

The University of Akron

IdeaExchange@UAkron

Williams Honors College, Honors Research
Projects

The Dr. Gary B. and Pamela S. Williams Honors
College

Spring 2023

Utilizing Computer-Aided Molecular Design with Azido-Ester Plasticizers

Justin Ebert
jme102@uakron.edu

Follow this and additional works at: https://ideaexchange.uakron.edu/honors_research_projects



Part of the [Other Chemical Engineering Commons](#), and the [Polymer Science Commons](#)

Please take a moment to share how this work helps you [through this survey](#). Your feedback will be important as we plan further development of our repository.

Recommended Citation

Ebert, Justin, "Utilizing Computer-Aided Molecular Design with Azido-Ester Plasticizers" (2023).
Williams Honors College, Honors Research Projects. 1659.
https://ideaexchange.uakron.edu/honors_research_projects/1659

This Dissertation/Thesis is brought to you for free and open access by The Dr. Gary B. and Pamela S. Williams Honors College at IdeaExchange@UAkron, the institutional repository of The University of Akron in Akron, Ohio, USA. It has been accepted for inclusion in Williams Honors College, Honors Research Projects by an authorized administrator of IdeaExchange@UAkron. For more information, please contact mjon@uakron.edu, uapress@uakron.edu.

Utilizing Computer-Aided Molecular Design with Azido-Ester Plasticizers

Justin Ebert

Department of Chemical, Biomolecular, & Corrosion
Engineering

Honors Research Project

Submitted to

*The Williams Honors College
The University of Akron*

Approved:

Donald Visco

4/15/23

Honors Project Sponsor (signed)

Dr. Donald P. Visco Jr.

Honors Project Sponsor (printed)

Date

Justin Khabaz

04/18/2023

Honors Project Reader (signed)

Dr. Fardin Khabaz

Honors Project Reader (printed)

Date

Ahmed Mohamed

Honors Project Reader (signed)

Ahmed Mohamed

04/18/2023

Honors Project Reader (printed)

Date

Accepted:

Donald Visco

4/15/23

Honors Faculty Advisor (signed)

Dr. Donald P. Visco Jr.

Honors Faculty Advisor (printed)

Date

Jie Zheng

04/17/2023

Department Chair (signed)

Dr. Jie Zheng

Department Chair (printed)

Date

Utilizing Computer-Aided Molecular Design with Azido-Ester Plasticizers

by

Justin Ebert

The University of Akron
Department of Chemical, Biomolecular, & Corrosion Engineering
Honors Research Project

Course: 4200-497

Recipient: The University of Akron Williams Honors College
Sponsored by: Dr. Donald P. Visco, Jr.

April 2023

Table of Contents

Cover Page	1
Table of Contents	2
Executive Summary	3
Introduction	6
Background	9
Methodology	11
Data and Results	12
Discussion	23
References	24
Appendix A	26
Appendix B	27

Executive Summary

Computer-Aided Molecular Design (CAMD) can be used in the engineering field to create new compounds from already existing compounds to optimize processes, increase sustainability, and improve their impact on the environment (Weis and Visco, 2009). The compounds being investigated in this research are energetic azido-ester plasticizers. Azido-ester plasticizers are used to modify the mechanical properties of propellant formulations (Zohari et al., 2017). The results from this research can be used to answer the following question: Can CAMD be used to design new, novel azido-ester plasticizers? The goal of this project is to deconstruct a set of azido-ester plasticizers into smaller fragments (called atomic Signatures), recombine the fragments in ways that satisfy various valence constraints, and then evaluate the newly designed compounds against various properties of interest (Weis and Visco, 2009; Visco and Krew, 2019).

The fitness functions used to evaluate the newly designed compounds were developed *via* a forward-stepping multiple linear regression approach. The equations, termed quantitative structure-property relationships, QSPRs, were based on decomposition temperature, T_D , heat of formation, ΔH°_f , glass transition temperature, T_G , and molecular weight, MW (Weis and Visco, 2009; Visco and Krew, 2019). Three QSPRs were developed in this study, one for each T_D , ΔH°_f , and T_G , where the coefficient of determination, R^2 , values were 0.984, 0.972, and 0.988, respectively, and the number of atomic Signatures used were 9, 9, and 8, respectively. The inverse design problem resulted in more than 2.5 million solutions. In order to screen the newly designed compounds for fitness, the following property ranges were used as filters:

QSPR filtering ranges to obtain desired azido-ester plasticizer molecular formulas.

T_D [°C]	ΔH°_f [kJ.mol ⁻¹]	T_G [°C]	MW [g.mol ⁻¹]
242 to 317	550 to 1400	-103 to -73	475 to 550

The structures were filtered to have a high positive decomposition temperature, high positive heat of formation, low glass transition temperature, and high molecular weight as this is desired for azido-ester plasticizers (Alexis, 2004; Schock and Bräse, 2020; Sheibani, 2019; Sheibani et. al., 2020; Zohari et. al., 2017). After these filters were applied, there were 512 potential solutions that satisfied these four ranges. Of the 512 potential solutions, 12 created structures that were feasible (*i.e.*, had connected graphs and satisfied aromatic constraints, such as Huckel's rule) (LibreTexts). One structure from each of the 12 were evaluated, though many isomers can exist, per structure.

As a way to quantify the novelty of each structure created through this approach, the Tanimoto Coefficient can be calculated with respect to each training set structure. The Tanimoto Coefficient quantifies the similarity between two structures by calculating a ratio of the number of similar structure features to the total number of features present (CCDC). The design structures have a minimum, average, and maximum coefficient range of 0.2407 to 0.3137, 0.4525 to 0.5167, and 0.6296 to 0.7547, respectively. Since a coefficient of 0.85 or greater signifies high similarity, it can be concluded that these 12 generated structures are novel and could potentially possess the optimal properties for azido-ester plasticizers in the propellant and explosive category due to the coefficients being significantly lower than 0.85 (Patterson et. al., 1996, p. 3057).

A chemical structure search was done in chemical databases and no designed structure exists. Since no structure exists, the next steps that should be taken for this project would be to synthesize and test these structures in a lab. The property values obtained from testing can then be compared to the QSPR predicted values. At this point, CAMD was successful in the design of new, novel azido-ester plasticizers, but next steps should be taken to validate the property

predictions through experimentation and testing. However, other properties that could potentially provide insight on the quality of these compounds are viscosity and moduli. During this research, these properties were omitted as data was not available for them to derive QSPRs. Instead of using experimental data to establish a QSPR for viscosity, it is possible to use group contribution with the Joback method; however, this was not investigated during the project since it is not utilizing CAMD.

Throughout this research, there were many technical and career skills obtained that have been applied to past internship experiences but can also be used in the future. Of these skills, the important things learned were proper scholarly research on polymers and their applications, investigating significant physical and chemical properties of these polymers, computer-aided molecular design using atomic Signature and quantitative structure-property relationships (which emphasized the application of multiple regression), and communication with a supervisor to guide the project being completed while allowing independent work. The results of this project could potentially supply the propellant formulation industry with azido-ester plasticizers that are more energetic, green, and have the more desired property values for T_D , ΔH°_f , T_G , and MW.

There are two recommendations that I would provide to other students that will work on a project of this type: 1) Begin CAMD work early before starting the official Honors Project and 2) Develop a background in coding, whether it be Python or C++. With this project in particular, Fortran 95 was used.

