



EDGEWOOD CHEMICAL BIOLOGICAL CENTER

U.S. ARMY RESEARCH, DEVELOPMENT AND ENGINEERING COMMAND
Aberdeen Proving Ground, MD 21010-5424

ECBC-CR-172

AN ANALYSIS OF DESCRIPTORS OF VOLATILE ORGANIC COMPOUNDS AND THEIR IMPACT ON RATE CONSTANT FOR REACTION WITH HYDROXYL RADICALS

Alida L. Hartwell

EXCET, INC.
Edgewood, MD 21040-1055

Jerry B. Cabalo

RESEARCH AND TECHNOLOGY DIRECTORATE

May 2018

Approved for public release: distribution unlimited.



Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

Blank

PREFACE

The work described in this report was authorized under contract no. W911SR-10-D-0020-0007, with funding from DTRA. The work was started in June 2017 and completed in August 2017.

The use of either trade or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for public release.

Acknowledgments

The authors acknowledge the following individuals and institutions for their mentoring and assistance with the execution of this technical program:

- Dr. Craig Knox (U.S. Army Edgewood Chemical Biological Center [ECBC], Aberdeen Proving Ground, MD) and
- ECBC personnel.

Blank

CONTENTS

1.	INTRODUCTION	1
1.1	Motivation.....	1
1.2	Background.....	1
2.	METHODS	2
2.1	Overview.....	2
2.2	Gaussian 09 Calculations.....	3
3.	RESULTS AND DISCUSSION.....	4
3.1	Whole Molecule Descriptors	4
3.1.1	Dipole Moment	5
3.1.2	Polarizability.....	6
3.1.3	Molar Volume	7
3.1.4	Molecular Mass.....	7
3.2	Normal Mode Descriptors.....	8
3.2.1	All Compounds	9
3.2.2	Alcohols	11
3.2.3	Aldehydes	11
3.2.4	Alkanes	12
3.2.5	Alkenes	13
3.2.6	Alkyl Nitrates.....	14
3.2.7	Aromatic Hydrocarbons.....	15
3.2.8	Dicarbonyls and Unsaturated Carbonyls	16
3.2.9	Ethers	17
3.2.10	Hydroperoxides.....	18
3.2.11	Ketones	19
4.	CONCLUSIONS.....	20
	LITERATURE CITED.....	23
	ACRONYMS AND ABBREVIATIONS.....	25
	APPENDIX: PYTHON PARSER PROGRAM.....	27

FIGURES

1.	Procedure diagram showing the general workflow of the study	3
2.	Rate constant vs dipole moment.....	5
3.	Rate constant vs polarizability	6
4.	Rate constant vs molar volume	7
5.	Rate constant vs molecular mass.....	8
6.	Rate constant vs frequencies for all compounds, band 1	10
7.	Rate constant vs frequencies for all compounds, band 2	10
8.	Rate constant vs frequencies for alcohols	11
9.	Rate constant vs frequencies for aldehydes.....	12
10.	Rate constant vs frequencies for alkanes.....	13
11.	Rate constant vs frequencies for alkenes.....	14
12.	Rate constant vs frequencies for alkyl nitrates	15
13.	Rate constant vs frequencies for aromatic hydrocarbons.....	16
14.	Rate constant vs frequencies for dicarbonyls and unsaturated carbonyls	17
15.	Rate constant vs frequencies for ethers	18
16.	Rate constant vs frequencies for hydroperoxides.....	19
17.	Rate constant vs frequencies for ketones	20

TABLE

1.	Frequency Band Ranges	9
----	-----------------------------	---

AN ANALYSIS OF DESCRIPTORS OF VOLATILE ORGANIC COMPOUNDS AND THEIR IMPACT ON RATE CONSTANT FOR REACTION WITH HYDROXYL RADICALS

1. INTRODUCTION

1.1 Motivation

This technical report serves as an initial analysis of 245 volatile organic compounds (VOCs). The goal of the study was to understand how different descriptors of VOCs impact the rate constants of gas-phase reactions at 298 K with OH radicals and create a predictive model based on these correlations. Although that goal was not achieved in full, considerable progress has been made, and there is potential for a quantitative structure–activity relationship (QSAR) model to be constructed in the future. The progress made during this study and the future potential have implications for the Army, especially in the development of technology to identify chemicals in the field as well as the leveraging of existing databases of chemical infrared (IR) and Raman spectra.

1.2 Background

The fate of chemicals used by the Army on the environment affects the future use of Army installations. The fate of these chemicals also impacts the surrounding communities, especially in terms of ground water or air quality. For installations located outside of the United States, the environmental effect on local communities will greatly influence future use where a foreign government is involved. As a result, understanding the environmental effects of the chemicals used by the Army is of utmost importance.

However, there are significant technical barriers to achieving this understanding. There are large numbers of chemicals that are used within the U.S. Department of Defense that are difficult to characterize. Furthermore, there are many possible reactions with chemicals that are commonly found in the environment, which can form chemicals that are potential hazards. As a result, it is not feasible to measure all possible combinations of chemicals to accurately assess the hazard. A predictive capability study is necessary.

QSARs are well-documented in the literature (1–8), where some physical property, such as boiling point, scales linearly with some aspect or descriptor of a molecule. QSAR approaches are primarily used for two things: summarizing previously known or suspected relationships between physical properties and chemical descriptors of molecules and providing predictive capabilities for these relationships. In the context of this study, a QSAR approach could be used to make predictions about chemical reactivity.

Several similar studies have been conducted that associate the rate constants of some reaction with different molecular descriptors. Such studies have focused on reactions with various combinations of different chemicals and sample groups, reaction phases, atmospheric reactants, and so on (9–16). The particular combination of hydroxyl radicals and VOCs in this study has not been previously analyzed, but it is of particular interest to the Army. Hydroxyl

radicals are one of the most reactive species present in the environment. They primarily form through either the decomposition of hydroperoxides or the reaction of oxygen atoms with water molecules. Because hydroxyl radicals are commonly found in the environment and are highly reactive, chemical reactions with compounds of interest and these radicals are a priority; therefore, we focused on reactions between VOCs (17) of interest and these radicals.

2. METHODS

2.1 Overview

This study used a five-step workflow, as shown in Figure 1. The workflow used a small cluster/supercomputer to run the programs necessary for data collection and an existing program written in Python (Python Software Foundation; Beaverton, OR), which is named the Python Pipeline, to manage the programs used in the calculation process. Using experimental data for 245 VOCs (18) for comparison, we calculated various descriptors and analyzed their impacts on the rate constant. The following descriptors were calculated: dipole moment, polarizability, molar volume, molecular mass, and vibrational frequencies.

Before entering the cluster and running the pipeline, we collected the names, Simplified Molecular Input Line Entry System (SMILES) formulas, and rate constants for the collection of VOCs. We organized this data into a three-column input file that we then passed to the Python Pipeline program with a header describing what calculations Gaussian 09 should execute (Section 2.2). Using Open Babel software, the SMILES code was converted into a molecular structure in Cartesian coordinates and combined with the header to form a Gaussian 09 input file, and the calculations were run on the small cluster. After the Gaussian 09 calculations were run, the output files were input to a parser program, which was coded in Python3, to extract the desired information (see the appendix). The data collected by the extractor program was then put into the LibreCalc software and analyzed.

Figure 1 shows a visual explanation of this workflow.

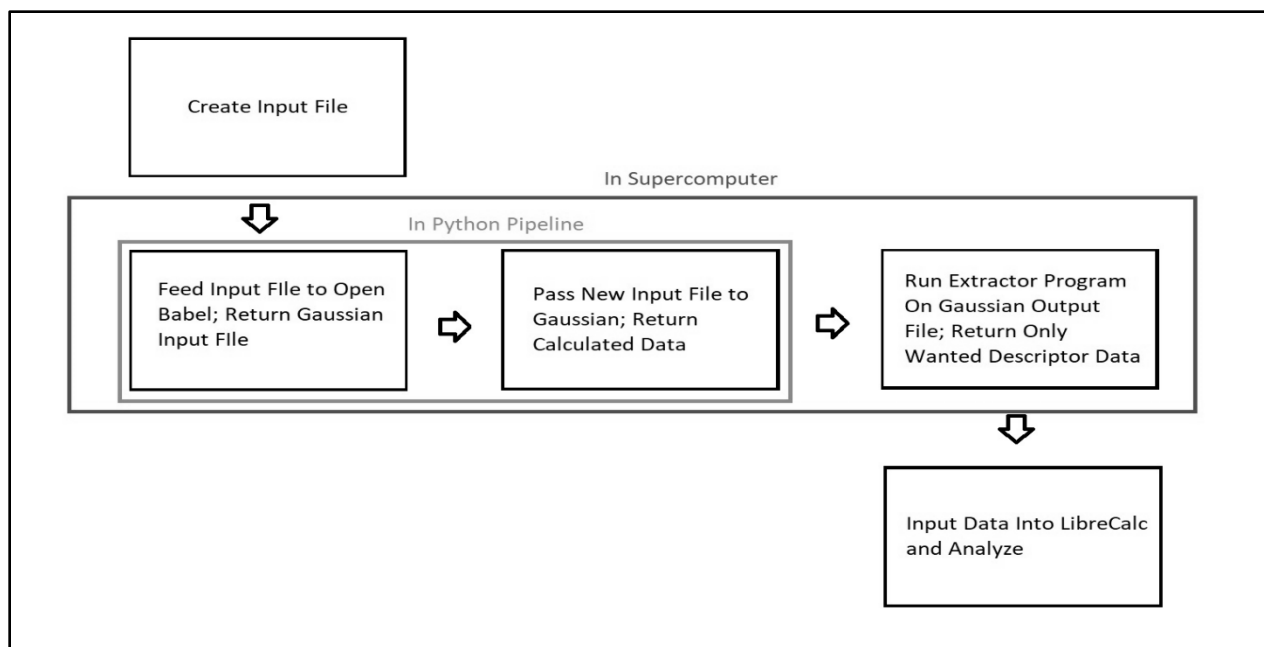


Figure 1. Procedure diagram showing the general workflow of the study.

2.2 Gaussian 09 Calculations

The main computing software used in this study was Gaussian 09 (18), which is a molecular modeling software suite that calculates electronic structure. As mentioned in Section 2.1, Gaussian 09 requires a header to describe what to calculate. The header includes three sections: method, basis set, and keyword(s). The header used for all calculations in this study was “HF/STO-3G opt freq polar volume”. In this case, the method is “HF”, the basis set is “STO-3G”, and the keywords are “opt freq polar volume”. Each of these is briefly explained below.

