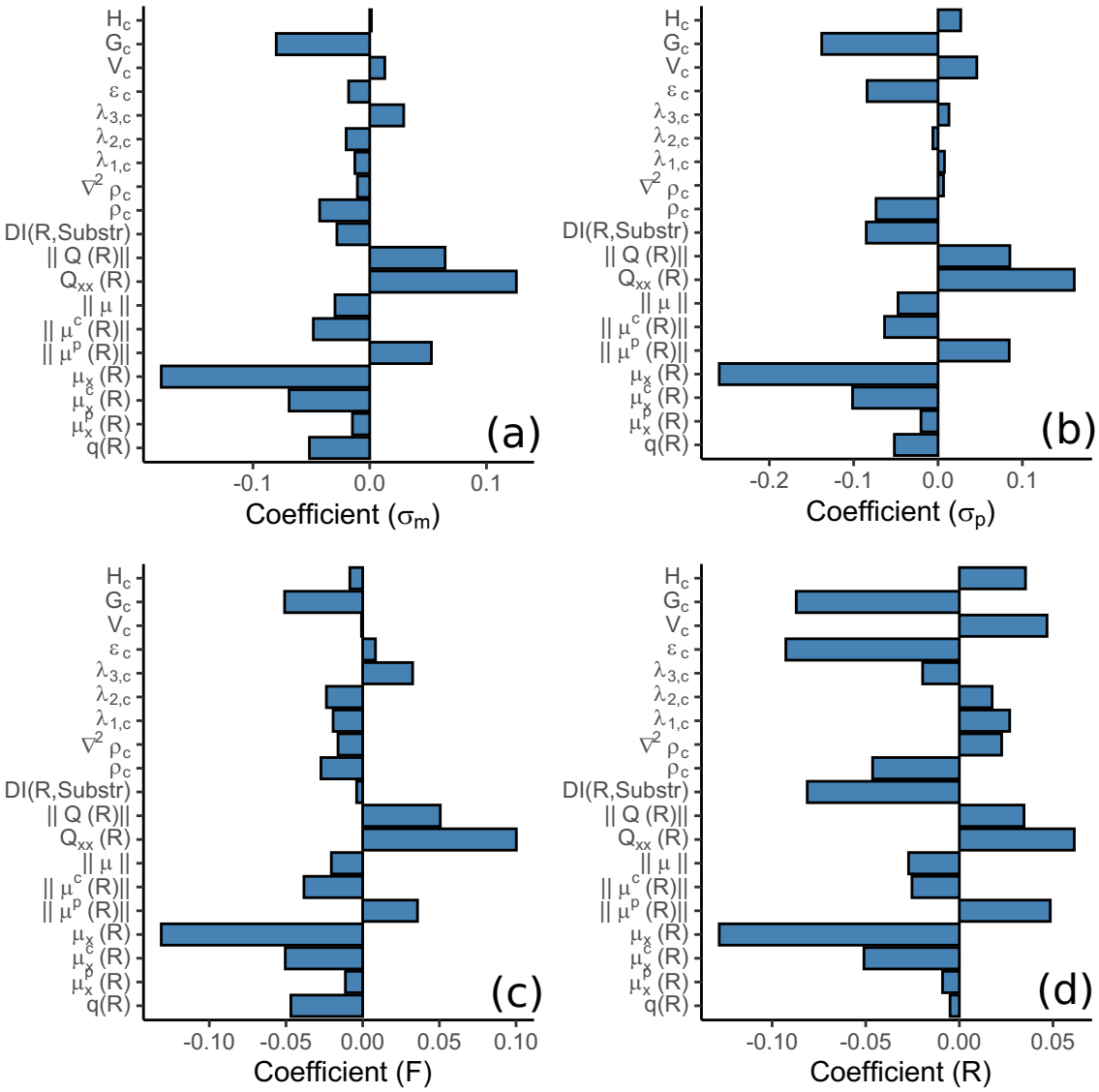

Figure 6.14: Importance of the various variables to the components

6.5 Conclusions

Multivariate data analysis techniques were applied to a data set containing proxies to the substituent effect and substituent descriptors from the Quantum Theory of Atoms in Molecules. MLR and PLSR models achieved acceptable accuracy in predicting the proxies from the substituent descriptors, while PCA illustrated the dimensionality of the data. In taking into account the limited success of these models, the following must be considered: (1) the substituent descriptors from RH molecules are applied to proxies calculated in aromatic systems, (2)

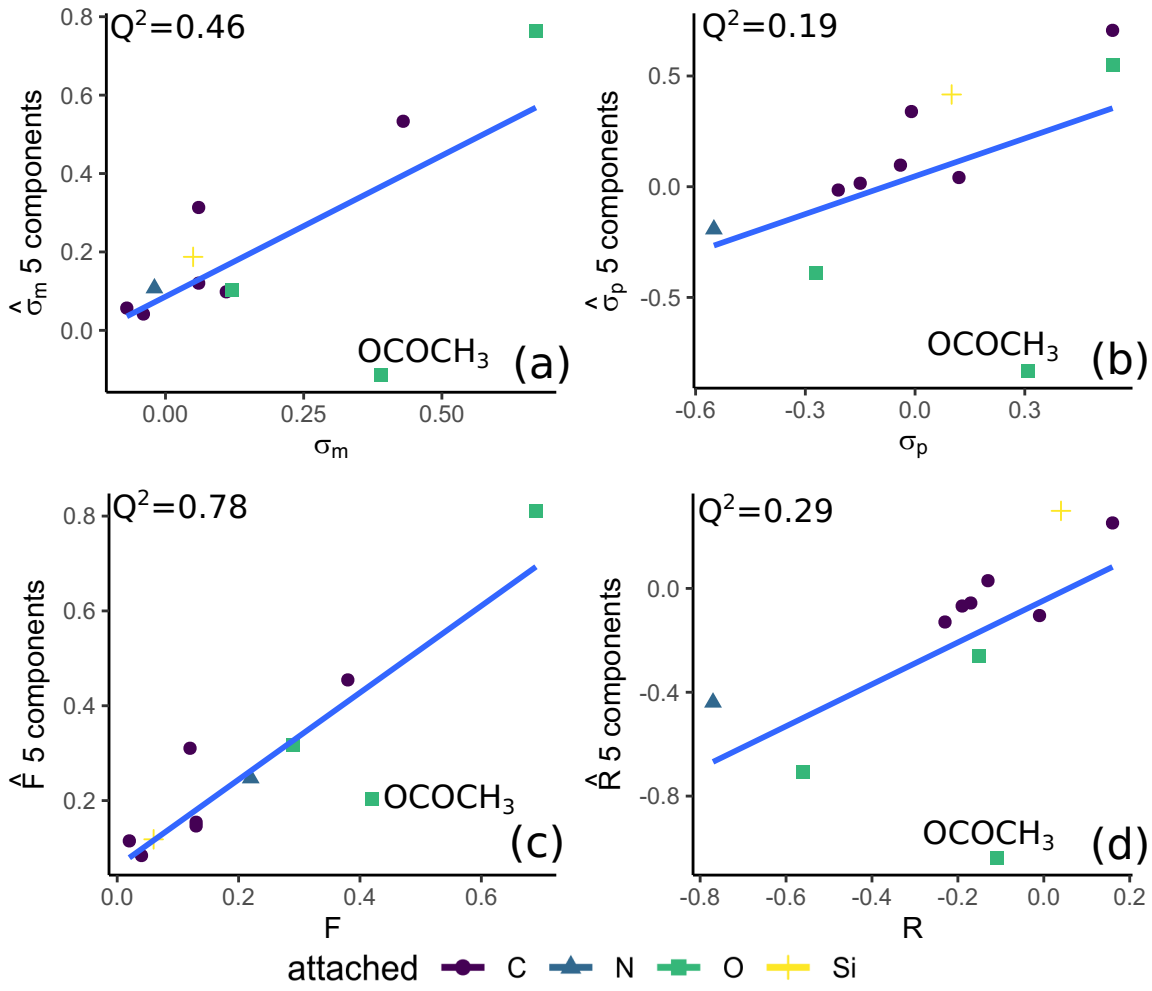

Figure 6.15: Prediction of test set proxies using the PLSR model calculated with 5 components, (A) σ_m, (B) σ_p, (C) F, (D) R

there is inherent uncertainty in the proxies and experimental data, (3) the proxies are also subject to solvation, whereas the substituent descriptors are in the gas phase.

Given these limitations on the data, the accuracy of the models is striking. Again, these are proxies to the substituent effect, not true measures of the substituent's properties, so accurately reproducing these proxies is only part of the bigger picture. From the acceptable PLSR modelling, we show that these descriptors can carry the same information as the proxies, but the lack of absolute quantification of the proxies is not a setback. The relation of the descriptors to proxies already applied in machine learning is evidence that QTAIM substituent

descriptors are viable for inclusion in future machine learning models.

6.5.1 Acknowledgements

This research was funded by the Ontario Graduate Scholarship, and the Lakehead University Faculty of Graduate Studies. This research was enabled in part by support provided by the Shared Hierarchical Academic Research Computing Network (www.sharcnet.ca) and Compute Canada (www.computecanada.ca).

References

[1] Ramakrishnan, R., Dral, P. O., Rupp, M. and von Lilienfeld, O. A., *J. Chem. Theor. Comput.* **2015**, *11*, 2087–2096.

[2] Dral, P. O., *J. Phys. Chem. Lett.* **2020**, *11*, 2336–2347.

[3] Beker, W., Gajewska, E. P., Badowski, T. and Grzybowski, B. A., *Angew. Chem. Int. Ed.* **2019**, *58*, 4515–4519.

[4] Hammett, L. P., *J. Am. Chem. Soc.* **1937**, *59*, 96–103.

[5] Galabov, B., Ilieva, S. and Schaefer, H. F., *J. Org. Chem.* **2006**, *71*, 6382–6387.

[6] Swain, C. G. and Lupton, E. C., *J. Am. Chem. Soc.* **1968**, *90*, 4328–4337.

[7] Topsom, R. D., in Taft, R. W., ed., Progress in Physical Organic Chemistry **1987**, pp. 125–191.
URL `http://doi.wiley.com/10.1002/9780470171950.ch3`

[8] Hansch, C., Leo, A., and Taft, R. W.„ *Chem. Rev.* **1991**, *91*, 165–195.

[9] Ozimiński, W. P. and Dobrowolski, J. C., *J. Phys. Org. Chem.* **2009**, *22*, 769–778.

[10] Bader, R., Atoms in Molecules: A Quantum Theory, Clarendon Press, Oxford, U.K. **1994**.

[11] Bader, R. F. W. and Matta, C. F., *Found. Chem.* **2013**, *15*, 253–276.

[12] Bader, R. F. W., *J. Chem. Phys.* **1972**, *56*, 3320.

[13] Matta, C F; Boyd, R. J., The Quantum Theory of Atoms in Molecules, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim **2007**.

[14] Bader, R. F. W. and Bayles, D., *J. Phys. Chem. A* **2000**, *104*, 5579–5589.

[15] Grabowski, S. J., Krygowski, T. M. and Leszczynski, J., *J. Phys. Chem. A* **2009**, *113*, 1105–1110.

[16] Matta, C. F. and Arabi, A. A., *Future Med. Chem.* **2011**, *3*, 969–994.

[17] Cortés-Guzmán, F. and Bader, R. F. W., *J. Phys. Org. Chem.* **2004**, *17*, 95–99.

[18] Popelier, P. L. A., *The J. Phys. Chem. A* **1999**, *103*, 2883–2890.

[19] O'Brien, S. E. and Popelier, P. L., *Can. J. Chem.* **1999**, *77*, 28–36.

[20] O'Brien, S. E. and Popelier, P. L. A., *J. Chem. Inform. Comput. Sci.* **2001**, *41*, 764–775.

[21] O'Brien, S. E. and Popelier, P. L. A., *J. Chem. Soc., Perkin Trans. 2* **2002**, *2*, 478–483.

[22] Popelier, P. L. A., Chaudry, U. A. and Smith, P. J., *J. Chem. Soc., Perkin Trans. 2* **2002**, *7*, 1231–1237.

[23] Adam, K. R., *J. Phys. Chem. A* **2002**, *106*, 11963–11972.

[24] Afonin, A. V., Vashchenko, A. V. and Sigalov, M. V., *Org. Biomol. Chem.* **2016**, *14*, 11199–11211.

[25] Duarte, L. J. and Bruns, R. E., *J. Phys. Chem. A* **2020**, *124*, 3407–3416, pMID: 32250118.

[26] Nikolova, V., Cheshmedzhieva, D., Ilieva, S. and Galabov, B., *J. Org. Chem.* **2019**, *84*, 1908–1915, pMID: 30620875.

[27] Tosso, R. D., Vettorazzi, M., Andujar, S. A., Gutierrez, L. J., Garro, J. C., Angelina, E., Rodríguez, R., Suvire, F. D., Nogueras, M., Cobo, J. and Enriz, R. D., *J. Mol. Struct.* **2017**, *1134*, 464–474.

[28] Smith, A. P., McKercher, A. E. and Mawhinney, R. C., *J. Phys. Chem.. A* **2011**, *115*, 12544–12554.

[29] Lefrancois-Gagnon, K. M. and Mawhinney, R. C., *J. Comput. Chem.* **2020**, *41*, 2485–2503.

[30] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A. J., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Keith, T., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J. and Fox, J. D., Gaussian 09 **2010**.

[31] Grimme, S., *J. Chem. Phys.* **2006**, *124*, 034108.

[32] Dunning Jr, T. H., *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[33] Kendall, R. A., Dunning, T. H. and Harrison, R. J., *J. Chem. Phys.* **1992**, *96*, 6796–6806.

[34] Dunning T.H., J., Peterson, K. A. and Wilson, A. K., *J. Chem. Phys.* **1993**, *98*, 1358–1371.

[35] Keith, T. A., AIMAll **2017**.
URL `aim.tkgristmill.com` Accessed Sept. 1, 2019

[36] Laidig, K., *Chem. Phys. Lett.* **1991**, *185*, 483–489.

[37] R Core Team, R: A Language and Environment for Statistical Computing **2016**.
URL `https://www.r-project.org/` Accessed Sept. 30, 2019

[38] Team, R., RStudio: Integrated Development Environment for R **2016**.
URL `http://www.rstudio.com/` Accessed Sept. 30, 2019

[39] Wickham, H., ggplot2: Elegant Graphics for Data Analysis **2009**.
URL `http://ggplot2.org` Accessed Sept. 30, 2019

[40] Lê, S., Josse, J. and Husson, F., *J. Stat. Softw.* **2008**, *25*, 1–18.

[41] Mevik, B.-H. and Wehrens, R., *J. Stat. Softw.* **2007**, *18*, 1–23.

Chapter 7

Conclusion

7.1 Overview

This thesis explored the potential of substituent descriptors from the Quantum Theory of Atoms in Molecules to replace the traditional substituent property proxies and be used as machine learning feature sets. Ideal descriptors for machine learning would be insensitive to choice of model chemistry (to enable data from multiple sources to be used simultaneously), and either be transferable between systems (enabling their use in a variety of chemical situations) or in lacking transferability if these errors can be easily explained by simple patterns (so that the patterns can be incorporated in a model, still allowing that descriptor's use). The model chemistry sensitivity and transferability of these integrated group and local critical point descriptors were studied. A proof of concept multivariate analysis was also performed. Each chapter's results are summarized here. Additionally, potential future work and a conclusion on the applicability of QTAIM descriptors in machine learning are presented.

7.2 Model Chemistry Sensitivity

QTAIM descriptors have long been proposed to be insensitive to the choice of model chemistry. For the first time, this was studied extensively for integrated and local properties simultaneously. This was done by varying DFT functionals and basis sets and comparing to high-level double hybrid DFT results. These comparisons conclusively proved the model chemistry insensitivity for integrated group properties. Local properties, however, require triple-ζ basis sets and hybrid functionals for acceptable accuracy.

Model chemistry sensitivity was assessed twofold. First, summary statistics such as the mean deviation and mean absolute deviation determined the absolute quantitative change in properties, which was low for group properties but not insignificant for local properties. Secondly, linear relationships between the various model chemistries and the reference values reveal high-quality linear relationships. Linear relationships reveal that properties that change with model chemistry are still applicable in uses that are built on linear models; it would simply change the determined regression that the model is built on without compromising accuracy.

Two exceptions prohibiting their model chemistry uses exist. The SVWN functional does not determine the correct structure for some substituents, disturbing its linear relationships greatly. Similarly, the double-ζ 6-31G(d) basis set is also not recommended due to high errors in test cases. Qualitative discrepancies in the model exist, and as such, this basis set should not be used since there is no certainty in the accuracy of its molecular structures. Overall this work gives considerable guidelines in selecting model chemistry for determining substituent descriptors. Additionally, it provides a definitive and thorough method for benchmarking DFT model performance for other QTAIM properties.

7.3 Group Property Transferability

This chapter quantitatively explores the theoretical definition of a functional group. In practical applications, it is clear that a functional group is transferable between molecular systems, imparting certain properties and behaving in a certain way in each system. Through summary statistics and linear relationships, the quantitative transferability of QTAIM properties is examined.

QTAIM functional group descriptors are not quantitatively transferred between molecular systems. Large mean deviations and mean absolute deviations are observed. Linear relationships, however, still adequately describe the data in the vast majority of cases. This linearity enables the functional group descriptor from RH molecules to be used in cases with varying molecular systems. This is ideal due to the smaller size and easier computation of the RH molecules than some substrates.

Linear relationship outliers are interpretable in terms of a changing structure in the substituent. Changes in bond angles, bond lengths, or dihedral angles of the studied molecules can significantly reduce the quality of the linear regression. Overall though, QTAIM functional group descriptors are linearly related well and the structural changes are readily accounted for using other descriptors.

7.4 BCP Error Parameterization

Unlike group properties, critical point properties are not intuitively transferable due to the changing substituent-substrate interaction. The substituent's changing interaction with the substrate is shown and quantified through a generalized electronegativity parameter describing the substituent's electronegativity in varying substrates.

Two readily determinable parameters explain the variance in BCP

properties in most cases. The first parameter is simply separating the substituents by the atom that is bonded to the substrate. Second, the change in electronegativity between the substituent in R-H and R-G(G≠H) molecules linearly relates to the change in BCP properties. In cases where the error is nearly constant for a given atom type, electronegativity is not needed to describe the change in BCP property. This is most often observed for the Laplacian of the electron density. Overall, this shows that even changes in BCP properties are easily parameterizable for inclusion in machine learning models.

7.5 Multivariate Data Analysis

The physical meaning of substituent descriptors from QTAIM is not as easily interpretable as the classic Hammett constants. The relationship between these groups and BCP descriptors was studied using multivariate data analysis to elucidate physical meaning in the descriptors.

Principal Component Analysis showed that the 19 included QTAIM descriptors could be reduced to 5 components while still capturing most of the variance of the data. One component relates to bond critical point, group charge, and the delocalization index descriptors. The second relates to the dipole moment descriptors. The third and fourth components relate to the quadrupole descriptors. The principal components separate the substituents based on type of atom bonded to the substituent.

Partial Least Squares was capable of describing 4 Hammett-style substituents simultaneously. Training and test sets are adequately represented. σ_m, σ_p and F relate to the dipole and quadrupole descriptors. The resonance parameter relates to the dipole, bond ellipticity, and kinetic energy density.

7.6 Topics for Future Study

A current drawback in using QTAIM descriptors to describe substituent properties is a lack of ability to describe resonance adequately. That is a drawback of the studied descriptors and not necessarily QTAIM descriptors in general. Future work could define a coordinate system to rigidly and consistently define physically meaningful y and z axes between substituents. This would likely focus on associating the z axis with the resonance effect, making z perpendicular to conjugation planes.

A further way to include the resonance effect in QTAIM is through the properties of lone pairs. QTAIM defines lone pairs as valence shell charge concentrations, critical points in $\nabla^2\rho$. Scalar properties can be evaluated at these points, similar to bond critical points. The inclusion of these and the axis system to better describe resonance effects is a topic of future study.

The substituent data set shown here is extensive and covers a wide variety of substituents. However, many more substituents exist. QTAIM properties for an even more extensive set of substituents can be included, leading toward the development of a substituent database. This database would then be suitable for machine learning methods, which typically require many data points, a larger data set than the current one. An alternative application of the database is to predict descriptors for new substituents. That is, a new substituent, composed of other substituents in the database could have its descriptors predicted based on the properties of its composite substituents.

7.7 General Conclusions

It has been shown that QTAIM functional group descriptors are appropriate for use in machine learning models. They are insensitive to the

model chemistry chosen. Additionally, values of the functional descriptors in RH molecules are transferable to other molecules. Errors for integrated group properties' transferability are interpretable based on geometry changes. One can envision a data structure incorporating multiple possible geometries for a substituent to be applied as necessary in a machine learning language. The non-transferability of BCP properties is overcome by including a parameter relating to the change in electronegativity, taking into account the atom of R bonded to the substrate. The BCP descriptors are worth incorporating in models still as they allow an easily calculable way of capturing resonance effects.

QTAIM functional group descriptors provide physically meaningful machine learning feature sets for inclusion in machine learning models based on linear relationships. The QTAIM descriptors presented are truly descriptive of the substituent's properties, form an *ab initio* perspective while being able to model the same data as traditional proxies despite major drawbacks.

Appendices

Appendix A

Appendix A

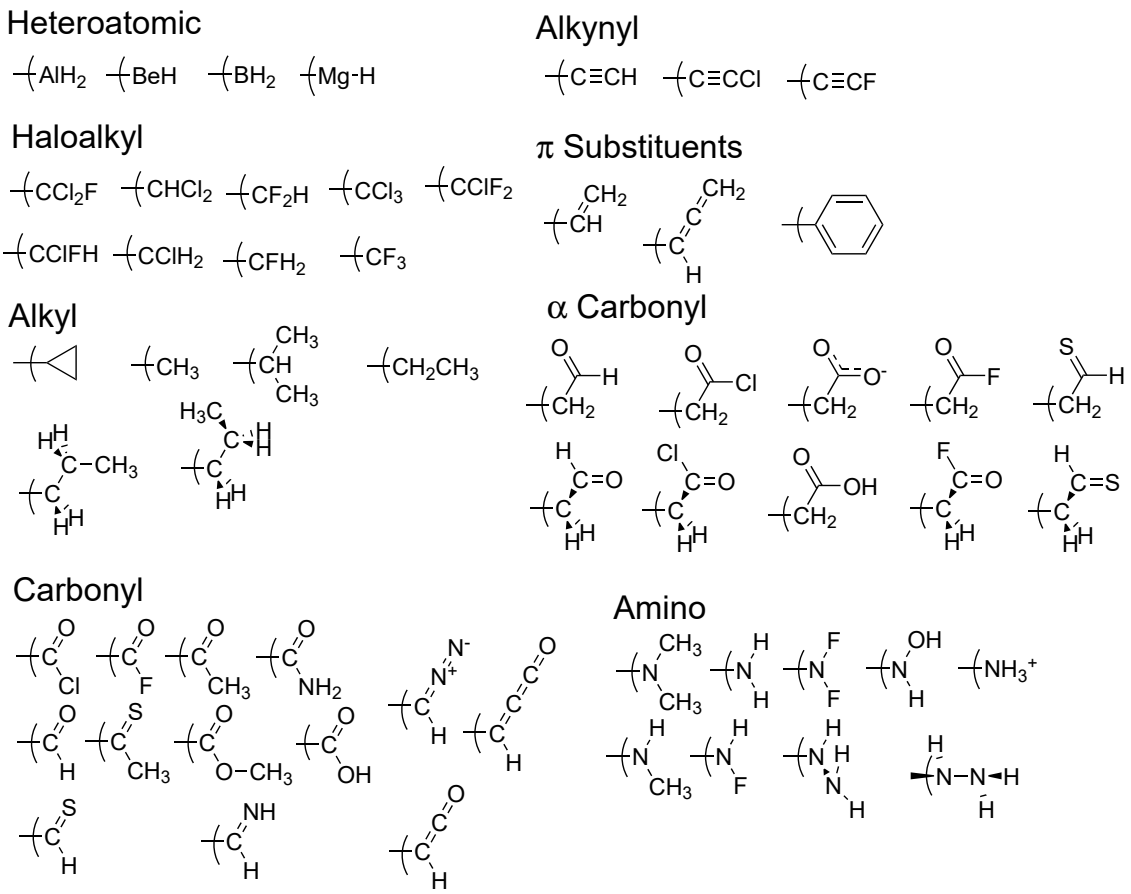

Figure A1: Part 1/2 of Substituents used in analysis

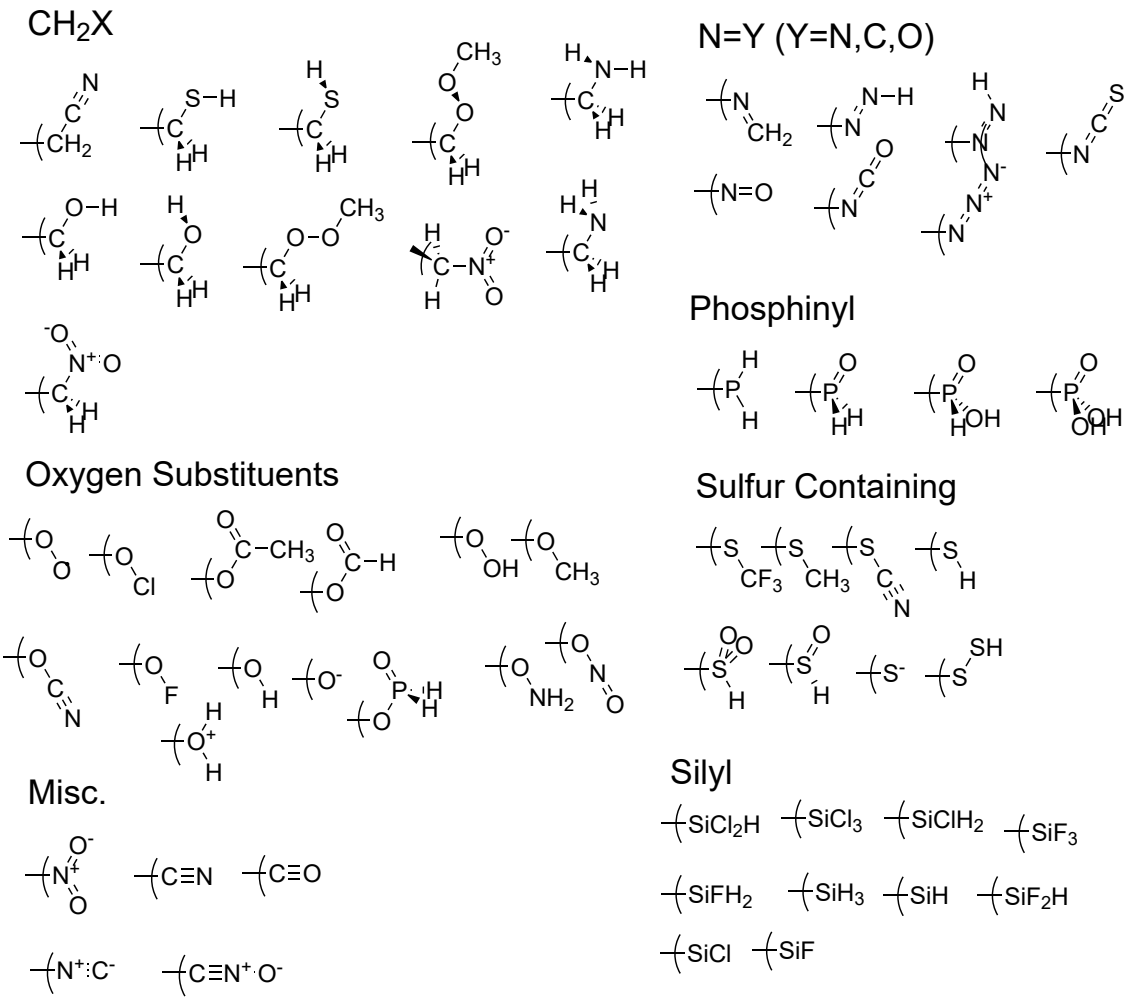

Figure A2: Part 2/2 of Substituents used in analysis

Table A1: Summary statistics for q(R)

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.060	0.065	0.049	0.088(O^-)	-0.205(PH_2)
SVWN/pc-1	-0.050	0.051	0.024	0.029(O^-)	-0.123(PH_2)
SVWN/aug-pc-1	-0.054	0.054	0.022	None	-0.121(PH_2O)
SVWN/pc-2	-0.038	0.039	0.030	0.007(NO_2)	-0.144(PHO(OH))
SVWN/aug-pc-2	-0.040	0.040	0.030	0.007($OO^{\bullet}$)	-0.150($PO(OH)_2$)
SVWN/def2-TZVPP	-0.047	0.048	0.031	0.024(O^-)	-0.123(PH_2)
SVWN/def2-TZVPPD	-0.049	0.049	0.030	0.012(NO)	-0.123(PH_2)
BLYP/6-31G(d)	-0.017	0.041	0.046	0.151(O^-)	-0.137(PH_2)
BLYP/pc-1	-0.004	0.017	0.027	0.088(O^-)	-0.077($PO(OH)_2$)
BLYP/aug-pc-1	-0.007	0.018	0.025	0.049(SCN)	-0.079($PO(OH)_2$)
BLYP/pc-2	0.010	0.027	0.030	0.054(O^-)	-0.090($PO(OH)_2$)
BLYP/aug-pc-2	0.008	0.026	0.030	0.050(COOH)	-0.094($PO(OH)_2$)
BLYP/def2-TZVPP	0.002	0.021	0.030	0.085(O^-)	-0.068(PH_2O)
BLYP/def2-TZVPPD	0.000	0.021	0.029	0.045(NCS)	-0.070($PO(OH)_2$)
PBE/6-31G(d)	-0.027	0.046	0.046	0.127(O^-)	-0.151(PH_2)
PBE/pc-1	-0.012	0.020	0.027	0.069(O^-)	-0.088($PO(OH)_2$)
PBE/aug-pc-1	-0.016	0.021	0.024	0.031(SCF_3)	-0.087($PO(OH)_2$)
PBE/pc-2	-0.002	0.018	0.028	0.037(O^-)	-0.097($PO(OH)_2$)
PBE/aug-pc-2	-0.004	0.017	0.028	0.032(NO)	-0.100($PO(OH)_2$)
PBE/def2-TZVPP	-0.010	0.025	0.028	0.064(O^-)	-0.077(PH_2)
PBE/def2-TZVPPD	-0.013	0.025	0.027	0.038(NO)	-0.078(PH_2O)
TPSS/6-31G(d)	-0.009	0.031	0.036	0.141(O^-)	-0.089(PH_2)
TPSS/pc-1	0.006	0.019	0.026	0.088(SH)	-0.048(SiF_3)
TPSS/aug-pc-1	0.003	0.016	0.023	0.082(SH)	-0.045(SiF_3)
TPSS/pc-2	0.017	0.023	0.020	0.060(SH)	-0.041($PO(OH)_2$)
TPSS/aug-pc-2	0.015	0.021	0.018	0.056(SH)	-0.041($PO(OH)_2$)
TPSS/def2-TZVPP	0.008	0.013	0.016	0.071(O^-)	-0.021(SiH)
TPSS/def2-TZVPPD	0.006	0.012	0.014	0.035(NO)	-0.021($SiCl_3$)
B3LYP/6-31G(d)	-0.020	0.032	0.033	0.120(O^-)	-0.099(PH_2)
B3LYP/pc-1	-0.004	0.018	0.026	0.085(SH)	-0.045(CO)
B3LYP/aug-pc-1	-0.007	0.017	0.023	0.079(SCH_3)	-0.043(SiF_3)
B3LYP/pc-2	0.009	0.018	0.021	0.061(SCF_3)	-0.049($PO(OH)_2$)
B3LYP/aug-pc-2	0.007	0.017	0.020	0.053(SH)	-0.052($PO(OH)_2$)
B3LYP/def2-TZVPP	0.000	0.010	0.015	0.051(O^-)	-0.032(SiH)
B3LYP/def2-TZVPPD	-0.002	0.010	0.013	0.021(NNN)	-0.032(SiH)
PBE0/6-31G(d)	-0.028	0.035	0.032	0.096(O^-)	-0.106(SCH_3)
PBE0/pc-1	-0.009	0.020	0.026	0.080(SH)	-0.050(SiF_3)
PBE0/aug-pc-1	-0.013	0.020	0.022	0.074(SCH_3)	-0.046(SiF_3)
PBE0/pc-2	0.000	0.011	0.017	0.047(S^-)	-0.044(PHO(OH))
PBE0/aug-pc-2	-0.002	0.011	0.015	0.035(SH)	-0.045($PO(OH)_2$)
PBE0/def2-TZVPP	-0.010	0.013	0.011	0.029(O^-)	-0.031(PH_2)
PBE0/def2-TZVPPD	-0.012	0.014	0.010	0.011(NO)	-0.031(PH_2)

Table A2: Linear regression parameters for q(R)

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.93	-0.0638	0.9901	0.039
SVWN/pc-1	0.9572	-0.0524	0.9984	0.0161
SVWN/aug-pc-1	0.9631	-0.0557	0.9984	0.0161
SVWN/pc-2	0.9461	-0.0411	0.9976	0.0195
SVWN/aug-pc-2	0.9475	-0.0428	0.9973	0.0207
SVWN/def2-TZVPP	0.9495	-0.0494	0.9969	0.0222
SVWN/def2-TZVPPD	0.9531	-0.051	0.9968	0.0228
BLYP/6-31G(d)	0.9218	-0.0208	0.9931	0.0323
BLYP/pc-1	0.952	-0.0066	0.9979	0.0182
BLYP/aug-pc-1	0.9579	-0.0096	0.998	0.018
BLYP/pc-2	0.9401	0.0067	0.998	0.0174
BLYP/aug-pc-2	0.9419	0.0052	0.998	0.0179
BLYP/def2-TZVPP	0.9389	-0.001	0.9984	0.0156
BLYP/def2-TZVPPD	0.9434	-0.0029	0.9983	0.0165
PBE/6-31G(d)	0.9257	-0.0312	0.992	0.0347
PBE/pc-1	0.9504	-0.015	0.9981	0.0173
PBE/aug-pc-1	0.9565	-0.0183	0.9983	0.0165
PBE/pc-2	0.9445	-0.005	0.9983	0.0162
PBE/aug-pc-2	0.9466	-0.0066	0.9982	0.0168
PBE/def2-TZVPP	0.9454	-0.0133	0.9982	0.017
PBE/def2-TZVPPD	0.9497	-0.0152	0.998	0.0178
TPSS/6-31G(d)	0.9467	-0.0122	0.9949	0.0283
TPSS/pc-1	0.9676	0.0046	0.9969	0.0226
TPSS/aug-pc-1	0.9753	0.0018	0.9974	0.0209
TPSS/pc-2	0.9672	0.0149	0.9988	0.014
TPSS/aug-pc-2	0.9704	0.0136	0.999	0.0131
TPSS/def2-TZVPP	0.9692	0.0063	0.9994	0.0098
TPSS/def2-TZVPPD	0.9738	0.0046	0.9995	0.0094
B3LYP/6-31G(d)	0.9531	-0.0225	0.9957	0.0263
B3LYP/pc-1	0.9799	-0.0055	0.9964	0.0247
B3LYP/aug-pc-1	0.9852	-0.008	0.997	0.0225
B3LYP/pc-2	0.9722	0.0075	0.9982	0.0175
B3LYP/aug-pc-2	0.9743	0.006	0.9984	0.0164
B3LYP/def2-TZVPP	0.9737	-0.0018	0.9994	0.01
B3LYP/def2-TZVPPD	0.978	-0.0036	0.9994	0.0098
PBE0/6-31G(d)	0.9621	-0.0302	0.9953	0.0278
PBE0/pc-1	0.9833	-0.0103	0.9963	0.025
PBE0/aug-pc-1	0.9891	-0.0131	0.9972	0.022
PBE0/pc-2	0.9825	-9e-04	0.9987	0.0151
PBE0/aug-pc-2	0.9851	-0.0026	0.9989	0.0139
PBE0/def2-TZVPP	0.9862	-0.0111	0.9995	0.0095
PBE0/def2-TZVPPD	0.9904	-0.0129	0.9995	0.0089

Table A3: Summary statistics for $\mu_x(R)$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.003	0.041	0.072	0.331($PO(OH)_2$)	-0.267(SCF_3)
SVWN/pc-1	-0.002	0.027	0.046	0.234($PO(OH)_2$)	-0.137(CHCCO)
SVWN/aug-pc-1	-0.002	0.014	0.026	0.095($PO(OH)_2$)	-0.109(CHCCO)
SVWN/pc-2	-0.004	0.019	0.036	0.155($PO(OH)_2$)	-0.235(CHCCO)
SVWN/aug-pc-2	-0.002	0.014	0.030	0.120($PO(OH)_2$)	-0.190(CHCCO)
SVWN/def2-TZVPP	-0.003	0.018	0.034	0.140($PO(OH)_2$)	-0.201(CHCCO)
SVWN/def2-TZVPPD	-0.002	0.015	0.031	0.102($PO(OH)_2$)	-0.204(CHCCO)
BLYP/6-31G(d)	0.002	0.034	0.060	0.284($PO(OH)_2$)	-0.231(SCF_3)
BLYP/pc-1	0.004	0.024	0.041	0.200($PO(OH)_2$)	-0.108(PHO(OH))
BLYP/aug-pc-1	0.004	0.014	0.026	0.116(ONO)	-0.108(CH_2COF—op)
BLYP/pc-2	0.002	0.016	0.028	0.119($PO(OH)_2$)	-0.096(CH_2COF—ip)
BLYP/aug-pc-2	0.003	0.013	0.024	0.125(ONO)	-0.091(CH_2COF—ip)
BLYP/def2-TZVPP	0.003	0.013	0.024	0.104($PO(OH)_2$)	-0.060(CH_2COH—ip)
BLYP/def2-TZVPPD	0.003	0.012	0.023	0.121(ONO)	-0.085(CH_2COF—op)
PBE/6-31G(d)	0.001	0.036	0.064	0.301($PO(OH)_2$)	-0.263(SCF_3)
PBE/pc-1	0.002	0.023	0.040	0.213($PO(OH)_2$)	-0.096(PHO(OH))
PBE/aug-pc-1	0.002	0.013	0.021	0.079($PO(OH)_2$)	-0.051(PHO(OH))
PBE/pc-2	0.001	0.013	0.023	0.139($PO(OH)_2$)	-0.053(CHCCO)
PBE/aug-pc-2	0.002	0.011	0.020	0.098($PO(OH)_2$)	-0.056(CHCCO)
PBE/def2-TZVPP	0.001	0.014	0.024	0.122($PO(OH)_2$)	-0.069(CH_2COH—op)
PBE/def2-TZVPPD	0.001	0.012	0.020	0.082($PO(OH)_2$)	-0.060(CHCCO)
TPSS/6-31G(d)	0.000	0.033	0.057	0.275($PO(OH)_2$)	-0.215(SCF_3)
TPSS/pc-1	0.002	0.022	0.039	0.194($PO(OH)_2$)	-0.084(PHO(OH))
TPSS/aug-pc-1	0.002	0.012	0.020	0.071(ONO)	-0.069(CH_2COF—op)
TPSS/pc-2	0.000	0.012	0.022	0.121($PO(OH)_2$)	-0.054(OCN)
TPSS/aug-pc-2	0.001	0.010	0.017	0.083($PO(OH)_2$)	-0.039(CH_2COF—op)
TPSS/def2-TZVPP	0.001	0.011	0.019	0.104($PO(OH)_2$)	-0.054(CH_2COH—op)
TPSS/def2-TZVPPD	0.001	0.009	0.015	0.070($PO(OH)_2$)	-0.045(OCN)
B3LYP/6-31G(d)	-0.002	0.031	0.052	0.214($PO(OH)_2$)	-0.182(SCF_3)
B3LYP/pc-1	0.000	0.020	0.035	0.135($PO(OH)_2$)	-0.086(Ocl)
B3LYP/aug-pc-1	0.000	0.008	0.014	0.045(OPH_2O)	-0.040(CH_2COF—ip)
B3LYP/pc-2	-0.002	0.012	0.020	0.061($PO(OH)_2$)	-0.064(CHCCO)
B3LYP/aug-pc-2	-0.001	0.005	0.010	0.028(SiF)	-0.065(CHCCO)
B3LYP/def2-TZVPP	-0.001	0.008	0.015	0.048($PO(OH)_2$)	-0.062(CHCCO)
B3LYP/def2-TZVPPD	-0.001	0.005	0.010	0.028(SiF)	-0.062(CHCCO)
PBE0/6-31G(d)	-0.004	0.033	0.057	0.215($PO(OH)_2$)	-0.202(SCF_3)
PBE0/pc-1	-0.002	0.022	0.038	0.131($PO(OH)_2$)	-0.114(ONO)
PBE0/aug-pc-1	-0.002	0.010	0.017	0.040(OPH_2O)	-0.092(ONO)
PBE0/pc-2	-0.004	0.014	0.024	0.066($PO(OH)_2$)	-0.102(ONO)
PBE0/aug-pc-2	-0.002	0.009	0.018	0.054(CH_2COF—ip)	-0.087(CHCCO)
PBE0/def2-TZVPP	-0.003	0.012	0.022	0.073(CH_2COF—ip)	-0.107(ONO)
PBE0/def2-TZVPPD	-0.003	0.010	0.018	0.059(CH_2COF—ip)	-0.095(ONO)

Table A4: Summary statistics for $\mu_y(R)$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.013	0.052	0.076	0.296(CHCCO)	-0.198($CClF_2$)
SVWN/pc-1	-0.011	0.028	0.047	0.234(CHCCO)	-0.167(NNN)
SVWN/aug-pc-1	-0.015	0.025	0.035	0.177(CHCCO)	-0.144(NNN)
SVWN/pc-2	-0.016	0.029	0.046	0.312(CHCCO)	-0.168(NNN)
SVWN/aug-pc-2	-0.018	0.031	0.045	0.252(CHCCO)	-0.160(NNN)
SVWN/def2-TZVPP	-0.021	0.034	0.049	0.241(CHCCO)	-0.173(NNN)
SVWN/def2-TZVPPD	-0.019	0.032	0.046	0.262(CHCCO)	-0.163(NNN)
BLYP/6-31G(d)	0.004	0.044	0.070	0.286(CH_2COO^-—ip)	-0.204($CClF_2$)
BLYP/pc-1	0.006	0.028	0.046	0.252(CH_2COO^-—ip)	-0.085(NCH_2)
BLYP/aug-pc-1	0.005	0.017	0.031	0.150(CH_2COCl—ip)	-0.106(SiF)
BLYP/pc-2	0.004	0.017	0.028	0.118(CH_2COCl—ip)	-0.057(SHO)
BLYP/aug-pc-2	0.003	0.015	0.025	0.114(CH_2COCl—ip)	-0.048(NCS)
BLYP/def2-TZVPP	-0.002	0.019	0.029	0.150(CH_2COO^-—ip)	-0.063(OPH_2O)
BLYP/def2-TZVPPD	0.002	0.015	0.024	0.106(CH_2COCl—ip)	-0.050(NCS)
PBE/6-31G(d)	-0.003	0.046	0.069	0.291(CH_2COO^-—ip)	-0.208($CClF_2$)
PBE/pc-1	0.000	0.024	0.040	0.246(CH_2COO^-—ip)	-0.080($CClF_2$)
PBE/aug-pc-1	-0.004	0.017	0.026	0.088(CH_2COO^-—ip)	-0.066(SiF)
PBE/pc-2	-0.005	0.016	0.025	0.132(CH_2COO^-—ip)	-0.074(NNN)
PBE/aug-pc-2	-0.006	0.018	0.027	0.107(CH_2COO^-—ip)	-0.064(NNN)
PBE/def2-TZVPP	-0.010	0.023	0.035	0.176(CH_2COO^-—ip)	-0.090(SiCl)
PBE/def2-TZVPPD	-0.007	0.019	0.027	0.095(SHO)	-0.070(SiCl)
TPSS/6-31G(d)	0.006	0.039	0.062	0.257(CH_2COO^-—ip)	-0.176($CClF_2$)
TPSS/pc-1	0.009	0.021	0.039	0.200(CH_2COO^-—ip)	-0.107(SHO)
TPSS/aug-pc-1	0.004	0.014	0.024	0.108(NCO)	-0.052(SiF)
TPSS/pc-2	0.003	0.014	0.023	0.104(CH_2COO^-—ip)	-0.050($PO(OH)_2$)
TPSS/aug-pc-2	0.002	0.014	0.022	0.088(CH_2COO^-—ip)	-0.056(SiCl)
TPSS/def2-TZVPP	-0.002	0.016	0.028	0.146(CH_2COO^-—ip)	-0.078(SiCl)
TPSS/def2-TZVPPD	0.001	0.014	0.022	0.078(SHO)	-0.065(SiCl)
B3LYP/6-31G(d)	0.003	0.033	0.051	0.203(CH_2COO^-—ip)	-0.144($CClF_2$)
B3LYP/pc-1	0.006	0.021	0.040	0.154(SiCl)	-0.187(SHO)
B3LYP/aug-pc-1	0.001	0.013	0.024	0.085(CHCCO)	-0.107(SiF)
B3LYP/pc-2	0.001	0.015	0.027	0.184(CHCCO)	-0.117(SHO)
B3LYP/aug-pc-2	0.000	0.009	0.019	0.161(CHCCO)	-0.044(NCO)
B3LYP/def2-TZVPP	-0.004	0.011	0.020	0.127(CHCCO)	-0.048(NCO)
B3LYP/def2-TZVPPD	-0.001	0.008	0.017	0.150(CHCCO)	-0.047(NCO)
PBE0/6-31G(d)	-0.001	0.033	0.050	0.195(CH_2COO^-—ip)	-0.140($CClF_2$)
PBE0/pc-1	0.002	0.020	0.038	0.154(CHCCO)	-0.199(SHO)
PBE0/aug-pc-1	-0.005	0.013	0.020	0.075(CHCCO)	-0.085(SHO)
PBE0/pc-2	-0.005	0.016	0.027	0.189(CHCCO)	-0.090(SHO)
PBE0/aug-pc-2	-0.007	0.015	0.023	0.158(CHCCO)	-0.046(CH_2COF—ip)
PBE0/def2-TZVPP	-0.010	0.018	0.026	0.125(CHCCO)	-0.065(CH_2COCl—ip)
PBE0/def2-TZVPPD	-0.008	0.016	0.023	0.145(CHCCO)	-0.057(CH_2COCl—ip)

Table A5: Summary statistics for $\mu_z(R)$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.007	0.039	0.067	0.209(PHO(OH))	-0.190(SH)
SVWN/pc-1	0.001	0.024	0.044	0.169(SSH)	-0.175(SH)
SVWN/aug-pc-1	0.000	0.016	0.028	0.110(CNO)	-0.111(CHNN)
SVWN/pc-2	-0.003	0.019	0.035	0.144(PHO(OH))	-0.095(S^-)
SVWN/aug-pc-2	-0.004	0.020	0.035	0.130(CNO)	-0.095(PH_2O)
SVWN/def2-TZVPP	-0.004	0.023	0.040	0.147(CNO)	-0.110(PH_2O)
SVWN/def2-TZVPPD	-0.004	0.022	0.036	0.136(CNO)	-0.093(PH_2O)
BLYP/6-31G(d)	-0.002	0.035	0.059	0.159(PHO(OH))	-0.156(PH_2O)
BLYP/pc-1	0.006	0.021	0.038	0.121($SiCl_2H$)	-0.127(SH)
BLYP/aug-pc-1	0.004	0.011	0.021	0.094($SiFH_2$)	-0.076(CHNN)
BLYP/pc-2	0.003	0.013	0.024	0.095(PHO(OH))	-0.075(SH)
BLYP/aug-pc-2	0.000	0.010	0.018	0.066(PHO(OH))	-0.068(PH_2O)
BLYP/def2-TZVPP	0.001	0.013	0.024	0.087(CNO)	-0.102(PH_2O)
BLYP/def2-TZVPPD	0.000	0.010	0.018	0.060(CNO)	-0.068(PH_2O)
PBE/6-31G(d)	-0.004	0.038	0.064	0.174(PHO(OH))	-0.166(SH)
PBE/pc-1	0.004	0.022	0.041	0.148(SSH)	-0.154(SH)
PBE/aug-pc-1	0.002	0.012	0.023	0.097(CNO)	-0.070(CHNN)
PBE/pc-2	0.000	0.014	0.028	0.115(CNO)	-0.081(SH)
PBE/aug-pc-2	-0.002	0.014	0.027	0.113(CNO)	-0.101(PH_2O)
PBE/def2-TZVPP	-0.001	0.018	0.035	0.144(CNO)	-0.132(PH_2O)
PBE/def2-TZVPPD	-0.002	0.015	0.028	0.119(CNO)	-0.102(PH_2O)
TPSS/6-31G(d)	-0.003	0.034	0.058	0.146(SSH)	-0.160(SH)
TPSS/pc-1	0.005	0.022	0.043	0.159(SSH)	-0.166(SH)
TPSS/aug-pc-1	0.002	0.009	0.017	0.070($SiFH_2$)	-0.071(CHNN)
TPSS/pc-2	0.000	0.012	0.024	0.089(PHO(OH))	-0.093(SH)
TPSS/aug-pc-2	-0.002	0.009	0.018	0.065(CNO)	-0.072(PH_2O)
TPSS/def2-TZVPP	-0.001	0.012	0.024	0.092(CNO)	-0.099(PH_2O)
TPSS/def2-TZVPPD	-0.002	0.009	0.019	0.071(CNO)	-0.074(PH_2O)
B3LYP/6-31G(d)	-0.003	0.030	0.052	0.142(SSH)	-0.155(SH)
B3LYP/pc-1	0.005	0.023	0.044	0.157($SiCl_2H$)	-0.159(SH)
B3LYP/aug-pc-1	0.002	0.010	0.023	0.102(SiF_2H)	-0.070($SiClH_2$)
B3LYP/pc-2	0.001	0.013	0.024	0.091(SSH)	-0.092(SH)
B3LYP/aug-pc-2	-0.001	0.005	0.009	0.022(SHO_2)	-0.039(CNO)
B3LYP/def2-TZVPP	0.000	0.007	0.012	0.040(CCCl)	-0.040(OH)
B3LYP/def2-TZVPPD	-0.001	0.004	0.008	0.023(SiH)	-0.037(CHCO)
PBE0/6-31G(d)	-0.004	0.032	0.055	0.168(SSH)	-0.176(SH)
PBE0/pc-1	0.003	0.024	0.048	0.186(SSH)	-0.189(SH)
PBE0/aug-pc-1	0.000	0.010	0.019	0.084($SiFH_2$)	-0.043(SH)
PBE0/pc-2	-0.001	0.012	0.023	0.100(SSH)	-0.100(SH)
PBE0/aug-pc-2	-0.003	0.007	0.012	0.039(CFH_2)	-0.036(CF_2H)
PBE0/def2-TZVPP	-0.003	0.010	0.018	0.055(CCF)	-0.053(CF_2H)
PBE0/def2-TZVPPD	-0.003	0.008	0.012	0.042(CFH_2)	-0.039(CF_2H)

Table A6: Summary statistics for $||\boldsymbol{\mu}(R)||$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.016	0.069	0.084	0.187(SH)	-0.277(CH_2COO^- —ip)
SVWN/pc-1	0.007	0.041	0.057	0.173(SH)	-0.224(CH_2COO^- —ip)
SVWN/aug-pc-1	-0.012	0.031	0.036	0.108(CHNN)	-0.110(CNO)
SVWN/pc-2	-0.002	0.035	0.045	0.103(NNN)	-0.121(CNO)
SVWN/aug-pc-2	-0.024	0.039	0.042	0.092(CHNN)	-0.130(CNO)
SVWN/def2-TZVPP	-0.025	0.042	0.048	0.095(NNN)	-0.156(CH_2COO^- —ip)
SVWN/def2-TZVPPD	-0.025	0.040	0.042	0.085(CHNN)	-0.136(CNO)
BLYP/6-31G(d)	-0.020	0.064	0.077	0.147(NH_2)	-0.286(CH_2COO^- —ip)
BLYP/pc-1	0.001	0.039	0.053	0.129(SH)	-0.251(CH_2COO^- —ip)
BLYP/aug-pc-1	-0.001	0.018	0.024	0.076(CHNN)	-0.093(ONO)
BLYP/pc-2	0.007	0.023	0.029	0.075(SH)	-0.104(CH_2COO^- —ip)
BLYP/aug-pc-2	-0.010	0.018	0.023	0.062(CHNN)	-0.101(ONO)
BLYP/def2-TZVPP	-0.018	0.027	0.030	0.049(CHNN)	-0.150(CH_2COO^- —ip)
BLYP/def2-TZVPPD	-0.012	0.019	0.022	0.055(CHNN)	-0.095(ONO)
PBE/6-31G(d)	-0.023	0.067	0.081	0.164(SH)	-0.292(CH_2COO^- —ip)
PBE/pc-1	0.000	0.037	0.053	0.154(SH)	-0.247(CH_2COO^- —ip)
PBE/aug-pc-1	-0.015	0.023	0.025	0.069(CHNN)	-0.097(CNO)
PBE/pc-2	-0.007	0.024	0.033	0.080(SH)	-0.132(CH_2COO^- —ip)
PBE/aug-pc-2	-0.024	0.028	0.028	0.056(CHNN)	-0.113(CNO)
PBE/def2-TZVPP	-0.031	0.037	0.037	0.047(NNN)	-0.176(CH_2COO^- —ip)
PBE/def2-TZVPPD	-0.026	0.030	0.028	0.049(CHNN)	-0.119(CNO)
TPSS/6-31G(d)	-0.006	0.061	0.074	0.158(SH)	-0.257(CH_2COO^- —ip)
TPSS/pc-1	0.021	0.039	0.050	0.168(SH)	-0.200(CH_2COO^- —ip)
TPSS/aug-pc-1	-0.002	0.015	0.021	0.070(CHNN)	-0.062(ONO)
TPSS/pc-2	0.006	0.021	0.028	0.093(SH)	-0.104(CH_2COO^- —ip)
TPSS/aug-pc-2	-0.010	0.018	0.022	0.053(CHNN)	-0.087(CH_2COO^- —ip)
TPSS/def2-TZVPP	-0.016	0.024	0.029	0.042(NH_2)	-0.146(CH_2COO^- —ip)
TPSS/def2-TZVPPD	-0.013	0.018	0.021	0.045(CHNN)	-0.080(SHO)
B3LYP/6-31G(d)	0.005	0.053	0.066	0.164(NCS)	-0.204(CH_2COO^- —ip)
B3LYP/pc-1	0.029	0.041	0.048	0.178(SHO)	-0.148(CH_2COO^- —ip)
B3LYP/aug-pc-1	0.014	0.018	0.022	0.104(SHO)	-0.039(S^-)
B3LYP/pc-2	0.023	0.026	0.024	0.112(SHO)	-0.031(CH_2COO^- —ip)
B3LYP/aug-pc-2	0.004	0.010	0.014	0.052(CHCCO)	-0.025(SiH)
B3LYP/def2-TZVPP	-0.001	0.014	0.019	0.043(ONO)	-0.082(CH_2COO^- —ip)
B3LYP/def2-TZVPPD	0.002	0.009	0.013	0.046(CHCCO)	-0.026(SiH)
PBE0/6-31G(d)	0.006	0.055	0.069	0.178(NCS)	-0.196(CH_2COO^- —ip)
PBE0/pc-1	0.034	0.043	0.050	0.199(SHO)	-0.128(CH_2COO^- —ip)
PBE0/aug-pc-1	0.005	0.017	0.023	0.089(SHO)	-0.049(S^-)
PBE0/pc-2	0.013	0.022	0.027	0.099(SH)	-0.045(CH_2COO^- —ip)
PBE0/aug-pc-2	-0.006	0.016	0.020	0.072(ONO)	-0.039(CFH_2)
PBE0/def2-TZVPP	-0.010	0.021	0.026	0.093(ONO)	-0.096(CH_2COO^- —ip)
PBE0/def2-TZVPPD	-0.008	0.017	0.020	0.081(ONO)	-0.044(SiCl)

Table A7: Summary statistics for $\mu_X^C(R)$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.004	0.046	0.093	0.323(CH_2SH-op)	-0.410(SCF_3)
SVWN/pc-1	-0.010	0.046	0.089	0.270($PO(OH)_2$)	-0.536(CHCCO)
SVWN/aug-pc-1	-0.011	0.042	0.072	0.194($PO(OH)_2$)	-0.365(CHCCO)
SVWN/pc-2	-0.009	0.043	0.094	0.232($CH_2CSH-op$)	-0.694(CHCCO)
SVWN/aug-pc-2	-0.009	0.042	0.086	0.249($CH_2CSH-op$)	-0.545(CHCCO)
SVWN/def2-TZVPP	-0.009	0.048	0.092	0.184($CH_2CSH-op$)	-0.584(CHCCO)
SVWN/def2-TZVPPD	-0.010	0.048	0.092	0.200($CH_2CSH-op$)	-0.575(CHCCO)
BLYP/6-31G(d)	-0.002	0.045	0.083	0.277($CH_2COF-ip$)	-0.446(SCF_3)
BLYP/pc-1	-0.005	0.047	0.089	0.301($PO(OH)_2$)	-0.363(SCF_3)
BLYP/aug-pc-1	-0.007	0.038	0.067	0.204($PO(OH)_2$)	-0.260(SCF_3)
BLYP/pc-2	-0.002	0.042	0.075	0.181(ONO)	-0.271(SCF_3)
BLYP/aug-pc-2	-0.003	0.039	0.072	0.192($CH_2CSH-op$)	-0.278(SCF_3)
BLYP/def2-TZVPP	-0.002	0.040	0.075	0.181($COCH_3$)	-0.344(SCF_3)
BLYP/def2-TZVPPD	-0.003	0.039	0.071	0.164($COCH_3$)	-0.324(SCF_3)
PBE/6-31G(d)	-0.002	0.044	0.086	0.301($CH_2COF-ip$)	-0.446(SCF_3)
PBE/pc-1	-0.005	0.047	0.090	0.299($PO(OH)_2$)	-0.386(SCF_3)
PBE/aug-pc-1	-0.007	0.038	0.067	0.206($PO(OH)_2$)	-0.283(SCF_3)
PBE/pc-2	-0.003	0.035	0.070	0.173($CH_2CSH-op$)	-0.296(SCF_3)
PBE/aug-pc-2	-0.004	0.033	0.067	0.189($CH_2CSH-op$)	-0.301(SCF_3)
PBE/def2-TZVPP	-0.003	0.037	0.072	0.165($CH_2COCl-op$)	-0.371(SCF_3)
PBE/def2-TZVPPD	-0.004	0.036	0.068	0.144($CH_2CSH-op$)	-0.350(SCF_3)
TPSS/6-31G(d)	0.000	0.035	0.067	0.232($CH_2COF-ip$)	-0.339(SCF_3)
TPSS/pc-1	-0.003	0.045	0.083	0.310($PO(OH)_2$)	-0.279(SCF_3)
TPSS/aug-pc-1	-0.004	0.035	0.062	0.223($PO(OH)_2$)	-0.200(CH_2SH-op)
TPSS/pc-2	-0.001	0.034	0.059	0.156(CH_2SH-ip)	-0.207(SCF_3)
TPSS/aug-pc-2	-0.002	0.031	0.055	0.143(CH_2SH-ip)	-0.206(SCF_3)
TPSS/def2-TZVPP	-0.001	0.028	0.053	0.145($PO(OH)_2$)	-0.271(SCF_3)
TPSS/def2-TZVPPD	-0.002	0.026	0.048	0.134($PO(OH)_2$)	-0.251(SCF_3)
B3LYP/6-31G(d)	-0.003	0.032	0.061	0.229($CH_2COF-ip$)	-0.255(SCF_3)
B3LYP/pc-1	-0.006	0.038	0.071	0.217(CH_2SH-ip)	-0.278(CHCCO)
B3LYP/aug-pc-1	-0.007	0.028	0.051	0.202(CH_2SH-ip)	-0.202(CH_2SH-op)
B3LYP/pc-2	-0.004	0.028	0.053	0.155(CH_2SH-ip)	-0.323(CHCCO)
B3LYP/aug-pc-2	-0.004	0.024	0.046	0.135(CH_2SH-ip)	-0.265(CHCCO)
B3LYP/def2-TZVPP	-0.005	0.022	0.044	0.105($CH_2COF-ip$)	-0.259(CHCCO)
B3LYP/def2-TZVPPD	-0.005	0.021	0.040	0.079($CH_2COF-ip$)	-0.249(CHCCO)
PBE0/6-31G(d)	-0.004	0.033	0.067	0.249($CH_2COF-ip$)	-0.285(CH_2SH-ip)
PBE0/pc-1	-0.006	0.038	0.071	0.223($CH_2COF-ip$)	-0.312(CHCCO)
PBE0/aug-pc-1	-0.007	0.029	0.050	0.185(CH_2SH-ip)	-0.193(CH_2SH-op)
PBE0/pc-2	-0.005	0.022	0.048	0.148($CH_2COF-op$)	-0.319(CHCCO)
PBE0/aug-pc-2	-0.005	0.020	0.040	0.131($CH_2COF-op$)	-0.237(CHCCO)
PBE0/def2-TZVPP	-0.006	0.019	0.041	0.149($CH_2COF-ip$)	-0.226(CHCCO)
PBE0/def2-TZVPPD	-0.006	0.020	0.038	0.125($CH_2COF-ip$)	-0.215(CHCCO)

Table A8: Summary statistics for $\mu_y^c(R)$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.012	0.071	0.109	0.339(OPH_2O)	-0.439(SiCl)
SVWN/pc-1	-0.014	0.075	0.118	0.419(NCS)	-0.448(SiCl)
SVWN/aug-pc-1	-0.011	0.061	0.100	0.354(NCS)	-0.355(SiCl)
SVWN/pc-2	-0.009	0.074	0.121	0.438(NCS)	-0.325(SiCl)
SVWN/aug-pc-2	-0.006	0.074	0.121	0.442(CH_2CSH-ip)	-0.274(SiCl)
SVWN/def2-TZVPP	-0.014	0.072	0.108	0.385(SCN)	-0.260(SiCl)
SVWN/def2-TZVPPD	-0.012	0.071	0.108	0.392(SCN)	-0.249(SiCl)
BLYP/6-31G(d)	-0.004	0.069	0.116	0.404(SHO)	-0.438(SiCl)
BLYP/pc-1	-0.010	0.090	0.153	0.572(SHO)	-0.553(CHCCO)
BLYP/aug-pc-1	-0.006	0.073	0.126	0.465(SHO)	-0.572(CHCCO)
BLYP/pc-2	-0.003	0.081	0.134	0.413(SCN)	-0.446(CHCCO)
BLYP/aug-pc-2	0.000	0.080	0.131	0.421(SCN)	-0.441(CHCCO)
BLYP/def2-TZVPP	-0.006	0.076	0.127	0.498(SCN)	-0.457(CHCCO)
BLYP/def2-TZVPPD	-0.004	0.075	0.124	0.502(SCN)	-0.438(CHCCO)
PBE/6-31G(d)	-0.007	0.065	0.105	0.342(SHO)	-0.420(SiCl)
PBE/pc-1	-0.011	0.086	0.145	0.522(SHO)	-0.532(CHCCO)
PBE/aug-pc-1	-0.008	0.067	0.116	0.420(SHO)	-0.554(CHCCO)
PBE/pc-2	-0.005	0.070	0.118	0.362(SCN)	-0.397(CHCCO)
PBE/aug-pc-2	-0.003	0.068	0.115	0.362(SCN)	-0.401(CHCCO)
PBE/def2-TZVPP	-0.009	0.067	0.108	0.395(SCN)	-0.410(CHCCO)
PBE/def2-TZVPPD	-0.008	0.065	0.105	0.397(SCN)	-0.392(CHCCO)
TPSS/6-31G(d)	-0.004	0.050	0.086	0.317(SHO)	-0.368(SiCl)
TPSS/pc-1	-0.006	0.079	0.134	0.517(SHO)	-0.540(CHCCO)
TPSS/aug-pc-1	-0.003	0.059	0.103	0.420(SHO)	-0.536(CHCCO)
TPSS/pc-2	0.001	0.058	0.094	0.322(SHO)	-0.315(CHCCO)
TPSS/aug-pc-2	0.003	0.055	0.090	0.306(SHO)	-0.316(CHCCO)
TPSS/def2-TZVPP	-0.005	0.051	0.084	0.272(SHO)	-0.335(CHCCO)
TPSS/def2-TZVPPD	-0.004	0.049	0.080	0.261(SHO)	-0.317(CHCCO)
B3LYP/6-31G(d)	-0.007	0.044	0.068	0.193(CHCCO)	-0.304(SiCl)
B3LYP/pc-1	-0.010	0.055	0.094	0.373(SHO)	-0.314(SiCl)
B3LYP/aug-pc-1	-0.007	0.039	0.069	0.283(SHO)	-0.280(CHCCO)
B3LYP/pc-2	-0.004	0.042	0.065	0.193(SHO)	-0.175(SiCl)
B3LYP/aug-pc-2	-0.002	0.038	0.060	0.175(SHO)	-0.134(SiCl)
B3LYP/def2-TZVPP	-0.011	0.033	0.050	0.133(CH_2COO^- —ip)	-0.131(NCO)
B3LYP/def2-TZVPPD	-0.009	0.031	0.046	0.119(SHO)	-0.120(NCO)
PBE0/6-31G(d)	-0.009	0.048	0.079	0.368(CHCCO)	-0.268(SiCl)
PBE0/pc-1	-0.010	0.048	0.081	0.307(SHO)	-0.281(SiCl)
PBE0/aug-pc-1	-0.008	0.034	0.056	0.220(SHO)	-0.228(CHCCO)
PBE0/pc-2	-0.005	0.027	0.044	0.113(SHO)	-0.137(SiCl)
PBE0/aug-pc-2	-0.004	0.024	0.039	0.108($CSCH_3$)	-0.115(NCO)
PBE0/def2-TZVPP	-0.012	0.022	0.032	0.090(CH_2COO^- —ip)	-0.122(NCO)
PBE0/def2-TZVPPD	-0.011	0.021	0.030	0.051(Ocl)	-0.111(NCO)