Introduction

Engineers, especially chemical engineers, are increasingly focused on optimizing processes and to create sustainable and environmentally friendly products (Weis and Visco, 2009). One way to do this is to design new chemical compounds. This can be done by using Computer-Aided Molecular Design (CAMD) (Weis and Visco, 2009). Instead of synthesizing and testing random chemical compounds, CAMD can create optimized compounds for specific applications. This will also assist in saved costs from testing insufficient compounds since the compound properties are unknown when synthesized. CAMD can take into consideration many physical and chemical properties of compounds, which are what engineers are mainly interested in when utilizing specific compounds for certain processes. A tool that can be used in CAMD is the Signature molecular descriptor (Weis and Visco, 2009; Visco and Krew, 2019). Signature can be described as molecular or atomic where the molecular Signature is the combination of all atomic Signatures present in a structure and uniquely identifies the structure and the atomic Signature shows the root atom and the atoms and bonds it is connected to (Weis and Visco, 2009; Visco and Krew, 2019). It can be used at different heights of 0, 1, or 2 where the height is how many atoms away from the root atom the Signature represents. “A height-0 atomic Signature is just the individual atom itself. A height-1 Signature includes the atom and its nearest neighboring atoms. A height-2 atomic Signature is the atom, its nearest and second-nearest neighbors (without backtracking)” (Visco and Krew, 2019, p. 231). An example of deriving the two unique atomic Signatures and a molecular Signature of a structure is shown in **Figure 1**.

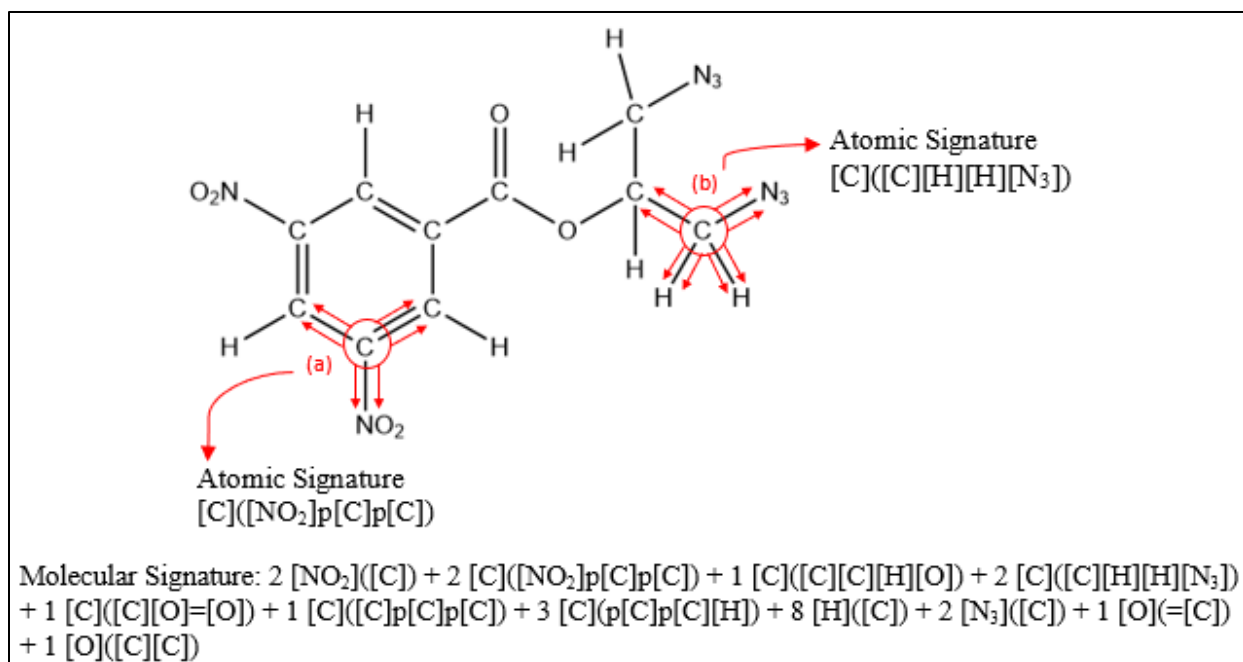


Figure 1: Above is a depiction on how to establish atomic and molecular Signatures for Compound 1 of **Table 1**. The notation for each atomic Signature is important when using inverse design. Symbols = and p represent the succeeding atom is connected to the root atom by a double bond and aromatic atom, respectively. The coefficient on the atomic Signature indicates the occurrence of that specific atomic Signature within the structure.

In **Figure (a)** the root atom is carbon (circled), C, which is bonded to two other carbons within an aromatic ring, pC (where “p” indicates the succeeding atom is part of an aromatic ring), and a nitro group, NO₂. Normally, the nitro group would be expanded and not condensed into NO₂ when creating an atomic Signature, however, for this specific project, NO₂ and N₃ were abbreviated for simplicity within the code by using halogens not used in any training set structures and then reverted back to their original functional groups upon inverse design. The atomic Signature at (a) can then be shown as [C]([NO₂)p[C]p[C]). **Figure 1 (b)** is another example for obtaining the atomic Signature at a given root atom. The molecular Signature is shown in **Figure 1** which is the sum of all unique atomic Signatures with their occurrence numbers and uniquely represents this structure from the other compounds used in the project.

Each atomic Signature has an impact on a compound's properties, which is why breaking a compound down to its atomic level will help in the design of compounds with desired qualities.

The goal of this project is to utilize these Signatures to design azido-ester plasticizers that will have the desired physical and chemical properties for these types of additives. Azido-ester plasticizers are used to modify the mechanical properties of propellant formulations (Zohari et. al., 2017). Physical properties analyzed during this project will be decomposition temperature (T_D), glass transition temperature (T_G), and molecular weight (MW), whereas the chemical property is heat of formation (ΔH°_f). The reason for focusing on these three properties is due to their importance in the realm of azido-ester plasticizers and will be discussed in the next section.

These properties can then be quantified with fitness functions using the atomic Signatures. The fitness functions used to evaluate the newly designed compounds were developed *via* a forward-stepping multiple linear regression approach. The equations, termed quantitative structure-property relationships (QSPRs), were based on T_D , ΔH°_f , and T_G . Molecular weight does not require a QSPR as it can be quantified from the molecular formula directly. These three QSPRs will use the atomic Signatures that have the greatest contribution to the respective property to predict numerically the property for the newly generated structures.

The inverse design process begins by deriving various constraint equations to ensure correct valency and aromaticity. Once these are derived, they can be used to recombine the atomic Signatures into potential, newly designed compounds. However, there will be a large number of potential solutions that will not have the properties of interest. The QSPRs will then be used to further filter these solutions with specified property ranges and solutions that satisfy the constraints and QSPR ranges will be inversely designed and their respective structures can potentially be generated.

This introduction establishes goals and objectives for this research project that will be used to answer this overall question: Can CAMD be used to design new, novel azido-ester plasticizers?

Background

The work being done for this specific project is not building off any existing work; however, the dataset has been obtained from two articles regarding the QSPR approach to investigate properties of these energetic azido-ester plasticizers (Sheibani, 2019 & Zohari et. al., 2017). These articles in particular used QSPR models to compare and contrast to the experimental values of T_D , ΔH°_f , and T_G . Sheibani and Zohari did not use atomic Signatures but instead utilized the number of carbon atoms, nitrogen atoms, ester functional groups, and azido functional groups. They also used two other structural parameters that account for non-additive contributions (Zohari et. al. 2017). In this work, the main focus is not to compare the modeling equations and their predictions to the experimental values, but to inverse design plasticizers that have the potential to perform better than the dataset. It was important to learn about which properties of these plasticizers are most important when designing new structures, which is why T_D , T_G , and ΔH°_f are the main properties being focused on in this project (Alexis, 2004; Schock and Bräse, 2020; Sheibani, 2019; Sheibani et. al., 2020; Zohari et. al., 2017). What will also be analyzed are the molecular weights of the designed compounds since molecular weight influences the properties and plasticizer itself (Stark et. al, 2005).

