NORTHWESTERN UNIVERSITY

Computational Modeling of Metal–Organic Frameworks for the Catalytic Hydrolysis of Nerve Agents and Their Simulants

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIRMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemical and Biological Engineering

By

Matthew L. Mendonca

EVANSTON, ILLINOIS

June 2020

Abstract

Computational Modeling of Metal–Organic Frameworks for the Catalytic Hydrolysis of Nerve Agents and Their Simulants

Matthew L. Mendonca

The effective capture and detoxification of chemical warfare agents (CWAs) is a pressing need in the modern world. Materials are needed for both the destruction of weapon stockpiles and personal protection via fabric coatings or respirators. Attractive candidates for these applications include metal–organic frameworks (MOFs) – highly crystalline materials composed of metal nodes connected by organic linkers – due to their high porosity, large surface area, high concentration of active sites, and chemical functionality that can be tailored towards specific target molecules. Previous experiments, performed in buffered solution, have shown that Zr(IV)-MOFs can catalytically degrade organophosphate-based nerve agents into nontoxic products within minutes via hydrolysis of the phosphate ester bond. This dissertation uses a molecular modeling approach to study the detailed reaction mechanisms and binding interactions involved in MOF-catalyzed nerve agent hydrolysis to help elucidate experimental observations and screen for promising candidate materials with potentially better performance for CWA detoxification.

By performing density functional theory (DFT) calculations, we explore the effects of temperature-induced node dehydration and distortion as well as varying node topologies, connectivities, and metal identities on the catalytic activity of M(IV)-MOFs for solution-phase organophosphate hydrolysis. To address the recent experimental observation of product inhibition in gas-phase nerve agent hydrolysis by Zr-MOFs, we examine the promising alternative of depositing single-atom transition-metal catalysts on MOF nodes to facilitate catalytic turnover.

Additionally, we perform a DFT screening to identify highly predictive nontoxic simulant molecules as candidates for safer and more accurate experimental studies of nerve agent hydrolysis. Throughout the dissertation, we also derive quantitative structure-activity relationship models and perform statistical analyses to determine the most important features for describing the hydrolysis barriers and binding energetics involved in organophosphate hydrolysis reactions. Broadly, the body of work described in this dissertation establishes design principles that can be used to guide future experimental testing for the optimization of MOF catalysts for nerve agent hydrolysis.

Acknowledgements

I would like to thank my advisor, Professor Randall Q. Snurr, for his mentorship, encouragement, and support throughout my PhD. The research in the Snurr group was what convinced me to come to Northwestern, and I am so grateful that I was able to work on such an interesting topic for the past few years with one of the world leaders in molecular modeling research. To the rest of my committee, Professors Linda J. Broadbelt, Omar K. Farha, and Justin M. Notestein, thank you for your time and valuable suggestions and for providing challenging and insightful discussions during my PhD milestones.

I am also grateful to all previous and current members of the Snurr group for their help. The presentations at group meetings and the discussions in the office have helped me become a better researcher and communicator. I would especially like to thank Dr. Haoyuan Chen for being an excellent collaborator and source of knowledge in our subgroup. I am very thankful to have had the opportunity to learn from Professors Omar K. Farha and Joseph T. Hupp; without their experimental expertise and their pioneering of the field of MOFs for CWA detoxification, my research would (literally) not have been possible. I would also like to thank Prof. Peilin Liao, Debmalya Ray, Prof. Christopher J. Cramer, Dr. Rungmai Limvorapitux, and Prof. SonBinh T. Nguyen for their collaboration and input on various manuscripts.

I would like to thank movies/TV shows/board games for keeping me sane and my Chicago Fire Rec Soccer team, Give ‘Em El, for keeping me healthy. I would especially like to thank my friends and family for all their support. Most importantly, I would not be here without my mother (Thelma Mendonca), my father (Daniel Mendonca), and my brother (Michael Mendonca); thank you for always believing in me. To Jessica (Jörp) Yu, this dissertation is dedicated to you.

Table of Contents

List of Figures

List of Tables

List of Schemes

Chapter 1: Introduction

1.1 Methods for Detoxification of Chemical Warfare Agents

Chemical warfare agents (CWAs) were first widely used over 100 years ago during World War I, yet these dangerous toxic compounds continue to be employed in today's world. Despite major international efforts to prohibit their development and use,[3] various CWAs have been used in recent years during the Syrian Civil War,[4] by the Islamic State terrorist group,[5] in an assassination in Malaysia,[6] and in a poisoning in the United Kingdom.[7] There are multiple different forms of CWAs, which vary in their degree of toxicity as well as their mode of action on the human body. The two main categories of CWAs are vesicant agents (e.g., sulfur mustard) and nerve agents. This dissertation will focus on the latter. Nerve agents can be divided into three classes: G-series agents, including sarin (GB), soman (GD), tabun (GA), and cyclosarin (GF);[8] V-series agents, including VX, VR, VE, VG, VS, and VM;[9] and Novichok agents, such as A-230, A-232, and A-234.[10] Organophosphate-based nerve agents function by inhibiting the enzyme acetylcholinesterase, which is responsible for breaking down the neurotransmitter acetylcholine. If acetylcholine is not broken down, it builds up in the synaptic cleft between nerve and muscle cells, causing continuous stimulation of muscles and glands, which eventually leads to asphyxiation, paralysis, and possibly death.[11] For context on the general toxicity of nerve agents, we note that exposure to GD in doses as small as 50 $\mu g/m^3$ can be instantly fatal.[12]

Since these highly toxic chemicals remain a serious global threat, materials are needed for both the destruction of weapon stockpiles and personal protection via fabric coatings or respirators to ensure the safety of military specialists and untrained citizens alike. Research into methods for nerve agent decomposition has been ongoing since the discovery of the first G-series agents in

1936. For example, incineration has commonly been used as a method for the bulk destruction of stockpiles, but this process has significant drawbacks because it can result in the formation of toxic gaseous byproducts.[13] Unfortunately, the time frame to apply effective treatment after exposure to nerve agents can be as short as minutes.[14] Thus, the most important measures to protect against airborne agents are through capture and degradation to nontoxic products before the chemicals reach their biological targets. Conventional solid adsorbents such as activated carbon are effective for adsorption, owing to their microporosity and high surface areas, but are inefficient at deactivation, which can lead to secondary emission once the materials become saturated and poses risks upon disposal.[15] For these reasons, reactive removal is preferred over physical adsorption.

A more promising method for nerve agent detoxification involves hydrolysis, with nucleophilic water or hydroxide substituting at the phosphorus atom of the agent, resulting in elimination of the toxic leaving group. Some metal oxides such as CaO,[16] MgO,[17] and Al_2O_3[18] were found to be reactive toward nerve agents, but they show poor stability in water and air and suffer from product inhibition, which reduces their activity over time. Nanomaterials based on TiO_2[19] and amorphous $Zr(OH)_4$[20] have also been found to be effective for hydrolysis due to their high surface areas and surface hydroxyl groups, but these materials are often tested stoichiometrically and may not be effective as catalysts.[21,22] Further, metal oxides offer a rather limited capacity for chemical functionalization and tunability, thus reducing the potential to improve their reactivities through rational design.

Most metal oxide-based systems for nerve agent hydrolysis are modeled after biological enzymes such as phosphotriesterase (PTE), which are capable of catalytic organophosphate detoxification.[23] The active center of PTE includes dimeric Zn(II) ions, in the form Zn–OH–Zn, which act as Lewis acids. One mode of their catalytic activity involves coordination of the agent

phosphoryl oxygen to a metal atom of the active site, rendering the agent phosphorus atom more vulnerable to nucleophilic attack, thereby enhancing hydrolysis rates.[24,25] Histidine residues surrounding the enzyme active site may also accelerate catalysis, in a synergistic fashion, through favorable hydrogen-bonding interactions. Unfortunately, poor stability outside of buffered media and deactivation after long-term storage limit the range of conditions in which enzymes such as PTE may find practical use.[26]

1.2 Metal–Organic Frameworks

In the last two decades, research interest in metal–organic frameworks (MOFs) has grown dramatically, driven largely by their potential applications in gas storage,[27,28] separations,[29] chemical sensing,[30–32] drug delivery,[33] and catalysis.[34] MOFs are highly crystalline porous materials composed of inorganic metal (or metal oxide) nodes connected by organic linkers. The organic linkers can be modified with functional groups to tune the properties of the MOF, for example, by introducing regions of strong electrostatic charge with electron-rich or poor groups. Some MOFs also have open metal sites, which can serve as catalytic or chemisorption sites, and these metals can be chosen judiciously for desired applications. Therefore, there is vast potential to design MOFs for specific purposes.

There are virtually endless arrangements of nodes, linkers, and functional groups that could be combined in different topologies to create a nearly infinite number of unique MOFs. Several thousand different MOFs[35,36] have already been synthesized and characterized due to the efforts of numerous research groups throughout the last two decades,[37,38] and yet this figure represents only a tiny fraction of the possible structures that remain to be discovered. Because of this, MOFs have become a ripe field for computationally-aided materials design and discovery,[39,40] where computational tools have great potential to accelate scientific advancement in the MOF field.

1.3 MOFs as Catalysts for Nerve Agent Hydrolysis

Among their many applications, MOFs are a promising class of catalysts for CWA degradation due to their high porosity, variable pore size, large surface area, high concentration of active sites, and chemical functionality that can be tailored to adsorb specific target molecules. Recent research has shown that MOFs can adsorb and catalytically degrade nerve agents into nontoxic products within a matter of minutes via hydrolysis of the phosphate ester bond.

Initial studies were directed at adsorption, where $[Zn_2Ca(BTC)_2(H_2O)_2](DMF)_2$ was the first MOF proven to be capable of capturing a nerve agent simulant, methylphosphonic acid (MPA).[41] Shortly after, a MOF-5 analogue was shown to adsorb diisopropyl fluorophosphate (DIFP), a surrogate for sarin.[42] Several MOFs containing open Lewis-acidic metal sites such as Al^{3+}, Cu^{2+}, and Cr^{3+} were then shown to be effective for hydrolysis of nerve agent simulants,[43–45] though they suffered from either poor stability or relatively low catalytic activity.

Currently, Zr(IV)-based MOFs are the most investigated class of MOFs for nerve agent hydrolysis,[46] in part due to their high chemical stability (in pH 1–12) and thermal stability (up to 500 ºC) afforded by their exceptionally strong Zr(IV)–O (node–linker) bonds.[47] The impressive catalytic ability of Zr-MOFs can be attributed to the periodic distribution of strongly Lewis-acidic Zr(IV) metal centers, giving rise to a large number of accessible Zr–OH–Zr active sites reminiscent of the dinuclear Zn-based active sites of the PTE enzyme. Importantly, the connectivity of the Zr_6O_8-cluster nodes (hereafter denoted as Zr_6 nodes) can be systematically tuned by using different organic linkers, thereby altering the number of well-defined[48] and quantifiable[49] Zr–OH_2, Zr–OH, and bridging hydroxide groups. Multiple studies have shown that lower Zr_6 node connectivity, which in turn yields larger numbers of potential binding and catalytic active sites, is directly correlated with accelerated hydrolysis rates.[22] Decreasing the node connectivity also yields larger

pore sizes within the framework, which makes it easier for agents and their hydrolyzed products to diffuse to and from active Zr sites.

According to Kirlikovali et al.,[46] there are three essential requirements to achieve sufficiently fast catalytic hydrolysis of organophosphate molecules, which are reproduced here due to their overall importance for this dissertation:

(1) Water, from either an aqueous solution (for solution-phase reactions) or a humid atmosphere (for gas-phase reactions)

(2) A Lewis-acidic site for activating the agent's phosphorus center

(3) A base, from either basic solutions or solid-phase bases, to ensure that a high concentration of hydroxyl groups are present for nucleophilic attack, for the displacement of hydrolysis products from the active site to yield catalytic turnover, and for neutralizing acidic byproducts (e.g., HF generated from sarin hydrolysis)

Fortunately, due to their robust nature and inherent chemical tunability, Zr-MOFs can be designed to meet all three catalytic requirements. The first experiments reporting Zr-MOFs to be capable of efficient hydrolysis of nerve agents and their simulants were performed in basic solutions, using a buffering agent such as *N*-ethylmorpholine to maintain pH values at ~8.5–10. Using these buffered solutions, various as-synthesized and functionalized Zr-MOFs yielded degradation half-lives on the order of minutes.[1,50–54] While promising for the catalytic destruction of nerve agent stockpiles, this detoxification method is ultimately not feasible for application in gas masks and protective fabrics.[55] Nonetheless, the use of basic solutions is still a valuable method for testing potential catalysts, as the quantification of gas-phase decontamination kinetics under relevant battlefield conditions is a rather difficult experimental challenge.[56]

To address this issue, recent studies have begun to investigate gas-phase organophosphate detoxification in various Zr-MOFs. In contrast to previous reports of catalytic turnover in buffered solution, initial experiments and computational studies showed that exposure to vapor-phase organophosphonates leads to strongly bound hydrolysis products on the Zr_6 nodes, albeit under ultrahigh-vacuum conditions, which may inhibit further reactions (i.e., product inhibition).[57–59] However, subsequent gas-phase experiments performed under varying relative humidity conditions showed that large amounts of moisture present in Zr-MOFs can result in a moderate to significant enhancement of nerve agent hydrolysis rates.[56] Further improvements could also come from base heterogenization, as was recently proved using Zr-MOF/polymer/fiber composite materials that showed similar catalytic activity under ambient humidity conditions compared to MOF powders in aqueous alkaline solution.[55,60] Going forward, an exhaustive investigation into the solid-state catalytic activity of MOFs toward nerve agents in the gas phase, under varying humidity levels[56,61] and in the presence of atmospheric contaminants, is warranted before MOFs may be applied in personal protective equipment such as gas-mask filters[62] and clothing.[63–65]

A more comprehensive overview of MOFs for detoxification of CWAs, in addition to toxic industrial chemicals, can be found in our 2017 review article[66] and other recent reviews.[46,67] References to other background literature, including both experimental and computational studies, on MOF-based nerve agent hydrolysis are provided in the Introduction sections of Chapters 3–5.

1.4 Density Functional Theory Calculations

Computational tools such as electronic structure methods are essential complements to experimental studies for characterizing the properties and reactivity of catalyst materials, especially MOFs.[68,69] The majority of the work in this dissertation involves the use of density functional theory (DFT), which is briefly summarized below. In this section, we only focus on the

basic aspects of performing DFT calculations that are necessary to understand the results presented in Chapters 2–5. A more extensive theoretical review of the assumptions/approximations made in quantum chemical methods (e.g., Born–Oppenheimer approximation) in general can be found in ref [70] and, for quantum chemical characterization of MOFs in particular, ref [68].

The exact electronic structure of any molecular system can be determined by solving the Schrödinger equation, which describes the system's quantum mechanical wave function. Unfortunately, solving the equation exactly for systems larger than a hydrogen atom is computationally impossible. Broadly speaking, DFT can be used to find an approximate solution to the Schrödinger equation. The theorems that give rise to DFT – devised by Hohenberg and Kohn – state that all of the information about a quantum system (e.g., the system's energy) can be derived from its electron density, $\rho(\boldsymbol{r})$, which is simply a function of the three-dimensional coordinates of a set of atoms. In the Kohn-Sham formulation of DFT,[71] the energy functional can be expressed as:

$$E[\rho(\boldsymbol{r})] = T_{fs}[\rho(\boldsymbol{r})] + V_{ne}[\rho(\boldsymbol{r})] + V_{ee}[\rho(\boldsymbol{r})] + E_{xc}[\rho(\boldsymbol{r})] \quad (1.1)$$

where T_{fs} is the kinetic energy of a fictitious system of non-interacting electrons, V_{ne} is the potential energy from nuclear-electron Coulombic attraction, V_{ee} corresponds to the Coulombic interaction of the electron density with itself, and E_{xc} represents the exchange-correlation energy of a real interacting system of electrons. Contained within E_{xc} are terms correcting for the classical electron-electron repulsion energy and the kinetic energy due to electron interactions in a real system. Unfortunately, E_{xc} cannot be determined exactly, and functional approximations to this energy are required, referred to as density functionals. Aside from choosing a density functional, the next most important decision for performing DFT calculations is choice of basis set, referring to the set of non-orthogonal single-particle functions that are used to construct molecular orbitals.[70]

Ultimately, the choice of a given functional and basis set is highly dependent on desired accuracy, the system of interest, and computational time restrictions.

B3LYP[72,73] is a widely used density functional due to its fairly accurate treatment of main-group thermochemistry, but it is known to underestimate reaction barrier heights.[74] The Pople basis sets (e.g., 6-311++G**) are commonly used for main-group elements in conjunction with B3LYP. For exploring reactions on MOF nodes, the M06-L functional[75] is commonly used because it is one of the most accurate functionals for transition metals and it is a local functional, which is more computationally affordable for large systems. For heavier elements such as transition metals that contain large numbers of electrons, and thus large numbers of basis functions, it is more computationally efficient to use effective core potentials (ECPs) to treat the electrons in core orbitals, which are only weakly affected by chemical bonding. Basis sets used with ECPs are typically of split valence (SV) or triple zeta valence (TZV) quality, referring to the number of basis functions used to represent valence orbitals, and are often supplemented by polarization functions (e.g., def2-SVP or def2-TZVPP).[76] Popular pseudopotentials include the Los Alamos National Laboratory (LANL2DZ)[77,78] and the Stuttgart-Dresden (SDD)[79,80] ECPs.

Although the large unit cells of MOFs make the use of reactive studies on the periodic scale difficult, valuable information can be obtained from calculations involving smaller clusters of the MOF nodes and linkers. When performing DFT calculations to model reactions on MOF node clusters, it is necessary to "cut" the node and surrounding linkers from its periodic structure found in a crystallographic information file (CIF), using capping groups such as formate or benzoate to represent the linkers and to maintain charge neutrality. By default, geometry optimizations (described below) are typically performed in the gas phase. However, to compare to experiments done in solution, solvation effects can be simulated implicitly, using models such as the polarizable

continuum model (PCM)[81] and SMD continuum solvation model,[82] or by including explicit solvent molecules in the model system.[83] In relatively large systems such as MOF-organophosphate complexes, it is important to accurately account for weak dispersion forces arising from medium-range non-covalent interactions.[84] One of the most widely used methods is the DFT-D3 dispersion correction,[85] commonly paired with the Becke-Johnson (BJ) damping function.[86,87]

Once a functional and basis set are chosen, in addition to other input parameters, one needs to actually perform the DFT calculations. Throughout this dissertation, the phrase "performing DFT calculations" generally refers to the process of computing the potential energy surface (PES) of a molecular system at various points along a reaction energy profile (or reaction coordinate diagram) connecting reactants to products, which is used to determine quantities such as activation barriers for a chosen reaction mechanism. For any given system, the geometry of its set of atoms can be represented by a vector ($\boldsymbol{r}$) of the atoms' coordinates. One can then introduce the concept of the system's electronic energy, $E(\boldsymbol{r})$, as a function of these atomic positions. Given these definitions, the first step in "performing DFT calculations" is a geometry optimization, using a chosen optimization algorithm (e.g., Berny algorithm[88]), where the goal is to find the value of $\boldsymbol{r}$ for which $E(\boldsymbol{r})$ is at a local minimum. Using an initial guess of the correct geometry, an iterative optimization procedure is followed to minimize the energy of the chosen system by adjusting the geometry until an optimal spatial arrangement of atoms is found such that the net force, $\partial E/\partial \boldsymbol{r}$, on each atom is effectively zero and the PES is at a stationary point. If the Hessian matrix (the second derivative matrix of the system) describing the curvature of the PES at $\boldsymbol{r}$ has all positive eigenvalues (i.e., minimum on the PES), then an intermediate along the reaction energy profile has been found. If the Hessian matrix contains exactly one negative eigenvalue (i.e., 1st order saddle point on the PES), then a transition state (TS) along the reaction energy profile has been located.

To verify the natures of all stationary points (both intermediates and TSs), the next step is the calculation of analytic vibrational frequencies, where minimized intermediate structures are characterized by zero imaginary frequencies and TS structures display exactly one imaginary frequency corresponding to the reaction path of interest. These frequencies can also be used to compute molecular partition functions, typically at standard conditions of 298.15 K and 1 atm, using the conventional particle in a box, rigid rotor, and quantum mechanical harmonic oscillator approximations.[70] From these partition functions, one can derive thermochemical properties such as enthalpies and Gibbs free energies.

1.5 Importance of Modeling for CWA Detoxification Research

Due to the danger involved in working with highly toxic CWAs, experiments are typically done using simulant molecules, which are safer to handle than actual CWAs but have similar chemical behavior and structure. The use of simulants greatly simplifies the execution of experiments pertaining to degradation of CWAs, and there are several simulants that give results similar to their respective CWAs. However, with any simulant there is a necessary tradeoff between accurately mimicking agent reactivity and being sufficiently nontoxic for researchers to study them in an academic setting. By its nature, computational research does not suffer from the same safety issues as experimental work. Therefore, computational scientists can simulate reactions of CWAs in MOFs using the real agents as well as their simulants. To explain how modeling can be used to optimize MOF catalysts for neutralizing CWAs, we present a case study of the first mechanistic analysis of organophosphate hydrolysis.

As a complement to their experimental work, Mondloch et al. performed DFT calculations to analyze the hydrolytic degradation of the simulant methyl-paraoxon (DMNP) and the nerve agents VX and GD on the Zr-MOF NU-1000, using a cluster model of the node and an implicit

solvation model.[1] The most favorable binding configuration for DMNP on the NU-1000 node (–26 kJ/mol, relative to free energy of separated reactants) involves stabilization due to hydrogen bonding with the hydroxide and water groups bound to the node, in addition to weak π–π stacking interactions between the benzene ring of the $TBAPy^{4-}$ linkers and the DMNP phenyl ring. They also considered the case of DMNP interacting directly with the Lewis-acidic Zr(IV) catalytic site by removing a terminal node-ligated water molecule. The dissociation of a coordinated water was predicted to be the rate-determining step, and the dominant interaction is the electrostatic attraction between the Zr atom and the P=O group of DMNP, where the bound configuration is 22 kJ/mol uphill in free energy. This provided insight into the experimental observation that dehydrating NU-1000 prior to reaction (by heating the MOF to remove terminal water ligands) reduces the half-life for DMNP hydrolysis to 1.5 minutes, compared to 15 minutes for the fully hydrated MOF. Nevertheless, the reaction is driven forward by the product stability, which is considerably downhill in free energy (–48 kJ/mol). These calculations also indicated that the mechanism on Zr_6 nodes is similar to those calculated for the PTE-catalyzed reaction of DMNP.[89] The authors then used the same DFT approach to investigate the binding and hydrolysis of VX and GD on the NU-1000 node, revealing qualitatively similar mechanisms as for DMNP hydrolysis. However, there were no hydrogen-bonding interactions present for the VX and GD analogues with the linkers; the calculations instead showed that attractive van der Waals interactions were most prevalent. Additionally, the replacement of a node-ligated water molecule had a much lower reaction barrier compared to DMNP (5.4 kJ/mol and 48.2 kJ/mol, respectively). The hydrolysis was again driven forward by the thermodynamically stable products for the VX and GD analogues (–123 kJ/mol and –83 kJ/mol, respectively), showing selective hydrolysis of the P–S bond in VX and the P–F bond in GD.

These thermodynamic results prompted experimental testing of the NU-1000-catalyzed hydrolysis of GD, which showed a reaction half-life of 3 min when run in an *N*-ethylmorpholine buffered solution and 36 min under 50% relative humidity. Also inspired by these DFT calculations, a later experimental study on a different Zr-MOF, UiO-67-$(NMe)_2$, showed selective hydrolysis of VX by cleavage of the P–S bond with a half-life of 1.8 minutes.[21] This is an important observation since cleavage of the P–O bond gives rise to EA-2192 (S-2-(diisopropylamino)ethyl O-hydrogen methylphosphonothioate), which is another highly toxic organophosphate molecule.[90] This case study serves as a model example for the power of combined theoretical and experimental studies for understanding reaction mechanisms and analyzing which degradation products will be obtained.

Computational modeling of nerve agent hydrolysis in MOFs is a relatively new research area compared to modeling of other MOF applications such as gas storage. However, the limited computational work that has been done since 2015 – in close conjunction with experiment – has shown promising results. As this field continues to develop, computational modeling can be used to guide collaborating experimental groups by screening candidate materials with potentially better performance for the degradation of nerve agents. If the modeling techniques applied are sufficiently accurate, this will save time and minimize the number of experiments that must be performed with these dangerous chemicals. Further, using techniques such as DFT calculations to explore reaction mechanisms in detail, valuable insights can be gained into the molecular-level interactions of the agents with the MOF surfaces that cannot be determined through experiments alone, which gives researchers a better understanding of the fundamental properties that affect catalyst performance. This is a good example of a research problem where computational science

has the potential to make significant contributions by probing systems that are difficult and dangerous to study experimentally.

1.6 Outline of Dissertation

This dissertation is a compilation of three peer-reviewed publications and one manuscript in preparation, organized as follows:

- Chapter 2: **M. L. Mendonca** and R. Q. Snurr, "Screening for Improved Nerve Agent Simulants and Insights into Organophosphate Hydrolysis Reactions from DFT and QSAR Modeling," *Chemistry – A European Journal*, 2019, *25*, 9217-9229.

- Chapter 3: H. Chen*, P. Liao*, **M. L. Mendonca***, and R. Q. Snurr, "Insights into Catalytic Hydrolysis of Organophosphate Warfare Agents by Metal–Organic Framework NU-1000," *The Journal of Physical Chemistry C*, 2018, *122*, 12362-12368. (* represents equal contribution)

- Chapter 4: **M. L. Mendonca**, D. Ray, C. J. Cramer, and R. Q. Snurr, "Exploring the Effects of Node Topology, Connectivity, and Metal Identity on the Binding of Nerve Agents and Their Hydrolysis Products in Metal–Organic Frameworks," *in preparation*.

- Chapter 5: **M. L. Mendonca** and R. Q. Snurr, "Computational Screening of Metal–Organic Framework-Supported Single-Atom Transition-Metal Catalysts for the Gas-Phase Hydrolysis of Nerve Agents," *ACS Catalysis*, 2020, *10*, 1310-1323.

Chapter 2, which describes a computational screening to identify highly predictive nontoxic simulant molecules as candidates for safer and more accurate experimental studies of nerve agent hydrolysis, does not explicitly concern MOFs. Chapter 3 describes the origins of the catalytic effects of the Zr-MOF NU-1000 for solution-phase organophosphate hydrolysis to provide insights on previous experiments, in addition to exploring the effects of temperature-induced node dehydration and distortion on the catalytic mechanism. Chapter 4 describes the use of DFT to examine the effects of MOF node topology, connectivity, and metal identity on the binding energies of multiple nerve agents and their corresponding hydrolysis products, again in the context of solution-phase reactions. Chapter 5 describes a study on gas-phase sarin hydrolysis

to determine if depositing single-atom catalysts on MOF nodes is a viable strategy to avoid product inhibition and accelerate gas-phase hydrolysis compared to unfunctionalized nodes. Finally, Chapter 6 summarizes the key findings of the dissertation and provides recommendations for future research directions.

Other articles that I contributed to and that were published over the course of my doctoral research are listed below. Several paragraphs from the review article were included in various sections of this introductory chapter. The other article, which was the result of a collaboration with an experimental group, is not directly related to nerve agent detoxification and is thus omitted from this dissertation.

- N. S. Bobbitt, **M. L. Mendonca**, A. J. Howarth, T. Islamoglu, J. T. Hupp, O. K. Farha, and R. Q. Snurr, "Metal–Organic Frameworks for the Removal of Toxic Industrial Chemicals and Chemical Warfare Agents," *Chemical Society Reviews*, 2017, *46*, 3357-3385.

- R. Limvorapitux, H. Chen, **M. L. Mendonca**, M. Liu, R. Q. Snurr, and S. T. Nguyen, "Elucidating the Mechanism of the UiO-66-Catalyzed Sulfide Oxidation: Activity and Selectivity Enhancements through Changes in the Node Coordination Environment and Solvent," *Catalysis Science & Technology*, 2019, *9*, 327-335.

Chapter 2: Screening for Improved Nerve Agent Simulants and Insights into Organophosphate Hydrolysis Reactions from DFT and QSAR Modeling

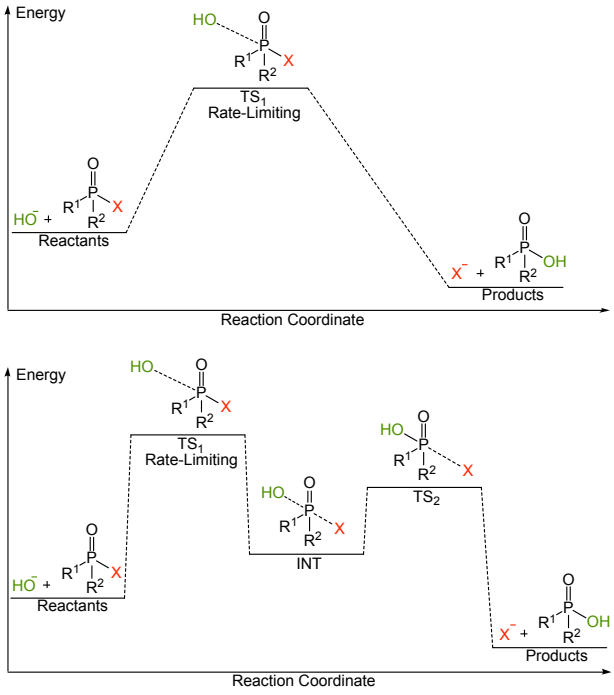
This chapter is a modified version of a published manuscript: Mendonca, M. L.; Snurr, R. Q. *Chem. Eur. J.* **2019**, *25*, 9217–9229.

2.1 Introduction

The first large-scale use of chemical warfare agents (CWAs) occurred during World War I, and dangerous toxic compounds continue to be employed in modern warfare. Despite global efforts to ban the stockpiling and use of CWAs, several countries are believed to still have operational production facilities. Various CWAs have been used as recently as 2018 during the Syrian Civil War,[4] by the terrorist group Islamic State,[5] in an assassination in Malaysia,[6] and in the recent poisoning in the United Kingdom.[7] In particular, organophosphate-based nerve agents are among the deadliest chemicals in the world. These agents function by inhibiting the enzyme acetylcholinesterase, causing accumulation of the neurotransmitter acetylcholine. This triggers continuous stimulation of muscles and glands, which can lead to asphyxiation and death. Nerve agents can be divided into three classes: G-series agents, including sarin (GB), soman (GD), tabun (GA), and cyclosarin (GF);[8] V-series agents, including VX, VR, VE, VG, VS, and VM;[9] and Novichok agents, such as A-230, A-232, and A-234.[10] Research into methods and materials for the capture and decomposition of these lethal agents has been ongoing since their discovery.[91] Detoxification of nerve agents usually involves hydrolysis, with nucleophilic water or hydroxide substituting at the phosphorus atom of the agent, resulting in elimination of the toxic leaving group

(see Scheme 2.1). Due to the growing threat of their use in modern warfare, there has been a large increase in research in the last few years to identify materials capable of destroying CWAs.[66,92,93]

Scheme 2.1. Schematic of Two Possible Organophosphate Hydrolysis Mechanisms, (Top) Concerted and (Bottom) Stepwise



Due to the high toxicity of CWAs and regulation under the Chemical Weapons Convention, simulants are often used in laboratory experiments to mimic the reactivity of nerve agents when screening materials for catalytic detoxification. There exist many simulants that give results similar to those of their respective agents, and their use streamlines studies of CWA degradation. However, there is a necessary tradeoff between accuracy in simulating agent reactivity and safety for academic study.[66] Further, there is little reason to believe that the current simulants used in the literature will have identical degradation behavior to all types of agents. Sometimes, using actual agents can reveal additional mechanisms not observed with simulants.[52] For example, dimethyl methylphosphonate (DMMP) is commonly used as a simulant[58,59] for GB because it exhibits similar adsorption behavior. However, it does not contain the P–F bond that is broken during sarin

hydrolysis, and is therefore not as effective for mimicking reactivity.[62] Choosing the most appropriate simulant is non-trivial because it depends on the physical or chemical application of interest (e.g., degradation, adsorption, or diffusion), but there are still many cases where researchers claim materials are active towards CWAs based on extrapolation of simulant data alone.[94,95] Conversely, the dismissal of catalytic or sorptive materials based on lack of activity towards simulants could result in missed opportunities. Thus, there is a pressing need for the discovery and validation of new simulants, or confirmation of current simulants, that are effective predictors of nerve agent reactivity under remediation conditions. As with the already widely used simulants, these new molecules must be safe enough for researchers to handle them in experiments.

In addition to determining improved simulants, it is interesting to study how molecular features affect organophosphate reaction energetics. Early studies in the nerve agent literature studied correlations between physiological action and chemical structure.[96–98] However, there have been limited reports of correlations between degradation behavior and chemical properties, which have only covered structural effects for a small set of compounds due to experimental limitations.[99–102] Quantitative structure-activity relationship (QSAR) modeling can be used to determine correlations between modifications in geometric and electronic structure and corresponding changes in a response variable, such as partition coefficient or reactivity.[103] QSAR models are applied in a diverse range of fields, from predicting biological toxicity based on structural descriptors,[104] to correlating reaction barriers with molecular orbital parameters.[105] With advances in computational techniques, quantum chemical methods are increasingly being used to calculate descriptors to build the models. Aside from providing valuable insight into the chemistry governing certain reaction classes or binding events, QSAR models can also drastically reduce the expense and time required to discover and test novel candidate molecules. To the best of our

knowledge, there has only been one previous QSAR model on organophosphate hydrolysis using molecular orbital descriptors,[106] but it only studied a small number of compounds and used outdated computational methods. Due to the threat of nerve agent attacks in today's world, along with the significant increase in research aimed at neutralizing these agents in recent years,[107,108] we believe there is a crucial need for a more thorough understanding of the basic chemistry underlying their detoxification mechanisms.

Overall, the main goal of this work is to generate important insights by investigating nerve agent hydrolysis reactions computationally, thereby reducing the number of experiments necessary with these dangerous chemicals. Herein, we perform a density functional theory (DFT) screening of over 100 organophosphate molecules, including hypothetical structures as well as previously synthesized chemicals, to determine improved simulants based on criteria such as low toxicity and similarity to nerve agent hydrolysis behavior. Both the pathway of degradation and the activation energy barrier are investigated as measures for identifying improved non-toxic analogs. We systematically vary the functional groups and leaving groups to observe the effects on overall reaction energetics and mechanism. Our analysis is focused on comparison to the G-series agents, sarin and soman, as they are the most commonly studied and used agents.

We also investigate correlations between molecular descriptors and reaction barriers for the alkaline hydrolysis of these organophosphates, and we use the results to construct QSAR models. Through careful consideration of the resulting models, a better understanding of the most important descriptors involved in organophosphate hydrolysis is formed. The generated models are then subjected to a thorough statistical analysis and validation procedure, in order to ensure accurate predictive capability. Through the assessment of applicability domain, we further show

that the QSAR models trained on G-series agents and simulants can reliably predict energetics for other organophosphate classes as well, including VX.

Caution: *Some of the organophosphate molecules mentioned herein are considered extremely dangerous and should only be handled in a lab with proper facilities and certified personnel.*

2.2 Computational Details

All electronic structure calculations were performed using the Gaussian 09 package (revision D.01).[109] Geometry optimizations were performed for all species using DFT with the B3LYP functional[72,73] and 6-311++G** basis set. Several other methods were tested on the uncatalyzed hydrolysis reaction of sarin to explore the effect of level of theory on the energy barrier. The DFT-D3 dispersion correction[85] with the Becke-Johnson damping function,[86,87] the M06-2X functional,[110] second-order Møller-Plesset theory (MP2),[111] the Hartree-Fock method (HF), and the modified Complete Basis Set method (CBS-QB3)[112,113] were tested and compared to available experimental data (see Table A.1 in Appendix A). In order to compare to experiments done in solution, the polarizable continuum model (PCM)[81,114] of water was used to model solvation effects implicitly. Harmonic vibrational analyses were performed, at the same level of theory, to verify the nature of all species and to calculate the thermochemical properties at standard conditions (298.15 K, 1 atm).

We used the EPA Toxicity Estimation Software Tool (T.E.S.T. v 4.2)[115] to compile experimental and predicted toxicity data for both nerve agents and all 117 simulants. T.E.S.T. allows users to estimate the toxicity of compounds using various QSAR methods such as hierarchical,[116] FDA,[117] and nearest neighbor[118] methods that predict values based on 45 molecular structure descriptors. We chose to use the consensus method, which takes an average of the

predicted toxicities from the above QSAR methods, because it was shown to achieve the best prediction results during external validation.[115]

Molecular orbital parameters, used as descriptors for building QSAR models, were calculated with the natural bond orbital (NBO) method[119] in Gaussian 09. All calculations required for QSAR model development and statistical testing were performed in MATLAB using built-in functions and a custom script to execute the leave-one-out cross-validation procedure.

2.3 Results and Discussion

2.3.1 DFT Screening for Improved Nerve Agent Simulants

Using DFT, we calculated the uncatalyzed alkaline hydrolysis mechanism for 117 potential organophosphate simulant molecules, as well as the nerve agents GB and GD. Functional groups and leaving groups were systematically varied to observe the effects on overall reaction energetics. For all molecules with a chiral phosphorus center, the S_P enantiomer was modeled for consistency (note that this isomer is generally more toxic than R_P),[120] although this convention should not affect the reaction energetics.

All molecules were observed to undergo S_N2 nucleophilic substitution pathways, with HO^- acting as the nucleophile.[121–123] The hydrolysis of nerve agents is known to proceed through a stepwise mechanism, which generally consists of two transition states connected by a pentacoordinated trigonal bipyramidal (TBP) intermediate.[123] Organophosphate compounds may also hydrolyze through a single-step concerted mechanism in which the transition state mostly involves nucleophilic attack but not the departure of the leaving group.[124,125] Scheme 2.1 depicts the differences between the stepwise and concerted mechanisms. In TBP phosphorus molecules, the most electronegative group is more stable in the axial position, where elimination is easiest.[126] Thus, the most electronegative group in each molecule (e.g., –F in sarin) was assumed to be the

preferred leaving group, and the nucleophilic hydroxide was positioned to attack each molecule directly opposite this leaving group.

Table 2.1. Effect of Conjugate Acid pK_a of Leaving Groups on Organophosphate Hydrolysis Mechanism

Leaving Group	Conjugate Acid pK_a	Mechanism
Cl^-	–8[a]	Concerted
F^-	3.17[a]	Stepwise
2,4-$(NO_2)_2$-$C_6H_3O^-$	4.07[b]	Concerted
2,4,6-Cl_3-$C_6H_2O^-$	5.99[c]	Concerted
2,4,5-Cl_3-$C_6H_2O^-$	6.72[c]	Concerted
4-NO_2-$C_6H_4O^-$	7.15[b]	Concerted
2-NO_2-$C_6H_4O^-$	7.23[b]	Concerted
3-CH_3-4-NO_2-$C_6H_3O^-$	7.33[d]	Concerted
2,4-Cl_2-$C_6H_3O^-$	7.85[c]	Concerted
4-CN-$C_6H_4O^-$	7.97[b]	Concerted
3-NO_2-$C_6H_4O^-$	8.36[b]	Concerted
3-Cl-$C_6H_4O^-$	8.56[b]	Concerted
4-Cl-$C_6H_4O^-$	9.41[b]	Concerted
4-SCH_3-$C_6H_4O^-$	9.53[b]	Stepwise
3-$N(CH_3)_2$-$C_6H_4O^-$	9.78[e]	Stepwise
$C_6H_5O^-$	9.99[b]	Concerted/Stepwise*
4-OCH_3-$C_6H_4O^-$	10.21[b]	Stepwise
C_6H_5-CH_2O^-	15.4[f]	Stepwise
CH_3O^-	15.54[f]	Stepwise
$CH_3CH_2O^-$	16[f]	Stepwise
$CH_3CH_2CH_2O^-$	16.1[f]	Stepwise
$(CH_3)_2CHO^-$	16.5[a]	Stepwise
$(CH_3)_3CCH_2O^-$	16.5[g]	Stepwise
$(CH_3CH_2)_2CHO^-$	18.2[h]	Stepwise

Experimental or predicted pK_a values of the conjugate acids of leaving groups taken from (a) ref [127], (b) ref [128], (c) ref [129], (d) ref [130], (e) ref [131], (f) ref [132], (g) estimated value for primary alcohols, and (h) ref [133]. *Simulants with the $C_6H_5O^-$ leaving group followed stepwise mechanisms when $R^1 = R^2 = –OCH_3$ or $R^1 = R^2 = –OCH_2CH_3$ and concerted when $R^1 = –OC_6H_5$ and $R^2 = –CH_3$.

We observed a relationship between the conjugate acid pK_a of the leaving groups and the reaction pathway that the molecules followed, in agreement with previous studies.[124,125] Simulants with a $(R)O^-$ leaving group where the conjugate acid pK_a is below ≈ 9.5–10 (“good” leaving groups) were found to hydrolyze in a concerted mechanism, whereas “poor” leaving groups reacted through a stepwise mechanism (Table 2.1). Both nerve agents and all simulants with a F^-

leaving group followed stepwise mechanisms, while all Cl^- leaving groups corresponded to concerted pathways. Aside from the conjugate acid pK_a of the leaving groups, the size of the leaving group may also play a role in the preferred decomposition mechanism due to varying steric interactions, although this effect should be less significant for uncatalyzed reactions.

For accurate comparisons to be made when performing experiments, an ideal simulant should have a similar hydrolysis rate as its corresponding nerve agent, while also being less toxic. To compile experimental and predicted toxicity data for both nerve agents and all 117 simulants, we used the T.E.S.T. software.[115] We chose oral rat LD_{50} $-\log_{10}(\text{mol}\bullet\text{kg}^{-1})$ as the toxicity endpoint,[134] which represents the mass of the compound per rat body weight that causes death in 50% of rats after oral ingestion, because it was the endpoint for which the most experimental data was available.

For those molecules that exhibit a stepwise mechanism, we found that the first transition state (TS_1) was rate-limiting.[135] For molecules that exhibit a concerted mechanism, there is only one transition state (also referred to here as TS_1). Thus, for all 119 molecules studied, the free energy barrier for TS_1 (ΔG_{TS1}) was chosen as a basis for comparing simulant to nerve agent behavior. Here, ΔG_{TS1} is defined as the difference in free energy between TS_1 and the separated reactants. For every stepwise mechanism observed, the second transition state (TS_2) – involving the departure of the leaving group – had a negligibly small energetic barrier, as its structure is essentially identical to the connecting TBP intermediate, and thus TS_2 does not significantly affect the reaction kinetics. In Table A.2, we show the hydrolysis free energy barriers for several organophosphate molecules obtained using our DFT procedure and compare them to available experimental values to put our screening results into perspective. Although DFT consistently under-predicts the experimental barriers, the ordering of the molecules is similar, which makes

ΔG_{TS1} a reliable comparison metric. The TS_1 free energy barriers from our DFT calculations are 59.74 kJ•mol^{-1} and 56.16 kJ•mol^{-1} for GD and GB, respectively.

Table A.3 contains ΔG_{TS1} and rat oral LD_{50} $-\log_{10}$ values for all organophosphate molecules studied. By plotting the DFT-predicted free energy barriers vs. rat oral LD_{50} values, we observed no correlation ($R^2 = 0.01$, Figure A.1) between organophosphate toxicity and hydrolysis reaction barrier.

We calculated the difference between the TS_1 free energy barrier for individual simulants relative to each nerve agent, $|\Delta\Delta G_{TS1}|$, to identify the best simulants. We chose $|\Delta\Delta G_{TS1}| < 3.25$ kJ•mol^{-1} as a criterion for an improved simulant to account for uncertainty in the DFT calculations,[136] and because this is roughly equal to the difference between the free energy barriers for soman and sarin. All simulants studied are indeed less toxic than the nerve agents, where rat oral LD_{50} $-\log_{10}$ values for GD[137] and GB[138] are 5.66 and 5.41 mol•kg^{-1}, respectively. For context, rat oral LD_{50} $-\log_{10} = 4.88$ mol•kg^{-1} for methyl-paraoxon (DMNP),[134] which is one of the more common simulants used when testing catalysts for detoxification.[1,50,139–141] For precautionary measures, we set the upper bound toxicity criterion at rat oral LD_{50} $-\log_{10} < 2.50$ mol•kg^{-1}. This upper bound on the toxicity endpoint was chosen to err on the side of caution, given any possible inaccuracies in the T.E.S.T. predictions. To our knowledge, DMNP, with its slightly higher toxicity endpoint, has been used safely without incident in many previous studies. One attractive feature of DMNP is its low vapor pressure (3.3×10^{-5} mmHg at 25 ºC),[134] which reduces the inhalation hazard. This is perhaps a more immediate concern for a research scientist, rather than oral exposure, and should also be taken into consideration when choosing simulants. However, we chose the oral toxicity endpoint because it was the indicator for which the most experimental data

was available. Rat oral LD_{50} $-\log_{10} < 2.50$ mol•kg^{-1} was used as a general safety measure, but it is unclear what the actual cutoff should be for an academic lab.

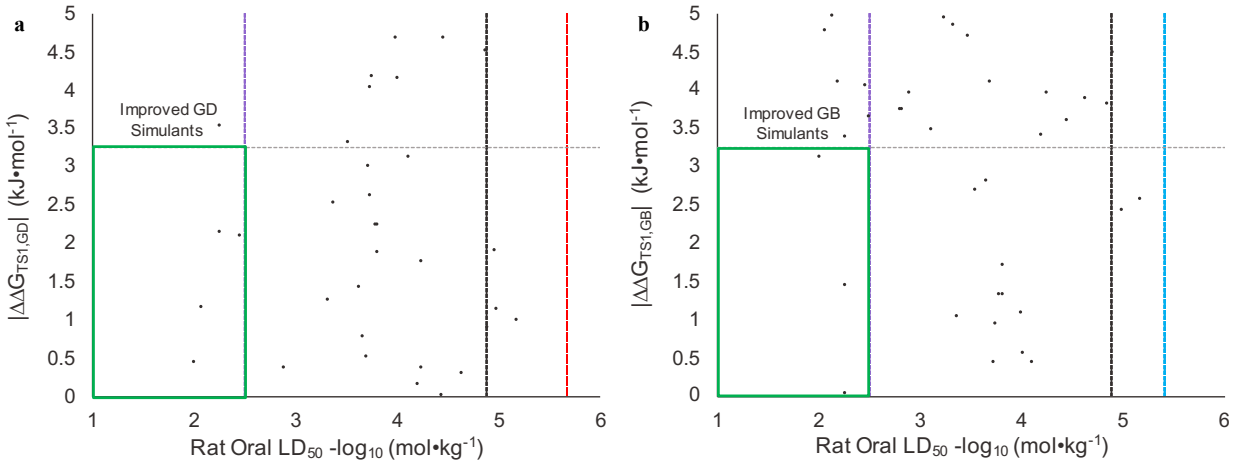
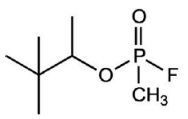
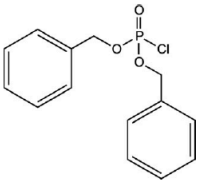
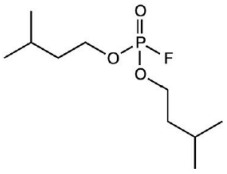
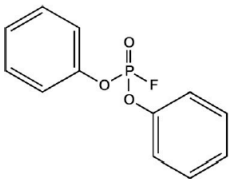
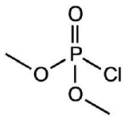
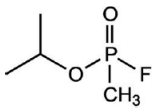
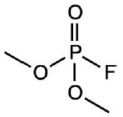
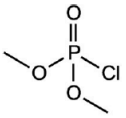
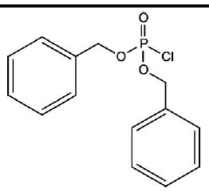


Figure 2.1. Differences in DFT-calculated hydrolysis free energy barriers for the first transition state, $|\Delta\Delta G_{TS1}|$, for simulants relative to (a) soman and (b) sarin vs. their toxicity endpoint. The horizontal grey line indicates $|\Delta\Delta G_{TS1}| <$ 3.25 kJ•mol^{-1} as the energy barrier criterion for an improved simulant. The vertical purple line indicates LD_{50} $-\log_{10}$ < 2.50 mol•kg^{-1} as the toxicity criterion. The black vertical line indicates LD_{50} $-\log_{10} = 4.88$ mol•kg^{-1} (toxicity of DMNP), the red vertical line indicates LD_{50} $-\log_{10} = 5.66$ mol•kg^{-1} (toxicity of soman), and the blue vertical line indicates LD_{50} $-\log_{10} = 5.41$ mol•kg^{-1} (toxicity of sarin).

As shown in Figure 2.1, these criteria suggest 4 improved GD simulants for soman and 3 improved GB simulants for sarin, based on the chosen energy barrier and toxicity bounds. Here, "improved" indicates that the simulant molecule has both $|\Delta\Delta G_{TS1}| < 3.25$ kJ•mol^{-1} and rat oral LD_{50} $-\log_{10} < 2.50$ mol•kg^{-1}. Note that Figure 2.1 only displays data for $|\Delta\Delta G_{TS1}|$ up to 5 kJ•mol^{-1}, for clarity and to highlight the top simulants, whereas all data points are included in Figure A.2. Table 2.2 displays the improved soman and sarin simulants, where each molecule is labeled using its unique chemical identifier that can be used for searching chemical databases. Further, if a particular application requires an identical degradation mechanism to that of nerve agents, then simulants that undergo stepwise hydrolysis should be chosen. When comparing the molecules in Table 2.2, it is clear that the simulants do not necessarily have a similar molecular structure as their corresponding nerve agents. However, it is apparent that molecules with F^- and Cl^- leaving

groups are generally better reactivity simulants compared to the other 22 $(R)O^-$ leaving groups considered. We believe that the primary reason molecules with F^- and Cl^- leaving groups are present among the improved simulants is because hydrolysis of the nerve agents involves breaking a P–F bond, which is better represented by a simulant with a halogen leaving group as opposed to an alkoxy leaving group.

Table 2.2. Improved Soman (Top) and Sarin (Bottom) Simulants, Labeled Using Their Unique Chemical Identifiers

				
Soman (GD) Toxicity[a] = 5.66 mol kg^{-1} ΔG_{TS1}[c] = 59.74 kJ mol^{-1} Mechanism[d]: Stepwise	CAS RN: 538-37-4 Toxicity[b] = 2.01 mol kg^{-1} $\lvert\Delta\Delta G_{TS1,GD}\rvert$[e] = 0.46 kJ mol^{-1} Mechanism[d]: Concerted	PubChem CID: 129642440 Toxicity[b] = 2.07 mol kg^{-1} $\lvert\Delta\Delta G_{TS1,GD}\rvert$[e] = 1.18 kJ mol^{-1} Mechanism[d]: Stepwise	CAS RN: 403-65-6 Toxicity[b] = 2.46 mol kg^{-1} $\lvert\Delta\Delta G_{TS1,GD}\rvert$[e] = 2.11 kJ mol^{-1} Mechanism[d]: Stepwise	CAS RN: 813-77-4 Toxicity[b] = 2.26 mol kg^{-1} $\lvert\Delta\Delta G_{TS1,GD}\rvert$[e] = 2.14 kJ mol^{-1} Mechanism[d]: Concerted
				
Sarin (GB) Toxicity[a] = 5.41 mol kg^{-1} ΔG_{TS1}[c] = 56.16 kJ mol^{-1} Mechanism[d]: Stepwise	CAS RN: 5954-50-7 Toxicity[b] = 2.26 mol kg^{-1} $\lvert\Delta\Delta G_{TS1,GB}\rvert$[e] = 0.05 kJ mol^{-1} Mechanism[d]: Stepwise	CAS RN: 813-77-4 Toxicity[b] = 2.26 mol kg^{-1} $\lvert\Delta\Delta G_{TS1,GB}\rvert$[e] = 1.45 kJ mol^{-1} Mechanism[d]: Concerted	CAS RN: 538-37-4 Toxicity[b] = 2.01 mol kg^{-1} $\lvert\Delta\Delta G_{TS1,GB}\rvert$[e] = 3.13 kJ mol^{-1} Mechanism[d]: Concerted	

(a) Experimental toxicity endpoint, which represents the rat oral LD_{50} $-\log_{10}$ value, obtained using T.E.S.T.[115] (b) Predicted toxicity endpoint values calculated using the consensus QSAR method in the T.E.S.T. software. (c) Free energy barrier for the first transition state in the soman or sarin hydrolysis reaction, calculated with DFT at T = 298.15 K using B3LYP/6-311++G**. (d) Mechanism of hydrolysis, either stepwise or concerted. (e) Difference between the TS_1 free energy barrier for the simulant molecule relative to soman or sarin.

Given the inherent uncertainty in DFT and the predicted toxicities, we hesitate to declare any one simulant as the best, but these 4 GD and 3 GB simulants may be worthwhile to study in experiments, specifically for testing catalyst ability in aqueous nerve agent hydrolysis under alkaline conditions. We note that some of these improved simulants may still be considered toxic by university standards, so we highly recommend seeking approval from institutional safety departments and further study of their toxicity prior to their utilization. Even though there is an

overlap between top soman and sarin simulants, it is important to note that there is no single best simulant for all types of nerve agents, so molecules should only be chosen based on similarity to the CWA of interest.

Interestingly, many commonly used simulants in the literature showed large energy barrier deviations from GD and GB. For example, ethyl-paraoxon and diisopropyl fluorophosphate both had $|\Delta\Delta G_{TS1}| > 10$ kJ•mol^{-1}, dimethyl methylphosphonate (DMMP) had $|\Delta\Delta G_{TS1}| > 20$ kJ•mol^{-1}, and diisopropyl methylphosphonate had $|\Delta\Delta G_{TS1}| > 30$ kJ•mol^{-1}. This should serve as a caution against selecting simulants based on literature precedent alone. A comparison of previous computational studies shows that this discrepancy also translates to reactions on metal–organic frameworks, where the hydrolysis energy barriers for sarin and DMMP are calculated as 55 kJ•mol^{-1} and 84 kJ•mol^{-1}, respectively, on UiO-66.[57,58]

Note that the above analysis was focused entirely on ΔG_{TS1} as an indicator of kinetic similarity, but the relative thermodynamics (ΔG_{rxn}) compared to nerve agents is also relevant.[1,21] If translating these results for experimental catalysis, simulant size would be another factor to consider, as smaller molecules may experience less difficulty accessing catalyst active sites.[52] A bulky simulant such as dibenzyl chlorophosphonate (CAS RN: 538-37-4, Table 2.2) may have significant difficulty diffusing into microporous catalysts, such as the UiO class of MOFs with relatively small pore diameters, and thus the degradation reaction may be limited to the exterior catalyst surface. In such a case, the kinetic diameter of a simulant should be taken into consideration. Also, cost and ease-of-synthesis are important; however, these factors are beyond the scope of this work.

2.3.2 QSAR Modeling – Calculation of Molecular Descriptors

In addition to identifying improved simulants, we were interested in studying how specific molecular features affect organophosphate reaction energetics. In the previous section, we briefly noted that simulants with F^- and Cl^- leaving groups have similar activation barriers to nerve agents. Below, we aim to extend beyond basic structural similarities to probe how electronic structure descriptors affect the energetics. Implicitly, this analysis will also help explain which electronic structure descriptors are most important when selecting an ideal reactivity simulant. To achieve a more comprehensive understanding of the fundamental chemistry underlying organophosphate hydrolysis mechanisms, we employed quantitative structure-activity relationship (QSAR) modeling.

The first step in the development of any QSAR model, besides selecting a molecular dataset, is the calculation of relevant molecular descriptors. We calculated molecular orbital parameters, using electron population analysis with the NBO method in Gaussian 09, for all 119 molecules studied in the DFT screening and used these as some of the descriptors to build our QSAR models. Based on their potential relevance to the hydrolysis of organophosphate molecules, we compiled 12 descriptors, 11 of which were obtained through DFT calculations. All quantum chemical descriptors were computed from structures of the individual reactant molecules, as opposed to computing them from the transition state structure. The free energy barriers obtained from the DFT screening in the previous section (ΔG_{TS1}) served as the response variable, which, when combined with the descriptors, were used to formulate QSAR models.

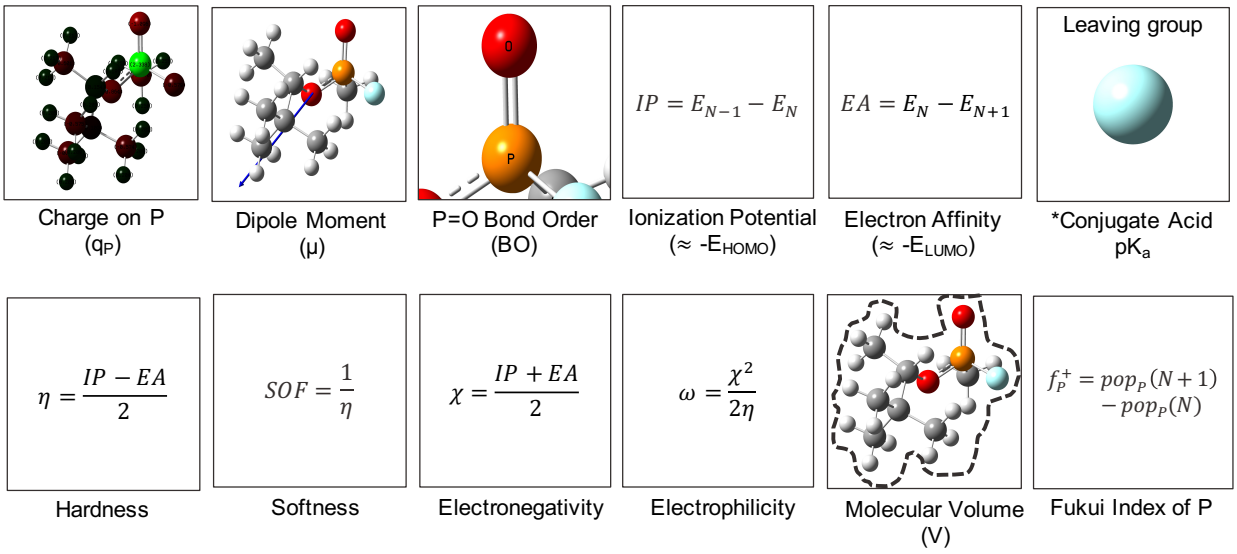


Figure 2.2. Calculated molecular descriptors used in the QSAR analysis, using the structure of soman as a representative example. *Note that all descriptors were obtained through DFT calculations, except for the conjugate acid pK_a of the leaving group. The experimental pK_a values were taken from available databases and correspond to those values listed in Table 2.1.

The descriptors are briefly summarized in Figure 2.2, using the structure of soman as a representative example. Based on basic chemical intuition and previous reports, we expected to see certain relationships (either direct or inverse correlations) between the selected descriptors and hydrolysis barriers, and the resulting QSAR model equations were used to confirm or deny our hypotheses. For example, we predicted that a higher charge on the central phosphorus atom (q_P) should facilitate nucleophilic attack through increased electrostatic attraction with the negatively charged hydroxide nucleophile.[100,135] The molecular dipole moment (μ) was chosen because it describes the overall polarity of the molecule. The bond order of the P=O bond (BO) was considered important, and we anticipated that the bond strength between P and the leaving group atom should weaken as the P=O bond order increases, resulting in a lower barrier. The conjugate acid pK_a of the leaving group is relevant because the overall hydrolysis rate should be dependent on the relative basicity of the leaving group compared to the nucleophile,[106] and pK_a was already shown in the previous section to have an effect on the degradation mechanism followed. The

molecular volume (*V*) was selected to investigate if the relative size has any effect on the approach of the nucleophile, and thus the reaction barrier.

The electron affinity (*EA*) was selected because the hydrolysis reaction consists of a pair of electrons on HO^- transferring into an unoccupied molecular orbital of the organophosphate. Koopmans' theorem[142] may be used to relate the ionization potential (*IP*) and *EA* to the HOMO and LUMO eigenvalues, respectively, where $IP \approx -E_{\mathrm{HOMO}}$ and $EA \approx -E_{\mathrm{LUMO}}$. Using this relationship, higher *EA* values should enable easier electron transfer, and thus lower energy barriers. Formally, *IP* and *EA* may be calculated as differences in electronic energies as follows:[143]

$$IP = E_{N-1} - E_N \tag{2.1}$$

$$EA = E_N - E_{N+1} \tag{2.2}$$

where the subscripts $N-1$, N, and $N+1$ denote the cationic, neutral, and anionic molecule, respectively. Here, the electronic energies are all calculated at the optimized geometry of the neutral molecule.

Molecular hardness (η) and softness (SOF) may be used to measure the stability and reactivity of molecules. Molecular electronegativity (χ) describes the ability of a species to attract electrons, and molecular electrophilicity (ω) measures the reactivity towards a nucleophile, where high values of these two descriptors should correspond to lower barriers. Using Koopmans' theorem, Parr and co-workers[144] derived equations for η, *SOF*, χ, and ω, which are reproduced here:

$$\eta = \frac{IP - EA}{2} \tag{2.3}$$

$$SOF = \frac{1}{\eta} \tag{2.4}$$

$$\chi = \frac{IP + EA}{2} \tag{2.5}$$

$$\omega = \frac{\chi^2}{2\eta} \tag{2.6}$$

To complement these global molecular descriptors, we applied the concept of Fukui indices, which are local reactivity descriptors that indicate preferred atomic sites for chemical reactions. Specifically, we used the atomic populations obtained from the NBO analysis to compute the Fukui index for nucleophilic attack on the phosphorus atom (f_P^+). f_P^+ describes the tendency of the electronic density to distort at the phosphorus atom upon accepting electrons from HO^-.[145] It can be defined as the difference in atomic populations between the molecule with an extra electron ($N + 1$) and the neutral molecule with N electrons:[146]

$$f_P^+ = pop_P(N + 1) - pop_P(N) \tag{2.7}$$

where the subscript P denotes that these are populations of the phosphorus atom. Here, populations for the anionic system are calculated at the optimized geometry of the neutral molecule. Molecules with larger f_P^+ values indicate more reactive phosphorus centers and should correlate with lower reaction barriers.

The unscaled values of the 12 descriptors for all 119 molecules are listed in Table A.4. Since these molecular descriptors have different units, all descriptors were scaled from 0 to 1 so that the weights of each descriptor in the developed QSAR models may be more clearly compared (see eq A.1 in Appendix A for scaling formula). The scaled descriptors for all molecules are listed in Table A.5. We note that there are a virtually infinite number of other chemical descriptors available in the literature, but we chose to limit our focus to these 12, whose selection relies only on basic chemical intuition. The scaled molecular descriptors, along with the ΔG_{TS1} response variable, were used to construct QSAR models by applying stepwise multiple linear regression (MLR), as described below.

2.3.3 QSAR Modeling – Linear QSAR Model Development

Before assessing the developed QSAR models, we first discuss uniparametric correlations for individual descriptors (e.g., R^2_{uni} for ΔG_{TS1} vs. q_{P}). The 5 most statistically significant descriptors for describing the alkaline hydrolysis of all 119 organophosphates are: pK_{a} ($R^2_{\text{uni}} = 0.52$, positive slope), q_{P} ($R^2_{\text{uni}} = 0.21$, positive slope), BO ($R^2_{\text{uni}} = 0.19$, negative slope), f_P^+ ($R^2_{\text{uni}} = 0.16$, negative slope), and V ($R^2_{\text{uni}} = 0.15$, positive slope). Clearly, the conjugate acid pK_{a} of the leaving group is the single most important descriptor (of the 12 considered) for describing the energetics of these reactions, and the sign of the term reveals that strongly basic leaving groups correlate with higher free energy barriers. Surprisingly, the sign of the q_{P} term indicates that higher charges on P correlate with higher barriers, which is opposite to our hypothesis in the previous section. A potential explanation is that although the increased charge on the phosphorus center could strengthen the interaction with the incoming HO^- nucleophile, attraction to the leaving group could also increase. The net effect would then be determined by the relative basicity of the leaving group compared with that of the nucleophile. The signs of the BO and f_P^+ terms confirm our expectations that higher P=O bond orders and more reactive phosphorus centers correlate with lower barriers. The sign of the molecular volume term indicates that smaller molecules generally correlate with lower barriers.

Out of the 119 molecules studied, there are several possible molecular datasets (i.e., sets of molecules with a unique characteristic in common) that contain enough molecules to create a statistically significant QSAR model including: all molecules ($N_{\text{m}} = 119$), those hydrolyzing through concerted mechanisms ($N_{\text{m}} = 57$) or stepwise mechanisms ($N_{\text{m}} = 62$), those with a F^- leaving group ($N_{\text{m}} = 37$) or Cl^- leaving group ($N_{\text{m}} = 25$), phosphono– molecules composed of alkyl and alkoxy substituents ($N_{\text{m}} = 58$), and phosphoro– molecules composed of two alkoxy

substituents (N_m = 54). Initially, we built non-predictive multi-parametric QSAR models for all subsets. The term non-predictive denotes that, for each molecular subset, all molecules were included in the model development as opposed to dividing them into training and test sets. This was done to gain a better understanding of the most important descriptors involved in the hydrolysis of organophosphates having particular attributes. Herein, we focus our discussion on the model created using all 119 molecules. Results for the other smaller subsets can be found in Appendix A.

For model development, we used stepwise forward-backward based feature selection combined with MLR using a 95% confidence interval, utilizing the built-in *stepwiselm* function in MATLAB. MLR allows for easy interpretation of the contribution each descriptor has on the model in terms of its coefficient weight and sign. This algorithm performs forward selection of descriptors if their corresponding *p*-values are less than 0.05, then uses backward elimination of the descriptor with the largest *p*-value if any descriptors in the model at the current step have *p*-values higher than 0.10. Here, the *p*-values are for an F-test of the change in the sum of squared error resulting from the addition or removal of a descriptor. Descriptors are also removed at each step if they are linearly dependent with other descriptors. The algorithm terminates whenever a single step cannot improve the model statistics. We note that there are a number of other more sophisticated methods that can be applied to build the models, such as artificial neural networks.[147] However, this added complexity is not warranted for our application here, as is evident by the promising statistical results detailed below using simple MLR.

The non-predictive multi-parametric QSAR model developed using all 119 molecules is:

$$\Delta G_{TS1} = (73.70 \pm 4.88) + (58.81 \pm 9.58)q_P - (80.68 \pm 11.99)BO - (12.33 \pm 7.40)\mathrm{p}K_a - (10.47 \pm 2.92)\omega + (10.60 \pm 4.38)V \quad (2.8)$$

$N_m = 119$, $Q^2_{LOO} = 0.64$, $R^2 = 0.72$, RMSE = 8.02 kJ•mol^{-1}, $F = 59.2$.

Here, ΔG_{TS1} is in kJ/mol and the descriptors are dimensionless (scaled from 0 to 1, as described earlier). In eq 2.8, N_m is the number of molecules, R^2 is the correlation coefficient between observed and predicted responses, and RMSE is the root mean square error. The leave-one-out cross-validation (LOO-CV) procedure is used to calculate Q^2_{LOO}, where every molecule is eliminated from the dataset once and then its response variable is predicted using the QSAR equation produced from the remaining set.

We described the most significant uniparametric correlation coefficients above for completeness and transparency, and to show that single descriptors are not sufficient to describe such a heterogeneous dataset. The results highlight the well-known fact that adding increasingly more terms to a regression model will improve the correlation statistics. However, the 5 descriptors in eq 2.8 are a reasonably small set of terms, allowing for a clear interpretation of their relationships with the free energy barrier. It is important to note that the absence of the other 7 descriptors from eq 2.8 does not imply that some of these parameters are not also important for describing organophosphate hydrolysis reactions. For example, ω and *EA* have an inter-descriptor correlation coefficient of $R^2_{id} = 0.98$, as detailed in the matrix below eq A.2 in Appendix A. Theoretically, *EA* could be substituted for ω in eq 2.8 to produce a slightly different QSAR model, without significantly affecting the model statistics.

Non-predictive multi-parametric QSAR models and uniparametric correlation coefficients for the other molecular subsets are described in eqs A.3–A.8 in Appendix A. Predictably, careful inspection of these supplementary results shows that the correlation coefficients are higher for the smaller molecular subsets because they are more homogeneous datasets, and the most significant descriptors vary between subsets.

The relatively high R^2 and Q^2_{LOO} statistics and the low RMSE value of eq 2.8 suggest that these 5 descriptors can produce QSAR models with acceptable accuracy for predicting the ΔG_{TS1} barriers. To confirm this, we performed statistical validations and predictions.

2.3.4 QSAR Modeling – QSAR Model Validation and Predictions

To determine the predictive capacity of the dataset ($N_m = 119$), we split the molecules into training and test sets, making sure that the training set spanned the entire response variable space[148] and that the test set included ≈15% of the total dataset.[149] To create a predictive QSAR model, we tested three different algorithms for dividing molecules: (i) random selection, (ii) cluster by rank in which molecules were sorted by ΔG_{TS1} values into a specific number of groups and then randomly selected from each group for the test set, and (iii) modified cluster by rank in which a rational division of molecules was executed to ensure a reasonable dispersion of conjugate acid pK_a values within the training and test sets. Each method included molecules with the highest and lowest ΔG_{TS1} values in the training set. More detailed descriptions of the processes used in algorithms (ii) and (iii) are included in Appendix A.

We ran each dataset division algorithm 100 times to examine the variability in model statistics. As seen in Figures A.3–A.6, the more rational algorithms (ii) and (iii) did not have a strong influence on the resulting QSAR model statistics and actually performed worse on average than (i) for certain statistics. The models that used methods (ii) and (iii) and resulted in the best statistics are highlighted in Appendix A. We restrict our statistical analysis and discussion herein to the best random selection model, using method (i).

Several tests for statistical significance were used to evaluate the predictive value of the models. To evaluate the training set, the Q^2_{LOO} correlation coefficient was used. Q^2_{LOO} can be calculated using the formula:

$$Q^2_{LOO} = 1 - \frac{\Sigma(y_{obs(train)} - y_{pred(train)})^2}{\Sigma(y_{obs(train)} - \bar{y}_{(train)})^2} \quad (2.9)$$

where $y_{obs(train)}$ are the actual ΔG_{TS1} responses, $y_{pred(train)}$ are responses estimated based on the LOO-CV procedure, $\bar{y}_{(train)}$ is the average response, and summations are over all training set molecules.

The possibility of chance correlation or structural redundancy in the developed QSAR models was measured by the *y*-randomization method. This involved randomizing the response values, while leaving the descriptor matrix unchanged, and repeating the entire process of statistical validation. Using the correlation coefficients for the randomized model (R^2_r) and non-random model (R^2_{nr}), the following parameter was calculated:

$$^{C}R^2_P = R_{nr} \times \sqrt{R^2_{nr} - R^2_r} \quad (2.10)$$

where $^{C}R^2_P > 0.5$ indicates that the QSAR model is not generated by chance only.[150]

To evaluate the test set, the test set correlation coefficient (R^2_{test}) was calculated as:

$$R^2_{test} = 1 - \frac{\Sigma(y_{obs(test)} - y_{pred(test)})^2}{\Sigma(y_{obs(test)} - \bar{y}_{(test)})^2} \quad (2.11)$$

where the summations are over all test set molecules. In general, models are deemed acceptable in the QSAR community if $Q^2_{LOO} > 0.5$ and $R^2_{test} > 0.6$.[151]

The prediction accuracy for the test set is also typically evaluated in terms of the root mean square error in prediction ($RMSE_{test}$). There is no definitive criterion for how low the $RMSE_{test}$ should be, as it depends on the application of interest and the user's error tolerance. Although R^2_{test} is important to calculate to validate models, the $RMSE_{test}$ is perhaps a better indicator of a model's practical usefulness.[152] Furthermore, the value of a QSAR model is not exclusively dependent on the quantitative accuracy of its predictions. The relative ranking of molecules, in terms of response variable, is also significant. Thus, the Spearman's rank correlation coefficient (ρ) may also be

calculated. For our study, accuracy in predicting the ΔG_{TS1} barriers is just as important as predicting trends in the barriers, in the context of determining which simulant molecules have predicted barriers closer to the actual nerve agents. Thus, when comparing computed statistics across all 100 models generated using the random selection algorithm, we identified the optimal model as the one that yielded the lowest $RMSE_{test}$.

The optimal predictive QSAR model developed using the random selection algorithm is:

$$\Delta G_{TS1} = (139.92 \pm 16.84) + (72.47 \pm 8.94)q_P - (93.91 \pm 11.86)BO - (19.75 \pm 7.33)\mathrm{p}K_a - (66.51 \pm 16.94)\eta - (76.38 \pm 18.23)SOF \quad (2.12)$$

Training set: $N_m = 101$, $Q^2_{LOO} = 0.64$, ${}^{C}R_P^2 = 0.68$, $R^2_{train} = 0.73$, $RMSE_{train} = 8.25$ kJ•mol^{-1}, $F = 50.3$

Test set: $N_m = 18$, $R^2_{test} = 0.87$, $RMSE_{test} = 4.36$ kJ•mol^{-1}, $\rho = 0.91$.

As before, ΔG_{TS1} is in kJ•mol^{-1} and the descriptors are dimensionless. Based on our established criteria, eq 2.12 represents a model with excellent predictive capacity, in terms of both quantitative ($RMSE_{test} = 4.36$ kJ•mol^{-1}) and ranking ($\rho = 0.91$) accuracy. The y-randomization statistic is ${}^{C}R_P^2 = 0.68$, which indicates that the QSAR model is not generated by chance only. Figures A.3–A.6 contain plots of Q^2_{LOO}, R^2_{test}, $RMSE_{test}$, and ρ for all 100 QSAR models generated using the random selection algorithm, highlighting the variability in model statistics.

We note that some descriptors in eq 2.8 selected when using the entire dataset ($N_m = 119$) are different from those in eq 2.12 that is built on only a training set ($N_m = 101$). Aside from the fact that the two models are constructed using slightly different datasets, the reason for the descriptor differences may be due to the stepwise MLR method itself. Stepwise models are locally, but not necessarily globally, optimal because a different initial model or a different sequence of steps may lead to slightly different fits, all of which may provide equally accurate predictions.

Regardless of the cause, these differences highlight the importance of generating both non-predictive and predictive QSAR models, depending on what the purpose of the model may be. While the non-predictive model in eq 2.8 is useful for gaining insight into the most significant descriptors involved in organophosphate hydrolysis, only the predictive model in eq 2.12 can be used to reliably predict free energy barriers for new molecules.

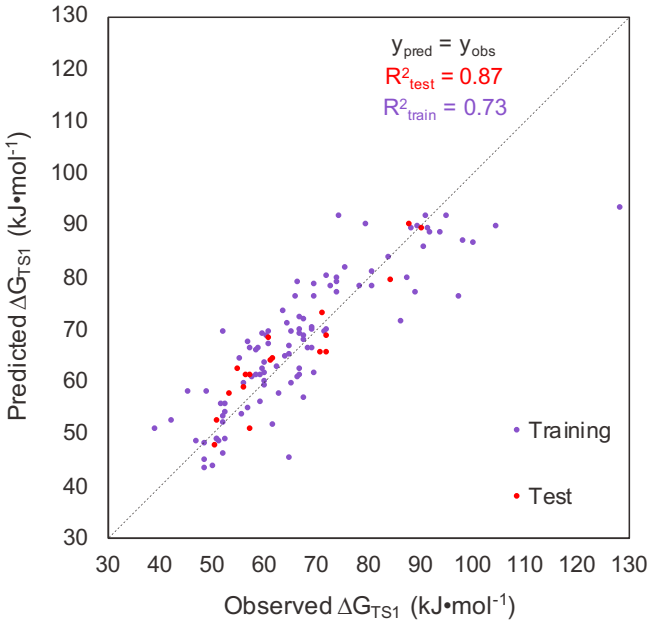


Figure 2.3. Comparison between observed (DFT-calculated) free energy barriers with those predicted by the optimal predictive QSAR model developed using the random selection algorithm, eq 2.12. The dashed line shows the relationship $y_{pred} = y_{obs}$, where data points for ideal models should lie close to this line.

ΔG_{TS1} barriers for all 119 molecules predicted using eq 2.12 are given in Table A.6. The regression plot of observed (DFT-calculated) and predicted barriers for training and test set molecules is represented in Figure 2.3, along with the corresponding linear correlation coefficients between observed and predicted barriers ($R^2_{train} = 0.73$, $R^2_{test} = 0.87$). The dashed line in this plot that shows the relationship $y_{pred} = y_{obs}$ is included to help assess the predictive power of the model, where data points for ideal models should lie close to this line.

From the values in Table A.6 and Figure 2.3, it is clear that the agreement between observed and predicted barriers is quite good. There is one noticeable outlier in the training set (identified as molecule 119 in Table A.3), although the statistical correlations are still significant even with this molecule included in building the model. Figure A.8 plots the distribution of raw residuals for the model, showing that the residuals have an acceptable normal distribution, with the one outlier. Further discussion of this outlier is continued in the next section.

2.3.5 QSAR Modeling – External Predictions and Assessment of Applicability Domain

The optimal predictive QSAR model in eq 2.12 was further applied to predict ΔG_{TS1} barriers for a set of 3 external molecules, not included in the training or test sets. Since our QSAR models were specifically developed using sarin and soman simulants, we wanted to determine if eq 2.12 could also reliably predict other nerve agent types such as cyclosarin (GF), tabun (GA), VX, and their related simulants whose structures were not involved in building the model. To accomplish this, we performed additional DFT calculations to generate observed ΔG_{TS1} values and molecular descriptors for GF, GA, and VX. Their ΔG_{TS1} barriers are listed at the end of Table A.3 (identified as molecules 120, 121, and 122), and the unscaled and scaled values of their 12 molecular descriptors are listed in Table A.4 and Table A.5, respectively.

To more easily assess how eq 2.12 performed for these external set molecules, we established the applicability domain (AD) of the optimal predictive QSAR model using the leverage-based method.[153,154] Predictions for molecules can only be considered reliable and not extrapolations if they lie within this AD. In other words, the AD represents the bounds in which a model tolerates a new molecule.[154] To visualize the AD, a Williams plot was constructed using the standardized cross-validated residuals and the leverage values with established thresholds.

Leverage (h) is simply a measure of the influence of a molecule's structure on the regression model. Leverages for individual molecules within the training and test sets, for example, may be calculated as the diagonals of the hat matrices (H), defined as:

$$H_{train} = X_{train}(X_{train}^T X_{train})^{-1} X_{train}^T \quad (2.13)$$

$$H_{test} = X_{test}(X_{train}^T X_{train})^{-1} X_{test}^T \quad (2.14)$$

where X_{train} and X_{test} are the design matrices containing the molecular descriptors for the training and test set molecules, respectively. More specifically, each row in the design matrices contains the 12 scaled descriptor values for an individual molecule. The cut-off leverage is defined as $h^* = 3(N_d + 1)/N_m$, where N_d is the number of descriptors used in the final model and N_m is the number of training set molecules used to build the model.[153] A prediction for a molecule is considered unreliable and outside the AD if its $h > h^*$.

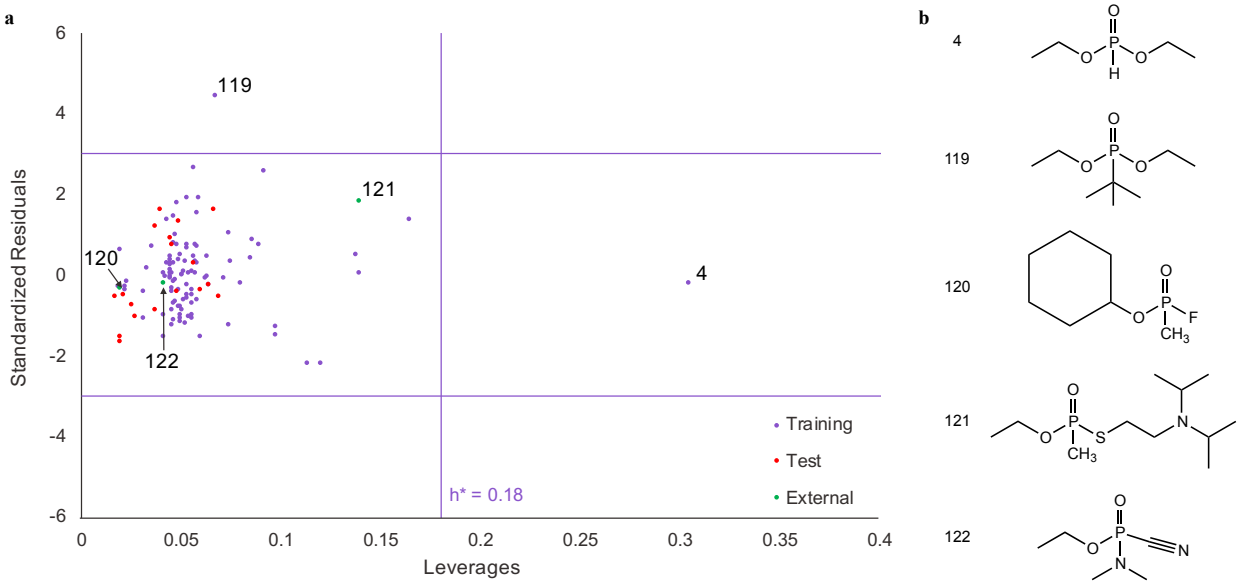


Figure 2.4. (a) Williams plot showing the applicability domain for the optimal predictive QSAR model developed using the random selection algorithm, eq 2.12. The horizontal purple lines signify the bounds for the standardized residuals (at ± 3 standard deviation units). The vertical purple line represents the cut-off leverage (h^*). The numbers correspond to molecule numbers, established in Table A.3. (b) The structures of the five numbered molecules.

From the Williams plot, it is easy to detect response outliers (with standardized residuals > 3 standard deviation units) and structurally influential molecules with high leverage ($h > h^*$). The AD for the optimal predictive QSAR model is represented in Figure 2.4a, where purple, red, and green points denote training, test, and external set molecules, respectively. The horizontal purple lines signify the bounds for the standardized residuals (at $\pm$ 3 standard deviation units) and the vertical purple line represents the cut-off leverage ($h^* = 0.18$). Only 2 out of the 122 total molecules are located outside of the AD, and so predictions for all other 120 organophosphates using eq 2.12 can be considered reliable. The 2 outliers (molecules 4 and 119) are both in the training set. Molecule 4, whose structure is presented in Figure 2.4b, is a structurally influential outlier. By comparing its structure with the rest of the molecules in Table A.3, it is apparent that molecule 4 is the only organophosphate considered with a hydrogen substituent, whereas all other molecules have methyl or larger substituents. However, molecule 4 can be said to have "good" leverage since it is in the training set, meaning it reinforces the strength of the model for any future predictions. Molecule 119, whose structure is represented in Figure 2.4b, is a response outlier, as is also evident from Figure 2.3. Molecule 119 contains a tert-butyl substituent, which is a very bulky moiety compared to all the other organophosphates. eq 2.12 wrongly predicts its free energy barrier and can be expected to perform poorly for other molecules with equally large alkyl groups.

Perhaps surprisingly, all 3 external set nerve agents fall within the AD. This inspires further confidence in our developed model because both VX and GA (molecules 121 and 122, highlighted in Figure 2.4b) contain leaving groups not included in the training or test sets. This result implies that eq 2.12 can also be used to make reasonable predictions for organophosphates with new leaving groups not considered herein.

2.4 Conclusions

In this work, we performed DFT calculations on the uncatalyzed alkaline hydrolysis mechanisms for 119 organophosphate molecules to determine improved simulants for the G-series nerve agents soman and sarin, based on criteria such as low toxicity and similarity to nerve agent hydrolysis energetics and degradation mechanism. All molecules were observed to undergo S_N2 nucleophilic substitution pathways, with hydroxide acting as the nucleophile. We observed a relationship between the conjugate acid pK_a of the leaving groups and the reaction pathway (concerted vs. stepwise) that molecules followed. For all 119 molecules, the rate-limiting free energy barrier for TS_1 was chosen as a basis for comparing simulants to nerve agents. To compare toxicities, we used the oral rat LD_{50} $-\log_{10}(\text{mol}\bullet\text{kg}^{-1})$ toxicity endpoint obtained from the EPA Toxicity Estimation Software Tool. Based on our established energy barrier and toxicity criteria, we identified 4 improved soman simulants and 3 improved sarin simulants. The improved simulants do not all have a similar molecular structure as their corresponding nerve agents, but it is evident that molecules with F^- and Cl^- leaving groups are generally better reactivity simulants. Although there is an overlap between top soman and sarin simulants, it is important to recognize that there is no single best simulant for all types of nerve agents. Many common simulants like DMMP showed large energy barrier deviations from the agents, serving as a caution against selecting simulants based on literature precedent alone.

Overall, we highlighted 5 unique molecules that have low estimated toxicities and are predicted to have nearly identical rate-determining hydrolysis reaction barriers to actual nerve agents. Our screening aims to narrow down the search space for potential simulants and to aid researchers in accurately and safely studying these reactions. If our simulant predictions are proven reliable, this could allow researchers to confidently predict kinetic degradation rates for promising

catalyst materials, potentially accelerating research in this field and improving overall lab safety. When translating our results for potential experimental validation, it is important to remember that, just as the choice of simulant depends on the desired application, it also heavily depends on the method of degradation. For example, the choice of solvent (water, methanol, etc.), catalyst identity (MOF, metal oxide, polyoxometalate, etc.), and the active site environment (elements present, steric interactions, etc.) are all important factors that may affect how a simulant degrades relative to actual nerve agents. Further experimental testing of the suggested simulants, in terms of toxicity, is also warranted if they are to be used in an academic environment. While validation with actual nerve agents is still necessary, using simulants with nearly identical reaction barriers will help to narrow down promising catalyst materials, as opposed to simply proving that a material is active towards organophosphate degradation without a true indicator of exactly how active. We believe that our determination of highly predictive and less toxic nerve agent simulants will be valuable for the broader scientific community pursuing the challenge of destroying chemical warfare agents.

To achieve an even more comprehensive understanding of the fundamental chemistry underlying organophosphate hydrolysis mechanisms, we employed quantitative structure-activity relationship modeling. Twelve molecular descriptors, obtained from quantum chemical calculations, were chosen as potentially relevant parameters for predicting the free energy barriers for these reactions. From our statistical analysis, we concluded that the most significant uniparametric descriptors for describing the alkaline hydrolysis of organophosphates are conjugate acid pK_a of the leaving group, charge on the phosphorus atom, P=O bond order, the Fukui index for nucleophilic attack on the phosphorus atom, and molecular volume. Weakly basic leaving groups, lower charges on P, higher P=O bond orders, more reactive phosphorus centers, and smaller molecules all correlate with lower free energy barriers. We then built multi-parametric

QSAR models using stepwise multiple linear regression, highlighting that single descriptors are not sufficient to describe such a heterogeneous dataset. To determine the predictive capacity of our models, we split the molecules into training and test sets using three different algorithms and performed statistical validations. Each dataset division algorithm was run 100 times, and we observed that simple random selection was sufficient for achieving acceptable QSAR model statistics. We identified the optimal model as the one that had the lowest root mean square error in prediction, which yielded a model that showed excellent predictive capacity, in terms of both quantitative and ranking accuracy, for determining free energy barriers. Since our model was developed using sarin and soman simulant data, we further tested its robustness by predicting barriers for 3 external set nerve agents (cyclosarin, tabun, and VX), whose structures were not involved in building the model. Through the assessment of its applicability domain using the leverage-based method, we showed that the optimal predictive QSAR model could reliably predict energetics for other organophosphate classes, even tabun and VX which contain leaving groups not included in the training or test sets. This implies that our model can be used to make reasonable predictions for organophosphates with other novel leaving groups.

The prediction of free energy barriers using DFT-calculated molecular descriptors is relatively simple and computationally cheap. Thus, our QSAR models generated using these descriptors can be utilized to predict alkaline organophosphate hydrolysis barriers without having to perform relatively expensive transition state calculations or dangerous experiments. In addition to predicting free energy barriers for uncatalyzed reactions, the molecular descriptors could potentially be used in new QSAR models for other applications. For instance, different response variables such as binding energies could be implemented to make predictions for uptake in porous materials[155] or for sensing and detection.

Chapter 3: Insights into Catalytic Hydrolysis of Organophosphate Warfare Agents by Metal–Organic Framework NU-1000

This chapter is a modified version of a published manuscript: Chen, H.; Liao, P.; Mendonca, M. L.; Snurr, R. Q. *J. Phys. Chem. C* **2018**, *122*, 12362–12368. Prof. Peilin Liao started this project and performed initial calculations before starting a faculty position at Purdue University. Dr. Haoyuan Chen and I finished the calculations, analyzed the data, and wrote the manuscript together.

3.1 Introduction

Chemical warfare agents (CWAs) were first used on a large scale in World War I but continue to be a major threat to the health and lives of many people in the world. CWAs that contain phosphate ester groups, called nerve agents, are extremely toxic because they can rapidly inactivate acetylcholinesterase, which is a key enzyme for nerve function.[1,156,157] Efficient destruction of phosphate CWAs has therefore drawn significant interest in the chemistry community. Although there are enzymes such as phosphotriesterase[156,158–160] that can break down phosphate ester bonds rapidly via catalyzed hydrolysis, the large-scale use of enzymes in nerve agent destruction is limited by the restrictive conditions under which the enzymes are active. Hence, solid materials that are more robust under different conditions have been applied to the destruction of CWAs, including heterogeneous catalysts such as zeolites,[161] polyoxometalates,[93,162,163] and metal–organic frameworks (MOFs).[1,21,25,57,62,141] In particular, MOFs have drawn significant attention because of their amenability to modular design and synthesis, which opens up the potential to catalytically decompose a broad spectrum of CWAs and improve the catalytic efficiency via chemical

modification.[1,57,62,66,164–167] Among them, zirconium-based MOFs have also been shown to possess high thermal (up to 500 °C) and chemical (pH 1–12) stabilities because of the strong Zr–O bonds in the nodes, which make them promising candidates for catalysis.[47,66] Further improvement of MOF catalysts can benefit from the knowledge of enzyme catalysis mechanisms,[164,165,168] as has already been demonstrated in the design of MOFs for CWA destruction to date.[25,169]

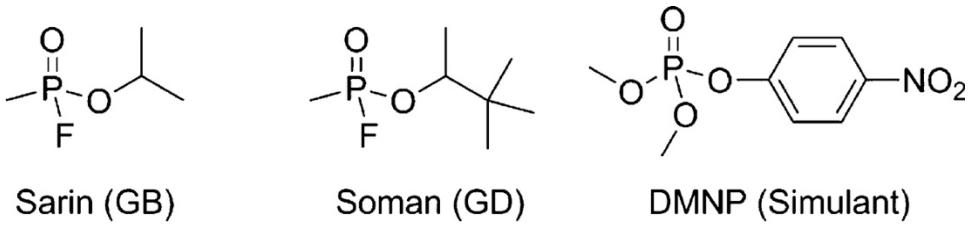


Figure 3.1. Chemical structures of two G-series nerve agents and a simulant that have been studied for catalytic hydrolysis on Zr-based MOFs.

In addition to experimental studies, theoretical calculations have also been used to study the catalytic destruction of nerve agents and simulants in MOFs with the goal of providing insights at the atomistic level and guiding the design of new materials with improved catalytic properties.[1,51,57–59] For example, Troya reported a gas-phase mechanistic study of the hydrolysis of the G-series nerve agent Sarin (GB, propan-2-yl methylphosphonofluoridate) and the simulant dimethyl methylphosphate on zirconium-based MOFs UiO-66 and MOF-808.[57–59] Mondloch et al. studied another Zr-based MOF NU-1000 both experimentally and computationally. In their work, dimethyl 4-nitrophenylphosphate (DMNP or methyl-paraoxon) was used as a simulant to mimic real nerve agents such as Soman (*O*-pinacolyl methylphosphonofluoridate, also known as GD, see Figure 3.1).[1] The use of simulants with similar structures but much lower toxicities than CWAs allows more comprehensive experiments to be performed in university laboratories. More recently, Islamoglu et al. performed calculations to investigate the effect of amino-functionalized linkers on UiO-66-catalyzed DMNP hydrolysis.[51]

In the present work, we aim to gain deeper insights into the phosphate ester hydrolysis reaction catalyzed by NU-1000 using density functional theory (DFT) calculations. In the previous work by Mondloch et al., the binding between the simulant and the MOF was studied, but the reaction free energy barrier and the complete reaction pathway were not investigated. Here, we map out the entire catalytic cycle of NU-1000-catalyzed hydrolysis of the simulant DMNP and calculate the free energy barrier. Detailed comparisons between the uncatalyzed and catalyzed transition states (TSs) are performed to shed light on the origin of the catalytic effect, revealing a resemblance to enzyme catalysis in terms of mechanistic fingerprints. The effects of temperature-induced node dehydration and distortion on catalytic efficiency are also discussed, along with a comparison to previous experimental kinetics data.

3.2 Computational Details

All electronic structure calculations were performed using the Gaussian 09 package (revision D.01).[109] The cluster used for representing Zr-based NU-1000 was taken from Mondloch et al.,[1] which corresponds to the mixed-staggered proton topology.[48] The bottom four benzene linkers were replaced by H atoms to speed up the calculations (see Figure B.1 in Appendix B for the geometry of the simplified cluster). Because those benzene linkers are far away from the reactive site and do not interact directly with the reactant, this simplification should have a minimal effect on the reaction on the other side of the node. The H atoms farthest from the Zr node in each benzene ring of the linkers were held fixed to mimic the constraints imposed by the surrounding MOF structure, whereas the rest of the benzene atoms were allowed to relax. The reactive species – including the DMNP molecule, the attacking hydroxide from solution, and nearby node OH groups – were allowed to relax, whereas the remaining atoms were held fixed. The atomic coordinates of

the distorted node were obtained from Platero-Prats et al.[170] (see Figure B.1 for the geometry of the simplified cluster).

