

# An Explosophore-Based Approach Towards the Prediction of Energetic Material Sensitivity Properties

Jonas Rein,<sup>†,a</sup> Jonathan M. Meinhardt,<sup>†,a</sup> Julie L. Hofstra Wahlman,<sup>b</sup> Matthew S. Sigman,<sup>\*,b</sup> Song Lin<sup>\*,a</sup>

<sup>a</sup> Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, United States.

<sup>b</sup> Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States.

---

**ABSTRACT:** Accurate prediction of the sensitivity properties of high-energy materials (HEMs) and the study of their decomposition mechanisms are two major focuses within energetics research. Due to the hazards associated with the synthesis and handling of energetic materials, predictive models for HEM sensitivity are of great importance in enabling the safe and efficient development of future HEMs. Traditional predictive modeling of HEM decomposition via machine learning algorithms generally displays limited interpretability, while mechanistic studies of HEMs typically focus on small subsets of structurally analogous compounds lacking generalizability. This study aims to bridge the gap between predictive modeling and computational mechanistic analysis of HEMs, with the goal of providing chemically interpretable models for HEM sensitivity property prediction. Herein, we disclose the use of multivariate linear regression (MLR) modeling for the prediction of the decomposition temperature and impact sensitivity of HEMs. We report an explosophore-based approach to sensitivity property prediction featuring an ensemble of quantum mechanical parameters and computational workflows that enable rapid parameterization and modeling of energetic functional groups. We then employ these methods to accurately predict sensitivity properties of nitrogen-rich tetrazole and azide HEMs. These statistical MLR models are readily interpreted based on the principles of physical organic chemistry, producing structure-property relationships to guide the rational design of new HEMs. Furthermore, we extend our explosophore-based approach to predict the sensitivity properties of HEMs containing multiple, non-equivalent energetic functional groups through the identification of molecular triggers for the bulk decomposition of HEMs. Finally, we showcase the viability of our methods towards *ab initio* virtual screening of HEMs through predictive modeling of external test sets of tetrazole HEMs using structures and parameters generated exclusively *in silico*.

---

## Introduction

An important consideration in the design of energetic materials is their sensitivity to decomposition induced by external stimuli. The practicality of any high energy material (HEM) is immediately constrained by its ability to be safely synthesized, formulated, and stored without significant risks of deflagration or detonation. In contrast to numerous studies that aim to predict the performance properties (detonation velocity, detonation pressure, etc.) of HEMs,<sup>1-3</sup> analogous methods for the prediction of the sensitivity properties of HEMs (decomposition temperature ( $T_{\text{dec}}$ ), impact sensitivity (IS), friction sensitivity, electric spark sensitivity) remain elusive. In the absence of robust predictive tools, the development of new energetic materials is often subject to time-consuming cycles of synthesis and characterization in which chemical insight is gained almost entirely through empirical trial-and-error.<sup>4</sup> As such, the ability to precisely tune the properties of HEM candidates based on their chemical structures offers the potential to accelerate the development of new HEMs for military and civilian applications. Streamlined methods for the *ab initio* design of HEMs can also minimize hazardous experimental work associated with the preparation of these materials. Benefits of such methods extend further to the synthesis of fine chemicals,<sup>5</sup> where energetic motifs such as nitrogen-containing heterocycles, azides, and nitro groups are commonly encountered in synthetic intermediates and in

bioactive molecules.<sup>6</sup> Therefore, understanding the decomposition thresholds of these motifs is also a component of improved safety in process development chemistry.

Numerous experimental and computational studies have provided valuable insight into the decomposition of HEMs.<sup>7-11</sup> However, the narrow scope of such studies often results in limited applicability for the streamlined screening of large numbers of HEM candidates. Mechanistic studies of HEM decomposition are usually performed on datasets consisting of a single molecule or a narrow set of related molecules.<sup>7,12-15</sup> Conclusions drawn from such studies can lead to poor generalizability when applied to structurally dissimilar compounds. Furthermore, the decomposition of energetic materials is often governed by multiple mechanisms with comparable activation energies, hindering rational design of HEMs based on a single mechanistic proposal.<sup>16</sup> Additionally, theoretical mechanistic studies involving excited states and transition state calculations remain a computationally expensive exercise, limiting their applicability for virtual screening of large numbers of structurally diverse compounds. Analogous challenges of limited generalizability and ballooning resource utilization are also encountered in other areas of chemistry such as asymmetric catalysis, organic methodology development, and battery research. These constraints are increasingly addressed by the implementation of data-driven approaches which aim to construct quantitative structure-

property relationships (QSPRs) to aid prediction and guide rational design. In this vein, multivariate linear regression (MLR) statistical modeling is an attractive approach for developing QSPRs due to its inherent interpretability when used in conjunction with physically relevant molecular descriptors.<sup>17</sup> We sought to utilize an MLR-based approach to construct sensitivity property models which would reveal key mechanistic features in HEM decomposition for energetic material families at the forefront of current research.

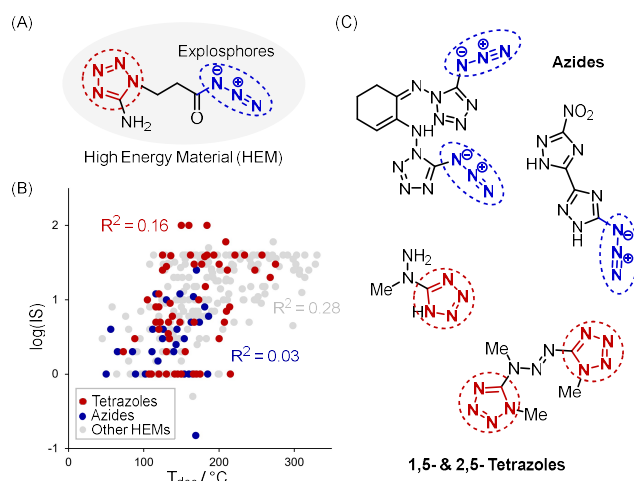
Many commercial HEMs suffer from a combination of harsh nitration conditions in synthesis (RDX, HMX, TNT, CL-20), significant toxicity (RDX,  $\text{Pb}(\text{N}_3)_2$ , AP), or deleterious environmental impacts in storage and use. In light of these factors, nitrogen-rich explosives have garnered significant attention as replacements for canonical primary and secondary explosives.<sup>18–20</sup> However, a lack of reported structure-property relationships for nitrogen-rich functionalities such as organic azides and nitrogen-containing heterocycles hampers the rational design of new heterocyclic explosives and structural tuning of existing HEMs.

In order to elucidate these QSPRs, we employed an *explosophore-based approach* to describe and classify HEMs based on the presence of energetic functional groups known as *explosophores* (Figure 1A), typical examples of which include nitro groups, nitramines, azides, and nitrogen-containing heterocycles. Explosophores are widely acknowledged to be responsible for the initial mechanistic steps leading to the deflagration or detonation of an HEM.<sup>21</sup> As such, detailed molecular description of explosophores is paramount in the development of chemically interpretable predictive models that can guide the design of next-generation HEMs. To enable this study, we compiled a highly diverse sensitivity property database containing over 400 HEMs reflecting prevailing trends in energetics research (Figure 1B).<sup>22</sup>

To capture the mechanistic importance of explosophores, we designed a general set of structural and quantum mechanical descriptors that can be applied to any type of explosophore, yielding a rich description of the chemical structures responsible for decomposition events. This descriptor ensemble includes natural bond orbital (NBO) analysis, Hirshfeld surface analysis, vibrational frequencies, charges, and the molecular geometry of explosophores. With this descriptor set in hand, we employed MLR modeling to develop chemically interpretable predictive models for the decomposition temperature and impact sensitivity of tetrazole and azide HEMs (Figure 1C). Through several case studies, we demonstrate the versatility of our explosophore-based approach for accurate prediction of HEM sensitivity properties and identification of structure-property correlations useful in rational design, as well as for corroboration of proposed decomposition mechanisms.

## Dataset Curation

We first compiled a sensitivity property dataset comprised of 386 energetic materials synthesized by the research groups of Klapötke and Shreeve.<sup>23</sup> The structural diversity contained within our dataset is extensive, featuring compounds representative of nearly every major explosophore class in



**Figure 1.** (A) Examples of explosophores contained within an HEM. (B) Plot of decomposition temperature ( $T_{\text{dec}}$ ) vs. impact sensitivity ( $\log(\text{IS})$ ) for our HEM database. The two properties are poorly correlated for major HEM classes. (C) Explosophore families modeled in this study with representative examples from each family.

numbers sufficiently large for the application of MLR modeling. For each of the HEMs in our dataset, sensitivity properties (i.e.,  $T_{\text{dec}}$  and IS) and experimental crystal structures were obtained from literature sources.<sup>24</sup> Notably, HEMs containing metal centers and ionic structures with associated energetic counterions and oxidizers (perchlorate, guanidinium, etc.) were excluded as the presence of these auxiliary components substantially alters the sensitivity properties of HEMs, precluding efforts to study the decomposition mechanisms of explosophores on organic molecules.<sup>25</sup>

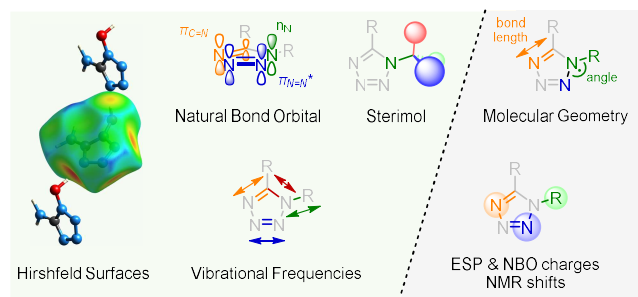
Consistency between the methods used to collect experimental HEM sensitivity properties is also a requirement for the compilation of high-quality datasets for predictive modeling. Differential scanning calorimetry (DSC) is typically employed to measure the decomposition temperature of energetic materials, the precision of which is dependent on the rate at which a sample is heated. As such, our dataset contains compounds subjected to the same DSC procedure.<sup>26</sup> Impact sensitivity measurements are typically conducted using mechanical tests and are often not reported with a high degree of precision.<sup>27</sup> In our study, we transformed experimental IS measurements to a logarithmic scale ( $\log(\text{IS})$ ).<sup>28</sup>

The diversity of the compounds in our dataset required several simple yet important filtering steps prior to the generation of data subsets appropriate for MLR modeling. Perhaps intuitively, different explosophores impart different levels of sensitivity to a given HEM. Therefore, it is challenging to develop an interpretable MLR model using a training set that includes molecules spanning several different explosophore classes. To address this issue, we utilized Python scripts to sort and categorize HEMs using SMARTS queries and SMILES keys.<sup>29</sup> The result of this selection was the creation of a number of subsets (i.e., monoazides, polyazides, tetrazoles, ditetrazoles) that maintained structural diversity about the constraint of a common explosophore. Statistical models generated from these datasets could then be used to extract chemical insight characteristic of a specific explosophore class.