The Gaussian 09 software has several different methods available, but the method that we used for the calculations in this study was Hartree–Fock (HF). Each method is an approximation to the Schrödinger equation by making various assumptions about what is happening to or within a molecule. HF uses the Born–Oppenheimer approximation, which assumes that nuclei are static and that the motion of the electrons can be separated from the motion of the nuclei. An additional assumption simplifies an extremely complex equation for the motion of many electrons to a set of equations for single electrons. For each individual equation, the motion of a single electron is considered, and the rest are treated as a static average charge. The equations are solved iteratively, where the minimum energy is sought according to the variational principle, in which the most-correct wavefunction approximations have the lowest energies. If there are 10 electrons in a molecule, there are 30 degrees of freedom because each electron moves in the x , y , and z directions. Typically, this would mean that Gaussian 09 has to solve one equation with 30 variables; but using the HF approach, Gaussian 09 solves 10 three-variable equations instead.

Basis sets are families of functions that are chosen for mathematical simplicity. The STO-3G (three Gaussian orbitals fitted to a Slater-type orbital [STO]) basis set assigns three functions that are centered on each atomic nucleus. The Gaussian 09 software uses these functions to form linear combinations or weighted sums of the functions to create an approximate wavefunction for the electrons at a given geometry of the nuclei in a molecule.

Keywords are the different tasks or commands that the user enters into the Gaussian 09 software. In this case, the following four commands were provided to the Gaussian 09 software: (1) The first command “opt” is an abbreviation for “optimization” that specifies a geometry optimization in which the Gaussian 09 software calculates the geometry of a molecule that has the lowest energy. (2) The command “Freq” is short for “frequencies”, and it instructs the Gaussian 09 software to calculate the normal modes of vibrational motion and the vibrational frequencies of the molecule. (3) The “polar” keyword is used to calculate the polarizability of a molecule. (4) The final keyword “volume” instructs the program to compute the molar volume for the compound.

All of the descriptors were calculated as specified by the header and then extracted from the output using a separate program (see the appendix for the readme.txt) and compiled. We then searched for correlations between the calculated properties from theory and the experimental measurements of reaction rate.

3. RESULTS AND DISCUSSION

This study calculated five descriptors, which were grouped into the following two categories: whole molecule descriptors and normal mode descriptors. Four out of five descriptors (dipole moment, polarizability, molar volume, and molecular mass) fall under the whole molecule descriptor category. The other descriptor (frequencies/wavenumbers) is categorized as an individual bond descriptor. The descriptors were compared to experimental rate constants as discussed in Section 2.

3.1 Whole Molecule Descriptors

Most of the descriptors examined in this study described some trait of a whole molecule rather than a property unique to individual atoms or bonds. The four descriptors of this type are dipole moment, polarizability, molar volume, and molecular mass. Most of these descriptors had no direct correlation to the rate of reaction with hydroxyl radicals.

Polarizability was the only descriptor that may have presented a correlation. We believe the other descriptors in this category were not correlated because they are properties that are averaged over the whole molecule. Vulnerability to reaction with -OH occurs at individual atoms and is not predicted by properties of whole molecules unless those properties are indicative of traits of individual atoms within the molecule.

3.1.1 Dipole Moment

Permanent dipole moment refers to the net polarity of a molecule. Different things will affect the permanent dipole moment of a molecule. Electronegativity plays a large role, for example, as a more highly electronegative atom will pull electrons toward it. This gives the more-electronegative atom a slightly negative charge and a less-electronegative atom a slightly positive charge.

In Figure 2, the coefficient of determination (R^2) has a low value, and weak negative correlation is shown between the rate of reaction with OH radicals and the calculated dipole moment. However, for low dipole moments, there is great variability in reaction rate. We believe this is because a strong dipole moment indicates at least one highly electronegative atom. Hydroxyl radicals are not able to pull electrons away from highly electronegative atoms, so they usually cannot react. However, there is a pattern in which the rate constant is low (between 0 and 50×10^{12} cm³/molecule/s) past a dipole moment of 1.00 debye, indicating that there was little-to-no reaction with -OH radicals. At a dipole moment of less than 1.00 debye, the reaction rates are highly variable. This is more likely correlated to the electronegativity of atoms within the molecule and not to the overall dipole moment of a molecule. For low dipole moment molecules, factors other than dipole moment may have a greater effect on reaction rate and result in greater variability.

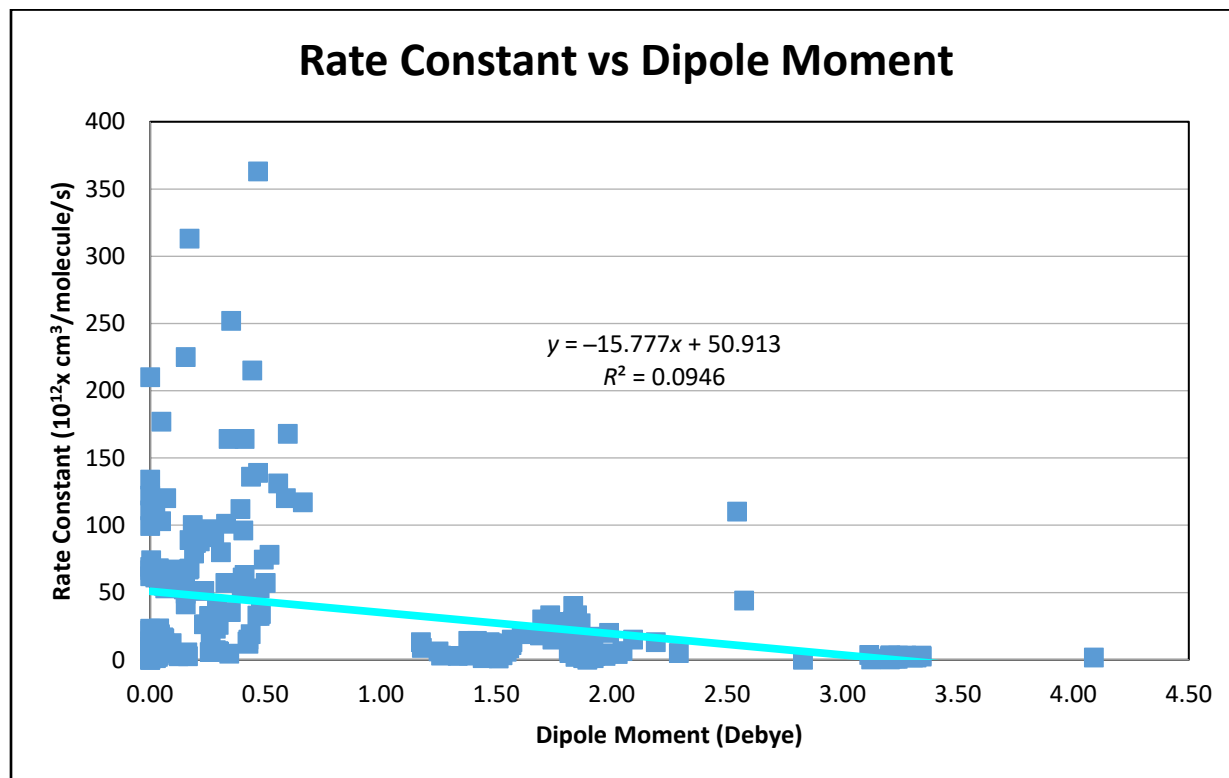


Figure 2. Rate constant vs dipole moment. Rate constants are variable with dipole moments <1.00 debye. Above 1.00 debye, they are fairly low relative to the maximum values.

3.1.2 Polarizability

Polarizability is the ability of a molecule to form an instantaneous dipole moment in response to an external electric field, such as electromagnetic radiation. In that case, electrons that are not tightly bound to the nuclei adjust to the external electric field. This is more likely to occur in molecules with greater electron densities, for example, as there are more electrons to shift.

Figure 3 shows a slight upward trend. The data around trend line do not correlate perfectly to it, so this effect would have to be examined using more compounds or concurrent correlations with other descriptors to determine a definitive cause. However, it does make sense that the data would have a positive correlation in this case. Molecules that have electrons moving around freely typically lack highly electronegative atoms that can pull electrons toward them and keep them close. Because of this, the hydroxyl radicals should be able to pull the loosely held electrons and react easily.

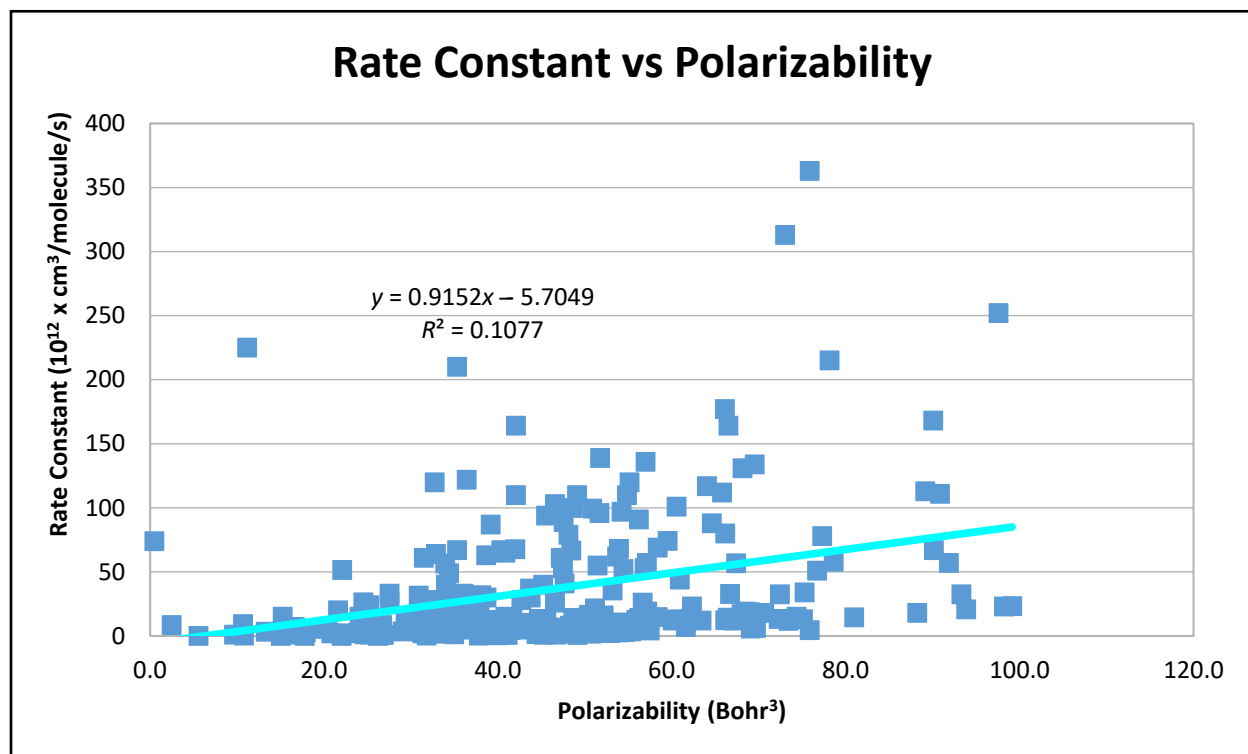


Figure 3. Rate constant vs polarizability. The graph shows an upward trend and some weak correlation.