Geometry optimizations were performed for all species using DFT with the B3LYP functional,[72,73] and the DFT-D3 dispersion correction[85] with the Becke–Johnson damping function[86,87] was applied for single-point energy calculations. The 6-31+G** basis set was used for the reactive species (defined above), whereas the 6-31G* basis set was used for the rest of the atoms in the node. The LANL2DZ[77,78,171] basis sets with effective core potentials were applied for Zr. Several other methods were tested on the uncatalyzed hydrolysis reaction of DMNP to explore the effect of level of theory on the energy barrier. The M06 and M06-2X functionals,[110] second-order Møller–Plesset theory (MP2),[111] Hartree–Fock method, and the modified complete basis set method (CBS-QB3)[112,113] were tested and compared to the available experimental data (see Table B.1). The polarizable continuum model[81,114] with default atomic cavity radii was used to model solvation effects implicitly. Harmonic vibrational analyses were performed to verify the nature of all stationary points and to calculate the thermochemical properties at standard conditions (298.15 K, 1 atm). Single-point energy calculations were also performed to refine the energies of all species, in which the 6-31G* and 6-31+G** basis sets were replaced by a larger 6-311++G** basis set.

3.3 Results and Discussion

3.3.1 Uncatalyzed Hydrolysis of DMNP

To investigate the catalytic effects of NU-1000 on DMNP hydrolysis, we first studied the baseline reaction, which is uncatalyzed hydrolysis of DMNP in aqueous solution. Phosphate ester hydrolysis reactions may have different mechanistic scenarios depending on various factors

including the pH.[124,172–174] Here, because the corresponding experiments[1] were performed at pH = 10, we focused on the alkaline hydrolysis mechanism where an OH^- anion acts as the nucleophile.

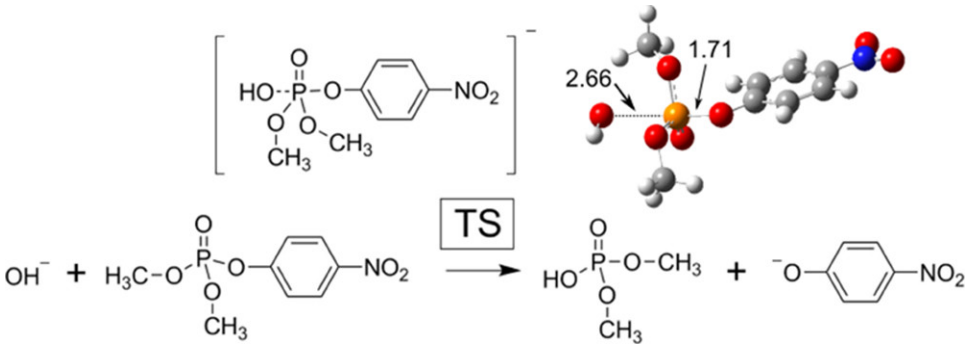


Figure 3.2. Illustration of the uncatalyzed alkaline hydrolysis mechanism of DMNP. Optimized TS structure and key bond lengths (in Å) are also shown.

As shown in Figure 3.2, the reaction is a single-step S_N2 nucleophilic substitution in which the TS mainly involves attack of the nucleophile but not the departure of the leaving group (also referred to as an "early TS"). This is consistent with previously reported calculations on similar systems.[124,125] The free energy barrier using B3LYP-D3(BJ)/6-311++G**//B3LYP/6-31+G** is 55 kJ/mol, which is in good agreement with the high-level CBS-QB3 results presented in Table B.1. The D3(BJ) dispersion correction is used for all calculations because the adsorbate has significant dispersion interactions with the MOF. A comparison of the enthalpic and free energy barriers across an extended set of methods is provided in Table B.1.

3.3.2 Catalytic Effects of NU-1000

In the previous work by Mondloch et al.,[1] the binding mode between DMNP and NU-1000 in the reactant state (RS) has been characterized by DFT calculations. Here, we also located the TS structure of the reaction and mapped out the complete reaction pathway, which is shown in Figure 3.3. First, the nonbridging phosphoryl oxygen atom in DMNP (O=P) forms a strong interaction with a Zr atom that has an open coordination site on the dehydrated node, as shown in the RS in Figure 3.4. Note that in the regular form of NU-1000, the coordination of that Zr atom is saturated

by a water molecule. We found that the creation of the open site through dehydration, which is essential for the catalysis, requires approximately 35 kJ/mol of energy (Figure 3.5). This node topology of partially dehydrated NU-1000 has also been confirmed by experimental IR spectroscopy and DFT calculations in previous work.[175]

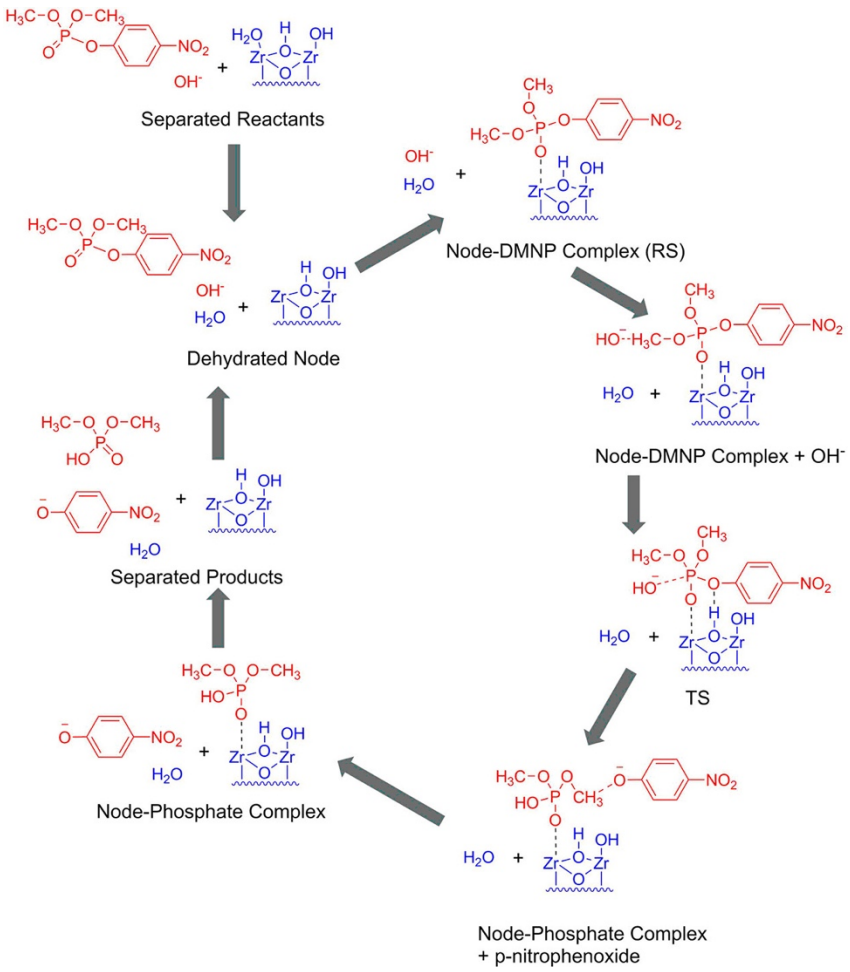


Figure 3.3. The catalytic cycle of DMNP hydrolysis on a NU-1000 node.

Then, similar to uncatalyzed alkaline hydrolysis, an OH^- anion from the solution attacks the phosphorous atom in a concerted mechanism, where only one TS is involved. This stands in contrast to the stepwise multi-TS mechanism observed for Sarin hydrolysis.[57] Upon desorption of the product, the catalytic site is regenerated, completing the cycle.

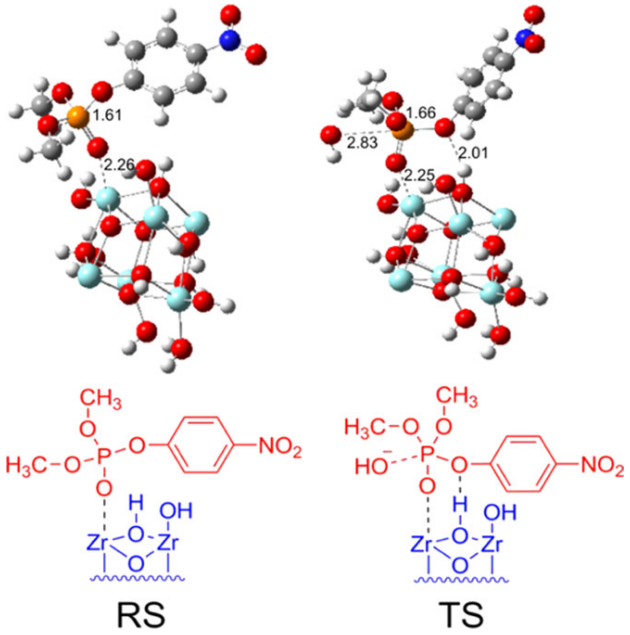


Figure 3.4. Schematic drawing (bottom) and optimized RS (node–DMNP complex) and TS structures (top) of the DMNP hydrolysis reaction catalyzed by the Zr_6 node in NU-1000. Key bond lengths are expressed in Å. The organic linker molecules in NU-1000 are also included in the computational model but omitted in the images for clarity.

A noticeable difference between the TS for the uncatalyzed reaction and the TS on the Zr node is the presence of a hydrogen bond between the leaving group oxygen and an OH group on the node (shown at 2.01 Å in Figure 3.4). This resembles a common theme in enzyme-catalyzed phosphate ester hydrolysis and phosphoryl transfer reactions, which is the stabilization of the negative charges that are building up on leaving group oxygen atoms to facilitate the cleavage of the P–O bond.[174,176–181] The TS structure here, compared to the TS in the uncatalyzed reaction, is even "earlier" (phosphorous is farther away from the nucleophile and closer to the leaving group) because the hydrogen bond effectively increases the leaving group ability. A trend has been demonstrated for similar types of reactions that the better the leaving group is, the "earlier" the rate-limiting TS is and the lower the reaction barrier.[124,182–185] Indeed, we also obtain a lower barrier for the MOF-catalyzed reaction than the uncatalyzed one. As shown in Figure 3.5, the barrier of the MOF-catalyzed reaction (defined as the free energy difference between TS and node–

DMNP complex) is 12 kJ/mol lower than the uncatalyzed one (54.7 vs 43.0 kJ/mol). This 43 kJ/mol DMNP hydrolysis barrier for NU-1000 compares well with the 47.3 kJ/mol barrier for UiO-66 calculated in a recent study using the M06-L functional and a continuum-cluster solvation scheme.[51] Also, note that the barrier for product desorption (defined as the free energy difference between separated products and node–phosphate complex) is only 10.5 kJ/mol, which suggests no product inhibition effects in the aqueous solution degradation reaction.

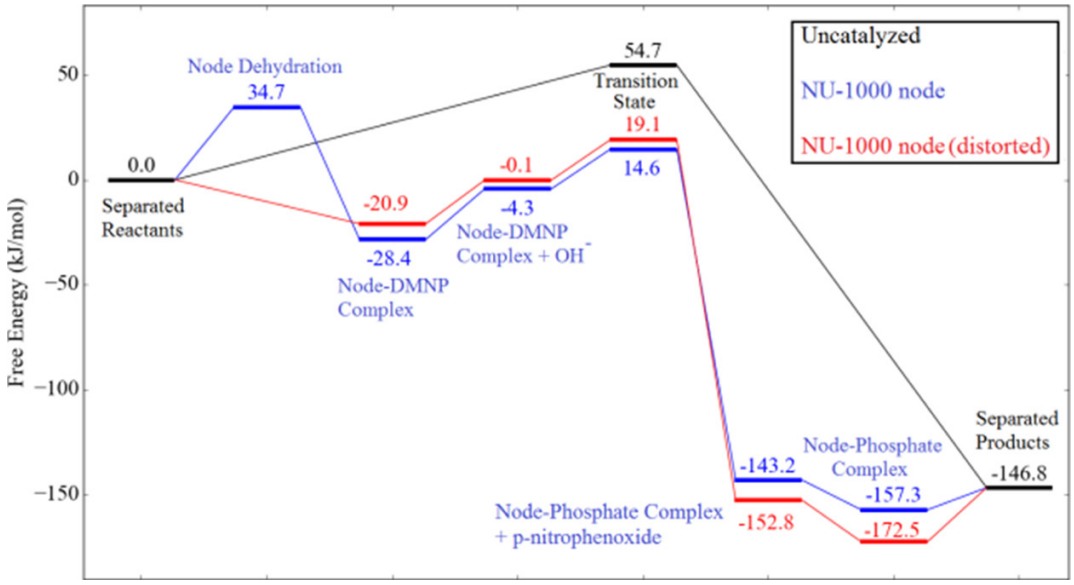


Figure 3.5. Comparison of the reaction free energy profiles of uncatalyzed (black), regular NU-1000 node-catalyzed (blue), and distorted NU-1000 node-catalyzed (red) DMNP hydrolysis reactions.

We further examined the catalytic effect of the Zr node by analyzing the electronic structure of key species using electron population analysis with the natural bond orbital (NBO) method.[119,186] We find that in the catalyzed TS, the phosphorous atom possesses more positive charge (2.65 vs 2.54) than in the uncatalyzed TS, as shown in Figure 3.6, which makes it more readily attacked by the negatively charged OH^- nucleophile. The extra amount of positive charge can be traced to the electron-withdrawing Zr atom, which exhibits a lower partial charge of 1.95 in the TS with DMNP compared to a partial charge of 2.29 in the metal node alone (Figure 3.6).

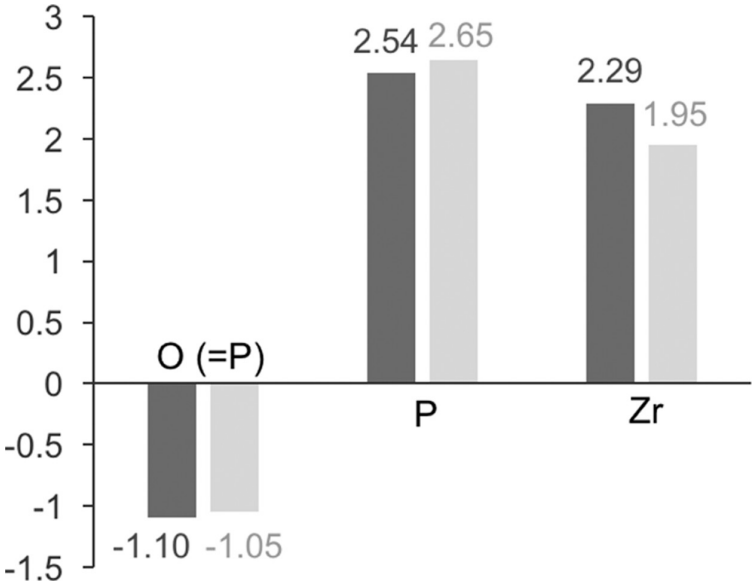


Figure 3.6. Partial charges on key atoms at the TS on the NU-1000 node (light grey bars) and at the uncatalyzed TS and metal node alone (dark grey bars), from NBO analysis.

The synergistic catalytic effect observed here, which involves a positively charged species acting as a Lewis acid to facilitate the nucleophilic attack together with another species enhancing the leaving group via electrostatics, is a common motif in the catalysis of phosphate ester hydrolysis and phosphoryl transfer reactions by enzymes,[174,178,179,181] synthetic dinuclear metal complexes,[187–190] and even stand-alone metal ions in solution.[191,192] Recognizing and building upon this concept could be useful in the development of new MOFs with even faster kinetics for catalyzing phosphate ester hydrolysis.

3.3.3 Effect of Node Distortion

Recently, Chapman and coworkers observed that the Zr_6 nodes in NU-1000 undergo structural distortion under high-temperature dehydration conditions (>130 °C) and remain in the distorted form for days to weeks upon return to ambient conditions.[170] Upon distortion, the distribution of Zr–Zr bonds changes from a single distance in the originally symmetric node to two different distances in the distorted structure. In spite of this, the long-range order in the crystal was maintained, and no conventional defects (e.g., missing linkers or nodes) were formed. The

transition was confirmed by DFT calculations and ab initio molecular dynamics (AIMD) simulations in their study. In the simulant hydrolysis experiments carried out by Mondloch et al.,[1] faster reaction kinetics were observed if the NU-1000 material was thermally treated at 300 °C before running the reaction at room temperature. Therefore, we were interested in investigating the reaction mechanism on the distorted NU-1000 node to determine if the temperature-induced node distortion enhances DMNP hydrolysis compared to the regular node, as seen in the experiment.

Starting from the optimized node structure determined by DFT and AIMD simulations in ref [170], we mapped out the reaction pathway on the distorted node. As shown in Figure 3.5, the reaction pathway is similar to the regular node, except that there is no dehydration step because the node in the starting structure does not have any coordinating water molecules. Notably, one of the core oxygen atoms is separated from the distorted core and leaves a vacancy in the node (see Figure B.1 for the geometry of the simplified cluster), which further reduces the coordination on the Zr atoms at the reactive site. DMNP binds to the distorted node in a different fashion, where the nitrobenzene ring is not situated between the MOF linkers as it is in the regular node–DMNP complex (see Figure B.2 for comparison). The dominant interactions on the distorted node are van der Waals attractions between the MOF linkers and the methoxy groups of DMNP, compared to slight π–π stacking between the nitrobenzene ring and linkers on the regular node. The binding free energy of a water molecule on the distorted node is +29 kJ/mol, which suggests that the distorted node is more favorable in its completely dehydrated form and that DMNP will not have to compete with water to bind to Zr. Because the distorted node is fully dehydrated, there is no longer a hydrogen-bonding interaction between a bridging node hydroxide and the leaving group

oxygen as in the TS structure on the regular node (Figure 3.7). Nevertheless, the 40 kJ/mol free energy barrier from RS to TS is lower than the 43 kJ/mol on the regular node (Figure 3.5).

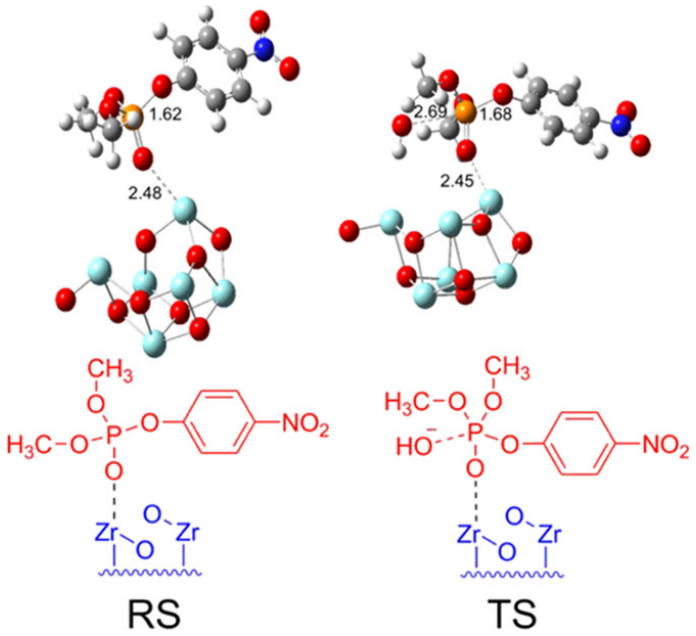


Figure 3.7. Schematic drawing (bottom) and optimized RS (node–DMNP complex) and TS structures (top) of the DMNP hydrolysis reaction catalyzed by the distorted Zr_6 node in NU-1000. Key bond lengths are expressed in Å. The organic linker molecules in NU-1000 are also included in the computational model but omitted in the images for clarity.

If we neglect the DFT-D3(BJ) dispersion correction, the barrier on the distorted node is calculated to be higher than that on the regular node (see Figure B.3 for reaction free energy profiles without dispersion included). Specifically, the inclusion of dispersion corrections results in a 37.4% reduction in barrier from RS to TS for the distorted node, compared to a 17.8% reduction for the regular node. This suggests that the distorted node has stronger dispersion interactions with the DMNP molecule, presumably because of the different binding geometry, that makes the hydrolysis reaction more favorable than on the regular node. The absence of a water-exchange step and the lower barrier on the distorted node suggest that the rate enhancement observed in the experiment upon high-temperature treatment of NU-1000 is due to both dehydration and node distortion.

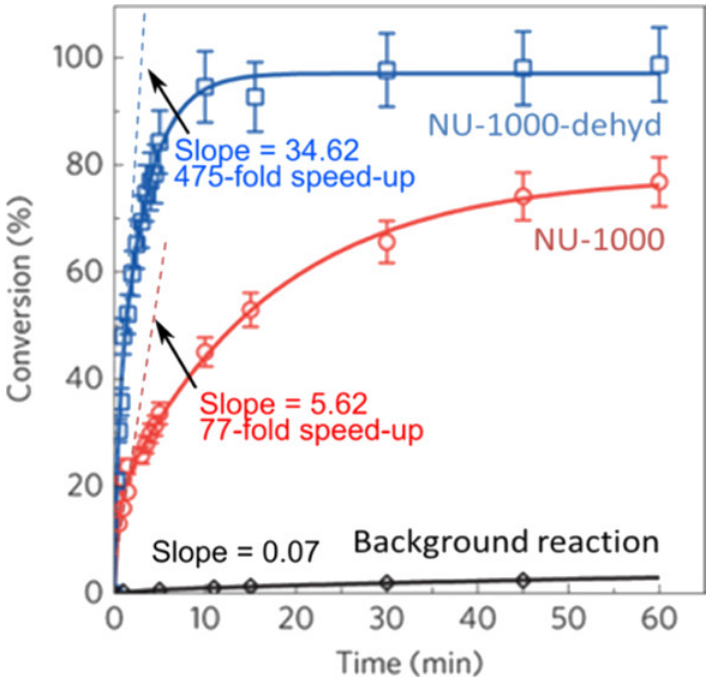


Figure 3.8. Fitting of experimental kinetics data from Mondloch et al.[1] of the catalyzed and uncatalyzed DMNP hydrolysis reactions. Adapted from ref 1 with permission from Springer Nature: Nature Materials. Copyright 2015.

By fitting the experimental kinetics data from ref 1, we extrapolated the rate enhancements of DMNP hydrolysis catalyzed by regular and "dehydrated" NU-1000 to be 77-fold and 475-fold, respectively, compared to the uncatalyzed reaction (Figure 3.8). From the calculated barrier heights (54.7 kJ/mol for uncatalyzed, 43.0 kJ/mol for regular node-catalyzed, and 40.0 kJ/mol for distorted node-catalyzed), we estimated the DFT-predicted rate enhancements using the equation $k_{\text{rel}} = e^{-\Delta\Delta G^{\ddagger}/RT}$. Here, $\Delta\Delta G^{\ddagger}$ represents the difference in the TS free energy barrier between the respective catalyzed and uncatalyzed reactions. Our results predict that the regular node and the distorted node will have a 112-fold and 382-fold rate increase over the uncatalyzed reaction, respectively, which is in good agreement with the experimental data. This suggests that the catalytic DMNP hydrolysis reaction under standard experimental conditions occurs through the regular NU-1000 node pathway, and the reaction under dehydrated conditions corresponds to the

distorted NU-1000 node pathway. Thus, these two computational node structures are reasonable models for studying catalytic reactions on NU-1000 under varying conditions.

3.4 Conclusions

In this work, we performed a mechanistic study on the hydrolysis of the CWA simulant DMNP catalyzed by the Zr-based MOF NU-1000 using DFT calculations. Computed barrier heights of the uncatalyzed and catalyzed reactions are in line with the experimental kinetics data. The origin of the catalytic effect was revealed by the analysis of the TS structures, in which a Zr atom interacts with the phosphate ester center and activates the phosphorous, whereas an OH group on the node enhances the leaving group on DMNP. This dual-stabilization scheme coincides with a common theme in enzymes that catalyze similar types of reactions. Investigation of the distorted node-catalyzed reaction suggests that dehydrated NU-1000 stays in the distorted form during catalysis, which supports previous experimental results. The absence of a water-exchange step and the lower barrier on the distorted node suggest that the rate enhancement observed in the experiment upon high-temperature treatment of NU-1000 is due to both dehydration and node distortion. The relatively small barriers for product desorption from the regular and distorted nodes suggest no product inhibition effects in the aqueous solution degradation reactions. The results and analyses presented here shed light on the mechanism of this catalytic reaction, and we hope that these insights may lead to the design of MOFs with greater catalytic activity for the destruction of dangerous organophosphate CWAs.

Chapter 4: Exploring the Effects of Node Topology, Connectivity, and Metal Identity on the Binding of Nerve Agents and Their Hydrolysis Products in Metal–Organic Frameworks

This chapter is the preliminary version of a manuscript: Mendonca, M. L.; Ray, D.; Cramer, C. J.; Snurr, R. Q. *in preparation*. Debmalya Ray (University of Minnesota) performed all periodic DFT calculations and performed cluster model DFT calculations for Zr-mono-defect UiO-66, Zr-bi(cis)-defect UiO-66, and M-MOF-808 (M = Zr, Hf, Ce, Th). I performed cluster model DFT calculations for Zr-NU-1000 (large pore), Zr-NU-1000 (c pore), and M-bi(trans)-defect UiO-66 (M = Zr, Hf, Ce, Th) and I performed all calculations required for QSAR model development and statistical testing. Additionally, I analyzed the data and wrote the manuscript.

4.1 Introduction

Despite being banned by the Chemical Weapons Convention,[3] chemical warfare agents (CWAs) have been used on both civilians and military personnel within the past few years according to various reports.[193,194] Nerve agents are an especially dangerous class of CWAs and can be divided into three main categories: G-series agents,[195] including tabun (GA), sarin (GB), and soman (GD); V-series agents,[195] such as VX; and Novichok agents,[10,196] such as A-230 and A-232 (see Figure 4.1a for structures). In the human body, organophosphate-based nerve agents inhibit acetylcholinesterase, the enzyme that controls the decomposition of the neurotransmitter acetylcholine. As a result, acetylcholine is not properly broken down, leading to continuous stimulation of muscle cells, asphyxiation, and possibly death.[197] Unfortunately, the time frame to apply effective treatment after exposure to CWAs can be as short as minutes.[14] Thus, the most

important measures to protect against nerve agents are through capture and/or degradation to nontoxic products before the chemicals reach their biological targets.

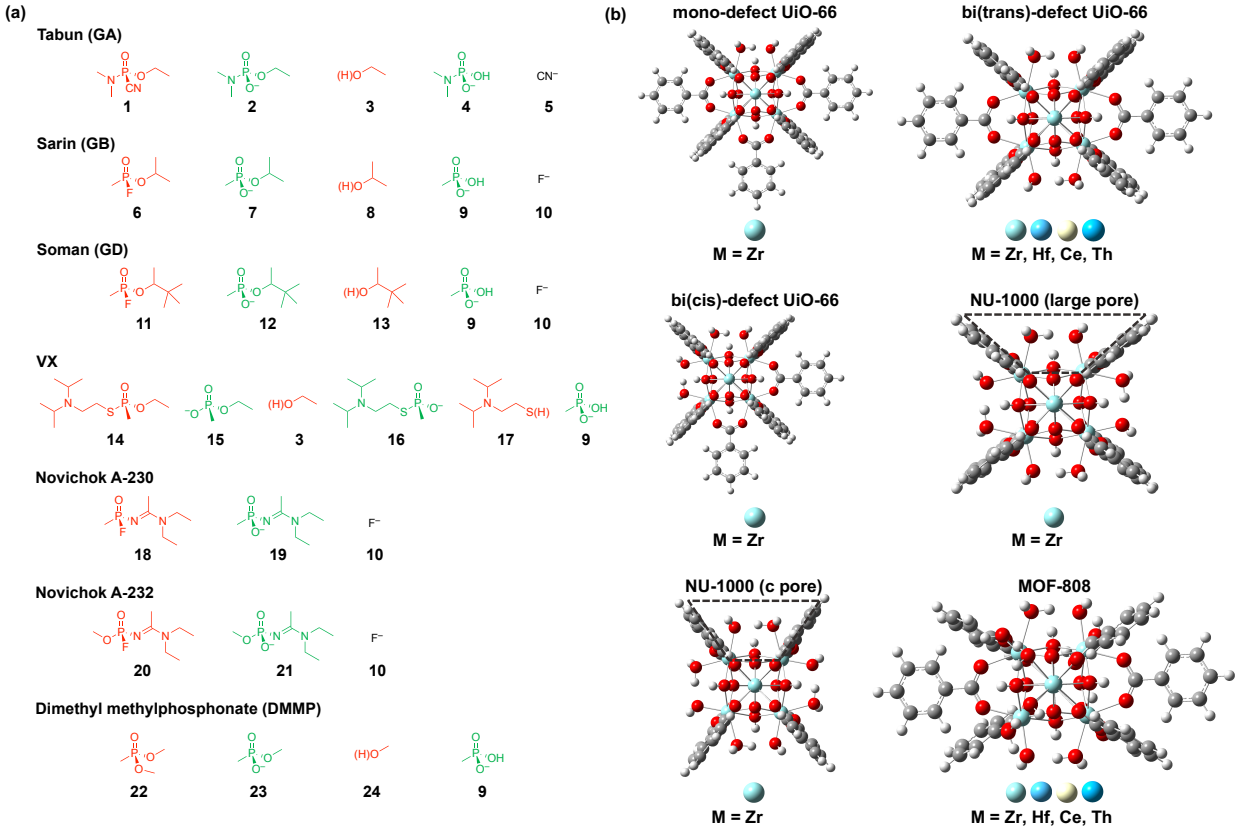


Figure 4.1. (a) Chemical structures of the 24 molecules explored as adsorbates in our study, including 6 nerve agents, a nerve agent simulant (DMMP), and their hydrolysis products. Structures in red, green, and black correspond to neutral molecules, bidentate anions, and monodentate anions, respectively. The numbering of molecules is used for simpler reference throughout the text. (b) Cluster model representations of the fully hydrated states of the 6 MOF node sites that we considered as binding sites for the 24 molecules. Dark gray, white, and red spheres represent C, H, and O atoms, respectively. The spheres below each node represent the different metals (M = Zr, Hf, Ce, Th) considered for that particular node. Dashed trapezoids are included for NU-1000 (large pore) and NU-1000 (c pore) to clarify which binding site was considered for these nodes.

In particular, metal–organic frameworks (MOFs), which are highly porous crystalline materials composed of metal nodes connected by organic linkers, have shown promise for efficient nerve agent adsorption and detoxification.[66] MOFs are attractive candidates for nerve agent degradation because of their high concentration of active sites and their large surface areas. Perhaps more importantly, MOFs also have much greater chemical tunability compared to

traditional filter materials like zeolites. Due to the substantial diversity of possible nodes, linkers, and topology combinations, MOFs can be rationally designed and continuously improved for chosen applications. The most common method of MOF-catalyzed nerve agent decomposition is through hydrolysis, with a nucleophile such as water or hydroxide attacking the agent phosphorus atom, resulting in the elimination of toxic leaving groups (see Scheme 4.1a).

Zr(IV)-based MOFs are the most investigated class of MOFs for nerve agent hydrolysis in part because of the high chemical and thermal stability afforded by the exceptionally strong Zr(IV)–O node-linker bonds. The impressive catalytic ability of Zr-MOFs can be attributed to the high concentration of strongly Lewis-acidic Zr(IV) metal centers. The connectivity of Zr_6O_8-cluster nodes (hereafter denoted as Zr_6 nodes) can be systematically tuned by using different organic linkers, thereby altering the number of terminal node water and hydroxyl groups. Previous studies have shown that lower Zr_6 node connectivity, which in turn yields larger numbers of potential binding and catalytic active sites, is directly correlated with accelerated hydrolysis rates.[22] One potential explanation is that the strength of the terminal Zr–OH_2 bonds get progressively weaker as the number of supporting carboxylate linkers decreases, and this water displacement was predicted to be the rate-limiting step for solution-phase sarin hydrolysis by several Zr-MOFs.[198] The modification of node connectivity can also yield different local steric environments, resulting in varying degrees of dispersion interactions between MOF linkers and the bound nerve agents, which may affect reaction energetics.[139] Additionally, the number and spatial orientation of missing linker defect sites, which generate catalytically active open metal sites, can strongly impact the reactivity of different MOFs for nerve agent hydrolysis.[199]

In addition to linker modification, there is potential to further improve the catalytic efficiency of Zr_6 MOFs by replacing Zr(IV) with other M(IV) cations possessing different

chemical properties (e.g., Lewis acid strength, electronegativity, etc.). For example, recent experimental and computational results showed that replacing Zr(IV) by Ce(IV) in UiO-66 leads to faster rates for the hydrolysis of nerve agents and simulants.[199,200] Moreover, there are numerous reports on the feasibility of synthesizing different isostructural series of microporous and mesoporous MOFs based on hexanuclear Zr(IV), Hf(IV), Ce(IV), and Th(IV) oxide cluster nodes, where the node metal identity is systematically changed while keeping the environment surrounding each active site constant.[170,201–203] The fact that the metal identity and connectivity of M_6 nodes can be readily tuned offers extensive opportunities to analyze structure-activity relationship trends. However, despite the steadily growing number of isostructural M(IV)-MOFs (UiO-66, MOF-808, NU-1000, NU-1008, and NU-1200 to name a few), only Ce-UiO-66 has been experimentally tested and compared to its Zr analogue for organophosphate hydrolysis to date.[200,204]

Thus, one major goal of the present work is to computationally explore the MOF topological/chemical space and to use the insights gained to guide future experimental testing. Herein, we use density functional theory (DFT) to perform a thorough analysis of the binding of multiple nerve agents and their corresponding hydrolysis products to M_6 nodes with varying metal identity and node connectivity. More explicitly, we investigate 24 molecules as adsorbates in our study, including 6 nerve agents (GA, GB, GD, VX, A-230, and A-232), a nerve agent simulant (dimethyl methylphosphonate (DMMP)), and 17 possible hydrolysis by-products including bidentate anions, monodentate anions, and neutral alcohols/thiols (Figure 4.1a). As adsorption sites for the molecules, we consider 12 different MOF node sites including Zr-mono-defect UiO-66, M-bi(trans)-defect UiO-66 (M = Zr, Hf, Ce, Th), Zr-bi(cis)-defect UiO-66, Zr-NU-1000 (large pore), Zr-NU-1000 (c pore), and M-MOF-808 (M = Zr, Hf, Ce, Th) (Figure 4.1b). For UiO-66, the

terminology “mono-defect” refers to UiO-66-11 (i.e., one missing linker) and “bi-defect” refers to UiO-66-10 (i.e., two missing linkers) where we consider both “trans” and “cis” isomers, denoting defect sites on opposite and adjacent faces of a M_6 node, respectively. We consider two distinct binding sites for NU-1000 where the open node face is either directed into the 31 Å hexagonal channel (“large pore”) or the 8 Å pore sited between nodes along the crystallographic *c* axis (“c pore”) because these different pore environments were previously shown to affect activation energies for sarin hydrolysis.[198] Further, we note that there is often a mixed valence of Ce(III) and Ce(IV) in Ce-based MOFs[204–206] but, for consistency and simpler comparison, we only model M_6 nodes where each metal has a +4 formal oxidation state.

By examining such a wide range of MOF-adsorbate combinations, we aim to gain a comprehensive perspective on how node topology, connectivity, and metal identity affect the binding energies of nerve agents and their hydrolysis products. At this stage of exploration, it is more computationally practical to calculate binding energies instead of full mechanistic pathways. Nonetheless, the displacement of a terminal node water, the subsequent adsorption of nerve agent, and the eventual desorption of product are all critical, and potentially rate-limiting, steps in hydrolytic degradation under various reaction conditions.[57,59,198,207–209] Thus, the computation of binding energies for key reaction species (i.e., water, agent, and products) may serve as a reliable method for predicting promising MOF catalysts, as suggested by the agreement between our results herein and observed experimental trends. Also, by exploring such a chemically diverse set of organophosphate molecules, we highlight the important result that no single metal or node topology is optimal for all possible nerve agents, in agreement with previous experiments that showed varying reactivity trends depending on the particular agent and MOF considered.[210]

Finally, using the large amount of data generated from this study, we also derive quantitative structure-activity relationship (QSAR) models based on intuitive molecular and node descriptors. The objective of this analysis is twofold: first, by performing simple linear and multiple linear regression on our entire dataset, we aim to understand the most important individual descriptors for describing the binding energy of both nerve agents and their hydrolysis products to M_6 nodes. Second, by splitting our data into training and test sets and performing a thorough statistical analysis, we aim to develop models capable of making quantitively accurate predictions for the binding energies of diverse organophosphate molecules in a wide variety of M(IV)-MOFs.

4.2 Computational Details

4.2.1 Periodic Calculations

Spin-polarized periodic DFT calculations were performed to optimize the unit cells of Zr-mono-defect UiO-66, M-bi(trans)-defect UiO-66 (M = Zr, Hf, Ce, Th), Zr-bi(cis)-defect UiO-66, Zr-NU-1000, and M-MOF-808 (M = Zr, Hf, Ce, Th) using the Vienna *ab initio* Simulation Package (VASP)[211–214] and projector-augmented wave potentials.[215,216] We used the PBE functional[217,218] along with Grimme's D3 dispersion correction[85] and Becke-Johnson damping.[87] A plane-wave kinetic energy cutoff of 520 eV and Γ-point sampling of the Brillouin zone was used for structural optimizations. Energy convergence criteria of 10^{-5} eV and force convergence criteria of 0.02 eV/Å were used for all periodic calculations.

4.2.2 Cluster Model Calculations

Cluster models, used to represent the M_6 nodes, were cut from the PBE-D3(BJ) optimized periodic unit cells of the various MOFs. To obtain these cluster models, the organic linkers around each metal node were truncated to benzoate groups (Figure 4.1b). We used benzoate capping groups, as opposed to formate, because this cluster model size showed improved fidelity with fully periodic

calculations in previous studies.[198,207] For the NU-1000 and MOF-808 nodes, we used the mixed-staggered proton topologies that exhibit alternating $–OH_2$ and –OH groups capping adjacent metal sites, as this was previously determined to be the most stable tautomer for these MOF nodes.[48,198]

All electronic structure calculations for the cluster models were performed using the Gaussian 09 package.[109] Geometry optimizations and frequency calculations were performed in the gas phase for all species using DFT with the M06-L functional.[75] The carbon atoms of the benzoate linker groups were held fixed to mimic the constrains imposed by the surrounding MOF structure, whereas all remaining atoms were allowed to relax. An automatic density-fitting set generated by Gaussian 09 was employed to reduce the computational cost. The def2-SVP basis set[76,219] was used for H, C, N, O, F, P, and S atoms. The SDD basis set and its associated effective core potential (ECP) was used for Zr, Hf, Ce, and Th atoms of the nodes.[79,80] The grid used for numerical integration in DFT was set to "ultrafine," i.e., a pruned grid of 99 radial shells and 590 angular points per shell.

The natures of all stationary points were determined by calculation of analytic vibrational frequencies. All minimized structures were characterized by zero imaginary frequencies. These frequencies were also used to compute molecular partition functions (at 298.15 K and 1 atm) using the conventional particle in a box, rigid rotor, and quantum mechanical harmonic oscillator approximations, except that all vibrational frequencies below 50 cm^{-1} were replaced with values of 50 cm^{-1} (the quasi-harmonic oscillator approximation).[70] Zero-point vibrational energies and thermal contributions to enthalpies and free energies were determined from these partition functions.

Electronic energies were further refined by performing single point calculations with the M06-2X meta-GGA hybrid density functional[74,110] on the gas phase optimized geometries, again

using the def2-SVP basis set for main group atoms and the SDD basis set with associated ECP for metal atoms. These single point calculations were performed using the SMD continuum solvation model[82] with parameters for water (ε = 78.355). Default convergence criteria for geometry optimizations and single point energy calculations were used. All reported free energies and enthalpies were computed by combining M06-2X(SMD) single point energies with thermochemical contributions obtained at the M06-L(gas phase) level. The energy values in the article correspond to standard-state Gibbs free energies, and self-consistent field (SCF) energies and enthalpies are provided in Appendix C along with Cartesian coordinates for all optimized structures.

4.2.3 QSAR Modeling Details

Molecular orbital parameters, used as several of the descriptors for building QSAR models, were calculated with the natural bond orbital (NBO) method[119] in Gaussian 09. All calculations required for QSAR model development and statistical testing were performed in MATLAB using built-in functions and a custom script to execute the leave-one-out cross-validation procedure.

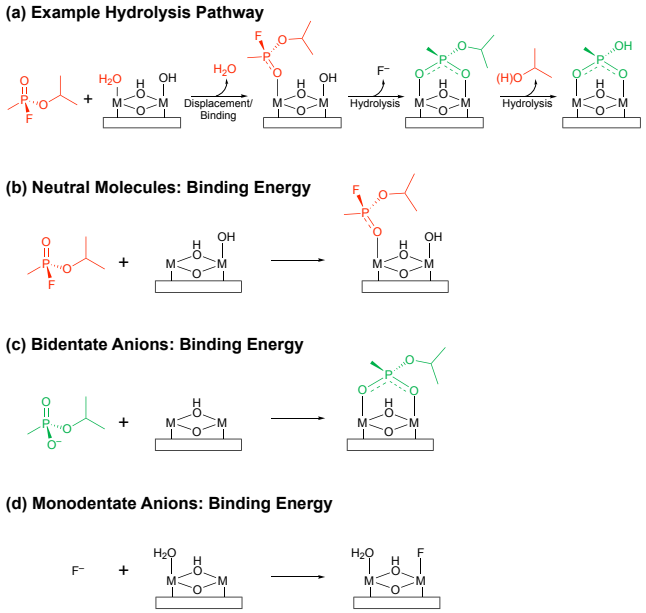
4.3 Results and Discussion

4.3.1 Rationale for Molecules Chosen for Study

To explain why we studied the binding energy of both neutral and anionic species, we first discuss a hypothetical organophosphate hydrolysis pathway on M(IV)-MOF nodes, using GB (molecule **6**) as an example (Scheme 4.1a). In their fully hydrated state (Figure 4.1b), all nodes that we considered contain at least one active (binding) site where adjacent metal atoms display capping –OH_2 and –OH groups. For catalytic hydrolysis to occur at an active site, the capping –OH_2 group must first be displaced to generate an open metal site where the nerve agent GB can bind. After binding to the open M atom, nucleophilic attack occurs on the P atom of GB, followed by the

elimination of a fluoride anion (**10**) into solution. We note that Scheme 4.1a depicts a simplified hydrolysis pathway, where each reaction arrow titled "hydrolysis" actually denotes several elementary steps (e.g., nucleophilic attack and fluoride elimination). Further, for the purpose of exploring the effect of node properties on product inhibition (explained below), we only consider the pathway in which the terminal –OH group capping the adjacent M atom acts as the nucleophile because this pathway results in hydrolysis products bound in a bidentate mode.[57] The product of this initial hydrolysis reaction is isopropyl methylphosphonic acid (IMPA, **7**), present in its anionic form and bound in a bidentate fashion to the node. In the presence of water and a base (required to generate hydroxide anions), IMPA could then either be displaced and desorb from the active site, remain bound to the node and cause product inhibition, or remain bound and undergo further hydrolysis. A subsequent hydrolysis reaction would result in the elimination of isopropanol (**8**) into solution, as all ionizing groups (alcohols and thiols) were treated as neutral molecules in our study. The product of this secondary hydrolysis is methylphosphonic acid (MPA, **9**), again modeled as a bidentate-bound anion. Finally, with the assistance of a hydroxide anion, MPA could then either be displaced and desorb from the active site or remain bound to the node, resulting in product inhibition.

Scheme 4.1. (a) Hypothetical Hydrolysis Pathway for GB (Molecule 6) on M(IV)-MOF Nodes and Chemical Equation Used to Calculate Binding Energies for (b) Neutral Molecules, (c) Bidentate Anions, and (d) Monodentate Anions



Overall, the hypothetical pathway depicted in Scheme 4.1a represents a worst-case scenario for nerve agent hydrolysis occurring on M_6 nodes. More specifically, this particular pathway involves both water displacement by a nerve agent, which corresponds to a large free energy barrier,[198,199,207] and hydrolysis products bound in a bidentate mode, which corresponds to very strong binding energies and may cause product inhibition.[57,208] We considered this hypothetical, extreme case so that we could analyze the effect of node properties on water displacement and product inhibition simultaneously, as both may be significant under varying reaction conditions. For example, water displacement may be the rate-limiting step for solution-phase reactions where nodes are fully hydrated, whereas product desorption/inhibition is more important for reactions occurring without a base to regenerate node active sites.[46,58,59] From an experimental design perspective, it is important to understand these limiting scenarios that may reduce a MOF's catalytic activity.

The hypothetical hydrolysis pathways for all nerve agent and simulant molecules we studied are given in Figure C.1. For the G-series agents (**1**, **6**, and **11**), we considered the most likely product molecules based on selectivities observed in previous MOF-catalyzed nerve agent hydrolysis experiments and computational mechanistic studies. For example, previous results indicate that Zr(IV)-MOFs selectively hydrolyze the P–F bond of GD (**11**),[1,52,210] eliminating fluoride and generating node-bound pinacolyl methylphosphonic acid (PMPA, **12**). Under our assumed framework, further hydrolysis would result in the cleavage of the P–O bond of PMPA (as P–C bonds are fairly inert and do not break readily), eliminating pinacolyl alcohol (**13**) and forming node-bound MPA (**9**). The common product molecules for DMMP (**22**) are similarly well-studied[58,59] and typically include methanol (**24**), node-bound methyl methylphosphonic acid (MMPA, **23**), and node-bound MPA. Conversely, the degradation of VX (**14**) is slightly more complex because it contains two bonds which can potentially be cleaved during initial hydrolysis. If its P–O bond is broken, ethanol (**3**) is eliminated and the highly toxic byproduct EA-2192 (**16**) is formed. Thus, it is preferable to selectively cleave the P–S bond of VX to generate the nontoxic products 2-(diisopropylamino)ethanethiol (DESH, **17**) and ethyl methylphosphonic acid (EMPA, **15**).[21,52] Then, depending on the hydrolysis conditions, both EA-2192 and EMPA could be further hydrolyzed to form node-bound MPA. Finally, arguably much less is known about the hydrolysis mechanisms of the Novichok agents A-230 (**18**) and A-232 (**20**), so we only considered initial hydrolysis of their most labile P–F bonds[220] to eliminate fluoride and form node-bound bidentate anions **19** and **21**, respectively.

4.3.2 Binding Energy Formulas

Now, we discuss the methods used to calculate binding energies throughout the study. For all nerve agent and simulant molecules (**1**, **6**, **11**, **14**, **18**, **20**, and **22**), we considered the binding geometry

in which the organophosphate binds to the node open metal site through a M–O(=P) bond, as this is known to be the most favorable binding mode.[57,209,221] The most favorable binding mode for all alcohol (**3**, **8**, **13**, and **24**) and thiol (**17**) product molecules was calculated to be through the M–OH(R) and M–SH(R) bonds, respectively. In their optimized node-bound structures, the hydrogen atom of each alcohol OH group (or thiol SH group) is effectively shared through hydrogen-bonding with the neighboring terminal node hydroxyl group. The binding free energies (ΔG_{bind}) for these neutral molecules were calculated via

$$\Delta G_{bind} = G_{node+mol} - G_{node-noOH_2} - G_{mol} \quad (4.1)$$

where the subscript *node+mol* denotes the molecule bound to the node, *node–noOH₂* denotes the bare node where the terminal –OH_2 group has been removed from the binding site, and *mol* denotes the individual molecule. Scheme 4.1b shows the chemical equation representation for eq 4.1.

All bidentate anion product molecules (**2**, **4**, **7**, **9**, **12**, **15**, **16**, **19**, **21**, and **23**) were considered to bind to the node in a bridging fashion through a M–O–P–O–M plane to two adjacent M atoms at the binding site. The ΔG_{bind} for bidentate anions were calculated via

$$\Delta G_{bind} = G_{node+mol} - G_{node-noOH_2,OH} - G_{mol} \quad (4.2)$$

where the subscript *node–noOH₂,OH* denotes the bare node where the terminal –OH_2 and –OH groups have been removed from the binding site. Scheme 4.1c shows the chemical equation representation for eq 4.2.

As their name suggests, the binding mode for the two monodentate anion product molecules (**5** and **10**) involved coordination to a single node M atom, where both anions exhibited hydrogen-bonding interactions with the neighboring terminal node water. The most favorable

binding mode for cyanide (**5**) was calculated to be through a M–C(≡N) bond, whereas fluoride (**10**) binds through a M–F bond. The ΔG_{bind} for monodentate anions were calculated via

$$\Delta G_{bind} = G_{node+mol} - G_{node-noOH} - G_{mol} \quad (4.3)$$

where the subscript *node–noOH* denotes the bare node where the terminal –OH group has been removed from the binding site. Scheme 4.1d shows the chemical equation representation for eq 4.3.

Optimized cluster models for the *node–noOH*$_2$, *node–noOH*, and *node–noOH*$_{2,}$*OH* bare node structures are shown in Figure C.2a–c, using the node of Zr-NU-1000 (large pore) as an example. Optimized cluster models that are representative of the general binding modes for node-bound nerve agent and simulant molecules, alcohol and thiol product molecules, bidentate anion products, and monodentate anion products are shown in Figure C.3a–d, again for Zr-NU-1000 (large pore).

4.3.3 Effects of Node Topology and Connectivity on Binding Energies for Zr(IV)-MOFs

First, we discuss the trends observed in binding free energies across the 6 different Zr(IV)-MOF node sites (Figure 4.2). The ΔG_{bind} values for water (Figure 4.2a) are calculated as –7.9, –28.1, –29.9, and –40.9 kJ/mol for MOF-808, NU-1000 (large pore), bi(trans)-defect UiO-66, and mono-defect UiO-66, respectively, where the number of linkers per node is 6, 8, 10, and 11, respectively. This trend agrees with previous results showing that the strength of terminal Zr–OH$_2$ bonds get progressively stronger as the number of supporting carboxylate linkers increases.[198] However, bi(cis)-defect UiO-66 (ΔG_{bind} = –54.5 kJ/mol) and NU-1000 (c pore) (ΔG_{bind} = –61.1 kJ/mol) notably deviate from this trend, implying that the local binding environment is also important. For bi-defect UiO-66, this indicates that the relative orientation of the two defect sites has a large impact on the node chemistry.[199] Similarly, it appears that the pore environment (large vs. c pore)

which the open node face of NU-1000 is directed towards is also influential. A more in-depth discussion on the variation in properties (e.g., binding site charges, bond lengths, etc.) across the different node sites is provided later in the QSAR modeling sections.

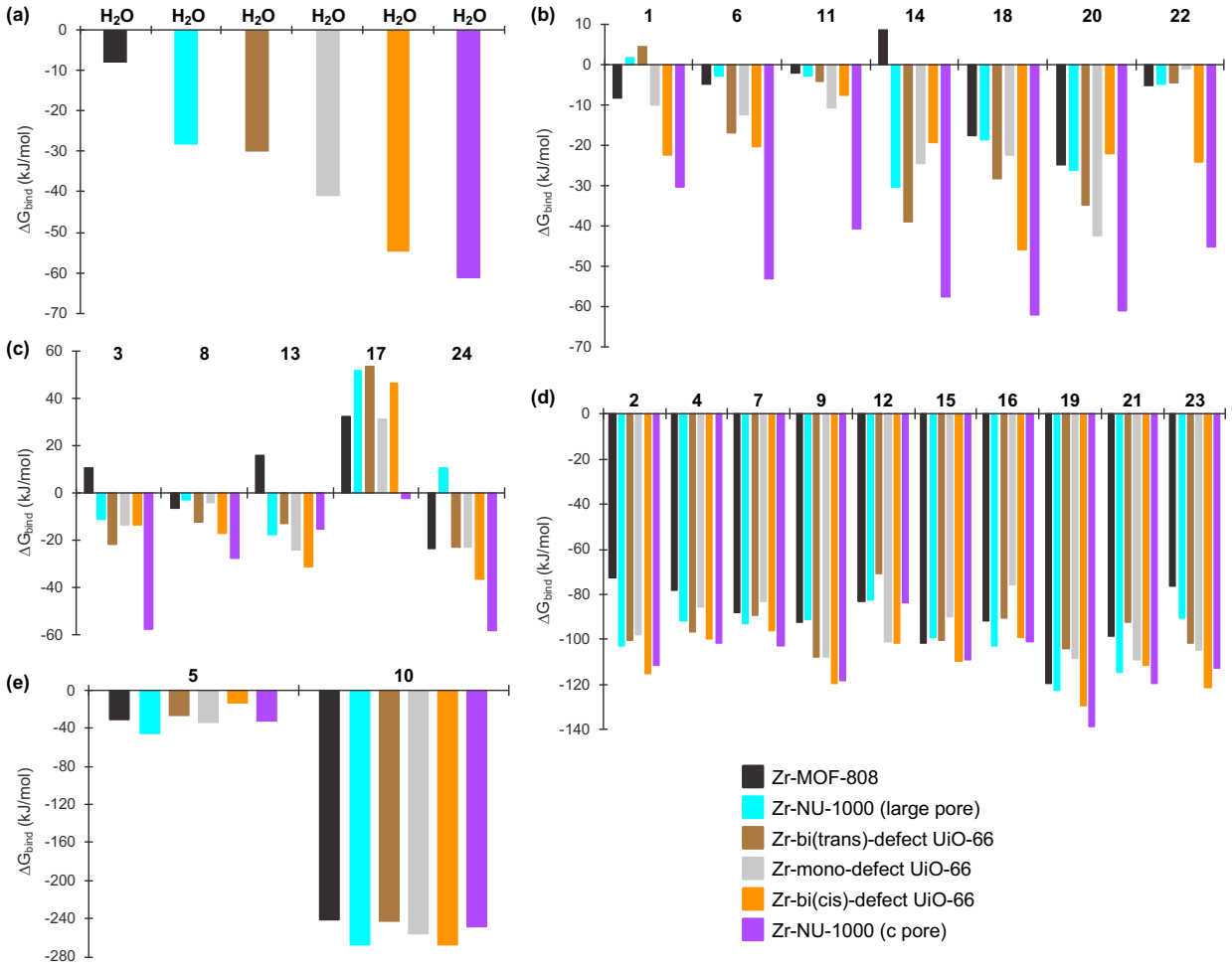


Figure 4.2. Binding free energies for (a) water, (b) nerve agent and simulant molecules, (c) alcohol and thiol hydrolysis product molecules, (d) bidentate anion products, and (e) monodentate anion products bound to 6 different Zr(IV)-MOF node sites.

As stated above, we modeled all nerve agent and simulant molecules as binding to node open metal sites through a M–O(=P) bond. In this binding geometry, we considered three possible orientations for each nerve agent (**1**, **6**, **11**, **14**, **18**, and **20**) and two orientations for DMMP (**22**), adopting the nomenclature introduced by Troya.[57] Briefly, each orientation is named according to the molecule's R group (e.g., F, CH_3, or OiPr for GB) that is approximately collinear with the

neighboring terminal node Zr–OH group in the optimized cluster model. An example is shown in Figure C.4 for the three orientations of GB bound to Zr-NU-1000 (large pore) along with their relative binding free energies. We performed this orientational analysis for every nerve agent and simulant molecule for all 6 Zr(IV)-MOF node sites and the results are listed in Table C.1. Notably, we observed that the most favorable molecular orientations vary across the different node sites, most likely due to slight changes in molecule-linker interactions and hydrogen-bonding with node oxo/hydroxyl groups.[209] This observation has important implications because the molecular orientation may strongly affect hydrolysis reaction barriers.[57] Therefore, we recommend a similar orientational analysis for all computational mechanistic studies. Only the ΔG_{bind} values for the most favorable molecular orientations on each node are shown in Figure 4.2b.

The general trend in binding free energies for the nerve agent and simulant molecules (Figure 4.2b) across the Zr(IV)-MOF node sites is MOF-808 < NU-1000 (large pore) < bi(trans)-defect UiO-66 ≈ mono-defect UiO-66 < bi(cis)-defect UiO-66 < NU-1000 (c pore). When examining these molecules as a group, it is challenging to go beyond a qualitative analysis because the trends change markedly for each molecule. The only constant across all 7 molecules is that NU-1000 (c pore) consistently displays the strongest binding free energies, which could be attributed to the relatively small pore size (8 Å) of this binding site resulting in closer proximity and increased interactions between the bound molecules and the benzoate linkers. In terms of the molecules themselves, the larger molecules (**14**, **18**, and **20**) generally show the strongest binding free energies (except for VX (**14**) bound to MOF-808), which may be caused by increased dispersion interactions with the linkers surrounding the binding sites.[139] To explore this hypothesis, we computed dispersion energies for the node-bound molecules using the DFT-D3 correction (with zero damping and parameters for the M06-2X functional)[85] and we computed solvent-

accessible surface areas (SA_{mol}), using VMD 1.9.3 (with default probe radius),[222] to quantify the size of individual molecules. As shown in Figure C.5a, there is a strong linear relationship between SA_{mol} and the computed dispersion energies ($R^2 \geq 0.97$ for each of the 6 Zr(IV)-MOF node sites). Further, the node connectivity has a clear influence on dispersion energies, where every molecule follows the trend MOF-808 < NU-1000 (large pore) ≈ NU-1000 (c pore) < bi(trans)-defect UiO-66 ≈ bi(cis)-defect UiO-66 < mono-defect UiO-66. However, when analyzing the effect of molecular size on the binding free energies, we found no discernible relationship for any of the nodes ($R^2 \approx 0$, Figure C.5b), indicating that the trends observed in the ΔG_{bind} values for the nerve agents are fairly complex and cannot be fully explained by dispersion interactions alone.

The general trend in binding free energies for the alcohols (Figure 4.2c) is essentially the same as that for the nerve agent and simulant molecules. In Figure 4.2c, the only thiol (DESH, **17**) is a clear outlier and exhibits unfavorable (positive) ΔG_{bind} values for every Zr(IV)-MOF node site except NU-1000 (c pore), which can be explained by the large Zr–SH(R) bond distance (≥ 2.87 Å) for all nodes, indicating weak physisorption. Additionally, both ethanol (**3**) and pinacolyl alcohol (**13**) bind unfavorably to MOF-808 and methanol (**24**) binds unfavorably to NU-1000 (large pore). The relative binding strengths of the considered alcohol products is important to take into account for the possibility of active site poisoning following organophosphate hydrolysis and may also be relevant in the context of MOF-based organophosphate degradation in non-aqueous environments (e.g., methanol and isopropanol (**8**)).[223,224]

Similar to the nerve agent and simulant molecules, we considered two possible binding orientations for each bidentate anion product. Here, each orientation is named according to the anion's R group that is directed towards the bridging node hydroxyl group at the active (binding) site. For example, the two orientations modeled for IMPA (**7**) are termed "CH_3–μ_3OH" and "OiPr–

μ_3OH," which are depicted in Figure C.6 for IMPA bound to Zr-NU-1000 (large pore). We performed this orientational analysis for every bidentate anion for all Zr(IV)-MOF node sites and the results are listed in Table C.2. As before, the most favorable orientations vary across the different node sites, most likely due to varying hydrogen-bonding interactions with the node μ_3OH group. This observation has implications for product inhibition because which orientation is preferred may affect the energy required to desorb these bidentate hydrolysis products from each node. Again, only the ΔG_{bind} values for the most favorable orientations on each node are shown in Figure 4.2d. Unlike the neutral molecules, it is more difficult to identify any clear trends for the bidentate anions aside from the observations that MOF-808 generally displays the weakest binding free energies and bi(cis)-defect UiO-66 and NU-1000 (c pore) display the strongest. The more important conclusion is that all of the bidentate anions bind strongly to each node (ΔG_{bind} < –70 kJ/mol), suggesting that product inhibition is likely to occur in the absence of a strong base to ensure a high concentration of hydroxide ions necessary to displace the products.[46] As before, the node connectivities and the bidentate anion sizes have strong influences on the node-bound dispersion energies but there is no correlation between these parameters and the binding free energies (Figure C.7).

Finally, the binding free energies for the two monodentate anion products are shown in Figure 4.2e. Cyanide (**5**) shows relatively modest ΔG_{bind} values. Conversely, fluoride (**10**) binds very strongly to every node site (ΔG_{bind} < –240 kJ/mol) due to the strong affinity of fluoride towards Zr(IV), which may lead to complete degradation of the MOF structures. Indeed, to better analyze solid-phase decontamination rates, Wang et al. used a HF digestion medium to displace unreacted nerve agents and their tightly bound hydrolysis products from Zr(IV)-MOF nodes, destroying the secondary building units themselves in the process.[56] Once again, this validates the

importance of using a base (either basic solutions or solid-phase bases) for MOF-catalyzed nerve agent hydrolysis, as the base can neutralize acidic reaction byproducts.[46] All binding free energy values used to make Figure 4.2 are listed in Table C.3.

The organophosphate molecules we studied vary widely in size and cleavable ester bond, among other properties, making it difficult to analyze our dataset as a whole. Moreover, no single MOF is likely to be the most active for hydrolyzing every agent, so it is more useful to consider the molecules individually. To reiterate, the displacement of a terminal water from the node active site, the subsequent adsorption of nerve agent, and the ultimate desorption of bidentate-bound products are all critical steps in hydrolytic degradation under various reaction conditions.[57,59,198,207–209] Thus, to help identify the optimal Zr(IV)-MOF node sites for each nerve agent and simulant molecule, we now focus the analysis on the binding energies for the key reaction species. Considering the hydrolysis pathways as a whole (Figure C.1), the optimal nodes for each agent will exhibit the weakest binding for water, the strongest binding for the nerve agent or simulant molecule, and the weakest binding for the corresponding bidentate anion products.

For each nerve agent (GA, GB, GD, VX, A-230, and A-232) and the simulant DMMP, we used the following procedure to rank the 6 Zr(IV)-MOF nodes sites. For each node, we separately tabulated the binding free energies for water and the agent ($\Delta G_{bind,water}$ and $\Delta G_{bind,agent}$, respectively). Additionally, of the considered bidentate anion products for each agent, we tabulated the binding free energy of the strongest-bound bidentate anion ($\Delta G_{bind,product}$) to each node. We used the minimum possible value (most negative) of $\Delta G_{bind,product}$ for each agent because this represents the most thermodynamically stable product, which would contribute the most to product inhibition. Then, we scaled each of the three ΔG_{bind} quantities from 0 to 1 using the formula

$$\Delta G_{ij}^{s} = \frac{\Delta G_{ij} - \Delta G_{i,min}}{\Delta G_{i,max} - \Delta G_{i,min}} \quad (4.4)$$

where ΔG_{ij} and ΔG_{ij}^{s} denote the unscaled and scaled binding free energy values for molecule *i* (water, agent, or product) bound to node *j*, respectively, and the subscripts *min* and *max* denote the minimum (most negative) and maximum (least negative) binding free energies for molecule *i* based on all 6 Zr(IV)-MOF nodes sites, respectively. According to the arguments above, we aim to maximize $\Delta G_{water,j}^{s}$, minimize $\Delta G_{agent,j}^{s}$, and maximize $\Delta G_{product,j}^{s}$. Thus, for each agent-node combination, we computed the following optimization metric:

$$\Delta G_{bind}^{opt} = \Delta G_{water,j}^{s} + \left(1 - \Delta G_{agent,j}^{s}\right) + \Delta G_{product,j}^{s} \quad (4.5)$$

where eq 4.5 assumes that each term contributes equally in terms of importance to the overall hydrolysis pathway. Finally, we ranked the ΔG_{bind}^{opt} values from maximum (best) to minimum (worst) to determine the optimal Zr(IV)-MOF node sites for hydrolyzing each agent. The results of this analysis are shown in Table 4.1.

Table 4.1. Ranking of Optimal Zr(IV)-MOF Node Sites for Hydrolyzing Nerve Agent and Simulant Molecules

Rank	GA	GB	GD	VX	A-230	A-232	DMMP
1	MOF-808	MOF-808	MOF-808	NU-1000 (large pore)	bi(trans)-defect UiO-66	bi(trans)-defect UiO-66	MOF-808
2	mono-defect UiO-66	NU-1000 (large pore)	NU-1000 (large pore)	MOF-808	MOF-808	MOF-808	NU-1000 (large pore)
3	NU-1000 (c pore)	bi(trans)-defect UiO-66	bi(trans)-defect UiO-66	bi(trans)-defect UiO-66	mono-defect UiO-66	mono-defect UiO-66	bi(trans)-defect UiO-66
4	NU-1000 (large pore)	NU-1000 (c pore)	NU-1000 (c pore)	mono-defect UiO-66	NU-1000 (large pore)	NU-1000 (c pore)	NU-1000 (c pore)
5	bi(trans)-defect UiO-66	mono-defect UiO-66	mono-defect UiO-66	NU-1000 (c pore)	bi(cis)-defect UiO-66	NU-1000 (large pore)	mono-defect UiO-66
6	bi(cis)-defect UiO-66	bi(cis)-defect UiO-66	bi(cis)-defect UiO-66	bi(cis)-defect UiO-66	NU-1000 (c pore)	bi(cis)-defect UiO-66	bi(cis)-defect UiO-66

Overall, we note that no single Zr(IV)-MOF node topology or connectivity is predicted to be optimal for hydrolyzing all considered organophosphate molecules, so we advocate that MOFs should be chosen on an agent-by-agent basis. However, with regard to bi-defect UiO-66, our results predict that the trans isomer should outperform the cis isomer for all considered agents, which may inform defect engineering opportunities.[225–227] Although the optimization metric we devised to rank the nodes cannot be used to quantitatively compare catalyst efficiencies, the results in Table 4.1 are in qualitative agreement with observed experimental trends that suggest MOF-808 to be the most active Zr(IV)-MOF catalyst for organophosphate hydrolysis, followed by NU-1000 and UiO-66,[1,25,66,141,199] if we consider our composite results for the large and c pores of NU-1000 as well as both mono-defect and bi-defect UiO-66. Thus, the simple calculation of binding energies for key reaction species (i.e., water, agent, and bidentate products) may serve as an adequate

substitute for full mechanistic studies as the initial computational step used to predict promising Zr_6 node sites for nerve agent hydrolysis.

4.3.4 Effect of Node Metal Identity on Binding Energies for M(IV)-MOFs

To model the binding for all molecules to the nodes of M-bi(trans)-defect UiO-66 and M-MOF-808 (M = Zr, Hf, Ce, Th), we used the most favorable orientations identified for each molecule on the Zr analogues (provided in Table C.1 and C.2) and reoptimized for the other three metals, as the metal identity is not likely to affect which orientations are most favorable. The binding free energies for M-bi(trans)-defect UiO-66 and M-MOF-808 are shown in Figure 4.3 and Figure C.8, respectively. The ΔG_{bind} values for water are calculated as –29.9, –35.4, –32.0, and –38.2 kJ/mol for Zr-, Hf-, Ce-, and Th-bi(trans)-defect UiO-66, respectively (Figure 4.3a). Previous experimental data suggests that the electronegativity of the metal in isostructural M(IV)-MOF nodes may influence catalytic performance.[203] To explore this, we calculated the electronegativity of the four M^{4+} cations according to the Mulliken scale:

$$\chi_{M^{4+},node} = \frac{IP+EA}{2} \tag{4.6}$$

where the values for the ionization potential (IP) for M^{4+} to M^{5+} and the electron affinity (EA) for M^{4+} to M^{3+} were obtained from NIST.[228] A plot of the relationship between $\chi_{M^{4+},node}$ and the binding free energy for water to M-bi(trans)-defect UiO-66 nodes (Figure C.9) shows a linear correlation (R^2 = 0.87) where metals with a higher electronegativity display weaker binding for water, which is important since water displacement from the node is a key hydrolysis step. The ΔG_{bind} values for water are calculated as –7.9, –14.8, –28.8, and –21.6 kJ/mol for Zr-, Hf-, Ce-, and Th-MOF-808, respectively (Figure C.8a). Surprisingly, a plot of the relationship between $\chi_{M^{4+},node}$ and the binding free energy for water to M-MOF-808 nodes (Figure C.10) does not

show a linear correlation (R^2 = 0.33), mainly due to Ce deviating from the linear trend, which suggests that the electronegativity of the metal cations in MOF-808 nodes may have a weaker influence on catalytic hydrolysis efficiency.

The general trend in binding free energies for the nerve agent and simulant molecules across the M-bi(trans)-defect UiO-66 nodes (Figure 4.3b) is Zr < Hf ≈ Th < Ce. As was the case for the different Zr(IV)-MOF node sites, the ordering of the four metals changes from molecule to molecule, making it difficult to analyze the molecules all together. Nonetheless, Ce-bi(trans)-defect UiO-66 consistently displays the strongest binding free energies. The general trend in binding free energies for the nerve agent and simulant molecules across the M-MOF-808 nodes (Figure C.8b) is Hf < Zr < Th < Ce, where Hf and Ce consistently display the weakest and strongest binding free energies, respectively. The main differences between the two node sites are that the ordering of the four metals is slightly different and the M-MOF-808 nodes show generally weaker binding free energies compared to their M-bi(trans)-defect UiO-66 analogues. Additionally, Zr- and Hf-MOF-808 exhibit unfavorable binding for VX (**14**) whereas Zr- and Hf-bi(trans)-defect UiO-66 exhibit strongly favorable binding (ΔG_{bind} ≈ –40 kJ/mol) for VX. Clearly, the effect of the node metal identity on the binding energies changes depending on the node topology/connectivity.

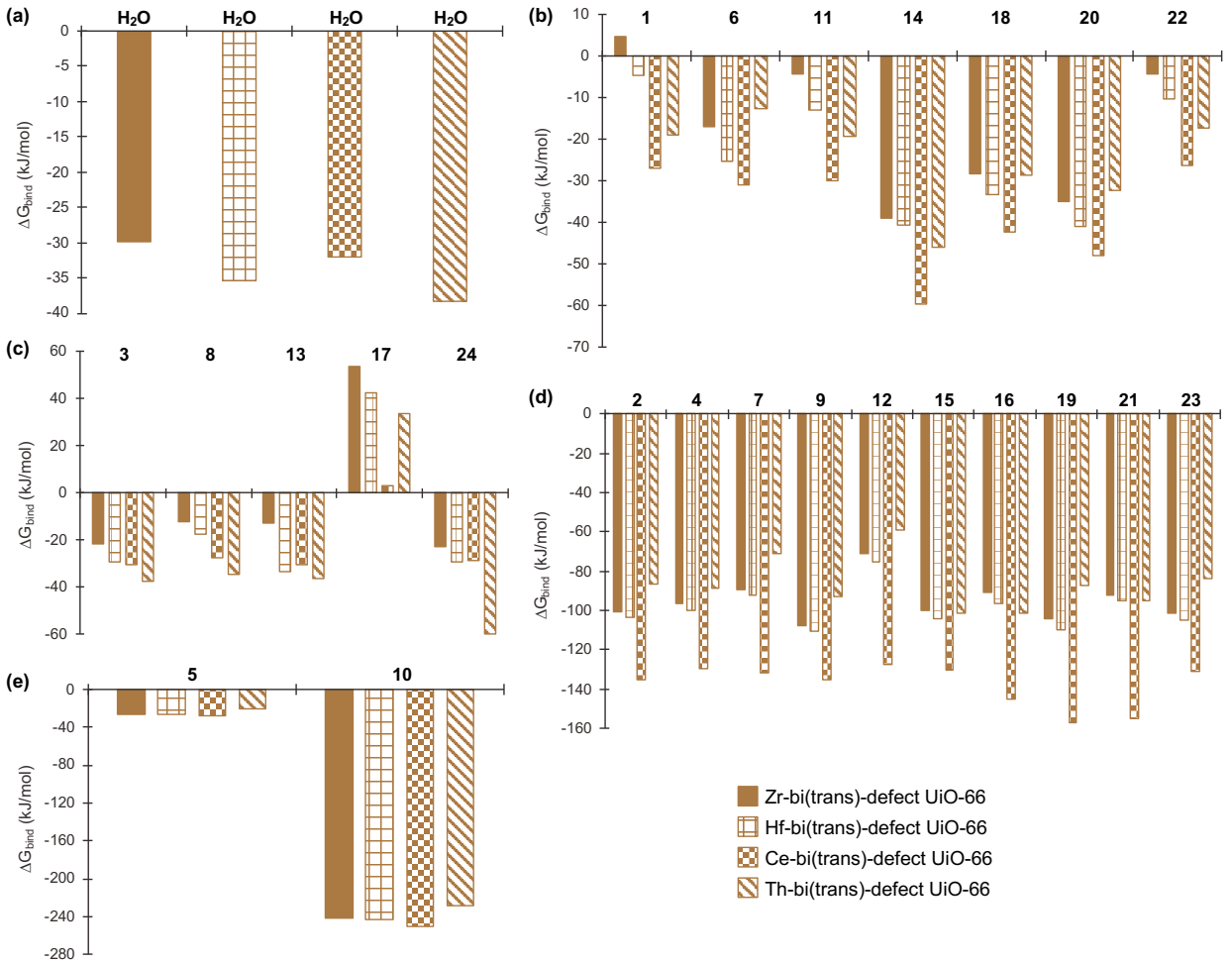


Figure 4.3. Binding free energies for (a) water, (b) nerve agent and simulant molecules, (c) alcohol and thiol hydrolysis product molecules, (d) bidentate anion products, and (e) monodentate anion products bound to the nodes of M-bi(trans)-defect UiO-66 (M = Zr, Hf, Ce, Th).

The general trend in binding free energies for the alcohols and thiols across the M-bi(trans)-defect UiO-66 nodes (Figure 4.3c) is Zr < Hf < Ce < Th and the trend across the M-MOF-808 nodes (Figure C.8c) is Hf < Zr < Th ≈ Ce. The thiol DESH (**17**) is the only neutral product molecule that displays unfavorable binding free energies to M-bi(trans)-defect UiO-66 nodes, similar to the observations made for the different Zr(IV)-MOF node sites. Conversely, ethanol (**3**), pinacolyl alcohol (**13**), and DESH all display unfavorable binding free energies to both Zr- and Hf-MOF-808 nodes, again implying a varying effect of the metal identity across different node sites.

The general trend in binding free energies for the bidentate anions across the M-bi(trans)-defect UiO-66 nodes (Figure 4.3d) is Th < Zr ≈ Hf < Ce and the trend across the M-MOF-808 nodes (Figure C.8d) is Zr ≈ Hf < Th < Ce, where Ce consistently displays the strongest binding for both node sites. Perhaps the most interesting observation here is that Th-bi(trans)-defect UiO-66 displays weaker binding free energies than the other three metals for the majority of the bidentate anions, while Th generally displays relatively strong binding for water and the nerve agents/simulant compared to the other metals. This highlights the importance of the holistic approach we used to analyze organophosphate hydrolysis pathways because considering the relative binding strengths of key reaction species from different steps in the hydrolysis mechanisms may reveal tradeoffs inherent to a particular node metal or topology (e.g., Th-bi(trans)-defect UiO-66 nodes may experience less product inhibition due to bidentate-bound products, but these nodes would require more energy for water displacement). Aside from this unique case, Hf- and Ce-bi(trans)-defect UiO-66 nodes as well as Hf-, Ce-, and Th-MOF-808 nodes all exhibit equally strong or stronger binding free energies for bidentate anions than their Zr analogues, suggesting that product inhibition may be a more general problem for most metals in the absence of a base.[208]

The binding free energies for the two monodentate anion products are shown in Figure 4.3e for M-bi(trans)-defect UiO-66 and in Figure C.8e for M-MOF-808. Similar to the different Zr(IV)-MOF node sites, cyanide (**5**) binds rather weakly to Hf, Ce, and Th, regardless of the node site. However, fluoride (**10**) again binds very strongly to both Hf and Ce nodes, presenting a potential problem for degradation of the secondary building units in the absence of a neutralizing base. In contrast, both Th node sites exhibit weaker fluoride binding than Zr and may be slightly less susceptible to post-hydrolysis node degradation. All binding free energy values for M-bi(trans)-

defect UiO-66 (M-MOF-808) nodes used to make Figure 4.3 (Figure C.8) are listed in Table C.4 (Table C.5).

Finally, to identify the optimal M-bi(trans)-defect UiO-66 and M-MOF-808 node sites for each nerve agent and simulant molecule, we used an identical procedure to rank the different metals as we used to rank the Zr(IV)-MOF nodes sites. Namely, we tabulated the binding free energies for water, agents, and the strongest-bound bidentate anion products for the M-bi(trans)-defect UiO-66 and M-MOF-808 node sites, scaled the three ΔG_{bind} quantities from 0 to 1 using eq 4.4 (where scaling was done separately for the four bi(trans)-defect UiO-66 metals and the four MOF-808 metals), and then computed the optimization metric (ΔG_{bind}^{opt}) for each agent-node combination using eq 4.5. As before, the ΔG_{bind}^{opt} values were ranked from maximum (best) to minimum (worst) to determine the optimal M-bi(trans)-defect UiO-66 and M-MOF-808 node sites for hydrolyzing each agent. The results of these analyses are shown in Table 4.2 and 4.3 for M-bi(trans)-defect UiO-66 and M-MOF-808, respectively.

Table 4.2. Ranking of Optimal M-bi(trans)-defect UiO-66 Node Sites for Hydrolyzing Nerve Agent and Simulant Molecules

Rank	GA	GB	GD	VX	A-230	A-232	DMMP
1	Ce	Zr	Ce	Zr	Zr	Zr	Ce
2	Zr	Ce	Zr	Ce	Ce	Hf	Zr
3	Th	Hf	Th	Th	Hf	Ce	Th
4	Hf	Th	Hf	Hf	Th	Th	Hf

For the bi(trans)-defect UiO-66 node topology, we note that no single metal is predicted to be optimal for hydrolyzing all considered organophosphate molecules, again supporting the notion that MOFs should only be selected based on the agent of interest. For example, our optimization metric predicts Ce-bi(trans)-defect UiO-66 to be the most active for hydrolyzing GA, GD, and DMMP and Zr-bi(trans)-defect UiO-66 to be the most active for hydrolyzing GB, VX, A-230, and

A-232. Although there is limited experimental data to compare with these computational predictions, our results are in qualitative agreement with experiments showing Ce-UiO-66 to be more active than Zr-UiO-66 for GD hydrolysis.[200] This agreement, combined with the optimization metric's correct prediction of Zr(IV)-MOF activities (MOF-808 > NU-1000 > UiO-66), suggests that our rather simple binding-energy-based approach may be a reasonable method for predicting promising M_6 node sites for nerve agent hydrolysis. Further, the results in Table 4.2 indicate that Hf- and Th-bi(trans)-defect UiO-66 nodes may be less than ideal catalysts, which, at the very least, could be used for prioritizing future experimental synthesis and testing.

Table 4.3. Ranking of Optimal M-MOF-808 Node Sites for Hydrolyzing Nerve Agent and Simulant Molecules

Rank	GA	GB	GD	VX	A-230	A-232	DMMP
1	Zr	Zr	Zr	Zr	Zr	Zr	Zr
2	Hf	Hf	Hf	Th	Th	Hf	Hf
3	Th	Ce	Th	Hf	Hf	Th	Th
4	Ce	Th	Ce	Ce	Ce	Ce	Ce

The rankings in Table 4.3 for the MOF-808 node topology are fairly surprising, where Zr is predicted to be the most active metal for hydrolyzing all 7 organophosphate molecules while Ce is predicted to be the "worst" metal for all agents, with the exception of GB. Notably, these predictions disagree with previous computational results that predicted Ce-MOF-808 to be more active than its Zr analogue for GB hydrolysis based on a mechanistic study.[199] There could be multiple reasons for this disagreement; for example, we used cluster models whereas the previous study used periodic models, our predictions are based on a hydrolysis pathway where the terminal M–OH group acts as a nucleophile and leads to bidentate-bound products whereas the previous study treated the M–OH group as a general base and considered monodentate-bound products, and our predictions are obtained by considering the binding free energies of water, agent, and bidentate products to be of equal importance to the overall hydrolysis pathway whereas the previous study

compares Ce- and Zr-MOF-808 on the basis of their computed activation electronic energies for water displacement. Put briefly, there are simply too many confounding variables to adequately compare the two studies; rather, we await experimental testing.

In a broader sense, the results in this section show that the effect of the node metal identity on binding energies varies depending on the node topology/connectivity, which implies that the trends observed for one isostructural M(IV)-MOF series may not necessarily hold for another series. Our results suggest that there is a complex underlying relationship between geometric and electronic properties that influences the binding strengths of organophosphate molecules and their products. We further explore this confluence of properties in the QSAR modeling sections below.

4.3.5 QSAR Modeling: Rationale for Selected Descriptors

Until now, our discussion of the trends in binding free energies for the different molecules across the various MOF node topologies, connectivities, and metals has been predominantly qualitative in nature. In the analysis below, we aim to rationalize some of the observed complexity by examining how specific molecular and node features affect the binding energetics. In addition to supporting the explanation of the trends predicted by DFT, this analysis will help clarify which geometric and electronic structure descriptors are most important for describing the binding of nerve agents and their hydrolysis products to M_6 nodes. Using the large amount of data generated from our DFT calculations, we employed QSAR modeling to achieve a more comprehensive understanding of the fundamental chemistry underlying these binding events. Throughout the following sections, the DFT-computed binding free energies (ΔG_{bind}) serve as the response variable that we attempt to describe or predict using molecular and node descriptors.

To develop the QSAR models, we first selected molecular descriptors based on their potential relevance to organophosphate binding. Due to their different binding modes, we treated

neutral molecules (including water, nerve agents/simulant, and alcohol/thiol products) and bidentate anions separately, each with their own unique set of descriptors. We considered 13 neutral molecules (**H_2O**, **1**, **3**, **6**, **8**, **11**, **13**, **14**, **17**, **18**, **20**, **22**, and **24**), and 10 bidentate anions (**2**, **4**, **7**, **9**, **12**, **15**, **16**, **19**, **21**, and **23**). We compiled 25 molecular descriptors for the neutral molecules and 22 for the bidentate anions, all of which were obtained either through DFT calculations or from evaluating descriptor equations derived in previous literature studies. All quantum chemical descriptors were computed for the optimized molecular structures (not including the nodes in the calculations) at the M06-2X(SMD) level of theory, some of which required the use of electron population analysis with the NBO method[119] to compute molecular orbital parameters. The molecular descriptors, along with their corresponding notations, units, and ranges of values are summarized in Table C.6, where values in red correspond to neutral molecules and those in green correspond to bidentate anions. Below, we describe the rationale for selecting particular descriptors.

For the neutral molecules, we selected the NBO-computed partial atomic charge and Wiberg bond index[229] of the binding O/S atom ($q_{O/S,mol}$ and $BI_{O/S,mol}$, respectively) because these are the atoms that directly coordinate to the open metal sites of the *node–noOH₂* structures (Scheme 4.1b), and thus the electronic properties of these atoms are expected to directly influence the molecular binding strengths. Here, "O/S atom" refers either to the O atom of water, the O(=P) atom of nerve agent and simulant molecules, the O(H) atom of alcohols, or the S(H) atom of thiols that binds to the node open M atom. Due to the different binding mode for the bidentate anions, involving a M–O–P–O–M plane formed on adjacent M atoms at the *node–noOH₂,OH* binding sites (Scheme 4.1c), we instead computed the average charge and average Wiberg bond index on the two binding O atoms ($Avq_{O,mol}$ and $AvBI_{O,mol}$, respectively). To investigate the effects of the relative

sizes of the neutral molecules and bidentate anions, which could influence the dispersion components of the binding energies, we selected several descriptors including molecular volume (V_{mol}, computed in Gaussian 09), solvent-accessible surface area (SA_{mol}, computed in VMD 1.9.3), total number of atoms (nAt_{mol}) and electrons ($nElec_{mol}$) in each molecule, molecular weight (MW_{mol}), and average molecular weight ($AMW_{mol} = MW_{mol} / nAt_{mol}$). We note that several of these size descriptors are correlated with each other, but we chose to initially consider them all to analyze which ones individually yielded the most accurate ΔG_{bind} predictions and then we removed the less-significant correlated descriptors before performing multiple linear regression analysis. The molecular dipole moment (μ_{mol}) was chosen because it describes the overall polarity of each molecule. To characterize the overall electronic structure of each molecule, we selected six descriptors starting with ionization potential (IP_{mol}) and electron affinity (EA_{mol}), computed using the $IP_{mol} \approx -E_{HOMO}$ and $EA_{mol} \approx -E_{LUMO}$ relations based on Koopmans' theorem.[142] These two descriptors were then used to compute molecular hardness (η_{mol}), softness (S_{mol}), electronegativity (χ_{mol}), and electrophilicity (ω_{mol}) using the equations derived by Parr and co-workers[144] (see eqs C.1–C.4 in Appendix C).

We also compiled constitutional descriptors including the number of hydrogen (nH_{mol}), carbon (nC_{mol}), nitrogen (nN_{mol}), oxygen (nO_{mol}), and non-hydrogen ($nNonH_{mol}$) atoms in each molecule to examine the effects of basic chemical composition. We did not use the nO_{mol} descriptor for the bidentate anions due to their limited chemical diversity (i.e., each anion has either two or three oxygens). To measure hydrogen-bonding abilities, important for interactions with node oxo/hydroxyl groups, we computed the number of donor and acceptor atoms for H-bonds in each molecule ($nHBd_{mol}$ and $nHBa_{mol}$, respectively). Here, $nHBd_{mol}$ refers to the number of H atoms

bonded to any N and O atoms in each molecule and $nHBa_{mol}$ refers to the number of N, O, and F atoms per molecule. The $nHBd_{mol}$ descriptor was not used for the bidentate anions because values were either zero or one for these molecules. The number of rotatable bonds per molecule (nRB_{mol}), defined as the number of single bonds bound to a nonterminal heavy (i.e., non-H) atom, was chosen as a measure of molecular flexibility.[230] To measure the degree of unsaturation for each neutral molecule, we computed the unsaturation index (UI_{mol}), as defined in eq C.5.[231] Again, UI_{mol} was not used for bidentate anions due to their more limited chemical diversity. Finally, for both neutral molecules and bidentate anions, we computed the hydrophilicity index (HyI_{mol}), as defined in eq C.6.[232]

Pearson's correlation coefficients between molecular descriptors for the neutral molecules and bidentate anions are presented in the form of heatmaps in Figure C.11 and C.12, respectively. The unscaled values of the descriptors for neutral molecules and bidentate anions are listed in Table C.7 and C.8, respectively. Since the descriptors have different units, all descriptors were scaled from 0 to 1 so that their weights in the developed QSAR models may be easily compared (see eq C.7 for scaling formula). The scaled values of the descriptors for neutral molecules and bidentate anions are listed in Table C.9 and C.10, respectively.

Next, we selected descriptors to describe the *node–noOH₂* and *node–noOH₂,OH* bare node structures which served as binding sites for the neutral molecules and bidentate anions, respectively, where we treated the two types of binding sites separately with their own unique set of descriptors. More explicitly, we considered the *node–noOH₂* sites (Figure C.2a) and the *node–noOH₂,OH* sites (Figure C.2c) of Zr-mono-defect UiO-66, M-bi(trans)-defect UiO-66 (M = Zr, Hf, Ce, Th), Zr-bi(cis)-defect UiO-66, Zr-NU-1000 (large pore), Zr-NU-1000 (c pore), and M-MOF-808 (M = Zr, Hf, Ce, Th). We compiled 24 node descriptors for the *node–noOH₂* sites and

20 for the *node–noOH₂,OH* sites, all of which were obtained either through DFT calculations or taken directly from experimental databases (e.g., NIST). All quantum chemical descriptors were computed for the optimized node structures. The node descriptors, along with their corresponding notations, units, and ranges of values are summarized in Table C.11, where values in red correspond to *node–noOH₂* sites and those in green correspond to *node–noOH₂,OH* sites.

For the *node–noOH₂* sites, we selected the NBO-computed partial atomic charge ($q_{M,node}$), Wiberg bond index ($BI_{M,node}$), and valence orbital population ($ValPop_{M,node}$) of the open M atom that serves as the binding site for neutral molecules. Similarly, for the *node–noOH₂,OH* sites, we computed the average charge ($Avq_{M,node}$), average Wiberg bond index ($AvBI_{M,node}$), and average valence orbital population ($AvValPop_{M,node}$) of the two adjacent M atoms that collectively serve as the binding site for bidentate anions. To further characterize the electronic structure of the M atom(s) at the node binding sites, we compiled six descriptors to describe the different M^{4+} cations including $IP_{M^{4+},node}$, $EA_{M^{4+},node}$, $\eta_{M^{4+},node}$, $S_{M^{4+},node}$, $\chi_{M^{4+},node}$, and $\omega_{M^{4+},node}$, where ionization potential and electron affinity values were obtained from NIST.[228] We also used elemental properties to describe the M atom(s) including the atomic number ($Z_{M,node}$), atomic weight ($AtWt_{M,node}$), period in the periodic table ($Pd_{M,node}$), and covalent radius ($CovR_{M,node}$).[233] From the DFT-optimized node structures, we also computed the distance between the two adjacent M atoms of the binding sites ($d_{M–M,node}$), which may be particularly relevant for the binding mode of bidentate anions.

Beyond the M atom(s) that directly bind the molecules, we also considered it important to characterize the geometric and electronic properties of the surrounding node oxo/hydroxyl groups, which may interact through hydrogen-bonding with the bound molecules. For both the *node–noOH₂* and *node–noOH₂,OH* sites, we selected the NBO-computed partial atomic charges of the

bridging node hydroxyl atoms ($q_{\mu_3 O(H),node}$ and $q_{\mu_3 (O)H,node}$) and the bridging node oxo atom ($q_{\mu_3 O,node}$). Exclusively for the *node–noOH₂* sites, we computed charges on the atoms of the neighboring terminal node hydroxyl group ($q_{tO(H),node}$ and $q_{t(O)H,node}$) as well as the bond length and Wiberg bond order of the neighboring terminal M–OH bond ($BL_{tM–OH,node}$ and $BO_{tM–OH,node}$, respectively). Figure 4.4, depicting the *node–noOH₂* binding site of Zr-NU-1000 (large pore) as an example, includes labels for the referenced metal, oxygen, and hydrogen atoms.

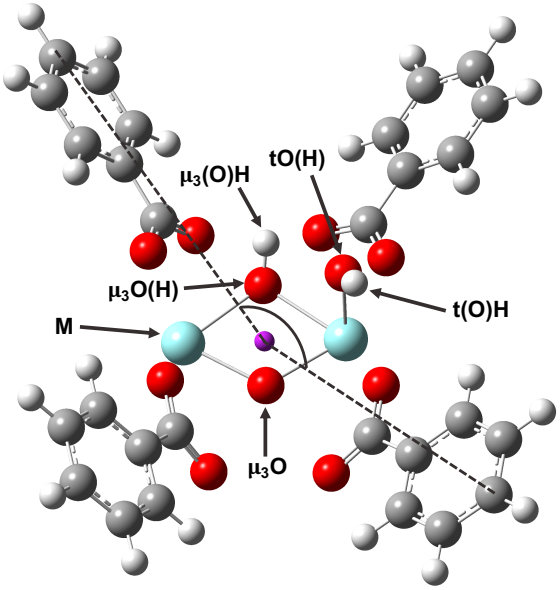


Figure 4.4. Top-down view of the *node–noOH₂* binding site of Zr-NU-1000 (large pore) with labels for the metal, bridging oxo, bridging hydroxyl, and terminal hydroxyl atoms referenced in the node descriptors used for QSAR modeling. Dark gray, white, red, and turquoise spheres represent C, H, O, and Zr atoms, respectively. The purple sphere denotes a dummy atom used to help visualize the linker bite angle, which is shown by the dashed lines.

To characterize the size of each node in its entirety, we chose the total number of electrons per node ($nElec_{node}$). Further, we selected the total number of linkers per node ($nLink_{node}$), as we previously showed that the node connectivity has a strong influence on dispersion energies (see Figure C.5 and C.7). Finally, to measure the steric effects of the local environments for *node–noOH₂* and *node–noOH₂,OH* sites, which could influence dispersion interactions between bound molecules and the benzoate linkers surrounding the binding sites, we introduce the concept of a

linker bite angle ($Bite_{node}$). We define the linker bite angle as the angle formed between two *para*-carbon atoms of diagonally opposite benzoate linkers surrounding the binding site and a point equidistant between the two adjacent M atoms of the binding site. For clarity, this angle is depicted with dashed lines in Figure 4.4. To provide context, the linker bite angle is 155.8º, 149.6º, 143.7º, 142.6º, 142.5º, and 138.2º for the optimized *node–noOH₂* cluster models of Zr-MOF-808, Zr-NU-1000 (large pore), Zr-bi(trans)-defect UiO-66, Zr-mono-defect UiO-66, Zr-bi(cis)-defect UiO-66, and Zr-NU-1000 (c pore), respectively.

Pearson's correlation coefficients between node descriptors for the *node–noOH₂* and *node–noOH₂,OH* sites are presented in the form of heatmaps in Figure C.13 and C.14, respectively. The unscaled values of the descriptors for the *node–noOH₂* and *node–noOH₂,OH* sites are listed in Table C.12 and C.13, respectively. Similar to the molecular descriptors, the node descriptors were scaled from 0 to 1 (using eq C.7) so that their weights in the developed QSAR models may be easily compared. The scaled values of the descriptors for the *node–noOH₂* and *node–noOH₂,OH* sites are listed in Table C.14 and C.15, respectively.

The distributions of binding free energy values for the neutral molecules bound to *node–noOH₂* sites and the bidentate anions bound to *node–noOH₂,OH* sites are shown in Figure C.15. Together, the scaled molecular descriptors for neutral molecules and the scaled node descriptors for *node–noOH₂* sites (49 total descriptors), along with the corresponding ΔG_{bind} response variable (156 total molecule-node combinations), were used to construct QSAR models (hereafter referred to simply as the "neutrals dataset"). Similarly, the scaled molecular descriptors for bidentate anions and the scaled node descriptors for *node–noOH₂,OH* sites (42 total descriptors), along with the corresponding ΔG_{bind} response variable (120 total molecule-node combinations), were used to construct QSAR models (hereafter referred to as the "bidentates dataset").