## Parameter Acquisition

Utilization of available experimental crystal structures for the HEMs in our compound library was initially convenient for featurization efforts in lieu of structures calculated using molecular mechanics. HEM sensitivity properties are also influenced by the solid state structure of the crystal lattice.<sup>30,31</sup> As later discussions will illustrate, we also demonstrate that our methods can be used to model compounds lacking experimental crystal structures in order to support streamlined *ab initio* screening of HEM candidates.

With the 3D structures of our HEMs as input geometries, we conducted DFT optimizations (B3LYP/6-31+G(d,p)) and single point calculations (M06-2X/def2-TZVP) in the gas-phase. Extraction of a suite of geometric and electronic parameters for the explosophores in these HEMs (Figure 2) was then conducted using automated computational tools (disclosed in the SI). These parameters included NBO charges, electrostatic potential (ESP) charges, predicted NMR shifts, NBO orbitals, Sterimol descriptors, vibrational frequencies, and a series of geometric descriptors such as bond angles, distances, and dihedral angles. Mathematical combinations of these parameters could provide additional parameters for evaluation such as bond orders and HOMO–LUMO gap energies (see SI for details). Lastly, parameters calculated from Hirshfeld isosurfaces were obtained to explore correlations between crystal packing and HEM sensitivity.<sup>32</sup>



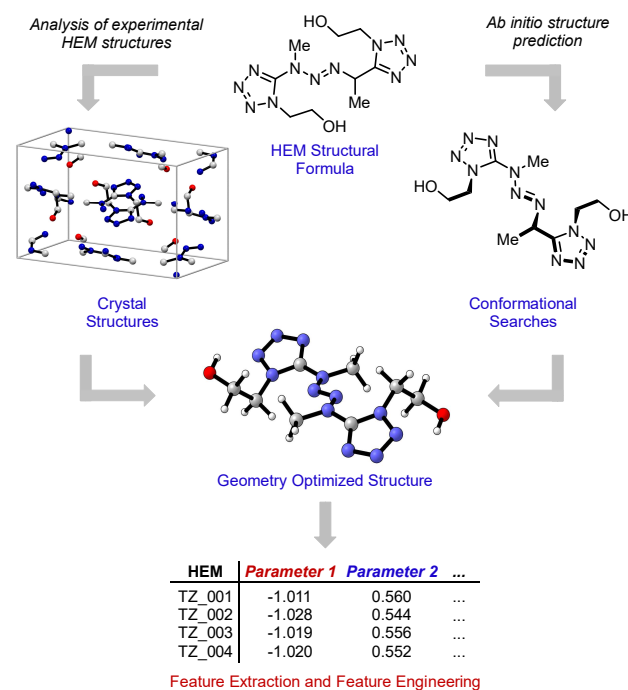
**Figure 2.** An explosophore-based approach to chemical featurization of energetic materials. The graphic above depicts geometric, quantum mechanical, and solid-state parameterization of tetrazole HEMs.

## QSPR for Tetrazole HEMs

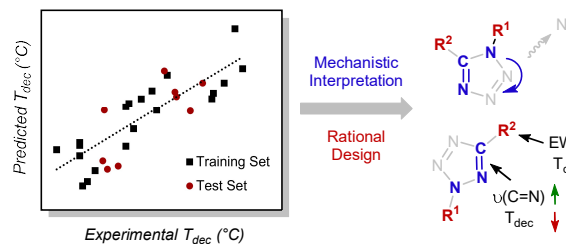
**Workflow:** Explosophores are the key reactive moieties responsible for the initiation of HEM deflagration and detonation. The viability of our explosophore-based approach to sensitivity prediction was initially explored using subsets of our larger dataset featuring HEMs containing only a single explosophore. Our workflow proceeds by first identifying an explosophore family of interest, followed by interrogation of a sensitivity property for prediction ( $T_{dec}$  or  $\log(IS)$ ). This selection process typically yielded datasets of 30–50 compounds suitable for MLR modeling. Of the HEMs in a selected explosophore class, compounds with multiple, non-equivalent instances of the explosophore are treated separately (*vide infra*). Once a dataset was curated, we then parameterized each molecule using our ensemble of chemical descriptors.

With experimental and descriptor datasets in hand, we applied MLR modeling to develop QSPRs for HEM sensitivity according to the general workflow described in Figure 3. First, an initial set of MLR models is generated for the sensitivity property of interest using forward stepwise regression. Multiple leading models are considered in terms of the number of parameters in the model, avoidance of cross-terms, and the recorded regression statistics (see SI for details). Notably, consideration of the univariate correlations between descriptors and sensitivity properties was a useful tool for model selection and the construction of additional descriptors to improve model performance. Secondly, univariate correlations can be explored to identify outlier compounds, i.e., compounds where the presence of a neighboring structure (e.g., acyl azides in an azide dataset) alters the properties of an explosophore in a manner that categorically differentiates it from the remainder of the test set. Inclusion of such compounds has the capacity to disproportionately influence the overall model. All suspected model outliers were verified using a one-tailed Grubbs outlier test at the 0.05 significance level.<sup>33</sup> When deemed appropriate, removal of such outliers was conducted and new models were generated to produce improved predictions. In the following sections, we apply this general workflow to generate accurate

### (A) Dataset Curation and HEM Featurization



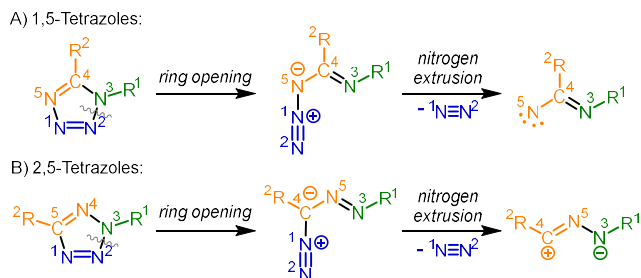
### (B) Modeling and Interpretation



**Figure 3.** Workflow schematic for (A) computational featurization of HEMs and (B) statistical modeling of HEMs for mechanistic study and rational design.

and interpretable structure-property relationships for nitrogen-rich HEMs.

**Tetrazoles:** Tetrazoles are emerging as an increasingly important class of energetic materials due to their potential to replace nitrated compounds and lead-based primary explosives.<sup>18,34</sup> Unlike their canonical nitro counterparts, tetrazoles do not require forcing nitration conditions for their synthesis and pose lower environmental toxicity.<sup>35</sup> The decomposition mechanisms of tetrazoles have been previously investigated in a number of experimental and computational studies, with the prevailing decomposition mechanism being a stepwise extrusion of molecular nitrogen for both 1,5- and 2,5-substituted tetrazoles (Figure 4A and 4B).<sup>7,8,36–39</sup> Among these studies, a number of dynamic isomerization events are observed for various tetrazole substituents (-H, -OR, -NHR), leading to a variety of possible decomposition pathways and rendering the development of generalizable reactivity trends challenging. Furthermore, the existence of two constitutional tetrazole isomers (i.e., 1,5- and 2,5-tetrazoles) requires separate mechanistic consideration in the context of a traditional mechanistic study, and the generality of trends observed in such studies would be challenged by the diversity of substituents encountered in tetrazole energetics.



**Figure 4:** (A) Atom numbering scheme used in tetrazole featurization and proposed decomposition pathways for 1,5- and (B) 2,5-tetrazoles. (C) Observed natural Lewis structures for the tetrazole HEMs studied in this work.

To address these challenges using our explosophore-based approach on tetrazole HEMs, several considerations were made to adequately parameterize the structural and electronic properties of the tetrazoles in our study: 1) tetrazoles have two different substitution patterns, 2) tetrazoles can bear a plethora of substituents that significantly influence their energetic properties, and 3) many of the tetrazoles in our dataset contain multiple tetrazole explosophores. To adequately parameterize both tetrazole substitution patterns, we constructed descriptors based on a mechanistic analysis of reported decomposition mechanisms of 1,5- and 2,5-tetrazoles and devised a numbering scheme for shared bonds and common features for tetrazoles of either substitution pattern (Figure 4A and 4B).<sup>7,8,36–39</sup> Notably, several key features are maintained regardless of the tetrazole substitution pattern: the bond broken in a stepwise extrusion of molecular nitrogen is the single bond between N<sup>2</sup> and N<sup>3</sup>, a triple bond is formed between N<sup>1</sup> and N<sup>2</sup> in the liberation of nitrogen, N<sup>3</sup> substituted with R<sup>1</sup> and has a lone pair, and there is a double bond between C<sup>4</sup> and N<sup>5</sup>. Recognition of these similarities allowed for a universal description of the tetrazole system through the framework of our descriptor collection workflow. In the case of symmetric ditetrazoles, both tetrazole moieties in each

molecule were treated as equivalent and thus only one of the two tetrazoles was parameterized for modeling. This assumption of symmetry is validated by the analysis of the experimental crystal structures of these symmetric ditetrazoles, revealing that ten out of eleven constitutionally symmetric ditetrazoles in our crystal structure database are also structurally symmetric in the solid state. With these considerations in mind, we curated and parameterized several distinct tetrazole datasets to develop QSPRs for the decomposition temperature and impact sensitivity of these HEMs.