3.1.3 Molar Volume

The Gaussian 09 software computes molar volume as the volume inside a contour of 0.001 electrons/bohr³. These data are returned in units of both cubic bohr per mole and cubic centimeters per mole. Data used in this study are in units of cubic centimeters per mole.

Figure 4 shows that the molar volume descriptor has no direct correlation and does not seem to be tied to any molecular property that contributed to the reaction rate with hydroxyl radicals.

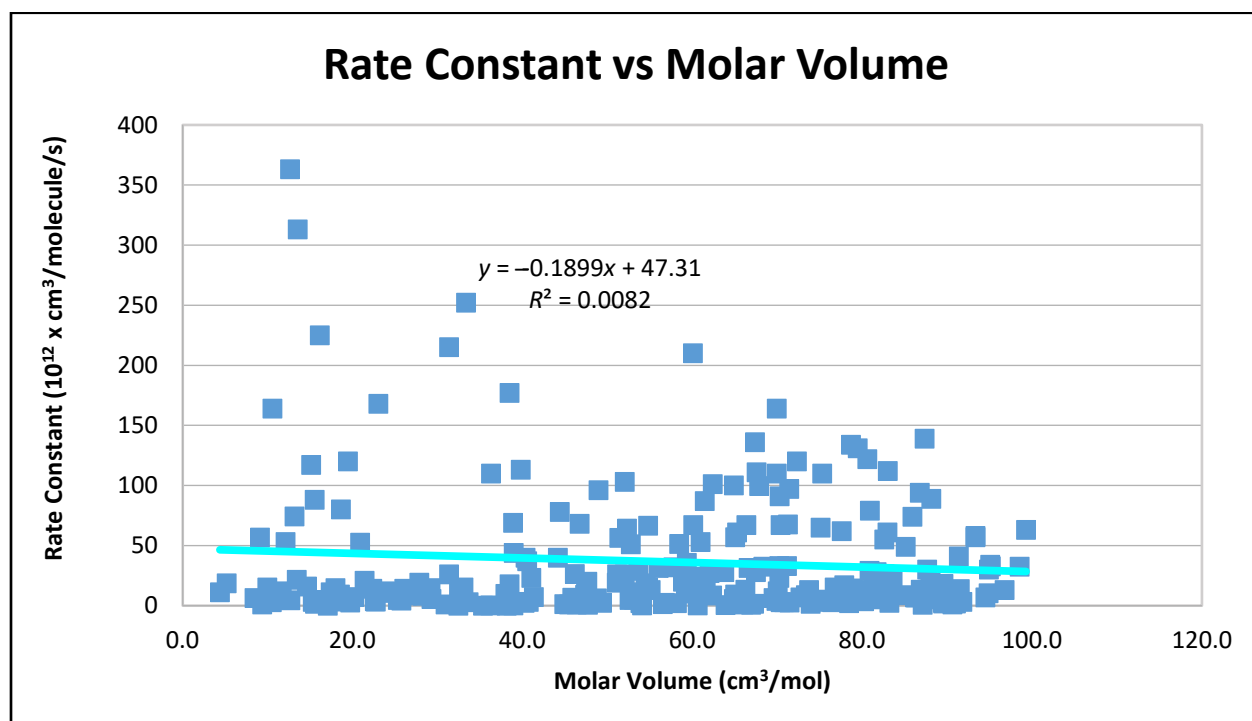


Figure 4. Rate constant vs molar volume. No correlation is present.

3.1.4 Molecular Mass

Molecular mass is calculated in atomic mass units and refers to the combined mass of all of the atoms in the molecule. Higher molecular mass could indicate more atoms with lower masses within a compound or fewer atoms with high masses in a compound. There appears to be no correlation, and an increase in molecular mass does not indicate a predictable change in rate constant.

No distinct correlation was indicated in the graph for rate constant versus molecular mass (Figure 5).

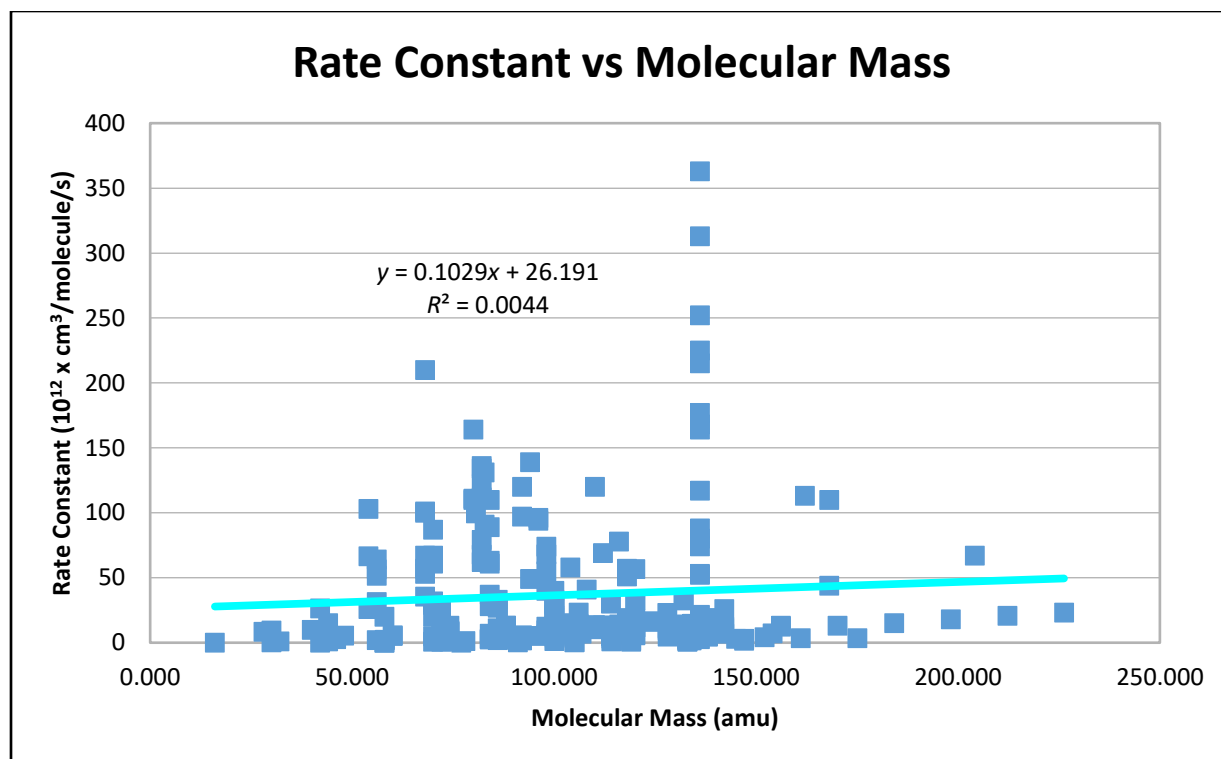


Figure 5. Rate constant vs molecular mass.

3.2 Normal Mode Descriptors

In this part of the study, the normal mode frequency was the only descriptor calculated that targeted specific bond motions. The Gaussian 09 software calculates several different frequencies for each molecule. It excludes six of the frequencies, which are the translational and rotational motions of the molecule in the x , y , and z directions. The remaining frequencies are those calculated when bonds bend, twist, and stretch or contract. Expressing this mathematically, we see that the Gaussian 09 software returns n frequencies for each compound, where

$$n = (3_{\text{frequency types}} \times \text{total number of atoms}) - 6_{\text{excluded frequencies}} \quad (1)$$

Methane, for example, has a central carbon atom and four hydrogen atoms bonded to it. The Gaussian 09 software returns nine frequencies for methane because $(3 \times 5) - 3 = 9$. Frequencies are given in units of wavenumbers or inverse centimeters.

To correlate multiple frequencies to one rate constant, we grouped them into bands by wavenumber and correlated the number of frequencies in each band to the rate constant. We grouped the frequencies in this way to correlate multiple frequencies to the single rate constant for each VOC. The bands used were indicative of alkanes or carbon–hydrogen bonds. We used these bands for multiple reasons as follows: First, we had reliable data about the bands that were indicative of alkanes (or C–H bonds). Second, most of the VOCs in the study contain C–H bonds, so we expected multiple results. The two bands used in the study are shown

in Table 1. The band ranges were expanded slightly on either end to account for small errors in calculation from the Gaussian 09 software.

Table 1. Frequency Band Ranges

Band Number	Frequency Band	
	Minimum (Wavenumbers)	Maximum (Wavenumbers)
1	1250	1550
2	2750	3050

The data collected was sorted and analyzed both overall and by compound type. The following compound groups were used: alcohols, aldehydes, alkanes, alkenes, alkyl nitrates, aromatic hydrocarbons, dicarbonyls and unsaturated carbonyls, ethers, hydroperoxides, and ketones.

3.2.1 All Compounds

Data for the comparisons of rate constants to numbers of frequencies in bands 1 and 2 for all VOCs in the study are shown in Figures 6 and 7. Overall, there was not a clear trend. When all functional groups were considered, there were multiple factors that contributed to the reaction rate. As a result, when only frequencies arising from the alkane vibrational modes were considered, there was not much correlation overall. Consideration of each functional group individually showed some interesting relationships between these vibrational frequencies, the presence of given functional groups, and the reaction rates.

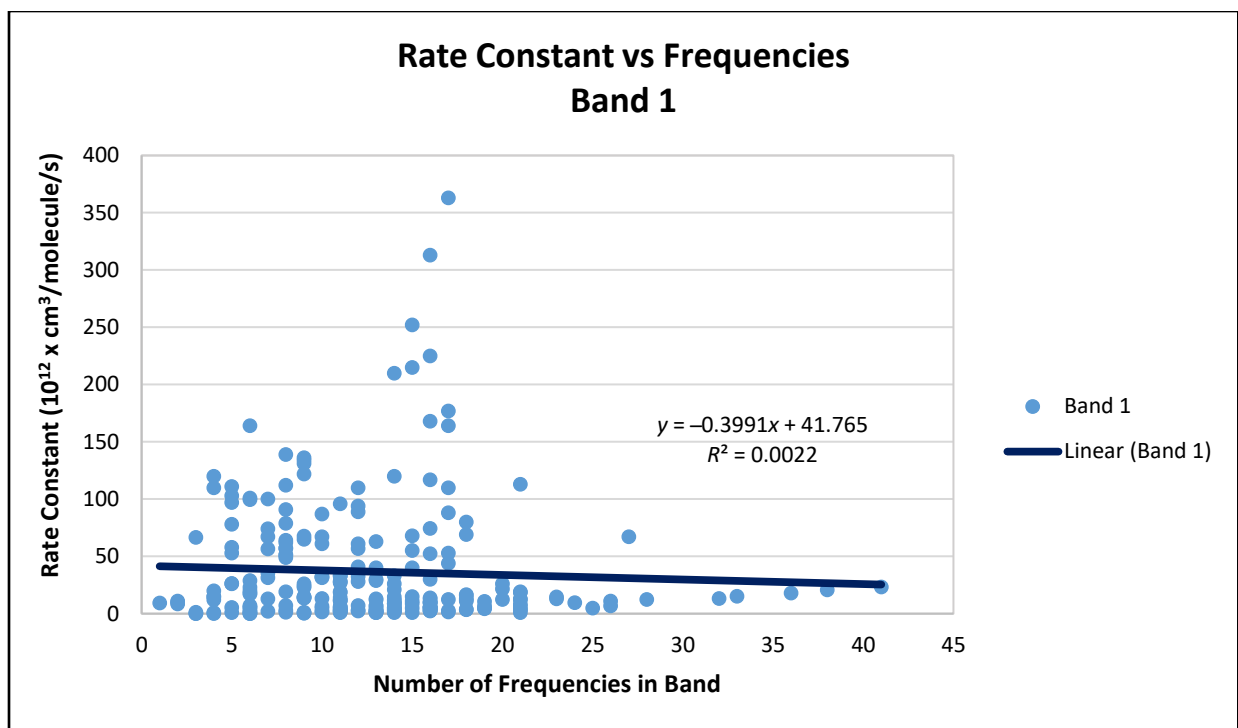


Figure 6. Rate constant vs frequencies for all compounds, band 1. A slightly negative slope on the trend line is shown for frequencies in band 1.