4.3.6 QSAR Modeling: Most Important Descriptors for Describing Organophosphate Binding to MOF Nodes

Before assessing the developed QSAR models, we first discuss uniparametric correlations based on simple linear regression between individual descriptors and the binding free energies (e.g., R^2_{uni} for ΔG_{bind} vs. $q_{O/S,mol}$) to highlight the most statistically significant (i.e., most important) descriptors. The most important molecular descriptor for describing the binding energetics of the neutrals dataset is $q_{O/S,mol}$ (R^2_{uni} = 0.25, positive correlation). The direction of the correlation indicates that lower (more negative) charges on the binding O/S atom correlate with more negative (stronger) binding free energies, which is intuitive because this should lead to a stronger attraction to the positively charged node M atom. The most important node descriptor for the neutrals dataset is $Bite_{node}$ (R^2_{uni} = 0.12, positive correlation), where the correlation direction reveals that smaller linker bite angles correlate with stronger binding free energies. This observation helps clarify one of the trends observed in Figure 4.2, where Zr-MOF-808 ($Bite_{node}$ = 155.8º) generally showed the weakest binding while Zr-NU-1000 (c pore) ($Bite_{node}$ = 138.2º) consistently showed the strongest binding, implying that the small pore environment of the Zr-NU-1000 (c pore) binding site leads to closer proximity and more favorable interactions between bound molecules and the surrounding benzoate linkers. The next most important molecular and node descriptors for the neutrals dataset are summarized in Figure C.16.

The most influential molecular descriptor for the bidentates dataset is nN_{mol} (R^2_{uni} = 0.10, negative correlation), where a larger number of nitrogen atoms per bidentate anion correlates with stronger binding free energies. The most important node descriptor for the bidentates dataset is $\omega_{M^{4+},node}$ (R^2_{uni} = 0.37, negative correlation), where more electrophilic M^{4+} cations of the *node–noOH₂,OH* sites correlate with stronger binding. This observation helps explain why the Ce

analogues of the M-bi(trans)-defect UiO-66 and M-MOF-808 node sites consistently showed the strongest binding free energies for bidentate anions (Figure 4.3d and C.8d, respectively), where Ce^{4+} displays the largest electrophilicity value ($\omega_{M^{4+},node}$ = 91.6 eV) of the four considered metals and thus has the strongest ability to bind the bidentate anions. The next most important molecular and node descriptors for the bidentates dataset are summarized in Figure C.17.

Despite the useful insights derived from simple linear regression, the correlations based on individual molecular and/or node descriptors are clearly very low, indicating that single descriptors are not sufficient to describe such complex systems, a similar conclusion drawn in our previous organophosphates QSAR study.[234] This result is not surprising, given the fairly complex trends observed in the binding free energy plots (Figure 4.2, 4.3, and C.8). Thus, we developed multi-parametric QSAR models for the neutrals and bidentates datasets by performing multiple linear regression (MLR), where MLR can improve the correlation statistics while still allowing for a simple interpretation of the contribution each descriptor has on the models in terms of its coefficient weight and sign. Initially, we built non-predictive models, where this terminology denotes that all molecule-node combinations were included for model development for each dataset (as opposed to dividing them into training and test sets). This was done to gain a deeper understanding of the most important descriptors for describing the binding energetics of our entire datasets before assessing the predictive capacity of our models using test sets.

Before model development, we removed all linearly dependent molecular and node descriptors from both datasets. To remove descriptors, we first identified any descriptors that showed a high Pearson's correlation coefficient ($|R| > 0.9$) with another descriptor (see Figure C.11–C.14). Then, for each highly correlated descriptor pair, we removed the descriptor that yielded the less accurate ΔG_{bind} prediction, as determined by comparing their R^2_{uni} and root mean

square error (RMSE) values. For development of the non-predictive multi-parametric QSAR models, we used stepwise forward-backward based feature selection combined with MLR using a 95% confidence interval, utilizing the built-in *stepwiselm* function in MATLAB. For a more detailed description of this algorithm, refer to Appendix C.

The non-predictive QSAR model developed using all 156 molecule-node combinations for the neutrals dataset is:

$$\Delta G_{bind} = -(33.05 \pm 11.67) + (55.02 \pm 4.26)q_{O/S,mol} + (17.85 \pm 3.91)IP_{mol} - (32.35 \pm 3.85)nHBd_{mol} - (29.39 \pm 13.46)q_{M,node} - (45.68 \pm 7.66)BI_{M,node} - (17.84 \pm 5.60)\omega_{M^{4+},node} + (27.02 \pm 6.03)nLink_{node} + (60.87 \pm 6.06)Bite_{node} \quad (4.7)$$

$N = 156$, $Q^2_{LOO} = 0.59$, $R^2_{adj} = 0.68$, RMSE = 12.29 kJ/mol

In eq 4.7, the descriptors are dimensionless (scaled from 0 to 1), so descriptors with larger weights can be said to have increased significance to the overall model. As evaluation metrics, we report the RMSE and the adjusted coefficient of determination (R^2_{adj}) between observed (i.e., DFT-computed) and predicted responses. We report the R^2_{adj}, which takes into consideration the number of model variables, to facilitate the comparison of models containing different numbers of terms (see eq C.8). Additionally, we report the Q^2_{LOO} statistic, calculated using the leave-one-out cross-validation (LOO-CV) procedure, in which every molecule-node combination is eliminated from the dataset once and then its response variable is predicted using the regression equation derived from the remaining set (see eq C.9).

By examining the descriptor weights in eq 4.7, we again observe that $q_{O/S,mol}$ and $Bite_{node}$ are the most significant molecular and node descriptors for the neutrals dataset. The $nHBd_{mol}$ descriptor is also important, where neutral molecules possessing more hydrogen-bonding donor

atoms correlate with stronger binding free energies, indicating that hydrogen-bonding interactions with node oxo/hydroxyl groups plays an important role in binding to M_6 node sites. Regarding the node properties, the $q_{M,node}$ descriptor is also fairly significant, where higher (more positive) charges on the open M atom correlate with stronger binding, serving as a complement to the $q_{O/S,mol}$ molecular descriptor.

The non-predictive QSAR model developed using all 120 molecule-node combinations for the bidentates dataset is:

$$\Delta G_{bind} = -(80.02 \pm 4.68) + (17.44 \pm 4.03)Avq_{O,mol} - (19.30 \pm 7.31)\mu_{mol}$$
$$- (12.20 \pm 4.90)nN_{mol} + (15.24 \pm 7.04)HyI_{mol} - (36.36 \pm 8.47)AMW_{mol}$$
$$- (25.34 \pm 4.79)\omega_{M^{4+},node} + (11.54 \pm 2.77)Bite_{node} - (16.63 \pm 5.05)q_{\mu_3 O,node} \quad (4.8)$$

$N = 120$, $Q^2_{LOO} = 0.55$, $R^2_{adj} = 0.65$, RMSE = 10.60 kJ/mol

In eq 4.8, the most influential molecular descriptor is AMW_{mol}, where bidentate anions with larger average molecular weights correlate with stronger binding free energies. Similar to the neutrals dataset, we also see that the charge on the binding O atoms is important for inducing a strong attraction to the two adjacent M atoms of the *node–noOH₂,OH* sites. As was determined by the uniparametric correlation analysis, the most influential node descriptor for the bidentates dataset is $\omega_{M^{4+},node}$. Also, the $q_{\mu_3 O,node}$ descriptor term appears to be moderately important, which is reasonable considering the typical binding mode of the bidentate anions (Figure C.6) that involves hydrogen-bonding interactions between the bridging node oxo atom and the bound molecule.

Overall, for both the neutrals and bidentates datasets, we observe that adding more descriptor terms results in improved correlations compared to simple linear regression using individual descriptors. This result is not purely a statistical artifact, but rather an indication that

both molecular and node properties, including both structural and chemical features, collectively contribute to the binding free energies of organophosphate molecules and their hydrolysis products to M_6 nodes. Considering the complex nature of the molecules used to build these models, compared to simple atomic or small molecule adsorbates, the RMSE values for both datasets are relatively low, especially when taking into consideration our use of MLR compared to more sophisticated machine learning techniques. To better evaluate the capacity of our descriptors for predicting accurate ΔG_{bind} values and to determine their robustness for describing new organophosphates and/or MOFs, we performed further statistical validations, as described below.

4.3.7 QSAR Modeling: Model Predictions and Applicability Domains

To develop predictive QSAR models, we divided each dataset into training and test sets using random selection, while ensuring that each training set spanned the entire response variable space for its respective dataset (i.e., we included the molecule-node combinations with the highest and lowest ΔG_{bind} values in the training sets). Each test set included 20% of its respective total dataset. Additionally, since we already gained a detailed understanding of the important molecular and node descriptors from the non-predictive models generated using standard MLR, we now allowed for molecular-node descriptor interaction terms (e.g., $q_{O/S,mol} \times q_{M,node}$) in the regression equations to further improve the models' predictive capabilities. However, the molecular descriptors and node descriptors are linearly independent and show no correlation ($R = 0$), so the resulting models developed using MLR with interaction terms are still physically interpretable.

Several tests for statistical significance were used to evaluate the predictive value of the models. To evaluate the training sets, we used the $RMSE_{train}$, R^2_{adj}, and Q^2_{LOO} values, as discussed earlier. Additionally, we used the *y*-randomization method to measure the possibility of chance correlation in the developed QSAR models, utilizing the ${}^{C}R^2_p > 0.5$ test (see eq C.10) to determine

that our models were not generated purely by chance.[150] To evaluate the test sets in terms of both quantitative and ranking accuracy, we used the root mean square error in prediction ($RMSE_{test}$, see eq C.11), the coefficient of determination (R^2_{test}, see eq C.12), and the Spearman's rank correlation coefficient (ρ_{test}, see eq C.13) values. For both the neutrals and bidentates datasets, we performed dataset division 100 times to examine the variability in model statistics (Figure C.18 and C.19). For each dataset, when comparing the statistics across all 100 generated models, the optimal model was identified as the one that yielded the lowest $RMSE_{test}$, while also satisfying the minimum QSAR criteria of $Q^2_{LOO} > 0.5$, ${}^{C}R^2_p > 0.5$, and $R^2_{test} > 0.6$.[151]

The optimal predictive QSAR model for the neutrals dataset is:

$$\Delta G_{bind} = -(71.49 \pm 5.09) + (29.85 \pm 6.11)nH_{mol} - (43.52 \pm 8.00)nHBa_{mol} + (62.99 \pm 4.80)Bite_{node} - (19.72 \pm 4.58)nElec_{node} + (45.30 \pm 12.18)(q_{O/S,mol} \times q_{M,node}) - (27.91 \pm 7.34)(q_{O/S,mol} \times CovR_{M,node}) + (46.51 \pm 10.93)(q_{O/S,mol} \times nLink_{node}) + (40.65 \pm 9.89)(q_{O/S,mol} \times q_{t(O)H,node}) + (45.53 \pm 7.21)(\chi_{mol} \times nLink_{node}) - (31.21 \pm 9.21)(nH_{mol} \times BO_{tM-OH,node}) - (21.42 \pm 9.74)(nHBd_{mol} \times nLink_{node}) + (24.45 \pm 9.82)(nHBa_{mol} \times BI_{M,node}) - (17.91 \pm 9.46)(HyI_{mol} \times q_{M,node}) + (42.40 \pm 8.64)(AMW_{mol} \times q_{M,node}) \quad (4.9)$$

$N_{train} = 125$, $Q^2_{LOO} = 0.59$, $R^2_{adj} = 0.83$, $RMSE_{train} = 9.40$ kJ/mol, ${}^{C}R^2_p = 0.79$

$N_{test} = 31$, $R^2_{test} = 0.75$, $RMSE_{test} = 8.25$ kJ/mol, $\rho_{test} = 0.82$

The optimal predictive QSAR model for the bidentates dataset is:

$$\Delta G_{bind} = -(84.81 \pm 4.39) + (14.28 \pm 4.85)Avq_{O,mol} - (22.20 \pm 3.49)nN_{mol} + (20.86 \pm 8.15)HyI_{mol} - (32.46 \pm 9.71)AMW_{mol} - (23.78 \pm 5.59)\omega_{M^{4+},node} + (10.27 \pm 3.32)Bite_{node} - (18.78 \pm 6.07)q_{\mu_3 O,node} \quad (4.10)$$

N_{train} = 96, Q^2_{LOO} = 0.51, R^2_{adj} = 0.61, $RMSE_{train}$ = 11.23 kJ/mol, ${}^{C}R^2_{p}$ = 0.61

N_{test} = 24, R^2_{test} = 0.70, $RMSE_{test}$ = 9.77 kJ/mol, ρ_{test} = 0.82

Based on the computed statistics, eqs 4.9 and 4.10 represent models with good predictive capacity in terms of both quantitative and ranking accuracy, where the optimal model for the neutrals dataset has slightly better predictive ability. Comparing eqs 4.7 and 4.9 for the neutrals dataset, we observe that using MLR with interaction terms results in models with better quantitative accuracy compared to standard MLR. In contrast, the optimal predictive QSAR model for the bidentates dataset (eq 4.10) includes no molecular-node descriptor interaction terms and is nearly identical to the non-predictive model (eq 4.8). Due to the many differences between the two datasets, we could not find a definitive explanation for this observation. The most likely reason for the presence of more terms in the regression equation for the neutrals dataset is that the neutral molecules are more chemically diverse (including water, nerve agents/simulant, and alcohol/thiol products) than the bidentate anions, thus requiring a larger number of descriptors to accurately describe their binding energetics. Nonetheless, both predictive models in eqs 4.9 and 4.10 can be used to make reliable binding free energy predictions for new organophosphate molecules and hydrolysis products and/or new M_6 nodes.

The ΔG_{bind} values for all 156 (120) molecule-node combinations predicted using eq 4.9 (eq 4.10) are given in Table C.16 (Table C.17). The regression plots of observed (DFT-calculated) and predicted binding free energies for the training and test sets are represented in Figure 4.5a and

4.5b for the neutrals and bidentates datasets, respectively. The dashed line in each plot that shows the relationship $y_{pred} = y_{obs}$ is included to help assess the predictive power of the models, where the agreement between observed and predicted ΔG_{bind} values is slightly better for the neutrals dataset. The distribution of raw residuals for the optimal models are shown in Figure C.20 and C.21 for the neutrals and bidentates datasets, respectively, where the residuals for both models have an acceptable normal distribution with no clear outliers.

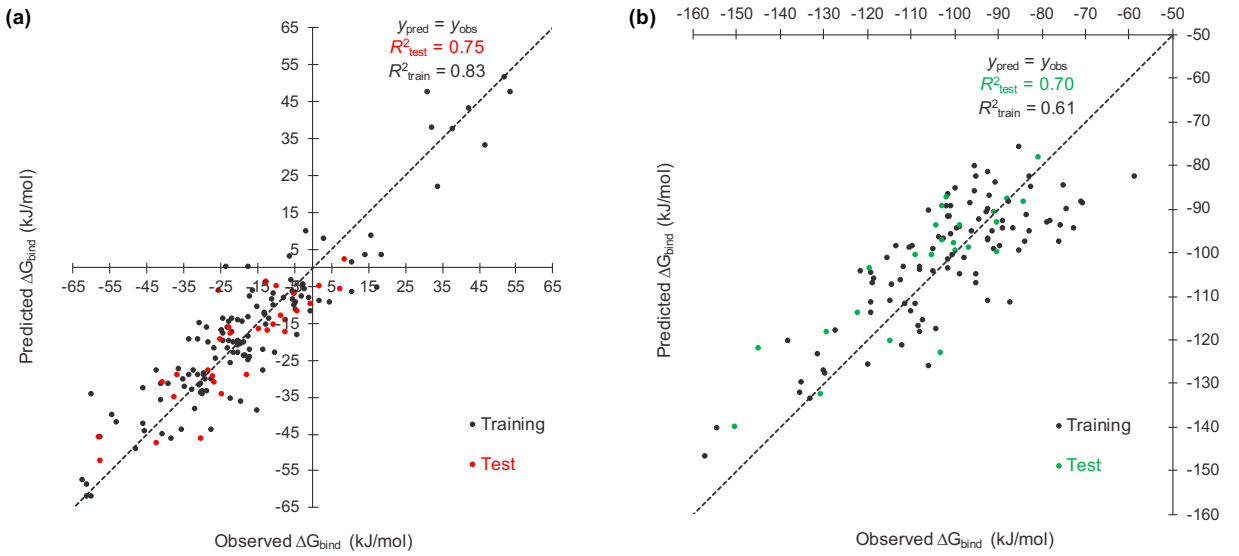


Figure 4.5. Comparison between observed (DFT-calculated) binding free energies with those predicted by the optimal predictive QSAR model developed for (a) the neutrals dataset (eq 4.9) and (b) the bidentates dataset (eq 4.10). The dashed line in each plot shows the relationship $y_{pred} = y_{obs}$, where data points for an ideal model lie close to this line.

To further evaluate the robustness of our models for describing new molecule-node combinations, we established the applicability domain (AD) of the optimal predictive QSAR model for each dataset using the leverage-based method.[153,154] Here, the AD represents the bounds in which a model tolerates a new molecule-node combination, where predictions are considered unreliable (i.e., extrapolations) if they lie outside the model's AD. To visualize the AD for each model, a Williams plot was constructed using standardized residuals (*st*) and leverage (*h*) values, where leverage is a measure of the influence of a molecule-node combination's properties on the

regression model. The formulas used to calculate *st* and *h* values are provided in Appendix C (eqs C.14–C.16). For each model, the cut-off leverage is defined as $h^* = 3p/N_{train}$, where p is the number of regression coefficients (which includes the intercept) and N_{train} is the number of training set observations.[154] Any prediction is considered unreliable and outside the AD if its $h > h^*$.

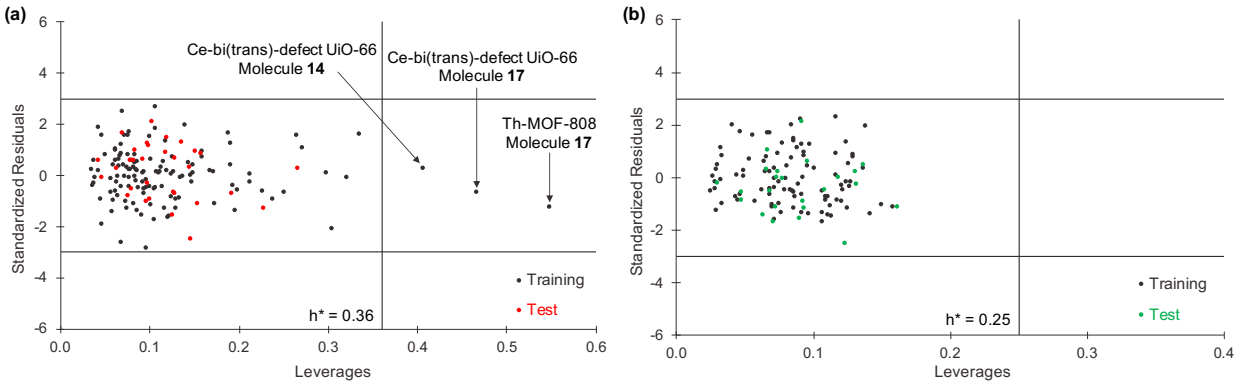


Figure 4.6. Williams plot showing the applicability domain for the optimal predictive QSAR model developed for (a) the neutrals dataset (eq 4.9) and (b) the bidentates dataset (eq 4.10). The horizontal lines in each plot signify the bounds for the standardized residuals (at ± 3 standard deviation units). The vertical line in each plot represents the cut-off leverage (h^*).

The Williams plots for the optimal predictive QSAR models developed for the neutrals and bidentates datasets are shown in Figure 4.6a and 4.6b, respectively. In these plots, it is simple to identify any response outliers (with $st > 3$ standard deviation units) and influential molecule-node combinations with high leverage ($h > h^*$). The horizontal lines in each plot signify the bounds for the standardized residuals (at ± 3 standard deviation units) and the vertical line in each plot represents the cut-off leverage (h^* = 0.36 and 0.25 for the neutrals and bidentates dataset, respectively). In Figure 4.6a for the neutrals dataset, only 3 out of the 156 molecule-node combinations are located outside of the AD, thus predictions for all other combinations can be considered reliable. The 3 high-leverage molecule-node combinations (molecule **14** bound to the *node–noOH*$_2$ site of Ce-bi(trans)-defect UiO-66, **17** bound to Ce-bi(trans)-defect UiO-66, and **17** bound to Th-MOF-808) are all in the training set. The most likely reason why these combinations

lie outside the AD is that molecules **14** (VX) and **17** (DESH) have relatively large steric bulk compared to the other neutral molecules and are the only ones that contain sulfur. However, these combinations can be said to have "good" leverage because they are in the training set, meaning they reinforce the strength of the model for making future predictions for similar molecule-node pairs. Conversely, all molecule-node combinations are located within the AD for the bidentates dataset (Figure 4.6b). While the distribution of residuals is similar for the two datasets, we note that the leverages are generally lower for the bidentates dataset, which we attribute to the more limited diversity of the bidentate anions.

Overall, the results in this section indicate that the optimal predictive QSAR models developed for the neutrals and bidentates datasets (eqs 4.9 and 4.10, respectively) can be used to make relatively accurate binding free energy predictions for different nerve agents and their hydrolysis products bound to M_6 nodes not considered herein.

4.4 Conclusions

In this work, we used DFT to analyze the binding of water, nerve agents (GA, GB, GD, VX, A-230, and A-232), a nerve agent simulant (DMMP), and their corresponding hydrolysis products (bidentate anions, monodentate anions, and neutral alcohols/thiols) to MOF nodes with varying node topology, connectivity, and metal identity. As binding sites for the molecules, we considered the hexanuclear M(IV) oxide cluster nodes of Zr-mono-defect UiO-66, M-bi(trans)-defect UiO-66 (M = Zr, Hf, Ce, Th), Zr-bi(cis)-defect UiO-66, Zr-NU-1000 (large pore), Zr-NU-1000 (c pore), and M-MOF-808 (M = Zr, Hf, Ce, Th). Overall, the trends we observed in binding free energies for the nerve agents and hydrolysis products are fairly complex due to the significant diversity of the molecules and nodes considered, which made it difficult to analyze our data all together. Instead, we chose to focus our analysis on identifying the optimal M_6 node sites for hydrolyzing

each nerve agent and simulant molecule. In MOF-catalyzed organophosphate hydrolysis, the displacement of a terminal water from the node active site, the subsequent adsorption of nerve agent, and the ultimate desorption of bidentate-bound products are all critical, and potentially rate-limiting, steps under various reaction conditions. Thus, we used a holistic approach towards analyzing hydrolysis pathways by computing an optimization metric for each agent-node pair based on the relative binding strengths of key reaction species (i.e., water, agents, and the strongest-bound bidentate anion products). This analysis showed that no single metal or node topology/connectivity is predicted to be optimal for hydrolyzing all possible organophosphate molecules, suggesting that MOFs should be selected based on the agent of interest. In general, our results are in qualitative agreement with observed experimental trends, implying that our simple binding-energy-based approach may be an adequate substitute for full mechanistic studies as the initial computational step used to predict promising M_6 node sites for nerve agent hydrolysis.

Using the large amount of data generated from our DFT calculations, we then derived QSAR models to better explain the complex trends observed in binding free energies for the different molecules across the various MOF node topologies, connectivities, and metals. Through simple linear and multiple linear regression, we identified the most important descriptors for describing the binding of nerve agents and their hydrolysis products to M_6 nodes. These results showed that both molecular and node properties, including both structural and chemical features, collectively contribute to the binding energetics. By splitting the data into training and test sets and performing a thorough statistical analysis, we showed that our QSAR models are capable of making quantitatively accurate binding free energy predictions for neutral molecules and bidentate anions in a wide variety of M(IV)-MOFs. To further evaluate the bounds in which our models tolerate new molecule-node combinations, we established applicability domains using the

leverage-based method, where this analysis showed that the optimal QSAR models can be used to make reliable predictions for different nerve agents and their hydrolysis products bound to M_6 nodes not considered herein. Ultimately, the insights gained from our study can be used to guide future experiments for the synthesis of MOF systems with enhanced catalytic activity.

Chapter 5: Computational Screening of Metal–Organic Framework-Supported Single-Atom Transition-Metal Catalysts for the Gas-Phase Hydrolysis of Nerve Agents

This chapter is a modified version of a published manuscript: Mendonca, M. L.; Snurr, R. Q. *ACS Catal.* **2020**, *10*, 1310–1323.

5.1 Introduction

The earliest documented use of chemical weapons dates back to ~400 BCE during the Peloponnesian War, where the Spartans used arsenical smoke against their enemies.[235] The largest deployment of chemical warfare agents (CWAs) occurred during World War I, causing an estimated 1.3 million casualties.[236] In 1993, the Chemical Weapons Convention was drafted to completely ban the development, stockpiling, and use of CWAs as well as destroy existing stockpiles.[3] Despite international efforts to prohibit their use, various CWAs have been used as recently as 2018 on both military and civilian populations.[193,196] Of particular concern are nerve agents, such as sarin (GB), soman (GD), and VX. In the human body, the acetylcholinesterase enzyme is responsible for breaking down the neurotransmitter acetylcholine. Nerve agents function by inhibiting this enzyme, which causes the neurotransmitters to keep sending signals, but none are broken down. This leads to a buildup of acetylcholine in the synaptic cleft between nerve and muscle cells and causes muscles to continuously contract, which can eventually lead to death.[197] Since these highly toxic chemicals remain a serious global threat, materials are needed for both the destruction of weapon stockpiles and personal protection via fabric coatings or respirators to ensure the safety of military specialists and untrained citizens alike.

Metal–organic frameworks (MOFs) are a promising class of catalysts for nerve agent degradation.[66] MOFs are highly crystalline materials composed of metal nodes connected by organic linkers, which combine to form highly porous frameworks. They are especially attractive for detoxification because of their extremely large surface areas, their high concentration of periodic active sites, and their chemical functionality that can be tuned to adsorb specific target molecules. Some of the most promising catalysts for this application are MOFs containing Zr_6 nodes, which display a periodic distribution of strongly Lewis-acidic Zr(IV) metal centers. Zr-based MOFs are highly stable under harsh thermal (up to 500 °C) and chemical (pH 1–12) conditions because of their strong Zr-oxo node-linker bonds.[22] In aqueous base solution (e.g., *N*-ethylmorpholine), these MOFs have shown degradation half-lives on the order of minutes for the hydrolysis of nerve agents and their simulants,[1,50–54] where the buffer moderates the reaction pH and regenerates Zr active sites. These buffered solutions are promising for the catalytic destruction of nerve agent stockpiles, but they are not feasible for application in gas masks and protective fabrics.[55]

To address this issue, recent studies have begun to investigate the gas-phase hydrolysis of nerve agents.[56] Initially, Troya performed a computational mechanistic study of the gas-phase hydrolysis of sarin on the Zr-based MOFs UiO-66 and MOF-808.[57] In contrast to previous experimental reports of catalytic turnover in buffered solution, the calculations showed extremely strong binding energies between the phosphonic acid product and the nodes. The product inhibition was predicted to be considerably worse when the hydrolysis product bound to the node through two adjacent Zr-oxo bonds in a bidentate fashion, compared to a monodentate binding mode on one Zr atom. Subsequent experiments were performed to probe the in situ capture and degradation of the nerve agent simulant dimethyl methylphosphonate on various Zr-MOFs.[58,59] These

experiments corroborated the theoretical prediction that exposure to vapor-phase organophosphonates leads to strongly bound products on the Zr_6 nodes, albeit under ultrahigh-vacuum conditions, which may inhibit further reactions.

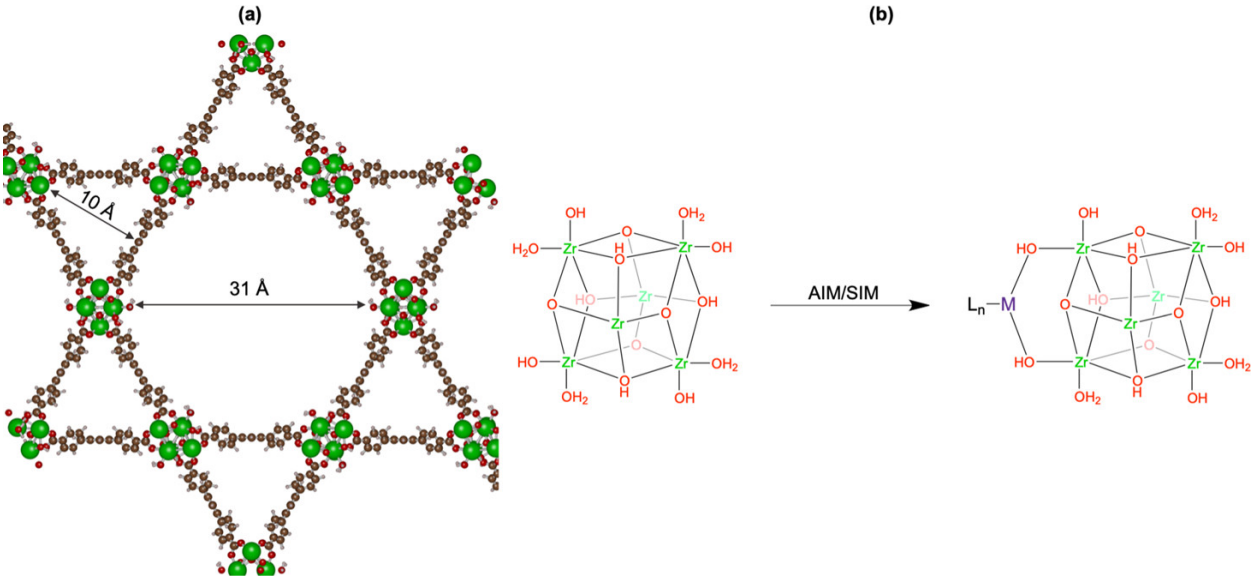


Figure 5.1. (a) Structure of NU-1000, viewed along the crystallographic *c* direction, highlighting the 31 Å hexagonal and 10 Å triangular channels (visualized using iRASPA[2]). Brown, white, red, and green spheres represent C, H, O, and Zr atoms, respectively. (b) Simplified representation of a NU-1000 node before and after metal deposition by AIM or SIM. For simplicity, only one single-atom site is depicted, although it is possible to deposit additional single-atom sites on the three remaining faces of the node.

A potential method to avoid this problem is to deposit metals as single-atom catalysts on the MOF nodes, such that the hydrolysis products bind to the active site in a more favorable monodentate fashion. Atomic layer deposition in MOFs (AIM)[237–240] or solvothermal deposition in MOFs (SIM)[203,241–244] can be used to synthesize such systems. NU-1000 (Figure 5.1), which is constructed from $[Zr_6(\mu_3\text{-}O)_4(\mu_3\text{-}OH)_4(OH)_4(OH_2)_4]^{8+}$ nodes and tetratopic 1,3,6,8-(*p*-benzoate)pyrene ($TBAPy^{4-}$) linkers, is an ideal platform for metalation. NU-1000 contains uniform hexagonal channels with 31 Å diameters that facilitate diffusion of metal precursors through the MOF pores to react with the periodic array of Zr_6 nodes, which contain reactive terminal –OH and $–OH_2$ groups that can substitute with the precursors, resulting in metal

deposition at each node face (see Figure 5.1b). Through a combination of various experimental techniques, such as single-crystal and powder X-ray diffraction (SCXRD and PXRD), and density functional theory (DFT) calculations,[245] it is possible to obtain atomically precise characterization of the deposited catalyst sites and their attachment to the MOF supports. For example, Hupp and Farha and co-workers have used SCXRD and PXRD to analyze single-atom V,[241] Mo,[246] Cr,[247] and Cu[248] species deposited on the nodes of NU-1000. Various postsynthetic strategies also allow experimentalists to prevent multiatom cluster growth and to direct the deposition of metal ions toward different-sized pores of NU-1000 depending on the desired grafting site.[249] In contrast to other solid supports, the MOF-supported catalytic active sites are uniformly sized, uniformly spaced, and are crystallographically well-defined, which enables detailed computational mechanistic studies to be performed to analyze structure–function relationships across an array of metals.[250–253]

Herein, we use DFT to perform a thorough screening of single-atom transition-metal catalysts, in varying oxidation states, deposited on NU-1000 nodes for the gas-phase hydrolysis of sarin. Through screening and calculation of molecular descriptors, we explore periodic trends for insights into the role of electronic structure in catalyzing organophosphonate hydrolysis. Our initial goal in this work is to investigate materials capable of circumventing the product inhibition previously observed during gas-phase degradation of nerve agents on Zr_6 nodes, by forcing the products to bind to the active site in a monodentate fashion, thus facilitating catalyst regeneration. By calculating the complete reaction pathways for selected M–NU-1000 systems, we highlight the need to consider more than a single reaction step when comparing catalytic cycles across a large group of diverse materials. Finally, using the energetic span model (ESM) to calculate relative

turnover frequencies (TOFs), we identify single-atom catalysts that are predicted to have higher TOFs than unfunctionalized NU-1000 for this reaction.

5.2 Computational Details

5.2.1 Cluster Models

A cluster model was used to represent the Zr_6 node of NU-1000, which was obtained from Mondloch et al.[1] and corresponds to the mixed-staggered proton topology.[48] A small "formate model" was constructed by truncating all eight pyrene-based linkers to capping formate groups ($[Zr_6(\mu_3\text{-}O)_4(\mu_3\text{-}OH)_4(OH)_4(OH_2)_4]^{8+}(HCOO^-)_8$, Figure 5.2b). This model was used to screen the supported single-metal catalysts. A larger "benzoate model" was constructed by modeling the bottom four linkers as formate groups and using benzoate groups for the four linkers located around the supported transition metals ($[Zr_6(\mu_3\text{-}O)_4(\mu_3\text{-}OH)_4(OH)_4(OH_2)_4]^{8+}(C_6H_5COO^-)_4(HCOO^-)_4$, Figure D.1 in Appendix D). This larger model was only used to test the effect of model size on the reaction energetics for the Ti^{IV}–NU-1000 system.

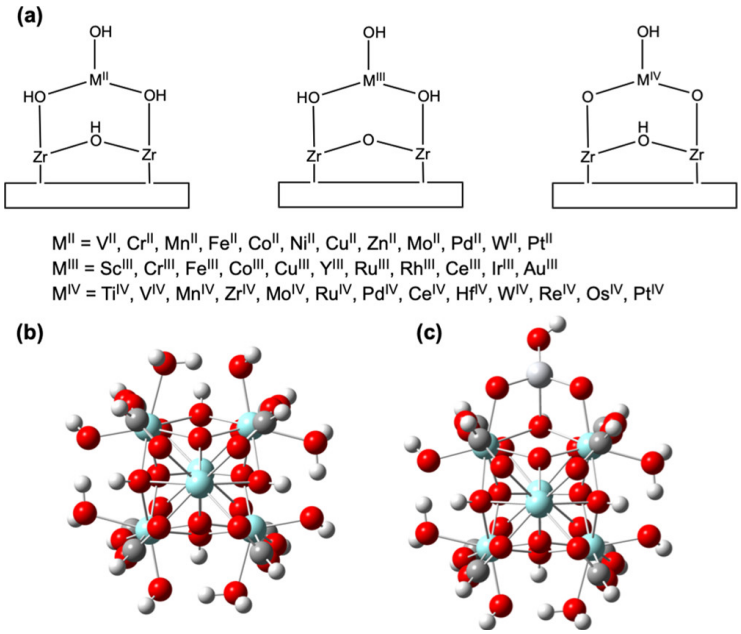


Figure 5.2. (a) Representation of M–NU-1000 systems with metals in +2, +3, and +4 oxidation states. (b) Optimized NU-1000 and (c) Ti^{IV}–NU-1000 (base catalyst denoted as **Ti^{IV}–OH**) formate cluster models. Dark gray, white, red, turquoise, and light gray spheres represent C, H, O, Zr, and Ti atoms, respectively.

Initially, we studied 36 M–NU-1000 systems with metals in +2, +3, and +4 oxidation states. We considered metals that have been previously deposited using AIM or SIM as well as new metal/oxidation state combinations to potentially guide future experiments.[254] The M^{II}–NU-1000 systems were V^{II}, Cr^{II}, Mn^{II}, Fe^{II}, Co^{II}, Ni^{II}, Cu^{II}, Zn^{II}, Mo^{II}, Pd^{II}, W^{II}, and Pt^{II}. The M^{III}–NU-1000 systems were Sc^{III}, Cr^{III}, Fe^{III}, Co^{III}, Cu^{III}, Y^{III}, Ru^{III}, Rh^{III}, Ce^{III}, Ir^{III}, and Au^{III}. The M^{IV}–NU-1000 systems were Ti^{IV}, V^{IV}, Mn^{IV}, Zr^{IV}, Mo^{IV}, Ru^{IV}, Pd^{IV}, Ce^{IV}, Hf^{IV}, W^{IV}, Re^{IV}, Os^{IV}, and Pt^{IV}. Following metal deposition by SIM or AIM, using water as a coreactant,[237] the original terminal –OH and $–OH_2$ groups on the Zr_6 node are replaced by a **M–OH** species bridging two oxo or hydroxo groups. This chelating mode has previously been determined as the most energetically favorable[255] and crystallographically dominant[241,246,247] configuration for metals anchored on NU-1000 nodes. The most stable proton topologies for the single-atom sites, where each structure is charge-neutral, were obtained from Ye et al.[251] To confirm that these topologies were indeed most stable, we reoptimized each system using the computational details described below, and the different topologies are depicted in Figure 5.2a. The optimized NU-1000 formate cluster models before and after deposition of Ti^{IV} are provided in Figure 5.2b,c, respectively. Although additional single-atom sites may be deposited on the three remaining faces of the node,[255] we only modeled one site for simplicity.

5.2.2 Electronic Structure Calculation Details

All electronic structure calculations were performed using the Gaussian 09 package (revision D.01).[109] Geometry optimizations and frequency calculations were performed for all species using DFT with the M06-L functional[75] and an ultrafine integration grid. We chose M06-L because it exhibits a reasonable balance between chemical accuracy and computational expense, it inherently

includes some dispersion corrections,[110] it performs well for transition-metal systems,[256] and it has provided reliable results in previous computational studies for NU-1000.[255,257,258] An automatic density-fitting set generated by Gaussian 09 was employed to reduce the computational cost. The def2-SVP basis set was used for H, C, and O atoms of the Zr_6 node, linkers, and capping groups. The def2-TZVP basis set was used for atoms of the reacting water and sarin molecules and the transition metals, including the SDD effective core potential (ECP) for second- and third-row transition-metal atoms.[76,79,219] For Ce, the SDD basis set with the associated ECP was applied.[80]

The C atoms of the formate and benzoate linker groups were held fixed to mimic the constraints imposed by the surrounding MOF structure, whereas all remaining atoms were allowed to relax. The nature of all stationary points, both intermediates and transition states, were verified by the calculation of analytic vibrational frequencies. All minima were characterized by zero imaginary frequencies, whereas all transition-state structures had exactly one imaginary frequency. These frequencies were used to calculate molecular partition functions using the particle in a box, rigid rotor, and quantum mechanical harmonic oscillator approximations, but all frequencies below 50 cm^{-1} were replaced with values of 50 cm^{-1} (quasi-harmonic-oscillator approximation).[70] All reported thermochemical properties were computed at standard conditions (298.15 K, 1 atm) in the gas phase (i.e., with no solvation model). The energy values in the article correspond to standard-state Gibbs free energies, and self-consistent field (SCF) energies and enthalpies are provided in Appendix D.

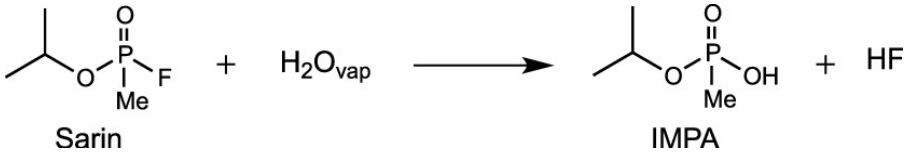
Geometry optimizations were performed for all possible spin states for each metal system, and the most stable spin structures (i.e., lowest-energy electronic states) were used to calculate the energetics for each species in a mechanism. Cartesian coordinates and spin states for all optimized structures are given in Appendix D.

5.3 Results and Discussion

5.3.1 Binding Energies

The overall reaction is shown in Scheme 5.1, where a vapor-phase water molecule hydrolyzes sarin into the products isopropyl methylphosphonic acid (IMPA) and HF. Before calculating the full catalytic mechanisms, we were interested in analyzing the binding energies of sarin, water, and IMPA to the single-atom catalysts. Because the ultimate goal is to use these materials for application in protective equipment such as gas masks, it is important to consider competitive adsorption with water that would be relevant in a realistic humid environment. Without a high affinity for selective adsorption of sarin under ambient conditions, high levels of catalytic activity are irrelevant.[155] Other atmospheric contaminants such as CO_2, SO_2, and NO_2 are also relevant[259,260] but are outside the scope of this initial study. Further, since product inhibition is known to be an issue for gas-phase nerve agent hydrolysis, we wanted to calculate the binding energies of IMPA to the 36 M–NU-1000 systems to potentially rule out any metals before investigating full degradation pathways.

Scheme 5.1. Overall Reaction for Gas-Phase Sarin Hydrolysis



We considered the binding geometry in which sarin (or IMPA) binds through a M–O(=P) bond, as this is known to be the most favorable binding mode.[57,221] All water binding geometries were initially modeled through a M–O bond and then optimized. We found that water either stayed in this chemisorbed geometry or converged to a physisorbed water molecule depending on the system. The binding free energies ($\Delta G_{binding}$) for the three adsorbate species are shown in Figure 5.3a–c for M^{II}, M^{III}, and M^{IV} systems, respectively. The $\Delta G_{binding}$ values are calculated via

$$\Delta G_{\text{binding}} = G_{\text{MOF+adsorbate}} - G_{\text{MOF}} - G_{\text{adsorbate}} \quad (5.1)$$

where a negative $\Delta G_{\text{binding}}$ indicates favorable binding. In Figure 5.3, the *x*-axes are labeled with elements from left to right along the periodic table, where the first-row metals appear first and then second-row metals and so on. The values from Figure 5.3 are listed in Tables D.1–D.3.

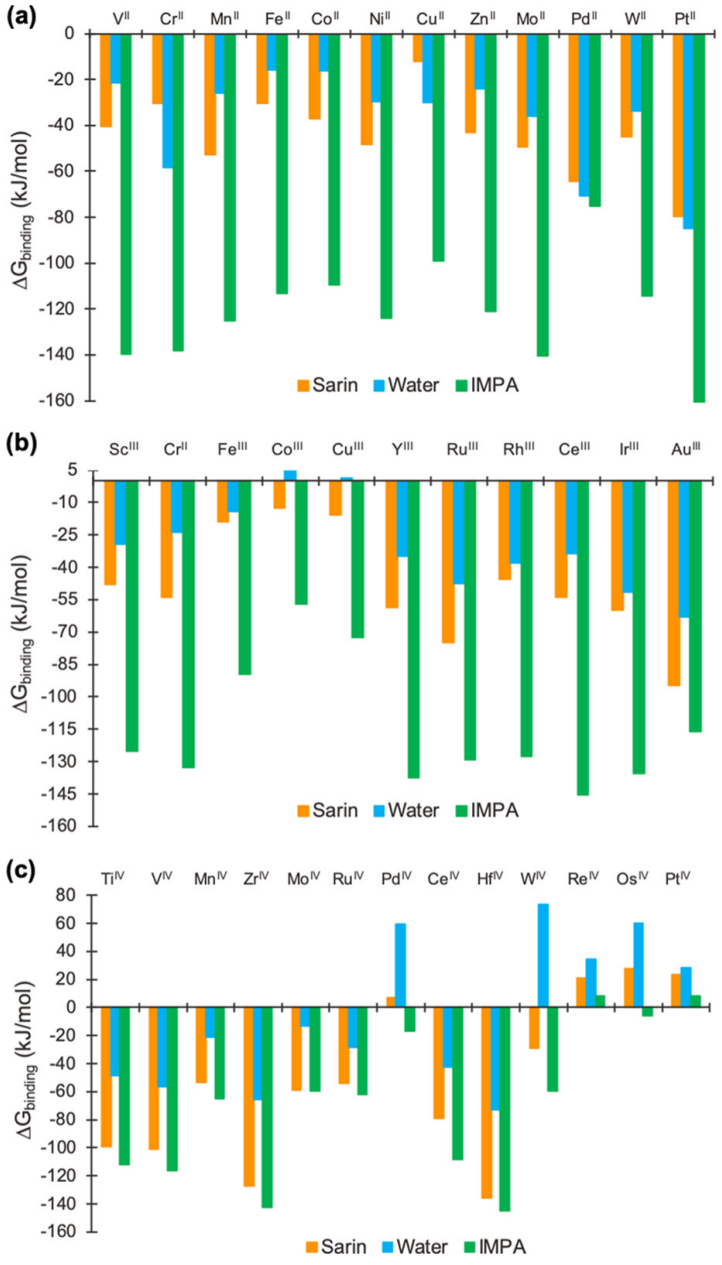


Figure 5.3. Binding free energies for sarin, water, and IMPA to (a) M^{II}–NU-1000, (b) M^{III}–NU-1000, and (c) M^{IV}–NU-1000 systems.

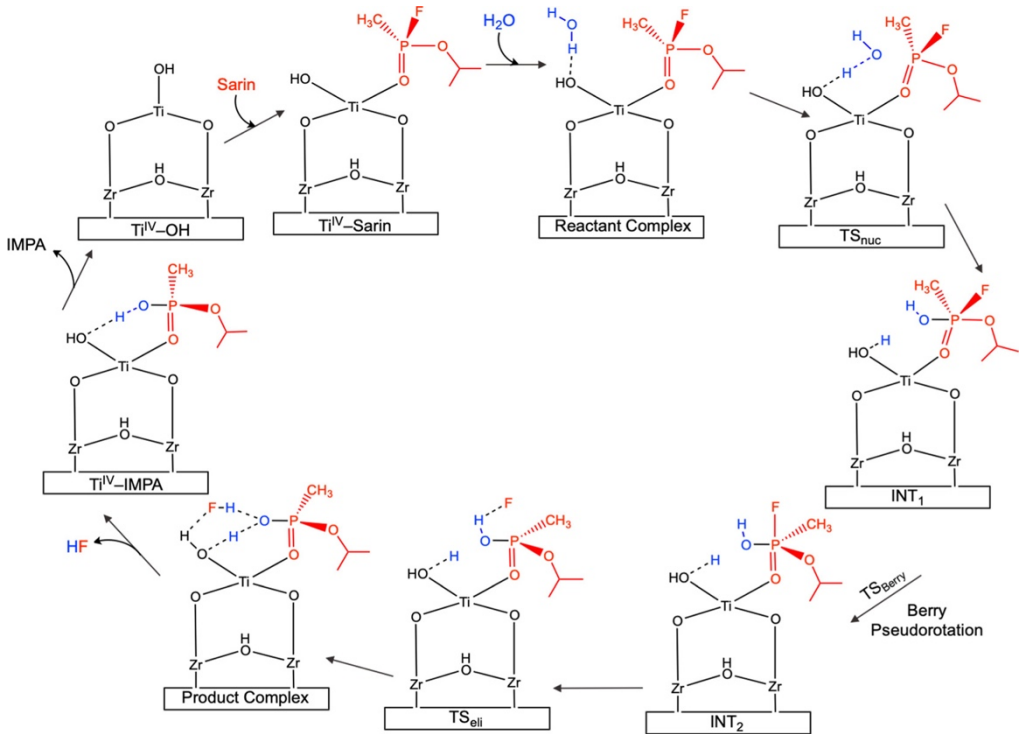
As mentioned above, it is ideal for sarin to have a stronger, or at least similar, binding energy to the catalysts compared to water. This is true for 32 out of the 36 systems, where 4 M^{II}–NU-1000 systems (Cr^{II}, Cu^{II}, Pd^{II}, and Pt^{II}) showered slight preferential adsorption for water. Twenty three out of the 36 systems showed IMPA binding energies larger than 100 kJ/mol and thus may be expected to suffer from product inhibition at room temperature. It is interesting to note that Zr^{IV}, whether isolated as a single-atom here or in the unfunctionalized parent MOF, exhibits very strong product binding, most likely due to its strong Lewis acidity.[139] By examining periodic trends, it is apparent that M^{IV}–NU-1000 systems generally exhibit the strongest binding energies for sarin, whereas M^{II}–NU-1000 systems show the strongest binding of IMPA. In general, most of the systems tend to bind IMPA much stronger than sarin, which is especially evident for the M^{II}–NU-1000 systems. The large difference in their binding energies can be attributed to the different binding geometry for the two adsorbate species on the single-atom sites. As shown in Figure D.2, the only interaction for sarin is through a relatively long M–O(=P) bond (e.g., 2.14 Å for **Fe^{II}-Sarin**), whereas for IMPA, there is a stabilizing hydrogen bond formed with the M–OH group in addition to a shorter M–O(=P) bond (e.g., 1.92 Å for **Fe^{II}-IMPA**). Surprisingly, we discovered that some of the M^{IV}–NU-1000 systems (Re^{IV}, Os^{IV}, and Pt^{IV}) showed unfavorable binding energies for sarin, and so we did not explore these metals any further. Pd^{IV}–NU-1000 also exhibited unfavorable binding, although its binding free energy for sarin was calculated to be only +7.1 kJ/mol, and thus we did not rule it out at this stage.

5.3.2 Catalytic Cycle and Validation of Formate Cluster Model for Ti^{IV}–NU-1000

To clearly explain each step of the proposed catalytic cycle for sarin hydrolysis on these systems, we use Ti^{IV}–NU-1000 as an example. We choose to discuss Ti^{IV} here because it is quite similar to the Zr^{IV} active sites in unfunctionalized NU-1000. Also, the only possible spin state for the Ti^{IV}–

NU-1000 system is a singlet, which simplifies the following discussion. The entire catalytic cycle is depicted in Scheme 5.2, starting from the deposited base catalyst, **Ti^{IV}–OH**, active site. Throughout the scheme, the sarin and ambient water molecules are colored red and blue, respectively, to clarify the movement of atoms during the course of the mechanism. First, sarin binds through a relatively strong Ti–O(=P) bond ($\Delta G_{binding}$ = –99.2 kJ/mol). In the **Reactant Complex**, a water molecule (that would be present in a humid environment) binds to the Ti–OH group through H-bonding. For direct comparison to Troya's work, this corresponds to a general base hydrolysis mechanism.[57] In a concerted step, the Ti–OH group abstracts a H atom from water, which then performs a nucleophilic attack on the P atom of sarin in **TS_{nuc}**. We only discuss one pathway herein, where nucleophilic attack occurs directly opposite the –OiPr group but note that there are two other pathways where attack can also occur opposite the –F or –CH_3 groups. A full explanation for this choice is given in Appendix D (see Figure D.3).

Scheme 5.2. Illustration of the Proposed Catalytic Cycle of Gas-Phase Sarin Hydrolysis on a Ti^{IV}–NU-1000 Cluster



Nucleophilic attack leads to the formation of a pentacoordinated trigonal bipyramidal (TBP) **INT$_1$** species. The hydroxide nucleophile and –OiPr group are axial substituents, whereas the –F and –CH_3 groups are equatorial. For the purposes of screening across many metals, we were only interested in the mechanism involving cleavage of the P–F bond to generate HF product as this is known to be the most favorable pathway.[57] In TBP phosphorus molecules, the most electronegative substituent is more stable in an axial position, from which elimination is more energetically favorable.[126] Thus, elimination of –F requires a Berry pseudorotation to direct it from an equatorial to an axial site,[122,261] forming **INT$_2$**. In **TS$_{eli}$**, elimination of HF involves proton transfer from the nucleophilic –OH group. The resulting **Product Complex** is stabilized by three hydrogen bonds, as depicted in Scheme 5.2. Once HF desorbs from the active site, the final **Ti^{IV}–IMPA** species is generated, where IMPA is bound in a monodentate fashion to the Ti atom. Finally, the IMPA product desorbs and regenerates the active site, completing the cycle.

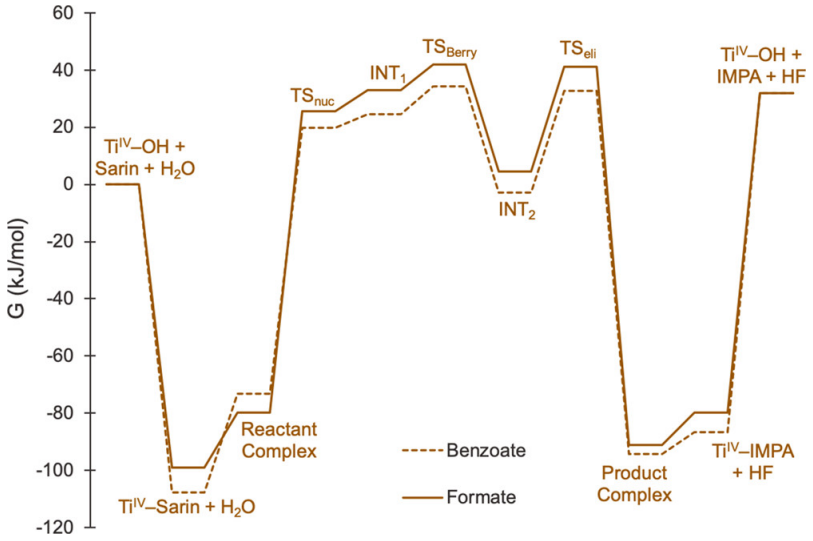


Figure 5.4. Comparison of the reaction free-energy profiles for gas-phase sarin hydrolysis catalyzed by the formate (solid line) and benzoate (dashed line) cluster models of Ti^{IV}–NU-1000.

The reaction free-energy profile for the catalytic cycle is shown in Figure 5.4. As mentioned before, we calculated the full mechanism for sarin hydrolysis on Ti^{IV}–NU-1000 using

both formate and benzoate linkers near the active site to test the effect of model size on the reaction energetics. Comparing the two profiles in Figure 5.4, we see that the results are quite similar for the two models. More explicitly, the free-energy barrier from the bound sarin molecule to the Berry pseudorotation transition state, $\Delta G = G(\mathbf{TS_{Berry}}) - G(\mathbf{Ti^{IV}–Sarin}) - G(\mathbf{H_2O})$, is calculated as 141.0 and 142.1 kJ/mol for the formate and benzoate models, respectively. The HF elimination barrier, $\Delta G = G(\mathbf{TS_{eli}}) - G(\mathbf{INT_2})$, is 36.7 and 35.6 kJ/mol for the formate and benzoate models, respectively. Finally, the product desorption barrier, $\Delta G = G(\mathbf{Ti^{IV}–OH}) + G(\mathbf{IMPA}) + G(\mathbf{HF}) - G(\mathbf{Product\ Complex})$, is 122.9 and 126.2 kJ/mol for the formate and benzoate models, respectively.

Since these barriers differ by only a few kJ/mol and are within the DFT error, our results suggest that the replacement of benzoate linkers by formate capping groups does not strongly affect the reaction energetics. This observation validates that our smaller formate model is acceptable for screening purposes. In addition to being relatively accurate, the formate model also considerably reduces the computational expense for screening across a wide variety of catalysts. The reason for the striking similarity between the differently sized models can be attributed to the effective isolation of the single-atom active site. The deposited metal is located far enough from the rest of the Zr_6 node support and linkers such that the inclusion of benzoate linkers does not affect the reaction, in contrast to the differences observed between formate and benzoate models when nerve agent hydrolysis takes place on unfunctionalized Zr_6 nodes.[198,207,221] Further, by using the smaller formate groups, we essentially limit the influence of dispersion interactions with the linkers[139] and only focus on the effects of metal identity and oxidation state on the reaction energetics.

5.3.3 Effect of Transition Metal on Key Barriers

To explore the effect of metal identity and oxidation state on the reaction mechanism described above for Ti^{IV}–NU-1000, we substituted Ti^{IV} in the single-atom catalyst model with the other 32 metals (omitting Re^{IV}, Os^{IV}, and Pt^{IV} based on the initial analysis of binding energies). Because the overall pathway is rather large, requiring the calculation of 10 species, we first sought to narrow down the number of possible candidates. Examining the profile in Figure 5.4 shows that there are two dominant barriers throughout the mechanism – nucleophilic attack and product desorption. The calculation of transition states is relatively expensive and more nuanced than optimizing minima structures. However, a careful analysis of Figure 5.4 reveals that **INT_1** is energetically and structurally similar to **TS_{nuc}**, and so the free-energy barrier to form **INT_1** serves as a good predictor of the nucleophilic attack barrier. The validity of this argument is proven in a later section, where we show a Brønsted–Evans–Polyani (BEP) relationship between the **TS_{nuc}** activation barrier and the reaction energy to form **INT_1**. Thus, we plotted $\Delta G_{\mathrm{INT}_1}$ versus $\Delta G_{\mathrm{IMPA,des.}}$ (Figure 5.5) to identify the most promising candidates to study in more detail. Here, $\Delta G_{\mathrm{INT}_1} = G(\mathbf{INT_1}) - G(\mathbf{Ti^{IV}}\text{–}\mathbf{Sarin}) - G(\mathbf{H_2O})$ and $\Delta G_{\mathrm{IMPA,des.}} = G(\mathbf{Ti^{IV}}\text{–}\mathbf{OH}) + G(\mathbf{IMPA}) - G(\mathbf{Ti^{IV}}\text{–}\mathbf{IMPA})$. Note that after identifying the most promising candidates below, the nucleophilic attack transition state (**TS_{nuc}**) is explicitly calculated in all following sections as part of the full hydrolysis mechanisms for the most promising catalysts.

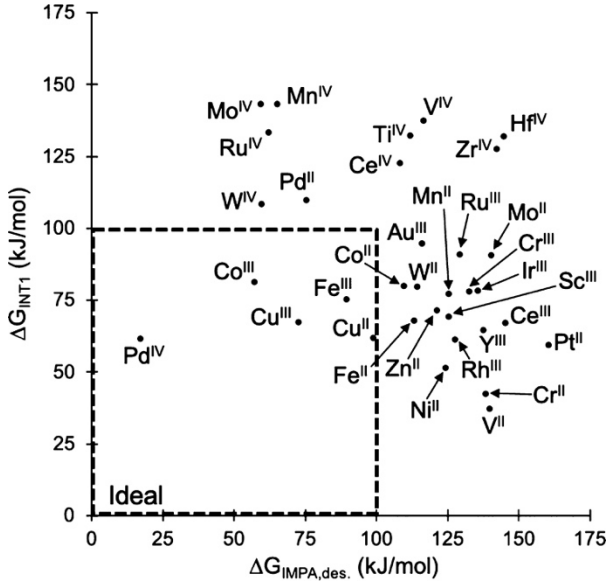


Figure 5.5. Effect of transition-metal identity and oxidation state on the free-energy barrier to form **INT1** and the free-energy barrier to desorb the IMPA product from the active site of M–NU-1000 systems, where the optimal metals are located closest to the origin. As a visual aid, we classify metals as "ideal" if they exhibit barriers lower than 100 kJ/mol.

In Figure 5.5, the optimal metals are located closest to the origin, showing low barriers for both nucleophilic attack and product desorption. As a visual aid, we classify metals as "ideal" here if they exhibit barriers lower than 100 kJ/mol, although this choice is arbitrary. The dashed box highlights Pd^{IV}, Cu^{III}, Co^{III}, Fe^{III}, and Cu^{II} as potentially good candidates, followed by W^{IV}, Co^{II}, and Pd^{II}. By removing the metal labels, which make it somewhat confusing to analyze the overall trends, we observe a pseudo-Pareto front (see Figure D.4). In other words, there is a trade-off between the energy barrier for nucleophilic attack and the product desorption energy such that no metal has perfectly low barriers for both steps.

In Table 5.1, we list the larger of the two barriers (either $\Delta G_{\mathrm{INT}_1}$ or $\Delta G_{\mathrm{IMPA,des.}}$) in order of increasing value to convey the magnitude of the key barriers for each M–NU-1000 system. The table cells are shaded white and green if the larger free-energy barrier corresponds to $\Delta G_{\mathrm{INT}_1}$ and $\Delta G_{\mathrm{IMPA,des.}}$, respectively. The values for both barriers from Figure 5.5 are listed in Table D.4. For 23 out of the 33 M–NU-1000 systems, the larger barrier is for desorption of the IMPA product

from the active site. This suggests that product inhibition may be a more general problem for most metals across the periodic table, regardless of the oxidation state, even though we consider IMPA binding in a monodentate fashion.

Table 5.1. Key Barriers for M–NU-1000 Systems

Metal	Larger Barrier[a] (kJ/mol)
Pd^{IV}	61.7
Cu^{III}	72.6
Co^{III}	81.3
Fe^{III}	89.4
Cu^{II}	98.8
W^{IV}	108.4
Co^{II}	109.6
Pd^{II}	109.8
Fe^{II}	113.3
W^{II}	114.1
Au^{III}	115.9
Zn^{II}	121.1
Ce^{IV}	122.7
Ni^{II}	124.1
Mn^{II}	125.3
Sc^{III}	125.4
Rh^{III}	127.6
Ru^{III}	129.3
Ti^{IV}	132.2
Cr^{III}	132.5
Ru^{IV}	133.3
Ir^{III}	135.6
Y^{III}	137.5
V^{IV}	137.5
Cr^{II}	138.3
V^{II}	139.6
Mo^{II}	140.3
Zr^{IV}	142.3
Mo^{IV}	143.2
Mn^{IV}	143.2
Hf^{IV}	144.8
Ce^{III}	145.3
Pt^{II}	160.4

[a]Larger barrier means the free-energy barrier to form **INT$_1$** or the free-energy barrier to desorb the IMPA product, whichever is larger. Cells shaded white and green signify that the value corresponds to $\Delta G_{\mathrm{INT}_1}$ and $\Delta G_{\mathrm{IMPA,des.}}$, respectively.

On the basis of the data in Figure 5.5 and Table 5.1, we decided to calculate the full pathways for gas-phase sarin hydrolysis for 19 of the 36 initially considered M–NU-1000 systems (Pd^{IV}, Cu^{III}, Co^{III}, Fe^{III}, Cu^{II}, W^{IV}, Co^{II}, Pd^{II}, Fe^{II}, W^{II}, Au^{III}, Zn^{II}, Ce^{IV}, Ni^{II}, Mn^{II}, Sc^{III}, Rh^{III}, Ru^{III}, and Ti^{IV}), representing a broad range of metals in different oxidation states. We decided to calculate the entire mechanism (including transition states) for such a large number of candidates because, as will be argued in the following sections, it is important to consider more than a single step when screening across a wide variety of catalysts. In other words, we wanted to include any potentially promising single-atom catalysts, even if they exhibited slightly large $\Delta G_{\mathrm{INT}_1}$ or $\Delta G_{\mathrm{IMPA,des.}}$ barriers. The remaining 17 M–NU-1000 systems showed free-energy barriers ranging from ∼130 to 160 kJ/mol and can be safely excluded, as the large barriers make them nonideal systems for this reaction.

5.3.4 Comparison of Full Pathways

With regard to the full mechanism for gas-phase sarin hydrolysis on a M–NU-1000 cluster, as depicted in Scheme 5.2, there are four elementary steps that must be considered when assessing the catalytic activity:

1. Nucleophilic attack, $\Delta G^{\ddagger}_{\mathrm{nuc}} = G(\mathbf{TS_{nuc}}) - G(\mathbf{Reactant\ Complex})$

2. Berry pseudorotation, $\Delta G^{\ddagger}_{\mathrm{Berry}} = G(\mathbf{TS_{Berry}}) - G(\mathbf{INT_1})$

3. HF elimination, $\Delta G^{\ddagger}_{\mathrm{eli}} = G(\mathbf{TS_{eli}}) - G(\mathbf{INT_2})$

4. IMPA desorption, $\Delta G_{\mathrm{IMPA,des.}} = G(\mathbf{M–OH}) + G(\mathbf{IMPA}) - G(\mathbf{M–IMPA})$

First, we analyze the free-energy barrier for each elementary step for the remaining 19 M–NU-1000 systems. As mentioned in the Computational Details section, all possible spin states for each system were considered. The structures, energies, and frequencies of the lowest-energy spin

states were used to calculate the Gibbs free energies for each species in a mechanism at T = 298.15 K. Note that, for application in personal protective equipment under realistic conditions, room temperature is the relevant temperature. The SCF energies and enthalpies corresponding to the most stable spin structures are given in Appendix D along with the SCF energies for higher-energy spin states. The free-energy barriers for the four elementary steps in the reaction are plotted in Figure D.5, and the values are listed in Table D.5.

The results indicate that the highest free-energy barriers along the reaction pathway vary for the 19 catalyst candidates. For M = Co^{III}, Cu^{III}, Pd^{IV}, Ce^{IV}, and W^{IV}, the largest barrier is for nucleophilic attack. For M = Sc^{III}, Ti^{IV}, Mn^{II}, Fe^{II}, Fe^{III}, Co^{II}, Ni^{II}, Cu^{II}, Zn^{II}, Ru^{III}, Rh^{III}, Pd^{II}, W^{II}, and Au^{III}, the largest barrier corresponds to IMPA desorption. Further, for M = Ti^{IV}, Cu^{III}, and Fe^{III}, the differences in free-energy barriers for nucleophilic attack and IMPA desorption are relatively small (6.3, 7.9, and 9.0 kJ/mol, respectively). Although it is clear that nucleophilic attack and IMPA desorption are the most influential elementary steps, it is impossible to state which single-atom catalyst model is the most promising based on this data alone. If we were instead considering only a small number of catalysts, where there was a single elementary step that dominated the reaction kinetics for each metal, then we could employ the philosophy of a "rate-determining step" (RDS)[262] to argue which catalyst is the most promising. Conversely, using the concept of an RDS for our study would inherently ignore barriers that are only slightly lower in energy than the largest barrier, resulting in misleading or erroneous predictions of the overall catalyst activities. Thus, it is important to consider more than a single reaction step when screening a large number of potential catalysts.[251] To illustrate this point, we discuss Pd^{II}–NU-1000. Figure D.5 and Table D.5 show that Pd^{II} has the smallest RDS barrier ($\Delta G_{IMPA,des.}$ = 75.2 kJ/mol) and might naively be predicted as the most promising catalyst for gas-phase sarin hydrolysis. However,

once we incorporate the relative free energies of all intermediates and transition states in the catalytic cycle into our kinetic analysis, we show that the opposite is true – surprisingly, Pd^{II}–NU-1000 is one of the worst catalysts out of the transition metals we examined.

Before moving on to address this issue, we also calculated the full pathway for gas-phase sarin hydrolysis on bare, unfunctionalized NU-1000 for comparison. To compute the reaction on unfunctionalized NU-1000 (hereafter referred to simply as NU-1000), we modeled the bottom four linkers as formate groups and used benzoate groups for the four linkers located around the active site. The bottom formate groups are located far away from the active site and do not directly interact with the reactants, and so this simplification should have a negligible effect on the reaction occurring on the opposite node face. However, in contrast to the functionalized M–NU-1000 systems, we considered it important to treat the linkers around the active site as benzoate groups because the linkers are in closer proximity to the reactants for the reaction on NU-1000.[198,207,221] We modeled a "partially dehydrated" node of NU-1000 ($[Zr_6(\mu_3\text{-}O)_4(\mu_3\text{-}OH)_4(OH)_4(OH_2)_3]^{8+}$-$(C_6H_5COO^-)_4(HCOO^-)_4$, Figure D.1b), in which one terminal H_2O moiety has desorbed, generating an open metal site. This node topology, confirmed by experimental infrared spectroscopy and DFT in a prior work,[175] is essential for the catalysis to occur and is commonly used for mechanistic studies.[57,139] We note that the cleavage of a Zr–OH_2 bond requires approximately 84 kJ/mol in free energy. We calculated the general base hydrolysis mechanism for NU-1000, in which sarin and IMPA bind to Zr in a monodentate fashion, for the most direct comparison to the 19 M–NU-1000 systems. An illustration of the proposed catalytic cycle for gas-phase sarin hydrolysis on NU-1000 is given in Figure D.6. Note that the energetics for a bidentate binding mode on NU-1000 is more unfavorable (see Figure D.7).

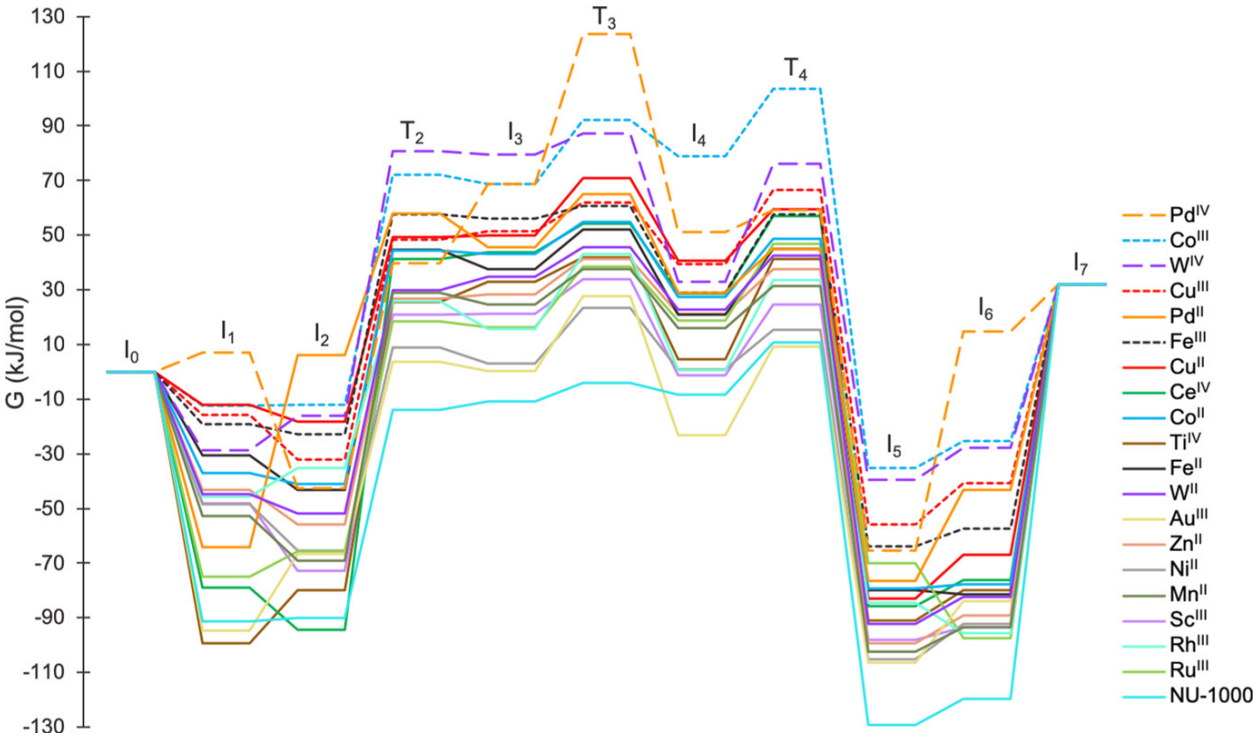


Figure 5.6. Comparison of the reaction free-energy profiles for gas-phase sarin hydrolysis catalyzed by single-atom M–NU-1000 systems as well as unfunctionalized NU-1000. The labels I_k and T_j correspond to the terminology used in the energetic span model for intermediates and transition states, respectively.

A comparison of the reaction free-energy profiles for NU-1000 and the 19 M–NU-1000 systems is shown in Figure 5.6. The individual free-energy profiles for each catalyst model are provided in Figure D.8. The most striking difference between the supported catalysts and NU-1000 is that the energy profile for the latter lies lower than all supported metals for nearly every species along the mechanism. This can be attributed to the strong Lewis acidity of the Zr^{IV} active site in NU-1000, which results in strong binding of all reactive intermediates. More importantly, we observe that all 19 of the single-atom catalysts display lower product desorption energies than NU-1000. This is true regardless of whether the IMPA product binds to NU-1000 in a monodentate or bidentate fashion, where the product desorption energy is $\Delta G_{\text{IMPA,des.}}$ = 151.7 and 168.4 kJ/mol for monodentate and bidentate modes, respectively (Figure D.7). This agrees with Troya's previous results[57] showing stronger binding energies for bidentate hydrolysis products. This is an

encouraging result, considering one primary aim of this study was to explore materials that can overcome the product inhibition previously observed during the gas-phase hydrolysis of nerve agents on Zr_6 nodes. We now integrate the relative free energies of all intermediates and transition states into our kinetic analysis for a more detailed comparison of all 20 catalysts.

5.3.5 Relative TOFs

The TOF of a catalytic reaction measures the number of cycles completed per active site per unit time, which is a useful quantification of the catalyst efficiency.[263] The ESM developed by Kozuch and Shaik[264,265] translates the energy profile derived from electronic structure calculations to the TOF obtained in experiments. A fundamental consequence of the ESM is that the kinetics of a catalytic cycle is determined by the relative free energies of all transition states and intermediates in the mechanism, instead of being controlled by a single reaction step. Using a DFT-calculated free-energy profile, the TOF may be computed as

$$TOF = \frac{k_B T}{h} \frac{exp\left(-\frac{\Delta G_r}{RT}\right) - 1}{\sum_{j=1,k=0}^{N} exp\left(\frac{T_j - I_k - \delta G'_{j,k}}{RT}\right)} \tag{5.2}$$

$$\delta G'_{j,k} = \begin{cases} \Delta G_r \text{ if } T_j \text{ follows } I_k \\ 0 \text{ if } T_j \text{ precedes } I_k \end{cases} \tag{5.3}$$

where k_B is the Boltzmann's constant, T is the temperature, h is the Planck's constant, R is the ideal gas constant, ΔG_r is the overall reaction free energy, N is the number of steps in the cycle, T_j is the free energy of transition state j (j = 1, 2, ...), and I_k is the free energy of intermediate k (k = 0, 1, ...). The largest $\delta G = (T_j - I_k - \delta G'_{j,k})$ is defined as the energetic span of the cycle, and the corresponding combination of states (j and k) are defined as the TOF-determining transition state (TDTS) and TOF-determining intermediate (TDI), respectively. Fundamentally, a smaller

energetic span results in a higher TOF and faster catalysis. In eq 5.2, the numerator is independent of the catalyst, whereas the denominator strongly depends on the catalyst properties.

Given the inherent uncertainty in DFT,[136] it is difficult to obtain reaction rate constants for direct comparison to experiments. For example, a small inaccuracy in the calculation of the energetic span results in an exponential error in the TOF calculation, and so it is challenging to predict quantitatively accurate absolute TOFs through DFT.[264] However, because of error compensation, relative TOF values can be useful for quantitatively comparing catalyst efficiencies. An additional reason to focus on relative TOFs is that the application of the ESM to compute TOFs relies on the assumption of ideal conditions without side reactions.[251,263] Whereas the development of a microkinetic model would be a more rigorous method to obtain TOFs, performing such an analysis would be computationally expensive for such a large-scale catalyst screening. Further, we note that eq 5.2 is generalized such that the concentrations or pressures of all reactant and product species are assumed to be 1 molar or 1 bar, which may not represent realistic battlefield conditions. Because these concentration effects would likely be consistent across all catalysts, our main focus in this section is on the relative performance of each catalyst. Thus, the discussion below is centered on the calculation of relative TOFs.

We used the procedures established by Ye et al.,[251] reproduced here for clarity, to apply eq 5.2 to calculate TOFs for all 20 catalysts. There are no corresponding transition states associated with the adsorption and desorption reaction steps throughout the mechanism. Thus, if no transition state is shown between two connecting intermediates in the free-energy profiles, then the free energy of that "missing" transition state is set equal to the energy of the higher of the two surrounding intermediates. Also, if a transition-state free energy is lower than the energy of either

of the intermediates it connects, then its energy is raised to that of the higher of the two surrounding intermediates.

Table 5.2. Calculated Relative TOFs for Gas-Phase Sarin Hydrolysis at 298.15 K and 1 atm

System	Relative TOF[a]	TDTS, TDI
Cu^{III}	1.00	T_4, I_5
Fe^{III}	2.66×10^{-1}	T_3, I_5
W^{IV}	1.65×10^{-1}	T_3, I_5
Ni^{II}	9.26×10^{-2}	T_3, I_5
Sc^{III}	2.04×10^{-2}	T_3, I_5
Au^{III}	9.66×10^{-3}	T_3, I_5
Ti^{IV}	8.88×10^{-3}	T_3, I_5
Fe^{II}	7.55×10^{-3}	T_3, I_6
Co^{II}	5.90×10^{-3}	T_3, I_5
W^{II}	1.60×10^{-3}	T_3, I_5
Co^{III}	1.39×10^{-3}	T_4, I_5
Rh^{III}	1.34×10^{-3}	T_3, I_6
NU-1000	8.48×10^{-4}	T_4, I_5
Mn^{II}	8.06×10^{-4}	T_3, I_5
Zn^{II}	5.95×10^{-4}	T_3, I_5
Pd^{II}	4.16×10^{-4}	T_3, I_5
Ce^{IV}	2.34×10^{-4}	T_4, I_5
Ru^{III}	1.68×10^{-4}	T_4, I_6
Cu^{II}	3.63×10^{-6}	T_3, I_5
Pd^{IV}	2.20×10^{-12}	T_3, I_5

[a]All values are relative to 2.03×10^{-9} s^{-1}.

The relative TOFs for NU-1000 and the 19 M–NU-1000 systems are listed in Table 5.2, where all values are relative to the largest calculated |TOF| (2.03×10^{-9} s^{-1} for Cu^{III}–NU-1000). The highest relative TOFs are calculated to be 1.00, 0.266, 0.165, 0.0926, and 0.0204 for Cu^{III}, Fe^{III}, W^{IV}, Ni^{II}, and Sc^{III}, respectively. After an exhaustive analysis, we could not find a single, definitive explanation for the observed activity across the series of metals, a similar conclusion drawn in related studies.[252] Rather, there is most likely a complex relationship between slight changes in the electronic structure, coordination geometry, and so forth that affects the catalyst efficiencies. Nonetheless, NU-1000 has a relative TOF of 8.48×10^{-4}, which means that this

method predicts Cu^{III}–NU-1000 to be over 1000 times more active than NU-1000 for gas-phase sarin hydrolysis. Moreover, we observe that 12 of the single-atom transition metal catalysts are predicted to be faster than NU-1000 for this reaction. This is most likely due to the lower product desorption energies calculated for the single-metal systems, as discussed in the previous section. Also, it is interesting to note that Pd^{II} exhibited a relative TOF of 4.16×10^{-4}, which is remarkably small, considering it showed the smallest RDS barrier (see Figure D.5 and Table D.5). This serves as a caution against using the RDS philosophy as the sole predictor of catalyst activity and enforces the notion that it is vital to incorporate the relative free energies of all intermediates and transition states into a proper kinetic analysis. We also found that the species that dominate the calculated TOFs (the TDTS and TDI) vary across the 20 catalyst systems, as seen in Table 5.2. In various combinations, states T_3, T_4, I_5, and I_6 (**TS_{Berry}**, **TS_{eli}**, **Product Complex**, and **M–IMPA + HF**, respectively) are predicted to be the most influential states on the catalytic efficiencies.

For an idea of how sensitive the absolute TOFs are to changes in energy, we mention that a 15 kJ/mol error in the free-energy value of the TDI for Cu^{III}–NU-1000 (which is a realistic uncertainty estimate for reaction free energies in transition-metal systems using DFT[251]) changes its absolute TOF by a factor of 213 (see Figure D.9). Although we noted that accurate absolute TOFs are currently unattainable through DFT calculations, we must emphasize that our predicted absolute TOFs are extremely small for these gas-phase reactions. For context, the experimental TOF is 1.3×10^{-2} s^{-1} for the hydrolysis of a related nerve agent, soman, by NU-1000 using a *N*-ethylmorpholine buffered solution.[1] Recent experiments were conducted to evaluate the gas-phase decontamination efficiency of three Zr-MOFs, including NU-1000, toward the nerve agents soman and VX under ambient conditions and showed considerably slower hydrolysis rates compared to rates in basic buffer solutions,[56] in agreement with our computational results. Thus, even though

we successfully identified multiple single-atom catalysts that have higher TOFs than NU-1000, the gas-phase hydrolysis of sarin is still quite slow (under the current modeling framework). One reason for the small absolute TOFs is that the overall reaction (Scheme 5.1) is thermodynamically unfavorable in the gas phase (ΔG_r = 31.9 kJ/mol). Nonetheless, our calculations predict that deposition of single-atom transition metal catalysts may be a viable strategy for improving gas-phase nerve agent hydrolysis by Zr_6 MOFs. The ranking of single-atom catalysts in Table 5.2 can also be used to suggest which metals should be tested first in future experiments. Once these experiments are performed, we will be able to gauge the accuracy of our computational predictions of the relative TOFs and refine our kinetic analysis, if necessary. However, it is a difficult experimental challenge to quantify the kinetics of gas-phase decontamination. To the best of our knowledge, only one such method has been reported. Wang et al.[56] used a digestion method to release unreacted nerve agents and tightly bound hydrolysis products from Zr-MOF catalysts to measure gas-phase reaction conversion profiles. Similar methods, in addition to accurate measurements of concentrations of catalyst active sites and adsorbed water under humid conditions, would be required to assess the reactivity of our proposed NU-1000-supported single-atom catalysts for gas-phase sarin hydrolysis.

As mentioned in the Computational Details section, we considered metals that have previously been deposited using AIM or SIM as well as novel metal/oxidation state combinations to guide future experiments and to explore periodic trends. However, from an experimental point of view, it is important to note that several metal oxidation states that we considered in this study may be challenging to stabilize. For example, it is possible that the W^{IV} single-atom catalyst would oxidize to W^{V} during the deposition process, similar to observations made when depositing a V^{IV} precursor on NU-1000 nodes.[241] This oxidation process would likely affect the resulting catalyst

activity and is thus an important consideration when selecting which metals to test in future experiments. Also, for realistic application under humid atmospheric conditions, it is essential for the node-supported metals to be relatively water-stable. This being said, several metals that were predicted to have higher TOFs than NU-1000 have previously been deposited on NU-1000 nodes and were found stable under neutral to basic conditions, including Ti^{IV},[266] Fe^{III},[266,267] Co^{II},[255,266] and Ni^{II}[255,257,266] (listed in order of decreasing stability).

In terms of methods to potentially improve these TOFs, we note that increases in temperature would only modestly improve the catalytic efficiency of these systems. For example, the |TOF| is predicted as 3.79×10^{-8} s^{-1} at 318.15 K for Cu^{III}–NU-1000 (Table D.6 and Figure D.10). Much higher temperatures may lead to distortion and/or degradation of the active sites. Also, in our current study, we only considered one explicit water molecule that acts as the nucleophile. The inclusion of multiple explicit water molecules near the active site, mimicking the reservoir of adsorbed water molecules found in MOF pores under humid conditions,[56,268] may facilitate product desorption,[269] stabilize the reaction energetics through H-bonding interactions, and/or enable alternative lower-energy pathways through proton-shuttling networks. It is unclear exactly how much the presence of atmospheric water would increase the TOFs, but we are currently investigating this matter in more detail. Recent gas-phase experiments performed under varying relative humidity conditions showed that large amounts of moisture present in Zr-MOFs can result in a moderate to significant enhancement of nerve agent hydrolysis rates.[56] Further improvements to the TOF could also come from base heterogenization, as was recently studied using a Zr_6 MOF/amino-functionalized dendrimer or polymer mixture for the hydrolysis of a nerve agent simulant.[55]

5.3.6 Catalytic Descriptor Relationships

Last, we discuss the calculation of catalytic descriptor relationships to estimate transition-state barriers for the final 19 M–NU-1000 systems. As mentioned, the calculation of transition states can be relatively expensive for larger systems. Identifying relationships to accurately predict transition-state free-energy barriers can facilitate DFT screening studies such as ours, especially in the early stages of selecting which catalysts to explore. More importantly, these relationships can help visualize trends in catalytic activity and can provide valuable insight when screening across diverse metals in varying oxidation states.

Linear BEP relationships can be used to estimate activation barriers based on the calculation of the reaction energy.[270,271] Figure 5.7a shows a plot of the activation free energy, $\Delta G^{\ddagger}_{\mathrm{nuc}} = G(\mathbf{TS_{nuc}}) - G(\mathbf{Reactant\ Complex})$, as a function of the reaction free energy, $\Delta G_{\mathrm{r,nuc}} = G(\mathbf{INT_1}) - G(\mathbf{Reactant\ Complex})$, for nucleophilic attack. The transition-state barrier was found to vary linearly with the free energy of reaction for nucleophilic attack for the 19 M–NU-1000 systems. A linear fit gives the BEP-type relationship

$$\Delta G^{\ddagger}_{\mathrm{nuc}} = 0.75\Delta G_{\mathrm{r,nuc}} + 21.02 \text{ kJ/mol}$$

with $R^2 = 0.86$. Pd^{IV}–NU-1000 is an outlier that deviates from the overall trend; excluding Pd^{IV} leads to a linear fit for the other 18 systems of

$$\Delta G^{\ddagger}_{\mathrm{nuc}} = 0.83\Delta G_{\mathrm{r,nuc}} + 15.57 \text{ kJ/mol}$$

with $R^2 = 0.96$. However, since we showed that the concept of a RDS is inadequate for our screening purposes where multiple elementary steps have similar energetic barriers, accurately predicting one activation barrier using this BEP-type relationship will not necessarily result in an accurate depiction of the overall catalyst efficiency for various metals. Nonetheless, one important conclusion from this BEP-type relationship is that our initial decision to use the reaction energy to

form **INT$_1$** as a predictor for the **TS$_{nuc}$** activation barrier to narrow down the 36 M–NU-1000 candidates was a reasonable choice.

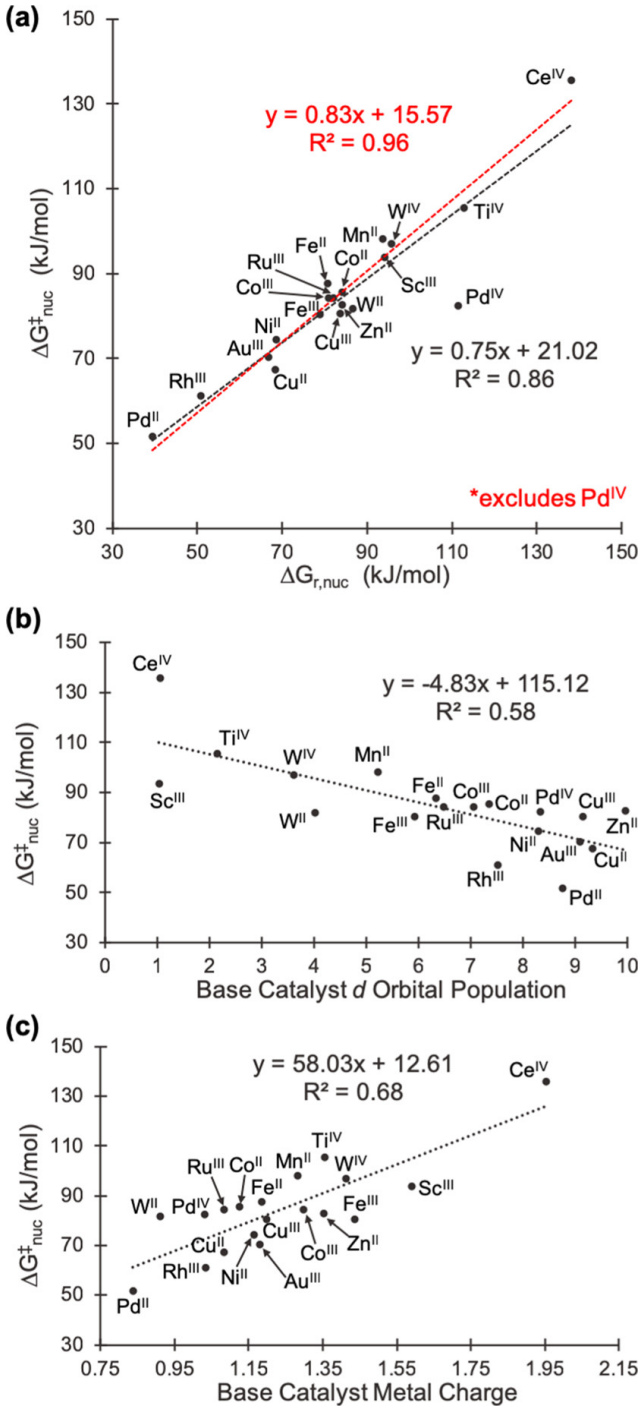


Figure 5.7. Calculated activation free energy ($\Delta G^{\ddagger}_{nuc}$) for nucleophilic attack as a function of (a) the reaction free energy ($\Delta G_{r,nuc}$) for nucleophilic attack, (b) the calculated NBO population of the d orbitals of the base catalyst (**M–OH**) metal atom, and (c) the calculated NBO charge of the base catalyst (**M–OH**) metal atom for 19 M–NU-1000 systems. The dashed lines represent linear regression relationships.

To explore possible electronic properties that are related to the catalytic activity, we also compared the activation free energy for nucleophilic attack to various molecular descriptors. All molecular orbital parameters were calculated with the natural bond orbital (NBO) method[119] in Gaussian 09. Figure 5.7b shows a plot of $\Delta G_{\mathrm{nuc}}^{\ddagger}$ as a function of the calculated NBO population of the d orbitals of the base catalyst (**M–OH**) metal atom for the 19 M–NU-1000 systems. A linear fit gives a coefficient of determination of $R^2 = 0.58$, where the $\mathbf{TS_{nuc}}$ free-energy barrier decreases with increasing electron population in the d orbitals. Figure 5.7c shows a plot of $\Delta G_{\mathrm{nuc}}^{\ddagger}$ as a function of the calculated NBO charge of the base catalyst (**M−OH**) metal atom. A linear fit gives $R^2 = 0.68$, where the $\mathbf{TS_{nuc}}$ free-energy barrier increases with increasing charge on the single metal. We also explored other possible molecular descriptors such as bond lengths, the sarin binding energy, the NBO charge on P in the **M–Sarin** species, and the P=O bond order in the **M–Sarin** species but found negligible correlations between these parameters and the nucleophilic attack barriers. Together, the plots in Figure 5.7b,c show that metals with nearly full d orbitals (~8–9 electrons) and lower atomic charges correlate with lower free-energy barriers for nucleophilic attack. Because of the relatively low R^2 values, the activation free energies computed using these molecular descriptors would have low accuracy. Still, the linear relationships are useful for exploring periodic trends for nucleophilic attack on sarin by M–NU-1000 systems. Although it would be more desirable to derive relationships between descriptors and the catalyst TOFs or energetic spans, we note that such correlations are difficult to obtain because these hydrolysis mechanisms are complex and consist of different TOF-determining species (the TDTS and TDI) across the metals studied; catalytic descriptor relationships for a single elementary step are more closely connected to the fundamental chemistry.

5.4 Conclusions

In this work, we utilized DFT to perform a comprehensive screening of single-atom transition-metal catalysts deposited on NU-1000 nodes for the gas-phase hydrolysis of sarin. To enhance MOF-catalyzed nerve agent hydrolysis, previous approaches have focused on either incorporating defect sites or synthesizing MOFs with different metal nodes and/or functionalized ligands.[272] This study represents one of the first reports to specifically consider single-atom catalysts for this application. The reason for exploring single-atom systems was to force the products to bind to the active site in a monodentate fashion, thereby avoiding unfavorable bidentate binding which possibly led to product inhibition in previous studies of gas-phase degradation of nerve agents on Zr_6 nodes. A similar concept was reported in a very recent study in which the monomerization of a Zr-substituted polyoxometalate resulted in the isolation of a single-site Zr that was predicted to be active for the hydrolysis of the nerve agent simulant dimethyl chlorophosphate, leading to a monodentate-bound phosphate product.[273]

Initially, we considered 36 M–NU-1000 systems with metals in +2, +3, and +4 oxidation states. We first analyzed the binding energies of sarin, water, and hydrolysis product IMPA to the metals to evaluate periodic trends. Thirty two out of the 36 systems showed preferential adsorption for sarin over water, which is ideal for realistic applications where competitive adsorption with water would be relevant in a humid atmosphere. M^{IV}–NU-1000 systems exhibited the strongest binding energies for sarin, whereas M^{II}–NU-1000 systems showed the strongest binding of IMPA. We then calculated the full catalytic cycle, using Ti^{IV}–NU-1000 as an example. To identify the most promising candidates to study in more detail, we then substituted Ti^{IV} with the other metals and plotted the free-energy barrier to form **INT$_1$** versus the free-energy barrier to desorb the IMPA product from the active site. This plot showed a tradeoff between the energy barrier for

nucleophilic attack and the product desorption energy such that no metal has perfectly low barriers for both steps. On the basis of this data, we chose the best 19 M–NU-1000 systems (Pd^{IV}, Cu^{III}, Co^{III}, Fe^{III}, Cu^{II}, W^{IV}, Co^{II}, Pd^{II}, Fe^{II}, W^{II}, Au^{III}, Zn^{II}, Ce^{IV}, Ni^{II}, Mn^{II}, Sc^{III}, Rh^{III}, Ru^{III}, and Ti^{IV}) and calculated the full pathways for gas-phase sarin hydrolysis and compared them to bare, unfunctionalized NU-1000. We observed that all 19 of the single-atom catalysts display lower product desorption energies than NU-1000, suggesting that they would experience less product inhibition than the parent MOF. The proposed catalytic cycle consists of four elementary steps: nucleophilic attack, Berry pseudorotation, HF elimination, and IMPA desorption. Our results indicate that the highest free-energy barriers, or "rate-determining steps", vary across the 20 catalysts, proving that it is important to consider more than a single reaction step when screening a large number of catalysts. Thus, we incorporated the relative free energies of all intermediates and transition states into our kinetic analysis and used the energetic span model to calculate relative TOFs to determine the efficiency of each catalyst. The highest relative TOFs were calculated to be 1.00, 0.266, 0.165, 0.0926, and 0.0204 for Cu^{III}, Fe^{III}, W^{IV}, Ni^{II}, and Sc^{III}, respectively. NU-1000 has a relative TOF of 8.48×10^{-4}, which means Cu^{III}–NU-1000 is predicted to be over 1000 times more active than NU-1000 for gas-phase sarin hydrolysis. Although we found that 12 of the single-atom transition-metal catalysts are predicted to be faster than NU-1000, we note that our calculated absolute TOFs are very small, partially due to the unfavorable thermodynamics of the overall reaction in the gas phase. We are currently investigating whether the inclusion of multiple explicit water molecules near the active site, simulating ambient moisture, can facilitate product desorption, stabilize the reaction energetics through H-bonding interactions, and/or enable alternative lower-energy pathways through proton-shuttling networks. Finally, we calculated catalytic descriptor relationships for the final 19 M–NU-1000 systems. We derived a BEP-type

relationship, where the activation free-energy barrier was found to vary linearly with the reaction free energy for nucleophilic attack. Also, we computed various molecular descriptors for the base catalysts and showed that metals with nearly full d orbitals and lower atomic charges correlate with lower free-energy barriers for nucleophilic attack on sarin.