Table A9: Summary statistics for $\mu_{c,z}(R)$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.015	0.058	0.120	0.438(PHO(OH))	-0.651(PH_2)
SVWN/pc-1	0.000	0.058	0.117	0.529(CCF)	-0.360(PH_2)
SVWN/aug-pc-1	0.003	0.057	0.110	0.546(CCF)	-0.336(PH_2)
SVWN/pc-2	0.001	0.053	0.108	0.442(CHCO)	-0.393(CSH)
SVWN/aug-pc-2	0.002	0.052	0.107	0.434(CHCO)	-0.415(CSH)
SVWN/def2-TZVPP	0.007	0.057	0.123	0.766(CCF)	-0.356(PH_2)
SVWN/def2-TZVPPD	0.006	0.058	0.122	0.757(CCF)	-0.357(PH_2)
BLYP/6-31G(d)	-0.013	0.054	0.103	0.316(PHO(OH))	-0.423(PH_2)
BLYP/pc-1	0.004	0.062	0.134	0.552(CHCO)	-0.525(SHO_2)
BLYP/aug-pc-1	0.007	0.052	0.116	0.539(CCF)	-0.384(SHO_2)
BLYP/pc-2	0.007	0.056	0.115	0.477(CHCO)	-0.402(SHO_2)
BLYP/aug-pc-2	0.008	0.052	0.111	0.469(CHCO)	-0.355(CSH)
BLYP/def2-TZVPP	0.012	0.056	0.128	0.805(CCF)	-0.278(SHO_2)
BLYP/def2-TZVPPD	0.011	0.053	0.124	0.793(CCF)	-0.269(CSH)
PBE/6-31G(d)	-0.015	0.054	0.103	0.343(PHO(OH))	-0.467(PH_2)
PBE/pc-1	0.005	0.061	0.130	0.563(CHCO)	-0.475(SHO_2)
PBE/aug-pc-1	0.008	0.053	0.112	0.534(CHCO)	-0.336(SHO_2)
PBE/pc-2	0.005	0.046	0.102	0.435(CHCO)	-0.337(SHO_2)
PBE/aug-pc-2	0.005	0.043	0.096	0.418(CHCO)	-0.329(CSH)
PBE/def2-TZVPP	0.009	0.050	0.112	0.718(CCF)	-0.213(SHO_2)
PBE/def2-TZVPPD	0.009	0.048	0.109	0.709(CCF)	-0.225(CSH)
TPSS/6-31G(d)	-0.017	0.046	0.087	0.243(PHO(OH))	-0.363(CCCl)
TPSS/pc-1	0.005	0.056	0.120	0.490(CHCO)	-0.535(SHO_2)
TPSS/aug-pc-1	0.007	0.046	0.098	0.455(CHCO)	-0.404(SHO_2)
TPSS/pc-2	0.002	0.041	0.081	0.305(CHCO)	-0.384(SHO_2)
TPSS/aug-pc-2	0.002	0.036	0.072	0.280(CHCO)	-0.337(SHO_2)
TPSS/def2-TZVPP	0.007	0.039	0.083	0.479(CCF)	-0.236(SHO_2)
TPSS/def2-TZVPPD	0.006	0.035	0.078	0.474(CCF)	-0.210(SHO_2)
B3LYP/6-31G(d)	-0.022	0.042	0.084	0.239(SSH)	-0.471(CCCl)
B3LYP/pc-1	-0.003	0.042	0.083	0.301(CHCO)	-0.380(SHO_2)
B3LYP/aug-pc-1	-0.001	0.034	0.068	0.280(CHCO)	-0.250(SHO_2)
B3LYP/pc-2	-0.003	0.029	0.056	0.189(CHCO)	-0.264(SHO_2)
B3LYP/aug-pc-2	-0.002	0.025	0.048	0.171(CHCO)	-0.208(SHO_2)
B3LYP/def2-TZVPP	0.004	0.025	0.062	0.401(CCF)	-0.103(CF_2H)
B3LYP/def2-TZVPPD	0.003	0.023	0.058	0.402(CCF)	-0.094(PH_2)
PBE0/6-31G(d)	-0.026	0.047	0.097	0.280(SSH)	-0.618(CCCl)
PBE0/pc-1	-0.005	0.038	0.073	0.263(CHCO)	-0.335(SHO_2)
PBE0/aug-pc-1	-0.004	0.033	0.062	0.223(CHCO)	-0.201(SHO_2)
PBE0/pc-2	-0.007	0.021	0.042	0.113(SH)	-0.183(SHO_2)
PBE0/aug-pc-2	-0.008	0.018	0.036	0.096(SH)	-0.146(CCCl)
PBE0/def2-TZVPP	-0.002	0.017	0.035	0.199(CCCl)	-0.081(PH_2)
PBE0/def2-TZVPPD	-0.003	0.016	0.033	0.192(CCF)	-0.082(PH_2)

Table A10: Summary statistics for $\mu_z^x(R)$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	0.001	0.047	0.080	0.316(SHO)	-0.269(CHCCO)
SVWN/pc-1	0.009	0.038	0.070	0.399(CHCCO)	-0.181(CH_2COF—ip)
SVWN/aug-pc-1	0.010	0.037	0.060	0.257(CHCCO)	-0.124(CH_2CSH—op)
SVWN/pc-2	0.005	0.037	0.074	0.458(CHCCO)	-0.191(CH_2CSH—op)
SVWN/aug-pc-2	0.007	0.037	0.071	0.355(CHCCO)	-0.222(CH_2CSH—op)
SVWN/def2-TZVPP	0.006	0.038	0.074	0.382(CHCCO)	-0.166(CH_2CSH—op)
SVWN/def2-TZVPPD	0.007	0.039	0.075	0.370(CHCCO)	-0.190(CH_2SH—op)
BLYP/6-31G(d)	0.004	0.044	0.069	0.235(SHO)	-0.148(CH_2SH—op)
BLYP/pc-1	0.009	0.049	0.087	0.341(SCF_3)	-0.239(CH_2COF—ip)
BLYP/aug-pc-1	0.010	0.039	0.068	0.273(SCF_3)	-0.180(CH_2COF—op)
BLYP/pc-2	0.004	0.046	0.080	0.284(SCF_3)	-0.206(CH_2COCl—op)
BLYP/aug-pc-2	0.006	0.038	0.072	0.283(SCF_3)	-0.177(CH_2CSH—op)
BLYP/def2-TZVPP	0.005	0.040	0.072	0.338(SCF_3)	-0.171(CH_2COCl—op)
BLYP/def2-TZVPPD	0.006	0.039	0.070	0.330(SCF_3)	-0.159(CH_2COCl—op)
PBE/6-31G(d)	0.003	0.042	0.067	0.252(SHO)	-0.171(CH_2SH—op)
PBE/pc-1	0.008	0.046	0.083	0.318(SCF_3)	-0.237(CH_2COF—ip)
PBE/aug-pc-1	0.009	0.037	0.062	0.241(SCF_3)	-0.175(CH_2COF—ip)
PBE/pc-2	0.004	0.035	0.065	0.252(SCF_3)	-0.186(CH_2COCl—ip)
PBE/aug-pc-2	0.006	0.029	0.059	0.249(SCF_3)	-0.167(CH_2CSH—op)
PBE/def2-TZVPP	0.004	0.032	0.060	0.308(SCF_3)	-0.148(CH_2COCl—op)
PBE/def2-TZVPPD	0.005	0.031	0.060	0.300(SCF_3)	-0.137(CH_2COF—ip)
TPSS/6-31G(d)	0.000	0.035	0.057	0.192(SHO)	-0.143(Ocl)
TPSS/pc-1	0.005	0.048	0.087	0.291(CH_2SH—op)	-0.294(CH_2SH—ip)
TPSS/aug-pc-1	0.006	0.035	0.061	0.203(CH_2SH—op)	-0.207(CH_2SH—ip)
TPSS/pc-2	0.001	0.035	0.060	0.182(SCF_3)	-0.194(CH_2SH—ip)
TPSS/aug-pc-2	0.003	0.027	0.049	0.175(SCF_3)	-0.141(CH_2SH—ip)
TPSS/def2-TZVPP	0.002	0.024	0.044	0.231(SCF_3)	-0.108(CH_2COF—op)
TPSS/def2-TZVPPD	0.003	0.021	0.041	0.223(SCF_3)	-0.099(CH_2COF—op)
B3LYP/6-31G(d)	0.002	0.038	0.062	0.200(SHO)	-0.152(CH_2COH—ip)
B3LYP/pc-1	0.006	0.042	0.079	0.291(CH_2SH—op)	-0.280(CH_2SH—ip)
B3LYP/aug-pc-1	0.007	0.030	0.053	0.207(CH_2SH—op)	-0.199(CH_2SH—ip)
B3LYP/pc-2	0.002	0.034	0.059	0.259(CHCCO)	-0.190(CH_2SH—ip)
B3LYP/aug-pc-2	0.004	0.024	0.044	0.200(CHCCO)	-0.129(CH_2SH—ip)
B3LYP/def2-TZVPP	0.003	0.022	0.041	0.197(CHCCO)	-0.097(CH_2COF—ip)
B3LYP/def2-TZVPPD	0.004	0.019	0.037	0.187(CHCCO)	-0.088(CH_2COF—ip)
PBE0/6-31G(d)	0.001	0.039	0.064	0.209(SHO)	-0.187(CH_2COH—ip)
PBE0/pc-1	0.005	0.040	0.078	0.303(CH_2SH—op)	-0.299(CH_2SH—ip)
PBE0/aug-pc-1	0.006	0.028	0.050	0.208(CH_2SH—op)	-0.206(CH_2SH—ip)
PBE0/pc-2	0.001	0.022	0.044	0.224(CHCCO)	-0.165(CH_2SH—ip)
PBE0/aug-pc-2	0.003	0.016	0.030	0.150(CHCCO)	-0.104(CH_2SH—ip)
PBE0/def2-TZVPP	0.003	0.013	0.027	0.142(CHCCO)	-0.076(CH_2COF—ip)
PBE0/def2-TZVPPD	0.004	0.013	0.025	0.131(CHCCO)	-0.066(CH_2COF—ip)

Table A11: Summary statistics for $\mu_z^y(R)$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.001	0.063	0.112	0.530(CHCCO)	-0.462(OPH_2O)
SVWN/pc-1	0.003	0.062	0.114	0.501(CHCCO)	-0.498(SHO)
SVWN/aug-pc-1	-0.004	0.047	0.094	0.511(CHCCO)	-0.355(NCS)
SVWN/pc-2	-0.007	0.060	0.115	0.479(CHCCO)	-0.403(NCS)
SVWN/aug-pc-2	-0.012	0.057	0.113	0.473(CHCCO)	-0.425(NCS)
SVWN/def2-TZVPP	-0.007	0.054	0.100	0.479(CHCCO)	-0.417(SCN)
SVWN/def2-TZVPPD	-0.006	0.053	0.101	0.485(CHCCO)	-0.423(SCN)
BLYP/6-31G(d)	0.008	0.058	0.103	0.500(SiCl)	-0.308(SHO)
BLYP/pc-1	0.016	0.088	0.151	0.563(SiCl)	-0.643(SHO)
BLYP/aug-pc-1	0.010	0.071	0.127	0.575(CHCCO)	-0.502(SHO)
BLYP/pc-2	0.007	0.079	0.135	0.510(CHCCO)	-0.444(SHO)
BLYP/aug-pc-2	0.003	0.073	0.128	0.509(CHCCO)	-0.434(SCN)
BLYP/def2-TZVPP	0.005	0.068	0.119	0.498(CHCCO)	-0.507(SCN)
BLYP/def2-TZVPPD	0.007	0.069	0.119	0.505(CHCCO)	-0.509(SCN)
PBE/6-31G(d)	0.004	0.053	0.092	0.428(SiCl)	-0.319(OPH_2O)
PBE/pc-1	0.011	0.079	0.138	0.555(CHCCO)	-0.586(SHO)
PBE/aug-pc-1	0.004	0.058	0.108	0.558(CHCCO)	-0.420(SHO)
PBE/pc-2	0.000	0.063	0.110	0.464(CHCCO)	-0.360(SCN)
PBE/aug-pc-2	-0.003	0.057	0.103	0.457(CHCCO)	-0.361(SCN)
PBE/def2-TZVPP	-0.001	0.051	0.092	0.441(CHCCO)	-0.390(SCN)
PBE/def2-TZVPPD	0.001	0.052	0.092	0.447(CHCCO)	-0.392(SCN)
TPSS/6-31G(d)	0.010	0.044	0.079	0.393(SiCl)	-0.252(SHO)
TPSS/pc-1	0.015	0.079	0.136	0.516(CHCCO)	-0.624(SHO)
TPSS/aug-pc-1	0.006	0.054	0.099	0.507(CHCCO)	-0.444(SHO)
TPSS/pc-2	0.002	0.053	0.090	0.363(CHCCO)	-0.356(SHO)
TPSS/aug-pc-2	-0.001	0.046	0.079	0.350(CHCCO)	-0.247(SHO)
TPSS/def2-TZVPP	0.003	0.038	0.067	0.329(CHCCO)	-0.219(SHO)
TPSS/def2-TZVPPD	0.005	0.038	0.067	0.332(CHCCO)	-0.204(SCN)
B3LYP/6-31G(d)	0.010	0.047	0.084	0.405(SiCl)	-0.225(NCH_2)
B3LYP/pc-1	0.016	0.064	0.112	0.468(SiCl)	-0.560(SHO)
B3LYP/aug-pc-1	0.008	0.043	0.079	0.365(CHCCO)	-0.389(SHO)
B3LYP/pc-2	0.006	0.045	0.074	0.254(SiCl)	-0.310(SHO)
B3LYP/aug-pc-2	0.002	0.037	0.061	0.237(CHCCO)	-0.193(SHO)
B3LYP/def2-TZVPP	0.007	0.030	0.048	0.235(CHCCO)	-0.158(SHO)
B3LYP/def2-TZVPPD	0.008	0.029	0.047	0.240(CHCCO)	-0.119(SHO)
PBE0/6-31G(d)	0.007	0.054	0.093	0.331(SiCl)	-0.263(NCH_2)
PBE0/pc-1	0.012	0.053	0.098	0.398(SiCl)	-0.507(SHO)
PBE0/aug-pc-1	0.003	0.031	0.059	0.303(CHCCO)	-0.305(SHO)
PBE0/pc-2	0.000	0.024	0.042	0.155(SiCl)	-0.203(SHO)
PBE0/aug-pc-2	-0.003	0.018	0.030	0.109(CHCCO)	-0.090(OPH_2O)
PBE0/def2-TZVPP	0.002	0.012	0.021	0.096(CHCCO)	-0.048(SHO)
PBE0/def2-TZVPPD	0.003	0.011	0.020	0.099(CHCCO)	-0.040(ONO)

Table A12: Summary statistics for $\mu_z^p(R)$

Model Chemistry	MD	MAD	SD	Max$^+$	Max$^-$
SVWN/6-31G(d)	0.009	0.050	0.095	0.501(PH_2)	-0.307(PH_2O)
SVWN/pc-1	0.000	0.050	0.108	0.332(SHO_2)	-0.511(CHCO)
SVWN/aug-pc-1	-0.004	0.049	0.101	0.343(PH_2)	-0.506(CHCO)
SVWN/pc-2	-0.004	0.042	0.094	0.425(CSH)	-0.476(CHCO)
SVWN/aug-pc-2	-0.006	0.043	0.100	0.458(CSH)	-0.469(CHCO)
SVWN/def2-TZVPP	-0.011	0.045	0.109	0.356(PH_2)	-0.664(CCF)
SVWN/def2-TZVPPD	-0.011	0.048	0.113	0.384(PH_2)	-0.673(CCF)
BLYP/6-31G(d)	0.010	0.049	0.091	0.380($SiCl_2H$)	-0.324($SiClH_2$)
BLYP/pc-1	0.002	0.064	0.139	0.540(SHO_2)	-0.566(CHCO)
BLYP/aug-pc-1	-0.003	0.050	0.116	0.404(SHO_2)	-0.573(CHCO)
BLYP/pc-2	-0.004	0.057	0.117	0.415(SHO_2)	-0.505(CHCO)
BLYP/aug-pc-2	-0.007	0.049	0.107	0.374(CSH)	-0.500(CHCO)
BLYP/def2-TZVPP	-0.011	0.052	0.119	0.294(CSH)	-0.761(CCF)
BLYP/def2-TZVPPD	-0.011	0.050	0.119	0.292(CSH)	-0.770(CCF)
PBE/6-31G(d)	0.011	0.047	0.085	0.333($SiCl_2H$)	-0.289($SiClH_2$)
PBE/pc-1	-0.001	0.059	0.129	0.484(SHO_2)	-0.570(CHCO)
PBE/aug-pc-1	-0.006	0.045	0.103	0.323(SHO_2)	-0.553(CHCO)
PBE/pc-2	-0.005	0.044	0.092	0.329(CSH)	-0.445(CHCO)
PBE/aug-pc-2	-0.007	0.035	0.085	0.352(CSH)	-0.431(CHCO)
PBE/def2-TZVPP	-0.011	0.040	0.095	0.255(CSH)	-0.632(CCF)
PBE/def2-TZVPPD	-0.010	0.039	0.096	0.253(CSH)	-0.644(CCF)
TPSS/6-31G(d)	0.014	0.042	0.083	0.360($SiCl_2H$)	-0.307($SiClH_2$)
TPSS/pc-1	0.000	0.061	0.135	0.585(SHO_2)	-0.515(CHCO)
TPSS/aug-pc-1	-0.005	0.041	0.096	0.413(SHO_2)	-0.481(CHCO)
TPSS/pc-2	-0.002	0.041	0.084	0.387(SHO_2)	-0.320(CHCO)
TPSS/aug-pc-2	-0.004	0.029	0.064	0.292(SHO_2)	-0.300(CHCO)
TPSS/def2-TZVPP	-0.008	0.031	0.070	0.164(SHO_2)	-0.421(CCF)
TPSS/def2-TZVPPD	-0.008	0.028	0.067	0.141(SHO_2)	-0.435(CCF)
B3LYP/6-31G(d)	0.019	0.044	0.086	0.421(CCCl)	-0.301($SiClH_2$)
B3LYP/pc-1	0.009	0.050	0.110	0.495(SHO_2)	-0.391(SH)
B3LYP/aug-pc-1	0.004	0.035	0.074	0.336(SHO_2)	-0.324(CHCO)
B3LYP/pc-2	0.004	0.035	0.072	0.334(SHO_2)	-0.256(SH)
B3LYP/aug-pc-2	0.001	0.024	0.051	0.230(SHO_2)	-0.208(CHCO)
B3LYP/def2-TZVPP	-0.004	0.026	0.060	0.103(SiH)	-0.378(CCF)
B3LYP/def2-TZVPPD	-0.004	0.023	0.058	0.108(PH_2)	-0.396(CCF)
PBE0/6-31G(d)	0.021	0.049	0.097	0.574(CCCl)	-0.274($SiClH_2$)
PBE0/pc-1	0.009	0.043	0.099	0.458(SHO_2)	-0.404(SH)
PBE0/aug-pc-1	0.005	0.030	0.061	0.269(SHO_2)	-0.248(CHCO)
PBE0/pc-2	0.006	0.025	0.053	0.236(SHO_2)	-0.213(SH)
PBE0/aug-pc-2	0.005	0.018	0.037	0.167(CCCl)	-0.118(SH)
PBE0/def2-TZVPP	-0.001	0.013	0.028	0.075(CNO)	-0.152(CCCl)
PBE0/def2-TZVPPD	-0.001	0.012	0.028	0.079(PH_2)	-0.157(CCCl)

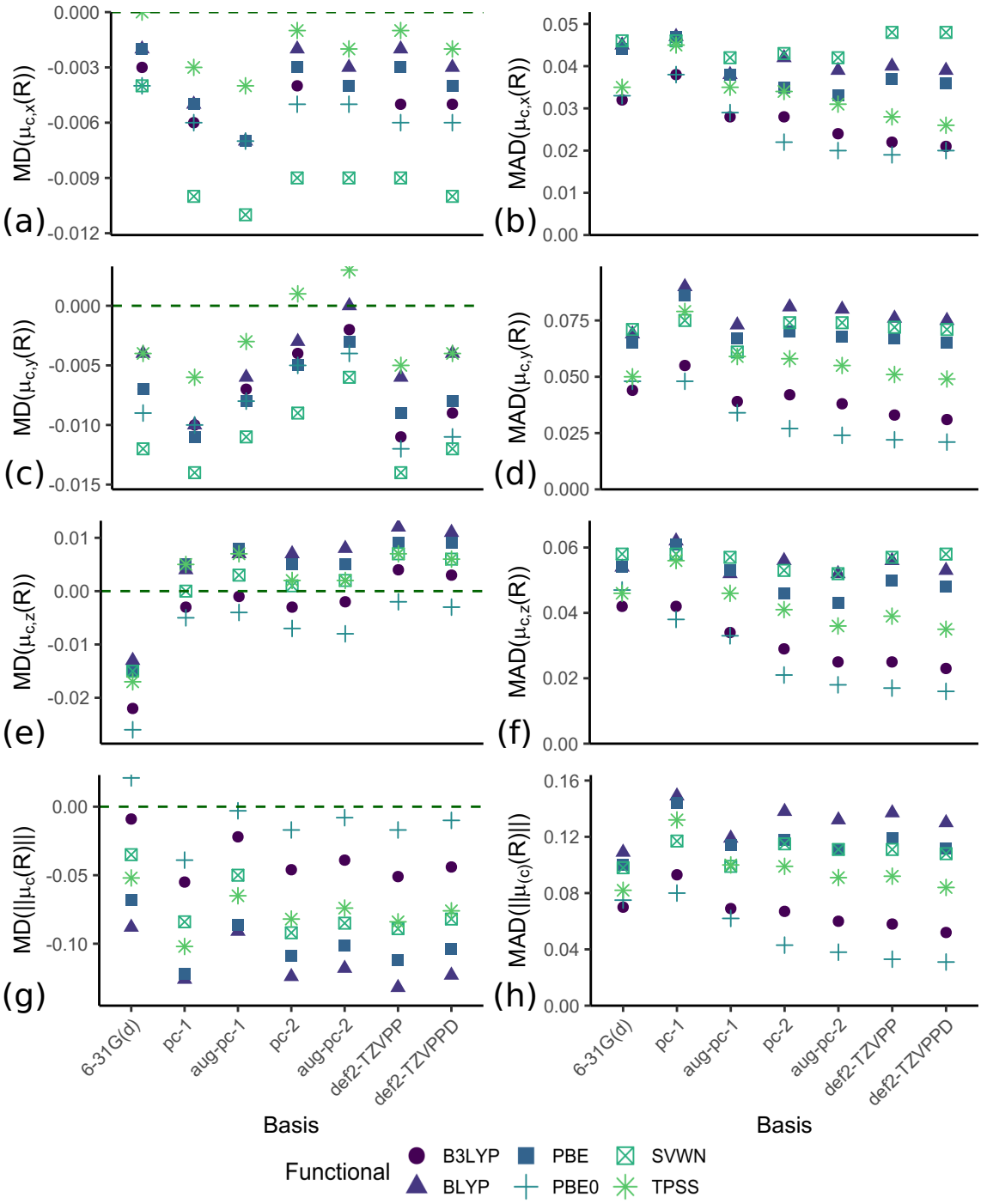

Figure A3: (a) MD(μ_x^c(R)) (b) MAD(μ_x^c(R)) (c) MD(μ_y^c(R)) (d) MAD(μ_y^c(R)) (e) MD(μ_z^c(R)) (f) MAD(μ_z^c(R)) (g) MD($||\mu^c(R)||$) (h) MAD($||\mu^c(R)||$) of tested model chemistries compared to reference B2PLYPD3-BJ/aug-cc-pV5Z values

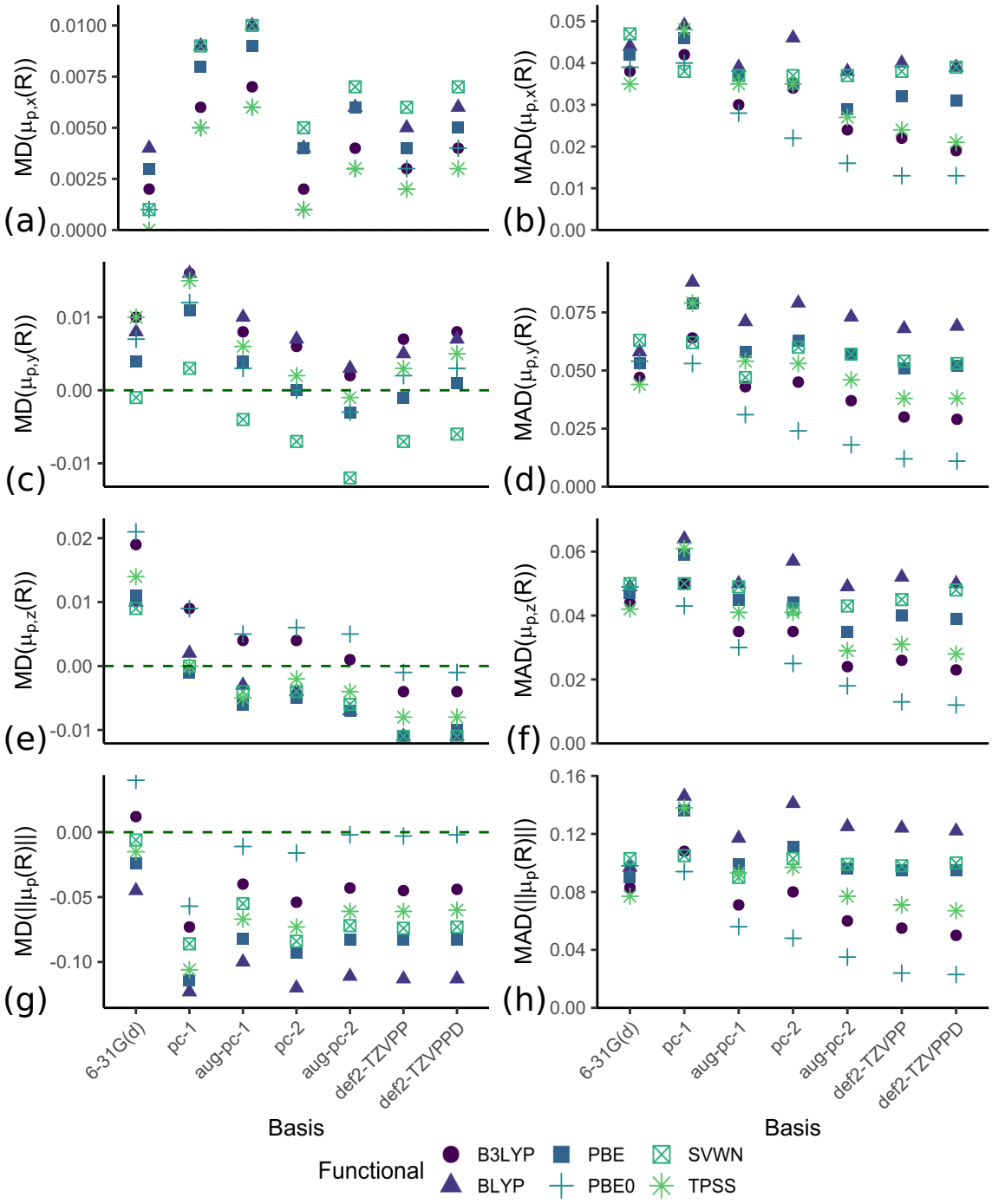

Figure A4: (a) MD(μ_x^p(R)) (b) MAD(μ_x^p(R)) (c) MD(μ_y^p(R)) (d) MAD(μ_y^p(R)) (e) MD(μ_z^p(R)) (f) MAD(μ_z^p(R)) (g) MD($||\mu^p(R)||$) (h) MAD($||\mu^p(R)||$) of tested model chemistries compared to reference B2PLYPD3-BJ/aug-cc-pV5Z values

Table A13: Linear regression parameters for μ_x(R)

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.9718	-0.0055	0.9762	0.0709
SVWN/pc-1	1.0075	-0.0011	0.9906	0.0457
SVWN/aug-pc-1	0.9925	-0.0022	0.997	0.0255
SVWN/pc-2	1.004	-0.0032	0.994	0.0365
SVWN/aug-pc-2	0.9822	-0.0033	0.9961	0.0286
SVWN/def2-TZVPP	0.9821	-0.0043	0.9947	0.0335
SVWN/def2-TZVPPD	0.9811	-0.0039	0.9959	0.0294
BLYP/6-31G(d)	0.9607	-0.001	0.9837	0.0578
BLYP/pc-1	0.992	0.0038	0.9923	0.0407
BLYP/aug-pc-1	0.9939	0.003	0.9968	0.0263
BLYP/pc-2	1.0075	0.0028	0.9966	0.0274
BLYP/aug-pc-2	0.9894	0.0024	0.9975	0.0233
BLYP/def2-TZVPP	0.9823	0.001	0.9975	0.023
BLYP/def2-TZVPPD	0.9884	0.0018	0.9977	0.0222
PBE/6-31G(d)	0.9618	-0.0027	0.9814	0.0618
PBE/pc-1	0.9946	0.0018	0.9927	0.0398
PBE/aug-pc-1	0.9819	5e-04	0.9982	0.0194
PBE/pc-2	0.9948	1e-04	0.9976	0.023
PBE/aug-pc-2	0.977	-2e-04	0.9987	0.0167
PBE/def2-TZVPP	0.9705	-0.0016	0.998	0.0203
PBE/def2-TZVPPD	0.9758	-9e-04	0.9988	0.0161
TPSS/6-31G(d)	0.9823	-0.0014	0.9849	0.0568
TPSS/pc-1	1.0185	0.0037	0.9934	0.0386
TPSS/aug-pc-1	0.9957	0.0019	0.9982	0.02
TPSS/pc-2	1.0087	0.001	0.9979	0.0218
TPSS/aug-pc-2	0.9915	8e-04	0.9988	0.0163
TPSS/def2-TZVPP	0.9865	-6e-04	0.9985	0.0178
TPSS/def2-TZVPPD	0.9896	2e-04	0.999	0.0147
B3LYP/6-31G(d)	0.9886	-0.0025	0.9874	0.0523
B3LYP/pc-1	1.0225	0.0022	0.9952	0.0333
B3LYP/aug-pc-1	1.0093	6e-04	0.9992	0.0132
B3LYP/pc-2	1.0232	2e-04	0.9987	0.0169
B3LYP/aug-pc-2	1.0033	-3e-04	0.9995	0.0104
B3LYP/def2-TZVPP	1	-0.0014	0.999	0.0148
B3LYP/def2-TZVPPD	1.0026	-9e-04	0.9995	0.01
PBE0/6-31G(d)	0.9931	-0.0041	0.9852	0.0568
PBE0/pc-1	1.0297	5e-04	0.9946	0.0353
PBE0/aug-pc-1	1.0009	-0.0016	0.9987	0.0171
PBE0/pc-2	1.0136	-0.0024	0.9977	0.0228
PBE0/aug-pc-2	0.9942	-0.0028	0.9985	0.0178
PBE0/def2-TZVPP	0.9914	-0.0039	0.9978	0.0219
PBE0/def2-TZVPPD	0.9931	-0.0034	0.9985	0.0181

Table A14: Linear regression parameters for μ_y(R)

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.9407	-0.0106	0.9815	0.0691
SVWN/pc-1	0.9803	-0.0101	0.9925	0.0455
SVWN/aug-pc-1	0.9906	-0.0149	0.9956	0.035
SVWN/pc-2	0.994	-0.0158	0.9926	0.0457
SVWN/aug-pc-2	0.9772	-0.0173	0.9931	0.0436
SVWN/def2-TZVPP	0.9683	-0.0193	0.9921	0.0462
SVWN/def2-TZVPPD	0.9766	-0.0176	0.9927	0.0447
BLYP/6-31G(d)	0.9309	0.0072	0.9858	0.0596
BLYP/pc-1	0.9646	0.0077	0.9932	0.0427
BLYP/aug-pc-1	0.9933	0.0051	0.9967	0.0306
BLYP/pc-2	0.9935	0.0041	0.9972	0.0279
BLYP/aug-pc-2	0.9838	0.0037	0.9981	0.0232
BLYP/def2-TZVPP	0.9665	-3e-04	0.9979	0.0235
BLYP/def2-TZVPPD	0.9834	0.0029	0.9982	0.0225
PBE/6-31G(d)	0.9262	4e-04	0.9868	0.0572
PBE/pc-1	0.9643	0.0019	0.9953	0.0354
PBE/aug-pc-1	0.9786	-0.0029	0.9981	0.0229
PBE/pc-2	0.9784	-0.0041	0.9982	0.0225
PBE/aug-pc-2	0.9682	-0.0045	0.9984	0.0209
PBE/def2-TZVPP	0.9523	-0.008	0.9977	0.0242
PBE/def2-TZVPPD	0.9675	-0.0054	0.9983	0.0214
TPSS/6-31G(d)	0.9458	0.0079	0.9882	0.0552
TPSS/pc-1	0.9864	0.0098	0.9946	0.0387
TPSS/aug-pc-1	0.9904	0.0042	0.998	0.0239
TPSS/pc-2	0.991	0.0033	0.9981	0.0229
TPSS/aug-pc-2	0.9799	0.0028	0.9986	0.0197
TPSS/def2-TZVPP	0.9664	-4e-04	0.9983	0.0215
TPSS/def2-TZVPPD	0.9778	0.0016	0.9987	0.0188
B3LYP/6-31G(d)	0.9677	0.0047	0.9913	0.0484
B3LYP/pc-1	1.0077	0.0054	0.9946	0.0395
B3LYP/aug-pc-1	1.0185	7e-04	0.9984	0.0219
B3LYP/pc-2	1.0212	5e-04	0.9979	0.025
B3LYP/aug-pc-2	1.008	-6e-04	0.9989	0.0182
B3LYP/def2-TZVPP	0.9946	-0.0035	0.9986	0.0198
B3LYP/def2-TZVPPD	1.006	-0.0014	0.999	0.0171
PBE0/6-31G(d)	0.9705	-1e-04	0.9916	0.0478
PBE0/pc-1	1.0152	0.0016	0.9953	0.0371
PBE0/aug-pc-1	1.0087	-0.0053	0.9987	0.0198
PBE0/pc-2	1.0116	-0.0058	0.9976	0.0263
PBE0/aug-pc-2	0.9975	-0.0068	0.9981	0.0234
PBE0/def2-TZVPP	0.9858	-0.0094	0.9978	0.0249
PBE0/def2-TZVPPD	0.9949	-0.0077	0.9981	0.0233

Table A15: Linear regression parameters for μ_z(R)

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.9857	-0.0064	0.9843	0.0669
SVWN/pc-1	1.0141	1e-04	0.9937	0.0434
SVWN/aug-pc-1	0.9866	0	0.9973	0.0275
SVWN/pc-2	0.9974	-0.0027	0.9957	0.0353
SVWN/aug-pc-2	0.9638	-0.0029	0.9967	0.0297
SVWN/def2-TZVPP	0.9649	-0.003	0.9954	0.0353
SVWN/def2-TZVPPD	0.9639	-0.0032	0.9966	0.0304
BLYP/6-31G(d)	0.9825	-0.0018	0.9879	0.0584
BLYP/pc-1	1.0078	0.0059	0.9952	0.0378
BLYP/aug-pc-1	1.0026	0.0038	0.9985	0.0208
BLYP/pc-2	1.0103	0.0022	0.9982	0.023
BLYP/aug-pc-2	0.9854	9e-04	0.9991	0.016
BLYP/def2-TZVPP	0.975	0.0019	0.9986	0.0199
BLYP/def2-TZVPPD	0.9836	8e-04	0.9991	0.0162
PBE/6-31G(d)	0.9795	-0.0033	0.9857	0.0635
PBE/pc-1	1.0069	0.0038	0.9943	0.0409
PBE/aug-pc-1	0.9846	0.0027	0.9984	0.0214
PBE/pc-2	0.9928	5e-04	0.9973	0.0278
PBE/aug-pc-2	0.9668	-4e-04	0.9985	0.0205
PBE/def2-TZVPP	0.9583	1e-04	0.9972	0.0272
PBE/def2-TZVPPD	0.9647	-6e-04	0.9983	0.0213
TPSS/6-31G(d)	0.9995	-0.0031	0.9884	0.0583
TPSS/pc-1	1.0313	0.0035	0.9949	0.0397
TPSS/aug-pc-1	1.0003	0.0018	0.9989	0.0175
TPSS/pc-2	1.0082	1e-04	0.9981	0.0238
TPSS/aug-pc-2	0.9835	-0.0011	0.9991	0.0155
TPSS/def2-TZVPP	0.9764	-5e-04	0.9984	0.0208
TPSS/def2-TZVPPD	0.9796	-0.0013	0.9992	0.0153
B3LYP/6-31G(d)	1.0165	-0.0034	0.9913	0.0513
B3LYP/pc-1	1.0447	0.0037	0.9957	0.0367
B3LYP/aug-pc-1	1.0229	0.0014	0.9988	0.0191
B3LYP/pc-2	1.0308	1e-04	0.9989	0.0181
B3LYP/aug-pc-2	1.0031	-0.0012	0.9997	0.0087
B3LYP/def2-TZVPP	0.9978	-4e-04	0.9995	0.0119
B3LYP/def2-TZVPPD	1.0017	-0.0014	0.9998	0.008
PBE0/6-31G(d)	1.0206	-0.0049	0.9905	0.0539
PBE0/pc-1	1.0515	0.0015	0.9952	0.0391
PBE0/aug-pc-1	1.0119	-1e-04	0.999	0.0176
PBE0/pc-2	1.0192	-0.0018	0.9986	0.0203
PBE0/aug-pc-2	0.9908	-0.0028	0.9996	0.0109
PBE0/def2-TZVPP	0.9871	-0.0023	0.999	0.0168
PBE0/def2-TZVPPD	0.989	-0.0031	0.9996	0.011

Table A16: Summary statistics for Q_{xx}(R)

Model Chemistry	MD	MAD	SD	Max$^+$	Max$^-$
SVWN/6-31G(d)	0.093	0.332	0.443	1.089(CH_2COO^-—ip)	-1.492(SCF_3)
SVWN/pc-1	0.067	0.217	0.288	0.902(C_6H_5)	-0.794(CHCCO)
SVWN/aug-pc-1	0.043	0.196	0.259	0.757(C_6H_5)	-0.652(CHCCO)
SVWN/pc-2	0.054	0.165	0.210	0.514(C_6H_5)	-0.682(CHCCO)
SVWN/aug-pc-2	0.045	0.160	0.214	0.652(PHO(OH))	-0.642(CHCCO)
SVWN/def2-TZVPP	0.050	0.210	0.283	0.759(CH_2COO^-—ip)	-0.741(SCF_3)
SVWN/def2-TZVPPD	0.041	0.190	0.259	0.679(CH_2OOCH_3—op)	-0.708(CHCCO)
BLYP/6-31G(d)	0.069	0.243	0.317	0.927(CH_2COO^-—ip)	-1.023(SCF_3)
BLYP/pc-1	0.043	0.142	0.215	0.982($PO(OH)_2$)	-0.794(NCO)
BLYP/aug-pc-1	0.013	0.094	0.148	0.705($PO(OH)_2$)	-0.733(NCO)
BLYP/pc-2	0.025	0.096	0.135	0.412($OCOCH_3$)	-0.277(C_6H_5)
BLYP/aug-pc-2	0.015	0.079	0.119	0.374($OCOCH_3$)	-0.350(NCO)
BLYP/def2-TZVPP	0.023	0.094	0.135	0.575(CH_2COO^-—ip)	-0.278(NCS)
BLYP/def2-TZVPPD	0.013	0.063	0.093	0.334(PHO(OH))	-0.282(C_6H_5)
PBE/6-31G(d)	0.076	0.274	0.354	1.008(CH_2COO^-—ip)	-1.216(SCF_3)
PBE/pc-1	0.048	0.147	0.211	0.903($PO(OH)_2$)	-0.633(NCO)
PBE/aug-pc-1	0.019	0.118	0.165	0.595($PO(OH)_2$)	-0.660(NCO)
PBE/pc-2	0.035	0.078	0.113	0.431(CH_2COO^-—ip)	-0.249(NCO)
PBE/aug-pc-2	0.024	0.076	0.110	0.480(PHO(OH))	-0.373(NCO)
PBE/def2-TZVPP	0.032	0.131	0.182	0.708(CH_2COO^-—ip)	-0.331($COOCH_3$)
PBE/def2-TZVPPD	0.022	0.097	0.132	0.510(PHO(OH))	-0.317(NCO)
TPSS/6-31G(d)	0.057	0.211	0.284	0.807(CH_2COO^-—ip)	-0.984(NCO)
TPSS/pc-1	0.032	0.143	0.217	1.008($PO(OH)_2$)	-0.811(NCO)
TPSS/aug-pc-1	0.004	0.118	0.178	0.716(SCF_3)	-0.848(NCO)
TPSS/pc-2	0.024	0.093	0.135	0.401($PO(OH)_2$)	-0.440(NCO)
TPSS/aug-pc-2	0.014	0.085	0.127	0.376(SCF_3)	-0.549(NCO)
TPSS/def2-TZVPP	0.021	0.101	0.142	0.504(CH_2COO^-—ip)	-0.448(NCO)
TPSS/def2-TZVPPD	0.010	0.069	0.101	0.401(PHO(OH))	-0.512(NCO)
B3LYP/6-31G(d)	0.054	0.190	0.248	0.708(CH_2COO^-—ip)	-0.930(SCF_3)
B3LYP/pc-1	0.030	0.135	0.195	0.709($PO(OH)_2$)	-0.490(OPH_2O)
B3LYP/aug-pc-1	0.005	0.075	0.123	0.689(SCF_3)	-0.404(SH)
B3LYP/pc-2	0.013	0.093	0.134	0.451(SCF_3)	-0.297(OPH_2O)
B3LYP/aug-pc-2	0.005	0.058	0.089	0.416(SCF_3)	-0.165(CH_2NO_2—op)
B3LYP/def2-TZVPP	0.009	0.059	0.098	0.464(S^-)	-0.198(CHCCO)
B3LYP/def2-TZVPPD	0.001	0.034	0.055	0.355(NCO)	-0.194(CHCCO)
PBE0/6-31G(d)	0.057	0.207	0.277	0.734(CH_2COO^-—ip)	-1.067(SCF_3)
PBE0/pc-1	0.029	0.140	0.199	0.630($PO(OH)_2$)	-0.509(ONO)
PBE0/aug-pc-1	0.007	0.100	0.138	0.544(SCF_3)	-0.444(SH)
PBE0/pc-2	0.019	0.072	0.111	0.518(NCO)	-0.346(CHCCO)
PBE0/aug-pc-2	0.011	0.050	0.074	0.386(NCO)	-0.290(CHCCO)
PBE0/def2-TZVPP	0.015	0.095	0.139	0.488(S^-)	-0.389(CHCCO)
PBE0/def2-TZVPPD	0.007	0.069	0.101	0.405(NCO)	-0.389(CHCCO)

Table A17: Summary statistics for Q_{xy}(R)

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	0.006	0.207	0.380	1.631($CH_2COCl-ip$)	-1.868(CHCCO)
SVWN/pc-1	-0.012	0.135	0.230	1.070($CH_2COCl-ip$)	-0.653(CHCCO)
SVWN/aug-pc-1	-0.012	0.144	0.233	0.724($CH_2COCl-ip$)	-0.835(CH_2OOCH_3-ip)
SVWN/pc-2	-0.003	0.099	0.176	0.666(SiCl)	-0.670(CHCCO)
SVWN/aug-pc-2	-0.002	0.115	0.195	0.688(SiCl)	-0.664(CH_2OOCH_3-ip)
SVWN/def2-TZVPP	-0.001	0.149	0.252	0.847($CH_2COCl-ip$)	-0.898(CH_2OOCH_3-ip)
SVWN/def2-TZVPPD	-0.002	0.148	0.249	0.769($CH_2COCl-ip$)	-0.908(CH_2OOCH_3-ip)
BLYP/6-31G(d)	0.013	0.125	0.222	0.986($CH_2COCl-ip$)	-0.602(CHCCO)
BLYP/pc-1	-0.014	0.091	0.173	0.642(NCO)	-0.708(SCF_3)
BLYP/aug-pc-1	-0.014	0.063	0.123	0.605(NCO)	-0.537(NNN)
BLYP/pc-2	-0.003	0.075	0.127	0.401(SiCl)	-0.421(SCF_3)
BLYP/aug-pc-2	-0.002	0.060	0.102	0.382(SiCl)	-0.284($CH_2COO^- - ip$)
BLYP/def2-TZVPP	-0.001	0.050	0.095	0.487(SiCl)	-0.396(SiF)
BLYP/def2-TZVPPD	-0.002	0.047	0.086	0.401(SiCl)	-0.289(SiF)
PBE/6-31G(d)	0.016	0.148	0.258	1.223($CH_2COCl-ip$)	-0.709(CH_2OOCH_3-ip)
PBE/pc-1	-0.010	0.091	0.168	0.539(NCO)	-0.624(SCF_3)
PBE/aug-pc-1	-0.009	0.076	0.133	0.505(NCO)	-0.405(NNN)
PBE/pc-2	0.000	0.051	0.089	0.439(SiCl)	-0.396(SiF)
PBE/aug-pc-2	0.001	0.046	0.091	0.443($PO(OH)_2$)	-0.278(SiF)
PBE/def2-TZVPP	0.002	0.077	0.136	0.537(SiCl)	-0.396(SiF)
PBE/def2-TZVPPD	0.001	0.071	0.124	0.459(SiCl)	-0.348(CH_2OOCH_3-ip)
TPSS/6-31G(d)	0.016	0.106	0.187	0.847($CH_2COCl-ip$)	-0.515(CCl_2F)
TPSS/pc-1	-0.013	0.096	0.194	0.680(NCO)	-0.974(SCF_3)
TPSS/aug-pc-1	-0.012	0.074	0.134	0.606(NCO)	-0.565(SCF_3)
TPSS/pc-2	-0.002	0.067	0.120	0.402(SSH)	-0.553(SCF_3)
TPSS/aug-pc-2	-0.002	0.054	0.096	0.308(NCO)	-0.369(SCF_3)
TPSS/def2-TZVPP	-0.001	0.043	0.075	0.262(PHO(OH))	-0.259(OCN)
TPSS/def2-TZVPPD	-0.002	0.035	0.066	0.232(NCO)	-0.316(OCN)
B3LYP/6-31G(d)	0.011	0.111	0.189	0.868($CH_2COCl-ip$)	-0.651(CHCCO)
B3LYP/pc-1	-0.018	0.082	0.175	0.677(SSH)	-0.967(SCF_3)
B3LYP/aug-pc-1	-0.017	0.043	0.090	0.323(SSH)	-0.589(SCF_3)
B3LYP/pc-2	-0.008	0.076	0.138	0.503(NNN)	-0.595(SCF_3)
B3LYP/aug-pc-2	-0.007	0.045	0.089	0.361(NNN)	-0.382(SCF_3)
B3LYP/def2-TZVPP	-0.006	0.033	0.083	0.503(NNN)	-0.502(NCO)
B3LYP/def2-TZVPPD	-0.007	0.024	0.066	0.442(NNN)	-0.395(NCO)
PBE0/6-31G(d)	0.012	0.126	0.219	1.026($CH_2COCl-ip$)	-0.678(CHCCO)
PBE0/pc-1	-0.015	0.085	0.180	0.763(SSH)	-0.971(SCF_3)
PBE0/aug-pc-1	-0.014	0.062	0.108	0.338(SSH)	-0.524(SCF_3)
PBE0/pc-2	-0.006	0.049	0.115	0.649(NNN)	-0.453(NCO)
PBE0/aug-pc-2	-0.005	0.040	0.081	0.496(NNN)	-0.386(NCO)
PBE0/def2-TZVPP	-0.005	0.062	0.119	0.617(NNN)	-0.562(NCO)
PBE0/def2-TZVPPD	-0.005	0.058	0.106	0.553(NNN)	-0.450(NCO)

Table A18: Summary statistics for Q_{xz}(R)

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	0.009	0.057	0.149	0.608(CH_2COCl—op)	-0.622(CH_2CSH—op)
SVWN/pc-1	0.007	0.038	0.099	0.447(PHO(OH))	-0.381(CH_2CSH—op)
SVWN/aug-pc-1	0.009	0.045	0.110	0.469($SiCl_2H$)	-0.428(CH_2CSH—op)
SVWN/pc-2	0.007	0.034	0.091	0.445($SiCl_2H$)	-0.347(CH_2CSH—op)
SVWN/aug-pc-2	0.007	0.036	0.091	0.464($SiCl_2H$)	-0.359(CH_2CSH—op)
SVWN/def2-TZVPP	0.007	0.042	0.107	0.454($SiCl_2H$)	-0.447(CH_2CSH—op)
SVWN/def2-TZVPPD	0.007	0.044	0.112	0.472($SiCl_2H$)	-0.466(CH_2CSH—op)
BLYP/6-31G(d)	0.006	0.030	0.080	0.527(PHO(OH))	-0.226(OPH_2O)
BLYP/pc-1	0.003	0.017	0.052	0.397(PHO(OH))	-0.239(SHO)
BLYP/aug-pc-1	0.006	0.011	0.027	0.157($SiCl_2H$)	-0.065(CF_2H)
BLYP/pc-2	0.003	0.019	0.050	0.352(PHO(OH))	-0.166(SHO_2)
BLYP/aug-pc-2	0.003	0.016	0.037	0.178(PHO(OH))	-0.093(CH_2COCl—op)
BLYP/def2-TZVPP	0.003	0.011	0.031	0.235(PHO(OH))	-0.086(SHO)
BLYP/def2-TZVPPD	0.004	0.010	0.030	0.214(PHO(OH))	-0.064(NHOH)
PBE/6-31G(d)	0.006	0.039	0.101	0.550(PHO(OH))	-0.367(CH_2CSH—op)
PBE/pc-1	0.004	0.017	0.053	0.427(PHO(OH))	-0.177(SHO)
PBE/aug-pc-1	0.007	0.015	0.045	0.293($SiCl_2H$)	-0.107(CH_2CSH—op)
PBE/pc-2	0.004	0.013	0.046	0.348(PHO(OH))	-0.071(SHO_2)
PBE/aug-pc-2	0.004	0.010	0.037	0.271($SiCl_2H$)	-0.075(SiH_3)
PBE/def2-TZVPP	0.004	0.018	0.050	0.270($SiCl_2H$)	-0.172(CH_2CSH—op)
PBE/def2-TZVPPD	0.005	0.019	0.052	0.273($SiCl_2H$)	-0.186(CH_2CSH—op)
TPSS/6-31G(d)	0.005	0.026	0.075	0.552(PHO(OH))	-0.225(OPH_2O)
TPSS/pc-1	0.002	0.020	0.059	0.437(PHO(OH))	-0.221(SHO)
TPSS/aug-pc-1	0.005	0.012	0.030	0.154($SiCl_2H$)	-0.077($CH_2CH_2CH_3$—op)
TPSS/pc-2	0.002	0.017	0.048	0.348(PHO(OH))	-0.193(SHO_2)
TPSS/aug-pc-2	0.002	0.014	0.035	0.198(PHO(OH))	-0.124(SHO_2)
TPSS/def2-TZVPP	0.002	0.008	0.028	0.237(PHO(OH))	-0.056(SHO)
TPSS/def2-TZVPPD	0.003	0.008	0.027	0.224(PHO(OH))	-0.039(SiH_3)
B3LYP/6-31G(d)	0.004	0.026	0.066	0.352(PHO(OH))	-0.218(CH_2CSH—op)
B3LYP/pc-1	0.002	0.016	0.043	0.263(PHO(OH))	-0.200(SHO)
B3LYP/aug-pc-1	0.003	0.010	0.025	0.124($SiCl_2H$)	-0.115(PHO(OH))
B3LYP/pc-2	0.001	0.014	0.038	0.200(PHO(OH))	-0.186(SHO_2)
B3LYP/aug-pc-2	0.001	0.011	0.027	0.108($SiCl_2H$)	-0.111(SHO_2)
B3LYP/def2-TZVPP	0.001	0.005	0.014	0.101($SiCl_2H$)	-0.053(SHO)
B3LYP/def2-TZVPPD	0.001	0.005	0.015	0.113($SiCl_2H$)	-0.031(NHOH)
PBE0/6-31G(d)	0.004	0.033	0.083	0.331(PHO(OH))	-0.337(CH_2CSH—op)
PBE0/pc-1	0.002	0.014	0.039	0.266(PHO(OH))	-0.147(SHO)
PBE0/aug-pc-1	0.003	0.014	0.038	0.221($SiCl_2H$)	-0.168(PHO(OH))
PBE0/pc-2	0.001	0.010	0.030	0.182($SiCl_2H$)	-0.114(SHO_2)
PBE0/aug-pc-2	0.001	0.008	0.025	0.193($SiCl_2H$)	-0.064(SiH_3)
PBE0/def2-TZVPP	0.001	0.014	0.037	0.187($SiCl_2H$)	-0.154(CH_2CSH—op)
PBE0/def2-TZVPPD	0.002	0.016	0.041	0.196($SiCl_2H$)	-0.170(CH_2CSH—op)

Table A19: Summary statistics for Q_{yy}(R)

Model Chemistry	MD	MAD	SD	Max$^+$	Max$^-$
SVWN/6-31G(d)	-0.024	0.323	0.435	0.900($COOCH_3$)	-1.358(C_6H_5)
SVWN/pc-1	-0.029	0.221	0.285	0.626(S^-)	-0.703(CH_2COH—ip)
SVWN/aug-pc-1	0.009	0.209	0.276	0.659(CCCl)	-0.751(CH_2OOCH_3—op)
SVWN/pc-2	-0.039	0.174	0.230	0.406(O^-)	-0.720(SiCl)
SVWN/aug-pc-2	-0.026	0.175	0.248	0.667($SiCl_2H$)	-0.961(C_6H_5)
SVWN/def2-TZVPP	-0.019	0.226	0.306	0.796($SiCl_2H$)	-0.918(C_6H_5)
SVWN/def2-TZVPPD	-0.030	0.206	0.290	0.710($SiCl_2H$)	-0.961(C_6H_5)
BLYP/6-31G(d)	0.007	0.227	0.309	1.167(NCO)	-0.779(C_6H_5)
BLYP/pc-1	0.005	0.128	0.192	0.695(NCO)	-0.644($SiCl_2H$)
BLYP/aug-pc-1	0.030	0.097	0.143	0.817(NCO)	-0.321(CH_2COO^-—ip)
BLYP/pc-2	-0.006	0.102	0.140	0.361(O^-)	-0.377(PHO(OH))
BLYP/aug-pc-2	-0.001	0.082	0.121	0.493(NCO)	-0.433(PHO(OH))
BLYP/def2-TZVPP	0.017	0.098	0.144	0.592(NCS)	-0.440(PHO(OH))
BLYP/def2-TZVPPD	-0.005	0.071	0.114	0.582(NCS)	-0.440(PHO(OH))
PBE/6-31G(d)	0.006	0.255	0.344	1.077(NCO)	-0.855(C_6H_5)
PBE/pc-1	0.004	0.133	0.189	0.579(S^-)	-0.480($SiCl_2H$)
PBE/aug-pc-1	0.045	0.128	0.172	0.765(NCO)	-0.386(CH_2COO^-—ip)
PBE/pc-2	-0.007	0.082	0.120	0.379(O^-)	-0.495(PHO(OH))
PBE/aug-pc-2	-0.001	0.086	0.135	0.507(NCO)	-0.549(PHO(OH))
PBE/def2-TZVPP	0.014	0.145	0.199	0.592($SiCl_2H$)	-0.546(PHO(OH))
PBE/def2-TZVPPD	-0.006	0.113	0.165	0.497(NCO)	-0.558(PHO(OH))
TPSS/6-31G(d)	0.020	0.194	0.271	1.172(NCO)	-0.634($SiCl_2H$)
TPSS/pc-1	0.016	0.134	0.206	0.707(NCO)	-0.727($SiCl_2H$)
TPSS/aug-pc-1	0.067	0.124	0.173	0.979(NCO)	-0.207(S^-)
TPSS/pc-2	0.006	0.095	0.135	0.511(NCO)	-0.396(SiCl)
TPSS/aug-pc-2	0.015	0.085	0.125	0.690(NCO)	-0.391(PHO(OH))
TPSS/def2-TZVPP	0.026	0.099	0.145	0.628(NCO)	-0.391(PHO(OH))
TPSS/def2-TZVPPD	0.011	0.073	0.119	0.696(NCO)	-0.403(PHO(OH))
B3LYP/6-31G(d)	-0.011	0.174	0.238	0.584(S^-)	-0.813($SiCl_2H$)
B3LYP/pc-1	-0.012	0.140	0.208	0.558(O^-)	-0.860($SiCl_2H$)
B3LYP/aug-pc-1	0.033	0.083	0.118	0.614(SSH)	-0.369(SHO_2)
B3LYP/pc-2	-0.014	0.104	0.149	0.398(O^-)	-0.548(NCO)
B3LYP/aug-pc-2	-0.002	0.060	0.083	0.228($CH_2CH_2CH_3$—ip)	-0.334(NCO)
B3LYP/def2-TZVPP	0.005	0.059	0.096	0.464(S^-)	-0.400(NCO)
B3LYP/def2-TZVPPD	-0.009	0.034	0.054	0.172(NCS)	-0.327(NCO)
PBE0/6-31G(d)	-0.013	0.191	0.262	0.592(S^-)	-0.749(SiCl)
PBE0/pc-1	-0.013	0.143	0.214	0.567(O^-)	-0.791($SiCl_2H$)
PBE0/aug-pc-1	0.049	0.108	0.138	0.623(SSH)	-0.304(NCO)
PBE0/pc-2	-0.016	0.077	0.127	0.419(O^-)	-0.648(NCO)
PBE0/aug-pc-2	-0.002	0.053	0.084	0.357(CHCCO)	-0.432(NCO)
PBE0/def2-TZVPP	0.002	0.099	0.139	0.488(S^-)	-0.487(NCO)
PBE0/def2-TZVPPD	-0.010	0.076	0.107	0.281(CHCCO)	-0.413(NCO)

Table A20: Summary statistics for Q_{yz}(R)

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.002	0.108	0.212	0.685($SiClH_2$)	-0.576(SSH)
SVWN/pc-1	-0.002	0.074	0.141	0.482($SiClH_2$)	-0.420($CHCCH_2$)
SVWN/aug-pc-1	-0.001	0.082	0.153	0.518($SiClH_2$)	-0.395($CHCCH_2$)
SVWN/pc-2	-0.001	0.064	0.125	0.551($SiClH_2$)	-0.396(PH_2O)
SVWN/aug-pc-2	0.000	0.066	0.127	0.558($SiClH_2$)	-0.406(PH_2O)
SVWN/def2-TZVPP	0.001	0.078	0.152	0.567($SiClH_2$)	-0.408($CHCCH_2$)
SVWN/def2-TZVPPD	0.000	0.079	0.153	0.568($SiClH_2$)	-0.402($CHCCH_2$)
BLYP/6-31G(d)	-0.002	0.047	0.088	0.336($SiClH_2$)	-0.327(SSH)
BLYP/pc-1	-0.001	0.020	0.043	0.173(SSH)	-0.224(PHO(OH))
BLYP/aug-pc-1	0.000	0.017	0.034	0.164(SSH)	-0.125(OPH_2O)
BLYP/pc-2	-0.001	0.033	0.059	0.159(SSH)	-0.160(OPH_2O)
BLYP/aug-pc-2	0.000	0.033	0.058	0.157(SSH)	-0.144(CHCO)
BLYP/def2-TZVPP	0.001	0.019	0.041	0.173($SiClH_2$)	-0.153($NHCH_3$)
BLYP/def2-TZVPPD	0.000	0.019	0.041	0.162(NNH—cis)	-0.148($NHCH_3$)
PBE/6-31G(d)	-0.002	0.068	0.128	0.465($SiClH_2$)	-0.419(SSH)
PBE/pc-1	-0.001	0.020	0.045	0.266($SiClH_2$)	-0.178(PHO(OH))
PBE/aug-pc-1	0.000	0.021	0.047	0.284($SiClH_2$)	-0.162(PH_2O)
PBE/pc-2	-0.001	0.017	0.045	0.306($SiClH_2$)	-0.216(PH_2O)
PBE/aug-pc-2	0.000	0.016	0.044	0.302($SiClH_2$)	-0.217(PH_2O)
PBE/def2-TZVPP	0.001	0.033	0.063	0.324($SiClH_2$)	-0.187($SiFH_2$)
PBE/def2-TZVPPD	0.000	0.033	0.062	0.309($SiClH_2$)	-0.181(PH_2O)
TPSS/6-31G(d)	-0.002	0.042	0.078	0.257($SiClH_2$)	-0.303(SSH)
TPSS/pc-1	0.000	0.025	0.051	0.255(SSH)	-0.253(PHO(OH))
TPSS/aug-pc-1	0.001	0.019	0.038	0.233(SSH)	-0.104(SH)
TPSS/pc-2	0.000	0.027	0.050	0.196(SSH)	-0.144(PHO(OH))
TPSS/aug-pc-2	0.001	0.027	0.048	0.195(SSH)	-0.108(CH_2NO_2—op)
TPSS/def2-TZVPP	0.001	0.013	0.031	0.139($SiClH_2$)	-0.111($NHCH_3$)
TPSS/def2-TZVPPD	0.001	0.013	0.031	0.123($SiClH_2$)	-0.109($NHCH_3$)
B3LYP/6-31G(d)	-0.002	0.044	0.083	0.274($SiClH_2$)	-0.322(SSH)
B3LYP/pc-1	0.000	0.018	0.041	0.254(SSH)	-0.165(SiH)
B3LYP/aug-pc-1	0.001	0.015	0.035	0.238(SSH)	-0.118(SH)
B3LYP/pc-2	0.000	0.024	0.045	0.214(SSH)	-0.102(CHCO)
B3LYP/aug-pc-2	0.001	0.023	0.045	0.204(SSH)	-0.114(CHNN)
B3LYP/def2-TZVPP	0.002	0.009	0.022	0.125($SiClH_2$)	-0.082($SiFH_2$)
B3LYP/def2-TZVPPD	0.001	0.009	0.022	0.122($SiClH_2$)	-0.074($SiFH_2$)
PBE0/6-31G(d)	-0.002	0.060	0.118	0.354($SiClH_2$)	-0.396(SSH)
PBE0/pc-1	0.000	0.020	0.041	0.193(SSH)	-0.115(PHO(OH))
PBE0/aug-pc-1	0.002	0.025	0.048	0.211($SiClH_2$)	-0.115(PH_2O)
PBE0/pc-2	0.000	0.014	0.034	0.218($SiClH_2$)	-0.155(PH_2O)
PBE0/aug-pc-2	0.001	0.015	0.036	0.215($SiClH_2$)	-0.151(PH_2O)
PBE0/def2-TZVPP	0.002	0.029	0.054	0.228($SiClH_2$)	-0.142($SiFH_2$)
PBE0/def2-TZVPPD	0.001	0.029	0.055	0.222($SiClH_2$)	-0.138($CHCCH_2$)

Table A21: Summary statistics for Q_{zz}(R)

