



DISSERTATION | DOCTORAL THESIS

Titel | Title

Modelling of Stability Trends in Chemical Space via Alchemical
Energy Decomposition

verfasst von | submitted by

Michael Jonathan Sahre B.Sc. M.Sc.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 796 605 419

Dissertationsgebiet lt. Studienblatt | Field of
study as it appears on the student record
sheet:

Chemie

Betreut von | Supervisor:

Univ.-Prof. Dr. Dr. h.c. Leticia Gonzalez Herrero
Univ.-Prof. Dr. Otto Anatole von Lilienfeld-Toal

CONTENTS

ACKNOWLEDGMENTS	1
ABSTRACT	2
ZUSAMMENFASSUNG	3
LIST OF ABBREVIATIONS	5
1 INTRODUCTION	6
2 THEORY	12
2.1 Electronic Structure Theory	12
2.2 Density Functional Theory	13
2.2.1 The Hohenberg-Kohn Theorems	13
2.2.2 Kohn-Sham Density Functional Theory	15
2.2.3 The Exchange-Correlation Functional	16
2.3 Electronic Structure Theory of Periodic Systems	19
2.3.1 Pseudopotentials	20
2.4 Quantum Alchemy	21
2.4.1 The Alchemical Hamiltonian	21
2.4.2 Alchemical Derivatives	23
2.5 Alchemical Atomic Energy Decomposition	25
2.5.1 Reference and path dependence of the decomposition	26
2.5.2 The uniform electron gas as reference	26
2.6 <i>Ab initio</i> atomic energy decompositions	28
2.6.1 Mean field decomposition	29
2.6.2 Interacting quantum atom	30
3 ALCHEMICAL ENERGY PARTITIONING OF PERIODIC SYSTEMS	32
3.1 An alchemical energy decomposition for periodic systems	32
3.2 Alchemical derivatives	35
3.3 Integration over the transformation path	37
3.4 Finite size effects	41
3.5 Summary	42
4 BINDING ENERGIES FROM ALCHEMICAL ENERGY DECOMPOSITION	44
4.1 Atomic Energy Notation	44
4.2 A model for bond dissociation energies	45
4.3 Performance	49
4.4 Comparison to Pauling's model	51
4.5 Comparison to Hammett's equation	53
4.6 Limitations to covalent bonds	54
4.7 Bond order	55
4.8 Electronic Configuration	55
4.9 Trends and dependence on period	56
5 TRANSFERABILITY OF ALCHEMICAL ATOMIC ENERGIES	58
5.1 Calculation of the atomic energies	58

5.2	Alchemical atomic energies of organic molecules	59
5.3	The relation between atomic energy and nuclear electrostatic repulsion potential	62
5.4	Comparison to other decomposition methods	64
6	CONCLUSION & OUTLOOK	67
	BIBLIOGRAPHY	71
A	APPENDIX: REPRINTED PUBLICATIONS	81
A.1	Transferability of atomic energies from alchemical decomposition	82
A.1.1	Transferability of atomic energies from alchemical de- composition	82
A.1.2	Supporting Information	96
A.2	Quantum Alchemy Based Bonding Trends and Their Link to Hammett's Equation and Pauling's Electronegativity Model . .	117
A.2.1	Quantum Alchemy Based Bonding Trends and Their Link to Hammett's Equation and Pauling's Electronega- tivity Model	117
A.2.2	Supporting Information	128

ACKNOWLEDGMENTS

The completion of my PhD would not have been possible without the support of my supervisors, colleagues and friends. Firstly, I would like to thank Anatole von Lilienfeld for giving me the opportunity to do a PhD in his group and for his support during my PhD. In particular for the paper on quantum alchemy based bonding trends, his guidance and creativity were invaluable. I would also like to thank Leticia González for becoming my supervisor after Anatole left the university of Vienna and for her advice and guidance during the last stage of my PhD. Furthermore, I would like to thank Guido von Rudorff for our fruitful collaborations. I enjoyed working with him due to his broad expertise and patience and greatly benefited from his advice in many scientific and technical problems. I am also grateful to Philipp Marquetand for our frequent scientific discussion that were the basis for my second paper on the transferability of alchemical energies. Furthermore, I would like to thank my fellow PhD students and postdocs in Anatoles and Leticias group. I have very fond memories of many interesting discussions with Giorgio and Danish. I am also grateful for the feedback I received on my thesis from Sylwia, Štěpán, Simon, Konstantin, Jan, Marco, Danish and Lukáš.

ABSTRACT

The design of chemical compounds with specific functionality is important to address challenges such as climate change or diseases. However, finding compounds with the desired properties is difficult due to the enormous size of chemical compound space. A physics-based understanding on how properties vary throughout chemical compound space would hence be highly beneficial for its efficient and reliable exploration. Quantum alchemy is a method describing property variations with respect to nuclear charge, i.e. with respect to composition, thus offering such physics-based framework. By annihilation and creation of nuclear charges, a reference system can be alchemically transformed into a compound, enabling the decomposition of the compound energy into atomic contributions. By partitioning the energy among the constituting atoms, the complexity, when considering all electrons and nuclei explicitly, is reduced. This could facilitate the development of simple models for structure-energy relationships. Hence, rational design efforts could be accelerated by enabling the rapid filtering of compounds for stability. In this thesis, a previously proposed alchemical atomic energy decomposition with respect to the uniform electron gas as reference is extended. By including the periodicity of the uniform electron gas and its external potential, a systematic error in the previously presented atomic energy expression is removed. Additionally, an efficient approximate procedure for atomic energy calculations with the extended decomposition scheme is proposed, enabling the application of the method to large datasets of molecules. Furthermore, the alchemical atomic energy decomposition is used to study stability trends in chemical compound space. As first application, the impact of the binding partners, A and B, on the bond dissociation energy E_{AB} is investigated, leading to the expression $E_{AB} \approx a - bZ_A Z_B + c(Z_A^{7/3} + Z_B^{7/3})$. After calibration of the parameters a, b, c , accurate energy estimates can be obtained for hydrogen saturated diatomics AB, connected via a single or double bond and being p -block elements from the same row $2 \leq n \leq 4$ of the periodic table. The model is of similar complexity and accuracy as Pauling’s electronegativity model, although of different functional form and origin. Furthermore, the change in bonding energy due to variation in nuclear charge is shown to be near-linear, in agreement with Hammett’s equation. In a second application, the stability of 1325 organic molecules with a larger variety of bonding patterns is modelled. Atomic energies are shown to be transferable across different molecules if the corresponding atoms share the same binding partners. This enables the estimation of atomisation energies with an error of 23 kcal/mol, comparable to simple statistical estimates but potentially more robust due to the connection to the physically meaningful energy decomposition. A comparison to other decomposition schemes indicates that alchemical atomic energies depend more strongly on the electrostatics in the local environment. Thus, atomic energies from the alchemical decomposition might be particularly useful as descriptors for processes with large variations in electrostatics, e.g. diffusion in solids.

ZUSAMMENFASSUNG

Die Entwicklung von Molekülen und Materialien mit bestimmten Eigenschaften spielt eine wichtige Rolle in der Bewältigung von Problemen wie Klimawandel oder Krankheiten. Die Suche nach chemischen Verbindungen mit den gewünschten Eigenschaften ist jedoch aufgrund der enormen Menge an möglichen Verbindungen schwierig, z. B. in der Größenordnung von 10^{60} für arzneimittelähnliche Moleküle. Ein physikalisch fundiertes Verständnis darüber, wie Eigenschaften innerhalb dieser Menge an möglichen Verbindungen variieren, wäre daher von großem Nutzen für ihre effiziente und zuverlässige Durchsuchung. Die Quantenalchemie bietet einen solchen physik-basierten Ansatz, indem sie Eigenschaftsänderungen in Bezug auf die Kernladung, d. h. in Bezug auf die chemische Zusammensetzung, beschreibt. Durch Annihilation und Erzeugung von Kernladungen kann ein Referenzsystem alchemistisch in ein Molekül oder Material umgewandelt werden. Das ermöglicht die Zerlegung der Energie einer Verbindung in atomare Beiträge.

Durch die Aufteilung der Energie auf die konstituierenden Atome wird die Komplexität, wenn alle Elektronen und Kerne explizit berücksichtigt werden, reduziert. Dies könnte die Entwicklung einfacher Modelle für Struktur-Energie-Beziehungen erleichtern. Entwicklungsprozesse könnten durch Anwendung solcher Modelle zur Filterung von stabilen Verbindungen beschleunigt werden.

In dieser Arbeit wird eine bestehenden alchemistische Energiezerlegung weiterentwickelt, in der das gleichförmigen Elektronengas als Referenz dient. Durch die Einbeziehung der Periodizität des gleichförmigen Elektronengases und seines externen Potentials wird ein systematischer Fehler in der Gleichung für die atomaren Energien beseitigt. Darüber hinaus wird ein effizientes Näherungsverfahren für Atomenergieberechnungen vorgeschlagen, das die Anwendung der Methode auf große Moleküldatensätze ermöglicht. Außerdem wird Energiezerlegungsmethode angewandt, um den Einfluss zweier Bindungspartner A und B auf die Bindungsdissoziationsenergie E_{AB} zu untersuchen, was zu der Gleichung $E_{AB} \approx a - bZ_A Z_B + c(Z_A^{7/3} + Z_B^{7/3})$ führt. Nach Kalibrierung der Parameter a, b, c kann damit die Bindungsenergie für wasserstoffgesättigte Moleküle AB genau bestimmt werden. A und B sind dabei p -Blockelemente aus der gleichen Reihe $2 \leq n \leq 4$ des Periodensystems und über eine Einfach- oder Doppelbindung verbunden. Das Modell ist von ähnlicher Komplexität und Genauigkeit wie das Pauling'sche Elektronegativitätsmodell, wenn auch anderer funktioneller Form und Ursprungs. Darüber hinaus wird gezeigt, dass die Änderung der Bindungsenergie aufgrund der Variation der Kernladung in Übereinstimmung mit der Hammett-Gleichung nahezu linear ist. Die Stabilität von Molekülen mit einer größeren Vielfalt von Bindungsmustern wird in einer zweiten Anwendung modelliert. Die Beziehung zwischen Atomenergie und lokaler Umgebung wird für 1325 organische Moleküle untersucht. Die Analyse zeigt, dass die Atomenergien zwischen verschiedenen Molekülen übertragbar sind, wenn die entsprechenden Atome die gleichen Bindungspartner haben. Dies ermöglicht die Abschätzung von Atomisierungsenergien mit einem Fehler von 23 kcal/mol,

vergleichbar mit einfachen statistischen Modellen, aber potentiell robuster da auf der physikalisch sinnvollen Energiezerlegung basierend. Im Vergleich zu anderen Energiezerlegungsschemata hängen die alchemistischen atomaren Energien stark von der Elektrostatik in der lokalen Umgebung ab. Daher könnte die alchemistische Energiezerlegung besonders nützlich für die Untersuchung von Prozessen mit großen Variationen in der Elektrostatik sein, z. B. der Diffusion in Festkörpern.

LIST OF ABBREVIATIONS

BDE	Bond Dissociation Energy
BOA	Born Oppenheimer Approximation
DFT	Density Functional Theory
MAE	Mean Absolute Error
MO	Molecular Orbital
QM	Quantum Mechanics
RDM1	One-Electron Reduced Density Matrix
UEG	Uniform Electron Gas

"Alles ist Chemie!"

This sentence, translating as "everything is chemistry", was brought up frequently by my high school chemistry teacher to emphasize the impact of the chemical sciences on our lives. Indeed, entire epochs in humanity like the copper, bronze and iron age have been named after materials that found widespread use in these periods and had enormous consequences for society. Also in modern times, chemical processes and compounds have continued to be of tremendous impact on our lives. For example, the invention of the Haber Bosch process or the development of antibiotics lead to a huge increase in population size and life expectancy.^{1,2}

Unfortunately, the discovery of new compounds with specific functionality is often based on luck or long and costly trial and error experiments.^{3,4} For example, penicillin was discovered due to accidental mold formation in a petri dish,⁵ while modern drug design can involve screening of millions of candidate molecules.⁶ However, pressing challenges such as climate change or pandemics make swifter development processes necessary.⁷

Computational modelling is a promising tool to accelerate the design of molecules and materials by assisting experimental synthesis and characterization with computational predictions. While the theoretical foundation for the exact quantum mechanical description of chemical systems was already laid out in the late 1920s,⁸ the complicated mathematical formulation in combination with lacking computing power prevented initially the application of theoretical models to practically relevant systems. Since then, accurate and efficient methods, such as density functional theory (DFT),^{9–11} have been developed and computing power has increased immensely.^{12,13} Consequently, computational methods are nowadays routinely used to support experiments, e.g. by verifying reaction mechanisms¹⁴ or reproducing experimental spectra.¹⁵

Besides aiding the interpretation of experiments, computational methods can also be used for property predictions of unknown compounds. Due to the often faster evaluation of properties by computation compared to experimental synthesis and characterization, computational screenings of large varieties of compounds have been performed.^{16–21} However, the huge amount of possible stoichiometries, bonding connectivity patterns and geometrical configurations leads to a gigantic space of possible compounds.²² For example, it was estimated that the space of small drug-like molecules is in the order of 10^{60} and that of inorganic four element compounds exceeds 10^{12} .^{23,24} Such large spaces cannot be sampled by brute-force enumeration with accurate *ab initio* methods like DFT.

Instead, stochastic algorithms¹⁶, as one example, have been employed to sample chemical subspaces of AlGaAs-alloys,²⁵ metal organic frameworks²⁰ or electrolytes.²⁶ Likewise machine learning has emerged as a promising tool for fast and accurate prediction of properties, such as atomization energies,^{27–31} dipole moments,³⁰ polarizabilities,^{28,30} or NMR-shifts.^{31–33}

It has been argued that stochastic algorithms are likely not sufficient for the effective design of compounds with specific functionality, so that physi-

cally motivated rules about how properties vary in chemical space are necessary.^{23,24} Additionally, machine learning techniques need large amounts of reference data and accurate out-of-sample predictions are not guaranteed.³¹ This could be in particular a problem in design applications where unknown parts of chemical space are explored.

But how can one find such physics-based rules? Arguably, a rigorous description of chemical compound space and of the variation of properties throughout this space is needed, that is consistent with the laws of physics. This could for example be helpful to encode more physics in machine learning models, thereby improving their data efficiency and reliability.^{31,34}

On a fundamental level, the information about the composition is encoded in the nuclear charges of the external potential and nuclear repulsion terms of the Hamiltonian. Consequently, the relationship between composition and properties can be modelled from first principles via the variation of properties due to a variation in the nuclear charges. In this regard, investigations about how the energy of a single atom changes as a function of the nuclear charge were already carried out in the early days of quantum chemistry.^{35–38} Hylleraas, for example, calculated energy changes in the series of helium-like atoms with perturbation theory.³⁵

The energy of a molecule as a functional of the electron density with the density itself being a function of the nuclear charges was expressed first by Wilson in 1962.³⁹ Building upon this work, Politzer and Parr developed approximate energy formulas for atoms and molecules by explicitly considering energy changes due to variations in the nuclear charges.⁴⁰ Furthermore, rigorous bounds on relative energies between polyatomic isoelectronic molecules were proven by studying the energy dependence on the nuclear charge.^{41–43}

A general description of chemical space was presented within a molecular grand canonical ensemble multi-component DFT framework.^{44,45} The molecular grand canonical ensemble has enabled the study of fluctuations in electron and also nuclear charge density. By allowing fluctuations in nuclear charge, a compound like lithium hydride can for example be converted into a single beryllium atom. The nuclear chemical potential $\mu_n(\mathbf{r})$ plays a central role in this framework. It describes the change in energy due to variation in nuclear charge density $Z(\mathbf{r})$, where $Z(\mathbf{r}) = \sum_I Z_I \delta(|\mathbf{r} - \mathbf{R}_I|)$ with Z_I being the number of protons at $\mathbf{R}_I$. The nuclear chemical potential $\mu_n(\mathbf{r})$ is then defined as

$$\mu_n(\mathbf{r}) = \frac{\delta E}{\delta Z(\mathbf{r})} = \sum_I \frac{Z_I \text{erf}[\sigma|\mathbf{R}_I - \mathbf{r}|]}{|\mathbf{r} - \mathbf{R}_I|} - \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (1)$$

where $\rho(\mathbf{r}')$ is the electron density and erf is the error function. The term $\text{erf}[\sigma|\mathbf{R}_I - \mathbf{r}|]$ switches off the divergence if μ_n is evaluated at $\mathbf{R}_I$. The nuclear chemical potential evaluated at the position $\mathbf{R}_I$ of a nucleus, is equal to $\partial_{Z_I} E$, i.e. the energy change with respect to the nuclear charge Z_I . This derivative has been named *alchemical potential*^{44,46} or *alchemical derivative*⁴⁷ analogous to the term *computational alchemy* in the context of free energy calculations⁴⁸ and in reminiscence of the medieval attempts to transform lead into gold. Higher order derivatives with respect to the nuclear charges can be calculated analytically with coupled perturbed self-consistent field theory.^{49,50}

These alchemical derivatives enable gradient-based optimization of chemical compounds. This has for example been exploited for the design of a

nonpeptidic inhibitor of an enzyme preventing cell death.⁴⁶ Furthermore, alchemical derivatives can be utilized to predict properties of a compound with respect to a reference system by a Taylor expansion in the following way: The Hamiltonians of the reference system, $\hat{H}^{\text{ref}}$, and the target system, $\hat{H}^{\text{tar}}$, are coupled linearly as

$$H(\lambda) = (1 - \lambda)\hat{H}^{\text{ref}} + \lambda\hat{H}^{\text{tar}} \quad (2)$$

where λ is a real number defined over the interval $[0, 1]$. Properties such as the energy can then be expressed as functions of λ , e.g. $E(\lambda) = \langle H(\lambda) \rangle$ and estimated by a Taylor expansion in λ around the reference. In case of reference and target being isoelectronic, the required derivatives $\partial_\lambda^n E$ can be expressed via the alchemical derivatives $\partial_{Z_i}^n E$ of the atoms in the reference. This allows the rapid screening of large chemical spaces based on only a few *ab initio* calculations to obtain the alchemical derivatives. For example, atomization energies for 3.1 million BN-doped fullerenes were estimated from alchemical derivatives calculated with couple cluster with single and double excitations in 8 core days, while self consistent field calculations with Møller-Plesset perturbation theory of second order would have required 977 core years.⁵¹ In similar spirit, deprotonation energies,^{52–54} stability of BN-doped graphene derivatives,⁵² metal alloys for heterogenous catalysis,^{55–57} or properties of semiconductor alloys^{58,59} were evaluated. Besides yielding quantitative predictions, quantum alchemy is also a conceptual tool to gain qualitative insights into structure-property relationships. In 1974, Politzer and Parr showed that an alchemical transformation can be used to define atoms in molecules.⁴⁰ In their approach, one starts from an infinitely spread electron density and no nuclear charges. The nuclear charges of the compound are then scaled up linearly to their respective value in the compound of interest at their respective position. The energy E^{mol} of the compound can be expressed as a sum of atomic contributions by integrating over the scaling path as

$$E^{\text{mol}} = \frac{1}{2} \sum_A Z_A V_{A,0} - \frac{1}{2} \sum_A \int_0^{Z_A} \left[Z'_A \left(\frac{\partial V_{A,0}}{\partial Z'_A} \right) - V_{A,0} \right] dZ'_A \quad (3)$$

where $\sum_A$ is a sum over all nuclei in the compound, Z_A is the nuclear charge and $V_{A,0}$ is the electrostatic potential

$$V_{A,0} = \sum_{B \neq A} \frac{Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} - \int d\mathbf{r} \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_A|}, \quad (4)$$

where $\mathbf{R}_{A/B}$ are positions of the nuclei A and B and $\rho(\mathbf{r})$ is the electron density.