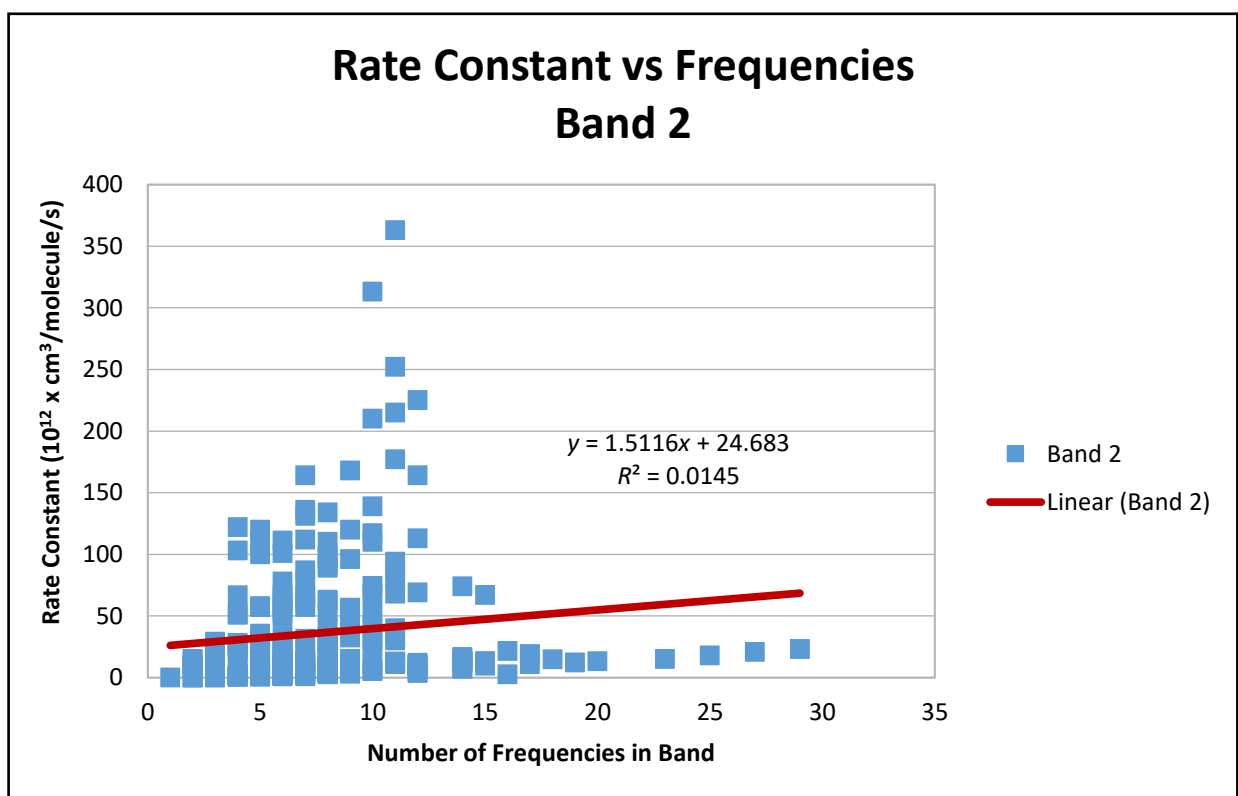


Figure 7. Rate constant vs frequencies for all compounds, band 2. A very small positive trend is shown for frequencies in band 2.

3.2.2 Alcohols

Alcohols are compounds with at least one –OH group bonded to a hydrocarbon. Hydrocarbons are made up of C–H bonds, so we would expect relatively high numbers of frequencies in both of the bands. Although alcohol functional groups are not likely to react with hydroxyl radicals, it is likely that as the molecule gets bigger and has more frequencies, there are more sites for the radical to attack. However, Figure 8 shows that the correlation for the alcohols is not a very strong predictor of reaction rate. Interestingly, the maximum of absolute rate constants in the chart is nearly the same value as the maximum rate constant in the alkane chart. The presence of the alcohol group undoubtedly introduces extra factors, in addition to dependence on the number of aliphatic carbons that remain to be understood. Most likely, this is due to the location of the alcohol group in the molecule.

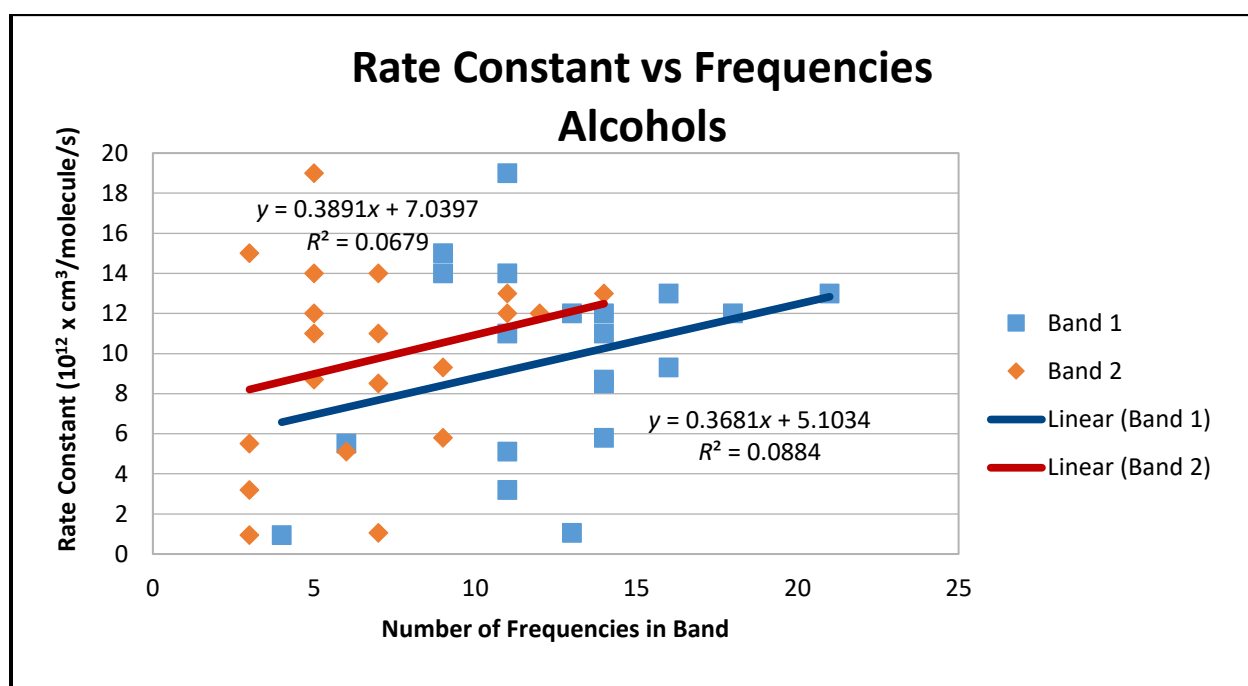


Figure 8. Rate constant vs frequencies for alcohols. Alcohols are fairly scattered, but the trend line still displays a positive slope.

3.2.3 Aldehydes

Aldehydes have a CHO structure in which the central carbon is double-bonded to the oxygen and single-bonded to both the hydrogen and some R group. Figure 9 also shows that this class of molecules has an ascending trend. Although the oxygen in the aldehyde is electronegative and might be considered electron density withdrawing, we see a definite correlation of the reaction rate with additional frequencies in bands 1 and 2 that indicates the presence of more carbon atoms. These carbon atoms also may provide additional sites for attack by the radical. Perhaps, because the aldehyde functional group is necessarily at the end of a carbon chain, the aldehyde group does not affect the ability of a radical to attack the rest of the

carbon chain. Given the increase in reaction rate relative to the alkanes, it is clear that the aldehyde group promotes a hydroxyl radical attack. A possible explanation may be that the radical site generated by hydrogen abstraction may transfer to the oxygen atom in the aldehyde.

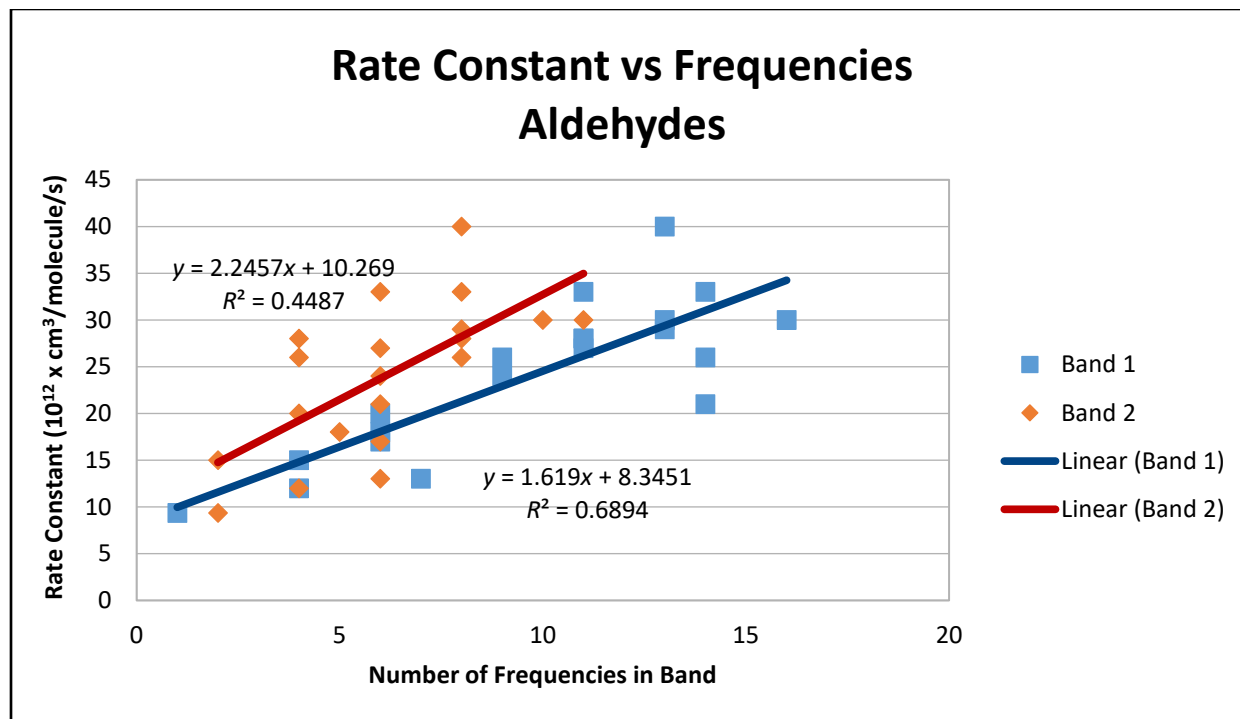


Figure 9. Rate constant vs frequencies for aldehydes. Aldehydes show a positive correlation.