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.069	0.301	0.413	0.919(MgH)	-1.480(CCCl)
SVWN/pc-1	-0.038	0.218	0.322	0.894(MgH)	-1.253(S^-)
SVWN/aug-pc-1	-0.052	0.187	0.272	0.746(MgH)	-1.317(CCCl)
SVWN/pc-2	-0.015	0.173	0.242	0.645(MgH)	-0.811(O^-)
SVWN/aug-pc-2	-0.019	0.163	0.231	0.662(MgH)	-0.847(CCCl)
SVWN/def2-TZVPP	-0.031	0.209	0.299	0.625(MgH)	-1.169(S^-)
SVWN/def2-TZVPPD	-0.011	0.192	0.270	0.652(MgH)	-0.952(CCCl)
BLYP/6-31G(d)	-0.076	0.227	0.333	0.869(C_6H_5)	-1.590($PO(OH)_2$)
BLYP/pc-1	-0.048	0.155	0.268	0.598($SiCl_3$)	-1.525($PO(OH)_2$)
BLYP/aug-pc-1	-0.043	0.114	0.171	0.532(S^-)	-1.015($PO(OH)_2$)
BLYP/pc-2	-0.020	0.099	0.155	0.412(CCCl)	-0.722(O^-)
BLYP/aug-pc-2	-0.014	0.095	0.131	0.475(C_6H_5)	-0.409($PO(OH)_2$)
BLYP/def2-TZVPP	-0.039	0.112	0.174	0.326(C_6H_5)	-0.975(S^-)
BLYP/def2-TZVPPD	-0.008	0.081	0.117	0.475(C_6H_5)	-0.401($PO(OH)_2$)
PBE/6-31G(d)	-0.082	0.233	0.334	0.656(PH_2O)	-1.492($PO(OH)_2$)
PBE/pc-1	-0.052	0.156	0.261	0.626(MgH)	-1.427($PO(OH)_2$)
PBE/aug-pc-1	-0.064	0.137	0.183	0.468(MgH)	-0.972($PO(OH)_2$)
PBE/pc-2	-0.027	0.089	0.146	0.405(SiCl)	-0.758(O^-)
PBE/aug-pc-2	-0.023	0.087	0.123	0.390(MgH)	-0.472($SiCl_3$)
PBE/def2-TZVPP	-0.046	0.130	0.198	0.377(MgH)	-1.057(S^-)
PBE/def2-TZVPPD	-0.017	0.100	0.136	0.383(MgH)	-0.515($SiCl_3$)
TPSS/6-31G(d)	-0.078	0.199	0.293	0.610(PH_2O)	-1.532($PO(OH)_2$)
TPSS/pc-1	-0.048	0.170	0.280	0.688($SiCl_3$)	-1.504($PO(OH)_2$)
TPSS/aug-pc-1	-0.071	0.155	0.205	0.413(S^-)	-1.092($PO(OH)_2$)
TPSS/pc-2	-0.030	0.104	0.160	0.392(SiCl)	-0.751(O^-)
TPSS/aug-pc-2	-0.029	0.099	0.127	0.310(CCF)	-0.556($PO(OH)_2$)
TPSS/def2-TZVPP	-0.046	0.108	0.169	0.298(PH_2O)	-0.934(S^-)
TPSS/def2-TZVPPD	-0.021	0.073	0.101	0.293(C_6H_5)	-0.542($PO(OH)_2$)
B3LYP/6-31G(d)	-0.044	0.191	0.287	0.584($SiCl_3$)	-1.168(S^-)
B3LYP/pc-1	-0.017	0.165	0.263	0.848($SiCl_3$)	-1.115(O^-)
B3LYP/aug-pc-1	-0.038	0.099	0.152	0.406(S^-)	-0.618($PO(OH)_2$)
B3LYP/pc-2	0.001	0.089	0.149	0.442(SiCl)	-0.795(O^-)
B3LYP/aug-pc-2	-0.003	0.055	0.078	0.263(CCCl)	-0.286(SCF_3)
B3LYP/def2-TZVPP	-0.014	0.067	0.140	0.190(CCCl)	-0.929(S^-)
B3LYP/def2-TZVPPD	0.008	0.043	0.057	0.198(C_6H_5)	-0.119($SiClH_2$)
PBE0/6-31G(d)	-0.044	0.190	0.283	0.588(SiCl)	-1.184(S^-)
PBE0/pc-1	-0.016	0.166	0.264	0.740($SiCl_3$)	-1.135(O^-)
PBE0/aug-pc-1	-0.056	0.120	0.167	0.297(S^-)	-0.612(C_6H_5)
PBE0/pc-2	-0.003	0.077	0.142	0.507(SiCl)	-0.838(O^-)
PBE0/aug-pc-2	-0.009	0.050	0.068	0.180(MgH)	-0.262($SiCl_3$)
PBE0/def2-TZVPP	-0.017	0.085	0.158	0.218(BeH)	-0.977(S^-)
PBE0/def2-TZVPPD	0.003	0.058	0.081	0.260(CH_2OOCH_3—ip)	-0.319($SiCl_3$)

Table A22: Summary statistics for ||**Q**(R)||

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.303	0.383	0.409	0.860($CH_2CSH-ip$)	-1.480(CCCl)
SVWN/pc-1	-0.136	0.251	0.329	0.714(CNO)	-1.253(S^-)
SVWN/aug-pc-1	-0.127	0.241	0.304	0.646(CNO)	-1.317(CCCl)
SVWN/pc-2	-0.110	0.200	0.253	0.420(CNO)	-0.811(O^-)
SVWN/aug-pc-2	-0.147	0.206	0.246	0.377(CNO)	-0.847(CCCl)
SVWN/def2-TZVPP	-0.206	0.270	0.299	0.489($CH_2CSH-ip$)	-1.169(S^-)
SVWN/def2-TZVPPD	-0.164	0.237	0.281	0.501($CH_2CSH-ip$)	-0.952(CCCl)
BLYP/6-31G(d)	-0.262	0.303	0.298	0.608(CCl_2F)	-1.196(S^-)
BLYP/pc-1	-0.078	0.167	0.245	0.598($SiCl_3$)	-1.139(S^-)
BLYP/aug-pc-1	-0.017	0.118	0.169	0.566(SSH)	-0.753(NCO)
BLYP/pc-2	-0.004	0.113	0.166	0.412(CCCl)	-0.722(O^-)
BLYP/aug-pc-2	-0.021	0.096	0.143	0.445(S^-)	-0.477(NCO)
BLYP/def2-TZVPP	-0.115	0.134	0.166	0.239(CCCl)	-0.975(S^-)
BLYP/def2-TZVPPD	-0.051	0.085	0.122	0.360(S^-)	-0.578(NCS)
PBE/6-31G(d)	-0.291	0.328	0.312	0.539(CCl_2F)	-1.227(S^-)
PBE/pc-1	-0.105	0.174	0.236	0.410(C_6H_5)	-1.158(S^-)
PBE/aug-pc-1	-0.086	0.161	0.190	0.533(SSH)	-0.713(NCO)
PBE/pc-2	-0.075	0.098	0.139	0.158(CCCl)	-0.758(O^-)
PBE/aug-pc-2	-0.093	0.113	0.130	0.268(O^-)	-0.564(PHO(OH))
PBE/def2-TZVPP	-0.179	0.188	0.187	0.129(CO)	-1.057(S^-)
PBE/def2-TZVPPD	-0.121	0.137	0.141	0.160(S^-)	-0.591(PHO(OH))
TPSS/6-31G(d)	-0.219	0.261	0.268	0.581(CCl_2F)	-1.156(S^-)
TPSS/pc-1	-0.026	0.174	0.266	0.688($SiCl_3$)	-1.077(S^-)
TPSS/aug-pc-1	-0.032	0.150	0.207	0.736(SSH)	-0.902(NCO)
TPSS/pc-2	-0.024	0.111	0.166	0.372(CCCl)	-0.751(O^-)
TPSS/aug-pc-2	-0.040	0.103	0.140	0.310(CCF)	-0.655(NCO)
TPSS/def2-TZVPP	-0.121	0.138	0.166	0.221(CCF)	-0.934(S^-)
TPSS/def2-TZVPPD	-0.072	0.093	0.115	0.189(S^-)	-0.668(NCO)
B3LYP/6-31G(d)	-0.169	0.229	0.267	0.596($SiCl_2H$)	-1.168(S^-)
B3LYP/pc-1	0.022	0.188	0.276	0.848($SiCl_3$)	-1.115(O^-)
B3LYP/aug-pc-1	0.031	0.102	0.157	0.659(SCF_3)	-0.351(CCl_3)
B3LYP/pc-2	0.051	0.123	0.176	0.535(NCO)	-0.795(O^-)
B3LYP/aug-pc-2	0.021	0.067	0.093	0.322(NCO)	-0.195(C_6H_5)
B3LYP/def2-TZVPP	-0.050	0.083	0.145	0.386(NCO)	-0.929(S^-)
B3LYP/def2-TZVPPD	-0.004	0.048	0.068	0.314(NCO)	-0.195(C_6H_5)
PBE0/6-31G(d)	-0.173	0.238	0.281	0.519($CH_2CSH-ip$)	-1.184(S^-)
PBE0/pc-1	0.022	0.191	0.285	0.816(C_6H_5)	-1.135(O^-)
PBE0/aug-pc-1	-0.015	0.134	0.184	0.623(SSH)	-0.523(CCl_3)
PBE0/pc-2	0.003	0.098	0.166	0.634(NCO)	-0.838(O^-)
PBE0/aug-pc-2	-0.027	0.063	0.088	0.419(NCO)	-0.338(CHCCO)
PBE0/def2-TZVPP	-0.091	0.121	0.169	0.473(NCO)	-0.977(S^-)
PBE0/def2-TZVPPD	-0.052	0.083	0.107	0.399(NCO)	-0.400(CHCCO)

Table A23: Summary statistics for atomic quadrupole moments, $Q(\Omega)_{xx}$

Model Chemistry	MD	MAD	SD	Max	Min
SVWN/6-31G(d)	-0.002	0.094	0.173	1.740	0.000
SVWN/pc-1	-0.013	0.083	0.148	1.238	0.000
SVWN/aug-pc-1	-0.017	0.081	0.133	0.770	0.000
SVWN/pc-2	-0.012	0.057	0.109	1.047	0.000
SVWN/aug-pc-2	-0.012	0.054	0.092	0.755	0.000
SVWN/def2-TZVPP	-0.010	0.069	0.124	1.200	0.000
SVWN/def2-TZVPPD	-0.013	0.065	0.106	0.787	0.000
BLYP/6-31G(d)	0.013	0.079	0.137	1.210	0.000
BLYP/pc-1	0.001	0.068	0.126	1.101	0.000
BLYP/aug-pc-1	-0.005	0.071	0.136	0.903	0.000
BLYP/pc-2	0.001	0.047	0.086	0.664	0.000
BLYP/aug-pc-2	0.000	0.045	0.077	0.458	0.000
BLYP/def2-TZVPP	0.003	0.046	0.085	0.968	0.000
BLYP/def2-TZVPPD	0.000	0.045	0.076	0.359	0.000
PBE/6-31G(d)	0.011	0.081	0.137	1.247	0.000
PBE/pc-1	-0.002	0.068	0.125	1.123	0.000
PBE/aug-pc-1	-0.008	0.066	0.125	0.824	0.000
PBE/pc-2	0.000	0.038	0.076	0.711	0.000
PBE/aug-pc-2	-0.002	0.033	0.059	0.338	0.000
PBE/def2-TZVPP	0.002	0.044	0.085	1.063	0.000
PBE/def2-TZVPPD	-0.001	0.039	0.064	0.326	0.000
TPSS/6-31G(d)	0.015	0.073	0.126	1.162	0.000
TPSS/pc-1	0.002	0.065	0.119	1.030	0.000
TPSS/aug-pc-1	-0.003	0.062	0.126	1.057	0.000
TPSS/pc-2	0.005	0.038	0.072	0.699	0.000
TPSS/aug-pc-2	0.004	0.031	0.055	0.330	0.000
TPSS/def2-TZVPP	0.007	0.034	0.070	0.907	0.000
TPSS/def2-TZVPPD	0.004	0.029	0.050	0.223	0.000
B3LYP/6-31G(d)	0.008	0.070	0.122	1.178	0.000
B3LYP/pc-1	-0.004	0.061	0.117	1.081	0.000
B3LYP/aug-pc-1	-0.009	0.058	0.120	1.042	0.000
B3LYP/pc-2	-0.002	0.034	0.074	0.747	0.000
B3LYP/aug-pc-2	-0.003	0.025	0.049	0.428	0.000
B3LYP/def2-TZVPP	-0.001	0.027	0.066	0.905	0.000
B3LYP/def2-TZVPPD	-0.003	0.023	0.042	0.389	0.000
PBE0/6-31G(d)	0.006	0.072	0.125	1.198	0.000
PBE0/pc-1	-0.007	0.063	0.118	1.105	0.000
PBE0/aug-pc-1	-0.010	0.056	0.114	1.035	0.000
PBE0/pc-2	-0.003	0.032	0.073	0.802	0.000
PBE0/aug-pc-2	-0.003	0.022	0.041	0.463	0.000
PBE0/def2-TZVPP	-0.002	0.031	0.074	0.959	0.000
PBE0/def2-TZVPPD	-0.004	0.021	0.041	0.427	0.000

Table A24: Summary statistics for atomic quadrupole moments, $Q(\Omega)_{yy}$

Model Chemistry	MD	MAD	SD	Max	Min
SVWN/6-31G(d)	0.010	0.090	0.165	1.335	0.000
SVWN/pc-1	0.027	0.080	0.154	1.238	0.000
SVWN/aug-pc-1	0.037	0.087	0.154	1.097	0.000
SVWN/pc-2	0.018	0.058	0.117	0.841	0.000
SVWN/aug-pc-2	0.014	0.056	0.104	0.728	0.000
SVWN/def2-TZVPP	0.015	0.067	0.126	1.200	0.000
SVWN/def2-TZVPPD	0.017	0.065	0.113	0.713	0.000
BLYP/6-31G(d)	-0.010	0.077	0.141	1.209	0.000
BLYP/pc-1	0.009	0.066	0.135	1.101	0.000
BLYP/aug-pc-1	0.024	0.073	0.145	1.210	0.000
BLYP/pc-2	0.004	0.052	0.098	0.664	0.000
BLYP/aug-pc-2	0.002	0.050	0.093	0.543	0.000
BLYP/def2-TZVPP	0.001	0.052	0.102	0.968	0.000
BLYP/def2-TZVPPD	0.004	0.049	0.091	0.609	0.000
PBE/6-31G(d)	-0.006	0.078	0.139	1.246	0.000
PBE/pc-1	0.014	0.066	0.133	1.122	0.000
PBE/aug-pc-1	0.026	0.069	0.138	1.103	0.000
PBE/pc-2	0.002	0.043	0.089	0.711	0.000
PBE/aug-pc-2	0.000	0.039	0.076	0.510	0.000
PBE/def2-TZVPP	-0.001	0.046	0.096	1.062	0.000
PBE/def2-TZVPPD	0.001	0.040	0.077	0.510	0.000
TPSS/6-31G(d)	-0.010	0.070	0.128	1.161	0.000
TPSS/pc-1	0.008	0.061	0.120	1.030	0.000
TPSS/aug-pc-1	0.019	0.063	0.124	1.042	0.000
TPSS/pc-2	-0.007	0.042	0.080	0.699	0.000
TPSS/aug-pc-2	-0.007	0.035	0.064	0.416	0.000
TPSS/def2-TZVPP	-0.009	0.037	0.080	0.906	0.000
TPSS/def2-TZVPPD	-0.006	0.030	0.056	0.278	0.000
B3LYP/6-31G(d)	-0.005	0.067	0.123	1.177	0.000
B3LYP/pc-1	0.013	0.057	0.114	1.081	0.000
B3LYP/aug-pc-1	0.026	0.058	0.119	1.114	0.000
B3LYP/pc-2	0.002	0.034	0.071	0.748	0.000
B3LYP/aug-pc-2	0.002	0.026	0.052	0.397	0.000
B3LYP/def2-TZVPP	0.000	0.029	0.069	0.905	0.000
B3LYP/def2-TZVPPD	0.003	0.023	0.044	0.356	0.000
PBE0/6-31G(d)	-0.001	0.070	0.126	1.197	0.000
PBE0/pc-1	0.016	0.058	0.114	1.105	0.000
PBE0/aug-pc-1	0.027	0.056	0.114	1.023	0.000
PBE0/pc-2	-0.001	0.032	0.069	0.802	0.000
PBE0/aug-pc-2	-0.002	0.021	0.042	0.440	0.000
PBE0/def2-TZVPP	-0.002	0.029	0.071	0.959	0.000
PBE0/def2-TZVPPD	0.000	0.018	0.036	0.390	0.000

Table A25: Summary statistics for atomic quadrupole moments, $Q(\Omega)_{zz}$

Model Chemistry	MD	MAD	SD	Max	Min
SVWN/6-31G(d)	-0.008	0.106	0.204	2.671	0.000
SVWN/pc-1	-0.014	0.096	0.203	2.476	0.000
SVWN/aug-pc-1	-0.021	0.087	0.153	1.076	0.000
SVWN/pc-2	-0.006	0.062	0.137	1.594	0.000
SVWN/aug-pc-2	-0.002	0.055	0.105	0.950	0.000
SVWN/def2-TZVPP	-0.004	0.073	0.166	2.400	0.000
SVWN/def2-TZVPPD	-0.004	0.061	0.107	0.922	0.000
BLYP/6-31G(d)	-0.003	0.091	0.184	2.419	0.000
BLYP/pc-1	-0.010	0.079	0.180	2.202	0.000
BLYP/aug-pc-1	-0.019	0.083	0.156	1.098	0.000
BLYP/pc-2	-0.005	0.054	0.114	1.328	0.000
BLYP/aug-pc-2	-0.002	0.051	0.100	0.916	0.000
BLYP/def2-TZVPP	-0.004	0.055	0.133	1.936	0.000
BLYP/def2-TZVPPD	-0.004	0.049	0.094	0.719	0.000
PBE/6-31G(d)	-0.005	0.093	0.187	2.493	0.000
PBE/pc-1	-0.011	0.081	0.184	2.245	0.000
PBE/aug-pc-1	-0.018	0.078	0.148	1.026	0.000
PBE/pc-2	-0.002	0.049	0.116	1.422	0.000
PBE/aug-pc-2	0.001	0.042	0.086	0.676	0.000
PBE/def2-TZVPP	-0.001	0.055	0.141	2.125	0.000
PBE/def2-TZVPPD	0.000	0.045	0.084	0.654	0.000
TPSS/6-31G(d)	-0.005	0.083	0.173	2.323	0.000
TPSS/pc-1	-0.010	0.075	0.167	2.060	0.000
TPSS/aug-pc-1	-0.017	0.071	0.134	0.938	0.000
TPSS/pc-2	0.001	0.044	0.099	1.399	0.000
TPSS/aug-pc-2	0.003	0.035	0.061	0.547	0.000
TPSS/def2-TZVPP	0.002	0.043	0.117	1.813	0.000
TPSS/def2-TZVPPD	0.002	0.032	0.056	0.404	0.000
B3LYP/6-31G(d)	-0.003	0.078	0.169	2.355	0.000
B3LYP/pc-1	-0.008	0.067	0.159	2.162	0.000
B3LYP/aug-pc-1	-0.017	0.065	0.128	0.991	0.000
B3LYP/pc-2	-0.001	0.032	0.089	1.495	0.000
B3LYP/aug-pc-2	0.001	0.022	0.041	0.489	0.000
B3LYP/def2-TZVPP	0.000	0.029	0.108	1.810	0.000
B3LYP/def2-TZVPPD	0.000	0.019	0.036	0.360	0.000
PBE0/6-31G(d)	-0.005	0.079	0.170	2.395	0.000
PBE0/pc-1	-0.009	0.068	0.160	2.210	0.000
PBE0/aug-pc-1	-0.017	0.061	0.121	0.921	0.000
PBE0/pc-2	0.004	0.034	0.097	1.605	0.000
PBE0/aug-pc-2	0.005	0.020	0.035	0.255	0.000
PBE0/def2-TZVPP	0.004	0.035	0.118	1.918	0.000
PBE0/def2-TZVPPD	0.004	0.020	0.036	0.340	0.000

Table A26: Summary statistics for atomic second radial moments, $\langle r^2 \rangle_\Omega$

Model Chemistry	MD	MAD	MD	Max	Min
SVWN/6-31G(d)	-0.083	0.436	0.671	6.578	0.002
SVWN/pc-1	0.002	0.278	0.533	5.797	0.000
SVWN/aug-pc-1	0.126	0.225	0.378	2.212	0.000
SVWN/pc-2	0.038	0.203	0.371	3.313	0.001
SVWN/aug-pc-2	0.066	0.197	0.332	2.222	0.001
SVWN/def2-TZVPP	0.016	0.262	0.453	4.172	0.001
SVWN/def2-TZVPPD	0.072	0.232	0.373	1.948	0.000
BLYP/6-31G(d)	-0.145	0.371	0.599	6.886	0.003
BLYP/pc-1	-0.063	0.241	0.501	6.122	0.001
BLYP/aug-pc-1	0.096	0.168	0.274	1.874	0.002
BLYP/pc-2	-0.002	0.164	0.302	3.582	0.005
BLYP/aug-pc-2	0.040	0.151	0.242	1.593	0.003
BLYP/def2-TZVPP	-0.033	0.193	0.375	4.492	0.008
BLYP/def2-TZVPPD	0.042	0.156	0.252	1.629	0.002
PBE/6-31G(d)	-0.125	0.387	0.606	6.752	0.000
PBE/pc-1	-0.052	0.228	0.492	5.982	0.000
PBE/aug-pc-1	0.084	0.158	0.278	1.856	0.000
PBE/pc-2	-0.009	0.153	0.303	3.576	0.001
PBE/aug-pc-2	0.032	0.134	0.232	1.596	0.000
PBE/def2-TZVPP	-0.035	0.208	0.385	4.449	0.000
PBE/def2-TZVPPD	0.035	0.162	0.260	1.409	0.001
TPSS/6-31G(d)	-0.167	0.327	0.529	6.809	0.001
TPSS/pc-1	-0.099	0.214	0.455	6.081	0.000
TPSS/aug-pc-1	0.013	0.127	0.211	1.364	0.002
TPSS/pc-2	-0.077	0.144	0.273	3.736	0.003
TPSS/aug-pc-2	-0.040	0.116	0.164	0.688	0.003
TPSS/def2-TZVPP	-0.098	0.162	0.339	4.732	0.000
TPSS/def2-TZVPPD	-0.037	0.109	0.168	0.666	0.000
B3LYP/6-31G(d)	-0.158	0.309	0.502	6.783	0.002
B3LYP/pc-1	-0.084	0.178	0.425	6.221	0.001
B3LYP/aug-pc-1	0.038	0.099	0.166	1.016	0.000
B3LYP/pc-2	-0.048	0.109	0.241	3.682	0.000
B3LYP/aug-pc-2	-0.018	0.083	0.125	0.665	0.000
B3LYP/def2-TZVPP	-0.070	0.119	0.303	4.672	0.000
B3LYP/def2-TZVPPD	-0.012	0.070	0.118	0.581	0.000
PBE0/6-31G(d)	-0.144	0.318	0.503	6.649	0.000
PBE0/pc-1	-0.076	0.155	0.403	6.032	0.000
PBE0/aug-pc-1	0.020	0.083	0.159	0.874	0.000
PBE0/pc-2	-0.061	0.093	0.236	3.713	0.000
PBE0/aug-pc-2	-0.032	0.065	0.105	0.588	0.000
PBE0/def2-TZVPP	-0.077	0.137	0.304	4.683	0.000
PBE0/def2-TZVPPD	-0.025	0.089	0.123	0.671	0.001

Table A27: Linear regression parameters for Q_{xx}(R)

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.8824	0.0655	0.967	0.3606
SVWN/pc-1	0.9571	0.057	0.9837	0.2729
SVWN/aug-pc-1	0.9582	0.0332	0.987	0.2436
SVWN/pc-2	0.9616	0.0446	0.9918	0.193
SVWN/aug-pc-2	0.9436	0.0314	0.9931	0.1744
SVWN/def2-TZVPP	0.9232	0.032	0.9878	0.2274
SVWN/def2-TZVPPD	0.9364	0.0263	0.989	0.2185
BLYP/6-31G(d)	0.9072	0.0469	0.9855	0.2438
BLYP/pc-1	0.9821	0.0389	0.9906	0.212
BLYP/aug-pc-1	1.0029	0.0139	0.9955	0.1488
BLYP/pc-2	1.0031	0.0261	0.9963	0.1356
BLYP/aug-pc-2	0.9918	0.0131	0.9971	0.1178
BLYP/def2-TZVPP	0.9611	0.0136	0.9976	0.105
BLYP/def2-TZVPPD	0.9828	0.0092	0.9985	0.0851
PBE/6-31G(d)	0.8956	0.0517	0.9817	0.2709
PBE/pc-1	0.9722	0.0411	0.9912	0.2032
PBE/aug-pc-1	0.9762	0.0138	0.9948	0.1567
PBE/pc-2	0.9763	0.0293	0.9978	0.1011
PBE/aug-pc-2	0.9649	0.016	0.9986	0.0786
PBE/def2-TZVPP	0.9358	0.0167	0.997	0.1142
PBE/def2-TZVPPD	0.9556	0.012	0.9982	0.0891
TPSS/6-31G(d)	0.9247	0.0398	0.9873	0.2319
TPSS/pc-1	1.0016	0.0329	0.9904	0.2176
TPSS/aug-pc-1	0.9955	0.0031	0.9935	0.1784
TPSS/pc-2	0.9952	0.0232	0.9962	0.1354
TPSS/aug-pc-2	0.9846	0.0104	0.9968	0.1229
TPSS/def2-TZVPP	0.9575	0.0108	0.9975	0.1071
TPSS/def2-TZVPPD	0.9743	0.0041	0.9985	0.0838
B3LYP/6-31G(d)	0.945	0.0416	0.9893	0.2175
B3LYP/pc-1	1.0218	0.035	0.993	0.1903
B3LYP/aug-pc-1	1.0219	0.0101	0.9975	0.1141
B3LYP/pc-2	1.0249	0.0191	0.9971	0.1226
B3LYP/aug-pc-2	1.01	0.0075	0.9985	0.0862
B3LYP/def2-TZVPP	0.9863	0.0058	0.9982	0.0934
B3LYP/def2-TZVPPD	1.0014	0.0014	0.9994	0.0554
PBE0/6-31G(d)	0.9433	0.0433	0.9861	0.2482
PBE0/pc-1	1.0222	0.034	0.9927	0.1941
PBE0/aug-pc-1	1.0034	0.008	0.9961	0.1383
PBE0/pc-2	1.0066	0.0206	0.9975	0.1109
PBE0/aug-pc-2	0.9914	0.009	0.9989	0.0722
PBE0/def2-TZVPP	0.9694	0.0073	0.9968	0.1222
PBE0/def2-TZVPPD	0.9824	0.0028	0.9981	0.0937

Table A28: Linear regression parameters for Q_{yy}(R)

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.8836	-0.0747	0.9719	0.3452
SVWN/pc-1	0.9582	-0.0471	0.9853	0.2692
SVWN/aug-pc-1	0.9423	-0.0158	0.9875	0.2433
SVWN/pc-2	0.9618	-0.0556	0.9908	0.2135
SVWN/aug-pc-2	0.9358	-0.0537	0.9914	0.2009
SVWN/def2-TZVPP	0.9193	-0.0544	0.9868	0.2446
SVWN/def2-TZVPPD	0.9258	-0.063	0.9878	0.2366
BLYP/6-31G(d)	0.9067	-0.0335	0.9885	0.2248
BLYP/pc-1	0.982	-0.0034	0.9931	0.1887
BLYP/aug-pc-1	0.9875	0.0247	0.9961	0.1412
BLYP/pc-2	1.0039	-0.004	0.9963	0.14
BLYP/aug-pc-2	0.9853	-0.007	0.9974	0.1164
BLYP/def2-TZVPP	0.9581	-0.0018	0.9976	0.1078
BLYP/def2-TZVPPD	0.9735	-0.0165	0.9981	0.0966
PBE/6-31G(d)	0.8922	-0.0412	0.9864	0.2413
PBE/pc-1	0.9696	-0.0093	0.9937	0.177
PBE/aug-pc-1	0.9572	0.0259	0.9958	0.1425
PBE/pc-2	0.9734	-0.0191	0.9978	0.1043
PBE/aug-pc-2	0.9544	-0.0207	0.9984	0.0866
PBE/def2-TZVPP	0.9286	-0.0171	0.9972	0.1138
PBE/def2-TZVPPD	0.9424	-0.0312	0.9979	0.0997
TPSS/6-31G(d)	0.9243	-0.0128	0.9904	0.2093
TPSS/pc-1	1.0002	0.016	0.992	0.2068
TPSS/aug-pc-1	0.9776	0.0571	0.9946	0.1658
TPSS/pc-2	0.9941	0.0032	0.9965	0.1351
TPSS/aug-pc-2	0.9757	0.0044	0.9975	0.1127
TPSS/def2-TZVPP	0.9517	0.0045	0.9981	0.0948
TPSS/def2-TZVPPD	0.9624	-0.0058	0.9986	0.0819
B3LYP/6-31G(d)	0.9536	-0.0311	0.9905	0.2143
B3LYP/pc-1	1.0291	2e-04	0.9931	0.1975
B3LYP/aug-pc-1	1.0138	0.0389	0.9976	0.1139
B3LYP/pc-2	1.0313	-1e-04	0.9969	0.1318
B3LYP/aug-pc-2	1.0092	0.0022	0.9988	0.0811
B3LYP/def2-TZVPP	0.9884	2e-04	0.9983	0.0924
B3LYP/def2-TZVPPD	0.9977	-0.01	0.9995	0.0535
PBE0/6-31G(d)	0.951	-0.0341	0.9883	0.2376
PBE0/pc-1	1.0283	-2e-04	0.9925	0.2048
PBE0/aug-pc-1	0.9937	0.0459	0.9964	0.138
PBE0/pc-2	1.0111	-0.0114	0.9971	0.1249
PBE0/aug-pc-2	0.9885	-0.0072	0.9988	0.0797
PBE0/def2-TZVPP	0.9694	-0.0113	0.9971	0.1208
PBE0/def2-TZVPPD	0.9766	-0.0205	0.9983	0.093

Table A29: Linear regression parameters for Q_{zz}(R)

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.8897	0.0047	0.9695	0.3399
SVWN/pc-1	0.9569	-0.0092	0.9778	0.3102
SVWN/aug-pc-1	0.9659	-0.0295	0.9842	0.2631
SVWN/pc-2	0.9572	0.014	0.9882	0.2248
SVWN/aug-pc-2	0.9451	0.0179	0.9904	0.1999
SVWN/def2-TZVPP	0.9204	0.0226	0.9848	0.2466
SVWN/def2-TZVPPD	0.9401	0.0295	0.9864	0.2381
BLYP/6-31G(d)	0.9038	-0.0113	0.9821	0.2625
BLYP/pc-1	0.9731	-0.0296	0.9845	0.2632
BLYP/aug-pc-1	1.0007	-0.0438	0.9937	0.1713
BLYP/pc-2	0.9883	-0.0118	0.9948	0.1538
BLYP/aug-pc-2	0.9827	-0.0029	0.9965	0.1258
BLYP/def2-TZVPP	0.9479	-0.0042	0.9957	0.1336
BLYP/def2-TZVPPD	0.9757	0.0081	0.9975	0.1057
PBE/6-31G(d)	0.8994	-0.0145	0.9828	0.2563
PBE/pc-1	0.968	-0.0301	0.9854	0.2533
PBE/aug-pc-1	0.9815	-0.0515	0.9929	0.1793
PBE/pc-2	0.9688	-0.0063	0.9961	0.1303
PBE/aug-pc-2	0.9627	0.0016	0.9979	0.094
PBE/def2-TZVPP	0.9303	9e-04	0.9958	0.1305
PBE/def2-TZVPPD	0.9556	0.0134	0.9978	0.0971
TPSS/6-31G(d)	0.9246	-0.0271	0.985	0.2459
TPSS/pc-1	0.9933	-0.0439	0.9831	0.2805
TPSS/aug-pc-1	0.9988	-0.0703	0.9909	0.2056
TPSS/pc-2	0.986	-0.0207	0.9945	0.1574
TPSS/aug-pc-2	0.9803	-0.0158	0.9968	0.1204
TPSS/def2-TZVPP	0.9514	-0.0137	0.9958	0.1335
TPSS/def2-TZVPPD	0.9728	-0.0025	0.9984	0.0828
B3LYP/6-31G(d)	0.9382	-0.0021	0.9842	0.2559
B3LYP/pc-1	1.0092	-0.0236	0.9855	0.2636
B3LYP/aug-pc-1	1.0182	-0.0501	0.9955	0.1479
B3LYP/pc-2	1.0101	-0.0062	0.9954	0.148
B3LYP/aug-pc-2	0.9989	-0.0026	0.9987	0.0779
B3LYP/def2-TZVPP	0.9734	0.0036	0.9963	0.1279
B3LYP/def2-TZVPPD	0.9935	0.0122	0.9993	0.0553
PBE0/6-31G(d)	0.9432	-0.0057	0.9843	0.2564
PBE0/pc-1	1.0136	-0.0254	0.9856	0.2635
PBE0/aug-pc-1	1.0074	-0.0609	0.9941	0.1666
PBE0/pc-2	0.999	-0.0021	0.9956	0.1429
PBE0/aug-pc-2	0.9871	-2e-04	0.9991	0.0622
PBE0/def2-TZVPP	0.9644	0.0073	0.9955	0.1389
PBE0/def2-TZVPPD	0.9815	0.0158	0.9989	0.0708

Table A30: Summary statistics for Vol(R)

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-7.502	9.843	11.812	11.823(AlH_2)	-82.072(O^-)
SVWN/pc-1	5.412	9.136	11.677	38.796($SiCl_3$)	-70.599(O^-)
SVWN/aug-pc-1	10.030	10.030	4.870	31.380($SiCl_3$)	None
SVWN/pc-2	4.468	5.050	4.590	17.350(SiF_3)	-18.914(O^-)
SVWN/aug-pc-2	4.036	4.469	3.883	13.135(MgH)	-4.078(CHCCO)
SVWN/def2-TZVPP	3.143	4.237	5.199	14.348(MgH)	-36.499(O^-)
SVWN/def2-TZVPPD	5.404	5.446	3.260	16.710(SiF_3)	-2.434(NCS)
BLYP/6-31G(d)	-7.157	9.238	11.558	19.979($SiCl_3$)	-84.387(O^-)
BLYP/pc-1	5.211	9.164	12.552	49.728($SiCl_3$)	-72.813(O^-)
BLYP/aug-pc-1	12.592	12.611	6.968	43.629($SiCl_3$)	-1.139(BH_2)
BLYP/pc-2	7.439	7.882	5.593	22.036(SiF_3)	-24.761(O^-)
BLYP/aug-pc-2	7.756	7.770	4.196	18.074(SiF_3)	-0.822(CCCl)
BLYP/def2-TZVPP	6.054	7.106	6.362	20.478(SiF_3)	-41.164(O^-)
BLYP/def2-TZVPPD	9.028	9.028	4.289	22.448(SiF_3)	None
PBE/6-31G(d)	-8.214	9.851	11.320	14.008($SiCl_3$)	-83.476(O^-)
PBE/pc-1	4.210	8.503	12.009	44.645($SiCl_3$)	-72.088(O^-)
PBE/aug-pc-1	8.478	8.517	5.595	33.507($SiCl_3$)	-2.293(O^-)
PBE/pc-2	2.275	3.792	4.895	15.231(SiF_3)	-25.067(O^-)
PBE/aug-pc-2	2.564	3.658	3.905	13.433(SiF_3)	-5.957(CCCl)
PBE/def2-TZVPP	1.236	3.402	5.657	16.236(SiF_3)	-41.090(O^-)
PBE/def2-TZVPPD	4.145	4.367	3.453	16.859(SiF_3)	-4.002(NCS)
TPSS/6-31G(d)	-11.018	11.447	10.508	9.171($SiCl_3$)	-83.751(O^-)
TPSS/pc-1	0.946	7.321	11.776	40.582($SiCl_3$)	-72.614(O^-)
TPSS/aug-pc-1	2.916	4.662	5.596	24.341($SiCl_3$)	-7.116(CH_2COO^- — ip)
TPSS/pc-2	-3.941	4.612	4.173	8.019(SiF_3)	-26.657(O^-)
TPSS/aug-pc-2	-4.285	4.592	3.373	5.521(SiF_3)	-13.389(CCCl)
TPSS/def2-TZVPP	-4.554	4.883	4.860	10.071(SiF_3)	-42.334(O^-)
TPSS/def2-TZVPPD	-2.311	3.179	3.134	9.994(SiF_3)	-11.149(C_6H_5)
B3LYP/6-31G(d)	-10.676	11.371	11.154	11.447($SiCl_3$)	-83.621(O^-)
B3LYP/pc-1	1.763	7.678	12.048	41.974($SiCl_3$)	-72.156(O^-)
B3LYP/aug-pc-1	5.737	6.211	6.423	33.082($SiCl_3$)	-9.829(O^-)
B3LYP/pc-2	0.269	2.230	3.659	10.322(SiF_3)	-25.836(O^-)
B3LYP/aug-pc-2	0.227	1.769	2.271	6.694(SiF_3)	-5.862(CHCCO)
B3LYP/def2-TZVPP	-0.637	2.260	4.934	9.465(SiF_3)	-41.655(O^-)
B3LYP/def2-TZVPPD	1.967	2.222	2.447	11.515(SiF_3)	-2.244(CNO)
PBE0/6-31G(d)	-11.895	12.372	11.066	5.390($SiCl_3$)	-82.851(O^-)
PBE0/pc-1	0.464	7.761	11.857	37.137($SiCl_3$)	-71.654(O^-)
PBE0/aug-pc-1	0.891	4.760	5.937	21.500($SiCl_3$)	-13.772(O^-)
PBE0/pc-2	-5.670	6.055	4.058	3.930(MgH)	-26.460(O^-)
PBE0/aug-pc-2	-5.840	6.108	3.582	3.922(MgH)	-15.900(CCl_3)
PBE0/def2-TZVPP	-6.058	6.253	4.816	4.583(MgH)	-41.914(O^-)
PBE0/def2-TZVPPD	-3.745	4.000	2.502	4.778(SiF_3)	-11.797(C_6H_5)

Table A31: Linear regression parameters for Vol(R)

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	1.0084	-10.6924	0.9906	11.8211
SVWN/pc-1	1.0434	-11.1066	0.993	10.5085
SVWN/aug-pc-1	1.0234	1.1064	0.9989	4.001
SVWN/pc-2	1.0053	2.4664	0.9986	4.5667
SVWN/aug-pc-2	0.9991	4.3646	0.9989	3.8991
SVWN/def2-TZVPP	1.0046	1.405	0.9981	5.1932
SVWN/def2-TZVPPD	1.0021	4.6122	0.9993	3.2648
BLYP/6-31G(d)	1.0231	-15.9399	0.9916	11.2733
BLYP/pc-1	1.0549	-15.6786	0.9928	10.7499
BLYP/aug-pc-1	1.0417	-3.2704	0.9985	4.8964
BLYP/pc-2	1.0238	-1.6228	0.9985	4.836
BLYP/aug-pc-2	1.0183	0.7838	0.9991	3.596
BLYP/def2-TZVPP	1.0246	-3.3059	0.9979	5.6676
BLYP/def2-TZVPPD	1.0219	0.686	0.9992	3.4119
PBE/6-31G(d)	1.0149	-13.8943	0.9916	11.2281
PBE/pc-1	1.0479	-14.0185	0.9929	10.6047
PBE/aug-pc-1	1.0257	-1.316	0.9985	4.6952
PBE/pc-2	1.0051	0.3249	0.9984	4.8782
PBE/aug-pc-2	1	2.5815	0.9989	3.9224
PBE/def2-TZVPP	1.0077	-1.7092	0.9979	5.6049
PBE/def2-TZVPPD	1.0051	2.2123	0.9992	3.4144
TPSS/6-31G(d)	1.0092	-14.5162	0.9926	10.4962
TPSS/pc-1	1.0406	-14.5059	0.9926	10.7775
TPSS/aug-pc-1	1.0125	-1.8443	0.998	5.4167
TPSS/pc-2	0.9913	-0.6205	0.9988	4.0585
TPSS/aug-pc-2	0.986	1.0317	0.9994	2.944
TPSS/def2-TZVPP	0.9963	-3.162	0.9984	4.8614
TPSS/def2-TZVPPD	0.9933	0.2223	0.9993	3.0446
B3LYP/6-31G(d)	1.0109	-14.8418	0.9917	11.1253
B3LYP/pc-1	1.0432	-14.6959	0.9924	10.932
B3LYP/aug-pc-1	1.0237	-3.2809	0.9978	5.7905
B3LYP/pc-2	1.0046	-1.4891	0.9991	3.6331
B3LYP/aug-pc-2	0.9995	0.4096	0.9996	2.2799
B3LYP/def2-TZVPP	1.0071	-3.3263	0.9984	4.8823
B3LYP/def2-TZVPPD	1.0064	-0.4723	0.9996	2.334
PBE0/6-31G(d)	1.003	-13.0244	0.9916	11.1084
PBE0/pc-1	1.0365	-13.4342	0.9921	11.0736
PBE0/aug-pc-1	1.006	-1.3825	0.9976	5.9196
PBE0/pc-2	0.9858	-0.2454	0.999	3.6992
PBE0/aug-pc-2	0.9809	1.4239	0.9994	2.775
PBE0/def2-TZVPP	0.9908	-2.5391	0.9984	4.7079
PBE0/def2-TZVPPD	0.9896	0.2136	0.9997	2.1815

Table A32: Summary statistics for ρ_c

Model Chemistry	MD	MAD	SD	Max$^+$	Max$^-$
SVWN/6-31G(d)	-0.032	0.032	0.011	None	-0.054($OO^{\bullet}$)
SVWN/pc-1	-0.029	0.029	0.010	None	-0.047(OH_2^+)
SVWN/aug-pc-1	-0.028	0.028	0.009	None	-0.044($OO^{\bullet}$)
SVWN/pc-2	-0.018	0.018	0.006	None	-0.031(NO)
SVWN/aug-pc-2	-0.018	0.018	0.006	None	-0.032(NO)
SVWN/def2-TZVPP	-0.019	0.019	0.006	None	-0.034(NO)
SVWN/def2-TZVPPD	-0.019	0.019	0.006	None	-0.033(NO)
BLYP/6-31G(d)	-0.028	0.028	0.010	None	-0.056(O^-)
BLYP/pc-1	-0.024	0.024	0.008	None	-0.049(O^-)
BLYP/aug-pc-1	-0.022	0.022	0.007	None	-0.036($OO^{\bullet}$)
BLYP/pc-2	-0.010	0.010	0.003	None	-0.022(NO)
BLYP/aug-pc-2	-0.010	0.010	0.004	None	-0.023(NO)
BLYP/def2-TZVPP	-0.011	0.011	0.004	None	-0.026(NO)
BLYP/def2-TZVPPD	-0.011	0.011	0.004	None	-0.024(NO)
PBE/6-31G(d)	-0.028	0.028	0.009	None	-0.050(O^-)
PBE/pc-1	-0.025	0.025	0.007	None	-0.043(O^-)
PBE/aug-pc-1	-0.023	0.023	0.006	None	-0.035(NO)
PBE/pc-2	-0.013	0.013	0.003	None	-0.025(NO)
PBE/aug-pc-2	-0.013	0.013	0.003	None	-0.025(NO)
PBE/def2-TZVPP	-0.014	0.014	0.004	None	-0.028(NO)
PBE/def2-TZVPPD	-0.014	0.014	0.004	None	-0.027(NO)
TPSS/6-31G(d)	-0.023	0.023	0.008	None	-0.051(O^-)
TPSS/pc-1	-0.020	0.020	0.008	None	-0.044(O^-)
TPSS/aug-pc-1	-0.019	0.019	0.006	None	-0.032(OOH)
TPSS/pc-2	-0.008	0.008	0.003	None	-0.018(NO)
TPSS/aug-pc-2	-0.008	0.008	0.003	None	-0.018(NO)
TPSS/def2-TZVPP	-0.010	0.010	0.003	None	-0.021(NO)
TPSS/def2-TZVPPD	-0.009	0.009	0.003	None	-0.020(NO)
B3LYP/6-31G(d)	-0.021	0.021	0.007	None	-0.043(O^-)
B3LYP/pc-1	-0.017	0.017	0.006	None	-0.034(O^-)
B3LYP/aug-pc-1	-0.016	0.016	0.004	None	-0.023($OO^{\bullet}$)
B3LYP/pc-2	-0.003	0.003	0.001	0.000(O^-)	-0.009(SHO_2)
B3LYP/aug-pc-2	-0.004	0.004	0.001	0.000(O^-)	-0.009(SHO_2)
B3LYP/def2-TZVPP	-0.005	0.005	0.001	None	-0.009(NO)
B3LYP/def2-TZVPPD	-0.005	0.005	0.001	None	-0.008(NO)
PBE0/6-31G(d)	-0.021	0.021	0.006	None	-0.037(OH_2^+)
PBE0/pc-1	-0.016	0.016	0.004	None	-0.027(OH_2^+)
PBE0/aug-pc-1	-0.015	0.015	0.003	None	-0.020(CO)
PBE0/pc-2	-0.005	0.005	0.002	0.003(O^-)	-0.009(SHO_2)
PBE0/aug-pc-2	-0.005	0.005	0.002	0.002(O^-)	-0.009(SHO_2)
PBE0/def2-TZVPP	-0.006	0.006	0.002	None	-0.010(CF_3)
PBE0/def2-TZVPPD	-0.006	0.006	0.002	0.001(O^-)	-0.010(CF_3)

Table A33: Linear regression parameters for ρ_c

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.8667	0.0061	0.9963	0.004
SVWN/pc-1	0.8762	0.0061	0.9979	0.003
SVWN/aug-pc-1	0.8881	0.0038	0.9987	0.0024
SVWN/pc-2	0.9303	0.0023	0.9986	0.0026
SVWN/aug-pc-2	0.9274	0.0029	0.9986	0.0026
SVWN/def2-TZVPP	0.9228	0.003	0.9986	0.0026
SVWN/def2-TZVPPD	0.9204	0.0037	0.9987	0.0025
BLYP/6-31G(d)	0.8924	0.003	0.9943	0.0051
BLYP/pc-1	0.9065	0.0023	0.9958	0.0044
BLYP/aug-pc-1	0.9242	-7e-04	0.9979	0.0032
BLYP/pc-2	0.9714	-0.0019	0.9988	0.0026
BLYP/aug-pc-2	0.9682	-0.0011	0.9987	0.0026
BLYP/def2-TZVPP	0.9644	-0.0014	0.999	0.0023
BLYP/def2-TZVPPD	0.9626	-7e-04	0.9991	0.0021
PBE/6-31G(d)	0.8961	0.0013	0.9968	0.0038
PBE/pc-1	0.9119	4e-04	0.998	0.0031
PBE/aug-pc-1	0.9278	-0.0027	0.999	0.0023
PBE/pc-2	0.9716	-0.0047	0.9988	0.0026
PBE/aug-pc-2	0.9685	-0.004	0.9988	0.0025
PBE/def2-TZVPP	0.9647	-0.0042	0.999	0.0023
PBE/def2-TZVPPD	0.9633	-0.0036	0.9991	0.0022
TPSS/6-31G(d)	0.9059	0.0039	0.9953	0.0047
TPSS/pc-1	0.9171	0.0036	0.9956	0.0046
TPSS/aug-pc-1	0.9345	-1e-04	0.9977	0.0034
TPSS/pc-2	0.9788	-0.002	0.9992	0.0021
TPSS/aug-pc-2	0.9764	-0.0015	0.9991	0.0022
TPSS/def2-TZVPP	0.9709	-0.0014	0.9993	0.002
TPSS/def2-TZVPPD	0.9699	-9e-04	0.9993	0.0019
B3LYP/6-31G(d)	0.9201	0.0017	0.9968	0.0039
B3LYP/pc-1	0.9385	4e-04	0.9976	0.0035
B3LYP/aug-pc-1	0.9559	-0.0031	0.999	0.0023
B3LYP/pc-2	1.0033	-0.0043	0.9997	0.0014
B3LYP/aug-pc-2	1.0006	-0.0037	0.9997	0.0014
B3LYP/def2-TZVPP	0.9946	-0.0034	0.9998	0.001
B3LYP/def2-TZVPPD	0.9927	-0.0028	0.9998	0.001
PBE0/6-31G(d)	0.929	-3e-04	0.9986	0.0026
PBE0/pc-1	0.9504	-0.002	0.999	0.0023
PBE0/aug-pc-1	0.9668	-0.0058	0.9995	0.0016
PBE0/pc-2	1.0109	-0.0079	0.9992	0.0022
PBE0/aug-pc-2	1.0084	-0.0074	0.9993	0.002
PBE0/def2-TZVPP	1.002	-0.0069	0.9995	0.0018
PBE0/def2-TZVPPD	1.0003	-0.0064	0.9995	0.0018

Table A34: Summary statistics for r(H-BCP)

Model Chemistry	MD	MAD	SD	Max^{+}	Max^{-}
SVWN/6-31G(d)	-0.019	0.026	0.024	0.033(O^{-})	-0.089(SCH_3)
SVWN/pc-1	-0.020	0.030	0.029	0.071(S^{-})	-0.055(CNO)
SVWN/aug-pc-1	-0.020	0.029	0.027	0.046(SCH_3)	-0.054(CNO)
SVWN/pc-2	-0.010	0.013	0.012	0.014(SHO_2)	-0.053($PO(OH)_2$)
SVWN/aug-pc-2	-0.011	0.013	0.012	0.012($OO^{\bullet}$)	-0.057($PO(OH)_2$)
SVWN/def2-TZVPP	-0.017	0.019	0.016	0.028(BeH)	-0.060(SCH_3)
SVWN/def2-TZVPPD	-0.018	0.020	0.016	0.022(BeH)	-0.071(S^{-})
BLYP/6-31G(d)	0.002	0.014	0.019	0.068(O^{-})	-0.052(SCH_3)
BLYP/pc-1	0.004	0.022	0.033	0.123(S^{-})	-0.038(CO)
BLYP/aug-pc-1	0.004	0.021	0.030	0.109(SCH_3)	-0.035(CO)
BLYP/pc-2	0.016	0.018	0.017	0.069(SCN)	-0.026($PO(OH)_2$)
BLYP/aug-pc-2	0.014	0.017	0.015	0.063(SCN)	-0.029($PO(OH)_2$)
BLYP/def2-TZVPP	0.009	0.010	0.012	0.043(O^{-})	-0.005(CH_3)
BLYP/def2-TZVPPD	0.007	0.009	0.010	0.036(NO)	-0.005(CH_2COO^{-} — ip)
PBE/6-31G(d)	0.000	0.015	0.021	0.056(O^{-})	-0.063(SCH_3)
PBE/pc-1	0.003	0.021	0.031	0.116(S^{-})	-0.034(CO)
PBE/aug-pc-1	0.003	0.020	0.029	0.101(SCF_3)	-0.031(CO)
PBE/pc-2	0.013	0.015	0.013	0.052(SCN)	-0.023($PO(OH)_2$)
PBE/aug-pc-2	0.012	0.014	0.011	0.046(SCN)	-0.024($PO(OH)_2$)
PBE/def2-TZVPP	0.006	0.010	0.011	0.037(NO)	-0.010(CH_3)
PBE/def2-TZVPPD	0.004	0.009	0.011	0.034(NO)	-0.014(S^{-})
TPSS/6-31G(d)	0.003	0.011	0.016	0.066(O^{-})	-0.047(SCH_3)
TPSS/pc-1	0.007	0.020	0.039	0.155(S^{-})	-0.039(CO)
TPSS/aug-pc-1	0.006	0.020	0.037	0.150(SCH_3)	-0.037(CO)
TPSS/pc-2	0.016	0.017	0.020	0.095(S^{-})	-0.013($PO(OH)_2$)
TPSS/aug-pc-2	0.014	0.015	0.018	0.084(SH)	-0.012($PO(OH)_2$)
TPSS/def2-TZVPP	0.008	0.009	0.012	0.081(S^{-})	-0.004(CH_3)
TPSS/def2-TZVPPD	0.006	0.008	0.010	0.046(S^{-})	-0.004(CH_3)
B3LYP/6-31G(d)	-0.007	0.014	0.016	0.050(O^{-})	-0.056(SCH_3)
B3LYP/pc-1	-0.002	0.025	0.040	0.154(S^{-})	-0.052(CO)
B3LYP/aug-pc-1	-0.002	0.025	0.038	0.143(SCH_3)	-0.050(CO)
B3LYP/pc-2	0.008	0.012	0.020	0.086(S^{-})	-0.022($PO(OH)_2$)
B3LYP/aug-pc-2	0.007	0.011	0.018	0.074(SH)	-0.024($PO(OH)_2$)
B3LYP/def2-TZVPP	0.000	0.008	0.011	0.066(S^{-})	-0.016(ONO)
B3LYP/def2-TZVPPD	-0.001	0.007	0.009	0.029(S^{-})	-0.017(ONO)
PBE0/6-31G(d)	-0.009	0.016	0.018	0.039(O^{-})	-0.065(SCH_3)
PBE0/pc-1	-0.003	0.026	0.041	0.158(S^{-})	-0.050(CO)
PBE0/aug-pc-1	-0.003	0.025	0.039	0.147(SCH_3)	-0.049(CO)
PBE0/pc-2	0.006	0.009	0.018	0.083(S^{-})	-0.016($PO(OH)_2$)
PBE0/aug-pc-2	0.005	0.009	0.016	0.065(SH)	-0.016($PO(OH)_2$)
PBE0/def2-TZVPP	-0.002	0.009	0.012	0.064(S^{-})	-0.027(ONO)
PBE0/def2-TZVPPD	-0.004	0.009	0.011	0.026(BeH)	-0.028(ONO)

Table A35: Linear regression parameters for r(H-BCP)

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	1.0003	-0.0196	0.9944	0.0242
SVWN/pc-1	1.0339	-0.0461	0.9933	0.0272
SVWN/aug-pc-1	1.0307	-0.0443	0.9942	0.0252
SVWN/pc-2	0.9822	0.0037	0.9989	0.0105
SVWN/aug-pc-2	0.981	0.0034	0.999	0.0099
SVWN/def2-TZVPP	0.998	-0.0155	0.9976	0.0157
SVWN/def2-TZVPPD	0.9982	-0.0169	0.9975	0.016
BLYP/6-31G(d)	0.9919	0.0088	0.9966	0.0186
BLYP/pc-1	1.0312	-0.02	0.991	0.0315
BLYP/aug-pc-1	1.0269	-0.0171	0.9922	0.0292
BLYP/pc-2	0.9771	0.0336	0.9976	0.0153
BLYP/aug-pc-2	0.9759	0.0331	0.9982	0.0133
BLYP/def2-TZVPP	0.9866	0.0192	0.9989	0.0107
BLYP/def2-TZVPPD	0.9872	0.0171	0.9991	0.0095
PBE/6-31G(d)	1	5e-04	0.9959	0.0207
PBE/pc-1	1.0327	-0.0225	0.9922	0.0293
PBE/aug-pc-1	1.0298	-0.0206	0.9933	0.0271
PBE/pc-2	0.9858	0.0244	0.9986	0.0119
PBE/aug-pc-2	0.9855	0.0232	0.999	0.0102
PBE/def2-TZVPP	0.9985	0.0071	0.9987	0.0114
PBE/def2-TZVPPD	0.9991	0.005	0.9988	0.0111
TPSS/6-31G(d)	0.9938	0.0082	0.9975	0.016
TPSS/pc-1	1.0269	-0.0142	0.987	0.0378
TPSS/aug-pc-1	1.0263	-0.0139	0.9881	0.0361
TPSS/pc-2	0.9918	0.0222	0.9961	0.02
TPSS/aug-pc-2	0.9928	0.02	0.9967	0.0182
TPSS/def2-TZVPP	1.0056	0.0036	0.9986	0.0122
TPSS/def2-TZVPPD	1.0062	0.0015	0.9991	0.0097
B3LYP/6-31G(d)	1.0018	-0.0086	0.9975	0.0162
B3LYP/pc-1	1.039	-0.0326	0.9867	0.0387
B3LYP/aug-pc-1	1.0356	-0.0301	0.9881	0.0365
B3LYP/pc-2	0.9934	0.0136	0.996	0.0203
B3LYP/aug-pc-2	0.9926	0.0128	0.9968	0.018
B3LYP/def2-TZVPP	1.0062	-0.0044	0.9989	0.0108
B3LYP/def2-TZVPPD	1.0064	-0.0062	0.9993	0.0086
PBE0/6-31G(d)	1.0104	-0.0175	0.997	0.0178
PBE0/pc-1	1.0423	-0.0357	0.9868	0.0387
PBE0/aug-pc-1	1.0407	-0.0344	0.9881	0.0366
PBE0/pc-2	1.0055	0.0022	0.9969	0.0181
PBE0/aug-pc-2	1.0057	5e-04	0.9975	0.016
PBE0/def2-TZVPP	1.0211	-0.0187	0.999	0.0103
PBE0/def2-TZVPPD	1.0212	-0.0204	0.9993	0.0085

Table A36: Summary statistics for $\nabla^2\rho_c$

Model Chemistry	MD	MAD	SD	Max$^+$	Max$^-$
SVWN/6-31G(d)	0.386	0.401	0.349	1.602(OH_2^+)	-0.120(PH_2)
SVWN/pc-1	0.317	0.320	0.270	0.997(O^-)	-0.045(SiF)
SVWN/aug-pc-1	0.312	0.315	0.265	0.976($OO^\bullet$)	-0.036(SiF)
SVWN/pc-2	0.173	0.194	0.144	0.548($OO^\bullet$)	-0.086(SiF_3)
SVWN/aug-pc-2	0.170	0.193	0.142	0.552($OO^\bullet$)	-0.092(SiF_3)
SVWN/def2-TZVPP	0.198	0.207	0.128	0.764(OH_2^+)	-0.059(BH_2)
SVWN/def2-TZVPPD	0.194	0.204	0.129	0.755(OH_2^+)	-0.060(BH_2)
BLYP/6-31G(d)	0.391	0.396	0.352	1.655(OH_2^+)	-0.068(PH_2)
BLYP/pc-1	0.324	0.324	0.299	1.105(O^-)	-0.012(SiF)
BLYP/aug-pc-1	0.313	0.313	0.290	1.064(OH_2^+)	-0.006(AlH_2)
BLYP/pc-2	0.161	0.170	0.168	0.586($OO^\bullet$)	-0.044(SiF)
BLYP/aug-pc-2	0.160	0.171	0.168	0.598($OO^\bullet$)	-0.049(SiF_3)
BLYP/def2-TZVPP	0.183	0.184	0.109	0.697(OH_2^+)	-0.015(AlH_2)
BLYP/def2-TZVPPD	0.176	0.177	0.106	0.691(OH_2^+)	-0.015(SiF)
PBE/6-31G(d)	0.390	0.395	0.343	1.623(OH_2^+)	-0.067(PH_2)
PBE/pc-1	0.319	0.320	0.283	1.051(O^-)	-0.023(SiF)
PBE/aug-pc-1	0.313	0.314	0.274	0.996(OH_2^+)	-0.021(AlH_2)
PBE/pc-2	0.177	0.187	0.151	0.562($OO^\bullet$)	-0.050(SiF_3)
PBE/aug-pc-2	0.176	0.188	0.150	0.574($OO^\bullet$)	-0.054(SiF_3)
PBE/def2-TZVPP	0.196	0.197	0.108	0.687(OH_2^+)	-0.020(AlH_2)
PBE/def2-TZVPPD	0.190	0.191	0.106	0.679(OH_2^+)	-0.020(AlH_2)
TPSS/6-31G(d)	0.363	0.364	0.334	1.589(OH_2^+)	-0.006(MgH)
TPSS/pc-1	0.300	0.300	0.285	1.080(O^-)	None
TPSS/aug-pc-1	0.294	0.294	0.274	0.974(ONH_2)	None
TPSS/pc-2	0.153	0.153	0.149	0.535(OF)	-0.004(AlH_2)
TPSS/aug-pc-2	0.151	0.151	0.146	0.538($OO^\bullet$)	-0.005(AlH_2)
TPSS/def2-TZVPP	0.173	0.173	0.094	0.703(OH_2^+)	None
TPSS/def2-TZVPPD	0.167	0.167	0.091	0.695(OH_2^+)	None
B3LYP/6-31G(d)	0.334	0.335	0.318	1.523(OH_2^+)	-0.028(PH_2)
B3LYP/pc-1	0.253	0.253	0.235	0.965(O^-)	None
B3LYP/aug-pc-1	0.247	0.247	0.229	0.800(ONH_2)	None
B3LYP/pc-2	0.076	0.079	0.079	0.299(O^-)	-0.023(SiF_3)
B3LYP/aug-pc-2	0.075	0.079	0.076	0.267($OO^\bullet$)	-0.029(SiF_3)
B3LYP/def2-TZVPP	0.107	0.108	0.069	0.575(OH_2^+)	-0.086(NC)
B3LYP/def2-TZVPPD	0.101	0.105	0.072	0.567(OH_2^+)	-0.109(O^-)
PBE0/6-31G(d)	0.324	0.325	0.302	1.468(OH_2^+)	-0.016(PH_2)
PBE0/pc-1	0.236	0.236	0.208	0.891(O^-)	-0.007(AlH_2)
PBE0/aug-pc-1	0.235	0.235	0.202	0.716(ONH_2)	-0.006(AlH_2)
PBE0/pc-2	0.076	0.079	0.051	0.229(O^-)	-0.023(SiF_3)
PBE0/aug-pc-2	0.074	0.078	0.049	0.183(NF_2)	-0.028(SiF_3)
PBE0/def2-TZVPP	0.109	0.113	0.074	0.536(OH_2^+)	-0.116(NC)
PBE0/def2-TZVPPD	0.104	0.111	0.078	0.530(OH_2^+)	-0.168(O^-)

Table A37: Summary statistics for $\lambda_{1,c}$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	0.184	0.184	0.119	0.527(OH_2^+)	None
SVWN/pc-1	0.178	0.178	0.096	0.474(O^-)	None
SVWN/aug-pc-1	0.171	0.171	0.086	0.383($OO^\bullet$)	None
SVWN/pc-2	0.104	0.104	0.061	0.278($OO^\bullet$)	None
SVWN/aug-pc-2	0.104	0.104	0.061	0.285($OO^\bullet$)	None
SVWN/def2-TZVPP	0.110	0.110	0.063	0.295($OO^\bullet$)	None
SVWN/def2-TZVPPD	0.108	0.108	0.061	0.293($OO^\bullet$)	None
BLYP/6-31G(d)	0.181	0.181	0.135	0.601(O^-)	None
BLYP/pc-1	0.173	0.173	0.110	0.562(O^-)	None
BLYP/aug-pc-1	0.160	0.160	0.094	0.388(OH_2^+)	None
BLYP/pc-2	0.082	0.082	0.068	0.269(OH_2^+)	None
BLYP/aug-pc-2	0.082	0.082	0.068	0.270(OH_2^+)	None
BLYP/def2-TZVPP	0.092	0.092	0.072	0.296(O^-)	None
BLYP/def2-TZVPPD	0.089	0.089	0.068	0.278($OO^\bullet$)	None
PBE/6-31G(d)	0.183	0.183	0.125	0.558(O^-)	None
PBE/pc-1	0.175	0.175	0.101	0.520(O^-)	None
PBE/aug-pc-1	0.166	0.166	0.087	0.368($OO^\bullet$)	None
PBE/pc-2	0.099	0.099	0.060	0.252($OO^\bullet$)	None
PBE/aug-pc-2	0.099	0.099	0.060	0.261($OO^\bullet$)	None
PBE/def2-TZVPP	0.107	0.107	0.063	0.278(O^-)	None
PBE/def2-TZVPPD	0.104	0.104	0.060	0.273($OO^\bullet$)	None
TPSS/6-31G(d)	0.164	0.164	0.126	0.580(O^-)	None
TPSS/pc-1	0.158	0.158	0.104	0.542(O^-)	None
TPSS/aug-pc-1	0.150	0.150	0.089	0.357(ONH_2)	None
TPSS/pc-2	0.080	0.080	0.061	0.240(OH_2^+)	None
TPSS/aug-pc-2	0.079	0.079	0.061	0.241(OH_2^+)	None
TPSS/def2-TZVPP	0.089	0.089	0.063	0.279(O^-)	None
TPSS/def2-TZVPPD	0.086	0.086	0.059	0.244($OO^\bullet$)	None
B3LYP/6-31G(d)	0.144	0.144	0.106	0.518(O^-)	None
B3LYP/pc-1	0.133	0.133	0.077	0.462(O^-)	None
B3LYP/aug-pc-1	0.124	0.124	0.063	0.271(OH_2^+)	None
B3LYP/pc-2	0.043	0.043	0.029	0.124(O^-)	None
B3LYP/aug-pc-2	0.042	0.042	0.028	0.119(OH_2^+)	None
B3LYP/def2-TZVPP	0.052	0.052	0.031	0.194(O^-)	None
B3LYP/def2-TZVPPD	0.049	0.049	0.027	0.126(OH_2^+)	None
PBE0/6-31G(d)	0.141	0.141	0.092	0.469(O^-)	None
PBE0/pc-1	0.128	0.128	0.064	0.412(O^-)	None
PBE0/aug-pc-1	0.123	0.123	0.051	0.226(ONH_2)	None
PBE0/pc-2	0.052	0.052	0.017	0.098(O^-)	None
PBE0/aug-pc-2	0.051	0.051	0.016	0.086(NF_2)	None
PBE0/def2-TZVPP	0.059	0.059	0.021	0.161(O^-)	None
PBE0/def2-TZVPPD	0.057	0.057	0.019	0.094(NF_2)	None