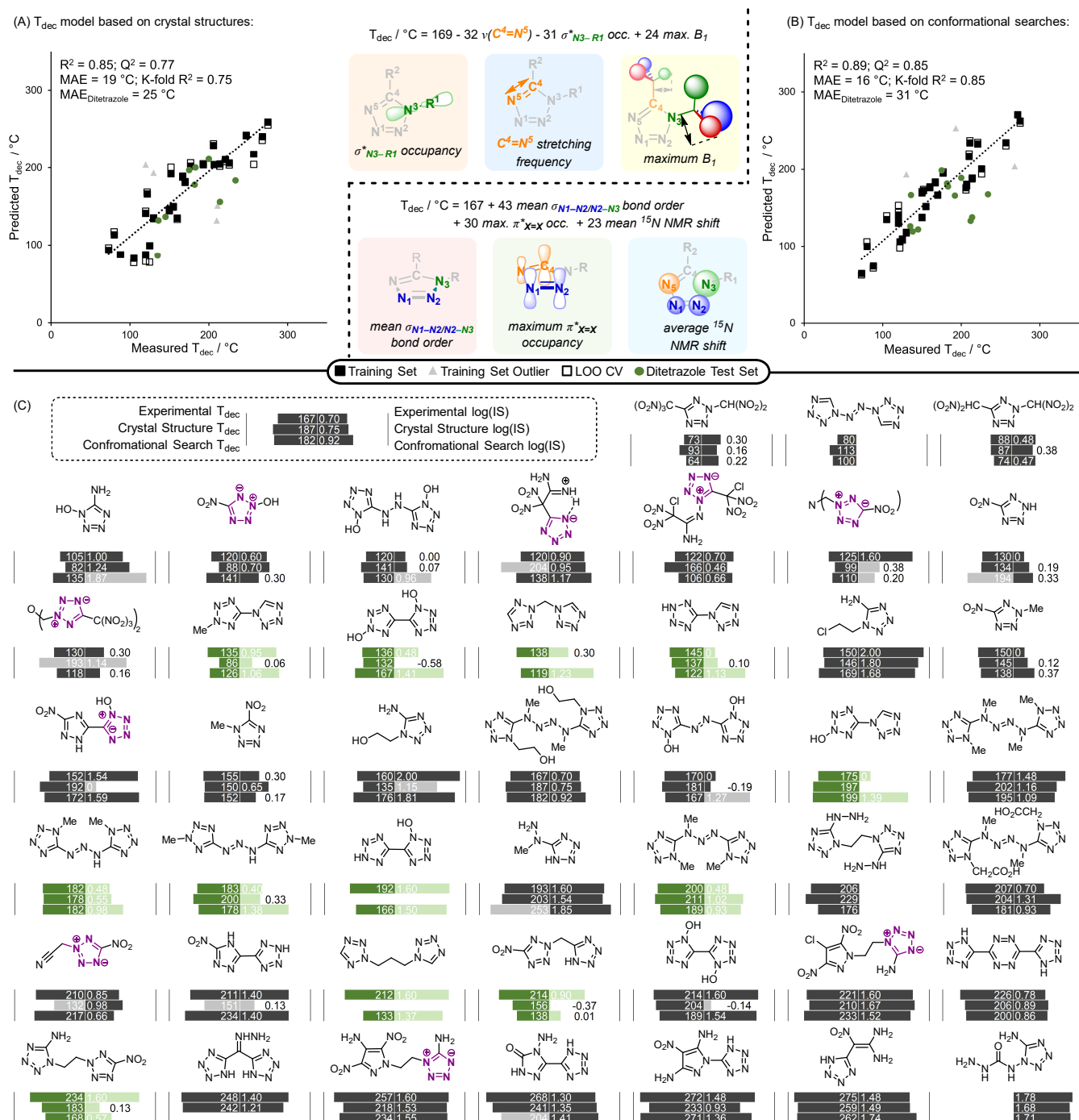
**Prediction of Tetrazole Sensitivity Properties from Crystal Structures:** Parameters were collected for a total of 33 tetrazole HEMs including twenty-four 1,5-tetrazoles and nine 2,5-tetrazoles from crystal structures. Using forward stepwise linear regression, we identified a 3-term statistical model for tetrazole decomposition temperature (Figure 5A). This model yielded predicted values of T<sub>dec</sub> within 44 °C of the experimental value for 87% of all 31 tetrazoles in the dataset with reported T<sub>dec</sub> values, with an MAE of 19 °C for the 27 tetrazoles comprising the training set.<sup>40</sup> The remaining four compounds that were predicted with larger errors are shown in Figure 5A and 5C (gray entries).

In this model, the term  $v(C^4=N^5)$  is the normalized vibrational stretching frequency of the C<sup>4</sup>=N<sup>5</sup> double bond. We observed that this frequency is influenced by the substituent on the C<sup>4</sup> carbon,<sup>41</sup> efficiently accounting for inductive effects and conjugation with the surrounding  $\pi$ -orbitals of the tetrazole. The negative sign of this descriptor in the model indicates that electron-withdrawing substituents lead to a lower decomposition temperature. The R<sup>1</sup>-N<sup>3</sup>  $\sigma^*$  occupancy descriptor is based on NBO analysis. As per the model equation, tetrazoles with a larger  $\sigma^*$  occupancy have a lower decomposition temperature. The R<sup>1</sup>-N<sup>3</sup>  $\sigma^*$  occupancy is controlled by the electronegativity of the substituent on nitrogen N<sup>3</sup> (H < C < N < O) and the tetrazole substitution pattern (1,5 < 2,5); the parameter correlates with the NBO charge on N<sup>3</sup> with a correlation coefficient of -0.78, demonstrating its relationship to the electronic properties of the ring substituent. Lastly, the *max R B<sub>1</sub>* descriptor is a Sterimol parameter that quantifies the maximal B<sub>1</sub> distance of either of the two ring substituents.<sup>42</sup> This statistical model is consistent with the main decomposition pathway proposed in literature, which features a rate-limiting cleavage of the N<sup>2</sup>-N<sup>3</sup> bond followed by nitrogen extrusion from the ring-opened intermediate (Figure 4A and 4B).<sup>7,8,36–39</sup> Electron-demanding substituents at the R<sup>1</sup> and R<sup>2</sup> positions stabilize the ensuing formal negative charge on the C<sup>4</sup>-N<sup>5</sup> motif and a sterically hindered R<sub>1</sub> substituent favors the rupture of the ring near N<sup>3</sup>.

As a next step, we sought to validate our tetrazole models using the trigger hypothesis, which states that the bulk decomposition of an energetic material is dominated by the initial decomposition of the explosophore with the lowest activation energy prior to the initiation of an exothermic chain reaction.<sup>43</sup> We explored this idea in the context of the ditetrazole HEMs in our dataset: in cases where an HEM contained two non-identical tetrazole units, we hypothesized that the molecule's bulk T<sub>dec</sub> would be closely related to the lower predicted T<sub>dec</sub> of either of the two tetrazole

explosophores. As such, we used our statistical model for  $T_{\text{dec}}$  (Figure 5A) to generate predictions for an external test set of 12 non-symmetric ditetrazoles and selected the lower of the two predicted values as the predicted  $T_{\text{dec}}$  for the HEM. This model predicted nine non-symmetric ditetrazoles with a mean absolute error of 25 °C (Figure 5A and 5C, green entries). This result serves as a good external validation for our original

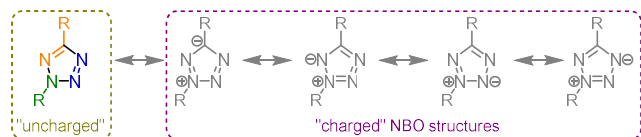
models and verifies the applicability of the trigger hypothesis for prediction of tetrazole decomposition temperature.



**Figure 5.** (A) Plot of predicted vs. experimental values of  $T_{\text{dec}}$  and the corresponding 3-term MLR model equation for the crystal structure tetrazole dataset. (B) Plot of predicted vs. experimental values of  $T_{\text{dec}}$  and the corresponding 3-term MLR model for the tetrazole dataset generated from *ab initio* conformational searches. (C) Table containing the structures and experimental and predicted values of  $T_{\text{dec}}$  and log(IS) for each tetrazole HEM in our initial tetrazole dataset. Listed predictions correspond to the models reported in (A), (B) and Figure 7.

Upon closer inspection of our models for  $T_{\text{dec}}$ , it became clear to us that several of the tetrazoles in the dataset possessed natural Lewis structures containing formal charges ("charged" NBO structures; highlighted in purple in Figure 6). The presence of these alternative Lewis structures had a significant impact on the accuracy of our  $T_{\text{dec}}$  predictions: while the MAE for tetrazoles with standard Lewis structures is 17 °C, the MAE for tetrazoles with "charged" Lewis structures is 46 °C. We thus sought to develop an expanded set of descriptors to better account for these electronic structures in future modeling efforts (*vide infra*).