As mentioned before, the properties being investigated are T_D , ΔH°_f , T_G , and MW and they were not chosen at random, therefore, they each have their importance toward azido-ester plasticizers. Thermal stability is very important for these energetic compounds, which makes

decomposition temperature one of the most significant properties to be investigated. To be thermally stable, the decomposition temperature needs to be high enough so the propellant does not degrade (Zohari et. al., 2017). The magnitude of this temperature will be investigated above or close to the training set values.

The glass transition temperature is the temperature at which the polymer changes state from a glassy material to a rubbery material. For azido-ester plasticizers, the glass transition temperature should be low so the glassy transition is essentially out of range in cold temperatures (Zohari et. al., 2017). Since azido-ester plasticizers are supposed to change the mechanical properties of propellant formulations into a more flexible and more processable polymeric material, the glassy state is undesired (Zohari et. al., 2017).

Alongside the decomposition temperature and glass transition temperature is the heat of formation. For this property, it is important that it be highly positive, as it creates the energy for the propellant or explosive (Sheibani, 2019).

Although there will not be a QSPR generated for molecular weight, it will still take part in the filtering of solutions in the inverse design. The molecular weight is investigated because it plays a role in the volatility of these energetic compounds. According to Sheibani, azido-ester plasticizers at low molecular weight tend to be volatile, causing it to become incompatible with the polymer and migrate from the polymer where hazardous combustion products are then created (Sheibani et. al., 2020). Since azido-ester plasticizers are also supposed to increase compatibility and hazardous combustion products are undesired, the solutions with higher molecular weight will be focused on in this project.

Lastly, as a way to quantify the similarity between the generated structures and the training set, the Tanimoto Coefficient will be calculated for each design structure with respect to

the training set (CCDC). The Tanimoto Coefficient calculates a ratio of the number similar structure features to the total number of features present (CCDC). Normally, when two chemical structures have a Tanimoto Coefficient of 0.85 or greater, they can be considered to be similar and have the same activity (Patterson et. al., 1996, p. 3057). Through CAMD, it is desired to obtain a lower Tanimoto Coefficient for generated structures since this will show their originality when compared to the training set.

Methodology

The first step to inverse design of molecules is to gather a set of related compounds from literature. In this case, the category of compounds is azido-ester plasticizers. It would not be reasonable to include compounds within this dataset that do not have an azido or ester group. Doing so would skew the QSPR and create inaccuracies when generating structures. The second step would be to create *.mol files in ChemDraw of the training set. The algorithm that reads these structures requires a *.mol file when dissecting the compounds into the unique height-1 atomic Signatures.

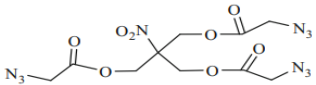
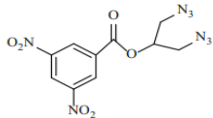
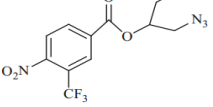
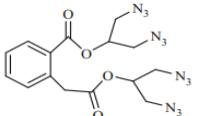
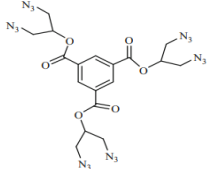
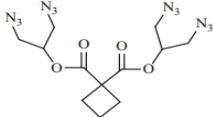
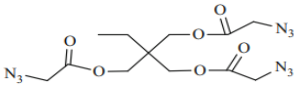
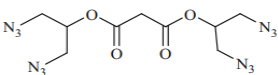
Once the training set is put through the algorithm, it will output a list of unique atomic Signatures found from the dataset where a QSPR can then be created for each property (T_D , T_G , and ΔH°_f). In these QSPRs, the dependent variable will be the respective property as it is a function of the atomic Signature, which is the independent variable. Using forward-stepping multiple linear regression, three final QSPRs can be established. Without using the QSPRs, inverse design of new plasticizers can be done with the atomic Signatures found in the beginning. These structures will have undergone filtering from constraint equations and a graphicality equation to assure correct bonding with valence electrons, a connected graph, and

correct aromaticity when atomic signatures are combined (Weis and Visco, 2009, p. 1021; Visco and Krew, 2019, p. 232). This filtering will still result in a large solution pool of structures. The QSPRs can then be used to further filter the designed structures into a range where the most optimal plasticizers are constructed, analyzed, and compared to those in the training set.

Data and Results

As described in the methodology, the first step of the process is to obtain the training set (Zohari et. al., 2017; Sheibani, 2019), which can be found in **Table 1** below:

Table 1: Azido-ester training set used for inverse design with respective property values (Zohari et. al., 2017; Sheibani, 2019).

Compound Number	Structure	MW [g.mol ⁻¹]	T _G [°C]	T _D [°C]	ΔH _f ^o [kJ.mol ⁻¹]
1		400.27	-34.1	217.2	-230.54
2		336.22	-38.98	249.63	-251.58
3		359.23	-42.24	250.83	-344.3
4		428.37	-55.48	237.23	888.1
5		582.46	-20.84	246.38	791.1
6		392.34	-67.49	243.04	547.3
7		383.33	-47	254.5	-157.5
8		352.28	-69	233.5	581.63

9		459.34	-34.38	209.83	645.31
10		272.22	-63.3	215	-328.86
11		281.19	-0.74	249.43	363.3
12		534.42	-55.66	242.48	827.6
13		352.28	-40.27	215.8	-
14		563.46	-45.99	242.34	1422.6
15		468.35	-35.4	221.5	-215.2
16		253.23	-84.27	241.36	595.36
17		253.23	-92.34	221.76	-

When the training set is dissected into atomic Signatures, the following unique atomic

Signature molecular descriptors are produced in **Table 2**:

Table 2: 28 Height-1 atomic Signatures were generated from the training set. The minimum and maximum occurrence within a structure between all training set compounds is provided.

Variable	Height-1 Signature	Min/Max Value	Variable	Height-1 Signature	Min/Max Value
X ₁	[NO ₂](C)	[0,2]	X ₁₅	Cp[C]p[N]	[0,1]
X ₂	[C](NO ₂)[C][C]	[0,1]	X ₁₆	[C](H)[H][H][O]	[0,1]
X ₃	[C](NO ₂)p[C]p[C]	[0,2]	X ₁₇	[C](H)[H][N ₃][O]	[0,2]
X ₄	C[C][C][C]	[0,1]	X ₁₈	[C](p[C][H]p[N])	[0,1]
X ₅	C[C][C][H]	[0,1]	X ₁₉	[C](p[C]p[C][H])	[0,4]
X ₆	C[C][H][H]	[0,3]	X ₂₀	[F](C)	[0,3]
X ₇	C[C][H][O]	[0,3]	X ₂₁	[H](C)	[6,21]
X ₈	C[F][F][F]	[0,1]	X ₂₂	[H](N)	[0,1]
X ₉	C[H][H][H]	[0,1]	X ₂₃	[N ₃](C)	[2,6]
X ₁₀	C[H][H][N ₃]	[2,6]	X ₂₄	[N](C)[C][C]	[0,1]
X ₁₁	C[H][H][N]	[0,3]	X ₂₅	[N](p[C][H]p[N])	[0,1]
X ₁₂	C[H][H][O]	[0,4]	X ₂₆	[N](p[C]p[N])	[0,1]
X ₁₃	C[O]=[O]	[1,4]	X ₂₇	[O](=[C])	[1,4]
X ₁₄	Cp[C]p[C]	[0,3]	X ₂₈	[O](C)[C]	[1,4]

From the Signatures generated, the QSPRs for each property being investigated were derived and are shown in **Equations 1, 2, and 3**. ΔH°_f and T_G utilize multiple linear regression while T_D uses the cross products of atomic Signatures as this property was nonlinear.

$$T_D = (-1.424)x_{12}x_{14} + 8.283x_9x_{14} + (-10.750)x_{17} + (-4.394)x_{10}x_{18} + (-20.904)x_{15}x_{15} + 53.344x_1 + (-3.054)x_3x_{23} + 0.186x_{13}x_{23} + 2.823x_{20} + 506.038 \quad (1)$$

$$\Delta H^\circ_f = (-189.437)x_7 + (-224.040)x_9 + (-735.489)x_8 + 122.473x_{11} + (-401.562)x_{12} + (-497.846)x_3 + (-167.850)x_6 + (-271.984)x_2 + 41.925x_{21} + 743.074 \quad (2)$$

$$T_G = 61.476x_{15} + (-10.279)x_{16} + 14.135x_{14} + 15.096x_{13} + 19.845x_1 + 13.362x_{17} + (-3.934)x_{19} + 5.928x_{24} + 175.983 \quad (3)$$

The stepwise regression process is shown in **Tables 3, 4, and 5** in **Appendix B** where the coefficient variation for each descriptor can be seen after adding one new descriptor at a time.