Table A38: Summary statistics for $\lambda_{2,c}$

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	0.178	0.178	0.116	0.517(O^-)	None
SVWN/pc-1	0.175	0.175	0.095	0.474(O^-)	None
SVWN/aug-pc-1	0.167	0.167	0.085	0.372($OO^{\bullet}$)	None
SVWN/pc-2	0.101	0.101	0.058	0.268($OO^{\bullet}$)	None
SVWN/aug-pc-2	0.100	0.100	0.059	0.276($OO^{\bullet}$)	None
SVWN/def2-TZVPP	0.107	0.107	0.061	0.283($OO^{\bullet}$)	None
SVWN/def2-TZVPPD	0.104	0.104	0.059	0.281($OO^{\bullet}$)	None
BLYP/6-31G(d)	0.176	0.176	0.132	0.601(O^-)	None
BLYP/pc-1	0.171	0.171	0.110	0.562(O^-)	None
BLYP/aug-pc-1	0.158	0.158	0.094	0.387(OH_2^+)	None
BLYP/pc-2	0.081	0.081	0.066	0.265(OH_2^+)	None
BLYP/aug-pc-2	0.080	0.080	0.066	0.267(OH_2^+)	None
BLYP/def2-TZVPP	0.090	0.090	0.070	0.296(O^-)	None
BLYP/def2-TZVPPD	0.087	0.087	0.066	0.271($OO^{\bullet}$)	None
PBE/6-31G(d)	0.178	0.178	0.122	0.558(O^-)	None
PBE/pc-1	0.173	0.173	0.100	0.520(O^-)	None
PBE/aug-pc-1	0.163	0.163	0.086	0.360($OO^{\bullet}$)	None
PBE/pc-2	0.097	0.097	0.058	0.247($OO^{\bullet}$)	None
PBE/aug-pc-2	0.097	0.097	0.058	0.256($OO^{\bullet}$)	None
PBE/def2-TZVPP	0.105	0.105	0.062	0.278(O^-)	None
PBE/def2-TZVPPD	0.102	0.102	0.058	0.266($OO^{\bullet}$)	None
TPSS/6-31G(d)	0.159	0.159	0.123	0.580(O^-)	None
TPSS/pc-1	0.156	0.156	0.104	0.542(O^-)	None
TPSS/aug-pc-1	0.148	0.148	0.088	0.352(ONH_2)	None
TPSS/pc-2	0.078	0.078	0.060	0.236(OH_2^+)	None
TPSS/aug-pc-2	0.077	0.077	0.059	0.237(OH_2^+)	None
TPSS/def2-TZVPP	0.087	0.087	0.061	0.279(O^-)	None
TPSS/def2-TZVPPD	0.084	0.084	0.057	0.239($OO^{\bullet}$)	None
B3LYP/6-31G(d)	0.140	0.140	0.103	0.518(O^-)	None
B3LYP/pc-1	0.133	0.133	0.077	0.462(O^-)	None
B3LYP/aug-pc-1	0.123	0.123	0.063	0.272(OH_2^+)	None
B3LYP/pc-2	0.043	0.043	0.028	0.124(O^-)	None
B3LYP/aug-pc-2	0.042	0.042	0.028	0.116(OH_2^+)	None
B3LYP/def2-TZVPP	0.051	0.051	0.029	0.194(O^-)	None
B3LYP/def2-TZVPPD	0.048	0.048	0.025	0.122(OH_2^+)	None
PBE0/6-31G(d)	0.137	0.137	0.090	0.469(O^-)	None
PBE0/pc-1	0.128	0.128	0.064	0.412(O^-)	None
PBE0/aug-pc-1	0.122	0.122	0.052	0.228(OPH_2O)	None
PBE0/pc-2	0.051	0.051	0.017	0.098(O^-)	None
PBE0/aug-pc-2	0.051	0.051	0.016	0.087(NF_2)	None
PBE0/def2-TZVPP	0.058	0.058	0.021	0.161(O^-)	None
PBE0/def2-TZVPPD	0.056	0.056	0.018	0.096(NF_2)	None

Table A39: Summary statistics for $\lambda_{3,c}$

Model Chemistry	MD	MAD	SD	Max$^+$	Max$^-$
SVWN/6-31G(d)	0.024	0.117	0.162	0.558(OH_2^+)	-0.298(PH_2O)
SVWN/pc-1	-0.037	0.085	0.090	0.202($OO^\bullet$)	-0.137(SiF)
SVWN/aug-pc-1	-0.026	0.092	0.104	0.230($OCOCH_3$)	-0.131(SiCl)
SVWN/pc-2	-0.033	0.037	0.051	0.048(S^-)	-0.191($PO(OH)_2$)
SVWN/aug-pc-2	-0.034	0.039	0.054	0.060(C_6H_5)	-0.205($PO(OH)_2$)
SVWN/def2-TZVPP	-0.018	0.081	0.095	0.256(OH_2^+)	-0.296(O^-)
SVWN/def2-TZVPPD	-0.018	0.082	0.096	0.255(OH_2^+)	-0.293(O^-)
BLYP/6-31G(d)	0.034	0.119	0.146	0.564(OH_2^+)	-0.229(PH_2O)
BLYP/pc-1	-0.020	0.078	0.095	0.245(OCN)	-0.146(NO)
BLYP/aug-pc-1	-0.005	0.083	0.112	0.303(OCN)	-0.109(NO)
BLYP/pc-2	-0.003	0.030	0.045	0.088(OPH_2O)	-0.115(PHO(OH))
BLYP/aug-pc-2	-0.002	0.031	0.046	0.083(OPH_2O)	-0.128($PO(OH)_2$)
BLYP/def2-TZVPP	0.001	0.068	0.084	0.161(OH_2^+)	-0.210(O^-)
BLYP/def2-TZVPPD	0.000	0.069	0.085	0.159(OH_2^+)	-0.248(O^-)
PBE/6-31G(d)	0.029	0.117	0.150	0.575(OH_2^+)	-0.233(PH_2O)
PBE/pc-1	-0.028	0.084	0.096	0.237(OCN)	-0.154(NO)
PBE/aug-pc-1	-0.016	0.089	0.112	0.291(OCN)	-0.124(NO)
PBE/pc-2	-0.019	0.034	0.045	0.063($OO^\bullet$)	-0.136($PO(OH)_2$)
PBE/aug-pc-2	-0.019	0.034	0.045	0.057($OO^\bullet$)	-0.148($PO(OH)_2$)
PBE/def2-TZVPP	-0.015	0.064	0.080	0.191(OH_2^+)	-0.246(O^-)
PBE/def2-TZVPPD	-0.015	0.065	0.081	0.190(OH_2^+)	-0.272(O^-)
TPSS/6-31G(d)	0.040	0.105	0.139	0.590(OH_2^+)	-0.220(O^-)
TPSS/pc-1	-0.014	0.076	0.095	0.245(OCN)	-0.127(NO)
TPSS/aug-pc-1	-0.003	0.082	0.111	0.297(OCN)	-0.102(COF)
TPSS/pc-2	-0.005	0.026	0.032	0.080($OO^\bullet$)	-0.046(NO)
TPSS/aug-pc-2	-0.006	0.025	0.030	0.073($OO^\bullet$)	-0.046(SiF_3)
TPSS/def2-TZVPP	-0.002	0.054	0.072	0.228(OH_2^+)	-0.241(O^-)
TPSS/def2-TZVPPD	-0.003	0.054	0.073	0.227(OH_2^+)	-0.265(O^-)
B3LYP/6-31G(d)	0.050	0.112	0.153	0.609(OH_2^+)	-0.165(O^-)
B3LYP/pc-1	-0.013	0.079	0.097	0.247($OO^\bullet$)	-0.106(COF)
B3LYP/aug-pc-1	0.000	0.088	0.115	0.297($OCOCH_3$)	-0.106(COF)
B3LYP/pc-2	-0.009	0.024	0.026	0.051(O^-)	-0.053(PHO(OH))
B3LYP/aug-pc-2	-0.010	0.023	0.026	0.048(C_6H_5)	-0.066($PO(OH)_2$)
B3LYP/def2-TZVPP	0.003	0.056	0.073	0.320(OH_2^+)	-0.252(O^-)
B3LYP/def2-TZVPPD	0.004	0.056	0.074	0.319(OH_2^+)	-0.263(O^-)
PBE0/6-31G(d)	0.047	0.108	0.158	0.622(OH_2^+)	-0.146(PH_2)
PBE0/pc-1	-0.021	0.083	0.097	0.245($OO^\bullet$)	-0.113(COF)
PBE0/aug-pc-1	-0.011	0.092	0.114	0.279($OCOCH_3$)	-0.116(COF)
PBE0/pc-2	-0.027	0.030	0.023	0.033(O^-)	-0.097(OH_2^+)
PBE0/aug-pc-2	-0.028	0.032	0.023	0.028(C_6H_5)	-0.105(OH_2^+)
PBE0/def2-TZVPP	-0.009	0.049	0.074	0.376(OH_2^+)	-0.282(O^-)
PBE0/def2-TZVPPD	-0.008	0.049	0.074	0.372(OH_2^+)	-0.274(O^-)

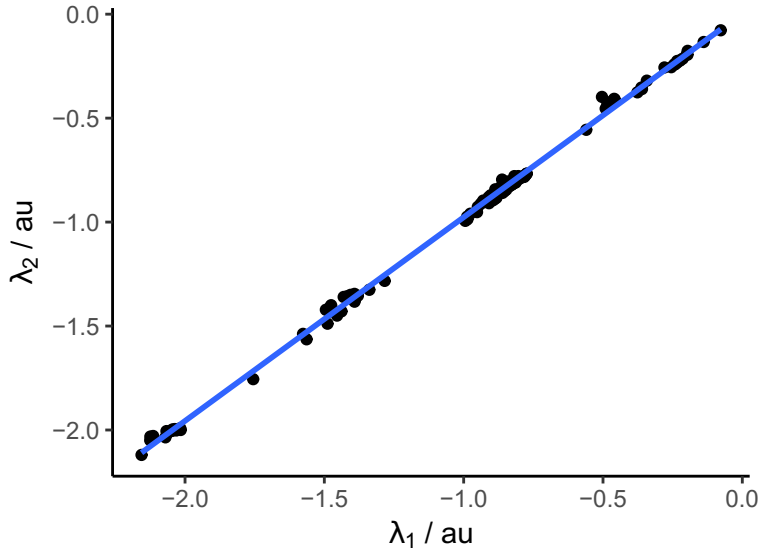

Figure A5: Comparison of λ_1 and λ_2 values determined using the B2PLYPD3-BJ/aug-cc-pV5Z model chemistry. As shown by the plot, they are essentially identical.

Table A40: Linear regression parameters for $\nabla^2\rho_c$

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.613	-0.1245	0.9385	0.1315
SVWN/pc-1	0.6955	-0.0848	0.9763	0.0908
SVWN/aug-pc-1	0.6984	-0.0853	0.9807	0.0821
SVWN/pc-2	0.8335	-0.047	0.9973	0.0364
SVWN/aug-pc-2	0.8363	-0.0456	0.9969	0.0388
SVWN/def2-TZVPP	0.8819	0.0428	0.9879	0.0817
SVWN/def2-TZVPPD	0.8852	0.0427	0.9866	0.0866
BLYP/6-31G(d)	0.6143	-0.1175	0.928	0.1435
BLYP/pc-1	0.6763	-0.103	0.9505	0.1295
BLYP/aug-pc-1	0.6847	-0.1029	0.9568	0.1219
BLYP/pc-2	0.8124	-0.0865	0.9917	0.0624
BLYP/aug-pc-2	0.8119	-0.0875	0.9924	0.0594
BLYP/def2-TZVPP	0.8841	0.03	0.9955	0.0497
BLYP/def2-TZVPPD	0.8898	0.0309	0.9952	0.0519
PBE/6-31G(d)	0.623	-0.107	0.936	0.1366
PBE/pc-1	0.6885	-0.0916	0.9641	0.1114
PBE/aug-pc-1	0.6954	-0.0889	0.9698	0.103
PBE/pc-2	0.8258	-0.0527	0.9967	0.0396
PBE/aug-pc-2	0.8257	-0.0533	0.997	0.0381
PBE/def2-TZVPP	0.8912	0.053	0.994	0.0583
PBE/def2-TZVPPD	0.8971	0.0547	0.993	0.0631
TPSS/6-31G(d)	0.6436	-0.1064	0.9261	0.1525
TPSS/pc-1	0.7032	-0.0911	0.946	0.1409
TPSS/aug-pc-1	0.7116	-0.0861	0.9531	0.1324
TPSS/pc-2	0.8395	-0.0588	0.9913	0.066
TPSS/aug-pc-2	0.8417	-0.0579	0.992	0.0634
TPSS/def2-TZVPP	0.907	0.0506	0.9952	0.053
TPSS/def2-TZVPPD	0.9133	0.0524	0.9948	0.0552
B3LYP/6-31G(d)	0.662	-0.1118	0.9338	0.1478
B3LYP/pc-1	0.759	-0.0651	0.9643	0.1225
B3LYP/aug-pc-1	0.7627	-0.0663	0.9688	0.1148
B3LYP/pc-2	0.9181	-0.032	0.9973	0.0399
B3LYP/aug-pc-2	0.9192	-0.032	0.9978	0.0358
B3LYP/def2-TZVPP	0.9636	0.0586	0.9941	0.0622
B3LYP/def2-TZVPPD	0.9678	0.0588	0.9933	0.0667
PBE0/6-31G(d)	0.6792	-0.0985	0.9422	0.1411
PBE0/pc-1	0.7857	-0.0465	0.9743	0.107
PBE0/aug-pc-1	0.7882	-0.0446	0.9787	0.0975
PBE0/pc-2	0.9533	0.0144	0.9983	0.0327
PBE0/aug-pc-2	0.9554	0.0151	0.9984	0.0325
PBE0/def2-TZVPP	0.9812	0.0837	0.9924	0.0722
PBE0/def2-TZVPPD	0.9853	0.0846	0.9912	0.0779

Table A41: Summary statistics for G_c

Model Chemistry	MD	MAD	SD	Max^+	Max^-
SVWN/6-31G(d)	-0.017	0.017	0.008	None	-0.050(PH_2O)
SVWN/pc-1	-0.013	0.013	0.006	0.004(S^-)	-0.027(PH_2)
SVWN/aug-pc-1	-0.013	0.013	0.005	0.001(SHO_2)	-0.027(PH_2)
SVWN/pc-2	-0.011	0.011	0.011	None	-0.057($PO(OH)_2$)
SVWN/aug-pc-2	-0.012	0.012	0.011	None	-0.060($PO(OH)_2$)
SVWN/def2-TZVPP	-0.011	0.011	0.006	None	-0.033(BH_2)
SVWN/def2-TZVPPD	-0.011	0.011	0.007	None	-0.033(BH_2)
BLYP/6-31G(d)	-0.014	0.014	0.006	None	-0.037(PH_2)
BLYP/pc-1	-0.009	0.011	0.006	0.010(S^-)	-0.018(PH_2)
BLYP/aug-pc-1	-0.010	0.011	0.005	0.008(SCH_3)	-0.018(PH_2)
BLYP/pc-2	-0.007	0.007	0.008	0.007(S^-)	-0.040($PO(OH)_2$)
BLYP/aug-pc-2	-0.007	0.007	0.008	0.003(SH)	-0.042($PO(OH)_2$)
BLYP/def2-TZVPP	-0.007	0.007	0.004	0.003(BeH)	-0.017(SiF)
BLYP/def2-TZVPPD	-0.007	0.007	0.004	0.002(BeH)	-0.017(PH_2O)
PBE/6-31G(d)	-0.014	0.014	0.006	None	-0.036(PH_2)
PBE/pc-1	-0.009	0.011	0.006	0.011(S^-)	-0.020(SiCl)
PBE/aug-pc-1	-0.010	0.011	0.005	0.007(SCH_3)	-0.019(PH_2)
PBE/pc-2	-0.007	0.008	0.009	0.006(S^-)	-0.043($PO(OH)_2$)
PBE/aug-pc-2	-0.008	0.008	0.009	0.001(SH)	-0.045($PO(OH)_2$)
PBE/def2-TZVPP	-0.007	0.007	0.005	None	-0.019(PH_2O)
PBE/def2-TZVPPD	-0.008	0.008	0.005	None	-0.019(PH_2O)
TPSS/6-31G(d)	-0.012	0.012	0.005	0.002($PO(OH)_2$)	-0.025(OCH_3)
TPSS/pc-1	-0.007	0.010	0.008	0.023(S^-)	-0.015(CNO)
TPSS/aug-pc-1	-0.008	0.010	0.007	0.020(SCH_3)	-0.016(O^-)
TPSS/pc-2	-0.006	0.007	0.005	0.018(S^-)	-0.022($PO(OH)_2$)
TPSS/aug-pc-2	-0.006	0.007	0.005	0.011(S^-)	-0.022($PO(OH)_2$)
TPSS/def2-TZVPP	-0.006	0.006	0.004	0.013(S^-)	-0.014(OCH_3)
TPSS/def2-TZVPPD	-0.006	0.007	0.004	0.007($PO(OH)_2$)	-0.015(OCH_3)
B3LYP/6-31G(d)	-0.013	0.013	0.005	None	-0.025(OCH_3)
B3LYP/pc-1	-0.007	0.010	0.008	0.025(S^-)	-0.017(CNO)
B3LYP/aug-pc-1	-0.008	0.010	0.007	0.019(SCH_3)	-0.016(CNO)
B3LYP/pc-2	-0.005	0.006	0.006	0.018(S^-)	-0.026($PO(OH)_2$)
B3LYP/aug-pc-2	-0.006	0.007	0.006	0.010(S^-)	-0.028($PO(OH)_2$)
B3LYP/def2-TZVPP	-0.005	0.006	0.004	0.011(S^-)	-0.014(OCH_3)
B3LYP/def2-TZVPPD	-0.006	0.006	0.003	0.005(BeH)	-0.014(OCH_3)
PBE0/6-31G(d)	-0.013	0.013	0.005	0.002($PO(OH)_2$)	-0.025(OCH_3)
PBE0/pc-1	-0.007	0.010	0.008	0.030(S^-)	-0.017(CNO)
PBE0/aug-pc-1	-0.008	0.010	0.007	0.022(S^-)	-0.016(CNO)
PBE0/pc-2	-0.006	0.007	0.006	0.019(S^-)	-0.026($PO(OH)_2$)
PBE0/aug-pc-2	-0.006	0.007	0.006	0.011(S^-)	-0.027($PO(OH)_2$)
PBE0/def2-TZVPP	-0.006	0.006	0.004	0.012(S^-)	-0.014(OCH_3)
PBE0/def2-TZVPPD	-0.006	0.006	0.003	0.004($PO(OH)_2$)	-0.015(OCH_3)

Table A42: Summary Statistics for V_C

Model Chemistry	MD	MAD	SD	Max$^+$	Max$^-$
SVWN/6-31G(d)	0.131	0.131	0.090	0.441(OH_2^+)	None
SVWN/pc-1	0.104	0.104	0.067	0.268($OO^\bullet$)	None
SVWN/aug-pc-1	0.105	0.105	0.067	0.272($OO^\bullet$)	None
SVWN/pc-2	0.066	0.066	0.030	0.157($OO^\bullet$)	None
SVWN/aug-pc-2	0.066	0.066	0.030	0.158($OO^\bullet$)	None
SVWN/def2-TZVPP	0.072	0.072	0.027	0.213(OH_2^+)	None
SVWN/def2-TZVPPD	0.071	0.071	0.027	0.211(OH_2^+)	None
BLYP/6-31G(d)	0.127	0.127	0.092	0.452(OH_2^+)	None
BLYP/pc-1	0.100	0.100	0.076	0.286(ONH_2)	None
BLYP/aug-pc-1	0.098	0.098	0.075	0.285(OH_2^+)	None
BLYP/pc-2	0.054	0.054	0.041	0.160(OF)	None
BLYP/aug-pc-2	0.055	0.055	0.040	0.162($OO^\bullet$)	None
BLYP/def2-TZVPP	0.059	0.059	0.026	0.191(OH_2^+)	None
BLYP/def2-TZVPPD	0.059	0.059	0.025	0.190(OH_2^+)	None
PBE/6-31G(d)	0.126	0.126	0.089	0.442(OH_2^+)	None
PBE/pc-1	0.099	0.099	0.071	0.270(ONH_2)	None
PBE/aug-pc-1	0.098	0.098	0.070	0.268($OO^\bullet$)	None
PBE/pc-2	0.059	0.059	0.034	0.152($OO^\bullet$)	None
PBE/aug-pc-2	0.060	0.060	0.034	0.156($OO^\bullet$)	None
PBE/def2-TZVPP	0.064	0.064	0.024	0.188(OH_2^+)	None
PBE/def2-TZVPPD	0.063	0.063	0.023	0.187(OH_2^+)	None
TPSS/6-31G(d)	0.115	0.115	0.092	0.434(OH_2^+)	None
TPSS/pc-1	0.089	0.089	0.076	0.274(ONH_2)	None
TPSS/aug-pc-1	0.089	0.089	0.074	0.270(ONH_2)	None
TPSS/pc-2	0.050	0.050	0.039	0.150(OF)	None
TPSS/aug-pc-2	0.050	0.050	0.039	0.150(OF)	None
TPSS/def2-TZVPP	0.055	0.055	0.029	0.194(OH_2^+)	-0.006(S^-)
TPSS/def2-TZVPPD	0.054	0.054	0.028	0.193(OH_2^+)	None
B3LYP/6-31G(d)	0.109	0.109	0.086	0.418(OH_2^+)	None
B3LYP/pc-1	0.078	0.078	0.063	0.241(O^-)	None
B3LYP/aug-pc-1	0.077	0.077	0.062	0.225(ONH_2)	None
B3LYP/pc-2	0.029	0.029	0.021	0.079($OO^\bullet$)	None
B3LYP/aug-pc-2	0.030	0.030	0.021	0.080($OO^\bullet$)	None
B3LYP/def2-TZVPP	0.037	0.038	0.020	0.162(OH_2^+)	-0.008(S^-)
B3LYP/def2-TZVPPD	0.037	0.037	0.020	0.160(OH_2^+)	-0.008(NC)
PBE0/6-31G(d)	0.106	0.106	0.082	0.403(OH_2^+)	None
PBE0/pc-1	0.073	0.073	0.055	0.217(O^-)	None
PBE0/aug-pc-1	0.074	0.074	0.054	0.201(ONH_2)	None
PBE0/pc-2	0.030	0.030	0.013	0.056($PO(OH)_2$)	0.000(S^-)
PBE0/aug-pc-2	0.030	0.030	0.012	0.057($PO(OH)_2$)	None
PBE0/def2-TZVPP	0.038	0.039	0.020	0.151(OH_2^+)	-0.015(NC)
PBE0/def2-TZVPPD	0.038	0.039	0.020	0.150(OH_2^+)	-0.018(O^-)

Table A43: Summary statistics for H_c

Model Chemistry	MD	MAD	SD	Max$^+$	Max$^-$
SVWN/6-31G(d)	0.114	0.114	0.088	0.421(OH_2^+)	None
SVWN/pc-1	0.092	0.092	0.067	0.258($OO^\bullet$)	None
SVWN/aug-pc-1	0.091	0.091	0.066	0.258($OO^\bullet$)	None
SVWN/pc-2	0.054	0.054	0.032	0.147($OO^\bullet$)	None
SVWN/aug-pc-2	0.054	0.054	0.031	0.148($OO^\bullet$)	None
SVWN/def2-TZVPP	0.061	0.061	0.029	0.202(OH_2^+)	None
SVWN/def2-TZVPPD	0.060	0.060	0.029	0.200(OH_2^+)	None
BLYP/6-31G(d)	0.112	0.112	0.090	0.433(OH_2^+)	None
BLYP/pc-1	0.090	0.090	0.075	0.280(O^-)	None
BLYP/aug-pc-1	0.088	0.088	0.074	0.275(OH_2^+)	None
BLYP/pc-2	0.047	0.047	0.041	0.153(OF)	None
BLYP/aug-pc-2	0.047	0.047	0.040	0.156($OO^\bullet$)	None
BLYP/def2-TZVPP	0.053	0.053	0.026	0.183(OH_2^+)	None
BLYP/def2-TZVPPD	0.051	0.051	0.025	0.181(OH_2^+)	None
PBE/6-31G(d)	0.112	0.112	0.087	0.424(OH_2^+)	None
PBE/pc-1	0.089	0.089	0.070	0.264(O^-)	None
PBE/aug-pc-1	0.088	0.088	0.069	0.258($OO^\bullet$)	None
PBE/pc-2	0.052	0.052	0.035	0.146($OO^\bullet$)	None
PBE/aug-pc-2	0.052	0.052	0.035	0.150($OO^\bullet$)	None
PBE/def2-TZVPP	0.057	0.057	0.025	0.180(OH_2^+)	None
PBE/def2-TZVPPD	0.055	0.055	0.024	0.178(OH_2^+)	None
TPSS/6-31G(d)	0.103	0.103	0.087	0.416(OH_2^+)	None
TPSS/pc-1	0.082	0.082	0.073	0.272(O^-)	None
TPSS/aug-pc-1	0.081	0.081	0.071	0.257(ONH_2)	None
TPSS/pc-2	0.044	0.044	0.038	0.142(OF)	None
TPSS/aug-pc-2	0.044	0.044	0.037	0.142($OO^\bullet$)	None
TPSS/def2-TZVPP	0.049	0.049	0.026	0.185(OH_2^+)	None
TPSS/def2-TZVPPD	0.048	0.048	0.025	0.183(OH_2^+)	None
B3LYP/6-31G(d)	0.096	0.096	0.083	0.399(OH_2^+)	None
B3LYP/pc-1	0.071	0.071	0.060	0.241(O^-)	None
B3LYP/aug-pc-1	0.070	0.070	0.059	0.212(ONH_2)	None
B3LYP/pc-2	0.024	0.024	0.020	0.075(O^-)	None
B3LYP/aug-pc-2	0.024	0.024	0.019	0.073($OO^\bullet$)	None
B3LYP/def2-TZVPP	0.032	0.032	0.018	0.153(OH_2^+)	-0.015(NC)
B3LYP/def2-TZVPPD	0.031	0.032	0.019	0.151(OH_2^+)	-0.015(O^-)
PBE0/6-31G(d)	0.094	0.094	0.079	0.385(OH_2^+)	None
PBE0/pc-1	0.066	0.066	0.053	0.220(O^-)	None
PBE0/aug-pc-1	0.066	0.066	0.052	0.190(ONH_2)	None
PBE0/pc-2	0.025	0.025	0.011	0.057(O^-)	None
PBE0/aug-pc-2	0.024	0.024	0.011	0.051(NF_2)	None
PBE0/def2-TZVPP	0.033	0.033	0.019	0.143(OH_2^+)	-0.022(NC)
PBE0/def2-TZVPPD	0.032	0.033	0.020	0.141(OH_2^+)	-0.030(O^-)

Table A44: Linear regression parameters for G_c

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.7756	-0.0032	0.9475	0.0053
SVWN/pc-1	0.8806	-0.0051	0.9679	0.0046
SVWN/aug-pc-1	0.8718	-0.0052	0.9793	0.0037
SVWN/pc-2	0.676	0.0093	0.9389	0.005
SVWN/aug-pc-2	0.6608	0.0099	0.9396	0.0048
SVWN/def2-TZVPP	0.7943	0.0018	0.9871	0.0026
SVWN/def2-TZVPPD	0.785	0.0021	0.9892	0.0024
BLYP/6-31G(d)	0.8517	-0.005	0.9694	0.0044
BLYP/pc-1	0.9574	-0.0068	0.9604	0.0056
BLYP/aug-pc-1	0.9519	-0.0069	0.9681	0.005
BLYP/pc-2	0.7634	0.0082	0.9599	0.0045
BLYP/aug-pc-2	0.7507	0.0086	0.9657	0.0041
BLYP/def2-TZVPP	0.8879	2e-04	0.9917	0.0023
BLYP/def2-TZVPPD	0.8795	4e-04	0.9943	0.0019
PBE/6-31G(d)	0.849	-0.0049	0.9718	0.0042
PBE/pc-1	0.9233	-0.0046	0.9595	0.0055
PBE/aug-pc-1	0.9154	-0.0047	0.9694	0.0047
PBE/pc-2	0.7369	0.0092	0.9561	0.0046
PBE/aug-pc-2	0.7266	0.0095	0.9622	0.0042
PBE/def2-TZVPP	0.8636	0.0012	0.9922	0.0022
PBE/def2-TZVPPD	0.8552	0.0013	0.9949	0.0018
TPSS/6-31G(d)	0.9995	-0.0119	0.9667	0.0054
TPSS/pc-1	1.0581	-0.0108	0.9398	0.0077
TPSS/aug-pc-1	1.0476	-0.0107	0.9444	0.0073
TPSS/pc-2	0.8935	0.001	0.9716	0.0044
TPSS/aug-pc-2	0.8887	9e-04	0.9778	0.0039
TPSS/def2-TZVPP	1.0399	-0.0082	0.9827	0.004
TPSS/def2-TZVPPD	1.0319	-0.0081	0.9839	0.0038
B3LYP/6-31G(d)	0.9683	-0.011	0.9705	0.0049
B3LYP/pc-1	1.0623	-0.0113	0.9398	0.0078
B3LYP/aug-pc-1	1.0543	-0.0113	0.9477	0.0072
B3LYP/pc-2	0.8581	0.0037	0.9671	0.0046
B3LYP/aug-pc-2	0.8456	0.0041	0.9749	0.0039
B3LYP/def2-TZVPP	0.9991	-0.0054	0.9852	0.0035
B3LYP/def2-TZVPPD	0.9912	-0.0052	0.9874	0.0032
PBE0/6-31G(d)	0.9892	-0.012	0.9698	0.005
PBE0/pc-1	1.0508	-0.0101	0.9322	0.0082
PBE0/aug-pc-1	1.0396	-0.0101	0.9423	0.0074
PBE0/pc-2	0.8532	0.0038	0.9684	0.0045
PBE0/aug-pc-2	0.8435	0.004	0.9764	0.0038
PBE0/def2-TZVPP	1.0008	-0.0057	0.9849	0.0036
PBE0/def2-TZVPPD	0.993	-0.0056	0.9875	0.0032

Table A45: Linear regression parameters for V_c

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.5756	-0.0624	0.9368	0.03
SVWN/pc-1	0.6698	-0.0461	0.9916	0.0123
SVWN/aug-pc-1	0.6697	-0.0459	0.9928	0.0114
SVWN/pc-2	0.8565	1e-04	0.9962	0.0106
SVWN/aug-pc-2	0.862	0.0028	0.9959	0.0111
SVWN/def2-TZVPP	0.8997	0.0263	0.9896	0.0184
SVWN/def2-TZVPPD	0.903	0.0273	0.989	0.0191
BLYP/6-31G(d)	0.5681	-0.0704	0.9261	0.0322
BLYP/pc-1	0.6268	-0.0703	0.9821	0.0169
BLYP/aug-pc-1	0.6334	-0.0692	0.9839	0.0162
BLYP/pc-2	0.8115	-0.0322	0.9912	0.0153
BLYP/aug-pc-2	0.8129	-0.0308	0.9912	0.0154
BLYP/def2-TZVPP	0.8862	0.0076	0.9951	0.0124
BLYP/def2-TZVPPD	0.8923	0.0094	0.9948	0.0129
PBE/6-31G(d)	0.5823	-0.0642	0.9327	0.0314
PBE/pc-1	0.653	-0.0595	0.9877	0.0146
PBE/aug-pc-1	0.6581	-0.0577	0.9892	0.0138
PBE/pc-2	0.8389	-0.0144	0.9953	0.0116
PBE/aug-pc-2	0.8404	-0.0131	0.9951	0.0118
PBE/def2-TZVPP	0.9038	0.0202	0.9941	0.0139
PBE/def2-TZVPPD	0.9101	0.0222	0.9936	0.0146
TPSS/6-31G(d)	0.5711	-0.081	0.9245	0.0327
TPSS/pc-1	0.6312	-0.0789	0.9817	0.0173
TPSS/aug-pc-1	0.6387	-0.0758	0.9846	0.016
TPSS/pc-2	0.8122	-0.036	0.9946	0.0119
TPSS/aug-pc-2	0.8155	-0.0342	0.9947	0.012
TPSS/def2-TZVPP	0.877	-0.0014	0.993	0.0148
TPSS/def2-TZVPPD	0.8841	9e-04	0.9927	0.0151
B3LYP/6-31G(d)	0.5985	-0.0738	0.9319	0.0324
B3LYP/pc-1	0.6959	-0.0608	0.9873	0.0158
B3LYP/aug-pc-1	0.6991	-0.0598	0.9897	0.0143
B3LYP/pc-2	0.9071	-0.0129	0.9967	0.0105
B3LYP/aug-pc-2	0.9112	-0.0105	0.9968	0.0104
B3LYP/def2-TZVPP	0.9463	0.013	0.9922	0.0168
B3LYP/def2-TZVPPD	0.9514	0.0147	0.9918	0.0174
PBE0/6-31G(d)	0.6182	-0.0677	0.9388	0.0316
PBE0/pc-1	0.7327	-0.0491	0.9895	0.0152
PBE0/aug-pc-1	0.7353	-0.0468	0.9922	0.0131
PBE0/pc-2	0.9541	0.0091	0.9976	0.0093
PBE0/aug-pc-2	0.959	0.0117	0.9976	0.0094
PBE0/def2-TZVPP	0.9699	0.0247	0.9906	0.019
PBE0/def2-TZVPPD	0.975	0.0266	0.99	0.0196

Table A46: Linear regression parameters for H_c

Model Chemistry	Slope	Intercept	r^2	SE
SVWN/6-31G(d)	0.5915	-0.0467	0.9383	0.0308
SVWN/pc-1	0.6792	-0.0342	0.9852	0.0169
SVWN/aug-pc-1	0.6809	-0.0339	0.9875	0.0155
SVWN/pc-2	0.8479	-0.0054	0.9982	0.0073
SVWN/aug-pc-2	0.8524	-0.0039	0.9979	0.0079
SVWN/def2-TZVPP	0.8924	0.0185	0.9889	0.0192
SVWN/def2-TZVPPD	0.8959	0.0191	0.9879	0.0201
BLYP/6-31G(d)	0.587	-0.0501	0.9277	0.0333
BLYP/pc-1	0.6465	-0.0485	0.9673	0.0241
BLYP/aug-pc-1	0.6543	-0.0477	0.971	0.0229
BLYP/pc-2	0.813	-0.0265	0.9921	0.0147
BLYP/aug-pc-2	0.8137	-0.0259	0.9924	0.0144
BLYP/def2-TZVPP	0.8851	0.0074	0.9954	0.0122
BLYP/def2-TZVPPD	0.8912	0.0086	0.995	0.0128
PBE/6-31G(d)	0.5986	-0.0458	0.9352	0.032
PBE/pc-1	0.6664	-0.0418	0.9774	0.0206
PBE/aug-pc-1	0.6728	-0.0403	0.9805	0.0193
PBE/pc-2	0.834	-0.0136	0.9969	0.0094
PBE/aug-pc-2	0.8349	-0.013	0.9969	0.0095
PBE/def2-TZVPP	0.8979	0.0165	0.9941	0.0141
PBE/def2-TZVPPD	0.9043	0.0178	0.9933	0.0151
TPSS/6-31G(d)	0.6011	-0.054	0.9279	0.034
TPSS/pc-1	0.6609	-0.0511	0.966	0.0252
TPSS/aug-pc-1	0.6693	-0.0487	0.9705	0.0237
TPSS/pc-2	0.8251	-0.0248	0.9932	0.0138
TPSS/aug-pc-2	0.828	-0.0238	0.9934	0.0137
TPSS/def2-TZVPP	0.8897	0.0057	0.9948	0.0131
TPSS/def2-TZVPPD	0.8966	0.0071	0.9943	0.0138
B3LYP/6-31G(d)	0.6248	-0.051	0.9344	0.0336
B3LYP/pc-1	0.7221	-0.0386	0.9781	0.0219
B3LYP/aug-pc-1	0.7258	-0.0382	0.9812	0.0204
B3LYP/pc-2	0.914	-0.0095	0.9976	0.0091
B3LYP/aug-pc-2	0.9168	-0.0084	0.9978	0.0087
B3LYP/def2-TZVPP	0.9544	0.0142	0.9934	0.0158
B3LYP/def2-TZVPPD	0.9594	0.0151	0.9926	0.0168
PBE0/6-31G(d)	0.6432	-0.0464	0.9426	0.0322
PBE0/pc-1	0.7545	-0.0305	0.9848	0.0191
PBE0/aug-pc-1	0.7573	-0.029	0.9878	0.0171
PBE0/pc-2	0.9558	0.0071	0.9987	0.0069
PBE0/aug-pc-2	0.9594	0.0085	0.9986	0.0072
PBE0/def2-TZVPP	0.9753	0.0231	0.9916	0.0182
PBE0/def2-TZVPPD	0.9801	0.0242	0.9907	0.0193

Appendix B

Appendix B

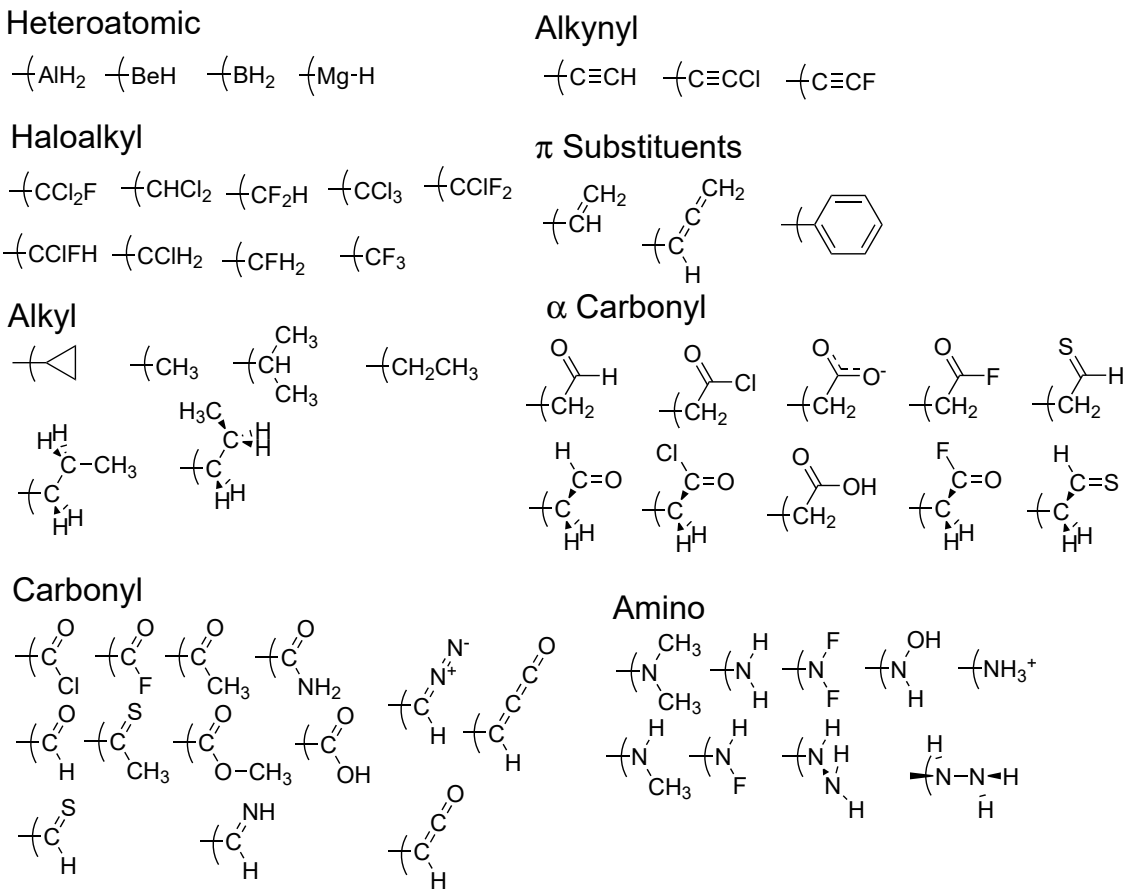

Figure B1: Part 1/2 of Substituents used in analysis

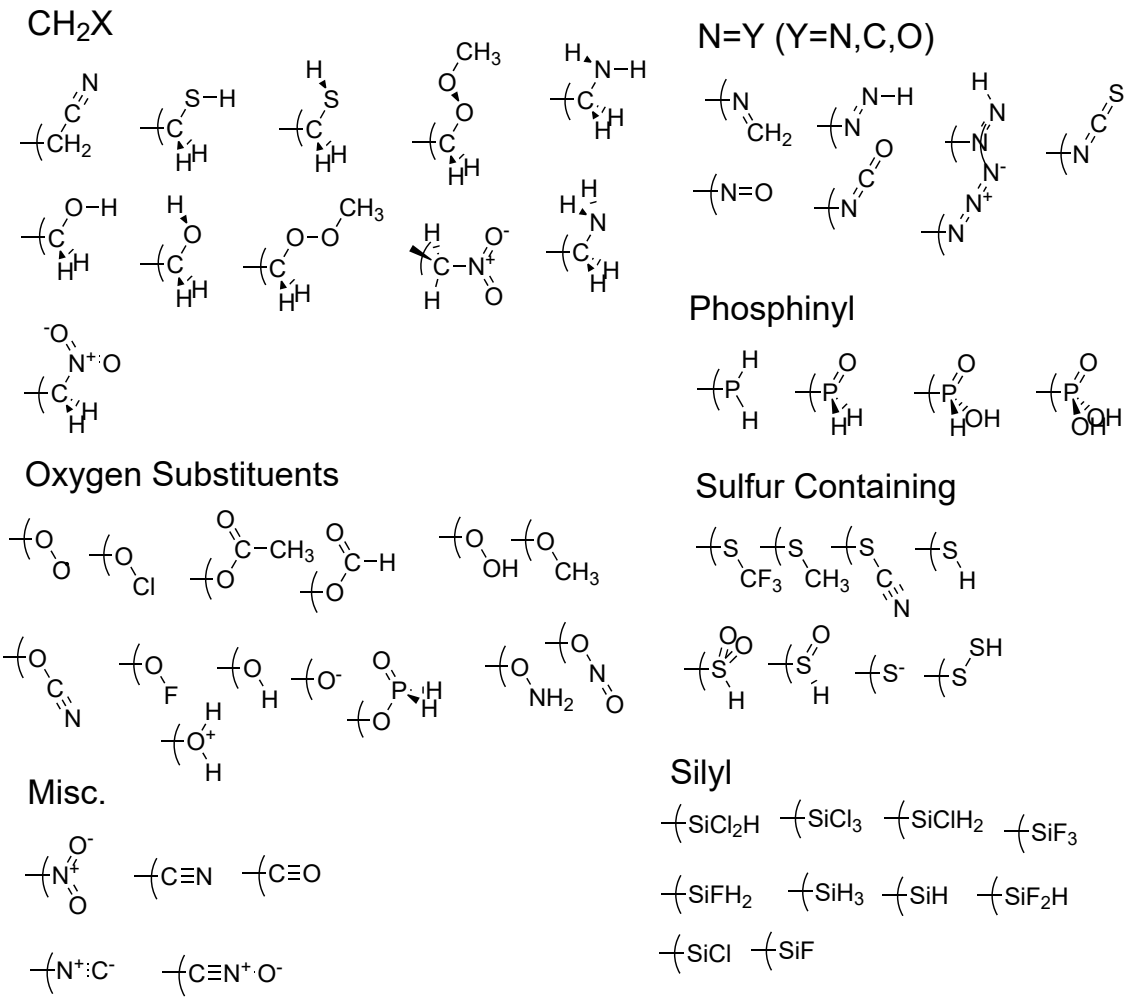

Figure B2: Part 2/2 of Substituents used in analysis

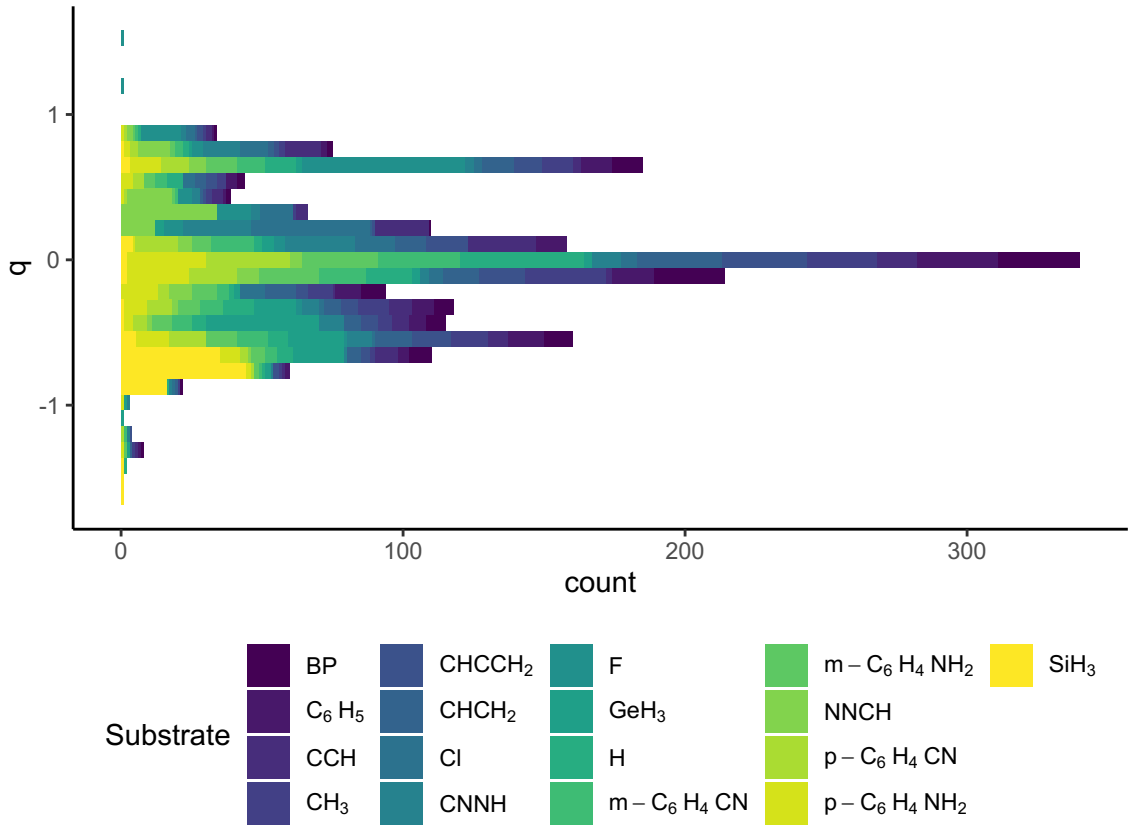

Figure B3: Histogram depicting range of q(R)

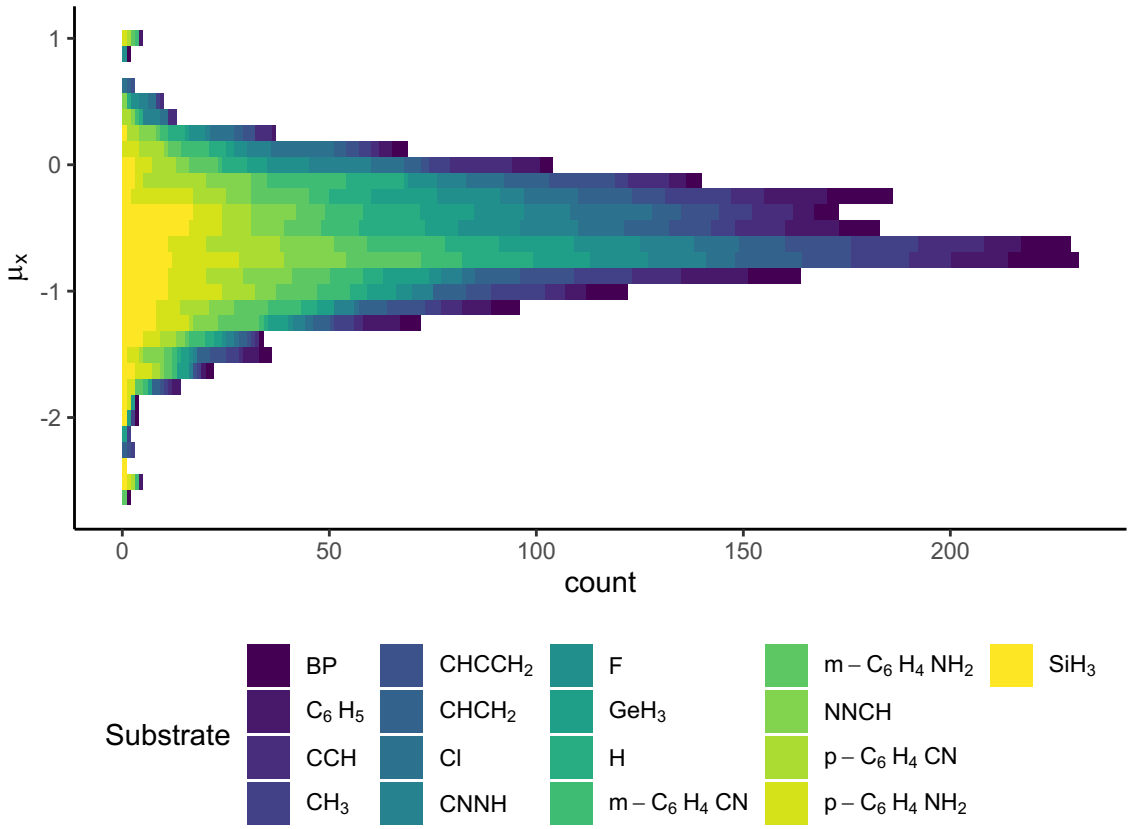

Figure B4: Histogram depicting range of μ_x(R)

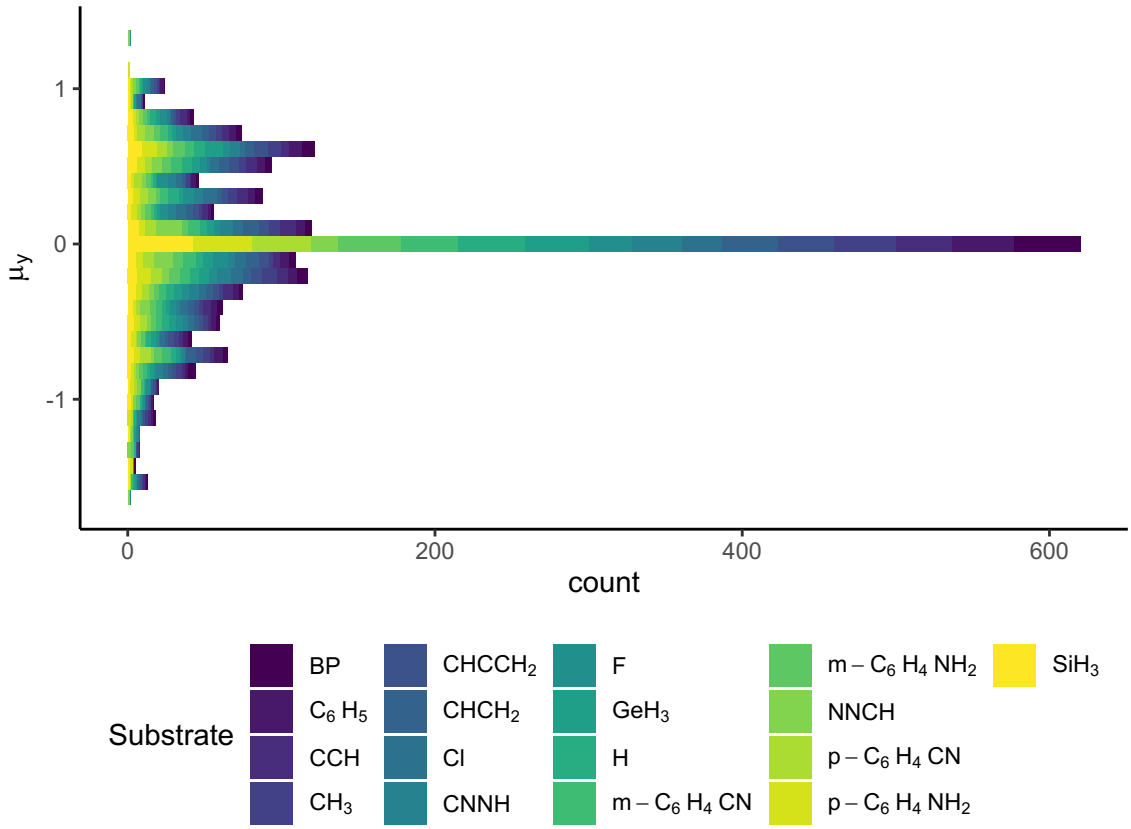

Figure B5: Histogram depicting range of μ_y(R)

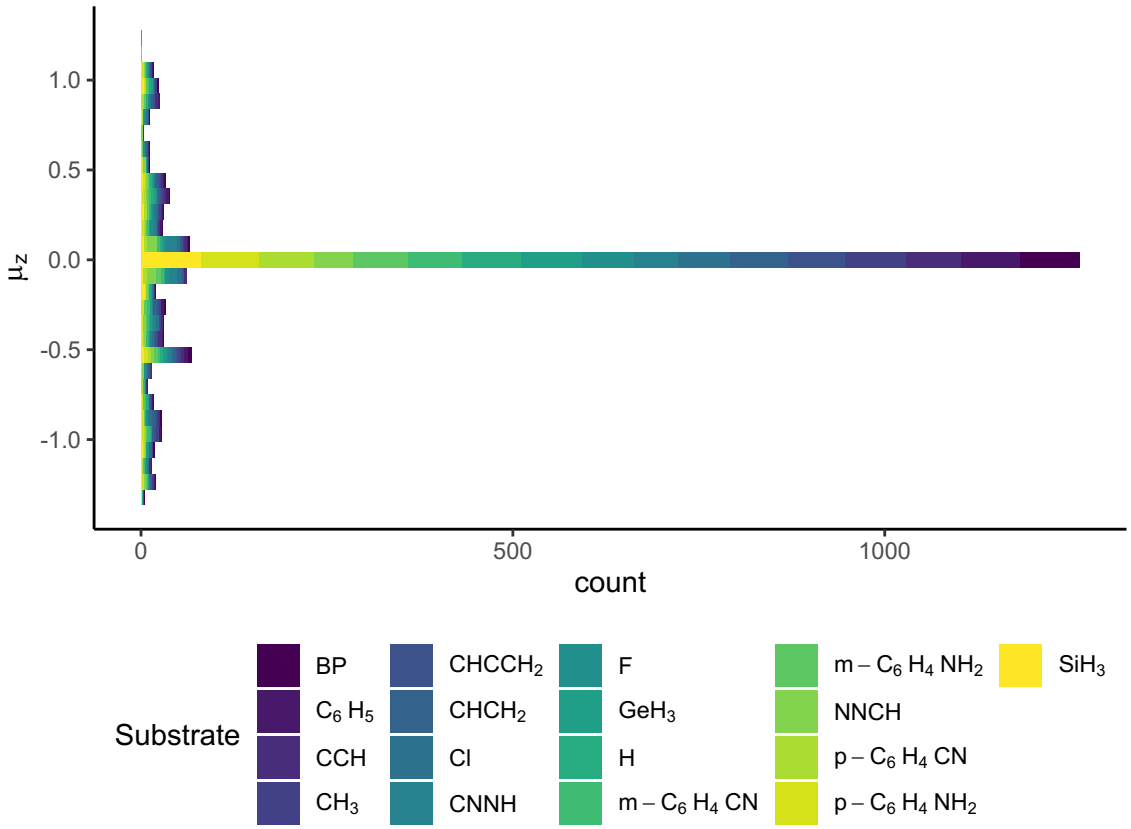

Figure B6: Histogram depicting range of μ_z(R)

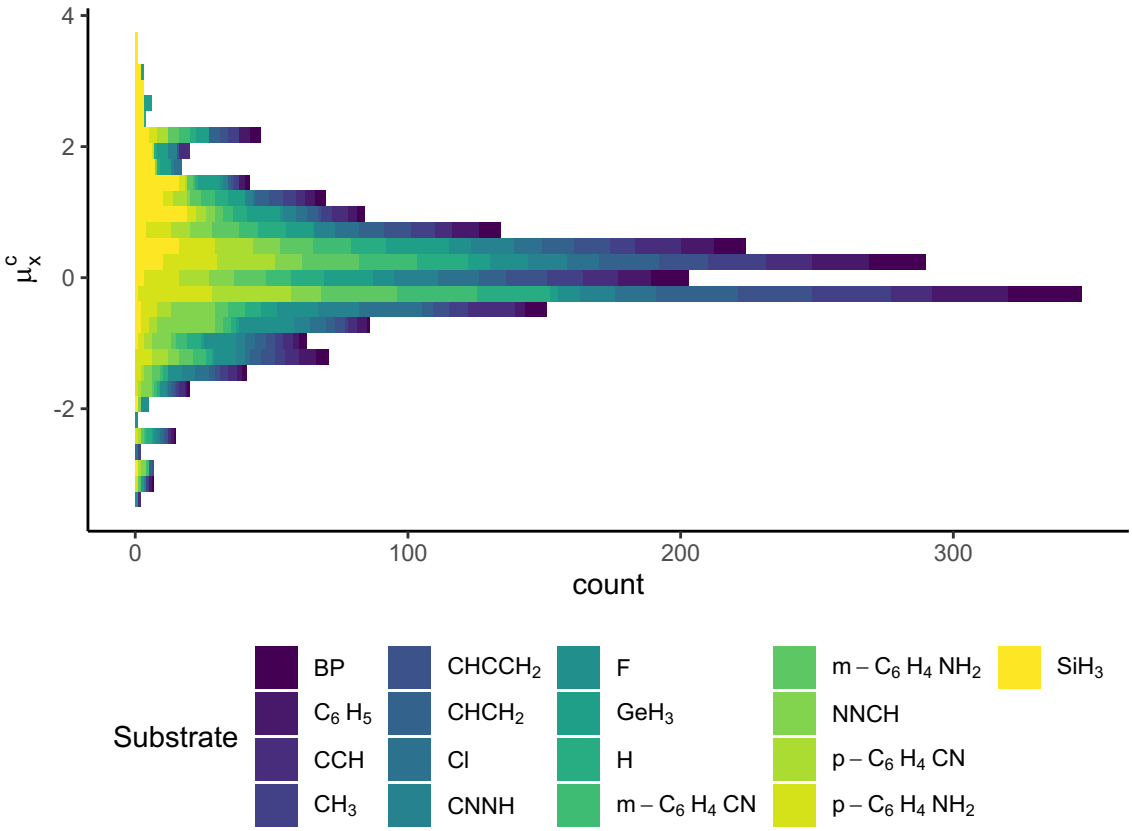

Figure B7: Histogram depicting range of μ_x^c(R)

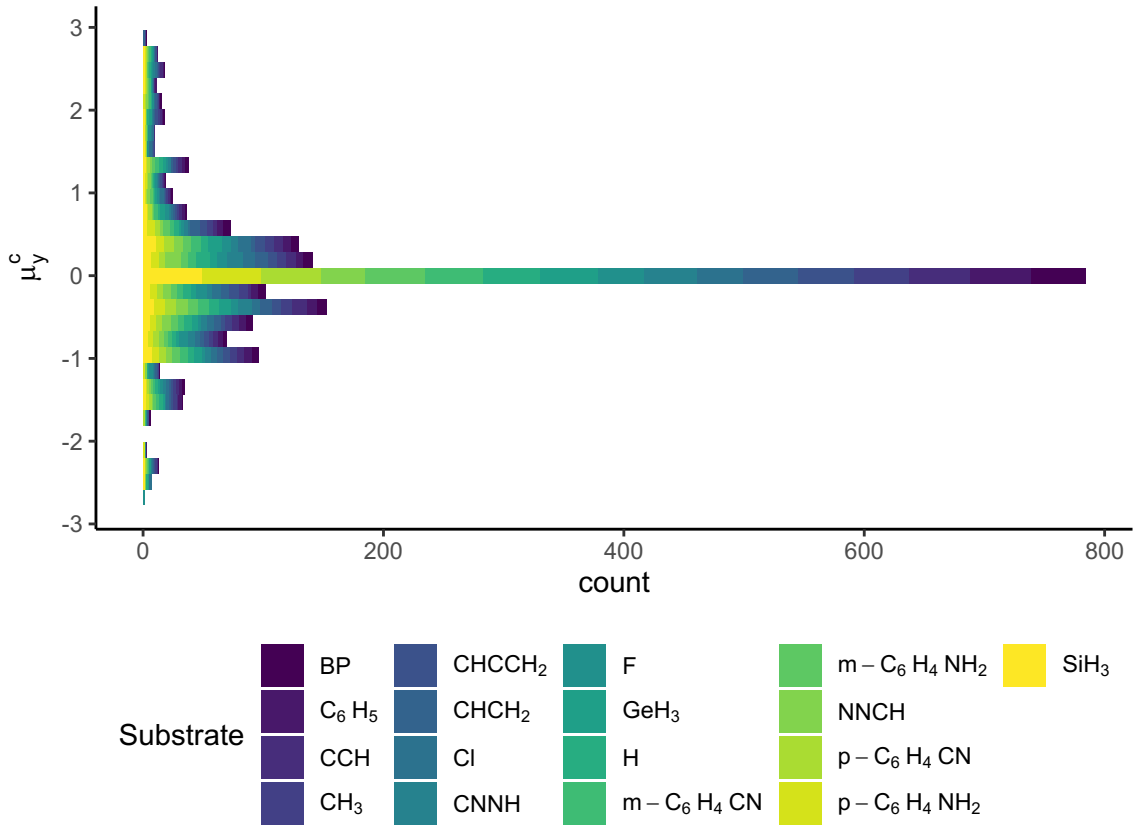

Figure B8: Histogram depicting range of μ_y^c(R)

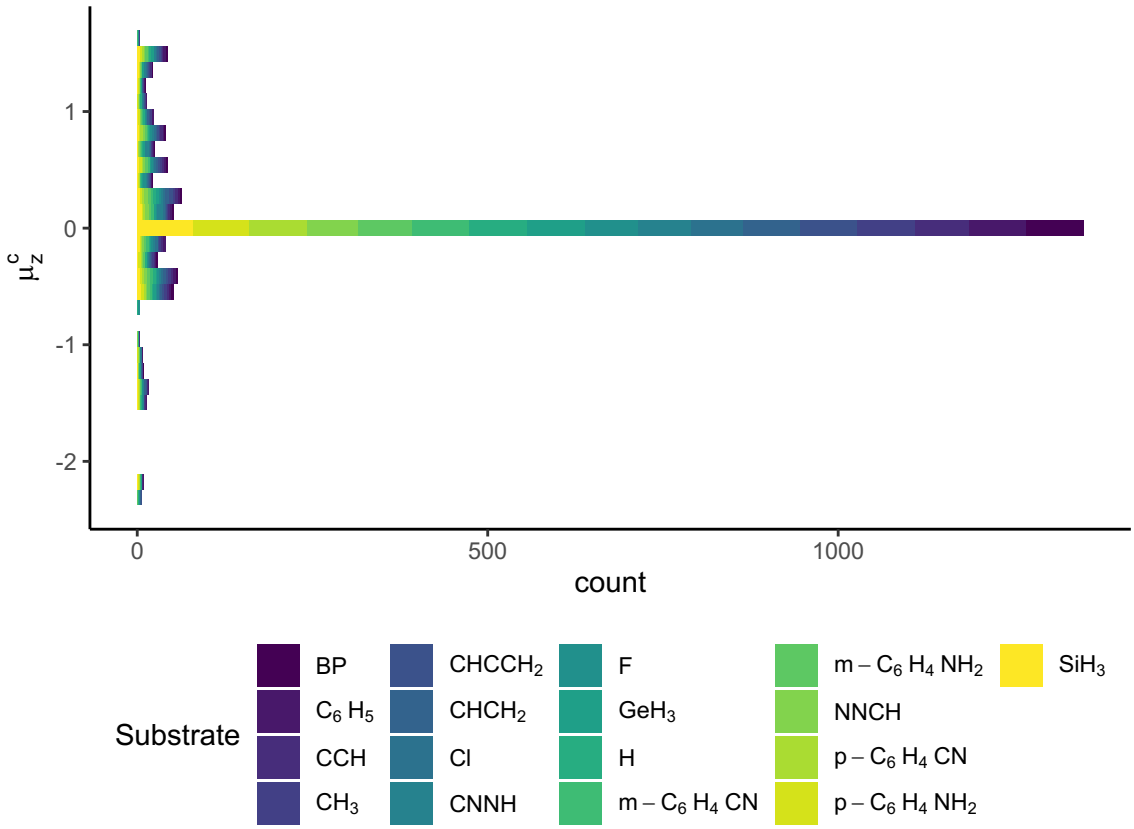

Figure B9: Histogram depicting range of μ_z^c(R)