Observed tetrazole Lewis structures:



**Figure 6.** Zwitterionic Lewis structures observed in the tetrazole HEMs studied in this work.

The effects of charged tetrazole Lewis structures were also observed when we applied this workflow to the prediction of tetrazole impact sensitivity, which yielded models with poorer performance. Our best model for  $\log(\text{IS})$  using the crystal structure dataset (Figure 5A) necessitated the removal of six compounds before satisfactory model statistics were obtained ( $R^2 = 0.86$ ,  $Q^2 = 0.83$ ,  $\text{MAE} = 0.17$ ,  $K\text{-fold } R^2 = 0.75$ ). This can largely be attributed to the contributions of 9 out of 31 tetrazole HEMs with "charged" Lewis structures; while the 22 tetrazoles with "uncharged" Lewis structures were modeled with an MAE of 0.34, the remaining 9 compounds with "charged" NBO structures had an MAE of 0.47. While the MLR equation of the  $\log(\text{IS})$  model retained good interpretability (Figure 7A), its shortcomings highlighted the necessity for construction of parameters that accounted for multiple natural Lewis structures.

**Tetrazole Decomposition Temperature Prediction from Conformational Searches:** Motivated by our initial success in identifying QSPR models for the decomposition temperature of structurally diverse tetrazole HEMs using descriptor sets constructed from crystallographic data, we sought to extend our methods to accommodate *ab initio* virtual screening of HEMs. Our reliance on crystal structure data would necessitate time-consuming and potentially hazardous laboratory work to test model predictions. To this end, we developed a second workflow for QSPR modeling of tetrazole sensitivity properties wherein 2D structural formulae were converted into 3D conformers using molecular mechanics with MacroModel version 11.8. All conformers within a 5.0 kcal/mol energy difference of the lowest energy conformer were subjected to the same DFT optimizations and single point calculations for parameter acquisition as described in the original workflow. With multiple conformations for each compound in hand, we generated two descriptor datasets for statistical applications: 1) the minimum energy conformer of each HEM according to DFT calculations and 2) Boltzmann averaging of descriptor values across all conformations.

As previously discussed, the multiple natural Lewis structures encountered within the tetrazole dataset (Figure 6) prompted us to construct additional NBO parameters. Due to the

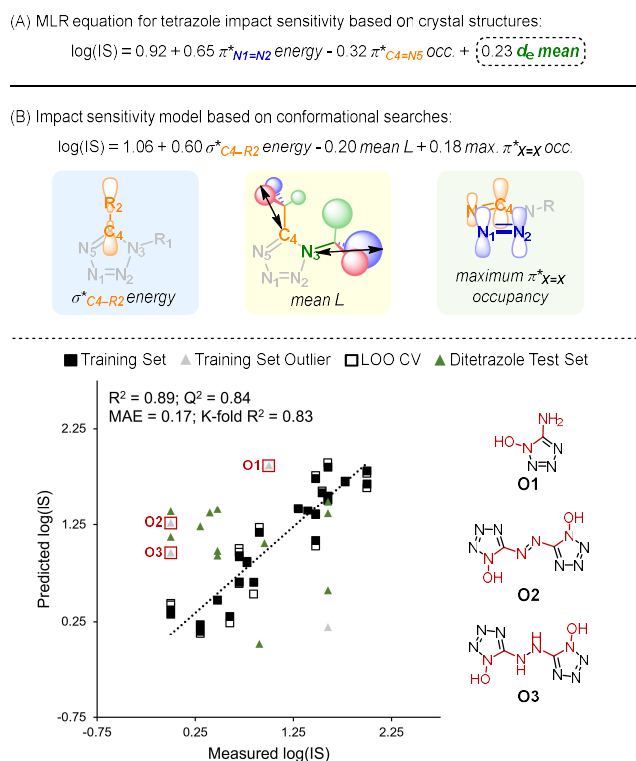
aromaticity of tetrazoles, we started by constructing minima, maxima, sums, and averages of NBO orbital occupancies, energies, bond orders, and energy gaps for the calculated localized  $\pi$  NBOs. Importantly, multiple Lewis structures were unified by identifying NBOs that are related by resonance. For instance, the natural Lewis structure of some tetrazoles contained an  $\text{N}^2=\text{N}^3$  double bond ("charged" NBO structures) whereas in the majority of tetrazoles it is localized as a  $\text{N}^1=\text{N}^2$  double bond ("uncharged" NBO structures). Hence, these  $\pi$  orbitals were treated as a single structural feature for the design of additional descriptors. Closer inspection of the  $\sigma$  NBO energies and orbital occupancies of tetrazoles with "charged" NBO structures revealed systemic alterations based on the calculated Lewis structure of the tetrazole. Analogous feature engineering was thus also performed on  $\sigma$  NBO orbitals. Taken together, these improvements offered a more uniform description of the  $\pi$ -system and the  $\sigma$  skeleton of tetrazoles in a manner agnostic of their Lewis structures.

Statistical models were generated with this new descriptor set following the same general workflow outlined in Figure 3. Intriguingly, the Boltzmann-averaged descriptor dataset provided better statistical models for thermal decomposition, while the minimum energy conformer dataset proved more effective for prediction of the impact sensitivity (*vide infra*). This result is consistent with the observation that many of the tetrazoles in our experimental dataset have a melting point below their decomposition temperature. Consequently, these tetrazoles have a high degree of conformational flexibility prior to decomposition, which is suitably approximated by the Boltzmann averaging of the descriptor set. However, decomposition initiated by impact occurs in the solid state, in which HEMs generally adopt a single conformation in a crystal lattice. The statistical model for  $T_{\text{dec}}$  derived from the Boltzmann-averaged parameters (Figure 5B) carries excellent statistical metrics, yielding predicted  $T_{\text{dec}}$  values within 30 °C of the experimental value for 90% of the 31 tetrazoles in the dataset, with an MAE of 16 °C for the 28 tetrazoles comprising the training set. The improved performance of this statistical model as compared to the model using descriptors derived from the crystal structure conformers is also attributed to the expanded set of NBO descriptors used in this second generation of modeling. This improvement is highlighted by the inclusion of the *average  $\text{N}^1\text{--N}^2/\text{N}^2\text{--N}^3$   $\sigma$  bond order* descriptor, which was constructed to account for tetrazoles (3 out of 31) that have a double bond between  $\text{N}^2$  and  $\text{N}^3$  in their natural Lewis structures. Closer inspection of this descriptor reveals a correlation coefficient with the  $\text{N}^2\text{--N}^3$   $\sigma$  bond order of 0.87, while the descriptor is not correlated with the  $\text{N}^1\text{--N}^2$   $\sigma$  bond order (correlation coefficient = 0.04). The large positive model coefficient for the *average  $\text{N}^1\text{--N}^2/\text{N}^2\text{--N}^3$   $\sigma$  bond order* in the MLR equation (Figure 5B) supports a rate-determining  $\text{N}^2\text{--N}^3$  bond breakage in the decomposition of both tetrazole classes. The  $X=X^*$  *maximum occupancy* descriptor is the highest occupancy in any  $\pi^*$  orbital within the tetrazole ring, accounting for any tetrazole resonance structures. The third descriptor in the model is the average calculated NMR shift of all 4 nitrogen atoms in the ring.

To validate this model with an external test set, the decomposition temperatures of our set of 12 non-symmetric ditetrazoles were predicted. The ditetrazoles were

parameterized in the same fashion as the training set using structures generated from conformational searches and subsequent DFT computations. As before, the lower predicted value between the two tetrazole moieties was adopted as the predicted property for the HEM in accordance with the trigger hypothesis. This resulted in good predictions for the decomposition temperature for 9 of the 12 ditetrazoles, giving a MAE of 16 °C. The total MAE for all 12 ditetrazoles was 31 °C.

**Tetrazole Impact Sensitivity Prediction from Conformational Searches:** Tetrazole impact sensitivity was successfully modeled using the minimum energy conformer descriptor set (Figure 7B). This model yielded log(IS) predictions for 87% of the 31 tetrazoles within 0.39 of the experimental value with a training set MAE of 0.17. As in the model for  $T_{\text{dec}}$  (Figure 5B), the parameter  $X=X^*$  maximum occupancy appears with a positive coefficient in the model for log(IS). The *ave R L* parameter is a Sterimol parameter that describes the average length of both substituents on each tetrazole. Three of the four log(IS) outliers (Figure 7B, gray points) contained a 5-amino-1*H*-tetrazol-1-ol subunit and were predicted with a higher impact sensitivity than the experimental value, suggesting that this motif may result in intermolecular interactions not captured in the conformational search model. This is further supported by the log(IS) model based on the crystal structure (Figure 7A), which gave better predictions for all these three compounds (Figure 5C). The



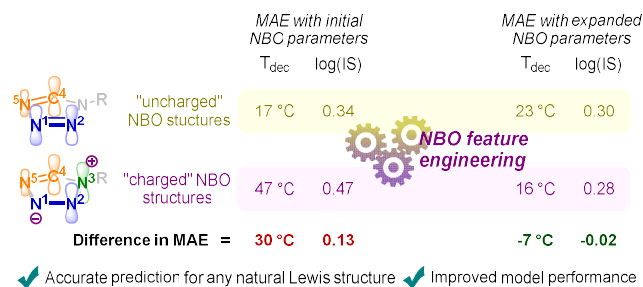
**Figure 7.** (A) 3-term MLR model for tetrazole impact sensitivity based on the crystal structure dataset. (B) 3-term MLR model and a corresponding plot of experimental vs. predicted log(IS) values for tetrazole impact sensitivity based on structures generated from *ab initio* conformational searches. The structures of the three training set outliers are displayed to the right of the plot.

improved performance of the crystal structure model may be attributed to the Hirshfeld parameter  $d_e \text{ mean}$  describing the average intermolecular distance between molecules in the solid state. The inclusion of the  $d_e \text{ mean}$  term in this model supports the notion that tighter crystal packing increases the sensitivity of tetrazole HEMs.