3.2.4 Alkanes

Alkanes are saturated hydrocarbons, where all bonds between the carbon atoms are single bonds, and all carbon-bonding sites are either occupied by bonds to hydrogen atoms or to other carbon atoms. Because we used wavenumber ranges that are associated with alkanes, it was not surprising that we had large numbers of frequencies in the bands, and that there was a very clear correlation between alkane vibrational frequencies and the reaction rate, as shown in Figure 10.

For the pure alkanes without any other functional groups, there was a visible correlation between the reaction rate and the number of frequencies. There was no influence other than the presence of additional carbon atoms that could have affected the reaction rate.

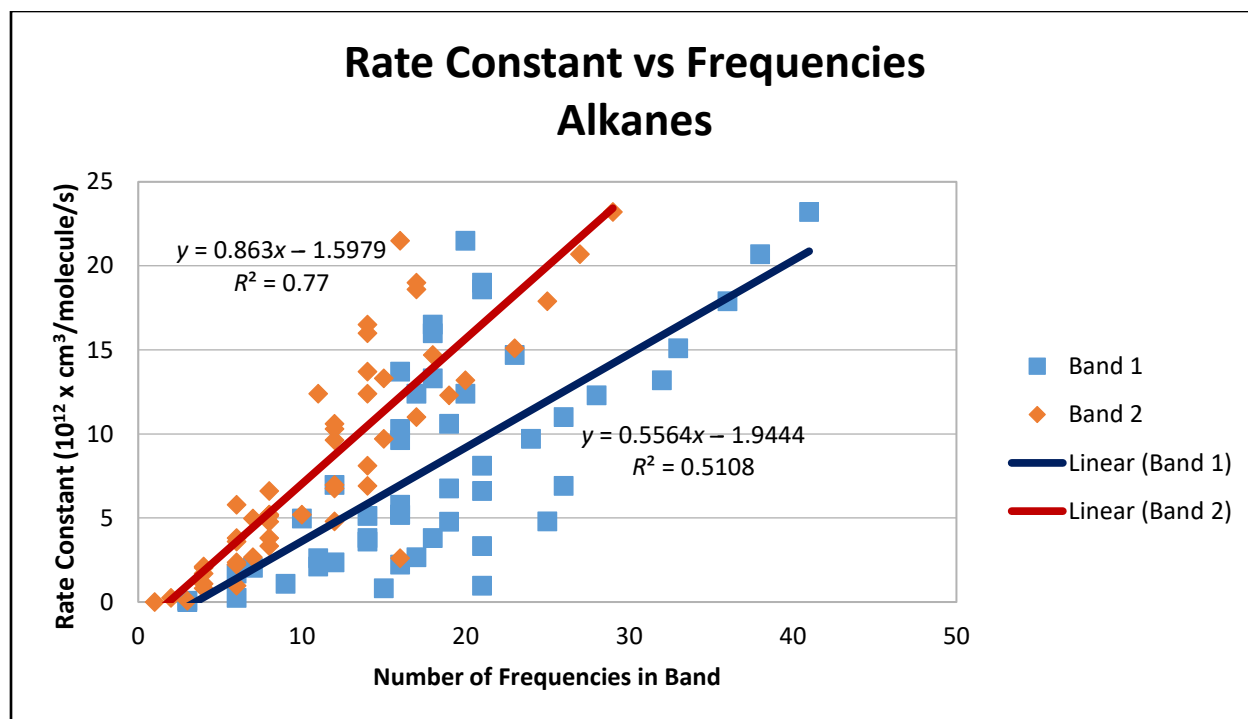


Figure 10. Rate constant vs frequencies for alkanes. Alkanes show an upward trend and strong positive correlation.

3.2.5 Alkenes

Alkenes are unsaturated hydrocarbons with at least one double bond between carbons. Not all possible bonding sites on the carbons are bonded to hydrogens in an alkene. As another type of hydrocarbon, it makes sense that there were fairly high numbers of frequencies in the bands. However, the dependence did not correlate as strongly as that of the alkanes, as shown in Figure 11. Because the alkene functional group is rich in electron density, it is a greater target for attack by the hydroxyl radicals. Thus, additional frequencies related to hydrocarbons do not greatly affect the reaction rate. The rate is dominated by the alkene groups, as illustrated by the fact that the rate constant absolute magnitudes are almost $10\times$ those of the alkanes. We expect that frequencies associated with the alkene groups should be more heavily weighted than those of groups associated with alkane groups.

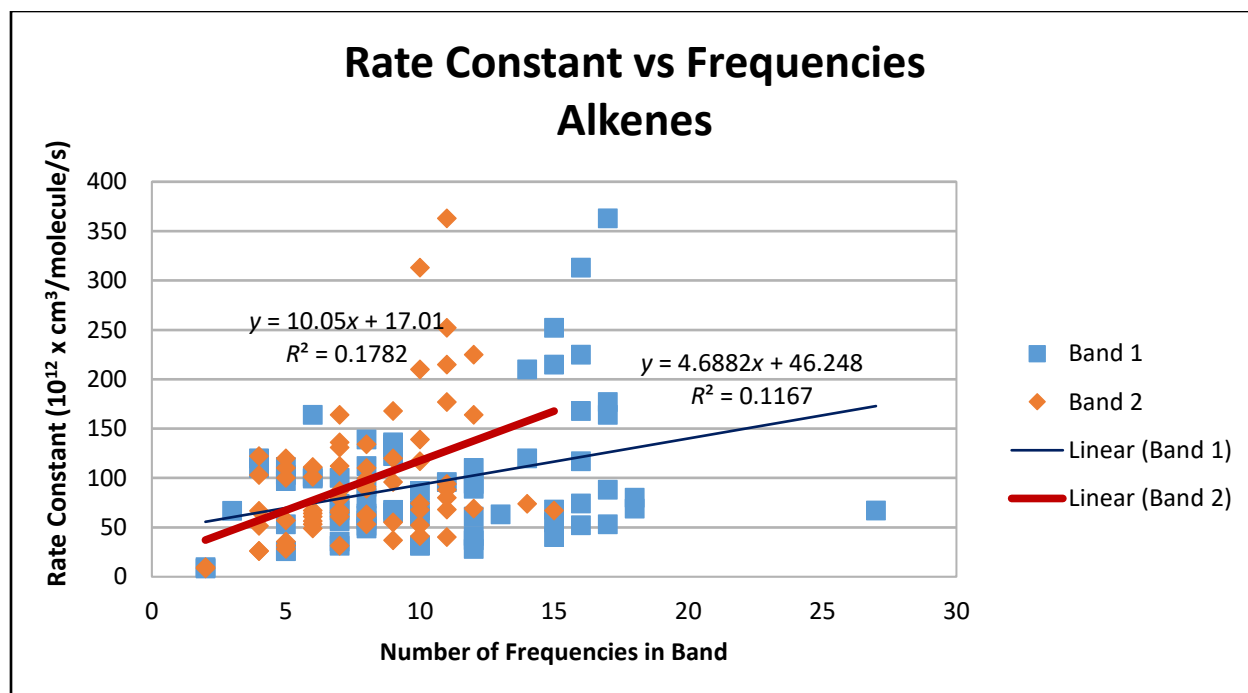


Figure 11. Rate constant vs frequencies for alkenes. Alkenes display a less uniform and pronounced upward trend than the other groups, but the data still show a positive correlation. We note that the reaction rates have a maxima almost an order of magnitude greater than that of the pure alkanes.

3.2.6 Alkyl Nitrates

Alkyl nitrates are composed of a nitrate (formula NO_3) with an R group (an alkyl) bonded to one of the oxygen molecules. The R group of these molecules contains C–H bonds, which is the reason that we see so many frequencies within the bands for this group.

As with the alkanes, the correlation was highly linear. Nitrate groups necessarily occur at the ends of carbon chains, and as with the aldehydes, they do not affect the increase in reactivity of the hydrocarbon part of the molecule, as shown in Figure 12. However, although there is a very strong positive correlation because there are more frequencies associated with alkanes, the absolute rate constant values were significantly reduced relative to those for the pure alkane molecules. We expected the nitrate group to function as an electron density withdrawing group. Additional aliphatic carbons increased the reaction rate, but the reaction barrier was still high, which resulted in an overall lower reaction rate. Thus, as the number of frequencies in the hydrocarbon chain increased with additional carbon atoms, the reaction rate increased linearly.