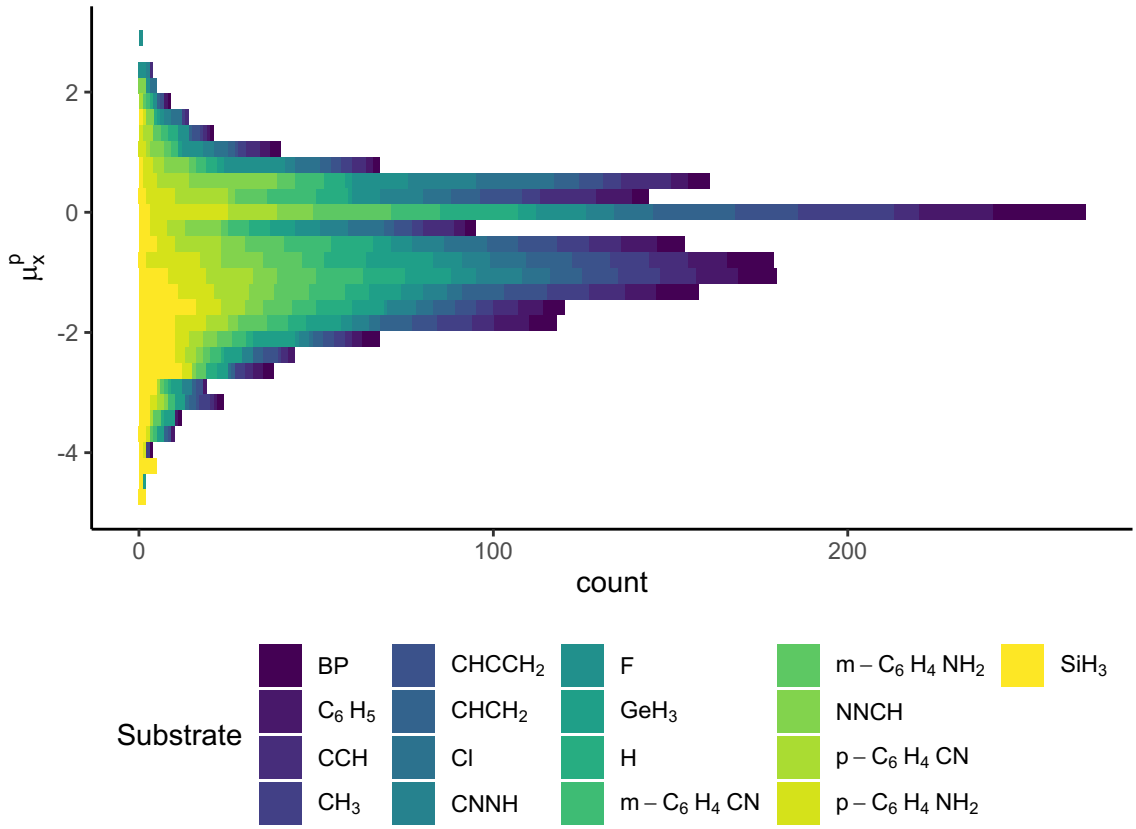

Figure B10: Histogram depicting range of μ_x^p(R)

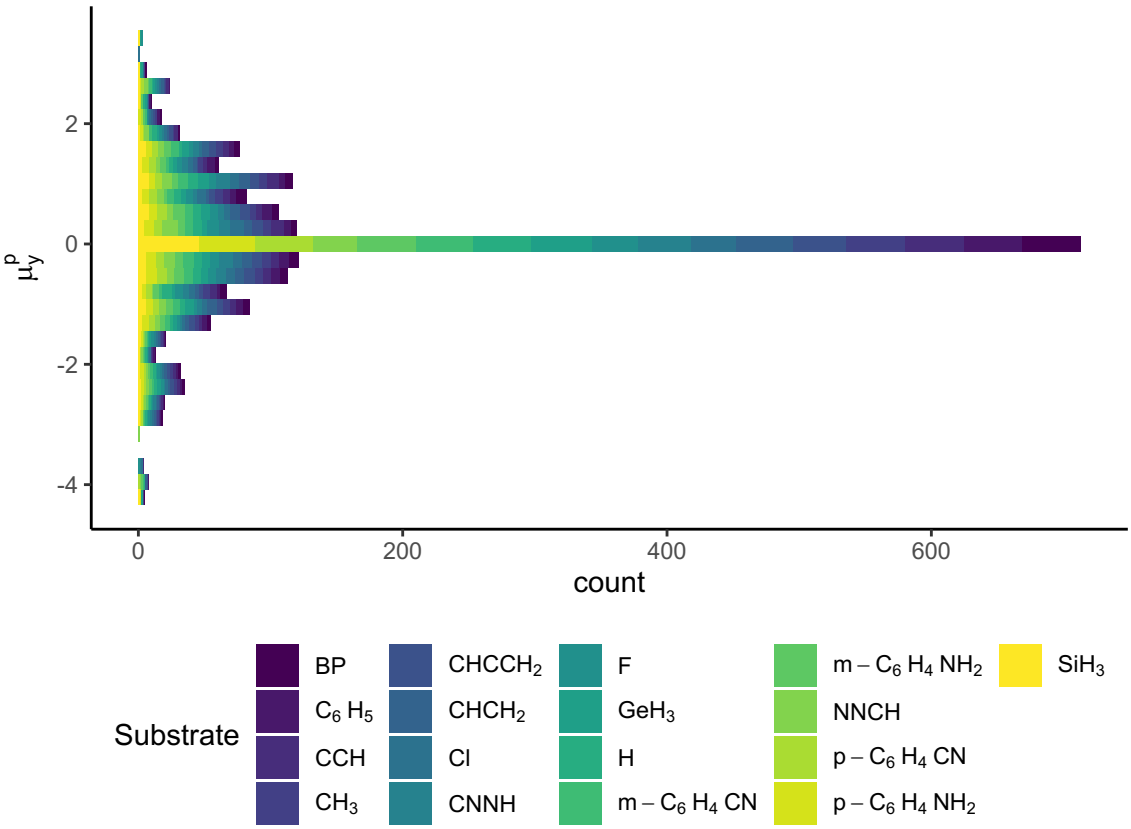

Figure B11: Histogram depicting range of μ_y^p(R)

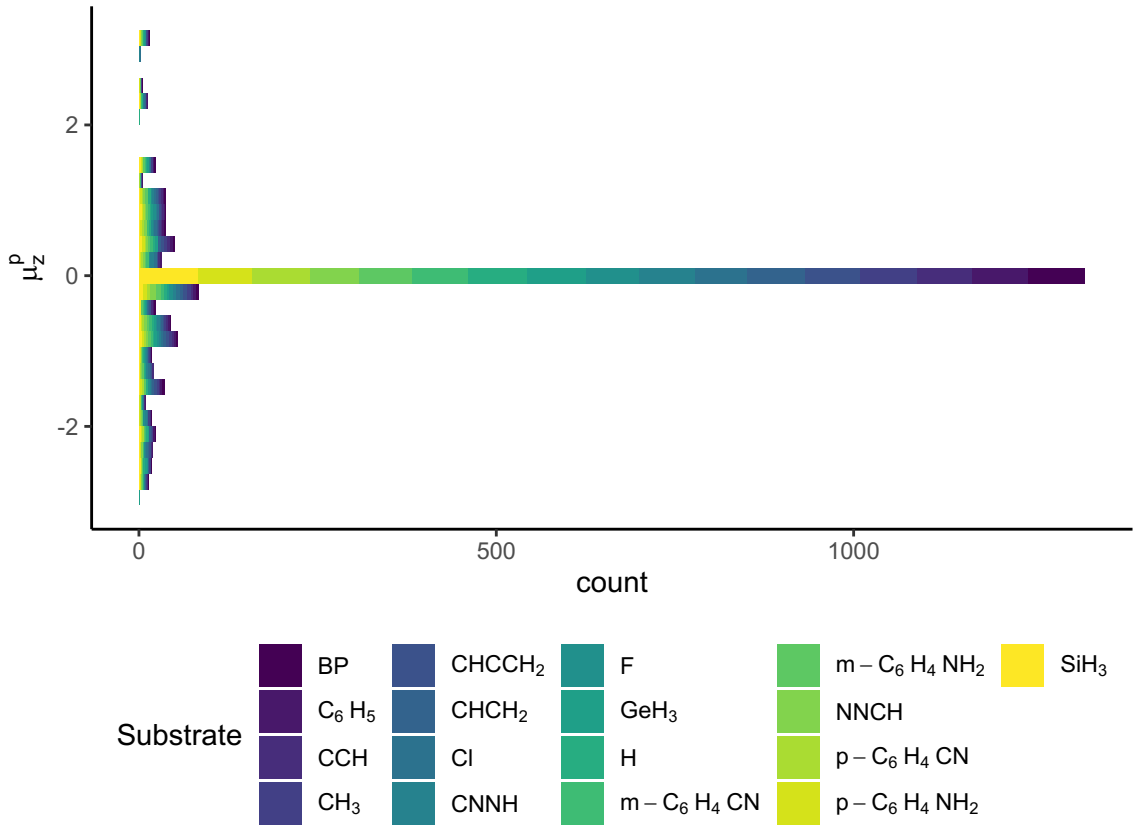

Figure B12: Histogram depicting range of μ_z^p(R)

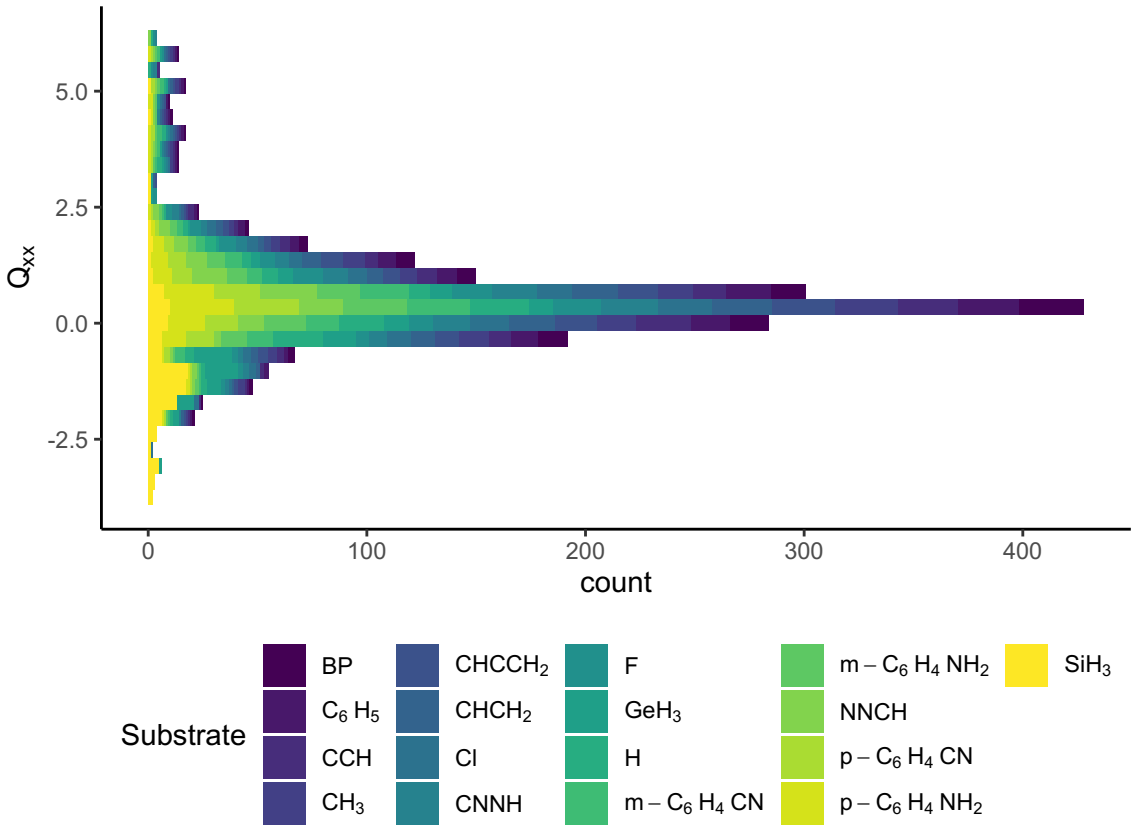

Figure B13: Histogram depicting range of Q_{xx}(R)

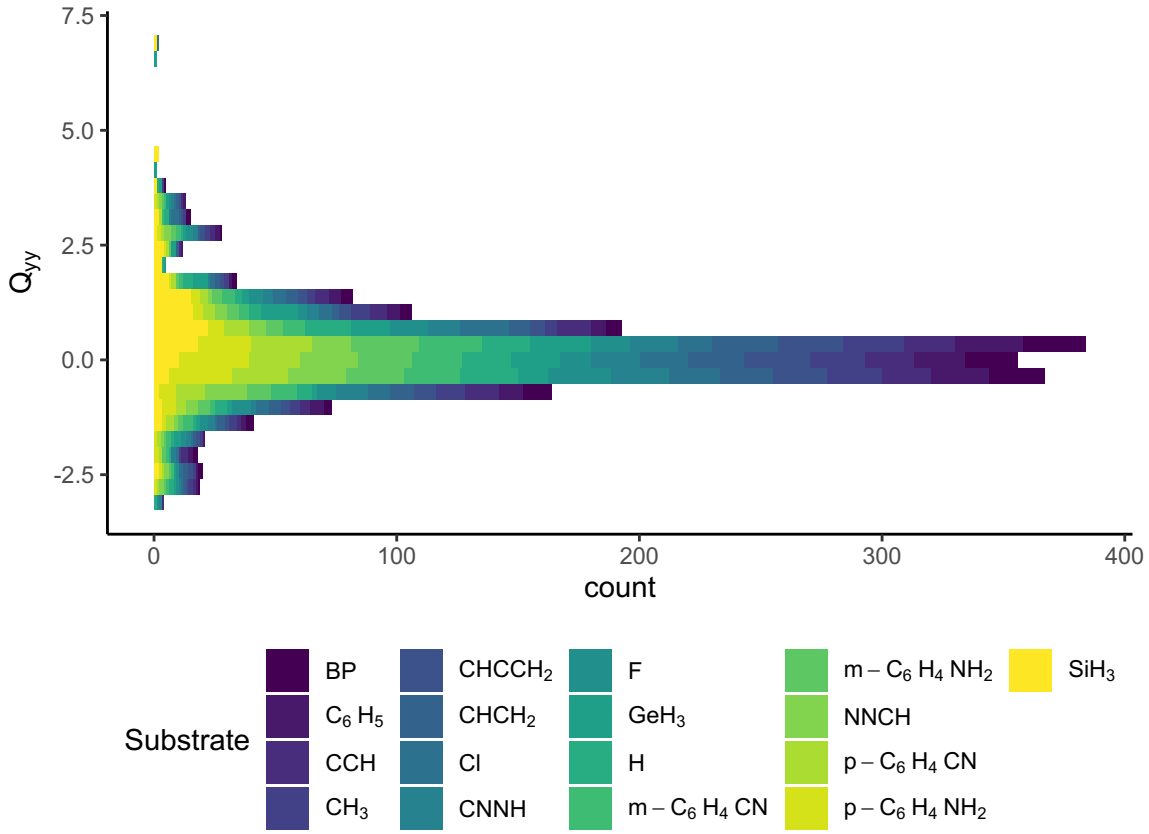

Figure B14: Histogram depicting range of Q_{yy}(R)

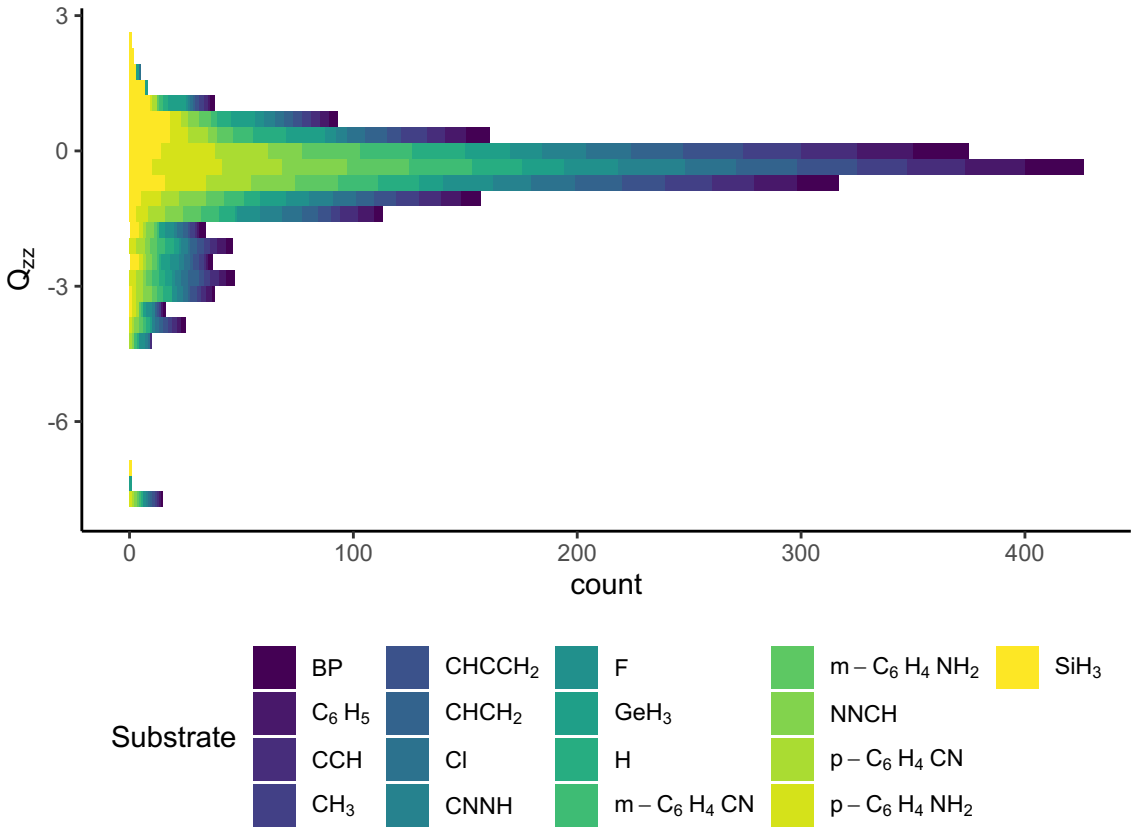

Figure B15: Histogram depicting range of Q_{zz}(R)

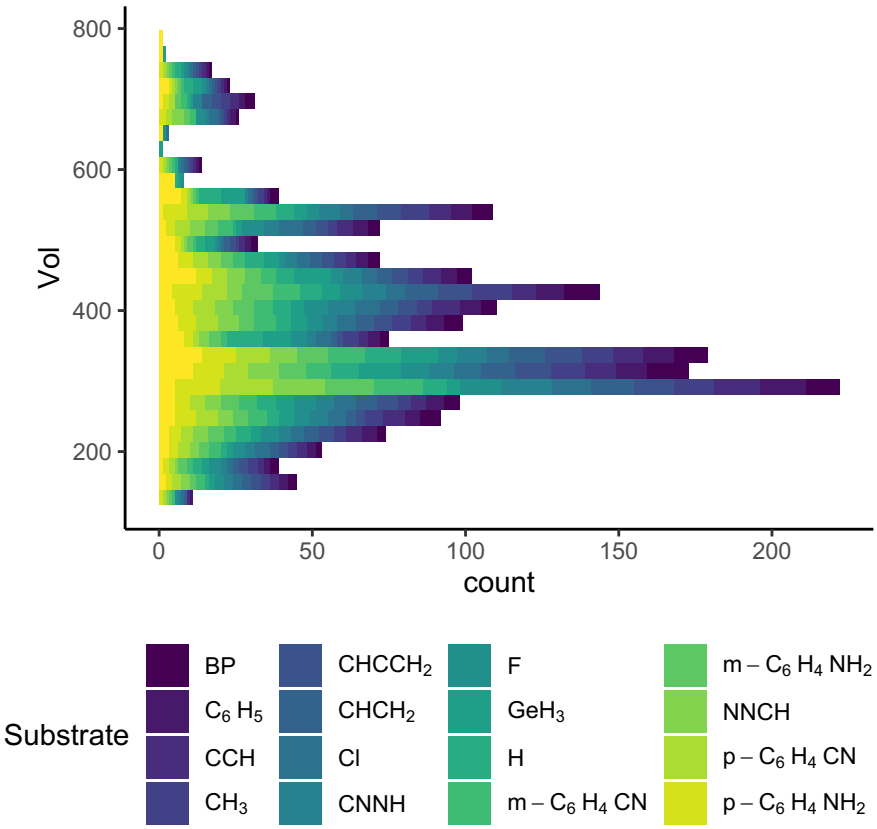

Figure B16: Histogram depicting range of Vol(R)

Table B1: Electronegativity, χ_G, of the substrates calculated using the model of Boyd and Edgecombe

Substrate	χ_G	Susbtrate	χ_G
F	4.0[a]	m-$C_6H_4NH_2$	2.6
NNCH	3.1	p-$C_6H_4NH_2$	2.6
Cl	3.1	$CHCH_2$	2.6
CCH	2.7	$CHCCH_2$	2.6
CNNH	2.6	BP	2.6
p-C_6H_4CN	2.6	CH_3	2.6
m-C_6H_4CN	2.6	GeH_3	2.1
C_6H_5	2.6	SiH_3	2.1

[a] by definition

Table B2: Summary statistics comparing q(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-0.038	0.046	0.055	0.140(S^-)	-0.373(CO)
C_6H_5	-0.022	0.046	0.069	0.279(S^-)	-0.424(CO)
CCH	0.106	0.133	0.116	0.588(S^-)	-0.212(CO)
CH_3	-0.034	0.045	0.047	0.145(O^-)	-0.249(CO)
$CHCCH_2$	0.005	0.046	0.065	0.249(S^-)	-0.350(CO)
$CHCH_2$	-0.015	0.041	0.061	0.242(S^-)	-0.335(CO)
Cl	0.255	0.255	0.123	0.742(OO)	0.007(CO)
CNNH	0.11	0.15	0.145	0.677(S^-)	-0.629(NO)
F	0.618	0.618	0.205	0.981(CO)	0.039(BeH)
GeH_3	-0.362	0.363	0.181	0.092(O^-)	-0.885($PO(OH)_2$)
m$-C_6H_4CN$	0.007	0.046	0.069	0.354(S^-)	-0.393(CO)
m$-C_6H_4NH_2$	-0.025	0.047	0.071	0.280(S^-)	-0.440(CO)
NNCH	0.323	0.324	0.134	0.724(O^-)	-0.087(NF_2)
p$-C_6H_4CN$	0.008	0.047	0.072	0.404(S^-)	-0.405(CO)
p$-C_6H_4NH_2$	-0.048	0.065	0.087	0.244(S^-)	-0.510(CO)
SiH_3	-0.594	0.594	0.231	0.000(SiF_2H)	-1.182($PO(OH)_2$)

Table B3: Linear regression parameters comparing q(R) for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	0.981	-0.038	0.981	0.013	0.005
C_6H_5	0.977	-0.023	0.97	0.016	0.006
CCH	1.051	0.109	0.931	0.027	0.011
CH_3	0.973	-0.036	0.986	0.011	0.004
$CHCCH_2$	0.978	0.004	0.974	0.015	0.006
$CHCH_2$	0.99	-0.016	0.977	0.014	0.006
Cl	0.74	0.245	0.953	0.016	0.006
CNNH	1.029	0.112	0.889	0.034	0.014
F	0.608	0.599	0.802	0.029	0.012
GeH_3	0.674	-0.378	0.824	0.029	0.012
$m-C_6H_4CN$	0.971	0.005	0.97	0.016	0.006
$m-C_6H_4NH_2$	0.974	-0.027	0.969	0.016	0.007
NNCH	0.75	0.311	0.919	0.021	0.009
$p-C_6H_4CN$	0.967	0.006	0.968	0.017	0.007
$p-C_6H_4NH_2$	0.989	-0.048	0.954	0.02	0.008
SiH_3	0.725	-0.607	0.672	0.047	0.019

Table B4: Summary statistics comparing μ_x(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-0.231	0.257	0.204	0.883(NH_3^+)	-1.121(CO_2^-)
C_6H_5	-0.21	0.239	0.215	0.898(NH_3^+)	-1.083(CO_2^-)
CCH	-0.009	0.095	0.136	0.524(ONO)	-0.489(OPH_2O)
CH_3	-0.233	0.249	0.165	0.424(NH_3^+)	-0.813(CO_2^-)
$CHCCH_2$	-0.149	0.175	0.168	0.623(NH_3^+)	-0.751(CO_2^-)
$CHCH_2$	-0.161	0.183	0.159	0.546(NH_3^+)	-0.824(CO_2^-)
Cl	0.184	0.224	0.208	1.031(OO)	-0.533(CO)
CNNH	-0.004	0.139	0.234	0.552(ONO)	-1.629(OCOH)
F	0.066	0.182	0.254	0.904(OO)	-0.660(CH_2OOCH_3-op)
GeH_3	-0.223	0.276	0.233	0.411(PHO(OH))	-0.831(NCS)
m$-C_6H_4CN$	-0.101	0.158	0.204	0.970(NH_3^+)	-0.984(CO_2^-)
m$-C_6H_4NH_2$	-0.229	0.259	0.222	0.932(NH_3^+)	-1.137(CO_2^-)
NNCH	-0.163	0.211	0.261	0.427(NH_3^+)	-1.630(OCOH)
p$-C_6H_4CN$	-0.086	0.164	0.231	1.079($NHNH_2$-ip)	-0.983(CO_2^-)
p$-C_6H_4NH_2$	-0.296	0.317	0.234	0.917(NH_3^+)	-0.983(CO_2^-)
SiH_3	-0.344	0.378	0.251	0.299(SiF)	-1.078(CO_2^-)

Table B5: Summary statistics comparing μ_Y(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-0.007	0.037	0.059	0.254(PHO(OH))	-0.265($OCOCH_3$)
C_6H_5	-0.013	0.039	0.061	0.168(CHCCO)	-0.209(OF)
CCH	0.016	0.054	0.105	0.710(ONO)	-0.180(CHNN)
CH_3	-0.006	0.039	0.062	0.290(PHO(OH))	-0.247($OCOCH_3$)
$CHCCH_2$	-0.019	0.056	0.071	0.241(PHO(OH))	-0.213(MgH)
$CHCH_2$	-0.015	0.044	0.059	0.119(CHCCO)	-0.197($OCOCH_3$)
Cl	0.038	0.093	0.138	0.508(NNH-cis)	-0.363(CH_2OOCH_3-ip)
CNNH	0.007	0.1	0.182	0.973(OCOH)	-0.781(OOH)
F	0.064	0.126	0.208	1.120(ONH_2)	-0.367(CH_2OOCH_3-ip)
GeH_3	-0.013	0.047	0.085	0.329(PHO(OH))	-0.460(OPH_2O)
m$-C_6H_4CN$	-0.008	0.046	0.072	0.272(OPH_2O)	-0.216($NHCH_3$)
m$-C_6H_4NH_2$	-0.009	0.047	0.073	0.252($CONH_2$)	-0.249($OCOCH_3$)
NNCH	0.004	0.117	0.177	0.834(OCOH)	-0.643(CH_2OH-ip)
p$-C_6H_4CN$	-0.003	0.046	0.081	0.409($NHNH_2$-ip)	-0.249($NHCH_3$)
p$-C_6H_4NH_2$	-0.022	0.061	0.102	0.293(OPH_2O)	-0.643(OF)
SiH_3	-0.004	0.05	0.096	0.408(CH_2COOH-ip)	-0.447(OPH_2O)

Table B6: Summary statistics comparing μ_z(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-0.028	0.036	0.232	0.067(SiH)	-2.477(CH_2NO_2-op)
C_6H_5	-0.022	0.044	0.234	0.323($PO(OH)_2$)	-2.440(CH_2NO_2-op)
CCH	-0.02	0.045	0.226	0.283(CH_2COH-op)	-2.346(CH_2NO_2-op)
CH_3	-0.026	0.033	0.225	0.044(NHF)	-2.421(CH_2NO_2-op)
$CHCCH_2$	-0.024	0.049	0.239	0.288($PO(OH)_2$)	-2.469(CH_2NO_2-op)
$CHCH_2$	-0.049	0.06	0.298	0.159($NHCH_3$)	-2.418(CH_2NO_2-op)
Cl	-0.022	0.078	0.249	0.347(CH_2COH-op)	-2.250(CH_2NO_2-op)
CNNH	-0.025	0.082	0.251	0.506($NHCH_3$)	-2.350(CH_2NO_2-op)
F	-0.031	0.073	0.252	0.317(NHOH)	-2.280(CH_2NO_2-op)
GeH_3	-0.027	0.05	0.25	0.714(OPH_2O)	-2.496(CH_2NO_2-op)
m$-C_6H_4CN$	-0.025	0.05	0.232	0.193($NHCH_3$)	-2.412(CH_2NO_2-op)
m$-C_6H_4NH_2$	-0.027	0.056	0.247	0.727($CONH_2$)	-2.444(CH_2NO_2-op)
NNCH	-0.024	0.085	0.255	0.260(CH_2CSH-op)	-2.368(CH_2NO_2-op)
p$-C_6H_4CN$	-0.012	0.053	0.236	0.434(NHOH)	-2.400(CH_2NO_2-op)
p$-C_6H_4NH_2$	-0.021	0.049	0.241	0.523($CONH_2$)	-2.479(CH_2NO_2-op)
SiH_3	-0.018	0.062	0.258	0.712(OPH_2O)	-2.469(CH_2NO_2-op)

Table B7: Summary statistics comparing $\mu_{c,x}(R)$ for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-0.149	0.202	0.227	0.984(NH_3^+)	-0.994(CO_2^-)
C_6H_5	-0.14	0.21	0.256	0.979(NH_3^+)	-1.015(CH_2OOCH_3-op)
CCH	0.14	0.233	0.247	0.817(NCS)	-0.868(CH_2OOCH_3-op)
CH_3	-0.163	0.19	0.171	0.652(PHO(OH))	-0.733(CO_2^-)
$CHCCH_2$	-0.059	0.156	0.197	0.719(NH_3^+)	-0.609(CO_2^-)
$CHCH_2$	-0.102	0.162	0.184	0.614(NH_3^+)	-0.698(CO_2^-)
Cl	0.422	0.467	0.349	2.279(CHCCO)	-1.161($PO(OH)_2$)
CNNH	0.117	0.301	0.508	0.752(CCF)	-4.174(OCOH)
F	0.731	0.751	0.376	1.692(CNO)	-0.818(CH_2OOCH_3-op)
GeH_3	-0.75	0.764	0.377	0.668(CHCCO)	-1.835($PO(OH)_2$)
m$-C_6H_4CN$	-0.029	0.178	0.239	1.049(NH_3^+)	-0.850(CO_2^-)
m$-C_6H_4NH_2$	-0.158	0.224	0.262	1.014(NH_3^+)	-1.107(OPH_2O)
NNCH	0.265	0.425	0.519	1.010(CCF)	-3.695(OCOH)
p$-C_6H_4CN$	-0.017	0.189	0.265	1.164($NHNH_2$-ip)	-0.849(CO_2^-)
p$-C_6H_4NH_2$	-0.219	0.267	0.273	0.998(NH_3^+)	-1.154(CH_2OOCH_3-op)
SiH_3	-1.329	1.33	0.587	0.030(BeH)	-2.753($PO(OH)_2$)

Table B8: Summary statistics comparing $\mu_{c,y}$(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-0.025	0.062	0.153	0.543(PHO(OH))	-1.235(NCS)
C_6H_5	-0.038	0.072	0.139	0.246(CH_2COCl-op)	-0.944(NCS)
CCH	-0.017	0.103	0.26	0.697(ONO)	-2.359(NCS)
CH_3	-0.022	0.058	0.135	0.610(PHO(OH))	-1.040(NCS)
$CHCCH_2$	-0.037	0.082	0.13	0.485(PHO(OH))	-0.731(NCS)
$CHCH_2$	-0.039	0.076	0.128	0.182(NCO)	-0.814(NCS)
Cl	0.028	0.125	0.195	0.925(CHCCO)	-0.739(CH_2OOCH_3-ip)
CNNH	-0.009	0.158	0.299	2.005(OCOH)	-0.957(NCS)
F	0.024	0.167	0.304	1.290(ONH_2)	-1.596(CNO)
GeH_3	-0.018	0.082	0.209	1.110(CHCCO)	-1.434(NCS)
m$-C_6H_4CN$	-0.032	0.079	0.138	0.316(ONO)	-0.687(NCS)
m$-C_6H_4NH_2$	-0.032	0.082	0.154	0.333(CH_2COOH-ip)	-1.014(NCS)
NNCH	-0.014	0.148	0.309	1.780(OCOH)	-1.358(CNO)
p$-C_6H_4CN$	-0.028	0.081	0.148	0.460($NHNH_2$-ip)	-0.705(NCS)
p$-C_6H_4NH_2$	-0.048	0.088	0.166	0.319(OPH_2O)	-1.101(NCS)
SiH_3	-0.013	0.11	0.289	1.102(CHCCO)	-2.363(NCS)

Table B9: Summary statistics comparing $\mu_{c,z}$(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-0.048	0.055	0.393	0.066(CH_2CSH-ip)	-4.209(CH_2NO_2-op)
C_6H_5	-0.033	0.072	0.397	0.635($PO(OH)_2$)	-4.156(CH_2NO_2-op)
CCH	-0.034	0.076	0.39	0.671(CH_2COH-op)	-4.033(CH_2NO_2-op)
CH_3	-0.041	0.049	0.379	0.106(CH_2COH-op)	-4.086(CH_2NO_2-op)
$CHCCH_2$	-0.042	0.077	0.406	0.571($PO(OH)_2$)	-4.185(CH_2NO_2-op)
$CHCH_2$	-0.08	0.096	0.482	0.221($NHCH_3$)	-4.126(CH_2NO_2-op)
Cl	-0.049	0.117	0.423	0.804(CH_2COH-op)	-3.889(CH_2NO_2-op)
CNNH	-0.046	0.124	0.413	0.806($NHCH_3$)	-3.918(CH_2NO_2-op)
F	-0.062	0.109	0.427	0.328(NHOH)	-3.955(CH_2NO_2-op)
GeH_3	-0.045	0.076	0.422	1.105(OPH_2O)	-4.248(CH_2NO_2-op)
m$-C_6H_4CN$	-0.044	0.08	0.397	0.278($NHCH_3$)	-4.129(CH_2NO_2-op)
m$-C_6H_4NH_2$	-0.044	0.093	0.416	1.032($CONH_2$)	-4.157(CH_2NO_2-op)
NNCH	-0.059	0.102	0.415	0.216(CH_2CSH-ip)	-4.008(CH_2NO_2-op)
p$-C_6H_4CN$	-0.015	0.094	0.421	1.492(NHOH)	-4.116(CH_2NO_2-op)
p$-C_6H_4NH_2$	-0.03	0.074	0.404	0.734($CONH_2$)	-4.191(CH_2NO_2-op)
SiH_3	-0.023	0.1	0.446	1.528(NHOH)	-4.186(CH_2NO_2-op)

Table B10: Summary statistics comparing $\mu_{p,x}$(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-0.082	0.1	0.128	0.183(SCF_3)	-0.860(NCS)
C_6H_5	-0.07	0.105	0.127	0.479(CH_2OOCH_3-op)	-0.714(NCS)
CCH	-0.149	0.191	0.191	0.415(CH_2OOCH_3-op)	-0.983(NCS)
CH_3	-0.07	0.094	0.125	0.133(SCF_3)	-0.805(NCS)
$CHCCH_2$	-0.09	0.103	0.101	0.134(CH_2OOCH_3-op)	-0.598(NCS)
$CHCH_2$	-0.059	0.084	0.101	0.141(SCN)	-0.635(NCS)
Cl	-0.238	0.288	0.288	0.719($PO(OH)_2$)	-2.166(CHCCO)
CNNH	-0.12	0.228	0.335	2.544(OCOH)	-0.742(NCS)
F	-0.666	0.676	0.368	0.194(SiCl)	-2.022(CCCl)
GeH_3	0.527	0.574	0.424	1.721($PO(OH)_2$)	-0.938(CHCCO)
m$-C_6H_4CN$	-0.072	0.1	0.11	0.198(SCN)	-0.600(NCS)
m$-C_6H_4NH_2$	-0.072	0.109	0.13	0.447(CH_2OOCH_3-op)	-0.749(NCS)
NNCH	-0.428	0.482	0.358	2.065(OCOH)	-1.342(CCF)
p$-C_6H_4CN$	-0.069	0.098	0.111	0.210(SCN)	-0.614(NCS)
p$-C_6H_4NH_2$	-0.076	0.109	0.132	0.532(CH_2OOCH_3-op)	-0.771(NCS)
SiH_3	0.986	1.004	0.554	2.449($PO(OH)_2$)	-0.623(NCS)

Table B11: Summary statistics comparing $\mu_{p,y}$(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	0.018	0.044	0.122	1.125(NCS)	-0.289(PHO(OH))
C_6H_5	0.025	0.05	0.11	0.861(NCS)	-0.199(CH_2COCl-op)
CCH	0.034	0.065	0.239	2.421(NCS)	-0.280(NCO)
CH_3	0.016	0.039	0.101	0.885(NCS)	-0.320(PHO(OH))
$CHCCH_2$	0.018	0.052	0.095	0.636(NCS)	-0.244(PHO(OH))
$CHCH_2$	0.025	0.055	0.101	0.749(NCS)	-0.181(NCO)
Cl	0.01	0.067	0.139	0.375(CH_2OOCH_3-ip)	-1.098(CHCCO)
CNNH	0.016	0.088	0.185	1.068(NCS)	-1.033(OCOH)
F	0.04	0.082	0.176	1.391(CNO)	-0.332(NCS)
GeH_3	0.005	0.052	0.171	1.360(NCS)	-1.041(CHCCO)
m$-C_6H_4CN$	0.024	0.048	0.1	0.677(NCS)	-0.176(CH_2COCl-op)
m$-C_6H_4NH_2$	0.023	0.053	0.116	0.921(NCS)	-0.203(CH_2COCl-op)
NNCH	0.018	0.1	0.201	1.118(CNO)	-0.946(OCOH)
p$-C_6H_4CN$	0.025	0.048	0.104	0.714(NCS)	-0.186(CH_2COCl-op)
p$-C_6H_4NH_2$	0.026	0.054	0.12	0.961(NCS)	-0.216(CH_2COCl-op)
SiH_3	0.009	0.075	0.259	2.424(NCS)	-1.054(CHCCO)

Table B12: Summary statistics comparing $\mu_{p,z}$(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	0.02	0.029	0.163	1.732(CH_2NO_2-op)	-0.137(CH_2CSH-ip)
C_6H_5	0.011	0.035	0.166	1.716(CH_2NO_2-op)	-0.312($PO(OH)_2$)
CCH	0.014	0.037	0.167	1.686(CH_2NO_2-op)	-0.387(CH_2COH-op)
CH_3	0.015	0.024	0.155	1.664(CH_2NO_2-op)	-0.089(CH_2CSH-ip)
$CHCCH_2$	0.017	0.037	0.169	1.716(CH_2NO_2-op)	-0.283($PO(OH)_2$)
$CHCH_2$	0.031	0.043	0.188	1.709(CH_2NO_2-op)	-0.119(CH_2CSH-ip)
Cl	0.027	0.051	0.184	1.639(CH_2NO_2-op)	-0.457(CH_2COH-op)
CNNH	0.02	0.062	0.171	1.567(CH_2NO_2-op)	-0.371(CH_2COH-op)
F	0.031	0.047	0.184	1.675(CH_2NO_2-op)	-0.145(CH_2CSH-ip)
GeH_3	0.018	0.037	0.176	1.752(CH_2NO_2-op)	-0.391(OPH_2O)
m$-C_6H_4CN$	0.019	0.036	0.166	1.716(CH_2NO_2-op)	-0.140(CH_2CSH-ip)
m$-C_6H_4NH_2$	0.017	0.046	0.173	1.714(CH_2NO_2-op)	-0.305($CONH_2$)
NNCH	0.035	0.057	0.174	1.640(CH_2NO_2-op)	-0.127(CF_3)
p$-C_6H_4CN$	0.002	0.044	0.193	1.716(CH_2NO_2-op)	-1.059(NHOH)
p$-C_6H_4NH_2$	0.01	0.036	0.166	1.712(CH_2NO_2-op)	-0.340($PO(OH)_2$)
SiH_3	0.005	0.049	0.199	1.717(CH_2NO_2-op)	-1.030(NHOH)

Table B13: Linear regression parameters comparing μ_x(R) for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	1.147	-0.162	0.839	0.047	0.028
C_6H_5	1.163	-0.133	0.83	0.049	0.03
CCH	1.06	0.019	0.905	0.032	0.019
CH_3	1.13	-0.172	0.886	0.038	0.023
$CHCCH_2$	1.128	-0.089	0.882	0.038	0.023
$CHCH_2$	1.152	-0.09	0.903	0.035	0.021
Cl	0.874	0.125	0.714	0.053	0.031
CNNH	0.998	-0.005	0.737	0.056	0.034
F	0.777	-0.042	0.61	0.06	0.037
GeH_3	1.031	-0.209	0.745	0.056	0.034
$m-C_6H_4CN$	1.17	-0.02	0.844	0.047	0.029
$m-C_6H_4NH_2$	1.167	-0.151	0.821	0.051	0.031
NNCH	0.951	-0.186	0.676	0.063	0.038
$p-C_6H_4CN$	1.154	-0.014	0.803	0.053	0.032
$p-C_6H_4NH_2$	1.179	-0.212	0.809	0.053	0.032
SiH_3	1.042	-0.324	0.724	0.06	0.036

Table B14: Linear regression parameters comparing μ_y(R) for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	1.048	-0.007	0.987	0.011	0.005
C_6H_5	1.016	-0.013	0.984	0.012	0.006
CCH	0.95	0.016	0.948	0.021	0.009
CH_3	1.043	-0.006	0.985	0.012	0.006
$CHCCH_2$	1.047	-0.019	0.98	0.014	0.006
$CHCH_2$	1.016	-0.015	0.984	0.012	0.005
Cl	0.925	0.037	0.912	0.027	0.013
CNNH	0.955	0.007	0.856	0.037	0.017
F	0.932	0.063	0.815	0.043	0.02
GeH_3	1.045	-0.013	0.972	0.017	0.008
$m-C_6H_4CN$	0.995	-0.008	0.976	0.015	0.007
$m-C_6H_4NH_2$	1.006	-0.009	0.976	0.015	0.007
NNCH	1.056	0.004	0.888	0.036	0.017
$p-C_6H_4CN$	0.982	-0.003	0.969	0.016	0.007
$p-C_6H_4NH_2$	1.045	-0.022	0.958	0.02	0.009
SiH_3	1.025	-0.004	0.961	0.019	0.009

Table B15: Linear regression parameters comparing μ_z(R) for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	0.839	-0.032	0.663	0.056	0.021
C_6H_5	0.82	-0.026	0.652	0.056	0.021
CCH	0.765	-0.025	0.654	0.052	0.019
CH_3	0.836	-0.03	0.676	0.054	0.02
$CHCCH_2$	0.815	-0.029	0.641	0.057	0.021
$CHCH_2$	0.672	-0.057	0.461	0.068	0.025
Cl	0.722	-0.028	0.591	0.057	0.021
CNNH	0.725	-0.032	0.59	0.057	0.021
F	0.711	-0.031	0.516	0.067	0.022
GeH_3	0.806	-0.031	0.613	0.06	0.022
$m-C_6H_4CN$	0.79	-0.029	0.645	0.055	0.02
$m-C_6H_4NH_2$	0.812	-0.031	0.623	0.059	0.022
NNCH	0.814	-0.025	0.586	0.065	0.023
$p-C_6H_4CN$	0.772	-0.018	0.632	0.055	0.02
$p-C_6H_4NH_2$	0.849	-0.024	0.647	0.058	0.022
SiH_3	0.761	-0.023	0.58	0.06	0.023

Table B16: Linear regression parameters comparing $\mu_{c,x}(R)$ for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	1.025	-0.134	0.95	0.022	0.025
C_6H_5	1.04	-0.116	0.94	0.024	0.028
CCH	1.012	0.147	0.94	0.024	0.027
CH_3	1.027	-0.147	0.972	0.016	0.018
$CHCCH_2$	1.024	-0.045	0.962	0.019	0.021
$CHCH_2$	1.037	-0.08	0.969	0.017	0.02
Cl	0.942	0.388	0.867	0.035	0.039
CNNH	0.937	0.079	0.765	0.049	0.055
F	0.917	0.68	0.855	0.037	0.042
GeH_3	1.042	-0.725	0.878	0.036	0.041
$m-C_6H_4CN$	1.042	-0.003	0.948	0.023	0.026
$m-C_6H_4NH_2$	1.043	-0.132	0.938	0.025	0.028
NNCH	0.937	0.229	0.755	0.051	0.057
$p-C_6H_4CN$	1.045	0.01	0.937	0.025	0.029
$p-C_6H_4NH_2$	1.045	-0.192	0.933	0.026	0.029
SiH_3	1.062	-1.292	0.755	0.056	0.064

Table B17: Linear regression parameters comparing $\mu_{c,y}(R)$ for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	0.993	-0.025	0.98	0.013	0.014
C_6H_5	0.978	-0.037	0.984	0.012	0.013
CCH	0.925	-0.016	0.942	0.021	0.023
CH_3	0.99	-0.022	0.984	0.012	0.013
$CHCCH_2$	0.994	-0.037	0.986	0.011	0.012
$CHCH_2$	0.982	-0.039	0.986	0.011	0.012
Cl	1.019	0.027	0.971	0.017	0.018
CNNH	0.926	-0.008	0.924	0.025	0.027
F	1.015	0.024	0.93	0.027	0.029
GeH_3	1.017	-0.019	0.965	0.018	0.019
$m-C_6H_4CN$	0.978	-0.032	0.984	0.012	0.013
$m-C_6H_4NH_2$	0.976	-0.032	0.98	0.013	0.014
NNCH	0.972	-0.013	0.924	0.027	0.029
$p-C_6H_4CN$	0.975	-0.027	0.981	0.013	0.013
$p-C_6H_4NH_2$	0.979	-0.048	0.977	0.014	0.015
SiH_3	0.997	-0.012	0.933	0.025	0.027

Table B18: Linear regression parameters comparing $\mu_{c,z}(R)$ for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	0.88	-0.054	0.762	0.046	0.035
C_6H_5	0.869	-0.04	0.755	0.046	0.036
CCH	0.826	-0.043	0.754	0.044	0.034
CH_3	0.878	-0.047	0.775	0.044	0.034
$CHCCH_2$	0.857	-0.049	0.743	0.047	0.036
$CHCH_2$	0.79	-0.091	0.646	0.055	0.042
Cl	0.803	-0.058	0.719	0.048	0.037
CNNH	0.795	-0.056	0.727	0.046	0.036
F	0.815	-0.065	0.677	0.054	0.039
GeH_3	0.858	-0.051	0.724	0.05	0.038
$m-C_6H_4CN$	0.855	-0.049	0.749	0.046	0.036
$m-C_6H_4NH_2$	0.858	-0.051	0.733	0.048	0.037
NNCH	0.845	-0.061	0.71	0.052	0.038
$p-C_6H_4CN$	0.836	-0.023	0.722	0.048	0.037
$p-C_6H_4NH_2$	0.884	-0.036	0.752	0.047	0.037
SiH_3	0.819	-0.032	0.692	0.051	0.039

Table B19: Linear regression parameters comparing $\mu_{p,x}$(R) for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	1.02	-0.085	0.976	0.015	0.012
C_6H_5	1.014	-0.072	0.976	0.015	0.012
CCH	0.998	-0.148	0.945	0.022	0.018
CH_3	1.03	-0.074	0.978	0.014	0.011
$CHCCH_2$	0.997	-0.089	0.984	0.012	0.01
$CHCH_2$	1.008	-0.06	0.984	0.012	0.009
Cl	0.878	-0.223	0.87	0.032	0.026
CNNH	0.927	-0.111	0.837	0.039	0.031
F	0.831	-0.643	0.794	0.041	0.033
GeH_3	1.028	0.523	0.791	0.049	0.04
m$-C_6H_4CN$	1.008	-0.073	0.982	0.013	0.01
m$-C_6H_4NH_2$	1.019	-0.074	0.975	0.015	0.012
NNCH	0.881	-0.417	0.801	0.042	0.033
p$-C_6H_4CN$	1.009	-0.07	0.981	0.013	0.01
p$-C_6H_4NH_2$	1.015	-0.078	0.974	0.015	0.012
SiH_3	1.052	0.979	0.697	0.065	0.052

Table B20: Linear regression parameters comparing $\mu_{p,y}(R)$ for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	0.957	0.017	0.975	0.014	0.011
C_6H_5	0.954	0.023	0.98	0.013	0.01
CCH	0.884	0.031	0.905	0.027	0.021
CH_3	0.961	0.015	0.983	0.012	0.009
$CHCCH_2$	0.963	0.017	0.986	0.011	0.008
$CHCH_2$	0.963	0.024	0.983	0.012	0.009
Cl	1.042	0.011	0.974	0.016	0.013
CNNH	0.912	0.014	0.944	0.021	0.016
F	1.019	0.04	0.954	0.022	0.017
GeH_3	0.987	0.005	0.953	0.021	0.016
$m-C_6H_4CN$	0.962	0.023	0.984	0.012	0.009
$m-C_6H_4NH_2$	0.954	0.022	0.978	0.013	0.01
NNCH	0.948	0.016	0.935	0.024	0.019
$p-C_6H_4CN$	0.962	0.024	0.982	0.012	0.009
$p-C_6H_4NH_2$	0.943	0.025	0.977	0.013	0.01
SiH_3	0.943	0.008	0.891	0.031	0.024

Table B21: Linear regression parameters comparing $\mu_{p,z}$(R) for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	0.947	0.021	0.893	0.03	0.015
C_6H_5	0.94	0.013	0.889	0.031	0.015
CCH	0.906	0.017	0.885	0.03	0.015
CH_3	0.946	0.016	0.902	0.029	0.014
$CHCCH_2$	0.931	0.019	0.884	0.032	0.015
$CHCH_2$	0.922	0.033	0.859	0.035	0.017
Cl	0.917	0.029	0.866	0.034	0.017
CNNH	0.891	0.023	0.881	0.031	0.015
F	0.951	0.032	0.849	0.039	0.018
GeH_3	0.944	0.02	0.875	0.033	0.016
$m-C_6H_4CN$	0.938	0.02	0.887	0.031	0.015
$m-C_6H_4NH_2$	0.928	0.019	0.879	0.032	0.016
NNCH	0.91	0.036	0.866	0.034	0.016
$p-C_6H_4CN$	0.915	0.005	0.85	0.036	0.018
$p-C_6H_4NH_2$	0.945	0.011	0.889	0.031	0.015
SiH_3	0.908	0.008	0.841	0.037	0.018

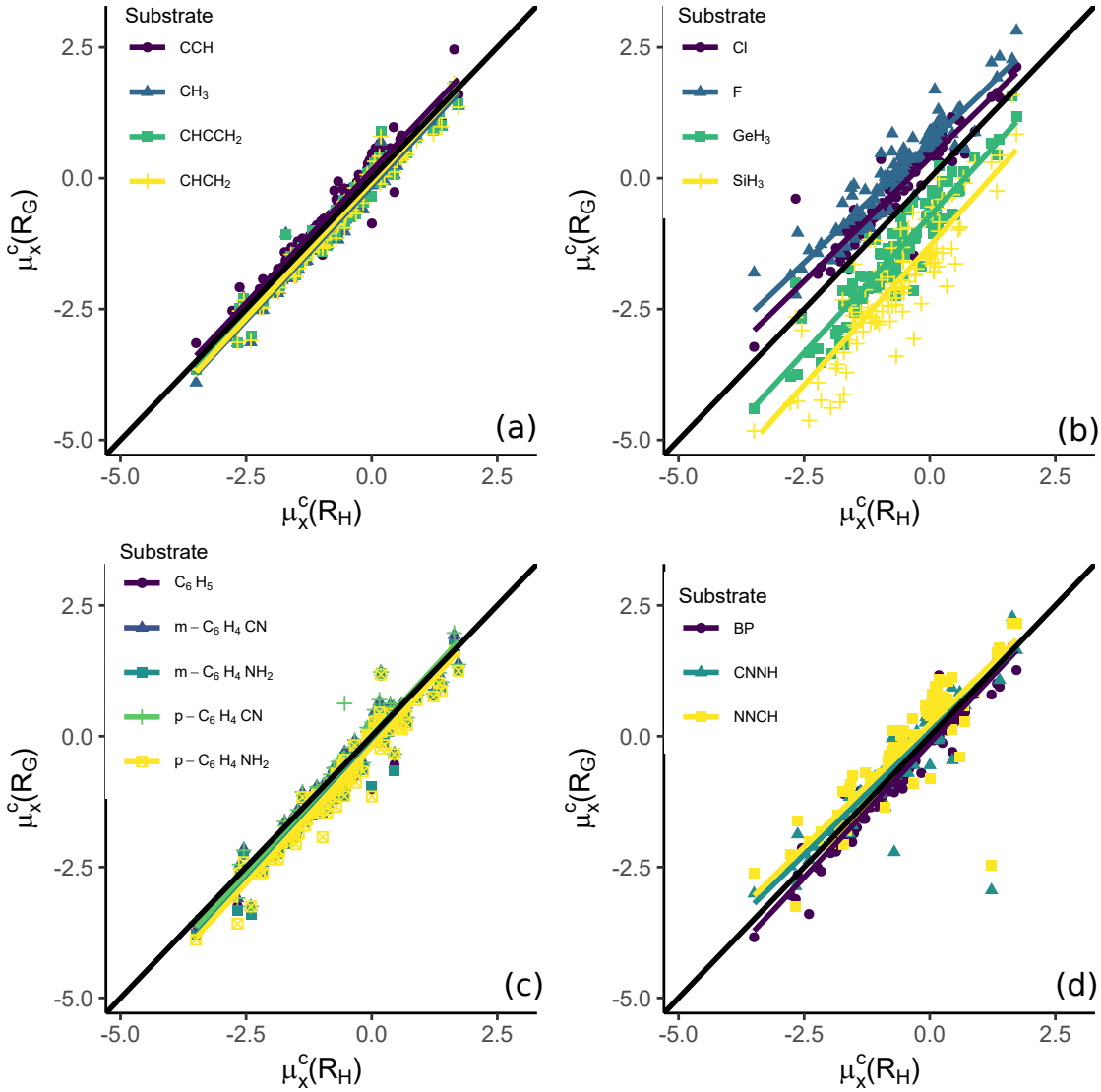

Figure B17: Linear plots of $\mu_x^c(R)$ for R in various R-G systems against $\mu_x^c(R)$. (A) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (B) G: Cl, F, SiH_3. GeH_3; (C) C_6H_5, m−C_6H_4CN, m−$C_6H_4NH_2$, p−C_6H_4CN, p−$C_6H_4NH_2$; D: BP, CNNH, NNCH

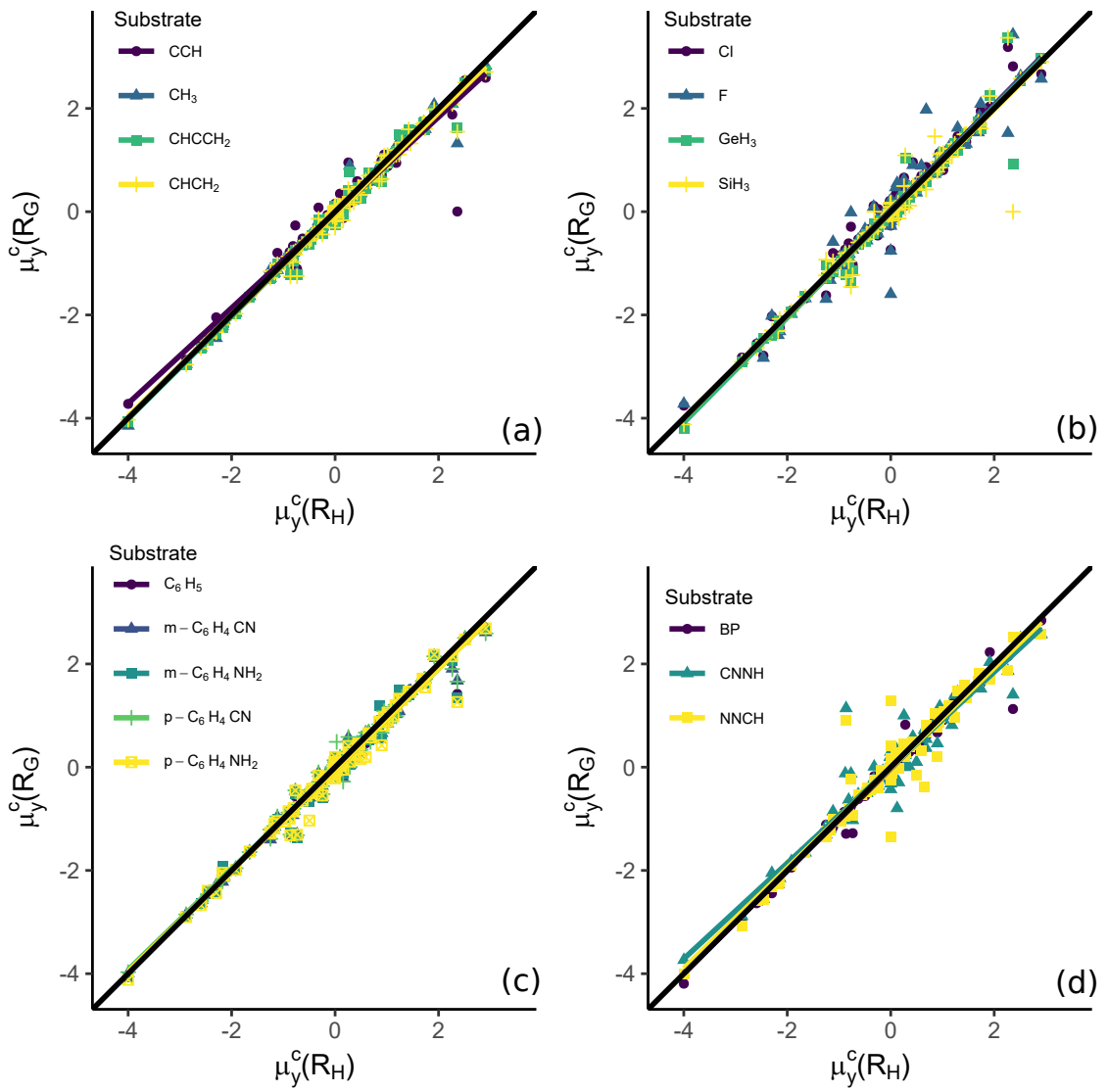

Figure B18: Linear plots of μ_y^c(R) for R in various R-G systems against μ_y^c(R). (A) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (B) G: Cl, F, SiH_3. GeH_3; (C) C_6H_5, m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; D: BP, CNNH, NNCH

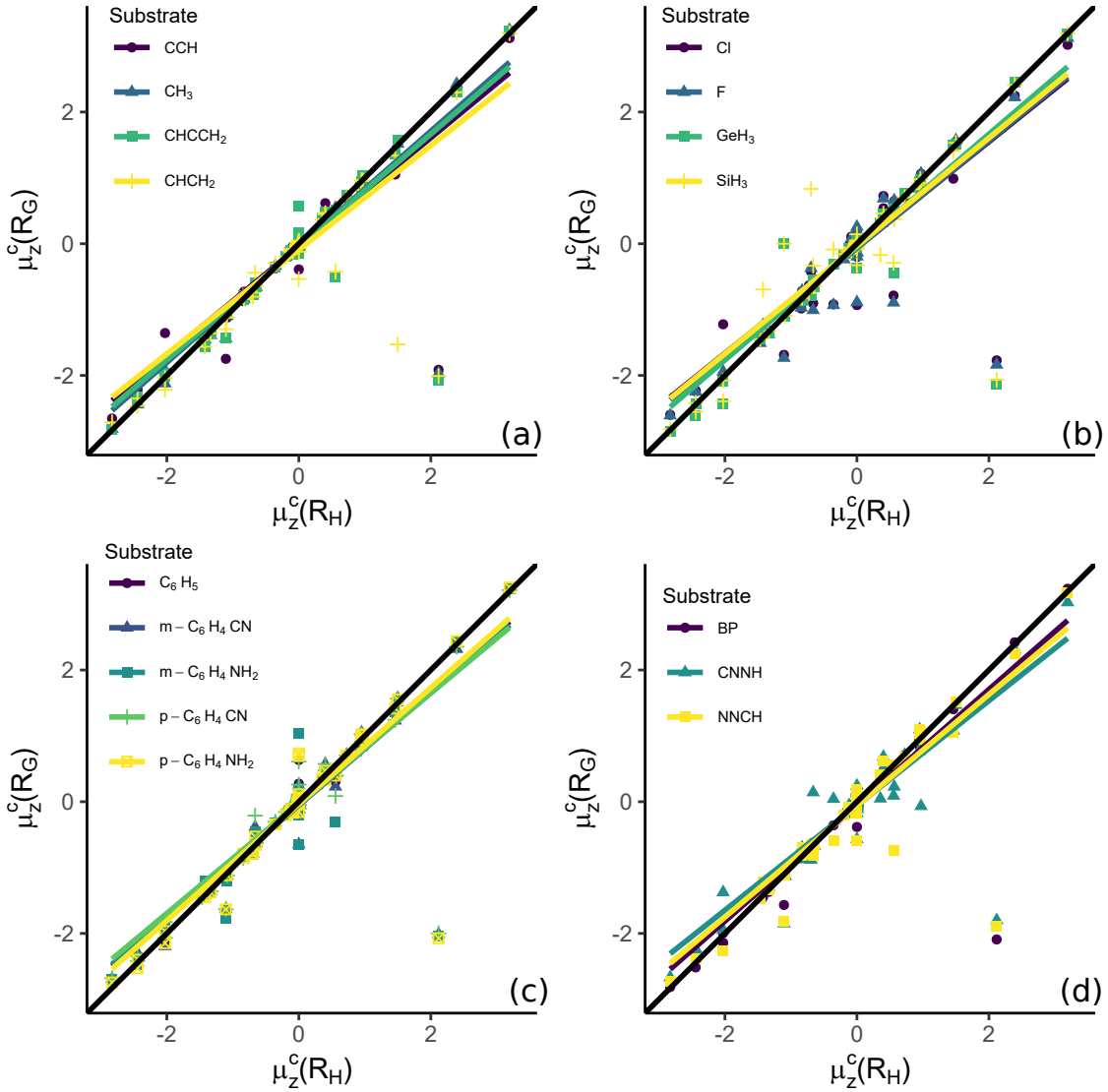

Figure B19: Linear plots of μ_z^c(R) for R in various R-G systems against μ_z^c(R). (A) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (B) G: Cl, F, SiH_3. GeH_3; (C) C_6H_5, m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; D: BP, CNNH, NNCH