Prediction of the ditetrazole test set using this impact sensitivity model in conjunction with the trigger hypothesis failed to produce good predictions (Figure 7B, green triangles). We propose two explanations for this outcome. Firstly, many of the ditetrazoles used in the external test set are fully conjugated. Our model for tetrazole impact sensitivity largely featured descriptors that describe the  $\pi$ -system of isolated tetrazole motifs. Consequently, the extended conjugation of the non-symmetric ditetrazole test set might produce cooperative effects that our model fails to explain. An alternative explanation is that the trigger hypothesis is simply not applicable for impact-initiated decomposition of tetrazoles. The investigation of this question is an area of ongoing research.

### Performance Improvements from NBO Feature Engineering

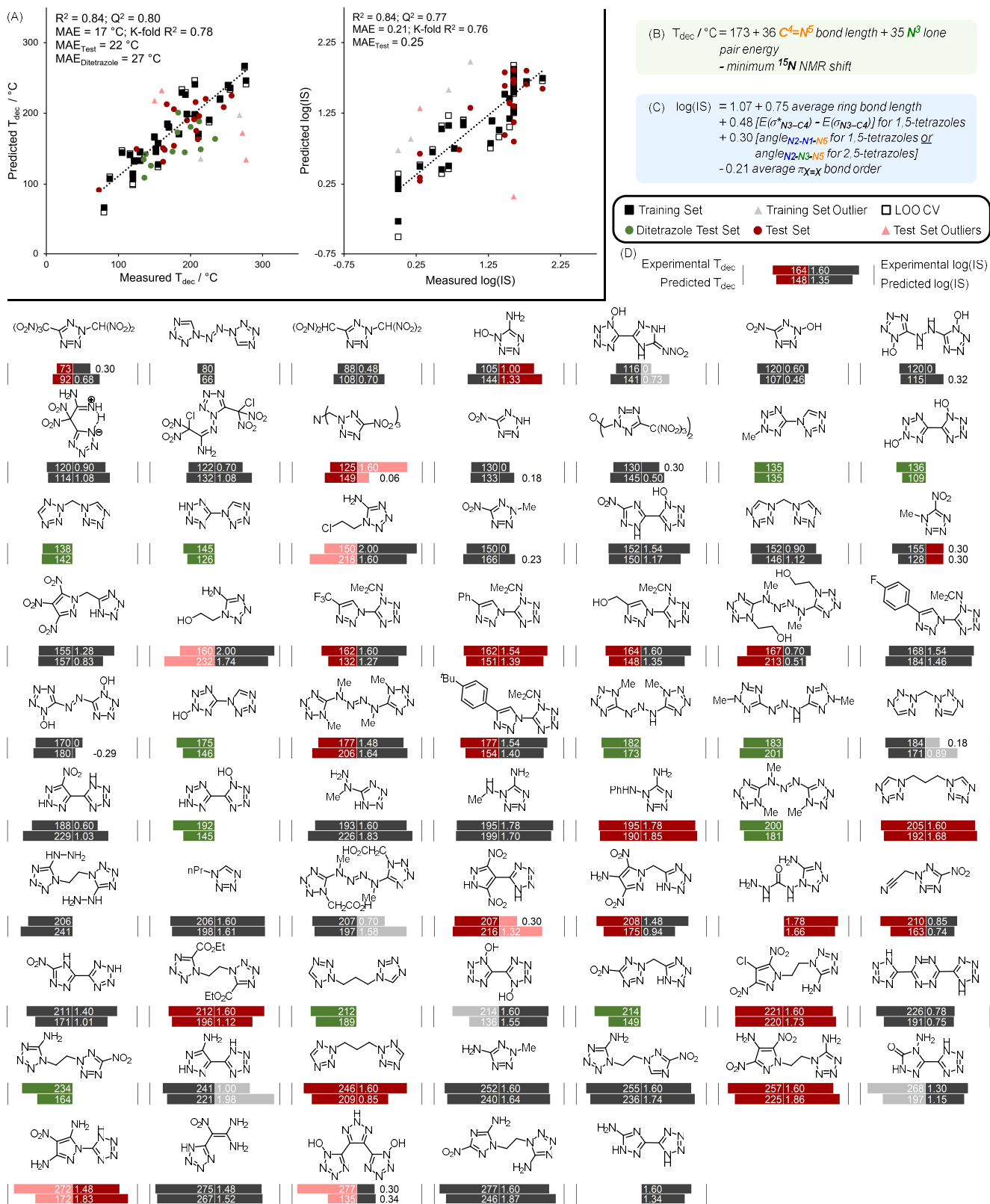
The influence of NBO feature engineering on model performance is illustrated through comparison between the performance of the best  $T_{\text{dec}}$  and log(IS) models generated using the initial parameter set with the performance of the best models developed with the extended parameter set. A useful metric for the ability of a descriptor set to account for “charged” NBO structures is the comparison of the MAE obtained for the dominant natural Lewis structure and that of the charged set (Figure 8). Our expanded suite of descriptors substantially reduced the difference in MAE between these two groups of tetrazoles for both the prediction of  $T_{\text{dec}}$  (30 to -7 °C) and log(IS) (0.13 to -0.02).



**Figure 8.** Improvements to tetrazole  $T_{\text{dec}}$  and log(IS) model performance upon the inclusion of engineered NBO descriptors. The difference in MAE between charged and uncharged Lewis structures is decreased for both  $T_{\text{dec}}$  and log(IS) upon the introduction of engineered descriptors.

### Tetrazole Decomposition Temperature and Impact Sensitivity Prediction Using an Expanded Tetrazole Dataset

After demonstrating proof-of-principle for *ab initio* prediction of tetrazole  $T_{\text{dec}}$  and log(IS), we conducted a comprehensive literature survey using *SciFinder*<sup>n</sup> for additional energetic tetrazoles disclosed by Klapötke and Shreeve that did not have reported crystal structure data. This allowed us to add an additional 24 structures to our library and produce an expanded dataset of 57 tetrazole HEMs (Figure 9). We again used molecular mechanics to obtain a series of conformers for each compound and then applied our



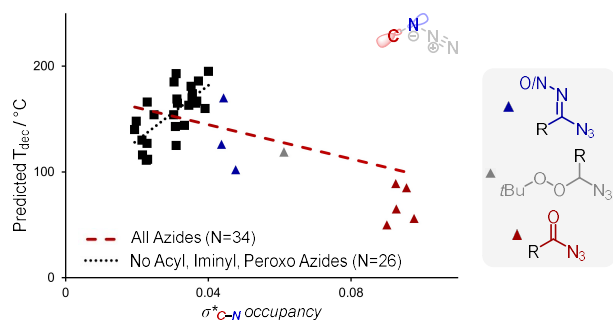
**Figure 9:** (A) Plot of predicted vs. experimental values of  $T_{\text{dec}}$  and  $\log(\text{IS})$  for the expanded tetrazole dataset using datasets generated from conformational searches. MLR models for (B)  $T_{\text{dec}}$ , 3-term, and (C)  $\log(\text{IS})$ , 4-term, for the expanded conformational search tetrazole dataset. (D) Table containing the structures and experimental and predicted values of  $T_{\text{dec}}$  and  $\log(\text{IS})$  for each tetrazole HEM in the expanded tetrazole dataset. Predictions correspond to the models reported in (B), (C) for the Kennard-Stone test and training set, as well as the non-symmetric ditetrazole dataset.

statistical modeling methods towards the prediction of  $T_{\text{dec}}$  (using a Boltzmann averaged dataset) and  $\log(\text{IS})$  (using a minimum energy conformer dataset). The 54 tetrazoles in this dataset with reported  $T_{\text{dec}}$  values were split into a 60:40 train:test set (33 tetrazoles in training set and 21 tetrazoles in test set) using the Kennard-Stone algorithm<sup>44</sup> to provide a reproducible uniform distribution of sensitivity property values for model development.

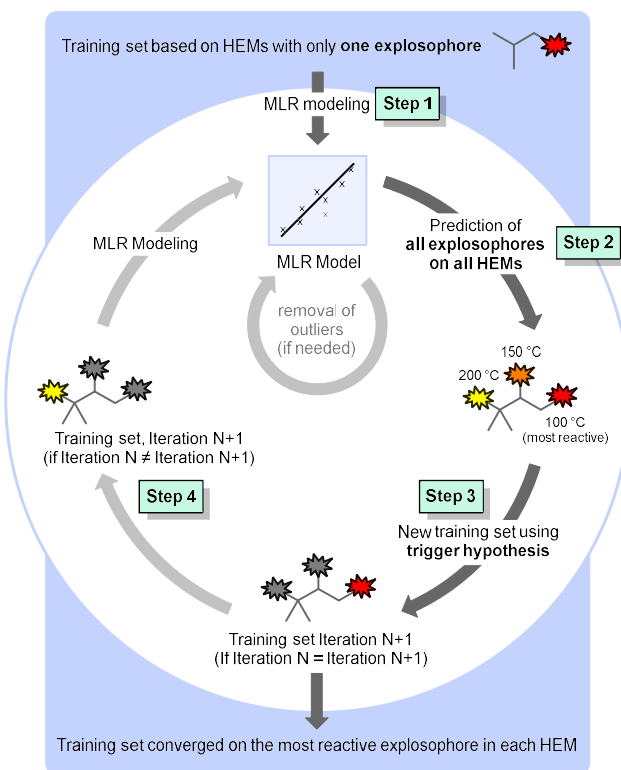
The training set of 33 tetrazoles resulted in a 3-term model for  $T_{\text{dec}}$  (Figure 9B) with good statistical metrics. This model yielded predicted  $T_{\text{dec}}$  values within 41 °C of the experimental value for 93% of the tetrazoles in the training set with an MAE of 17 °C. Two compounds were excluded from the training set due to their outlier status as evidenced by large prediction errors (70 °C, 78 °C). This model was then validated with two external validation sets: the remaining test set of 21 tetrazoles and the test set of 12 non-symmetric ditetrazoles. Prediction of the 33 compounds in these validation sets resulted in an MAE of 25 °C (four outliers with prediction errors of 68 °C, 72 °C, 100 °C, and 142 °C). The experimental decomposition temperatures of tetrazoles used in this model span a range of over 200 °C, highlighting the utility of this predictive model.