Overall, our calculations predict that deposition of single-atom transition-metal catalysts may be a viable strategy for improving gas-phase nerve agent hydrolysis rates by Zr-based MOFs. Our results establish design principles for enhancing the gas-phase decontamination of CWAs that can be used to guide future experiments. Previous studies of MOFs for nerve agent degradation have almost entirely focused on Zr(IV) frameworks. Going forward, the exploration of non-Zr-based active sites for these reactions[200] could have an important influence in the field of CWA detoxification. Ultimately, an exhaustive investigation into the solid-state catalytic activity of MOFs toward nerve agents in the gas phase, under varying humidity levels[56,61] and in the presence of atmospheric contaminants, is warranted before MOFs may be applied in personal protective equipment such as gas-mask filters.

Chapter 6: Conclusions and Future Directions

6.1 Summary

Due to the high toxicity of CWAs such as nerve agents, simulants are often used in experiments as substitutes for the agents. However, there is little reason to believe that the current simulants used in the literature are optimal predictors of nerve agent reactivity. Thus, in Chapter 2, we performed DFT calculations on the alkaline hydrolysis of over 100 organophosphate molecules to identify improved simulants for the G-series nerve agents soman and sarin, based on low toxicity and similarity to nerve agent hydrolysis energetics and degradation mechanism. This screening highlighted 5 molecules that have nearly identical reaction barriers to the actual agents, while being far less toxic. We also derived QSAR models to determine the most significant molecular descriptors for describing the hydrolysis free energy barriers of these reactions. The optimal QSAR model was subjected to a thorough statistical analysis and validation procedure to confirm its predictive capacity, showing excellent quantitative and ranking accuracy. It was further shown that the model trained on G-series agents can reliably predict energetics for other organophosphate classes as well, including VX. Through these computational insights, experimentalists may be aided in accurately and safely studying these reactions with less toxic simulants.

In Chapter 3, we performed DFT calculations to explore the catalytic hydrolysis of the nerve agent simulant DMNP on the Zr-based MOF NU-1000. The energy barriers computed in this study are in quantitative agreement with previous experimental kinetics data on the same reaction system. A comparison between uncatalyzed aqueous hydrolysis and the MOF-catalyzed reaction revealed the origin of the catalytic effects of NU-1000, where a node Zr atom activates the phosphate center to facilitate nucleophilic attack and a node hydroxyl group stabilizes the

negative charges building up on the leaving group oxygen of DMNP to facilitate the cleavage of its P–O bond, which resembles enzymatic catalysis of similar reactions. The effects of temperature-induced node dehydration and distortion on the catalytic efficiency were also examined, and the results are consistent with experimental findings, where the distorted node of NU-1000 shows an increase in the rate of DMNP hydrolysis compared to the completely hydrated regular form of NU-1000.

As described throughout this dissertation, MOFs built from hexanuclear M(IV) oxide cluster nodes are effective catalysts for nerve agent hydrolysis, where the properties of the active sites on the nodes can strongly influence the reaction energetics. Importantly, the connectivity and metal identity of these M_6 nodes can be easily tuned, offering extensive opportunities for computational screening to predict promising new materials. Thus, in Chapter 4, we used DFT to examine the effects of node topology, connectivity, and metal identity on the binding energies of multiple nerve agents and their corresponding hydrolysis products. By computing an optimization metric based on the relative binding strengths of key hydrolysis reaction species (water, agent, and bidentate-bound products), we predicted optimal M_6 nodes for hydrolyzing specific nerve agent and simulant molecules, where our results are in qualitative agreement with observed experimental trends. This analysis highlighted the notion that no single metal or node topology is optimal for all possible organophosphates, suggesting that MOFs should be selected based on the agent of interest. Using the large amount of data generated from our DFT calculations, we then derived QSAR models to help explain the complex trends observed in the binding energies. Through linear regression, we identified the most important descriptors for describing the binding of nerve agents and their hydrolysis products to M_6 nodes. These results suggested that both molecular and node properties, including both structural and chemical features, collectively contribute to the binding

energetics. By performing a thorough statistical analysis, we showed that our QSAR models are capable of making quantitatively accurate binding energy predictions for nerve agents and their hydrolysis products in a wide variety of M(IV)-MOFs. The insights gained in this study can be used to guide future experiments for the synthesis of MOFs with enhanced catalytic activity for organophosphate hydrolysis.

Some recent studies have suggested that the gas-phase hydrolysis of nerve agents by Zr-based MOFs may be limited by product inhibition resulting from strong bidentate binding of the hydrolysis products to the Zr_6 nodes. A potential method to avoid this problem is to deposit single-atom catalysts on the nodes so that the products bind in a more favorable monodentate fashion. Such catalytic active sites can be characterized with atomic precision, enabling detailed computational mechanistic studies. Thus, in Chapter 5, we used DFT to perform a comprehensive screening of single-atom transition-metal catalysts, in varying oxidation states, deposited on NU-1000 nodes for the gas-phase hydrolysis of the nerve agent sarin. By calculating the complete reaction pathways for M–NU-1000 systems, we discovered that the highest reaction barrier varies between catalysts, highlighting the need to consider more than a single reaction step when screening a large number of diverse materials. Importantly, the single-metal catalysts are predicted to exhibit lower product desorption energies than unfunctionalized NU-1000. By comparing their relative turnover frequencies using the energetic span model, we identified several catalysts that are predicted to be more active than the parent MOF for this reaction. Finally, we explored periodic trends and molecular descriptors for their effect on catalytic activity.

Overall, this dissertation establishes design principles that can be used to guide future experimental testing for the optimization of MOF catalysts for CWA detoxification. As computers grow more powerful and molecular modeling techniques are refined for better accuracy and

efficiency, computational studies will continue to play a vital role in this rapidly growing research area. From providing valuable mechanistic insights into experimental observations to screening candidate materials with potentially better performance for the degradation of nerve agents, computational modeling has great potential to accelerate scientific advancement in the field of MOF-catalyzed nerve agent hydrolysis.

To further advance the findings of this dissertation, we suggest the research directions detailed below.

6.2 Recommendations for Future Research

In Chapter 5, we sought to improve MOF-based catalysts for gas-phase nerve agent hydrolysis by depositing single-atom transition-metal catalysts on the nodes of NU-1000, with the goal of forcing hydrolysis products to weakly bind to the active sites in a monodentate fashion, thus facilitating catalyst regeneration. Since the DFT calculations in that study were performed in the gas phase (i.e., with no solvation model), we considered one explicit water molecule (that would be present in a humid environment) near the single-atom active sites acting as the nucleophile. However, in a realistic battlefield environment, the amount of moisture present in the air would likely be much higher. Further, recent studies have shown that Zr-MOFs are capable of considerable water adsorption under ambient conditions. Using this observation, preliminary solid-phase experiments have shown that the water present inside MOF pores under varying relative humidity conditions is sufficient for catalytic nerve agent hydrolysis to occur, thereby circumventing the need for liquid water.[56,60] Since these results were reported less than a year ago, there is still much to be explored concerning the exact mechanism behind the observed enhancement of gas-phase hydrolysis rates with large amounts of moisture. This presents a great opportunity to study the effects of increased humidity on the reaction mechanism through DFT calculations. As a continuation of our work in

Chapter 5, we suggest the exploration of gas-phase sarin hydrolysis catalyzed by single-atom catalysts supported on the nodes of NU-1000, in addition to various amino-functionalized systems as described below. For cluster-model DFT calculations, the inclusion of multiple explicit water molecules near the active site, mimicking the reservoir of adsorbed water molecules found in Zr-MOF pores under ambient conditions, may facilitate product desorption, stabilize the computed reaction energetics through hydrogen-bonding interactions, and/or enable alternative lower-energy pathways through proton-shuttling networks. Depending on the humidity level chosen for study, it may also be worthwhile to perform periodic DFT calculations to fully capture the effects of large amounts of water present in the pores.

Additionally, amino-functionalized linkers have been previously used in experiments to enhance the rate of nerve agent hydrolysis in basic buffered solutions, where higher pK_as of the amine functionalities (i.e., higher basicity) are predicted to enable faster hydrolysis.[21,51,169] This rate enhancement has also been observed with the use of heterogenized bases (e.g., amine-based dendrimers and polymers) to catalyze organophosphate hydrolysis in pure water[55,274] and under ambient humidity conditions.[60] To date, there has been limited computational analysis of the effects of amino groups on solution-phase hydrolysis,[51,198] which did not consist of a full mechanistic study and kinetic analysis. Thus, we recommend further exploration of this topic, especially in the context of gas-phase nerve agent hydrolysis. To do so, we suggest performing cluster-model DFT calculations using different amino-functionalized linkers on NU-1000 (or other Zr-MOFs), with varying pK_as, to act as Brønsted bases to aid the proton-transfer process during the degradation mechanism, again modeling multiple explicit water molecules near the active site. Periodic calculations could also be used to model larger amine-based dendrimers or polymeric bases in the MOF pores, although this would present many computational challenges.

Combining the hypotheses detailed above, we recommend exploring the synergistic effects of ambient moisture and amino-functionalized linkers, to aid proton-transfer and lower nucleophilic attack barriers, together with node-supported single-atom catalysts to facilitate product desorption. Together, these effects have the potential to significantly accelerate gas-phase nerve agent hydrolysis, with the ultimate goal of achieving sufficiently fast kinetics for practical application in personal protective equipment such as gas masks and protective suits. By using DFT to map out the complete catalytic pathways for the systems listed above and comparing their relative degradation kinetics to those obtained with unfunctionalized Zr-MOFs, some very promising catalysts could be discovered that warrant experimental testing. In general, future computational mechanistic studies should be performed in the gas phase in the presence of moisture, heterogenized base, and atmospheric contaminants to better compare with experiments as the field of MOF-based CWA detoxification moves forward into conducting studies under realistic operating conditions.

References

(1) Mondloch, J. E.; Katz, M. J.; Isley III, W. C.; Ghosh, P.; Liao, P.; Bury, W.; Wagner, G. W.; Hall, M. G.; DeCoste, J. B.; Peterson, G. W.; Snurr, R. Q.; Cramer, C. J.; Hupp, J. T.; Farha, O. K. Destruction of Chemical Warfare Agents Using Metal–Organic Frameworks. *Nat. Mater.* **2015**, *14*, 512–516.

(2) Dubbeldam, D.; Calero, S.; Vlugt, T. J. H. IRASPA: GPU-Accelerated Visualization Software for Materials Scientists. *Mol. Simul.* **2018**, *44*, 653–676.

(3) Walker, P. F. A Century of Chemical Warfare: Building a World Free of Chemical Weapons. In *One Hundred Years of Chemical Warfare: Research, Deployment, Consequences.*; Friedrich, B., Hoffmann, D., Renn, J., Schmaltz, F., Wolf, M., Eds.; Springer: Cham, 2017; pp 379–400.

(4) Kingsley, P.; Barnard, A. Banned Nerve Agent Sarin Used in Syria Chemical Attack, Turkey Says. https://www.nytimes.com/2017/04/06/world/middleeast/chemical-attack-syria.html (accessed Jan 4, 2020).

(5) Hennessy-Fiske, M. Syria has faced international condemnation for using chemical weapons. But Islamic State has them too. http://www.latimes.com/world/middleeast/la-fg-isis-chemical-weapons-2017-story.html (accessed Jan 4, 2020).

(6) Paddock, R. C.; Sang-Hun, C. Kim Jong-nam Was Killed by VX Nerve Agent, Malaysians Say. https://www.nytimes.com/2017/02/23/world/asia/kim-jong-nam-vx-nerve-agent-.html (accessed Jan 4, 2020).

(7) Chai, P. R.; Hayes, B. D.; Erickson, T. B.; Boyer, E. W. Novichok Agents: A Historical, Current, and Toxicological Perspective. *Toxicol. Commun.* **2018**, *2*, 45–48.

(8) Arduini, F.; Palleschi, G. Disposable Electrochemical Biosensor Based on Cholinesterase Inhibition with Improved Shelf-Life and Working Stability for Nerve Agent Detection. In *Portable Chemical Sensors: Weapons Against Bioterrorism*; Nikolelis, D., Ed.; Springer Netherlands: Dordrecht, 2012; pp 261–278.

(9) Jang, Y. J.; Kim, K.; Tsay, O. G.; Atwood, D. A.; Churchill, D. G. Update 1 of: Destruction and Detection of Chemical Warfare Agents. *Chem. Rev.* **2015**, *115*, PR1–PR76.

(10) Costanzi, S.; Machado, J.-H.; Mitchell, M. Nerve Agents: What They Are, How They Work, How to Counter Them. *ACS Chem. Neurosci.* **2018**, *9*, 873–885.

(11) Mercey, G.; Verdelet, T.; Renou, J.; Kliachyna, M.; Baati, R.; Nachon, F.; Jean, L.; Renard, P.-Y. Reactivators of Acetylcholinesterase Inhibited by Organophosphorus Nerve Agents. *Acc. Chem. Res.* **2012**, *45*, 756–766.

(12) Vucinic, S.; Antonijevic, B.; Tsatsakis, A. M.; Vassilopoulou, L.; Docea, A. O.; Nosyrev, A. E.; Izotov, B. N.; Thiermann, H.; Drakoulis, N.; Brkic, D. Environmental Exposure to Organophosphorus Nerve Agents. *Environ. Toxicol. Pharmacol.* **2017**, *56*, 163–171.

(13) Nawała, J.; Jóźwik, P.; Popiel, S. Thermal and Catalytic Methods Used for Destruction of Chemical Warfare Agents. *Int. J. Environ. Sci. Technol.* **2019**, *16*, 3899–3912.

(14) Wiener, S. W.; Hoffman, R. S. Nerve Agents: A Comprehensive Review. *J. Intensive Care Med.* **2004**, *19*, 22–37.

(15) Štengl, V.; Henych, J.; Janoš, P.; Skoumal, M. Nanostructured Metal Oxides for Stoichiometric Degradation of Chemical Warfare Agents. *Rev. Environ. Contam. Toxicol.* **2016**, *236*, 239–258.

(16) Wagner, G. W.; Koper, O. B.; Lucas, E.; Decker, S.; Klabunde, K. J. Reactions of VX, GD, and HD with Nanosize CaO: Autocatalytic Dehydrohalogenation of HD. *J. Phys. Chem. B* **2000**, *104*, 5118–5123.

(17) Wagner, G. W.; Bartram, P. W.; Koper, O.; Klabunde, K. J. Reactions of VX, GD, and HD with Nanosize MgO. *J. Phys. Chem. B* **1999**, *103*, 3225–3228.

(18) Wagner, G. W.; Procell, L. R.; O'Connor, R. J.; Munavalli, S.; Carnes, C. L.; Kapoor, P. N.; Klabunde, K. J. Reactions of VX, GB, GD, and HD with Nanosize Al2O3. Formation of Aluminophosphonates. *J. Am. Chem. Soc.* **2001**, *123*, 1636–1644.

(19) Wagner, G. W.; Peterson, G. W.; Mahle, J. J. Effect of Adsorbed Water and Surface Hydroxyls on the Hydrolysis of VX, GD, and HD on Titania Materials: The Development of Self-Decontaminating Paints. *Ind. Eng. Chem. Res.* **2012**, *51*, 3598–3603.

(20) Bandosz, T. J.; Laskoski, M.; Mahle, J.; Mogilevsky, G.; Peterson, G. W.; Rossin, J. A.; Wagner, G. W. Reactions of VX, GD, and HD with Zr(OH)4: Near Instantaneous Decontamination of VX. *J. Phys. Chem. C* **2012**, *116*, 11606–11614.

(21) Moon, S.-Y.; Wagner, G. W.; Mondloch, J. E.; Peterson, G. W.; DeCoste, J. B.; Hupp, J. T.; Farha, O. K. Effective, Facile, and Selective Hydrolysis of the Chemical Warfare Agent VX Using Zr6-Based Metal–Organic Frameworks. *Inorg. Chem.* **2015**, *54*, 10829–10833.

(22) Liu, Y.; Howarth, A. J.; Vermeulen, N. A.; Moon, S.-Y.; Hupp, J. T.; Farha, O. K. Catalytic Degradation of Chemical Warfare Agents and Their Simulants by Metal-Organic Frameworks. *Coord. Chem. Rev.* **2017**, *346*, 101–111.

(23) Prokop, Z.; Opluštil, F.; DeFrank, J.; Damborský, J. Enzymes Fight Chemical Weapons. *Biotechnol. J.* **2006**, *1*, 1370–1380.

(24) Bigley, A. N.; Xiang, D. F.; Ren, Z.; Xue, H.; Hull, K. G.; Romo, D.; Raushel, F. M. Chemical Mechanism of the Phosphotriesterase from Sphingobium Sp. Strain TCM1, an Enzyme Capable of Hydrolyzing Organophosphate Flame Retardants. *J. Am. Chem. Soc.* **2016**, *138*, 2921–2924.

(25) Katz, M. J.; Mondloch, J. E.; Totten, R. K.; Park, J. K.; Nguyen, S. T.; Farha, O. K.; Hupp, J. T. Simple and Compelling Biomimetic Metal–Organic Framework Catalyst for the Degradation of Nerve Agent Simulants. *Angew. Chemie - Int. Ed.* **2014**, *53*, 497–501.

(26) Hoskin, F. C.; Roush, A. H. Hydrolysis of Nerve Gas by Squid-Type Diisopropyl Phosphorofluoridate Hydrolyzing Enzyme on Agarose Resin. *Science* **1982**, *215*, 1255–1257.

(27) Peng, Y.; Krungleviciute, V.; Eryazici, I.; Hupp, J. T.; Farha, O. K.; Yildirim, T. Methane Storage in Metal–Organic Frameworks: Current Records, Surprise Findings, and Challenges. *J. Am. Chem. Soc.* **2013**, *135*, 11887–11894.

(28) Mason, J. A.; Veenstra, M.; Long, J. R. Evaluating Metal–Organic Frameworks for Natural Gas Storage. *Chem. Sci.* **2014**, *5*, 32–51.

(29) Li, J.-R.; Sculley, J.; Zhou, H.-C. Metal–Organic Frameworks for Separations. *Chem. Rev.* **2012**, *112*, 869–932.

(30) Hu, Z.; Deibert, B. J.; Li, J. Luminescent Metal–Organic Frameworks for Chemical Sensing and Explosive Detection. *Chem. Soc. Rev.* **2014**, *43*, 5815–5840.

(31) Ha, S.; Lee, M.; Seo, H. O.; Song, S. G.; Kim, K.; Park, C. H.; Kim, I. H.; Kim, Y. D.; Song, C. Structural Effect of Thioureas on the Detection of Chemical Warfare Agent Simulants. *ACS Sensors* **2017**, *2*, 1146–1151.

(32) Stassen, I.; Bueken, B.; Reinsch, H.; Oudenhoven, J. F. M.; Wouters, D.; Hajek, J.; Van Speybroeck, V.; Stock, N.; Vereecken, P. M.; Van Schaijk, R.; De Vos, D.; Ameloot, R. Towards Metal–Organic Framework Based Field Effect Chemical Sensors: UiO-66-NH2 for Nerve Agent Detection. *Chem. Sci.* **2016**, *7*, 5827–5832.

(33) Vilela, S. M. F.; Salcedo-Abraira, P.; Colinet, I.; Salles, F.; de Koning, M. C.; Joosen, M. J. A.; Serre, C.; Horcajada, P. Nanometric MIL-125-NH2 Metal–Organic Framework as a Potential Nerve Agent Antidote Carrier. *Nanomaterials* **2017**, *7*, 321.

(34) Yang, D.; Gates, B. C. Catalysis by Metal Organic Frameworks: Perspective and Suggestions for Future Research. *ACS Catal.* **2019**, *9*, 1779–1798.

(35) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal−Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26*, 6185–6192.

(36) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64*, 5985–5998.

(37) Howarth, A. J.; Peters, A. W.; Vermeulen, N. A.; Wang, T. C.; Hupp, J. T.; Farha, O. K. Best Practices for the Synthesis, Activation, and Characterization of Metal−Organic Frameworks. *Chem. Mater.* **2017**, *29*, 26–39.

(38) Zhou, H.-C.; Long, J. R.; Yaghi, O. M. Introduction to Metal–Organic Frameworks. *Chem. Rev.* **2012**, *112*, 673–674.

(39) Moghadam, P. Z.; Islamoglu, T.; Goswami, S.; Exley, J.; Fantham, M.; Kaminski, C. F.; Snurr, R. Q.; Farha, O. K.; Fairen-Jimenez, D. Computer-Aided Discovery of a Metal–Organic Framework with Superior Oxygen Uptake. *Nat. Commun.* **2018**, *9*, 1378.

(40) Kwon, O.; Kim, J. Y.; Park, S.; Lee, J. H.; Ha, J.; Park, H.; Moon, H. R.; Kim, J. Computer-Aided Discovery of Connected Metal-Organic Frameworks. *Nat. Commun.* **2019**, *10*, 3620.

(41) Zou, R.; Zhong, R.; Han, S.; Xu, H.; Burrell, A. K.; Henson, N.; Cape, J. L.; Hickmott, D. D.; Timofeeva, T. V.; Larson, T. E.; Zhao, Y. A Porous Metal–Organic Replica of α-PbO2 for Capture of Nerve Agent Surrogate. *J. Am. Chem. Soc.* **2010**, *132*, 17996–17999.

(42) Montoro, C.; Linares, F.; Quartapelle Procopio, E.; Senkovska, I.; Kaskel, S.; Galli, S.; Masciocchi, N.; Barea, E.; Navarro, J. A. R. Capture of Nerve Agents and Mustard Gas Analogues by Hydrophobic Robust MOF-5 Type Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2011**, *133*, 11888–11891.

(43) Bromberg, L.; Klichko, Y.; Chang, E. P.; Speakman, S.; Straut, C. M.; Wilusz, E.; Hatton, T. A. Alkylaminopyridine-Modified Aluminum Aminoterephthalate Metal-Organic Frameworks As Components of Reactive Self-Detoxifying Materials. *ACS Appl. Mater. Interfaces* **2012**, *4*, 4595–4602.

(44) Wang, S.; Bromberg, L.; Schreuder-Gibson, H.; Hatton, T. A. Organophophorous Ester Degradation by Chromium(III) Terephthalate Metal−Organic Framework (MIL-101) Chelated to N,N-Dimethylaminopyridine and Related Aminopyridines. *ACS Appl. Mater. Interfaces* **2013**, *5*, 1269–1278.

(45) Roy, A.; Srivastava, A. K.; Singh, B.; Shah, D.; Mahato, T. H.; Srivastava, A. Kinetics of

Degradation of Sulfur Mustard and Sarin Simulants on HKUST-1 Metal Organic Framework. *Dalt. Trans.* **2012**, *41*, 12346–12348.

(46) Kirlikovali, K. O.; Chen, Z.; Islamoglu, T.; Hupp, J. T.; Farha, O. K. Zirconium-Based Metal−Organic Frameworks for the Catalytic Hydrolysis of Organophosphorus Nerve Agents. *ACS Appl. Mater. Interfaces* **2020**, *12*, 14702–14720.

(47) Howarth, A. J.; Liu, Y.; Li, P.; Li, Z.; Wang, T. C.; Hupp, J. T.; Farha, O. K. Chemical, Thermal and Mechanical Stabilities of Metal–Organic Frameworks. *Nat. Rev. Mater.* **2016**, *1*, 15018.

(48) Planas, N.; Mondloch, J. E.; Tussupbayev, S.; Borycz, J.; Gagliardi, L.; Hupp, J. T.; Farha, O. K.; Cramer, C. J. Defining the Proton Topology of the Zr6-Based Metal-Organic Framework NU-1000. *J. Phys. Chem. Lett.* **2014**, *5*, 3716–3723.

(49) Klet, R. C.; Liu, Y.; Wang, T. C.; Hupp, J. T.; Farha, O. K. Evaluation of Brønsted Acidity and Proton Topology in Zr- and Hf-Based Metal–Organic Frameworks Using Potentiometric Acid–Base Titration. *J. Mater. Chem. A* **2016**, *4*, 1479–1485.

(50) Liu, Y.; Moon, S.-Y.; Hupp, J. T.; Farha, O. K. Dual-Function Metal–Organic Framework as a Versatile Catalyst for Detoxifying Chemical Warfare Agent Simulants. *ACS Nano* **2015**, *9*, 12358–12364.

(51) Islamoglu, T.; Ortuño, M. A.; Proussaloglou, E.; Howarth, A. J.; Vermeulen, N. A.; Atilgan, A.; Asiri, A. M.; Cramer, C. J.; Farha, O. K. Presence versus Proximity: The Role of Pendant Amines in the Catalytic Hydrolysis of a Nerve Agent Simulant. *Angew. Chemie - Int. Ed.* **2018**, *57*, 1949–1953.

(52) de Koning, M. C.; van Grol, M.; Breijaert, T. Degradation of Paraoxon and the Chemical Warfare Agents VX, Tabun, and Soman by the Metal−Organic Frameworks UiO-66-NH2, MOF-808, NU-1000, and PCN-777. *Inorg. Chem.* **2017**, *56*, 11804–11809.

(53) Palomba, J. M.; Credille, C. V.; Kalaj, M.; DeCoste, J. B.; Peterson, G. W.; Tovar, T. M.; Cohen, S. M. High-Throughput Screening of Solid-State Catalysts for Nerve Agent Degradation. *Chem. Commun.* **2018**, *54*, 5768–5771.

(54) Lee, D. T.; Zhao, J.; Peterson, G. W.; Parsons, G. N. A Catalytic "MOF-Cloth" Formed via Directed Supramolecular Assembly of UiO-66-NH2 Crystals on ALD-Coated Textiles for Rapid Degradation of Chemical Warfare Agent Simulants. *Chem. Mater.* **2017**, *29*, 4894–4903.

(55) Chen, Z.; Islamoglu, T.; Farha, O. K. Toward Base Heterogenization: A Zirconium Metal–Organic Framework/Dendrimer or Polymer Mixture for Rapid Hydrolysis of a Nerve-Agent Simulant. *ACS Appl. Nano Mater.* **2019**, *2*, 1005–1008.

(56) Wang, H.; Mahle, J. J.; Tovar, T. M.; Peterson, G. W.; Hall, M. G.; DeCoste, J. B.; Buchanan, J. H.; Karwacki, C. J. Solid-Phase Detoxification of Chemical Warfare Agents Using Zirconium-Based Metal Organic Frameworks and the Moisture Effects: Analyze via Digestion. *ACS Appl. Mater. Interfaces* **2019**, *11*, 21109–21116.

(57) Troya, D. Reaction Mechanism of Nerve-Agent Decomposition with Zr-Based Metal Organic Frameworks. *J. Phys. Chem. C* **2016**, *120*, 29312–29323.

(58) Wang, G.; Sharp, C.; Plonka, A. M.; Wang, Q.; Frenkel, A. I.; Guo, W.; Hill, C.; Smith, C.; Kollar, J.; Troya, D.; Morris, J. R. Mechanism and Kinetics for Reaction of the Chemical Warfare Agent Simulant, DMMP(g), with Zirconium(IV) MOFs: An Ultrahigh-Vacuum and DFT Study. *J. Phys. Chem. C* **2017**, *121*, 11261–11272.

(59) Plonka, A. M.; Wang, Q.; Gordon, W. O.; Balboa, A.; Troya, D.; Guo, W.; Sharp, C. H.;

Senanayake, S. D.; Morris, J. R.; Hill, C. L.; Frenkel, A. I. In Situ Probes of Capture and Decomposition of Chemical Warfare Agent Simulants by Zr-Based Metal Organic Frameworks. *J. Am. Chem. Soc.* **2017**, *139*, 599–602.
(60) Chen, Z.; Ma, K.; Mahle, J. J.; Wang, H.; Syed, Z. H.; Atilgan, A.; Chen, Y.; Xin, J. H.; Islamoglu, T.; Peterson, G. W.; Farha, O. K. Integration of Metal–Organic Frameworks on Protective Layers for Destruction of Nerve Agents under Relevant Conditions. *J. Am. Chem. Soc.* **2019**, *141*, 20016–20021.
(61) Cho, K. Y.; Seo, J. Y.; Kim, H.-J.; Pai, S. J.; Do, X. H.; Yoon, H. G.; Hwang, S. S.; Han, S. S.; Baek, K.-Y. Facile Control of Defect Site Density and Particle Size of UiO-66 for Enhanced Hydrolysis Rates: Insights into Feasibility of Zr(IV)-Based Metal-Organic Framework (MOF) Catalysts. *Appl. Catal. B Environ.* **2019**, *245*, 635–647.
(62) DeCoste, J. B.; Peterson, G. W. Metal−Organic Frameworks for Air Purification of Toxic Chemicals. *Chem. Rev.* **2014**, *114*, 5695−5727.
(63) Lu, A. X.; McEntee, M.; Browe, M. A.; Hall, M. G.; DeCoste, J. B.; Peterson, G. W. MOFabric: Electrospun Nanofiber Mats from PVDF/UiO-66-NH2 for Chemical Protection and Decontamination. *ACS Appl. Mater. Interfaces* **2017**, *9*, 13632–13636.
(64) Zhao, J.; Lee, D. T.; Yaga, R. W.; Hall, M. G.; Barton, H. F.; Woodward, I. R.; Oldham, C. J.; Walls, H. J.; Peterson, G. W.; Parsons, G. N. Ultra-Fast Degradation of Chemical Warfare Agents Using MOF–Nanofiber Kebabs. *Angew. Chemie - Int. Ed.* **2016**, *55*, 13224–13228.
(65) Lopez-Maya, E.; Montoro, C.; Rodriguez-Albelo, L. M.; Aznar Cervantes, S. D.; Lozano-Perez, A. A.; Cenis, J. L.; Barea, E.; Navarro, J. A. R. Textile/Metal–Organic-Framework Composites as Self-Detoxifying Filters for Chemical-Warfare Agents. *Angew. Chemie* **2015**, *127*, 6894–6898.
(66) Bobbitt, N. S.; Mendonca, M. L.; Howarth, A. J.; Islamoglu, T.; Hupp, J. T.; Farha, O. K.; Snurr, R. Q. Metal–Organic Frameworks for the Removal of Toxic Industrial Chemicals and Chemical Warfare Agents. *Chem. Soc. Rev.* **2017**, *46*, 3357–3385.
(67) Islamoglu, T.; Chen, Z.; Wasson, M. C.; Buru, C. T.; Kirlikovali, K. O.; Afrin, U.; Mian, M. R.; Farha, O. K. Metal–Organic Frameworks against Toxic Chemicals. *Chem. Rev.* **2020**.
(68) Odoh, S. O.; Cramer, C. J.; Truhlar, D. G.; Gagliardi, L. Quantum-Chemical Characterization of the Properties and Reactivities of Metal–Organic Frameworks. *Chem. Rev.* **2015**, *115*, 6051–6111.
(69) Nørskov, J. K.; Abild-Pedersen, F.; Studt, F.; Bligaard, T. Density Functional Theory in Surface Chemistry and Catalysis. *PNAS* **2011**, *108*, 937–943.
(70) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons: Chichester, 2004.
(71) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
(72) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(73) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.
(74) Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41*, 157–167.

(75) Zhao, Y.; Truhlar, D. G. A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* **2006**, *125*, 194101.
(76) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
(77) Hay, P. J.; Wadt, W. R. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985**, *82*, 270–283.
(78) Hay, P. J.; Wadt, W. R. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for K to Au Including the Outermost Core Orbitals. *J. Chem. Phys.* **1985**, *82*, 299–310.
(79) Andrae, D.; HäuBermann, U.; Dolg, M.; Stoll, H.; PreuB, H. Energy-Adjusted Ab Initio Pseudopotentials for the Second and Third Row Transition Elements. *Theor. Chim. Acta.* **1990**, *77*, 123–141.
(80) Dolg, M.; Stoll, H.; Preuss, H. Energy-Adjusted Ab Initio Pseudopotentials for the Rare Earth Elements. *J. Chem. Phys.* **1989**, *90*, 1730–1734.
(81) Miertus, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of AB Initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981**, *55*, 117–129.
(82) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
(83) Basdogan, Y.; Keith, J. A. A Paramedic Treatment for Modeling Explicitly Solvated Chemical Reaction Mechanisms. *Chem. Sci.* **2018**, *9*, 5341–5346.
(84) Marom, N.; Tkatchenko, A.; Rossi, M.; Gobre, V. V.; Hod, O.; Scheffler, M.; Kronik, L. Dispersion Interactions with Density-Functional Theory: Benchmarking Semiempirical and Interatomic Pairwise Corrected Density Functionals. *J. Chem. Theory Comput.* **2011**, *7*, 3944–3951.
(85) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(86) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
(87) Becke, A. D.; Johnson, E. R. A Density-Functional Model of the Dispersion Interaction. *J. Chem. Phys.* **2005**, *123*, 154101.
(88) Schlegel, H. B. Optimization of Equilibrium Geometries and Transition Structures. *J. Comput. Chem.* **1982**, *3*, 214–218.
(89) Chen, S.-L.; Fang, W.-H.; Himo, F. Theoretical Study of the Phosphotriesterase Reaction Mechanism. *J. Phys. Chem. B* **2007**, *111*, 1253–1255.
(90) Smith, B. M. Catalytic Methods for the Destruction of Chemical Warfare Agents under Ambient Conditions. *Chem. Soc. Rev.* **2008**, *37*, 470–478.
(91) Yang, Y.-C.; Baker, J. A.; Ward, J. R. Decontamination of Chemical Warfare Agents. *Chem. Rev.* **1992**, *92*, 1729–1743.
(92) Yang, Y.-C. Chemical Detoxification of Nerve Agent VX. *Acc. Chem. Res.* **1999**, *32*,

109–115.
(93) Wang, Q.; Chapleski Jr., R. C.; Plonka, A. M.; Gordon, W. O.; Guo, W.; Nguyen-Phan, T.-D.; Sharp, C. H.; Marinkovic, N. S.; Senanayake, S. D.; Morris, J. R.; Hill, C. L.; Troya, D.; Frenkel, A. I. Atomic-Level Structural Dynamics of Polyoxoniobates during DMMP Decomposition. *Sci. Rep.* **2017**, *7*, 773.
(94) Chen, S.; Ruan, Y.; Brown, J. D.; Gallucci, J.; Maslak, V.; Hadad, C. M.; Badjic, J. D. Assembly of Amphiphilic Baskets into Stimuli-Responsive Vesicles. Developing a Strategy for the Detection of Organophosphorus Chemical Nerve Agents. *J. Am. Chem. Soc.* **2013**, *135*, 14964–14967.
(95) Pascual, L.; Sayed, S. El; Martinez-Manez, R.; Costero, A. M.; Gil, S.; Gavina, P.; Sancenon, F. Acetylcholinesterase-Capped Mesoporous Silica Nanoparticles That Open in the Presence of Diisopropylfluorophosphate (a Sarin or Soman Simulant). *Org. Lett.* **2016**, *18*, 5548–5551.
(96) McCombie, H.; Saunders, B. C. Alkyl Fluorophosphonates: Preparation and Physiological Properties. *Nature* **1946**, *157*, 287–289.
(97) Cook, H. G.; Saunders, B. C.; Smith, F. E. Esters Containing Phosphorus. Part VIII. Structural Requirements for High Toxicity and Miotic Action of Esters of Fluorophosphonic Acid. *J. Chem. Soc.* **1949**, *0*, 635–638.
(98) Fukuto, T. R.; Metcalf, R. L. Structure and Insecticidal Activity of Some Diethyl Substituted Phenyl Phosphates. *J. Agric. Food Chem.* **1956**, *4*, 930–935.
(99) Hartley, F. R. The Chemistry or Organophosphorus Compounds: Ter- and Quinque-Valent Phosphorus Acids and Their Derivatives. In *The Chemistry or Organophosphorus Compounds: Ter- and Quinque-Valent Phosphorus Acids and Their Derivatives, Volume 4*; Patai, S., Ed.; Wiley, 1996; pp 894–930.
(100) Hudson, R. F.; Keay, L. The Mechanism of Hydrolysis of Phosphonochloridates and Related Compounds. Part I. The Effect of Substituents. *J. Chem. Soc.* **1960**, *0*, 1859–1864.
(101) Aldridge, W. N.; N., D. A. The Inhibition of Erythrocyte Cholinesterase by Tri-Esters of Phosphoric Acid. 1. Diethyl p-Nitrophenyl Phosphate (E600) and Analogues. *Biochem. J.* **1952**, *51*, 62–70.
(102) Fukuto, T. R.; Metcalf, R. L. The Effect of Structure on the Reactivity of Alkylphosphonate Esters. *J. Am. Chem. Soc.* **1959**, *81*, 372–377.
(103) Randic, M. On Characterization of Molecular Branching. *J . Am. Chem. Soc.* **1975**, *97*, 6609–6615.
(104) Wang, D.; Shi, J.; Xiong, Y.; Hu, J.; Lin, Z.; Qiu, Y.; Cheng, J. A QSAR-Based Mechanistic Study on the Combined Toxicity of Antibiotics and Quorum Sensing Inhibitors against Escherichia Coli. *J. Hazard. Mater.* **2018**, *341*, 438–447.
(105) Nandi, S.; Monesi, A.; Drgan, V.; Merzel, F.; Novic, M. Quantitative Structure-Activation Barrier Relationship Modeling for Diels-Alder Ligations Utilizing Quantum Chemical Structural Descriptors. *Chem. Cent. J.* **2013**, *7*, 171.
(106) Johnson, H.; Kenleyb, R. A.; Rynard, C. QSAR for Cholinesterase Inhibition by Organophosphorus Esters and CNDO/2 Calculations for Organophosphorus Ester Hydrolysis. *Quant. Struct.-Act. Relat.* **1985**, *4*, 172–180.
(107) Manco, G.; Porzio, E.; Suzumoto, Y. Enzymatic Detoxification: A Sustainable Means of Degrading Toxic Organophosphate Pesticides and Chemical Warfare Nerve Agents. *J. Chem. Technol. Biotechnol.* **2018**, *93*, 2064–2082.

(108) Giannakoudakis, D. A.; Bandosz, T. J. *Detoxification of Chemical Warfare Agents: From WWI to Multifunctional Nanocomposite Approaches*; Springer, 2017.
(109) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Fox, D. J.; et al. Gaussian 09, Revision D.01. Gaussian, Inc.: Wallingford, CT 2013.
(110) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
(111) Moller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
(112) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A Complete Basis Set Model Chemistry. VI. Use of Density Functional Geometries and Frequencies. *J. Chem. Phys.* **1999**, *110*, 2822–2827.
(113) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A Complete Basis Set Model Chemistry. VII. Use of the Minimum Population Localization Method. *J. Chem. Phys.* **2000**, *112*, 6532–6542.
(114) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. Ab Initio Study of Solvated Molecules: A New Implementation of the Polarizable Continuum Model. *Chem. Phys. Lett.* **1996**, *255*, 327–335.
(115) US EPA. Toxicity Estimation Software Tool: A Program to Estimate Toxicity from Molecular Structure (Version 4.2). 2016.
(116) Romesburg, H. C. *Cluster Analysis for Researchers*; Lifetime Learning Publications: Belmont, CA, 1984.
(117) Contrera, J. F.; Matthews, E. J.; Benz, R. D. Predicting the Carcinogenic Potential of Pharmaceuticals in Rodents Using Molecular Structural Similarity and E-State Indices. *Regul. Toxicol. Pharmacol.* **2003**, *38*, 243–259.
(118) Martin, T. M.; Harten, P.; Venkatapathy, R.; Das, S.; Young, D. M. A. A Hierarchical Clustering Methodology for the Estimation of Toxicity. *Toxicol. Mech. Method.* **2008**, *18*, 251–266.
(119) Glendening, E. D.; Landis, C. R.; Weinhold, F. Natural Bond Orbital Methods. *WIREs Comput. Mol. Sci.* **2012**, *2*, 1–42.
(120) Benschop, H. P.; De Jong, L. P. A. Nerve Agent Stereoisomers: Analysis, Isolation, and Toxicology. *Acc. Chem. Res.* **1988**, *21*, 368–374.
(121) Ashkenazi, N.; Segall, Y.; Chen, R.; Sod-Moriah, G.; Fattal, E. The Mechanism of Nucleophilic Displacements at Phosphorus in Chloro-Substituted Methylphosphonate Esters: P–O vs P–C Bond Cleavage: A DFT Study. *J. Org. Chem.* **2010**, *75*, 1917–1926.
(122) Šečkute, J.; Menke, J. L.; Emnett, R. J.; Patterson, E. V.; Cramer, C. J. Ab Initio Molecular Orbital and Density Functional Studies on the Solvolysis of Sarin and O,S-Dimethyl Methylphosphonothiolate, a VX-like Compound. *J. Org. Chem.* **2005**, *70*, 8649–8660.
(123) Zheng, F.; Zhan, C.-G.; Ornstein, R. L. Theoretical Studies of Reaction Pathways and Energy Barriers for Alkaline Hydrolysis of Phosphotriesterase Substrates Paraoxon and

Related Toxic Phosphofluoridate Nerve Agents. *J. Chem. Soc. Perkin Trans. 2* **2001**, 2355–2363.
(124) Chen, H.; Giese, T. J.; Huang, M.; Wong, K.-Y.; Harris, M. E.; York, D. M. Mechanistic Insights into RNA Transphosphorylation from Kinetic Isotope Effects and Linear Free Energy Relationships of Model Reactions. *Chem. Eur. J.* **2014**, *20*, 14336–14343.
(125) Tarrat, N. Alkaline Hydrolysis of Phosphate Triesters in Solution: Stepwise or Concerted? A Theoretical Study. *J. Mol. Struct. THEOCHEM* **2010**, *941*, 56–60.
(126) Kesharwani, M. K.; Khan, M. A. S.; Bandyopadhyay, T.; Ganguly, B. Solvolysis Process of Organophosphorus Compound P-[2-(Dimethylamino)Ethyl]-N,N-Dimethylphosphonamidic Fluoride with Simple and α-Nucleophiles: A DFT Study. *Theor. Chem. Acc.* **2010**, *127*, 39–47.
(127) Evans, D. A. pKa Table. evans.rc.fas.harvard.edu/pdf/evans_pKa_table.pdf (accessed Jan 4, 2020).
(128) *CRC Handbook of Chemistry and Physics*, 90th ed.; Lide, D. R., Ed.; CRC Press LLC: Boca Raton, FL, 2010.
(129) Bourne, N.; Chrystiuk, E.; Davis, A. M.; Williams, A. A Single Transition State in the Reaction of Aryl Diphenylphosphinate Esters with Phenolate Ions in Aqueous Solution. *J . Am. Chem. Soc.* **1988**, *110*, 1890–1895.
(130) Schwarzenbach, R. P.; Stierli, R.; Folsom, B. R.; Zeyer, J. Compound Properties Relevant for Assessing the Environmental Partitioning of Nitrophenols. *Environ. Sci. Technol.* **1988**, *22*, 83–92.
(131) ChemAxon Ltd. Chemicalize. https://chemicalize.com (accessed Jan 4, 2020).
(132) Serjeant, P.; Dempsey, B. *Ionisation Constants of Organic Acids in Aqueous Solution*; Pergamon Press: New York, 1979.
(133) The Metabolomics Innovation Centre. FooDB. foodb.ca/compounds/FDB029632 (accessed Jan 4, 2020).
(134) U.S. National Library of Medicine. ChemIDplus. http://chem.sis.nlm.nih.gov/chemidplus/chemidheavy.jsp (accessed Jan 4, 2020).
(135) Cox, J. R.; Ramsay, O. B. Mechanisms of Nucleophilic Substitution in Phosphate Esters. *Chem. Rev.* **1964**, *64*, 317–352.
(136) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289–320.
(137) APPENDIX C: HEALTH RISK ASSESSMENT FOR THE NERVE AGENT GD (SOMAN). *J. Toxicol. Environ. Heal. Part A* **2000**, *59*, 417–438.
(138) National Technical Information Service (Springfield, VA 2216). *PB158-508*.
(139) Chen, H.; Liao, P.; Mendonca, M. L.; Snurr, R. Q. Insights into Catalytic Hydrolysis of Organophosphate Warfare Agents by Metal–Organic Framework NU-1000. *J. Phys. Chem. C* **2018**, *122*, 12362–12368.
(140) Katz, M. J.; Klet, R. C.; Moon, S.-Y.; Mondloch, J. E.; Hupp, J. T.; Farha, O. K. One Step Backward Is Two Steps Forward: Enhancing the Hydrolysis Rate of UiO-66 by Decreasing [OH–]. *ACS Catal.* **2015**, *5*, 4637–4642.
(141) Moon, S.-Y.; Liu, Y.; Hupp, J. T.; Farha, O. K. Instantaneous Hydrolysis of Nerve-Agent Simulants with a Six-Connected Zirconium-Based Metal–Organic Framework. *Angew. Chemie - Int. Ed.* **2015**, *54*, 6795–6799.
(142) Koopmans, T. Über Die Zuordnung von Wellenfunktionen Und Eigenwerten Zu Den

Einzelnen Elektronen Eines Atoms. *Physica* **1934**, *1*, 104–113.
(143) Padmanabhan, J.; Parthasarathi, R.; Elango, M.; Subramanian, V.; Krishnamoorthy, B. S.; Gutierrez-Olivia, S.; Toro-Labbe, A.; Roy, D. R.; Chattaraj, P. K. Multiphilic Descriptor for Chemical Reactivity and Selectivity. *J. Phys. Chem. A* **2007**, *111*, 9130–9138.
(144) Parr, R. G.; Chattaraj, P. K. Principle of Maximum Hardness. *J. Am. Chem. Soc.* **1991**, *113*, 1854–1855.
(145) Ayers, P. W.; Parr, R. G. Variational Principles for Describing Chemical Reactions: The Fukui Function and Chemical Hardness Revisited. *J . Am. Chem. Soc.* **2000**, *122*, 2010–2018.
(146) Yang, W.; Mortier, W. J. The Use of Global and Local Molecular Parameters for the Analysis of the Gas-Phase Basicity of Amines. *J. Am. Chem. Soc.* **1986**, *108*, 5708–5711.
(147) Ajay. A Unified Framework for Using Neural Networks To Build QSARs. *J. Med. Chem.* **1993**, *36*, 3565–3571.
(148) Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.-D.; Lee, K.-H.; Tropsha, A. Rational Selection of Training and Test Sets for the Development of Validated QSAR Models. *J. Comput. Aided Mol. Des.* **2003**, *117*, 241–253.
(149) Cao, L.; Zhu, P.; Zhao, Y.; Zhao, J. Using Machine Learning and Quantum Chemistry Descriptors to Predict the Toxicity of Ionic Liquids. *J. Hazard. Mater.* **2018**, *352*, 17–26.
(150) Mitra, I.; Saha, A.; Roy, K. Exploring Quantitative Structure-Activity Relationship (QSAR) Studies of Antioxidant Phenolic Compounds Obtained from Traditional Chinese Medicinal Plants. *Mol. Simul.* **2010**, *36*, 1067–1079.
(151) Golbraikh, A.; Tropsha, A. Beware of Q2! *J. Mol. Graph. Model.* **2002**, *20*, 269–276.
(152) Alexander, D. L. J.; Tropsha, A.; Winkler, D. A. Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. *J. Chem. Inf. Model.* **2015**, *55*, 1316–1322.
(153) Minovski, N.; Zuperl, S.; Drgan, V.; Novic, M. Assessment of Applicability Domain for Multivariate Counter-Propagation Artificial Neural Network Predictive Models by Minimum Euclidean Distance Space Analysis: A Case Study. *Anal. Chim. Acta* **2013**, *759*, 28–42.
(154) Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environ. Health Perspect.* **2003**, *111*, 1361–1375.
(155) Agrawal, M.; Sava Gallis, D. F.; Greathouse, J. A.; Sholl, D. S. How Useful Are Common Simulants of Chemical Warfare Agents at Predicting Adsorption Behavior? *J. Phys. Chem. C* **2018**, *122*, 26061–26069.
(156) Raushel, F. M. Catalytic Detoxification. *Nat. Chem. Biol.* **2011**, *469*, 310–311.
(157) Kim, K.; Tsay, O. G.; Atwood, D. A.; Churchill, D. G. Destruction and Detection of Chemical Warfare Agents. *Chem. Rev.* **2011**, *111*, 5345–5403.
(158) Dumas, D. P.; Caldwell, S. R.; Wild, J. R.; Raushel, F. M. Purification and Properties of the Phosphotriesterase from Pseudomonas Diminuta. *J. Biol. Chem.* **1989**, *264*, 19659–19665.
(159) Li, W.-S.; Lum, K. T.; Chen-Goodspeed, M.; Sogorb, M. A.; Raushel, F. M. Stereoselective Detoxification of Chiral Sarin and Soman Analogues by Phosphotriesterase. *Bioorganic Med. Chem.* **2001**, *9*, 2083–2091.

(160) Bigley, A. N.; Raushel, F. M. Catalytic Mechanisms for Phosphotriesterases. *Biochim. Biophys. Acta* **2013**, *1834*, 443–453.
(161) Yang, S. W.; Doetschman, D. C.; Schulte, J. T.; Sambur, J. B.; Kanyi, C. W.; Fox, J. D.; Kowenje, C. O.; Jones, B. R.; Sherma, N. D. Sodium X-Type Faujasite Zeolite Decomposition of Dimethyl Methylphosphonate (DMMP) to Methylphosphonate: Nucleophilic Zeolite Reactions I. *Microporous Mesoporous Mater.* **2006**, *92*, 56–60.
(162) Kinnan, M. K.; Creasy, W. R.; Fullmer, L. B.; Schreuder-Gibson, H. L.; Nyman, M. Nerve Agent Degradation with Polyoxoniobates. *Eur. J. Inorg. Chem.* **2014**, 2361–2367.
(163) Guo, W.; Lv, H.; Sullivan, K. P.; Gordon, W. O.; Balboa, A.; Wagner, G. W.; Musaev, D. G.; Bacsa, J.; Hill, C. L. Broad-Spectrum Liquid- and Gas-Phase Decontamination of Chemical Warfare Agents by One-Dimensional Heteropolyniobates. *Angew. Chemie - Int. Ed.* **2016**, *55*, 7403–7407.
(164) Lee, J.; Farha, O. K.; Roberts, J.; Scheidt, K. A.; Nguyen, S. T.; Hupp, J. T. Metal-Organic Framework Materials as Catalysts. *Chem. Soc. Rev.* **2009**, *38*, 1450–1459.
(165) Chughtai, A. H.; Ahmad, N.; Younus, H. A.; Laypkov, A.; Verpoort, F. Metal-Organic Frameworks: Versatile Heterogeneous Catalysts for Efficient Catalytic Organic Transformations. *Chem. Soc. Rev.* **2015**, *44*, 6804–6849.
(166) Barea, E.; Montoro, C.; Navarro, J. A. R. Toxic Gas Removal – Metal–Organic Frameworks for the Capture and Degradation of Toxic Gases and Vapours. *Chem. Soc. Rev.* **2014**, *43*, 5419–5430.
(167) Bai, Y.; Dou, Y.; Xie, L.-H.; Rutledge, W.; Li, J.-R.; Zhou, H.-C. Zr-Based Metal–Organic Frameworks: Design, Synthesis, Structure, and Applications. *Chem. Soc. Rev.* **2016**, *45*, 2327–2367.
(168) Corma, A.; García, H.; Llabrés i Xamena, F. X. Engineering Metal Organic Frameworks for Heterogeneous Catalysis. *Chem. Rev.* **2010**, *110*, 4606–4655.
(169) Katz, M. J.; Moon, S.-Y.; Mondloch, J. E.; Beyzavi, M. H.; Stephenson, C. J.; Hupp, J. T.; Farha, O. K. Exploiting Parameter Space in MOFs: A 20-Fold Enhancement of Phosphate-Ester Hydrolysis with UiO-66-NH2. *Chem. Sci.* **2015**, *6*, 2286–2291.
(170) Platero-Prats, A. E.; Mavrandonakis, A.; Gallington, L. C.; Liu, Y.; Hupp, J. T.; Farha, O. K.; Cramer, C. J.; Chapman, K. W. Structural Transitions of the Metal-Oxide Nodes within Metal-Organic Frameworks: On the Local Structures of NU-1000 and UiO-66. *J. Am. Chem. Soc.* **2016**, *138*, 4178–4185.
(171) Wadt, W. R.; Hay, P. J. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for Main Group Elements Na to Bi. *J. Chem. Phys.* **1985**, *82*, 284–298.
(172) Hengge, A. C. Isotope Effects in the Study of Phosphoryl and Sulfuryl Transfer Reactions. *Acc. Chem. Res.* **2002**, *35*, 105–112.
(173) Klähn, M.; Rosta, E.; Warshel, A. On the Mechanism of Hydrolysis of Phosphate Monoesters Dianions in Solutions and Proteins. *J. Am. Chem. Soc.* **2006**, *128*, 15310–15323.
(174) Lassila, J. K.; Zalatan, J. G.; Herschlag, D. Biological Phosphoryl-Transfer Reactions: Understanding Mechanism and Catalysis. *Annu. Rev. Biochem.* **2011**, *80*, 669–702.
(175) Yang, D.; Bernales, V.; Islamoglu, T.; Farha, O. K.; Hupp, J. T.; Cramer, C. J.; Gagliardi, L.; Gates, B. C. Tuning the Surface Chemistry of Metal Organic Framework Nodes: Proton Topology of the Metal-Oxide-Like Zr6 Nodes of UiO-66 and NU-1000. *J. Am. Chem. Soc.* **2016**, *138*, 15189–15196.

(176) Westheimer, F. H. Why Nature Chose Phosphates. *Science* **1987**, *235*, 1173–1178.
(177) Oivanen, M.; Kuusela, S.; Lönnberg, H. Kinetics and Mechanisms for the Cleavage and Isomerization of the Phosphodiester Bonds of RNA by Bronsted Acids and Bases. *Chem. Rev.* **1998**, *98*, 961–990.
(178) Blaskó, A.; Bruice, T. C. Recent Studies of Nucleophilic, General-Acid, and Metal Ion Catalysis of Phosphate Diester Hydrolysis. *Acc. Chem. Res.* **1999**, *32*, 475–484.
(179) Cleland, W. W.; Hengge, A. C. Enzymatic Mechanisms of Phosphate and Sulfate Transfer. *Chem. Rev.* **2006**, *106*, 3252–3278.
(180) Kamerlin, S. C. L.; Sharma, P. K.; Prasad, R. B.; Warshel, A. Why Nature Really Chose Phosphate. *Q. Rev. Biophys.* **2013**, *46*, 1–132.
(181) Kellerman, D. L.; York, D. M.; Piccirilli, J. A.; Harris, M. E. Altered (Transition) States: Mechanisms of Solution and Enzyme Catalyzed RNA 2'-O-Transphosphorylation. *Curr. Opin. Chem. Biol.* **2014**, *21*, 96–102.
(182) Lönnberg, H.; Strömberg, R.; Williams, A. Compelling Evidence for a Stepwise Mechanism of the Alkaline Cyclisation of Uridine 3′-Phosphate Esters. *Org. Biomol. Chem.* **2004**, *2*, 2165–2167.
(183) Huang, M.; York, D. M. Linear Free Energy Relationships in RNA Transesterification: Theoretical Models to Aid Experimental Interpretations. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15846–15855.
(184) Roston, D.; Cui, Q. Substrate and Transition State Binding in Alkaline Phosphatase Analyzed by Computation of Oxygen Isotope Effects. *J. Am. Chem. Soc.* **2016**, *138*, 11946–11957.
(185) Roston, D.; Demapan, D.; Cui, Q. Leaving Group Ability Observably Affects Transition State Structure in a Single Enzyme Active Site. *J. Am. Chem. Soc.* **2016**, *138*, 7386–7394.
(186) Reed, A. E.; Weinhold, F. Natural Bond Orbital Analysis of Near-Hartree–Fock Water Dimer. *J. Chem. Phys.* **1983**, *78*, 4066–4073.
(187) Cowan, J. A. Chemical Nucleases. *Curr. Opin. Chem. Biol.* **2001**, *5*, 634–642.
(188) Sreedhara, A.; Cowan, J. A. Catalytic Hydrolysis of DNA by Metal Ions and Complexes. *J. Biol. Inorg. Chem.* **2001**, *6*, 337–347.
(189) Liu, C.; Wang, M.; Zhang, T.; Sun, H. DNA Hydrolysis Promoted by Di- and Multi-Nuclear Metal Complexes. *Coord. Chem. Rev.* **2004**, *248*, 147–168.
(190) Humphry, T.; Iyer, S.; Iranzo, O.; Morrow, J. R.; Richard, J. P.; Paneth, P.; Hengge, A. C. Altered Transition State for the Reaction of an RNA Model Catalyzed by a Dinuclear Zinc(II) Catalyst. *J. Am. Chem. Soc.* **2008**, *130*, 17858–17866.
(191) Chen, H.; Piccirilli, J. A.; Harris, M. E.; York, D. M. Effect of Zn2+ Binding and Enzyme Active Site on the Transition State for RNA 2′-O-Transphosphorylation Interpreted through Kinetic Isotope Effects. *Biochim. Biophys. Acta* **2015**, *1854*, 1795–1800.
(192) Zhang, S.; Gu, H.; Chen, H.; Strong, E.; Ollie, E. W.; Kellerman, D.; Liang, D.; Miyagi, M.; Anderson, V. E.; Piccirilli, J. A.; York, D. M.; Harris, M. E. Isotope Effect Analyses Provide Evidence for an Altered Transition State for RNA 2′-O-Transphosphorylation Catalyzed by Zn2+. *Chem. Commun.* **2016**, *52*, 4462–4465.
(193) *Note by the Technical Secretariat: Report of the Fact-Finding Mission Regarding the Incident of Alleged Use of Toxic Chemicals as a Weapon in Douma, Syrian Arab Republic, on 7 April 2018*; 2019; Vol. S/1731/201.
(194) Stone, R. Obscure Cold War Nerve Agents Set to Be Banned. *Science* **2019**, *366*, 404–

405.
(195) Jang, Y. J.; Kim, K.; Tsay, O. G.; Atwood, D. A.; Churchill, D. G. Update 1 of: Destruction and Detection of Chemical Warfare Agents. *Chem. Rev.* **2015**, *115*, PR1–PR76.
(196) Carlsen, L. After Salisbury Nerve Agents Revisited. *Mol. Inf.* **2018**, *37*, 1800106.
(197) Schwenk, M. Chemical Warfare Agents. Classes and Targets. *Toxicol. Lett.* **2018**, *293*, 253–263.
(198) Momeni, M. R.; Cramer, C. J. Dual Role of Water in Heterogeneous Catalytic Hydrolysis of Sarin by Zirconium-Based Metal–Organic Frameworks. *ACS Appl. Mater. Interfaces* **2018**, *10*, 18435–18439.
(199) Momeni, M.; Cramer, C. J. Computational Screening of Roles of Defects and Metal Substitution on Reactivity of Different Single- vs Double-Node Metal−Organic Frameworks for Sarin Decomposition. *J. Phys. Chem. C* **2019**, *123*, 15157–15165.
(200) Islamoglu, T.; Atilgan, A.; Moon, S.-Y.; Peterson, G. W.; DeCoste, J. B.; Hall, M.; Hupp, J. T.; Farha, O. K. Cerium(IV) vs Zirconium(IV) Based Metal–Organic Frameworks for Detoxification of a Nerve Agent. *Chem. Mater.* **2017**, *29*, 2672–2675.
(201) Islamoglu, T.; Ray, D.; Li, P.; Majewski, M. B.; Akpinar, I.; Zhang, X.; Cramer, C. J.; Gagliardi, L.; Farha, O. K. From Transition Metals to Lanthanides to Actinides: Metal-Mediated Tuning of Electronic Properties of Isostructural Metal–Organic Frameworks. *Inorg. Chem.* **2018**, *57*, 13246–13251.
(202) Lyu, J.; Zhang, X.; Li, P.; Wang, X.; Buru, C. T.; Bai, P.; Guo, X.; Farha, O. K. Exploring the Role of Hexanuclear Clusters as Lewis Acidic Sites in Isostructural Metal−Organic Frameworks. *Chem. Mater.* **2019**, *31*, 4166–4172.
(203) Wang, X.; Zhang, X.; Li, P.; Otake, K.; Cui, Y.; Lyu, J.; Krzyaniak, M. D.; Zhang, Y.; Li, Z.; Liu, J.; Buru, C. T.; Islamoglu, T.; Wasielewski, M. R.; Li, Z.; Farha, O. K. Vanadium Catalyst on Isostructural Transition Metal, Lanthanide, and Actinide Based Metal−Organic Frameworks for Alcohol Oxidation. *J . Am. Chem. Soc.* **2019**, *141*, 8306–8314.
(204) Son, F. A.; Atilgan, A.; Idrees, K. B.; Islamoglu, T.; Farha, O. K. Solvent-Assisted Linker Exchange Enabled Preparation of Cerium-Based Metal–Organic Frameworks Constructed from Redox Active Linkers. *Inorg. Chem. Front.* **2020**, *7*, 984–990.
(205) Xiong, Y.; Chen, S.; Ye, F.; Su, L.; Zhang, C.; Shen, S.; Zhao, S. Synthesis of a Mixed Valence State Ce-MOF as an Oxidase Mimetic for the Colorimetric Detection of Biothiols. *Chem. Commun.* **2015**, *51*, 4635–4638.
(206) Smolders, S.; Lomachenko, K. A.; Bueken, B.; Struyf, A.; Bugaev, A. L.; Atzori, C.; Stock, N.; Lamberti, C.; Roeffaers, M. B. J.; De Vos, D. E. Unravelling the Redox-Catalytic Behavior of Ce4+ Metal–Organic Frameworks by X-Ray Absorption Spectroscopy. *ChemPhysChem* **2018**, *19*, 373–378.
(207) Momeni, M. R.; Cramer, C. J. Structural Characterization of Pristine and Defective [Zr12(M3–O)8(M3–OH)8(M2–OH)6]18+ Double–Node Metal–Organic Framework and Predicted Applications for Single-Site Catalytic Hydrolysis of Sarin. *Chem. Mater.* **2018**, *30*, 4432–4439.
(208) Mendonca, M. L.; Snurr, R. Q. Computational Screening of Metal−Organic Framework-Supported Single-Atom Transition-Metal Catalysts for the Gas-Phase Hydrolysis of Nerve Agents. *ACS Catal.* **2020**, *10*, 1310–1323.

(209) Harvey, J. A.; McEntee, M. L.; Garibay, S. J.; Durke, E. M.; DeCoste, J. B.; Greathouse, J. A.; Sava Gallis, D. F. Spectroscopically Resolved Binding Sites for the Adsorption of Sarin Gas in a Metal−Organic Framework: Insights beyond Lewis Acidity. *J. Phys. Chem. Lett.* **2019**, *10*, 5142–5147.
(210) Ploskonka, A. M.; DeCoste, J. B. Insight into Organophosphate Chemical Warfare Agent Simulant Hydrolysis in Metal-Organic Frameworks. *J. Hazard. Mater.* **2019**, *375*, 191–197.
(211) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169–11186.
(212) Kresse, G.; Furthmüller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.
(213) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B* **1993**, *47*, 558–561.
(214) Kresse, G.; Hafner, J. Ab Initio Molecular-Dynamics Simulation of the Liquid-Metal–Amorphous-Semiconductor Transition in Germanium. *Phys. Rev. B* **1994**, *49*, 14251–14269.
(215) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953–17979.
(216) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59*, 1758–1775.
(217) Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Phys. Rev. B* **1992**, *45*, 13244–13249.
(218) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
(219) Weigend, F. Accurate Coulomb-Fitting Basis Sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.
(220) Harvey, S. P.; McMahon, L. R.; Berg, F. J. Hydrolysis and Enzymatic Degradation of Novichok Nerve Agents. *Heliyon* **2020**, *6*, e03153.
(221) Harvey, J. A.; Greathouse, J. A.; Sava Gallis, D. F. Defect and Linker Effects on the Binding of Organophosphorous Compounds in UiO-66 and Rare-Earth MOFs. *J. Phys. Chem. C* **2018**, *122*, 26889–26896.
(222) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.
(223) Sava Gallis, D. F.; Harvey, J. A.; Pearce, C. J.; Hall, M. G.; DeCoste, J. B.; Kinnan, M. K.; Greathouse, J. A. Efficient MOF-Based Degradation of Organophosphorus Compounds in Non-Aqueous Environments. *J. Mater. Chem. A* **2018**, *6*, 3038–3045.
(224) Harvey, J. A.; Pearce, C. J.; Hall, M. G.; Bruni, E. J.; DeCoste, J. B.; Sava Gallis, D. F. Insights into the Solvent-Assisted Degradation of Organophosphorus Compounds by a Zr-Based Metal–Organic Framework. *Dalt. Trans.* **2019**, *48*, 16153–16157.
(225) Bueken, B.; Van Velthoven, N.; Krajnc, A.; Smolders, S.; Taulelle, F.; Mellot-Draznieks, C.; Mali, G.; Bennett, T. D.; De Vos, D. Tackling the Defect Conundrum in UiO-66: A Mixed-Linker Approach to Engineering Missing Linker Defects. *Chem. Mater.* **2017**, *29*, 10478–10486.
(226) Firth, F. C. N.; Cliffe, M. J.; Vulpe, D.; Aragones-Anglada, M.; Moghadam, P. Z.; Fairen-Jimenez, D.; Slater, B.; Grey, C. P. Engineering New Defective Phases of UiO Family Metal–Organic Frameworks with Water. *J. Mater. Chem. A* **2019**, *7*, 7459–7469.

(227) Vandichel, M.; Hajek, J.; Vermoortele, F.; Waroquier, M.; De Vos, D. E.; Van Speybroeck, V. Active Site Engineering in UiO-66 Type Metal–Organic Frameworks by Intentional Creation of Defects: A Theoretical Rationalization. *CrystEngComm* **2015**, *17*, 395–406.
(228) Kramida, A.; Ralchenko, Y.; Reader, J. NIST Atomic Spectra Database (ver. 5.7.1). https://physics.nist.gov/asd (accessed Mar 15, 2020).
(229) Wiberg, K. B. Application of the Pople-Santry-Segal CNDO Method to the Cyclopropylcarbinyl and Cyclobutyl Cation and to Bicyclobutane. *Tetrahedron* **1968**, *24*, 1083–1096.
(230) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623.
(231) Talete. Dragon Version 6.0. 2010. https://doi.org/http://www.talete.mi.it/.
(232) Todeschini, R.; Vighi, M.; Finizio, A.; Gramatica, P. 3D-Modelling and Prediction by WHIM Descriptors. Part 8. Toxicity and Physico-Chemical Properties of Environmental Priority Chemicals by 2D-TI and 3D-WHIM Descriptors. *SAR QSAR Environ. Res.* **1997**, *7*, 173–193.
(233) Cordero, B.; Gómez, V.; Platero-Prats, A. E.; Revés, M.; Echeverría, J.; Cremades, E.; Barragán, F.; Alvarez, S. Covalent Radii Revisited. *Dalt. Trans.* **2008**, 2832–2838.
(234) Mendonca, M. L.; Snurr, R. Q. Screening for Improved Nerve Agent Simulants and Insights into Organophosphate Hydrolysis Reactions from DFT and QSAR Modeling. *Chem. Eur. J.* **2019**, *25*, 9217–9229.
(235) Partington, J. R. *A History of Greek Fire and Gunpowder*; Johns Hopkins University Press: Baltimore, 1999.
(236) Fitzgerald, G. J. Chemical Warfare and Medical Response During World War I. *Am. J. Public Heal.* **2008**, *98*, 611–625.
(237) Mondloch, J. E.; Bury, W.; Fairen-Jimenez, D.; Kwon, S.; DeMarco, E. J.; Weston, M. H.; Sarjeant, A. A.; Nguyen, S. T.; Stair, P. C.; Snurr, R. Q.; Farha, O. K.; Hupp, J. T. Vapor-Phase Metalation by Atomic Layer Deposition in a Metal-Organic Framework. *J. Am. Chem. Soc.* **2013**, *135*, 10294–10297.
(238) Kim, I. S.; Li, Z.; Zheng, J.; Platero-Prats, A. E.; Mavrandonakis, A.; Pellizzeri, S.; Ferrandon, M.; Vjunov, A.; Gallington, L. C.; Webber, T. E.; Vermeulen, N. A.; Penn, R. L.; Getman, R. B.; Cramer, C. J.; Chapman, K. W.; Camaioni, D. M.; Fulton, J. L.; Lercher, J. A.; Farha, O. K.; Hupp, J. T.; Martinson, A. B. F. Sinter-Resistant Platinum Catalyst Supported by Metal–Organic Framework. *Angew. Chemie - Int. Ed.* **2018**, *57*, 909–913.
(239) Ikuno, T.; Zheng, J.; Vjunov, A.; Sanchez-Sanchez, M.; Ortuño, M. A.; Pahls, D. R.; Fulton, J. L.; Camaioni, D. M.; Li, Z.; Ray, D.; Mehdi, B. L.; Browning, N. D.; Farha, O. K.; Hupp, J. T.; Cramer, C. J.; Gagliardi, L.; Lercher, J. A. Methane Oxidation to Methanol Catalyzed by Cu-Oxo Clusters Stabilized in NU-1000 Metal−Organic Framework. *J. Am. Chem. Soc.* **2017**, *139*, 10294–10301.
(240) Islamoglu, T.; Goswami, S.; Li, Z.; Howarth, A. J.; Farha, O. K.; Hupp, J. T. Postsynthetic Tuning of Metal−Organic Frameworks for Targeted Applications. *Acc. Chem. Res.* **2017**, *50*, 805–813.
(241) Otake, K.; Cui, Y.; Buru, C. T.; Li, Z.; Hupp, J. T.; Farha, O. K. Single-Atom-Based

Vanadium Oxide Catalysts Supported on Metal−Organic Frameworks: Selective Alcohol Oxidation and Structure−Activity Relationship. *J. Am. Chem. Soc.* **2018**, *140*, 8652–8656.
(242) Ahn, S.; Thornburg, N. E.; Li, Z.; Wang, T. C.; Gallington, L. C.; Chapman, K. W.; Notestein, J. M.; Hupp, J. T.; Farha, O. K. Stable Metal−Organic Framework-Supported Niobium Catalysts. *Inorg. Chem.* **2016**, *55*, 11954–11961.
(243) Li, Z.; Peters, A. W.; Bernales, V.; Ortuño, M. A.; Schweitzer, N. M.; DeStefano, M. R.; Gallington, L. C.; Platero-Prats, A. E.; Chapman, K. W.; Cramer, C. J.; Gagliardi, L.; Hupp, J. T.; Farha, O. K. Metal−Organic Framework Supported Cobalt Catalysts for the Oxidative Dehydrogenation of Propane at Low Temperature. *ACS Cent. Sci.* **2017**, *3*, 31–38.
(244) Noh, H.; Cui, Y.; Peters, A. W.; Pahls, D. R.; Ortuño, M. A.; Vermeulen, N. A.; Cramer, C. J.; Gagliardi, L.; Hupp, J. T.; Farha, O. K. An Exceptionally Stable Metal-Organic Framework Supported Molybdenum(VI) Oxide Catalyst for Cyclohexene Epoxidation. *J. Am. Chem. Soc.* **2016**, *138*, 14720–14726.
(245) Ye, J.; Gagliardi, L.; Cramer, C. J.; Truhlar, D. G. Single Ni Atoms and Ni4 Clusters Have Similar Catalytic Activity for Ethylene Dimerization. *J. Catal.* **2017**, *354*, 278–286.
(246) Noh, H.; Kung, C.-W.; Otake, K.; Peters, A. W.; Li, Z.; Liao, Y.; Gong, X.; Farha, O. K.; Hupp, J. T. Redox-Mediator-Assisted Electrocatalytic Hydrogen Evolution from Water by a Molybdenum Sulfide-Functionalized Metal−Organic Framework. *ACS Catal.* **2018**, *8*, 9848–9858.
(247) Goetjen, T. A.; Zhang, X.; Liu, J.; Hupp, J. T.; Farha, O. K. Metal–Organic Framework Supported Single Site Chromium(III) Catalyst for Ethylene Oligomerization at Low Pressure and Temperature. *ACS Sustain. Chem. Eng.* **2019**, *7*, 2553–2557.
(248) Zheng, J.; Ye, J.; Ortuño, M. A.; Fulton, J. L.; Gutiérrez, O. Y.; Camaioni, D. M.; Motkuri, R. K.; Li, Z.; Webber, T. E.; Mehdi, B. L.; Browning, N. D.; Penn, R. L.; Farha, O. K.; Hupp, J. T.; Truhlar, D. G.; Cramer, C. J.; Lercher, J. A. Selective Methane Oxidation to Methanol on Cu-Oxo Dimers Stabilized by Zirconia Nodes of an NU-1000 Metal−Organic Framework. *J. Am. Chem. Soc.* **2019**, *141*, 9292–9304.
(249) Peters, A. W.; Otake, K.; Platero-Prats, A. E.; Li, Z.; DeStefano, M. R.; Chapman, K. W.; Farha, O. K.; Hupp, J. T. Site-Directed Synthesis of Cobalt Oxide Clusters in a Metal−Organic Framework. *ACS Appl. Mater. Interfaces* **2018**, *10*, 15073–15078.
(250) Pahls, D. R.; Ortuño, M. A.; Winegar, P. H.; Cramer, C. J.; Gagliardi, L. Computational Screening of Bimetal-Functionalized Zr6O8 MOF Nodes for Methane C–H Bond Activation. *Inorg. Chem.* **2017**, *56*, 8739–8743.
(251) Ye, J.; Gagliardi, L.; Cramer, C. J.; Truhlar, D. G. Computational Screening of MOF-Supported Transition Metal Catalysts for Activity and Selectivity in Ethylene Dimerization. *J. Catal.* **2018**, *360*, 160–167.
(252) Ortuño, M. A.; Bernales, V.; Gagliardi, L.; Cramer, C. J. Computational Study of First-Row Transition Metals Supported on MOF NU-1000 for Catalytic Acceptorless Alcohol Dehydrogenation. *J. Phys. Chem. C* **2016**, *120*, 24697–24705.
(253) Simons, M. C.; Ortuño, M. A.; Bernales, V.; Gaggioli, C. A.; Cramer, C. J.; Bhan, A.; Gagliardi, L. C–H Bond Activation on Bimetallic Two-Atom Co-M Oxide Clusters Deposited on Zr-Based MOF Nodes: Effects of Doping at the Molecular Level. *ACS Catal.* **2018**, *8*, 2864–2869.
(254) Bernales, V.; Ortuño, M. A.; Truhlar, D. G.; Cramer, C. J.; Gagliardi, L. Computational

Design of Functionalized Metal−Organic Framework Nodes for Catalysis. *ACS Cent. Sci.* **2018**, *4*, 5–19.
(255) Klet, R. C.; Wang, T. C.; Fernandez, L. E.; Truhlar, D. G.; Hupp, J. T.; Farha, O. K. Synthetic Access to Atomically Dispersed Metals in Metal−Organic Frameworks via a Combined Atomic-Layer-Deposition-in-MOF and Metal-Exchange Approach. *Chem. Mater.* **2016**, *28*, 1213–1219.
(256) Zhao, Y.; Truhlar, D. G. Applications and Validations of the Minnesota Density Functionals. *Chem. Phys. Lett.* **2011**, *502*, 1–13.
(257) Li, Z.; Schweitzer, N. M.; League, A. B.; Bernales, V.; Peters, A. W.; Getsoian, A. "Bean"; Wang, T. C.; Miller, J. T.; Vjunov, A.; Fulton, J. L.; Lercher, J. A.; Cramer, C. J.; Gagliardi, L.; Hupp, J. T.; Farha, O. K. Sintering-Resistant Single-Site Nickel Catalyst Supported by Metal-Organic Framework. *J. Am. Chem. Soc.* **2016**, *138*, 1977–1982.
(258) Kim, I. S.; Borycz, J.; Platero-Prats, A. E.; Tussupbayev, S.; Wang, T. C.; Farha, O. K.; Hupp, J. T.; Gagliardi, L.; Chapman, K. W.; Cramer, C. J.; Martinson, A. B. F. Targeted Single-Site MOF Node Modification: Trivalent Metal Loading via Atomic Layer Deposition. *Chem. Mater.* **2015**, *27*, 4772–4778.
(259) Balow, R. B.; Lundin, J. G.; Daniels, G. C.; Gordon, W. O.; McEntee, M.; Peterson, G. W.; Wynne, J. H.; Pehrsson, P. E. Environmental Effects on Zirconium Hydroxide Nanoparticles and Chemical Warfare Agent Decomposition: Implications of Atmospheric Water and Carbon Dioxide. *ACS Appl. Mater. Interfaces* **2017**, *9*, 39747–39757.
(260) Kaledin, A. L.; Driscoll, D. M.; Troya, D.; Collins-Wildman, D. L.; Hill, C. L.; Morris, J. R.; Musaev, D. G. Impact of Ambient Gases on the Mechanism of [Cs8Nb6O19]-Promoted Nerve-Agent Decomposition. *Chem. Sci.* **2018**, *9*, 2147–2158.
(261) Daniel, K. A.; Kopff, L. A.; Patterson, E. V. Computational Studies on the Solvolysis of the Chemical Warfare Agent VX. *J. Phys. Org. Chem.* **2008**, *21*, 321–328.
(262) Kozuch, S.; Martin, J. M. L. The Rate-Determining Step Is Dead. Long Live the Rate-Determining State! *ChemPhysChem* **2011**, *12*, 1413–1418.
(263) Falivene, L.; Kozlov, S. M.; Cavallo, L. Constructing Bridges between Computational Tools in Heterogeneous and Homogeneous Catalysis. *ACS Catal.* **2018**, *8*, 5637–5656.
(264) Kozuch, S.; Shaik, S. How to Conceptualize Catalytic Cycles? The Energetic Span Model. *Acc. Chem. Res.* **2011**, *44*, 101–110.
(265) Kozuch, S.; Shaik, S. Kinetic-Quantum Chemical Model for Catalytic Cycles: The Haber-Bosch Process and the Effect of Reagent Concentration. *J. Phys. Chem. A* **2008**, *112*, 6032–6041.
(266) Yuan, S.; Peng, J.; Zhang, Y.; Shao-Horn, Y. Stability Trend of Metal−Organic Frameworks with Heterometal-Modified Hexanuclear Zr Building Units. *J. Phys. Chem. C* **2019**, *123*, 28266–28274.
(267) Beyzavi, M. H.; Vermeulen, N. A.; Howarth, A. J.; Tussupbayev, S.; League, A. B.; Schweitzer, N. M.; Gallagher, J. R.; Platero-Prats, A. E.; Hafezi, N.; Sarjeant, A. A.; Miller, J. T.; Chapman, K. W.; Stoddart, J. F.; Cramer, C. J.; Hupp, J. T.; Farha, O. K. A Hafnium-Based Metal-Organic Framework as a Nature-Inspired Tandem Reaction Catalyst. *J. Am. Chem. Soc.* **2015**, *137*, 13624–13631.
(268) Chen, Z.; Wang, X.; Noh, H.; Ayoub, G.; Peterson, G. W.; Buru, C. T.; Islamoglu, T.; Farha, O. K. Scalable, Room Temperature, and Water-Based Synthesis of Functionalized Zirconium-Based Metal–Organic Frameworks for Toxic Chemical Removal.

CrystEngComm **2019**, *21*, 2409–2415.
(269) Doan, H. A.; Li, Z.; Farha, O. K.; Hupp, J. T.; Snurr, R. Q. Theoretical Insights into Direct Methane to Methanol Conversion over Supported Dicopper Oxo Nanoclusters. *Catal. Today* **2018**, *312*, 2–9.
(270) Bligaard, T.; Nørskov, J. K.; Dahl, S.; Matthiesen, J.; Christensen, C. H.; Sehested, J. The Brønsted–Evans–Polanyi Relation and the Volcano Curve in Heterogeneous Catalysis. *J. Catal.* **2004**, *224*, 206–217.
(271) Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H. Towards the Computational Design of Solid Catalysts. *Nat. Chem.* **2009**, *1*, 37–46.
(272) Kalaj, M.; Momeni, M. R.; Bentz, K. C.; Barcus, K. S.; Palomba, J. M.; Paesani, F.; Cohen, S. M. Halogen Bonding in UiO-66 Frameworks Promotes Superior Chemical Warfare Agent Simulant Degradation. *Chem. Commun.* **2019**, *55*, 3481–3484.
(273) Tian, Y.; Plonka, A. M.; Ebrahim, A. M.; Palomino, R. M.; Senanayake, S. D.; Balboa, A.; Gordon, W. O.; Troya, D.; Musaev, D. G.; Morris, J. R.; Mitchell, M. B.; Collins-Wildman, D. L.; Hill, C. L.; Frenkel, A. I. Correlated Multimodal Approach Reveals Key Details of Nerve-Agent Decomposition by Single-Site Zr-Based Polyoxometalates. *J. Phys. Chem. Lett.* **2019**, *10*, 2295–2299.
(274) Garibay, S. J.; Farha, O. K.; DeCoste, J. B. Single-Component Frameworks for Heterogeneous Catalytic Hydrolysis of Organophosphorous Compounds in Pure Water. *Chem. Commun.* **2019**, *55*, 7005–7008.

Appendix A: Supporting Information for Chapter 2

This appendix is a modified version of the Supporting Information for the manuscript in Chapter 2 (*Chem. Eur. J.* **2019**, *25*, 9217–9229).

A.1 Density Functional Sensitivity

Table A.1. Free energy barriers for the first transition state (ΔG_{TS1}) in the alkaline hydrolysis reaction of sarin, obtained at T = 298.15 K and P = 1 atm.

Single point energy//geometry	ΔG_{TS1} (kJ/mol)
B3LYP/6-311++G**//B3LYP/6-311++G**	56.16
B3LYP-D3(BJ)/6-311++G**//B3LYP/6-311++G**	44.32
MP2/6-311++G**//B3LYP/6-311++G**	46.17
HF/6-31+G**//HF/6-31+G**	79.01
M06-2X/6-311++G**//M06-2X/6-311++G**	29.20
CBS-QB3	50.54
Experiment[(a)]	65.53

(a) Experimental data obtained from Larsson, L. *Acta. Chem. Scand*. **1957**, *11*, 1131-1142.

Comparing the free energy barriers using different levels of theory in Table A.1, we see that B3LYP corresponds to the best agreement with experiment. The B3LYP method also performs similar to the high-level CBS-QB3 method, which further inspires confidence in using B3LYP to study uncatalyzed organophosphate hydrolysis reactions. Thus, this method was used to screen for improved nerve agent simulants in our study.

A.2 Comparison to Experimental Results

Next, we calculated the free energy barriers for several other molecules for which experimental data was available. This was done to put our screening results into perspective and to obtain a rough estimate of expected deviations from experiment.

Table A.2. Free energy barriers for the first transition state (ΔG_{TS1}) in the alkaline hydrolysis reaction of several organophosphate molecules, obtained at T = 298.15 K and P = 1 atm. All energies are in kJ/mol.

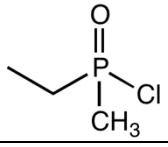
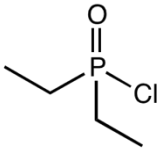
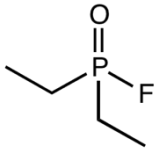
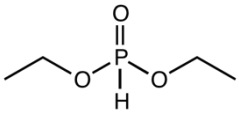
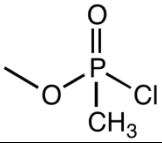
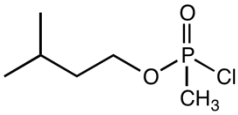
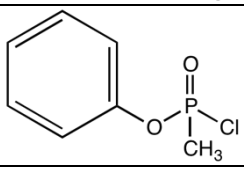
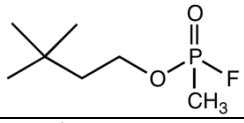
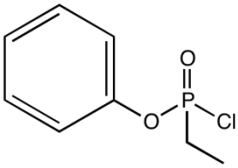
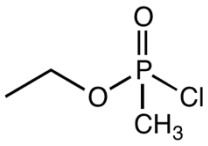
Molecule	B3LYP	Expt.[a]	Deviation	Rank B3LYP	Rank Expt.
*iso*Propoxy-methyl-phosphoryl fluoride (Sarin)	56.16	65.53	-9.38	3	3
Methoxy-methyl-phosphoryl fluoride	51.18	61.4	-10.23	6	7
Ethoxy-methyl-phosphoryl fluoride	52.05	63.09	-11.05	5	6
n-Propoxy-methyl-phosphoryl fluoride	52.76	63.49	-10.73	4	5
3,3-Dimethylbutoxy-methyl-phosphoryl fluoride	49.08	63.95	-14.87	7	4
*iso*Propoxy-ethyl-phosphoryl fluoride	56.61	67.24	-10.63	2	2
*iso*Propoxy-*iso*propyl-phosphoryl fluoride	60.12	70.94	-10.82	1	1

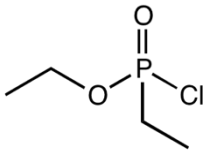
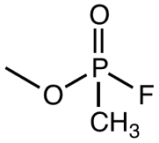
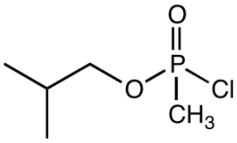
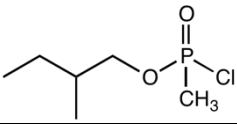
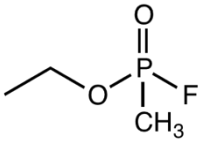
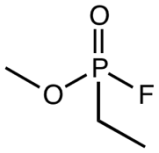
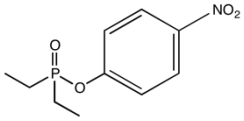
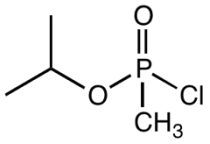
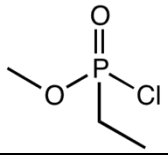
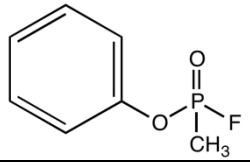
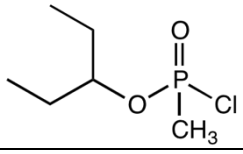
(a) Experimental data obtained from Larsson, L. *Acta. Chem. Scand*. **1957**, *11*, 1131-1142.

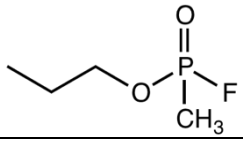
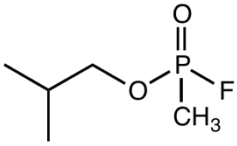
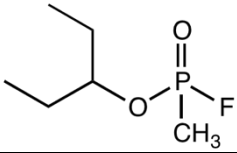
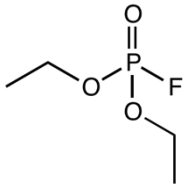
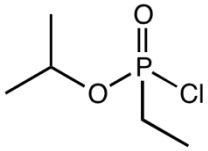
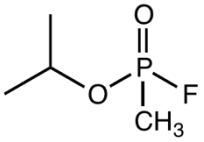
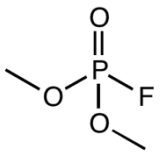
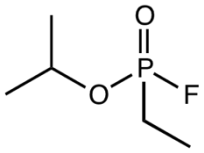
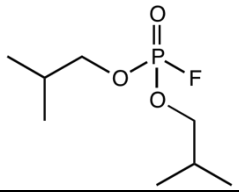
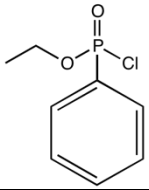
Comparing the B3LYP free energy barriers to those from experiment in Table A.2, the average deviation is -11.10 $\pm$ 1.62 kJ/mol, with the calculated barriers lower than experiment in all cases. Comparing the ordering of the molecules, the Spearman's rank correlation coefficient is 0.79.

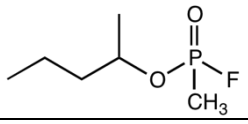
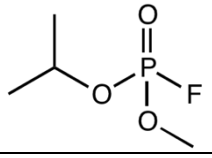
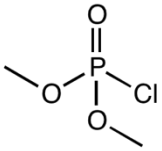
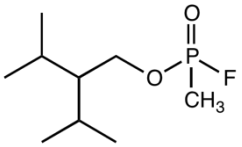
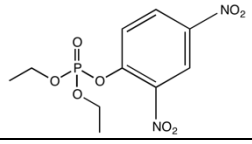
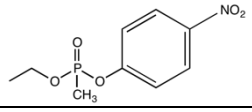
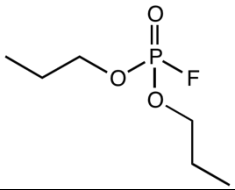
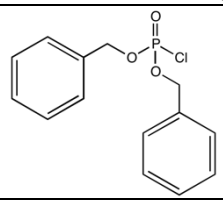
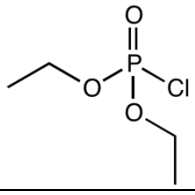
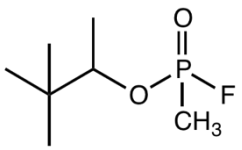
A.3 DFT Screening Results

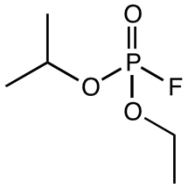
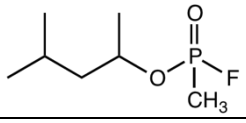
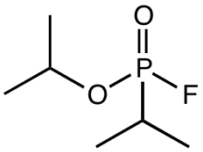
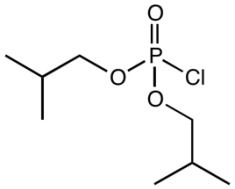
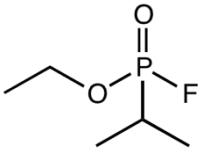
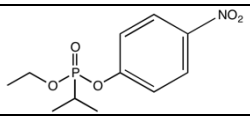
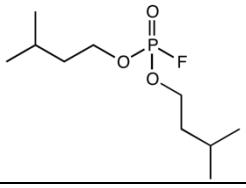
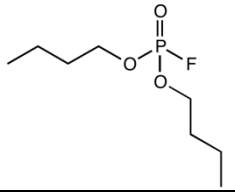
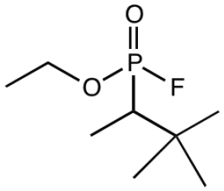
Table A.3. Organophosphate molecules analyzed in this study.