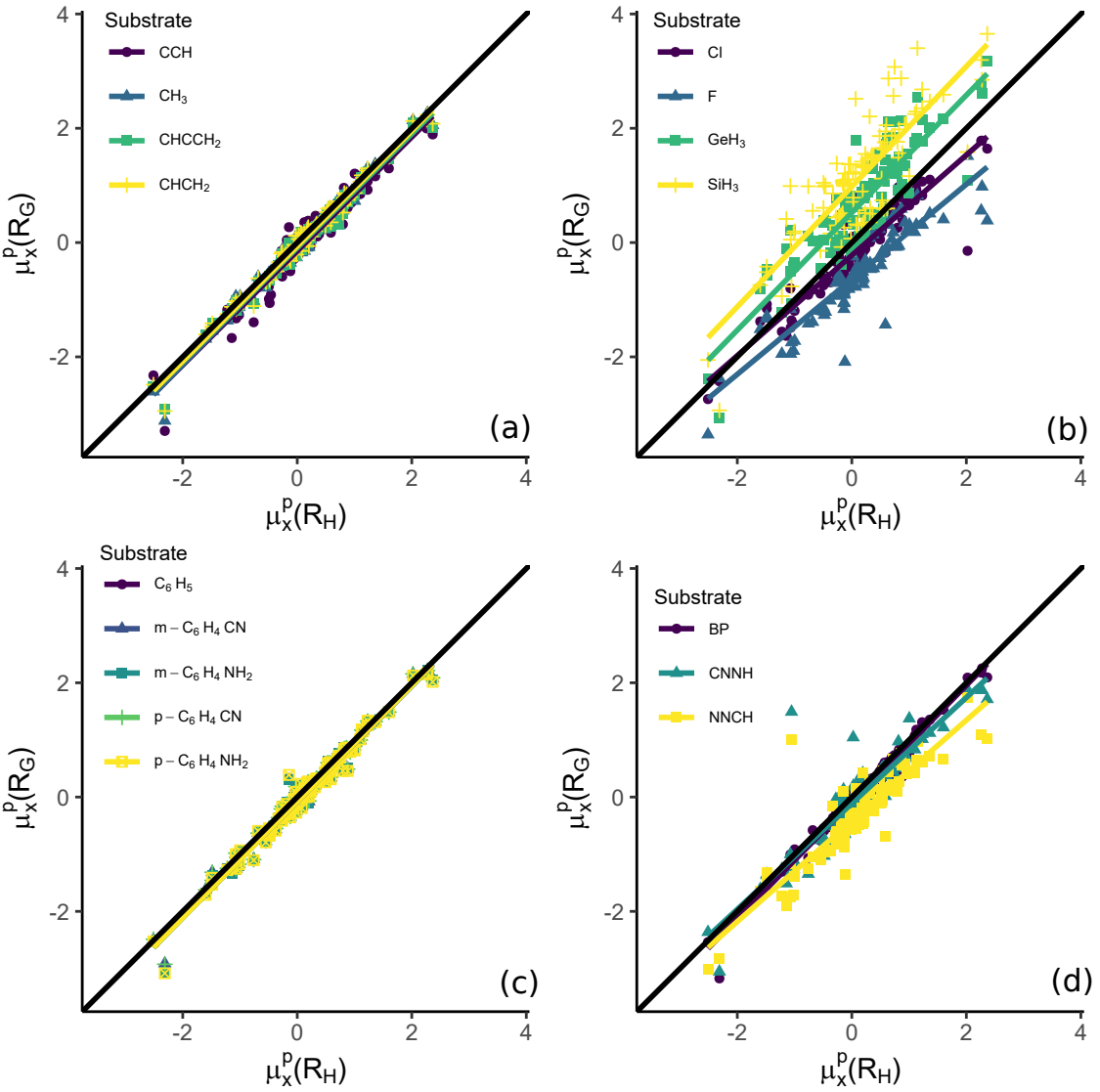

Figure B20: Linear plots of $\mu_x^p(R)$ for R in various R-G systems against $\mu_x^p(R)$. (A) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (B) G: Cl, F, SiH_3. GeH_3; (C) C_6H_5, m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; D: BP, CNNH, NNCH

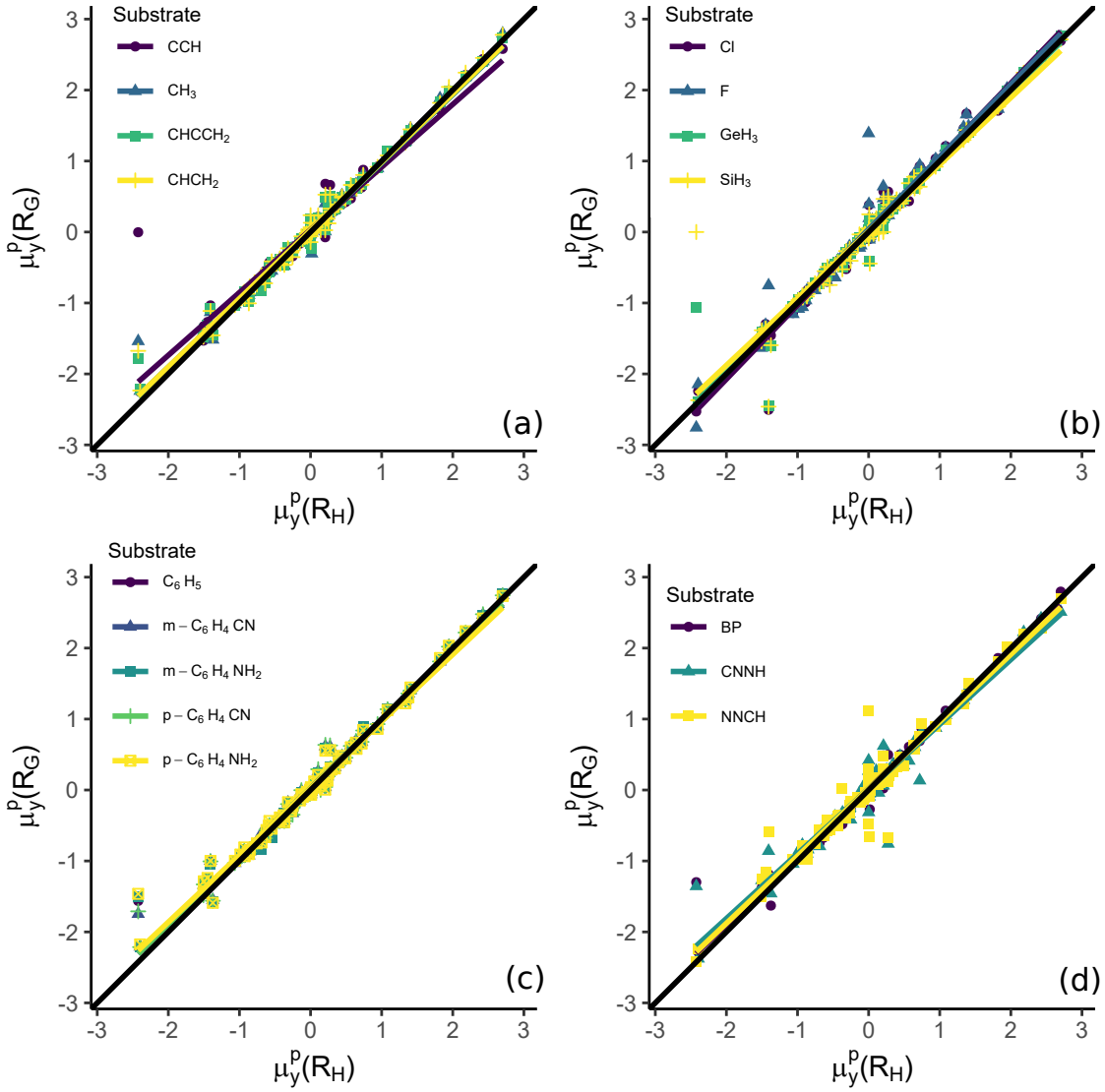

Figure B21: Linear plots of μ_y^p(R) for R in various R-G systems against μ_y^p(R). (A) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (B) G: Cl, F, SiH_3. GeH_3; (C) C_6H_5, m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; D: BP, CNNH, NNCH

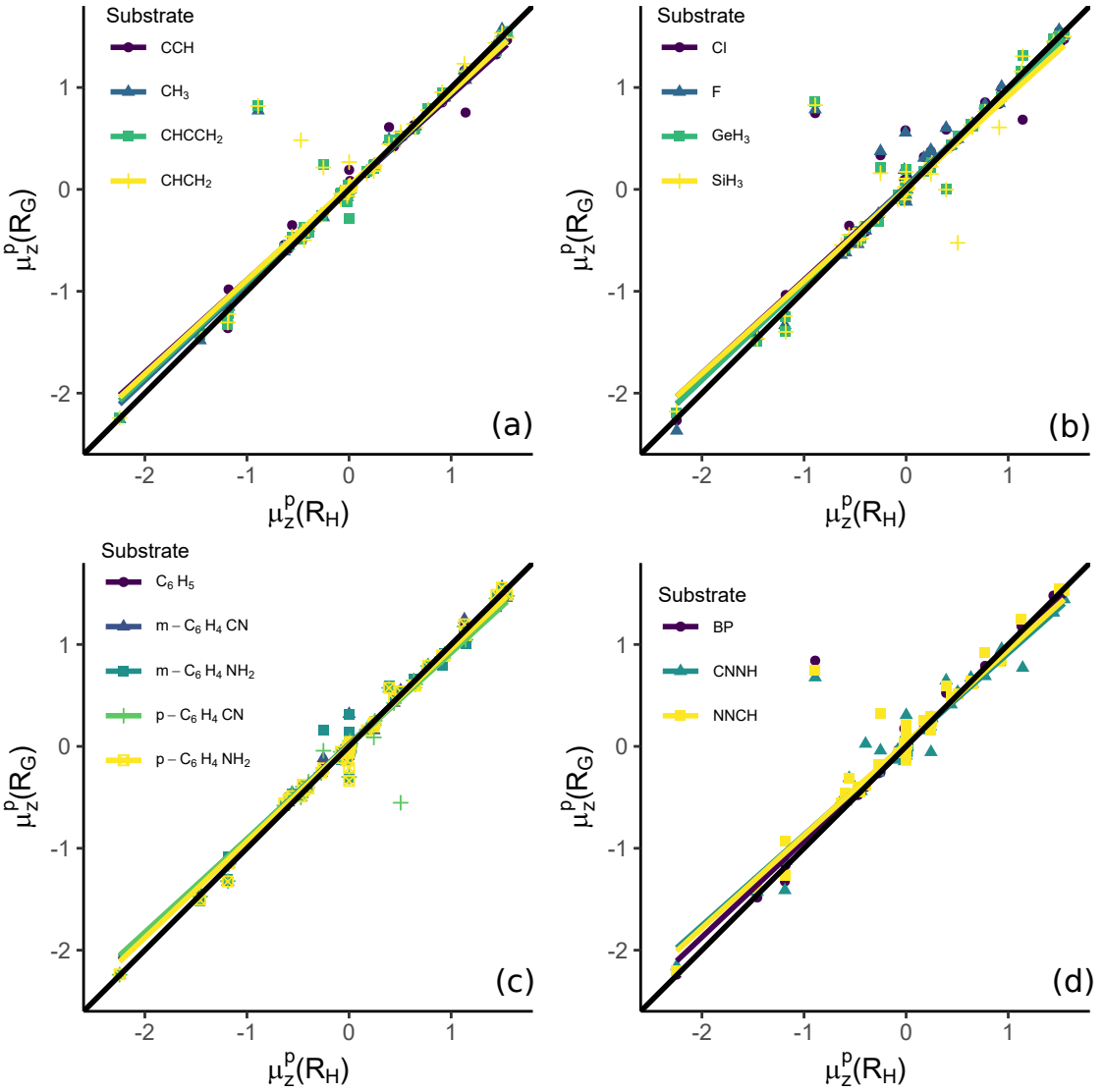

Figure B22: Linear plots of μ_z^p(R) for R in various R-G systems against μ_z^p(R). (A) G: CCH, CH_3, $CHCCH_2$, $CHCH_2$; (B) G: Cl, F, SiH_3. GeH_3; (C) C_6H_5, m$-C_6H_4CN$, m$-C_6H_4NH_2$, p$-C_6H_4CN$, p$-C_6H_4NH_2$; D: BP, CNNH, NNCH

Table B22: Summary statistics comparing Q_{xx}(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	0.149	0.247	0.312	0.984(PHO(OH))	-1.804(NCS)
C_6H_5	0.171	0.283	0.34	1.272(CHCCO)	-1.431(NCS)
CCH	0.177	0.291	0.369	1.430(CHCCO)	-2.065(NCS)
CH_3	0.15	0.227	0.271	1.017(CHCCO)	-1.485(NCS)
$CHCCH_2$	0.171	0.254	0.287	1.173(CHCCO)	-1.271(O^-)
$CHCH_2$	0.15	0.248	0.296	1.116(CHCCO)	-1.271(NCS)
Cl	0.123	0.31	0.676	0.968(ONO)	-6.370(CHCCO)
CNNH	0.256	0.349	0.407	1.783(CHCCO)	-1.397(NCS)
F	0.459	0.526	0.415	2.149(CHCCO)	-0.590(ONH_2)
GeH_3	-0.589	0.67	0.777	0.641(NCO)	-6.552(CHCCO)
m—C_6H_4CN	0.181	0.275	0.331	1.594(CHCCO)	-1.318(O^-)
m—$C_6H_4NH_2$	0.168	0.286	0.35	1.269(CHCCO)	-1.562(NCS)
NNCH	0.476	0.534	0.415	2.514(CHCCO)	-0.713(CH_2OH-ip)
p—C_6H_4CN	0.194	0.291	0.362	1.582(CHCCO)	-1.367(O^-)
p—$C_6H_4NH_2$	0.196	0.312	0.367	1.305(CHCCO)	-1.671(NCS)
SiH_3	-1.295	1.33	1.051	0.907(NCO)	-7.387(CHCCO)

Table B23: Summary statistics comparing Q_{yy}(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-0.123	0.219	0.3	1.426(CH_2OOCH_3-op)	-0.994(CHCCO)
C_6H_5	-0.114	0.234	0.323	1.396(CH_2OOCH_3-op)	-1.364(CHCCO)
CCH	-0.106	0.229	0.337	1.675(NCS)	-1.405(CHCCO)
CH_3	-0.13	0.207	0.281	1.449(CH_2OOCH_3-op)	-1.168(CHCCO)
$CHCCH_2$	-0.109	0.21	0.281	1.462(CH_2OOCH_3-op)	-1.245(CHCCO)
$CHCH_2$	-0.094	0.21	0.286	1.431(CH_2OOCH_3-op)	-1.156(CHCCO)
Cl	-0.009	0.213	0.667	6.596(CHCCO)	-0.465(Ocl)
CNNH	-0.118	0.261	0.37	1.344(CH_2OOCH_3-op)	-1.767(CHCCO)
F	-0.176	0.287	0.351	1.314(CH_2OOCH_3-op)	-1.856(CHCCO)
GeH_3	0.309	0.405	0.663	6.164(CHCCO)	-0.940(NCO)
m$-C_6H_4CN$	-0.103	0.221	0.311	1.467(CH_2OOCH_3-op)	-1.504(CHCCO)
m$-C_6H_4NH_2$	-0.115	0.25	0.345	1.362(CH_2OOCH_3-op)	-1.350(CHCCO)
NNCH	-0.275	0.351	0.387	1.282(CH_2OOCH_3-op)	-2.591(CHCCO)
p$-C_6H_4CN$	-0.113	0.232	0.334	1.452(CH_2OOCH_3-op)	-1.487(CHCCO)
p$-C_6H_4NH_2$	-0.148	0.27	0.367	1.378(CH_2OOCH_3-op)	-1.475(CHCCO)
SiH_3	0.673	0.729	0.76	6.656(CHCCO)	-1.486(NCO)

Table B24: Summary statistics comparing Q_{zz}(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-0.025	0.131	0.204	0.546(O^-)	-1.497(CH_2OOCH_3-op)
C_6H_5	-0.058	0.147	0.215	0.633(O^-)	-1.511(CH_2OOCH_3-op)
CCH	-0.071	0.144	0.224	0.700(O^-)	-1.518(CH_2OOCH_3-op)
CH_3	-0.02	0.116	0.192	0.407(NCS)	-1.515(CH_2OOCH_3-op)
$CHCCH_2$	-0.062	0.134	0.207	0.675(O^-)	-1.531(CH_2OOCH_3-op)
$CHCH_2$	-0.056	0.131	0.214	0.793(S^-)	-1.498(CH_2OOCH_3-op)
Cl	-0.113	0.195	0.256	0.905($PO(OH)_2$)	-1.372(CH_2OOCH_3-op)
CNNH	-0.138	0.187	0.245	0.630(O^-)	-1.456(CH_2OOCH_3-op)
F	-0.283	0.318	0.283	0.407($NHNH_2$-ip)	-1.546(CH_2OOCH_3-op)
GeH_3	0.28	0.32	0.284	1.151(S^-)	-1.025(CH_2OOCH_3-op)
m$-C_6H_4CN$	-0.079	0.143	0.204	0.645(O^-)	-1.484(CH_2OOCH_3-op)
m$-C_6H_4NH_2$	-0.052	0.158	0.232	0.816($CONH_2$)	-1.490(CH_2OOCH_3-op)
NNCH	-0.202	0.237	0.23	0.394(CH_2CSH-op)	-1.549(CH_2OOCH_3-op)
p$-C_6H_4CN$	-0.08	0.142	0.204	0.658(O^-)	-1.462(CH_2OOCH_3-op)
p$-C_6H_4NH_2$	-0.047	0.178	0.249	0.626(O^-)	-1.551(CH_2OOCH_3-op)
SiH_3	0.622	0.653	0.472	2.049(S^-)	-0.605(CH_2OOCH_3-op)

Table B25: Linear regression parameters comparing Q_{xx}(R) for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	0.922	0.187	0.941	0.022	0.029
C_6H_5	0.902	0.219	0.93	0.023	0.031
CCH	0.957	0.198	0.918	0.027	0.036
CH_3	0.932	0.183	0.956	0.019	0.026
$CHCCH_2$	0.912	0.214	0.951	0.019	0.026
$CHCH_2$	0.906	0.196	0.948	0.02	0.027
Cl	0.828	0.204	0.741	0.047	0.064
CNNH	0.954	0.279	0.903	0.029	0.04
F	0.952	0.482	0.9	0.031	0.042
GeH_3	0.867	-0.525	0.683	0.055	0.076
$m-C_6H_4CN$	0.919	0.221	0.934	0.023	0.031
$m-C_6H_4NH_2$	0.897	0.218	0.926	0.024	0.032
NNCH	0.886	0.531	0.895	0.029	0.04
$p-C_6H_4CN$	0.905	0.24	0.92	0.025	0.034
$p-C_6H_4NH_2$	0.892	0.248	0.918	0.025	0.034
SiH_3	0.848	-1.222	0.524	0.075	0.103

Table B26: Linear regression parameters comparing Q_{yy}(R) for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	0.904	-0.107	0.914	0.026	0.027
C_6H_5	0.864	-0.091	0.901	0.027	0.027
CCH	0.879	-0.085	0.891	0.029	0.03
CH_3	0.912	-0.115	0.924	0.024	0.025
$CHCCH_2$	0.885	-0.089	0.926	0.023	0.024
$CHCH_2$	0.889	-0.075	0.923	0.024	0.025
Cl	0.949	0	0.687	0.061	0.064
CNNH	0.893	-0.101	0.869	0.033	0.033
F	0.916	-0.161	0.884	0.032	0.033
GeH_3	0.967	0.315	0.69	0.061	0.063
$m-C_6H_4CN$	0.876	-0.08	0.908	0.026	0.027
$m-C_6H_4NH_2$	0.842	-0.089	0.888	0.028	0.029
NNCH	0.853	-0.246	0.857	0.033	0.035
$p-C_6H_4CN$	0.863	-0.09	0.895	0.028	0.028
$p-C_6H_4NH_2$	0.835	-0.12	0.873	0.03	0.031
SiH_3	0.97	0.678	0.63	0.069	0.071

Table B27: Linear regression parameters comparing Q_{zz}(R) for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	0.928	-0.072	0.975	0.014	0.019
C_6H_5	0.926	-0.106	0.972	0.015	0.02
CCH	0.936	-0.113	0.969	0.016	0.022
CH_3	0.934	-0.064	0.978	0.013	0.018
$CHCCH_2$	0.929	-0.109	0.974	0.014	0.02
$CHCH_2$	0.928	-0.103	0.973	0.015	0.02
Cl	0.946	-0.149	0.96	0.019	0.026
CNNH	0.931	-0.183	0.963	0.017	0.024
F	0.964	-0.306	0.951	0.021	0.03
GeH_3	0.961	0.254	0.948	0.021	0.03
$m-C_6H_4CN$	0.936	-0.121	0.975	0.014	0.02
$m-C_6H_4NH_2$	0.916	-0.107	0.968	0.015	0.022
NNCH	0.918	-0.257	0.97	0.015	0.022
$p-C_6H_4CN$	0.935	-0.123	0.975	0.014	0.02
$p-C_6H_4NH_2$	0.913	-0.105	0.963	0.017	0.023
SiH_3	0.946	0.587	0.864	0.035	0.049

Table B28: Summary statistics comparing Vol(R) for R-G molecules to those in R-H molecules

Substrate	MD	MAD	SD	Max	Min
BP	-11.852	11.941	8.663	1.886(SiCl)	-70.162(O^-)
C_6H_5	-16.997	16.997	10.242	-0.924(BeH)	-78.882(O^-)
CCH	-18.601	18.601	9.145	-3.048(BeH)	-79.067(O^-)
CH_3	-12.527	12.527	7.194	-0.943(SiCl)	-61.164(O^-)
$CHCCH_2$	-14.637	14.637	9.245	-0.570(SiCl)	-78.102(O^-)
$CHCH_2$	-15.159	15.159	9.43	-0.391(SiF)	-75.460(O^-)
Cl	-24.056	24.056	7.924	-3.141(BeH)	-70.236(O^-)
CNNH	-19.17	19.309	10.619	6.501(OOH)	-82.814(O^-)
F	-28.967	28.967	8.66	-3.712(BeH)	-63.353(O^-)
GeH_3	6.315	9.589	11.161	32.112(SiF_3)	-47.731(O^-)
m$-C_6H_4CN$	-19.488	19.488	10.598	-1.773(BeH)	-82.013(O^-)
m$-C_6H_4NH_2$	-16.803	16.803	10.384	-0.210(CO)	-79.087(O^-)
NNCH	-21.841	21.869	10.792	1.379(BeH)	-86.755(O^-)
p$-C_6H_4CN$	-19.617	19.617	10.943	-1.615(CO)	-83.767(O^-)
p$-C_6H_4NH_2$	-14.566	14.788	10.608	3.300(CO)	-78.781(O^-)
SiH_3	11.482	14.137	13.026	48.160(SiF_3)	-43.315(O^-)

Table B29: Linear regression parameters comparing Vol(R) for R-G molecules to those in R-H molecules

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	1.002	-12.649	0.995	0.006	2.6
C_6H_5	1	-17.068	0.994	0.007	3.076
CCH	1.002	-19.221	0.995	0.007	2.746
CH_3	1.005	-14.364	0.997	0.005	2.153
$CHCCH_2$	1.004	-16.106	0.995	0.007	2.773
$CHCH_2$	1.004	-16.882	0.995	0.007	2.827
Cl	1.002	-24.723	0.996	0.006	2.421
CNNH	0.997	-18.129	0.993	0.008	3.23
F	1.011	-33.183	0.996	0.006	2.605
GeH_3	1.01	2.22	0.993	0.008	3.33
$m-C_6H_4CN$	0.999	-19.123	0.993	0.008	3.185
$m-C_6H_4NH_2$	1	-16.936	0.993	0.008	3.118
NNCH	1.006	-24.058	0.993	0.008	3.354
$p-C_6H_4CN$	1	-19.768	0.993	0.008	3.286
$p-C_6H_4NH_2$	0.997	-13.564	0.993	0.008	3.184
SiH_3	1.02	3.654	0.99	0.009	3.836

Table B30: Outlier summary depicting substituents qualitatively detected as outliers, the properties for whose linear relationships they were outliers, the substrates for which said outlying behaviour occured, and the reason assessed for the outlying behaviour. Angle means a change in the angle brought about the outlier, Dihedral means a change in dihedral is the source of the outlying behaviour

Substituent	Property(Substrates)	Reason
CO_2^-	μ_x(CH_3,aro,$CHCH_2$,$CHCCH_2$,SiH_3,BP)	Charged
S^-	μ_x(SiH_3),Q_{xx}(SiH_3,GeH_3), Q_{yy}(SiH_3)	Charged
O^-	Q_{xx}($CHCH_2$,$CHCCH_2$,CCH,aro	Charged
CO^+	μ_x(CH_3)	Charged
NH_3^+	μ_x(carbon attached)	Charged
NCS	μ_x(CH_3), μ_y(F), Q_{xx}(carbon attached,NNCH) Q_{yy}(carbon attached,F,CNNH)	Angle
NCO	Q_{yy}(CCH,F,SiH_3)	Angle
OPH_2O	μ_x(aro,BP), μ_y(CCH,SiH_3,GeH_3,m$-C_6H_4$CN,CNNH,NNCH) μ_z(CCH,F,Cl,SiH_3,aro,BCP,CNNH,NNCH)	Dihedral
OCOH	μ_x(CNNH),μ_y(CNNH,NNCH)	Dihedral
OOH	μ_y(CNNH) ,μ_z(CNNH)	Dihedral
CH_2OOCH_3-op	μ_x(CCH), Q_{yy}(all) Q_{zz}(CH_3,$CHCH_2$,$CHCCH_2$,CCH)	Dihedral
$PO(OH)_2$	μ_z($CHCCH_2$,aromatic,NNCH)	Dihedral
CH_2OOCH_3-ip	μ_y(NNCH)	Dihedral
CHCCO	Q_{xx}(all) Q_{yy}(all not CCH, CH_3, CNNH)	Angle
CH_2NO_2	μ_z(All)	Dihedral

Table B31: Outlier summary depicting substituents qualitatively detected as outliers, the properties for whose linear relationships they were outliers, the substrates for which said outlying behaviour occured, and the reason assessed for the outlying behaviour. Angle means a change in the angle brought about the outlier, Dihedral means a change in dihedral is the source of the outlying behaviour. The aro abbreviation refres to all aromatic substrates

Substituent	Property(Substrates)	Reason
ONH_2	μ_y(F),Q_{yy}(F)	Interaction
$OO\cdot$	μ_x(F,Cl)	Interaction
ONO	μ_x(CCH), μ_y(CCH,CNNH)	Interaction
$NHNH_2$−ip	μ_y(p−C_6H_4CN),μ_z(F,Cl,CNNH),Q_{yy}(CNNH)	Planarity
$N(CH_3)_2$	μ_z($CHCH_2$,$CHCCH_2$,F,Cl,m−$C_6H_4NH_2$,NNCH)	Planarity
$NHCH_3$	μ_z(p−C_6H_4CN,CNNH)	Planarity
NHOH	μ_z(SiH_3,p−C_6H_4CN)	Planarity
$CONH_2$	μ_z(aro)	Planarity
CH_2COOH−ip	μ_x(SiH_3), μ_y(SiH_3,NNCH)	Misc
COCl	μ_y(GeH_3)	Misc
CH_2OH−ip	μ_y(NNCH),Q_{yy}(NNCH)	Misc
PHO(OH)	μ_x(CH_3),μ_y(SiH_3,GeH_3)	Misc
PH_2O	μ_z($CHCH_2$)	Misc

Appendix C

Appendix C

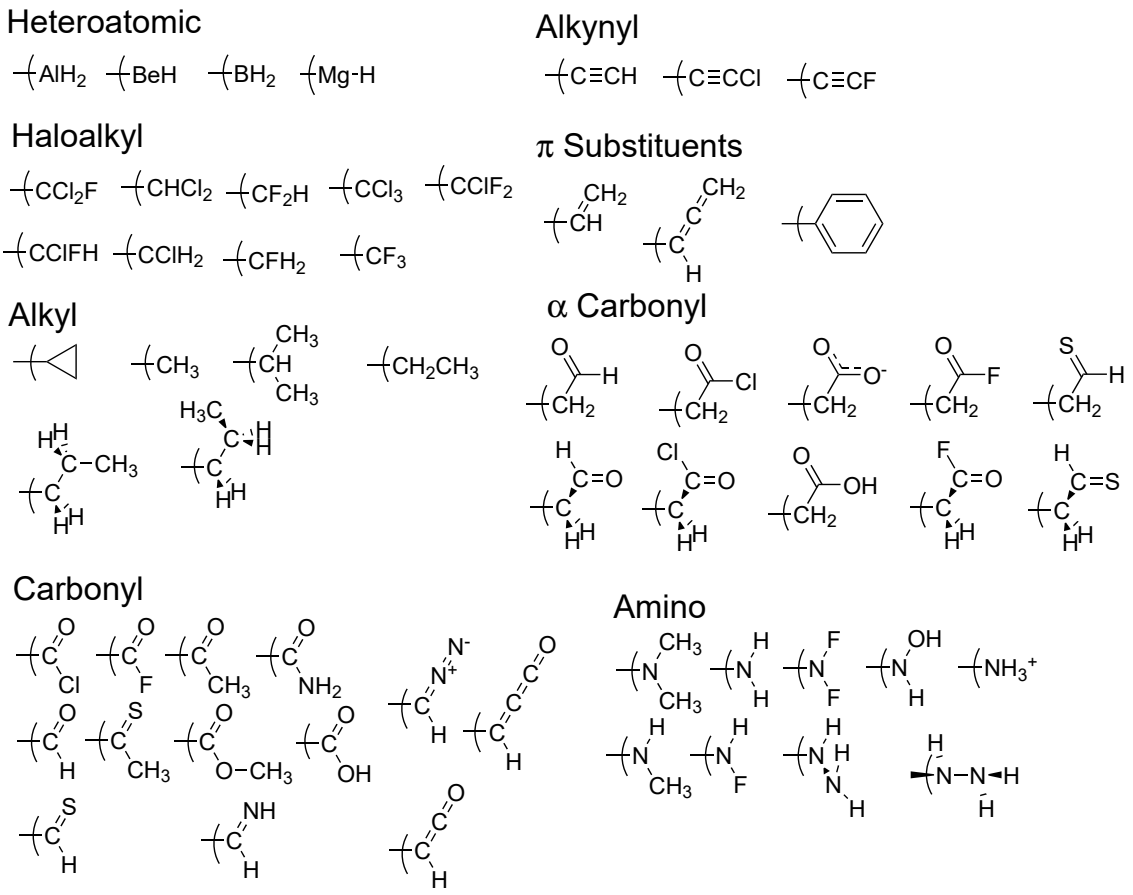

Figure C1: Part 1/2 of Substituents used in analysis

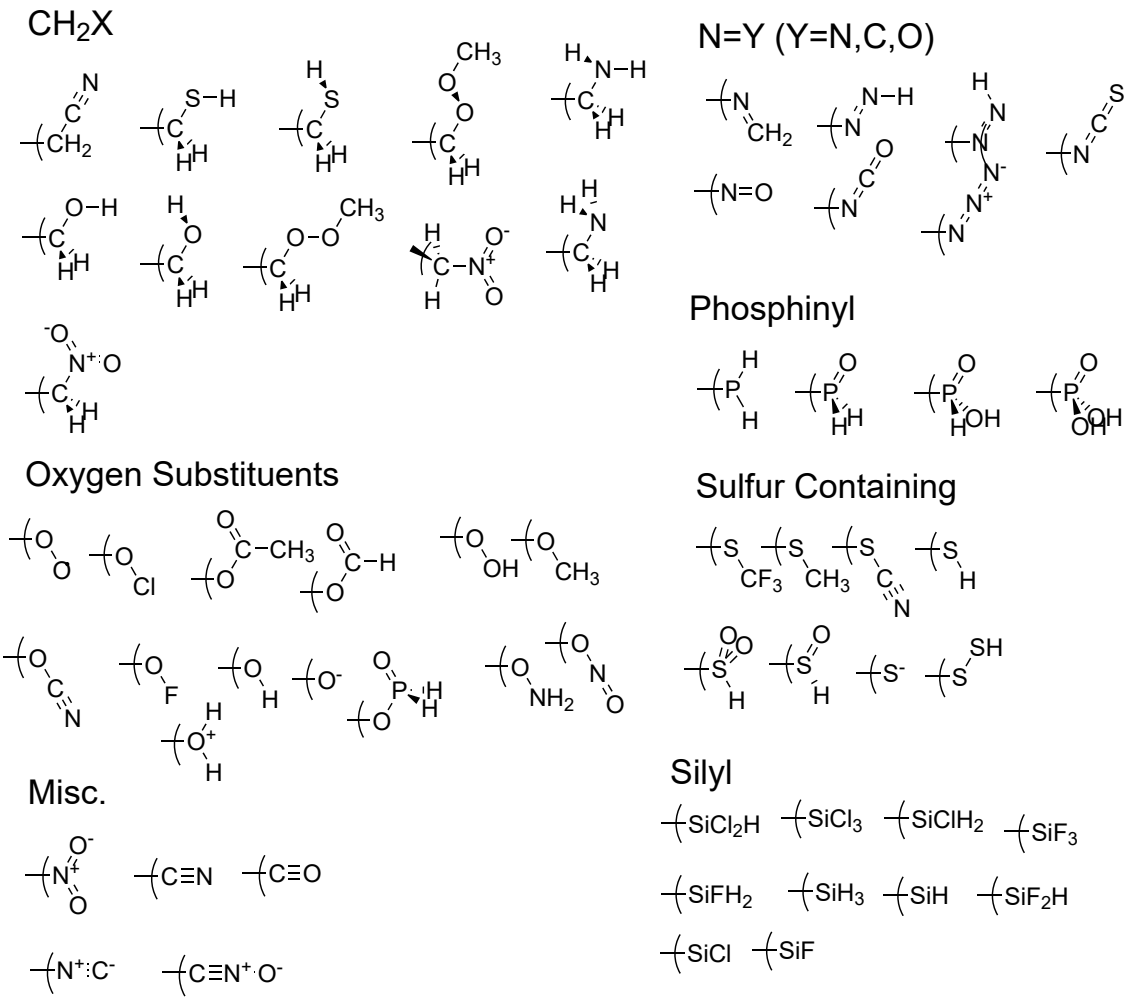

Figure C2: Part 2/2 of Substituents used in analysis

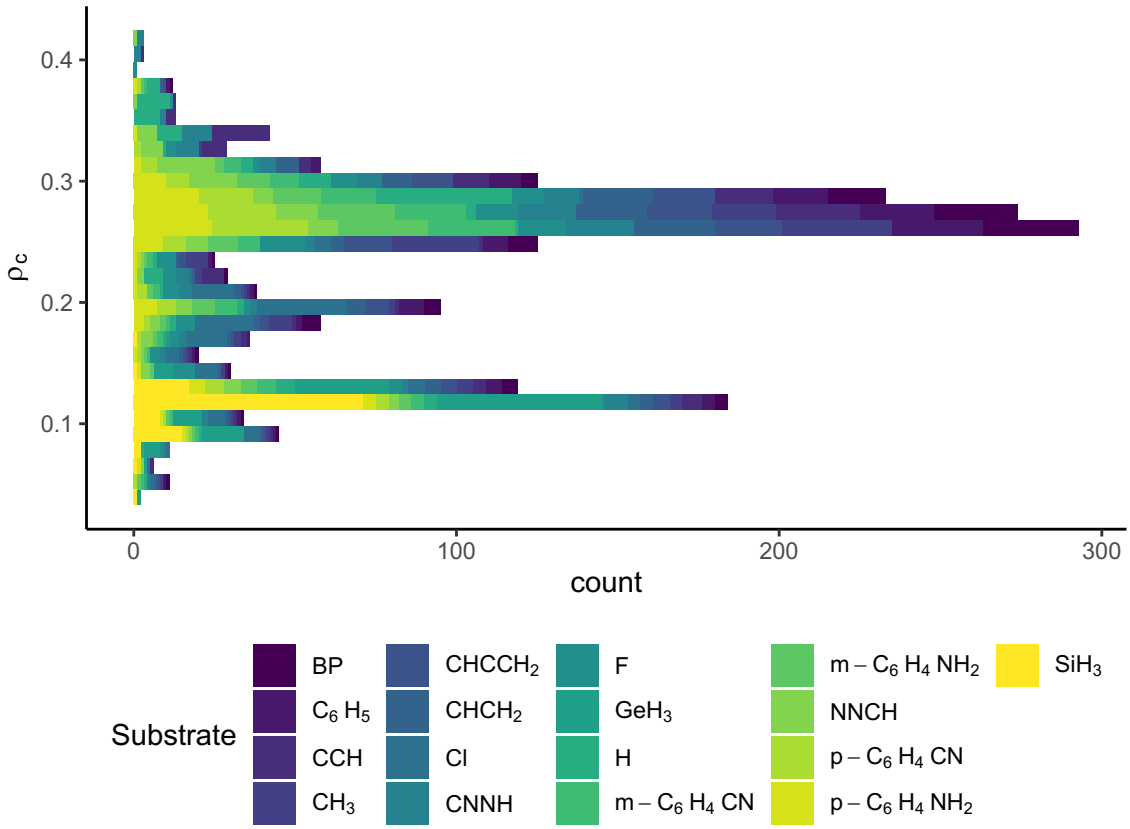

Figure C3: Histogram depicting range of ρ_c

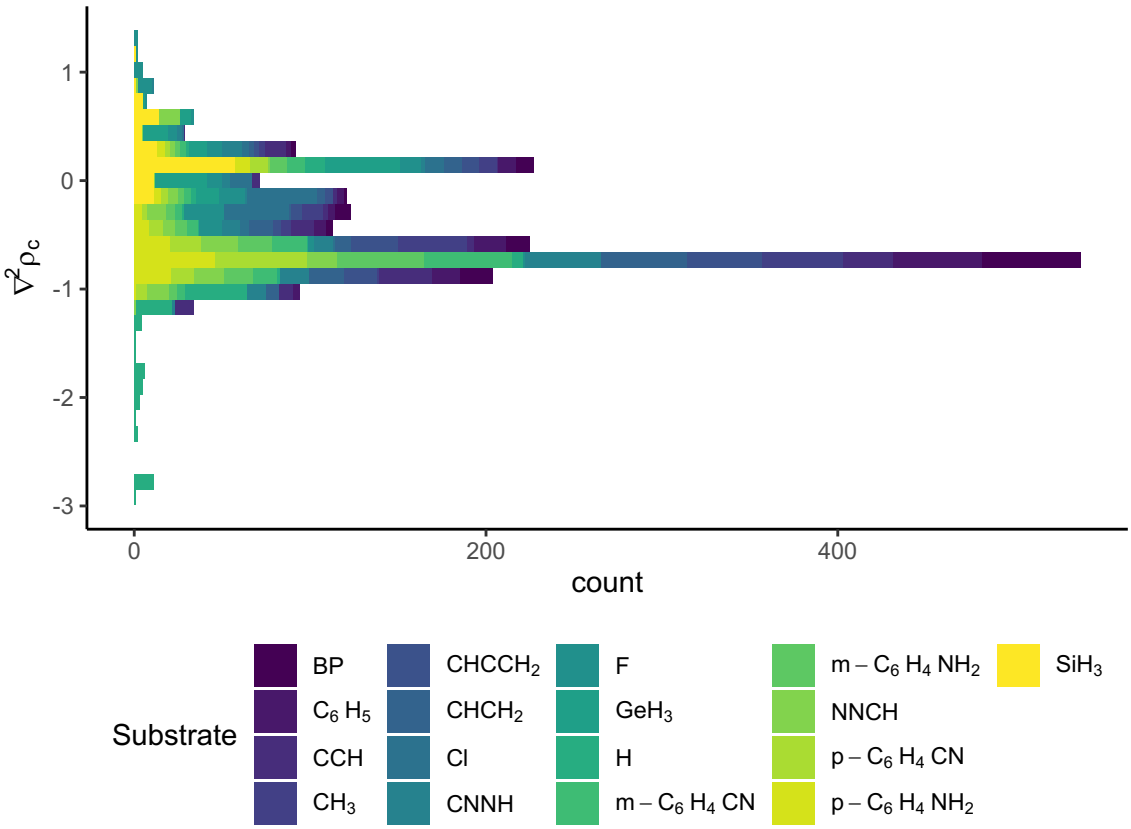

Figure C4: Histogram depicting range of $\nabla^2\rho_c$

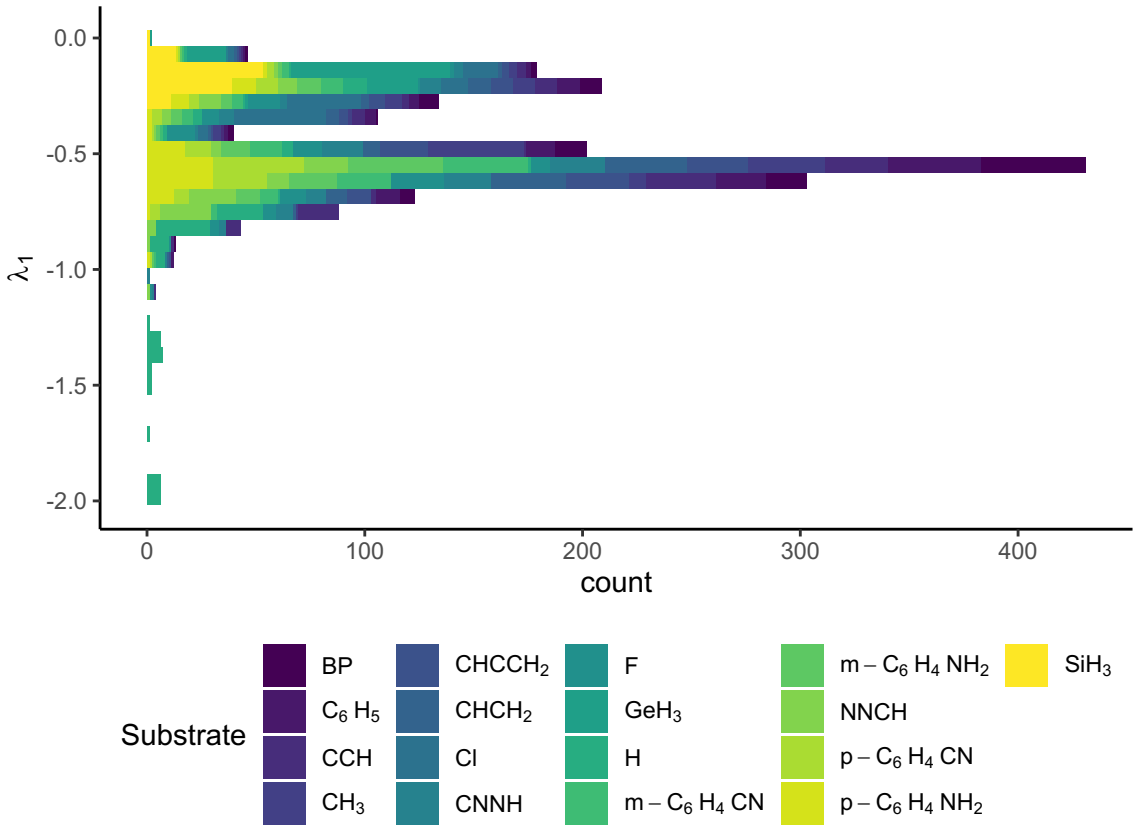

Figure C5: Histogram depicting range of $\lambda_{1,c}$

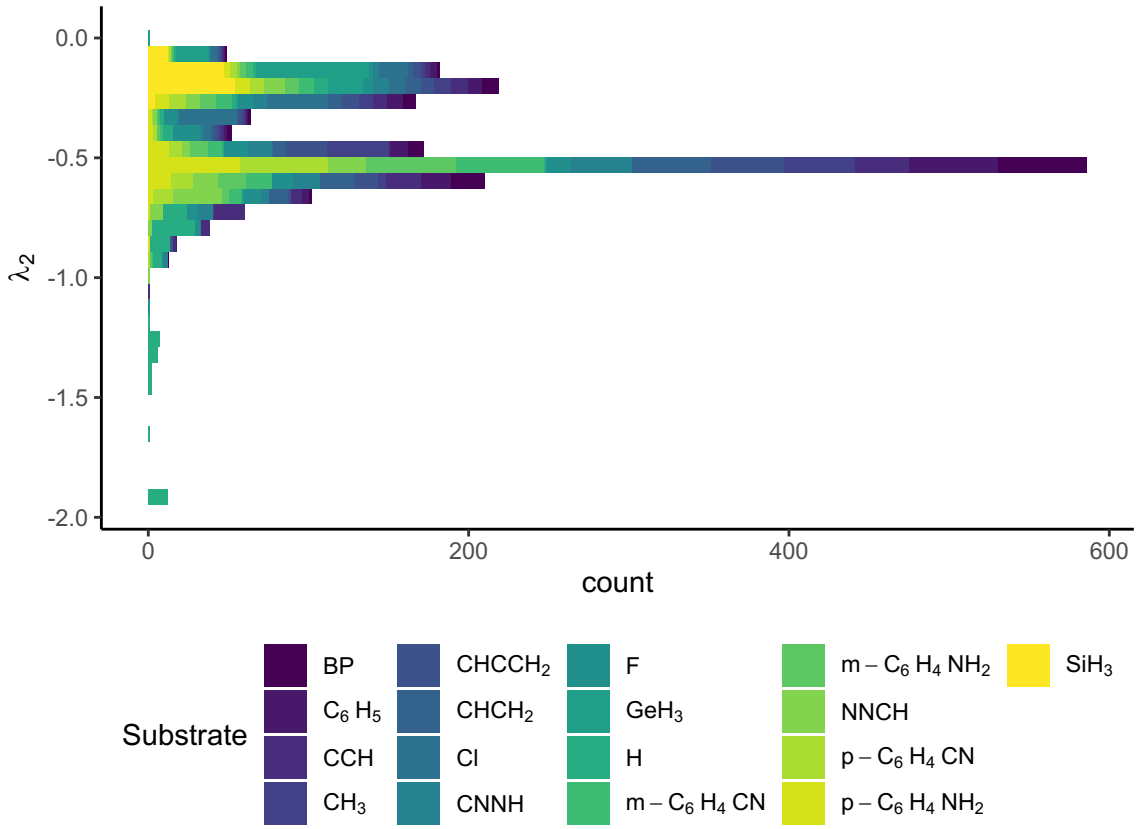

Figure C6: Histogram depicting range of $\lambda_{2,c}$

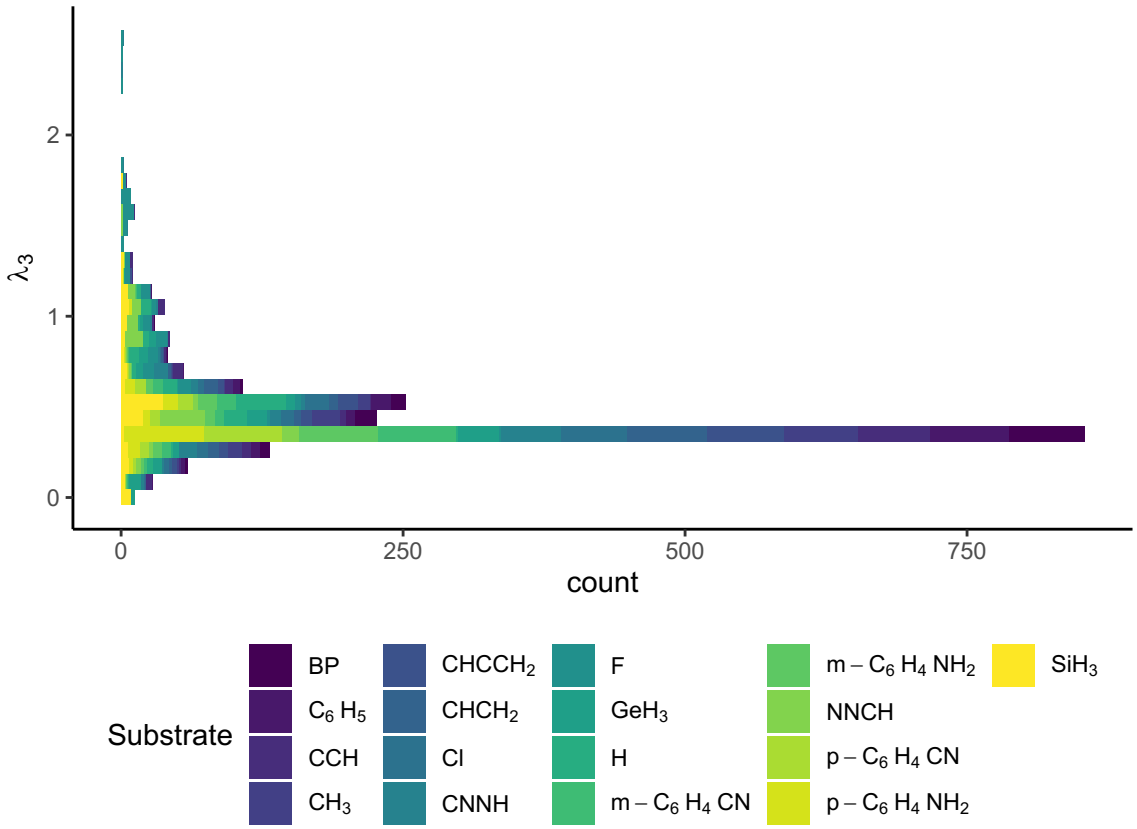

Figure C7: Histogram depicting range of $\lambda_{3,c}$

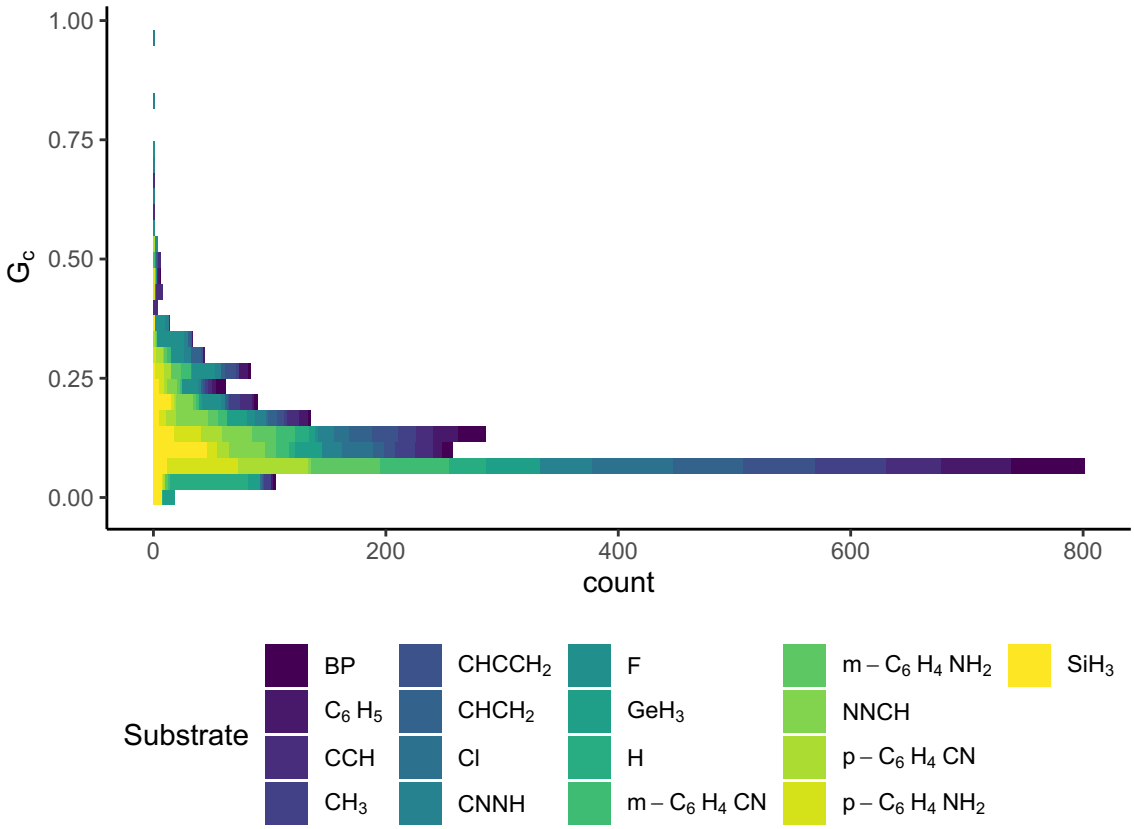

Figure C8: Histogram depicting range ofG_c

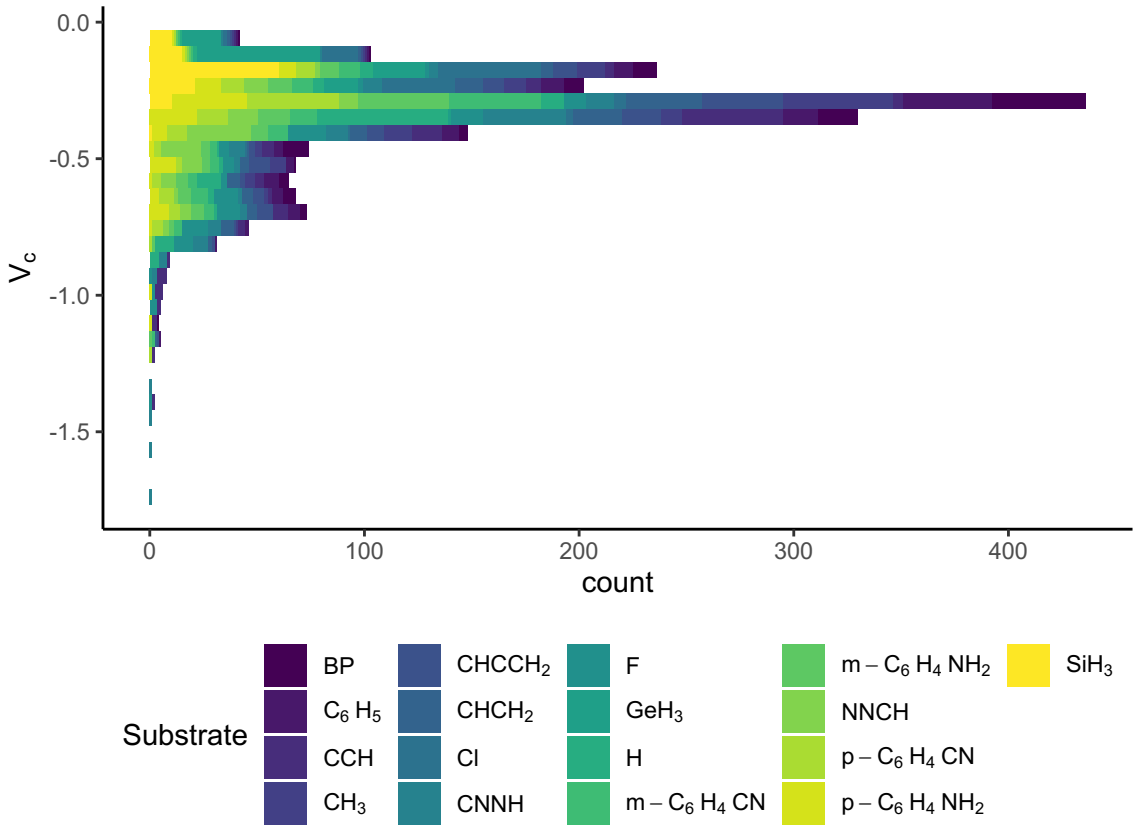

Figure C9: Histogram depicting range ofV_c

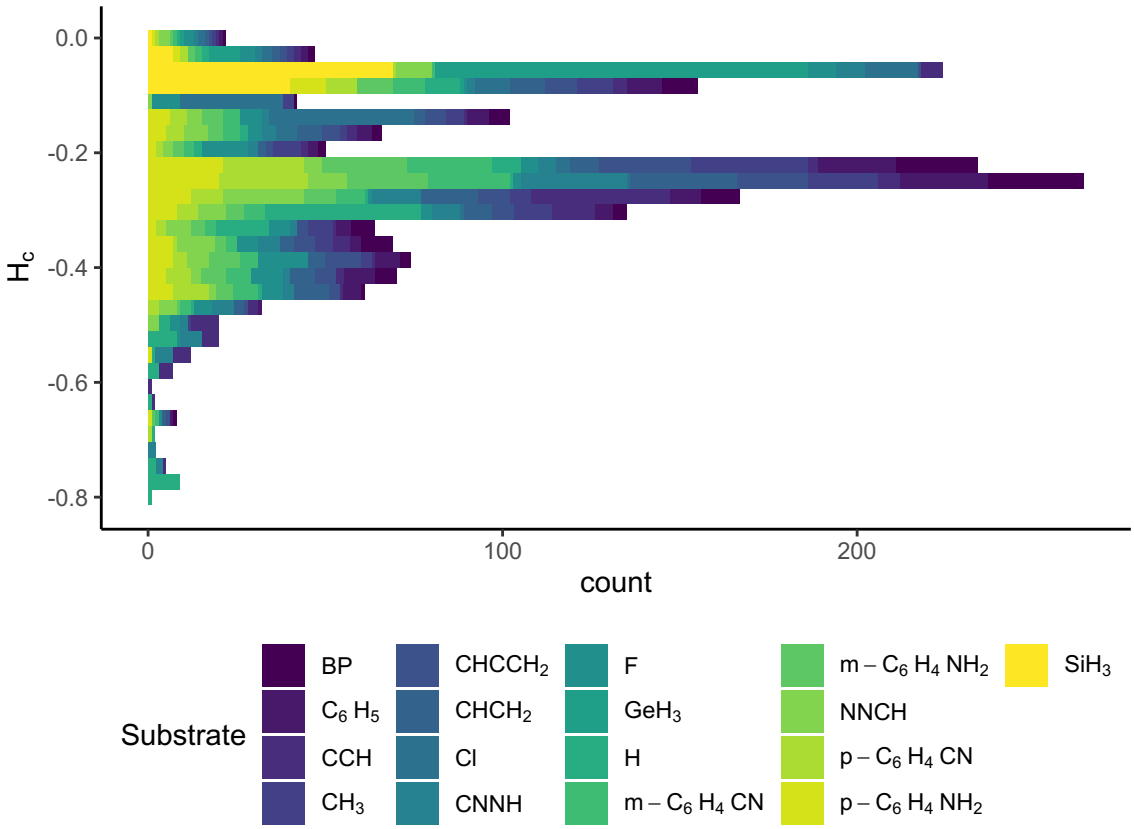

Figure C10: Histogram depicting range ofH_c

Table C1: Summary statistics showing error in ρ over all substrates compared to G=H

Substrate	MD	MAD	SD	Max	Min
BP	-0.034	0.035	0.032	0.01	-0.127
C_6H_5	-0.031	0.032	0.029	0.023	-0.115
CCH	-0.007	0.012	0.016	0.053	-0.056
CH_3	-0.044	0.045	0.034	0.008	-0.135
$CHCCH_2$	-0.032	0.033	0.028	0.02	-0.117
$CHCH_2$	-0.026	0.028	0.028	0.015	-0.111
Cl	-0.104	0.104	0.053	0.007	-0.307
CNNH	-0.014	0.023	0.027	0.092	-0.093
F	-0.038	0.047	0.049	0.11	-0.22
GeH_3	-0.163	0.163	0.06	NA	-0.248
$m-C_6H_4CN$	-0.03	0.031	0.027	0.021	-0.109
$m-C_6H_4NH_2$	-0.031	0.032	0.029	0.023	-0.116
NNCH	-0.015	0.027	0.034	0.049	-0.137
$p-C_6H_4CN$	-0.029	0.03	0.027	0.022	-0.108
$p-C_6H_4NH_2$	-0.029	0.031	0.03	0.034	-0.119
SiH_3	-0.164	0.164	0.063	0.002	-0.25

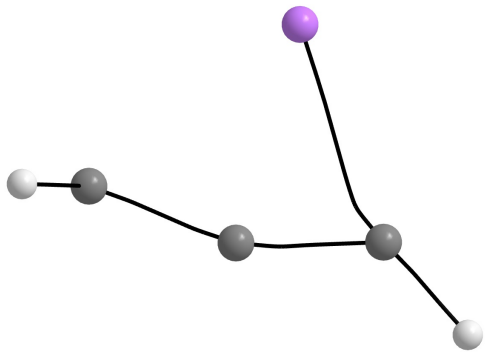

Figure C11: Optimized Li-$CHCCH_2$ structure visualized in AIMStudio showing bond paths

Table C2: Linear regression statistics relating ρ at the R-G bond to that in the R-H bond

Substrate	Slope	Intercept	r^2	Slope SE	IntSE
BP	0.682	0.055	0.844	0.027	0.008
C_6H_5	0.718	0.048	0.875	0.025	0.007
CCH	0.924	0.014	0.953	0.019	0.006
CH_3	0.624	0.06	0.84	0.025	0.007
$CHCCH_2$	0.737	0.041	0.881	0.025	0.007
$CHCH_2$	0.736	0.047	0.883	0.025	0.007
Cl	0.414	0.059	0.542	0.036	0.01
CNNH	0.913	0.01	0.869	0.033	0.01
F	0.579	0.079	0.586	0.047	0.014
GeH_3	0.199	0.061	0.7	0.012	0.004
$m-C_6H_4CN$	0.739	0.043	0.889	0.024	0.007
$m-C_6H_4NH_2$	0.713	0.049	0.871	0.026	0.007
NNCH	0.819	0.035	0.797	0.039	0.011
$p-C_6H_4CN$	0.744	0.042	0.888	0.025	0.007
$p-C_6H_4NH_2$	0.718	0.049	0.867	0.026	0.008
SiH_3	0.166	0.068	0.601	0.013	0.004

Table C3: Summary statistics showing error in $\nabla^2\rho$ over all substrates compared to G=H

Substrate	MD	MAD	SD	Max	Min
BP	0.62	0.624	0.634	2.302	-0.098
C_6H_5	0.621	0.623	0.642	2.361	-0.065
CCH	0.576	0.576	0.688	2.654	NA
CH_3	0.675	0.68	0.652	2.372	-0.137
$CHCCH_2$	0.634	0.635	0.639	2.343	-0.048
$CHCH_2$	0.609	0.61	0.658	2.376	-0.051
Cl	1.058	1.059	0.783	2.906	-0.05
CNNH	0.619	0.619	0.667	3.194	NA
F	1.278	1.278	0.797	3.348	NA
GeH_3	1.346	1.402	0.979	3.415	-0.314
$m-C_6H_4CN$	0.619	0.621	0.637	2.336	-0.051
$m-C_6H_4NH_2$	0.622	0.624	0.643	2.363	-0.068
NNCH	0.71	0.71	0.713	2.809	NA
$p-C_6H_4CN$	0.614	0.616	0.639	2.34	-0.048
$p-C_6H_4NH_2$	0.625	0.626	0.648	2.398	-0.056
SiH_3	1.429	1.484	1.027	3.61	-0.342

Table C4: Linear regression statistics relating $\nabla^2\rho$ at the R-G bond to that in the R-H bond

Substrate	Slope	Intercept	r^2	Slope SE	IntSE
BP	0.26	-0.267	0.431	0.028	0.04
C_6H_5	0.258	-0.269	0.399	0.03	0.042
CCH	0.267	-0.302	0.253	0.043	0.062
CH_3	0.228	-0.251	0.407	0.026	0.037
$CHCCH_2$	0.26	-0.253	0.407	0.029	0.042
$CHCH_2$	0.247	-0.294	0.34	0.032	0.046
Cl	0.058	-0.07	0.064	0.021	0.03
CNNH	0.291	-0.211	0.259	0.046	0.065
F	0.2	0.316	0.116	0.053	0.077
GeH_3	-0.223	-0.122	0.862	0.008	0.012
$m-C_6H_4CN$	0.266	-0.262	0.422	0.029	0.042
$m-C_6H_4NH_2$	0.256	-0.27	0.394	0.03	0.043
NNCH	0.29	-0.123	0.205	0.055	0.077
$p-C_6H_4CN$	0.262	-0.271	0.404	0.03	0.043
$p-C_6H_4NH_2$	0.254	-0.269	0.375	0.031	0.044
SiH_3	-0.286	-0.114	0.847	0.011	0.016

Table C5: Summary statistics showing error in λ_1 over all substrates compared to G=H

Substrate	MD	MAD	SD	Max	Min
BP	0.419	0.419	0.406	1.526	-0.007
C_6H_5	0.407	0.407	0.399	1.508	-0.014
CCH	0.348	0.348	0.334	1.287	-0.013
CH_3	0.455	0.455	0.422	1.574	NA
$CHCCH_2$	0.41	0.41	0.394	1.502	-0.005
$CHCH_2$	0.396	0.396	0.398	1.501	-0.011
Cl	0.637	0.637	0.463	1.864	NA
CNNH	0.363	0.363	0.332	1.406	-0.015
F	0.392	0.423	0.43	1.684	-0.232
GeH_3	0.763	0.763	0.466	1.808	NA
$m-C_6H_4CN$	0.405	0.406	0.394	1.486	-0.012
$m-C_6H_4NH_2$	0.408	0.408	0.4	1.511	-0.015
NNCH	0.32	0.327	0.366	1.483	-0.088
$p-C_6H_4CN$	0.402	0.402	0.392	1.485	-0.011
$p-C_6H_4NH_2$	0.403	0.404	0.397	1.516	-0.018
SiH_3	0.739	0.739	0.454	1.764	NA

Table C6: Linear regression statistics relating λ_1 at the R-G bond to that in the R-H bond

Substrate	Slope	Intercept	r^2	Slope SE	IntSE
BP	0.21	-0.299	0.485	0.02	0.021
C_6H_5	0.225	-0.297	0.52	0.02	0.021
CCH	0.351	-0.241	0.789	0.017	0.017
CH_3	0.172	-0.296	0.426	0.019	0.019
$CHCCH_2$	0.235	-0.285	0.542	0.02	0.021
$CHCH_2$	0.228	-0.305	0.515	0.021	0.021
Cl	0.091	-0.188	0.311	0.013	0.013
CNNH	0.339	-0.225	0.719	0.02	0.02
F	0.204	-0.333	0.351	0.027	0.028
GeH_3	0.065	-0.086	0.758	0.003	0.004
$m-C_6H_4CN$	0.238	-0.287	0.554	0.02	0.021
$m-C_6H_4NH_2$	0.223	-0.298	0.513	0.02	0.021
NNCH	0.302	-0.302	0.533	0.027	0.027
$p-C_6H_4CN$	0.239	-0.289	0.552	0.02	0.021
$p-C_6H_4NH_2$	0.231	-0.295	0.527	0.02	0.021
SiH_3	0.085	-0.092	0.781	0.004	0.004