Note that this is not the only way to define atomic contributions to the energy. Another definition is for example the interacting quantum atom.^{60,61} In this approach, the electron density is partitioned into sub-volumes around the nuclei based on its topology. Atomic energies are subsequently calculated from the interactions between electrons and the nucleus within one sub-volume and from their interactions with the electron density and nuclei in the other sub-volumes.

It is also possible to obtain atomic energies by first partitioning the energy among basis functions of the wavefunction in combination with an partitioning of the basis functions among the atoms. Recently, this was realized with

localized intrinsic bond orbitals as basis functions and a partitioning of these basis functions among the atoms via atomic charges from intrinsic atomic orbitals.⁶²

While atomic energies are not uniquely defined, they can still be useful to get a better understanding of structure-energy relationships throughout chemical compound space. The interacting quantum atom decomposition was for example used to study binding energy trends of first row diatomics,⁶³ to investigate the role of electrostatics on torsional energy profiles⁶⁴, analyze the energy barriers of S_N2 reactions⁶⁵ or to characterize carbon atoms in aromatic, anti-aromatic and non-aromatic systems.⁶⁶ In addition to such conceptual applications, atomic energies from the interacting quantum atom method were also used for the development of machine learning force fields,^{67–71} which were subsequently used, for example, for geometry optimization of single molecules,⁷² dimers⁷³ and crystals⁷⁴ and molecular dynamics simulations of liquid water.⁷⁵

The usefulness of atomic energies from intrinsic bond and intrinsic atomic orbitals as fingerprints of atoms in different local environments was recently investigated⁷⁶ and also employed to classify types of electronic excitations⁷⁷ or to predict atomization energies in combination with machine learning models.⁷⁸

In the alchemical energy partitioning, Eq. (3), the energy depends on the sum of atomic electrostatic potentials and their derivatives with respect to the nuclear charges, which has motivated simple approximate formulas for the energy of atoms and molecules.^{40,79,80} However, a challenge in the calculation of atomic energies from Eq. (3) is the evaluation of the electrostatic potential in Eq. (4) and its derivative with respect to the nuclear charge for the states along the transformation path from an infinitely spread out electron density up to the density in the compound, i.e. the evaluation of the integrand in Eq. (4). The variation of this electrostatic potential and of the energy due to a change in the nuclear charge has only been studied for single atoms^{81–85} and between pairs of isoelectronic diatomics.⁸⁵ However, an explicit decomposition of the energy of a molecule into atomic contributions based on Eq. (3) has not been carried out yet to the best of my knowledge.

An important step towards the practical calculation of atomic energies from an alchemical transformation was provided by von Rudorff and von Lilienfeld in 2019.⁸⁶ Utilizing the Hellmann-Feynman theorem and the chain rule, they proposed a decomposition of the electronic energy difference between a reference and a target system into site specific contributions similar to Eq. (3) as

$$E^{\text{el, tar}} - E^{\text{el, ref}} = - \sum_I \underbrace{\Delta Z_I \int_0^1 d\lambda \int d\mathbf{r} \frac{\rho(\mathbf{r}, \lambda)}{|\mathbf{r} - \mathbf{R}_I|}}_{:=\Delta E_I} \quad (5)$$

where $E^{\text{el, tar}}$ and $E^{\text{el, ref}}$ are the electronic energies (e.g. without nuclear repulsion) of the target and the reference system, $\sum_I$ is the sum over all unique nuclei in reference and target, $\Delta Z_I = Z_I^{\text{tar}} - Z_I^{\text{ref}}$ is the change in nuclear charge at site I from reference to target, $\mathbf{R}_I$ is the position of nucleus I , λ is a coupling variable linearly connecting the Hamiltonians of reference and target (see Eq. (2)) and $\rho(\mathbf{r}, \lambda)$ is the electron density as a function of spatial coordinate $\mathbf{r}$

and the coupling variable λ . The alchemical transformation is performed by continuously changing λ from 0 to 1 and the atomic contributions are calculated by integrating over the electrostatic potential along the transformation path. It should be further noted that it is assumed that the external potential in the Hamiltonians of reference and target is only due to a set of point charges and that reference and target must be isoelectronic.

It was further demonstrated, that the atomic energies in Eq. (5) depend on the reference system and the transformation path. For example, atomic energies differences ΔE_I between compound A and C can differ for the direct transformation ($A \rightarrow C$) from atomic energies for the transformation via an intermediate compound B ($A \rightarrow B \rightarrow C$). This reference and path-dependence makes it difficult to obtain atomic energies that are comparable throughout chemical compound space. Thus, an infinitely extended uniform electron gas (UEG) was suggested as common reference from which any compound isoelectronic to the UEG can be generated by the same transformation path via linear up-scaling of its nuclear charge. Such an infinitely extended UEG is quite similar to the infinitely extended electron density employed by Wilson and Politzer. However, the difference is the presence of an uniform positive background charge density in case of the UEG. The background charge integrated over space is equal to the number of electrons in the system at $\lambda = 0$ and is linearly scaled down along the transformation path to 0 at $\lambda = 1$, such that the system is neutral for any λ -value over the interval $[0, 1]$. This λ -dependent infinitely extended background charge density can be modelled via periodic boundary conditions. Since the evaluation of the electron density from DFT with periodic boundary conditions is implemented in various electronic structure codes, the sampling of the electron density along the λ -path and subsequent calculation of atomic energies seems to be straightforward. However, in Eq. (5) it is assumed that the external potential of the reference is due to point charges while it is uniform if the UEG is used. Furthermore, if periodic boundary conditions are employed also the contributions from periodic images to the external potential should be taken into account. Thus, it would be desirable to modify Eq. (5) accordingly to recover the exact energy difference between reference and target from the sum of atomic energies.

Then, atomic energies from the alchemical energy decomposition might be useful to get a better understanding of structure-energy relationships because the complicated many body interactions between nuclei and electrons could be simplified by projecting the energy onto the constituting atoms. This could be helpful to explore the large chemical space of molecules emerging from the many possible combinations of the same local bonding patterns. If the local environment dependent atomic energies could be efficiently and accurately calculated, molecular energies could be estimated by summing up the atomic energies for the local environments within a specific molecule. The atomic energies from the alchemical decomposition scheme could thus be potentially useful to advance efforts in the computational design of compounds by providing swift, physically motivated estimates of stability.

Thus, the goal of this thesis is to investigate whether atomic energies can be modelled by just considering their local environment and whether they are transferable throughout the chemical compound space. Furthermore, the thesis will explore in how far these atomic energies can be utilized for the prediction of compound stability.

The necessary theory and methods for the calculation of atomic energies are reviewed in chapter 2.

In chapters 3-5, the results of the thesis are presented. In chapter 3, the alchemical decomposition scheme Eq. (5) is extended by including an uniform external potential for the reference system and periodic boundary conditions. Furthermore, approximations for the efficient calculation of atomic energies are investigated. The atomic contributions to bond dissociation energies (BDEs) of simple molecules are calculated with the alchemical energy decomposition in chapter 4. From the relation between these atomic binding contributions and the nuclear charge of the binding partner, a model for BDEs is constructed and tested for various compositions and bond orders. Furthermore, the relation to other simple models describing binding trends in chemical compound space is explored. In chapter 5, atomic energies are calculated for atoms in more diverse local environments and for a larger set of molecules than in the previous chapter. The dependence of the atomic energies on the local environment is analysed and an approach to estimate atomisation energies is proposed. Furthermore, the atomic energy-local structure relationship emerging from the alchemical decomposition scheme is compared to the same relationship for other atomic energy decomposition schemes.

A summary and an outlook is given in chapter 6.

The main results of this thesis have been already presented in two papers that can be found in the sections A.1.1 and A.2.1.

The following section will introduce the relevant theoretical background to the research presented in this thesis. First, the basics of electronic structure theory, in particular DFT, will be summarized. Then, quantum alchemy will be introduced and the alchemical atomic energy decomposition from Ref 86 will be derived. Finally, two other atomic decomposition schemes are briefly presented, to which the alchemical one will be compared in this thesis.

2.1 Electronic Structure Theory

The non-relativistic time-independent Schrödinger equation is defined as

$$\hat{H}\Psi = E\Psi \quad (6)$$

with $\hat{H}$ and Ψ being Hamiltonian and wavefunction of the system of interest, respectively.⁸⁷

The Hamiltonian $\hat{H}$ of a molecule (in atomic units) has the general form

$$\hat{H} = \underbrace{-\sum_{i=1}^N \frac{1}{2} \nabla_i^2}_{:=\hat{T}} - \underbrace{\sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2}_{:=\hat{T}^{\text{nuc}}} - \underbrace{\sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|}}_{:=\hat{V}_{e,N}} + \underbrace{\sum_{i=1}^N \sum_{j>i}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{:=\hat{V}_{ee}} + \underbrace{\sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}}_{:=\hat{V}_{NN}}, \quad (7)$$

with N and M being the number of electrons and nuclei, respectively. Z_A and $\mathbf{R}_A$ denote the nuclear charge and position of nucleus A and $\mathbf{r}_i$ the position of the i -th electron.⁸⁷ The different terms describe the kinetic energy of the electrons $\hat{T}$ and nuclei $\hat{T}^{\text{nuc}}$, the electrostatic interactions between electrons and nuclei $\hat{V}_{e,N}$, between electrons $\hat{V}_{ee}$ and between nuclei $\hat{V}_{NN}$, respectively. The corresponding wavefunction $\Psi(\{\mathbf{r}\}, \{\mathbf{R}\})$ is a function of the positions of all electrons and nuclei.

Since the masses of the nuclei are significantly higher than the one of the electrons, the nuclei move slower and as an approximation, the movements of electrons and nuclei can be decoupled. This simplification, known as the Born-Oppenheimer approximation (BOA),⁸⁸ can accurately describe systems, for which only one electronic state is important, e.g. ground state properties of closed shell molecules.⁸⁹ In this approximation, the wavefunction is determined for fixed positions of the nuclei, meaning the electronic wavefunction $\Psi(\{\mathbf{r}\}; \{\mathbf{R}\})$ depends only on the electron coordinates explicitly, but parametrically on the nuclear coordinates. The Schrödinger equation is thus solved for the modified hamiltonian $\hat{H}^{\text{BOA}}$ in which $\hat{T}^{\text{nuc}}$ is omitted and $\hat{V}_{NN}$ is a constant:

$$\hat{H}^{\text{BOA}} = \underbrace{-\sum_{i=1}^N \frac{1}{2} \nabla_i^2}_{:=\hat{T}} - \underbrace{\sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|}}_{:=\hat{V}_{e,N}} + \underbrace{\sum_{i=1}^N \sum_{j>i}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{:=\hat{V}_{ee}} + \underbrace{\sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}}_{:=\hat{V}_{NN}}. \quad (8)$$

The corresponding electronic wavefunction has to be antisymmetric with respect to the exchange of two electrons, which is fulfilled if the wavefunction is a linear combination of so-called Slater-determinants.⁸⁷ A single Slater-determinant for N electrons has the general form

$$\Psi^{\text{SL}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \dots & \phi_N(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \dots & \phi_N(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \phi_1(\mathbf{x}_N) & \phi_2(\mathbf{x}_N) & \dots & \phi_N(\mathbf{x}_N) \end{vmatrix} \quad (9)$$

The wavefunction is expanded in functions or orbitals ϕ_i depending on the position and spin $\mathbf{x}$ of only one electron with $\frac{1}{\sqrt{N!}}$ as a normalization factor. There are in total N orbitals for N electrons. The orbitals in i -th row in the determinant depend on position and spin $\mathbf{x}_i$ of the i -th electron with the electron occupying all orbitals $\phi_1 \dots \phi_N$. The exchange of two electrons corresponds to the exchange of two rows in the determinant. Swapping rows changes the sign of determinant and hence the required antisymmetry of the wavefunction is fulfilled.

According to the variational theorem, the wavefunction of a system can be determined by minimizing the energy of a trial wavefunction. If one chooses a single Slater-determinant as trial wavefunction, the best approximation to the true wavefunction can be obtained by minimization of the energy with respect to the orbitals ϕ_i . This is known as the Hartree-Fock method.^{87,90–92} The result can be improved by inclusion of more determinants but this leads to an increase in computational cost.⁹³

A different approach, yielding generally more accurate results⁹⁴ than Hartree-Fock at comparable cost is DFT.

2.2 Density Functional Theory

The high computational cost of solving the Schrödinger equation is related to the high dimensionality of the wavefunction which depends on $4N$ variables for N electrons (per electron three for position and one for spin). This begs the question, whether it is possible to reduce the dimensionality, while still maintaining an exact description of the electronic structure and molecular properties.

2.2.1 The Hohenberg-Kohn Theorems

Hohenberg and Kohn proved that indeed the electron density, depending only on three variables, uniquely determines the Hamilton operator of a system and therefore all properties of the system.⁹ The electron density $\rho(\mathbf{r})$ of a N -electron wavefunction $\Psi(\mathbf{x}, \mathbf{x}_2, \dots, \mathbf{x}_N)$ is defined as

$$\rho(\mathbf{r}) = N \int d\mathbf{x}_1 d\mathbf{x}_2 \dots d\mathbf{x}_N \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N), \quad (10)$$

where the integral is over all spatial and spin coordinates for all electrons except of the first one. For the first electrons $\mathbf{x}_1 = (\mathbf{r}_1, \sigma_1)$ integration is only carried out over the spin σ_1 . $\rho(\mathbf{r})$ can be interpreted as the probability density

to find any electron with arbitrary spin at position $\mathbf{r}$ while the other $N - 1$ electrons have arbitrary positions and spins.⁹⁵ The original proof by Hohenberg and Kohn assumes two external potentials V_{ext} and V'_{ext} that differ by more than a constant but lead to the same electron density $\rho(\mathbf{r})$. This electron density corresponds to a non-degenerate ground state for N electrons.^{9,96} The ground state wavefunction, Hamiltonians and corresponding energies are denoted as Ψ , Ψ' , $\hat{H}$, $\hat{H}'$ and E_0 , E'_0 , respectively, with $E_0 \neq E'_0$ due to the non-degeneracy of the systems. Then from the variational principle follows

$$\begin{aligned} E_0 < \langle \Psi' | \hat{H} | \Psi' \rangle &= \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle \\ E_0 < E'_0 + \langle \Psi' | \hat{V}_{\text{ext}} - \hat{V}'_{\text{ext}} | \Psi' \rangle &= E'_0 + \int d\mathbf{r} \rho(\mathbf{r}) (\hat{V}_{\text{ext}} - \hat{V}'_{\text{ext}}) \end{aligned} \quad (11)$$

where $\hat{H} - \hat{H}' = \hat{V}_{\text{ext}} - \hat{V}'_{\text{ext}}$ because the systems are isoelectronic. Analogously, one can define

$$E'_0 < E_0 - \int d\mathbf{r} \rho(\mathbf{r}) (\hat{V}_{\text{ext}} - \hat{V}'_{\text{ext}}) \quad (12)$$

and adding of Eqs. (11) and (12) results in the contradiction

$$E_0 + E'_0 < E_0 + E'_0. \quad (13)$$

This proof shows that there is a one-to-one correspondence between electron density and external potential. Since the electron density determines the number of electrons as well, all information about the Hamiltonian is contained in the electron density. Hence, it is possible to construct the Hamiltonian for a given ground state density and therefore also the associated wavefunction and all molecular ground state properties. Furthermore, one can express the energy as a functional of the density as

$$E[\rho] = \int d\mathbf{r} \hat{V}_{\text{ext}} + F_{\text{HK}}[\rho] \quad (14)$$

with the first term accounting for the electrostatic interaction between electrons and nuclei and the second term $F_{\text{HK}}[\rho]$ being a universal functional describing kinetic energy and electron-electron interactions as a functional of the density.

The second Hohenberg-Kohn theorem⁹ states that the ground state density ρ_0 of a N -electron system yields the minimum energy $E[\rho]$ for all densities fulfilling

$$\int d\mathbf{r} \rho = N. \quad (15)$$

The Hamiltonian $\hat{H}$ and the associated wavefunction Ψ_0 can be determined from ρ_0 according to the first HK-theorem and the energy E_0 can be obtained as

$$E_0 = \langle \Psi_0[\rho_0] | \hat{H} | \Psi_0[\rho_0] \rangle. \quad (16)$$

For some other density $\tilde{\rho}$ fulfilling Eq. (15) the corresponding wavefunction $\tilde{\Psi}$ can be found analogously. Then from the variational theorem follows

$$E_0 = \langle \Psi_0[\rho_0] | \hat{H} | \Psi_0[\rho_0] \rangle < \langle \tilde{\Psi}[\tilde{\rho}] | \hat{H} | \tilde{\Psi}[\tilde{\rho}] \rangle \quad (17)$$

if $\tilde{\rho} \neq \rho_0$ proving that ρ_0 yields the minimum energy $E[\rho_0]$.

While these results show that it is possible to calculate molecular properties from the density, they do not give an explicit formulation of the universal functional F_{HK} . Approaches to evaluate the energy in the framework of DFT will be addressed in the next sections.

2.2.2 Kohn-Sham Density Functional Theory

The universal functional F_{HK} can be split up into

$$F_{\text{HK}}[\rho] = T[\rho] + J[\rho] + E_{\text{ncl}}[\rho] \quad (18)$$

with the kinetic energy functional $T[\rho]$, the classical coulomb interaction between electrons $J[\rho]$ and non-classical contributions due to exchange, as well as a self-interaction correction and correlation effects described by the last term $E_{\text{ncl}}[\rho]$.⁹⁷ $T[\rho]$ and $E_{\text{ncl}}[\rho]$ are unknown.

Thomas and Fermi proposed to use the local kinetic energy functional of the UEG as an approximation of $T[\rho]$, which leads to errors in the order of 10-20%.⁹⁸⁻¹⁰⁰ Since the kinetic energy is in the order of the total energy for coulombic systems this leads to a huge absolute error when applied to molecules and materials.¹⁰⁰

Thus, Kohn and Sham proposed to split the kinetic energy into two terms. One term is the kinetic energy of system of non-interacting particles with the same density as the system of interacting electrons.^{10,97,100} The second term is a correction which accounts for the difference between the non-interacting kinetic energy and the energy of the interacting system. The non-interacting kinetic energy can be calculated exactly and is a very good approximation to the true kinetic energy such that chemical systems can be accurately described even if the correction term is neglected.¹⁰⁰ In the following, the details of the Kohn-Sham approach will be laid out.

The system of N non-interacting particles can be represented by a single Slater determinant with an associated density

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\phi_i(\mathbf{r})|^2 \quad (19)$$

with the ϕ_i being one-electron orbitals. The Hamiltonian of the non-interacting system is given by

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 + V_S(\mathbf{r}_i) \quad (20)$$

with the kinetic energy being exactly $-\frac{1}{2} \sum_i \langle \phi_i | \nabla_i^2 | \phi_i \rangle$ and the effective potential $V_S(\mathbf{r}_i)$ chosen such that the density of the non-interacting system (Eq. (19)) is identical to the density of the interacting system. It is noted, that the existence of a non-interacting system as defined in Eq.(19) and the corresponding potential V_S is only assumed to exist. However, this is not proven to be generally true.¹⁰¹

The orbitals ϕ_i can be found by solving

$$\left(-\frac{1}{2} \sum_i \nabla_i^2 + V_S(\mathbf{r}_i) \right) \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \quad (21)$$

for all ϕ_i .¹⁰² To solve these equations, the potential V_S must be known. In order to obtain an expression for V_S , it is helpful to go back to the exact energy functional as defined in Eq. (14) and Eq. (18) for the interacting system. The functional can be reformulated utilizing the orbital-representation of the density (Eq. (19)) as

$$\begin{aligned} E[\rho] &= T_S[\rho] + \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' \frac{\rho(\mathbf{r}')\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} - \int d\mathbf{r} \rho(\mathbf{r}) \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} + E_{xc}[\rho(\mathbf{r})] \\ &= -\frac{1}{2} \sum_i \langle \phi_i(\mathbf{r}_i) | \nabla_i^2 | \phi_i(\mathbf{r}_i) \rangle + \frac{1}{2} \sum_{i,j} \int d\mathbf{r} \int d\mathbf{r}' |\phi_i(\mathbf{r})|^2 \frac{1}{|\mathbf{r} - \mathbf{r}'|} |\phi_j(\mathbf{r}')|^2 \\ &\quad - \sum_i \int d\mathbf{r} |\phi_i(\mathbf{r})|^2 \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} + E_{xc}[\rho(\mathbf{r})]. \end{aligned} \quad (22)$$

$T_S[\rho]$ is the kinetic energy of the non-interacting system and $E_{xc}[\rho] \equiv E_{\text{ncl}}[\rho] + T[\rho] - T_S[\rho]$ accounts for the non-classical electron-electron interactions including a correction term $T[\rho] - T_S[\rho]$ for the difference in kinetic energy between the interacting and non-interacting systems.