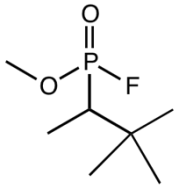
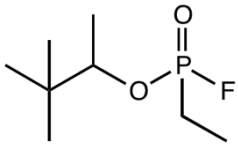
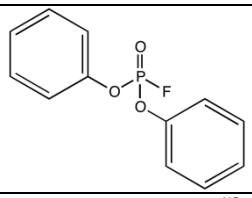
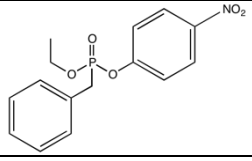
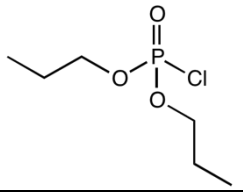
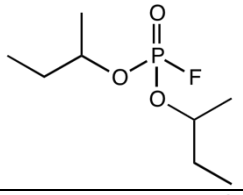
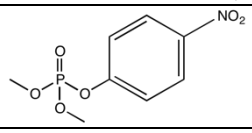
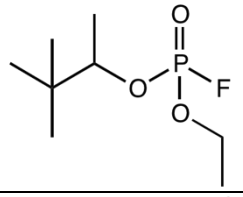
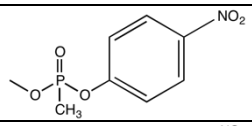
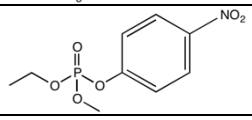
Molecule #[a]	Database ID[b]	Structure[c]	ΔG_{TS1} (kJ/mol)[d]	Rat Oral LD_{50} log_{10} (mol/kg)[e]
1	CAS RN: 13213-38-2		39.20	2.40[g]
2	CAS RN: 1112-37-4		42.30	2.68[g]
3	CAS RN: 756-78-5		45.36	2.38[g]
4	CAS RN: 762-04-9		47.06	1.55[h]
5	CAS RN: 1066-52-0		48.57	2.45[g]
6	"Hypothetical"		48.83	2.80[g]
7	CAS RN: 14235-74-6		48.86	2.63[g]
8	CAS RN: 660-21-9		49.08	5.05[g]
9	"Hypothetical"		50.47	3.26[g]
10	CAS RN: 5284-09–3		50.73	2.93[g]

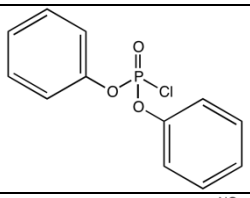
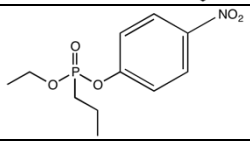
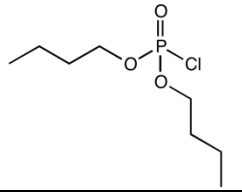
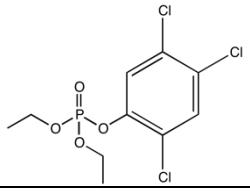
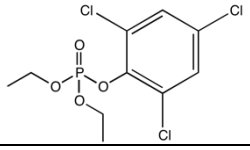
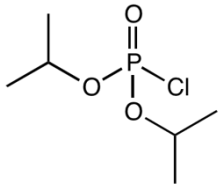
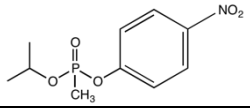
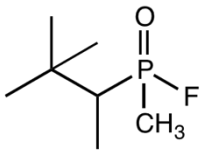
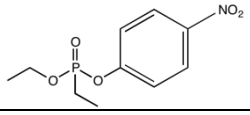
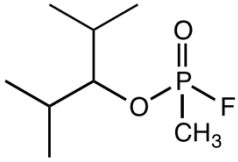
11	CAS RN: 5284-10–6		51.12	3.44[g]
12	CAS RN: 353-88-8		51.18	2.13[g]
13	CAS RN: 18359-05-2		51.22	3.23[g]
14	“Hypothetical”		51.45	3.47[g]
15	CAS RN: 673-97-2		52.05	2.19[g]
16	CAS RN: 665-03-2		52.10	2.46[g]
17	CAS RN: 7531-39-7		52.33	4.84[g]
18	CAS RN: 1445-76-7		52.40	2.80[g]
19	CAS RN: 21502-57-8		52.41	2.83[g]
20	CAS RN: 133826-40-1		52.51	2.50[g]
21	“Hypothetical”		52.67	3.11[g]

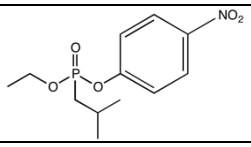
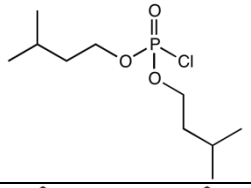
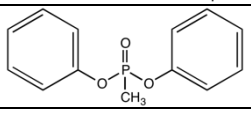
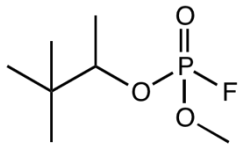
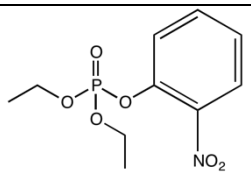
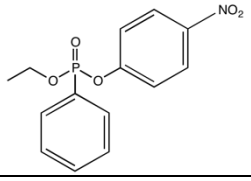
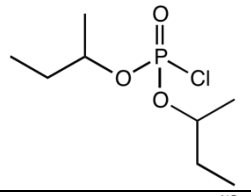
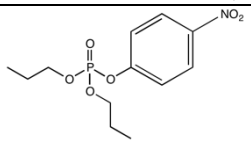
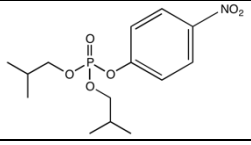
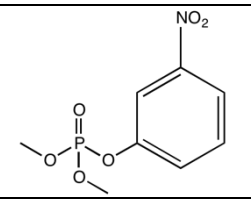
22	CAS RN: 763-14-4		52.76	2.27[g]
23	CAS RN: 2053-81-8		53.47	3.55[g]
24	CAS RN: 66348-71-8		55.06	3.99[g]
25	CAS RN: 358-74-7		55.58	4.01[g]
26	CAS RN: 28829-95-0		55.72	3.73[g]
27	CAS RN: 107-44-8		56.16	5.41[h]
28	CAS RN: 5954-50-7		56.21	2.26[g]
29	CAS RN: 1189-87-3		56.61	4.11[g]
30	CAS RN: 563-22-4		57.11	3.74[g]
31	CAS RN: 5284-12-8		57.20	3.37[g]

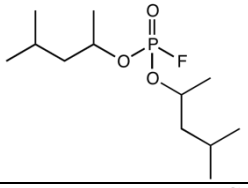
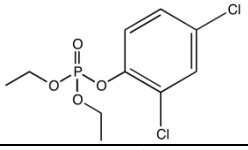
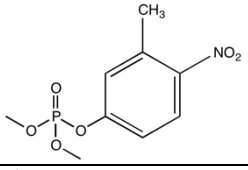
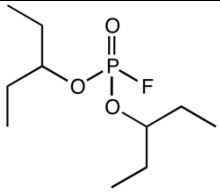
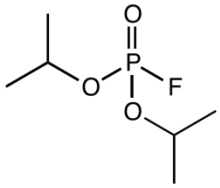
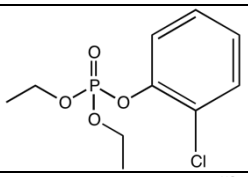
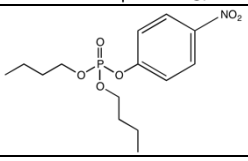
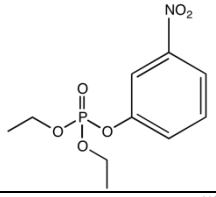
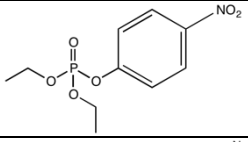
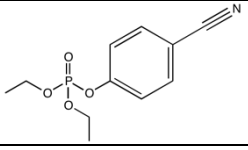
32	CAS RN: 761-93-3		57.50	3.81[g]
33	CAS RN: 2276-27-9		57.50	3.79[g]
34	CAS RN: 813-77-4		57.60	2.26[g]
35	CAS RN: 113548-88-2		57.87	3.81[g]
36	CAS RN: 54436-53-2		58.59	4.99[g]
37	CAS RN: 3735-98-6		58.74	5.18[g]
38	CAS RN: 381-45-3		58.96	3.66[g]
39	CAS RN: 538-37-4		59.29	2.01[g]
40	CAS RN: 814-49-3		59.57	4.20[h]
41	CAS RN: 96-64-0		59.74	5.66[h]

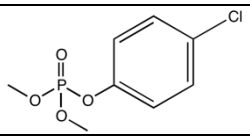
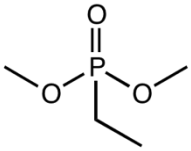
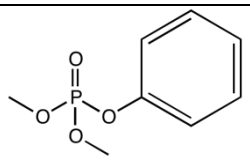
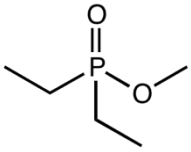
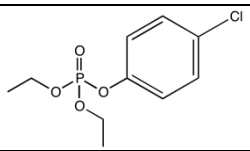
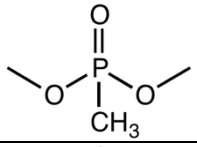
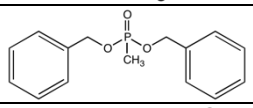
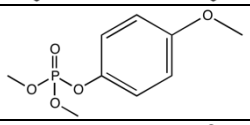
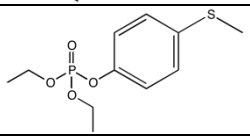
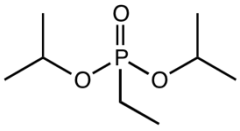
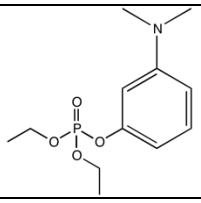
42	ChemSpider ID: 21252683		59.77	4.45[g]
43	CAS RN: 352-53-4		60.05	4.64[g]
44	CAS RN: 665-33-8		60.12	4.25[g]
45	CAS RN: 17158-87-1		60.12	2.89[g]
46	CAS RN: 1426-08-0		60.26	3.70[g]
47	CAS RN: 13538-10-8		60.65	4.90[g]
48	PubChem CID: 129642440[f]		60.93	2.07[g]
49	CAS RN: 674-48-6		61.02	3.33[g]
50	“Hypothetical”		61.17	3.63[g]

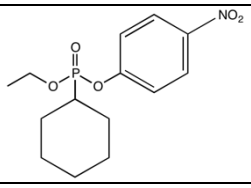
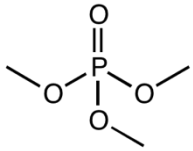
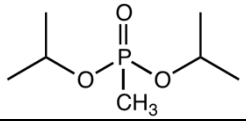
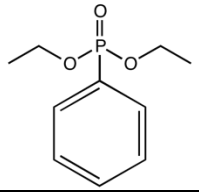
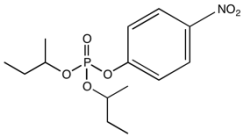
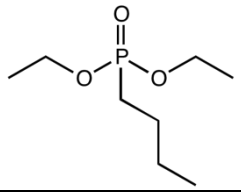
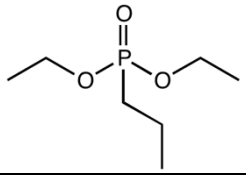
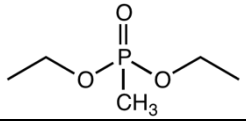
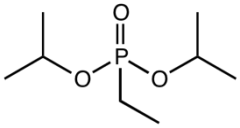
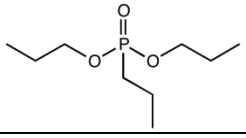
51	"Hypothetical"		61.52	4.24[g]
52	CAS RN: 97931-20-9		61.67	4.96[g]
53	CAS RN: 403-65-6		61.86	2.46[g]
54	CAS RN: 3015-70-1		62.76	3.72[g]
55	CAS RN: 2510-89-6		63.07	3.52[g]
56	CAS RN: 625-17-2		63.93	3.75[g]
57	CAS RN: 950-35-6		64.26	4.88[h]
58	"Hypothetical"		64.42	4.46[g]
59	ChemSpider ID: 9228613		64.90	4.73[g]
60	CAS RN: 1021-47-2		64.98	4.92[g]

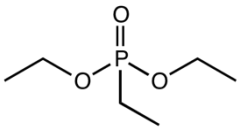
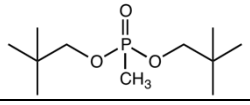
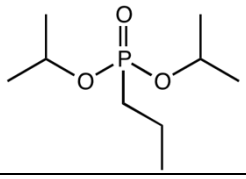
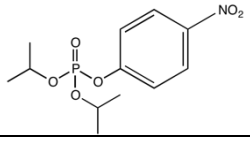
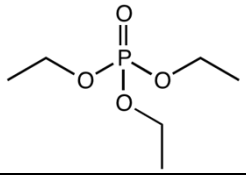
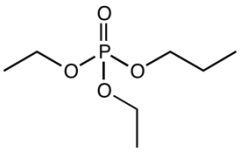
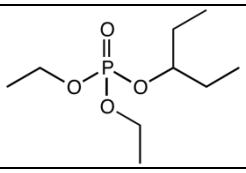
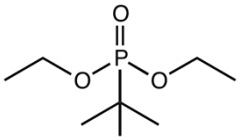
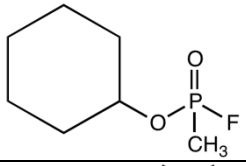
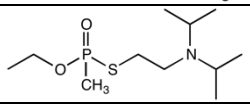
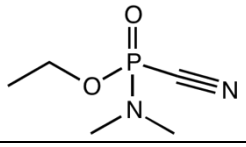
61	CAS RN: 2524-64-3		65.15	2.08[g]
62	CAS RN: 3015-73-4		65.23	5.02[g]
63	CAS RN: 819-43-2		65.28	2.87[g]
64	CAS RN: 20362-80-5		66.17	3.16[g]
65	“Hypothetical”		66.67	2.75[g]
66	CAS RN: 2574-25-6		66.68	3.51[g]
67	CAS RN: 3735-97-5		66.79	4.85[g]
68	PubChem CID: 58861548		66.80	4.31[g]
69	CAS RN: 546-71-4		66.87	5.08[g]
70	CAS RN: 113548-85-9		66.99	4.00[g]

71	CAS RN: 13538-11-9		67.02	4.72[g]
72	“Hypothetical”		67.55	2.07[g]
73	CAS RN: 7526-26-3		67.70	3.03[h]
74	“Hypothetical”		67.92	3.92[g]
75	CAS RN: 4532-02–9		67.93	4.12[g]
76	CAS RN: 2012-00-2		68.38	4.24[g]
77	CAS RN: 54757-38-9		69.21	3.77[g]
78	CAS RN: 1153-30-6		69.27	4.44[g]
79	“Hypothetical”		69.30	3.87[g]
80	InChI Key: WVAUFDIKPLFXQL-UHFFFAOYSA-N		69.54	3.50[g]

81	CAS RN: 311-60-4		69.54	2.23[g]
82	CAS RN: 3279-62-7		69.84	3.48[h]
83	CAS RN: 2255-17-6		70.83	4.04[h]
84	"Hypothetical"		71.27	2.72[g]
85	CAS RN: 55-91-4		71.67	4.57[h]
86	CAS RN: 16462-86-5		71.91	2.92[g]
87	CAS RN: 2255-19-8		71.93	3.83[g]
88	CAS RN: 4532-06–3		72.24	4.04[g]
89	CAS RN: 311-45-5		72.29	5.18[h]
90	CAS RN: 6132-16-7		73.00	4.17[g]

91	CAS RN: 18264-30-7		74.03	3.28[g]
92	CAS RN: 6163-75-3		74.16	2.45[g]
93	CAS RN: 10113-28-7		74.16	2.53[g]
94	CAS RN: 5689-41-8		74.46	2.24[g]
95	CAS RN: 5076-63-1		75.75	3.29[g]
96	CAS RN: 756-79-6		78.38	1.18[h]
97	CAS RN: 19236-58-9		79.80	1.87[g]
98	CAS RN: 7357-14-4		80.79	3.01[g]
99	CAS RN: 3070-13-1		80.93	4.36[g]
100	CAS RN: 2510-86-3		84.09	2.71[g]
101	CAS RN: 4619-09–4		84.24	2.74[g]

102	CAS RN: 13538-15-3		86.19	4.42[g]
103	CAS RN: 512-56-1		87.69	2.22[h]
104	CAS RN: 1445-75-6		88.04	2.34[h]
105	CAS RN: 1754-49-0		88.24	3.50[g]
106	“Hypothetical”		89.08	4.48[g]
107	CAS RN: 2404-75-3		89.59	3.05[g]
108	CAS RN: 18812-51-6		90.27	2.61[g]
109	CAS RN: 683-08-9		90.80	2.90[g]
110	CAS RN: 1067-69-2		91.11	2.86[g]
111	CAS RN: 1789-95-3		91.67	3.01[g]

112	CAS RN: 78-38-6		91.97	1.85[h]
113	CAS RN: 53803-21-7		93.92	2.83[g]
114	CAS RN: 18812-55-0		95.30	2.88[g]
115	CAS RN: 3254-66-8		97.30	4.67[g]
116	CAS RN: 78-40-0		98.39	2.19[h]
117	CAS RN: 814-22-2		100.41	3.30[g]
118	“Hypothetical”		104.55	3.27[g]
119	CAS RN: 19935-93-4		128.28	2.66[g]
120	CAS RN: 329-99-7		57.83	N/A[i]
121	CAS RN: 50782-69-9		83.91	N/A[i]
122	CAS RN: 77-81-6		72.32	N/A[i]

(a) Numbers used to identify molecules for QSAR analysis. Molecules 1-119 are numbered in order of increasing ΔG_{TS1} values, and 120-122 correspond to external set molecules in the QSAR analysis. (b) Identifiers, such as CAS

Registry Numbers, that can be used to easily look up the molecules in chemical databases. “Hypothetical” indicates a hypothetical molecule for which a unique chemical identifier could not be found. (c) All molecules are drawn with their considered leaving group on the right-hand side. For molecules that have a stereocenter at the phosphorous atom, the S_P-enantiomer was studied. (d) Free energy barriers for the first (or only) transition state in the hydrolysis reactions, calculated with DFT at T = 298.15 K using B3LYP/6-311++G**. (e) Toxicity endpoint, which represents the mass of the compound per rat body weight that causes death in 50% of rats after oral ingestion. Values were obtained using the EPA Toxicity Estimation Software Tool (T.E.S.T. v 4.2).[1] (f) Molecule originally referenced in Cook et al.[2] (g) Predicted toxicity values calculated using the consensus QSAR method in the T.E.S.T. software. (h) Experimental toxicity values from the T.E.S.T. database. (i) These nerve agents were only used as an external set in QSAR analysis, so toxicities are omitted here.

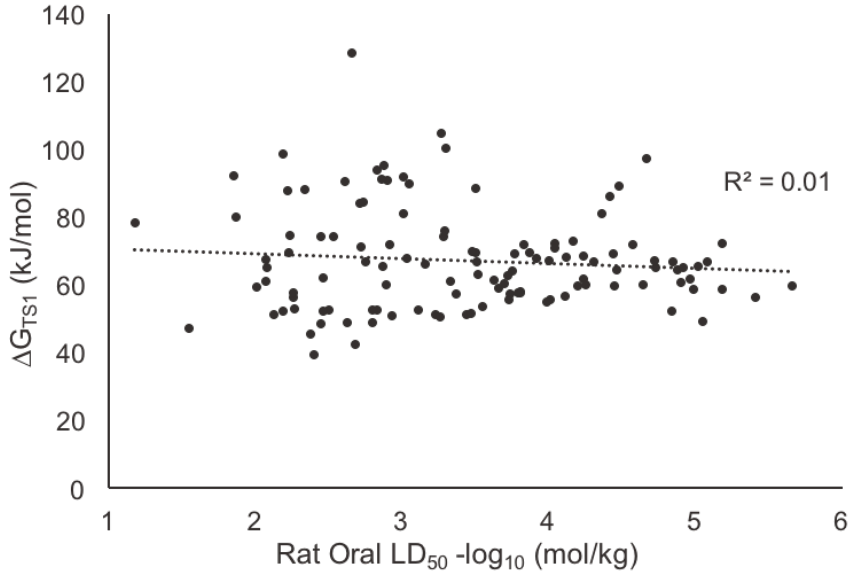


Figure A.1. DFT-calculated free energy barriers for the first (or only) transition state in the hydrolysis reactions for 119 organophosphate molecules vs. their toxicity endpoint. The negligible correlation coefficient indicates no correlation.

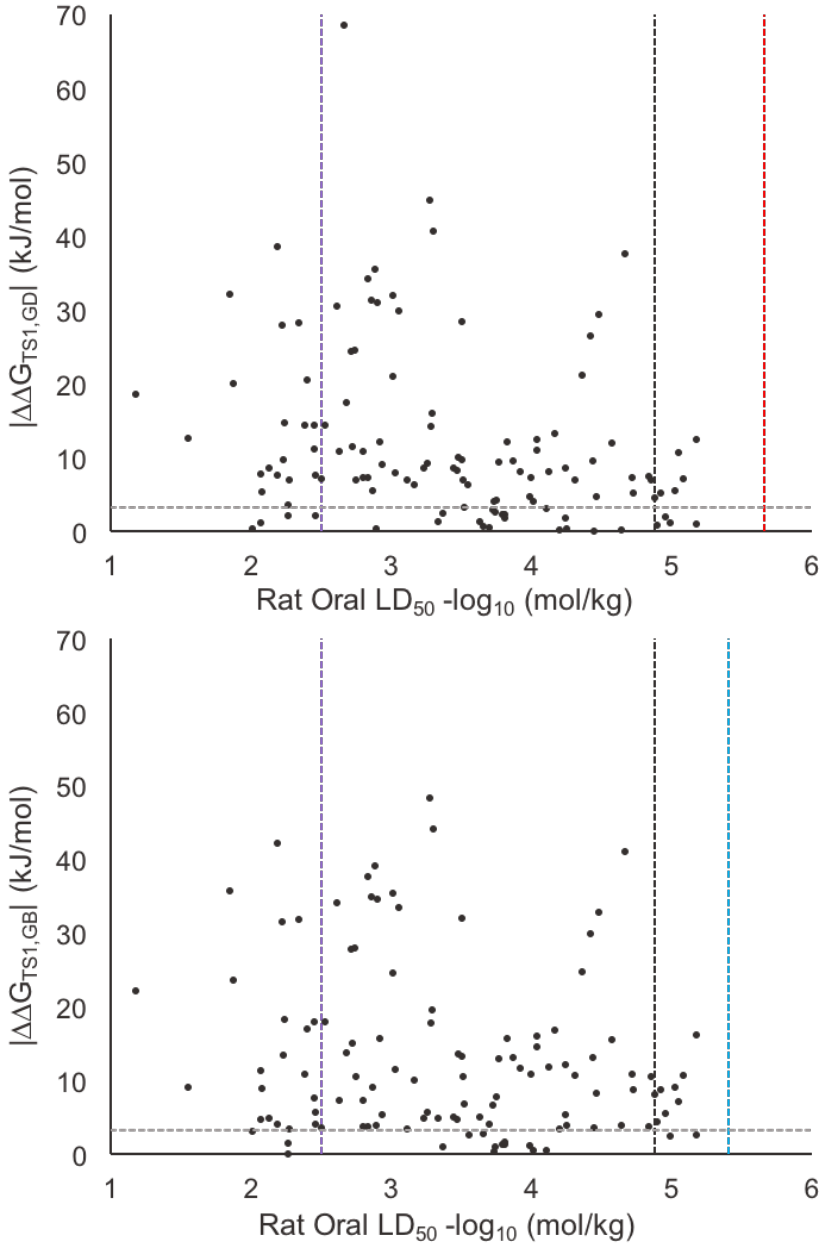


Figure A.2. Differences in DFT-calculated hydrolysis free energy barriers for the first transition state, $|\Delta\Delta G_{TS1}|$, for simulants relative to (top) soman and (bottom) sarin vs. their toxicity endpoint. The horizontal grey line indicates $|\Delta\Delta G_{TS1}| < 3.25$ kJ/mol as the energy barrier criterion for an improved simulant. The vertical purple line indicates $LD_{50} - \log_{10} < 2.50$ mol/kg as the toxicity criterion. The black vertical line indicates $LD_{50} - \log_{10} = 4.88$ mol/kg (toxicity of DMNP). The red vertical line indicates $LD_{50} - \log_{10} = 5.66$ mol/kg (toxicity of soman). The blue vertical line indicates $LD_{50} - \log_{10} = 5.41$ mol/kg (toxicity of sarin).

A.4 QSAR Analysis – Molecular Descriptors

Table A.4. Unscaled molecular descriptors used for QSAR analysis, calculated using DFT with B3LYP/6-311++G**.

Molecule #	q_p[a]	μ[b]	IP[c]	EA[d]	BO[e]	pK_a[f]	η[g]	SOF[h]	χ[i]	ω[j]	V[k]	f_p^{+1}
1	1.825	6.132	0.301	0.038	1.240	-8.00[m]	0.131	7.612	0.169	0.109	92.541	0.201
2	1.835	6.139	0.297	0.039	1.240	-8.00	0.129	7.746	0.168	0.110	98.724	0.196
3	2.121	6.029	0.297	0.032	1.259	3.17	0.133	7.543	0.165	0.102	93.178	0.062
4	2.162	6.740	0.319	0.026	1.258	16.00	0.146	6.843	0.173	0.102	106.886	0.002
5	2.042	4.170	0.314	0.038	1.276	-8.00	0.138	7.240	0.176	0.112	86.690	0.218
6	2.048	4.497	0.312	0.037	1.273	-8.00	0.137	7.290	0.174	0.111	149.606	0.202
7	2.035	4.041	0.263	0.037	1.284	-8.00	0.113	8.849	0.150	0.100	130.833	-0.001
8	2.332	4.174	0.309	0.027	1.284	3.17	0.141	7.070	0.168	0.100	132.891	0.089
9	2.043	4.059	0.264	0.037	1.285	-8.00[m]	0.113	8.840	0.150	0.100	143.710	-0.002
10	2.047	4.445	0.312	0.037	1.273	-8.00[m]	0.137	7.275	0.175	0.111	99.021	0.206
11	2.056	4.441	0.312	0.038	1.273	-8.00[m]	0.137	7.312	0.175	0.112	98.906	0.199
12	2.328	4.087	0.325	0.032	1.289	3.17[m]	0.146	6.843	0.179	0.109	80.506	0.144
13	2.050	4.597	0.311	0.037	1.272	-8.00[m]	0.137	7.301	0.174	0.111	132.260	0.196
14	2.049	4.591	0.311	0.037	1.272	-8.00[m]	0.137	7.291	0.174	0.111	143.990	0.201
15	2.332	4.335	0.322	0.032	1.286	3.17[m]	0.145	6.895	0.177	0.108	89.353	0.110
16	2.332	4.094	0.326	0.032	1.288	3.17[m]	0.147	6.806	0.179	0.109	89.669	0.089
17	2.101	10.150	0.267	0.113	1.223	7.15[n]	0.077	13.029	0.190	0.235	173.265	-0.005
18	2.053	4.473	0.308	0.037	1.268	-8.00[m]	0.135	7.380	0.173	0.110	117.064	0.200
19	2.051	4.191	0.314	0.039	1.276	-8.00[m]	0.137	7.280	0.176	0.113	98.186	0.212
20	2.330	3.941	0.264	0.038	1.305	3.17[m]	0.113	8.846	0.151	0.101	122.275	0.001
21	2.055	4.614	0.304	0.037	1.267	-8.00[m]	0.133	7.492	0.171	0.109	147.963	0.194
22	2.332	4.449	0.321	0.031	1.286	3.17[m]	0.145	6.897	0.176	0.107	105.493	0.101
23	2.333	4.478	0.319	0.032	1.284	3.17[m]	0.144	6.963	0.176	0.108	115.552	0.071
24	2.336	4.475	0.310	0.032	1.280	3.17[m]	0.139	7.178	0.171	0.105	133.889	0.069
25	2.544	4.863	0.324	0.030	1.302	3.17[m]	0.147	6.783	0.177	0.106	101.585	0.246
26	2.062	4.457	0.308	0.038	1.268	-8.00[m]	0.135	7.405	0.173	0.111	123.914	0.195
27	2.335	4.343	0.320	0.031	1.281	3.17[m]	0.144	6.934	0.176	0.107	99.326	0.096
28	2.540	4.500	0.329	0.033	1.309	3.17[m]	0.148	6.764	0.181	0.111	82.956	0.429
29	2.339	4.324	0.318	0.031	1.280	3.17[m]	0.144	6.965	0.175	0.106	112.523	0.053
30	2.547	4.709	0.318	0.030	1.299	3.17[m]	0.144	6.936	0.174	0.105	161.116	0.412
31	2.074	5.071	0.277	0.068	1.273	-8.00[m]	0.105	9.562	0.172	0.142	147.365	0.079
32	2.335	4.403	0.316	0.031	1.280	3.17[m]	0.142	7.019	0.173	0.106	135.942	0.082
33	2.546	4.427	0.323	0.032	1.300	3.17[m]	0.146	6.870	0.178	0.108	111.084	0.181
34	2.250	4.200	0.321	0.041	1.295	-8.00[m]	0.140	7.123	0.181	0.117	79.340	0.257

35	2.333	4.632	0.303	0.031	1.283	3.17[m]	0.136	7.362	0.167	0.103	175.072	0.066
36	2.537	7.356	0.289	0.126	1.298	4.07[n]	0.082	12.243	0.207	0.263	208.747	-0.003
37	2.313	11.125	0.272	0.117	1.265	7.15[n]	0.078	12.902	0.195	0.244	176.173	-0.003
38	2.545	4.773	0.322	0.029	1.300	3.17[m]	0.146	6.846	0.176	0.105	146.187	0.443
39	2.256	5.141	0.265	0.054	1.290	-8.00[m]	0.105	9.493	0.160	0.121	207.542	0.081
40	2.261	4.704	0.317	0.039	1.290	-8.00[m]	0.139	7.174	0.178	0.114	112.930	0.228
41	2.338	4.516	0.309	0.031	1.280	3.17[m]	0.139	7.206	0.170	0.104	150.855	0.054
42	2.546	4.913	0.320	0.031	1.297	3.17[m]	0.144	6.929	0.175	0.107	120.939	0.037
43	2.336	4.448	0.313	0.031	1.280	3.17[m]	0.141	7.079	0.172	0.105	122.578	0.082
44	2.351	4.480	0.315	0.030	1.279	3.17[m]	0.142	7.030	0.173	0.105	134.645	0.021
45	2.265	4.714	0.312	0.039	1.286	-8.00[m]	0.137	7.308	0.176	0.113	180.447	0.215
46	2.347	4.487	0.316	0.028	1.284	3.17[m]	0.144	6.955	0.172	0.103	121.712	0.168
47	2.336	9.027	0.270	0.114	1.258	7.15[n]	0.078	12.790	0.192	0.235	178.320	-0.002
48	2.548	5.044	0.313	0.029	1.298	3.17[m]	0.142	7.044	0.171	0.102	184.604	0.215
49	2.546	4.928	0.318	0.029	1.300	3.17[m]	0.145	6.915	0.173	0.104	175.150	0.493
50	2.358	4.319	0.310	0.029	1.274	3.17[m]	0.140	7.119	0.169	0.102	148.036	0.007
51	2.354	4.327	0.312	0.028	1.279	3.17[m]	0.142	7.052	0.170	0.102	144.605	0.130
52	2.343	4.476	0.309	0.031	1.278	3.17[m]	0.139	7.205	0.170	0.104	155.309	0.036
53	2.557	5.845	0.264	0.041	1.333	3.17[m]	0.111	8.979	0.153	0.104	175.290	0.032
54	2.332	8.750	0.256	0.114	1.256	7.15[n]	0.071	14.073	0.185	0.240	261.064	-0.003
55	2.263	4.837	0.317	0.039	1.289	-8.00[m]	0.139	7.198	0.178	0.114	154.169	0.222
56	2.552	5.328	0.314	0.029	1.293	3.17[m]	0.142	7.033	0.172	0.104	165.698	0.006
57	2.530	9.364	0.277	0.115	1.294	7.15[n]	0.081	12.410	0.196	0.238	151.576	-0.002
58	2.545	2.424	0.318	0.030	1.294	3.17[m]	0.144	6.933	0.174	0.105	163.164	0.221
59	2.315	8.261	0.271	0.114	1.261	7.15[n]	0.078	12.748	0.192	0.236	153.900	-0.002
60	2.534	9.484	0.275	0.115	1.291	7.15[n]	0.080	12.500	0.195	0.238	163.962	-0.008
61	2.257	5.869	0.263	0.052	1.312	-8.00[m]	0.105	9.479	0.158	0.118	168.932	0.165
62	2.331	8.940	0.269	0.113	1.257	7.15[n]	0.078	12.847	0.191	0.235	181.732	-0.004
63	2.263	4.965	0.314	0.038	1.286	-8.00[m]	0.138	7.255	0.176	0.113	171.165	0.214
64	2.536	6.440	0.256	0.060	1.290	6.72[o]	0.098	10.229	0.158	0.128	222.255	-0.002
65	2.540	5.473	0.263	0.060	1.288	5.99[o]	0.102	9.829	0.161	0.128	223.966	-0.003
66	2.286	6.601	0.313	0.040	1.288	-8.00[m]	0.137	7.318	0.176	0.114	151.930	0.212
67	2.326	10.819	0.271	0.114	1.256	7.15[n]	0.079	12.719	0.193	0.236	180.574	-0.003
68	2.137	6.009	0.287	0.031	1.258	3.17[m]	0.128	7.797	0.159	0.099	130.217	0.012
69	2.327	8.820	0.269	0.114	1.256	7.15[n]	0.078	12.835	0.191	0.235	194.223	-0.003
70	2.341	4.474	0.308	0.031	1.281	3.17[m]	0.139	7.214	0.169	0.103	146.473	0.044
71	2.335	8.826	0.269	0.113	1.253	7.15[n]	0.078	12.845	0.191	0.235	225.847	-0.003
72	2.279	6.320	0.312	0.039	1.293	-8.00[m]	0.137	7.326	0.175	0.113	219.825	0.215

73	2.330	6.846	0.257	0.037	1.266	9.99[n]	0.110	9.102	0.147	0.098	180.295	0.000
74	2.544	2.614	0.319	0.031	1.299	3.17[m]	0.144	6.950	0.175	0.106	148.507	0.301
75	2.538	5.762	0.273	0.111	1.290	7.23[n]	0.081	12.376	0.192	0.228	183.150	-0.003
76	2.338	8.635	0.268	0.114	1.261	7.15[n]	0.077	13.003	0.191	0.237	214.095	-0.002
77	2.273	5.325	0.310	0.038	1.279	-8.00[m]	0.136	7.347	0.174	0.111	163.263	0.202
78	2.538	9.945	0.273	0.115	1.286	7.15[n]	0.079	12.593	0.194	0.237	222.522	-0.002
79	2.539	9.985	0.274	0.115	1.286	7.15[n]	0.079	12.585	0.194	0.237	244.859	-0.002
80	2.523	10.123	0.275	0.122	1.292	8.36[n]	0.077	12.991	0.198	0.256	154.531	-0.003
81	2.553	5.572	0.305	0.029	1.291	3.17[m]	0.138	7.251	0.167	0.102	224.931	0.002
82	2.537	5.637	0.255	0.053	1.287	7.85[o]	0.101	9.898	0.154	0.118	189.851	-0.001
83	2.530	8.594	0.271	0.109	1.293	7.33[p]	0.081	12.392	0.190	0.224	172.395	-0.002
84	2.561	6.593	0.312	0.030	1.295	3.17[m]	0.141	7.099	0.171	0.104	191.981	0.155
85	2.558	6.539	0.321	0.030	1.298	3.17[m]	0.145	6.884	0.176	0.106	142.657	0.196
86	2.538	4.860	0.257	0.044	1.284	8.56[n]	0.106	9.399	0.151	0.107	194.695	-0.001
87	2.538	9.824	0.274	0.115	1.286	7.15[n]	0.079	12.589	0.194	0.237	255.973	-0.002
88	2.530	10.491	0.274	0.121	1.286	8.36[n]	0.076	13.125	0.197	0.256	198.800	-0.003
89	2.537	9.891	0.275	0.115	1.288	7.15[n]	0.080	12.530	0.195	0.238	190.080	-0.002
90	2.537	9.372	0.266	0.070	1.286	7.97[n]	0.098	10.219	0.168	0.144	191.477	0.000
91	2.530	6.132	0.255	0.047	1.286	9.41[n]	0.104	9.607	0.151	0.109	163.257	-0.001
92	2.335	6.782	0.303	0.029	1.243	15.54[s]	0.137	7.302	0.166	0.101	102.097	0.012
93	2.531	5.351	0.259	0.037	1.284	9.99[n]	0.111	9.035	0.148	0.099	129.549	0.000
94	2.119	7.823	0.281	0.030	1.201	15.54[s]	0.125	7.979	0.155	0.096	106.817	0.003
95	2.537	6.493	0.254	0.046	1.281	9.41[n]	0.104	9.637	0.150	0.108	166.250	-0.001
96	2.328	6.866	0.307	0.030	1.244	15.54[s]	0.139	7.216	0.169	0.103	92.118	0.026
97	2.319	3.676	0.262	0.042	1.233	15.40[s]	0.110	9.083	0.152	0.104	222.923	0.014
98	2.524	5.189	0.233	0.040	1.282	10.21[n]	0.097	10.318	0.136	0.096	170.948	0.001
99	2.537	6.081	0.219	0.040	1.277	9.53[n]	0.090	11.166	0.130	0.094	205.441	-0.001
100	2.538	5.548	0.257	0.037	1.278	9.99[n]	0.110	9.069	0.147	0.098	169.552	0.000
101	2.538	5.015	0.201	0.032	1.275	9.78[r]	0.085	11.815	0.117	0.080	200.019	0.001
102	2.349	9.130	0.268	0.113	1.256	7.15[n]	0.077	12.944	0.191	0.235	249.466	-0.003
103	2.528	5.501	0.311	0.029	1.268	15.54[s]	0.141	7.087	0.170	0.102	103.635	0.074
104	2.344	7.121	0.300	0.030	1.230	16.50[m]	0.135	7.402	0.165	0.101	159.461	0.006
105	2.343	4.140	0.272	0.055	1.237	16.00[s]	0.109	9.202	0.163	0.123	164.326	0.036
106	2.543	10.212	0.272	0.114	1.276	7.15[n]	0.079	12.688	0.193	0.236	240.149	-0.002
107	2.337	3.852	0.296	0.030	1.231	16.00[s]	0.133	7.529	0.163	0.100	175.527	0.000
108	2.337	3.877	0.298	0.030	1.231	16.00[s]	0.134	7.467	0.164	0.101	140.191	-0.001
109	2.323	3.530	0.302	0.030	1.233	16.00[s]	0.136	7.339	0.166	0.101	121.148	0.011
110	2.352	6.964	0.298	0.030	1.229	16.50[m]	0.134	7.467	0.164	0.100	157.127	0.003

111	2.337	4.029	0.297	0.029	1.230	16.10[s]	0.134	7.451	0.163	0.099	178.866	0.002
112	2.332	3.769	0.299	0.030	1.231	16.00[s]	0.135	7.412	0.165	0.100	128.071	-0.001
113	2.326	3.718	0.300	0.030	1.229	16.50[t]	0.135	7.415	0.165	0.101	202.369	0.016
114	2.357	6.981	0.297	0.029	1.229	16.50[m]	0.134	7.460	0.163	0.099	171.583	0.001
115	2.542	10.188	0.272	0.114	1.277	7.15[n]	0.079	12.684	0.193	0.236	210.362	-0.002
116	2.539	5.863	0.306	0.027	1.260	16.00[s]	0.139	7.177	0.167	0.100	136.844	0.005
117	2.539	5.949	0.306	0.025	1.260	16.10[s]	0.141	7.107	0.165	0.097	149.825	0.001
118	2.542	5.913	0.303	0.027	1.255	18.20[u]	0.138	7.240	0.165	0.099	181.294	0.002
119	2.364	3.346	0.292	0.025	1.229	16.00[s]	0.133	7.504	0.158	0.094	162.370	0.094
120	2.335	4.479	0.303	0.026	1.280	3.17[m]	0.138	7.220	0.164	0.097	127.046	0.000
121	1.947	3.455	0.217	0.032	1.219	9.60[q]	0.093	10.808	0.125	0.084	225.971	0.014
122	2.258	5.680	0.274	0.036	1.252	9.40[m]	0.119	8.412	0.155	0.101	121.313	0.118

(a) Charge on the phosphorus atom calculated using electron population analysis with the natural bond orbital (NBO) method. (b) Molecular dipole moment, in Debye. (c) Ionization potential, in Hartrees. (d) Electron affinity, in Hartrees. (e) Bond order of the P=O bond. (f) Experimental pK_a values for the conjugate acid of the leaving group. (g) Molecular hardness, in Hartrees. (h) Molecular softness, in Hartrees^{-1}. (i) Molecular electronegativity, in Hartrees. (j) Molecular electrophilicity, in Hartrees. (k) Molecular volume, in cm^3/mol. (l) Fukui index for nucleophilic attack on the phosphorus atom. pK_a values obtained from (m) Evans,[3] (n) Lide,[4] (o) Bourne et al.,[5] (p) Schwarzenbach et al.,[6] (q) Wille et al.,[7] (r) Chemicalize,[8] (s) Sarjeant et al.,[9] (t) estimated value for primary alcohols, and (u) FooDB.[10]

The formula used to scale molecular descriptors was as follows:

$$x_{ij}^{s} = \frac{x_{ij} - x_{j,min}}{x_{j,max} - x_{j,min}} \tag{A.1}$$

where x_{ij} and x_{ij}^{s} are the unscaled and scaled j^{th} descriptor values for molecule i, respectively, and $x_{j,min}$ and $x_{j,max}$ are the minimum and maximum values for the j^{th} descriptor. For all descriptors, $\min(x_{ij}^{s}) = 0$ and $\max(x_{ij}^{s}) = 1$.

Table A.5. Scaled molecular descriptors used for QSAR analysis, calculated using DFT with B3LYP/6-311++G**.

Molecule #	q_p	μ	IP	EA	BO	pK_a	η	SOF	χ	ω	V	f_p^+
1	0.000	0.426	0.779	0.132	0.298	0.000	0.786	0.116	0.580	0.157	0.073	0.417
2	0.014	0.427	0.753	0.144	0.299	0.000	0.756	0.134	0.568	0.160	0.107	0.407
3	0.403	0.414	0.753	0.075	0.440	0.426	0.801	0.107	0.530	0.120	0.076	0.139
4	0.458	0.496	0.921	0.019	0.433	0.916	0.978	0.011	0.617	0.118	0.152	0.020
5	0.296	0.201	0.887	0.134	0.571	0.000	0.873	0.065	0.657	0.175	0.040	0.451
6	0.304	0.238	0.865	0.125	0.545	0.000	0.861	0.072	0.636	0.166	0.387	0.419
7	0.286	0.186	0.488	0.128	0.634	0.000	0.546	0.285	0.373	0.108	0.283	0.014
8	0.689	0.201	0.848	0.020	0.632	0.426	0.917	0.042	0.566	0.106	0.295	0.194
9	0.297	0.188	0.488	0.126	0.637	0.000	0.548	0.284	0.372	0.107	0.354	0.012
10	0.303	0.232	0.871	0.127	0.548	0.000	0.865	0.070	0.642	0.169	0.108	0.427
11	0.314	0.232	0.866	0.134	0.545	0.000	0.856	0.075	0.642	0.172	0.108	0.414
12	0.684	0.191	0.968	0.078	0.671	0.426	0.978	0.011	0.683	0.157	0.006	0.303
13	0.307	0.250	0.863	0.127	0.539	0.000	0.858	0.073	0.636	0.167	0.291	0.408
14	0.305	0.249	0.864	0.124	0.542	0.000	0.861	0.072	0.635	0.166	0.356	0.417
15	0.689	0.220	0.945	0.070	0.644	0.426	0.963	0.018	0.662	0.149	0.055	0.235
16	0.689	0.192	0.977	0.074	0.664	0.426	0.988	0.006	0.687	0.156	0.057	0.194
17	0.375	0.888	0.513	0.877	0.169	0.578	0.074	0.857	0.808	0.846	0.517	0.006
18	0.310	0.235	0.841	0.127	0.512	0.000	0.839	0.084	0.621	0.164	0.208	0.415
19	0.307	0.203	0.881	0.141	0.568	0.000	0.864	0.071	0.656	0.178	0.104	0.438
20	0.687	0.174	0.495	0.136	0.793	0.426	0.547	0.285	0.383	0.115	0.236	0.019
21	0.313	0.252	0.806	0.124	0.504	0.000	0.813	0.100	0.595	0.157	0.378	0.403
22	0.689	0.233	0.943	0.068	0.644	0.426	0.963	0.018	0.659	0.148	0.144	0.217
23	0.691	0.236	0.927	0.076	0.633	0.426	0.945	0.027	0.653	0.149	0.199	0.158
24	0.695	0.236	0.855	0.070	0.598	0.426	0.889	0.057	0.598	0.134	0.300	0.153
25	0.978	0.280	0.966	0.050	0.765	0.426	0.995	0.003	0.666	0.142	0.122	0.506
26	0.322	0.234	0.838	0.133	0.508	0.000	0.833	0.088	0.622	0.167	0.245	0.405
27	0.693	0.221	0.930	0.067	0.607	0.426	0.953	0.023	0.650	0.145	0.110	0.207
28	0.972	0.239	1.000	0.084	0.819	0.426	1.000	0.000	0.709	0.166	0.020	0.871
29	0.699	0.218	0.919	0.066	0.598	0.426	0.944	0.027	0.641	0.143	0.183	0.121
30	0.982	0.263	0.918	0.054	0.747	0.426	0.952	0.024	0.634	0.136	0.450	0.839
31	0.339	0.304	0.594	0.428	0.549	0.000	0.437	0.383	0.615	0.338	0.374	0.173
32	0.694	0.227	0.899	0.063	0.601	0.426	0.930	0.035	0.626	0.138	0.311	0.179
33	0.980	0.230	0.956	0.073	0.754	0.426	0.970	0.015	0.671	0.153	0.175	0.377
34	0.578	0.204	0.943	0.159	0.717	0.000	0.903	0.049	0.710	0.199	0.000	0.528
35	0.691	0.254	0.798	0.067	0.626	0.426	0.844	0.082	0.557	0.123	0.527	0.149
36	0.968	0.567	0.688	1.000	0.734	0.461	0.138	0.750	1.000	1.000	0.712	0.011

37	0.663	1.000	0.555	0.915	0.491	0.578	0.084	0.840	0.859	0.896	0.533	0.011
38	0.979	0.270	0.944	0.048	0.752	0.426	0.977	0.011	0.649	0.137	0.368	0.899
39	0.586	0.312	0.501	0.296	0.673	0.000	0.446	0.373	0.476	0.223	0.705	0.177
40	0.593	0.262	0.912	0.140	0.676	0.000	0.890	0.056	0.677	0.183	0.185	0.472
41	0.697	0.240	0.844	0.066	0.603	0.426	0.882	0.061	0.589	0.130	0.394	0.124
42	0.981	0.286	0.930	0.065	0.731	0.426	0.954	0.023	0.649	0.144	0.229	0.089
43	0.695	0.233	0.880	0.063	0.601	0.426	0.914	0.043	0.613	0.134	0.238	0.180
44	0.715	0.236	0.891	0.057	0.592	0.426	0.927	0.036	0.617	0.133	0.304	0.058
45	0.598	0.263	0.872	0.140	0.646	0.000	0.857	0.074	0.650	0.177	0.556	0.446
46	0.710	0.237	0.900	0.038	0.629	0.426	0.947	0.026	0.613	0.124	0.233	0.352
47	0.695	0.759	0.539	0.881	0.437	0.578	0.093	0.824	0.829	0.848	0.545	0.011
48	0.982	0.301	0.873	0.040	0.738	0.426	0.923	0.038	0.594	0.121	0.579	0.445
49	0.980	0.288	0.916	0.043	0.749	0.426	0.958	0.021	0.627	0.129	0.527	1.000
50	0.725	0.218	0.851	0.042	0.557	0.426	0.904	0.049	0.581	0.119	0.378	0.031
51	0.720	0.219	0.865	0.034	0.594	0.426	0.921	0.039	0.586	0.116	0.359	0.275
52	0.705	0.236	0.843	0.065	0.588	0.426	0.882	0.060	0.587	0.129	0.418	0.088
53	0.995	0.393	0.491	0.164	1.000	0.426	0.525	0.303	0.395	0.131	0.528	0.079
54	0.690	0.727	0.427	0.881	0.422	0.578	0.000	1.000	0.750	0.874	1.000	0.009
55	0.596	0.277	0.904	0.140	0.666	0.000	0.884	0.059	0.672	0.181	0.412	0.459
56	0.988	0.334	0.883	0.049	0.697	0.426	0.926	0.037	0.607	0.127	0.475	0.028
57	0.958	0.798	0.591	0.899	0.706	0.578	0.124	0.773	0.875	0.865	0.398	0.012
58	0.979	0.000	0.916	0.049	0.708	0.426	0.953	0.023	0.630	0.133	0.461	0.457
59	0.666	0.671	0.545	0.883	0.455	0.578	0.096	0.819	0.834	0.850	0.410	0.011
60	0.964	0.811	0.579	0.896	0.684	0.578	0.116	0.785	0.865	0.862	0.466	0.000
61	0.587	0.396	0.486	0.274	0.844	0.000	0.448	0.372	0.454	0.206	0.493	0.345
62	0.689	0.749	0.532	0.879	0.424	0.578	0.088	0.832	0.823	0.846	0.563	0.008
63	0.595	0.292	0.885	0.137	0.647	0.000	0.870	0.067	0.657	0.177	0.505	0.444
64	0.967	0.462	0.427	0.353	0.677	0.562	0.348	0.474	0.456	0.259	0.786	0.013
65	0.972	0.350	0.485	0.347	0.663	0.534	0.400	0.419	0.494	0.261	0.796	0.010
66	0.627	0.480	0.875	0.148	0.661	0.000	0.854	0.076	0.657	0.182	0.399	0.439
67	0.681	0.965	0.549	0.885	0.418	0.578	0.098	0.815	0.838	0.852	0.557	0.011
68	0.424	0.412	0.676	0.063	0.436	0.426	0.745	0.141	0.469	0.101	0.280	0.040
69	0.682	0.735	0.534	0.880	0.423	0.578	0.089	0.831	0.825	0.847	0.632	0.009
70	0.702	0.236	0.836	0.060	0.609	0.426	0.880	0.062	0.580	0.126	0.369	0.105
71	0.693	0.736	0.532	0.879	0.395	0.578	0.088	0.832	0.823	0.846	0.806	0.010
72	0.618	0.448	0.867	0.141	0.699	0.000	0.852	0.077	0.647	0.176	0.773	0.445
73	0.687	0.508	0.435	0.124	0.493	0.687	0.505	0.320	0.334	0.098	0.556	0.016
74	0.978	0.022	0.923	0.065	0.744	0.426	0.949	0.025	0.644	0.143	0.381	0.617

75	0.970	0.384	0.560	0.856	0.674	0.581	0.127	0.768	0.829	0.806	0.571	0.010
76	0.698	0.714	0.523	0.886	0.454	0.578	0.076	0.854	0.820	0.858	0.742	0.012
77	0.610	0.333	0.852	0.129	0.594	0.000	0.847	0.080	0.630	0.167	0.462	0.420
78	0.969	0.864	0.566	0.891	0.648	0.578	0.109	0.797	0.853	0.857	0.788	0.012
79	0.971	0.869	0.567	0.891	0.645	0.578	0.109	0.796	0.854	0.857	0.911	0.012
80	0.949	0.885	0.582	0.959	0.689	0.624	0.077	0.852	0.902	0.960	0.414	0.010
81	0.990	0.362	0.815	0.048	0.685	0.426	0.871	0.067	0.559	0.115	0.801	0.020
82	0.968	0.369	0.425	0.285	0.651	0.605	0.390	0.429	0.416	0.206	0.608	0.014
83	0.959	0.709	0.545	0.839	0.698	0.585	0.126	0.770	0.810	0.785	0.512	0.013
84	1.000	0.479	0.868	0.056	0.713	0.426	0.909	0.046	0.601	0.129	0.620	0.325
85	0.997	0.473	0.939	0.057	0.735	0.426	0.966	0.016	0.651	0.141	0.348	0.407
86	0.969	0.280	0.437	0.194	0.633	0.632	0.460	0.360	0.375	0.143	0.635	0.014
87	0.970	0.850	0.567	0.891	0.648	0.578	0.109	0.797	0.854	0.858	0.972	0.012
88	0.958	0.927	0.567	0.956	0.646	0.624	0.067	0.870	0.891	0.959	0.657	0.010
89	0.968	0.858	0.575	0.894	0.663	0.578	0.114	0.789	0.861	0.860	0.609	0.012
90	0.968	0.799	0.505	0.448	0.647	0.610	0.349	0.473	0.563	0.347	0.617	0.016
91	0.959	0.426	0.420	0.218	0.650	0.665	0.430	0.389	0.376	0.157	0.462	0.014
92	0.694	0.501	0.800	0.048	0.321	0.898	0.858	0.074	0.548	0.113	0.125	0.039
93	0.960	0.336	0.450	0.126	0.629	0.687	0.516	0.311	0.345	0.101	0.276	0.017
94	0.400	0.620	0.623	0.054	0.000	0.898	0.707	0.166	0.426	0.087	0.151	0.022
95	0.968	0.468	0.411	0.214	0.607	0.665	0.426	0.393	0.367	0.153	0.478	0.013
96	0.684	0.510	0.832	0.055	0.328	0.898	0.879	0.062	0.574	0.122	0.070	0.068
97	0.672	0.144	0.474	0.168	0.246	0.893	0.508	0.317	0.385	0.131	0.790	0.044
98	0.951	0.318	0.251	0.148	0.613	0.695	0.337	0.486	0.218	0.085	0.504	0.018
99	0.968	0.420	0.142	0.155	0.578	0.669	0.241	0.602	0.145	0.075	0.694	0.014
100	0.969	0.359	0.440	0.121	0.585	0.687	0.511	0.315	0.336	0.097	0.496	0.016
101	0.970	0.298	0.000	0.074	0.566	0.679	0.177	0.691	0.000	0.000	0.664	0.019
102	0.713	0.771	0.523	0.879	0.417	0.578	0.081	0.846	0.816	0.847	0.936	0.010
103	0.956	0.354	0.861	0.043	0.509	0.898	0.912	0.044	0.588	0.120	0.134	0.164
104	0.705	0.540	0.777	0.055	0.221	0.935	0.834	0.087	0.536	0.113	0.441	0.029
105	0.704	0.197	0.553	0.296	0.276	0.916	0.490	0.334	0.513	0.230	0.468	0.087
106	0.976	0.895	0.553	0.886	0.574	0.578	0.101	0.811	0.841	0.853	0.885	0.012
107	0.696	0.164	0.743	0.057	0.228	0.916	0.804	0.105	0.512	0.108	0.529	0.016
108	0.696	0.167	0.759	0.056	0.228	0.916	0.819	0.096	0.524	0.111	0.335	0.014
109	0.677	0.127	0.791	0.050	0.242	0.916	0.849	0.079	0.543	0.113	0.230	0.038
110	0.717	0.522	0.756	0.052	0.215	0.935	0.819	0.096	0.519	0.108	0.428	0.022
111	0.697	0.185	0.753	0.042	0.225	0.920	0.822	0.094	0.512	0.102	0.548	0.021
112	0.689	0.155	0.770	0.050	0.228	0.916	0.832	0.089	0.528	0.109	0.268	0.014

113	0.682	0.149	0.772	0.053	0.219	0.935	0.831	0.089	0.531	0.111	0.677	0.048
114	0.723	0.524	0.751	0.043	0.217	0.935	0.820	0.095	0.511	0.102	0.508	0.018
115	0.975	0.892	0.554	0.887	0.580	0.578	0.101	0.810	0.842	0.853	0.721	0.012
116	0.971	0.395	0.822	0.028	0.446	0.916	0.889	0.057	0.552	0.106	0.316	0.026
117	0.971	0.405	0.821	0.000	0.447	0.920	0.907	0.047	0.536	0.091	0.388	0.018
118	0.974	0.401	0.799	0.024	0.409	1.000	0.873	0.065	0.534	0.100	0.561	0.021
119	0.732	0.106	0.708	0.004	0.216	0.916	0.810	0.101	0.458	0.074	0.457	0.203
120	0.694	0.236	0.795	0.010	0.604	0.426	0.878	0.062	0.523	0.092	0.263	0.016
121	0.166	0.118	0.124	0.074	0.140	0.672	0.280	0.553	0.087	0.019	0.807	0.044
122	0.589	0.374	0.567	0.111	0.387	0.664	0.623	0.225	0.419	0.111	0.231	0.251

A.5 QSAR Analysis – Non-Predictive QSAR Models for Molecular Subsets

Before dividing the molecular dataset into training and test sets, non-predictive QSAR models were generated for different molecular subsets. The term non-predictive denotes that, for each molecular subset (e.g., molecules with leaving group = F^-), all molecules were included in the model development. This was done in order to gain a better understanding of the most important descriptors involved in organophosphate hydrolysis, especially since some descriptors selected when using the entire dataset may be different from those when using only a training set.

Out of the 119 molecules studied, there are several possible molecular subsets that contain enough molecules to create a statistically significant QSAR model including: all molecules (N_m = 119), those hydrolyzing through concerted mechanisms (N_m = 57) or stepwise mechanisms (N_m = 62), those with a F^- leaving group (N_m = 37) or Cl^- leaving group (N_m = 25), phosphono– molecules composed of alkyl and alkoxy substituents (N_m = 58), and phosphoro– molecules composed of two alkoxy substituents (N_m = 54). The non-predictive multi-parametric QSAR models for each of these subsets is included below, along with their corresponding model statistics. As described in the main text, ΔG_{TS1} is in kJ/mol and the descriptors are dimensionless (scaled from 0 to 1). Uniparametric correlation coefficients (R^2_{uni}) for individual descriptors are also listed, in order of decreasing statistical significance.

Included below each model is a matrix containing correlation coefficients between descriptors (R^2_{id}) within that molecular subset. The diagonal is blacked out, omitting correlations between descriptors and themselves, for clarity. Inter-descriptor correlations that have $R^2_{id} > 0.5$ are considered significant and are highlighted in yellow. Theoretically, these highly correlated descriptors could be substituted for one another in the non-predictive multi-parametric QSAR models without significantly affecting the model statistics.

<u>All molecules:</u>

$$\Delta G_{TS1} = (73.70 \pm 4.88) + (58.81 \pm 9.58)q_P - (80.68 \pm 11.99)BO - (12.33 \pm 7.40)pK_a - (10.47 \pm 2.92)\omega + (10.60 \pm 4.38)V \quad (A.2)$$

N_m = 119, Q^2_{LOO} = 0.64, R^2 = 0.72, F = 59.2, RMSE = 8.02 kJ/mol

Uniparametric correlation coefficients and sign of their slope (R^2_{uni}, sign): pK_a (0.52, positive), q_P (0.21, positive), BO (0.19, negative), f_P^+ (0.16, negative), V (0.15, positive), IP (0.08, negative), χ (0.05, negative), η (0.03, negative), SOF (0.02, positive), μ (0.02, positive), EA (0, N/A), ω (0, N/A).

Descriptor	q_p	μ	IP	EA	BO	pK_a	η	SOF	χ	ω	V	f_p^+
q_p		0.05	0.04	0.04	0.16	0.29	0.05	0.06	0.00	0.04	0.19	0.04
μ	0.05		0.22	0.71	0.01	0.06	0.61	0.67	0.28	0.70	0.25	0.19
IP	0.04	0.22		0.26	0.01	0.08	0.69	0.58	0.09	0.15	0.35	0.37
EA	0.04	0.71	0.26		0.00	0.01	0.81	0.89	0.46	0.98	0.31	0.16
BO	0.16	0.01	0.01	0.00		0.23	0.00	0.00	0.01	0.00	0.00	0.13
pK_a	0.29	0.06	0.08	0.01	0.23		0.05	0.05	0.02	0.01	0.06	0.34
η	0.05	0.61	0.69	0.81	0.00	0.05		0.98	0.08	0.70	0.43	0.31
SOF	0.06	0.67	0.58	0.89	0.00	0.05	0.98		0.15	0.80	0.42	0.28
χ	0.00	0.28	0.09	0.46	0.01	0.02	0.08	0.15		0.58	0.01	0.01
ω	0.04	0.70	0.15	0.98	0.00	0.01	0.70	0.80	0.58		0.26	0.12
V	0.19	0.25	0.35	0.31	0.00	0.06	0.43	0.42	0.01	0.26		0.14
f_p^+	0.04	0.19	0.37	0.16	0.13	0.34	0.31	0.28	0.01	0.12	0.14	

Concerted mechanism:

$$\Delta G_{TS1} = (41.68 \pm 2.06) + (24.41 \pm 3.83)q_P + (10.52 \pm 4.47)V \quad \text{(A.3)}$$

$N_m = 57$, $Q^2_{LOO} = 0.65$, $R^2 = 0.71$, F = 64.7, RMSE = 5.97 kJ/mol

Uniparametric correlation coefficients and sign of their slope (R^2_{uni}, sign): q_P (0.68, positive), V (0.48, positive), pK_a (0.43, positive), μ (0.32, positive), η (0.32, negative), SOF (0.30, positive), f_P^+ (0.30, negative), EA (0.25, positive), IP (0.24, negative), ω (0.22, positive), BO (0.08, positive), χ (0.06, positive).

Descriptor	q_p	μ	IP	EA	BO	pK_a	η	SOF	χ	ω	V	f_p^+
q_p		0.35	0.34	0.36	0.23	0.64	0.45	0.41	0.09	0.31	0.48	0.47
μ	0.35		0.29	0.81	0.04	0.62	0.74	0.79	0.46	0.79	0.36	0.50
IP	0.34	0.29		0.29	0.00	0.62	0.65	0.53	0.01	0.20	0.38	0.84
EA	0.36	0.81	0.29		0.02	0.59	0.87	0.93	0.64	0.99	0.36	0.57
BO	0.23	0.04	0.00	0.02		0.01	0.01	0.02	0.02	0.03	0.01	0.00
pK_a	0.64	0.62	0.62	0.59	0.01		0.77	0.74	0.12	0.52	0.43	0.80
η	0.45	0.74	0.65	0.87	0.01	0.77		0.98	0.29	0.80	0.47	0.85
SOF	0.41	0.79	0.53	0.93	0.02	0.74	0.98		0.39	0.88	0.45	0.76
χ	0.09	0.46	0.01	0.64	0.02	0.12	0.29	0.39		0.73	0.08	0.06
ω	0.31	0.79	0.20	0.99	0.03	0.52	0.80	0.88	0.73		0.32	0.48
V	0.48	0.36	0.38	0.36	0.01	0.43	0.47	0.45	0.08	0.32		0.43
f_p^+	0.47	0.50	0.84	0.57	0.00	0.80	0.85	0.76	0.06	0.48	0.43	

Stepwise mechanism:

$$\Delta G_{TS1} = (73.74 \pm 7.30) + (33.67 \pm 6.28)q_P - (60.14 \pm 14.36)BO + (12.29 \pm 6.00)pK_a + (8.17 \pm 4.46)V \quad (A.4)$$

$N_m = 62$, $Q^2_{LOO} = 0.69$, $R^2 = 0.84$, F = 75.8, RMSE = 7.11 kJ/mol

Uniparametric correlation coefficients and sign of their slope (R^2_{uni}, sign): pK_a (0.68, positive), BO (0.45, negative), ω (0.22, negative), V (0.19, positive), χ (0.18, negative), IP (0.15, negative), f_P^+ (0.14, negative), η (0.11, negative), SOF (0.09, positive), q_P (0.02, positive), μ (0.02, positive), EA (0, N/A).

Descriptor	q_P	μ	IP	EA	BO	pK_a	η	SOF	χ	ω	V	f_p^+
q_P		0.00	0.01	0.00	0.31	0.03	0.01	0.02	0.01	0.00	0.13	0.16
μ	0.00		0.03	0.01	0.04	0.08	0.02	0.01	0.04	0.07	0.00	0.06
IP	0.01	0.03		0.30	0.04	0.11	0.98	0.97	0.97	0.37	0.23	0.20
EA	0.00	0.01	0.30		0.01	0.00	0.43	0.39	0.16	0.11	0.03	0.02
BO	0.31	0.04	0.04	0.01		0.71	0.03	0.01	0.06	0.10	0.01	0.28
pK_a	0.03	0.08	0.11	0.00	0.71		0.09	0.06	0.14	0.14	0.04	0.23
η	0.01	0.02	0.98	0.43	0.03	0.09		0.98	0.91	0.24	0.21	0.18
SOF	0.02	0.01	0.97	0.39	0.01	0.06	0.98		0.91	0.26	0.20	0.14
χ	0.01	0.04	0.97	0.16	0.06	0.14	0.91	0.91		0.53	0.24	0.21
ω	0.00	0.07	0.37	0.11	0.10	0.14	0.24	0.26	0.53		0.14	0.14
V	0.13	0.00	0.23	0.03	0.01	0.04	0.21	0.20	0.24	0.14		0.02
f_p^+	0.16	0.06	0.20	0.02	0.28	0.23	0.18	0.14	0.21	0.14	0.02	

Leaving group = F^-:

$$\Delta G_{TS1} = (52.12 \pm 1.46) + (18.19 \pm 3.26)V \quad (A.5)$$

$N_m = 37$, $Q^2_{LOO} = 0.34$, $R^2 = 0.47$, F = 31.1, RMSE = 4.49 kJ/mol

Uniparametric correlation coefficients and sign of their slope (R^2_{uni}, sign): V (0.47, positive), q_P (0.23, positive), μ (0.07, positive), ω (0.06, negative), EA (0.03, negative), BO (0.03, positive), χ (0.01, negative), IP (0, N/A), f_P^+ (0, N/A), η (0, N/A), SOF (0, N/A).

Descriptor	q_P	μ	IP	EA	BO	η	SOF	χ	ω	V	f_p^+
q_P		0.00	0.06	0.00	0.65	0.05	0.03	0.07	0.07	0.21	0.29
μ	0.00		0.07	0.01	0.00	0.06	0.05	0.08	0.07	0.07	0.02
IP	0.06	0.07		0.38	0.01	0.98	0.96	0.98	0.39	0.09	0.20
EA	0.00	0.01	0.38		0.19	0.51	0.56	0.24	0.05	0.03	0.06
BO	0.65	0.00	0.01	0.19		0.03	0.05	0.00	0.12	0.05	0.19
η	0.05	0.06	0.98	0.51	0.03		0.99	0.92	0.27	0.06	0.19
SOF	0.03	0.05	0.96	0.56	0.05	0.99		0.89	0.22	0.05	0.17
χ	0.07	0.08	0.98	0.24	0.00	0.92	0.89		0.54	0.14	0.21
ω	0.07	0.07	0.39	0.05	0.12	0.27	0.22	0.54		0.31	0.12
V	0.21	0.07	0.09	0.03	0.05	0.06	0.05	0.14	0.31		0.00
f_p^+	0.29	0.02	0.20	0.06	0.19	0.19	0.17	0.21	0.12	0.00	

Leaving group = Cl^-:

$$\Delta G_{TS1} = (38.33 \pm 1.24) + (24.34 \pm 1.68)q_P + (4.74 \pm 1.69)\mu \quad (A.6)$$

$N_m = 25$, $Q^2_{LOO} = 0.89$, $R^2 = 0.91$, F = 116.0, RMSE = 2.40 kJ/mol

Uniparametric correlation coefficients and sign of their slope (R^2_{uni}, sign): q_P (0.88, positive), BO (0.57, positive), V (0.43, positive), ω (0.10, positive), μ (0.09, positive), EA (0.05, positive), χ (0.04, positive), f_P^+ (0.03, positive), IP (0.01, positive), SOF (0, N/A), η (0, N/A).

Descriptor	q_P	μ	IP	EA	BO	η	SOF	χ	ω	V	f_p^+
q_P		0.02	0.01	0.03	0.74	0.00	0.00	0.05	0.06	0.36	0.03
μ	0.02		0.00	0.05	0.01	0.01	0.01	0.00	0.04	0.14	0.04
IP	0.01	0.00		0.30	0.03	0.94	0.92	0.87	0.01	0.09	0.80
EA	0.03	0.05	0.30		0.06	0.55	0.59	0.04	0.77	0.09	0.14
BO	0.74	0.01	0.03	0.06		0.05	0.06	0.01	0.03	0.24	0.01
η	0.00	0.01	0.94	0.55	0.05		1.00	0.65	0.11	0.11	0.69
SOF	0.00	0.01	0.92	0.59	0.06	1.00		0.61	0.13	0.11	0.66
χ	0.05	0.00	0.87	0.04	0.01	0.65	0.61		0.09	0.05	0.77
ω	0.06	0.04	0.01	0.77	0.03	0.11	0.13	0.09		0.03	0.00
V	0.36	0.14	0.09	0.09	0.24	0.11	0.11	0.05	0.03		0.05
f_p^+	0.03	0.04	0.80	0.14	0.01	0.69	0.66	0.77	0.00	0.05	

Phosphono– molecules:

$$\Delta G_{TS1} = (75.17 \pm 4.00) - (10.10 \pm 3.41)\mu - (37.20 \pm 4.74)BO + (20.36 \pm 3.72)pK_a \quad (A.7)$$

$N_m = 58$, $Q^2_{LOO} = 0.81$, $R^2 = 0.86$, F = 110.0, RMSE = 6.44 kJ/mol

Uniparametric correlation coefficients and sign of their slope (R^2_{uni}, sign): BO (0.77, negative), pK_a (0.68, positive), f_P^+ (0.24, negative), q_P (0.23, positive), V (0.16, positive), χ (0.07, negative), IP (0.06, negative), η (0.01, negative), μ (0, N/A), SOF (0, N/A), ω (0, N/A), EA (0, N/A).

Descriptor	q_p	μ	IP	EA	BO	pK_a	η	SOF	χ	ω	V	f_p^+
q_p		0.06	0.01	0.01	0.07	0.66	0.01	0.02	0.01	0.02	0.09	0.45
μ	0.06		0.33	0.74	0.05	0.05	0.66	0.71	0.44	0.76	0.30	0.20
IP	0.01	0.33		0.47	0.09	0.06	0.78	0.67	0.00	0.39	0.53	0.38
EA	0.01	0.74	0.47		0.01	0.00	0.90	0.95	0.56	0.99	0.38	0.14
BO	0.07	0.05	0.09	0.01		0.55	0.04	0.03	0.01	0.01	0.20	0.18
pK_a	0.66	0.05	0.06	0.00	0.55		0.02	0.02	0.02	0.01	0.14	0.52
η	0.01	0.66	0.78	0.90	0.04	0.02		0.98	0.25	0.84	0.51	0.26
SOF	0.02	0.71	0.67	0.95	0.03	0.02	0.98		0.35	0.92	0.50	0.24
χ	0.01	0.44	0.00	0.56	0.01	0.02	0.25	0.35		0.64	0.03	0.00
ω	0.02	0.76	0.39	0.99	0.01	0.01	0.84	0.92	0.64		0.36	0.13
V	0.09	0.30	0.53	0.38	0.20	0.14	0.51	0.50	0.03	0.36		0.30
f_p^+	0.45	0.20	0.38	0.14	0.18	0.52	0.26	0.24	0.00	0.13	0.30	

Phosphoro– molecules:

$$\Delta G_{TS1} = (83.40 \pm 3.70) - (46.07 \pm 5.72)BO + (15.26 \pm 3.53)pK_a \quad (A.8)$$

$N_m = 54$, $Q^2_{LOO} = 0.66$, $R^2 = 0.75$, F = 77.1, RMSE = 5.89 kJ/mol

Uniparametric correlation coefficients and sign of their slope (R^2_{uni}, sign): BO (0.66, negative), pK_a (0.43, positive), f_P^+ (0.27, negative), IP (0.12, negative), q_P (0.08, positive), μ (0.07, positive), χ (0.06, negative), η (0.05, negative), V (0.05, positive), SOF (0.04, positive), EA (0, N/A), ω (0, N/A).

Descriptor	q_p	μ	IP	EA	BO	pK_a	η	SOF	χ	ω	V	f_p^+
q_p		0.05	0.05	0.03	0.00	0.70	0.06	0.07	0.00	0.03	0.02	0.10
μ	0.05		0.13	0.73	0.03	0.09	0.57	0.64	0.27	0.71	0.24	0.29
IP	0.05	0.13		0.14	0.05	0.16	0.64	0.52	0.22	0.05	0.26	0.47
EA	0.03	0.73	0.14		0.00	0.03	0.73	0.82	0.42	0.97	0.22	0.24
BO	0.00	0.03	0.05	0.00		0.24	0.02	0.01	0.03	0.00	0.03	0.19
pK_a	0.70	0.09	0.16	0.03	0.24		0.12	0.12	0.02	0.03	0.01	0.30
η	0.06	0.57	0.64	0.73	0.02	0.12		0.98	0.02	0.58	0.34	0.49
SOF	0.07	0.64	0.52	0.82	0.01	0.12	0.98		0.07	0.69	0.32	0.44
χ	0.00	0.27	0.22	0.42	0.03	0.02	0.02	0.07		0.57	0.00	0.01
ω	0.03	0.71	0.05	0.97	0.00	0.03	0.58	0.69	0.57		0.16	0.16
V	0.02	0.24	0.26	0.22	0.03	0.01	0.34	0.32	0.00	0.16		0.21
f_p^+	0.10	0.29	0.47	0.24	0.19	0.30	0.49	0.44	0.01	0.16	0.21	

A.6 QSAR Analysis – Comparison of Dataset Division Algorithms for Predictive QSAR Models

As stated in the main text, we also used more rational algorithms to divide our data into training and test sets, which we refer to as cluster by rank and modified cluster by rank. As with the random selection approach, both of these methods include the molecules with the highest and lowest ΔG_{TS1} values in the training set so that it spans the entire response variable space. Again, 18 out of the 119 total molecules (~15 %) were included in the test set.

For cluster by rank, we first ranked all 119 molecules by ΔG_{TS1} values, in order from lowest to highest. Once the molecules with the highest and lowest ΔG_{TS1} values were placed in the training set, we divided the remaining 117 molecules into 17 groups. Due to the odd number, the first and last groups each had one less molecule. Then, one molecule was randomly selected from each group (two for the middle group) until the test set included 18 molecules.

For modified cluster by rank, the molecules with the highest and lowest ΔG_{TS1} values were first placed in the training set. Out of the 119 molecules, there are a total of 23 unique leaving groups that have different conjugate acid pK_a values. At least one molecule with each specific leaving group was included in the training set. For leaving groups that were present for multiple molecules, one molecule was randomly chosen for the training set. This was done so that the training set spans the entire pK_a descriptor space, because we already know that the pK_a affects the type of hydrolysis mechanism that each molecule follows. Finally, we ranked the remaining molecules and divided them into 18 groups, using the same procedure as in cluster by rank, and assigned molecules randomly from each group until the test set included 18 molecules.

We ran each dataset division algorithm 100 times to examine the variability in model statistics. As seen in Figures A.3–A.6 below, the more rational algorithms do not have a strong influence on the resulting QSAR model statistics. Aside from highlighting the optimal cluster by rank and modified cluster by rank models, we restricted our analysis to the optimal random selection model. Here, "optimal" denotes the models that yielded the lowest $RMSE_{test}$.

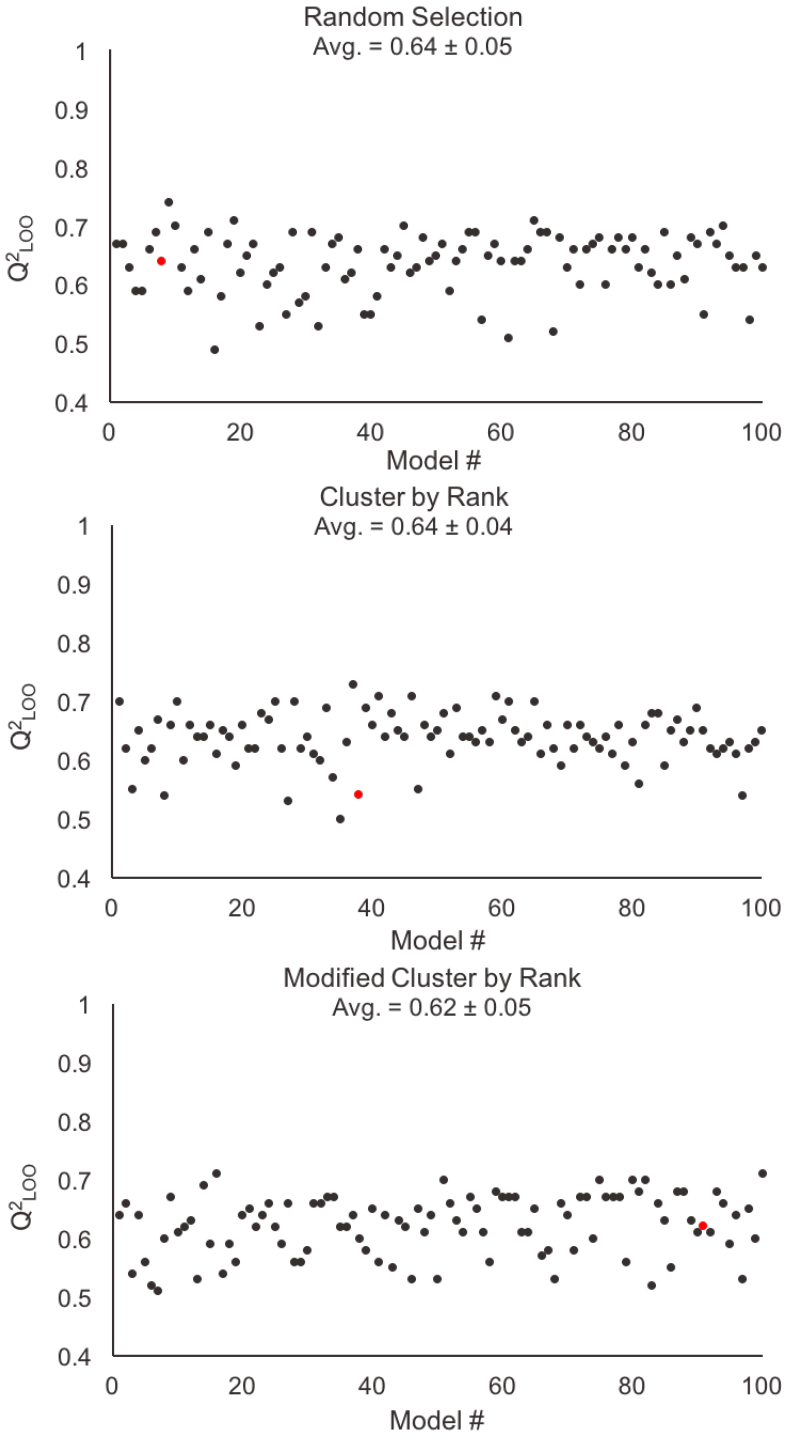


Figure A.3. Training set leave-one-out correlation coefficients for the 100 QSAR models generated using (top) random selection, (middle) cluster by rank, and (bottom) modified cluster by rank algorithms. Red points indicate the optimal model for each dataset division algorithm.

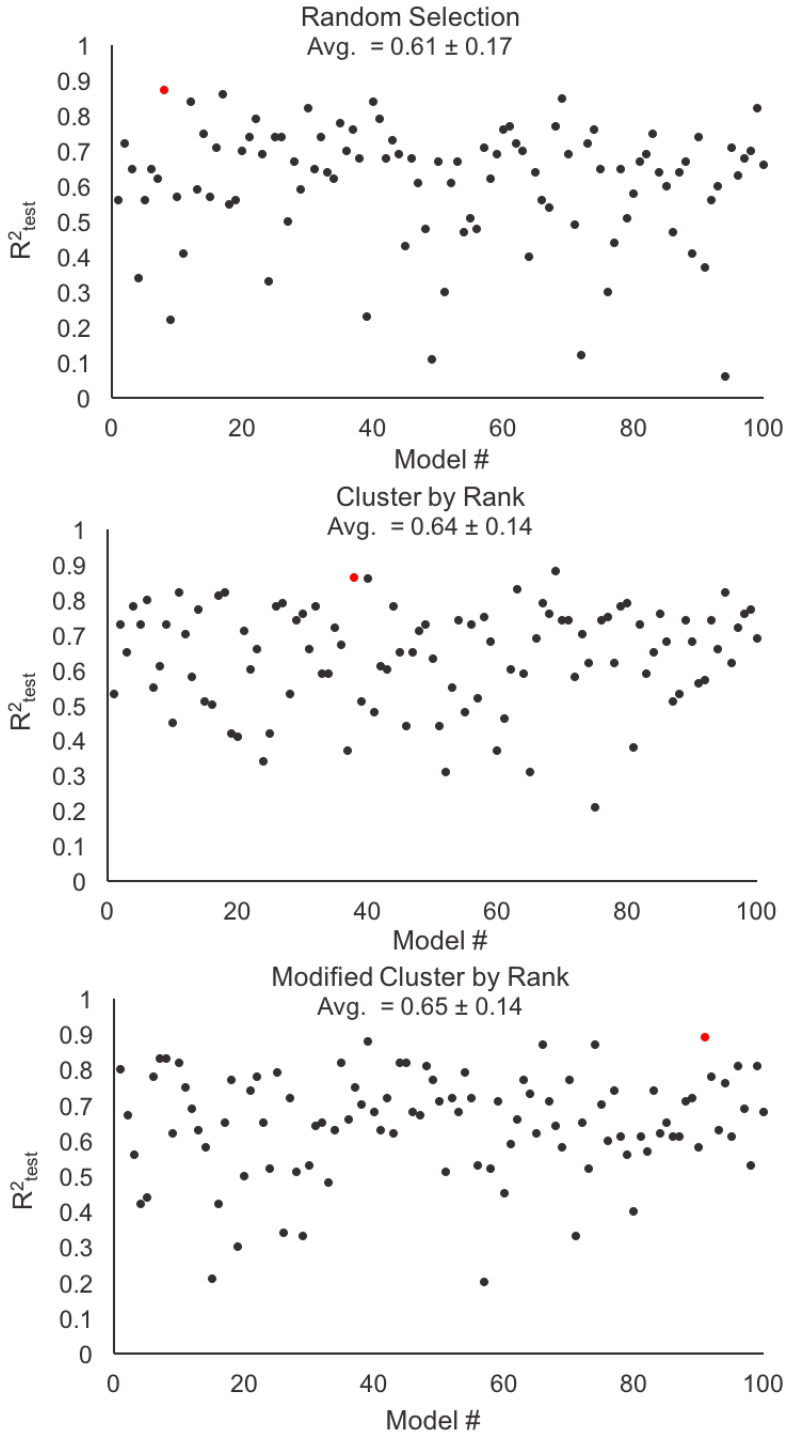


Figure A.4. Test set correlation coefficients for the 100 QSAR models generated using (top) random selection, (middle) cluster by rank, and (bottom) modified cluster by rank algorithms. Red points indicate the optimal model for each dataset division algorithm.

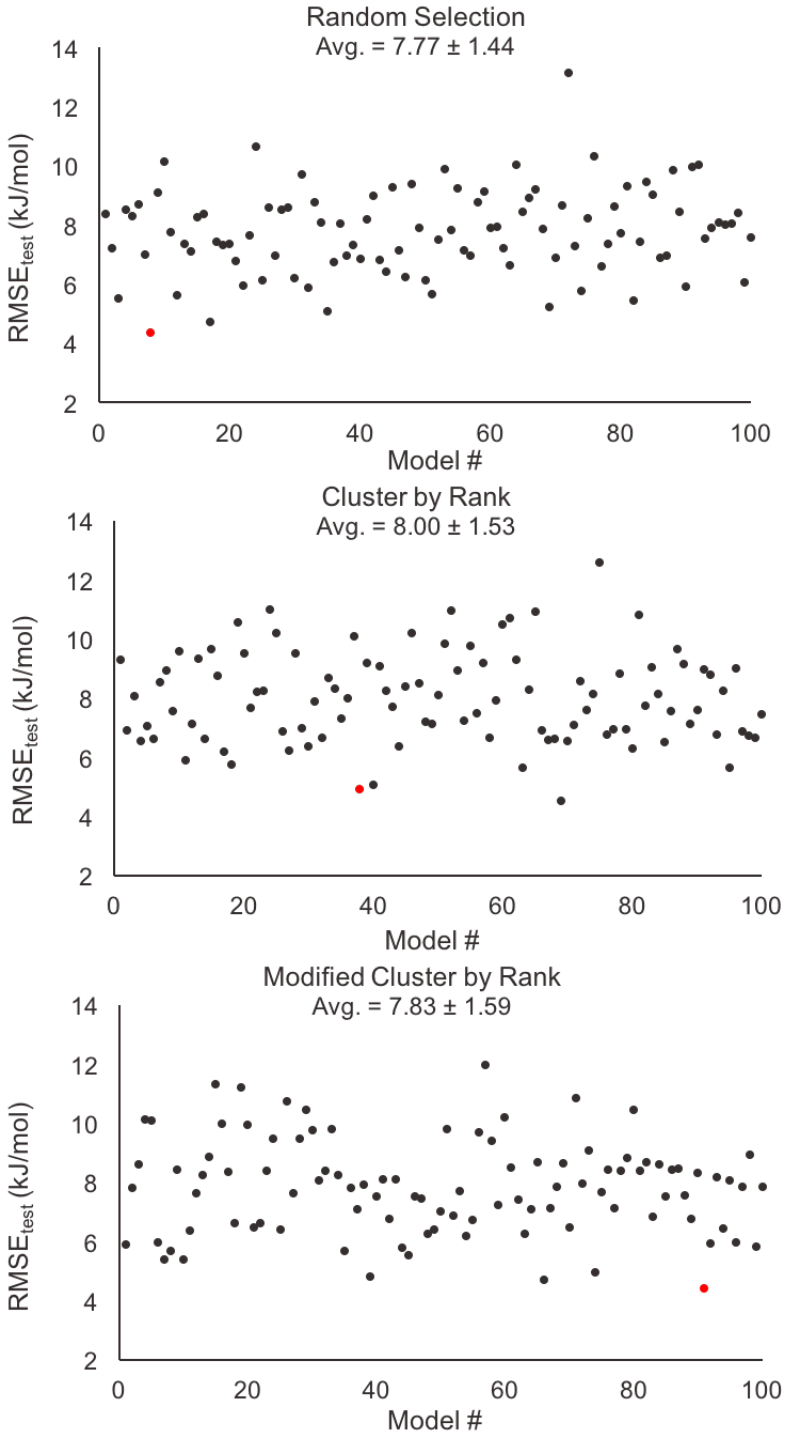


Figure A.5. Test set root-mean-square errors for the 100 QSAR models generated using (top) random selection, (middle) cluster by rank, and (bottom) modified cluster by rank algorithms. Red points indicate the optimal model for each dataset division algorithm.

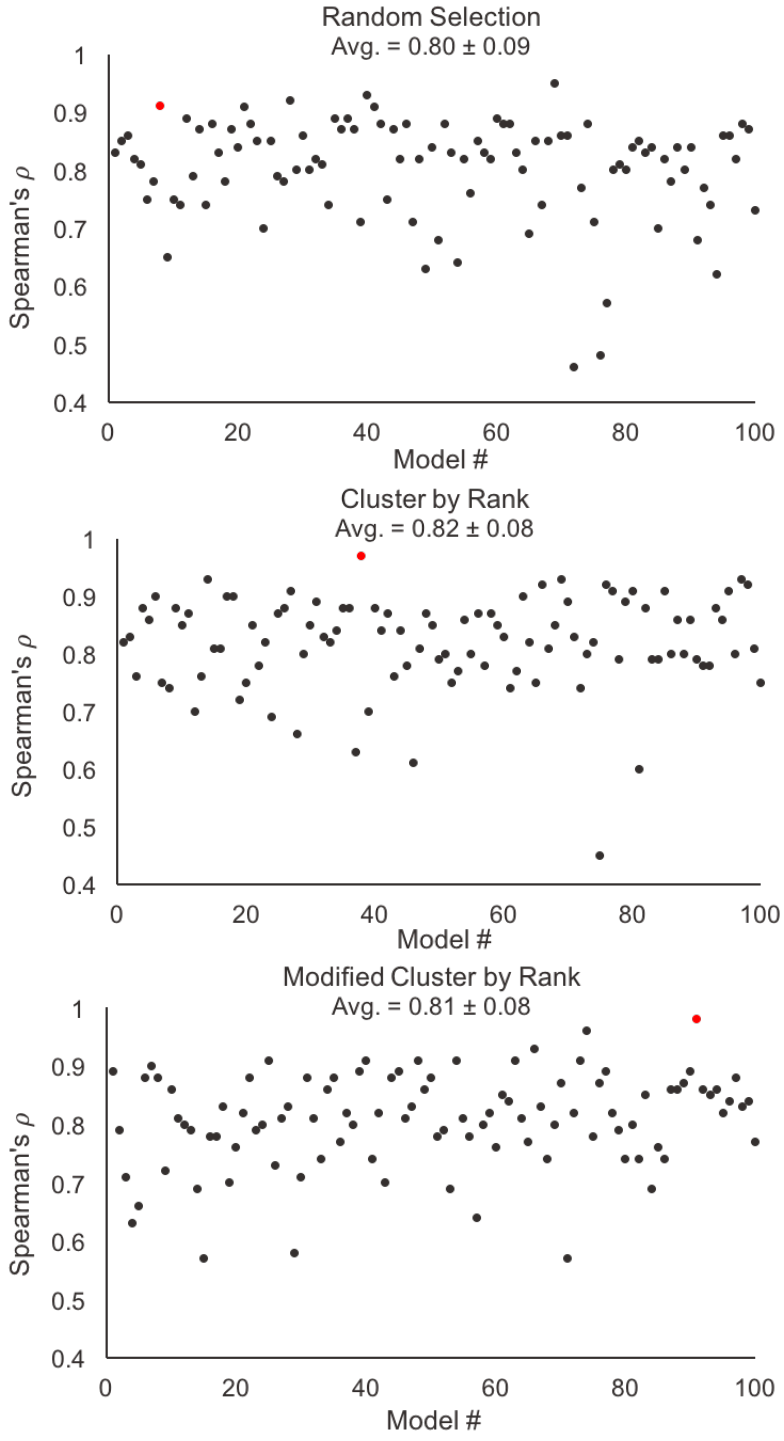


Figure A.6. Test set Spearman's rank correlation coefficients for the 100 QSAR models generated using (top) random selection, (middle) cluster by rank, and (bottom) modified cluster by rank algorithms. Red points indicate the optimal model for each dataset division algorithm.

A.7 QSAR Analysis – Optimal Random Selection Predictive QSAR Model

Table A.6. Predictions, residuals, and leverages for training, test, and external set molecules using the optimal random selection QSAR model (eq 2.12 in the main text).

Molecule #	Set	Observation (kJ/mol)	Prediction (kJ/mol)	Raw Residual (kJ/mol)	Standardized Residual	Leverage
1	Training	39.20	50.83	-11.63	-1.48	0.098
119	Training	128.28	92.98	35.30	4.43	0.067
57	Training	64.26	64.43	-0.17	-0.02	0.051
14	Training	51.45	48.43	3.02	0.37	0.045
80	Training	69.54	61.43	8.11	1.02	0.075
118	Training	104.55	89.29	15.27	1.90	0.053
86	Training	71.91	80.06	-8.15	-1.02	0.056
9	Training	50.47	43.52	6.95	0.88	0.086
35	Training	57.87	60.46	-2.60	-0.32	0.023
44	Training	60.12	63.23	-3.11	-0.38	0.023
90	Training	73.00	77.98	-4.99	-0.62	0.052
100	Training	84.09	83.55	0.54	0.07	0.056
6	Training	48.83	47.98	0.85	0.11	0.045
116	Training	98.39	86.82	11.57	1.44	0.047
73	Training	67.70	71.79	-4.09	-0.51	0.056
99	Training	80.93	80.60	0.33	0.04	0.042
70	Training	66.99	61.95	5.04	0.62	0.020
114	Training	95.30	91.62	3.68	0.46	0.064
74	Training	67.92	67.49	0.43	0.05	0.045
111	Training	91.67	89.19	2.49	0.31	0.058
49	Training	61.02	66.91	-5.89	-0.73	0.047
8	Training	49.08	57.96	-8.88	-1.09	0.032
54	Training	62.76	62.50	0.26	0.03	0.140
38	Training	58.96	66.02	-7.06	-0.88	0.051
105	Training	88.24	88.87	-0.63	-0.08	0.072
40	Training	59.57	55.89	3.68	0.46	0.045
56	Training	63.93	73.17	-9.25	-1.15	0.050
106	Training	89.08	76.71	12.37	1.54	0.059
75	Training	67.93	68.36	-0.42	-0.05	0.043
19	Training	52.41	46.01	6.40	0.79	0.047
21	Training	52.67	53.65	-0.98	-0.12	0.047
47	Training	60.65	68.65	-8.00	-1.00	0.050
48	Training	60.93	69.08	-8.16	-1.01	0.042

79	Training	69.30	70.13	-0.84	-0.10	0.045
5	Training	48.57	44.66	3.91	0.49	0.050
110	Training	91.11	91.41	-0.30	-0.04	0.063
46	Training	60.26	58.91	1.35	0.17	0.033
82	Training	69.84	78.27	-8.43	-1.05	0.055
102	Training	86.19	71.08	15.10	1.89	0.059
63	Training	65.28	59.36	5.92	0.74	0.054
58	Training	64.42	70.81	-6.39	-0.79	0.049
26	Training	55.72	53.49	2.23	0.28	0.047
41	Training	59.75	62.12	-2.38	-0.29	0.020
31	Training	57.20	54.64	2.57	0.32	0.076
39	Training	59.29	60.98	-1.69	-0.21	0.080
107	Training	89.59	89.43	0.16	0.02	0.058
72	Training	67.55	56.51	11.04	1.37	0.044
55	Training	63.07	57.26	5.81	0.72	0.048
22	Training	52.76	55.50	-2.75	-0.34	0.046
97	Training	79.80	89.83	-10.03	-1.26	0.075
113	Training	93.92	88.25	5.67	0.71	0.059
3	Training	45.36	57.96	-12.60	-1.56	0.041
109	Training	90.80	85.65	5.15	0.64	0.054
85	Training	71.67	69.17	2.50	0.31	0.052
112	Training	91.97	88.30	3.67	0.46	0.057
25	Training	55.58	64.19	-8.61	-1.07	0.056
42	Training	59.77	68.70	-8.93	-1.11	0.047
11	Training	51.12	48.85	2.27	0.28	0.044
77	Training	69.21	65.91	3.30	0.42	0.085
115	Training	97.30	76.09	21.20	2.65	0.057
7	Training	48.86	43.00	5.86	0.74	0.089
62	Training	65.23	69.19	-3.96	-0.49	0.053
33	Training	57.50	66.08	-8.58	-1.07	0.050
91	Training	74.03	76.93	-2.90	-0.36	0.056
2	Training	42.30	52.32	-10.03	-1.28	0.098
64	Training	66.17	75.98	-9.81	-1.22	0.053
71	Training	67.02	72.21	-5.18	-0.65	0.058
4	Training	47.06	48.45	-1.39	-0.20	0.305
81	Training	69.54	75.94	-6.40	-0.80	0.050
36	Training	58.59	65.57	-6.97	-0.87	0.046
61	Training	65.15	44.99	20.16	2.57	0.092

43	Training	60.05	61.33	-1.29	-0.16	0.023
18	Training	52.40	52.03	0.37	0.05	0.046
45	Training	60.12	59.93	0.20	0.02	0.054
67	Training	66.79	69.76	-2.97	-0.37	0.050
98	Training	80.79	77.93	2.86	0.36	0.050
92	Training	74.16	79.65	-5.50	-0.68	0.048
30	Training	57.11	67.32	-10.21	-1.27	0.046
103	Training	87.69	79.62	8.07	1.00	0.048
27	Training	56.16	59.55	-3.39	-0.42	0.032
117	Training	100.41	86.20	14.20	1.76	0.048
60	Training	64.98	66.43	-1.45	-0.18	0.047
17	Training	52.33	69.44	-17.11	-2.20	0.113
69	Training	66.87	68.87	-2.00	-0.25	0.053
95	Training	75.75	81.63	-5.88	-0.73	0.056
87	Training	71.93	69.84	2.09	0.26	0.045
76	Training	68.38	66.20	2.18	0.27	0.058
15	Training	52.05	55.55	-3.50	-0.43	0.046
93	Training	74.16	78.82	-4.65	-0.58	0.053
78	Training	69.27	69.81	-0.53	-0.07	0.046
59	Training	64.90	65.08	-0.18	-0.02	0.051
16	Training	52.10	52.90	-0.80	-0.10	0.062
37	Training	58.74	60.75	-2.01	-0.25	0.064
94	Training	74.46	91.50	-17.03	-2.20	0.120
96	Training	78.38	77.78	0.60	0.07	0.052
53	Training	61.86	51.60	10.26	1.36	0.165
68	Training	66.80	61.00	5.79	0.72	0.036
65	Training	66.67	78.95	-12.28	-1.54	0.061
66	Training	66.68	60.66	6.02	0.75	0.058
13	Training	51.22	48.86	2.36	0.29	0.045
20	Training	52.51	48.68	3.83	0.50	0.138
52	Test	61.67	64.10	-2.44	-0.56	0.017
88	Test	72.24	65.46	6.77	1.61	0.067
32	Test	57.50	60.81	-3.31	-0.77	0.026
24	Test	55.06	62.21	-7.15	-1.66	0.020
23	Test	53.47	57.23	-3.76	-0.88	0.037
34	Test	57.60	50.69	6.91	1.62	0.040
84	Test	71.27	73.01	-1.74	-0.41	0.048
29	Test	56.61	61.13	-4.52	-1.05	0.028

28	Test	56.21	58.51	-2.31	-0.55	0.070
108	Test	90.27	89.01	1.26	0.30	0.057
51	Test	61.52	63.62	-2.10	-0.49	0.022
83	Test	70.83	65.13	5.70	1.34	0.049
10	Test	50.73	47.58	3.15	0.74	0.046
50	Test	61.17	67.90	-6.73	-1.56	0.020
12	Test	51.18	52.19	-1.01	-0.24	0.064
104	Test	88.04	89.70	-1.66	-0.39	0.060
89	Test	72.29	68.51	3.79	0.89	0.045
101	Test	84.24	79.12	5.11	1.19	0.038
120	External	57.83	61.82	-3.99	-0.35	0.020
121	External	83.91	64.73	19.18	1.82	0.140
122	External	72.32	74.56	-2.24	-0.20	0.042

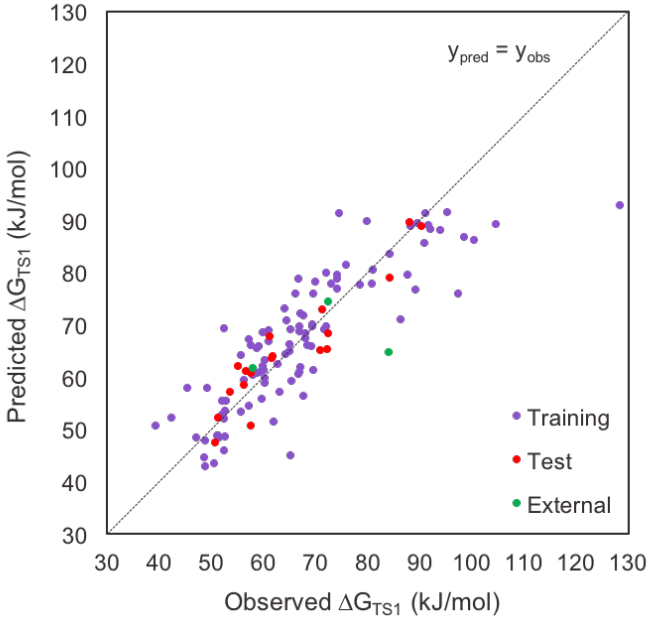


Figure A.7. Comparison between observed (DFT-calculated) free energy barriers with those predicted by the optimal predictive QSAR model developed using the random selection algorithm (eq 2.12 in the main text). The dashed line shows the relationship $y_{pred} = y_{obs}$, where data points for ideal models should lie close to this line.