What also changes when adding one descriptor at a time is the R^2 value. The R^2 value is important for each QSPR because it conveys how accurate the property relationship is. So, while creating these QSPRs, **Figures 2, 3, and 4** were used to determine the number of Signatures sufficient enough to obtain an accurate QSPR. Those figures are displayed below and show how the R^2 value changes for T_D , ΔH°_f , and T_G , respectively.

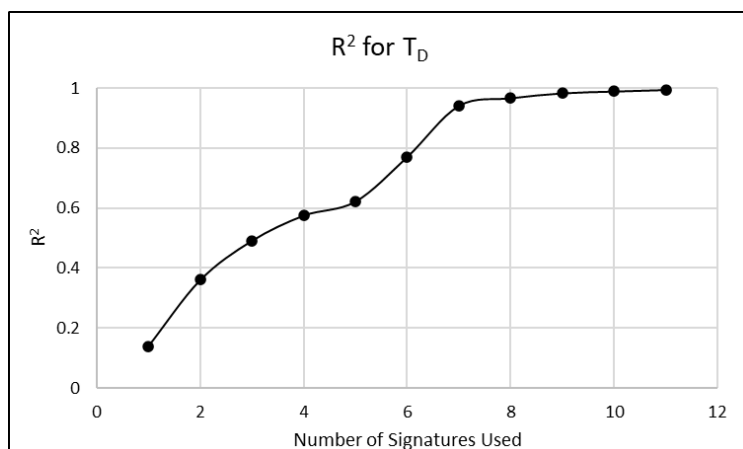


Figure 2: R^2 trend for T_D when using stepwise regression. 9 descriptors were used for this QSPR.

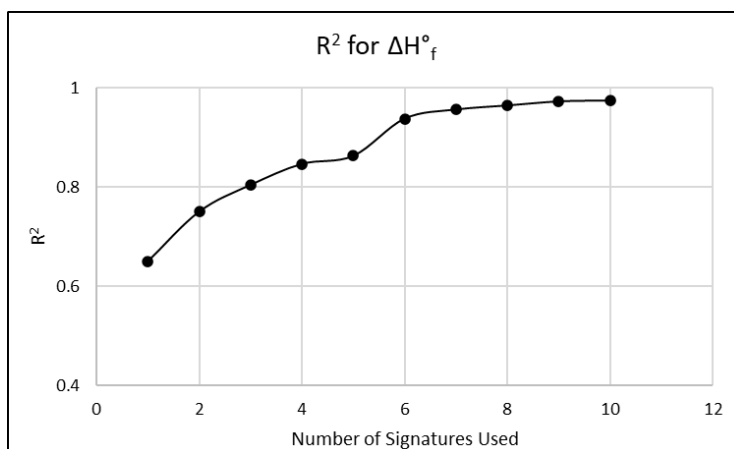


Figure 3: R^2 trend for ΔH°_f when using stepwise regression. 9 descriptors were used for this QSPR.

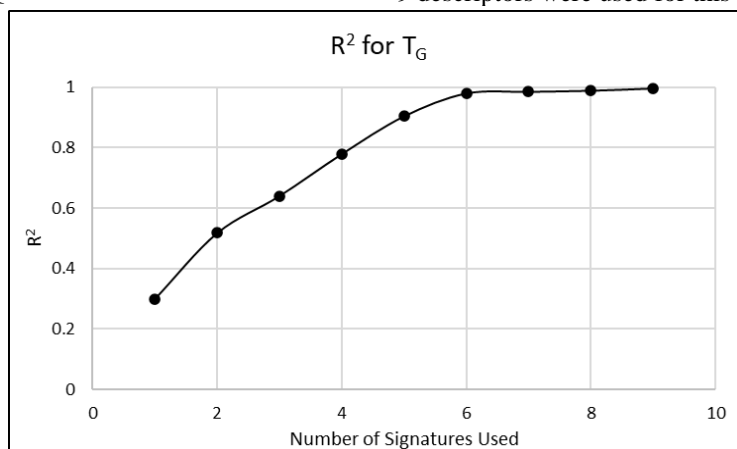


Figure 4: R^2 trend for T_G when using stepwise regression. 8 descriptors were used for this QSPR.

Once a certain number of descriptors are being used, the R^2 value will reach its asymptote at 1 and continue until the model will become overfit. The number of Signatures used for T_D , ΔH°_f ,

and T_G were 9, 9, and 8, respectively. It is essential to not use an overfit model as it will skew the predicted value and its performance will decrease. Therefore, R^2 values around 0.98 were chosen to avoid overfitting each model.

After creating the QSPR equations, it is necessary to derive constraint equations that ensure correct bonding, valency, and aromaticity (consistency equations), and a connected graph (graphicality equation). The consistency equations are comprised of one equation for each type of bond occurring in the dataset and ensure for any bond in an atomic Signature there is a corresponding bond in another atomic Signature. The graphicality equation utilizes the height-0 atomic Signatures and ensures the total vertex degrees of a solution is even, creating a connected graph if satisfied. These equations will confirm the Signature combinations construct actual structures (Weis and Visco, 2009, p. 1021; Visco and Krew, 2019, p. 232). Once the MATLAB algorithm is run, the following equations derived can be seen in **Figure 5**:

C-Br Bonding	$0 = -x_1 + x_2 + x_3$
C-H Bonding	$0 = -x_5 - 2x_6 - x_7 - 3x_9 - 2x_{10} - 2x_{11} - 2x_{12} - 3x_{16} - 2x_{17} - x_{18} - x_{19} + x_{21}$
C-O Bonding	$0 = -x_7 - x_{12} - x_{13} - x_{16} - x_{17} + 2x_{28}$
C-F Bonding	$0 = -3x_8 + x_{20}$
C-I Bonding	$0 = -x_{10} - x_{17} + x_{23}$
C-N Bonding	$0 = -x_{11} + 3x_{24}$
C=O Bonding	$0 = -x_{13} + x_{27}$
C-pN Bonding	$0 = -x_{15} - x_{18} + x_{25} + x_{26}$
N-H Bonding	$0 = -x_{22} + x_{25}$
N-pN Bonding	$0 = (x_{25} + x_{26}) \% 2$
C-pC Bonding	$0 = (2x_3 + 2x_{14} + x_{15} + x_{18} + 2x_{19}) \% 2$
C-C Bonding	$0 = (3x_2 + 4x_4 + 3x_5 + 2x_6 + 2x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14} + x_{15}) \% 2$
Graphicality	$0 = (-x_1 + 2x_2 + x_3 + 2x_4 + 2x_5 + 2x_6 + 2x_7 + 2x_8 + 2x_9 + 2x_{10} + 2x_{11} + 2x_{12} + x_{13} + x_{14} + x_{15} + 2x_{16} + 2x_{17} + x_{18} + x_{19} - x_{20} - x_{21} - x_{22} - x_{23} + x_{24} + x_{25} - x_{27} + 2) \% 2$

Figure 5: Constraint equations derived from Height-1 atomic Signatures from the training set in **Table 2**. “% 2” indicates a modulus equation.