According to the second Hohenberg-Kohn theorem, the ground state density can be found by minimizing this expression with respect to the density via minimization with respect to the orbitals ϕ_i under the constraint $\langle \phi_i | \phi_j \rangle = \delta_{ij}$.

This corresponds to solving

$$\begin{aligned} \left(-\frac{1}{2} \nabla^2 + \underbrace{\left[\int d\mathbf{r}_2 \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_2 - \mathbf{r}_1|} + V_{xc}(\mathbf{r}_1) - \sum_A^M \frac{Z_A}{|\mathbf{r}_1 - \mathbf{R}_A|} \right]}_{:= V_{\text{eff}}(\mathbf{r}_1)} \right) \phi_i(\mathbf{r}_1) &= \epsilon_i \phi_i(\mathbf{r}_1) \\ \left(-\frac{1}{2} \nabla^2 + V_{\text{eff}}(\mathbf{r}_1) \right) \phi_i(\mathbf{r}_1) &= \epsilon_i \phi_i(\mathbf{r}_1) \end{aligned} \quad (23)$$

for all orbitals ϕ_i with the exchange-correlation potential $V_{xc} = \frac{\delta E_{xc}}{\delta \rho}$.¹⁰³ From comparison to Eq. (21), it follows that $V_{\text{eff}} = V_S$, meaning that the density of the non-interacting system (and thus also the interacting one) can be determined from Eq. (23).

Application of Eq. (23) to obtain the density of specific molecules or materials requires an explicit expression for E_{xc} and V_{xc} . Approximations for these terms will be introduced in the next section.

2.2.3 The Exchange-Correlation Functional

Wavefunction approaches to quantum chemistry can be systematically improved by including more Slater determinants in the trial wavefunction. Constructing increasingly accurate exchange-correlation functionals is not as straightforward. Nevertheless, the "Jacobs ladder of density functional approximations for the exchange-correlation energy" was introduced to classify and order different approaches by accuracy.⁹⁴ The different rungs of the ladder correspond to increasingly sophisticated approximations of the exchange-correlation energy and will be presented in the following.

The Local Density Approximation

The basis for most approximate exchange-correlation functionals is a UEG exposed to a uniform positive background potential.^{104,105} While such a model resembles the electronic structure of metals quite well, it is quite different from molecules with a finite strongly varying electron density. The advantage of the model is however, that its exchange functional is known exactly and the correlation functional can be determined with high accuracy.¹⁰⁶

The exchange-correlation energy of the UEG is assumed to be

$$E_{XC} = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{XC}(\rho(\mathbf{r})). \quad (24)$$

$\epsilon_{XC}(\rho(\mathbf{r}))$ is the exchange-correlation energy per particle of the UEG, which is weighted by the probability ρ to find a particle at $\mathbf{r}$.¹⁰⁶ Since the energy contribution at each point $\mathbf{r}$ depends only on the density at $\mathbf{r}$, the expression is known as the local density approximation. $\epsilon_{XC}(\rho(\mathbf{r}))$ is split further into an exchange and a correlation contribution

$$\epsilon_{XC}(\rho(\mathbf{r})) = \epsilon_X(\rho(\mathbf{r})) + \epsilon_C(\rho(\mathbf{r})). \quad (25)$$

The exchange term^{107–109} is given by

$$\epsilon_X(\rho(\mathbf{r})) = -\frac{3}{4} \left(\frac{3\rho(\mathbf{r})}{\pi} \right)^{1/3}. \quad (26)$$

While the exact expression for $\epsilon_C(\rho(\mathbf{r}))$ is unknown, analytical expressions were proposed based on accurate quantum Monte Carlo simulations of the UEG.^{110–112}

The Generalized Gradient Approximation

As mentioned earlier, the local density approximation does not give accurate results for molecules due to their inhomogeneous density. A better description of molecules is achieved by including information about the gradient of the density $\nabla\rho$ in the exchange-correlation functional. Such a gradient expansion approximation can be expressed as

$$E_{XC} = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{XC}(\rho_\alpha, \rho_\beta) + \sum_{\sigma, \sigma'} \int d\mathbf{r} C_{XC}^{\sigma, \sigma'}(\rho_\alpha, \rho_\beta) \frac{\nabla\rho_\sigma}{\rho_\sigma^{2/3}} \frac{\nabla\rho_{\sigma'}}{\rho_{\sigma'}^{2/3}} d\mathbf{r} + \dots \quad (27)$$

with σ, σ' indicating α or β spin and ρ_α, ρ_β being the spin densities.¹¹³ However, this expansion can even perform worse than the local density approximation because the associated exchange-correlation hole^a has lost many physically correct features incorporated in the local density approximation hole.¹¹³ This problem can be solved by omitting gradient expansion terms, which lead to unphysical behaviour of the exchange-correlation hole.¹¹³ The

^a The exchange-correlation hole is the difference between the conditional electron pair densities of correlated and uncorrelated electrons. It describes how much the probability to find an electron in some distance to another electron differs if the electrons are correlated or uncorrelated (see Ref 114 for details).

resulting functionals are known as generalized gradient approximation (GGA) functionals

$$E_{\text{XC}}^{\text{GGA}} = \int d\mathbf{r} f(\rho_\alpha, \rho_\beta, \nabla_\alpha, \nabla_\beta). \quad (28)$$

Approaches to construct GGA-functionals can be divided in two classes. Either functionals are fitted to reproduce reference data for specific chemical compounds or the functionals are constructed to obey strict mathematical conditions. In the latter case, they rely on reference systems, such as the uniform and slowly varying electron gases.¹¹⁵ An example of the former class is the B86 exchange functional and PBE is an example of the latter.^{116,117}

Meta-GGA functionals

Meta-GGA functionals (MGGA) are an extension of the GGA formalism taking also the kinetic energy density $\tau(\mathbf{r})$ or higher order derivatives of the density into account.⁹⁴ Inclusion of $\tau(\mathbf{r})$ leads to functionals of the general form^{94,118}

$$E_{\text{XC}}^{\text{MGGA}}[\rho] = \int d\mathbf{r} \rho \epsilon_{\text{XC}}^{\text{MGGA}}(\rho, \nabla\rho, \tau) \quad (29)$$

with the kinetic energy density being a sum over the Kohn-Sham orbitals ϕ_i

$$\tau(\mathbf{r}) = \frac{1}{2} \sum_i [\nabla\phi_i(\mathbf{r})][\nabla\phi_i(\mathbf{r})]. \quad (30)$$

Hybrid-Functionals

The exchange energy tends to be significantly larger than the correlation energy. Thus, in particular a better description of exchange could improve xc-functionals.¹¹⁹ Since the exchange energy can be calculated exactly for a single Slater determinant, using the Fock exchange energy seems to be a way to improve the exchange functional. However, complete replacement of the exchange term in GGA-functionals with Fock exchange leads to a worse performance for calculations of molecules because the localized exact xc-hole is poorly reproduced.^{94,119} In approximate xc-functionals, the exchange and correlation hole are both localized and the approximate xc-hole is hence localized as well. In contrast, the Fock exchange hole is non-local, and the combination with the approximate local correlation hole does not recover the exact local xc-hole.

Thus, it was proposed to mix exact exchange with the one from approximate exchange functionals to better preserve the features of the exact xc-hole and the obtained expressions are known as hybrid functionals.¹²⁰ The fraction of exact exchange is often determined from a fit to molecular reference data, such as in the B3LYP functional but can also be determined from non-empirical considerations as in the PBE0 functional.^{121–123}

Other Approximations

Random phase approximations and double hybrid functionals are placed on the highest rung of Jacobs ladder.¹²⁴ In random phase approximations, unoccupied orbitals are included in the calculation of the energy as well, enabling

the modelling of long-range electron correlation.^{125,126} In double hybrid functionals, the correlation energy is described by mixing GGA correlation with correlation energy from second order perturbation theory.¹²⁷

2.3 Electronic Structure Theory of Periodic Systems

Electronic structure calculations cannot be carried out for arbitrarily large systems by explicit modelling of all electrons since the computing effort increases with system size. An approach to deal with extended systems is instead the use of periodic boundary conditions, where a unit cell is repeated infinitely in all directions. This requires only the explicit modelling of the particles in one unit cell. The interactions with particles in other unit cells are included via the periodic boundary conditions. This is a convenient way to describe periodic systems like crystals but can also be applied to disordered systems or even molecules. In the latter two cases however, the unit cell should be large enough such that the interaction of an atom with its own images in the adjacent unit cells can be neglected.

The periodicity of the system is described by introducing lattice vectors $\mathbf{L}$. They are defined as

$$\mathbf{L}(n_1, n_2, n_3) = \sum_{i=1}^3 n_i \mathbf{a}_i, \quad (31)$$

where n_i are integers and $\mathbf{a}_i$ are the basis vectors of the unit cell. A vector $\mathbf{r}$ can then be translated to a symmetrically identical vector $\mathbf{r} + \mathbf{L}(n_1, n_2, n_3)$ by addition of a lattice vector.¹²⁸

A convenient basis for the electron density or the associated Kohn-Sham orbitals of a periodic system are plane waves, $e^{i\mathbf{q}\cdot\mathbf{r}}$, with the wavevectors $\mathbf{q}$. The plane wave basis must recover the periodicity of the lattice. This is the case, if their wavevectors $\mathbf{q}$ obey the equation $\mathbf{L} \cdot \mathbf{q} = 2\pi n$ for any $\mathbf{L}$ where n is an integer. These wavevectors, in the following denoted as $\mathbf{G}$, form the so-called reciprocal lattice and are the Fourier transform of the lattice vectors $\mathbf{L}$ in real space.¹²⁹ The vectors $\mathbf{G}$ can be expressed in a basis $\{\mathbf{b}_i\}$

$$\mathbf{G} = \sum_{i=1}^3 m_i \mathbf{b}_i, \quad (32)$$

where m_i are integers. The reciprocal basis is related to the basis in real space as

$$\begin{aligned} \mathbf{b}_1 &= 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}, \\ \mathbf{b}_2 &= 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}, \\ \mathbf{b}_3 &= 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}. \end{aligned} \quad (33)$$

The KS orbitals ψ of a periodic system can be expanded in a basis of plane waves as

$$\psi_{i,\mathbf{k}} = \sum_m c_{i,m}(\mathbf{k}) \times \frac{1}{\Omega} \exp\{i(\mathbf{k} + \mathbf{G}_m) \cdot \mathbf{r}\} \quad (34)$$

where $\mathbf{k}$ is a vector in the reciprocal space, Ω is the volume of the system, $\sum_m$ is a sum over all reciprocal lattice vectors $\mathbf{G}_m$ and $c_{i,m}$ are expansion coefficients.¹³⁰ While the KS-orbitals of a non-periodic system are characterized by a single quantum number i , the orbitals dependent on two quantum numbers, i and $\mathbf{k}$ if periodic boundary conditions are applied.

2.3.1 Pseudopotentials

The expansion of the KS-orbitals in plane waves allows the efficient solution of the KS equations via the fast Fourier transform. However, the accurate description of the localized core electrons would require a large set of plane waves, which makes this approach inefficient in practice. Thus, normally only the valence electrons are modelled explicitly and their interactions with nuclear charges and core electrons are modelled via an effective potential. The valence state ψ_i , obtained if all electrons are considered, is then described by a pseudo-orbital ψ_i^{PS} instead. One type of such effective potentials are norm-conserving pseudopotentials.¹³¹ Commonly, they are constructed such that they reproduce certain properties of the all-electron wavefunctions of a single atom. In particular, the pseudopotentials should fulfill following conditions to ensure transferability to other systems than single atoms:^{131,132}

1. The eigenvalues of all-electron orbitals ψ_i and corresponding pseudo-valence orbitals ψ_i^{PS} should agree for the chosen atomic reference configuration.
2. ψ_i and ψ_i^{PS} should coincide beyond a defined cutoff radius r_c .
3. The logarithmic derivatives should agree, e.g. $\frac{d}{dr} \ln \psi_i = \frac{d}{dr} \ln \psi_i^{\text{PS}}$,
4. The respective charges Q_l within the cutoff region should be identical, e.g. $Q_l = \int_0^{r_c} dr r^2 |\psi_i|^2 = \int_0^{r_c} dr r^2 |\psi_i^{\text{PS}}|^2$
5. The first order energy derivatives of the logarithmic derivatives should agree at r_c and hence also for $r > r_c$,

where it can be shown that the 5-th point is implied by the 4-th condition.

In this thesis, norm-conserving pseudopotentials as proposed by Goedecker, Teter and Hutter will be used.^{133–135} These pseudopotentials can be split into a local and a non-local part, V_{loc} and V_{nl} , respectively as

$$V(\mathbf{r}) = V_{\text{loc}}(\mathbf{r}) + V_{\text{nl}}(\mathbf{r}). \quad (35)$$

The local part is defined as

$$V_{\text{loc}}(r) = -\frac{Z_{\text{ion}}}{r} \text{erf}\left(\frac{r}{\sqrt{2}r_c}\right) + \exp\left\{\left[-\frac{1}{2}\left(\frac{r}{r_c}\right)^2\right]\right\} \times \left[\sum_{i=1}^4 C_i \left(\frac{r}{r_c}\right)^{2i-2}\right], \quad (36)$$

where r is the distance to the nucleus, Z_{ion} is the nuclear valence charge, erf is the error function, r_c is the cutoff radius and C_i are coefficients to model the short range part of the potential. The non-local part is given by

$$V_{\text{nl}}(\mathbf{r}, \mathbf{r}') = \sum_l \sum_{m=-l}^{+l} \sum_{i=1}^3 \sum_{j=1}^3 Y_{lm}(\hat{\mathbf{r}}) p_i^l(r) h_{ij}^l p_j^l(r') Y_{lm}^*(\hat{\mathbf{r}}') \quad (37)$$

with the angular momentum quantum number l , the spherical harmonics Y_{lm} and the Gaussian-type projectors p_i^l . They are defined as

$$p_i^l(r) = \frac{\sqrt{2} r^{l+2(i-1)} \exp\left\{\left(-\frac{r^2}{2r_i^2}\right)\right\}}{r_i^{l+(4i-1)/2} \sqrt{\Gamma\left(l + \frac{4i-1}{2}\right)}} \quad (38)$$

where Γ is the Gamma-function. The parameters h_{ij}^l and the coefficients C_i are determined from a fit to reference data and are tabulated in Refs. 134,135.

2.4 Quantum Alchemy

In the following sections, a detailed derivation of the alchemical atomic energy decomposition, that is investigated and applied in this thesis, will be given.

For this purpose, the term alchemical transformation will be defined and it will be described how it can be used to transform a system A into a system B.

Then, it will be shown how the energy change along such an alchemical transformation path can be expressed as a sum of atomic terms.

Based on these results, a decomposition of the energy into atomic terms will be introduced.

2.4.1 The Alchemical Hamiltonian

Two systems can be connected by linear coupling of their Hamiltonians as

$$\hat{H}(\lambda) = \hat{H}_A + \lambda(\hat{H}_B - \hat{H}_A), \quad (39)$$

with $\hat{H}_A$ and $\hat{H}_B$ being the Hamiltonians of system A and system B, respectively. λ is a coupling variable defined over the interval from 0 to 1 with $\hat{H}(\lambda = 0) = \hat{H}_A$ and $\hat{H}(\lambda = 1) = \hat{H}_B$. By changing the coupling variable λ , A is transformed into B. This is referred to as an *alchemical* transformation because nuclear charges can be transmuted along the transformation path.

As an example, the transformation of N_2 into CO is illustrated in Figure 1. Along the transformation path, the nuclear charges change linearly from seven to six and eight, respectively (Figure 1A). This is accompanied by a lowering of the energy (Figure 1B) and a depletion of electrons at the site transformed into a carbon atom (R_1) and an increase of electron density at the site mutated

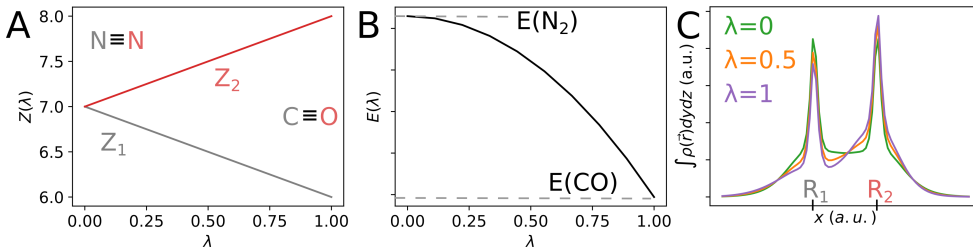


Figure 1: The alchemical transformation of N_2 into CO. The nuclear charges Z_1 and Z_2 change linearly along the transformation path (Panel A). This is accompanied by a change in energy (Panel B) and a change of the electron density illustrated for the 1D-projection of the density on the bonding axis (Panel C).

into oxygen ($\mathbf{R}_2$)(Figure 1C). Since the nuclear charges are fractional for $0 < \lambda < 1$, such a transformation cannot be observed experimentally. Nevertheless, the intermediate systems are well defined within the Born-Oppenheimer approximation via the λ -dependent Hamiltonian in Eq. (39). Hence, the λ -dependent wavefunction and observables can be obtained by solving the Schrödinger equation for $H(\lambda)$.

Furthermore, $\hat{H}(\lambda)$ can be simplified if the systems A and B are isoelectronic, like N_2 and CO . In this case, only the external potential depends explicitly on λ but kinetic energy and electron-electron operators, $\hat{T}$ and $\hat{V}_{\text{ee}}$ do not. This is evident from Eq. (39) and the definitions

$$\hat{T} = \frac{1}{2} \sum_i^N \nabla_i^2 \quad (40)$$

and

$$\hat{V}_{\text{ee}} = \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (41)$$

which yield identical expressions for the same number of electrons N . Consequently, these terms cancel in the λ -dependent part $\lambda(\hat{H}_{\text{B}} - \hat{H}_{\text{A}})$ of the Hamiltonian Eq. (39) and only the electron-nuclei and nucleus-nucleus interaction, $\hat{V}_{\text{e,N}}$ and $\hat{V}_{\text{N,N}}$, remain such that

$$\hat{H}(\lambda) = \hat{H}_{\text{A}} + \lambda(\Delta \hat{V}_{\text{e,N}} + \Delta \hat{V}_{\text{N,N}}), \quad (42)$$

with $\Delta \hat{V}_{\text{e,N}} = \hat{V}_{\text{e,N}}^{\text{B}} - \hat{V}_{\text{e,N}}^{\text{A}}$ and $\Delta \hat{V}_{\text{N,N}} = \hat{V}_{\text{N,N}}^{\text{B}} - \hat{V}_{\text{N,N}}^{\text{A}}$ denoting the differences between the respective operators of systems B and A. $\Delta \hat{V}_{\text{e,N}}$ is defined as

$$\Delta \hat{V}_{\text{e,N}} = \sum_{i=1}^N \sum_{I=1}^M \frac{-Z_I^{\text{B}} + Z_I^{\text{A}}}{|\mathbf{r}_i - \mathbf{R}_I|}, \quad (43)$$

where N is the number of electrons and $\mathbf{r}_i$ the position of the i -th electron. $\mathbf{R}_I$ is the position of the I -th nucleus and Z_I^{A} and Z_I^{B} are the nuclear charges at this site in the systems A and B, respectively. M denotes the number of all unique sites in A and B.