## QSPR for Azide HEMs

**Prediction of Decomposition Temperature and Impact Sensitivity of Azide HEMs:** Organic azides are of interest for their promise as replacements for toxic metal-based primary explosives.<sup>18</sup> In addition to their direct use as energetics, azides are ubiquitous handles in the preparation of heterocyclic HEMs, bioactive molecules, and materials.<sup>45–48</sup> However, to the best of our knowledge, no QSPRs have been reported describing the sensitivity properties of small molecule azide HEMs. To further extend the application of our methods for sensitivity property prediction, we turned our attention to the 35 azide-containing HEMs in our compound library (Figure 12D). Many of these HEMs included multiple azide functional groups per molecule (10 of 35 HEMs, for a total of 54 azide explosophores) in addition to non-azide explosophores. As was



**Figure 10.** Univariate correlation between the lowest predicted azide  $T_{\text{dec}}$  and the  $\sigma^*(\text{C}-\text{N}_3)$  occupancy descriptor. Categorical differences between acyl azides (red), iminyl azides (blue), peroxo azides (gray), and the remaining azide HEMs (black) are apparent. The linear fit corresponding to the entire azide dataset of 34 HEMs given by the red trace is driven by outlier azides. The linear fit corresponding to the remaining 26 azides is given by the black trace.



**Figure 11.** An iterative approach to modeling HEMs with multiple explosophores. Based on the trigger hypothesis, each HEM is represented by one of its explosophores. After an initial phase of MLR modeling, the explosophore that is predicted to be most reactive (lowest  $T_{\text{dec}}$  or lowest  $\log(\text{IS})$ ) in the  $N^{\text{th}}$  iteration is taken forward for dataset of the next iteration  $N+1$ . During the MLR modeling steps, judicious removal of outliers may significantly improve the model performance. A quantitative metric for outlier selection is a decreased difference between  $R^2$  and  $Q^2$  upon removal of a suspected outlier HEM.

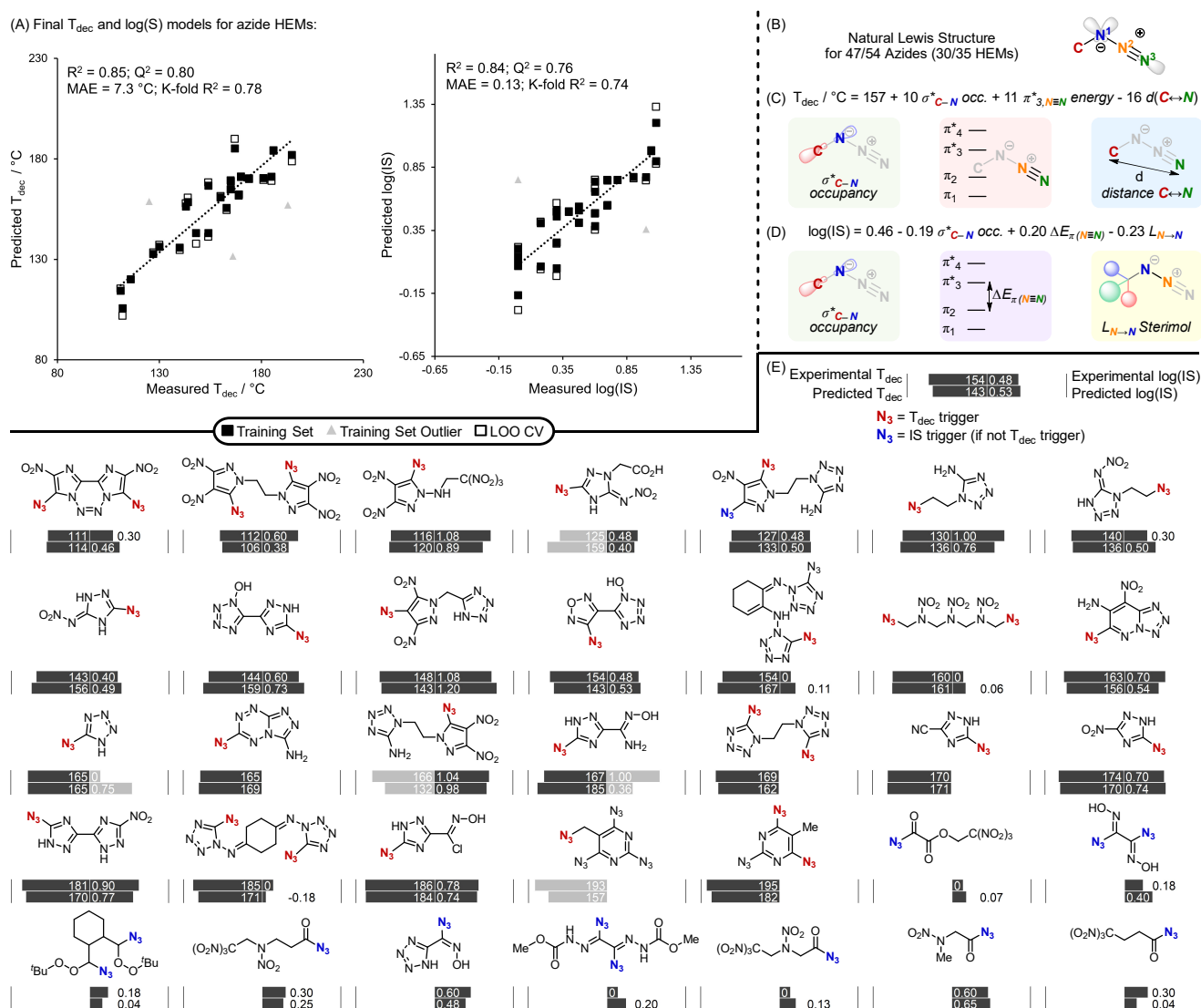
the case with the tetrazole/ditetrazole datasets, we initially modeled azides containing only one azide explosophore, and then expanded the scope of our prediction to encompass HEMs with multiple azide explosophores.

During curation of the azide dataset, it became apparent that acyl azides possess  $T_{\text{dec}}$  values that are categorically lower than the  $T_{\text{dec}}$  of azides attached to other functionalities. The best  $T_{\text{dec}}$  model generated (*vide infra* for modeling workflow) for the full azide dataset ( $N = 34$ ,  $R^2 = 0.75$ ,  $Q^2 = 0.69$ ) contained an appreciable univariate correlation ( $R^2 = 0.47$ ) between  $T_{\text{dec}}$  and the  $\sigma^*(\text{C}-\text{N}_3)$  occupancy descriptor (Figure 9). For the full azide dataset, the correlation coefficient corresponding to this plot is given as -0.69; however, upon removal of the nine acyl, iminyl, and peroxy azides, this correlation coefficient is given as +0.68. This categorical difference is easily visualized in Figure 10 and carries an intuitive interpretation. In analogy to a concave upward nonlinear Hammett plot, this indicates that these azides follow a different decomposition mechanism than the remainder of the azide HEMs.<sup>49</sup> This finding is in agreement with the literature, as acyl azides can readily undergo a Curtius rearrangement to furnish the corresponding isocyanate with

the liberation of molecular nitrogen.<sup>50</sup> This concerted reaction mechanism is not accessible to other classes of azides, which most likely decompose via transient nitrene species and the liberation of molecular nitrogen.<sup>44</sup> As such, models generated including these azides are a useful tool in mechanistic classification but fail to provide the detailed chemical interpretability we desired. This mechanistic divergence prompted us to exclude the nine acyl, peroxy, and iminyl azides in our dataset from further  $T_{\text{dec}}$  modeling efforts.

After this initial exploration, we systematically developed models for  $T_{\text{dec}}$  for the remaining 26 azide HEMs in our study. We first generated an initial training model using the 19 HEMs containing only a single azide explosophore (Figure 11, Step 1) see SI for model equation and statistics). We then applied an iterative workflow to expand the scope of our prediction to the remaining 7 HEMs with multiple azide explosophores per

molecule. This iterative process is described as follows and is demonstrated in detail in the SI. First, the training model for HEMs with one azide explosophore was used to generate  $T_{\text{dec}}$  predictions for *each* azide explosophore in *every* HEM in the dataset (Figure 11, Step 2). For HEMs with multiple azide explosophores, the azide group predicted to be the most sensitive (i.e., lowest predicted  $T_{\text{dec}}$ ) is added to the training dataset for a subsequent generation of modeling (Figure 11, Step 3). At this point, all HEMs were now included in this updated training set. A new generation of models was produced from this training set, and the azide explosophore in each molecule with the lowest predicted  $T_{\text{dec}}$  was again compiled into a new dataset. This process was repeated until dataset convergence was reached, i.e., once explosophore selection for every HEM in the dataset was the same as in the previous iteration of modeling (Figure 11, Step 4). In our experience, convergence was achieved within 2–4 iterations of



**Figure 12:** (A) Plot of predicted vs. experimental values of  $T_{\text{dec}}$  and  $\log(\text{IS})$  for the azide dataset generated from experimental crystal structures. (B) Observed NBO natural Lewis structure for the majority of azide compounds in the dataset. 3-term MLR models for (C)  $T_{\text{dec}}$  and (D)  $\log(\text{IS})$  for the crystal structure azide dataset. (E) Table containing the structures and experimental and predicted values of  $T_{\text{dec}}$  and  $\log(\text{IS})$  for each azide HEM studied in this work. Predictions correspond to the models reported in (C) and (D).

this workflow. After three iterations of modeling, our workflow had converged on a final training set and a corresponding predictive model for azide decomposition temperature (Figure 12C).<sup>51</sup> This model yielded  $T_{\text{dec}}$  predictions with 18 °C of the experimental value for 88% of the 26 azides in the training set with a MAE of 7.3 °C.