Upon using these constraint equations in the inverse design/CAMD, there were over 2.5 million total solutions created. Solutions at this part of the process are unique molecular Signatures where each solution may have many isomers. At this point, the QSPRs previously derived can be inputted into the inverse design algorithm and quantify the properties being investigated. To get a better understanding of the property values quantified from the QSPRs for each solution, histograms were prepared. **Figures 6, 7, 8, and 9** display the distribution of solutions between the minimum and maximum of their respective properties. Since the initial set of solutions was over 2.5 million, the histograms used only 0.5% of the data to understand the distribution of each solution's properties as it was not possible to analysis the entire dataset.

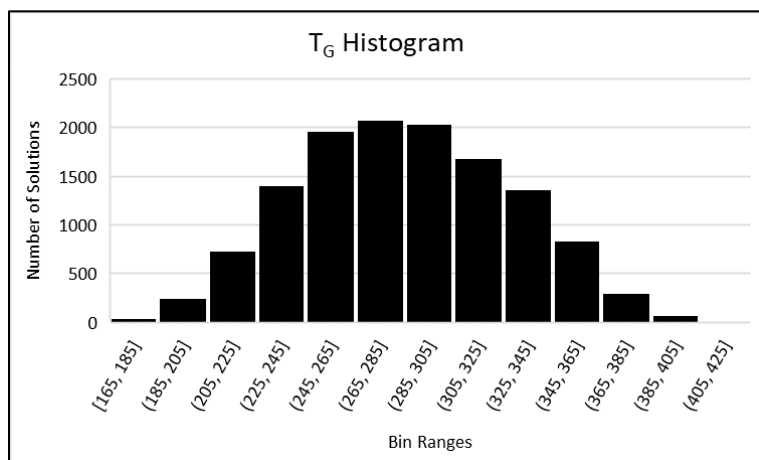


Figure 6: Distribution of 0.5% of total solutions for T_G .

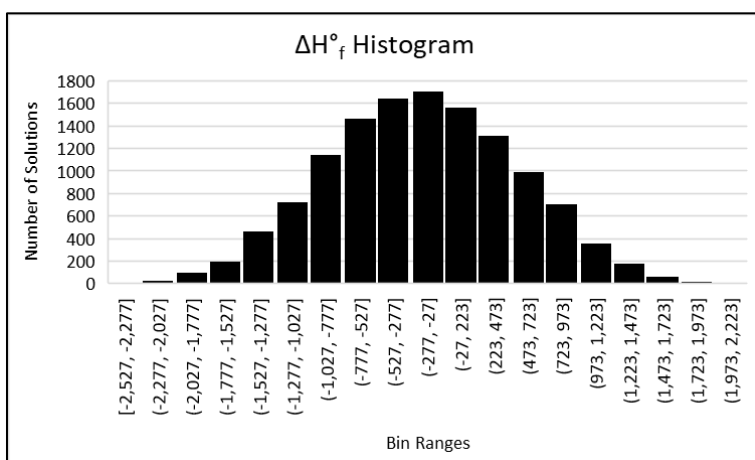


Figure 7: Distribution of 0.5% of total solutions for ΔH°_f .

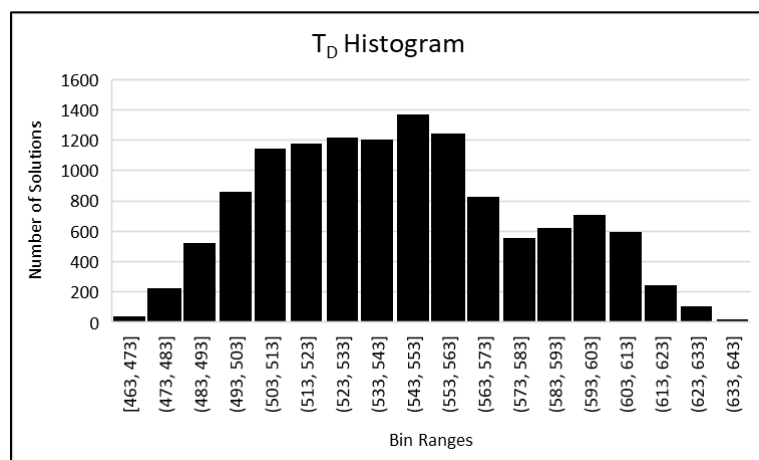


Figure 8: Distribution of 0.5% of total solutions for T_D .

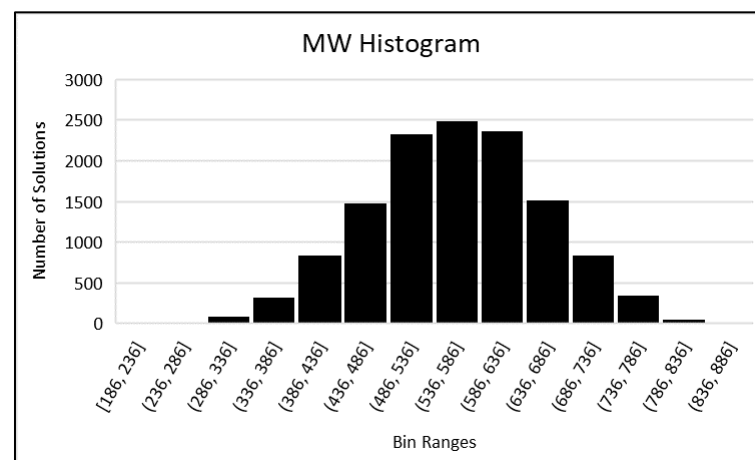


Figure 9: Distribution of 0.5% of total solutions for MW.

From analyzing these distributions and taking into account the background on the importance of each property provided earlier (whether the property should be high or low, in this case, with respect to the training set), the ranges in **Table 6** were chosen. These ranges were then used to further filter the solution set to yield potentially new, novel azido-ester plasticizers that possess desired property values.

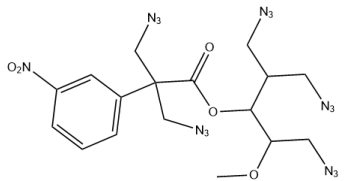
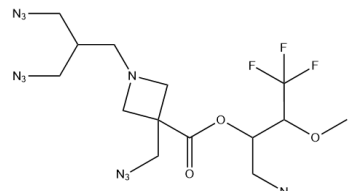
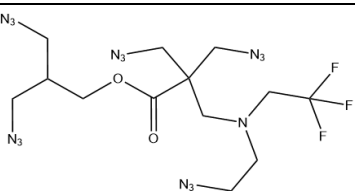
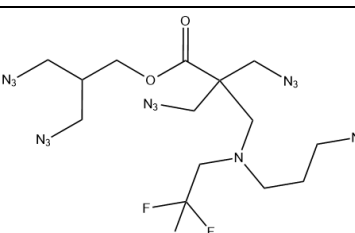
Table 6: QSPR filtering ranges to obtain desired azido-ester plasticizer molecular formulas.