$\Delta \hat{V}_{\text{N,N}}$ is defined as

$$\Delta \hat{V}_{\text{N,N}} = \frac{1}{2} \sum_{I=1}^M \sum_{J \neq I}^M \frac{Z_I^{\text{B}} Z_J^{\text{B}} - Z_I^{\text{A}} Z_J^{\text{A}}}{|\mathbf{R}_I - \mathbf{R}_J|}. \quad (44)$$

The nuclear charge at site I can change due to annihilation, i.e. $Z_I^{\text{A}} > 0$ and $Z_I^{\text{B}} = 0$, due to mutation, i.e. $Z_I^{\text{A}} \neq Z_I^{\text{B}}$ and $Z_I^{\text{A}}, Z_I^{\text{B}} < 0$ or due to nuclear charge creation, i.e. $Z_I^{\text{A}} = 0$ and $Z_I^{\text{B}} > 0$.

The three possibilities are illustrated for an alchemical transformation of LiH to He_2 in Figure 2. In the starting system A, lithium is on position $\mathbf{R}_1$ and hydrogen is on position $\mathbf{R}_2$. In the final system, the lithium atom at $\mathbf{R}_1$ has vanished, the hydrogen at $\mathbf{R}_2$ is transformed into a helium atom, and a helium atom is created at position $\mathbf{R}_3$. Three unique sites at $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ exist for this transformation and hence $M = 3$.

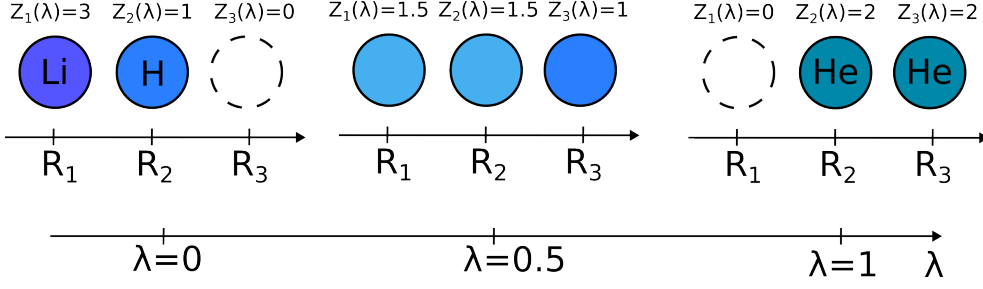


Figure 2: The three possible changes of atomic sites, (1) annihilation (R_1), (2) mutation (R_2), (3) creation (R_3), during an isoelectronic alchemical transformation illustrated for the transformation of LiH to He_2 .

2.4.2 Alchemical Derivatives

The dependence of the energy on λ is given by its definition as the expectation value of $\hat{H}$

$$E(\lambda) = \langle \Psi_\lambda | \hat{H}(\lambda) | \Psi_\lambda \rangle, \quad (45)$$

where Ψ_λ denotes the wavefunction of $\hat{H}(\lambda)$ with parametric dependence of Ψ on λ .

According to the Hellmann-Feynman theorem,^{136,137} the derivative $\partial_\lambda E$ can be evaluated from the expectation value of $\partial_\lambda \hat{H}$ as

$$\frac{\partial E}{\partial \lambda} = \frac{\partial}{\partial \lambda} \langle \Psi_\lambda | \hat{H}(\lambda) | \Psi_\lambda \rangle = \left\langle \Psi_\lambda \left| \frac{\partial \hat{H}(\lambda)}{\partial \lambda} \right| \Psi_\lambda \right\rangle. \quad (46)$$

Using the definition of $\hat{H}(\lambda)$ in Eq. (42), one obtains

$$\frac{\partial \hat{H}}{\partial \lambda} = \frac{\partial}{\partial \lambda} (\hat{H}_A + \lambda(\Delta \hat{V}_{e,N} + \Delta \hat{V}_{N,N})) = \Delta \hat{V}_{e,N} + \Delta \hat{V}_{N,N}, \quad (47)$$

meaning the derivative depends only on the change in external potential between the two systems A and B. Insertion of this expression into Eq. (46) yields,

$$\frac{\partial E}{\partial \lambda} = \langle \Psi_\lambda | \Delta \hat{V}_{e,N} + \Delta \hat{V}_{N,N} | \Psi_\lambda \rangle = \Delta \hat{V}_{N,N} + \int d\mathbf{r}_1 \dots d\mathbf{r}_N \Delta \hat{V}_{e,N} |\Psi_\lambda(\mathbf{r}_1, \dots, \mathbf{r}_N)|^2, \quad (48)$$

where I exploit that $\Delta \hat{V}_{N,N}$ does not depend on the electron coordinates $\mathbf{r}_i$ and that the wavefunction is normalized. The term $\Delta \hat{V}_{N,N}$ is the difference between nuclear repulsion in system B and system A. Since the calculation of this term is trivial, only the term $\int d\mathbf{r}_1 \dots d\mathbf{r}_N \Delta \hat{V}_{e,N} |\Psi_\lambda|^2$ will be considered in the following. Then only the derivative of the electronic energy $\partial_\lambda E^{\text{el}}$, i.e. without nuclear repulsion, is considered. This derivative,

$$\frac{\partial E^{\text{el}}}{\partial \lambda} = \int d\mathbf{r}_1 \dots d\mathbf{r}_N \Delta \hat{V}_{e,N} |\Psi_\lambda(\mathbf{r}_1, \dots, \mathbf{r}_N)|^2, \quad (49)$$

describes the difference in electrostatic interaction between electrons and nuclei for the wavefunction Ψ_λ and systems B and A. Using the definition of $\Delta\hat{V}_{e,N}$ in Eq. (43), it can be written as a sum over site-specific terms as

$$\frac{\partial E^{\text{el}}}{\partial \lambda} = \int d\mathbf{r} \Delta\hat{V}_{e,N} |\Psi_\lambda|^2 = - \sum_I \int d\mathbf{r} \frac{(Z_I^{\text{B}} - Z_I^{\text{A}})}{|\mathbf{r} - \mathbf{R}_I|} \rho_\lambda(\mathbf{r}), \quad (50)$$

where $Z_I^{\text{A/B}}$ are the nuclear charges of site I in systems A and B, respectively and $\mathbf{R}_I$ is the position of the nucleus at site I . $\rho_\lambda(\mathbf{r})$ is the electron density defined as

$$\rho_\lambda(\mathbf{r}) = N \int d\mathbf{r}_2 \dots d\mathbf{r}_N |\Psi_\lambda(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2, \quad (51)$$

where N is the number of electrons. Note that the sum over all electron coordinates $\mathbf{r}_i$ in $\Delta\hat{V}_{e,N}$ and integration over all these coordinates in Eq. (49) is replaced by the integral over the electron density depending only on $\mathbf{r}$ because electrons are indistinguishable.

In Ref 86, the partitioning of $\partial_\lambda E$ into a sum of atomic terms was achieved by defining λ -dependent nuclear charges $Z_I(\lambda)$ and by consecutive application of the chain rule. This alternative approach will be reviewed in the following. Starting from Eq. (42) (and omitting the nuclear repulsion $\Delta\hat{V}_{\text{NN}}$), the electronic Hamiltonian $\hat{H}^{\text{el}}(\lambda)$ can be expanded as

$$\begin{aligned} \hat{H}^{\text{el}}(\lambda) &= \hat{T} + \hat{V}_{\text{ee}} + \lambda \Delta\hat{V}_{e,N} \\ &= \hat{T} + \hat{V}_{\text{ee}} - \sum_{i=1}^N \sum_I \frac{Z_I^{\text{A}} + \lambda(Z_I^{\text{B}} - Z_I^{\text{A}})}{|\mathbf{r}_i - \mathbf{R}_I|} \\ &= \hat{T} + \hat{V}_{\text{ee}} - \sum_i \sum_I \frac{Z_I(\lambda)}{|\mathbf{r}_i - \mathbf{R}_I|} \end{aligned} \quad (52)$$

with

$$Z_I(\lambda) = Z_I^{\text{A}} + \lambda(Z_I^{\text{B}} - Z_I^{\text{A}}). \quad (53)$$

Due to the implicit dependence of the Hamiltonian on λ in Eq. (52), the derivative of the electronic energy E^{el} as a function of λ can be calculated via the chain rule as

$$\frac{\partial E^{\text{el}}}{\partial \lambda} = \left\langle \Psi_\lambda \left| \frac{\partial \hat{H}^{\text{el}}}{\partial \lambda} \right| \Psi_\lambda \right\rangle = \left\langle \Psi_\lambda \left| \sum_I \frac{\partial \hat{H}^{\text{el}}}{\partial Z_I} \frac{\partial Z_I}{\partial \lambda} \right| \Psi_\lambda \right\rangle = \sum_{I=1}^M \frac{\partial E^{\text{el}}}{\partial Z_I} \frac{\partial Z_I}{\partial \lambda} \quad (54)$$

with $\frac{\partial E^{\text{el}}}{\partial Z_I} \frac{\partial Z_I}{\partial \lambda} = \frac{\partial Z_I}{\partial \lambda} \left\langle \Psi_\lambda \left| \frac{\partial \hat{H}^{\text{el}}}{\partial Z_I} \right| \Psi_\lambda \right\rangle$.

Furthermore, from Eq. (52) follows

$$\frac{\partial H^{\text{el}}}{\partial Z_I} = - \sum_{i=1}^N \frac{1}{|\mathbf{r}_i - \mathbf{R}_I|} \quad (55)$$

and from Eq. (53) follows

$$\frac{\partial Z_I}{\partial \lambda} = Z_I^{\text{B}} - Z_I^{\text{A}} \quad (56)$$

such that

$$\frac{\partial E^{\text{el}}}{\partial Z_I} \frac{\partial Z_I}{\partial \lambda} = -(Z_I^{\text{B}} - Z_I^{\text{A}}) \left\langle \Psi_\lambda \left| \sum_{i=1}^N \frac{1}{|\mathbf{r}_i - \mathbf{R}_I|} \right| \Psi_\lambda \right\rangle = -(Z_I^{\text{B}} - Z_I^{\text{A}}) \int d\mathbf{r} \frac{\rho_\lambda(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|}, \quad (57)$$

which is identical to the site-specific contributions to $\partial_\lambda E$ in Eq. (50).

The derivative $\frac{\partial E}{\partial Z_I}$ of the energy with respect to Z_I can be interpreted as the electrostatic potential due to the electron density ρ_λ acting on a positive test charge at position $\mathbf{R}_I$.⁸⁶

Since the derivative $\frac{\partial E}{\partial Z_I}$ measures the tendency of a system to transmute the nuclear charge at position $\mathbf{R}_I$, it has been named the alchemical potential μ_I .^{39,44,86}

2.5 Alchemical Atomic Energy Decomposition

The final result from the previous section can be used to decompose the energy difference between system A and system B into atomic contributions. The electronic energy difference between systems A and B, $\Delta E^{\text{el}} = E^{\text{el}, \text{B}} - E^{\text{el}, \text{A}}$, can be calculated by integration over the transformation path as

$$\Delta E^{\text{el}} = \int_0^1 d\lambda \frac{\partial E^{\text{el}}}{\partial \lambda}. \quad (58)$$

Using the definitions of $\frac{\partial E^{\text{el}}}{\partial \lambda}$ in Eq. (54) and of $\frac{\partial E^{\text{el}}}{\partial Z_I} \frac{\partial Z_I}{\partial \lambda}$ in Eq. (57), ΔE^{el} can be expressed as a sum of atomic terms as

$$\Delta E^{\text{el}} = \sum_{I=1}^M \underbrace{\Delta Z_I \int_0^1 d\lambda \frac{\partial E^{\text{el}}}{\partial Z_I}}_{:=\Delta E_I^{\text{el}}} = \sum_{I=1}^M \underbrace{-\Delta Z_I \int_0^1 d\lambda \int d\mathbf{r} \frac{\rho(\lambda, \mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|}}_{:=\Delta E_I^{\text{el}}}, \quad (59)$$

where $\Delta Z_I = Z_I^{\text{B}} - Z_I^{\text{A}}$.

In order to obtain atomic contributions to the total energy, the nuclear repulsion must be partitioned as well. In Ref 86 an equal splitting of each two body term was proposed. The atomic energies for systems A and B respectively are then defined as

$$E_I^{\text{NN}, \text{A/B}} = \frac{Z_I^{\text{A/B}}}{2} \sum_{J \neq I} \frac{Z_J^{\text{A/B}}}{|\mathbf{R}_I - \mathbf{R}_J|}, \quad (60)$$

where $Z_I^{\text{A/B}}$ is the nuclear charge at site I in either system A or system B. The resulting atomic contribution to the nuclear repulsion difference is hence

$$E_I^{\text{NN}, \text{B}} - E_I^{\text{NN}, \text{A}} = \sum_{J \neq I} \frac{Z_I^{\text{B}} Z_J^{\text{B}} - Z_I^{\text{A}} Z_J^{\text{A}}}{2|\mathbf{R}_I - \mathbf{R}_J|}. \quad (61)$$

and the total atomic energy is

$$\Delta E_I = -\Delta Z_I \int_0^1 d\lambda \int d\mathbf{r} \frac{\rho(\lambda, \mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} + \sum_{J \neq I} \frac{Z_I^{\text{B}} Z_J^{\text{B}} - Z_I^{\text{A}} Z_J^{\text{A}}}{2|\mathbf{R}_I - \mathbf{R}_J|} \quad (62)$$

such that the total energy difference between systems A and B is

$$\Delta E = \sum_I \Delta E_I. \quad (63)$$

2.5.1 Reference and path dependence of the decomposition

The previous section showed, that alchemical transformations can be used to define an exact partitioning of the energy difference between two systems A and B into atomic contributions. However, if these atomic energies should be used to understand energy-structure relationships throughout chemical compound space, they should also be comparable throughout this chemical space. In particular, it would be desirable that atomic energies depend on the local environment within a molecule. For example, the carbon atom in a carboxyl-group should have similar atomic energy values across different carboxylic acids.

In Ref 86, it was demonstrated that this not necessarily the case. Even for a single site I , different atomic energies ΔE_I can be obtained depending on the transformation path connecting the systems A and B. This was illustrated for the transformation of hexazine (N_6) into borazine ($N_3B_3H_6$) with identical geometries for the 6-membered ring in N_6 and $B_3N_3H_6$ (see Figure 3). The atomic energies can be obtained by coupling the respective Hamiltonians and integration of Eq. (59) along the associated transformation path. However, another option is to transform N_6 first to benzene and then to $N_3B_3H_6$.⁸⁶

As shown in Figure 3, the site-specific contributions are different for the two transformation paths. For example, all hydrogens have identical atomic energies for the path $N_6 \rightarrow C_6H_6 \rightarrow B_3N_3H_6$. This is because the hydrogens are symmetric in benzene and from benzene to borazine, the nuclear charges of these sites remain constant. Thus, $\Delta Z_I = 0$, resulting in no contribution from these sites according to Eq. (59). In contrast, the direct transformation from hexazine to borazine yields atomic energies differing for H bonded either to N (H-N) or to B (H-B). I note, that the atomic energies in Figure 3 do not contain the nuclear repulsion term. If this term is added, the atomic energies for the path $N_6 \rightarrow C_6H_6 \rightarrow B_3N_3H_6$ would differ for the hydrogens in H-N and H-B as well. However, the atomic nuclear repulsion ΔE_I^{NN} as defined in Eq. (61) is identical for the paths $N_6 \rightarrow C_6H_6 \rightarrow B_3N_3H_6$ and $N_6 \rightarrow B_3N_3H_6$. Hence, the total atomic energies ΔE_I would differ for the two transformation paths even in nuclear repulsion is included.

Besides this path dependence, the spatial orientation of systems A and B influences the atomic energies. In case of A and B containing planar 6-membered rings, as in Figure 3, it might be a reasonable choice to arrange them to overlap. However, if one system contained a 5-membered ring or was non-cyclic, it would be unclear how to arrange the systems A and B.

Consequently, it is crucial to find an unbiased reference system A and a mutual alchemical transformation path to isoelectronic, but otherwise arbitrary systems B for comparable, and thus meaningful, atomic energies throughout chemical space.

2.5.2 The uniform electron gas as reference

The UEG was proposed as unbiased reference system for all molecules and materials isoelectronic to it.⁸⁶ The UEG is a hypothetical model of electrons in an infinite volume, exposed to a uniform positive background charge density.¹⁰⁵ It can be modeled using periodic boundary conditions, with the

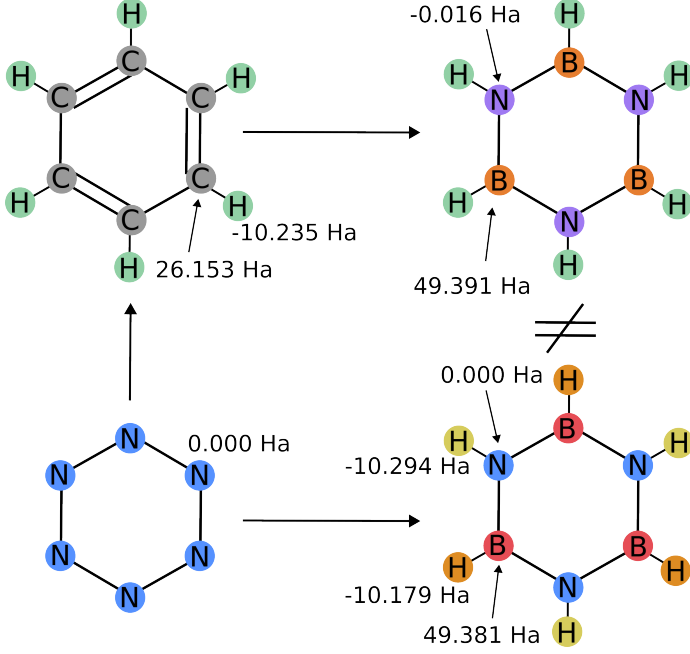


Figure 3: Path-dependent relative atomic energies from Eq. (59) of borazine (BNBNBH₆) with hexazine as a reference. Colors encode atomic energies. All coordinates correspond to the geometry of benzene. Calculations were performed at HF/STO-6G level. (Figure is redrawn based on Fig 2 in Ref. 86)

volume of the unit cell and the number of electrons within the cell defining the average electron density. The positive background charge density is identical to the average electron density to obtain a neutral unit cell. This is necessary because the energy of a periodic system would diverge for a charged unit cell.

The UEG can be transformed into an arbitrary isoelectronic compound by linearly growing nuclear charges at the positions of the nuclei of the compound while simultaneously reducing the positive background charge density. This resolves the dependence of the atomic energies on the relative spatial orientation of systems A and B, discussed in the previous section. Since the external potential of the UEG is uniform, the atomic energies ΔE_I are invariant with respect to translations or rotations of any system B. Furthermore, atomic energies for any system are calculated via a linear transformation of the UEG into the respective compound according to Eq. (39). This prevents inconsistent atomic energies due to different transformation paths, such as the transformations $N_6 \rightarrow C_6H_6 \rightarrow B_3N_3H_6$ vs $N_6 \rightarrow B_3N_3H_6$ in Figure 3. Hence, the UEG should enable the calculation of atomic energies that are comparable throughout chemical compound space.

However, the choice of the UEG as reference requires a modification of the atomic energy decomposition scheme.