The impact sensitivity of the azide HEMs was modeled following the same principles. Experimental values of  $\log(\text{IS})$  for 30 of the 35 compounds were distributed between 0 and 1.1, providing a well-behaved dataset for modeling efforts. The five remaining azides outside of this sensitivity range were excluded from modeling, as extreme datapoints exert disproportionate influence on the models. To establish a starting point for modeling, a  $\log(\text{IS})$  model based only on HEMs bearing one azide explosophore was first constructed. Subsequent modeling efforts reached convergence after two iterations of modeling. The final model for impact sensitivity contained 28 compounds; this model yielded  $\log(\text{IS})$  predictions within 0.25 of the experimental value for 93% of the 30 azides in the training set with a MAE of 0.13 (Figure 12D). Notably, acyl, iminyl, and peroxy azides were successfully predicted in the final model.

Analysis of calculated NBO orbitals for the azide dataset reveals that most azides share a common natural Lewis structure (Figure 12B). The models for  $T_{\text{dec}}$  and  $\log(\text{IS})$  can both be readily understood and share analogous parameters. Both statistical models contain  $\sigma^*(\text{C}-\text{N}_3)$  occupancy descriptor; the coefficient for this descriptor is positive for  $T_{\text{dec}}$  and negative for  $\log(\text{IS})$ . This parameter serves as a good readout for the inductive effects of substituents on the carbon of the azide group. Furthermore, both models contain a parameter that describes the NBO LUMO orbital of the  $\text{N}_2\equiv\text{N}_3$  triple bond. The  $T_{\text{dec}}$  model indicates that azides with lower  $\text{N}_2\equiv\text{N}_3$  NBO LUMO energies are more sensitive, and the  $\log(\text{IS})$  model suggests that compounds with a smaller  $\text{N}_2\equiv\text{N}_3$  NBO HOMO/LUMO energy gap are more sensitive. These observations provide a consistent mechanistic picture of azide decomposition via extrusion of molecular nitrogen. Lastly, both models contain a parameter that describes the geometry/steric environment of the azide. In our model for  $T_{\text{dec}}$ , azides with a longer azide carbon-terminal nitrogen are predicted to have a lower  $T_{\text{dec}}$ . In our model for  $\log(\text{IS})$ , azides with bulkier substituents are more sensitive.

Notably, very low mean absolute errors were achieved for both  $T_{\text{dec}}$  and  $\log(\text{IS})$  (7.3 °C and 0.13, respectively), testifying to the accuracy of this method for predictive modeling of azide-containing HEMs. Good statistical performance was achieved with removal of very few outliers, despite the diversity of the dataset with regards to the presence of various additional explosophores. Furthermore, the statistical models were constructed from descriptors unique to the azide moiety and were thus largely agnostic of other explosophores contained in the molecule. This corroborates that the azides likely serve as triggers for explosive decomposition, even in the presence of other explosophores.

## Conclusion

In summary, we report a series of computational and statistical methods that enabled prediction of decomposition temperature and impact sensitivity across a structurally diverse scope of tetrazole and azide HEMs. The key conceptual innovation in this approach is the decision to interrogate the structural and quantum mechanical properties of HEMs at the level of their explosophores, providing the ability to describe and interpret mechanistic events leading to decomposition while also modeling sensitivity properties. The advantages of this descriptor suite are compounded by the fact that descriptors are localized at the level of individual chemical bonds but can readily be expanded to larger chemical motifs, allowing a common set of parameters to be generated for an explosophore of user-specified structure and a sensitivity property of interest. A comprehensive account of the computational methods used to construct these models is disclosed in the SI.

To the best of our knowledge, this work contains the first QSPRs for the sensitivity properties of tetrazole and azide HEMs in the reported literature. The models disclosed herein contain interpretable terms such as charges, NBO orbitals, and steric parameters, thereby offering chemical insight to guide the development of new HEMs with improved sensitivity properties. In contrast to literature precedence, the models in this work were trained and tested on new HEM datasets with high structural diversity, demonstrating the generality of our methods for chemical featurization of HEMs. Internal and external validation of our models further supports their validity as useful heuristics for rational design and mechanistic study.

The marriage of our explosophore-based lens for sensitivity property prediction with the trigger hypothesis subsequently allowed us to predict sensitivity properties for HEMs of increased structural complexity. Accurate sensitivity property predictions for HEMs containing multiple azide and tetrazole explosophores were obtained through simultaneous modeling of each of the explosophores in these HEMs, an ability unique to our explosophore-based approach. Successful modeling of sensitivity properties through this workflow lends support to the notion that specific explosophores act as molecular triggers for the bulk decomposition of HEMs. However, our efforts to predict the impact sensitivity of conjugated non-symmetric ditetrazoles in accordance with the trigger hypothesis proved unsuccessful, highlighting a limitation of this approach. Impact sensitivity predictions for these molecules may require the construction of additional descriptors that better capture the electronic structures characteristic of these HEMs.

Finally, we demonstrated the application of our methods towards streamlined virtual screening of HEMs. In contrast to our earlier models that utilized parameters calculated from experimental crystal structures, we successfully predicted the sensitivity properties of tetrazoles whose 3D structures were obtained from 2D structural formulae alone. Future expansions of this computational workflow to other compound classes, while not demonstrated in this work, have the potential to enable virtual screening of various other HEM families to improve the identification of viable synthetic targets prior to any experimental work. Derivatization of our workflow to include HEMs with ionic structures or even composite energetic formulations would further support the rational design of future energetic materials.

## Author Information

### Corresponding Authors

**Song Lin** – Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, United States; orcid.org/0000-0002-8880-6476; Email: songlin@cornell.edu

**Matthew S. Sigman** – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States; orcid.org/0000-0002-5746-8830; Email: sigman@chem.utah.edu

### Authors

**Jonas Rein** – Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, United States; orcid.org/0000-0001-8237-6519

**Jonathan M. Meinhardt** – Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, United States; orcid.org/0000-0001-6880-5938

**Julie L. Hofstra Wahlman** – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States; orcid.org/0000-0001-9558-4317

### Author Contributions

† J.R. and J.M.M. contributed equally to this work.

### Notes

The authors declare no competing financial interest.

## Acknowledgements

Financial support was provided by Office of Naval Research's Young Investigator Program (S.L.) and the National Science Foundation (CHE-1763436 (M.S.S)). Computational resources were provided from the Center for High Performance Computing (CHPC) at the University of Utah. S.L. is grateful to the Camille and Henry Dreyfus Foundation for a Teacher-Scholar Award. J.R. acknowledges support from the ERP Fellowship from the Studienstiftung des Deutschen Volkes. J.L.H.W. acknowledges support from the National Institutes of Health for a postdoctoral research fellowship (F32GM134613). We thank Dr. Christopher Sandford for helpful discussions relating to DFT calculations and chemical featurization.

## Abbreviations

HEM, high-energy material; MLR, multivariate linear regression; QSPR, quantitative structure-property relationship;  $T_{\text{dec}}$ , decomposition temperature; IS, impact sensitivity.

## References

- (1) Sucasca, M.; Dobrilovic, M.; Bohanek, V.; Stimac, B. Estimation of Explosive Energy Output by EXPLO5 Thermochemical Code. *Zeitschrift für Anorg. und Allg. Chemie* **2021**, 647 (4), 231–238.
- (2) Chandrasekaran, N.; Oommen, C.; Kumar, V. R. S.; Lukin, A. N.; Abrukov, V. S.; Anufrieva, D. A. Prediction of Detonation Velocity and N–O Composition of High Energy C–H–N–O Explosives by Means of Artificial Neural Networks. *Propellants, Explos. Pyrotech.* **2019**, 44 (5), 579–587.
- (3) Keshavarz, M. H. Simple Correlation for Predicting Detonation Velocity of Ideal and Non-Ideal Explosives. *J. Hazard. Mater.* **2009**, 166 (2–3), 762–769.
- (4) Elton, D. C.; Boukouvalas, Z.; Butrico, M. S.; Fuge, M. D.; Chung, P. W. Applying Machine Learning Techniques to Predict the Properties of Energetic Materials. *Sci. Rep.* **2018**, 8 (1), 9059.
- (5) Green, S. P.; Wheelhouse, K. M.; Payne, A. D.; Hallett, J. P.; Miller, P. W.; Bull, J. A. Thermal Stability and Explosive Hazard Assessment of Diazo Compounds and Diazo Transfer Reagents. *Org. Process Res. Dev.* **2020**, 24 (1), 67–84.
- (6) Scriven, E. F. V.; Turnbull, K. Azides: Their Preparation and Synthetic Uses. *Chem. Rev.* **1988**, 88 (2), 297–368.
- (7) Kiselev, V. G.; Cheblakov, P. B.; Gritsan, N. P. Tautomerism and Thermal Decomposition of Tetrazole: High-Level Ab Initio Study. *J. Phys. Chem. A* **2011**, 115 (9), 1743–1753.
- (8) Yuan, B.; Bernstein, E. R. Initial Mechanisms for the Unimolecular Decomposition of Electronically Excited Nitrogen-Rich Energetic Materials with Tetrazole Rings: 1-DTE, 5-DTE, BTA, and BTH. *J. Chem. Phys.* **2016**, 144 (23), 234302.
- (9) Brill, T. B.; James, K. J. Kinetics and Mechanisms of Thermal Decomposition of Nitroaromatic Explosives. *Chem. Rev.* **2002**, 93 (8), 2667–2692.
- (10) Xu, J.; Zhu, L.; Fang, D.; Wang, L.; Xiao, S.; Liu, L.; Xu, W. QSPR Studies of Impact Sensitivity of Nitro Energetic Compounds Using Three-Dimensional Descriptors. *J. Mol. Graph. Model.* **2012**, 36, 10–19.
- (11) Mathieu, D.; Alaime, T. Impact Sensitivities of Energetic Materials: Exploring the Limitations of a Model Based Only on Structural Formulas. *J. Mol. Graph. Model.* **2015**, 62 (2), 81–86.
- (12) Yuan, B.; Bernstein, E. R. Initial Mechanisms for the Unimolecular Decomposition of Electronically Excited Nitrogen-Rich Energetic Materials with Tetrazole Rings: 1-DTE, 5-DTE, BTA, and BTH. *J. Chem. Phys.* **2016**, 144 (23), 234302.
- (13) Furman, D.; Kosloff, R.; Dubnikova, F.; Zybin, S. V.; William A. Goddard, I.; Rom, N.; Hirshberg, B.; Zeiri, Y. Decomposition of Condensed Phase Energetic Materials: Interplay between Uni- and Bimolecular Mechanisms. *J. Am. Chem. Soc.* **2014**, 136 (11), 4192–4200.
- (14) Ghule, V. D. Computational Studies on Energetic Properties of Trinitro-Substituted Imidazole-Triazole and Pyrazole-Triazole Derivatives. *J. Phys. Chem. A* **2012**, 116 (37), 9391–9397.
- (15) Hao, L.; Wang, J.; Zhai, D.; Ma, P.; Ma, C.; Pan, Y.; Jiang, J. Theoretical Study on CL-20-Based Cocrystal Energetic