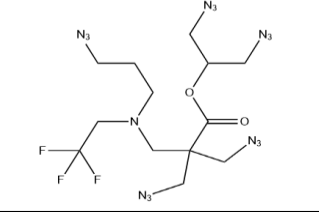
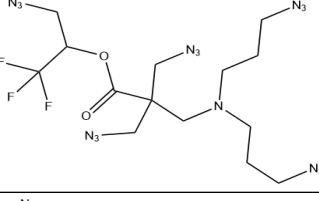
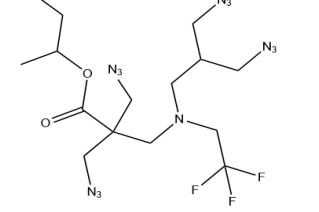
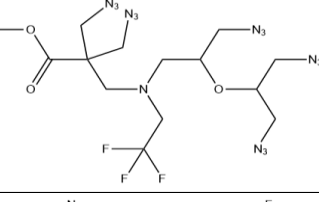
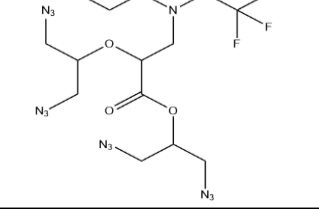
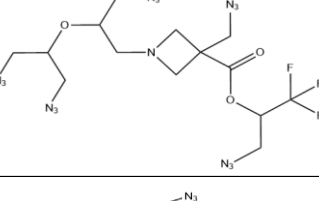
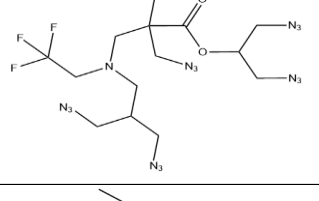
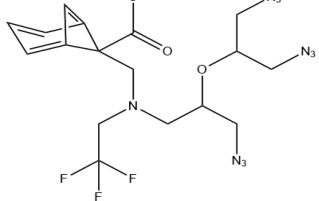
T_D [°C]	ΔH°_f [kJ.mol ⁻¹]	T_G [°C]	MW [g.mol ⁻¹]
242 to 317	550 to 1400	-103 to -73	475 to 550

After these filters were applied, there were 512 potential solutions that satisfied these four ranges. Of the 512 potential solutions, 12 created structures that were feasible (*i.e.*, had connected graphs and satisfied aromatic constraints, such as Huckel's rule) (LibreTexts). One structure from each of the twelve were evaluated, though many isomers can exist, per structure.

Table 7 displays one structure from each of the 12 solutions below:

Table 7: Sample set of the 12 inversely designed azido-ester plasticizers and quantified properties.

Cmpd #	Formula	Structure	MW [g.mol ⁻¹]	T_G [°C]	T_D [°C]	ΔH°_f [kJ.mol ⁻¹]
1	C ₁₇ H ₂₀ N ₁₆ O ₅ 102 isomers		528.5	-74.0	272.0	704.9
2	C ₁₄ H ₂₀ F ₃ N ₁₃ O ₃ 151 isomers		475.4	-86.3	242.3	834.6
3	C ₁₃ H ₁₉ F ₃ N ₁₆ O ₂ 58 isomers		488.4	-76.0	242.4	770.0
4	C ₁₄ H ₂₁ F ₃ N ₁₆ O ₂ 62 isomers		502.5	-76.0	242.4	686.0

5	$C_{13}H_{19}F_3N_{16}O_2$ 89 isomers		488.4	-76.0	242.4	814.3
6	$C_{14}H_{21}F_3N_{16}O_2$ 96 isomers		502.5	-76.0	242.4	730.3
7	$C_{14}H_{21}F_3N_{16}O_2$ 104 isomers		502.5	-76.0	242.4	842.0
8	$C_{14}H_{21}F_3N_{16}O_3$ 120 isomers		518.5	-86.3	242.4	876.6
9	$C_{13}H_{19}F_3N_{16}O_3$ 37 isomers		504.4	-76.0	242.4	603.3
10	$C_{14}H_{19}F_3N_{16}O_3$ 91 isomers		516.5	-76.0	242.4	603.3
11	$C_{14}H_{20}F_3N_{19}O_2$ 94 isomers		543.5	-76.0	242.6	1024.1
12	$C_{18}H_{21}F_3N_{10}O_3$ 153 isomers		482.4	-73.7	242.1	876.6

To better understand how one of these structures was generated from the training set, an example showing how Compound 1 from **Table 7** was constructed from three training set compounds can be found in **Figure 10** below. Training set compounds 4, 9, and 13 are labeled as such in the figure.

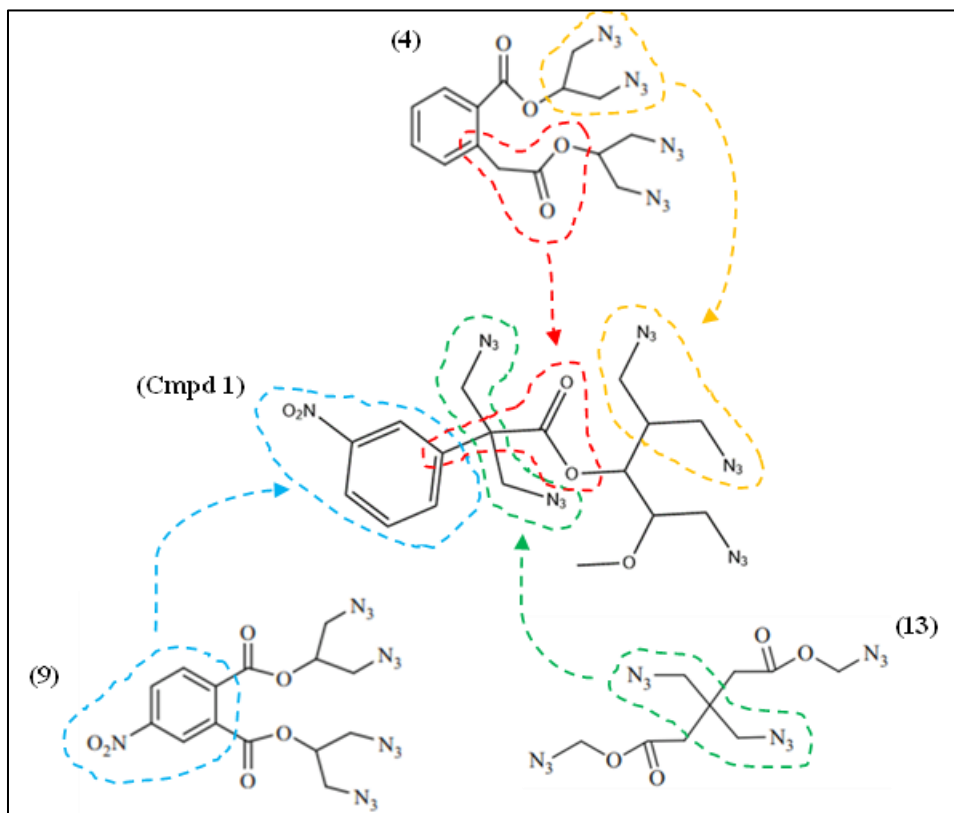


Figure 10: Example of how the inversely designed Cmpd 1 is formed from three training set compounds 4, 9, and 13.

Four main components from the three training set compounds were used: nitrobenzene group (light blue), 1,3-diazidopropane group (yellow and green), and ester group (red). It can be seen that the generated structure is not an exact replica of any training set compound, but a combination of specific groups of atomic Signatures from the training set while also maintaining its novelty among the training set.

Once the 12 generated structures were established, their Tanimoto coefficients were calculated with respect to the 17 training set compounds to measure their similarity and novelty.

Normally, when two chemical structures have a Tanimoto Coefficient of 0.85 or greater, they can be considered to be similar and have the same activity (Patterson et. al., 1996, p. 3057).

Overall, the generated plasticizers can be considered novel and unique when compared to the training set with their average Tanimoto Coefficient ranging from approximately 0.4525 to 0.5167. The average maximum Tanimoto Coefficient calculated from all structures was approximately 0.7547. Below in **Table 8** the minimum, average, and maximum Tanimoto Coefficients for each designed structure can be found to understand their novelty when compared to the training set compounds.

Table 8: Minimum, average, and maximum Tanimoto Coefficients for the 12 designed azido-ester plasticizers with respect to the 17 training set compounds.