The external potential of the UEG can be defined using periodic boundary conditions (see section 2.3) as

$$V_{\text{ext}}^{\text{UEG}}(\mathbf{r}) = - \sum_{\mathbf{L}} \rho^{\text{bg}} \int_{\Omega_{\text{cell}}} d\mathbf{r}' \frac{1}{|\mathbf{r} - \mathbf{r}' - \mathbf{L}|}, \quad (64)$$

where $\mathbf{L}$ are the lattice vectors of the periodic system, ρ^{bg} is the positive uniform background charge density and the integral is carried out over the

volume of a single unit cell Ω_{cell} .¹⁰⁵ Note that the potential is due to a continuous charge density instead of a set of discrete point charges.

For the linear transformation of the UEG into a compound, the external potential $V_{e,N}(\lambda)$ is given by

$$V_{e,N}(\lambda) = \sum_{i=1}^N \sum_{\mathbf{L}} \left(- \sum_I \frac{Z_I(\lambda)}{|\mathbf{r}_i - \mathbf{R}_I - \mathbf{L}|} - (1 - \lambda) \int_{\Omega_{\text{cell}}} d\mathbf{r}' \frac{\rho^{\text{bg}}}{|\mathbf{r}_i - \mathbf{r}' - \mathbf{L}|} \right), \quad (65)$$

where $Z_I(\lambda)$ is defined in Eq. (53) with $Z_I^A = 0$.

In order to obtain atomic energies, $\partial_\lambda E^{\text{el}}$ was expressed as a sum of site-specific terms (see Eq. (54)) and these terms were integrated with respect to λ in the previous sections. Following the same approach for the external potential Eq. (65) results in

$$\Delta E^{\text{el}} = \sum_{I=1}^M \left(\sum_{\mathbf{L}} -\Delta Z_I \int_0^1 d\lambda \int d\mathbf{r} \frac{\rho(\lambda, \mathbf{r})}{|\mathbf{r} - \mathbf{R}_I - \mathbf{L}|} \right) + \sum_{\mathbf{L}} \int_0^1 d\lambda \int d\mathbf{r} \int d\mathbf{r}' \frac{\rho^{\text{bg}} \rho(\lambda, \mathbf{r})}{|\mathbf{r} - \mathbf{r}' - \mathbf{L}|}. \quad (66)$$

Since the potential of the UEG is continuous, ΔE^{el} contains the term

$$\sum_{\mathbf{L}} \int_0^1 d\lambda \int d\mathbf{r} \int d\mathbf{r}' \frac{\rho^{\text{bg}} \rho(\lambda, \mathbf{r})}{|\mathbf{r} - \mathbf{r}' - \mathbf{L}|} \quad (67)$$

which is not site-specific but has contributions from the whole unit cell. The question is if and how this term should be partitioned among the constituting atoms. Additionally and in contrast to Eq. (59) the energy expression depends on the summation over the lattice vectors $\mathbf{L}$.

In Ref 86, the summation over $\mathbf{L}$ and the term Eq. (67) were neglected. This was justified by the application of the decomposition scheme to small molecules (8-14 valence electrons), which were placed in large unit cells (15-20 Å). It was assumed, that the hence interactions with the periodic images in the other unit cells could be neglected and that the uniform potential of the UEG could be approximated as being zero. In this case, the energy difference between a compound B and the UEG can be expressed as

$$E^{\text{B}} - E^{\text{UEG}} = \Delta E = \underbrace{\sum_I Z_I^{\text{B}} \left(- \int_0^1 d\lambda \int d\mathbf{r} \frac{\rho(\lambda, \mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} + \sum_{j \neq I} \frac{Z_j^{\text{B}}}{2|\mathbf{R}_I - \mathbf{R}_j|} \right)}_{:=\Delta E_I}. \quad (68)$$

The first term in the brackets is the simplified expression of Eq. (66) with the term Eq. (67) and the summation over $\mathbf{L}$ omitted. Note that $\Delta Z_I = Z_I^{\text{B}} - Z_I^{\text{UEG}} = Z_I^{\text{B}}$ because no point charges are present in the UEG. The second term is the contribution due to nuclear repulsion. Since the summation over $\mathbf{L}$ is neglected, the decomposition in Eq. (61) can be used with all $Z_I^A = Z_I^{\text{UEG}} = 0$.

2.6 *Ab initio* atomic energy decompositions

Besides the alchemical decomposition presented in the previous section, other approaches to calculate atomic energies have been developed. In this section, two of these methods will be described to which the alchemical decomposition will be compared in chapter 5.

2.6.1 Mean field decomposition

It was proposed to decompose the energy into molecular orbital (MO) contributions and partition these MO contributions among the atoms based on atomic weights to the MOs. Atomic energies can then be obtained by summing up all MO contributions for one specific atom.⁶²

In detail this is achieved in the following way. The energy can be expressed as a function of one-electron reduced density matrices (RDM1s) for mean-field methods (Hartree-Fock or DFT) as

$$E(\mathbf{D}) = \sum_{\sigma=\alpha,\beta} \left(\text{Tr}[\mathbf{h}_{\text{core}} \mathbf{D}_{\sigma}] + \frac{1}{2} \text{Tr}[\mathbf{G}_{\sigma}(\mathbf{D}) \mathbf{D}_{\sigma}] \right) + h_{\text{nuc}} + (E_{\text{xc}}(\mathbf{D})), \quad (69)$$

where $\mathbf{D}_{\sigma} = \mathbf{C}_{\sigma} \mathbf{C}_{\sigma}^T$ is the spin-dependent RDM1, constructed from the molecular spin-orbitals $\mathbf{C}_{\sigma}$. $\mathbf{h}_{\text{core}} = \mathbf{L}_{\text{kin}} + \mathbf{V}_{\text{nuc}}$ is the one-particle Hamiltonian depending on kinetic energy $\mathbf{L}_{\text{kin}}$ and interaction between electrons and nuclei $\mathbf{V}_{\text{nuc}}$. $\mathbf{G}_{\sigma}$ describes the electron-electron interaction within Hartree-Fock (Coulomb and Fock-exchange), h_{nuc} is the nuclear repulsion and E_{xc} is the exchange-correlation energy. The RDM1 can be expressed as a sum of orbital-RDM1s (orb-RDM1s) $\mathbf{d}_{i,\sigma} = \mathbf{C}_{i,\sigma} \mathbf{C}_{i,\sigma}^T$ such that the Hartree-Fock energy E^{HF} is partitioned into orbital contributions as

$$E^{\text{HF}}(\mathbf{D}, \mathbf{d}) = \sum_{\sigma} \sum_i^{N_{\sigma}} \left(\text{Tr}[\mathbf{h}_{\text{core}} \mathbf{d}_{i,\sigma}] + \frac{1}{2} \text{Tr}[\mathbf{G}_{\sigma}(\mathbf{D}) \mathbf{d}_{i,\sigma}] \right) + h_{\text{nuc}}, \quad (70)$$

where N_{σ} is the number of spin-orbitals with spin σ . The energy of the exchange-correlation functional can be expressed in orbital dependent densities, which can be obtained from the $\mathbf{d}_{i,\sigma}$ (see Ref. 62). It was further proposed to generate the orb-RDM1s from a localized MO-basis such as intrinsic bond orbitals¹³⁸ instead of the canonical MOs.

The orb-RDM1s are partitioned into atomic contributions δ_K as

$$\delta_K = \sum_{\sigma} \delta_{K,\sigma} = \sum_{\sigma} \sum_i^{N_{\sigma}} \mathbf{d}_{i,\sigma} \mathbf{p}_{i,\sigma}^K, \quad (71)$$

where K is the index of the K -th atom and $\mathbf{p}_{i,\sigma}^K$ is the weighting coefficient. The weights can be determined with different methods such as Hirshfeld,¹³⁹ Bader¹⁴⁰ or Becke¹⁴¹ partitionings or the population analysis from Mulliken¹⁴² or intrinsic atomic orbitals¹³⁸.

The atomic contribution to the electronic energy is consequently defined as

$$E_{\text{el},K} = \text{Tr}[\mathbf{L}_{\text{kin}} \delta_K] + \frac{1}{2} (\text{Tr}[\mathbf{V}_K \mathbf{D}] + \text{Tr}[\mathbf{V}_{\text{nuc}} \delta_K]) + \frac{1}{2} \sum_{\sigma} \text{Tr}[\mathbf{G}_{\sigma}(\mathbf{D}) \delta_{K,\sigma}], \quad (72)$$

where $\mathbf{L}_{\text{kin}}$ is the kinetic energy part of the Hamiltonian, $\text{Tr}[\mathbf{V}_K \mathbf{D}]$ is the atomic energy due to the interaction of the K -th nucleus with the electrons and $\text{Tr}[\mathbf{V}_{\text{nuc}} \delta_K]$ is the interaction of the RDM1 contributed to the K -th atom with the external potential. The nuclear repulsion energy is partitioned as

$$E_{\text{nuc},K} = \frac{Z_K}{2} \sum_{L \neq K}^M \frac{Z_L}{|\mathbf{R}_K - \mathbf{R}_L|} \quad (73)$$

where $Z_{K/L}$ and $\mathbf{R}_{K/L}$ are nuclear charges and positions of atoms K and L respectively and the sum runs over all atoms M in the compound.¹⁴³

2.6.2 Interacting quantum atom

The interacting quantum atom approach is based on a partitioning of space into atomic subspaces $\{\Omega_A\}$, where each subspace is assigned to exactly one atom.^{60,61} The energy due to interactions within one of these subspaces Ω_A is the so-called intra-atomic energy E_A^{intra} and the energy due to pairwise interactions between the particles in different subspaces, Ω_A and Ω_B , yields the inter-atomic energy E_{AB}^{inter} . The energy of a compound, E^{mol} , can then be written as a sum of atomic contributions as

$$E^{\text{mol}} = \sum_A^M E_A^{\text{intra}} + \frac{1}{2} \sum_A^M \sum_{B \neq A}^M E_{AB}^{\text{inter}}, \quad (74)$$

where M is the number of atoms in the compound and the inter-atomic energies E_{AB}^{inter} is split equally between atoms A and B .

In the following, it will be briefly presented how the subspaces Ω_A are defined and how the energy contributions to these subspaces can be calculated.

ATOMIC SUBSPACES Space can be partitioned into non-penetrating or penetrating (also known as fuzzy) fragments.¹⁴⁴ In this thesis, the non-penetrating subspaces from the theory of quantum atoms in molecules will be used.¹⁴⁰ In this theory, the subspace Ω_A is enclosed by the surface with zero local flux of the electron density gradient field $\nabla\rho(\mathbf{r})$. This surface can be defined as $\nabla\rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = 0$ for all $\mathbf{r} \in S(\Omega_A)$, where $\mathbf{n}(\mathbf{r})$ is a vector normal to the surface $S(\Omega_A)$.

SPATIAL ENERGY CONTRIBUTIONS Energy contributions to subspace Ω_A and hence atom A can be defined with the help of the one- and two-particle density matrices as derived in more detail in Ref 61. Here, only the final expressions will be presented.

The kinetic energy and electron-nuclei interaction energy can be obtained from the one-particle matrix $\rho_1(1;1)$. The energy is divided into intra- and inter-atomic contributions by piecewise integration over the subspaces $\{\Omega_A\}$ as

$$\sum_A \int_{\Omega_A} d\mathbf{x}_1 \left(\hat{T} + \sum_B \hat{V}_{\text{en}}^B \right) \rho_1(1;1) = \sum_A T_A + \sum_A \sum_B V_{\text{en}}^{AB}, \quad (75)$$

where $\hat{T}$ is the kinetic energy operator and $\hat{V}_{\text{en}}^B$ is the external potential due to nucleus B , e.g. $V_{\text{en}}^B = \frac{Z_B}{|\mathbf{r} - \mathbf{R}_B|}$. T_A is the intra-atomic kinetic energy contribution of subspace Ω_A and V_{en}^{AB} is the interaction energy between electrons in subspace Ω_A and the potential of nucleus B . If $A = B$ it is intra-atomic and if $A \neq B$ it is inter-atomic. The inter-atomic contributions are divided equally between A and B , e.g. atom A has the energy contributions $\frac{1}{2} V_{\text{en}}^{AB}$ and $\frac{1}{2} V_{\text{en}}^{BA}$ due to the interaction of the electrons in Ω_A with nucleus B and of nucleus A with the electrons in Ω_B , respectively.

The energy contribution from the two-particle density matrix $\rho_2(1;2)$ can be divided into inter-atomic contributions by integration over subspaces Ω_A and Ω_B as

$$V_{\text{ee}} = \sum_A V_{\text{ee}}^{AA} + \sum_A \sum_{B \neq A} \frac{1}{2} V_{\text{ee}}^{AB}, \quad (76)$$

where V_{ee}^{AA} is the intra-atomic electron-electron repulsion in Ω_A and V_{ee}^{AB} is the inter-atomic electron-electron repulsion between the electrons in Ω_A and Ω_B . They are defined as

$$V_{ee}^{AA} = \frac{1}{2} \int_{\Omega_A} \int_{\Omega_A} \rho_2(1;2) \frac{1}{r_{12}} d\mathbf{x}_1 d\mathbf{x}_2 \quad (77)$$

and

$$V_{ee}^{AB} = \int_{\Omega_A} \int_{\Omega_B} \rho_2(1;2) \frac{1}{r_{12}} d\mathbf{x}_1 d\mathbf{x}_2, \quad (78)$$

where r_{12} is the distance between electron 1 and electron 2.

Furthermore, the nuclear repulsion energy is decomposed into inter-atomic energies as

$$V_{nn} = \frac{1}{2} \sum_A \sum_{A \neq B} V_{nn}^{AB} = \frac{1}{2} \sum_A \sum_{A \neq B} \frac{Z_A Z_B}{R_{AB}}, \quad (79)$$

where R_{AB} is the distance between A and B.

The atomic energy of atom A is then defined as

$$E_A = T_A + V_{en}^{AA} + V_{ee}^{AA} + \frac{1}{2} \sum_{B \neq A} (V_{en}^{AB} + V_{en}^{BA} + V_{ee}^{AB} + V_{nn}^{AB}) \quad (80)$$

with the definitions of the individual terms given above.

In this chapter, atomic energies will be calculated for 15 small molecules with the approximate atomic energy expression Eq. (68) (see section 2.5). Furthermore, a modified alchemical energy decomposition scheme, explicitly considering the periodic boundary conditions and the external potential of the UEG will be derived. Atomic energies from the modified model will then be compared to the ones from Eq. (68). The impact of the approximations in Eq. (68) on atomic energy trends will be discussed.

In the second part of the chapter, the practical computation of the atomic energies with the modified decomposition scheme will be investigated. In particular, different numerical techniques for the calculation of the necessary derivatives and integrals are compared. Based on this analysis a computationally efficient approach for the calculation of atomic energies is proposed. Furthermore, the impact of finite size effects on the atomic energies will be discussed.

3.1 An alchemical energy decomposition for periodic systems

Atomic energies were calculated with Eq. (68) for the molecules HCN, ethene, ethyne and for twelve molecules of structure $\text{CH}_3\text{-YH}_\gamma$ with $\text{Y} = \text{C}, \text{N}, \text{O}, \text{F}, \text{Si}, \text{P}, \text{S}, \text{Cl}, \text{Ge}, \text{As}, \text{Se}, \text{Br}$ and γ being the number of hydrogens such that all valencies of Y are saturated (15 molecules in total). In order to obtain the atomic energies, the electron density was calculated with PBE¹¹⁷ for 21 equally spaced points along the λ -path ($\Delta\lambda = 0.05$) from the UEG ($\lambda = 0$) to the respective compound ($\lambda = 1$). The integral over $\mathbf{r}$ in Eq. (68) was only evaluated over a single unit cell for the electron densities $\rho(\lambda, \mathbf{r})$. Since the PBE calculations were carried out with pseudopotentials, the nuclear valence charges $Z_{I,V}$ were used instead of the full nuclear charges. The atomic energies were then summed up for each molecule and are plotted against the exact energy difference^a between molecule and UEG, ΔE^{exact} , in Figure 4. The sum of atomic energies $\sum_I \Delta E_I$ is on average about 8 Ha lower than the exact energy difference. The main reason for the systematic shift is the neglected uniform potential of the UEG. This potential has a positive contribution to ΔE (see Eq. (66) and (67)) and would shift the sum of atomic energies to more positive values. Furthermore, the deviation between $\sum_I \Delta E_I$ and ΔE^{exact} differs slightly for the 15 molecules. This is likely due to the neglected interaction with the periodic images in the other unit cells in Eq. (68). This interaction is not the same for all molecules leading to slightly different errors as observed in Figure 4.

In order to include the potential of the UEG and the interaction with periodic images, the atomic energy decomposition should be derived using the

^a The exact energy difference is obtained from the total energies of the PBE calculations at $\lambda = 1$ for the molecule and $\lambda = 0$ for the UEG.

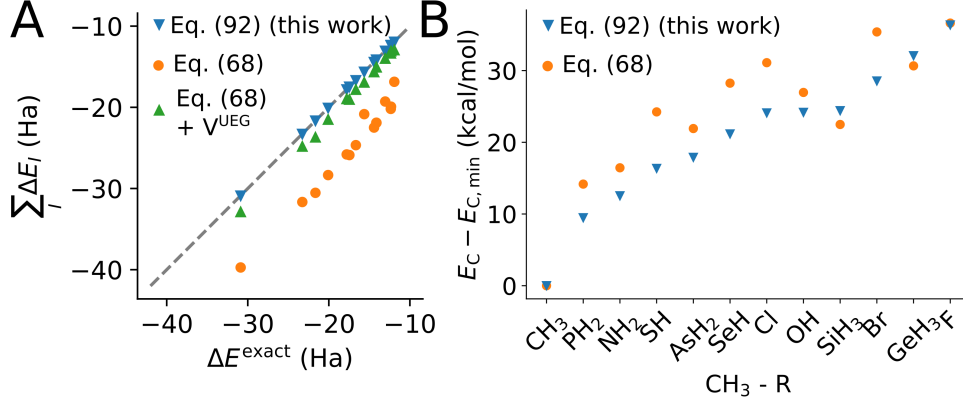


Figure 4: Panel A: The sum of atomic energies for 15 different molecules vs the exact molecular energy with respect to the uniform electron gas, ΔE^{exact} , computed including periodic boundary conditions and background potential via Eq. (92) (this work), with Eq. (68) or with Eq. (68) including the interaction with the uniform background charge. Panel B: Relative atomic energies of carbon in $\text{CH}_3\text{-R}$ calculated for 12 different residues R with Eq. (92) and (68). Adapted from *J. Chem. Phys.* 160, 054106 (2024), with the permission of AIP Publishing.

external potential as defined in Eq. (65). In this work, we propose to generalize this potential to

$$V_{\text{ext}}^{\text{pbc}}(\mathbf{r}, \lambda) = \sum_{\mathbf{L}} \sum_I V_I(\lambda, \mathbf{r} - \mathbf{R}_I - \mathbf{L}) + \rho^{\text{bg}}(\lambda) V^{\text{bg}}(\mathbf{r} - \mathbf{L}), \quad (81)$$

where the superscript pbc indicates the use of periodic boundary conditions, $\mathbf{L}$ is the lattice vector, $V_I(\lambda, \mathbf{r} - \mathbf{R}_I - \mathbf{L})$ is the λ - and lattice vector dependent potential due to nucleus I and its periodic images, $\rho^{\text{bg}}(\lambda)$ is the uniform background charge of the UEG and V^{bg} is the normalized potential due to a positive uniform charge distribution.