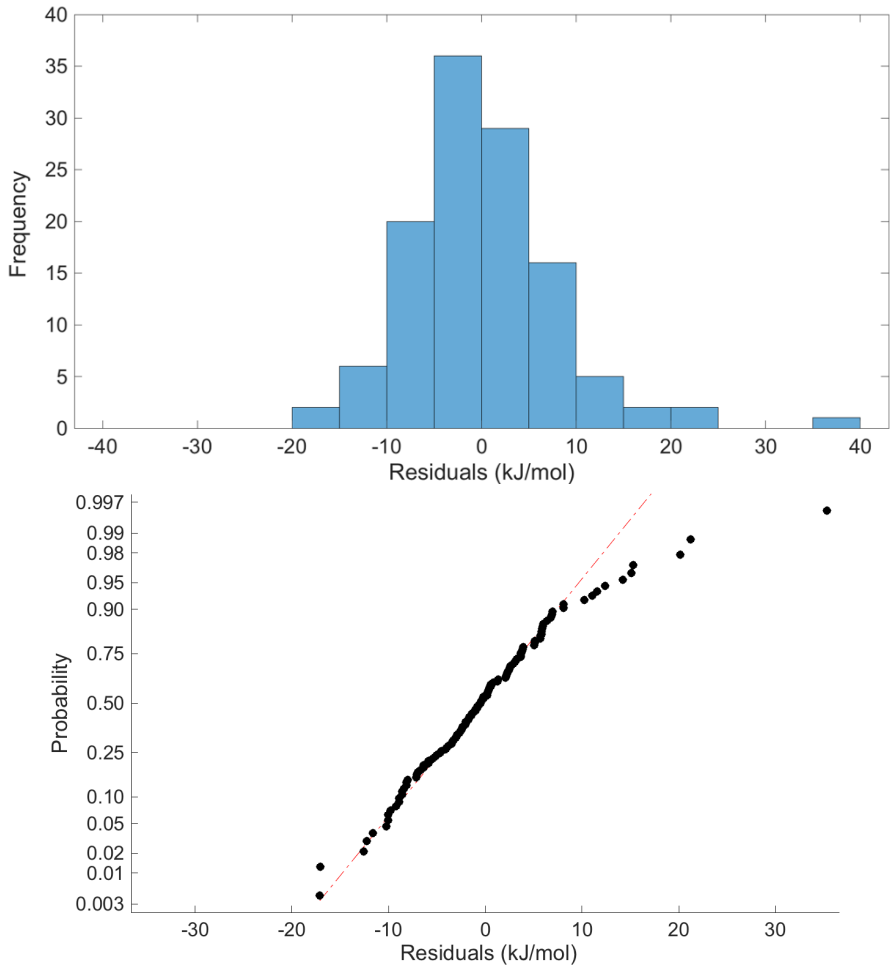


Figure A.8. Distribution of residuals for the optimal predictive QSAR model developed using the random selection algorithm (eq 2.12 in the main text). The (top) histogram of raw residuals and (bottom) normal probability plot of raw residuals show that the residuals have an acceptable normal distribution, with one outlier (molecule 119).

A.8 QSAR Analysis – Optimal Cluster by Rank Predictive QSAR Model

$$\Delta G_{TS1} = (72.86 \pm 5.46) + (60.29 \pm 10.80)q_P - (80.76 \pm 13.40)BO - (13.58 \pm 8.17)pK_a - (9.18 \pm 3.43)\omega + (10.75 \pm 5.11)V \quad (A.9)$$

Training set: $N_m = 101$, $Q^2_{LOO} = 0.54$, ${}^C R_P^2 = 0.68$, $R^2_{train} = 0.70$, $F = 45.0$, $RMSE_{train} = 8.50$ kJ/mol
Test set: $N_m = 18$, $R^2_{test} = 0.86$, $RMSE_{test} = 4.91$ kJ/mol, Spearman's $\rho = 0.97$

A.9 QSAR Analysis – Optimal Modified Cluster by Rank Predictive QSAR Model

$$\Delta G_{TS1} = (139.75 \pm 17.72) + (75.14 \pm 10.46)q_P - (95.61 \pm 13.37)BO - (21.56 \pm 8.68)pK_a - (67.00 \pm 17.57)\eta - (74.55 \pm 19.06)SOF \quad (A.10)$$

Training set: $N_m = 101$, $Q^2_{LOO} = 0.62$, ${}^C R_P^2 = 0.72$, $R^2_{train} = 0.72$, $F = 48.9$, $RMSE_{train} = 8.24$ kJ/mol
Test set: $N_m = 18$, $R^2_{test} = 0.89$, $RMSE_{test} = 4.40$ kJ/mol, Spearman's $\rho = 0.98$

A.10 References for Appendix A

1. US EPA. Toxicity Estimation Software Tool: A Program to Estimate Toxicity from Molecular Structure, **2016**.
2. Cook, H. G.; Saunders, B. C.; Smith, F. E. *J. Chem. Soc.* **1949**, *0*, 635–638.
3. Evans, D. A. "pKa Table." can be found under evans.rc.fas.harvard.edu/pdf/evans_pKa_table.pdf, **2005**.
4. *CRC Handbook of Chemistry and Physics*, CRC Press LLC, **2010**.
5. Bourne, N.; Chrystiuk, E.; Davis, A. M.; Williams, A. *J. Am. Chem. Soc.* **1988**, *110*, 1890–1895.
6. Schwarzenbach, R. P.; Stierli, R.; Folsom, B. R.; Zeyer, J. *Environ. Sci. Technol.* **1988**, *22*, 83–92.
7. Wille, T.; Scott, C.; Thierman, H.; Worek, F. *Biocatl. Biotransform.* **2011**, *30*, 203–208.
8. ChemAxon Ltd. Chemicalize, can be found under https://chemicalize.com.
9. Serjeant, P.; Dempsey, B. in *Ionisation Constants of Organic Acids in Aqueous Solution*, Pergamon Press, **1979**.
10. The Metabolomics Innovation Centre. FooDB, can be found under foodb.ca/compounds/FDB029632.

Appendix B: Supporting Information for Chapter 3

This appendix is a modified version of the Supporting Information for the manuscript in Chapter 3 (*J. Phys. Chem. C* **2018**, *122*, 12362–12368).

B.1 Benchmarking Different Levels of Theory for the Uncatalyzed Reaction

Table B.1. Free energy barriers ($\Delta G^{\ddagger}$), enthalpy barriers ($\Delta H^{\ddagger}$), and reaction free energies (ΔG_r) of the uncatalyzed DMNP hydrolysis reaction obtained at T = 298.15 K and P = 1 atm. All energies are in kJ/mol. Experimental data obtained from: Ginjaar, L.; Vel, S. *Rec. trav. chim.* **1958**, *77*, 956.

Single point energy//geometry	$\Delta G^{\ddagger}$	$\Delta H^{\ddagger}$	ΔG_r
B3LYP/6-311++G**//B3LYP/6-31+G**	68	23	-161
B3LYP-D3(BJ)/6-311++G**//B3LYP/6-31+G**	55	9	-147
M06/6-311++G**//B3LYP/6-31+G**	44	-1	-167
M06-D3/6-311++G**//B3LYP/6-31+G**	42	-4	–
M06-2X/6-311++G**//B3LYP/6-31+G**	44	-1	–
M06/6-311++G**//M06/6-31+G**	42	-3	-166
M06-D3/6-311++G**//M06/6-31+G**	40	-5	–
M06-2X/6-311++G**//M06/6-31+G**	42	-3	–
HF/6-31+G*//HF/6-31+G*	102	53	–
MP2/6-311++G**//B3LYP/6-31+G**	57	12	–
MP2/6-311++G**//M06/6-31+G**	56	12	–
MP2/6-311++G**//HF/6-31+G*	63	15	–
CBS-QB3	68	19	–
Experiment	81	49	–

Comparing the enthalpy barriers in Table B.1, the "lowest" level of theory HF/6-31+G* corresponds to the best agreement with experiment. We believe that HF performs so well in this situation due to a fortuitous error cancellation. The B3LYP method performs remarkably similar to CBS-QB3, which inspires confidence in using B3LYP to study these reactions. For modeling the NU-1000-catalyzed reactions, the inclusion of dispersion corrections is necessary since the adsorbate has significant dispersion interactions with the MOF, and thus the B3LYP-D3(BJ) level of theory was chosen for all calculations. We note that a recent computational study using NU-1000 clusters found good agreement among the M06-L, PBE-D3, B3LYP-D3, and M06 functionals for describing trends between energy barriers and molecular descriptors.[1]

Table B.2. Electronic energies, free energies, and enthalpies at different levels of theory for the optimized species of the uncatalyzed DMNP hydrolysis reaction obtained at T = 298.15 K and P = 1 atm.

Single point energy//geometry	Species	E (hartree)	G (hartree)	H (hartree)
B3LYP/6-311++G**//B3LYP/6-31+G**	DMNP	-1158.384872	-1158.432389	-1158.367401
	OH	-75.950644	-75.966902	-75.947339
	TS	-1234.325195	-1234.373353	-1234.306065
B3LYP-D3(BJ)/6-311++G**//B3LYP/6-31+G**	DMNP	-1158.435474	-1158.482991	-1158.418003
	OH	-75.950890	-75.967148	-75.947585
	TS	-1234.381147	-1234.429305	-1234.362017
M06/6-311++G**//B3LYP/6-31+G**	DMNP	-1157.870851	-1157.918368	-1157.853380
	OH	-75.907261	-75.923519	-75.903956
	TS	-1233.777007	-1233.825165	-1233.757877
M06-D3/6-311++G**//B3LYP/6-31+G**	DMNP	-1157.874476	-1157.921993	-1157.857005
	OH	-75.907280	-75.923538	-75.903975
	TS	-1233.781453	-1233.829611	-1233.762323
M06-2X/6-311++G**//B3LYP/6-31+G**	DMNP	-1158.026654	-1158.074171	-1158.009183
	OH	-75.909315	-75.925573	-75.906010
	TS	-1233.934878	-1233.983036	-1233.915748
M06/6-311++G**//M06/6-31+G**	DMNP	-1157.873435	-1157.920735	-1157.856125
	OH	-75.907088	-75.923336	-75.903783
	TS	-1233.779900	-1233.828047	-1233.760911
M06-D3/6-311++G**//M06/6-31+G**	DMNP	-1157.877084	-1157.924384	-1157.859774
	OH	-75.907088	-75.923336	-75.903783
	TS	-1233.784262	-1233.832409	-1233.765273
M06-2X/6-311++G**//M06/6-31+G**	DMNP	-1158.028983	-1158.076283	-1158.011673
	OH	-75.909107	-75.925355	-75.905802
	TS	-1233.937633	-1233.985780	-1233.918644
HF/6-31+G*//HF/6-31+G*	DMNP	-1152.916345	-1152.963684	-1152.899761
	OH	-75.502386	-75.518611	-75.499082
	TS	-1228.396591	-1228.443567	-1228.378676
MP2/6-311++G**//B3LYP/6-31+G**	DMNP	-1155.748443	-1155.795960	-1155.730972
	OH	-75.763154	-75.779412	-75.759849
	TS	-1231.505447	-1231.553605	-1231.486317
MP2/6-311++G**//M06/6-31+G**	DMNP	-1155.748380	-1155.795680	-1155.731070
	OH	-75.762906	-75.779154	-75.759601
	TS	-1231.505204	-1231.553351	-1231.486215
MP2/6-311++G**//HF/6-31+G*	DMNP	-1155.725739	-1155.773078	-1155.709156
	OH	-75.762514	-75.778738	-75.759210
	TS	-1231.480710	-1231.527686	-1231.462796
CBS-QB3	DMNP	-1156.770051	-1156.817424	-1156.752574
	OH	-75.846046	-75.862308	-75.842742
	TS	-1232.606791	-1232.653686	-1232.588209

B.2 Comparison of Regular and Distorted Node Structures

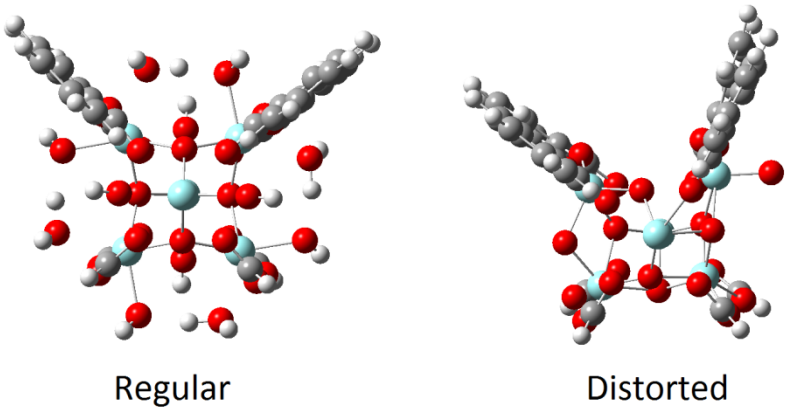


Figure B.1. Starting structures of (left) regular and (right) distorted NU-1000 node used in this work. Color scheme: H - white, C - grey, O - red, Zr - cyan.

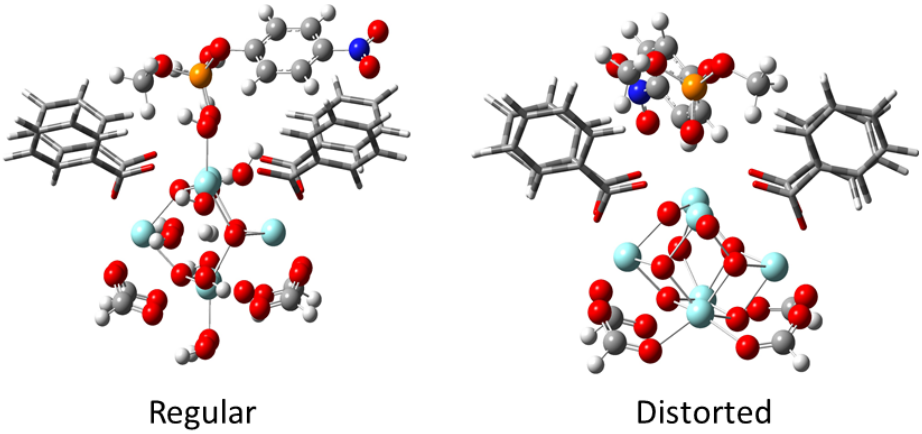


Figure B.2. Comparison of the binding mode of DMNP on (left) regular and (right) distorted NU-1000 node. Color scheme: H - white, C - grey, N - blue, O - red, P - orange, Zr - cyan.

B.3 Effect of Dispersion Corrections on the Reaction Free Energy Profiles

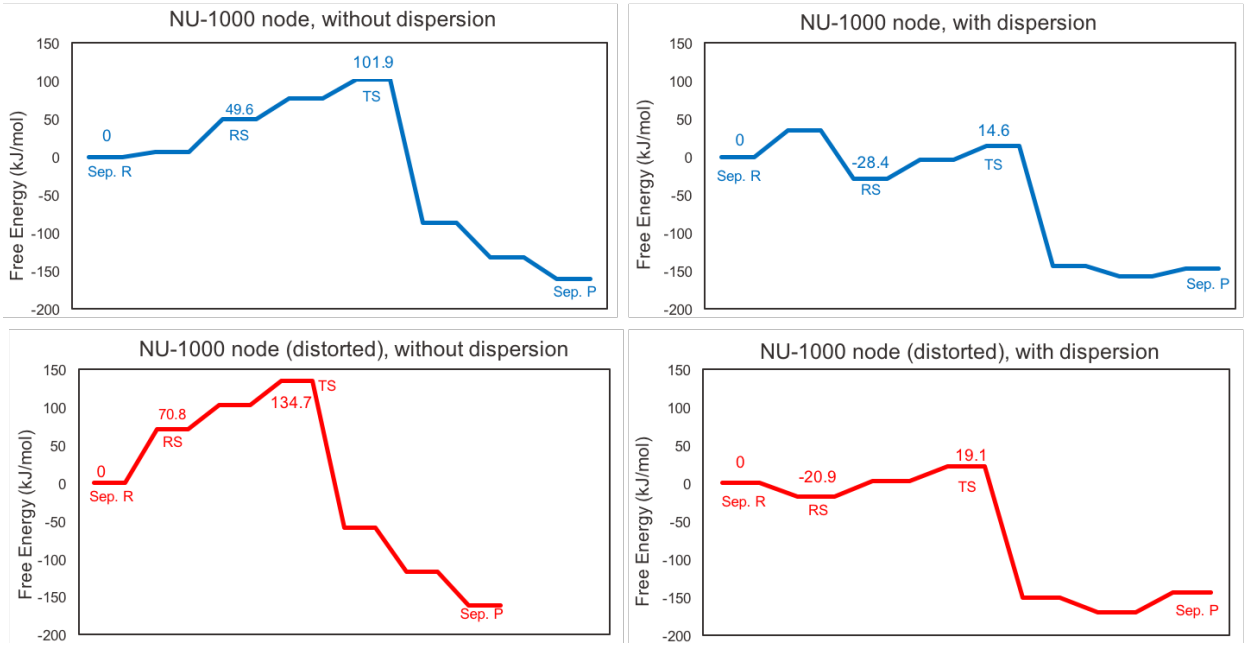


Figure B.3. Comparison of the free energy profiles of the DMNP hydrolysis reaction catalyzed by (top) regular and (bottom) distorted NU-1000 node obtained (left) without and (right) with dispersion corrections. “Sep. R”, “RS”, “TS”, and “Sep. P” correspond to separated reactants, node–DMNP complex, transition state, and separated products, respectively.

B.4 Natural Bond Orbital Population Analysis

Table B.3. Partial atomic charges for the uncatalyzed transition state, calculated using the NBO method at B3LYP-D3(BJ)/6-311++G**//B3LYP/6-31+G**. The ordering of atoms is the same as in the Cartesian coordinates (provided in the Supporting Information document but omitted here for conciseness).

Atom	NBO Charge
C	-0.237180
C	-0.161960
C	0.037130
C	-0.155000
C	-0.265940
C	0.373370
H	0.232820
H	0.245160
H	0.244780
H	0.238150
N	0.483120
O	-0.426610
O	-0.426600
O	-0.788480
P	2.536730
O	-0.842310
O	-0.829190
O	-1.096920
C	-0.206340
H	0.196200
H	0.185390
H	0.184920
C	-0.207260
H	0.172320
H	0.232600
H	0.184700
O	-1.316250
H	0.412630

Table B.4. Partial atomic charges for the NU-1000 node, calculated using the NBO method at B3LYP-D3(BJ)/6-311++G**//B3LYP/6-31+G**. The ordering of atoms is the same as in the Cartesian coordinates (provided in the Supporting Information document but omitted here for conciseness).

Atom	NBO Charge
Zr	1.969840
Zr	1.945990
Zr	1.974020
Zr	2.292270
Zr	1.989040
Zr	1.986110
O	-0.692390
O	-0.689190
C	0.725530
O	-0.705920
O	-0.690440
H	0.220680
O	-0.698410
O	-0.692910
H	0.221310
O	-0.685510
O	-0.685800
C	0.729070
O	-1.052750
H	0.495680
O	-0.700390
O	-0.662560
C	0.726070
O	-0.706680
O	-0.693170
H	0.221050
O	-0.692330
O	-0.717260
H	0.221440
O	-0.678760
O	-0.702060
C	0.726670
O	-1.051630
H	0.497100

O	-1.073920
H	0.495170
O	-0.963830
H	0.517890
O	-1.083010
O	-1.099110
O	-0.978800
H	0.516160
O	-0.966200
H	0.518720
O	-1.085420
O	-0.974580
O	-1.095690
H	0.535140
O	-0.907310
H	0.516930
H	0.535030
O	-0.905090
H	0.512370
O	-0.908210
H	0.537870
H	0.518160
H	0.135430
H	0.134490
H	0.134950
H	0.133020
C	-0.195700
C	-0.195380
C	-0.195890
C	-0.183540
C	-0.167270
H	0.226990
C	-0.146150
C	0.847730
C	-0.176560
H	0.226200
C	-0.193340
H	0.215640

C	-0.211820
H	0.216320
C	-0.166700
H	0.227030
C	-0.145970
C	0.847770
C	-0.175960
H	0.226390
C	-0.193380
H	0.215710
C	-0.211790
H	0.216390
C	-0.167530
H	0.227020
C	-0.145920
C	0.842780
C	-0.176540
H	0.226220
C	-0.193530
H	0.215570
C	-0.211890
H	0.216330
C	-0.163790
H	0.227820
C	-0.148210
C	0.853540
C	-0.165440
H	0.226200
C	-0.211840
H	0.217160
C	-0.211500
H	0.217100
O	-1.029170
H	0.484180
H	0.520920

Table B.5. Partial atomic charges for the transition state on the NU-1000 node, calculated using the NBO method at B3LYP-D3(BJ)/6-311++G**//B3LYP/6-31+G**. The ordering of atoms is the same as in the Cartesian coordinates (provided in the Supporting Information document but omitted here for conciseness).

Atom	NBO Charge
Zr	1.974430
Zr	1.949940
Zr	1.980050
Zr	1.953890
Zr	1.994450
Zr	1.987020
O	-0.693660
O	-0.689400
C	0.721700
O	-0.745680
O	-0.695530
H	0.219650
O	-0.693710
O	-0.705020
H	0.219800
O	-0.682850
O	-0.689650
C	0.728890
O	-1.054230
H	0.493750
O	-0.699210
O	-0.664020
C	0.725990
O	-0.745450
O	-0.696670
H	0.219410
O	-0.732700
O	-0.728000
H	0.219540
O	-0.679820
O	-0.702980
C	0.724390
O	-1.054300
H	0.495560

O	-1.053710
H	0.494080
O	-0.964270
H	0.514330
O	-1.084150
O	-1.084920
O	-0.963900
H	0.515450
O	-0.966440
H	0.516420
O	-1.085050
O	-0.980240
O	-1.078650
H	0.509690
O	-0.904240
H	0.533770
H	0.535850
O	-0.906860
H	0.511160
O	-0.909420
H	0.537760
H	0.517070
H	0.134470
H	0.133010
H	0.131990
H	0.130780
C	-0.190180
C	-0.197070
C	-0.191300
C	-0.187440
C	-0.173310
H	0.225270
C	-0.183810
C	0.985740
C	-0.171210
H	0.225480
C	-0.213090
H	0.215550

C	-0.213070
H	0.215430
C	-0.168150
H	0.226640
C	-0.145020
C	0.842310
C	-0.177910
H	0.225910
C	-0.193730
H	0.215160
C	-0.212160
H	0.215890
C	-0.172870
H	0.224470
C	-0.183620
C	0.973890
C	-0.170070
H	0.226790
C	-0.213960
H	0.214890
C	-0.214810
H	0.214940
C	-0.169860
H	0.226080
C	-0.187620
C	0.995710
C	-0.170820
H	0.221290
C	-0.211210
H	0.216430
C	-0.211420
H	0.216330
O	-1.020690
H	0.484600
H	0.531230
O	-1.046090
C	-0.273910
C	-0.163390

C	0.053310
C	-0.162500
C	-0.275030
C	0.482870
H	0.237360
H	0.246770
H	0.245550
H	0.266060
N	0.487670
O	-0.413720
O	-0.411480
O	-0.862920
P	2.648180
O	-0.845830
O	-0.834520
C	-0.211740
H	0.239870
H	0.185330
H	0.176240
C	-0.208430
H	0.185590
H	0.175820
H	0.241300
O	-1.323580
H	0.426870

B.5 References for Appendix B

1. Simons, M. C.; Ortuño, M. A.; Bernales, V.; Gaggioli, C. A.; Cramer, C. J.; Bhan, A.; Gagliardi, L. C–H Bond Activation on Bimetallic Two-Atom Co-M Oxide Clusters Deposited on Zr-Based MOF Nodes: Effects of Doping at the Molecular Level. *ACS Catal.* **2018**, *8*, 2864-2869.

B.6 Optimized Cartesian Coordinates (in Å) and Free Energies of Important Species

For conciseness, the Cartesian coordinates are omitted here, but they are available free of charge in the Supporting Information at https://doi.org/10.1021/acs.jpcc.8b03641.

Appendix C: Supporting Information for Chapter 4

This appendix is the preliminary version of the Supporting Information for the manuscript in Chapter 4 (*in preparation*).

C.1 Hypothetical Hydrolysis Pathways on M(IV)-MOF Nodes

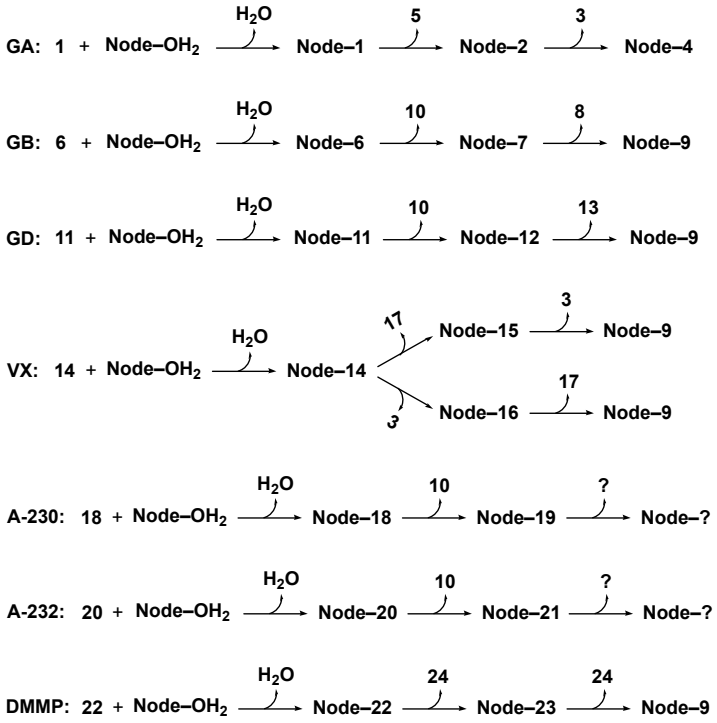


Figure C.1. Hypothetical hydrolysis pathways for the nerve agent and simulant molecules occurring on M(IV)-MOF nodes, using the molecule numbering established in Figure 4.1a of the main text. The terminology "Node–**N**" represents molecule **N** bound to the node. Numbers listed above the curved reaction arrows represent hydrolysis product molecules (monodentate anions, alcohols, and thiols) released into solution. The question marks in the pathways for A-230 and A-232 indicate that the full hydrolysis mechanisms for these Novichok agents are currently unknown.

C.2 Representative Binding Modes

In Figure C.2a, the *node–noOH*$_2$ structure represents a bare node where the terminal –OH$_2$ group has been removed from the active (binding) site, which was used in the formula to calculate the binding energies of neutral molecules throughout the study. In Figure C.2b, the *node–noOH* structure represents a bare node where the terminal –OH group has been removed from the binding site, which was used in the formula to calculate the binding energies of monodentate anions. In Figure C.2c, the *node–noOH*$_2$*,OH* structure represents a bare node where the terminal –OH$_2$ and –OH groups have been removed from the binding site, which was used in the formula to calculate the binding energies of bidentate anions.

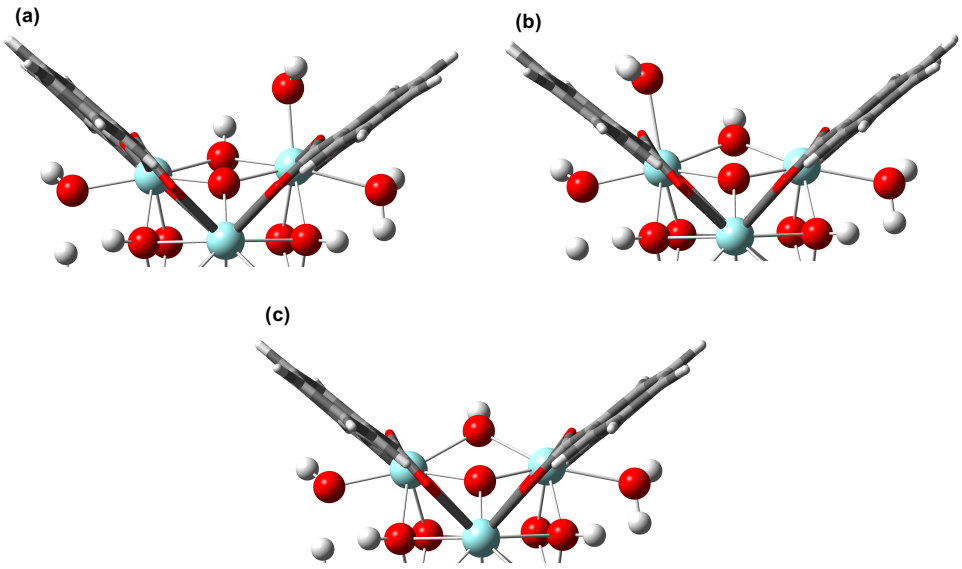


Figure C.2. Optimized cluster models for the (a) *node–noOH*$_2$, (b) *node–noOH*, and (c) *node–noOH*$_2$*,OH* bare node structures of Zr-NU-1000 (large pore). The benzoate linkers around the binding site are shown in tube format for clarity, and only the top half of the node is shown to highlight the site where binding occurs. Dark gray, white, red, and turquoise spheres represent C, H, O, and Zr atoms, respectively.

Figure C.3 shows structures that are representative of the binding modes for (a) nerve agent and simulant molecules, (b) alcohol and thiol product molecules, (c) bidentate anions, and (d) monodentate anions.

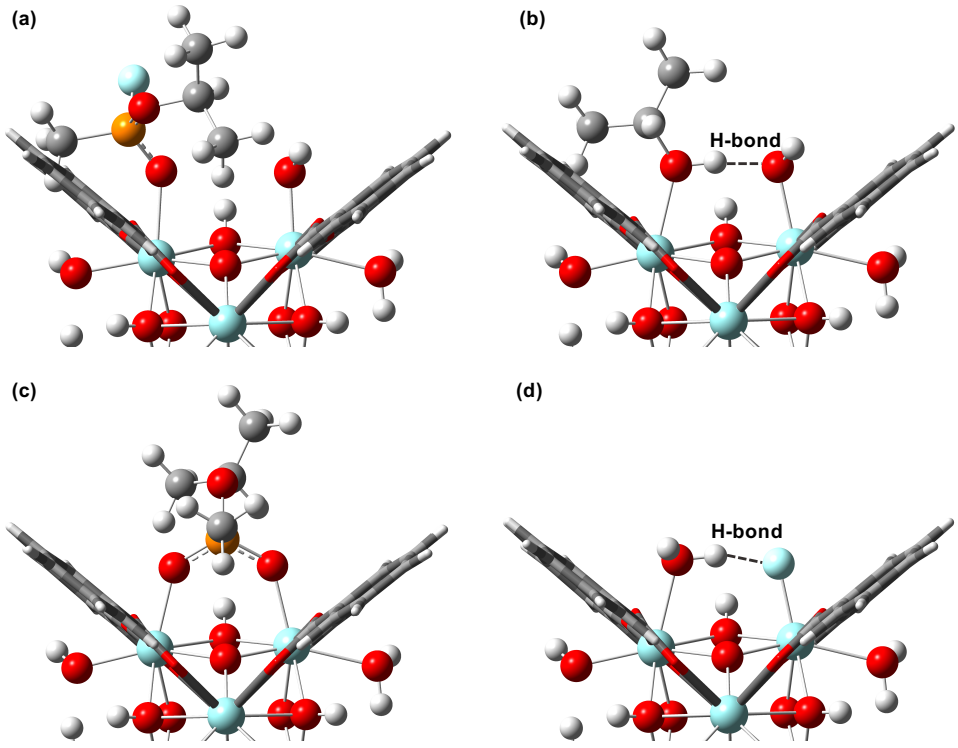


Figure C.3. Optimized cluster models for molecules (a) **6** (sarin, GB), (b) **8** (isopropanol), (c) **7** (isopropyl methylphosphonic acid, IMPA), and (d) **10** (fluoride anion) bound to the node of Zr-NU-1000 (large pore). The benzoate linkers around the binding site are shown in tube format for clarity, and only the top half of the node is shown to highlight the site where binding occurs. Dark gray, white, red, orange, light blue, and turquoise spheres represent C, H, O, P, F, and Zr atoms, respectively.

C.3 Relative Stabilities of Different Binding Orientations: Nerve Agent/Simulant Molecules

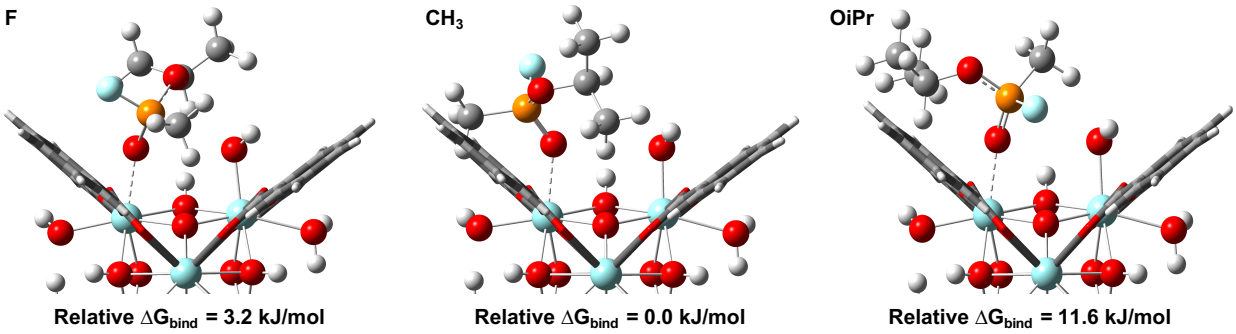


Figure C.4. Optimized cluster models for the three possible orientations of GB (**6**) bound to the node of Zr-NU-1000 (large pore), along with their relative binding free energies. Each orientation is named according to the R group (F, CH_3, or OiPr) that is approximately collinear with the neighboring terminal node Zr–OH group. The benzoate linkers around the binding site are shown in tube format for clarity, and only the top half of the node is shown to highlight the site where binding occurs. Dark gray, white, red, orange, light blue, and turquoise spheres represent C, H, O, P, F, and Zr atoms, respectively.

Table C.1. Relative binding free energies (in kJ/mol) for different orientations of nerve agent and simulant molecules bound to Zr(IV)-MOF node sites. Values in red indicate the most favorable molecular orientation for each node site.

Molecule	Orientation	MOF-808	NU-1000 (large pore)	bi(trans)-defect UiO-66	mono-defect UiO-66	bi(cis)-defect UiO-66	NU-1000 (c pore)
1	OEt	0.0	4.6	0.0	14.8	30.8	3.1
	NMe_2	7.3	0.0	4.2	0.0	0.0	0.0
	CN	43.6	12.5	11.6	18.8	24.2	3.4
6	OiPr	0.0	11.6	29.7	12.6	0.4	17.7
	CH_3	29.5	0.0	0.0	0.0	14.8	19.7
	F	39.5	3.2	12.5	4.4	0.0	0.0
11	OR	0.0	16.8	9.3	5.3	0.0	0.0
	CH_3	21.9	0.0	0.0	3.8	2.9	35.2
	F	53.4	19.0	18.3	0.0	18.1	6.6
14	CH_3	0.0	0.0	0.0	0.0	0.0	6.8
	SR	2.4	30.4	68.6	16.4	20.9	0.0
	OEt	5.8	44.8	N/A[a]	17.9	13.4	11.8
18	CH_3	0.0	0.0	1.3	0.0	0.0	10.3
	NR	9.2	14.9	19.1	1.9	23.9	25.2
	F	20.3	4.9	0.0	3.9	7.2	0.0
20	F	0.0	23.8	19.3	21.2	0.0	12.8
	NR	6.2	12.4	24.2	20.3	1.6	39.0
	OMe	19.1	0.0	0.0	0.0	12.3	0.0
22	OMe	0.0	0.0	7.3	0.0	0.0	0.9
	CH3	10.7	14.5	0.0	9.6	28.1	0.0

[a]N/A indicates that the OEt orientation for molecule **14** could not be optimized on the Zr-bi(trans)-defect UiO-66 node, despite multiple attempts.

C.4 Effects of Molecular Size and Dispersion on Binding Energy: Neutral Molecules

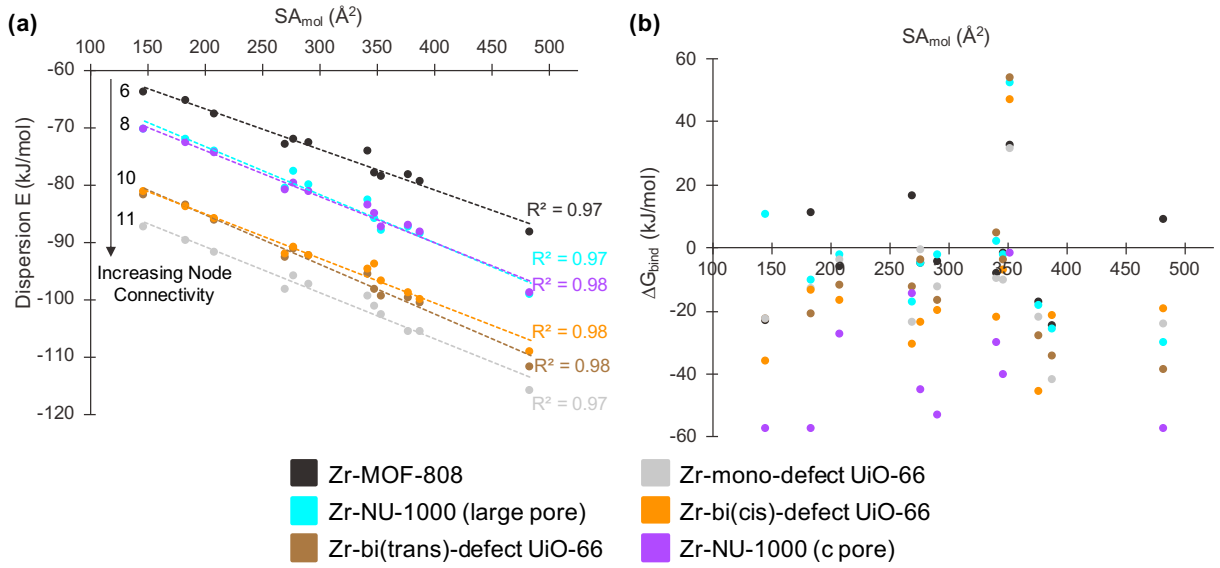


Figure C.5. The effect of molecular solvent-accessible surface area (SA_{mol}) on (a) the dispersion energy and (b) the binding free energy computed for node-bound neutral molecules. Neutral molecules refer to molecules **1**, **3**, **6**, **8**, **11**, **13**, **14**, **17**, **18**, **20**, **22**, **24**, and **H_2O**.

C.5 Relative Stabilities of Different Binding Orientations: Bidentate Anions

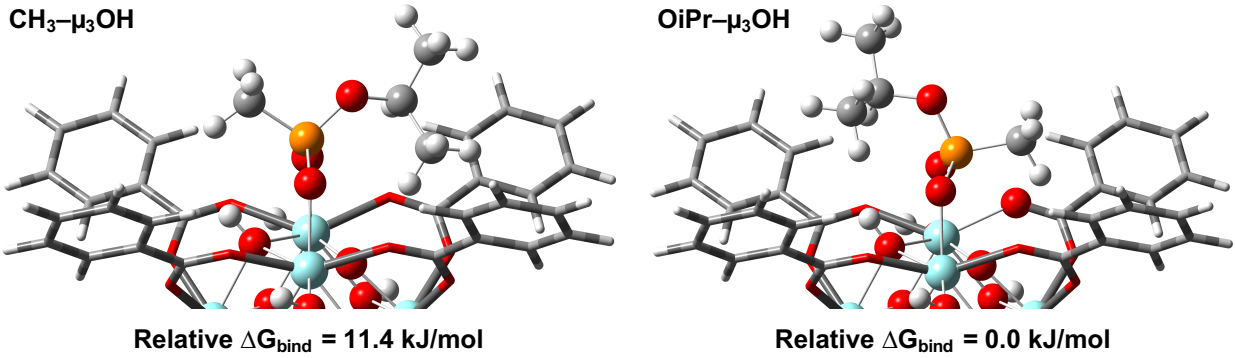


Figure C.6. Optimized cluster models for the two possible orientations of IMPA (**7**) bound to the node of Zr-NU-1000 (large pore), along with their relative binding free energies. Each orientation is named according to the R group (CH_3 or OiPr) that is directed towards the bridging node hydroxyl group at the binding site. The benzoate linkers around the binding site are shown in tube format for clarity, and only the top half of the node is shown to highlight the site where binding occurs. Dark gray, white, red, orange, and turquoise spheres represent C, H, O, P, and Zr atoms, respectively.

Table C.2. Relative binding free energies (in kJ/mol) for different orientations of bidentate anion products bound to Zr(IV)-MOF node sites. Values in green indicate the most favorable orientation for each node site.

Molecule	Orientation	MOF-808	NU-1000 (large pore)	bi(trans)-defect UiO-66	mono-defect UiO-66	bi(cis)-defect UiO-66	NU-1000 (c pore)
2	NMe_2–μ_3OH	0.0	0.0	0.0	9.1	12.8	0.0
	OEt–μ_3OH	12.9	14.5	5.1	0.0	0.0	7.8
4	NMe_2–μ_3OH	0.0	4.0	3.4	0.0	0.0	8.7
	OH–μ_3OH	4.2	0.0	0.0	1.7	7.1	0.0
7	CH_3–μ_3OH	0.0	11.4	0.1	11.6	3.7	0.0
	OiPr–μ_3OH	23.9	0.0	0.0	0.0	0.0	4.5
9	CH_3–μ_3OH	0.0	0.0	0.5	0.2	0.0	0.0
	OH–μ_3OH	9.4	8.9	0.0	0.0	10.8	0.2
12	CH_3–μ_3OH	0.0	10.8	4.4	11.0	11.6	3.5
	OR–μ_3OH	7.6	0.0	0.0	0.0	0.0	0.0
15	CH_3–μ_3OH	0.0	0.0	0.03	2.0	0.0	0.0
	OEt–μ_3OH	30.0	35.8	0.0	0.0	2.8	0.4
16	CH_3–μ_3OH	0.0	24.3	16.7	11.1	5.1	15.5
	SR–μ_3OH	11.0	0.0	0.0	0.0	0.0	0.0
19	CH_3–μ_3OH	0.0	0.0	0.2	5.1	24.9	4.5
	NR–μ_3OH	24.8	6.6	0.0	0.0	0.0	0.0
21	OMe–μ_3OH	0.0	0.0	0.0	0.0	9.1	0.0
	NR–μ_3OH	17.0	7.8	5.6	26.5	0.0	1.0
23	CH_3–μ_3OH	0.0	10.9	1.9	12.2	4.3	3.4
	OMe–μ_3OH	4.8	0.0	0.0	0.0	0.0	0.0

C.6 Effects of Molecular Size and Dispersion on Binding Energy: Bidentate Anions

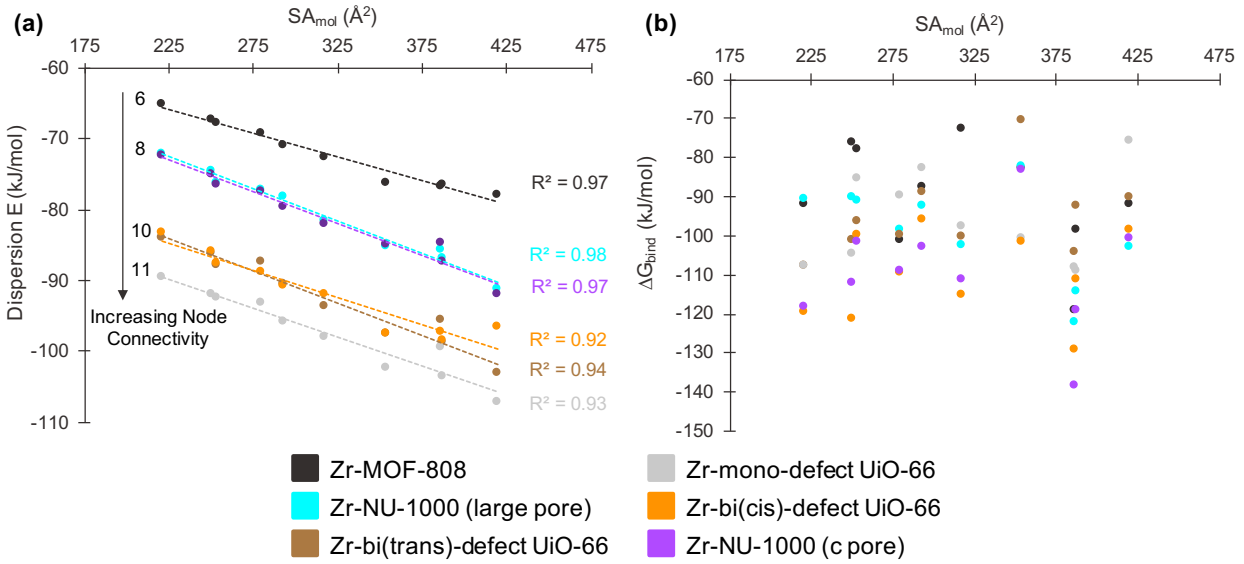


Figure C.7. The effect of molecular solvent-accessible surface area (SA_{mol}) on (a) the dispersion energy and (b) the binding free energy computed for node-bound bidentate anions. Bidentate anions refer to molecules **2**, **4**, **7**, **9**, **12**, **15**, **16**, **19**, **21**, and **23**.

C.7 Binding Free Energy Values for All Molecules: Zr(IV)-MOF Nodes

Table C.3. Binding free energies (in kJ/mol) for molecules bound to Zr(IV)-MOF node sites in their most favorable orientations, corresponding to results in Figure 4.2. Negative and positive values indicate favorable and unfavorable binding, respectively.

Molecule	MOF-808	NU-1000 (large pore)	bi(trans)-defect UiO-66	mono-defect UiO-66	bi(cis)-defect UiO-66	NU-1000 (c pore)
H_2O	-7.9	-28.1	-29.9	-40.9	-54.5	-61.1
1	-8.5	1.9	4.6	-10.0	-22.6	-30.3
2	-72.9	-102.6	-100.4	-97.9	-115.4	-111.6
3	10.8	-10.8	-21.4	-13.4	-13.6	-57.7
4	-78.0	-91.5	-96.7	-85.4	-100.0	-101.6
5	-30.7	-45.3	-26.4	-33.6	-13.5	-32.2
6	-4.8	-2.7	-17.0	-12.5	-20.3	-53.3
7	-87.8	-92.7	-89.1	-82.9	-96.0	-102.9
8	-6.3	-2.8	-12.1	-4.2	-17.0	-27.7
9	-92.3	-90.9	-107.9	-108.0	-119.7	-118.6
10	-241.3	-266.8	-242.2	-255.0	-267.0	-248.9
11	-2.2	-2.8	-4.2	-10.8	-7.5	-40.6
12	-83.0	-82.7	-70.7	-101.0	-101.9	-83.5
13	15.9	-17.5	-12.8	-24.2	-31.1	-15.1
14	8.6	-30.5	-39.1	-24.6	-19.4	-57.5
15	-101.4	-98.9	-100.2	-89.8	-109.6	-109.2
16	-92.0	-102.9	-90.4	-75.8	-98.9	-100.8
17	32.4	52.1	53.6	31.1	46.8	-2.0
18	-17.5	-18.7	-28.5	-22.4	-45.8	-62.2
19	-119.4	-122.3	-104.3	-108.2	-129.3	-138.4
20	-24.9	-26.1	-34.9	-42.3	-22.1	-61.2
21	-98.8	-114.6	-92.5	-109.1	-111.3	-119.3
22	-5.3	-5.0	-4.4	-1.1	-24.2	-45.4
23	-76.3	-90.2	-101.6	-104.9	-121.6	-112.4
24	-23.6	10.5	-22.7	-22.9	-36.4	-57.9

C.8 Binding Free Energies on M-MOF-808 Nodes

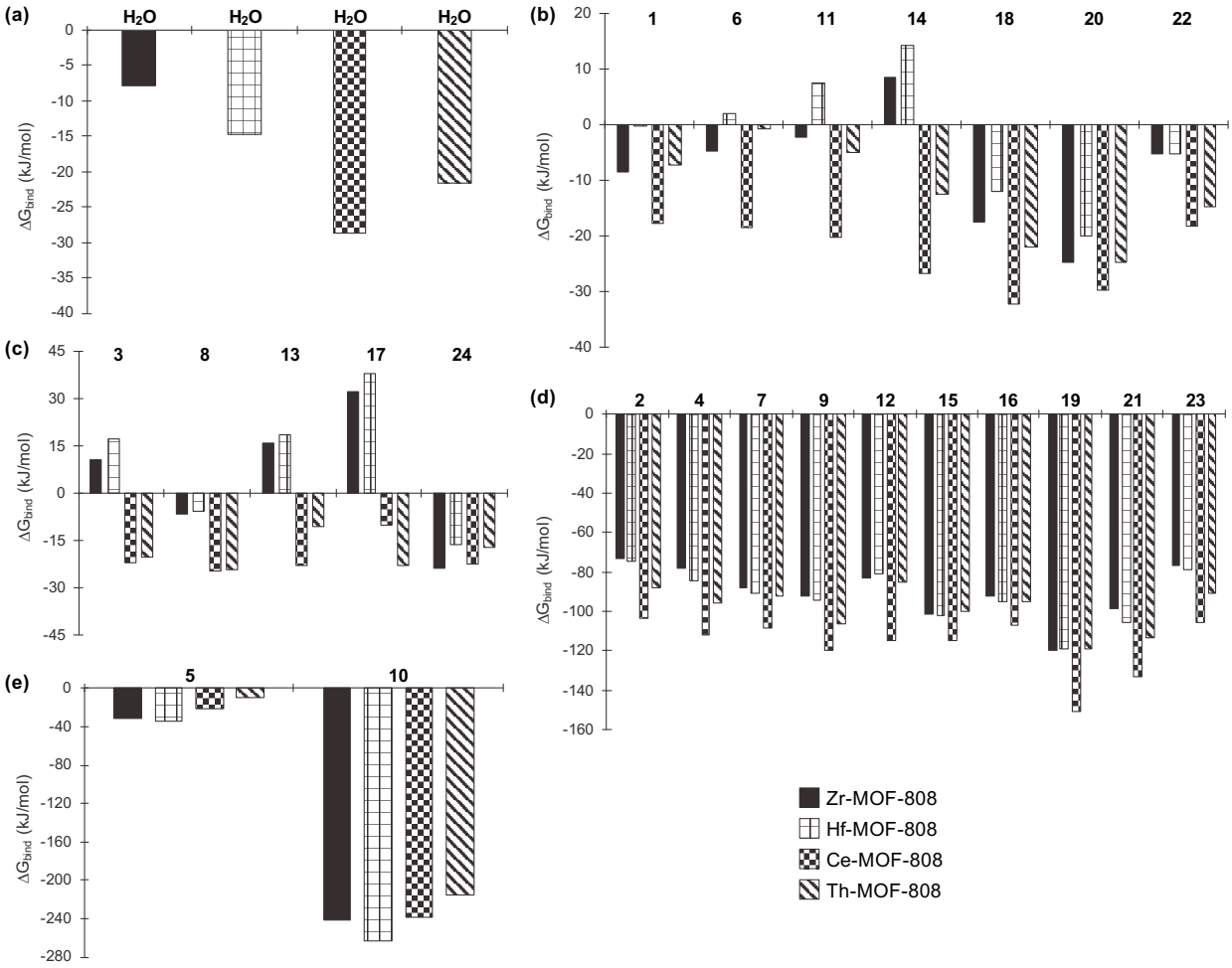


Figure C.8. Binding free energies for (a) water, (b) nerve agent and simulant molecules, (c) alcohol and thiol hydrolysis product molecules, (d) bidentate anion products, and (e) monodentate anion products bound to the nodes of M-MOF-808 (M = Zr, Hf, Ce, Th).

C.9 Effect of Metal Electronegativity on Water Binding Energy: M(IV)-MOF Nodes

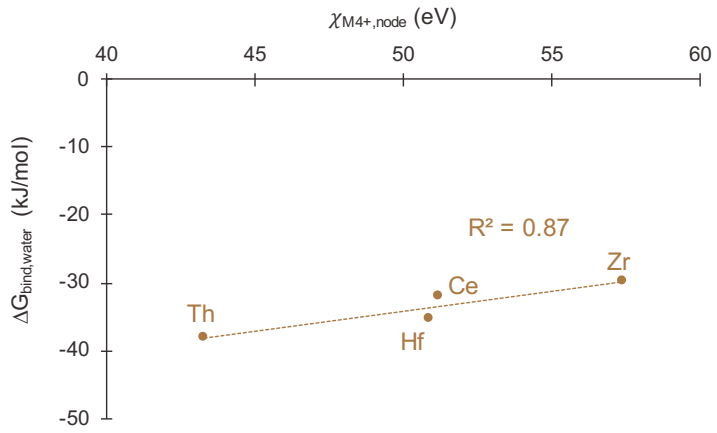


Figure C.9. The effect of the electronegativity of M^{4+} cations in the nodes of M-bi(trans)-defect UiO-66 on the binding free energy for water.

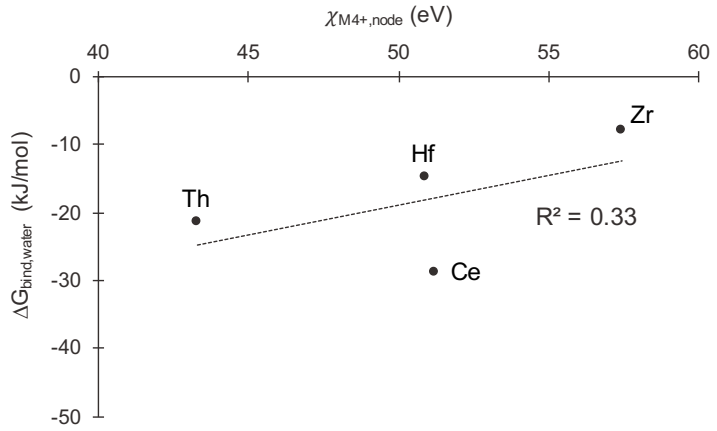


Figure C.10. The effect of the electronegativity of M^{4+} cations in the nodes of M-MOF-808 on the binding free energy for water.

C.10 Binding Free Energy Values for All Molecules: M(IV)-MOF Nodes

Table C.4. Binding free energies (in kJ/mol) for molecules bound to M-bi(trans)-defect UiO-66 node sites, corresponding to results in Figure 4.3. Negative and positive values indicate favorable and unfavorable binding, respectively.

Molecule	Zr	Hf	Ce	Th
H_2O	-29.9	-35.4	-32.0	-38.2
1	4.6	-4.6	-27.2	-19.0
2	-100.4	-103.6	-135.3	-86.8
3	-21.4	-29.5	-30.6	-37.6
4	-96.7	-99.7	-129.9	-89.0
5	-26.4	-26.6	-27.8	-21.1
6	-17.0	-25.5	-31.1	-12.7
7	-89.1	-92.4	-131.4	-71.0
8	-12.1	-17.6	-27.5	-34.5
9	-107.9	-110.6	-135.5	-92.5
10	-242.2	-242.8	-250.1	-229.1
11	-4.2	-13.2	-30.2	-19.3
12	-70.7	-75.0	-127.4	-58.9
13	-12.8	-33.5	-30.3	-36.5
14	-39.1	-40.7	-59.9	-45.9
15	-100.2	-104.4	-130.1	-101.6
16	-90.4	-96.5	-145.0	-101.5
17	53.6	42.3	2.9	33.9
18	-28.5	-33.5	-42.4	-28.9
19	-104.3	-110.0	-157.4	-87.4
20	-34.9	-41.1	-48.0	-32.5
21	-92.5	-95.2	-154.7	-95.1
22	-4.4	-10.5	-26.5	-17.4
23	-101.6	-105.1	-130.9	-84.0
24	-22.7	-29.3	-28.9	-60.2

Table C.5. Binding free energies (in kJ/mol) for molecules bound to M-MOF-808 node sites, corresponding to results in Figure C.8. Negative and positive values indicate favorable and unfavorable binding, respectively.

Molecule	Zr	Hf	Ce	Th
H_2O	-7.9	-14.8	-28.8	-21.6
1	-8.5	-0.4	-17.9	-7.4
2	-72.9	-74.6	-103.4	-88.0
3	10.8	17.4	-21.8	-20.2
4	-78.0	-84.1	-112.1	-95.5
5	-30.7	-34.6	-20.6	-9.5
6	-4.8	1.9	-18.5	-0.7
7	-87.8	-90.8	-108.3	-92.5
8	-6.3	-5.8	-24.6	-24.1
9	-92.3	-94.4	-120.0	-106.0
10	-241.3	-263.1	-239.2	-214.9
11	-2.2	7.4	-20.2	-5.0
12	-83.0	-80.8	-115.0	-85.3
13	15.9	18.7	-23.0	-10.6
14	8.6	14.2	-26.8	-12.5
15	-101.4	-102.1	-114.8	-100.0
16	-92.0	-95.0	-107.2	-95.3
17	32.4	37.9	-10.3	-22.7
18	-17.5	-12.0	-32.3	-22.0
19	-119.4	-118.9	-150.4	-119.2
20	-24.9	-20.0	-29.8	-24.8
21	-98.8	-105.3	-133.1	-113.5
22	-5.3	-5.2	-18.3	-14.8
23	-76.3	-79.0	-105.9	-90.9
24	-23.6	-16.4	-22.2	-17.0

C.11 QSAR Modeling: Molecular Descriptors

Table C.6. List of molecular descriptors used in the QSAR models for the prediction of binding free energies.

Descriptor	Notation	Unit	Range[a]
NBO partial atomic charge on binding O/S atom[b]	$q_{O/S,mol}$	a.u.	(–1.149) – (–0.148)
average NBO partial atomic charge on binding O atoms[c]	$Avq_{O,mol}$	a.u.	(–1.226) – (–1.189)
Wiberg bond index of binding O/S atom[b]	$BI_{O/S,mol}$		1.424 – 2.041
average Wiberg bond index of binding O atoms[c]	$AvBI_{O,mol}$		1.308 – 1.368
molecular volume	V_{mol}	cm^3/mol	17.289 – 205.526, 63.697 – 184.317
solvent–accessible surface area	SA_{mol}	$Å^2$	113.827 – 483.313, 220.756 – 420
molecular dipole moment	μ_{mol}	Debye	1.704 – 7.723, 4.911 – 16.278
total number of atoms in molecule	nAt_{mol}		3 – 42, 9 – 35
number of electrons in molecule	$nElec_{mol}$		10 – 146, 50 – 130
molecular ionization potential	IP_{mol}	hartree	0.261 – 0.383, 0.254 – 0.318
molecular electron affinity	EA_{mol}	hartree	(–0.112) – (–0.019), (–0.115) – (–0.057)
molecular hardness	η_{mol}	hartree	0.160 – 0.246, 0.162 – 0.217
molecular softness	S_{mol}	$hartree^{-1}$	4.057 – 6.259, 4.611 – 6.180
molecular electronegativity	χ_{mol}	hartree	0.092 – 0.147, 0.081 – 0.108
molecular electrophilicity	ω_{mol}	hartree	0.025 – 0.065, 0.018 – 0.034
number of hydrogen atoms in molecule	nH_{mol}		2 – 26, 4 – 21
number of carbon atoms in molecule	nC_{mol}		0 – 11, 1 – 9
number of nitrogen atoms in molecule	nN_{mol}		0 – 2, 0 – 2
number of oxygen atoms in molecule[b]	nO_{mol}		0 – 3
number of non-hydrogen atoms in molecule	$nNonH_{mol}$		1 – 16, 5 – 14
number of rotatable bonds in molecule	nRB_{mol}		0 – 8, 0 – 6
number of donor atoms for H-bonds in molecule[b]	$nHBd_{mol}$		0 – 2
number of acceptor atoms for H-bonds in molecule	$nHBa_{mol}$		1 – 5, 3 – 5
molecular unsaturation index[b]	UI_{mol}		0 – 1.585
molecular hydrophilicity index	HyI_{mol}		(–0.673) – 6.169, (–0.626) – 0.671
molecular weight	MW_{mol}	g/mol	18.015 – 267.368, 95.014 – 238.306
average molecular weight	AMW_{mol}	g/mol	4.866 – 7.783, 6.636 – 10.557

[a]The values in red and green correspond to neutral molecules and bidentate anions, respectively. [b]These descriptors were only computed for neutral molecules. [c]These descriptors were only computed for bidentate anions.

Below, we provide more explicit definitions and equations for some of the molecular descriptors described in the main text and shown in Table C.6.

Molecular ionization potential (IP_{mol}): refers to the amount of energy required to remove the valence electron from an isolated neutral species. Koopmans' theorem[1] may be used to relate the ionization potential to the energy of the highest occupied molecular orbital, where $IP_{mol} \approx -E_{HOMO}$.

Molecular electron affinity (EA_{mol}): refers to the difference in energy of a neutral species and its anion. Koopmans' theorem[1] may be used to relate the electron affinity to the energy of the lowest unoccupied molecular orbital, where $EA_{mol} \approx -E_{LUMO}$.

Molecular hardness (η_{mol}): measures the stability of a molecule.

Molecular softness (S_{mol}): measures the reactivity of a molecule.

Molecular electronegativity (χ_{mol}): describes the strength with which a species attracts electrons.

Molecular electrophilicity (ω_{mol}): measures the reactivity of a species towards attracting electrons from a nucleophile.

Using Koopmans' theorem, Parr and co-workers[2] derived η_{mol}, S_{mol}, χ_{mol}, and ω_{mol} as:

$$\eta_{mol} = \frac{IP_{mol} - EA_{mol}}{2} \tag{C.1}$$

$$S_{mol} = \frac{1}{\eta_{mol}} \tag{C.2}$$

$$\chi_{mol} = \frac{IP_{mol} + EA_{mol}}{2} \tag{C.3}$$

$$\omega_{mol} = \frac{{\chi_{mol}}^2}{2\eta_{mol}} \tag{C.4}$$

Molecular unsaturation index (UI_{mol}): measures the extent of unsaturated bonds in a molecule. The definition as described in the user manual for Dragon[3] is:

$$UI_{mol} = \log_2(1 + nDB_{mol} + nTB_{mol} + nAB_{mol}) \tag{C.5}$$

where nDB_{mol}, nTB_{mol}, and nAB_{mol} denote the number of double, triple, and aromatic bonds in the molecule, respectively.

Molecular hydrophilicity index (HyI_{mol}): measures the extent of hydrophilicity of a molecule.[4]

$$HyI_{mol} = \frac{(1+nHy_{mol})\log_2(1+nHy_{mol}) + nC_{mol}\left(\frac{1}{nNonH_{mol}}\log_2\frac{1}{nNonH_{mol}}\right) + \sqrt{\frac{nHy_{mol}}{(nNonH_{mol})^2}}}{\log_2(1+nNonH_{mol})} \tag{C.6}$$

where nHy_{mol} denotes the number of hydrophilic groups (number of H atoms bonded to O, S, or N atoms) in the molecule.

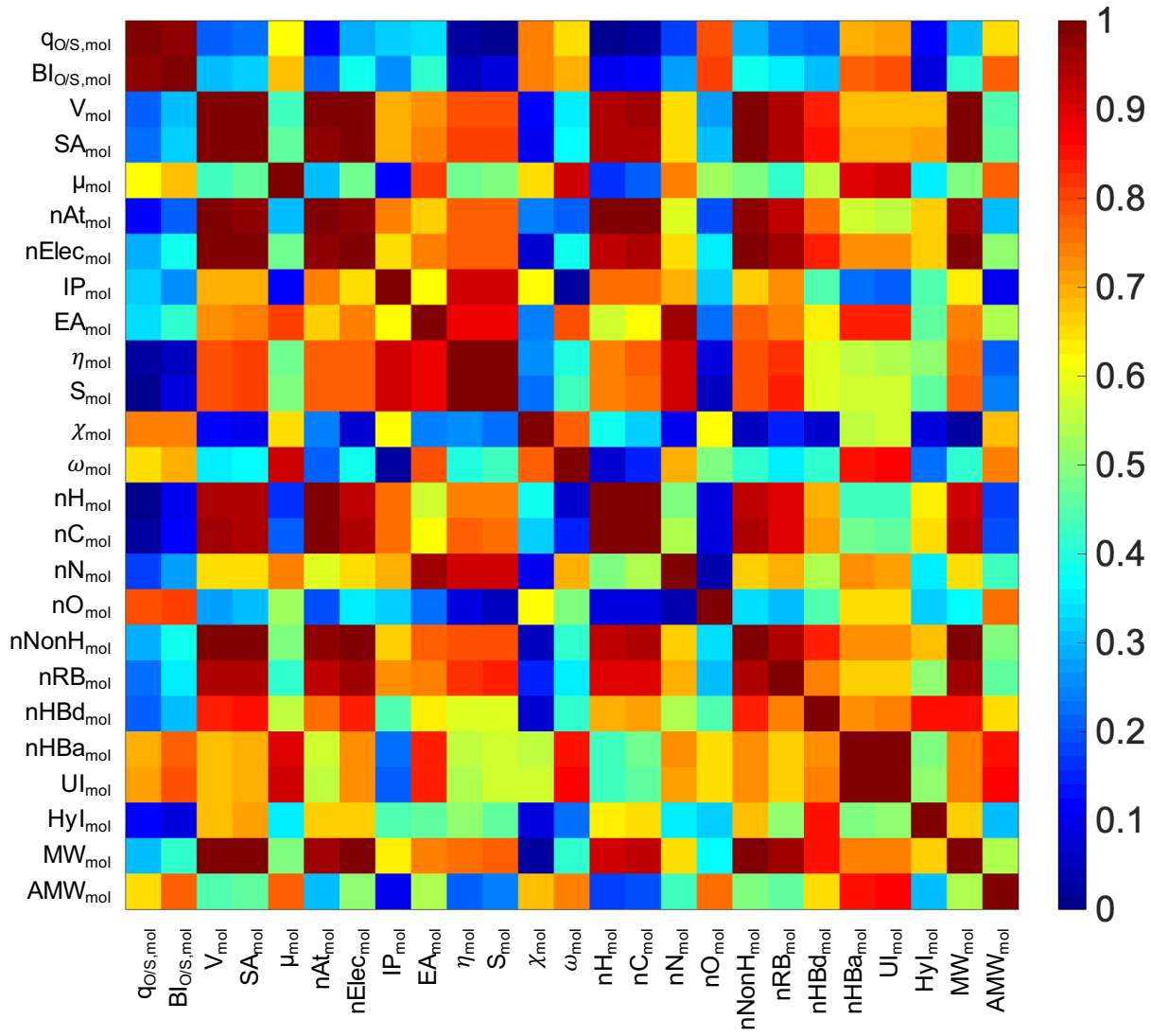


Figure C.11. Heatmap of absolute value Pearson's correlation coefficients between molecular descriptors for the neutral molecules.

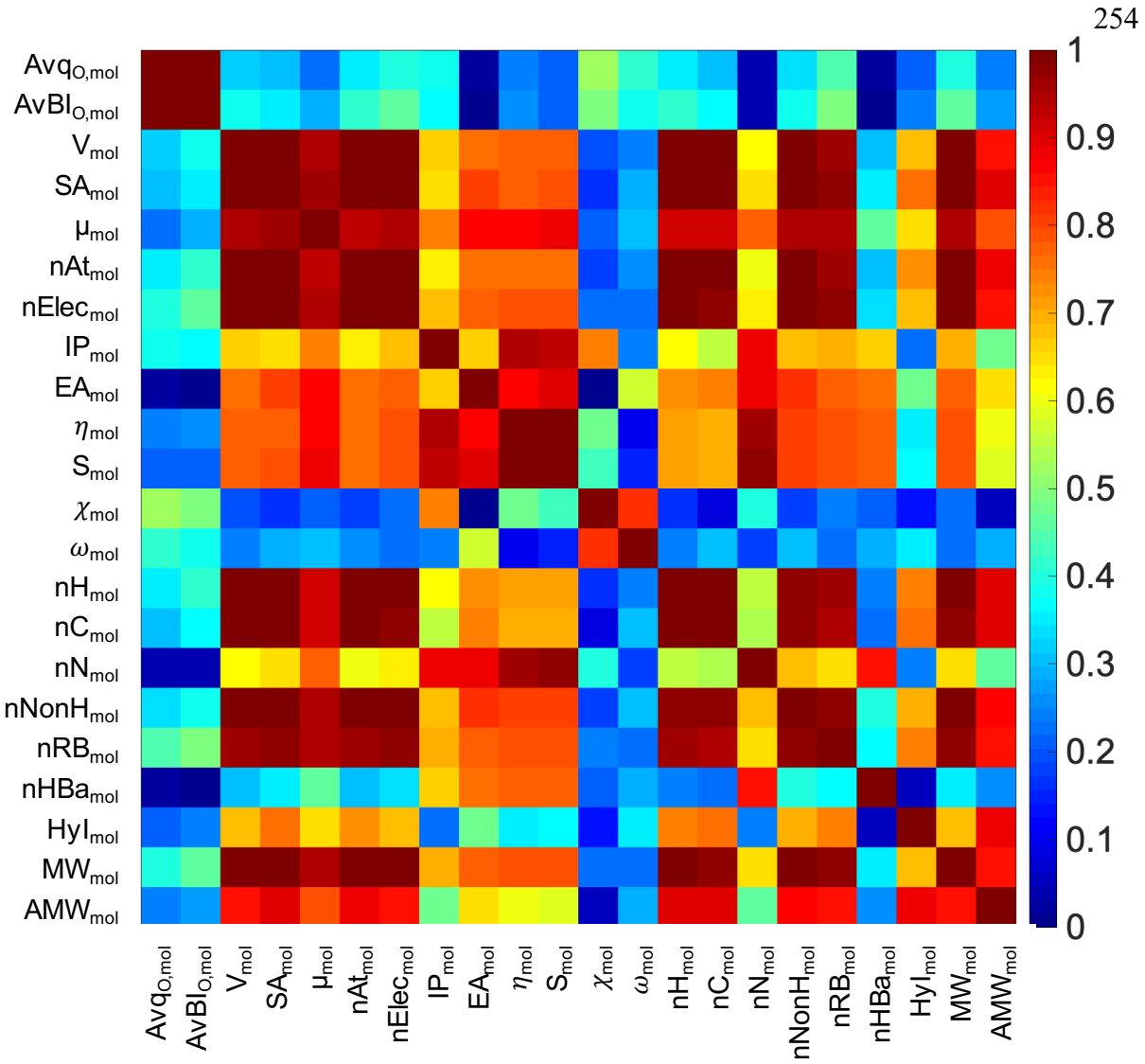


Figure C.12. Heatmap of absolute value Pearson's correlation coefficients between molecular descriptors for the bidentate anions.

Table C.7. The unscaled values of the molecular descriptors for the neutral molecules used for QSAR modeling. For descriptor units, refer to Table C.6.

Molecule	$q_{O/S,mol}$	$BI_{O/S,mol}$	V_{mol}	SA_{mol}	μ_{mol}	nAt_{mol}	$nElec_{mol}$	IP_{mol}	EA_{mol}	η_{mol}
H_2O	-0.972	1.533	17.289	113.827	2.440	3	10	0.383	-0.110	0.246
1	-1.098	1.492	125.453	342.293	7.046	21	86	0.313	-0.019	0.166
3	-0.808	1.735	40.278	184.254	2.104	9	26	0.338	-0.107	0.222
6	-1.127	1.459	102.993	291.424	3.721	18	74	0.372	-0.094	0.233
8	-0.814	1.736	63.600	209.187	2.184	12	34	0.338	-0.100	0.219
11	-1.125	1.460	145.404	347.746	3.749	27	98	0.366	-0.086	0.226
13	-0.809	1.745	89.137	270.280	2.269	21	58	0.330	-0.094	0.212
14	-1.127	1.460	205.526	483.313	2.713	42	146	0.261	-0.058	0.160
17	-0.148	2.041	144.837	353.589	1.704	29	90	0.262	-0.079	0.170
18	-1.149	1.424	152.558	377.772	7.723	28	104	0.292	-0.030	0.161
20	-1.137	1.445	155.593	388.342	6.867	29	112	0.298	-0.027	0.162
22	-1.136	1.441	91.313	277.482	5.697	16	66	0.370	-0.096	0.233
24	-0.793	1.745	28.765	146.818	2.123	6	18	0.343	-0.112	0.227

Molecule	S_{mol}	χ_{mol}	ω_{mol}	nH_{mol}	nC_{mol}	nN_{mol}	nO_{mol}	$nNonH_{mol}$	nRB_{mol}	$nHBd_{mol}$
H_2O	4.057	0.136	0.038	2	0	0	1	1	0	2
1	6.031	0.147	0.065	11	5	2	2	10	4	0
3	4.498	0.115	0.030	6	2	0	1	3	0	1
6	4.291	0.139	0.041	10	4	0	2	8	2	0
8	4.566	0.119	0.033	8	3	0	1	4	0	1
11	4.429	0.140	0.043	16	7	0	2	11	3	0
13	4.723	0.118	0.033	14	6	0	1	7	1	1
14	6.259	0.101	0.032	26	11	1	2	16	8	0
17	5.873	0.092	0.025	19	8	1	0	10	4	0
18	6.208	0.131	0.053	16	7	2	1	12	4	0
20	6.158	0.135	0.056	16	7	2	2	13	5	0
22	4.286	0.137	0.040	9	3	0	3	7	2	0
24	4.402	0.116	0.029	4	1	0	1	2	0	1

Table C.7 (continued).

Molecule	$nHBa_{mol}$	UI_{mol}	HyI_{mol}	MW_{mol}	AMW_{mol}
H_2O	1	0	6.169	18.015	6.005
1	4	1.585	-0.480	162.129	7.720
3	1	0	0.638	46.069	5.119
6	3	1	-0.473	140.094	7.783
8	1	0	0.323	60.096	5.008
11	3	1	-0.614	182.175	6.747
13	1	0	-0.088	102.177	4.866
14	3	1	-0.673	267.368	6.366
17	1	0	-0.161	161.307	5.562
18	4	1.585	-0.565	194.190	6.935
20	5	1.585	-0.523	210.189	7.248
22	3	1	-0.401	124.076	7.755
24	1	0	1.262	32.042	5.340

Table C.8. The unscaled values of the molecular descriptors for the bidentate anions used for QSAR modeling. For descriptor units, refer to Table C.6.

Molecule	$Avq_{O,mol}$	$AvBI_{O,mol}$	V_{mol}	SA_{mol}	μ_{mol}	nAt_{mol}	$nElec_{mol}$	IP_{mol}	EA_{mol}	η_{mol}
2	-1.200	1.345	107.803	317.303	8.968	20	82	0.264	-0.102	0.183
4	-1.204	1.339	78.980	253.376	7.104	14	66	0.271	-0.105	0.188
7	-1.210	1.332	94.664	293.255	8.660	18	74	0.316	-0.103	0.210
9	-1.217	1.321	63.697	220.756	4.911	9	50	0.318	-0.115	0.217
12	-1.210	1.333	138.440	354.065	9.347	27	98	0.315	-0.098	0.206
15	-1.210	1.332	89.112	279.764	7.724	15	66	0.318	-0.113	0.215
16	-1.189	1.368	184.317	420.152	16.278	35	130	0.254	-0.091	0.173
19	-1.226	1.308	152.276	386.074	15.554	28	104	0.258	-0.070	0.164
21	-1.204	1.341	152.309	387.380	15.213	29	112	0.267	-0.057	0.162
23	-1.209	1.332	66.722	249.963	6.559	12	58	0.318	-0.114	0.216

Table C.8 (continued).

Molecule	S_{mol}	χ_{mol}	ω_{mol}	nH_{mol}	nC_{mol}	nN_{mol}	$nNonH_{mol}$	nRB_{mol}	$nHBa_{mol}$	HyI_{mol}
2	5.470	0.081	0.018	11	4	1	9	3	4	-0.424
4	5.324	0.083	0.018	7	2	1	7	1	4	0.447
7	4.772	0.106	0.027	10	4	0	8	2	3	-0.473
9	4.611	0.101	0.024	4	1	0	5	0	3	0.671
12	4.847	0.108	0.029	16	7	0	11	3	3	-0.614
15	4.643	0.103	0.024	8	3	0	7	2	3	-0.401
16	5.792	0.082	0.019	21	9	1	14	6	3	-0.626
19	6.082	0.094	0.027	16	7	2	12	4	4	-0.565
21	6.180	0.105	0.034	16	7	2	13	5	5	-0.523
23	4.632	0.102	0.024	6	2	0	6	1	3	-0.307

Molecule	MW_{mol}	AMW_{mol}
2	152.110	7.606
4	124.056	8.861
7	137.095	7.616
9	95.014	10.557
12	179.176	6.636
15	123.068	8.205
16	238.306	6.809
19	191.191	6.828
21	207.190	7.144
23	109.041	9.087

The formula used to scale descriptors was as follows:

$$x_{ij}^{s} = \frac{x_{ij} - x_{j,min}}{x_{j,max} - x_{j,min}} \tag{C.7}$$

where x_{ij} and x_{ij}^{s} are the unscaled and scaled j^{th} descriptor values for molecule *i*, respectively, and $x_{j,min}$ and $x_{j,max}$ are the minimum and maximum values for the j^{th} descriptor. For all descriptors, $\min(x_{ij}^{s}) = 0$ and $\max(x_{ij}^{s}) = 1$. Scaling was performed separately for the neutral molecules descriptors and the bidentate anions descriptors.

Table C.9. The scaled values of the molecular descriptors for the neutral molecules used for QSAR modeling.

Molecule	$q_{O/S,mol}$	$BI_{O/S,mol}$	V_{mol}	SA_{mol}	μ_{mol}	nAt_{mol}	$nElec_{mol}$	IP_{mol}	EA_{mol}	η_{mol}
H_2O	0.177	0.176	0.000	0.000	0.122	0.000	0.000	1.000	0.013	1.000
1	0.051	0.110	0.575	0.618	0.887	0.462	0.559	0.423	1.000	0.070
3	0.341	0.504	0.122	0.191	0.066	0.154	0.118	0.631	0.050	0.722
6	0.023	0.057	0.455	0.481	0.335	0.385	0.471	0.912	0.189	0.845
8	0.335	0.505	0.246	0.258	0.080	0.231	0.176	0.635	0.128	0.683
11	0.024	0.060	0.681	0.633	0.340	0.615	0.647	0.862	0.279	0.761
13	0.340	0.521	0.382	0.423	0.094	0.462	0.353	0.563	0.191	0.599
14	0.022	0.060	1.000	1.000	0.168	1.000	1.000	0.000	0.575	0.000
17	1.000	1.000	0.678	0.649	0.000	0.667	0.588	0.005	0.355	0.121
18	0.000	0.000	0.719	0.714	1.000	0.641	0.691	0.254	0.880	0.015
20	0.012	0.034	0.735	0.743	0.858	0.667	0.750	0.301	0.914	0.030
22	0.013	0.028	0.393	0.443	0.663	0.333	0.412	0.898	0.164	0.848
24	0.355	0.520	0.061	0.089	0.070	0.077	0.059	0.672	0.000	0.778

Molecule	S_{mol}	χ_{mol}	ω_{mol}	nH_{mol}	nC_{mol}	nN_{mol}	nO_{mol}	$nNonH_{mol}$	nRB_{mol}	$nHBd_{mol}$
H_2O	0.000	0.807	0.322	0.000	0.000	0.000	0.333	0.000	0.000	1.000
1	0.896	1.000	1.000	0.375	0.455	1.000	0.667	0.600	0.500	0.000
3	0.200	0.433	0.133	0.167	0.182	0.000	0.333	0.133	0.000	0.500
6	0.106	0.858	0.416	0.333	0.364	0.000	0.667	0.467	0.250	0.000
8	0.231	0.503	0.195	0.250	0.273	0.000	0.333	0.200	0.000	0.500
11	0.169	0.879	0.467	0.583	0.636	0.000	0.667	0.667	0.375	0.000
13	0.302	0.476	0.203	0.500	0.545	0.000	0.333	0.400	0.125	0.500
14	1.000	0.180	0.189	1.000	1.000	0.500	0.667	1.000	1.000	0.000
17	0.825	0.000	0.000	0.708	0.727	0.500	0.000	0.600	0.500	0.000
18	0.977	0.714	0.709	0.583	0.636	1.000	0.333	0.733	0.500	0.000
20	0.954	0.794	0.788	0.583	0.636	1.000	0.667	0.800	0.625	0.000
22	0.104	0.821	0.385	0.292	0.273	0.000	1.000	0.400	0.250	0.000
24	0.156	0.436	0.119	0.083	0.091	0.000	0.333	0.067	0.000	0.500

Table C.9 (continued).

Molecule	$nHBa_{mol}$	UI_{mol}	HyI_{mol}	MW_{mol}	AMW_{mol}
H_2O	0.000	0.000	1.000	0.000	0.391
1	0.750	1.000	0.028	0.578	0.979
3	0.000	0.000	0.192	0.113	0.087
6	0.500	0.631	0.029	0.490	1.000
8	0.000	0.000	0.146	0.169	0.049
11	0.500	0.631	0.009	0.658	0.645
13	0.000	0.000	0.085	0.338	0.000
14	0.500	0.631	0.000	1.000	0.514
17	0.000	0.000	0.075	0.575	0.239
18	0.750	1.000	0.016	0.707	0.709
20	1.000	1.000	0.022	0.771	0.817
22	0.500	0.631	0.040	0.425	0.990
24	0.000	0.000	0.283	0.056	0.163

Table C.10. The scaled values of the molecular descriptors for the bidentate anions used for QSAR modeling.

Molecule	$Avq_{O,mol}$	$AvBI_{O,mol}$	V_{mol}	SA_{mol}	μ_{mol}	nAt_{mol}	$nElec_{mol}$	IP_{mol}	EA_{mol}	η_{mol}
2	0.702	0.618	0.366	0.484	0.357	0.423	0.400	0.143	0.227	0.381
4	0.577	0.513	0.127	0.164	0.193	0.192	0.200	0.252	0.176	0.472
7	0.418	0.392	0.257	0.364	0.330	0.346	0.300	0.959	0.204	0.867
9	0.228	0.209	0.000	0.000	0.000	0.000	0.000	1.000	0.000	1.000
12	0.409	0.405	0.620	0.669	0.390	0.692	0.600	0.944	0.299	0.808
15	0.433	0.397	0.211	0.296	0.247	0.231	0.200	0.995	0.046	0.973
16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.418	0.197
19	0.000	0.000	0.734	0.829	0.936	0.731	0.675	0.061	0.765	0.047
21	0.584	0.550	0.735	0.836	0.906	0.769	0.775	0.196	1.000	0.000
23	0.441	0.404	0.025	0.146	0.145	0.115	0.100	0.993	0.026	0.982

Table C.10 (continued).

Molecule	S_{mol}	χ_{mol}	ω_{mol}	nH_{mol}	nC_{mol}	nN_{mol}	$nNonH_{mol}$	nRB_{mol}	$nHBa_{mol}$	HyI_{mol}
2	0.548	0.000	0.000	0.412	0.375	0.500	0.444	0.500	0.500	0.156
4	0.455	0.072	0.024	0.176	0.125	0.500	0.222	0.167	0.500	0.827
7	0.103	0.916	0.554	0.353	0.375	0.000	0.333	0.333	0.000	0.118
9	0.000	0.747	0.361	0.000	0.000	0.000	0.000	0.000	0.000	1.000
12	0.151	1.000	0.654	0.706	0.750	0.000	0.667	0.500	0.000	0.010
15	0.020	0.790	0.406	0.235	0.250	0.000	0.222	0.333	0.000	0.174
16	0.753	0.038	0.095	1.000	1.000	0.500	1.000	1.000	0.000	0.000
19	0.938	0.476	0.552	0.706	0.750	1.000	0.778	0.667	0.500	0.047
21	1.000	0.880	1.000	0.706	0.750	1.000	0.889	0.833	1.000	0.079
23	0.013	0.766	0.383	0.118	0.125	0.000	0.111	0.167	0.000	0.246

Molecule	MW_{mol}	AMW_{mol}
2	0.398	0.247
4	0.203	0.567
7	0.294	0.250
9	0.000	1.000
12	0.587	0.000
15	0.196	0.400
16	1.000	0.044
19	0.671	0.049
21	0.783	0.130
23	0.098	0.625

C.12 QSAR Modeling: Node Descriptors

Table C.11. List of node descriptors used in the QSAR models for the prediction of binding free energies.

Descriptor	Notation	Unit	Range[a]
distance between 2 binding site metal atoms	$d_{M–M,node}$	Å	3.584 – 3.997, 3.559 – 4.017
NBO partial atomic charge on binding site M atom[b]	$q_{M,node}$	a.u.	1.921 – 2.341
average NBO partial atomic charge on binding site M atoms[c]	$Avq_{M,node}$	a.u.	1.990 – 2.317
Wiberg bond index of binding site M atom[b]	$BI_{M,node}$		3.212 – 4.092
average Wiberg bond index of binding site M atoms[c]	$AvBI_{M,node}$		3.235 – 3.903
NBO population of valence orbitals of binding site M atom[b]	$ValPop_{M,node}$		1.590 – 2.072
average NBO population of valence orbitals of binding site M atoms[c]	$AvValPop_{M,node}$		1.622 – 2.000
experimental ionization potential of M^{4+} atom	$IP_{M4+,node}$	eV	58.000 – 80.348, 58.000 – 80.348
experimental electron affinity of M^{4+} atom	$EA_{M4+,node}$	eV	28.648 – 36.906, 28.648 – 36.906
experimental hardness of M^{4+} atom	$\eta_{M4+,node}$	eV	14.322 – 22.965, 14.322 – 22.965
experimental softness of M^{4+} atom	$S_{M4+,node}$	eV^{-1}	0.044 – 0.070, 0.044 – 0.070
experimental electronegativity of M^{4+} atom	$\chi_{M4+,node}$	eV	43.324 – 57.383, 43.324 – 57.383
experimental electrophilicity of M^{4+} atom	$\omega_{M4+,node}$	eV	63.947 – 91.618, 63.947 – 91.618
atomic number of binding site M atom	$Z_{M,node}$		40 – 90, 40 – 90
atomic weight of binding site M atom	$AtWt_{M,node}$	g/mol	91.224 – 232.040, 91.224 – 232.040
period in the periodic table of binding site M atom	$Pd_{M,node}$		5 – 7, 5 – 7
covalent radius of binding site M atom	$CovR_{M,node}$	Å	1.75 – 2.06, 1.75 – 2.06
total number of linkers per node	$nLink_{node}$		6 – 11, 6 – 11
linker bite angle	$Bite_{node}$	degree	138.194 – 157.337, 138.127 – 156.311
number of electrons in node	$nElec_{node}$		790 – 1266, 780 – 1256
NBO partial atomic charge on binding site μ_3O(H) atom	$q_{\mu 3O(H),node}$	a.u.	(–0.974) – (–0.867), (–0.985) – (–0.888)
NBO partial atomic charge on binding site μ_3(O)H atom	$q_{\mu 3(O)H,node}$	a.u.	0.514 – 0.529, 0.509 – 0.520
NBO partial atomic charge on binding site μ_3O atom	$q_{\mu 3O,node}$	a.u.	(–1.137) – (–0.916), (–1.163) – (–0.930)
NBO partial atomic charge on binding site terminal O(H) atom[b]	$q_{tO(H),node}$	a.u.	(–1.081) – (–0.976)
NBO partial atomic charge on binding site terminal (O)H atom[b]	$q_{t(O)H,node}$	a.u.	0.500 – 0.508
bond length of binding site terminal M–OH bond[b]	$BL_{tM–OH,node}$	Å	2.030 – 2.210
Wiberg bond order of binding site terminal M–OH bond[b]	$BO_{tM–OH,node}$		0.700 – 0.923

[a]The values in red correspond to node binding sites with the terminal –OH_2 group removed (*node–noOH₂*) and values in green correspond to node binding sites with the terminal –OH_2 and –OH groups removed (*node–noOH₂,OH*). [b]These descriptors were only computed for *node–noOH₂* structures. [c]These descriptors were only computed for *node–noOH₂,OH* structures.