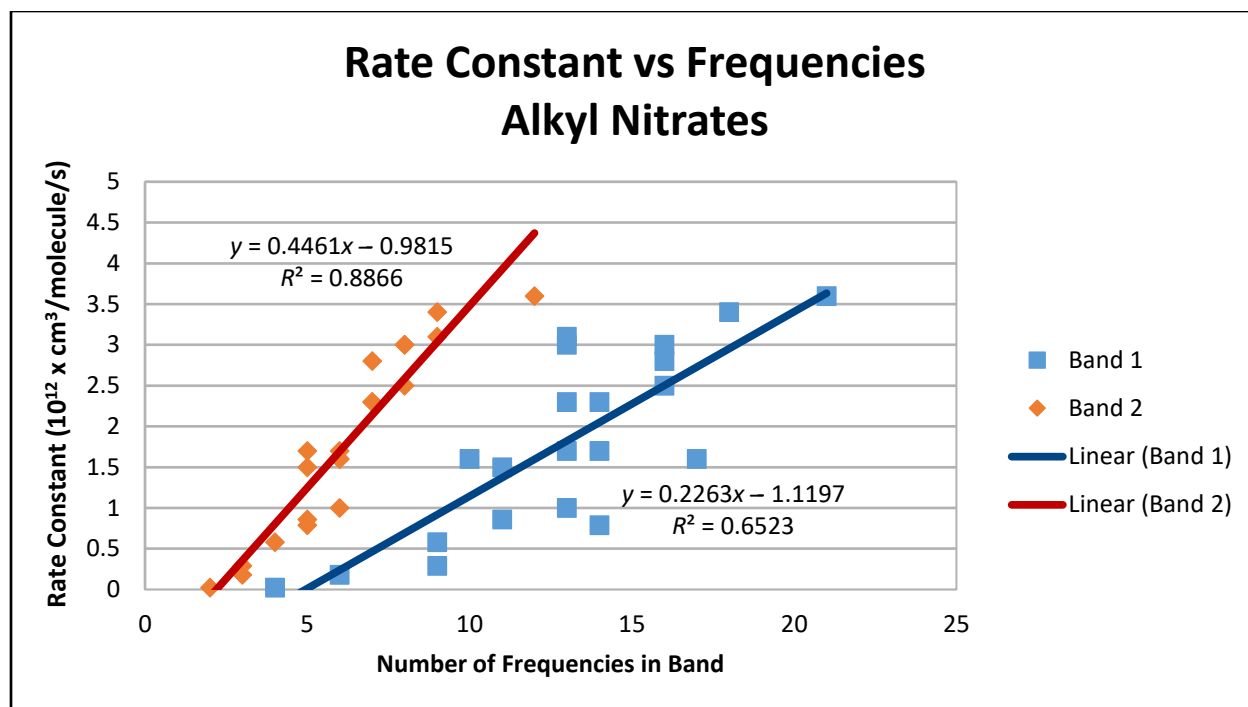


Figure 12. Rate constant vs frequencies for alkyl nitrates. Alkyl nitrates show a much greater correlation than the other molecular groups. The R^2 value for band 2 is one of the highest among all of the data analyzed.

3.2.7 Aromatic Hydrocarbons

Aromatic hydrocarbons are compounds that contain a benzene ring-shaped hydrocarbon in their structure. Again, this is a type of unsaturated hydrocarbon like the alkenes. Figure 13 shows a weak correlation between the number of vibrational frequencies in the bands and reaction rate, although there is a similar pattern with a positive slope.

As with alkenes, aromatic molecules have electron-rich regions above and below the aromatic ring. This may provide a greater target for attack by the hydroxyl radical than the monitored frequencies that are associated with saturated carbons. When predicting reaction rate, we concluded that the frequencies associated with the aromatic ring should be more heavily weighted than the frequencies associated with aliphatic carbon atoms. Additional frequencies from additional carbon atoms do not contribute to the reaction rate with the hydroxyl radicals because the contributions from the aromatic rings are much greater.

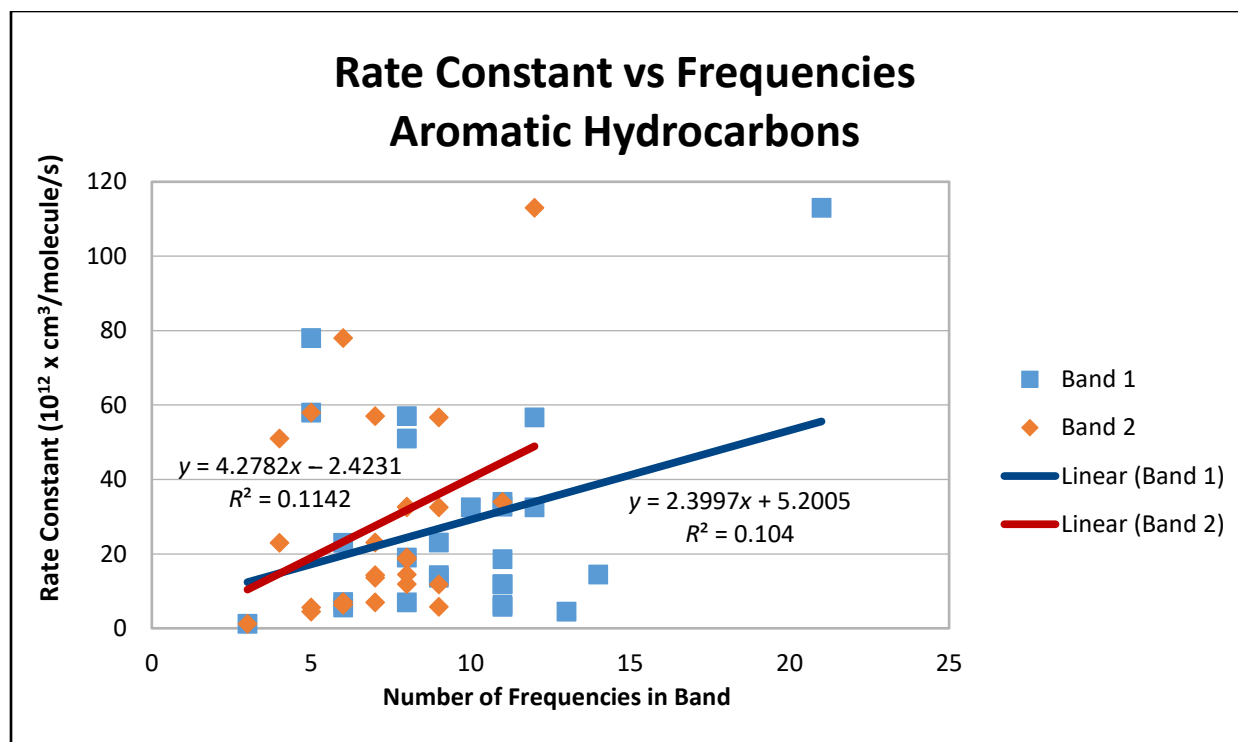


Figure 13. Rate constant vs frequencies for aromatic hydrocarbons. This group shows dramatic oscillations and more shallow slopes on the upward trend than did the other groups.

3.2.8 Dicarboxyls and Unsaturated Carbonyls

In general, carbonyls are any compounds containing a carbonyl group that is carbon double-bonded to oxygen. Dicarboxyls are compounds that contain two carbonyl groups in each molecule, and the term “unsaturated carbonyls” refers to carbonyls that are bound to an alkene.

Carbonyls have a large number of C–H bonds, as do the pure alkanes. Figure 14 also shows a strong correlation between the number of frequencies associated with aliphatic carbons and the reaction rate. However, the sample size for this compound group was relatively small and could be expanded in future research.

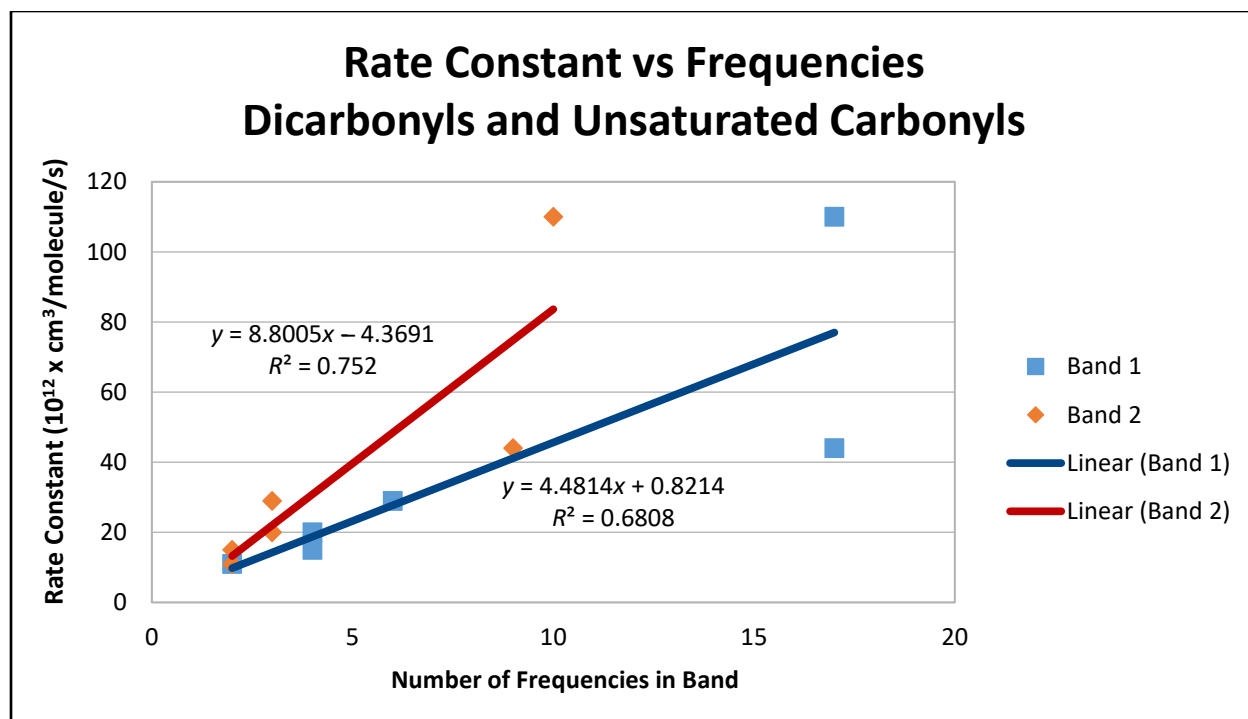


Figure 14. Rate constant vs frequencies for dicarbonyls and unsaturated carbonyls. This sample size is very small but does display a positive slope on both trend lines.

3.2.9 Ethers

Ethers are compounds with the general formula $R-O-R'$ where R and R' represent alkyl or aryl groups. This was also a small sample group, and the structure of ethers can vary widely with the different alkyl or aryl groups; therefore, the data could be investigated further as well.

However, this result was still interesting because as the numbers of alkane frequencies increase in the ether materials, the reaction rate does not really change, as shown in Figure 15. This indicates that the ether functional group counteracts the increase in reactivity to the hydroxyl radicals. Unlike the ketone group (Section 3.2.11), which are also electronegative, there is no dependence on reactivity as the number of aliphatic carbons increase. Without the electron-rich π bonds as in the aldehydes or ketones, when the radical abstracts a hydrogen atom leaving a radical site on the molecule, the ether prevents any stabilization of the radical site. It is also possible that the level of theory was inadequate to describe the system. Either the basis set was inadequate, or the HF treatment was insufficient. A density functional theory (DFT) or a post-HF method with a basis set consisting of more polarization functions may be necessary. Lastly, it is possible the data set is too small to draw strong conclusions.