Table C7: Summary statistics showing error in λ_2 over all substrates compared to G=H

Substrate	MD	MAD	SD	Max	Min
BP	0.411	0.412	0.398	1.507	-0.007
C_6H_5	0.416	0.416	0.389	1.487	-0.009
CCH	0.349	0.349	0.326	1.274	-0.013
CH_3	0.446	0.446	0.413	1.55	NA
$CHCCH_2$	0.426	0.426	0.392	1.508	NA
$CHCH_2$	0.404	0.404	0.386	1.474	-0.004
Cl	0.627	0.627	0.456	1.819	NA
CNNH	0.376	0.376	0.339	1.442	-0.001
F	0.402	0.433	0.424	1.641	-0.232
GeH_3	0.748	0.748	0.46	1.773	NA
$m-C_6H_4CN$	0.415	0.415	0.384	1.467	0
$m-C_6H_4NH_2$	0.415	0.416	0.389	1.487	-0.01
NNCH	0.345	0.35	0.365	1.497	-0.088
$p-C_6H_4CN$	0.41	0.41	0.382	1.462	0
$p-C_6H_4NH_2$	0.418	0.418	0.392	1.513	-0.021
SiH_3	0.725	0.725	0.451	1.736	NA

Table C8: Linear regression statistics relating λ_2 at the R-G bond to that in the R-H bond

Substrate	Slope	Intercept	r^2	Slope SE	IntSE
BP	0.211	-0.287	0.497	0.02	0.02
C_6H_5	0.227	-0.269	0.566	0.019	0.019
CCH	0.355	-0.223	0.803	0.016	0.017
CH_3	0.175	-0.285	0.44	0.018	0.019
$CHCCH_2$	0.22	-0.266	0.546	0.019	0.019
$CHCH_2$	0.233	-0.275	0.568	0.019	0.019
Cl	0.09	-0.179	0.311	0.013	0.013
CNNH	0.311	-0.221	0.685	0.02	0.02
F	0.202	-0.309	0.343	0.027	0.028
GeH_3	0.059	-0.087	0.715	0.003	0.004
$m-C_6H_4CN$	0.238	-0.261	0.598	0.018	0.019
$m-C_6H_4NH_2$	0.226	-0.271	0.56	0.019	0.019
NNCH	0.283	-0.279	0.539	0.025	0.025
$p-C_6H_4CN$	0.242	-0.262	0.6	0.018	0.019
$p-C_6H_4NH_2$	0.221	-0.273	0.551	0.019	0.019
SiH_3	0.075	-0.094	0.748	0.004	0.004

Table C9: Summary statistics showing error in λ_3 over all substrates compared to G=H

Substrate	MD	MAD	SD	Max	Min
BP	-0.21	0.228	0.192	0.128	-0.731
C_6H_5	-0.201	0.222	0.179	0.178	-0.633
CCH	-0.12	0.209	0.213	0.818	-0.469
CH_3	-0.226	0.24	0.197	0.098	-0.752
$CHCCH_2$	-0.201	0.222	0.181	0.239	-0.667
$CHCH_2$	-0.191	0.212	0.171	0.213	-0.6
Cl	-0.207	0.224	0.16	0.116	-0.777
CNNH	-0.12	0.251	0.305	1.623	-0.53
F	0.484	0.484	0.399	2.002	-0.022
GeH_3	-0.165	0.188	0.159	0.316	-0.556
$m-C_6H_4CN$	-0.201	0.22	0.176	0.217	-0.618
$m-C_6H_4NH_2$	-0.201	0.222	0.18	0.175	-0.635
NNCH	0.045	0.158	0.232	1.092	-0.456
$p-C_6H_4CN$	-0.198	0.218	0.175	0.317	-0.607
$p-C_6H_4NH_2$	-0.196	0.22	0.183	0.229	-0.646
SiH_3	-0.036	0.138	0.206	0.759	-0.573

Table C10: Linear regression statistics relating λ_3 at the R-G bond to that in the R-H bond

Substrate	Slope	Intercept	r^2	Slope SE	IntSE
BP	0.249	0.237	0.282	0.037	0.024
C_6H_5	0.342	0.19	0.374	0.041	0.026
CCH	0.991	-0.115	0.526	0.088	0.056
CH_3	0.187	0.257	0.268	0.029	0.018
$CHCCH_2$	0.343	0.19	0.362	0.042	0.027
$CHCH_2$	0.412	0.159	0.433	0.044	0.028
Cl	0.377	0.165	0.596	0.03	0.019
CNNH	0.896	-0.059	0.293	0.131	0.082
F	0.937	0.521	0.229	0.166	0.107
GeH_3	0.884	-0.096	0.623	0.064	0.041
$m-C_6H_4CN$	0.363	0.178	0.401	0.042	0.027
$m-C_6H_4NH_2$	0.34	0.191	0.368	0.042	0.026
NNCH	0.95	0.074	0.453	0.1	0.063
$p-C_6H_4CN$	0.387	0.167	0.403	0.044	0.028
$p-C_6H_4NH_2$	0.372	0.178	0.35	0.047	0.03
SiH_3	1.15	-0.125	0.623	0.083	0.053

Table C11: Summary statistics showing error in *G* over all substrates compared to G=H

Substrate	MD	MAD	SD	Max	Min
BP	0.048	0.049	0.059	0.353	-0.032
C_6H_5	0.057	0.058	0.071	0.378	-0.026
CCH	0.103	0.103	0.123	0.549	NA
CH_3	0.04	0.042	0.047	0.179	-0.048
$CHCCH_2$	0.057	0.058	0.07	0.401	-0.027
$CHCH_2$	0.064	0.065	0.078	0.388	-0.02
Cl	0.023	0.03	0.028	0.135	-0.072
CNNH	0.107	0.107	0.143	0.922	NA
F	0.229	0.229	0.104	0.73	NA
GeH_3	0.032	0.061	0.062	0.166	-0.112
$m-C_6H_4CN$	0.057	0.058	0.073	0.398	-0.024
$m-C_6H_4NH_2$	0.057	0.058	0.071	0.38	-0.026
NNCH	0.095	0.095	0.037	0.211	NA
$p-C_6H_4CN$	0.059	0.06	0.076	0.423	-0.021
$p-C_6H_4NH_2$	0.061	0.062	0.075	0.382	-0.022
SiH_3	0.063	0.088	0.072	0.251	-0.114

Table C12: Linear regression statistics relating *G* at the R-G bond to that in the R-H bond

Substrate	Slope	Intercept	r^2	Slope SE	IntSE
BP	0.975	0.049	0.182	0.193	0.012
C_6H_5	1.054	0.054	0.153	0.231	0.015
CCH	1.504	0.075	0.11	0.398	0.025
CH_3	0.796	0.051	0.19	0.153	0.01
$CHCCH_2$	1.028	0.056	0.149	0.229	0.015
$CHCH_2$	1.12	0.057	0.144	0.255	0.016
Cl	0.393	0.059	0.212	0.072	0.005
CNNH	1.352	0.087	0.069	0.467	0.03
F	-0.562	0.32	0.031	0.305	0.02
GeH_3	-0.4	0.112	0.054	0.156	0.01
$m-C_6H_4CN$	1.065	0.054	0.15	0.237	0.015
$m-C_6H_4NH_2$	1.056	0.054	0.154	0.231	0.015
NNCH	0.938	0.099	0.35	0.122	0.008
$p-C_6H_4CN$	1.099	0.054	0.146	0.247	0.016
$p-C_6H_4NH_2$	1.065	0.057	0.141	0.245	0.016
SiH_3	-0.477	0.147	0.051	0.191	0.012

Table C13: Summary statistics showing error in *V* over all substrates compared to G=H

Substrate	MD	MAD	SD	Max	Min
BP	0.06	0.072	0.084	0.335	-0.306
C_6H_5	0.041	0.059	0.071	0.23	-0.334
CCH	-0.063	0.066	0.096	0.018	-0.583
CH_3	0.089	0.094	0.088	0.346	-0.164
$CHCCH_2$	0.044	0.06	0.072	0.262	-0.359
$CHCH_2$	0.024	0.046	0.063	0.204	-0.339
Cl	0.218	0.22	0.175	0.781	-0.097
CNNH	-0.059	0.09	0.171	0.195	-1.235
F	-0.138	0.307	0.325	0.67	-1.054
GeH_3	0.273	0.273	0.145	0.629	NA
$m-C_6H_4CN$	0.04	0.058	0.07	0.212	-0.368
$m-C_6H_4NH_2$	0.042	0.06	0.072	0.233	-0.339
NNCH	-0.013	0.136	0.187	0.547	-0.36
$p-C_6H_4CN$	0.035	0.057	0.072	0.204	-0.403
$p-C_6H_4NH_2$	0.034	0.059	0.074	0.242	-0.329
SiH_3	0.232	0.232	0.141	0.587	-0.006

Table C14: Linear regression statistics relating *V* at the R-G bond to that in the R-H bond

Substrate	Slope	Intercept	r^2	Slope SE	IntSE
BP	0.465	-0.084	0.788	0.023	0.009
C_6H_5	0.525	-0.072	0.827	0.022	0.009
CCH	0.752	-0.048	0.909	0.022	0.009
CH_3	0.393	-0.088	0.798	0.018	0.007
$CHCCH_2$	0.52	-0.07	0.829	0.022	0.009
$CHCH_2$	0.552	-0.072	0.844	0.022	0.009
Cl	0.082	-0.087	0.118	0.021	0.009
CNNH	0.748	-0.04	0.766	0.039	0.015
F	-0.001	-0.268	0	0.075	0.031
GeH_3	0.004	-0.051	0.006	0.005	0.002
$m-C_6H_4CN$	0.542	-0.066	0.836	0.023	0.009
$m-C_6H_4NH_2$	0.52	-0.073	0.822	0.023	0.009
NNCH	0.241	-0.184	0.145	0.056	0.022
$p-C_6H_4CN$	0.553	-0.065	0.83	0.023	0.009
$p-C_6H_4NH_2$	0.53	-0.073	0.827	0.023	0.009
SiH_3	-0.014	-0.068	0.044	0.006	0.002

Table C15: Summary statistics showing error in *V* over all substrates compared to G=H

Substrate	MD	MAD	SD	Max	Min
BP	0.107	0.108	0.112	0.455	-0.048
C_6H_5	0.098	0.1	0.102	0.41	-0.078
CCH	0.041	0.048	0.066	0.271	-0.09
CH_3	0.129	0.129	0.122	0.47	-0.01
$CHCCH_2$	0.101	0.102	0.102	0.424	-0.058
$CHCH_2$	0.088	0.089	0.097	0.399	-0.052
Cl	0.241	0.241	0.183	0.754	NA
CNNH	0.048	0.064	0.089	0.355	-0.313
F	0.091	0.179	0.249	0.754	-0.324
GeH_3	0.305	0.305	0.191	0.737	-0.006
$m-C_6H_4CN$	0.097	0.099	0.099	0.398	-0.068
$m-C_6H_4NH_2$	0.099	0.1	0.103	0.412	-0.08
NNCH	0.082	0.123	0.179	0.624	-0.169
$p-C_6H_4CN$	0.094	0.095	0.098	0.395	-0.071
$p-C_6H_4NH_2$	0.095	0.097	0.101	0.419	-0.1
SiH_3	0.295	0.295	0.194	0.736	-0.007

Table C16: Linear regression statistics relating *V* at the R-G bond to that in the R-H bond

Substrate	Slope	Intercept	r^2	Slope SE	IntSE
BP	0.719	-0.057	0.812	0.032	0.015
C_6H_5	0.85	-0.021	0.857	0.032	0.015
CCH	1.332	0.075	0.92	0.037	0.017
CH_3	0.597	-0.078	0.862	0.022	0.01
$CHCCH_2$	0.838	-0.023	0.853	0.032	0.015
$CHCH_2$	0.922	-0.008	0.89	0.03	0.014
Cl	0.125	-0.146	0.164	0.027	0.012
CNNH	1.304	0.064	0.677	0.085	0.038
F	-0.254	-0.661	0.048	0.11	0.05
GeH_3	0.241	-0.042	0.756	0.013	0.006
$m-C_6H_4CN$	0.878	-0.011	0.861	0.033	0.015
$m-C_6H_4NH_2$	0.844	-0.023	0.851	0.033	0.015
NNCH	0.219	-0.332	0.098	0.063	0.028
$p-C_6H_4CN$	0.906	-0.004	0.856	0.035	0.016
$p-C_6H_4NH_2$	0.865	-0.022	0.844	0.035	0.016
SiH_3	0.271	-0.07	0.69	0.017	0.008

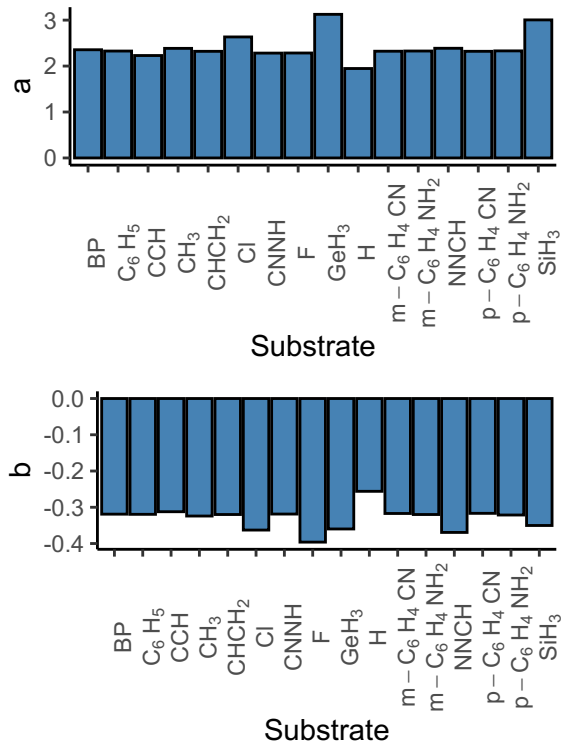

Figure C12: Electronegativity factor fit parameters (A) *a* and (B) *b* Fit parameters for 17 different substrates calculated at the B3LYP/def2-TZVPPD level of theory

Table C17: MD and MAD of χ_R(G) compared to χ_R(H)

Substrate	MD	MAD	Max	Min	SD
BP	0.358	0.359	0.766	-0.029	0.106
C_6H_5	0.337	0.337	0.729	0.032	0.100
CCH	0.246	0.246	0.614	0.035	0.090
CH_3	0.369	0.369	0.584	0.008	0.102
$CHCH_2$	0.351	0.351	0.719	0.050	0.100
Cl	0.231	0.280	0.752	-1.326	0.230
CNNH	0.308	0.308	1.004	0.039	0.151
F	0.124	0.173	0.444	-0.736	0.178
GeH_3	0.503	0.506	1.045	-0.086	0.245
m$-C_6H_4CN$	0.321	0.321	0.725	0.040	0.099
m$-C_6H_4NH_2$	0.338	0.338	0.738	0.032	0.101
NNCH	0.404	0.406	0.899	-0.146	0.137
p$-C_6H_4CN$	0.324	0.324	0.737	0.042	0.101
p$-C_6H_4NH_2$	0.355	0.355	0.748	0.028	0.100
SiH_3	0.549	0.551	1.157	-0.059	0.273

Table C18: Linear regression parameters comparing χ_R(G) to χ_R(H)

Substrate	Slope	Intercept	r^2	Slope SE	Intercept SE
BP	0.976	0.423	0.953	0.02	0.056
C_6H_5	1.013	0.302	0.961	0.019	0.053
CCH	1.103	-0.035	0.981	0.014	0.04
CH_3	0.956	0.489	0.956	0.019	0.053
$CHCH_2$	1.032	0.263	0.963	0.019	0.052
Cl	0.784	0.818	0.785	0.039	0.108
CNNH	1.16	-0.125	0.947	0.026	0.071
F	0.793	0.687	0.884	0.028	0.077
GeH_3	0.656	1.439	0.765	0.034	0.094
m$-C_6H_4CN$	1.019	0.269	0.963	0.019	0.052
m$-C_6H_4NH_2$	1.011	0.308	0.96	0.019	0.053
NNCH	1.032	0.318	0.931	0.027	0.074
p$-C_6H_4CN$	1.024	0.258	0.961	0.019	0.053
p$-C_6H_4NH_2$	1.023	0.293	0.962	0.019	0.053
SiH_3	0.626	1.567	0.694	0.039	0.107

Table C19: Linear relationships between $\Delta\rho$(G - H) and $\Delta\chi$(G - H), part 1/4

Substrate	attached	r^2	Slope	Intercept
BP	C	0.286	0.064	-0.216
CH_3	C	0.341	0.056	-0.205
C_6H_5	C	0.431	0.095	-0.307
CCH	C	0.606	0.137	-0.395
$CHCH_2$	C	0.354	0.085	-0.273
BP	N	0.022	0.029	-0.164
CH_3	N	0.003	-0.009	-0.047
C_6H_5	N	0.100	0.056	-0.252
CCH	N	0.068	0.040	-0.150
$CHCH_2$	N	0.079	0.058	-0.254
BP	O	0.971	0.125	-0.581
CH_3	O	0.877	0.130	-0.612
C_6H_5	O	0.917	0.142	-0.640
CCH	O	0.904	0.214	-0.868
$CHCH_2$	O	0.925	0.139	-0.625
BP	P	0.542	0.041	-0.109
CH_3	P	0.709	0.055	-0.153
C_6H_5	P	0.252	0.018	-0.050
CCH	P	0.966	-0.089	0.211
$CHCH_2$	P	0.224	0.018	-0.049
BP	S	0.538	0.081	-0.284
CH_3	S	0.719	0.090	-0.320
C_6H_5	S	0.189	0.034	-0.132
CCH	S	0.605	-0.072	0.209
$CHCH_2$	S	0.004	0.006	-0.038
BP	Si	0.335	-0.029	0.068
CH_3	Si	0.159	-0.017	0.039
C_6H_5	Si	0.325	-0.030	0.068
CCH	Si	0.439	-0.028	0.056
$CHCH_2$	Si	0.408	-0.030	0.069

Table C20: Linear relationships between $\Delta\rho$(G - H) and $\Delta\chi$(G - H), part 2/4

Substrate	attached	r^2	Slope	Intercept
F	C	0.904	0.218	-0.626
Cl	C	0.836	0.144	-0.521
GeH_3	C	0.251	0.072	-0.398
SiH_3	C	0.339	0.089	-0.459
F	N	0.957	0.215	-0.777
Cl	N	0.944	0.120	-0.550
GeH_3	N	0.241	0.041	-0.361
SiH_3	N	0.187	0.045	-0.378
F	O	0.990	0.179	-0.744
Cl	O	0.985	0.105	-0.551
GeH_3	O	0.363	0.025	-0.337
SiH_3	O	0.143	0.017	-0.306
F	P	0.286	0.016	-0.049
Cl	P	0.567	-0.058	0.108
GeH_3	P	0.949	-0.112	0.283
SiH_3	P	0.931	-0.224	0.678
F	S	0.802	0.092	-0.305
Cl	S	0.321	0.054	-0.249
GeH_3	S	0.120	0.056	-0.321
SiH_3	S	0.128	0.066	-0.363
F	Si	0.013	-0.005	0.026
Cl	Si	0.188	-0.009	-0.004
GeH_3	Si	0.425	-0.018	0.014
SiH_3	Si	0.035	-0.012	0.005

Table C21: Linear relationships between $\Delta\rho$(G - H) and $\Delta\chi$(G - H), part 3/4

Substrate	attached	r^2	Slope	Intercept
m−C_6H_4CN	C	0.418	0.096	-0.307
m−$C_6H_4NH_2$	C	0.429	0.095	-0.307
p−C_6H_4CN	C	0.421	0.098	-0.314
p−$C_6H_4NH_2$	C	0.518	0.108	-0.344
m−C_6H_4CN	N	0.224	0.077	-0.321
m−$C_6H_4NH_2$	N	0.113	0.059	-0.261
p−C_6H_4CN	N	0.293	0.085	-0.349
p−$C_6H_4NH_2$	N	0.000	0.003	-0.061
m−C_6H_4CN	O	0.902	0.138	-0.620
m−$C_6H_4NH_2$	O	0.916	0.142	-0.639
p−C_6H_4CN	O	0.901	0.134	-0.604
p−$C_6H_4NH_2$	O	0.958	0.181	-0.793
m−C_6H_4CN	P	0.043	0.006	-0.022
m−$C_6H_4NH_2$	P	0.013	0.006	-0.021
p−C_6H_4CN	P	0.080	-0.007	0.013
p−$C_6H_4NH_2$	P	0.226	0.022	-0.060
m−C_6H_4CN	S	0.044	0.017	-0.079
m−$C_6H_4NH_2$	S	0.182	0.035	-0.134
p−C_6H_4CN	S	0.022	0.014	-0.068
p−$C_6H_4NH_2$	S	0.428	0.059	-0.208
m−C_6H_4CN	Si	0.309	-0.025	0.054
m−$C_6H_4NH_2$	Si	0.311	-0.030	0.068
p−C_6H_4CN	Si	0.316	-0.025	0.054
p−$C_6H_4NH_2$	Si	0.292	-0.033	0.077

Table C22: Linear relationships between $\Delta\rho$(G - H) and $\Delta\chi$(G - H), part 4/4

Substrate	attached	r^2	Slope	Intercept
CNNH	C	0.660	0.177	-0.525
NNCH	C	0.801	0.205	-0.625
CNNH	N	0.791	0.210	-0.787
NNCH	N	0.218	0.065	-0.267
CNNH	O	0.973	0.232	-0.968
NNCH	O	0.962	0.159	-0.698
CNNH	P	0.723	-0.135	0.326
NNCH	P	0.680	0.038	-0.109
CNNH	S	0.306	-0.146	0.432
NNCH	S	0.943	0.139	-0.465
CNNH	Si	0.123	-0.023	0.042
NNCH	Si	0.307	0.018	-0.034

Table C23: Linear relationships between $\Delta\nabla^2\rho$(G - H) and $\Delta\chi$(G - H), part 1/4

Substrate	attached	r^2	Slope	Intercept
BP	C	0.146	0.249	-0.388
CH_3	C	0.335	0.329	-0.573
C_6H_5	C	0.009	0.065	0.174
CCH	C	0.009	-0.060	0.440
$CHCH_2$	C	0.031	0.128	-0.039
BP	N	0.183	1.670	-4.868
CH_3	N	0.423	2.382	-7.246
C_6H_5	N	0.039	0.714	-1.525
CCH	N	0.565	3.699	-12.303
$CHCH_2$	N	0.013	0.484	-0.758
BP	O	0.349	-0.418	3.776
CH_3	O	0.316	-0.468	4.040
C_6H_5	O	0.247	-0.385	3.703
CCH	O	0.400	1.092	-1.859
$CHCH_2$	O	0.180	-0.336	3.579
BP	P	0.651	0.382	-1.070
CH_3	P	0.764	0.525	-1.484
C_6H_5	P	0.697	0.354	-0.953
CCH	P	0.755	0.589	-1.238
$CHCH_2$	P	0.714	0.406	-1.069
BP	S	0.075	-0.131	0.769
CH_3	S	0.183	-0.155	0.874
C_6H_5	S	0.128	0.173	-0.206
CCH	S	0.164	-0.225	0.853
$CHCH_2$	S	0.226	0.327	-0.722
BP	Si	0.398	-0.460	1.059
CH_3	Si	0.181	-0.232	0.537
C_6H_5	Si	0.405	-0.581	1.323
CCH	Si	0.175	-0.215	0.612
$CHCH_2$	Si	0.392	-0.489	1.128

Table C24: Linear relationships between $\Delta\nabla^2\rho$(G - H) and $\Delta\chi$(G - H), part 2/4

Substrate	attached	r^2	Slope	Intercept
F	C	0.286	0.998	-1.910
Cl	C	0.133	-0.183	1.360
GeH_3	C	0.034	-0.291	2.115
SiH_3	C	0.168	-0.633	3.302
F	N	0.024	-0.093	2.135
Cl	N	0.057	0.108	1.445
GeH_3	N	0.091	0.846	-0.745
SiH_3	N	0.001	0.127	1.903
F	O	0.764	-0.314	4.264
Cl	O	0.536	-0.151	3.282
GeH_3	O	0.804	0.796	0.345
SiH_3	O	0.789	1.080	-0.556
F	P	0.992	2.299	-5.074
Cl	P	0.914	0.974	-2.590
GeH_3	P	0.999	0.963	-3.018
SiH_3	P	0.984	1.935	-6.440
F	S	0.403	1.099	-2.449
Cl	S	0.012	0.070	0.409
GeH_3	S	0.047	0.236	-0.084
SiH_3	S	0.020	0.177	0.132
F	Si	0.856	1.255	-1.978
Cl	Si	0.446	0.218	-0.426
GeH_3	Si	0.471	-0.318	0.612
SiH_3	Si	0.023	-0.094	0.005

Table C25: Linear relationships between $\Delta\nabla^2\rho$(G - H) and $\Delta\chi$(G - H), part 3/4

Substrate	attached	r^2	Slope	Intercept
m−C_6H_4CN	C	0.009	0.066	0.173
m−$C_6H_4NH_2$	C	0.007	0.058	0.195
p−C_6H_4CN	C	0.005	0.051	0.214
p−$C_6H_4NH_2$	C	0.000	-0.001	0.364
m−C_6H_4CN	N	0.004	0.181	0.362
m−$C_6H_4NH_2$	N	0.028	0.594	-1.094
p−C_6H_4CN	N	0.001	-0.094	1.330
p−$C_6H_4NH_2$	N	0.282	2.780	-8.927
m−C_6H_4CN	O	0.263	-0.421	3.823
m−$C_6H_4NH_2$	O	0.212	-0.353	3.588
p−C_6H_4CN	O	0.184	-0.373	3.652
p−$C_6H_4NH_2$	O	0.030	0.117	1.796
m−C_6H_4CN	P	0.696	0.321	-0.853
m−$C_6H_4NH_2$	P	0.964	0.476	-1.271
p−C_6H_4CN	P	0.630	0.293	-0.782
p−$C_6H_4NH_2$	P	0.808	0.402	-1.069
m−C_6H_4CN	S	0.200	0.265	-0.490
m−$C_6H_4NH_2$	S	0.120	0.173	-0.206
p−C_6H_4CN	S	0.172	0.283	-0.554
p−$C_6H_4NH_2$	S	0.003	0.022	0.256
m−C_6H_4CN	Si	0.411	-0.477	1.082
m−$C_6H_4NH_2$	Si	0.398	-0.585	1.332
p−C_6H_4CN	Si	0.423	-0.481	1.090
p−$C_6H_4NH_2$	Si	0.360	-0.701	1.616

Table C26: Linear relationships between $\Delta\nabla^2\rho$(G - H) and $\Delta\chi$(G - H), part 4/4

Substrate	attached	r^2	Slope	Intercept
CNNH	C	0.104	-0.244	1.041
NNCH	C	0.620	-0.903	3.040
CNNH	N	0.223	1.607	-4.928
NNCH	N	0.014	-0.119	1.720
CNNH	O	0.831	1.517	-3.602
NNCH	O	0.838	-0.652	5.105
CNNH	P	0.100	0.152	-0.180
NNCH	P	0.974	1.610	-3.934
CNNH	S	0.090	0.451	-1.107
NNCH	S	0.846	-0.687	2.545
CNNH	Si	0.112	-0.302	0.847
NNCH	Si	0.199	0.407	-0.490

Table C27: Linear relationships between ΔG(G - H) and $\Delta\chi$(G - H), part 1/4

Substrate	attached	r^2	Slope	Intercept
BP	C	0.807	0.259	-0.757
CH_3	C	0.811	0.246	-0.723
C_6H_5	C	0.840	0.264	-0.760
CCH	C	0.854	0.234	-0.626
$CHCH_2$	C	0.834	0.254	-0.730
BP	N	0.538	0.528	-1.790
CH_3	N	0.658	0.543	-1.847
C_6H_5	N	0.483	0.453	-1.507
CCH	N	0.814	1.131	-3.837
$CHCH_2$	N	0.417	0.450	-1.496
BP	O	0.759	0.267	-0.858
CH_3	O	0.434	0.216	-0.687
C_6H_5	O	0.754	0.287	-0.899
CCH	O	0.843	0.801	-2.729
$CHCH_2$	O	0.791	0.297	-0.915
BP	P	0.606	0.147	-0.406
CH_3	P	0.741	0.205	-0.575
C_6H_5	P	0.534	0.102	-0.282
CCH	P	0.021	-0.008	0.051
$CHCH_2$	P	0.543	0.114	-0.306
BP	S	0.817	0.129	-0.401
CH_3	S	0.866	0.133	-0.421
C_6H_5	S	0.731	0.134	-0.411
CCH	S	0.634	-0.371	1.145
$CHCH_2$	S	0.703	0.135	-0.415
BP	Si	0.376	-0.120	0.278
CH_3	Si	0.177	-0.064	0.148
C_6H_5	Si	0.384	-0.145	0.329
CCH	Si	0.284	-0.076	0.187
$CHCH_2$	Si	0.394	-0.128	0.293

Table C28: Linear relationships between ΔG(G - H) and $\Delta\chi$(G - H), part 2/4

Substrate	attached	r^2	Slope	Intercept
F	C	0.711	0.695	-1.616
Cl	C	0.875	0.130	-0.351
GeH_3	C	0.013	0.022	-0.027
SiH_3	C	0.001	0.005	0.057
F	N	0.805	0.162	-0.382
Cl	N	0.620	0.067	-0.181
GeH_3	N	0.419	0.159	-0.474
SiH_3	N	0.363	0.177	-0.499
F	O	0.939	0.120	-0.266
Cl	O	0.934	0.059	-0.166
GeH_3	O	0.862	0.145	-0.436
SiH_3	O	0.814	0.198	-0.592
F	P	0.974	0.446	-0.995
Cl	P	0.480	0.152	-0.466
GeH_3	P	0.068	0.009	-0.132
SiH_3	P	0.722	0.072	-0.329
F	S	0.856	0.526	-1.378
Cl	S	0.092	0.044	-0.129
GeH_3	S	0.013	-0.016	0.059
SiH_3	S	0.087	0.037	-0.109
F	Si	0.713	0.229	-0.340
Cl	Si	0.096	0.022	-0.060
GeH_3	Si	0.822	-0.101	0.175
SiH_3	Si	0.000	0.002	-0.097

Table C29: Linear relationships between ΔG(G - H) and $\Delta\chi$(G - H), part 3/4

Substrate	attached	r^2	Slope	Intercept
m−C_6H_4CN	C	0.834	0.255	-0.729
m−$C_6H_4NH_2$	C	0.837	0.265	-0.763
p−C_6H_4CN	C	0.836	0.257	-0.735
p−$C_6H_4NH_2$	C	0.850	0.268	-0.776
m−C_6H_4CN	N	0.517	0.391	-1.284
m−$C_6H_4NH_2$	N	0.478	0.441	-1.465
p−C_6H_4CN	N	0.548	0.382	-1.247
p−$C_6H_4NH_2$	N	0.606	0.756	-2.598
m−C_6H_4CN	O	0.774	0.280	-0.867
m−$C_6H_4NH_2$	O	0.764	0.295	-0.930
p−C_6H_4CN	O	0.768	0.289	-0.895
p−$C_6H_4NH_2$	O	0.830	0.484	-1.657
m−C_6H_4CN	P	0.433	0.077	-0.216
m−$C_6H_4NH_2$	P	0.575	0.114	-0.313
p−C_6H_4CN	P	0.263	0.052	-0.150
p−$C_6H_4NH_2$	P	0.670	0.118	-0.320
m−C_6H_4CN	S	0.667	0.125	-0.380
m−$C_6H_4NH_2$	S	0.726	0.134	-0.411
p−C_6H_4CN	S	0.648	0.123	-0.375
p−$C_6H_4NH_2$	S	0.855	0.136	-0.419
m−C_6H_4CN	Si	0.392	-0.121	0.272
m−$C_6H_4NH_2$	Si	0.375	-0.145	0.330
p−C_6H_4CN	Si	0.402	-0.122	0.273
p−$C_6H_4NH_2$	Si	0.337	-0.169	0.389

Table C30: Linear relationships between ΔG(G - H) and $\Delta\chi$(G - H), part 4/4

Substrate	attached	r^2	Slope	Intercept
CNNH	C	0.822	0.249	-0.678
NNCH	C	0.540	0.281	-0.757
CNNH	N	0.715	1.141	-3.979
NNCH	N	0.817	0.158	-0.480
CNNH	O	0.965	0.934	-3.343
NNCH	O	0.839	0.093	-0.254
CNNH	P	0.471	-0.161	0.426
NNCH	P	0.947	0.360	-0.902
CNNH	S	0.374	-0.439	1.374
NNCH	S	0.923	0.338	-1.020
CNNH	Si	0.133	-0.088	0.219
NNCH	Si	0.213	0.095	-0.130

Table C31: Linear relationships between ΔG(V - H) and $\Delta\chi$(G - H), part 1/4

Substrate	attached	r^2	Slope	Intercept
BP	C	0.718	-0.456	1.417
CH_3	C	0.739	-0.409	1.302
C_6H_5	C	0.761	-0.512	1.564
CCH	C	0.717	-0.482	1.361
$CHCH_2$	C	0.744	-0.476	1.451
BP	N	0.598	-0.639	2.364
CH_3	N	0.468	-0.490	1.883
C_6H_5	N	0.746	-0.727	2.632
CCH	N	0.838	-1.336	4.599
$CHCH_2$	N	0.763	-0.778	2.803
BP	O	0.922	-0.638	2.661
CH_3	O	0.691	-0.550	2.384
C_6H_5	O	0.908	-0.670	2.724
CCH	O	0.917	-1.328	4.992
$CHCH_2$	O	0.928	-0.677	2.725
BP	P	0.581	-0.198	0.544
CH_3	P	0.728	-0.279	0.779
C_6H_5	P	0.424	-0.116	0.326
CCH	P	0.981	0.163	-0.411
$CHCH_2$	P	0.428	-0.127	0.345
BP	S	0.744	-0.291	0.994
CH_3	S	0.803	-0.306	1.060
C_6H_5	S	0.627	-0.224	0.770
CCH	S	0.654	0.685	-2.078
$CHCH_2$	S	0.454	-0.188	0.648
BP	Si	0.355	0.126	-0.291
CH_3	Si	0.173	0.070	-0.162
C_6H_5	Si	0.362	0.145	-0.327
CCH	Si	0.372	0.098	-0.220
$CHCH_2$	Si	0.393	0.133	-0.304

Table C32: Linear relationships between ΔG(V - H) and $\Delta\chi$(G - H), part 2/4

Substrate	attached	r^2	Slope	Intercept
F	C	0.838	-1.140	2.754
Cl	C	0.851	-0.305	1.042
GeH_3	C	0.529	-0.117	0.582
SiH_3	C	0.686	-0.169	0.712
F	N	0.795	-0.347	1.299
Cl	N	0.531	-0.107	0.724
GeH_3	N	0.089	-0.106	0.762
SiH_3	N	0.340	-0.321	1.473
F	O	0.970	-0.319	1.599
Cl	O	0.899	-0.156	1.154
GeH_3	O	0.541	-0.091	0.958
SiH_3	O	0.579	-0.127	1.046
F	P	0.918	-0.317	0.722
Cl	P	0.059	-0.061	0.284
GeH_3	P	0.900	0.223	-0.490
SiH_3	P	0.835	0.339	-0.952
F	S	0.978	-0.777	2.144
Cl	S	0.194	-0.071	0.361
GeH_3	S	0.323	0.091	-0.139
SiH_3	S	0.024	-0.031	0.250
F	Si	0.442	-0.144	0.186
Cl	Si	0.020	0.011	0.012
GeH_3	Si	0.983	0.123	-0.197
SiH_3	Si	0.013	-0.028	0.194

Table C33: Linear relationships between ΔV(G - H) and $\Delta\chi$(G - H), part 3/4

Substrate	attached	r^2	Slope	Intercept
m−C_6H_4CN	C	0.743	-0.493	1.501
m−$C_6H_4NH_2$	C	0.761	-0.515	1.574
p−C_6H_4CN	C	0.747	-0.502	1.524
p−$C_6H_4NH_2$	C	0.786	-0.537	1.643
m−C_6H_4CN	N	0.808	-0.738	2.659
m−$C_6H_4NH_2$	N	0.758	-0.733	2.656
p−C_6H_4CN	N	0.850	-0.787	2.826
p−$C_6H_4NH_2$	N	0.623	-0.817	2.964
m−C_6H_4CN	O	0.925	-0.665	2.690
m−$C_6H_4NH_2$	O	0.913	-0.678	2.758
p−C_6H_4CN	O	0.931	-0.671	2.704
p−$C_6H_4NH_2$	O	0.913	-0.938	3.764
m−C_6H_4CN	P	0.257	-0.074	0.219
m−$C_6H_4NH_2$	P	0.281	-0.109	0.308
p−C_6H_4CN	P	0.063	-0.031	0.105
p−$C_6H_4NH_2$	P	0.509	-0.136	0.374
m−C_6H_4CN	S	0.531	-0.184	0.638
m−$C_6H_4NH_2$	S	0.620	-0.224	0.770
p−C_6H_4CN	S	0.480	-0.176	0.612
p−$C_6H_4NH_2$	S	0.735	-0.266	0.901
m−C_6H_4CN	Si	0.371	0.123	-0.273
m−$C_6H_4NH_2$	Si	0.350	0.144	-0.327
p−C_6H_4CN	Si	0.380	0.124	-0.273
p−$C_6H_4NH_2$	Si	0.313	0.162	-0.375

Table C34: Linear relationships between ΔV(G - H) and $\Delta\chi$(G - H), part 4/4

Substrate	attached	r^2	Slope	Intercept
CNNH	C	0.688	-0.560	1.616
NNCH	C	0.576	-0.787	2.274
CNNH	N	0.839	-1.881	6.727
NNCH	N	0.573	-0.346	1.389
CNNH	O	0.980	-1.489	5.785
NNCH	O	0.920	-0.349	1.784
CNNH	P	0.672	0.360	-0.897
NNCH	P	0.898	-0.318	0.820
CNNH	S	0.429	0.991	-3.026
NNCH	S	0.927	-0.847	2.676
CNNH	Si	0.149	0.100	-0.226
NNCH	Si	0.228	-0.089	0.137

Appendix D

Appendix D

Table D1: Substituent descriptors data used as the **X** matrix in multivariate modeling, part 1/6

Substituent	q(R)	μ_x^p(R)	μ_x^c(R)	μ_x(R)	Q_{xx}(R)	$\|\|\boldsymbol{\mu}^p\|\|$(R)	$\|\|\boldsymbol{\mu}^c\|\|$(R)
C_6H_5	-0.005	-0.136	0.004	-0.132	3.150	0.136	0.004
CCH	-0.147	-0.121	0.103	-0.019	4.552	0.121	0.103
CCl_2H	-0.077	0.104	-0.541	-0.437	0.985	0.339	1.034
CCl_3	-0.118	0.249	-0.717	-0.467	0.734	0.249	0.717
$CClF_2$	-0.098	0.899	-1.528	-0.628	1.537	1.280	1.816
$CClH_2$	-0.032	-0.034	-0.315	-0.349	0.899	0.272	1.068
CF_2H	-0.043	0.623	-1.147	-0.524	1.683	1.292	2.133
CF_3	-0.089	1.216	-1.926	-0.710	1.860	1.216	1.926
CFH_2	-0.009	0.210	-0.574	-0.363	1.122	1.019	1.807
$CH(CH_2)_2$	0.001	-0.139	-0.001	-0.141	1.336	0.140	0.002
$CH(CH_3)_2$	0.041	-0.142	-0.056	-0.197	0.888	0.142	0.064
$CH_2CH_2CH_3$-ip	0.031	-0.141	-0.046	-0.187	0.030	0.141	0.056
CH_2CH_3	0.032	-0.142	-0.024	-0.166	0.667	0.142	0.024
CH_2CN	-0.045	0.559	-1.190	-0.631	3.192	1.953	3.560
CH_2NH_2-ip	0.017	-0.082	0.126	0.043	2.339	0.316	0.745
CH_2OH-ip	-0.003	-0.293	0.562	0.269	3.818	0.871	1.474
CH_3	0.018	-0.139	-0.014	-0.152	0.411	0.139	0.014
$CHCH_2$	-0.006	-0.134	0.005	-0.130	1.499	0.135	0.005
CN	-0.199	1.632	-2.803	-1.171	-1.370	1.632	2.803
$COCH_3$	-0.017	0.774	-1.196	-0.422	1.113	1.608	2.765
COCl	-0.100	1.150	-1.915	-0.765	0.576	1.561	2.273
COF	-0.098	1.424	-2.291	-0.867	0.704	1.445	2.329
COH	-0.012	0.721	-1.348	-0.627	0.538	1.544	2.568
$CONH_2$	-0.012	1.023	-1.543	-0.520	1.583	1.195	2.640
$COOCH_3$	-0.055	0.371	-0.394	-0.023	3.650	0.509	1.173
COOH	-0.064	0.829	-1.024	-0.195	3.572	0.864	1.323

Table D2: Substituent descriptors data used as the **X** matrix in multivariate modeling, part 2/6

Substituent	q(R)	μ_x^p(R)	μ_x^c(R)	μ_x(R)	Q_{xx}(R)	$\|\|\boldsymbol{\mu}^p\|\|$(R)	$\|\|\boldsymbol{\mu}^c\|\|$(R)
$N(CH_3)_2$	-0.276	-0.262	0.172	-0.091	0.703	0.324	0.757
NC	-0.540	-2.525	1.386	-1.138	-2.179	2.525	1.386
NCO	-0.482	0.544	-1.289	-0.745	0.182	0.565	1.316
NCS	-0.494	-2.360	1.693	-0.667	0.642	3.449	2.977
NH_2	-0.358	-0.136	-0.100	-0.236	0.016	0.189	0.714
$NHCH_3$	-0.347	-0.264	0.058	-0.206	0.449	0.405	0.743
$NHNH_2$−IP	-0.342	-0.125	-0.584	-0.709	-2.112	0.130	0.695
NHOH	-0.366	-0.253	0.163	-0.089	1.028	0.593	0.737
NNN	-0.432	-0.131	-0.535	-0.667	1.560	0.227	0.547
NO	-0.299	0.045	-0.736	-0.691	0.449	0.736	1.162
NO_2	-0.427	0.561	-1.545	-0.985	0.573	0.561	1.545
OCH_3	-0.569	-0.042	-0.301	-0.343	0.377	0.836	1.418
OCN	-0.641	0.476	-1.521	-1.045	2.630	0.478	1.866
$OCOCH_3$	-0.631	-1.267	1.822	0.556	3.538	1.295	1.944
OH	-0.580	0.167	-0.568	-0.401	-0.078	0.459	1.166
PH_2	0.540	-1.103	0.521	-0.581	2.175	2.427	2.001
$PO(OH)_2$	0.586	0.099	-0.360	-0.261	5.268	0.360	0.413
SCF_3	0.020	-0.242	-0.187	-0.429	2.984	2.010	2.483
SCH_3	0.076	-0.472	0.146	-0.325	1.754	0.663	0.188
SCN	-0.012	-0.016	-0.536	-0.552	6.300	2.808	4.126
SH	0.050	-0.442	0.072	-0.369	1.787	0.610	0.144
SiH_3	0.702	0.403	-1.001	-0.599	2.836	0.403	1.001

Table D3: Substituent descriptors data used as the **X** matrix in multivariate modeling, part 3/6

Substituent	$\|\|\boldsymbol{\mu}\|\|$(R)	$\|\|\mathbf{Q}\|\|$(R)	DI(R,H)	ρ_c	$\nabla^2\rho_c$	ε_c	V_c
C_6H_5	0.132	5.792	1.018	0.296	-1.219	0.015	-0.400
CCH	0.019	4.552	0.972	0.302	-1.336	0.000	-0.420
CCl_2H	0.709	1.186	0.992	0.306	-1.319	0.016	-0.411
CCl_3	0.467	0.734	0.980	0.312	-1.387	0.000	-0.422
$CClF_2$	0.632	1.738	0.968	0.317	-1.429	0.014	-0.420
$CClH_2$	0.827	1.809	1.007	0.297	-1.230	0.028	-0.399
CF_2H	0.848	1.837	0.979	0.311	-1.364	0.024	-0.408
CF_3	0.710	1.860	0.962	0.319	-1.450	0.000	-0.420
CFH_2	0.804	2.183	1.001	0.300	-1.253	0.052	-0.398
$CH(CH_2)_2$	0.141	1.345	1.026	0.295	-1.197	0.023	-0.403
$CH(CH_3)_2$	0.201	0.908	1.032	0.288	-1.125	0.006	-0.387
$CH_2CH_2CH_3$-ip	0.188	0.344	1.029	0.287	-1.124	0.007	-0.387
CH_2CH_3	0.166	0.795	1.030	0.287	-1.126	0.008	-0.387
CH_2CN	1.613	7.919	1.002	0.290	-1.168	0.013	-0.388
CH_2NH_2-ip	0.492	2.519	1.019	0.292	-1.176	0.032	-0.392
CH_2OH-ip	0.605	4.112	1.007	0.298	-1.233	0.042	-0.398
CH_3	0.152	0.411	1.029	0.287	-1.129	0.000	-0.389
$CHCH_2$	0.130	2.571	1.020	0.295	-1.213	0.011	-0.399
CN	1.171	1.370	0.937	0.301	-1.356	0.000	-0.413
$COCH_3$	1.163	1.944	1.988	0.288	-1.164	0.015	-0.367
COCl	0.784	1.866	0.968	0.299	-1.287	0.023	-0.391
COF	0.885	2.600	0.962	0.303	-1.324	0.037	-0.396
COH	1.032	1.587	0.987	0.291	-1.194	0.016	-0.373
$CONH_2$	1.611	2.071	0.993	0.293	-1.206	0.028	-0.376
$COOCH_3$	0.756	4.144	0.978	0.300	-1.279	0.037	-0.389
COOH	0.626	4.286	0.975	0.301	-1.294	0.034	-0.391

Table D4: Substituent descriptors data used as the **X** matrix in multivariate modeling, part 4/6

Substituent	$\|\|\boldsymbol{\mu}\|\|$(R)	$\|\|\mathbf{Q}\|\|$(R)	DI(R,H)	ρ_c	$\nabla^2\rho_c$	ε_c	V_c
$N(CH_3)_2$	0.559	1.910	2.820	0.357	-1.900	0.046	-0.618
NC	1.138	2.179	0.682	0.350	-2.234	0.000	-0.665
NCO	0.754	1.827	0.742	0.344	-2.007	0.006	-0.628
NCS	0.670	0.703	0.728	0.344	-2.040	0.003	-0.632
NH_2	0.622	0.646	0.856	0.351	-1.875	0.035	-0.608
$NHCH_3$	0.547	0.467	0.863	0.354	-1.886	0.042	-0.613
$NHNH_2$−IP	0.790	2.181	0.865	0.357	-1.909	0.052	-0.614
NHOH	0.249	2.227	0.841	0.360	-1.986	0.051	-0.622
NNN	0.671	2.681	0.786	0.342	-1.895	0.028	-0.601
NO	0.710	1.778	0.870	0.335	-1.701	0.001	-0.531
NO_2	0.985	2.988	0.777	0.345	-1.987	0.036	-0.588
OCH_3	0.649	0.888	0.651	0.384	-2.858	0.024	-0.897
OCN	1.474	6.388	0.563	0.369	-3.017	0.017	-0.909
$OCOCH_3$	0.690	5.421	1.617	0.373	-2.902	0.015	-0.887
OH	0.715	1.514	0.640	0.380	-2.867	0.021	-0.896
PH_2	0.625	2.176	1.046	0.169	-0.162	0.090	-0.306
$PO(OH)_2$	0.607	9.512	0.910	0.195	-0.319	0.001	-0.356
SCF_3	0.644	3.008	1.055	0.225	-0.707	0.110	-0.287
SCH_3	0.668	3.036	1.079	0.224	-0.698	0.130	-0.294
SCN	1.397	9.064	1.039	0.225	-0.716	0.115	-0.281
SH	0.474	3.009	1.074	0.223	-0.695	0.116	-0.291
SiH_3	0.599	2.836	0.803	0.126	0.145	0.000	-0.213

Table D5: Substituent descriptors data used as the **X** matrix in multivariate modeling, part 5/6

Substituent	G_c	$\lambda_{1,c}$	$\lambda_{2,c}$	$\lambda_{3,c}$	H_c
C_6H_5	0.048	-0.841	-0.829	0.452	-0.352
CCH	0.043	-0.896	-0.896	0.455	-0.377
CCl_2H	0.041	-0.915	-0.900	0.496	-0.370
CCl_3	0.037	-0.952	-0.952	0.518	-0.384
$CClF_2$	0.031	-0.988	-0.974	0.532	-0.389
$CClH_2$	0.046	-0.861	-0.838	0.469	-0.353
CF_2H	0.034	-0.949	-0.927	0.512	-0.375
CF_3	0.029	-0.994	-0.994	0.538	-0.391
CFH_2	0.042	-0.887	-0.842	0.476	-0.356
$CH(CH_2)_2$	0.052	-0.832	-0.813	0.448	-0.351
$CH(CH_3)_2$	0.053	-0.786	-0.782	0.443	-0.334
$CH_2CH_2CH_3$-ip	0.053	-0.785	-0.779	0.440	-0.334
CH_2CH_3	0.053	-0.786	-0.780	0.441	-0.334
CH_2CN	0.048	-0.816	-0.806	0.454	-0.340
CH_2COO^--ip	0.053	-0.779	-0.776	0.437	-0.333
CH_2NH_2-ip	0.049	-0.828	-0.803	0.455	-0.343
CH_2OH-ip	0.045	-0.868	-0.833	0.468	-0.353
CH_3	0.053	-0.783	-0.783	0.437	-0.335
$CHCH_2$	0.048	-0.836	-0.827	0.450	-0.351
CN	0.037	-0.909	-0.909	0.463	-0.376
$COCH_3$	0.038	-0.829	-0.816	0.480	-0.329
COCl	0.035	-0.902	-0.883	0.498	-0.356
COF	0.032	-0.930	-0.896	0.503	-0.363
COH	0.037	-0.844	-0.831	0.481	-0.336
$CONH_2$	0.037	-0.856	-0.833	0.484	-0.339
$COOCH_3$	0.034	-0.903	-0.871	0.494	-0.354
COOH	0.034	-0.910	-0.880	0.497	-0.357

Table D6: Substituent descriptors data used as the **X** matrix in multivariate modeling, part 6/6

Substituent	G_c	$\lambda_{1,c}$	$\lambda_{2,c}$	$\lambda_{3,c}$	H_c
$N(CH_3)_2$	0.072	-1.419	-1.357	0.876	-0.547
NC	0.053	-1.564	-1.564	0.894	-0.612
NCO	0.063	-1.438	-1.429	0.860	-0.565
NCS	0.061	-1.454	-1.450	0.864	-0.571
NH_2	0.070	-1.392	-1.345	0.863	-0.538
$NHCH_3$	0.071	-1.406	-1.350	0.870	-0.542
$NHNH_2-$IP	0.069	-1.431	-1.360	0.882	-0.546
NHOH	0.063	-1.478	-1.407	0.899	-0.559
NNN	0.064	-1.393	-1.356	0.854	-0.538
NO	0.053	-1.283	-1.283	0.865	-0.478
NO_2	0.046	-1.472	-1.421	0.906	-0.543
OCH_3	0.091	-2.045	-1.997	1.185	-0.806
OCN	0.077	-2.070	-2.035	1.089	-0.832
$OCOCH_3$	0.081	-2.031	-2.002	1.131	-0.806
OH	0.089	-2.037	-1.995	1.166	-0.806
PH_2	0.133	-0.279	-0.256	0.374	-0.173
$PO(OH)_2$	0.138	-0.376	-0.376	0.433	-0.218
SCF_3	0.055	-0.470	-0.423	0.186	-0.232
SCH_3	0.060	-0.460	-0.407	0.169	-0.234
SCN	0.051	-0.485	-0.435	0.204	-0.230
SH	0.058	-0.458	-0.410	0.173	-0.232
SiH_3	0.125	-0.214	-0.214	0.573	-0.088

Table D7: Proxy data used as the **Y** matrix in multivariate modeling, part 1/2

Substituent	σ_m	σ_p	F	R
C_6H_5	0.06	-0.01	0.12	-0.13
CCH	0.21	0.23	0.22	0.01
CCl_2H	0.31	0.32	0.31	0.01
CCl_3	0.4	0.46	0.38	0.09
$CClF_2$	0.42	0.46	0.4	0.06
$CClH_2$	0.11	0.12	0.13	-0.01
CF_2H	0.29	0.32	0.29	0.03
CF_3	0.43	0.54	0.38	0.16
CFH_2	0.12	0.11	0.15	-0.04
$CH(CH_2)_2$	-0.07	-0.21	0.02	-0.23
$CH(CH_3)_2$	-0.04	-0.15	0.04	-0.19
$CH_2CH_2CH_3$-ip	-0.06	-0.13	0.01	-0.14
CH_2CH_3	-0.07	-0.15	0	-0.15
CH_2CN	0.16	0.18	0.17	0.01
CH_2NH_2-ip	-0.03	-0.11	0.04	-0.15
CH_2OH-ip	0	0	0.03	-0.03
CH_3	-0.07	-0.17	0.01	-0.18
$CHCH_2$	0.06	-0.04	0.13	-0.17
CN	0.56	0.66	0.51	0.15
$COCH_3$	0.38	0.5	0.33	0.17
COCl	0.51	0.61	0.46	0.15
COF	0.55	0.7	0.48	0.22
COH	0.35	0.42	0.33	0.09
$CONH_2$	0.28	0.36	0.26	0.1
$COOCH_3$	0.37	0.45	0.34	0.11
COOH	0.37	0.45	0.34	0.11

Table D8: Proxy data used as the **Y** matrix in multivariate modeling, part 2/2

Substituent	σ_m	σ_p	F	R
$N(CH_3)_2$	-0.16	-0.83	0.15	-0.98
NC	0.48	0.49	0.47	0.02
NCO	0.27	0.19	0.31	-0.12
NCS	0.48	0.38	0.51	-0.13
NH_2	-0.16	-0.66	0.08	-0.74
$NHCH_3$	-0.21	-0.7	0.03	-0.73
$NHNH_2$−IP	-0.02	-0.55	0.22	-0.77
NHOH	-0.04	-0.34	0.11	-0.45
NNN	0.37	0.08	0.48	-0.4
NO	0.62	0.91	0.49	0.42
NO_2	0.71	0.78	0.65	0.13
OCH_3	0.12	-0.27	0.29	-0.56
OCN	0.67	0.54	0.69	-0.15
$OCOCH_3$	0.39	0.31	0.42	-0.11
OH	0.12	-0.37	0.33	-0.7
PH_2	0.06	0.05	0.09	-0.04
$PO(OH)_2$	0.36	0.42	0.34	0.08
SCF_3	0.4	0.5	0.36	0.14
SCH_3	0.15	0	0.23	-0.23
SCN	0.51	0.52	0.49	0.03
SH	0.25	0.15	0.3	-0.15
SiH_3	0.05	0.1	0.06	0.04

Appendix E

Appendix E

Table E1: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 1

Substituent	q(R)	μ_x^p(R)	μ_x^c(R)	μ_x(R)	μ_y^p(R)	μ_y^c(R)
AlH_2	7.58E-01	5.83E-05	-4.70E-05	1.13E-05	-3.88E-01	1.13E+00
BeH	8.40E-01	8.43E-06	-2.27E-15	8.43E-06	-9.51E-07	-1.87E-15
BH_2	6.51E-01	4.30E-05	-3.06E-05	1.24E-05	-3.37E-01	8.19E-01
C_6H_5	-5.09E-03	-2.33E-05	-3.08E-04	-3.31E-04	1.36E-01	-3.42E-03
CCCl	-1.58E-01	-2.80E-06	0.00E+00	-2.80E-06	-5.62E-06	0.00E+00
CCF	-1.60E-01	1.44E-05	0.00E+00	1.44E-05	-3.31E-07	0.00E+00
CCH	-1.47E-01	7.50E-06	-9.76E-16	7.50E-06	2.14E-06	-7.79E-16
CCl_2F	-1.07E-01	-9.93E-01	1.49E+00	5.02E-01	4.26E-01	-1.95E-01
CCl_2H	-7.73E-02	-9.55E-02	4.58E-02	-4.97E-02	7.48E-05	4.37E-07
CCl_3	-1.17E-01	-5.92E-05	-1.08E-04	-1.67E-04	2.24E-04	5.39E-04
$CClF_2$	-9.75E-02	-2.03E-01	6.73E-01	4.70E-01	-1.26E+00	1.69E+00
CClFH	-6.18E-02	-1.04E+00	1.47E+00	4.28E-01	-3.06E-01	9.64E-01
$CClH_2$	-3.24E-02	-1.44E-04	7.69E-07	-1.43E-04	1.18E-01	-2.29E-02
CF_2H	-4.32E-02	9.96E-02	-2.61E-02	7.35E-02	1.21E-04	1.17E-05
CF_3	-8.96E-02	5.94E-04	-4.92E-04	1.02E-04	1.57E-03	-1.88E-03
CFH_2	-9.40E-03	-3.56E-04	2.96E-06	-3.53E-04	1.20E-01	-6.58E-03
$CH_2CH_2CH_3$—ip	3.08E-02	2.01E-03	3.37E-05	2.04E-03	1.23E-01	1.93E-02
$CH_2CH_2CH_3$—op	3.35E-02	1.41E-02	8.59E-04	1.49E-02	-9.20E-02	2.31E-02
CH_2CH_3	3.17E-02	-1.08E-04	1.00E-05	-9.83E-05	1.29E-01	2.28E-02
CH_2CN	-4.49E-02	-4.99E-04	1.53E-07	-4.99E-04	1.17E-01	-3.17E-02
CH_2COCl—ip	-3.30E-02	1.58E+00	-2.71E+00	-1.13E+00	9.94E-01	-8.18E-01
CH_2COCl—op	-4.01E-02	1.57E+00	-2.71E+00	-1.14E+00	8.21E-01	-7.80E-01
CH_2COF—ip	-3.42E-02	9.67E-01	-1.77E+00	-8.02E-01	1.46E+00	-2.37E+00
CH_2COF—op	-3.61E-02	1.09E+00	-1.80E+00	-7.01E-01	1.35E+00	-2.34E+00
CH_2CHO—ip	-2.20E-02	-1.69E+00	2.76E+00	1.07E+00	-1.96E-01	-2.95E-02
CH_2CHO—op	-9.81E-03	-1.56E+00	2.75E+00	1.19E+00	-6.50E-02	-4.42E-02
CH_2COOH—ip	-2.56E-02	-5.17E-01	5.24E-01	7.33E-03	1.02E+00	-1.66E+00
CH_2COO——ip	-9.43E-01	7.84E-02	1.20E-01	1.98E-01	1.05E+00	-3.95E+00
CH_2CSH—ip	-1.79E-02	-9.46E-01	2.39E-01	-7.06E-01	1.39E+00	-7.80E-01
CH_2CSH—op	-1.01E-02	9.19E-01	-2.36E-01	6.82E-01	1.58E+00	-7.96E-01
CH_2NH_2—ip	1.68E-02	2.36E-01	-4.09E-01	-1.73E-01	-1.10E-01	-1.03E-02
CH_2NH_2—op	4.42E-02	-3.36E-02	-6.18E-01	-6.52E-01	-2.31E-01 6	4.77E-01
CH_2NO_2—ip	-5.64E-02	-6.83E-02	-4.62E-02	-1.15E-01	1.08E+00	-2.52E+00
CH_2NO_2—op	-5.48E-02	9.89E-02	-1.07E-01	-7.88E-03	1.08E+00	-2.52E+00
CH_2OH—ip	-3.60E-03	5.17E-01	-9.30E-01	-4.12E-01	-7.01E-01	1.14E+00
CH_2OH—op	2.28E-02	-3.31E-01	9.35E-01	6.04E-01	-6.87E-01	1.15E+00
CH_2OOCH_3—ip	-1.58E-02	1.24E-01	-1.19E-02	1.12E-01	4.17E-02	-3.77E-03

Table E2: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 2

Substituent	q(R)	μ_x^p(R)	μ_x^c(R)	μ_x(R)	μ_y^p(R)	μ_y^c(R)	μ_y(R)
CH_2OOCH_3—op	-1.91E-03	5.54E-02	2.29E-05	5.54E-02	7.01E-02	-1.62E-03	6.85E-02
CH_2SH—ip	-1.69E-02	-3.46E-01	1.91E-01	-1.55E-01	4.50E-01	1.40E-01	5.90E-01
CH_2SH—op	-1.12E-02	5.31E-01	-2.07E-01	3.23E-01	4.73E-01	1.40E-01	6.14E-01
CH_3	1.76E-02	8.03E-02	7.88E-03	8.81E-02	8.02E-02	7.86E-03	8.81E-02
$CHCCH_2$	-4.20E-02	1.25E-04	7.07E-06	1.32E-04	1.12E-01	-2.70E-02	8.46E-02
CHCCO	-4.15E-02	-8.76E-01	1.72E+00	8.48E-01	-2.70E+00	3.28E+00	5.86E-01
$CH(CH_2)_2$	1.54E-03	-1.58E-04	-1.03E-05	-1.68E-04	8.45E-02	1.69E-04	8.47E-02
$CHCH_2$	-5.95E-03	8.63E-05	-2.16E-17	8.63E-05	1.12E-01	-3.85E-03	1.08E-01
$CH(CH_3)_2$	4.10E-02	1.11E-01	2.55E-02	1.37E-01	6.95E-04	-4.61E-04	2.34E-04
CHCO	-8.11E-02	4.91E-05	-1.67E-17	4.91E-05	1.16E-01	-5.22E-02	6.37E-02
CHNH	-1.82E-03	-4.80E-02	-5.95E-01	-6.43E-01	-1.03E+00	1.71E+00	6.84E-01
CHNN	-7.93E-02	-1.95E-04	-1.94E-16	-1.95E-04	1.20E-01	-5.11E-02	6.88E-02
CN	-1.99E-01	3.08E-06	0.00E+00	3.08E-06	-1.79E-05	0.00E+00	-1.79E-05
CNO	-2.32E-01	7.35E-06	0.00E+00	7.35E-06	-1.86E-05	0.00E+00	-1.86E-05
CO	5.40E-01	8.20E-06	0.00E+00	8.20E-06	-1.14E-05	0.00E+00	-1.14E-05
$COCH_3$	4.54E-03	1.55E+00	-2.76E+00	-1.20E+00	5.11E-02	-4.07E-02	1.04E-02
COCl	-9.98E-02	1.55E+00	-2.27E+00	-7.22E-01	1.94E-01	1.10E-01	3.03E-01
COF	-9.78E-02	8.18E-01	-1.34E+00	-5.26E-01	-1.19E+00	1.90E+00	7.12E-01
COH	-1.16E-02	6.12E-05	-2.45E-18	6.12E-05	1.08E-01	-7.40E-03	1.00E-01
CONH2	-1.14E-02	9.75E-01	-2.58E+00	-1.61E+00	-6.89E-01	5.55E-01	-1.35E-01
$COOCH_3$	-5.54E-02	1.90E-01	1.06E-01	2.96E-01	-4.72E-01	1.17E+00	6.96E-01
COOH	-6.72E-01	5.78E-01	-1.22E+00	-6.44E-01	-8.38E-01	8.75E-01	3.68E-02
$CSCH_3$	-3.63E-02	-8.27E-01	2.21E-01	-6.07E-01	1.64E+00	-8.15E-01	8.30E-01
CSH	-4.97E-02	-1.46E-05	-1.11E-17	-1.46E-05	1.02E-01	-3.10E-02	7.06E-02
MgH	7.70E-01	9.71E-06	-5.44E-16	9.71E-06	-3.18E-07	-2.45E-17	-3.18E-07
NC	-5.40E-01	3.35E-05	0.00E+00	3.35E-05	6.34E-07	0.00E+00	6.34E-07
NCH_2	-3.48E-01	-1.30E-01	-4.31E-01	-5.61E-01	-1.18E+00	1.78E+00	6.06E-01
$N(CH_3)_2$	-3.43E-01	9.81E-02	2.98E-01	3.96E-01	5.00E-01	-6.57E-01	-1.57E-01
NCO	-4.82E-01	-4.93E-01	1.19E+00	7.00E-01	2.76E-01	-5.56E-01	-2.80E-01
NCS	-4.94E-01	3.66E-01	-7.50E-01	-3.84E-01	3.43E+00	-2.88E+00	5.49E-01
NF_2	-4.12E-01	2.98E-02	5.76E-01	6.06E-01	-8.96E-01	1.31E+00	4.13E-01
NH_2	-3.58E-01	5.37E-05	1.31E-06	5.50E-05	1.75E-01	-1.69E-01	5.89E-03
NH_3^+	4.85E-01	7.65E-02	-1.07E-01	-3.05E-02	7.65E-02	-1.07E-01	-3.05E-02
$NHCH_3$	-3.47E-01	-2.07E-02	-5.04E-01	-5.25E-01	-3.74E-01	5.26E-01	1.52E-01
NHF	-3.82E-01	-2.25E-02	7.31E-01	7.09E-01	-7.21E-01	1.30E+00	5.75E-01
$NHNH_2$—ip	-3.42E-01	-2.57E-02	4.15E-02	1.58E-02	7.94E-02	-6.82E-02	1.12E-02
$NHNH_2$—op	-3.64E-01	1.83E-01	-1.66E-01	1.61E-02	6.38E-02	-5.68E-02	7.03E-03
NHOH	-3.66E-01	-4.28E-01	3.60E-01	-6.82E-02	-3.87E-01	6.26E-01	2.39E-01
NNH-trans	-3.42E-01	1.83E-01	-1.68E-01	1.55E-02	5.41E-02	-5.33E-02	8.07E-04
NNH-cis	-3.17E-01	-7.95E-05	-2.23E-17	-7.95E-05	7.41E-02	-6.75E-02	6.57E-03