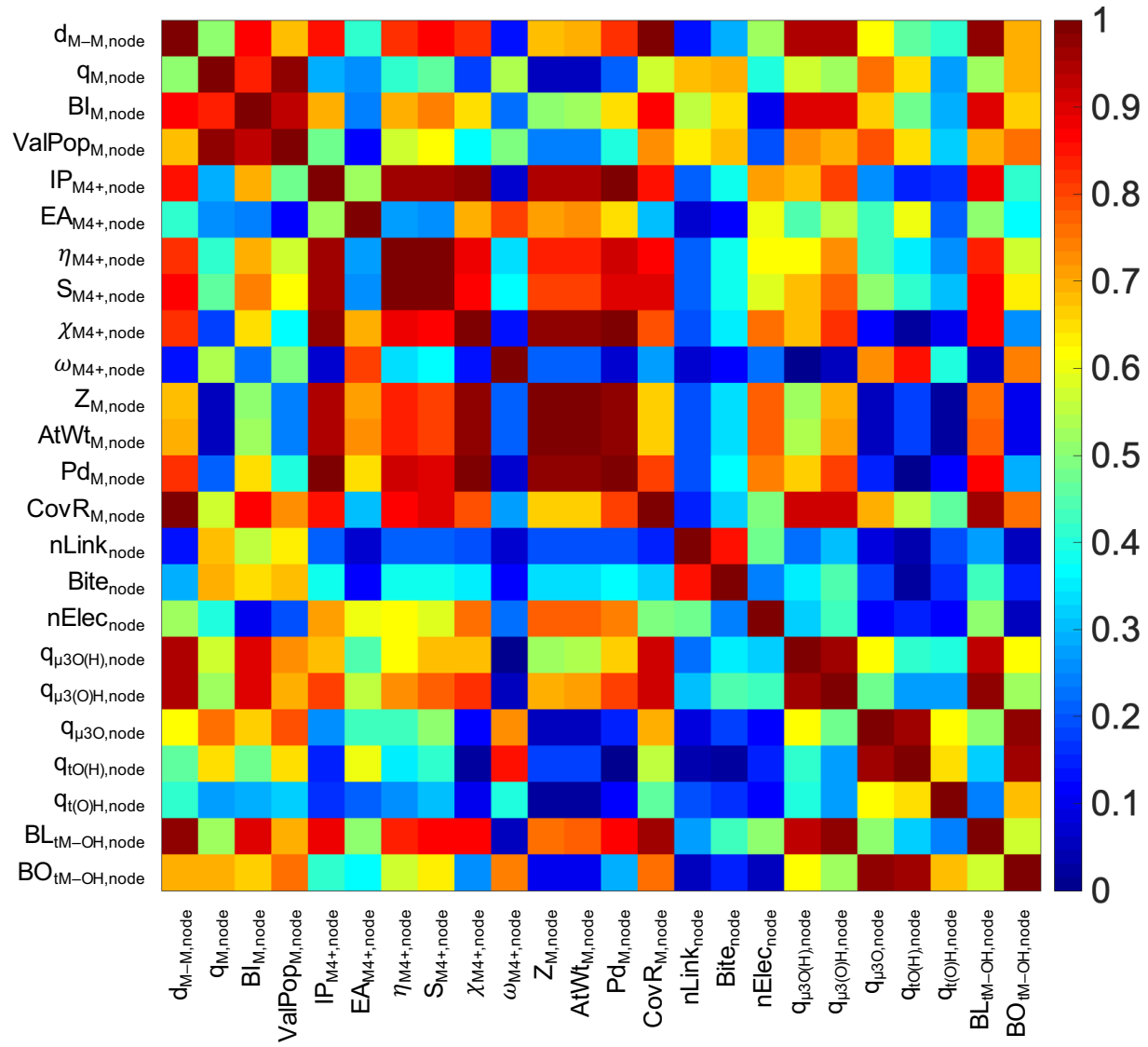


Figure C.13. Heatmap of absolute value Pearson's correlation coefficients between node descriptors for the *node–noOH₂* binding sites.

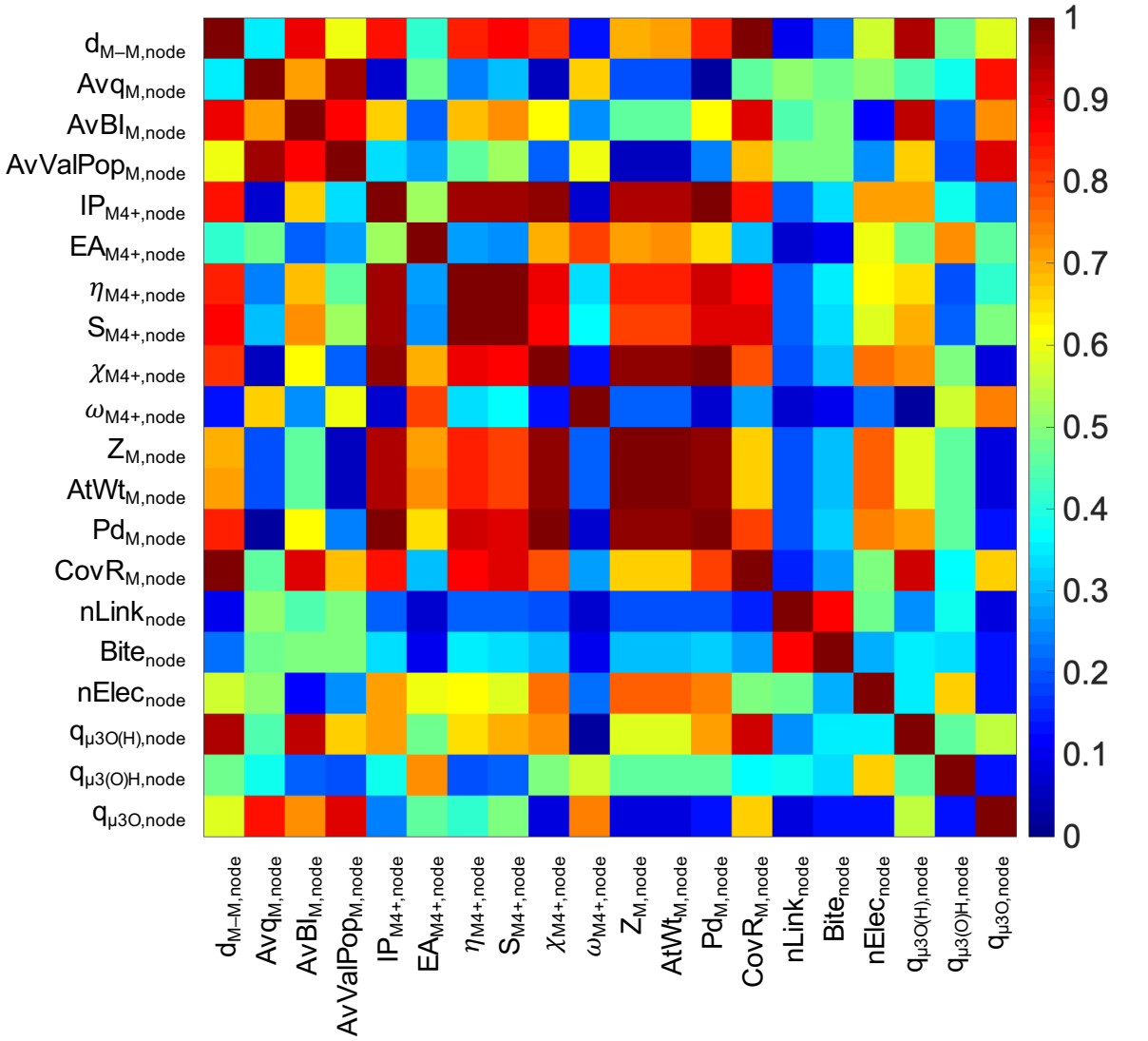


Figure C.14. Heatmap of absolute value Pearson's correlation coefficients between node descriptors for the *node–noOH₂,OH* binding sites.

Table C.12. The unscaled values of the node descriptors for the *node–noOH₂* binding sites used for QSAR modeling. For descriptor units, refer to Table C.11.

node–noOH₂ **Site**	$d_{M-M,node}$	$q_{M,node}$	$BI_{M,node}$	$ValPop_{M,node}$	$IP_{M4+,node}$	$EA_{M4+,node}$	$\eta_{M4+,node}$	$S_{M4+,node}$	$\chi_{M4+,node}$
Zr-MOF-808	3.600	2.056	3.667	1.876	80.348	34.418	22.965	0.044	57.383
Zr-NU-1000 (large pore)	3.611	2.192	3.432	1.735	80.348	34.418	22.965	0.044	57.383
Zr-bi(trans)-defect UiO-66	3.617	2.233	3.378	1.700	80.348	34.418	22.965	0.044	57.383
Zr-mono-defect UiO-66	3.606	2.244	3.360	1.689	80.348	34.418	22.965	0.044	57.383
Zr-bi(cis)-defect UiO-66	3.623	2.203	3.414	1.725	80.348	34.418	22.965	0.044	57.383
Zr-NU-1000 (c pore)	3.610	2.196	3.426	1.731	80.348	34.418	22.965	0.044	57.383
Hf-bi(trans)-defect UiO-66	3.595	2.341	3.212	1.590	68.370	33.370	17.500	0.057	50.870
Ce-bi(trans)-defect UiO-66	3.898	2.068	3.768	1.918	65.550	36.906	14.322	0.070	51.228
Th-bi(trans)-defect UiO-66	3.990	2.224	3.781	1.781	58.000	28.648	14.676	0.068	43.324
Hf-MOF-808	3.584	2.158	3.517	1.777	68.370	33.370	17.500	0.057	50.870
Ce-MOF-808	3.896	1.921	4.033	2.072	65.550	36.906	14.322	0.070	51.228
Th-MOF-808	3.997	2.065	4.092	1.949	58.000	28.648	14.676	0.068	43.324

Table C.12 (continued).

***node–noOH₂* Site**	$\omega_{M4+,node}$	$Z_{M,node}$	$AtWt_{M,node}$	$Pd_{M,node}$	$CovR_{M,node}$	$nLink_{node}$	$Bite_{node}$	$nElec_{node}$	$q_{\mu 3O(H),node}$
Zr-MOF-808	71.693	40	91.224	5	1.75	6	155.759	790	-0.933
Zr-NU-1000 (large pore)	71.693	40	91.224	5	1.75	8	149.637	878	-0.942
Zr-bi(trans)-defect UiO-66	71.693	40	91.224	5	1.75	10	143.714	966	-0.944
Zr-mono-defect UiO-66	71.693	40	91.224	5	1.75	11	142.567	1010	-0.944
Zr-bi(cis)-defect UiO-66	71.693	40	91.224	5	1.75	10	142.463	966	-0.942
Zr-NU-1000 (c pore)	71.693	40	91.224	5	1.75	8	138.194	878	-0.943
Hf-bi(trans)-defect UiO-66	73.936	72	178.490	6	1.75	10	143.415	1158	-0.974
Ce-bi(trans)-defect UiO-66	91.618	58	140.120	6	2.04	10	144.622	1074	-0.905
Th-bi(trans)-defect UiO-66	63.947	90	232.040	7	2.06	10	144.456	1266	-0.882
Hf-MOF-808	73.936	72	178.490	6	1.75	6	155.763	982	-0.961
Ce-MOF-808	91.618	58	140.120	6	2.04	6	156.912	898	-0.902
Th-MOF-808	63.947	90	232.040	7	2.06	6	157.337	1090	-0.867

Table C.12 (continued).

***node–noOH*$_2$ Site**	$q_{\mu 3(O)H,node}$	$q_{\mu 3O,node}$	$q_{tO(H),node}$	$q_{t(O)H,node}$	$BL_{tM-OH,node}$	$BO_{tM-OH,node}$
Zr-MOF-808	0.525	-1.053	-1.055	0.503	2.050	0.743
Zr-NU-1000 (large pore)	0.526	-1.064	-1.049	0.504	2.030	0.754
Zr-bi(trans)-defect UiO-66	0.527	-1.061	-1.052	0.502	2.038	0.743
Zr-mono-defect UiO-66	0.528	-1.062	-1.052	0.501	2.043	0.737
Zr-bi(cis)-defect UiO-66	0.527	-1.063	-1.051	0.501	2.044	0.739
Zr-NU-1000 (c pore)	0.527	-1.062	-1.049	0.503	2.035	0.751
Hf-bi(trans)-defect UiO-66	0.529	-1.137	-1.077	0.500	2.042	0.700
Ce-bi(trans)-defect UiO-66	0.523	-0.916	-0.976	0.508	2.123	0.923
Th-bi(trans)-defect UiO-66	0.515	-1.052	-1.058	0.506	2.170	0.785
Hf-MOF-808	0.527	-1.129	-1.081	0.501	2.053	0.701
Ce-MOF-808	0.519	-0.917	-0.991	0.503	2.148	0.891
Th-MOF-808	0.514	-1.040	-1.061	0.500	2.210	0.775

Table C.13. The unscaled values of the node descriptors for the *node–noOH$_2$,OH* binding sites used for QSAR modeling. For descriptor units, refer to Table C.11.

***node–noOH$_2$,OH* Site**	$d_{M-M,node}$	$Avq_{M,node}$	$AvBI_{M,node}$	$AvValPop_{M,node}$	$IP_{M4+,node}$	$EA_{M4+,node}$	$\eta_{M4+,node}$	$S_{M4+,node}$	$\chi_{M4+,node}$
Zr-MOF-808	3.578	2.117	3.550	1.823	80.348	34.418	22.965	0.044	57.383
Zr-NU-1000 (large pore)	3.567	2.193	3.418	1.744	80.348	34.418	22.965	0.044	57.383
Zr-bi(trans)-defect UiO-66	3.597	2.214	3.392	1.725	80.348	34.418	22.965	0.044	57.383
Zr-mono-defect UiO-66	3.588	2.211	3.396	1.729	80.348	34.418	22.965	0.044	57.383
Zr-bi(cis)-defect UiO-66	3.591	2.182	3.437	1.755	80.348	34.418	22.965	0.044	57.383
Zr-NU-1000 (c pore)	3.563	2.181	3.437	1.756	80.348	34.418	22.965	0.044	57.383
Hf-bi(trans)-defect UiO-66	3.575	2.317	3.235	1.622	68.370	33.370	17.500	0.057	50.870
Ce-bi(trans)-defect UiO-66	3.911	2.097	3.704	1.889	65.550	36.906	14.322	0.070	51.228
Th-bi(trans)-defect UiO-66	4.017	2.254	3.705	1.755	58.000	28.648	14.676	0.068	43.324
Hf-MOF-808	3.559	2.218	3.400	1.723	68.370	33.370	17.500	0.057	50.870
Ce-MOF-808	3.884	1.990	3.899	2.000	65.550	36.906	14.322	0.070	51.228
Th-MOF-808	3.990	2.152	3.903	1.854	58.000	28.648	14.676	0.068	43.324

Table C.13 (continued).

***node–noOH₂,OH* Site**	$\omega_{M4+,node}$	$Z_{M,node}$	$AtWt_{M,node}$	$Pd_{M,node}$	$CovR_{M,node}$	$nLink_{node}$	$Bite_{node}$	$nElec_{node}$	$q_{\mu 3O(H),node}$
Zr-MOF-808	71.693	40	91.224	5	1.75	6	155.406	780	-0.947
Zr-NU-1000 (large pore)	71.693	40	91.224	5	1.75	8	149.531	868	-0.965
Zr-bi(trans)-defect UiO-66	71.693	40	91.224	5	1.75	10	143.341	956	-0.962
Zr-mono-defect UiO-66	71.693	40	91.224	5	1.75	11	142.088	1000	-0.961
Zr-bi(cis)-defect UiO-66	71.693	40	91.224	5	1.75	10	142.144	956	-0.959
Zr-NU-1000 (c pore)	71.693	40	91.224	5	1.75	8	138.127	868	-0.964
Hf-bi(trans)-defect UiO-66	73.936	72	178.490	6	1.75	10	143.053	1148	-0.985
Ce-bi(trans)-defect UiO-66	91.618	58	140.120	6	2.04	10	143.581	1064	-0.927
Th-bi(trans)-defect UiO-66	63.947	90	232.040	7	2.06	10	143.308	1256	-0.902
Hf-MOF-808	73.936	72	178.490	6	1.75	6	155.416	972	-0.973
Ce-MOF-808	91.618	58	140.120	6	2.04	6	155.935	888	-0.922
Th-MOF-808	63.947	90	232.040	7	2.06	6	156.311	1080	-0.888

Table C.13 (continued).

***node–noOH$_2$,OH* Site**	$q_{\mu 3(O)H,node}$	$q_{\mu 3O,node}$
Zr-MOF-808	0.520	-1.076
Zr-NU-1000 (large pore)	0.516	-1.088
Zr-bi(trans)-defect UiO-66	0.514	-1.084
Zr-mono-defect UiO-66	0.515	-1.084
Zr-bi(cis)-defect UiO-66	0.515	-1.085
Zr-NU-1000 (c pore)	0.515	-1.087
Hf-bi(trans)-defect UiO-66	0.514	-1.163
Ce-bi(trans)-defect UiO-66	0.518	-0.935
Th-bi(trans)-defect UiO-66	0.509	-1.079
Hf-MOF-808	0.520	-1.155
Ce-MOF-808	0.517	-0.930
Th-MOF-808	0.511	-1.070

Table C.14. The scaled values of the node descriptors for the *node–noOH₂* binding sites used for QSAR modeling.

***node–noOH₂* Site**	$d_{M-M,node}$	$q_{M,node}$	$BI_{M,node}$	$ValPop_{M,node}$	$IP_{M4+,node}$	$EA_{M4+,node}$	$\eta_{M4+,node}$	$S_{M4+,node}$	$\chi_{M4+,node}$
Zr-MOF-808	0.039	0.320	0.517	0.594	1.000	0.699	1.000	0.000	1.000
Zr-NU-1000 (large pore)	0.064	0.646	0.250	0.301	1.000	0.699	1.000	0.000	1.000
Zr-bi(trans)-defect UiO-66	0.079	0.742	0.189	0.228	1.000	0.699	1.000	0.000	1.000
Zr-mono-defect UiO-66	0.052	0.768	0.168	0.206	1.000	0.699	1.000	0.000	1.000
Zr-bi(cis)-defect UiO-66	0.094	0.672	0.230	0.280	1.000	0.699	1.000	0.000	1.000
Zr-NU-1000 (c pore)	0.062	0.655	0.243	0.293	1.000	0.699	1.000	0.000	1.000
Hf-bi(trans)-defect UiO-66	0.025	1.000	0.000	0.000	0.464	0.572	0.368	0.517	0.537
Ce-bi(trans)-defect UiO-66	0.758	0.350	0.632	0.681	0.338	1.000	0.000	1.000	0.562
Th-bi(trans)-defect UiO-66	0.983	0.723	0.647	0.397	0.000	0.000	0.041	0.936	0.000
Hf-MOF-808	0.000	0.564	0.346	0.389	0.464	0.572	0.368	0.517	0.537
Ce-MOF-808	0.755	0.000	0.934	1.000	0.338	1.000	0.000	1.000	0.562
Th-MOF-808	1.000	0.344	1.000	0.746	0.000	0.000	0.041	0.936	0.000

Table C.14 (continued).

***node–noOH₂* Site**	$\omega_{M4+,node}$	$Z_{M,node}$	$AtWt_{M,node}$	$Pd_{M,node}$	$CovR_{M,node}$	$nLink_{node}$	$Bite_{node}$	$nElec_{node}$	$q_{\mu 3O(H),node}$
Zr-MOF-808	0.280	0.000	0.000	0.000	0.000	0.000	0.918	0.000	0.390
Zr-NU-1000 (large pore)	0.280	0.000	0.000	0.000	0.000	0.400	0.598	0.185	0.305
Zr-bi(trans)-defect UiO-66	0.280	0.000	0.000	0.000	0.000	0.800	0.288	0.370	0.284
Zr-mono-defect UiO-66	0.280	0.000	0.000	0.000	0.000	1.000	0.228	0.462	0.281
Zr-bi(cis)-defect UiO-66	0.280	0.000	0.000	0.000	0.000	0.800	0.223	0.370	0.304
Zr-NU-1000 (c pore)	0.280	0.000	0.000	0.000	0.000	0.400	0.000	0.185	0.297
Hf-bi(trans)-defect UiO-66	0.361	0.640	0.620	0.500	0.000	0.800	0.273	0.773	0.000
Ce-bi(trans)-defect UiO-66	1.000	0.360	0.347	0.500	0.935	0.800	0.336	0.597	0.644
Th-bi(trans)-defect UiO-66	0.000	1.000	1.000	1.000	1.000	0.800	0.327	1.000	0.868
Hf-MOF-808	0.361	0.640	0.620	0.500	0.000	0.000	0.918	0.403	0.123
Ce-MOF-808	1.000	0.360	0.347	0.500	0.935	0.000	0.978	0.227	0.678
Th-MOF-808	0.000	1.000	1.000	1.000	1.000	0.000	1.000	0.630	1.000

Table C.14 (continued).

***node–noOH*$_2$ Site**	$q_{\mu 3(O)H,node}$	$q_{\mu 3O,node}$	$q_{tO(H),node}$	$q_{t(O)H,node}$	$BL_{tM-OH,node}$	$BO_{tM-OH,node}$
Zr-MOF-808	0.766	0.381	0.245	0.429	0.115	0.195
Zr-NU-1000 (large pore)	0.834	0.330	0.296	0.487	0.000	0.245
Zr-bi(trans)-defect UiO-66	0.886	0.343	0.275	0.342	0.046	0.194
Zr-mono-defect UiO-66	0.959	0.339	0.275	0.202	0.076	0.169
Zr-bi(cis)-defect UiO-66	0.868	0.333	0.285	0.169	0.078	0.175
Zr-NU-1000 (c pore)	0.897	0.336	0.305	0.365	0.032	0.230
Hf-bi(trans)-defect UiO-66	1.000	0.000	0.033	0.000	0.070	0.000
Ce-bi(trans)-defect UiO-66	0.587	1.000	1.000	1.000	0.517	1.000
Th-bi(trans)-defect UiO-66	0.065	0.384	0.213	0.766	0.777	0.380
Hf-MOF-808	0.899	0.036	0.000	0.184	0.128	0.008
Ce-MOF-808	0.305	0.994	0.857	0.377	0.655	0.857
Th-MOF-808	0.000	0.438	0.185	0.046	1.000	0.339

Table C.15. The scaled values of the node descriptors for the *node–noOH$_2$,OH* binding sites used for QSAR modeling.

***node–noOH$_2$,OH* Site**	$d_{M-M,node}$	$Avq_{M,node}$	$AvBI_{M,node}$	$AvValPop_{M,node}$	$IP_{M4+,node}$	$EA_{M4+,node}$	$\eta_{M4+,node}$	$S_{M4+,node}$	$\chi_{M4+,node}$
Zr-MOF-808	0.041	0.389	0.472	0.531	1.000	0.699	1.000	0.000	1.000
Zr-NU-1000 (large pore)	0.017	0.620	0.274	0.321	1.000	0.699	1.000	0.000	1.000
Zr-bi(trans)-defect UiO-66	0.082	0.686	0.235	0.272	1.000	0.699	1.000	0.000	1.000
Zr-mono-defect UiO-66	0.064	0.676	0.241	0.282	1.000	0.699	1.000	0.000	1.000
Zr-bi(cis)-defect UiO-66	0.069	0.588	0.302	0.351	1.000	0.699	1.000	0.000	1.000
Zr-NU-1000 (c pore)	0.007	0.584	0.302	0.353	1.000	0.699	1.000	0.000	1.000
Hf-bi(trans)-defect UiO-66	0.034	1.000	0.000	0.000	0.464	0.572	0.368	0.517	0.537
Ce-bi(trans)-defect UiO-66	0.768	0.328	0.701	0.704	0.338	1.000	0.000	1.000	0.562
Th-bi(trans)-defect UiO-66	1.000	0.808	0.703	0.350	0.000	0.000	0.041	0.936	0.000
Hf-MOF-808	0.000	0.698	0.247	0.267	0.464	0.572	0.368	0.517	0.537
Ce-MOF-808	0.710	0.000	0.993	1.000	0.338	1.000	0.000	1.000	0.562
Th-MOF-808	0.941	0.495	1.000	0.612	0.000	0.000	0.041	0.936	0.000

Table C.15 (continued).

***node–noOH$_2$,OH* Site**	$\omega_{M4+,node}$	$Z_{M,node}$	$AtWt_{M,node}$	$Pd_{M,node}$	$CovR_{M,node}$	$nLink_{node}$	$Bite_{node}$	$nElec_{node}$	$q_{\mu 3O(H),node}$
Zr-MOF-808	0.280	0.000	0.000	0.000	0.000	0.000	0.950	0.000	0.398
Zr-NU-1000 (large pore)	0.280	0.000	0.000	0.000	0.000	0.400	0.627	0.185	0.212
Zr-bi(trans)-defect UiO-66	0.280	0.000	0.000	0.000	0.000	0.800	0.287	0.370	0.240
Zr-mono-defect UiO-66	0.280	0.000	0.000	0.000	0.000	1.000	0.218	0.462	0.247
Zr-bi(cis)-defect UiO-66	0.280	0.000	0.000	0.000	0.000	0.800	0.221	0.370	0.271
Zr-NU-1000 (c pore)	0.280	0.000	0.000	0.000	0.000	0.400	0.000	0.185	0.220
Hf-bi(trans)-defect UiO-66	0.361	0.640	0.620	0.500	0.000	0.800	0.271	0.773	0.000
Ce-bi(trans)-defect UiO-66	1.000	0.360	0.347	0.500	0.935	0.800	0.300	0.597	0.602
Th-bi(trans)-defect UiO-66	0.000	1.000	1.000	1.000	1.000	0.800	0.285	1.000	0.856
Hf-MOF-808	0.361	0.640	0.620	0.500	0.000	0.000	0.951	0.403	0.128
Ce-MOF-808	1.000	0.360	0.347	0.500	0.935	0.000	0.979	0.227	0.648
Th-MOF-808	0.000	1.000	1.000	1.000	1.000	0.000	1.000	0.630	1.000

Table C.15 (continued).

***node–noOH$_2$,OH* Site**	$q_{\mu 3(O)H,node}$	$q_{\mu 3O,node}$
Zr-MOF-808	0.930	0.376
Zr-NU-1000 (large pore)	0.615	0.323
Zr-bi(trans)-defect UiO-66	0.432	0.339
Zr-mono-defect UiO-66	0.509	0.338
Zr-bi(cis)-defect UiO-66	0.489	0.337
Zr-NU-1000 (c pore)	0.557	0.328
Hf-bi(trans)-defect UiO-66	0.461	0.000
Ce-bi(trans)-defect UiO-66	0.821	0.978
Th-bi(trans)-defect UiO-66	0.000	0.361
Hf-MOF-808	1.000	0.036
Ce-MOF-808	0.688	1.000
Th-MOF-808	0.174	0.400

C.13 QSAR Modeling: Distributions of Binding Free Energies

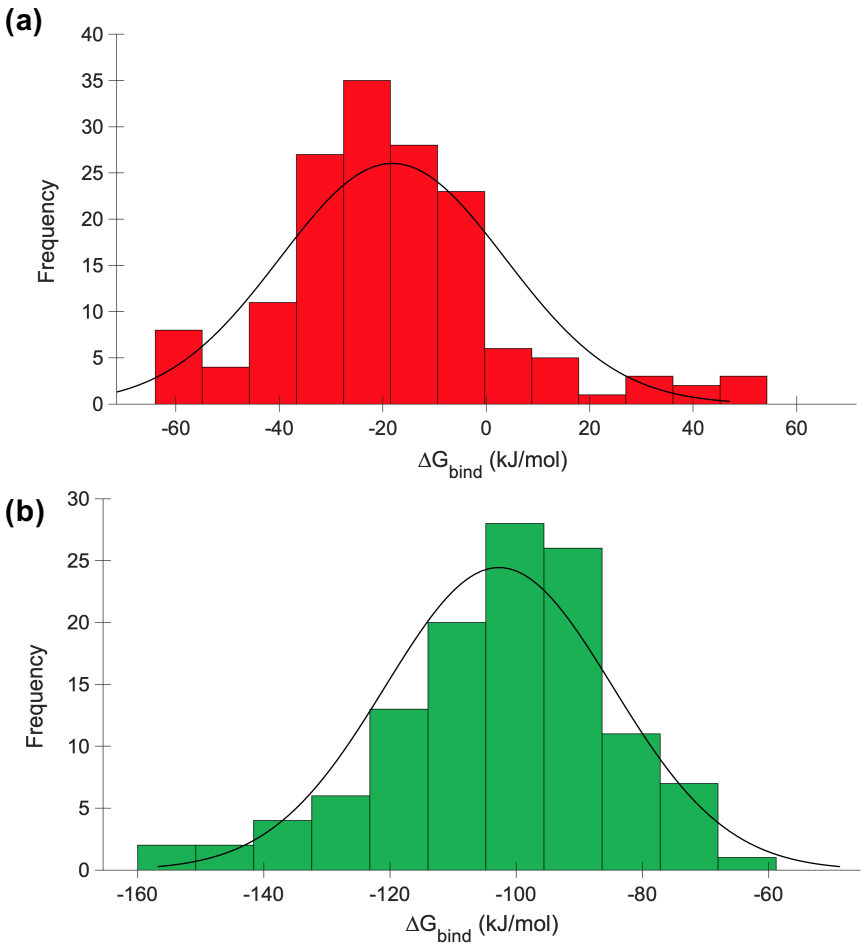


Figure C.15. Distribution of binding free energies for (a) the neutral molecules bound to *node–noOH₂* sites and (b) the bidentate anions bound to *node–noOH₂,OH* sites.

C.14 QSAR Modeling: Most Important Individual Descriptors

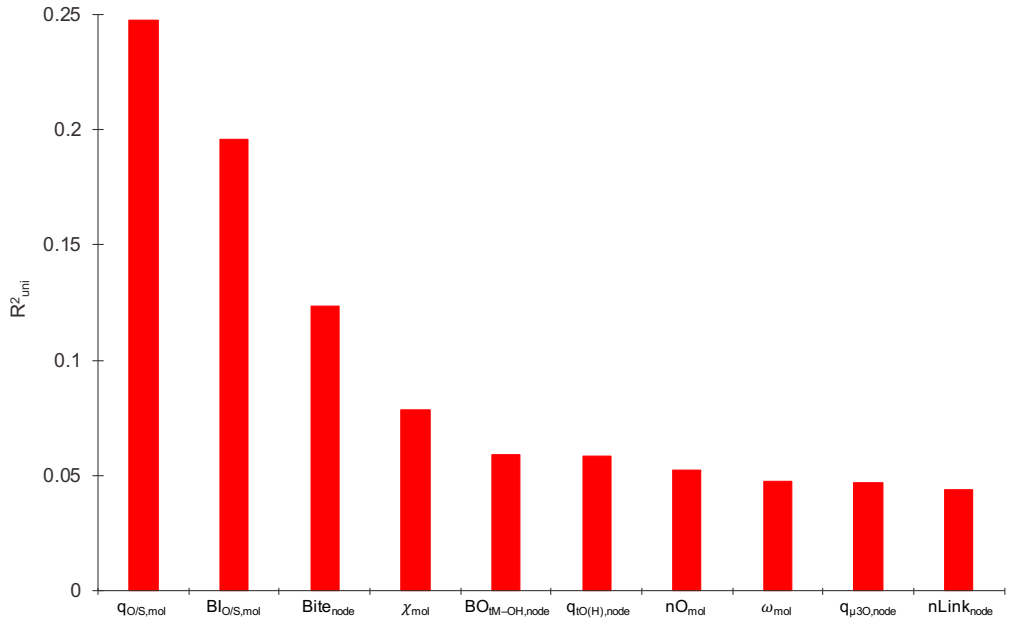


Figure C.16. The 10 most important descriptors for describing the binding free energies of neutral molecules to *node–noOH₂* sites, as identified by uniparametric coefficients of determination (R^2_{uni}) for simple linear regression using individual descriptors.

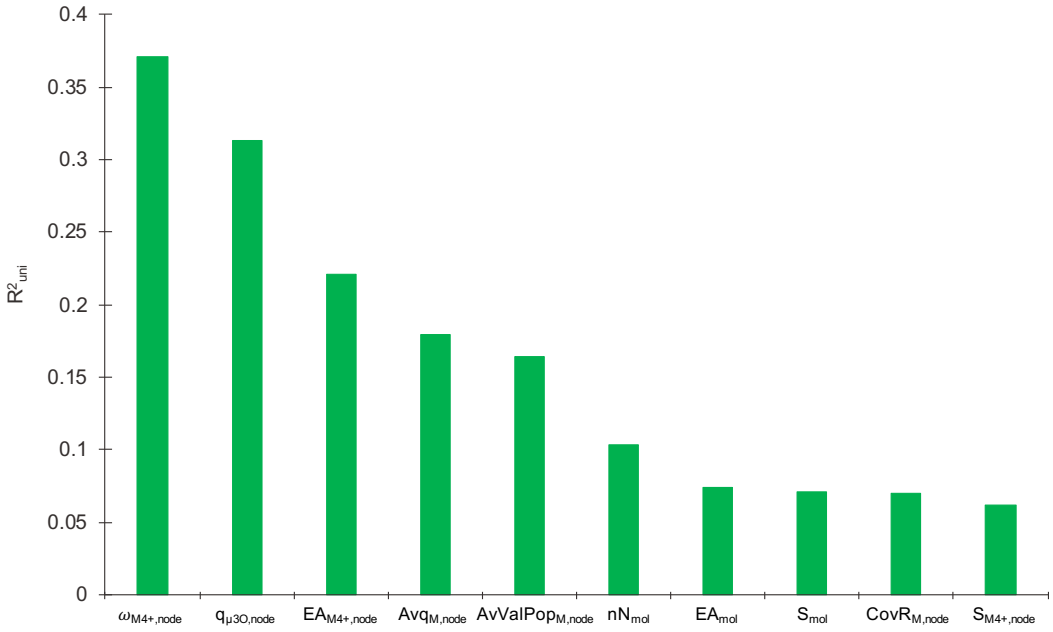


Figure C.17. The 10 most important descriptors for describing the binding free energies of bidentate anions to *node–noOH₂,OH* sites, as identified by uniparametric coefficients of determination (R^2_{uni}) for simple linear regression using individual descriptors.

C.15 QSAR Modeling: Description of Stepwise MLR Algorithm

As stated in the main text, we used stepwise forward-backward based feature selection combined with MLR using a 95% confidence interval for development of the multi-parametric QSAR models (for both the neutrals and bidentates datasets). In MATLAB, we utilized the built-in *stepwiselm* function. This algorithm performs forward selection of descriptors if their corresponding *p*-values are less than 0.05, then uses backward elimination of the descriptor with the largest *p*-value if any descriptors in the model at the current step have *p*-values higher than 0.10. Here, the *p*-values are for an F-test of the change in the sum of squared error resulting from the addition or removal of a descriptor. Finally, the algorithm terminates whenever a single step cannot improve the model statistics.

C.16 QSAR Modeling: Tests for Statistical Significance

Training set:

$$R^2_{adj} = 1 - \left(\frac{N_{train}-1}{N_{train}-p}\right)\frac{\sum\left(y_{obs(train)} - y_{pred(train)}\right)^2}{\sum\left(y_{obs(train)} - \bar{y}_{(train)}\right)^2} \tag{C.8}$$

In eq C.8, $y_{obs(train)}$ are the observed ΔG_{bind} responses computed by DFT for the training set molecule-node combinations, $y_{pred(train)}$ are the responses predicted by the regression equation (developed using the entire training set), $\bar{y}_{(train)}$ is the average response for the training set, and summations are over all training set molecule-node combinations. Further, N_{train} is the number of training set observations (i.e., the number of molecule-node combinations used to train the QSAR model) and p is the number of regression coefficients (which includes the intercept).

$$Q^2_{LOO} = 1 - \frac{\sum\left(y_{obs(train)} - y_{pred(train)}\right)^2}{\sum\left(y_{obs(train)} - \bar{y}_{(train)}\right)^2} \tag{C.9}$$

In eq C.9, $y_{obs(train)}$ are the observed ΔG_{bind} responses computed by DFT, $y_{pred(train)}$ are the responses predicted by the leave-one-out cross-validation (LOO-CV) procedure, $\bar{y}_{(train)}$ is the average response, and summations are over all training set molecule-node combinations. In the LOO-CV procedure, every molecule-node combination is eliminated from the dataset once and then its response variable is predicted using the regression equation derived from the remaining set.

$${}^{C}R^2_P = R_{nr} \times \sqrt{R^2_{nr} - R^2_r} \tag{C.10}$$

The possibility of chance correlation or structural redundancy in the developed QSAR models was measured by the *y*-randomization method. This involved randomizing the training set response values, while leaving the training set descriptor matrix unchanged and repeating the entire process of statistical validation. Using the coefficients of determination for the randomized model (R^2_r) and non-random model (R^2_{nr}), the ${}^{C}R^2_P$ parameter was calculated using eq C.10. If ${}^{C}R^2_P > 0.5$, then that QSAR model is not generated purely by chance.[5]

Test set:

$$RMSE_{test} = \sqrt{\frac{\sum\left(y_{obs(test)} - y_{pred(test)}\right)^2}{N_{test}}} \tag{C.11}$$

In eq C.11, $y_{obs(test)}$ are the observed ΔG_{bind} responses computed by DFT for the test set molecule-node combinations, $y_{pred(test)}$ are the test set responses predicted by the regression equation (developed using the training set), N_{test} is the number of test set observations, and summations are over all test set molecule-node combinations.

$$R^2_{test} = 1 - \frac{\sum\left(y_{obs(test)} - y_{pred(test)}\right)^2}{\sum\left(y_{obs(test)} - \bar{y}_{(test)}\right)^2} \tag{C.12}$$

In eq C.12, $\bar{y}_{(test)}$ is the average response for the test set.

$$\rho_{test}\left(Y_{obs(test)}, Y_{pred(test)}\right) = 1 - \frac{6\sum d^2}{\mathrm{n}(n^2-1)} \tag{C.13}$$

To compute the Spearman's rank correlation coefficient for the test set (ρ_{test}), the observed responses and predicted responses are ranked in order of value and then the two columns ($Y_{obs(test)}$ and $Y_{pred(test)}$) are compared. In eq C.13, d is the difference between the ranks of the two columns and n is the length of each column (i.e., the number of test set observations).

C.17 QSAR Modeling: Variability in Predictive QSAR Model Statistics

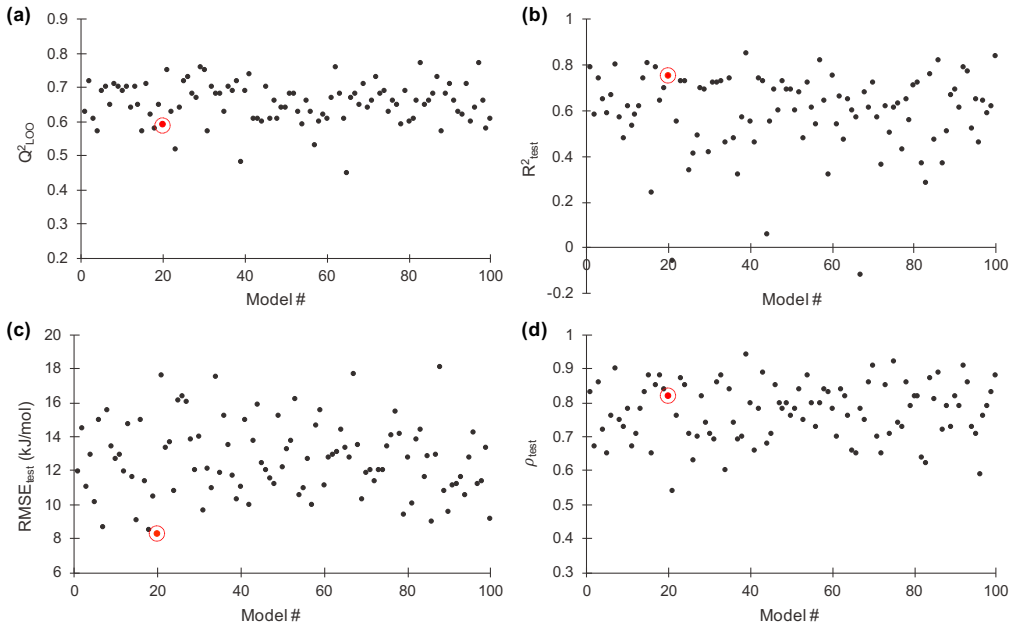


Figure C.18. (a) Training set Q^2_{LOO}, (b) test set coefficients of determination, (c) test set root mean square errors, and (d) test set Spearman's rank correlation coefficients for the 100 predictive QSAR models generated for the neutrals dataset. Red points indicate the optimal model (i.e., the model with the lowest $RMSE_{test}$, which also satisfies the minimum QSAR criteria of $Q^2_{LOO} > 0.5$, ${}^{C}R^2_p > 0.5$, and $R^2_{test} > 0.6$).

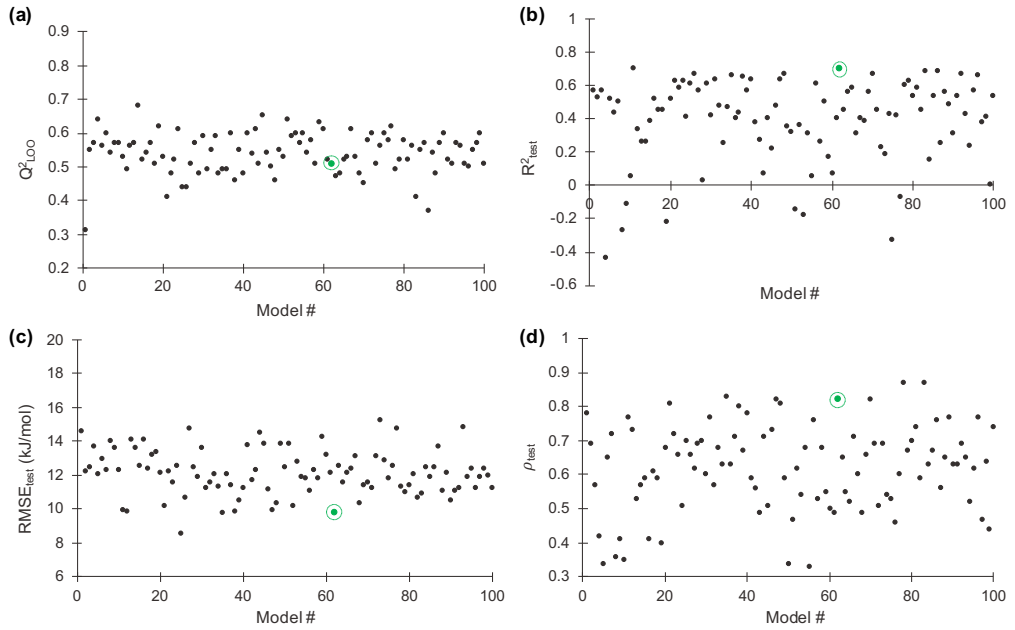


Figure C.19. (a) Training set Q^2_{LOO}, (b) test set coefficients of determination, (c) test set root mean square errors, and (d) test set Spearman's rank correlation coefficients for the 100 predictive QSAR models generated for the bidentates dataset. Green points indicate the optimal model (i.e., the model with the lowest $RMSE_{test}$, which also satisfies the minimum QSAR criteria of $Q^2_{LOO} > 0.5$, ${}^{C}R^2_p > 0.5$, and $R^2_{test} > 0.6$).

C.18 QSAR Modeling: Leverages and Standardized Residuals

As stated in the main text, leverage (h) is a measure of the influence of a molecule-node combination's properties on a regression model. Leverages for individual molecule-node combinations within the training and test sets are calculated as the diagonals of the hat matrices (H), defined as:

$$H_{train} = X_{train}(X_{train}^T X_{train})^{-1} X_{train}^T \quad \text{(C.14)}$$

$$H_{test} = X_{test}(X_{train}^T X_{train})^{-1} X_{train}^T \quad \text{(C.15)}$$

where X_{train} and X_{test} are the design matrices containing the molecular/node descriptors for the training and test sets, respectively. More specifically, each row in the design matrices contains the 49 (42) scaled descriptor values for an individual molecule-node combination in the neutrals (bidentates) dataset.

The standardized residuals for response i (st_i) are computed as raw residuals (r_i) divided by their estimated standard deviation:

$$st_i = \frac{r_i}{\sqrt{MSE(1 - h_{ii})}} = \frac{y_{obs,i} - y_{pred,i}}{\sqrt{MSE(1 - h_{ii})}} \quad \text{(C.16)}$$

where MSE is the mean squared error and h_{ii} is the leverage value for ΔG_{bind} response i.

C.19 QSAR Modeling: Optimal Predictive QSAR Model for Neutrals Dataset

Table C.16. Predictions, residuals, and leverages for training and test set molecule-node combinations using the optimal predictive QSAR model for the neutrals dataset (eq 4.9 in the main text).

Node/Molecule	Set	Observation (kJ/mol)	Prediction (kJ/mol)	Raw Residual (kJ/mol)	Standardized Residual	Leverage
Zr-NU-1000 (c pore)/18	Training	-62.23	-57.57	-4.66	-0.52	0.104
Zr-bi(trans)-defect UiO-66/17	Training	53.55	47.49	6.06	0.70	0.157
Zr-mono-defect UiO-66/24	Training	-22.86	-21.82	-1.04	-0.11	0.055
Zr-MOF-808/3	Training	10.75	1.24	9.51	1.04	0.060
Ce-bi(trans)-defect UiO-66/3	Training	-30.60	-32.02	1.42	0.16	0.084
Zr-mono-defect UiO-66/17	Training	31.15	47.45	-16.30	-2.08	0.305
Hf-bi(trans)-defect UiO-66/22	Training	-10.52	-10.34	-0.18	-0.02	0.144
Zr-NU-1000 (c pore)/20	Training	-61.17	-61.84	0.67	0.08	0.147
Th-bi(trans)-defect UiO-66/14	Training	-45.92	-42.56	-3.36	-0.40	0.193
Hf-MOF-808/14	Training	14.21	3.47	10.74	1.27	0.191
Zr-mono-defect UiO-66/13	Training	-24.21	-13.70	-10.51	-1.18	0.100
Hf-bi(trans)-defect UiO-66/H_2O	Training	-35.37	-44.01	8.64	1.08	0.271
Ce-bi(trans)-defect UiO-66/20	Training	-48.02	-49.29	1.27	0.15	0.170
Zr-NU-1000 (c pore)/13	Training	-15.07	-38.95	23.88	2.69	0.105
Zr-MOF-808/11	Training	-2.17	-5.79	3.62	0.40	0.058
Zr-NU-1000 (large pore)/18	Training	-18.65	-20.35	1.70	0.18	0.042
Hf-MOF-808/17	Training	37.89	37.30	0.59	0.08	0.298
Ce-bi(trans)-defect UiO-66/24	Training	-28.88	-30.25	1.37	0.15	0.092
Zr-NU-1000 (large pore)/8	Training	-2.77	-4.76	1.99	0.22	0.034
Zr-bi(cis)-defect UiO-66/14	Training	-19.44	-36.46	17.02	1.95	0.138
Th-MOF-808/18	Training	-22.03	-13.76	-8.28	-0.94	0.129
Zr-bi(trans)-defect UiO-66/24	Training	-22.74	-19.85	-2.89	-0.31	0.036
Zr-MOF-808/13	Training	15.91	8.56	7.35	0.81	0.074
Hf-MOF-808/20	Training	-20.04	-19.72	-0.32	-0.04	0.136
Hf-bi(trans)-defect UiO-66/20	Training	-41.13	-31.53	-9.60	-1.12	0.172
Ce-bi(trans)-defect UiO-66/13	Training	-30.32	-31.57	1.25	0.14	0.125
Ce-MOF-808/11	Training	-20.18	-23.18	3.00	0.33	0.073
Zr-NU-1000 (c pore)/8	Training	-27.65	-43.81	16.16	1.82	0.105
Hf-MOF-808/3	Training	17.40	-5.31	22.71	2.50	0.068
Zr-NU-1000 (large pore)/17	Training	52.11	51.55	0.56	0.07	0.213
Hf-bi(trans)-defect UiO-66/24	Training	-29.28	-28.60	-0.68	-0.07	0.075
Th-bi(trans)-defect UiO-66/18	Training	-28.85	-33.37	4.52	0.50	0.082

Ce-MOF-808/22	Training	-18.32	-23.96	5.64	0.62	0.064
Zr-bi(trans)-defect UiO-66/20	Training	-34.95	-30.29	-4.66	-0.52	0.093
Hf-bi(trans)-defect UiO-66/13	Training	-33.49	-19.33	-14.16	-1.58	0.089
Ce-MOF-808/17	Training	-10.26	-22.97	12.71	1.58	0.265
Zr-NU-1000 (large pore)/24	Training	10.52	-6.76	17.28	1.88	0.041
Zr-NU-1000 (large pore)/13	Training	-17.51	0.01	-17.52	-1.91	0.046
Zr-NU-1000 (large pore)/20	Training	-26.13	-24.56	-1.57	-0.18	0.095
Ce-MOF-808/14	Training	-26.85	-21.86	-4.99	-0.60	0.226
Ce-MOF-808/13	Training	-23.03	-16.49	-6.54	-0.73	0.092
Zr-bi(cis)-defect UiO-66/13	Training	-31.11	-19.47	-11.64	-1.29	0.073
Hf-bi(trans)-defect UiO-66/17	Training	42.31	42.88	-0.57	-0.07	0.321
Th-bi(trans)-defect UiO-66/H_2O	Training	-38.22	-46.35	8.13	0.94	0.157
Th-bi(trans)-defect UiO-66/6	Training	-12.66	-15.31	2.65	0.30	0.112
Zr-NU-1000 (c pore)/6	Training	-53.29	-41.87	-11.42	-1.30	0.128
Ce-bi(trans)-defect UiO-66/14	Training	-59.86	-61.84	1.98	0.27	0.407
Zr-bi(cis)-defect UiO-66/17	Training	46.79	32.95	13.84	1.63	0.188
Hf-bi(trans)-defect UiO-66/18	Training	-33.49	-28.99	-4.50	-0.50	0.095
Hf-bi(trans)-defect UiO-66/11	Training	-13.15	-12.71	-0.44	-0.05	0.082
Th-MOF-808/3	Training	-20.22	-21.20	0.98	0.11	0.091
Ce-bi(trans)-defect UiO-66/H_2O	Training	-31.97	-38.33	6.36	0.72	0.124
Zr-bi(cis)-defect UiO-66/20	Training	-22.08	-35.62	13.54	1.52	0.099
Zr-bi(trans)-defect UiO-66/H_2O	Training	-29.93	-34.33	4.40	0.51	0.142
Zr-bi(trans)-defect UiO-66/22	Training	-4.41	-11.46	7.05	0.78	0.066
Ce-MOF-808/18	Training	-32.30	-28.09	-4.21	-0.47	0.105
Hf-bi(trans)-defect UiO-66/1	Training	-4.56	-9.36	4.80	0.54	0.114
Th-bi(trans)-defect UiO-66/24	Training	-60.15	-34.47	-25.68	-2.87	0.095
Hf-MOF-808/8	Training	-5.76	-3.48	-2.28	-0.25	0.062
Zr-bi(cis)-defect UiO-66/24	Training	-36.40	-27.68	-8.72	-0.95	0.044
Ce-MOF-808/3	Training	-21.76	-17.54	-4.22	-0.46	0.067
Zr-NU-1000 (c pore)/H_2O	Training	-61.08	-58.70	-2.38	-0.29	0.213
Zr-mono-defect UiO-66/18	Training	-22.41	-26.05	3.64	0.40	0.069
Zr-bi(trans)-defect UiO-66/14	Training	-39.10	-31.65	-7.45	-0.85	0.123
Zr-MOF-808/22	Training	-5.32	-8.56	3.24	0.36	0.064
Ce-bi(trans)-defect UiO-66/11	Training	-30.19	-33.80	3.61	0.42	0.148
Zr-mono-defect UiO-66/20	Training	-42.32	-27.76	-14.56	-1.65	0.120
Zr-bi(cis)-defect UiO-66/8	Training	-16.97	-24.31	7.34	0.80	0.050
Zr-NU-1000 (c pore)/22	Training	-45.37	-44.48	-0.89	-0.10	0.134
Hf-MOF-808/22	Training	-5.23	-6.82	1.59	0.18	0.118

Hf-MOF-808/18	Training	-11.96	-13.86	1.90	0.21	0.084
Zr-bi(cis)-defect UiO-66/3	Training	-13.59	-27.92	14.33	1.56	0.046
Zr-NU-1000 (c pore)/3	Training	-57.65	-46.09	-11.56	-1.31	0.116
Zr-MOF-808/H_2O	Training	-7.91	-8.47	0.56	0.06	0.123
Zr-NU-1000 (large pore)/6	Training	-2.67	-4.58	1.91	0.21	0.077
Zr-mono-defect UiO-66/11	Training	-10.84	-8.52	-2.32	-0.27	0.146
Zr-bi(trans)-defect UiO-66/1	Training	4.59	-9.67	14.26	1.58	0.075
Ce-bi(trans)-defect UiO-66/6	Training	-31.15	-29.18	-1.97	-0.22	0.083
Th-bi(trans)-defect UiO-66/20	Training	-32.48	-33.30	0.82	0.09	0.109
Th-bi(trans)-defect UiO-66/1	Training	-18.98	-14.85	-4.13	-0.47	0.113
Zr-NU-1000 (large pore)/22	Training	-5.00	-7.21	2.21	0.24	0.076
Ce-bi(trans)-defect UiO-66/17	Training	2.90	7.58	-4.68	-0.68	0.466
Zr-MOF-808/17	Training	32.39	37.92	-5.53	-0.68	0.251
Ce-MOF-808/8	Training	-24.64	-17.22	-7.42	-0.82	0.063
Zr-bi(cis)-defect UiO-66/22	Training	-24.18	-17.96	-6.22	-0.68	0.056
Zr-NU-1000 (large pore)/H_2O	Training	-28.06	-20.24	-7.82	-0.89	0.132
Hf-bi(trans)-defect UiO-66/8	Training	-17.57	-25.26	7.69	0.85	0.067
Th-MOF-808/20	Training	-24.76	-17.13	-7.63	-0.93	0.238
Zr-bi(trans)-defect UiO-66/6	Training	-17.01	-7.88	-9.13	-1.01	0.072
Zr-mono-defect UiO-66/H_2O	Training	-40.93	-36.06	-4.87	-0.58	0.198
Zr-bi(cis)-defect UiO-66/6	Training	-20.29	-14.49	-5.80	-0.64	0.061
Zr-mono-defect UiO-66/8	Training	-4.21	-18.34	14.13	1.56	0.071
Zr-MOF-808/18	Training	-17.53	-13.45	-4.08	-0.45	0.076
Th-MOF-808/H_2O	Training	-21.56	-23.25	1.69	0.20	0.167
Th-MOF-808/24	Training	-17.00	-22.41	5.41	0.61	0.101
Th-bi(trans)-defect UiO-66/8	Training	-34.53	-32.26	-2.27	-0.25	0.091
Th-MOF-808/22	Training	-14.84	-10.80	-4.04	-0.45	0.099
Zr-NU-1000 (large pore)/3	Training	-10.80	-6.98	-3.82	-0.41	0.035
Zr-bi(trans)-defect UiO-66/13	Training	-12.81	-12.27	-0.54	-0.06	0.063
Th-MOF-808/8	Training	-24.14	-19.80	-4.34	-0.48	0.086
Hf-MOF-808/13	Training	18.72	3.52	15.20	1.68	0.076
Ce-MOF-808/H_2O	Training	-28.77	-16.30	-12.47	-1.39	0.084
Zr-bi(trans)-defect UiO-66/3	Training	-21.40	-20.40	-1.00	-0.11	0.037
Zr-bi(cis)-defect UiO-66/18	Training	-45.77	-32.93	-12.84	-1.41	0.056
Zr-mono-defect UiO-66/3	Training	-13.37	-22.43	9.06	0.99	0.061
Zr-NU-1000 (large pore)/14	Training	-30.54	-15.14	-15.40	-1.73	0.104
Hf-MOF-808/1	Training	-0.40	-12.03	11.63	1.32	0.117
Zr-NU-1000 (c pore)/11	Training	-40.63	-45.35	4.72	0.53	0.085

Zr-NU-1000 (c pore)/17	Training	-1.96	9.83	-11.79	-1.40	0.196
Zr-MOF-808/8	Training	-6.35	2.78	-9.13	-1.00	0.058
Th-bi(trans)-defect UiO-66/17	Training	33.93	21.64	12.29	1.60	0.335
Th-MOF-808/11	Training	-5.02	-10.14	5.12	0.57	0.078
Zr-NU-1000 (large pore)/1	Training	1.87	-9.06	10.93	1.20	0.066
Ce-MOF-808/6	Training	-18.51	-23.94	5.43	0.60	0.061
Zr-mono-defect UiO-66/22	Training	-1.06	-8.47	7.41	0.82	0.083
Th-bi(trans)-defect UiO-66/11	Training	-19.34	-20.56	1.22	0.14	0.107
Ce-bi(trans)-defect UiO-66/8	Training	-27.51	-30.26	2.75	0.31	0.081
Hf-MOF-808/24	Training	-16.35	-6.41	-9.94	-1.10	0.081
Zr-MOF-808/24	Training	-23.57	0.22	-23.79	-2.62	0.067
Ce-MOF-808/20	Training	-29.83	-33.40	3.57	0.42	0.190
Hf-bi(trans)-defect UiO-66/3	Training	-29.49	-29.02	-0.47	-0.05	0.071
Zr-NU-1000 (large pore)/11	Training	-2.78	-8.05	5.27	0.57	0.036
Th-MOF-808/17	Training	-22.68	-14.72	-7.96	-1.26	0.548
Th-bi(trans)-defect UiO-66/22	Training	-17.42	-18.53	1.11	0.12	0.110
Zr-bi(cis)-defect UiO-66/H_2O	Training	-54.46	-40.17	-14.29	-1.62	0.120
Th-MOF-808/14	Test	-12.52	-3.94	-8.58	-1.13	0.152
Zr-mono-defect UiO-66/6	Test	-12.54	-4.46	-8.08	-1.03	0.095
Zr-NU-1000 (c pore)/1	Test	-30.26	-46.51	16.25	2.08	0.102
Ce-MOF-808/1	Test	-17.88	-29.29	11.41	1.47	0.117
Hf-MOF-808/11	Test	7.37	-5.76	13.13	1.65	0.068
Zr-NU-1000 (c pore)/14	Test	-57.52	-52.29	-5.23	-0.70	0.192
Zr-bi(cis)-defect UiO-66/11	Test	-7.50	-17.39	9.89	1.26	0.097
Zr-NU-1000 (c pore)/24	Test	-57.90	-45.96	-11.94	-1.55	0.125
Zr-MOF-808/14	Test	8.60	2.33	6.27	0.83	0.156
Hf-bi(trans)-defect UiO-66/6	Test	-25.52	-6.37	-19.15	-2.51	0.144
Zr-bi(trans)-defect UiO-66/18	Test	-28.45	-27.79	-0.66	-0.08	0.045
Th-MOF-808/13	Test	-10.58	-15.39	4.81	0.61	0.091
Th-bi(trans)-defect UiO-66/3	Test	-37.58	-35.31	-2.27	-0.29	0.096
Hf-MOF-808/6	Test	1.91	-4.94	6.85	0.88	0.117
Zr-MOF-808/20	Test	-24.90	-19.36	-5.54	-0.72	0.127
Hf-MOF-808/H_2O	Test	-14.76	-16.55	1.79	0.25	0.266
Ce-bi(trans)-defect UiO-66/22	Test	-26.55	-31.32	4.77	0.60	0.077
Zr-mono-defect UiO-66/14	Test	-24.55	-34.42	9.87	1.29	0.135
Zr-bi(trans)-defect UiO-66/11	Test	-4.23	-11.94	7.71	0.98	0.082
Ce-bi(trans)-defect UiO-66/18	Test	-42.44	-47.55	5.11	0.66	0.126
Ce-bi(trans)-defect UiO-66/1	Test	-27.20	-29.46	2.26	0.30	0.144

Th-MOF-808/6	Test	-0.70	-9.89	9.19	1.17	0.097
Th-MOF-808/1	Test	-7.41	-14.46	7.05	0.93	0.150
Hf-bi(trans)-defect UiO-66/14	Test	-40.74	-31.28	-9.46	-1.30	0.228
Zr-mono-defect UiO-66/1	Test	-9.99	-4.94	-5.05	-0.65	0.125
Th-bi(trans)-defect UiO-66/13	Test	-36.50	-29.08	-7.42	-0.95	0.099
Zr-MOF-808/1	Test	-8.47	-13.17	4.70	0.59	0.081
Zr-MOF-808/6	Test	-4.81	-7.08	2.27	0.28	0.062
Zr-bi(trans)-defect UiO-66/8	Test	-12.14	-16.96	4.82	0.60	0.042
Zr-bi(cis)-defect UiO-66/1	Test	-22.57	-16.23	-6.34	-0.80	0.074
Ce-MOF-808/24	Test	-22.22	-17.96	-4.26	-0.54	0.078

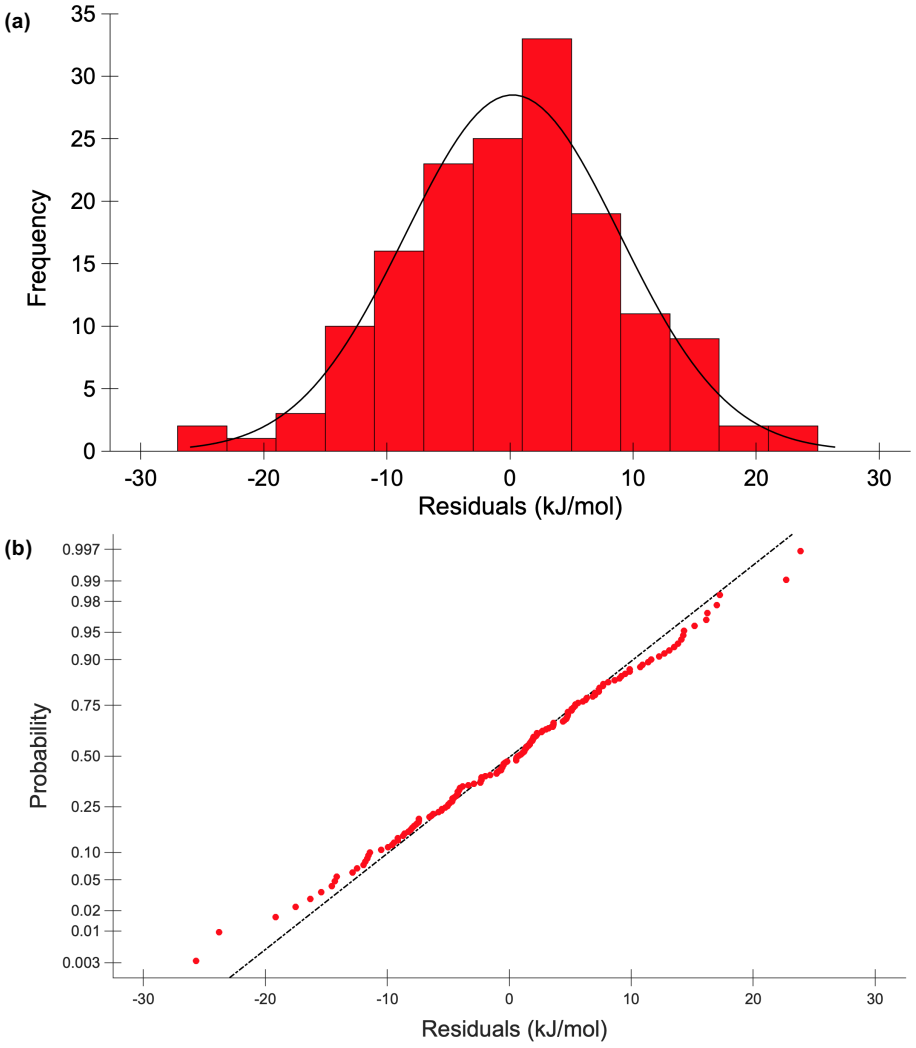


Figure C.20. Distribution of residuals for the optimal predictive QSAR model developed for the neutrals dataset (eq 4.9 in the main text). The (a) histogram of raw residuals and (b) normal probability plot of raw residuals show that the residuals have an acceptable normal distribution, with no clear outliers.

C.20 QSAR Modeling: Optimal Predictive QSAR Model for Bidentates Dataset

Table C.17. Predictions, residuals, and leverages for training and test set molecule-node combinations using the optimal predictive QSAR model for the bidentates dataset (eq 4.10 in the main text).

Node/Molecule	Set	Observation (kJ/mol)	Prediction (kJ/mol)	Raw Residual (kJ/mol)	Standardized Residual	Leverage
Ce-bi(trans)-defect UiO-66/19	Training	-157.38	-146.68	-10.70	-1.03	0.149
Th-bi(trans)-defect UiO-66/12	Training	-58.87	-82.62	23.75	2.22	0.091
Zr-NU-1000 (large pore)/7	Training	-92.74	-90.77	-1.97	-0.18	0.028
Zr-bi(trans)-defect UiO-66/12	Training	-70.69	-88.84	18.15	1.67	0.067
Zr-NU-1000 (large pore)/15	Training	-98.88	-94.27	-4.61	-0.42	0.028
Zr-bi(trans)-defect UiO-66/7	Training	-89.10	-94.56	5.46	0.49	0.030
Hf-MOF-808/16	Training	-94.97	-82.54	-12.43	-1.18	0.125
Zr-NU-1000 (c pore)/9	Training	-118.57	-105.96	-12.61	-1.17	0.086
Zr-NU-1000 (c pore)/19	Training	-138.38	-120.43	-17.95	-1.69	0.106
Zr-NU-1000 (c pore)/16	Training	-100.82	-95.87	-4.95	-0.46	0.083
Th-bi(trans)-defect UiO-66/9	Training	-92.55	-97.01	4.46	0.42	0.096
Ce-MOF-808/7	Training	-108.30	-116.98	8.68	0.81	0.093
Hf-MOF-808/19	Training	-118.91	-107.10	-11.81	-1.15	0.158
Zr-NU-1000 (c pore)/12	Training	-83.49	-91.58	8.09	0.75	0.084
Ce-MOF-808/9	Training	-120.02	-125.64	5.62	0.54	0.125
Th-MOF-808/19	Training	-119.16	-104.86	-14.30	-1.37	0.141
Zr-mono-defect UiO-66/4	Training	-85.40	-99.60	14.20	1.32	0.077
Hf-bi(trans)-defect UiO-66/19	Training	-109.98	-113.41	3.43	0.33	0.136
Zr-NU-1000 (large pore)/12	Training	-82.68	-85.05	2.37	0.22	0.065
Zr-mono-defect UiO-66/23	Training	-104.92	-104.42	-0.50	-0.05	0.071
Hf-MOF-808/7	Training	-90.80	-83.98	-6.82	-0.64	0.091
Hf-bi(trans)-defect UiO-66/21	Training	-95.16	-107.02	11.86	1.11	0.088
Th-bi(trans)-defect UiO-66/16	Training	-101.46	-86.92	-14.54	-1.36	0.092
Zr-bi(trans)-defect UiO-66/21	Training	-92.50	-111.29	18.79	1.71	0.046
Th-MOF-808/7	Training	-92.50	-81.73	-10.77	-1.00	0.075
Zr-NU-1000 (c pore)/21	Training	-119.30	-114.03	-5.27	-0.48	0.062
Zr-NU-1000 (c pore)/23	Training	-112.45	-106.47	-5.98	-0.56	0.085
Th-bi(trans)-defect UiO-66/19	Training	-87.40	-111.47	24.07	2.28	0.116
Zr-bi(cis)-defect UiO-66/7	Training	-96.01	-95.21	-0.80	-0.07	0.033
Th-MOF-808/16	Training	-95.33	-80.30	-15.03	-1.42	0.110
Zr-MOF-808/9	Training	-92.27	-97.11	4.84	0.45	0.080
Zr-MOF-808/15	Training	-101.41	-91.94	-9.47	-0.86	0.045

Zr-bi(trans)-defect UiO-66/23	Training	-101.57	-103.73	2.16	0.20	0.068
Th-MOF-808/15	Training	-99.97	-85.23	-14.74	-1.37	0.077
Ce-MOF-808/21	Training	-133.05	-133.71	0.66	0.06	0.116
Ce-MOF-808/4	Training	-112.07	-121.32	9.25	0.88	0.128
Th-bi(trans)-defect UiO-66/4	Training	-88.95	-92.69	3.74	0.35	0.100
Zr-MOF-808/4	Training	-77.99	-92.79	14.80	1.38	0.086
Zr-mono-defect UiO-66/12	Training	-100.97	-89.53	-11.44	-1.06	0.070
Zr-mono-defect UiO-66/19	Training	-108.18	-118.38	10.20	0.95	0.093
Ce-bi(trans)-defect UiO-66/15	Training	-130.05	-127.05	-3.00	-0.28	0.089
Zr-MOF-808/23	Training	-76.28	-97.61	21.33	1.98	0.084
Ce-MOF-808/16	Training	-107.23	-115.55	8.32	0.79	0.129
Zr-mono-defect UiO-66/2	Training	-97.92	-101.42	3.50	0.32	0.029
Zr-MOF-808/2	Training	-72.86	-94.61	21.75	1.98	0.041
Zr-bi(cis)-defect UiO-66/21	Training	-111.31	-111.94	0.63	0.06	0.049
Zr-MOF-808/21	Training	-98.84	-105.18	6.34	0.58	0.064
Zr-bi(cis)-defect UiO-66/2	Training	-115.40	-101.37	-14.03	-1.27	0.029
Zr-NU-1000 (large pore)/9	Training	-90.93	-99.43	8.50	0.78	0.065
Zr-bi(trans)-defect UiO-66/19	Training	-104.28	-117.69	13.41	1.25	0.091
Th-bi(trans)-defect UiO-66/15	Training	-101.63	-91.85	-9.78	-0.90	0.057
Hf-bi(trans)-defect UiO-66/16	Training	-96.51	-88.85	-7.66	-0.72	0.108
Ce-bi(trans)-defect UiO-66/7	Training	-131.37	-123.55	-7.82	-0.73	0.086
Zr-mono-defect UiO-66/21	Training	-109.14	-111.98	2.84	0.26	0.049
Th-bi(trans)-defect UiO-66/21	Training	-95.08	-105.08	10.00	0.93	0.074
Th-bi(trans)-defect UiO-66/23	Training	-83.96	-97.52	13.56	1.27	0.097
Hf-MOF-808/9	Training	-94.41	-92.64	-1.77	-0.17	0.127
Zr-NU-1000 (large pore)/2	Training	-102.57	-96.94	-5.63	-0.51	0.025
Hf-MOF-808/2	Training	-74.56	-90.14	15.58	1.45	0.086
Zr-bi(cis)-defect UiO-66/12	Training	-101.87	-89.48	-12.39	-1.14	0.069
Ce-bi(trans)-defect UiO-66/12	Training	-127.39	-117.83	-9.56	-0.91	0.120
Zr-bi(cis)-defect UiO-66/15	Training	-109.65	-98.70	-10.95	-0.99	0.033
Hf-MOF-808/23	Training	-78.99	-93.15	14.16	1.35	0.124
Ce-bi(trans)-defect UiO-66/4	Training	-129.87	-127.89	-1.98	-0.19	0.124
Th-MOF-808/9	Training	-106.03	-90.39	-15.64	-1.48	0.113
Hf-bi(trans)-defect UiO-66/12	Training	-74.95	-84.56	9.61	0.91	0.116
Hf-bi(trans)-defect UiO-66/9	Training	-110.57	-98.95	-11.62	-1.10	0.112
Zr-mono-defect UiO-66/15	Training	-89.83	-98.75	8.92	0.81	0.033
Hf-bi(trans)-defect UiO-66/2	Training	-103.62	-96.46	-7.16	-0.66	0.068
Zr-NU-1000 (large pore)/4	Training	-91.49	-95.11	3.62	0.33	0.071

Zr-bi(trans)-defect UiO-66/2	Training	-100.44	-100.73	0.29	0.03	0.026
Zr-mono-defect UiO-66/9	Training	-108.00	-103.91	-4.09	-0.38	0.071
Ce-bi(trans)-defect UiO-66/9	Training	-135.51	-132.21	-3.30	-0.31	0.121
Th-MOF-808/12	Training	-85.27	-76.01	-9.26	-0.87	0.109
Ce-MOF-808/12	Training	-114.96	-111.25	-3.71	-0.35	0.126
Th-MOF-808/4	Training	-95.47	-86.08	-9.39	-0.89	0.117
Th-bi(trans)-defect UiO-66/7	Training	-70.96	-88.35	17.39	1.59	0.056
Zr-MOF-808/19	Training	-119.41	-111.57	-7.84	-0.74	0.110
Zr-MOF-808/16	Training	-91.96	-87.01	-4.95	-0.46	0.079
Zr-MOF-808/12	Training	-83.00	-82.72	-0.28	-0.03	0.080
Ce-bi(trans)-defect UiO-66/2	Training	-135.32	-129.72	-5.60	-0.52	0.084
Th-MOF-808/21	Training	-113.47	-98.46	-15.01	-1.41	0.096
Hf-bi(trans)-defect UiO-66/7	Training	-92.40	-90.29	-2.11	-0.20	0.075
Zr-NU-1000 (large pore)/21	Training	-114.58	-107.50	-7.08	-0.65	0.046
Ce-MOF-808/23	Training	-105.88	-126.15	20.27	1.94	0.138
Zr-mono-defect UiO-66/7	Training	-82.91	-95.25	12.34	1.12	0.033
Th-bi(trans)-defect UiO-66/2	Training	-86.81	-94.52	7.71	0.71	0.053
Zr-NU-1000 (c pore)/2	Training	-111.63	-103.47	-8.16	-0.74	0.043
Zr-mono-defect UiO-66/16	Training	-75.76	-93.82	18.06	1.67	0.068
Zr-bi(cis)-defect UiO-66/23	Training	-121.62	-104.37	-17.25	-1.59	0.071
Hf-bi(trans)-defect UiO-66/23	Training	-105.05	-99.46	-5.59	-0.53	0.106
Hf-bi(trans)-defect UiO-66/4	Training	-99.71	-94.63	-5.08	-0.48	0.122
Zr-NU-1000 (c pore)/4	Training	-101.60	-101.64	0.04	0.00	0.092
Zr-MOF-808/7	Training	-87.76	-88.45	0.69	0.06	0.044
Ce-bi(trans)-defect UiO-66/21	Training	-154.70	-140.28	-14.42	-1.36	0.106
Zr-bi(trans)-defect UiO-66/9	Training	-107.89	-103.22	-4.67	-0.43	0.068
Hf-MOF-808/21	Test	-105.27	-100.71	-4.56	-0.49	0.108
Zr-bi(trans)-defect UiO-66/4	Test	-96.71	-98.90	2.19	0.23	0.074
Zr-bi(cis)-defect UiO-66/16	Test	-98.87	-93.77	-5.10	-0.54	0.068
Ce-bi(trans)-defect UiO-66/16	Test	-145.02	-122.12	-22.90	-2.50	0.123
Zr-bi(cis)-defect UiO-66/19	Test	-129.35	-118.33	-11.02	-1.18	0.093
Zr-NU-1000 (large pore)/23	Test	-90.24	-99.94	9.70	1.03	0.067
Zr-bi(cis)-defect UiO-66/4	Test	-100.00	-99.55	-0.45	-0.05	0.077
Hf-MOF-808/4	Test	-84.10	-88.32	4.22	0.46	0.136
Ce-bi(trans)-defect UiO-66/23	Test	-130.88	-132.72	1.84	0.20	0.130
Zr-NU-1000 (large pore)/16	Test	-102.89	-89.34	-13.55	-1.43	0.063
Hf-MOF-808/15	Test	-102.06	-87.48	-14.58	-1.56	0.089
Zr-NU-1000 (large pore)/19	Test	-122.33	-113.90	-8.43	-0.91	0.092

Ce-MOF-808/15	Test	-114.78	-120.48	5.70	0.61	0.096
Zr-bi(trans)-defect UiO-66/16	Test	-90.44	-93.13	2.69	0.28	0.065
Hf-MOF-808/12	Test	-80.79	-78.25	-2.54	-0.28	0.132
Ce-MOF-808/19	Test	-150.40	-140.10	-10.30	-1.15	0.161
Hf-bi(trans)-defect UiO-66/15	Test	-104.35	-93.79	-10.56	-1.12	0.072
Zr-bi(trans)-defect UiO-66/15	Test	-100.24	-98.06	-2.18	-0.23	0.030
Zr-NU-1000 (c pore)/7	Test	-102.89	-97.30	-5.59	-0.59	0.047
Zr-bi(cis)-defect UiO-66/9	Test	-119.69	-103.87	-15.82	-1.68	0.071
Ce-MOF-808/2	Test	-103.36	-123.15	19.79	2.12	0.092
Th-MOF-808/23	Test	-90.91	-90.90	-0.01	0.00	0.118
Zr-NU-1000 (c pore)/15	Test	-109.16	-100.80	-8.36	-0.88	0.047
Th-MOF-808/2	Test	-87.97	-87.90	-0.07	-0.01	0.073

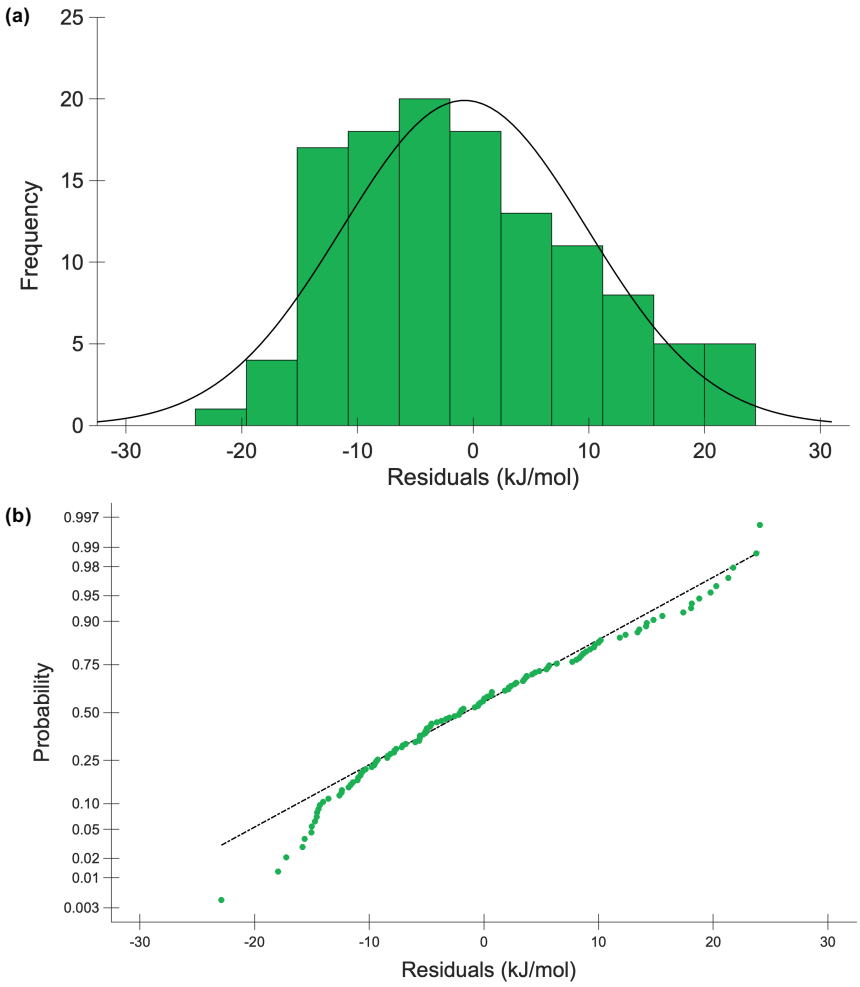


Figure C.21. Distribution of residuals for the optimal predictive QSAR model developed for the bidentates dataset (eq 4.10 in the main text). The (a) histogram of raw residuals and (b) normal probability plot of raw residuals show that the residuals have an acceptable normal distribution, with no clear outliers.

C.21 References for Appendix C

1. Koopmans, T. Über Die Zuordnung von Wellenfunktionen Und Eigenwerten Zu Den Einzelnen Elektronen Eines Atoms. *Physica* **1934**, *1*, 104–113.
2. Parr, R. G.; Chattaraj, P. K. Principle of Maximum Hardness. *J. Am. Chem. Soc.* **1991**, *113*, 1854–1855.
3. Talete. *Dragon Version 6.0.* **2010**, http://www.talete.mi.it/
4. Todeschini, R.; Vighi, M.; Finizio, A.; Gramatica, P. 3D Modeling and Prediction by WHIM Descriptors. Part 8. Toxicity and Physico-Chemical Properties of Environmental Priority Chemicals by 2D-TI and 3D-WHIM Descriptors. *SAR & QSAR Environ. Res.* **1997**, *7*, 173–193.
5. Mitra, I.; Saha, A.; Roy, K. Exploring Quantitative Structure-Activity Relationship (QSAR) Studies of Antioxidant Phenolic Compounds Obtained from Traditional Chinese Medicinal Plants. *Mol. Simul.* **2010**, *36*, 1067–1079.

C.22 Cartesian Coordinates and Raw Energy Values

The optimized cartesian coordinates and raw energy values for every system (molecules, nodes, and node-bound molecules) will be provided in the supplementary files of the published manuscript.

Appendix D: Supporting Information for Chapter 5

This appendix is a modified version of the Supporting Information for the manuscript in Chapter 5 (*ACS Catal.* **2020**, *10*, 1310–1323).

D.1 Benzoate Cluster Models

Different "benzoate cluster models" were used at various points throughout the study. They were constructed by modeling the bottom 4 linkers as formate groups and using benzoate groups for the 4 linkers located around the active site. Fig. D.1a corresponds to the "fully hydrated" NU-1000 node ($[Zr_6(\mu_3\text{-}O)_4(\mu_3\text{-}OH)_4(OH)_4(OH_2)_4]^{8+}(C_6H_5COO^-)_4(HCOO^-)_4$). Fig. D.1b corresponds to the "partially dehydrated" NU-1000 node ($[Zr_6(\mu_3\text{-}O)_4(\mu_3\text{-}OH)_4(OH)_4(OH_2)_3]^{8+}(C_6H_5COO^-)_4(HCOO^-)_4$), in which one terminal H_2O moiety has desorbed to generate an open metal site. This "partially dehydrated" node topology was used to calculate the full pathway for gas-phase sarin hydrolysis on bare, unfunctionalized NU-1000. The cleavage of a Zr-OH_2 bond ("fully hydrated" → "partially dehydrated") requires approximately 84 kJ/mol in free energy. Fig. D.1c represents the Ti^{IV}-NU-1000 benzoate cluster model, which was only used to test the effect of model size on the reaction energetics for the Ti^{IV}-NU-1000 system.

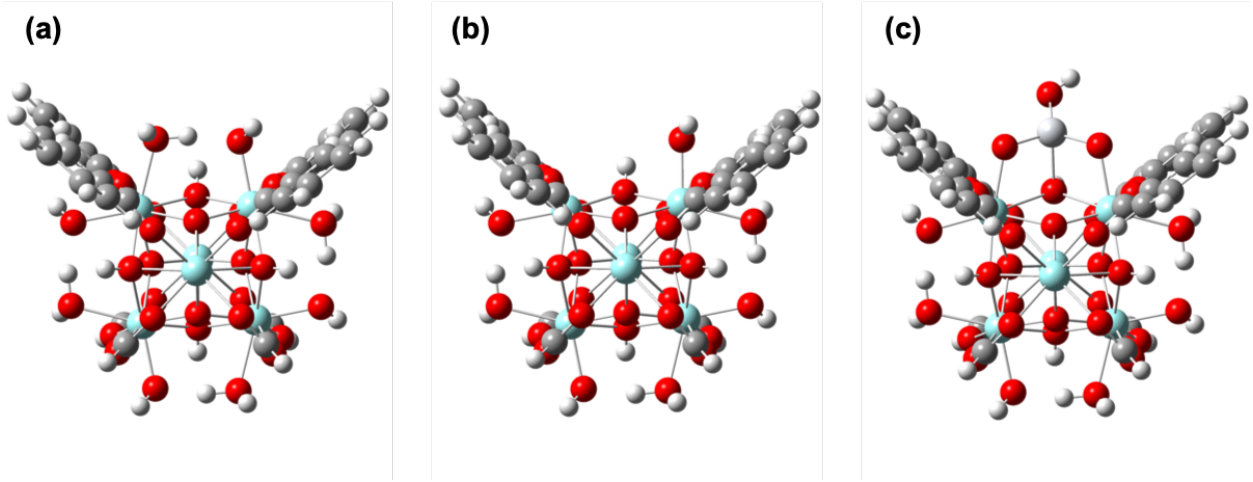


Figure D.1. (a) Optimized NU-1000 ("fully hydrated"), (b) "partially dehydrated" NU-1000, and (c) Ti^{IV}-NU-1000 benzoate cluster models. Dark grey, white, red, turquoise, and light grey spheres represent C, H, O, Zr, and Ti atoms, respectively.

D.2 Binding Energies for Sarin, Water, and IMPA for All 36 M-NU-1000 Systems

Table D.1. Binding free energies for sarin, water, and IMPA to M^{II}-NU-1000 systems. Values (in kJ/mol) are calculated at T = 298.15 K and P = 1 atm. Negative and positive energies indicate favorable and unfavorable binding, respectively.

Metal	Sarin	Water	IMPA
V^{II}	-40.4	-21.2	-139.6
Cr^{II}	-30.2	-58.5	-138.3
Mn^{II}	-52.7	-26.0	-125.3
Fe^{II}	-30.4	-15.8	-113.3
Co^{II}	-36.9	-16.0	-109.6
Ni^{II}	-48.3	-29.6	-124.1
Cu^{II}	-12.0	-29.9	-98.8
Zn^{II}	-43.1	-24.0	-121.1
Mo^{II}	-49.6	-36.1	-140.3
Pd^{II}	-64.2	-70.5	-75.2
W^{II}	-44.9	-33.8	-114.1
Pt^{II}	-79.6	-84.7	-160.4

Table D.2. Binding free energies for sarin, water, and IMPA to M^{III}-NU-1000 systems. Values (in kJ/mol) are calculated at T = 298.15 K and P = 1 atm. Negative and positive energies indicate favorable and unfavorable binding, respectively.

Metal	Sarin	Water	IMPA
Sc^{III}	-48.1	-29.5	-125.4
Cr^{III}	-54.0	-23.7	-132.5
Fe^{III}	-19.1	-14.0	-89.4
Co^{III}	-12.4	8.8	-57.2
Cu^{III}	-15.7	1.4	-72.6
Y^{III}	-58.8	-34.8	-137.5
Ru^{III}	-74.9	-47.4	-129.3
Rh^{III}	-45.6	-37.9	-127.6
Ce^{III}	-53.9	-33.6	-145.3
Ir^{III}	-59.8	-51.5	-135.6
Au^{III}	-94.6	-62.8	-115.9

Table D.3. Binding free energies for sarin, water, and IMPA to M^{IV}-NU-1000 systems. Values (in kJ/mol) are calculated at T = 298.15 K and P = 1 atm. Negative and positive energies indicate favorable and unfavorable binding, respectively.

Metal	Sarin	Water	IMPA
Ti^{IV}	-99.2	-48.6	-111.7
V^{IV}	-100.8	-56.4	-116.4
Mn^{IV}	-53.7	-21.1	-65.2
Zr^{IV}	-127.3	-65.3	-142.3
Mo^{IV}	-59.1	-13.4	-59.4
Ru^{IV}	-54.0	-28.7	-62.0
Pd^{IV}	7.1	59.1	-17.0
Ce^{IV}	-79.0	-42.4	-108.2
Hf^{IV}	-135.5	-72.9	-144.8
W^{IV}	-28.8	73.4	-59.7
Re^{IV}	21.0	34.6	8.2
Os^{IV}	27.6	59.8	-5.6
Pt^{IV}	23.2	28.2	8.3

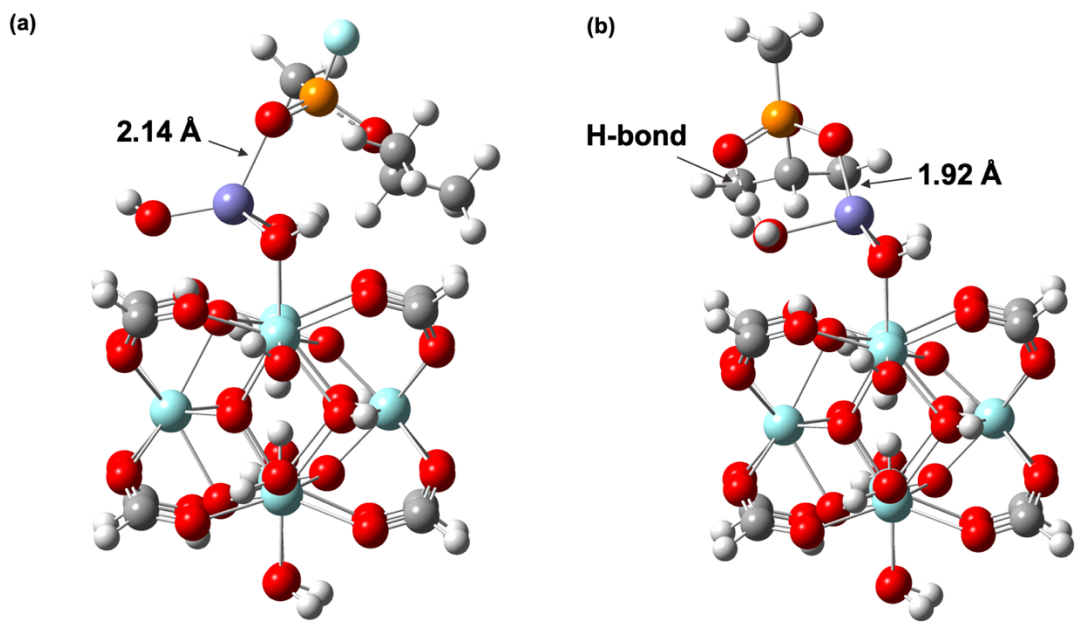


Figure D.2. (a) Optimized **Fe^{II}-Sarin** and (b) **Fe^{II}-IMPA** species, highlighting the different binding geometry for each adsorbate species that contributes to the stronger binding energies for IMPA compared to sarin. Dark grey, white, red, turquoise, and purple spheres represent C, H, O, Zr, and Fe atoms, respectively.

D.3 Discussion of Mechanistic Pathway Chosen

In the main text, we used a formate cluster model of Ti^{IV}-NU-1000 as an example to clearly explain each step of the proposed catalytic cycle for gas-phase sarin hydrolysis on these systems (see Scheme 5.2). In the first transition state (**TS_{nuc}**) of the mechanism, the Ti–OH group abstracts a H atom from water, which then performs a nucleophilic attack on the P atom of sarin in a concerted step. We only discussed one pathway, where nucleophilic attack occurs directly opposite the –OiPr group. However, we noted that there are two other pathways where attack can also occur opposite the $–CH_3$ or –F groups of sarin. For the discussion below, we will refer to these as the **OiPr**, **CH_3**, and **F** pathways, respectively, where all pathways are calculated using formate cluster models. For each pathway, we only modeled the mechanism involving cleavage of the P–F bond in sarin to generate HF product, as this is known to be the most favorable pathway.[1]

First, we compare the **OiPr** and **CH_3** pathways. For both pathways, nucleophilic attack leads to the formation of a pentacoordinated trigonal bipyramidal (TBP) **INT_1** species where the –F group is in an equatorial position. For the purposes of screening across many metals, we were only interested in the mechanism involving cleavage of the P–F bond to generate HF product. Thus, elimination of –F requires a Berry pseudorotation to direct it from an equatorial to an axial site, forming **INT_2**. Up until this point, the **OiPr** and **CH_3** pathways are unique (i.e., each pathway has unique **Reactant Complex**, **TS_{nuc}**, **INT_1**, and **TS_{Berry}** species). Upon the formation of **INT_2**, both pathways are exactly the same (i.e., each pathway has identical **INT_2**, **TS_{eli}**, **Product Complex**, and **Ti^{IV}–IMPA** species). In **TS_{eli}**, elimination of HF involves proton transfer from the nucleophilic –OH group and the resulting **Product Complex** is stabilized by hydrogen bonds. Once HF desorbs from the active site, the final **Ti^{IV}–IMPA** species is generated. Finally, the IMPA product desorbs and regenerates the active site, completing the cycle. As seen in Fig. D.3a, the **OiPr** and **CH_3** pathways have essentially identical reaction free energy profiles. Since the pathways differ by only a few kJ/mol and are within DFT error, we chose to only focus on the **OiPr** pathway, although this choice is arbitrary. For the purposes of screening a large number of catalysts, it was more computationally feasible to only consider one pathway for comparison. Further, in theory, the relative turnover frequencies (TOFs) between catalysts should be similar regardless of the hydrolysis pathway considered, given the similarity in the energy profiles.

The **F** pathway differs significantly from the other two pathways. After nucleophilic attack occurs directly opposite the –F group in **TS_{nuc}**, the **INT_1** species is formed. Here, the –F group is already in an axial position and does not require a Berry pseudorotation, in contrast to the other pathways. Further, we could not optimize a stable **TS_{eli}** species involving elimination of HF. We attempted a relaxed potential energy scan by increasing the distance of the P–F bond to simulate bond cleavage, but there was no maximum on the potential energy surface identifying a transition state, unlike in the other pathways; the F atom simply "flies away" from the sarin molecule and the energy decreases monotonically. Thus, the next species after **INT_1** along the reaction coordinate is **Ti^{IV}–IMPA**. The **F** pathway is "missing" four species (**TS_{Berry}**, **INT_2**, **TS_{eli}**, and **Product Complex**) overall compared to the **OiPr** and **CH_3** pathways. Due to the orientation of sarin in the **F** pathway, it is impossible for there to be a proton transfer from the nucleophilic –OH group to form HF. Further, in the modeling scheme we used, we only considered one ambient water molecule in the gas-phase reaction. Thus, there are no other surrounding sources of H to participate in proton transfer to the leaving F atom. For this reason, the **F** pathway is an "incomplete" mechanism

(indicated in Fig. D.3b), and it is difficult to directly compare it to the other two pathways. Since the energetic span model considers the relative free energies of *all* intermediates and transition states in the catalytic cycle to calculate the TOF, we needed a "complete" mechanism in order to accurately compare each catalyst. Thus, we chose to only focus on the **OiPr** pathway in the main text.

Finally, comparing the profiles in Fig. D.3, we note that the free energy barriers for the nucleophilic attack elementary step ($\Delta G^{\ddagger}_{nuc}$ = G(**TS$_{nuc}$**) – G(**Reactant Complex**)) are nearly identical for all three pathways. Explicitly, $\Delta G^{\ddagger}_{nuc}$ is 105.4, 107.3, and 109.2 kJ/mol for the **OiPr**, **CH$_3$**, and **F** pathways, respectively. Since all three pathways have identical **Ti^{IV}–IMPA** species, the free energy barriers for the IMPA desorption elementary step are equivalent for each pathway. In the main text, we observed that nucleophilic attack and IMPA desorption are the most influential elementary steps in the overall reaction. Since the barriers for these key steps are similar for each pathway, we chose the **OiPr** pathway for screening all catalysts.