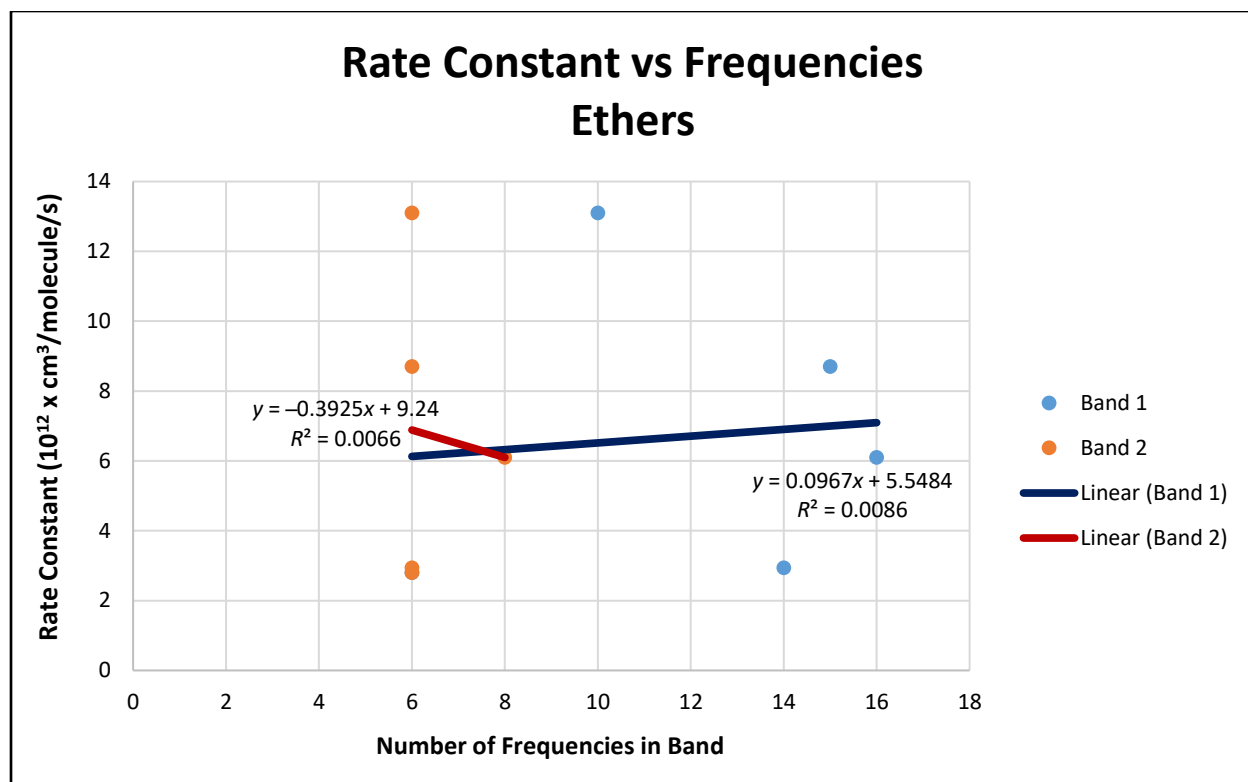


Figure 15. Rate constant vs frequencies for ethers. No correlation was noted for the compounds examined.

3.2.10 Hydroperoxides

Hydroperoxides are chemicals with the formula ROOR' , where R' is a hydrogen atom. This group included only two VOCs, which was not large enough to produce a usable trend line in Figure 16. However, the observed reaction rates were even lower for hydroperoxides than they were for ethers. The hydroperoxides undoubtedly acted in a way that was similar to ethers in impeding the radical chemistry on the hydrocarbon portion of the molecule.

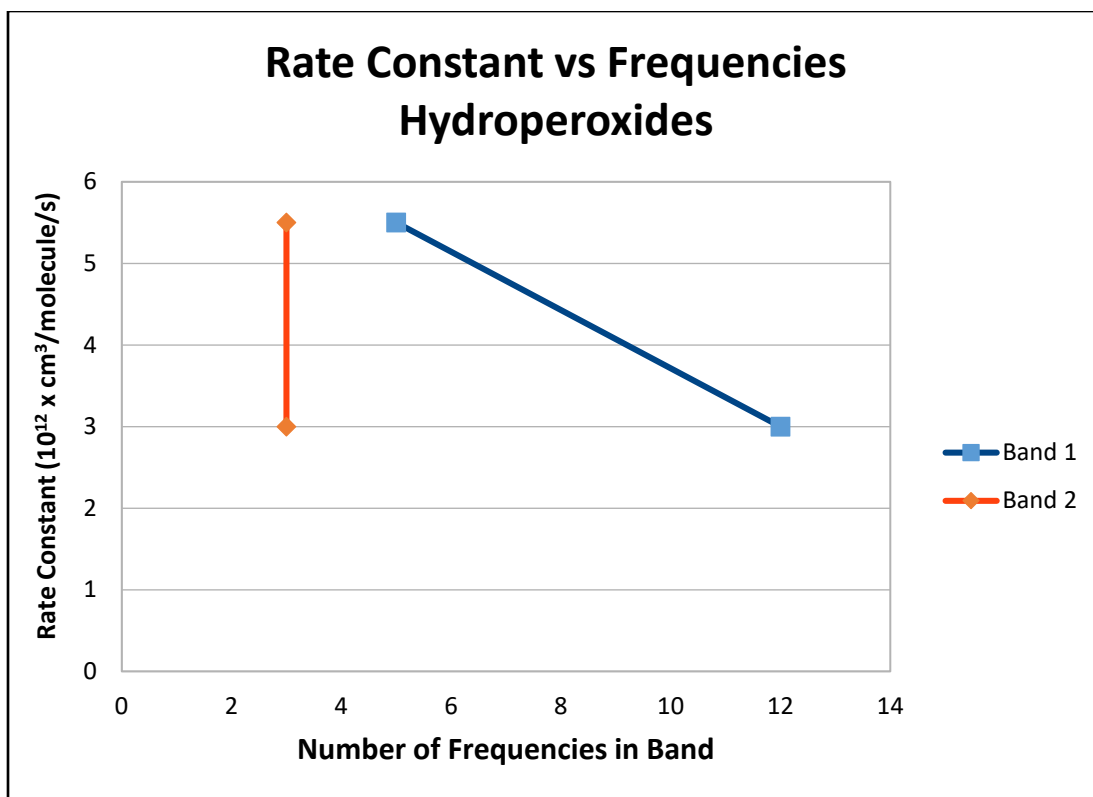


Figure 16. Rate constant vs frequencies for hydroperoxides.

3.2.11 Ketones

Ketones are compounds with the general formula $RC(=O)R'$, where R and R' may be a number of substituents that contain carbon. This group also displays a trend upwards, as shown in Figure 17. Although there was a definite correlation between the reaction rate and the number of frequencies associated with additional aliphatic carbons, we observed that the overall reaction rate of the ketone-containing compounds relative to the pure alkane compounds was about half of those tested. This indicates that the electron-withdrawing ketone functional group limits the availability of electrons that can bind with an approaching hydroxyl radical.

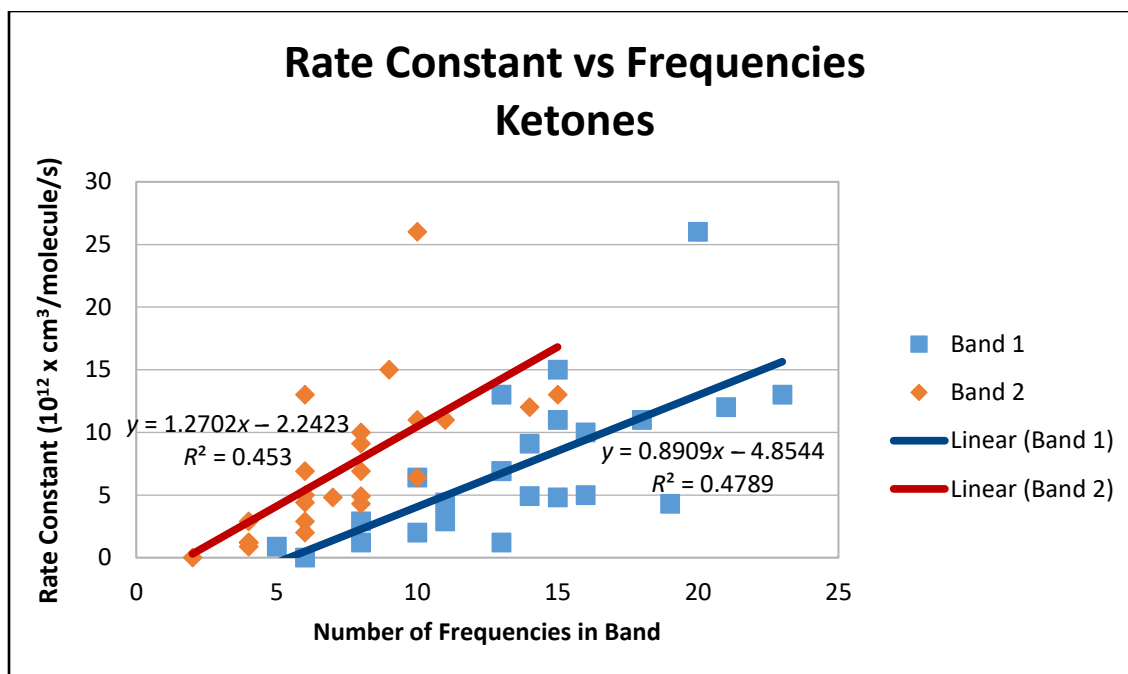


Figure 17. Rate constants vs frequencies for ketones.

4. CONCLUSIONS

Using the Gaussian 09 software, five descriptors for 245 volatile organic compounds were calculated. Those five descriptors were dipole moment, polarizability, molar volume, molecular mass, and bond frequencies (wavenumbers). All of the descriptors were categorized as whole molecule descriptors, except for frequencies, which was described as an individual bond descriptor. We then compared each of these descriptors to experimental rate constants for the reactions between the VOCs and hydroxyl radicals.

The resulting data led us to conclude that, for most of the whole molecule descriptors, there were no correlations to the rate constants. Polarizability was the one exception. The data for polarizability indicated a slight relationship with the rate constant, so this quantity did not help with predictability.

Data calculated and analyzed for the relationships between frequencies and rate constants were more promising. Overall, it did not appear that there was much correlation, but there appeared to be relationships when we examined each of the nine different groups of VOCs individually. Every group that had enough compounds to produce a significant trend line showed a positive linear correlation to the rate constant. We believe this ascending pattern is connected to the number of aliphatic carbons and is not necessarily linked to different functional groups. The aliphatic carbon idea is consistent, unless there are more electron-dense regions in a molecule or group of molecules that can take away from this effect, as in the aromatic hydrocarbons.

There are promising implications for the U.S. Army with further research on this topic, including the potential to identify compounds in the field by taking a measurement of frequencies or other descriptors, such as IR and Raman readings. Expanding on the information in this initial study could also produce a useful quantitative structure–activity relationship model for the prediction of rate constants or other descriptors that rely purely on the vibrational structure of a molecule.