Compound #	Min Tanimoto Coefficient	Avg Tanimoto Coefficient	Max Tanimoto Coefficient
1	0.3137	0.5167	0.7000
2	0.2692	0.4791	0.6786
3	0.2549	0.4635	0.6429
4	0.2407	0.4530	0.6667
5	0.2800	0.4824	0.6727
6	0.2642	0.4626	0.6964
7	0.2642	0.4525	0.6964
8	0.2593	0.4714	0.7455
9	0.2745	0.4880	0.7547
10	0.2692	0.4890	0.7407
11	0.2642	0.4573	0.7273
12	0.2500	0.4576	0.6296

Analyzing the numerical value of the Tanimoto Coefficients is satisfactory, however, an example of how Compound 1 received its maximum coefficient with the training set Compound 14 can be seen in the visual comparison in **Figure 11** below:

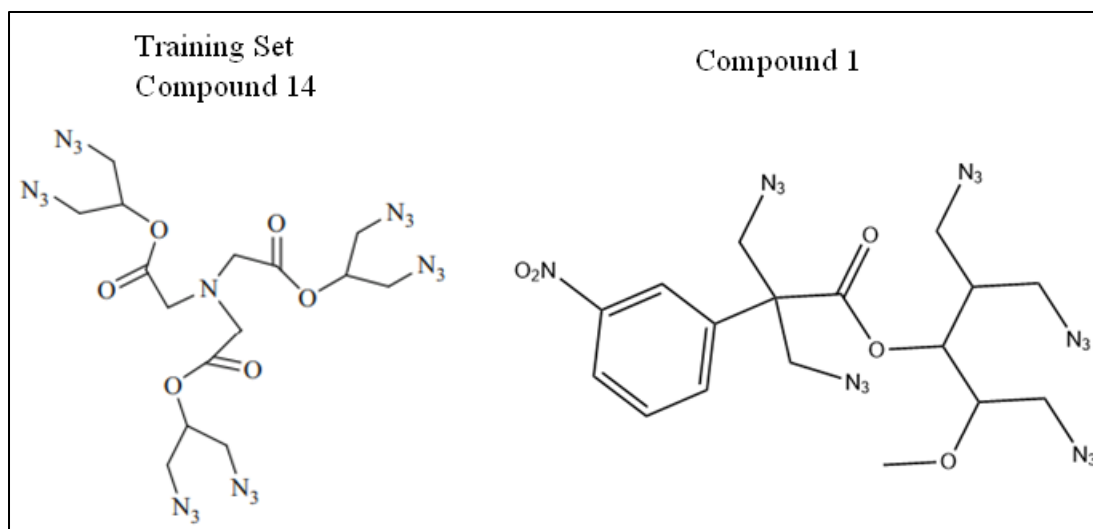


Figure 11: Visual comparison of Compound 1's more similar structure from the training set (Training set compound 14).

Visually, it is easier to see many differences between the two compounds, however, through the Tanimoto Coefficient calculation, the generated compound has a similarity of 0.7000 to the training set compound. Even though there was one nitrobenzene group added, one ether group added, one less 1,3-diazidopropane group, and two less ester groups, these two structures are moderately similar intermolecularly. This demonstrates the importance of calculating the Tanimoto Coefficient since it can quantify the similarity of two structures instead of a visual comparison.

Discussion

Upon the structure generation of the inverse designed solutions, there were 12 azido-ester plasticizers with different molecular Signatures created. Due to their novelty among the training set compounds, these new compounds have the potential to have optimal properties for such plasticizers within the propellant and explosive formulation category. The initial, large solution pool was filtered so solutions would have a high positive decomposition temperature, high positive heat of formation, low glass transition temperature, and high molecular weight. The similarities between the training set and the design set can be seen when analyzed visually, however, the Tanimoto Coefficient was used as a numerical metric to examine the actual magnitude of similarity. With an average Tanimoto Coefficient range of 0.4525 to 0.5167, it can be concluded that the design set has its own originality. A chemical structure search was done in chemical databases and no designed structure exists. Since no structure exists, the next steps that should be taken for this project would be to synthesize and test these structures in a lab. The property values obtained from testing can then be compared to the QSPR predicted values. Other properties that could potentially provide insight on the quality of these compounds could be their respective viscosity and moduli. During this research, these properties were omitted as data was not available for them to derive QSPRs. Instead of using experimental data to establish a QSPR for viscosity, it is possible to use group contribution with the Joback method; however, this was not investigated during the project since it is not utilizing CAMD. At this point, CAMD was successful in the design of new, novel azido-ester plasticizers, but next steps should be taken to validate the property predictions.

References

- Alexis, F. (2004). Factors affecting the degradation and drug-release mechanism of poly(lactic acid) and poly[(lactic acid)-co-(glycolic acid)]. *Polymer International*, 54(1), 36–46. <https://doi.org/10.1002/pi.1697>
- CCDC, Cambridge Crystallographic Data Centre. (n.d.). What are the differences between the Tanimoto and Dice similarity coefficients? Retrieved April 10, 2023, from <https://www.ccdc.cam.ac.uk/support-and-resources/support/case/?caseid=899a6a77-e379-4981-84f4-07de67f39016#:~:text=The%20Tanimoto%20coefficient%20is%20the,is%200%20to%201%20inclusive.>
- LibreTexts. (n.d.). Aromaticity and Huckel's Rule. In Organic Chemistry (Wade) Complete and Semesters I and II. [https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Map%3A_Organic_Chemistry_\(Wade\)_Complete_and_Semesters_I_and_II/Map%3A_Organic_Chemistry_\(Wade\)/17%3A_Aromatic_Compounds/17.05%3A_Aromaticity_and_Huckel's_Rule.](https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Map%3A_Organic_Chemistry_(Wade)_Complete_and_Semesters_I_and_II/Map%3A_Organic_Chemistry_(Wade)/17%3A_Aromatic_Compounds/17.05%3A_Aromaticity_and_Huckel's_Rule.)
- Patterson, D. E., Cramer, R. D., Ferguson, A. M., Clark, R. D., & Weinberger, L. E. (1996). Neighborhood Behavior: A Useful Concept for Validation of “Molecular Diversity” Descriptors. *Journal of Medicinal Chemistry*, 39(16), 3049–3059. <https://doi.org/10.1021/jm960290n>
- Schock, M., & Bräse, S. (2020). Reactive & Efficient: Organic azides as cross-linkers in Material Sciences. *Molecules*, 25(4), 1009. <https://doi.org/10.3390/molecules25041009>
- Sheibani, N. (2019). Heat of formation assessment of Organic Azido compounds used as green energetic plasticizers by QSPR approaches. *Propellants, Explosives, Pyrotechnics*, 44(10), 1254–1262. <https://doi.org/10.1002/prep.201900082>
- Sheibani, N., Zohari, N., & Fareghi-Alamdari, R. (2020). Rational design, synthesis and evaluation of new Azido-ester structures as green energetic plasticizers. *Dalton Transactions*, 49(36), 12695–12706. <https://doi.org/10.1039/d0dt02250k>
- Stark, T. D., Choi, H., & Diebel, P. W. (2005). Influence of plasticizer molecular weight on plasticizer retention in PVC Geomembranes. *Geosynthetic International*, 12(2), 99–110. <https://doi.org/10.1680/gein.2005.12.2.99>
- Visco, D. P., & Krew, J. A. (2019). Molecular Design Using the Signature Molecular Descriptor. *Chemical Engineering Education*, 53(4), 230–236.
- Weis, D. C., & Visco, D. P. (2009). Computer-aided molecular design using the Signature Molecular Descriptor: Application to Solvent Selection. *Computers & Chemical Engineering*, 34(7), 1018–1029. <https://doi.org/10.1016/j.compchemeng.2009.10.017>

Zohari, N., Sheibani, N., & Chavoshi, H. Z. (2017). Investigation of the most effective molecular descriptors on the thermal behaviour of energetic Azido-ester plasticizers through QSPR approach. *Journal of Thermal Analysis and Calorimetry*, 131(3), 3157–3167.
<https://doi.org/10.1007/s10973-017-6809-z>

Appendix A

Sample Calculations for **Equations 1, 2, and 3** using **Table 7** - Cmpd 1:

Table 9: Atomic Signatures' occurrences for Cmpd 1 which are used to calculate QSPR values.