In this work, the coulombic nuclear potentials V_I are approximated by pseudopotentials V_I^{PP} from Goedecker, Teter and Hutter.^{134,135} They can be expressed analytically as

$$V_I^{\text{PP}}(\lambda, r) = V_{I,\text{loc}}(\lambda, r) + V_{I,\text{nl}}(\lambda, \mathbf{r}, \mathbf{r}') \quad (82)$$

with

$$V_{I,\text{loc}}(\lambda, r) = -\frac{Z_V(\lambda)}{r} \text{erf}(\alpha^{\text{PP}} r) + \sum_i C_i^{\text{PP}}(\lambda) (\sqrt{2} \alpha^{\text{PP}} r) \exp\left\{-\left(\alpha^{\text{PP}} r\right)^2\right\} \quad (83)$$

and

$$V_{I,\text{nl}}(\lambda, \mathbf{r}, \mathbf{r}') = \sum_{lm} \sum_{ij} \langle \mathbf{r} | p_i^{lm} \rangle h_{ij}^l(\lambda) \langle p_j^{lm} | \mathbf{r}' \rangle. \quad (84)$$

For a definition of the parameters see section 2.3.1.

The λ -dependence of the pseudopotential parameters follows from the linearly coupled Hamiltonians of molecule and UEG (see Eq. (39)) as: $Z_V(\lambda) = \bar{Z}_V \lambda$, $C_i^{\text{PP}}(\lambda) = \bar{C}_i^{\text{PP}} \lambda$ and $h_{ij}^l(\lambda) = \bar{h}_{ij}^l \lambda$. $\bar{Z}_V$, $\bar{C}_i^{\text{PP}}$ and $\bar{h}_{ij}^l$ denote the optimized

parameters reported in Ref. 134,135. With this definition, the pseudopotentials are zero for $\lambda = 0$ and equal to the potentials of the respective elements at $\lambda = 1$.

The background charge $\rho^{\text{bg}}(\lambda)$ is equal to N_e/Ω_{cell} at $\lambda = 0$, where N_e is the number of valence electrons in the molecule and Ω_{cell} is the volume of the unit cell. It is scaled down linearly to zero at $\lambda = 1$, while the nuclear valence charges $Z_{I,V}(\lambda)$ are simultaneously scaled up. Thus, $\rho^{\text{bg}}(\lambda)$ can be expressed as a function of the $Z_{I,V}(\lambda)$ as

$$\rho^{\text{bg}}(\lambda) = \frac{N_e - \sum_I Z_{I,V}(\lambda)}{\Omega_{\text{cell}}} \quad (85)$$

with $Z_{I,V}(\lambda) = \bar{Z}_{I,V}\lambda$ and $\bar{Z}_{I,V}$ being the nuclear valence charge of the element at site I in the compound. The background potential V^{bg} is defined per unit cell

$$V^{\text{bg}}(\mathbf{r}) = - \int_{\Omega_{\text{cell}}} d\mathbf{r}' \frac{1}{|\mathbf{r} - \mathbf{r}'|}. \quad (86)$$

The nuclear repulsion energy E^{nuc} depends on the interaction between the nuclei E^{NN} , between nuclei and background potential E^{Nbg} and the self-interaction of the background E^{bg} such that

$$E^{\text{nuc}} = E^{\text{NN}} + E^{\text{Nbg}} + E^{\text{bg}} \quad (87)$$

with

$$E^{\text{NN}} = \frac{1}{2} \sum_{I,J} \sum_{\mathbf{L}} \frac{Z_{I,V}(\lambda) Z_{J,V}(\lambda)}{|\mathbf{R}_I - \mathbf{R}_J - \mathbf{L}|}, \quad (88)$$

where the self-interaction terms for $\mathbf{R}_I = \mathbf{R}_J$ at $\mathbf{L} = 0$ are excluded,

$$E^{\text{Nbg}} = \sum_{I,\mathbf{L}} -Z_{I,V}(\lambda) \rho^{\text{bg}}(\lambda) V^{\text{bg}}(\mathbf{R}_I - \mathbf{L}) \quad (89)$$

and

$$E^{\text{bg}} = \frac{1}{2} \rho^{\text{bg}}(\lambda)^2 \sum_{\mathbf{L}} - \int_{\Omega_{\text{cell}}} d\mathbf{r} V^{\text{bg}}(\mathbf{r} - \mathbf{L}). \quad (90)$$

Note, that

$$E^{\text{Nbg}}(\lambda = 1) - E^{\text{Nbg}}(\lambda = 0) = \int_0^1 d\lambda \frac{\partial E^{\text{Nbg}}}{\partial \lambda} = 0 \quad (91)$$

since the either nuclear charges or the background charge and hence also their interaction energy are zero at $\lambda = 0$ respectively $\lambda = 1$. Hence, only E^{bg} and E^{NN} contribute to E^{nuc} .

The atomic energy expression for the periodic external potential Eq. (81) and the nuclear repulsion as defined in Eq. (87) can be derived analogously to the derivation in sections 2.4.2 and 2.5 by utilizing the Hellmann-Feynman theorem^{136,137} and chain rule. This results in

$$E^{\text{mol}} - E^{\text{UEG}} = \sum_I \int_0^1 \underbrace{\left(\frac{\partial E^{\text{nuc}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} + \int d\mathbf{r} \Delta V_I \rho_\lambda(\mathbf{r}) \right)}_{:= \frac{\partial E_I}{\partial \lambda}} d\lambda \quad (92)$$

where $\Delta V_I = \sum_{\mathbf{L}} V_I^{\text{PP}}(\lambda = 1, \mathbf{r} - \mathbf{R}_I - \mathbf{L}) - \bar{Z}_{I,V}/\Omega_{\text{cell}} V^{\text{bg}}(\mathbf{r} - \mathbf{L})$ and $\bar{Z}_{I,V}$ is the nuclear valence charge of the I -th nucleus at $\lambda = 1$. A detailed derivation is given in A.1.2 section I. The practical calculation of atomic energies with Eq. (92) will be described in the following sections.

However, before delving into these details, the atomic energies obtained from Eq. (92) and Eq. (68) are compared.

The mean absolute error (MAE) in the sum of atomic energies with respect to the exact energy difference is greatly reduced if Eq. (92) is used (MAE = 0.004 Ha vs 8 Ha with Eq. (68)). As mentioned before, the main error source in Eq. (68) is the neglected interaction between background potential and electron density. Hence, the MAE can be reduced to 1.3 Ha if the interaction between the electron density and the background charge in the same unit cell and of the background charge with itself is added to Eq. (68) (see Figure 4, Eq. (68) + V^{UEG}). This correction is identical to evaluating Eq. (92) only for $\mathbf{L} = 0$. Thus, the remaining deviation between the Eq. (92) and Eq. (68) should be related to the interaction with periodic images, which is only considered in Eq. (92) by summation over all lattice vectors $\mathbf{L}$. In Figure 4B, the relative atomic energies of carbon in the 12 molecules of structure $\text{CH}_3\text{-YH}_y$ are plotted as obtained from Eq. (68) and (92). The atomic energies are shifted as $E_{\text{C}} - E_{\text{C, min}}$, where $E_{\text{C, min}}$ is the atomic energy of carbon in ethane, which has the lowest energy of all carbon atoms. The relative atomic energy trends can already be recovered to a large extent by Eq. (68) as observable in the plot.

3.2 Alchemical derivatives

The calculation of analytical derivatives $\partial_{\lambda} E_I$, requires the summation over the lattice vectors $\mathbf{L}$. However, the individual nuclear repulsion and electron-nucleus attraction terms in Eq. (92) diverge for the summation over an infinite number of lattice vectors. The attractive and repulsive terms must be added up in the correct order such that the divergences cancel as has been described for example in Refs 145–148. The correct implementation of these equations is technically complicated, in particular if pseudopotentials are used instead of coulomb potentials. Hence in this thesis, the derivatives are evaluated numerically via finite differences or perturbatively.

$\partial_{\lambda} E_I$ can be calculated via central finite differences as

$$\frac{\partial E_I}{\partial \lambda} \approx \frac{E[V_I(\lambda + \Delta\lambda)] - E[V_I(\lambda - \Delta\lambda)]}{2\Delta\lambda}. \quad (93)$$

$E[V_I(\lambda \pm \Delta\lambda)]$ is the energy of a system where $V_I(\mathbf{r}, \lambda \pm \Delta\lambda)$ is the pseudopotential of the I -th nucleus (see Eq. (81)-(84)) and $V_J(\mathbf{r}, \lambda)$ denotes the pseudopotentials of the other nuclei. The pseudopotentials are rescaled by $\lambda + \Delta\lambda$ and λ , respectively, by scaling the λ -dependent parameters in Eqs. 83 and 84 ($\Delta\lambda = 5 \cdot 10^{-5}$ in this work).

A drawback of the finite difference approach is that for each atom and integration point along the λ -path two additional self-consistent calculations must be carried out. To reduce the computational cost, the derivatives can also be evaluated perturbatively as

$$\frac{\partial E_I}{\partial \lambda} = \frac{E[V_I(\lambda + \Delta\lambda), \rho_{\lambda}] - E[V_I(\lambda - \Delta\lambda), \rho_{\lambda}]}{2\Delta\lambda}. \quad (94)$$

In the term $E[V_I(\lambda + \Delta\lambda), \rho_\lambda]$, the potential of the I -th nucleus is perturbed as for the finite difference approach. However, the energy is evaluated for the density ρ_λ at $H(\lambda)$ instead of reoptimizing it for the perturbed external potential. Hence, the same density ρ_λ can be used to obtain the derivatives $\partial_\lambda E_I$ for all atoms. This perturbative ansatz has been employed before^{46,59} and should recover the analytical derivative independent of the chosen $\Delta\lambda$ -value as shown in A.1.2 section II. However, in practice derivatives from Eq. (94) are $\Delta\lambda$ -dependent when evaluated with the electronic structure code CPMD.¹⁴⁹ As larger $\Delta\lambda$ -values are chosen the derivatives deviate more and more from finite difference derivatives evaluated for $\Delta\lambda = 5 \cdot 10^{-5}$. The reason for this might be the practical implementation of the energy expression in presence of a uniform background potential in CPMD. Nevertheless, for $\Delta\lambda = 5 \cdot 10^{-5}$ derivatives from finite difference or the perturbative approach yield almost the same atomic energies. The mean absolute difference in atomic energy for the 91 atoms in the 15 molecules mentioned above is 0.03 mHa. Eq. (93) and (94) yield also similar results for 55 molecules with 10-20 atoms and 38 valence electrons from the QM9 database.¹⁹ The mean absolute deviation between atomic energies amounts to 0.02 mHa, indicating that the accuracy of the perturbative approach is not affected by the size of the molecule.

Although out of the scope of this work, the finite difference approach might also be useful to calculate higher order derivatives $\partial_\lambda^n E$ with $n > 1$. For example, second order derivatives $\partial_\lambda^2 E$ can be obtained from the second order derivatives with respect to the nuclear charges (or from the derivatives with respect to pseudopotential parameters if pseudopotentials are used).⁵⁰ These derivatives could then be used to optimize heterogeneous catalysts by providing adsorption energy estimates of a molecule on the catalyst surface for various catalyst compositions. This is achieved by estimating adsorption energies via a Taylor expansion of the energy of the system in λ , where variations in λ correspond to variations in the nuclear charges of the catalyst (see Ref 57 for details). It was demonstrated that this approach yields adsorption energy estimates with errors in the order of 0.02-0.24 eV for metallic surfaces by using only first order derivatives $\partial_\lambda E$.⁵⁵ However, the predictions were shown to be significantly less accurate for catalysts with a band gap.¹⁵⁰ It was argued that predictions could be improved by inclusion of the second order derivatives $\partial_\lambda^2 E$.¹⁵⁰ However, to the best of my knowledge, such calculations have not been performed yet. The finite difference approach presented in this thesis provides a straightforward route to obtain second order derivatives with respect to the pseudopotential parameters. Hence, it could possibly lead to improved adsorption energy predictions for heterogeneous catalysts with a band gap.

λ -dependence of atomic derivatives

Atomic derivatives $\partial_\lambda E_I$ evaluated from central finite difference as a function of λ are shown for selected elements X in Figure 5, where X has a single bond to a methyl group and remaining valencies are saturated with hydrogens, e.g. the compounds are of the form $\text{CH}_3\text{-XH}_x$. The left panel shows the derivatives for different elements in the second row of the periodic table and the right panel shows the derivatives for elements in the VI-th main group of the periodic table. The shape of the curves is similar for all elements. The

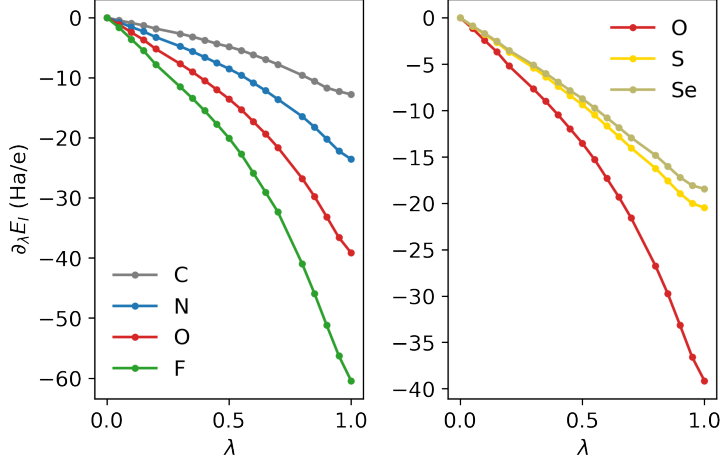


Figure 5: The alchemical atomic derivatives $\partial_\lambda E_I$ as function of λ for elements in the second row (left panel) or the VI-th main group (right panel) of the periodic table. The derivatives of nucleus I are evaluated for structures of type $\text{CH}_3\text{-IH}_i$, where i is chosen such that all valencies of I are saturated.

derivatives decrease monotonously starting from $\partial_\lambda E_I = 0$ at $\lambda = 0$ with the second order derivative $\partial_\lambda^2 E_I$ reaching a minimum at $\lambda \approx 0.9$. A comparison of $\partial_\lambda E_I(\lambda)$ for the elements in the second row of the periodic table shows that for a fixed value of λ , $\partial_\lambda E_I(\lambda)$ is more negative for a higher nuclear charge of the element. The opposite trend is observed for the elements in the same main group - from oxygen to selen $\partial_\lambda E_I(\lambda)$ becomes more positive for a given value of λ .

These characteristics can be understood qualitatively from the definition of $\partial_\lambda E_I(\lambda)$ in Eq. (92). The term $\int d\mathbf{r} \Delta V_I \rho(\lambda, \mathbf{r})$ in this equation becomes more negative as more density accumulates around the I -th nucleus resulting in the observed behaviour with increasing λ . Furthermore, more electron density accumulates around elements with a higher electronegativity which is responsible for the more negative $\partial_\lambda E_I$ for increased nuclear charge in the same row (Figure 5 left panel) and lower nuclear charge in the same main group (Figure 5 right panel). It should be noted that the alchemical derivative within the same main group would become more negative with increasing nuclear charge for the all electron case, while the observed opposite trend is due to the use of pseudopotentials.

It is also noteworthy, that all derivatives $\partial_\lambda E_I = 0$ at $\lambda = 0$. This can be explained by the the variation in energy due to a variation in nuclear charge density at some point $\mathbf{r}$ being equal to the electrostatic potential at $\mathbf{r}$ as shown in Ref 44. At $\lambda = 0$ the system consist of an uniform electron density and the positive uniform background charge. Consequently, the electrostatic potential is zero at every point in space and $\partial_\lambda E_I = 0$.

3.3 Integration over the transformation path

The integral over λ in Eq. (92) is calculated by numerical integration. The performance of four different integration methods, the trapezoidal rule, Simpson's rule, cubic splines and gaussian quadrature, have been compared. The mean absolute deviation between the sum of atomic energies and the exact

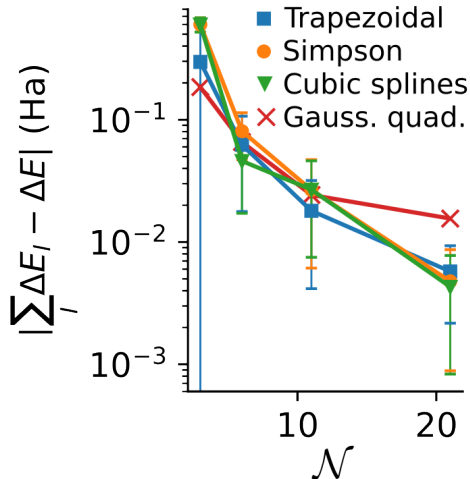


Figure 6: Mean absolute error (MAE) of $\sum_I \Delta E_I$ vs ΔE for the 15 molecules mentioned in the text for an increasing number of integration points $\mathcal{N}$ along the λ -path using different integration methods. For gaussian quadrature only the MAE of ethane is plotted. Adapted from *J. Chem. Phys.* 160, 054106 (2024), with the permission of AIP Publishing.

energy difference between UEG and molecule, $|\sum_I \Delta E_I - \Delta E^{\text{exact}}|$, is used to assess the accuracy of the integration method since the deviation should be zero if the integration (and calculation of derivatives) is exact. Figure 6 shows the MAE, $|\sum_I \Delta E_I - \Delta E^{\text{exact}}|$, for the different numerical integration methods and increasing number of integration points $\mathcal{N}$. The MAE is averaged over the 15 test molecules for trapezoidal rule, Simpson’s rule and cubic splines but was only calculated for ethane for gaussian quadrature. For the trapezoidal rule, Simpson’s rule, cubic splines the integration points are placed equidistantly between 0 and 1 and $\partial_\lambda E_I$ is set to zero at $\lambda = 0$ as justified in the previous section. In case of the gaussian quadrature, the integration points are the roots of the Legendre polynomials shifted from the interval $[0,1]$ to the interval $[-1, 1]$.

The MAE decreases for all methods as the number of grid points is increased. Trapezoidal rule, Simpson rule and cubic splines yield similar errors (e.g. 5.9, 4.5 and 4.4 mHa for 21 grid points). Gaussian quadrature shows a similar performance to the other method except for 21 grid points, where it has a significantly larger error. The reason for this is the higher error in the nuclear repulsion term $\int_0^1 d\lambda \frac{\partial E}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda}$. This error is in the order of 10^{-4} Ha for the other methods but 10^{-2} Ha for gaussian quadrature. The errors for the integral $\int d\lambda \int d\mathbf{r} \Delta V_I \rho(\lambda, \mathbf{r})$ are similar for all four integration methods.

Even for 21 integration points the error is above chemical accuracy (1.6 mHa) for all integration methods. It should be possible to reduce the error further by inclusion of more integration points. However, this approach is computationally not feasible if the atomic energies should be calculated for a large dataset of molecules. In particular, the slow convergence of the calculations for $\lambda < 1$ (often more than 100 SCF iterations) make this procedure computationally expensive. Fortunately, trends in atomic energy for varying local environment seem to converge faster than the overall MAE. This is illustrated in Figure 7 by ranking the atomic energies of carbon in the molecules in the test set by

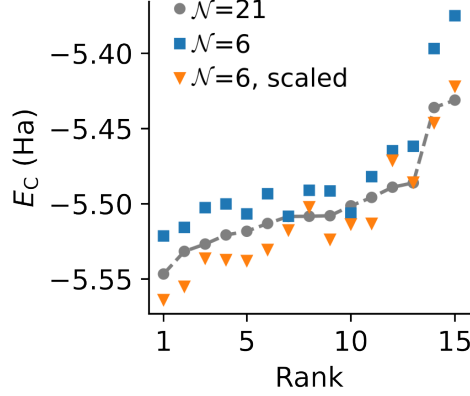


Figure 7: Comparison of atomic energy rankings for carbon atoms in 15 different molecular environments with varying number of integration points $\mathcal{N}$ and rescaling according to Eq. (95). Adapted from *J. Chem. Phys.* 160, 054106 (2024), with the permission of AIP Publishing.

increasing atomic energy. The energy ranking for 21 grid points and integration with Simpson’s rule appears similar to the one for 6 integration points. The ranking similarity can be quantified by the Kendall rank coefficient τ .¹⁵¹ A coefficient of $\tau = 1$ would indicate that the atomic energies are ranked identically for 21 and 6 integration points. $\tau = -1$ would indicate maximally different rankings, e.g. a negative correlation between the rankings. $\tau = 0$ would indicate that no correlation exists between the rankings at all. In case $\tau = 0.77$ indicating a good agreement between the atomic energy rankings for the coarse and fine grid.