Table E3: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 3

Substituent	q(R)	μ_x^p(R)	μ_x^c(R)	μ_x(R)	μ_y^p(R)	μ_y^c(R)	μ_y(R)
NNN	-4.32E-01	-8.01E-02	-5.47E-01	-6.27E-01	-2.12E-01	-2.63E-02	-2.39E-01
NO_2	-4.27E-01	7.68E-05	1.13E-16	7.68E-05	-2.78E-05	-1.43E-05	-4.21E-05
NO	-2.99E-01	-1.90E-01	-4.13E-01	-6.03E-01	-7.11E-01	1.09E+00	3.75E-01
OCH_3	-5.69E-01	-2.29E-01	7.29E-01	5.01E-01	-8.04E-01	1.22E+00	4.12E-01
OCl	-6.01E-01	3.28E-01	-8.56E-01	-5.28E-01	5.84E-01	-4.77E-01	1.07E-01
OCN	-6.41E-01	4.08E-01	-8.80E-01	-4.71E-01	2.49E-01	-1.65E+00	-1.40E+00
$OCOCH_3$	-6.05E-01	-7.42E-01	6.85E-01	-5.65E-02	1.10E+00	-1.82E+00	-7.22E-01
OCOH	-6.08E-01	6.34E-01	-1.24E+00	-6.03E-01	-9.56E-01	9.18E-01	-3.77E-02
OF	-6.03E-01	2.76E-01	-8.78E-01	-6.02E-01	-2.93E-01	6.39E-01	3.46E-01
OH_2^+	2.70E-01	1.45E-05	1.58E-05	3.03E-05	9.54E-02	-1.78E-01	-8.23E-02
OH	-5.80E-01	-1.25E-05	-1.22E-16	-1.25E-05	1.28E-01	-1.70E-01	-4.27E-02
O-	-1.43E+00	6.75E-06	0.00E+00	6.75E-06	-2.86E-05	0.00E+00	-2.86E-05
ONH_2	-5.83E-01	1.48E-01	-5.49E-02	9.29E-02	-4.80E-01	7.21E-01	2.41E-01
ONO	-5.93E-01	7.59E-01	-1.43E+00	-6.74E-01	-7.99E-01	1.02E+00	2.20E-01
OOH	-5.87E-01	1.32E-01	-1.75E-01	-4.25E-02	3.40E-02	-4.08E-02	-6.81E-03
$OO\cdot$	-5.91E-01	2.40E-01	-8.42E-01	-6.02E-01	1.18E-02	-5.72E-01	-5.60E-01
OPH_2O	-6.27E-01	7.49E-02	-9.70E-01	-8.96E-01	-3.19E-01	8.39E-01	5.20E-01
PH_2	5.40E-01	2.10E-04	-6.28E-06	2.04E-04	-2.37E-01	6.01E-01	3.64E-01
PH_2O	5.80E-01	-6.13E-04	-2.53E-05	-6.38E-04	-3.30E-01	7.06E-01	3.76E-01
PHO(OH)	5.87E-01	-1.17E-01	1.06E+00	9.42E-01	-4.26E-01	1.16E+00	7.34E-01
$PO(OH)_2$	5.87E-01	-1.07E-02	-4.05E-01	-4.15E-01	-3.59E-01	-8.25E-02	-4.41E-01
SCF_3	1.98E-02	-1.95E+00	2.48E+00	5.27E-01	-4.70E-01	9.63E-02	-3.73E-01
SCH_3	7.63E-02	5.27E-01	-1.30E-01	3.97E-01	4.03E-01	1.35E-01	5.38E-01
SCN	-1.21E-02	3.62E-01	2.68E-02	3.89E-01	2.78E+00	-4.13E+00	-1.34E+00
SH	4.98E-02	1.59E-05	-1.63E-18	1.59E-05	2.81E-02	3.39E-02	6.20E-02
SHO_2	5.26E-02	-3.06E-02	-3.63E-02	-6.68E-02	-1.52E-04	3.29E-05	-1.19E-04
SHO	7.26E-02	-6.81E-01	1.99E-01	-4.82E-01	2.29E+00	-3.58E+00	-1.29E+00
$SiCl_2H$	6.84E-01	3.67E-01	-8.05E-01	-4.38E-01	-3.06E-04	1.29E-05	-2.93E-04
$SiCl_3$	6.70E-01	2.75E-04	5.15E-06	2.80E-04	8.23E-04	-4.80E-06	8.18E-04
$SiClH_2$	6.94E-01	1.24E-04	-3.63E-05	8.77E-05	-4.01E-01	9.35E-01	5.35E-01
SiCl	7.13E-01	-1.75E+00	9.82E-01	-7.66E-01	-2.41E+00	2.92E+00	5.10E-01
SiF_2H	6.99E-01	3.74E-01	-8.34E-01	-4.60E-01	1.29E-04	2.45E-05	1.53E-04
SiF_3	6.83E-01	-4.24E-04	2.26E-03	1.84E-03	-8.07E-04	8.48E-04	4.05E-05
$SiFH_2$	7.05E-01	-8.21E-04	-2.99E-05	-8.51E-04	-3.98E-01	9.53E-01	5.55E-01
SiF	7.21E-01	-1.78E+00	9.94E-01	-7.88E-01	2.06E+00	-2.55E+00	-4.91E-01
SiH_3	7.02E-01	-2.32E-01	5.78E-01	3.46E-01	-2.32E-01	5.78E-01	3.46E-01
SiH	7.12E-01	3.63E-04	1.86E-16	3.63E-04	-2.36E-01	7.34E-01	4.97E-01
S^-	-7.98E-01	-2.79E-05	0.00E+00	-2.79E-05	-1.40E-06	0.00E+00	-1.40E-06
SSH	4.25E-02	3.94E-02	2.79E-02	6.74E-02	2.10E-02	6.31E-03	2.73E-02

Table E4: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 4

Substituent	μ_z^p(R)	μ_z^c(R)	μ_z(R)	$\|\|\boldsymbol{\mu}^p$(R)$\|\|$	$\|\|\boldsymbol{\mu}^c$(R)$\|\|$	$\|\|\boldsymbol{\mu}$(R)$\|\|$	Q_{xx}(R)
AlH_2	-5.51E-06	-1.70E-16	-5.51E-06	3.88E-01	1.13E+00	7.44E-01	-4.54E+00
BeH	-5.43E-01	1.19E+00	6.48E-01	5.43E-01	1.19E+00	6.48E-01	9.52E-01
BH_2	-2.43E-06	3.79E-18	-2.43E-06	3.37E-01	8.19E-01	4.81E-01	-1.65E+00
C_6H_5	-6.09E-05	-8.17E-17	-6.09E-05	1.36E-01	3.43E-03	1.33E-01	2.33E+00
CCCl	9.17E-01	-1.09E+00	-1.73E-01	9.17E-01	1.09E+00	1.73E-01	-7.70E-01
CCF	2.77E+00	-3.04E+00	-2.70E-01	2.77E+00	3.04E+00	2.70E-01	-3.52E-01
CCH	1.22E-01	-1.03E-01	1.90E-02	1.22E-01	1.03E-01	1.90E-02	-1.98E+00
CCl_2F	3.80E-04	-9.36E-06	3.70E-04	1.08E+00	1.51E+00	5.53E-01	-1.67E-01
CCl_2H	-3.26E-01	1.03E+00	7.07E-01	3.40E-01	1.03E+00	7.09E-01	3.67E-01
CCl_3	-2.49E-01	7.17E-01	4.68E-01	2.49E-01	7.17E-01	4.68E-01	-5.93E-01
$CClF_2$	2.86E-04	6.36E-06	2.93E-04	1.28E+00	1.82E+00	6.32E-01	2.42E+00
CClFH	9.74E-02	-3.68E-02	6.06E-02	1.09E+00	1.76E+00	7.87E-01	-2.78E+00
$CClH_2$	2.46E-01	-1.07E+00	-8.22E-01	2.72E-01	1.07E+00	8.27E-01	-7.41E-01
CF_2H	-1.29E+00	2.13E+00	8.45E-01	1.29E+00	2.13E+00	8.48E-01	1.51E+00
CF_3	-1.22E+00	1.93E+00	7.10E-01	1.22E+00	1.93E+00	7.10E-01	-1.29E+00
CFH_2	1.01E+00	-1.81E+00	-7.95E-01	1.02E+00	1.81E+00	8.03E-01	1.17E-01
$CH_2CH_2CH_3$—ip	6.94E-02	5.24E-02	1.22E-01	1.41E-01	5.59E-02	1.87E-01	-1.56E-01
$CH_2CH_2CH_3$—op	-1.14E-01	-2.12E-02	-1.35E-01	1.47E-01	3.14E-02	1.52E-01	-5.51E-01
CH_2CH_3	5.92E-02	9.19E-03	6.84E-02	1.42E-01	2.46E-02	1.66E-01	-1.12E-01
CH_2CN	1.95E+00	-3.56E+00	-1.61E+00	1.95E+00	3.56E+00	1.61E+00	1.25E+00
CH_2COCl—ip	1.13E-04	-3.59E-06	1.09E-04	1.87E+00	2.83E+00	1.15E+00	9.02E-01
CH_2COCl—op	1.04E-01	-2.44E-02	7.93E-02	1.77E+00	2.82E+00	1.14E+00	9.69E-01
CH_2COF—ip	-2.15E-04	-3.28E-06	-2.18E-04	1.75E+00	2.96E+00	1.21E+00	4.79E-01
CH_2COF—op	1.03E-01	-2.21E-02	8.11E-02	1.74E+00	2.95E+00	1.22E+00	-6.16E-02
CH_2CHO—ip	-1.06E-04	9.93E-07	-1.05E-04	1.70E+00	2.76E+00	1.10E+00	-2.84E+00
CH_2CHO—op	1.06E-01	-6.07E-03	9.98E-02	1.56E+00	2.75E+00	1.20E+00	-2.22E+00
CH_2COOH—ip	6.48E-05	-4.39E-06	6.04E-05	1.14E+00	1.74E+00	6.46E-01	4.25E+00
CH_2COO——ip	2.89E-04	-9.82E-07	2.88E-04	1.05E+00	3.95E+00	2.91E+00	-5.12E+00
CH_2CSH—ip	3.84E-04	4.05E-06	3.88E-04	1.68E+00	8.16E-01	9.35E-01	-4.57E-01
CH_2CSH—op	1.04E-01	-6.21E-03	9.80E-02	1.83E+00	8.30E-01	1.04E+00	-7.69E-01
CH_2NH_2—ip	-1.78E-01	6.22E-01	4.44E-01	3.16E-01	7.45E-01	4.92E-01	1.46E-01
CH_2NH_2—op	-2.40E-04	2.39E-06	-2.38E-04	2.33E-01	7.81E-01	6.97E-01	-1.04E+00
CH_2NO_2—ip	-2.65E-04	5.29E-08	-2.65E-04	1.08E+00	2.52E+00	1.45E+00	-3.22E+00
CH_2NO_2—op	1.00E-01	-3.30E-02	6.75E-02	1.09E+00	2.52E+00	1.44E+00	-3.25E+00
CH_2OH—ip	-1.14E-04	2.97E-06	-1.11E-04	8.72E-01	1.47E+00	6.04E-01	1.59E+00
CH_2OH—op	1.08E-01	1.42E-02	1.22E-01	7.70E-01	1.48E+00	7.71E-01	1.17E+00
CH_2OOCH_3—ip	1.67E-05	-1.30E-06	1.54E-05	1.30E-01	1.24E-02	1.18E-01	-1.21E+00

Table E5: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 5

Substituent	μ_z^p(R)	μ_z^c(R)	μ_z(R)	$\|\|\boldsymbol{\mu}^p(R)\|\|$	$\|\|\boldsymbol{\mu}^c(R)\|\|$	$\|\|\boldsymbol{\mu}(R)\|\|$	Q_{xx}(R)
CH_2OOCH_3—op	1.11E-01	-1.26E-03	1.10E-01	1.42E-01	2.05E-03	1.41E-01	-2.05E+00
CH_2SH—ip	1.06E-03	2.01E-06	1.07E-03	5.68E-01	2.37E-01	6.10E-01	1.87E+00
CH_2SH—op	1.03E-01	-7.07E-03	9.61E-02	7.18E-01	2.51E-01	7.00E-01	1.50E+00
CH_3	8.02E-02	7.86E-03	8.80E-02	1.39E-01	1.36E-02	1.53E-01	-9.69E-05
$CHCCH_2$	7.54E-02	-1.61E-02	5.94E-02	1.35E-01	3.14E-02	1.03E-01	-1.41E+00
CHCCO	-2.81E-04	-3.93E-19	-2.81E-04	2.84E+00	3.71E+00	1.03E+00	1.12E+00
$CH(CH_2)_2$	1.11E-01	9.84E-04	1.12E-01	1.40E-01	9.98E-04	1.41E-01	-1.18E+00
$CHCH_2$	7.62E-02	-2.27E-03	7.40E-02	1.35E-01	4.47E-03	1.31E-01	-2.64E+00
$CH(CH_3)_2$	8.90E-02	5.84E-02	1.47E-01	1.43E-01	6.37E-02	2.01E-01	5.81E-01
CHCO	2.89E+00	-3.48E+00	-5.92E-01	2.89E+00	3.48E+00	5.96E-01	-1.43E+00
CHNH	-5.66E-05	0.00E+00	-5.66E-05	1.03E+00	1.81E+00	9.39E-01	1.10E+00
CHNN	6.24E-01	-1.27E+00	-6.46E-01	6.36E-01	1.27E+00	6.50E-01	-1.27E+00
CN	1.63E+00	-2.80E+00	-1.17E+00	1.63E+00	2.80E+00	1.17E+00	-8.32E-02
CNO	2.31E+00	-3.51E+00	-1.21E+00	2.31E+00	3.51E+00	1.21E+00	6.13E-01
CO	2.30E+00	-2.55E+00	-2.49E-01	2.30E+00	2.55E+00	2.49E-01	-3.46E-02
$COCH_3$	1.86E-04	1.13E-06	1.87E-04	1.55E+00	2.76E+00	1.20E+00	-2.65E+00
COCl	-3.50E-05	0.00E+00	-3.50E-05	1.56E+00	2.27E+00	7.83E-01	-1.62E+00
COF	-1.24E-04	8.42E-17	-1.24E-04	1.44E+00	2.33E+00	8.85E-01	-2.89E+00
COH	1.54E+00	-2.57E+00	-1.03E+00	1.54E+00	2.57E+00	1.03E+00	1.18E-02
CONH2	-5.56E-05	9.92E-18	-5.56E-05	1.19E+00	2.64E+00	1.61E+00	-1.36E+00
$COOCH_3$	-4.29E-04	-1.70E-06	-4.31E-04	5.09E-01	1.17E+00	7.56E-01	4.68E+00
COOH	-1.01E-04	-6.49E-17	-1.01E-04	1.02E+00	1.50E+00	6.45E-01	-4.89E+00
$CSCH_3$	3.92E-04	4.05E-06	3.96E-04	1.84E+00	8.45E-01	1.03E+00	-9.44E-01
CSH	-1.91E+00	1.15E+00	-7.56E-01	1.91E+00	1.16E+00	7.59E-01	-7.07E-01
MgH	-3.20E-01	1.20E+00	8.80E-01	3.20E-01	1.20E+00	8.80E-01	2.21E+00
NC	2.52E+00	-1.39E+00	1.14E+00	2.52E+00	1.39E+00	1.14E+00	9.77E-01
NCH_2	-6.00E-05	-1.18E-16	-6.00E-05	1.18E+00	1.83E+00	8.26E-01	1.28E+00
$N(CH_3)_2$	-7.97E-04	1.05E-04	-6.92E-04	5.10E-01	7.22E-01	4.26E-01	2.68E-01
NCO	-1.73E-05	0.00E+00	-1.73E-05	5.65E-01	1.32E+00	7.54E-01	1.76E+00
NCS	-2.92E-06	-5.33E-17	-2.92E-06	3.45E+00	2.98E+00	6.70E-01	2.94E+00
NF_2	1.69E-05	-5.84E-06	1.11E-05	8.96E-01	1.43E+00	7.33E-01	1.03E+00
NH_2	7.20E-02	-6.94E-01	-6.22E-01	1.89E-01	7.14E-01	6.22E-01	1.43E+00
NH_3^+	7.65E-02	-1.07E-01	-3.05E-02	1.33E-01	1.85E-01	5.28E-02	-1.13E-05
$NHCH_3$	1.51E-01	-1.42E-01	8.99E-03	4.04E-01	7.43E-01	5.47E-01	-3.85E-01
NHF	1.33E-01	-1.51E-01	-1.84E-02	7.33E-01	1.50E+00	9.13E-01	-2.90E-01
$NHNH_2$—ip	9.95E-02	6.91E-01	7.90E-01	1.30E-01	6.95E-01	7.90E-01	1.12E+00
$NHNH_2$—op	-1.09E-01	8.90E-01	7.81E-01	2.22E-01	9.08E-01	7.82E-01	5.51E-01
NHOH	1.39E-01	-1.47E-01	-7.79E-03	5.93E-01	7.37E-01	2.49E-01	1.27E+00
NNH-trans	4.40E-06	-6.35E-17	4.40E-06	1.91E-01	1.76E-01	1.56E-02	1.44E+00
NNH-cis	1.70E-01	9.92E-01	1.16E+00	1.85E-01	9.94E-01	1.16E+00	5.58E-01

Table E6: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 6

Substituent	μ_z^p(R)	μ_z^c(R)	μ_z(R)	$\|\|\boldsymbol{\mu}^p(R)\|\|$	$\|\|\boldsymbol{\mu}^c(R)\|\|$	$\|\|\boldsymbol{\mu}(R)\|\|$	Q_{xx}(R)
NNN	1.67E-05	1.09E-16	1.67E-05	2.27E-01	5.47E-01	6.71E-01	2.97E+00
NO_2	-5.61E-01	1.55E+00	9.84E-01	5.61E-01	1.55E+00	9.84E-01	2.12E+00
NO	-2.98E-05	1.54E-16	-2.98E-05	7.36E-01	1.16E+00	7.10E-01	9.62E-02
OCH_3	2.22E-04	1.99E-06	2.24E-04	8.36E-01	1.42E+00	6.49E-01	1.50E+00
OCl	5.81E-05	-4.34E-16	5.81E-05	6.70E-01	9.80E-01	5.39E-01	1.26E+00
OCN	7.13E-06	-1.25E-16	7.13E-06	4.78E-01	1.87E+00	1.47E+00	3.50E+00
$OCOCH_3$	6.28E-05	-4.40E-06	5.84E-05	1.32E+00	1.94E+00	7.24E-01	-2.98E+00
OCOH	-1.06E-04	-4.34E-16	-1.06E-04	1.15E+00	1.54E+00	6.04E-01	-4.89E+00
OF	-2.05E-06	-2.99E-16	-2.05E-06	4.03E-01	1.09E+00	6.94E-01	7.60E-01
OH_2^+	4.00E-01	-6.74E-01	-2.74E-01	4.11E-01	6.97E-01	2.86E-01	2.39E+00
OH	4.40E-01	-1.15E+00	-7.14E-01	4.59E-01	1.17E+00	7.15E-01	-1.21E+00
O-	1.62E-01	-1.55E+00	-1.39E+00	1.62E-01	1.55E+00	1.39E+00	-1.02E+00
ONH_2	1.53E-04	2.58E-06	1.56E-04	5.02E-01	7.23E-01	2.58E-01	2.64E-01
ONO	-3.70E-05	-2.11E-16	-3.70E-05	1.10E+00	1.76E+00	7.09E-01	-2.67E+00
OOH	-3.83E-01	1.04E+00	6.56E-01	4.07E-01	1.05E+00	6.57E-01	1.45E+00
$OO\cdot$	-6.67E-05	0.00E+00	-6.67E-05	2.40E-01	1.02E+00	8.23E-01	3.30E-01
OPH_2O	-2.43E-01	5.30E-01	2.87E-01	4.08E-01	1.39E+00	1.07E+00	-8.16E+00
PH_2	-2.42E+00	1.91E+00	-5.08E-01	2.43E+00	2.00E+00	6.25E-01	-6.13E-02
PH_2O	-8.01E-01	2.46E+00	1.65E+00	8.66E-01	2.55E+00	1.70E+00	-8.08E-02
PHO(OH)	-1.51E-02	3.22E-02	1.71E-02	4.42E-01	1.57E+00	1.19E+00	-6.86E+00
$PO(OH)_2$	-1.86E-03	-6.50E-05	-1.92E-03	3.59E-01	4.13E-01	6.06E-01	2.78E+00
SCF_3	4.33E-04	6.62E-06	4.40E-04	2.01E+00	2.48E+00	6.46E-01	1.46E+00
SCH_3	-4.99E-04	8.30E-07	-4.98E-04	6.63E-01	1.88E-01	6.68E-01	2.05E+00
SCN	1.82E-04	6.55E-19	1.82E-04	2.81E+00	4.13E+00	1.40E+00	3.75E+00
SH	-6.10E-01	1.40E-01	-4.70E-01	6.10E-01	1.44E-01	4.74E-01	-2.89E+00
SHO_2	-2.02E+00	3.39E+00	1.37E+00	2.02E+00	3.39E+00	1.37E+00	3.48E+00
SHO	3.04E-02	4.65E-02	7.69E-02	2.39E+00	3.58E+00	1.38E+00	4.53E-01
$SiCl_2H$	-1.11E+00	1.91E+00	7.98E-01	1.17E+00	2.07E+00	9.11E-01	-2.82E-01
$SiCl_3$	-1.21E+00	2.04E+00	8.30E-01	1.21E+00	2.04E+00	8.30E-01	-3.57E+00
$SiClH_2$	8.46E-01	-1.58E+00	-7.30E-01	9.36E-01	1.83E+00	9.05E-01	-3.74E+00
SiCl	-2.57E-05	0.00E+00	-2.57E-05	2.98E+00	3.08E+00	9.20E-01	-1.13E+00
SiF_2H	-5.94E-01	1.51E+00	9.18E-01	7.02E-01	1.73E+00	1.03E+00	1.81E+00
SiF_3	-7.52E-01	1.79E+00	1.04E+00	7.52E-01	1.79E+00	1.04E+00	-3.46E+00
$SiFH_2$	-4.19E-01	1.13E+00	7.12E-01	5.78E-01	1.48E+00	9.03E-01	-1.68E+00
SiF	5.22E-05	0.00E+00	5.22E-05	2.72E+00	2.74E+00	9.28E-01	5.15E-01
SiH_3	-2.33E-01	5.78E-01	3.46E-01	4.02E-01	1.00E+00	5.99E-01	-5.17E-04
SiH	-2.61E+00	2.08E+00	-5.31E-01	2.62E+00	2.21E+00	7.28E-01	2.18E+00
S^-	-7.16E-01	-7.05E-01	-1.42E+00	7.16E-01	7.05E-01	1.42E+00	-2.44E+00
SSH	6.31E-01	-1.20E-01	5.11E-01	6.33E-01	1.24E-01	5.16E-01	-2.78E-01

Table E7: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 7

Substituent	Q_{xy}(R)	Q_{xz}(R)	Q_{yy}(R)	Q_{yz}(R)	Q_{zz}(R)	$\|\mathbf{Q}(R)\|$	Vol(R)
AlH_2	2.11E-05	-5.15E-05	1.36E+00	-5.01E-05	3.18E+00	4.66E+00	3.01E+02
BeH	-9.97E-07	-5.24E-05	9.52E-01	2.65E-06	-1.90E+00	1.90E+00	1.59E+02
BH_2	-8.36E-06	1.80E-06	5.81E-01	-1.99E-05	1.06E+00	1.67E+00	2.04E+02
C_6H_5	1.11E-03	-1.75E-03	3.77E+00	-2.13E-04	-6.09E+00	6.15E+00	7.42E+02
CCCl	-9.09E-06	-1.14E-04	-7.70E-01	-3.11E-05	1.54E+00	1.54E+00	4.38E+02
CCF	7.15E-06	-4.59E-05	-3.52E-01	-2.95E-05	7.04E-01	7.04E-01	3.27E+02
CCH	9.53E-06	-4.92E-05	-1.98E+00	2.11E-05	3.97E+00	3.97E+00	2.96E+02
CCl_2F	2.01E+00	-2.87E-03	1.89E-01	5.44E-04	-2.24E-02	2.07E-01	5.71E+02
CCl_2H	1.28E-05	3.65E-02	-2.42E+00	1.08E-04	2.05E+00	2.61E+00	5.35E+02
CCl_3	2.97E-04	-8.40E-04	-5.90E-01	4.89E-04	1.18E+00	1.18E+00	6.82E+02
$CClF_2$	1.01E+00	-5.02E-04	-6.75E-01	-2.34E-05	-1.74E+00	2.50E+00	4.58E+02
CClFH	1.61E+00	2.86E-02	1.89E+00	5.68E-02	8.94E-01	2.84E+00	4.21E+02
$CClH_2$	1.16E-03	2.07E-05	-5.33E-01	-3.31E-01	1.27E+00	1.28E+00	3.84E+02
CF_2H	3.63E-04	1.57E-01	-2.97E+00	2.56E-04	1.46E+00	2.97E+00	3.05E+02
CF_3	-5.19E-04	9.36E-04	-1.29E+00	-2.07E-03	2.58E+00	2.58E+00	3.44E+02
CFH_2	1.35E-04	-3.03E-04	4.67E-01	-3.68E-01	-5.84E-01	6.18E-01	2.67E+02
$CH_2CH_2CH_3$—ip	3.95E-04	-6.47E-04	8.62E-01	5.51E-01	-7.06E-01	9.19E-01	5.50E+02
$CH_2CH_2CH_3$—op	-5.73E-01	-6.65E-01	-6.44E-03	4.98E-01	5.58E-01	6.40E-01	5.50E+02
CH_2CH_3	5.87E-04	-4.63E-04	5.02E-01	6.92E-01	-3.90E-01	5.27E-01	3.92E+02
CH_2CN	4.71E-04	6.07E-04	1.40E+00	-1.81E-01	-2.65E+00	2.65E+00	3.98E+02
CH_2COCl—ip	-3.20E+00	-5.73E-05	-2.01E+00	-3.69E-04	1.11E+00	2.01E+00	5.59E+02
CH_2COCl—op	-2.70E+00	-2.44E-01	-2.37E+00	1.71E-02	1.40E+00	2.39E+00	5.59E+02
CH_2COF—ip	1.56E+00	-5.02E-04	-2.41E+00	6.42E-04	1.93E+00	2.55E+00	4.43E+02
CH_2COF—op	1.62E+00	-1.29E-01	-2.07E+00	-2.18E-01	2.13E+00	2.43E+00	4.43E+02
CH_2CHO—ip	9.32E-01	-3.19E-04	1.72E+00	-2.47E-04	1.12E+00	2.86E+00	4.13E+02
CH_2CHO—op	8.07E-01	4.52E-01	9.44E-01	-1.67E-01	1.27E+00	2.22E+00	4.12E+02
CH_2COOH—ip	-3.99E+00	-3.25E-05	-4.62E+00	-1.63E-04	3.71E-01	5.13E+00	4.82E+02
CH_2COO——ip	-1.65E+00	5.41E-04	-4.56E-01	-2.03E-04	5.58E+00	6.20E+00	5.27E+02
CH_2CSH—ip	-1.46E-01	-8.45E-04	5.85E-01	-6.23E-04	-1.28E-01	6.15E-01	5.40E+02
CH_2CSH—op	1.07E+00	4.97E-01	8.31E-01	3.61E-01	-6.21E-02	9.26E-01	5.40E+02
CH_2NH_2—ip	5.00E-01	2.53E+00	1.69E+00	1.96E-01	-1.84E+00	2.04E+00	3.51E+02
CH_2NH_2—op	1.59E+00	-7.37E-04	-1.08E-01	-1.19E-04	1.14E+00	1.26E+00	3.49E+02
CH_2NO_2—ip	1.87E-01	-5.71E-04	1.10E+00	4.64E-04	2.12E+00	3.27E+00	4.50E+02
CH_2NO_2—op	3.80E-02	2.30E-02	1.06E+00	-9.20E-02	2.19E+00	3.31E+00	4.49E+02
CH_2OH—ip	2.75E+00	5.87E-04	-4.90E-01	1.16E-04	-1.10E+00	1.63E+00	3.10E+02
CH_2OH—op	-2.06E+00	2.20E-01	-3.48E-01	5.27E-01	-8.17E-01	1.20E+00	3.07E+02
CH_2OOCH_3—ip	3.82E+00	6.46E-04	3.55E+00	-7.37E-04	-2.34E+00	3.61E+00	5.60E+02

Table E8: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 8

Substituent	Q_{xy}(R)	Q_{xz}(R)	Q_{yy}(R)	Q_{yz}(R)	Q_{zz}(R)	$\|\mathbf{Q}(R)\|$	Vol(R)
CH_2OOCH_3—op	-3.07E+00	-4.48E-02	4.10E+00	7.05E-01	-2.05E+00	4.10E+00	5.58E+02
CH_2SH—ip	1.77E+00	4.22E-04	3.81E-01	2.03E-03	-2.25E+00	2.41E+00	4.36E+02
CH_2SH—op	-1.07E+00	1.31E-01	5.75E-01	4.68E-01	-2.07E+00	2.14E+00	4.35E+02
CH_3	2.05E-01	2.05E-01	-6.07E-05	2.05E-01	1.58E-04	1.59E-04	2.34E+02
$CHCCH_2$	-9.53E-05	-9.93E-04	-1.31E+00	2.67E-01	2.72E+00	2.72E+00	4.50E+02
CHCCO	8.87E-01	-1.37E-03	-8.74E-01	-3.86E-04	-2.49E-01	1.18E+00	4.63E+02
$CH(CH_2)_2$	-8.75E-05	3.27E-03	-5.67E-01	5.46E-01	1.75E+00	1.78E+00	4.69E+02
$CHCH_2$	-9.88E-05	-1.81E-03	1.24E+00	4.14E-01	1.40E+00	2.64E+00	3.43E+02
$CH(CH_3)_2$	4.67E-04	7.05E-01	-9.66E-01	1.54E-03	3.84E-01	9.73E-01	5.50E+02
CHCO	-3.60E-04	-4.90E-04	2.58E+00	1.00E-01	-1.15E+00	2.59E+00	3.59E+02
CHNH	1.89E+00	-2.22E-04	1.60E-01	-4.28E-05	-1.26E+00	1.37E+00	2.96E+02
CHNN	2.68E-04	1.81E-05	2.80E+00	6.13E-02	-1.54E+00	2.81E+00	3.64E+02
CN	-1.66E-06	-5.05E-05	-8.32E-02	-9.30E-05	1.66E-01	1.66E-01	2.48E+02
CNO	2.94E-06	-5.11E-05	6.13E-01	-6.87E-05	-1.23E+00	1.23E+00	3.24E+02
CO	-4.66E-07	-4.55E-05	-3.46E-02	-5.31E-05	6.93E-02	6.93E-02	1.89E+02
$COCH_3$	-1.33E+00	-2.67E-04	1.68E+00	3.76E-04	9.65E-01	2.68E+00	4.09E+02
COCl	-2.92E+00	-7.77E-04	6.01E-01	1.22E-04	1.02E+00	1.64E+00	4.06E+02
COF	-1.32E+00	5.95E-05	1.50E+00	2.47E-04	1.40E+00	2.89E+00	2.89E+02
COH	-2.85E-05	-1.39E-04	3.12E-01	-3.92E-01	-3.24E-01	3.67E-01	2.51E+02
CONH2	-4.63E-01	-2.46E-04	3.19E+00	6.52E-04	-1.83E+00	3.20E+00	3.66E+02
$COOCH_3$	-1.52E+00	-1.63E-04	-5.39E+00	-8.53E-04	7.10E-01	5.85E+00	4.84E+02
COOH	-3.02E+00	1.10E-04	2.41E+00	-3.97E-04	2.47E+00	4.89E+00	3.05E+02
$CSCH_3$	-5.80E-01	-8.52E-04	1.10E+00	-6.10E-04	-1.61E-01	1.19E+00	5.41E+02
CSH	5.21E-05	-1.01E-04	-1.43E+00	-1.92E-01	2.14E+00	2.18E+00	3.82E+02
MgH	-8.82E-07	-7.39E-05	2.21E+00	3.66E-06	-4.42E+00	4.42E+00	2.39E+02
NC	-1.07E-06	-1.63E-04	9.77E-01	7.95E-06	-1.95E+00	1.95E+00	2.74E+02
NCH_2	1.25E+00	-2.14E-04	-1.02E+00	-4.37E-05	-2.66E-01	1.35E+00	3.13E+02
$N(CH_3)_2$	1.49E+00	-6.96E-04	-2.04E+00	-2.69E-03	1.78E+00	2.22E+00	5.27E+02
NCO	3.02E+00	-1.23E-04	-4.14E+00	-1.04E-04	2.38E+00	4.16E+00	3.27E+02
NCS	-9.82E-01	-3.74E-04	-5.34E+00	-3.73E-05	2.39E+00	5.34E+00	4.54E+02
NF_2	6.96E-01	-1.32E-04	-2.44E-01	1.63E-04	-7.83E-01	1.07E+00	2.93E+02
NH_2	-1.26E-04	-3.81E-04	5.93E-01	8.54E-02	-2.03E+00	2.08E+00	2.09E+02
NH_3^+	-7.21E-01	-7.21E-01	5.32E-06	-7.21E-01	6.00E-06	1.13E-05	1.78E+02
$NHCH_3$	1.91E+00	2.20E-01	-1.47E+00	1.07E+00	1.86E+00	1.96E+00	3.67E+02
NHF	1.49E+00	-4.08E-01	-1.16E+00	-1.05E+00	1.45E+00	1.54E+00	2.49E+02
$NHNH_2$—ip	3.24E+00	2.14E-01	-1.19E+00	-9.25E-01	6.57E-02	1.33E+00	3.26E+02
$NHNH_2$—op	1.72E+00	3.88E-01	-1.02E+00	5.08E-01	4.69E-01	1.02E+00	3.28E+02
NHOH	-3.61E+00	4.04E-01	-1.64E+00	-1.02E+00	3.71E-01	1.72E+00	2.87E+02
NNH-trans	2.93E+00	-2.47E-05	-1.85E+00	4.68E-05	4.05E-01	1.94E+00	2.70E+02
NNH-cis	4.29E-06	-5.43E-04	-1.19E+00	-7.34E-01	6.31E-01	1.19E+00	2.70E+02

Table E9: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 9

Substituent	Q_{xy}(R)	Q_{xz}(R)	Q_{yy}(R)	Q_{yz}(R)	Q_{zz}(R)	$\|\mathbf{Q}(R)\|$	Vol(R)
NNN	-7.84E-01	-1.64E-04	-4.87E+00	4.89E-06	1.90E+00	4.91E+00	3.31E+02
NO_2	7.82E-05	-8.78E-04	-2.96E+00	-1.12E-04	8.46E-01	3.05E+00	3.07E+02
NO	-9.83E-01	1.52E-06	-1.30E+00	-7.16E-06	1.21E+00	1.45E+00	2.24E+02
OCH_3	-1.86E-01	7.67E-04	-2.13E+00	1.32E-04	6.32E-01	2.19E+00	3.35E+02
OCl	3.30E-01	-4.99E-04	-1.84E+00	1.05E-04	5.82E-01	1.88E+00	3.28E+02
OCN	1.53E-01	-1.26E-04	-6.51E+00	3.35E-05	3.01E+00	6.52E+00	3.39E+02
$OCOCH_3$	-2.78E+00	-6.72E-05	-9.51E-01	-1.63E-04	3.93E+00	4.10E+00	5.08E+02
OCOH	-2.92E+00	1.27E-04	2.45E+00	-4.28E-04	2.44E+00	4.89E+00	3.52E+02
OF	1.98E-01	-4.06E-05	-8.18E-01	1.36E-07	5.74E-02	9.13E-01	2.19E+02
OH_2^+	-2.47E-05	-7.12E-05	-5.20E-01	5.80E-01	-1.87E+00	2.52E+00	1.51E+02
OH	2.86E-05	2.44E-06	1.30E+00	6.65E-01	-8.65E-02	1.45E+00	1.73E+02
O-	7.06E-05	-7.24E-05	-1.02E+00	1.51E-05	2.04E+00	2.04E+00	2.36E+02
ONH_2	1.99E+00	3.48E-04	-2.04E+00	1.80E-04	1.77E+00	2.22E+00	2.97E+02
ONO	-1.82E+00	-1.08E-04	1.45E-02	9.86E-05	2.66E+00	3.08E+00	3.22E+02
OOH	1.45E+00	-7.62E-01	-1.46E+00	-1.03E+00	3.86E-03	1.68E+00	2.57E+02
$OO\cdot$	-4.04E-02	2.37E-04	-1.15E+00	-1.37E-05	8.17E-01	1.18E+00	2.29E+02
OPH_2O	-4.25E-02	1.92E-01	3.38E+00	-3.79E-01	4.77E+00	8.20E+00	4.70E+02
PH_2	1.85E-04	-5.91E-04	2.12E+00	-1.07E+00	-2.05E+00	2.41E+00	3.06E+02
PH_2O	-6.82E-04	3.89E-04	2.17E+00	2.38E+00	-2.09E+00	2.46E+00	3.50E+02
PHO(OH)	-4.91E+00	1.31E+00	5.36E+00	-3.66E+00	1.50E+00	7.21E+00	4.16E+02
$PO(OH)_2$	-7.93E+00	3.91E-03	-6.57E+00	-3.34E-03	3.78E+00	6.59E+00	4.81E+02
SCF_3	-2.47E+00	-2.25E-03	1.16E+00	1.91E-03	-2.61E+00	2.62E+00	5.34E+02
SCH_3	-1.86E+00	-2.45E-04	3.28E-01	-4.67E-04	-2.38E+00	2.58E+00	4.30E+02
SCN	-1.57E+00	6.00E-04	-3.79E+00	1.07E-04	3.65E-02	4.35E+00	4.33E+02
SH	-4.14E-05	1.11E-04	2.14E+00	-2.33E-01	7.49E-01	3.00E+00	2.77E+02
SHO_2	-1.45E-04	-4.17E-01	-6.31E+00	-4.00E-04	2.84E+00	6.32E+00	3.88E+02
SHO	1.45E+00	-3.77E-01	-2.38E+00	-4.53E-01	1.93E+00	2.53E+00	3.32E+02
$SiCl_2H$	7.26E-05	-4.42E+00	-5.17E+00	-7.39E-04	5.46E+00	6.14E+00	6.01E+02
$SiCl_3$	-1.06E-03	1.39E-03	-3.57E+00	1.13E-03	7.14E+00	7.14E+00	7.34E+02
$SiClH_2$	8.59E-04	-2.30E-04	-1.46E-01	-5.29E+00	3.88E+00	4.40E+00	4.66E+02
SiCl	-6.62E+00	-7.37E-05	2.05E+00	-5.95E-05	-9.19E-01	2.05E+00	4.25E+02
SiF_2H	-1.42E-03	-3.42E+00	-4.71E+00	-2.69E-03	2.89E+00	4.75E+00	3.66E+02
SiF_3	4.51E-03	9.53E-03	-3.47E+00	4.44E-03	6.93E+00	6.93E+00	3.86E+02
$SiFH_2$	2.71E-03	9.38E-04	2.04E+00	3.37E+00	-3.55E-01	2.17E+00	3.48E+02
SiF	4.35E+00	-8.90E-05	-1.61E+00	2.14E-04	1.09E+00	1.64E+00	3.06E+02
SiH_3	1.42E+00	1.42E+00	-5.60E-04	1.42E+00	1.08E-03	1.08E-03	3.33E+02
SiH	-1.88E-04	-1.68E-03	5.82E-01	-2.14E+00	-2.77E+00	2.92E+00	2.91E+02
S^-	-3.98E-05	-3.70E-05	-2.44E+00	-9.23E-05	4.87E+00	4.87E+00	3.48E+02
SSH	2.67E+00	2.11E-01	8.03E-01	4.17E-01	-5.25E-01	8.15E-01	4.68E+02

Table E10: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 10

Substituent	ρ_c	r(H-BCP)	$\nabla^2\rho_c$	$\lambda_{1,c}$	$\lambda_{2,c}$	$\lambda_{3,c}$	V_c
AlH_2	8.70E-02	1.49E+00	2.17E-01	-1.39E-01	-1.34E-01	4.89E-01	-1.27E-01
BeH	1.03E-01	1.42E+00	1.45E-01	-2.23E-01	-2.23E-01	5.91E-01	-1.50E-01
BH_2	1.93E-01	1.26E+00	-4.45E-01	-5.04E-01	-3.98E-01	4.57E-01	-3.53E-01
C_6H_5	2.96E-01	7.53E-01	-1.22E+00	-8.41E-01	-8.29E-01	4.52E-01	-4.00E-01
CCCl	3.01E-01	6.94E-01	-1.33E+00	-8.90E-01	-8.90E-01	4.53E-01	-4.21E-01
CCF	3.01E-01	6.94E-01	-1.32E+00	-8.85E-01	-8.85E-01	4.50E-01	-4.21E-01
CCH	3.02E-01	6.96E-01	-1.34E+00	-8.96E-01	-8.96E-01	4.55E-01	-4.20E-01
CCl_2F	3.15E-01	6.95E-01	-1.41E+00	-9.76E-01	-9.60E-01	5.26E-01	-4.21E-01
CCl_2H	3.06E-01	7.18E-01	-1.32E+00	-9.15E-01	-9.00E-01	4.96E-01	-4.11E-01
CCl_3	3.12E-01	6.96E-01	-1.39E+00	-9.52E-01	-9.52E-01	5.18E-01	-4.22E-01
$CClF_2$	3.17E-01	6.95E-01	-1.43E+00	-9.88E-01	-9.74E-01	5.32E-01	-4.20E-01
CClFH	3.09E-01	7.18E-01	-1.34E+00	-9.38E-01	-9.11E-01	5.05E-01	-4.10E-01
$CClH_2$	2.97E-01	7.44E-01	-1.23E+00	-8.61E-01	-8.38E-01	4.69E-01	-3.99E-01
CF_2H	3.11E-01	7.21E-01	-1.36E+00	-9.49E-01	-9.27E-01	5.12E-01	-4.08E-01
CF_3	3.19E-01	6.95E-01	-1.45E+00	-9.94E-01	-9.94E-01	5.38E-01	-4.20E-01
CFH_2	3.00E-01	7.47E-01	-1.25E+00	-8.87E-01	-8.42E-01	4.76E-01	-3.98E-01
$CH_2CH_2CH_3$—ip	2.87E-01	7.58E-01	-1.12E+00	-7.85E-01	-7.79E-01	4.40E-01	-3.87E-01
$CH_2CH_2CH_3$—op	2.86E-01	7.81E-01	-1.12E+00	-7.79E-01	-7.74E-01	4.39E-01	-3.85E-01
CH_2CH_3	2.87E-01	7.78E-01	-1.13E+00	-7.86E-01	-7.80E-01	4.41E-01	-3.87E-01
CH_2CN	2.90E-01	7.79E-01	-1.17E+00	-8.16E-01	-8.06E-01	4.54E-01	-3.88E-01
CH_2COCl—ip	2.92E-01	7.78E-01	-1.18E+00	-8.24E-01	-8.16E-01	4.55E-01	-3.92E-01
CH_2COCl—op	2.88E-01	7.52E-01	-1.15E+00	-8.04E-01	-7.93E-01	4.48E-01	-3.86E-01
CH_2COF—ip	2.94E-01	7.52E-01	-1.20E+00	-8.29E-01	-8.22E-01	4.55E-01	-3.96E-01
CH_2COF—op	2.87E-01	7.56E-01	-1.14E+00	-7.98E-01	-7.90E-01	4.46E-01	-3.86E-01
CH_2CHO—ip	2.91E-01	7.51E-01	-1.17E+00	-8.13E-01	-8.10E-01	4.51E-01	-3.91E-01
CH_2CHO—op	2.83E-01	7.58E-01	-1.10E+00	-7.74E-01	-7.67E-01	4.37E-01	-3.80E-01
CH_2COOH—ip	2.93E-01	7.56E-01	-1.19E+00	-8.26E-01	-8.19E-01	4.53E-01	-3.96E-01
CH_2COO——ip	2.86E-01	7.71E-01	-1.12E+00	-7.79E-01	-7.76E-01	4.37E-01	-3.86E-01
CH_2CSH—ip	2.92E-01	7.54E-01	-1.18E+00	-8.21E-01	-8.14E-01	4.55E-01	-3.92E-01
CH_2CSH—op	2.83E-01	7.81E-01	-1.10E+00	-7.75E-01	-7.68E-01	4.39E-01	-3.79E-01
CH_2NH_2—ip	2.92E-01	7.55E-01	-1.18E+00	-8.28E-01	-8.03E-01	4.55E-01	-3.92E-01
CH_2NH_2—op	2.88E-01	7.71E-01	-1.13E+00	-8.04E-01	-7.79E-01	4.52E-01	-3.82E-01
CH_2NO_2—ip	3.00E-01	7.67E-01	-1.26E+00	-8.80E-01	-8.56E-01	4.73E-01	-4.03E-01
CH_2NO_2—op	2.96E-01	7.78E-01	-1.23E+00	-8.56E-01	-8.34E-01	4.65E-01	-3.96E-01
CH_2OH—ip	2.98E-01	7.34E-01	-1.23E+00	-8.68E-01	-8.33E-01	4.68E-01	-3.98E-01
CH_2OH—op	2.93E-01	7.40E-01	-1.18E+00	-8.41E-01	-8.05E-01	4.63E-01	-3.88E-01
CH_2OOCH_3—ip	2.95E-01	7.53E-01	-1.21E+00	-8.51E-01	-8.22E-01	4.62E-01	-3.92E-01

Table E11: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 11

Substituent	ρ_c	r(H-BCP)	$\nabla^2\rho_c$	$\lambda_{1,c}$	$\lambda_{2,c}$	$\lambda_{3,c}$	V_c
CH_2OOCH_3—op	2.95E-01	7.65E-01	-1.20E+00	-8.54E-01	-8.21E-01	4.70E-01	-3.90E-01
CH_2SH—ip	2.91E-01	7.54E-01	-1.17E+00	-8.14E-01	-8.06E-01	4.53E-01	-3.91E-01
CH_2SH—op	2.93E-01	7.54E-01	-1.18E+00	-8.26E-01	-8.13E-01	4.57E-01	-3.94E-01
CH_3	2.87E-01	7.58E-01	-1.13E+00	-7.83E-01	-7.83E-01	4.37E-01	-3.89E-01
$CHCCH_2$	2.91E-01	7.56E-01	-1.18E+00	-8.25E-01	-8.07E-01	4.47E-01	-3.94E-01
CHCCO	2.95E-01	7.73E-01	-1.22E+00	-8.47E-01	-8.44E-01	4.68E-01	-3.93E-01
$CH(CH_2)_2$	2.95E-01	7.47E-01	-1.20E+00	-8.32E-01	-8.13E-01	4.48E-01	-4.03E-01
$CHCH_2$	2.95E-01	7.39E-01	-1.21E+00	-8.36E-01	-8.27E-01	4.50E-01	-3.99E-01
$CH(CH_3)_2$	2.88E-01	7.54E-01	-1.12E+00	-7.86E-01	-7.82E-01	4.43E-01	-3.87E-01
CHCO	2.87E-01	7.40E-01	-1.16E+00	-8.18E-01	-7.79E-01	4.36E-01	-3.96E-01
CHNH	2.93E-01	7.54E-01	-1.20E+00	-8.37E-01	-8.28E-01	4.63E-01	-3.86E-01
CHNN	2.94E-01	7.31E-01	-1.21E+00	-8.62E-01	-7.96E-01	4.48E-01	-4.06E-01
CN	3.01E-01	6.80E-01	-1.36E+00	-9.09E-01	-9.09E-01	4.63E-01	-4.13E-01
CNO	2.97E-01	6.71E-01	-1.31E+00	-8.82E-01	-8.82E-01	4.53E-01	-4.20E-01
CO	2.86E-01	5.78E-01	-1.47E+00	-9.86E-01	-9.86E-01	5.02E-01	-3.98E-01
$COCH_3$	2.88E-01	7.56E-01	-1.16E+00	-8.29E-01	-8.16E-01	4.80E-01	-3.67E-01
COCl	2.99E-01	7.12E-01	-1.29E+00	-9.02E-01	-8.83E-01	4.98E-01	-3.91E-01
COF	3.03E-01	7.08E-01	-1.32E+00	-9.30E-01	-8.96E-01	5.03E-01	-3.96E-01
COH	2.91E-01	7.49E-01	-1.19E+00	-8.44E-01	-8.31E-01	4.81E-01	-3.73E-01
CONH2	2.93E-01	7.46E-01	-1.21E+00	-8.56E-01	-8.33E-01	4.84E-01	-3.76E-01
$COOCH_3$	3.00E-01	7.26E-01	-1.28E+00	-9.03E-01	-8.71E-01	4.94E-01	-3.89E-01
COOH	3.01E-01	7.22E-01	-1.29E+00	-9.10E-01	-8.80E-01	4.97E-01	-3.91E-01
$CSCH_3$	2.96E-01	7.38E-01	-1.23E+00	-8.53E-01	-8.51E-01	4.77E-01	-3.91E-01
CSH	2.97E-01	7.33E-01	-1.25E+00	-8.61E-01	-8.60E-01	4.75E-01	-3.95E-01
MgH	5.70E-02	1.56E+00	2.01E-01	-7.75E-02	-7.75E-02	3.56E-01	-6.93E-02
NC	3.50E-01	5.10E-01	-2.23E+00	-1.56E+00	-1.56E+00	8.94E-01	-6.65E-01
NCH_2	3.50E-01	4.36E-01	-1.87E+00	-1.38E+00	-1.35E+00	8.68E-01	-5.97E-01
$N(CH_3)_2$	3.57E-01	5.12E-01	-1.90E+00	-1.42E+00	-1.36E+00	8.76E-01	-6.18E-01
NCO	3.44E-01	4.66E-01	-2.01E+00	-1.44E+00	-1.43E+00	8.60E-01	-6.28E-01
NCS	3.44E-01	4.60E-01	-2.04E+00	-1.45E+00	-1.45E+00	8.64E-01	-6.32E-01
NF_2	3.66E-01	4.80E-01	-2.17E+00	-1.58E+00	-1.54E+00	9.49E-01	-6.36E-01
NH_2	3.51E-01	5.07E-01	-1.87E+00	-1.39E+00	-1.35E+00	8.63E-01	-6.08E-01
NH_3^+	3.45E-01	4.59E-01	-2.10E+00	-1.49E+00	-1.49E+00	8.74E-01	-6.18E-01
$NHCH_3$	3.54E-01	5.09E-01	-1.89E+00	-1.41E+00	-1.35E+00	8.70E-01	-6.13E-01
NHF	3.59E-01	4.94E-01	-2.01E+00	-1.49E+00	-1.42E+00	9.05E-01	-6.20E-01
$NHNH_2$—ip	3.57E-01	5.09E-01	-1.91E+00	-1.43E+00	-1.36E+00	8.82E-01	-6.14E-01
$NHNH_2$—op	3.60E-01	4.99E-01	-1.98E+00	-1.48E+00	-1.40E+00	8.94E-01	-6.28E-01
NHOH	3.60E-01	4.99E-01	-1.99E+00	-1.48E+00	-1.41E+00	8.99E-01	-6.22E-01
NNH-trans	3.52E-01	5.25E-01	-1.89E+00	-1.39E+00	-1.38E+00	8.84E-01	-5.87E-01
NNH-cis	3.45E-01	5.15E-01	-1.79E+00	-1.34E+00	-1.33E+00	8.70E-01	-5.68E-01

Table E12: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 12

Substituent	ρ_c	r(H-BCP)	$\nabla^2\rho_c$	$\lambda_{1,c}$	$\lambda_{2,c}$	$\lambda_{3,c}$	V_c
NNN	3.42E-01	4.87E-01	-1.90E+00	-1.39E+00	-1.36E+00	8.54E-01	-6.01E-01
NO_2	3.45E-01	5.38E-01	-1.99E+00	-1.47E+00	-1.42E+00	9.06E-01	-5.88E-01
NO	3.35E-01	4.84E-01	-1.70E+00	-1.28E+00	-1.28E+00	8.65E-01	-5.31E-01
OCH_3	3.84E-01	3.70E-01	-2.86E+00	-2.05E+00	-2.00E+00	1.18E+00	-8.97E-01
OCl	3.77E-01	3.61E-01	-2.92E+00	-2.07E+00	-2.01E+00	1.15E+00	-8.96E-01
OCN	3.69E-01	3.48E-01	-3.02E+00	-2.07E+00	-2.04E+00	1.09E+00	-9.09E-01
$OCOCH_3$	3.73E-01	3.60E-01	-2.90E+00	-2.03E+00	-2.00E+00	1.13E+00	-8.87E-01
OCOH	3.71E-01	3.60E-01	-2.88E+00	-2.02E+00	-1.99E+00	1.12E+00	-8.81E-01
OF	3.79E-01	3.58E-01	-3.00E+00	-2.12E+00	-2.03E+00	1.15E+00	-9.04E-01
OH_2^+	3.48E-01	3.66E-01	-3.33E+00	-2.16E+00	-2.12E+00	9.46E-01	-9.42E-01
OH	3.80E-01	3.13E-01	-2.87E+00	-2.04E+00	-2.00E+00	1.17E+00	-8.96E-01
O-	3.70E-01	4.07E-01	-2.30E+00	-1.76E+00	-1.76E+00	1.21E+00	-7.98E-01
ONH_2	3.87E-01	3.63E-01	-2.99E+00	-2.12E+00	-2.05E+00	1.18E+00	-9.18E-01
ONO	3.76E-01	3.63E-01	-2.90E+00	-2.05E+00	-2.00E+00	1.15E+00	-8.87E-01
OOH	3.83E-01	3.63E-01	-2.97E+00	-2.11E+00	-2.03E+00	1.18E+00	-9.06E-01
$OO\cdot$	3.77E-01	3.65E-01	-2.91E+00	-2.06E+00	-2.01E+00	1.16E+00	-8.77E-01
OPH_2O	3.68E-01	3.54E-01	-2.92E+00	-2.02E+00	-2.00E+00	1.10E+00	-8.96E-01
PH_2	1.69E-01	1.36E+00	-1.62E-01	-2.79E-01	-2.56E-01	3.74E-01	-3.06E-01
PH_2O	1.82E-01	1.36E+00	-2.56E-01	-3.43E-01	-3.20E-01	4.06E-01	-3.27E-01
PHO(OH)	1.89E-01	1.35E+00	-2.95E-01	-3.62E-01	-3.53E-01	4.21E-01	-3.42E-01
$PO(OH)_2$	1.95E-01	1.34E+00	-3.19E-01	-3.76E-01	-3.76E-01	4.33E-01	-3.56E-01
SCF_3	2.25E-01	9.45E-01	-7.07E-01	-4.70E-01	-4.23E-01	1.86E-01	-2.87E-01
SCH_3	2.24E-01	9.69E-01	-6.98E-01	-4.60E-01	-4.07E-01	1.69E-01	-2.94E-01
SCN	2.25E-01	9.26E-01	-7.16E-01	-4.85E-01	-4.35E-01	2.04E-01	-2.81E-01
SH	2.23E-01	9.60E-01	-6.95E-01	-4.58E-01	-4.10E-01	1.73E-01	-2.91E-01
SHO_2	2.42E-01	9.39E-01	-8.52E-01	-5.60E-01	-5.56E-01	2.64E-01	-2.76E-01
SHO	2.22E-01	9.03E-01	-7.01E-01	-4.90E-01	-4.54E-01	2.44E-01	-2.55E-01
$SiCl_2H$	1.34E-01	1.48E+00	1.18E-01	-2.34E-01	-2.34E-01	5.86E-01	-2.29E-01
$SiCl_3$	1.38E-01	1.42E+00	1.08E-01	-2.42E-01	-2.42E-01	5.91E-01	-2.36E-01
$SiClH_2$	1.30E-01	1.41E+00	1.31E-01	-2.26E-01	-2.23E-01	5.80E-01	-2.21E-01
SiCl	1.22E-01	1.42E+00	7.42E-02	-1.99E-01	-1.93E-01	4.66E-01	-1.94E-01
SiF_2H	1.35E-01	1.50E+00	1.14E-01	-2.44E-01	-2.41E-01	6.00E-01	-2.30E-01
SiF_3	1.40E-01	1.42E+00	1.08E-01	-2.55E-01	-2.55E-01	6.19E-01	-2.40E-01
$SiFH_2$	1.31E-01	1.40E+00	1.29E-01	-2.34E-01	-2.25E-01	5.88E-01	-2.22E-01
SiF	1.20E-01	1.42E+00	5.94E-02	-1.97E-01	-1.95E-01	4.52E-01	-1.88E-01
SiH_3	1.26E-01	1.48E+00	1.45E-01	-2.14E-01	-2.14E-01	5.73E-01	-2.13E-01
SiH	1.19E-01	1.43E+00	9.66E-02	-1.96E-01	-1.76E-01	4.69E-01	-1.90E-01
S^-	2.09E-01	1.02E+00	-5.78E-01	-3.60E-01	-3.60E-01	1.42E-01	-2.94E-01
SSH	2.25E-01	9.46E-01	-7.06E-01	-4.76E-01	-4.24E-01	1.94E-01	-2.85E-01

Table E13: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 13

Substituent	G_c	H_c
AlH_2	9.05E-02	-3.63E-02
BeH	9.33E-02	-5.72E-02
BH_2	1.21E-01	-2.32E-01
C_6H_5	4.77E-02	-3.52E-01
CCCl	4.44E-02	-3.76E-01
CCF	4.56E-02	-3.76E-01
CCH	4.31E-02	-3.77E-01
CCl_2F	3.41E-02	-3.87E-01
CCl_2H	4.06E-02	-3.70E-01
CCl_3	3.74E-02	-3.84E-01
$CClF_2$	3.12E-02	-3.89E-01
CClFH	3.70E-02	-3.73E-01
$CClH_2$	4.57E-02	-3.53E-01
CF_2H	3.37E-02	-3.75E-01
CF_3	2.89E-02	-3.91E-01
CFH_2	4.25E-02	-3.56E-01
$CH_2CH_2CH_3$—ip	5.29E-02	-3.34E-01
$CH_2CH_2CH_3$—op	5.31E-02	-3.32E-01
CH_2CH_3	5.30E-02	-3.34E-01
CH_2CN	4.83E-02	-3.40E-01
CH_2COCl—ip	4.81E-02	-3.44E-01
CH_2COCl—op	4.95E-02	-3.37E-01
CH_2COF—ip	4.84E-02	-3.47E-01
CH_2COF—op	5.02E-02	-3.36E-01
CH_2CHO—ip	4.90E-02	-3.42E-01
CH_2CHO—op	5.22E-02	-3.28E-01
CH_2COOH—ip	4.91E-02	-3.47E-01
CH_2COO——ip	5.33E-02	-3.33E-01
CH_2CSH—ip	4.85E-02	-3.44E-01
CH_2CSH—op	5.15E-02	-3.27E-01
CH_2NH_2—ip	4.87E-02	-3.43E-01
CH_2NH_2—op	4.98E-02	-3.32E-01
CH_2NO_2—ip	4.35E-02	-3.60E-01
CH_2NO_2—op	4.50E-02	-3.51E-01
CH_2OH—ip	4.47E-02	-3.53E-01
CH_2OH—op	4.63E-02	-3.42E-01
CH_2OOCH_3—ip	4.47E-02	-3.47E-01

Table E14: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 14

Substituent	G_c	H_c
CH_2OOCH_3—op	4.43E-02	-3.46E-01
CH_2SH—ip	4.99E-02	-3.42E-01
CH_2SH—op	4.94E-02	-3.45E-01
CH_3	5.32E-02	-3.35E-01
$CHCCH_2$	4.87E-02	-3.45E-01
CHCCO	4.34E-02	-3.49E-01
$CH(CH_2)_2$	5.18E-02	-3.51E-01
$CHCH_2$	4.78E-02	-3.51E-01
$CH(CH_3)_2$	5.28E-02	-3.34E-01
CHCO	5.30E-02	-3.43E-01
CHNH	4.30E-02	-3.43E-01
CHNN	5.16E-02	-3.54E-01
CN	3.70E-02	-3.76E-01
CNO	4.58E-02	-3.74E-01
CO	1.51E-02	-3.83E-01
$COCH_3$	3.79E-02	-3.29E-01
COCl	3.46E-02	-3.56E-01
COF	3.24E-02	-3.63E-01
COH	3.73E-02	-3.36E-01
CONH2	3.72E-02	-3.39E-01
$COOCH_3$	3.45E-02	-3.54E-01
COOH	3.36E-02	-3.57E-01
$CSCH_3$	4.23E-02	-3.49E-01
CSH	4.21E-02	-3.53E-01
MgH	5.98E-02	-9.49E-03
NC	5.34E-02	-6.12E-01
NCH_2	6.52E-02	-5.32E-01
$N(CH_3)_2$	7.16E-02	-5.47E-01
NCO	6.31E-02	-5.65E-01
NCS	6.09E-02	-5.71E-01
NF_2	4.74E-02	-5.89E-01
NH_2	6.97E-02	-5.38E-01
NH_3^+	4.62E-02	-5.72E-01
$NHCH_3$	7.09E-02	-5.42E-01
NHF	5.86E-02	-5.62E-01
$NHNH_2$—ip	6.85E-02	-5.46E-01
$NHNH_2$—op	6.64E-02	-5.62E-01
NHOH	6.27E-02	-5.59E-01
NNH-trans	5.70E-02	-5.30E-01
NNH-cis	5.95E-02	-5.08E-01

Table E15: B2PLYPD3-BJ/aug-cc-pV5Z reference QTAIM properties in Gaussian09 standard orientation, part 15

Substituent	G_c	H_c
NNN	6.38E-02	-5.38E-01
NO_2	4.58E-02	-5.43E-01
NO	5.27E-02	-4.78E-01
OCH_3	9.12E-02	-8.06E-01
OCl	8.28E-02	-8.13E-01
OCN	7.74E-02	-8.32E-01
$OCOCH_3$	8.08E-02	-8.06E-01
OCOH	7.98E-02	-8.01E-01
OF	7.69E-02	-8.27E-01
OH_2^+	5.46E-02	-8.87E-01
OH	8.95E-02	-8.06E-01
O-	1.12E-01	-6.86E-01
ONH_2	8.56E-02	-8.33E-01
ONO	8.10E-02	-8.06E-01
OOH	8.21E-02	-8.24E-01
$OO\cdot$	7.45E-02	-8.02E-01
OPH_2O	8.35E-02	-8.13E-01
PH_2	1.33E-01	-1.73E-01
PH_2O	1.32E-01	-1.96E-01
PHO(OH)	1.34E-01	-2.08E-01
$PO(OH)_2$	1.38E-01	-2.18E-01
SCF_3	5.53E-02	-2.32E-01
SCH_3	5.98E-02	-2.34E-01
SCN	5.10E-02	-2.30E-01
SH	5.84E-02	-2.32E-01
SHO_2	3.13E-02	-2.44E-01
SHO	4.00E-02	-2.15E-01
$SiCl_2H$	1.29E-01	-9.96E-02
$SiCl_3$	1.31E-01	-1.04E-01
$SiClH_2$	1.27E-01	-9.42E-02
SiCl	1.06E-01	-8.76E-02
SiF_2H	1.29E-01	-1.01E-01
SiF_3	1.34E-01	-1.07E-01
$SiFH_2$	1.27E-01	-9.47E-02
SiF	1.01E-01	-8.66E-02
SiH_3	1.25E-01	-8.84E-02
SiH	1.07E-01	-8.32E-02
S^-	7.46E-02	-2.19E-01
SSH	5.42E-02	-2.31E-01