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄
Cmpd 1	1	0	1	1	1	0	2	0	0	5	0	0	1	1
	X ₁₅	X ₁₆	X ₁₇	X ₁₈	X ₁₉	X ₂₀	X ₂₁	X ₂₂	X ₂₃	X ₂₄	X ₂₅	X ₂₆	X ₂₇	X ₂₈
Cmpd 1	0	1	0	0	4	0	20	0	5	0	0	0	1	2

$$T_D = (-1.424)X_{12}X_{14} + 8.283X_9X_{14} + (-10.750)X_{17} + (-4.394)X_{10}X_{18} + (-20.904)X_{15}X_{15} + 53.344X_1 + (-3.054)X_3X_{23} + 0.186X_{13}X_{23} + 2.823X_{20} + 506.038 \quad (1)$$

$$T_D = (-1.424)*(0)*(1) + 8.283*(0)*(1) + (-10.750)*(0) + (-4.394)*(5)*(0) + (-20.904)*(1)*(0) + 53.344*(1) + (-3.054)*(1)(5) + 0.186*(1)(5) + 2.823*(0) + 506.038$$

$$T_D = 545.0 \text{ K}$$

$$T_D = 272.0 \text{ }^\circ\text{C}$$

$$\Delta H^\circ_f = (-189.437)X_7 + (-224.040)X_9 + (-735.489)X_8 + 122.473X_{11} + (-401.562)X_{12} + (-497.846)X_3 + (-167.850)X_6 + (-271.984)X_2 + 41.925X_{21} + 743.074 \quad (2)$$

$$\Delta H^\circ_f = (-189.437)*(2) + (-224.040)*(0) + (-735.489)*(0) + 122.473*(0) + (-401.562)*(0) + (-497.846)*(1) + (-167.850)*(0) + (-271.984)*(0) + 41.925*(20) + 743.074$$

$$\Delta H^\circ_f = 704.9 \text{ kJ.mol}^{-1}$$

$$T_G = 61.476X_{15} + (-10.279)X_{16} + 14.135X_{14} + 15.096X_{13} + 19.845X_1 + 13.362X_{17} + (-3.934)X_{19} + 5.928X_{24} + 175.983 \quad (3)$$

$$T_G = 61.476*(0) + (-10.279)*(1) + 14.135*(1) + 15.096*(1) + 19.845*(1) + 13.362*(0) + (-3.934)*(4) + 5.928*(0) + 175.983$$

$$T_G = 199 \text{ K}$$

$$T_G = -74.0 \text{ }^\circ\text{C}$$

Appendix B

Table 10: Stepwise Regression for the T_D QSPR showing the progression of the coefficients for each Signature when a new Signature is added to the equation. Highlighted row signifies the QSPR chosen.

# of Signatures Used	R^2	Constant	Variable Coefficient										
			$x_{12}^*x_{14}$	$x_9^*x_{14}$	x_{17}	$x_{10}^*x_{18}$	$x_1^*x_{15}$	x_1	$x_3^*x_{23}$	$x_{13}^*x_{23}$	x_{20}	$x_4^*x_7$	$x_{14}^*x_{14}$
1	0.140	510.875	-0.964	-	-	-	-	-	-	-	-	-	-
2	0.360	509.518	-1.254	9.465	-	-	-	-	-	-	-	-	-
3	0.492	511.265	-1.371	9.092	-11.133	-	-	-	-	-	-	-	-
4	0.576	512.788	-1.472	8.768	-11.894	-5.963	-	-	-	-	-	-	-
5	0.623	515.164	-1.470	7.812	-13.082	-6.755	-3.546	-	-	-	-	-	-
6	0.772	512.435	-1.110	7.893	-11.717	-5.845	-13.909	19.247	-	-	-	-	-
7	0.942	513.431	-1.347	8.159	-12.215	-6.177	-20.283	49.695	-3.005	-	-	-	-
8	0.968	507.325	-1.431	8.182	-11.106	-4.735	-21.627	54.952	-3.086	0.162	-	-	-
9	0.984	506.038	-1.424	8.283	-10.750	-4.394	-20.904	53.344	-3.054	0.186	2.823	-	-
10	0.990	504.939	-1.385	8.458	-10.315	-4.063	-20.903	53.799	-3.032	0.195	2.951	2.503	-
11	0.995	504.396	-1.309	8.724	-9.852	-3.824	-21.385	55.701	-3.193	0.180	2.440	3.030	0.519

Table 11: Stepwise Regression for the ΔH_f° QSPR showing the progression of the coefficients for each Signature when a new Signature is added to the equation. Highlighted row signifies the QSPR chosen.

# of Signatures Used	R^2	Constant	Variable Coefficient										
			x_7	x_9	x_8	x_{11}	x_{12}	x_3	x_6	x_2	x_{21}	x_5	x_{14}
1	0.650	-166.007	381.218	-	-	-	-	-	-	-	-	-	-
2	0.751	-344.867	458.983	563.797	-	-	-	-	-	-	-	-	-
3	0.805	-278.281	441.056	497.211	-507.075	-	-	-	-	-	-	-	-
4	0.847	-245.906	395.191	464.836	-493.586	160.977	-	-	-	-	-	-	-
5	0.863	-27.964	298.328	400.994	-614.665	185.193	-77.050	-	-	-	-	-	-
6	0.938	583.396	77.954	102.423	-635.991	201.780	-233.445	-369.660	-	-	-	-	-
7	0.956	715.297	70.647	88.173	-678.740	165.121	-264.614	-451.504	-110.625	-	-	-	-
8	0.964	841.848	25.328	-0.199	-713.371	168.256	-280.620	-498.105	-122.956	-230.487	-	-	-
9	0.972	743.074	-189.437	-224.040	-735.489	122.473	-401.562	-497.846	-167.850	-271.984	41.925	-	-
10	0.974	786.447	-243.635	-275.762	-750.512	131.202	-426.976	-507.439	-173.545	-292.279	46.355	111.216	-
11	0.976	993.597	-262.089	-245.617	-891.961	251.717	-393.062	-551.787	-99.343	-307.835	21.910	325.226	96.328

Table 12: Stepwise Regression for the T_G QSPR showing the progression of the coefficients for each Signature when a new Signature is added to the equation. Highlighted row signifies the QSPR chosen.

# of Signatures Used	R ²	Constant	Variable Coefficient								
			x ₁₅	x ₁₆	x ₁₄	x ₁₃	x ₁	x ₁₇	x ₁₉	x ₂₄	x ₂₃
1	0.300	221.506	50.894	-	-	-	-	-	-	-	-
2	0.517	224.220	48.180	-43.420	-	-	-	-	-	-	-
3	0.640	218.842	53.558	-38.042	8.067	-	-	-	-	-	-
4	0.779	194.142	67.891	-23.709	9.870	10.367	-	-	-	-	-
5	0.903	180.427	61.531	-14.313	7.881	14.685	15.756	-	-	-	-
6	0.979	173.712	64.584	-9.151	8.946	16.239	17.865	13.355	-	-	-
7	0.985	175.820	61.493	-10.451	13.720	15.431	19.656	13.109	-3.772	-	-
8	0.988	175.983	61.476	-10.279	14.135	15.096	19.845	13.362	-3.934	5.928	-
9	0.996	181.404	61.298	-9.762	15.908	16.641	18.046	14.095	-4.358	10.838	-2.494