While a qualitative interpretation of the relation between atomic energy and local environment is possible for coarse integration grids, the results might not be sufficient to develop of quantitatively accurate models, e.g. if the atomic energies should be used as training labels of a machine learning model. Exact agreement between the sum of atomic energies and ΔE^{exact} can be achieved by rescaling of the atomic energies as

$$\Delta \tilde{E}_I^{\mathcal{N}} = \Delta E^{\text{exact}} \frac{\Delta E_I^{\mathcal{N}}}{\sum_J \Delta E_J^{\mathcal{N}}}, \quad (95)$$

where $\Delta E_I^{\mathcal{N}}$ is the atomic energy obtained after integration over $\mathcal{N}$ points along λ , $\sum_J \Delta E_J^{\mathcal{N}}$ is the sum of all atomic energies in the molecule numerically integrated over $\mathcal{N}$ points and ΔE^{exact} is the exact energy difference between the molecule and the UEG. The rescaling of atomic energies integrated over 6 points does not significantly affect the atomic energy ranking (see Figure 7 but even slightly improves the agreement with the energy ranking at $\mathcal{N} = 21$ ($\tau = 0.85$) compared to unscaled energies. Thus, the combination of a coarse integration grid and rescaling might be a viable option for the calculation of atomic energies for large datasets and the utilization of these energies for development of quantitative models.

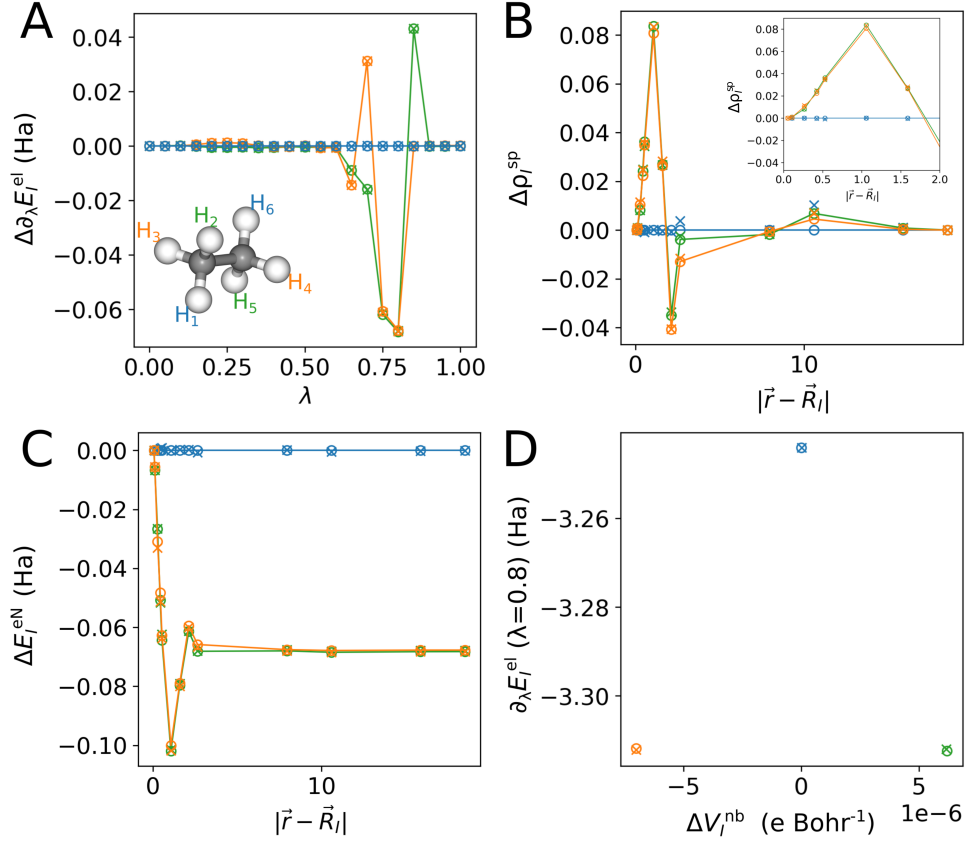


Figure 8: The origin of different atomic energies for the symmetrical hydrogen atoms in ethane. Colors in the plots correspond to the hydrogens H₁-H₆ as illustrated in the inset of A. Δ indicates that the respective quantities are with respect to H₁. **Panel A:** $\partial_{\lambda} E_I^{\text{el}}$ along the transformation path. **Panel B:** The density in spheres with increasing radius $|\vec{r} - \vec{R}_I|$ around the hydrogens (see Eq.(96)). **Panel C:** Interaction energy between electron density and potential V_I in spheres with radius $|\vec{r} - \vec{R}_I|$ (see Eq.(97)). **Panel D:** The relation between $\partial_{\lambda} E_I^{\text{el}}$ in the central unit cell and the external potential V_I^{nb} due to the nuclei in the 26 surrounding unit cells.

Atom	E_I (mHa)	$E_I - E_1$ (mHa)
H ₁	-506.0	0.0
H ₂	-511.9	-5.9
H ₃	-511.8	-5.8
H ₄	-511.8	-5.8
H ₅	-511.8	-5.9
H ₆	-506.3	-0.4

Table 1: Atomic energies of the symmetrical hydrogen atoms in ethane. The indices are the same as for the hydrogens in Figure 8.

3.4 Finite size effects

In the previous sections, it was investigated to which extent the sum of atomic energies agrees with the exact energy difference between a molecule and the UEG. The decomposition scheme should also recover the symmetry of the compound, e.g. the atomic energies of symmetrically identical atoms should be identical, as well.

Exemplarily, this is studied for the atomic energies of the six symmetrically identical hydrogen atoms in ethane (see Table 1). The values can differ by up to 6 mHa in apparent violation of the symmetry of the molecule. This error should not be caused by the numerical integration in λ , since this does not affect the spatial symmetry of the electron density and the external potential.

To identify the reason for the unsymmetrical atomic energies, the difference in derivatives $\Delta\partial_\lambda E_I^{\text{el}} = \partial_\lambda E_I^{\text{el}} - \partial_\lambda E_1^{\text{el}}$ is plotted as a function of λ in Figure 8A. $\partial_\lambda E_1^{\text{el}}$ denotes the derivative of hydrogen H_1 (see inset of Figure 8A). The derivatives differ strongly in the interval $\lambda = [0.6, 0.85]$. In this range only the inversion symmetry is retained, resulting in three pairs of hydrogens with (nearly) identical derivatives (H_1/H_6 , H_2/H_5 , H_3/H_4). Since the derivatives are calculated as $\partial_\lambda E_I^{\text{el}} = \int d\mathbf{r} \Delta V_I \rho(\lambda, \mathbf{r})$, deviations should be either due to unsymmetrical densities $\rho(\lambda, \mathbf{r})$ or potentials ΔV_I . To investigate possible deviations in $\rho(\lambda, \mathbf{r})$ from the expected symmetry, the electron density $\rho(\lambda = 0.8)$ is spherically integrated up to a certain distance $|\mathbf{r} - \mathbf{R}_I|$ around each hydrogen atom, e.g.

$$\rho^{\text{sp}}(r) = \int_0^{|\mathbf{r}-\mathbf{R}_I|} d\mathbf{r}' \rho(\lambda = 0.8, \mathbf{r}'). \quad (96)$$

The difference $\Delta\rho^{\text{sp}}$ between $\rho^{\text{sp}}(r)$ and the same integral for H_1 is plotted in Figure 8B for distances $|\mathbf{r} - \mathbf{R}_I|$ in the range of 0.05 to 18.5 Å. The number of electrons within a radius of ≈ 1.5 Å, e.g. close to the core, is higher for the hydrogens H_2 , H_3 , H_4 and H_5 than for H_1 and H_6 s. Hence, the derivatives $\partial_\lambda E_I^{\text{el}} = \int \Delta V_I \rho d\mathbf{r}$ of the hydrogens H_{2-5} have a more negative contribution from this region. This is shown in Figure 8C by plotting

$$\Delta E_I^{\text{eN}}(r) = \int_0^{|\mathbf{r}-\mathbf{R}_I|} d\mathbf{r}' V_I \rho(\lambda = 0.8, \mathbf{r}') - \int_0^{|\mathbf{r}-\mathbf{R}_1|} d\mathbf{r}' V_1 \rho(\lambda = 0.8, \mathbf{r}') \quad (97)$$

for distances $|\mathbf{r} - \mathbf{R}_I|$ in the range of 0.05 to 18.5 Å. Consequently, the different atomic energies for the symmetrical hydrogens in ethane can be attributed to differences in the electron density around the hydrogen atoms.

The reason for the different electron distribution are likely different interactions with the periodic images. Since the unit cell has not the same symmetry as the molecule, the hydrogen atoms are in fact not symmetrical anymore if the periodic system is considered. The external potential V_I^{nb} at the positions of the hydrogen atoms due to the nuclei in the 26 unit cells surrounding the cell at $\mathbf{L} = 0$ is not identical for the hydrogen atoms. This is illustrated in Figure 8D, where $\Delta V_I^{\text{nb}} = V_I^{\text{nb}} - V_1^{\text{nb}}$ is plotted against $\partial_\lambda E_I^{\text{el}}$ (V_1^{nb} is the potential at the position of H_1 due to the nuclei in the neighboring unit cells). The hydrogen atoms are exposed to three slightly different external potentials with the potential of each of the pairs H_1/H_6 , H_2/H_5 and H_3/H_4 being nearly identical. The same pairs of hydrogen have also the same derivatives

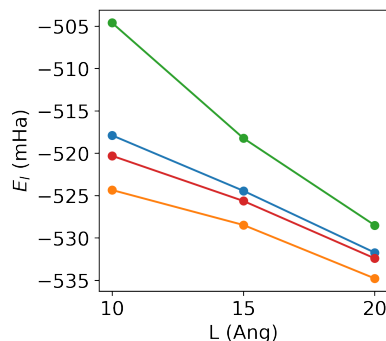


Figure 9: The atomic energies calculated with the alchemical decomposition scheme of the four symmetrical hydrogen atoms in methane as a function of the cubic cell length L .

with respect to λ (see Figure 8A) and the same electron distributions ρ^{sp} (see Figure 8B). It is surprising, that the small difference in external potential in the order of $10^{-6} \text{ e Bohr}^{-1}$ could lead to differences in atomic energies of 6 mHa. Furthermore, it is unclear why the derivatives differ only in the interval $\lambda = [0.6, 0.85]$ although the differences in external potentials V_I^{nb} also exist at other λ -values. A more in depth study of this behaviour is however out of the scope of this work.

An approach to mitigate finite size effects is to increase the cell size. The dependence of the atomic energies on the cell size is studied exemplarily for methane in Figure 9. The atomic energies of the four symmetrical hydrogen atoms become more similar as the length of the cubic cell is increased from $10 \text{ \AA}$ over $15 \text{ \AA}$ to $20 \text{ \AA}$. However, even at $20 \text{ \AA}$ the difference between the atomic energies is rather large (up to 7 mHa). It might be possible to extrapolate to infinite cell size by studying the atomic energies as function of the cell size. However, this is quite expensive as it requires the calculation of atomic energies for various infinite sizes. Thus, an increase of the cell size alone does not seem to be a promising strategy to deal with finite effects, in particular if the alchemical decomposition scheme should be applied to a large dataset of molecules.

Nevertheless, the variation in atomic energy due to different local environments can often still be resolved since it tends to be larger than the energy difference between the symmetrical hydrogen atoms. For example, the average atomic energy of hydrogen in ethane is 14 mHa lower than the atomic energy of hydrogen in the NH_2 group and the atomic energy of hydrogen in NH_2 -group is 16 mHa lower than of hydrogen in the OH-group of CH_3OH . In contrast, the atomic energies of the hydrogens in ethane differ by at most 6 mHa.

3.5 Summary

In this chapter, the atomic energy decomposition scheme from Ref. 86 (see Eq. (68)) has been extended to include periodic boundary conditions and the potential of the UEG. The extended decomposition was compared to the model in Ref. 86. In particular, only the extended model recovers the energy difference between a molecule and the UEG. Furthermore, the numerical as-

pects of atomic energy calculations have been thoroughly investigated. Based on this analysis, approximations for the efficient calculation of atomic energies have been proposed.

In the following two chapters, the alchemical energy decomposition scheme will be used to investigate structure energy relationships. The goal of these applications is the development of simple models to estimate the stability of compounds throughout chemical space.

In this chapter, the alchemical energy decomposition scheme is used to investigate the impact of the elemental composition on the stability of chemical bonds. This is done by calculating atomic contributions to the bond dissociation energy (BDE) for homolytic bond cleavage reactions of type



with A and B being any combination of the elements carbon, nitrogen, oxygen and fluorine and x, y chosen such that all remaining valencies are saturated by hydrogen atoms.

The bond dissociation energy E_{AB} is defined as

$$E_{\text{AB}} = E^{\text{tot}}(\text{A}) + E^{\text{tot}}(\text{B}) - E^{\text{tot}}(\text{AB}) \quad (99)$$

where $E^{\text{tot}}(\text{A})$, $E^{\text{tot}}(\text{B})$ and $E^{\text{tot}}(\text{AB})$ are the total energies of the fragments $\text{H}_x\text{A}^\bullet$, $\text{H}_x\text{B}^\bullet$ and the molecule $\text{H}_x\text{A} - \text{BH}_y$, respectively.

The contribution of, for example atom A, to the BDE can be calculated as

$$E_{\text{AB}}(\text{A}) = E^{\text{tot}}(\text{A/A}) - E^{\text{tot}}(\text{A/AB}) \quad (100)$$

where $E_{\text{AB}}(\text{A})$ is the atomic contribution of A to the BDE, $E^{\text{tot}}(\text{A/A})$ is the atomic energy of A in the fragment $\text{H}_x\text{A}^\bullet$ and $E^{\text{tot}}(\text{A/AB})$ is the atomic energy of atom A in the compound $\text{H}_x\text{A} - \text{BH}_y$.

The relationship between this atomic BDE $E_{\text{AB}}(\text{A})$ and the binding partner B will be analysed in this chapter. However, before the results of this analysis are presented, in the following, a notation for the relevant energy terms will be introduced.

4.1 Atomic Energy Notation

The total energy, $E^{\text{tot}}(\text{AB})$, of a compound $\text{H}_x\text{A} - \text{BH}_y$ can be partitioned in the nuclear repulsion contribution, $E^{\text{nuc}}(\text{AB})$ and the electronic energy, $E^{\text{el}}(\text{AB})$ as

$$E^{\text{tot}}(\text{AB}) = E^{\text{el}}(\text{AB}) + E^{\text{nuc}}(\text{AB}). \quad (101)$$

The electronic energy can be partitioned into atomic contributions

$$\begin{aligned} E^{\text{el}}(\text{AB}) = & E^{\text{el}}(\text{A/AB}) + E^{\text{el}}(\text{B/AB}) \\ & + E^{\text{el}}(\text{AH/AB}) + E^{\text{el}}(\text{BH/AB}) \end{aligned} \quad (102)$$

where $E^{\text{el}}(\text{A/AB})$ and $E^{\text{el}}(\text{B/AB})$ are the respective contributions of atom A and B. $E^{\text{el}}(\text{AH/AB})$ and $E^{\text{el}}(\text{BH/AB})$ denote the sum of atomic energies of the hydrogens bonded to A and B, respectively. The same notation is then used to describe the atomic energies of the fragments, e.g. $E^{\text{el}}(\text{A/A})$ and $E^{\text{el}}(\text{AH/A})$ are the respective atomic contributions of A and the hydrogens bonded to A to

the electronic energy in the fragment $H_xA^\bullet$. The atomic energies in fragment $^\bullet BH_y$ are analogously defined as $E^{\text{el}}(B/B)$ and $E^{\text{el}}(BH/B)$. As an example, the different atomic energy terms for methylamine and its fragments are on display in Figure 10.

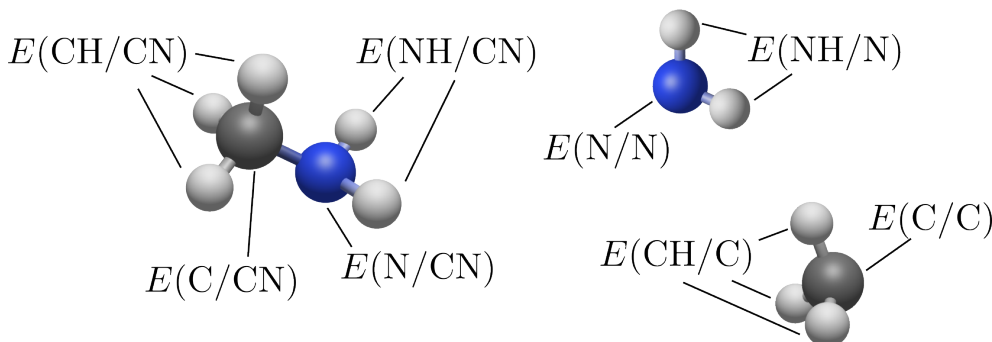


Figure 10: Atomic energy notation exemplified for methylamine and its fragments after homolytic cleavage of the C-N bond. Adapted with permission from *J. Am. Chem. Soc.* 2023, 145, 5899-5908. Copyright 2023 American Chemical Society

Finally, the atomic contribution to the BDE can be defined as already given in Eq. (100). The atomic contribution to only the electronic energy term of the BDE, $E_{AB}^{\text{el}} = E^{\text{el}}(A) + E^{\text{el}}(B) - E^{\text{el}}(AB)$, is defined as

$$E_{AB}^{\text{el}}(A) = E^{\text{el}}(A/A) - E^{\text{el}}(A/AB) \quad (103)$$

for atom A and analogously for atom B and the hydrogen atoms.

In the following analysis, I will focus only on the electronic contribution to the BDE, E_{AB}^{el} , because the calculation of the electronic energy is the computationally expensive step. If the variation in electronic energy with binding partner can be accurately described with help of the atomic energy decomposition scheme, it is trivial to add the nuclear repulsion energy.

Using Eq. (103) the electronic contribution to the BDE can be expressed as a sum of atomic terms as

$$E_{AB}^{\text{el}} = E_{AB}^{\text{el}}(A) + E_{AB}^{\text{el}}(AH) + E_{AB}^{\text{el}}(B) + E_{AB}^{\text{el}}(BH). \quad (104)$$

with $E_{AB}^{\text{el}}(A)$, $E_{AB}^{\text{el}}(AH)$, $E_{AB}^{\text{el}}(B)$, $E_{AB}^{\text{el}}(BH)$ referring to the atomic contributions of atom A, the hydrogens attached to atom A, atom B and the hydrogens attached to atom B, respectively.