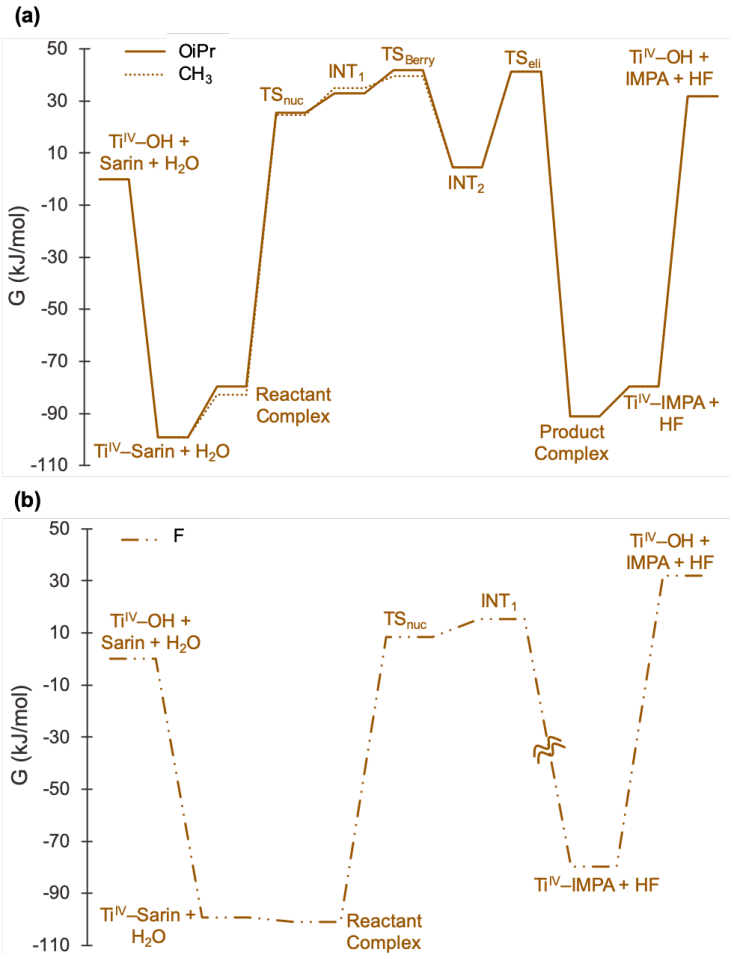


Figure D.3. Comparison of the reaction free energy profiles for gas-phase sarin hydrolysis catalyzed by the formate cluster model of Ti^{IV}-NU-1000, following the (a) **OiPr** and **CH$_3$** pathways and (b) the **F** pathway.

D.4 Effect of Transition Metal on Key Barriers – Values

Table D.4. The effect of transition metal identity and oxidation state on the free energy barrier to form **INT1** (ΔG_{INT1}) and the free energy barrier to desorb the IMPA product from the active site ($\Delta G_{IMPA,des.}$) of M-NU-1000 systems. Values (in kJ/mol) are calculated at T = 298.15 K and P = 1 atm.

Metal	ΔG_{INT1}	$\Delta G_{IMPA,des.}$
Pd^{IV}	61.7	17.0
Cu^{III}	67.3	72.6
Co^{III}	81.3	57.2
Fe^{III}	75.3	89.4
Cu^{II}	62.0	98.8
W^{IV}	108.4	59.7
Co^{II}	80.0	109.6
Pd^{II}	109.8	75.2
Fe^{II}	68.0	113.3
W^{II}	79.5	114.1
Au^{III}	94.7	115.9
Zn^{II}	71.5	121.1
Ce^{IV}	122.7	108.2
Ni^{II}	51.5	124.1
Mn^{II}	77.2	125.3
Sc^{III}	69.3	125.4
Rh^{III}	61.3	127.6
Ru^{III}	91.0	129.3
Ti^{IV}	132.2	111.7
Cr^{III}	78.0	132.5
Ru^{IV}	133.3	62.0
Ir^{III}	78.4	135.6
Y^{III}	64.7	137.5
V^{IV}	137.5	116.4
Cr^{II}	42.3	138.3
V^{II}	37.1	139.6
Mo^{II}	90.6	140.3
Zr^{IV}	127.5	142.3
Mo^{IV}	143.2	59.4
Mn^{IV}	143.2	65.2
Hf^{IV}	131.9	144.8
Ce^{III}	67.1	145.3
Pt^{II}	59.3	160.4

D.5 Effect of Transition Metal on Key Barriers – Pareto Front

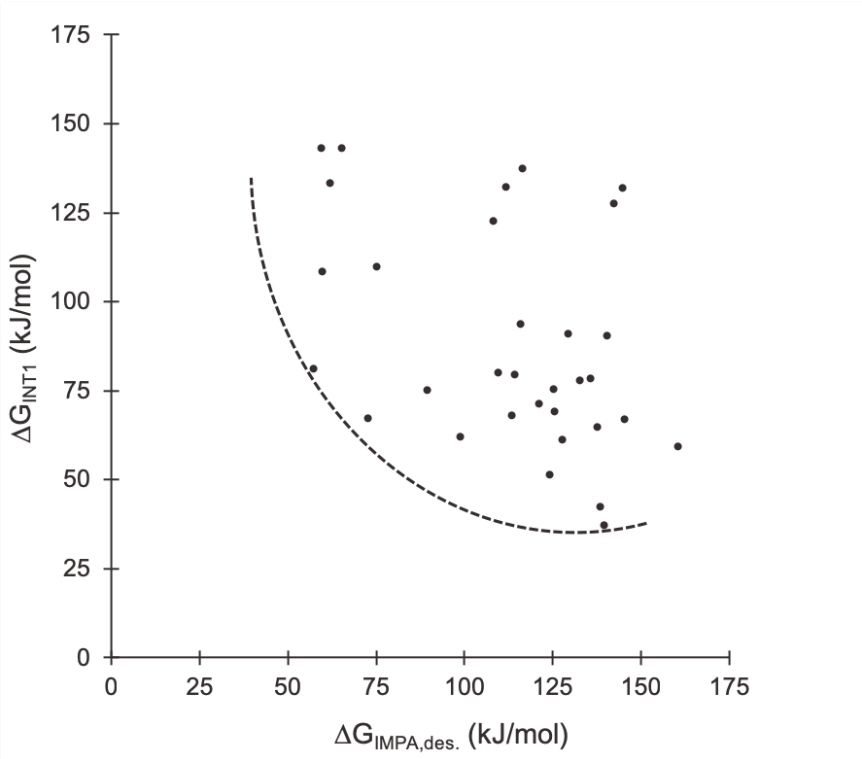


Figure D.4. The effect of transition metal identity and oxidation state (shown in Fig. 5.5) on the free energy barrier to form **INT**$_1$ and the free energy barrier to desorb the IMPA product from the active site of 32 M-NU-1000 systems, where the optimal metals are located closest to the origin. A pseudo-Pareto front curve shows the tradeoff between the energy barrier for nucleophilic attack and the product desorption energy, such that no metal has perfectly low barriers for both steps. Pd^{IV}-NU-1000 is omitted from this plot, as it falls outside the Pareto-front curve for the other 32 metals.

D.6 Comparison of Full Pathways – Free Energy Barriers for Four Elementary Steps

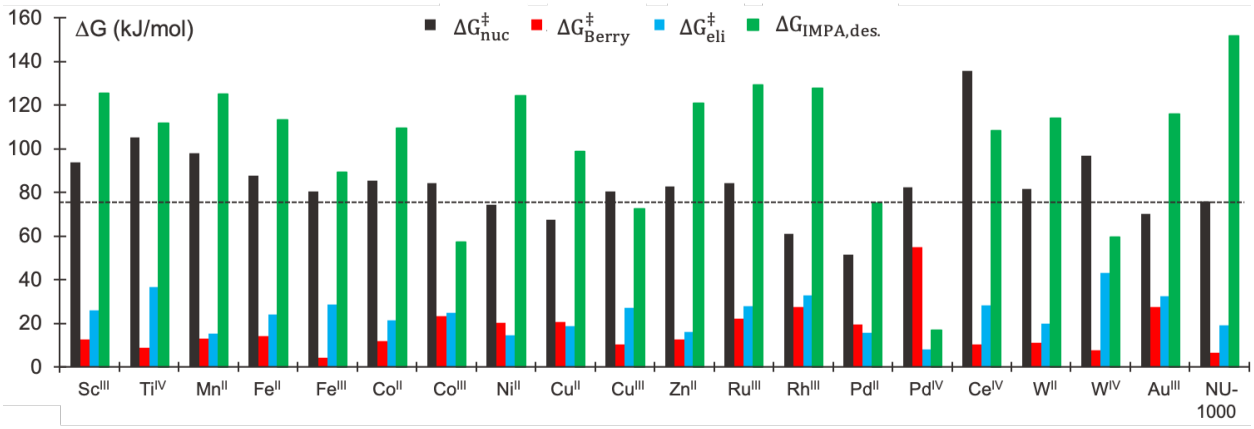


Figure D.5. Free energy barriers for the four elementary steps in gas-phase sarin hydrolysis: nucleophilic attack, Berry pseudorotation, HF elimination, and IMPA desorption. The dashed line indicates the smallest "rate-determining step" barrier for all systems ($\Delta G_{IMPA,des.}$ = 75.2 kJ/mol for Pd^{II}-NU-1000).

Table D.5. Free energy barriers for the four elementary steps for gas-phase sarin hydrolysis catalyzed by NU-1000 and M-NU-1000 systems. Values (in kJ/mol) are calculated at T = 298.15 K and P = 1 atm.

System	$\Delta G^{\ddagger}_{nuc}$	$\Delta G^{\ddagger}_{Berry}$	$\Delta G^{\ddagger}_{eli}$	$\Delta G_{IMPA,des.}$
Sc^{III}	93.7	12.6	25.9	125.4
Ti^{IV}	105.4	8.8	36.7	111.7
Mn^{II}	98.1	13.0	15.5	125.3
Fe^{II}	87.6	14.4	24.2	113.3
Fe^{III}	80.4	4.4	28.8	89.4
Co^{II}	85.6	11.8	21.3	109.6
Co^{III}	84.2	23.4	24.8	57.2
Ni^{II}	74.4	20.2	14.6	124.1
Cu^{II}	67.4	20.8	18.9	98.8
Cu^{III}	80.5	10.4	27.0	72.6
Zn^{II}	82.7	12.9	16.2	121.1
Ru^{III}	84.3	22.3	28.1	129.3
Rh^{III}	61.2	27.5	32.9	127.6
Pd^{II}	51.7	19.5	15.8	75.2
Pd^{IV}	82.3	54.9	8.1	17.0
Ce^{IV}	135.9	10.6	28.4	108.2
W^{II}	81.7	11.0	19.8	114.1
W^{IV}	96.9	7.7	43.2	59.7
Au^{III}	70.3	27.5	32.4	115.9
NU-1000	76.0	6.7	19.0	151.7

D.7 Proposed Catalytic Cycle for Gas-Phase Sarin Hydrolysis on NU-1000

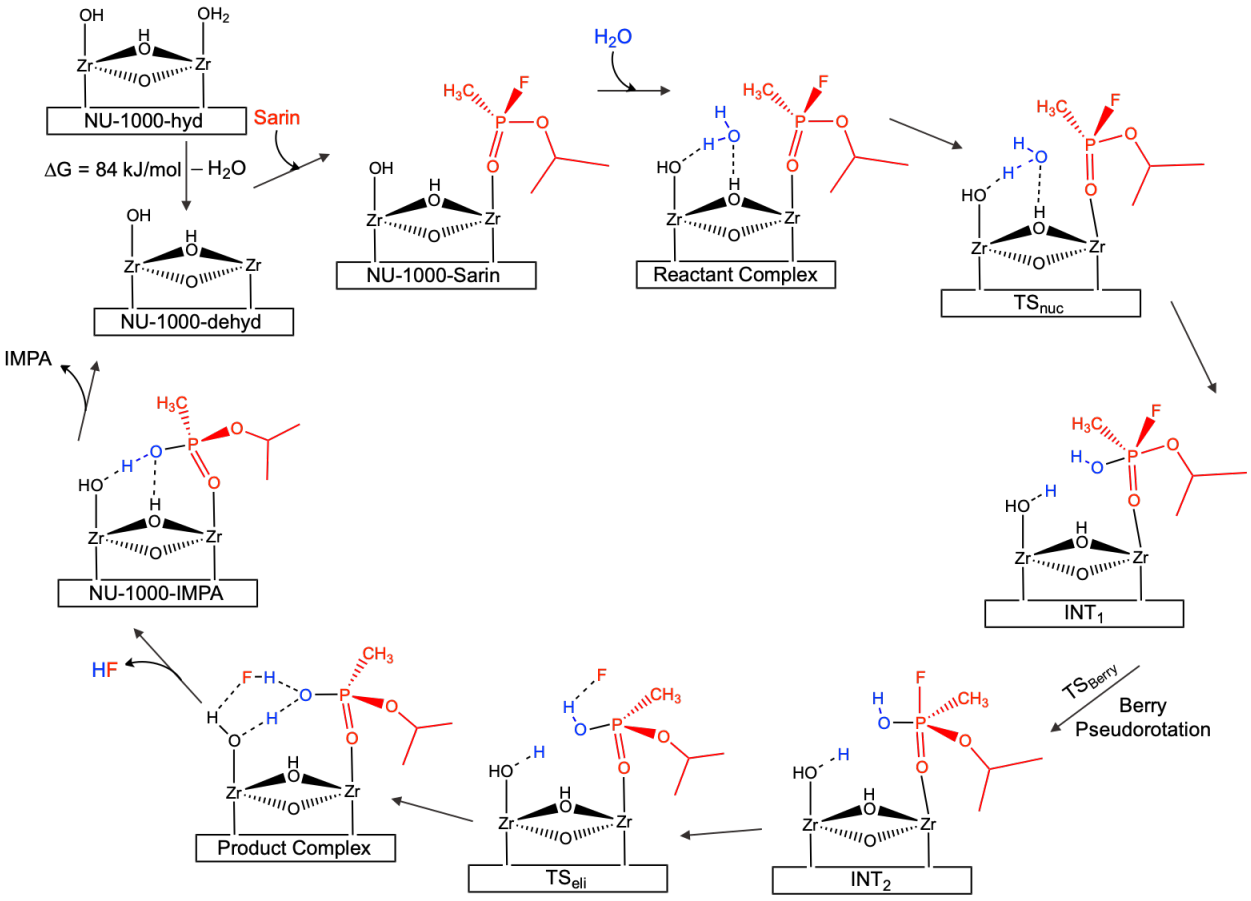


Figure D.6. Illustration of the proposed catalytic cycle of gas-phase sarin hydrolysis on a NU-1000 cluster. The sarin and ambient water molecules are colored red and blue, respectively, to clarify the movement of atoms during the course of the reaction mechanism.

D.8 Comparison of Monodentate and Bidentate Product Binding on NU-1000

In the main text, and throughout most of the SI, we present results for unfunctionalized NU-1000 assuming a general base hydrolysis mechanism, in which sarin and IMPA bind to Zr in a monodentate fashion, for the most direct comparison to the 19 M-NU-1000 catalysts. Here, "general base hydrolysis mechanism" refers to the case when the nucleophile is an explicit water molecule (that would be present from ambient moisture). However, we also performed a calculation to confirm that the energetics for a bidentate binding mode on NU-1000 are more unfavorable, as shown below. As seen in Fig. D.7, the free energy barrier to desorb the IMPA product and regenerate the **NU-1000-dehyd** active site is $\Delta G_{IMPA,des.}$ = 151.7 and 168.4 kJ/mol for monodentate and bidentate modes, respectively. This agrees with previous results[1] showing stronger binding energies for a bidentate hydrolysis product where IMPA is bound through two Zr-oxo bonds, compared to only one Zr-oxo bond in the monodentate mode.

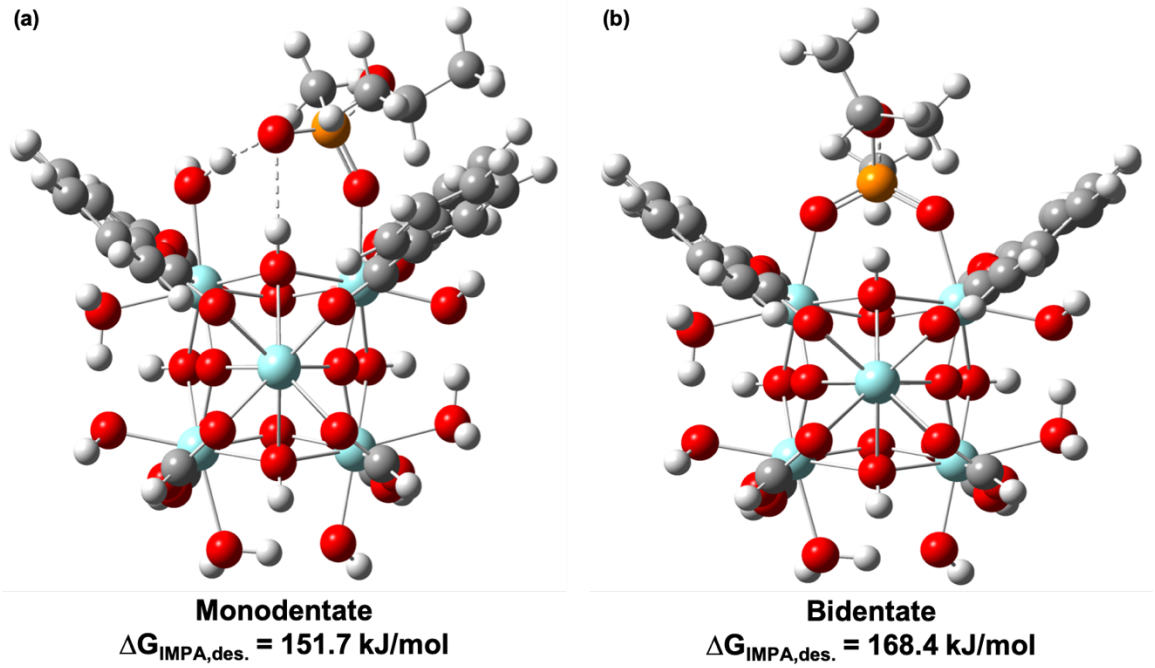
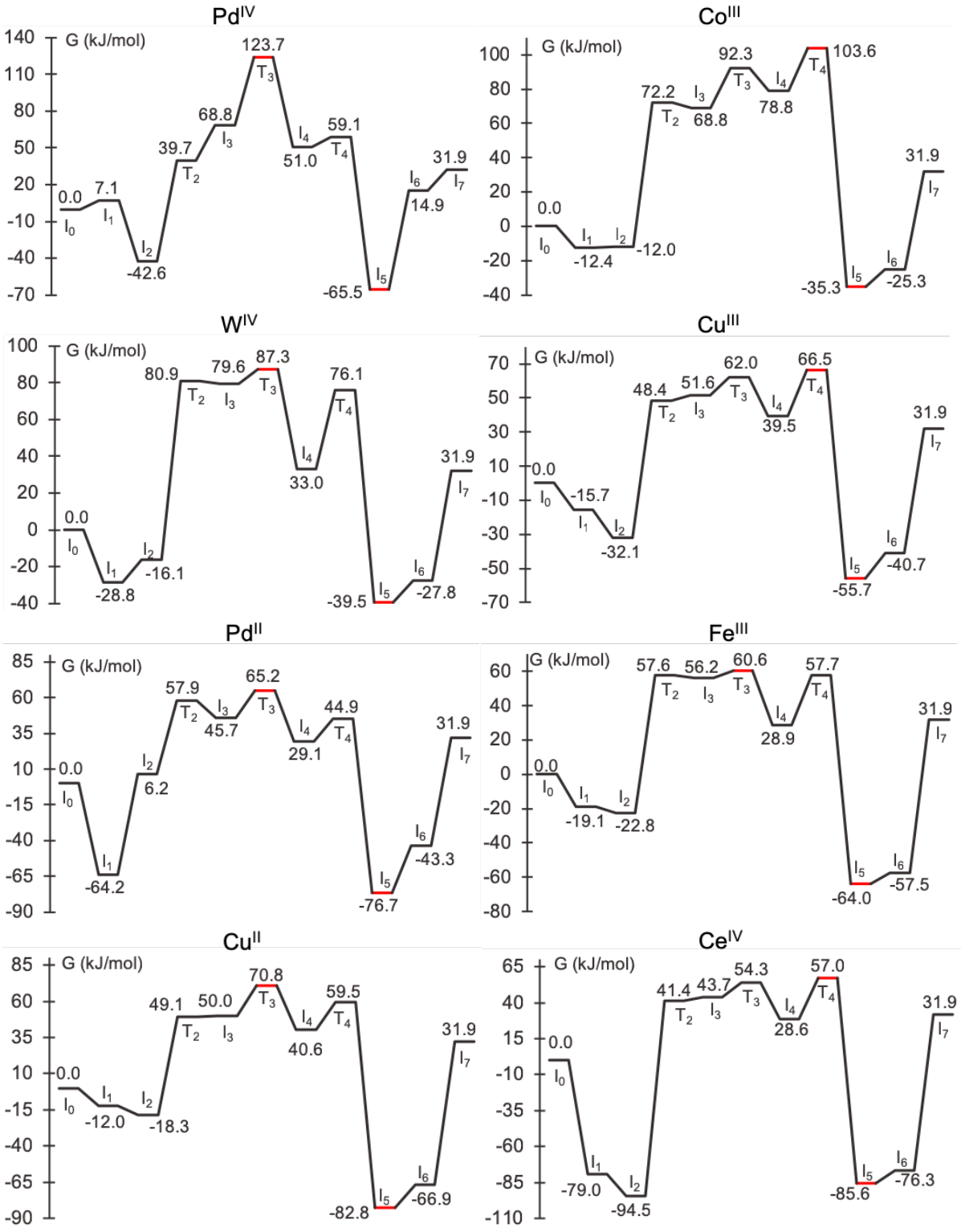
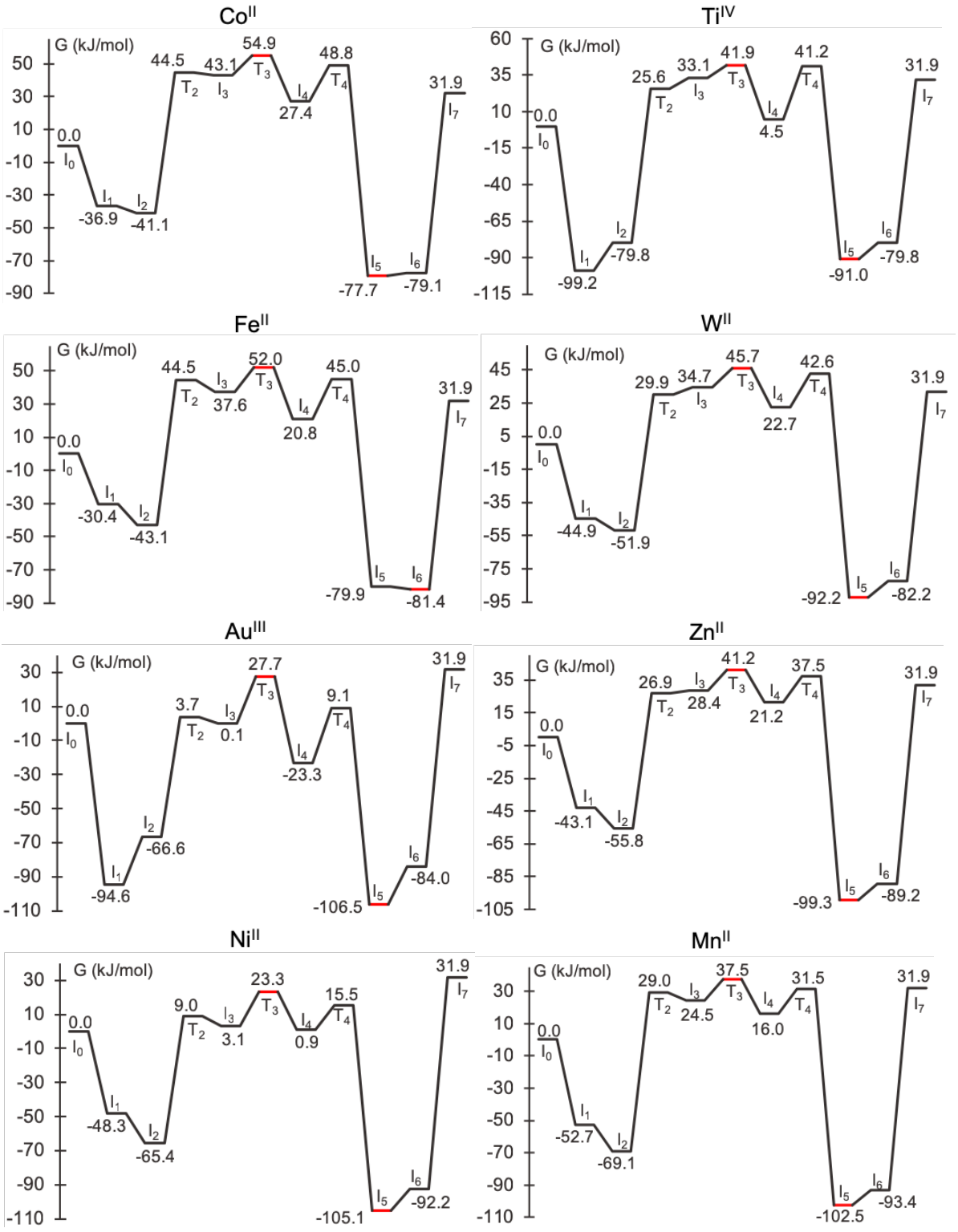


Figure D.7. Optimized **NU-1000–IMPA** species where IMPA is bound in a (a) monodentate and (b) bidentate fashion, highlighting the different binding geometries that contribute to the stronger binding energy for bidentate-bound IMPA compared to monodentate-bound IMPA. Dark grey, white, red, and turquoise spheres represent C, H, O, and Zr atoms, respectively.

D.9 Individual Free Energy Profiles for NU-1000 and 19 M-NU-1000 Catalysts




Co^II
G (kJ/mol)
50
30
10
-10
-30
-50
-70
-90
0.0
I0
I1
-36.9
I2
-41.1
44.5
T2
43.1
I3
54.9
T3
I4
27.4
48.8
T4
I5
-77.7
I6
-79.1
31.9
I7
Ti^IV
G (kJ/mol)
60
35
10
-15
-40
-65
-90
-115
0.0
I0
I1
-99.2
I2
-79.8
25.6
T2
33.1
I3
41.9
T3
I4
4.5
41.2
T4
I5
-91.0
I6
-79.8
31.9
I7
Fe^II
G (kJ/mol)
50
30
10
-10
-30
-50
-70
-90
0.0
I0
I1
-30.4
I2
-43.1
44.5
T2
I3
37.6
52.0
T3
I4
20.8
45.0
T4
I5
-79.9
I6
-81.4
31.9
I7
W^II
G (kJ/mol)
45
25
5
-15
-35
-55
-75
-95
0.0
I0
I1
-44.9
I2
-51.9
29.9
T2
34.7
I3
45.7
T3
I4
22.7
42.6
T4
I5
-92.2
I6
-82.2
31.9
I7
Au^III
G (kJ/mol)
30
10
-10
-30
-50
-70
-90
-110
0.0
I0
I1
-94.6
I2
-66.6
3.7
T2
I3
0.1
27.7
T3
I4
-23.3
9.1
T4
I5
-106.5
I6
-84.0
31.9
I7
Zn^II
G (kJ/mol)
35
15
-5
-25
-45
-65
-85
-105
0.0
I0
I1
-43.1
I2
-55.8
26.9
T2
I3
28.4
41.2
T3
I4
21.2
37.5
T4
I5
-99.3
I6
-89.2
31.9
I7
Ni^II
G (kJ/mol)
30
10
-10
-30
-50
-70
-90
-110
0.0
I0
I1
-48.3
I2
-65.4
9.0
T2
I3
3.1
23.3
T3
I4
0.9
15.5
T4
I5
-105.1
I6
-92.2
31.9
I7
Mn^II
G (kJ/mol)
30
10
-10
-30
-50
-70
-90
-110
0.0
I0
I1
-52.7
I2
-69.1
29.0
T2
I3
24.5
37.5
T3
I4
16.0
31.5
T4
I5
-102.5
I6
-93.4
31.9
I7

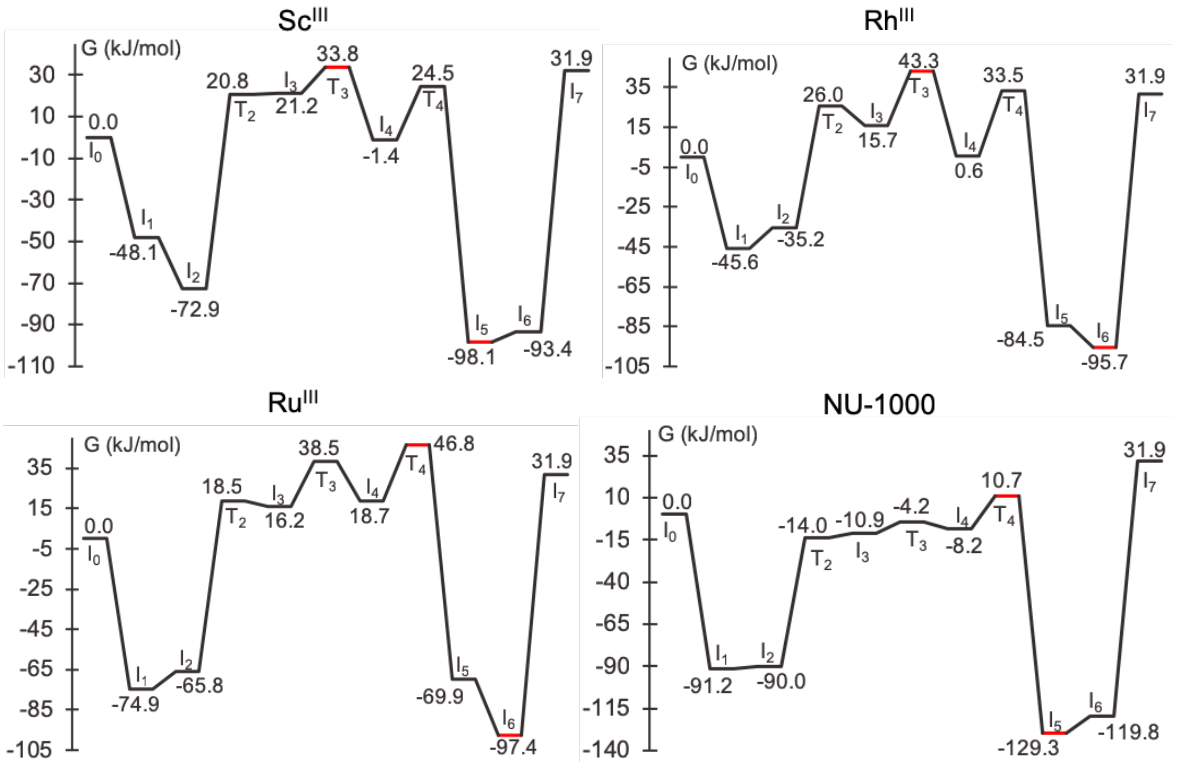


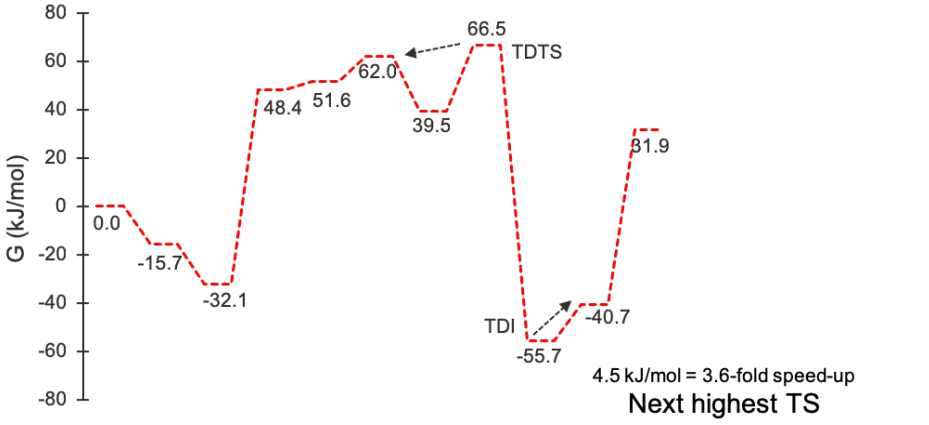
Figure D.8. Reaction free energy profiles (at T = 298.15 K and P = 1 atm) for gas-phase sarin hydrolysis catalyzed by single-atom M-NU-1000 systems, as well as unfunctionalized NU-1000. The TDTS (turnover-frequency-determining transition state) and TDI (turnover-frequency-determining intermediate) for each system are colored red.

The labels I_k and T_j in Fig. D.8 correspond to the terminology used in the energetic span model for intermediates and transition states, respectively:

I_0: M–OH + Sarin + H_2O
I_1: M–Sarin + H_2O
I_2: Reactant Complex
T_2: TS_{nuc}
I_3: INT_1
T_3: TS_{Berry}
I_4: INT_2
T_4: TS_{eli}
I_5: Product Complex
I_6: M–IMPA + HF
I_7: M–OH + IMPA + HF

D.10 Influence of TDTS/TDI on Absolute TOFs for Cu^{III}-NU-1000

To get an idea of how sensitive the absolute TOFs are to changes in energy, we calculated absolute TOFs for the Cu^{III}-NU-1000 system using different free energy values for the TDTS and TDI. As seen in Fig. D.9, lowering the TDTS free energy to that of the next highest TS and raising the TDI free energy to that of the next lowest INT leads to a >750x increase in predicted |TOF|. Thus, a small inaccuracy in the calculation of the energetic span results in a large error in the absolute TOF calculation, which is why we used relative TOFs to compare the catalysts in our study.



15 kJ/mol = 213-fold speed-up
Next lowest INT

TDI	TDTS	|TOF| (s^{-1})
-55.7	66.5	2.03×10^{-9}
-52.7	66.5	6.77×10^{-9}
-49.7	66.5	2.23×10^{-8}
-46.7	66.5	7.06×10^{-8}
-43.7	66.5	1.99×10^{-7}
-40.7	66.5	4.33×10^{-7}

TDI	TDTS	|TOF| (s^{-1})
-55.7	65.0	3.38×10^{-9}
-52.7	65.0	1.13×10^{-8}
-49.7	65.0	3.71×10^{-8}
-46.7	65.0	1.17×10^{-7}
-43.7	65.0	3.31×10^{-7}
-40.7	65.0	7.21×10^{-7}

TDI	TDTS	|TOF| (s^{-1})
-55.7	63.5	5.20×10^{-9}
-52.7	63.5	1.73×10^{-8}
-49.7	63.5	5.71×10^{-8}
-46.7	63.5	1.81×10^{-7}
-43.7	63.5	5.08×10^{-7}
-40.7	63.5	1.11×10^{-6}

TDI	TDTS	|TOF| (s^{-1})
-55.7	62.0	7.36×10^{-9}
-52.7	62.0	2.45×10^{-8}
-49.7	62.0	8.08×10^{-8}
-46.7	62.0	2.56×10^{-7}
-43.7	62.0	7.20×10^{-7}
-40.7	62.0	1.57×10^{-6}

Figure D.9. Calculated absolute TOFs for gas-phase sarin hydrolysis catalyzed by Cu^{III}-NU-1000 using different free energy values for the TDTS and TDI.

D.11 Influence of Temperature on Absolute TOFs for Cu^{III}-NU-1000

Table D.6. Calculated absolute TOFs for gas-phase sarin hydrolysis catalyzed by Cu^{III}-NU-1000 at P = 1 atm and temperatures ranging from T = 298.15 K to T = 500.15 K.

T (K)	|TOF| (s^{-1})
298.15	2.03×10^{-9}
308.15	9.21×10^{-9}
318.15	3.79×10^{-8}
344.15	1.00×10^{-6}
370.15	1.52×10^{-5}
396.15	1.34×10^{-4}
422.15	7.05×10^{-4}
448.15	2.46×10^{-3}
474.15	6.72×10^{-3}
500.15	1.56×10^{-2}

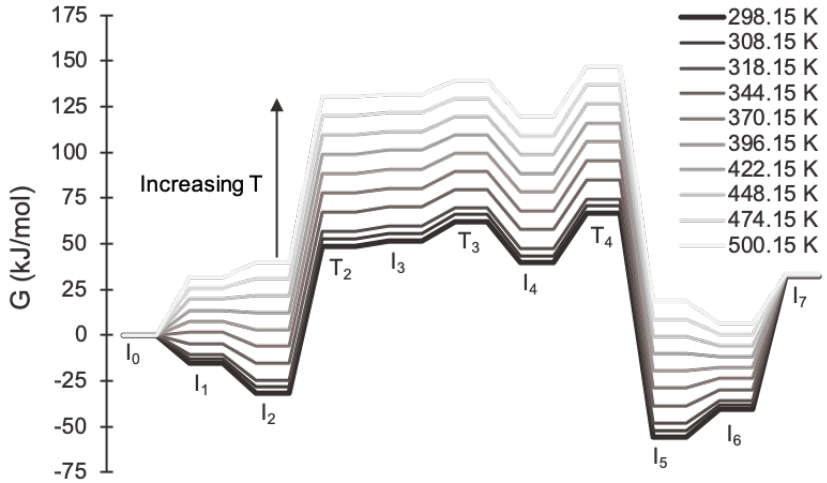


Figure D.10. Reaction free energy profiles for gas-phase sarin hydrolysis catalyzed by Cu^{III}-NU-1000 at P = 1 atm and temperatures ranging from 298.15 K to 500.15 K. The labels I_k and T_j correspond to the terminology used in the energetic span model for intermediates and transition states, respectively.

D.12 Raw Energy Values for All Spin States for NU-1000 and 19 M-NU-1000 Catalysts

In the tables below, the raw energy values are given for every species involved in the reactions catalyzed by NU-1000 and 19 M-NU-1000 catalysts.

Table D.7. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Pd^{IV}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Pd^{IV} ($3d^6$)	E ($S^2 = 6$)	E ($S^2 = 2$)	E ($S^2 = 0$)
Pd^{IV}–OH	-3210.8430750	-3210.8651153	-3210.8750646
Pd^{IV}–Sarin	-3961.1998650	-3961.2081855	-3961.2302080
Reactant Complex	-4037.6609152	-4037.7095648	-4037.6957955
TS_{nuc}	-4037.6248014	-4037.6776932	-4037.6603385
INT_1	-4037.6411866	-4037.6708251	-4037.6720228
TS_{Berry}	-4037.6313367	-4037.6375535	-4037.6523767
INT_2	-4037.6508120	-4037.6789776	-4037.6792904
TS_{eli}	-4037.6321329	-4037.6712782	-4037.6613149
Product Complex	-4037.6754169	-4037.7180369	-4037.7056577
Pd^{IV}–IMPA	-3937.1907149	-3937.1938282	-3937.2037384

Table D.8. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Pd^{IV}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Pd^{IV} ($3d^6$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Pd^{IV}–OH	0	-3210.875065	-3210.445848	-3210.572717	0.0	0.0
Pd^{IV}–Sarin	0	-3961.230208	-3960.642428	-3960.792716	-53.9	7.1
Reactant Complex	2	-4037.709565	-4037.093603	-4037.251200	-140.6	-42.6
TS_{nuc}	2	-4037.677693	-4037.066459	-4037.219851	-69.3	39.7
INT_1	0	-4037.672023	-4037.055806	-4037.208759	-41.4	68.8
TS_{Berry}	0	-4037.652377	-4037.037030	-4037.187845	7.9	123.7
INT_2	0	-4037.679290	-4037.062817	-4037.215546	-59.8	51.0
TS_{eli}	2	-4037.671278	-4037.058491	-4037.212474	-48.4	59.1
Product Complex	2	-4037.718037	-4037.102963	-4037.259902	-165.2	-65.5
Pd^{IV}–IMPA	0	-3937.203738	-3936.603769	-3936.753896	-51.0	14.9

Table D.9. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Co^{III}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Co^{III} ($3d^6$)	E ($S^2 = 6$)	E ($S^2 = 2$)	E ($S^2 = 0$)
Co^{III}–OH	-4466.3950953	-4466.3728753	-4466.3585509
Co^{III}–Sarin	-5216.7581002	-5216.7487010	-5216.7383390
Reactant Complex	-5293.2190805	-5293.2224817	-5293.2097424
TS_{nuc}	-5293.1841795	-5293.1911892	-5293.1703468
INT_1	-5293.1867137	-5293.1957839	-5293.1767376
TS_{Berry}	-5293.1852953	-5293.1774286	-5293.1693849
INT_2	-5293.1915577	-5293.1930882	-5293.1804416
TS_{eli}	-5293.1757665	-5293.1803865	-5293.1664801
Product Complex	-5293.2222821	-5293.2322124	-5293.2101493
Co^{III}–IMPA	-5192.7423753	-5192.7308345	-5192.7275449

Table D.10. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Co^{III}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Co^{III} ($3d^6$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Co^{III}–OH	6	-4466.3950953	-4465.9552110	-4466.0848640	0.0	0.0
Co^{III}–Sarin	6	-5216.7581002	-5216.1590530	-5216.3122930	-72.9	-12.4
Reactant Complex	2	-5293.2224817	-5292.5946680	-5292.7517030	-118.8	-12.0
TS_{nuc}	2	-5293.1911892	-5292.5673600	-5292.7196250	-47.1	72.2
INT_1	2	-5293.1957839	-5292.5677380	-5292.7209010	-48.1	68.8
TS_{Berry}	6	-5293.1852953	-5292.5591880	-5292.7119710	-25.7	92.3
INT_2	2	-5293.1930882	-5292.5647490	-5292.7171030	-40.3	78.8
TS_{eli}	2	-5293.1803865	-5292.5552280	-5292.7076500	-15.3	103.6
Product Complex	2	-5293.2322124	-5292.6048660	-5292.7605460	-145.6	-35.3
Co^{III}–IMPA	6	-5192.7423753	-5192.1302080	-5192.2813360	-95.8	-25.3

Table D.11. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by W^{IV}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = W^{IV} ($3d^2$)	E (S^2 = 2)	E (S^2 = 0)
W^{IV}–OH	-3150.2079499	-3150.2476876
W^{IV}–Sarin	-3900.5945250	-3900.6176382
Reactant Complex	-3977.0482218	-3977.0730616
TS_{nuc}	-3977.0066691	-3977.0402773
INT_1	-3977.0148274	-3977.0400134
TS_{Berry}	-3977.0124686	-3977.0384787
INT_2	-3977.0276113	-3977.0608454
TS_{eli}	-3977.0114971	-3977.0408096
Product Complex	-3977.0561772	-3977.0836774
W^{IV}–IMPA	-3876.5687815	-3876.5926877

Table D.12. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by W^{IV}–NU-1000, using the most stable spin state for each species in the mechanism.

M = W^{IV} ($3d^2$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
W^{IV}–OH	0	-3150.2476876	-3149.8181820	-3149.9458440	0.0	0.0
W^{IV}–Sarin	0	-3900.6176382	-3900.0292540	-3900.1795090	-91.9	-28.8
Reactant Complex	0	-3977.0730616	-3976.4567240	-3976.6142210	-116.4	-16.1
TS_{nuc}	0	-3977.0402773	-3976.4266580	-3976.5772960	-37.5	80.9
INT_1	0	-3977.0400134	-3976.4236040	-3976.5777590	-29.5	79.6
TS_{Berry}	0	-3977.0384787	-3976.4224350	-3976.5748270	-26.4	87.3
INT_2	0	-3977.0608454	-3976.4428570	-3976.5955310	-80.0	33.0
TS_{eli}	0	-3977.0408096	-3976.4267360	-3976.5790950	-37.7	76.1
Product Complex	0	-3977.0836774	-3976.4670760	-3976.6231550	-143.6	-39.5
W^{IV}–IMPA	0	-3876.5926877	-3875.9917390	-3876.1432770	-92.0	-27.8

Table D.13. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Cu^{III}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = Cu^{III} ($3d^8$)	E ($S^2 = 2$)	E ($S^2 = 0$)
Cu^{III}–OH	-4724.0438253	-4724.0353133
Cu^{III}–Sarin	-5474.4110371	-5474.4130803
Reactant Complex	-5550.8773118	-5550.8793263
TS_{nuc}	-5550.8386628	-5550.8496361
INT_1	-5550.8396075	-5550.8511393
TS_{Berry}	-5550.8365357	-5550.8453644
INT_2	-5550.8491326	-5550.8566456
TS_{eli}	-5550.8345063	-5550.8429633
Product Complex	-5550.8814257	-5550.8884917
Cu^{III}–IMPA	-5450.3935131	-5450.3963485

Table D.14. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Cu^{III}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Cu^{III} ($3d^8$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Cu^{III}–OH	2	-4724.0438253	-4723.6042660	-4723.7333960	0.0	0.0
Cu^{III}–Sarin	0	-5474.4130803	-5473.8131800	-5473.9620620	-86.2	-15.7
Reactant Complex	0	-5550.8793263	-5550.2516780	-5550.4078700	-139.7	-32.1
TS_{nuc}	0	-5550.8496361	-5550.2263990	-5550.3772100	-73.3	48.4
INT_1	0	-5550.8511393	-5550.2242080	-5550.3759970	-67.6	51.6
TS_{Berry}	0	-5550.8453644	-5550.2185460	-5550.3720470	-52.7	62.0
INT_2	0	-5550.8566456	-5550.2285230	-5550.3806010	-78.9	39.5
TS_{eli}	0	-5550.8429633	-5550.2184770	-5550.3703020	-52.5	66.5
Product Complex	0	-5550.8884917	-5550.2616480	-5550.4168710	-165.9	-55.7
Cu^{III}–IMPA	0	-5450.3963485	-5449.7850960	-5449.9357520	-111.1	-40.7

Table D.15. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Pd^{II}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = Pd^{II} ($3d^8$)	E ($S^2 = 2$)	E ($S^2 = 0$)
Pd^{II}–OH	-3212.1766871	-3212.1852764
Pd^{II}–Sarin	-3962.5330759	-3962.5664566
Reactant Complex	-4038.9980691	-4039.0019177
TS_{nuc}	-4038.9675883	-4038.9840283
INT_1	-4038.9765386	-4038.9919381
TS_{Berry}	-4038.9688104	-4038.9842440
INT_2	-4038.9821156	-4038.9993732
TS_{eli}	-4038.9726789	-4038.9896402
Product Complex	-4039.0193253	-4039.0352409
Pd^{II}–IMPA	-3938.5260573	-3938.5369233

Table D.16. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Pd^{II}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Pd^{II} ($3d^8$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Pd^{II}–OH	0	-3212.1852764	-3211.7327920	-3211.8615360	0.0	0.0
Pd^{II}–Sarin	0	-3962.5664566	-3961.9550800	-3962.1086750	-121.3	-64.2
Reactant Complex	0	-4039.0019177	-4038.3629140	-4038.5214090	-94.3	6.2
TS_{nuc}	0	-4038.9840283	-4038.3485480	-4038.5017210	-56.6	57.9
INT_1	0	-4038.9919381	-4038.3525790	-4038.5063930	-67.2	45.7
TS_{Berry}	0	-4038.9842440	-4038.3458240	-4038.4989700	-49.4	65.2
INT_2	0	-4038.9993732	-4038.3591350	-4038.5127190	-84.4	29.1
TS_{eli}	0	-4038.9896402	-4038.3525890	-4038.5066960	-67.2	44.9
Product Complex	0	-4039.0352409	-4038.3962560	-4038.5529820	-181.8	-76.7
Pd^{II}–IMPA	0	-3938.5369233	-3937.9127310	-3938.0648890	-108.8	-43.3

Table D.17. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Fe^{III}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Fe^{III} ($3d^6$)	E (S^2 = 8.75)	E (S^2 = 3.75)	E (S^2 = 0.75)
Fe^{III}–OH	-4347.3838677	-4347.3368886	-4347.3129595
Fe^{III}–Sarin	-5097.7501282	-5097.7110450	-5097.6855229
Reactant Complex	-5174.2134160	-5174.1790995	-5174.1416041
TS_{nuc}	-5174.1850588	-5174.1496836	-5174.1154438
INT_1	-5174.1879790	-5174.1612754	-5174.1209738
TS_{Berry}	-5174.1879131	-5174.1499891	-5174.1285989
INT_2	-5174.1987323	-5174.1580019	-5174.1286996
TS_{eli}	-5174.1844163	-5174.1460121	-5174.1132674
Product Complex	-5174.2304841	-5174.1961158	-5174.1533482
Fe^{III}–IMPA	-5073.7442940	-5073.7064787	-5073.6741621

Table D.18. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Fe^{III}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Fe^{III} ($3d^6$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Fe^{III}–OH	8.75	-4347.3838677	-4346.9445730	-4347.0756670	0.0	0.0
Fe^{III}–Sarin	8.75	-5097.7501282	-5097.1514760	-5097.3056420	-81.0	-19.1
Reactant Complex	8.75	-5174.2134160	-5173.5866940	-5173.7465990	-125.8	-22.8
TS_{nuc}	8.75	-5174.1850588	-5173.5621930	-5173.7159690	-61.5	57.6
INT_1	8.75	-5174.1879790	-5173.5609910	-5173.7165210	-58.3	56.2
TS_{Berry}	8.75	-5174.1879131	-5173.5615180	-5173.7148500	-59.7	60.6
INT_2	8.75	-5174.1987323	-5173.5712170	-5173.7269140	-85.2	28.9
TS_{eli}	8.75	-5174.1844163	-5173.5604410	-5173.7159440	-56.9	57.7
Product Complex	8.75	-5174.2304841	-5173.6041710	-5173.7622890	-171.7	-64.0
Fe^{III}–IMPA	8.75	-5073.7442940	-5073.1327270	-5073.2844290	-130.3	-57.5

Table D.19. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Cu^{II}–NU-1000, using the only possible spin state (doublet) for each species in the mechanism.

M = Cu^{II} ($3d^9$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Cu^{II}–OH	0.75	-4724.7043261	-4724.2519800	-4724.3813910	0.0	0.0
Cu^{II}–Sarin	0.75	-5475.0662397	-5474.4553570	-5474.6086650	-71.7	-12.0
Reactant Complex	0.75	-5551.5307208	-5550.8913710	-5551.0506040	-118.6	-18.3
TS_{nuc}	0.75	-5551.5064524	-5550.8710680	-5551.0249240	-65.3	49.1
INT_1	0.75	-5551.5084828	-5550.8694190	-5551.0246010	-61.0	50.0
TS_{Berry}	0.75	-5551.5017716	-5550.8628220	-5551.0166670	-43.7	70.8
INT_2	0.75	-5551.5130234	-5550.8727180	-5551.0281660	-69.7	40.6
TS_{eli}	0.75	-5551.5029137	-5550.8658510	-5551.0209640	-51.6	59.5
Product Complex	0.75	-5551.5594887	-5550.9198550	-5551.0751950	-193.4	-82.8
Cu^{II}–IMPA	0.75	-5451.0672065	-5450.4425140	-5450.5937260	-136.6	-66.9

Table D.20. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Ce^{IV}–NU-1000, using the only possible spin state (singlet) for each species in the mechanism.

M = Ce^{IV} ($3d^0$, $3f^0$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Ce^{IV}–OH	0	-3558.3913560	-3557.9635690	-3558.0922030	0.0	0.0
Ce^{IV}–Sarin	0	-4308.7794759	-4308.1928370	-4308.3449980	-139.7	-79.0
Reactant Complex	0	-4385.2463231	-4384.6319810	-4384.7904290	-194.8	-94.5
TS_{nuc}	0	-4385.1947364	-4384.5839910	-4384.7386810	-68.8	41.4
INT_1	0	-4385.1962200	-4384.5813850	-4384.7378030	-62.0	43.7
TS_{Berry}	0	-4385.1932339	-4384.5787820	-4384.7337810	-55.2	54.3
INT_2	0	-4385.2033787	-4384.5875240	-4384.7435520	-78.1	28.6
TS_{eli}	0	-4385.1908808	-4384.5777740	-4384.7327530	-52.5	57.0
Product Complex	0	-4385.2448949	-4384.6300820	-4384.7870600	-189.8	-85.6
Ce^{IV}–IMPA	0	-4284.7552818	-4284.1560920	-4284.3081210	-141.8	-76.3

Table D.21. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Co^{II}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = Co^{II} ($3d^7$)	E (S^2 = 3.75)	E (S^2 = 0.75)
Co^{II}–OH	-4467.0131991	-4466.9750198
Co^{II}–Sarin	-5217.3860745	-5217.3509167
Reactant Complex	-5293.8508905	-5293.8077928
TS_{nuc}	-5293.8189284	-5293.7904255
INT_1	-5293.8233291	-5293.7984626
TS_{Berry}	-5293.8205395	-5293.7944577
INT_2	-5293.8294248	-5293.8042252
TS_{eli}	-5293.8191773	-5293.7941206
Product Complex	-5293.8672042	-5293.8383582
Co^{II}–IMPA	-5193.3821318	-5193.3461695

Table D.22. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Co^{II}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Co^{II} ($3d^7$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Co^{II}–OH	3.75	-4467.0131991	-4466.5613060	-4466.6926030	0.0	0.0
Co^{II}–Sarin	3.75	-5217.3860745	-5216.7748430	-5216.9293600	-98.4	-36.9
Reactant Complex	3.75	-5293.8508905	-5293.2114750	-5293.3705080	-146.9	-41.1
TS_{nuc}	3.75	-5293.8189284	-5293.1828520	-5293.3379200	-71.8	44.5
INT_1	3.75	-5293.8233291	-5293.1827860	-5293.3384420	-71.6	43.1
TS_{Berry}	3.75	-5293.8205395	-5293.1805140	-5293.3339480	-65.7	54.9
INT_2	3.75	-5293.8294248	-5293.1884000	-5293.3443980	-86.4	27.4
TS_{eli}	3.75	-5293.8191773	-5293.1808970	-5293.3362680	-66.7	48.8
Product Complex	3.75	-5293.8672042	-5293.2272900	-5293.3849970	-188.5	-79.1
Co^{II}–IMPA	3.75	-5193.3821318	-5192.7572480	-5192.9090440	-150.8	-77.7

Table D.23. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Ti^{IV}–NU-1000 (*formate cluster model*, **OiPr** pathway), using the only possible spin state (singlet) for each species in the mechanism.

M = Ti^{IV} ($3d^0$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Ti^{IV}–OH	0	-3932.6207000	-3932.1914030	-3932.3175390	0.0	0.0
Ti^{IV}–Sarin	0	-4683.0167169	-4682.4285820	-4682.5780080	-160.4	-99.2
Reactant Complex	0	-4759.4713551	-4758.8545710	-4759.0101780	-181.1	-79.8
TS_{nuc}	0	-4759.4320839	-4758.8191570	-4758.9700490	-88.1	25.6
INT_1	0	-4759.4335375	-4758.8160750	-4758.9671920	-80.0	33.1
TS_{Berry}	0	-4759.4303222	-4758.8137610	-4758.9638460	-73.9	41.9
INT_2	0	-4759.4438831	-4758.8261690	-4758.9780710	-106.5	4.5
TS_{eli}	0	-4759.4282916	-4758.8137040	-4758.9641110	-73.8	41.2
Product Complex	0	-4759.4787509	-4758.8618290	-4759.0144610	-200.1	-91.0
Ti^{IV}–IMPA	0	-4658.9878242	-4658.3867040	-4658.5347750	-149.1	-79.8

Table D.24. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Ti^{IV}–NU-1000 (*benzoate cluster model*, **OiPr** pathway), using the only possible spin state (singlet) for each species in the mechanism.

M = Ti^{IV} ($3d^0$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Ti^{IV}–OH	0	-4856.1180405	-4855.5006950	-4855.6273380	0.0	0.0
Ti^{IV}–Sarin	0	-5606.5171789	-5605.7412980	-5605.8911040	-169.4	-107.8
Reactant Complex	0	-5682.9662768	-5682.1619510	-5682.3174410	-176.0	-73.1
TS_{nuc}	0	-5682.9326569	-5682.1313720	-5682.2820860	-95.8	19.7
INT_1	0	-5682.9330755	-5682.1284520	-5682.2802810	-88.1	24.4
TS_{Berry}	0	-5682.9298327	-5682.1255400	-5682.2765260	-80.5	34.3
INT_2	0	-5682.9436789	-5682.1382160	-5682.2906720	-113.7	-2.8
TS_{eli}	0	-5682.9285579	-5682.1262420	-5682.2771270	-82.3	32.7
Product Complex	0	-5682.9760505	-5682.1717100	-5682.3254850	-201.7	-94.2
Ti^{IV}–IMPA	0	-5582.4876430	-5581.6985500	-5581.8472340	-155.8	-86.8

Table D.25. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Ti^{IV}–NU-1000 (*formate cluster model*, **CH_3** pathway), using the only possible spin state (singlet) for each species in the mechanism.

M = Ti^{IV} ($3d^0$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Ti^{IV}–OH	0	-3932.6207000	-3932.1914030	-3932.3175390	0.0	0.0
Ti^{IV}–Sarin	0	-4683.0167169	-4682.4285820	-4682.5780080	-160.4	-99.2
Reactant Complex	0	-4759.4728164	-4758.8563470	-4759.0112830	-185.7	-82.7
TS_{nuc}	0	-4759.4329460	-4758.8203160	-4758.9704170	-91.1	24.6
INT_1	0	-4759.4334479	-4758.8159420	-4758.9665270	-79.6	34.8
TS_{Berry}	0	-4759.4320945	-4758.8154100	-4758.9646840	-78.3	39.7
INT_2	0	-4759.4438831	-4758.8261690	-4758.9780710	-106.5	4.5
TS_{eli}	0	-4759.4282916	-4758.8137040	-4758.9641110	-73.8	41.2
Product Complex	0	-4759.4787509	-4758.8618290	-4759.0144610	-200.1	-91.0
Ti^{IV}–IMPA	0	-4658.9878242	-4658.3867040	-4658.5347750	-149.1	-79.8

Table D.26. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Ti^{IV}–NU-1000 (*formate cluster model*, **F** pathway), using the only possible spin state (singlet) for each species in the mechanism.

M = Ti^{IV} ($3d^0$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Ti^{IV}–OH	0	-3932.6207000	-3932.1914030	-3932.3175390	0.0	0.0
Ti^{IV}–Sarin	0	-4683.0167169	-4682.4285820	-4682.5780080	-160.4	-99.2
Reactant Complex	0	-4759.4807358	-4758.8638470	-4759.0182030	-205.4	-100.9
TS_{nuc}	0	-4759.4380729	-4758.8255120	-4758.9766100	-104.8	8.3
INT_1	0	-4759.4384851	-4758.8223280	-4758.9739820	-96.4	15.2
Ti^{IV}–IMPA	0	-4658.9878242	-4658.3867040	-4658.5347750	-149.1	-79.8

Table D.27. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Fe^{II}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Fe^{II} ($3d^6$)	E (S^2 = 6)	E (S^2 = 2)	E (S^2 = 0)
Fe^{II}–OH	-4347.9700167	-4347.9174260	-4347.8828451
Fe^{II}–Sarin	-5098.3414529	-5098.2864624	-5098.2550917
Reactant Complex	-5174.8069331	-5174.7416793	-5174.7176193
TS_{nuc}	-5174.7760746	-5174.7253175	-5174.7037578
INT_1	-5174.7804670	-5174.7342305	-5174.7147384
TS_{Berry}	-5174.7751548	-5174.7304964	-5174.7097325
INT_2	-5174.7875306	-5174.7392591	-5174.7184799
TS_{eli}	-5174.7766519	-5174.7295917	-5174.7094851
Product Complex	-5174.8240315	-5174.7740510	-5174.7544111
Fe^{II}–IMPA	-5074.3394859	-5074.2860639	-5074.2522963

Table D.28. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Fe^{II}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Fe^{II} ($3d^6$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Fe^{II}–OH	6	-4347.9700167	-4347.5182660	-4347.6497590	0.0	0.0
Fe^{II}–Sarin	6	-5098.3414529	-5097.7302150	-5097.8840470	-94.2	-30.4
Reactant Complex	6	-5174.8069331	-5174.1680200	-5174.3284210	-145.8	-43.1
TS_{nuc}	6	-5174.7760746	-5174.1401120	-5174.2950470	-72.6	44.5
INT_1	6	-5174.7804670	-5174.1409350	-5174.2976910	-74.7	37.6
TS_{Berry}	6	-5174.7751548	-5174.1363090	-5174.2922020	-62.6	52.0
INT_2	6	-5174.7875306	-5174.1473100	-5174.3040700	-91.5	20.8
TS_{eli}	6	-5174.7766519	-5174.1391050	-5174.2948690	-69.9	45.0
Product Complex	6	-5174.8240315	-5174.1842860	-5174.3424530	-188.6	-79.9
Fe^{II}–IMPA	6	-5074.3394859	-5073.7151120	-5073.8676140	-153.2	-81.4

Table D.29. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by W^{II}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = W^{II} ($3d^4$)	E ($S^2 = 6$)	E ($S^2 = 2$)	E ($S^2 = 0$)
W^{II}–OH	-3151.3406086	-3151.3425521	-3151.3261634
W^{II}–Sarin	-3901.7078305	-3901.7169470	-3901.6888250
Reactant Complex	-3978.1756922	-3978.1815055	-3978.1479653
TS_{nuc}	-3978.1500894	-3978.1498218	-3978.1144369
INT_1	-3978.1518231	-3978.1481804	-3978.1220840
TS_{Berry}	-3978.1486631	-3978.1398636	-3978.1190754
INT_2	-3978.1572687	-3978.1560926	-3978.1397754
TS_{eli}	-3978.1469519	-3978.1401228	-3978.1169050
Product Complex	-3978.1976799	-3978.1855234	-3978.1590634
W^{II}–IMPA	-3877.7077578	-3877.6983433	-3877.6754308

Table D.30. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by W^{II}–NU-1000, using the most stable spin state for each species in the mechanism.

M = W^{II} ($3d^4$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
W^{II}–OH	2	-3151.3425521	-3150.8910990	-3151.0221010	0.0	0.0
W^{II}–Sarin	2	-3901.7169470	-3901.1073240	-3901.2618870	-105.4	-44.9
Reactant Complex	2	-3978.1815055	-3977.5439720	-3977.7041010	-154.0	-51.9
TS_{nuc}	6	-3978.1500894	-3977.5159220	-3977.6729680	-80.4	29.9
INT_1	6	-3978.1518231	-3977.5133720	-3977.6711450	-73.7	34.7
TS_{Berry}	6	-3978.1486631	-3977.5105990	-3977.6669450	-66.4	45.7
INT_2	6	-3978.1572687	-3977.5179170	-3977.6756910	-85.6	22.7
TS_{eli}	6	-3978.1469519	-3977.5106130	-3977.6681400	-66.5	42.6
Product Complex	6	-3978.1976799	-3977.5593220	-3977.7194730	-194.3	-92.2
W^{II}–IMPA	6	-3877.7077578	-3877.0850270	-3877.2402750	-145.5	-82.2

Table D.31. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Au^{III}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = Au^{III} ($3d^8$)	E ($S^2 = 2$)	E ($S^2 = 0$)
Au^{III}–OH	-3219.3132699	-3219.3215683
Au^{III}–Sarin	-3969.6714759	-3969.7170265
Reactant Complex	-4046.1271143	-4046.1656189
TS_{nuc}	-4046.0865172	-4046.1414271
INT_1	-4046.0975496	-4046.1448286
TS_{Berry}	-4046.0928477	-4046.1363846
INT_2	-4046.1115276	-4046.1563074
TS_{eli}	-4046.1017434	-4046.1398804
Product Complex	-4046.1404225	-4046.1824970
Au^{III}–IMPA	-3945.6567854	-3945.6890300

Table D.32. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Au^{III}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Au^{III} ($3d^8$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Au^{III}–OH	0	-3219.3215683	-3218.8814720	-3219.0099910	0.0	0.0
Au^{III}–Sarin	0	-3969.7170265	-3969.1173310	-3969.2687080	-157.0	-94.6
Reactant Complex	0	-4046.1656189	-4045.5387790	-4045.6976150	-165.7	-66.6
TS_{nuc}	0	-4046.1414271	-4045.5177290	-4045.6708310	-110.4	3.7
INT_1	0	-4046.1448286	-4045.5181310	-4045.6721870	-111.5	0.1
TS_{Berry}	0	-4046.1363846	-4045.5102260	-4045.6617010	-90.7	27.7
INT_2	0	-4046.1563074	-4045.5289640	-4045.6811000	-139.9	-23.3
TS_{eli}	0	-4046.1398804	-4045.5164630	-4045.6687660	-107.1	9.1
Product Complex	0	-4046.1824970	-4045.5564230	-4045.7127920	-212.0	-106.5
Au^{III}–IMPA	0	-3945.6890300	-3945.0778320	-3945.2288200	-151.9	-84.0

Table D.33. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Zn^{II}–NU-1000, using the only possible spin state (singlet) for each species in the mechanism.

M = Zn^{II} ($3d^{10}$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Zn^{II}–OH	0	-4863.6096356	-4863.1569020	-4863.2853790	0.0	0.0
Zn^{II}–Sarin	0	-5613.9848854	-5613.3732770	-5613.5245050	-105.8	-43.1
Reactant Complex	0	-5690.4512550	-5689.8115120	-5689.9688770	-158.6	-55.8
TS_{nuc}	0	-5690.4197551	-5689.7834950	-5689.9373830	-85.0	26.9
INT_1	0	-5690.4219874	-5689.7817800	-5689.9368280	-80.5	28.4
TS_{Berry}	0	-5690.4187246	-5689.7788060	-5689.9319270	-72.7	41.2
INT_2	0	-5690.4258424	-5689.7846660	-5689.9395470	-88.1	21.2
TS_{eli}	0	-5690.4175693	-5689.7792330	-5689.9333620	-73.9	37.5
Product Complex	0	-5690.4684910	-5689.8285450	-5689.9854440	-203.3	-99.3
Zn^{II}–IMPA	0	-5589.9792838	-5589.3548180	-5589.5061860	-156.0	-89.2

Table D.34. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Ni^{II}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = Ni^{II} ($3d^8$)	E (S^2 = 2)	E (S^2 = 0)
Ni^{II}–OH	-4592.5469644	-4592.5193883
Ni^{II}–Sarin	-5342.9221573	-5342.8993843
Reactant Complex	-5419.3909122	-5419.3391699
TS_{nuc}	-5419.3621548	-5419.3187816
INT_1	-5419.3674778	-5419.3210980
TS_{Berry}	-5419.3603872	-5419.3145374
INT_2	-5419.3695754	-5419.3232403
TS_{eli}	-5419.3604452	-5419.3134641
Product Complex	-5419.4053964	-5419.3602855
Ni^{II}–IMPA	-5318.9161992	-5318.8966087

Table D.35. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Ni^{II}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Ni^{II} ($3d^8$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Ni^{II}–OH	2	-4592.5469644	-4592.0940640	-4592.2228260	0.0	0.0
Ni^{II}–Sarin	2	-5342.9221573	-5342.3112670	-5342.4639350	-108.0	-48.3
Reactant Complex	2	-5419.3909122	-5418.7515980	-5418.9099840	-166.3	-65.4
TS_{nuc}	2	-5419.3621548	-5418.7267500	-5418.8816560	-101.0	9.0
INT_1	2	-5419.3674778	-5418.7280910	-5418.8838910	-104.6	3.1
TS_{Berry}	2	-5419.3603872	-5418.7214650	-5418.8761880	-87.2	23.3
INT_2	2	-5419.3695754	-5418.7291750	-5418.8847330	-107.4	0.9
TS_{eli}	2	-5419.3604452	-5418.7230450	-5418.8791640	-91.3	15.5
Product Complex	2	-5419.4053964	-5418.7661630	-5418.9250980	-204.5	-105.1
Ni^{II}–IMPA	2	-5318.9161992	-5318.2921760	-5318.4448100	-156.5	-92.2

Table D.36. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Mn^{II}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Mn^{II} ($3d^5$)	E (S^2 = 8.75)	E (S^2 = 3.75)	E (S^2 = 0.75)
Mn^{II}–OH	-4235.2866729	-4235.2225562	-4235.1820103
Mn^{II}–Sarin	-4985.6654078	-4985.5985106	-4985.5521447
Reactant Complex	-5062.1340502	-5062.0660409	-5062.0065010
TS_{nuc}	-5062.0952865	-5062.0204075	-5061.9891755
INT_1	-5062.1011292	-5062.0396015	-5061.9980681
TS_{Berry}	-5062.0967176	-5062.0185808	-5061.9941940
INT_2	-5062.1047038	-5062.0430331	-5062.0033164
TS_{eli}	-5062.0968034	-5062.0285219	-5061.9924925
Product Complex	-5062.1465775	-5062.0741293	-5062.0364385
Mn^{II}–IMPA	-4961.6590769	-4961.5940189	-4961.5505515

Table D.37. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Mn^{II}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Mn^{II} ($3d^5$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Mn^{II}–OH	8.75	-4235.2866729	-4234.8347940	-4234.9654060	0.0	0.0
Mn^{II}–Sarin	8.75	-4985.6654078	-4985.0544110	-4985.2081550	-114.3	-52.7
Reactant Complex	8.75	-5062.1340502	-5061.4951390	-5061.6539610	-173.7	-69.1
TS_{nuc}	8.75	-5062.0952865	-5061.4603450	-5061.6166100	-82.3	29.0
INT_1	8.75	-5062.1011292	-5061.4613980	-5061.6183190	-85.1	24.5
TS_{Berry}	8.75	-5062.0967176	-5061.4579500	-5061.6133840	-76.0	37.5
INT_2	8.75	-5062.1047038	-5061.4641230	-5061.6215510	-92.2	16.0
TS_{eli}	8.75	-5062.0968034	-5061.4586220	-5061.6156490	-77.8	31.5
Product Complex	8.75	-5062.1465775	-5061.5069810	-5061.6666870	-204.7	-102.5
Mn^{II}–IMPA	8.75	-4961.6590769	-4961.0348730	-4961.1878170	-161.7	-93.4

Table D.38. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Sc^{III}–NU-1000, using the only possible spin state (singlet) for each species in the mechanism.

M = Sc^{III} ($3d^0$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Sc^{III}–OH	0	-3844.5298542	-3844.0900380	-3844.2189460	0.0	0.0
Sc^{III}–Sarin	0	-4594.9063759	-4594.3077070	-4594.4599530	-109.2	-48.1
Reactant Complex	0	-4671.3798923	-4670.7531600	-4670.9089610	-180.9	-72.9
TS_{nuc}	0	-4671.3442646	-4670.7211120	-4670.8732730	-96.8	20.8
INT_1	0	-4671.3475708	-4670.7202550	-4670.8731270	-94.6	21.2
TS_{Berry}	0	-4671.3424592	-4670.7159420	-4670.8683260	-83.2	33.8
INT_2	0	-4671.3563800	-4670.7284830	-4670.8817290	-116.2	-1.4
TS_{eli}	0	-4671.3438951	-4670.7191730	-4670.8718520	-91.7	24.5
Product Complex	0	-4671.3896934	-4670.7628090	-4670.9185460	-206.3	-98.1
Sc^{III}–IMPA	0	-4570.9029133	-4570.2913820	-4570.4413880	-165.0	-93.4

Table D.39. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Rh^{III}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Rh^{III} ($3d^6$)	E ($S^2 = 6$)	E ($S^2 = 2$)	E ($S^2 = 0$)
Rh^{III}–OH	-3194.2009712	-3194.2112161	-3194.2163645
Rh^{III}–Sarin	-3944.5629644	-3944.5898047	-3944.5901100
Reactant Complex	-4021.0174055	-4021.0464189	-4021.0454991
TS_{nuc}	-4020.9891287	-4021.0211057	-4021.0269678
INT_1	-4020.9953698	-4021.0289580	-4021.0336090
TS_{Berry}	-4020.9917920	-4021.0208676	-4021.0252362
INT_2	-4021.0046047	-4021.0394188	-4021.0385238
TS_{eli}	-4020.9848747	-4021.0222971	-4021.0249823
Product Complex	-4021.0306441	-4021.0649702	-4021.0693722
Rh^{III}–IMPA	-3920.5419303	-3920.5753404	-3920.5884757

Table D.40. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Rh^{III}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Rh^{III} ($3d^6$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Rh^{III}–OH	0	-3194.2163645	-3193.7752860	-3193.9020940	0.0	0.0
Rh^{III}–Sarin	0	-3944.5901100	-3943.9907090	-3944.1421690	-103.3	-45.6
Reactant Complex	2	-4021.0464189	-4020.4194460	-4020.5777570	-131.2	-35.2
TS_{nuc}	0	-4021.0269678	-4020.4024180	-4020.5544550	-86.5	26.0
INT_1	0	-4021.0336090	-4020.4052220	-4020.5583610	-93.8	15.7
TS_{Berry}	0	-4021.0252362	-4020.3976260	-4020.5478700	-73.9	43.3
INT_2	2	-4021.0394188	-4020.4110330	-4020.5641020	-109.1	0.6
TS_{eli}	0	-4021.0249823	-4020.3995400	-4020.5515660	-78.9	33.5
Product Complex	0	-4021.0693722	-4020.4414770	-4020.5965290	-189.0	-84.5
Rh^{III}–IMPA	0	-3920.5884757	-3919.9768370	-3920.1253940	-165.5	-95.7

Table D.41. The SCF energies (in hartrees) for every species involved in the reaction catalyzed by Ru^{III}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Ru^{III} ($3d^5$)	E (S^2 = 8.75)	E (S^2 = 3.75)	E (S^2 = 0.75)
Ru^{III}–OH	-3178.5470301	-3178.5614276	-3178.5634276
Ru^{III}–Sarin	-3928.9134596	-3928.9473897	-3928.9391351
Reactant Complex	-4005.3726351	-4005.4058563	-4005.3919897
TS_{nuc}	-4005.3450139	-4005.3765277	-4005.3687907
INT_1	-4005.3501319	-4005.3799519	-4005.3750821
TS_{Berry}	-4005.3513354	-4005.3693849	-4005.3620610
INT_2	-4005.3608397	-4005.3806617	-4005.3828561
TS_{eli}	-4005.3489403	-4005.3633050	-4005.3676791
Product Complex	-4005.3962134	-4005.4108066	-4005.4113244
Ru^{III}–IMPA	-3904.9079100	-3904.9345978	-3904.9257275

Table D.42. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by Ru^{III}–NU-1000, using the most stable spin state for each species in the mechanism.

M = Ru^{III} ($3d^5$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Ru^{III}–OH	0.75	-3178.5634276	-3178.1223400	-3178.2501250	0.0	0.0
Ru^{III}–Sarin	3.75	-3928.9473897	-3928.3485580	-3928.5013330	-131.7	-74.9
Reactant Complex	3.75	-4005.4058563	-4004.7789710	-4004.9374400	-163.9	-65.8
TS_{nuc}	3.75	-4005.3765277	-4004.7528800	-4004.9053200	-95.4	18.5
INT_1	3.75	-4005.3799519	-4004.7522090	-4004.9062230	-93.6	16.2
TS_{Berry}	3.75	-4005.3693849	-4004.7431290	-4004.8977120	-69.8	38.5
INT_2	0.75	-4005.3828561	-4004.7535880	-4004.9052680	-97.3	18.7
TS_{eli}	0.75	-4005.3676791	-4004.7424640	-4004.8945680	-68.1	46.8
Product Complex	0.75	-4005.4113244	-4004.7835800	-4004.9390120	-176.0	-69.9
Ru^{III}–IMPA	3.75	-3904.9345978	-3904.3224660	-3904.4740530	-161.8	-97.4

Table D.43. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for each species involved in the reaction catalyzed by NU-1000, using the only possible spin state (singlet) for each species in the mechanism.

NU-1000	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
NU-1000-hyd	0	-3932.6102671	-3931.8192600	-3931.9759530	-135.1	-83.9
NU-1000-dehyd	0	-3856.1123160	-3855.3496670	-3855.5044400	0.0	0.0
NU-1000–Sarin	0	-4606.5074788	-4605.5852600	-4605.7618570	-156.3	-91.2
Reactant Complex	0	-4682.9700750	-4682.0192840	-4682.2009840	-198.0	-90.0
TS_{nuc}	0	-4682.9409172	-4681.9941010	-4682.1720340	-131.9	-14.0
INT_1	0	-4682.9430214	-4681.9923250	-4682.1708320	-127.2	-10.9
TS_{Berry}	0	-4682.9417274	-4681.9915940	-4682.1682790	-125.3	-4.2
INT_2	0	-4682.9435354	-4681.9922650	-4682.1698300	-127.1	-8.2
TS_{eli}	0	-4682.9327312	-4681.9846070	-4682.1625970	-107.0	10.7
Product Complex	0	-4682.9852259	-4682.0354070	-4682.2159510	-240.3	-129.3
NU-1000–IMPA (monodentate)	0	-4582.4945568	-4581.5602460	-4581.7369160	-189.2	-119.8
NU-1000–IMPA (bidentate)	0	-4506.0394998	-4505.1321820	-4505.3037400	-163.2	-136.5

Table D.44. Summary of the open shell species that has the maximum spin contamination for specific M-NU-1000 systems.*

Metal	Species	S^2	S^2_{cal}	$\Delta S^2 = S^2_{cal} - S^2$
Pd^{IV}	TS_{nuc}	2.000	2.014	0.014
Co^{III}	INT_1	2.000	2.079	0.079
Cu^{III}	Cu^{III}–OH	2.000	2.008	0.008
Fe^{III}	Fe^{III}–OH	8.750	8.760	0.010
Cu^{II}	Cu^{II}–OH	0.750	0.753	0.003
Co^{II}	Co^{II}–OH	3.750	3.762	0.012
Fe^{II}	Fe^{II}–OH	6.000	6.040	0.040
W^{II}	W^{II}–OH	2.000	2.091	0.091
Ni^{II}	Ni^{II}–OH	2.000	2.006	0.006
Mn^{II}	Mn^{II}–OH	8.750	8.760	0.010
Rh^{III}	INT_2	2.000	2.017	0.017
Ru^{III}	INT_2	0.750	0.769	0.019

*S^2 represents the theoretical value and S^2_{cal} is the calculated value.

For the M-NU-1000 systems containing open-shell spin states, we quantified the species throughout the respective mechanisms that had the maximum spin contamination and tabulated these values in Table D.44. If there is no spin contamination, then the calculated value of the total spin, S^2_{cal}, should equal the theoretical value, S^2. A general rule of thumb is that the spin contamination can be considered negligible if the calculated value of the total spin (S^2_{cal}) differs from the theoretical value of the total spin (S^2) by less than 10%. As seen by the ΔS^2 values in the table, there is no significant spin contamination for any of the open-shell systems that we considered.

D.13 Raw Energy Values for All Spin States for Other 17 M-NU-1000 Systems

In the tables below, the raw energy values are given for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for the other 17 M-NU-1000 systems. In the main text, these 17 systems were considered non-ideal and excluded from further analysis based on their relatively high free energy barriers to form **INT$_1$** and desorb IMPA. In other words, we did not calculate the full catalytic pathways for these 17 M-NU-1000 systems.

Table D.45. The SCF energies (in hartrees) for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Cr^{III}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = Cr^{III} ($3d^3$)	E (S^2 = 3.75)	E (S^2 = 0.75)
Cr^{III}–OH	-4128.1589235	-4128.1147536
Cr^{III}–Sarin	-4878.5380321	-4878.4859003
INT_1	-4954.9753057	-4954.9185592
Cr^{III}–IMPA	-4854.5322127	-4854.4748627

Table D.46. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Cr^{III}–NU-1000, using the most stable spin state for each species.

M = Cr^{III} ($3d^3$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Cr^{III}–OH	3.75	-4128.1589235	-4127.7187260	-4127.847237	0.0	0.0
Cr^{III}–Sarin	3.75	-4878.5380321	-4877.9386780	-4878.090508	-115.2	-54.0
INT_1	3.75	-4954.9753057	-4954.3471880	-4954.500347	-89.9	24.0
Cr^{III}–IMPA	3.75	-4854.5322127	-4853.9210090	-4854.07241	-167.4	-100.6

Table D.47. The SCF energies (in hartrees) for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Ru^{IV}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Ru^{IV} ($3d^4$)	E (S^2 = 6)	E (S^2 = 2)	E (S^2 = 0)
Ru^{IV}–OH	-3177.9138694	-3177.9414559	-3177.9257416
Ru^{IV}–Sarin	-3928.2980099	-3928.3095846	-3928.3230826
INT_1	-4004.7337744	-4004.7348302	-4004.7181709
Ru^{IV}–IMPA	-3904.2798100	-3904.2872277	-3904.2725004

Table D.48. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Ru^{IV}–NU-1000, using the most stable spin state for each species.

M = Ru^{IV} ($3d^4$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Ru^{IV}–OH	2	-3177.9414559	-3177.5118840	-3177.63979	0.0	0.0
Ru^{IV}–Sarin	0	-3928.3230826	-3927.7340070	-3927.88304	-120.9	-54.0
INT_1	2	-4004.7348302	-4004.1185470	-4004.27181	-32.7	79.4
Ru^{IV}–IMPA	2	-3904.2872277	-3903.6868880	-3903.83812	-95.8	-30.1

Table D.49. The SCF energies (in hartrees) for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Ir^{III}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Ir^{III} ($3d^6$)	E (S^2 = 6)	E (S^2 = 2)	E (S^2 = 0)
Ir^{III}–OH	-3187.9692489	-3187.9812500	-3187.9916957
Ir^{III}–Sarin	-3938.3175082	-3938.3660811	-3938.3731898
INT_1	-4014.7512083	-4014.8042118	-4014.8086038
Ir^{III}–IMPA	-3914.3025457	-3914.3532170	-3914.3680014

Table D.50. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Ir^{III}–NU-1000, using the most stable spin state for each species.

M = Ir^{III} (3d^6)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Ir^{III}–OH	0	-3187.9916957	-3187.5505690	-3187.67876	0.0	0.0
Ir^{III}–Sarin	0	-3938.3731898	-3937.7733120	-3937.92425	-122.5	-59.8
INT$_1$	0	-4014.8086038	-4014.1802470	-4014.33394	-93.1	18.6
Ir^{III}–IMPA	0	-3914.3680014	-3913.7562120	-3913.90512	-176.3	-103.7

Table D.51. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Y^{III}–NU-1000, using the only possible spin state (singlet) for each species.

M = Y^{III} (3d^0)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Y^{III}–OH	0	-3122.1698029	-3121.7305700	-3121.86124	0.0	0.0
Y^{III}–Sarin	0	-3872.5506820	-3871.9524890	-3872.10632	-120.4	-58.8
INT$_1$	0	-3948.9930356	-3948.3663690	-3948.52123	-109.2	5.9
Y^{III}–IMPA	0	-3848.5485345	-3847.9371040	-3848.08831	-178.6	-105.6

Table D.52. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for V^{IV}–NU-1000, using the only possible spin state (doublet) for each species.

M = V^{IV} (3d^1)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
V^{IV}–OH	0.75	-4027.0937857	-4026.6641040	-4026.7901800	0.0	0.0
V^{IV}–Sarin	0.75	-4777.4910431	-4776.9023750	-4777.0512610	-163.3	-100.8
INT$_1$	0.75	-4853.9018659	-4853.2852400	-4853.4384410	-70.7	36.7
V^{IV}–IMPA	0.75	-4753.4598592	-4752.8593980	-4753.0092220	-149.1	-84.5

Table D.53. The SCF energies (in hartrees) for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Cr^{II}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Cr^{II} (3d^4)	E (S^2 = 6)	E (S^2 = 2)	E (S^2 = 0)
Cr^{II}–OH	-4128.7351163	-4128.6846896	-4128.6458371
Cr^{II}–Sarin	-4879.1051682	-4879.0561057	-4879.0188296
INT$_1$	-4955.5545434	-4955.5072365	-4955.4490910
Cr^{II}–IMPA	-4855.1111020	-4855.0611353	-4855.0020983

Table D.54. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Cr^{II}–NU-1000, using the most stable spin state for each species.

M = Cr^{II} (3d^4)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Cr^{II}–OH	6	-4128.7351163	-4128.2827630	-4128.41318	0.0	0.0
Cr^{II}–Sarin	6	-4879.1051682	-4878.4941890	-4878.64738	-92.8	-30.2
INT$_1$	6	-4955.5545434	-4954.9152150	-4955.07081	-100.4	12.1
Cr^{II}–IMPA	6	-4855.1111020	-4854.4872230	-4854.64056	-173.2	-106.4

Table D.55. The SCF energies (in hartrees) for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for V^{II}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = V^{II} (3d^3)	E (S^2 = 3.75)	E (S^2 = 0.75)
V^{II}–OH	-4028.2639965	-4028.2380387
V^{II}–Sarin	-4778.6387803	-4778.6178205
INT$_1$	-4855.0910918	-4855.0533496
V^{II}–IMPA	-4754.6421460	-4754.5986672

Table D.56. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for V^{II}–NU-1000, using the most stable spin state for each species.

M = V^{II} ($3d^3$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
V^{II}–OH	3.75	-4028.2639965	-4027.8121210	-4027.9435080	0.0	0.0
V^{II}–Sarin	3.75	-4778.6387803	-4778.0278750	-4778.1815820	-104.2	-40.4
INT_1	3.75	-4855.0910918	-4854.4517250	-4854.6069950	-119.2	-3.2
V^{II}–IMPA	3.75	-4754.6421460	-4754.0183010	-4754.1713600	-177.7	-107.7

Table D.57. The SCF energies (in hartrees) for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Mo^{II}–NU-1000, for three possible spin states. The lower of the energies is colored red.

M = Mo^{II} ($3d^4$)	E (S^2 = 6)	E (S^2 = 2)	E (S^2 = 0)
Mo^{II}–OH	-3152.4687820	-3152.4606937	-3152.4372446
Mo^{II} –Sarin	-3902.8445127	-3902.8326193	-3902.8225576
INT_1	-3979.2776699	-3979.2686870	-3979.2301978
Mo^{II} –IMPA	-3878.8458581	-3878.8228157	-3878.7913820

Table D.58. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Mo^{II}–NU-1000, using the most stable spin state for each species.

M = Mo^{II} ($3d^4$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Mo^{II}–OH	6	-3152.4687820	-3152.0172240	-3152.14874	0.0	0.0
Mo^{II} –Sarin	6	-3902.8445127	-3902.2341210	-3902.3903	-107.2	-49.6
INT_1	6	-3979.2776699	-3978.6384360	-3978.79536	-70.9	41.0
Mo^{II} –IMPA	6	-3878.8458581	-3878.2227630	-3878.37689	-176.0	-108.4

Table D.59. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Zr^{IV}–NU-1000, using the only possible spin state (singlet) for each species.

M = Zr^{IV} ($3d^0$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Zr^{IV}–OH	0	-3130.2527616	-3129.8240310	-3129.9522340	0.0	0.0
Zr^{IV}–Sarin	0	-3880.6602171	-3880.0727420	-3880.2233990	-190.7	-127.3
INT_1	0	-3957.0760056	-3956.4604040	-3956.6143900	-110.7	0.2
Zr^{IV}–IMPA	0	-3856.6315512	-3856.0315330	-3856.1811250	-181.1	-110.4

Table D.60. The SCF energies (in hartrees) for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Mo^{IV}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = Mo^{IV} ($3d^2$)	E (S^2 = 2)	E (S^2 = 0)
Mo^{IV}–OH	-3151.3063474	-3151.3244422
Mo^{IV}–Sarin	-3901.7039415	-3901.6948838
INT_1	-3978.1143996	-3978.1150485
Mo^{IV}–IMPA	-3877.6689025	-3877.6680229

Table D.61. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Mo^{IV}–NU-1000, using the most stable spin state for each species.

M = Mo^{IV} ($3d^2$)	S^2	E (hartree)	H (hartree)	G (hartree)	Rel. H (kJ/mol)	Rel. G (kJ/mol)
Mo^{IV}–OH	0	-3151.3244422	-3150.8951240	-3151.0223720	0.0	0.0
Mo^{IV}–Sarin	2	-3901.7039415	-3901.1159320	-3901.2675760	-117.5	-59.1
INT_1	0	-3978.1150485	-3977.4994730	-3977.6526070	-26.6	84.1
Mo^{IV}–IMPA	2	-3877.6689025	-3877.0684230	-3877.2196990	-91.3	-27.5

Table D.62. The SCF energies (in hartrees) for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Mn^{IV}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = Mn^{IV} ($3d^3$)	**E (S^2 = 3.75)**	**E (S^2 = 0.75)**
Mn^{IV}–OH	-4234.0117536	-4233.9781473
Mn^{IV}–Sarin	-4984.3916517	-4984.3688596
INT_1	-5060.8011273	-5060.7673469
Mn^{IV}–IMPA	-4960.3600302	-4960.3218325

Table D.63. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Mn^{IV}–NU-1000, using the most stable spin state for each species.

M = Mn^{IV} ($3d^3$)	**S^2**	**E (hartree)**	**H (hartree)**	**G (hartree)**	**Rel. H (kJ/mol)**	**Rel. G (kJ/mol)**
Mn^{IV}–OH	3.75	-4234.0117536	-4233.5825880	-4233.7102110	0.0	0.0
Mn^{IV}–Sarin	3.75	-4984.3916517	-4983.8029100	-4983.9533500	-116.2	-53.7
INT_1	3.75	-5060.8011273	-5060.1841780	-5060.3383700	-19.4	89.5
Mn^{IV}–IMPA	3.75	-4960.3600302	-4959.7590440	-4959.9097500	-99.6	-33.3

Table D.64. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Hf^{IV}–NU-1000, using the only possible spin state (singlet) for each species.

M = Hf^{IV} ($3d^0$)	**S^2**	**E (hartree)**	**H (hartree)**	**G (hartree)**	**Rel. H (kJ/mol)**	**Rel. G (kJ/mol)**
Hf^{IV}–OH	0	-3131.2377160	-3130.8091250	-3130.9383540	0.0	0.0
Hf^{IV}–Sarin	0	-3881.6482787	-3881.0609500	-3881.2126420	-198.9	-135.5
INT_1	0	-3958.0621999	-3957.4470500	-3957.6019400	-114.8	-3.5
Hf^{IV}–IMPA	0	-3857.6178118	-3857.0179940	-3857.1681950	-184.7	-112.9

Table D.65. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Ce^{III}–NU-1000, using the only possible spin state (doublet) for each species.

M = Ce^{III} ($3d^1$)	**S^2**	**E (hartree)**	**H (hartree)**	**G (hartree)**	**Rel. H (kJ/mol)**	**Rel. G (kJ/mol)**
Ce^{III}–OH	0.75	-3559.0327095	-3558.5939520	-3558.7256430	0.0	0.0
Ce^{III}–Sarin	0.75	-4309.4107538	-4308.8131390	-4308.9688630	-113.2	-53.9
INT_1	0.75	-4385.8523936	-4385.2261270	-4385.3828650	-99.7	13.2
Ce^{III}–IMPA	0.75	-4285.4133691	-4284.8015840	-4284.9556830	-181.5	-113.4

Table D.66. The SCF energies (in hartrees) for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Pt^{II}–NU-1000, for two possible spin states. The lower of the energies is colored red.

M = Pt^{II} ($3d^8$)	**E (S^2 = 2)**	**E (S^2 = 0)**
Pt^{II}–OH	-3203.6035043	-3203.6107565
Pt^{II}–Sarin	-3953.9633274	-3953.9977450
INT_1	-4030.3847838	-4030.4424506
Pt^{II}–IMPA	-3929.9409981	-3929.9963182

Table D.67. The SCF energies, enthalpies, free energies, relative enthalpies, and relative free energies for the **M–OH**, **M–Sarin**, **INT$_1$**, and **M–IMPA** species for Pt^{II}–NU-1000, using the most stable spin state for each species.

M = Pt^{II} ($3d^8$)	**S^2**	**E (hartree)**	**H (hartree)**	**G (hartree)**	**Rel. H (kJ/mol)**	**Rel. G (kJ/mol)**
Pt^{II}–OH	0	-3203.6107565	-3203.1579780	-3203.2875160	0.0	0.0
Pt^{II}–Sarin	0	-3953.9977450	-3953.3862280	-3953.5405360	-137.0	-79.6
INT_1	0	-4030.4424506	-4029.8028420	-4029.9574900	-133.0	-20.3
Pt^{II}–IMPA	0	-3929.9963182	-3929.3721320	-3929.5233120	-198.6	-128.5

D.14 Raw Energy Values for Sarin, H_2O, IMPA and HF

Table D.68. The SCF energies, enthalpies, and free energies for the gas-phase **Sarin**, **H_2O**, **IMPA**, and **HF** species.

Species	E (hartree)	H (hartree)	G (hartree)
Sarin	-750.332840	-750.176069	-750.222695
H_2O	-76.443285	-76.418134	-76.439555
IMPA	-726.296194	-726.127386	-726.174698
HF	-100.468436	-100.455696	-100.475401

D.15 References for Appendix D

1. Troya, D. Reaction Mechanism of Nerve-Agent Decomposition with Zr-Based Metal Organic Frameworks. *J. Phys. Chem. C* **2016**, *120*, 29312–29323.

D.16 Cartesian Coordinates for Each System

The optimized cartesian coordinates (in Å) for each species, using the most stable spin state (lowest E), involved in the gas-phase hydrolysis of sarin catalyzed by each system are provided in the supplementary file (*Coordinates.zip*), available free of charge at https://doi.org/10.1021/acscatal.9b03594.