- Compounds in an External Electric Field. *ACS Omega* **2020**, *5* (24), 14767–14775.
- (16) Adams, G. F.; Shaw R. W. Chemical Reactions in Energetic Materials. *Annu. Rev. Phys. Chem.* **2003**, *43* (1), 311–340.
- (17) Santiago, C. B.; Guo, J. Y.; Sigman, M. S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**, *9* (9), 2398–2412.
- (18) Mehta, N.; Oyler, K.; Cheng, G.; Shah, A.; Marin, J.; Yee, K. Primary Explosives. *Zeitschrift für Anorg. und Allg. Chemie* **2014**, *640* (7), 1309–1313.
- (19) Huynh, M. H. V.; Coburn, M. D.; Meyer, T. J.; Wetzler, M. Green Primary Explosives: 5-Nitrotetrazolato-N2-Ferrate Hierarchies. *Proc. Natl. Acad. Sci.* **2006**, *103* (27), 10322–10327.
- (20) Zhang, Q.; Chen, D.; Jing, D.; Fan, G.; He, L.; Li, H.; Wang, W.; Nie, F. Access to Green Primary Explosives via Constructing Coordination Polymers Based on Bis-Tetrazole Oxide and Non-Lead Metals. *Green Chem.* **2019**, *21* (8), 1947–1955.
- (21) Lothrop, W. C.; Handrick, G. R. The Relationship between Performance and Constitution of Pure Organic Explosive Compounds. *Chem. Rev.* **2002**, *44* (3), 419–445.
- (22) Many previous machine learning approaches to QPSRs for HEMs utilize a sensitivity property dataset based primarily on data acquired by Storm et. al. (258 HEMs) (Storm, C. B.; Stine J. R.; Kramer, J. F. *Chemistry and Physics of Energetic Materials* **1990**, 605-639). The original storm dataset has since been expanded by Mathieu (308 HEMs), (Mathieu, D. *Ind. Eng. Chem. Res.* **2017**, *56*, 8191–8201). Our dataset is composed of compounds reported after 2000 which are not included in these datasets.
- (23) The compounds in our dataset were compiled from a minimal number of research groups with the intention of ensuring maximum consistency between sensitivity measurements prior to statistical modeling efforts.
- (24) Crystal structures of these materials were collected to avoid the resource intensive step of conformational searching. We later demonstrate that our methods can be accurately applied for compounds lacking experimental crystal structures.
- (25) Badgujar, D. M.; Talawar, M. B.; Asthana, S. N.; Mahulikar, P. P. Advances in Science and Technology of Modern Energetic Materials: An Overview. *J. Hazard. Mater.* **2008**, *151* (2–3), 289–305.
- (26) Decomposition temperatures for all compounds in the dataset were measured via DSC with heating rates of 5 °C/min.
- (27) Impact sensitivities for all compounds in the dataset were measured using the BAM Fall Hammer apparatus. Details of these measurements are provided in the publications listed in our datasets.
- (28) Kamlet, M. J.; Adolph, H. G. The Relationship of Impact Sensitivity with Structure of Organic High Explosives. II. Polynitroaromatic Explosives. *Propellants, Explos. Pyrotech.* **1979**, *4* (2), 30–34.
- (29) SMARTS Theory Manual, Daylight Chemical Information Systems **2019**. Available at: <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>.
- (30) Reichel, M.; Dosch, D.; Klapötke, T.; Karaghiosoff, K. Correlation between Structure and Energetic Properties of Three Nitroaromatic Compounds: Bis(2,4-Dinitrophenyl) Ether, Bis(2,4,6-Trinitrophenyl) Ether, and Bis(2,4,6-Trinitrophenyl) Thioether. *J. Am. Chem. Soc.* **2019**, *141* (50), 19911–19916.
- (31) Politzer, P.; Murray, J. S. Impact Sensitivity and Crystal Lattice Compressibility/Free Space. *J. Mol. Model.* **2014**, *20* (5), 1–8.
- (32) CrystalExplorer (Version 3.1), Wolff S. K.; Grimwood, D. J.; McKinnon, J. J.; Turner, M. J.; Jayatilaka, D. Spackman, M. A. University of Western Australia, **2012**.
- (33) Grubbs, F. E.; Beck, G. Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations. *Technometrics* **1972**, *14* (4), 847.
- (34) Yin, P.; Zhang, Q.; Shreeve, J. M. Dancing with Energetic Nitrogen Atoms: Versatile N-Functionalization Strategies for N-Heterocyclic Frameworks in High Energy Density Materials. *Acc. Chem. Res.* **2015**, *49* (1), 4–16.
- (35) Brinck, T. Green Energetic Materials. *Green Energetic Materials*. In: Hoboken: John Wiley & Sons, Inc., **2014**; 1–290.
- (36) Lesnikovich, A. I.; Levchik, S. V.; Balabanovich, A. I.; Ivashkevich, O. A.; Gaponik, P. N. The Thermal Decomposition of Tetrazoles. *Thermochim. Acta* **1992**, *200* (C), 427–441.
- (37) Ni, L.; Yao, H.; Yao, X.; Jiang, J.; Wang, Z.; Shu, C. M.; Pan, Y. Thermal Hazard Analysis and Initial Decomposition Mechanism of 5-(4-Pyridyl)Tetrazolate-Methylene Tetrazole. *Fuel* **2020**, *269*, 117434.
- (38) Sebris, A.; Turks, M. Recent Investigations and Applications of Azidoazomethine-Tetrazole Tautomeric Equilibrium. *Chem. Heterocycl. Compd.* **2019** (11), 1041–1043.
- (39) Prokudin, V. G.; Poplavsky, V. S.; Ostrovskii, V. A. Mechanism of the Monomolecular Thermal

Decomposition of 1,5- and 2,5-Disubstituted Tetrazoles. *Russ. Chem. Bull.* **1996**, *45* (9), 2094–2100.

- (40) Of the 33 tetrazoles in the crystal structure dataset, 31 possessed reported decomposition temperatures.
- (41) Billes, F.; Endrédi, H.; Keresztury, G. Vibrational Spectroscopy of Triazoles and Tetrazole. *Theochem* **2000**, *530* (1–2), 183–200.
- (42) Brethomé, A. V.; Fletcher, S. P.; Paton, R. S. Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **2019**, *9* (3), 2312–2323.
- (43) Jensen, T. L.; Moxnes, J. F.; Unneberg, E.; Christensen, D. Models for Predicting Impact Sensitivity of Energetic Materials Based on the Trigger Linkage Hypothesis and Arrhenius Kinetics. *J. Mol. Model.* **2020**, *26* (4), 65.
- (44) Shoaf, A. L.; Bayse, C. A. The Effect of Nitro Groups on N<sub>2</sub> Extrusion from Aromatic Azide-Based Energetic Materials. *New J. Chem.* **2019**, *43* (38), 15326–15334.
- (45) Schilling, C. I.; Jung, N.; Biskup, M.; Schepers, U.; Bräse, S. Bioconjugation Viaazide–Staudinger Ligation: An Overview. *Chem. Soc. Rev.* **2011**, *40* (9), 4840–4871.
- (46) Xie, S.; Sundhoro, M.; Houk, K. N.; Yan, M. Electrophilic Azides for Materials Synthesis and Chemical Biology. *Acc. Chem. Res.* **2020**, *53* (4), 937–948.
- (47) Bräse, S.; Gil, C.; Knepper, K.; Zimmermann, V. Organic Azides: An Exploding Diversity of a Unique Class of Compounds. *Angew. Chem. Int. Ed.* **2005**, *44* (33), 5188–5240.
- (48) Döhler, D.; Michael, P.; Binder, W. H. CuAAC-Based Click Chemistry in Self-Healing Polymers. *Acc. Chem. Res.* **2017**, *50* (10), 2610–2620.
- (49) Schreck, J. O. Nonlinear Hammett Relationships. *J. Chem. Educ.* **1971**, *48* (2), 103–107.
- (50) Ghosh, A. K.; Sarkar, A.; Brindisi, M. The Curtius Rearrangement: Mechanistic Insight and Recent Applications in Natural Product Syntheses. *Org. Biomol. Chem.* **2018**, *16* (12), 2006–2027.
- (51) For details of data selection and the modeling process, a tutorial of the iterative approach to the modeling of explosive properties, and a detailed discussion of model outliers, see the SI.