Blank

LITERATURE CITED

1. Andrade-Ochoa, S.; Garcia-Machorro, J.; Bello, M.; Rodriguez-Valdez, L.M.; Flores-Sandoval, C.A.; Correa-Basurto, J. QSAR, DFT and Molecular Modeling Studies of Peptides from HIV-1 to Describe Their Recognition Properties by MHC-I. *J. Biomol. Struct. Dyn.* [Online]. DOI: 10.1080/07391102.2017.1352538. Published online: 3 August 2017. <https://www.tandfonline.com/doi/abs/10.1080/07391102.2017.1352538?journalCode=tbsd20> (accessed 11 April 2018).
2. Batista, V.; Ahmed, L. MoD-QM/MM Smells Good: Models of Olfactory Receptors Validated by Mutagenesis and Activity Profiles. Presented at the 251st American Chemical Society National Meeting and Exposition, San Diego, CA, 13–17 March, 2016.
3. Briard, J.G.; Fernandez, M.; De Luna, P.; Woo, T.K.; Ben, R.N. QSAR Accelerated Discovery of Potent Ice Recrystallization Inhibitors. *Sci. Rep.* **2016**, *6* (26403).
4. Dhail, S. QSAR Models for CDK/Cyclin Inhibitors: A Comparison between QM and Classical Descriptors. Presented at the 251st American Chemical Society National Meeting and Exposition, San Diego, CA, 13–17 March 2016.
5. Ginex, T.; Munoz-Muriedas, J.; Herrero, E.; Gilbert, E.; Cozzini, P.; Luque, F.J. Application of the Quantum Mechanical IEF/PCM-MST Hydrophobic Descriptors to Selectivity in Ligand Binding. *J. Mol. Model.* **2016**, *22* (6), 136.
6. Lukas, L.; Goss K.-U.; Endo, S. 3D-QSAR Predictions for α -Cyclodextrin Binding Constants Using Quantum Mechanically Based Descriptors. *Chemosphere* **2017**, *169*, 693–699.
7. Lukas, L.; Goss K.-U.; Endo, S. 3D-QSAR Predictions for Bovine Serum Albumin-Water Partition Coefficients of Organic Anions Using Quantum Mechanically Based Descriptors. *Environ. Sci. Proc. Impacts* **2017**, *19* (3), 261–269.
8. Toropova, A.P.; Schultz, T.W.; Toropov, A.A. Building Up a QSAR Model for Toxicity toward *Tetrahymena pyriformis* by the Monte Carlo Method: A Case of Benzene Derivatives. *Environ. Toxicol. Pharmacol.* **2016**, *42*, 135–145.
9. Fu, C.-w.; Lin, T.-H. Predicting the Metabolic Sites by Flavin-Containing Monooxygenase on Drug Molecules Using SVM Classification on Computed Quantum Mechanics and Circular Fingerprints Molecular Descriptors. *PLoS ONE* **2017**, *12* (1).
10. Jin, X.; Peldszus, S.; Huck, P.M. Predicting the Reaction Rate Constants of Micropollutants with Hydroxyl Radicals in Water Using QSPR Modeling. *Chemosphere* **2015**, *138*, 1–9.

11. Luo, S.; Wei, Z.; Spinney, R.; Yang, Z.; Chai, L.; Xiao, R. A Novel Model to Predict Gas-Phase Hydroxyl Radical Oxidation Kinetics of Polychlorinated Compounds. *Chemosphere* **2017**, *172*, 333–340.
12. Markelj, J.; Pompe, M. Modeling of Atmospheric OH Reaction Rates Using Newly Developed Variable Distance Weighted Zero Order Connectivity Index. *Atmos. Environ.* **2016**, *131*, 418–423.
13. Schindler, M. A QSAR for the Prediction of Rate Constants for the Reaction of VOCs with Nitrate Radicals. *Chemosphere* **2016**, *154*, 23–33.
14. Sudhakaran, S.; Amy, G.L. QSAR Models for Oxidation of Organic Micropollutants in Water Based on Ozone and Hydroxyl Radical Rate Constants and Their Chemical Classification. *Water Res.* **2013**, *47* (3), 1111–1122.
15. Wang, B.-c.; Qian, J.-z.; Fan, Y.; Tan, J. The QSAR Study of Flavonoid-Metal Complexes Scavenging •OH Free Radical. *J. Mol. Struct.* **2014**, *1075*, 204–212.
16. Yang, Z.; Luo, S.; Wei, Z.; Ye, T.; Spinney, R.; Chen, D.; Xiao, R. Rate Constants of Hydroxyl Radical Oxidation of Polychlorinated Biphenyls in the Gas Phase: A Single-Descriptor Based QSAR and DFT Study. *Environ. Pollut.* **2016**, *211*, 157–164.
17. Atkinson, R.; Arey, J. Atmospheric Degradation of Volatile Organic Compounds. *Chem. Rev.* **2003**, *103* (12), 4605–4638.
18. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, G.; Comperts, R.; Mennucci, B.; Hratchian, H.P.; Ortiz, J.V.; Izmaylov, A.F.; Sonnenberg, J.L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V.G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J.A., Jr.; 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.; Millam, J.M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J.W.; Martin, R.L.; Morokuma, K.; Farkas, O.; Foresman, J.B.; Fox, D.J. *Gaussian 09, Revision A.02*. Gaussian, Inc.: Wallingford, CT, 2009.

ACRONYMS AND ABBREVIATIONS

DFT	density functional theory
HF	Hartree–Fock
IR	infrared
QSAR	quantitative structure–activity relationship
R^2	coefficient of determination
SMILES	Simplified Molecular Input Line Entry System
STO	Slater-type orbital
VOC	volatile organic compound

Blank

APPENDIX PYTHON PARSER PROGRAM

Parser Program readme.txt

This program was coded in the Python3 language (Python Software Foundation; Beaverton, OR) to extract the desired data from the Gaussian 09 output files produced by the Python Pipeline. This is the readme.txt file that explains how to use the pipeline program.

This file contains instructions for the program *infoExtractor.py* and its partner programs.

Summary of each program:

makeList.py:

This program should be run BEFORE *infoExtractor.py*, if you are extracting from more than one file. It will create a separate input file that lists the relative pathway to each Gaussian 09 output file.

infoExtractor.py:

The main program. Handles command line input and does the actual info extraction.

orgLines.py:

This program was written after *infoExtractor.py* and makes the output file more visually appealing and easy to read.

extractorVars.py:

This program contains all of the descriptors that *infoExtractor.py* can handle and has functions to format each one.

Steps to using the program(s):

1. Create the input file
 - If you are extracting from a single Gaussian 09 output file, then you will use that as your input file for the *infoExtractor.py*.
 - If you are extracting from more than one file, then you will have to use *makeList.py*.

Directions for using *makeList.py*:

1. Locate the relative pathway to the files you want to extract from the extractor program.
 - The default pathway is: `"../*/*/job*.log"`.
 - If the default pathway will not collect your data, you may call another path by typing `"-i 'pathway/' "` into the command line when you run *makeList.py* (Note the quotes around the pathway; it must be a single string).
- b. Determine your output file name.
 - This can be anything and is required by the program. Call this with `"-o FileName.txt"`.

- c. If you are not using the default pathway, make sure the organize function does not run.
- d. If you need to group by certain numbers of jobs, use the `-n` call.
 - The default is 100, so *makeList.py* will output a file with 100 jobs and then start a new one.
 - If you want less or more than that add "`-n number`" to your command line entry (where number is an actual integer).
- e. After choosing all of these options, call *makeList.py*.
 - Here are examples of calls with the default pathway:
 1. `python3 makeList.py -o myOutFile.txt`
 2. `python3 makeList.py -o testout.csv -n 15`
 - Here are some examples of calls with a different pathway:
 1. `python3 makeList.py -i "../*" -o output.txt -or False`
 2. `python3 makeList.py -i "../test/*/*.csv" -o test.csv -or False -n 15`

2. Run *infoExtractor.py*.

- This program will run *orgLines.py* and *extractorVars.py* on its own.
- However, new descriptors must be added into *extractorVars.py*, if they are not included in this list:
 - Polarizability (P)
 - Molar Volume (M)
 - Molecular Mass (m)
 - Frequencies (F)
 - Note: Dipole moment is extracted and placed in a separate output file through *runsmiles.py* (the original pipeline program).
 - Adding a new descriptor to the list:
 1. Whenever a new descriptor is added, a function must be added as well to properly extract and format the data.
 2. To figure out the best way to extract, manually find the data in the Gaussian output file and look for specific characters, words, etc. that signal the data.
 3. Write the function to extract this data and format as desired.
- Selecting various options in *infoExtractor.py*:
 - `-d` or `--descriptor`:
 1. This option determines which descriptors will be extracted.
 2. Choices are anything from the list in *extractorVars.py*.
 3. Examples:
 - a. Enter one choice as: `-d M`.
 - b. Enter multiple choices as: `-d {M,P}`.
 - `-i` or `--input`:
 1. This option determines the input file.
 2. If not in current directory, you may enter a pathway.
 3. Examples:
 - a. Enter: `-i myInput.txt`
 - b. Enter: `-i ../Test/myInput.txt` (Do not enter as a string.)
 - `-o` or `--output`:

1. This option determines the output file.
 2. Follow same instructions as `-i`.
- `-n` or `--numFile`:
 1. This option is only used when the input file is the actual file to extract from and not a *makeList.py* file.
 2. Example:
 - a. Enter: `-n 1` (must type this exact thing).
 - `-c` or `--calc`:
 1. This option is used when the descriptor is something that needs to be calculated or collected multiple times.
 2. Currently, this only works for the frequency descriptor as it will collect all frequencies and determine how many are in the bands (given in *extractorVars.py*).
 3. To use this option, designate `-c` as true and `-d` as frequency.
 4. Example:
 - a. Enter: `-c True -d F`
 5. This is the only command of this type that currently works.
- Running *infoExtractor.py* in the command line:
 - Once the desired choices have been selected, simply call the program with all of the choices.
 - Examples:
 1. `python3 infoExtractor.py -i myInput.txt -o myOutput.txt -d {P,M,m,F}`.
 2. `python3 infoExtractor.py -i chemical_1.log -o chem1_out.csv -n 1 -c True -d F`.

DISTRIBUTION LIST

The following individuals and organizations were provided with one Adobe portable document format (pdf) electronic version of this report:

U.S. Army Edgewood Chemical
Biological Center (ECBC)
RDCB-DRI-T
ATTN: Cabalo, J.
Goad, A.

ECBC Technical Library
RDCB-DRB-BL
ATTN: Foppiano, S.
Stein, J.

Defense Threat Reduction Agency, R&D
J9-CBS
ATTN: Cronce, D.

Office of the Chief Counsel
AMSRD-CC
ATTN: Upchurch, V.

Department of Homeland Security
RDCB-PI-CSAC
ATTN: Mearns, H.

Defense Technical Information Center
ATTN: DTIC OA

G-3 History Office
U.S. Army RDECOM
ATTN: Smart, J.

ECBC Rock Island
RDCB-DES
ATTN: Lee, K.
RDCB-DEM
ATTN: Grodecki, J.