4.2 A model for bond dissociation energies

These atomic energies were calculated for molecules $H_xA - BH_y$ and their fragments, $H_xA^\bullet$ and $^\bullet BH_y$ with the alchemical atomic energy decomposition scheme Eq. (68). In this equation, the first term was neglected since it accounts for the nuclear repulsion, while here only the electronic energy is of interest. A and B were chosen from the elements C, N, O and F such that ten different molecules and four different fragments could be generated. Note that Eq. (68) is the decomposition scheme proposed in Ref. 86. As discussed in the previous chapter, the equation does not include the periodic boundary conditions and the potential of the UEG. This decomposition scheme was used nevertheless

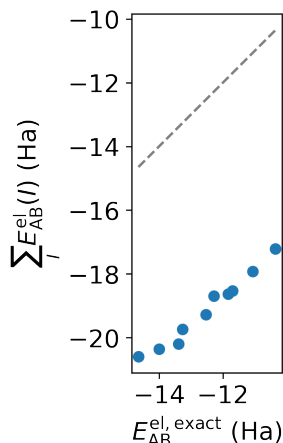


Figure 11: The sum of atomic contributions to the electronic part of the BDE, $\sum_I E_{AB}^{\text{el}}$ Eq. (104), plotted against the exact electronic BDE for 10 molecules with structure $\text{H}_x\text{A} - \text{BH}_y$. The gray dashed line indicates exact agreement between $\sum_I E_{AB}^{\text{el}}$ and $E_{AB}^{\text{el, exact}}$.

because the modified version had not been developed. The sum of atomic energies evaluated from Eq. (68) deviates from the exact energy between the respective compound and the reference. However, as for the compounds in the previous chapter, this deviation is systematic as illustrated in Figure 11. The sum of atomic contributions from Eq. (104) is shifted downwards by approximately 6 Ha for each of the ten compounds but the energy trend across the different systems is mostly recovered. In addition, the atomic energy trend for carbon bonded to C, N, O or F is already correctly reproduced with Eq. (68) as shown in chapter 3 Figure 4B. Thus, the decomposition scheme should be suitable to investigate binding energy trends for varying compositions.

In Figure 12A, the atomic contribution to the electronic BDE, $E_{AB}^{\text{el}}(\text{A})$, is plotted as a function of the nuclear charge of the binding partner, Z_B , to investigate the impact of the binding partner on the atomic BDE. $E_{AB}^{\text{el}}(\text{A})$ increases as the nuclear charge Z_B increases for A being either carbon, nitrogen, oxygen or fluorine. Figure 12B shows the sum of atomic energies of the hydrogen atoms attached to atom A, $E_{AB}^{\text{el}}(\text{AH})$, as a function of the binding partner B. The energies are less sensitive to a change in binding partner compared to the atomic energies of A and remain approximately constant. The lower sensitivity is inline with chemical intuition. Atom A should be affected more by a variation in Z_B than the hydrogens because it is directly bonded to atom B.

The observed relationship between atomic energy and Z_B motivates the approximation of $E_{AB}^{\text{el}}(\text{A})$ as a linear function in Z_B and of $E_{AB}^{\text{el}}(\text{AH})$ as a constant in Z_B . We propose to express the sum of the atomic energies Eq. (104) as a sum of two linear functions

$$E_{AB}^{\text{el}} \approx \underbrace{\beta_A Z_B + \alpha_A}_{=E_{AB}^{\text{el}}(\text{A}) + E_{AB}^{\text{el}}(\text{AH})} + \underbrace{\beta_B Z_A + \alpha_B}_{=E_{AB}^{\text{el}}(\text{B}) + E_{AB}^{\text{el}}(\text{BH})} \quad (105)$$

The slopes β_A and β_B describes the sensitivity of $E_{AB}^{\text{el}}(\text{A})$ and $E_{AB}^{\text{el}}(\text{B})$ to a variation in Z_B and Z_A , respectively. The constants α_A and α_B account for the

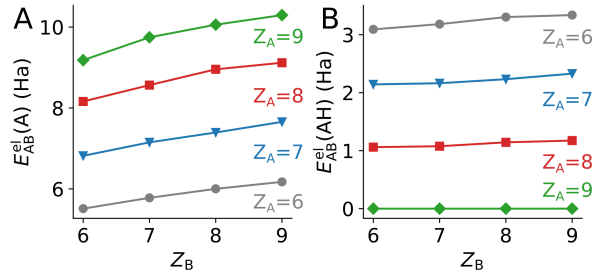


Figure 12: Atomic electronic BDEs (see Eq. (103)) of atom A as a function of the nuclear charge Z_B of the binding partner for carbon, nitrogen, oxygen and fluorine (Panel A) and the sum of the atomic electronic BDEs of the hydrogens attached to the different atoms A (Panel B). Reprinted with permission from *J. Am. Chem. Soc.* 2023, 145, 5899-5908. Copyright 2023 American Chemical Society

constant contributions of the hydrogen atoms attached to A and B, respectively, as well as the constant contributions of the atoms A and B.

In order to obtain BDEs the nuclear repulsion must also be included. Here we approximate the nuclear repulsion by only considering the nuclear repulsion between the binding partners A and B as $Z_A Z_B / \bar{d}$, where $\bar{d}$ is the average distance between A and B in all considered molecules $H_x A - B H_y$.

Consequently, we can model the bond dissociation energy E_{AB} as

$$E_{AB} \approx \underbrace{\beta_A Z_B + \alpha_A}_{=E_{AB}^{el}(A)+E_{AB}^{el}(AH)} + \underbrace{\beta_B Z_A + \alpha_B}_{=E_{AB}^{el}(B)+E_{AB}^{el}(BH)} - \underbrace{\frac{Z_A Z_B}{\bar{d}}}_{=E_{AB}^{nuc}}. \quad (106)$$

The optimal parameters α, β can be determined from a least squares fit to $E_{AB} + \frac{Z_A Z_B}{\bar{d}}$ for then ten compounds $H_x A - B H_y$. E_{AB} is the BDE of the total molecule, which can in principle be obtained from experiments or highly accurate reference calculations. However, here the BDEs were calculated with DFT (PBE0/def2-TZVP) for practical convenience. With optimized parameters α, β , these BDEs can be accurately reproduced by Eq. (106) with a MAE of 0.5 kcal/mol. To test the transferability of the model, it is also applied to compounds $H_x A - B H_y$ where A and B are from the third or fourth row and the fourth to seventh main group of the periodic table (e.g. the elements Si, P, S, Cl or Ge, As, Se, Br). After reoptimizing the parameters α, β with respect to reference BDEs for these compounds, Eq. (106) can reproduce the BDEs with a MAE of 0.4 and 0.3 kcal/mol, respectively.

Eq. (106) can accurately describe the variation in BDE for varying binding partners in different rows of the periodic table. However, the equation requires eight parameters for each row of the periodic table since a pair of parameters α, β exists for each element. If these parameters could instead be expressed as functions of the nuclear charge, the total number of parameters could be reduced. Thus, the relationship between α_A and β_A and the corresponding nuclear charge Z_A is analysed next.

This is done by plotting the optimized parameters, α_A, β_A as function of the nuclear charge Z_A in Figure 13.

The optimized parameter β_A is approximately linear in Z_A , e.g.

$$\beta_A \approx b' Z_A. \quad (107)$$

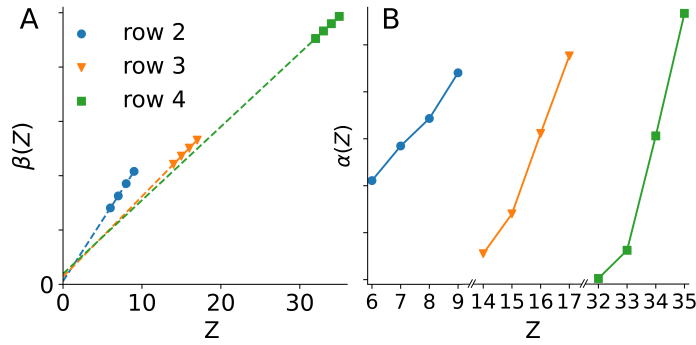


Figure 13: Optimized parameters β (Panel A) and α (Panel B) of Eq. (106) as a function of Z after a fit to bond dissociation energies calculated with DFT (PBE0/def2-TZVP). Reprinted with permission from *J. Am. Chem. Soc.* 2023, 145, 5899-5908. Copyright 2023 American Chemical Society

with the row-dependent slope b' . Consequently, $\beta_{A/B}$ can be expressed as function of $Z_{A/B}$, so that the terms $\beta_A Z_B$ and $\beta_B Z_A$ in Eq. (106) can be simplified as

$$\beta_A Z_B + \beta_B Z_A \approx 2b' Z_A Z_B. \quad (108)$$

The relationship of the optimized offset α_A and Z_A is non-linear as can be observed in Figure 13B.

Thus, we choose to account for the non-linearity with following ansatz for α_A

$$\alpha_A \approx c Z_A^\gamma + a', \quad (109)$$

where the exponent γ accounts for the non-linearity and a' is an offset.

Insertion of the definitions of β and α given in Eq. (107) and Eq. (109) in Eq. (106) and rearrangement of the resulting expression leads to

$$E_{AB} \approx a - b Z_A Z_B + c (Z_A^\gamma + Z_B^\gamma) \quad (110)$$

with $a = 2a'$ and $b = \left(\frac{1}{d} - 2b'\right)$.

The parameters a, b, c are optimized by non-linear least squares and γ via a line scan. For the compounds in second, third and fourth row an individual set of parameters a, b, c is determined while a single γ -value is optimized for all rows together. The optimal value for γ in this optimization procedure is $\gamma_{\text{opt}} = 2.36$. The terms $c Z_{A/B}^{2.36}$ in Eq. (110) resemble the leading term in an expression for the energy of free atoms,^{152,153} which is equal to $-0.768745 \cdot Z^{7/3}$. This is an interesting result which might offer a physical interpretation of the binding energy expression which can accordingly be written as

$$\begin{aligned} E_{AB} &\approx a - b Z_A Z_B + c (Z_A^{7/3} + Z_B^{7/3}) \\ &\approx a - b Z_A Z_B + c' (E_A^{\text{atom}} + E_B^{\text{atom}}) \end{aligned} \quad (111)$$

with the energies of the free atoms E_A^{atom} and E_B^{atom} and $c' = -\frac{c}{0.768745}$. The optimized parameters for this equation with $\gamma = 7/3$ are given in Table 2. I note that the MAE for $\gamma = 7/3$ ($= 1.87$ kcal/mol) is almost identical to the one

Table 2: Coefficients and MAEs of Eq. (111) for different rows n . b and c are scaled such that bZ_AZ_B and $c(Z_A^{7/3} + Z_B^{7/3})$ are given in kcal/mol if Z_A and Z_B are given in atomic units of charge e .

n	a (kcal/mol)	b (630 Bohr ⁻¹)	c (630 Bohr ⁻¹ $e^{-1/3}$)	MAE (kcal/mol)
2	215.7	10.470	1.987	1.0
3	392.7	8.823	1.496	2.4
4	1109.2	6.180	0.813	2.2

for $\gamma = 2.36$ ($= 1.84$ kcal/mol) and in the following I will present only results for $\gamma = 7/3$.

The parameter a depends strongly on the row of the periodic table, respectively the principal quantum number n . In particular, the optimized a varies by ≈ 180 kcal/mol from second to third but by ≈ 700 kcal/mol from third to fourth row. This might be associated with the addition of only 8 vs 18 core electrons when going from second to third or third to fourth row of the periodic table.

The parameter b has the unit of an inverse distance and the overall term $-bZ_AZ_B$ is destabilizing and might be related to the nuclear repulsion between A and B. Finally, the parameter c has a unit of inverse distance times inverse cubic root of electric charge. The term $c(Z_A^{7/3} + Z_B^{7/3})$ could be interpreted as the sum of electronic energies of atoms in a molecule due to the aforementioned similarity with the energy expression in Z for free atoms.

4.3 Performance

The BDE predictions of Eq. (111) (in the following called "our model") for the different compounds in the second, third and fourth row of the periodic table are presented in Figure 14 together with the DFT reference data and the optimal parameters are given in Table 2. Our model predicts BDEs with an overall

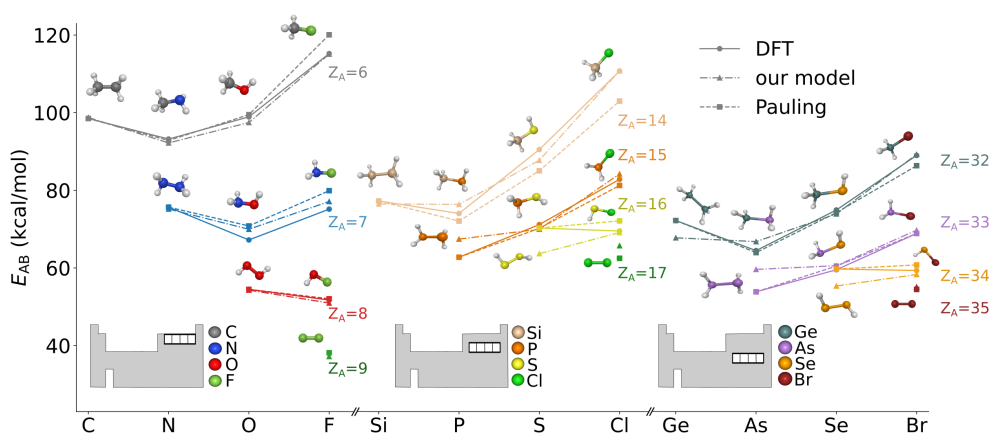


Figure 14: Calculated bond dissociation energies from density functional theory (DFT, PBE0/def2-TZVP), our quantum alchemy based chemical bond model (Eq. (111)), and Pauling's electronegativity model in Eq. (112). MAEs with respect to DFT amount to 1.8 and 1.4 kcal/mol for our and Pauling's model, respectively. Reprinted with permission from *J. Am. Chem. Soc.* 2023, 145, 5899-5908. Copyright 2023 American Chemical Society

MAE of just 1.8 kcal/mol with respect to the DFT reference. Furthermore, the BDE trends for varying composition are reproduced well, as observable from similar shapes of the curves for our model and the reference. It is promising that such a simple model can accurately described bonding trends in chemical space. However, it should be noted that the sampled space is rather small with fixed equilibrium geometries, only *p*-block elements from the same row of the periodic table and only single bonds between A and B.

The prediction errors of our model are also compared to the errors of other well established methods. For this comparison, the quantum chemistry W4-17 dataset¹⁵⁴ is chosen as reference because it provides highly accurate bond dissociation energies at CCSDTQ5/CBS or CCSDTQ56/CBS level. Our model is recalibrated to the W4-17 data and leave-one-out predictions are shown in Figure 15 together with leave-one-out predictions from a machine learning model. The machine learning method is kernel ridge regression with a gaussian kernel.¹⁵⁵ The model is trained with molecular structures of the compounds H_xA-BH_y as input and the corresponding BDEs as labels. The molecular structures are encoded via the bag of bonds representation.²⁹ This representation is a vector with the components being all pairwise combinations $Z_I Z_J / |\mathbf{R}_I - \mathbf{R}_J|$ ordered and padded by zeros in a specific way, where $Z_{I/J}$ and $\mathbf{R}_{I/J}$ are the nuclear charges and positions of the atoms *I* and *J*, respectively (see Ref. 29 for details). The machine learning model yields a substantially higher error (MAE = 10.4 kcal/mol) than our model (MAE = 1.3 kcal/mol) with respect to the W4-17 reference. This result is not surprising since the highly flexible machine learning model depending on all pairwise interactions needs more training data for reliable out of sample predictions than our simple constrained model depending only on the nuclear charges and three parameters. Errors of various QM methods with respect to W4-17 are 9.5 kcal/mol for the semi-empirical PM7 method¹⁵⁶, 1.2 kcal/mol for DFT (PBE0/def2-TZVP), and 0.6 kcal/mol for CCSD(T)/F12 calculations (taken from G2 Ref.¹⁵⁷). These results show that our model can achieve comparable accuracy or even outperform (in case of PM7) these methods for the given set of compounds. However, our model is biased due to the calibration and would likely perform significantly worse when applied to other compounds.

In addition, a comparison of leave-one-out prediction errors of our model fitted either to PBE0, B3LYP or the W4-17 reference data yields MAEs of 1.3, 1.6 and 1.3 kcal/mol, respectively. This indicates that the performance of our model is robust with respect to the choice of the reference method.

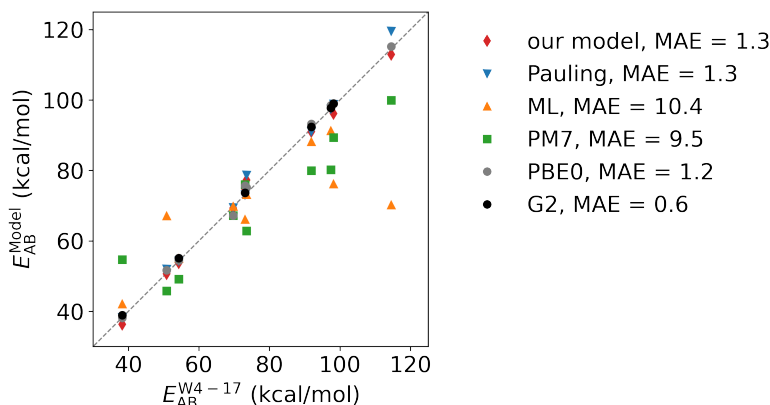


Figure 15: Calculated BDEs using our model, Pauling’s, various quantum methods (PM7, DFT (PBE0/DEF2-TZVP), G2) and a quantum machine learning model. Bond dissociation energies are scattered against the W4 entries¹⁵⁴ for ten saturated single bond diatomics composed of carbon, nitrogen, oxygen or fluorine. Reprinted with permission from *J. Am. Chem. Soc.* 2023, 145, 5899-5908. Copyright 2023 American Chemical Society

4.4 Comparison to Pauling’s model

Already in the 1930s, Pauling developed another simple model to describe bonding trends for varying composition. He decomposed the wavefunction of a diatomic molecule, A-B, into a covalent and an ionic part. He further argued, that the bond energy, E_{AB} , could be split into a covalent and ionic contribution, E^{cov} and E^{ion} , as well.¹⁵⁸ The covalent contribution was then approximated as the mean bond energy of the compounds A-A and B-B, $E^{\text{cov}} = (E_{AA} + E_{BB})/2$. The ionic term was described as a function of the electronegativity difference $E^{\text{ion}} = 23(\chi_A - \chi_B)^2$ with $\chi_{A/B}$ being the electronegativities of A and B. This yields the bond energy equation

$$E_{AB} = \frac{E_{AA} + E_{BB}}{2} + 23(\chi_A - \chi_B)^2 \quad (112)$$

where the electronegativities were determined from a fit to experimental bond dissociation enthalpies (the 23 in the second term is a conversion factor to obtain the energy in kcal/mol if the electronegativities reported by Pauling are used).¹⁵⁸

The predictions with this equation are shown in Figure 14 as well. The MAE is 1.4 kcal/mol, slightly below the one for our model (1.8 kcal/mol) but of comparable magnitude. Eq. (112) predicts BDEs for symmetrical molecules, E_{AA} , exactly by construction, however it performs worse than our model for large electronegativity differences, e.g. for C-F or Si-Cl (see Figure 14).

Both approaches are coarse grained models based on a decomposition of the energy into physically meaningful contributions. The difference is that these contributions are two-body interactions in Paulings model, while our model emerges from one-body terms corresponding to the atoms in the compound. Furthermore, Paulings model depends directly on the energy of the homonuclear bonds E_{AA} and E_{BB} and introduces the electronegativity as additional concept to describe the ionic bond contribution. These choices lead to a very robust model, that can also give decent bond energy estimates for A and B

coming from different rows. In contrast, our model uses the nuclear charges as input and might thus be closer connected to the underlying physics. However, our model cannot be directly applied to molecules where A and B come from different rows because the parameters a, b, c are optimized for one specific row of the periodic table. An extension might be possible by inspecting how our model is constructed from Eq. (106). In Eq. (106), the fragments A and B are both modeled by a linear function with element specific parameters $\alpha_{A/B}, \beta_{A/B}$. By inserting the ansatz $\alpha_{A/B} = a'_{n/m} + c'_{n/m} Z_{A/B}^{7/3}$ and $\beta_{A/B} = b'_{n/m} Z_{A/B}$ into Eq. (106), one obtains

$$E_{AB} = a'_n + c'_n Z_A^{7/3} + b'_n Z_A Z_B + a'_m + c'_m Z_B^{7/3} + b'_m Z_A Z_B - \frac{Z_A Z_B}{\bar{d}}, \quad (113)$$

where n and m are the principal quantum numbers of Z_A and Z_B , respectively. This expression should enable the calculation of BDEs for A and B coming from different rows if the parameters $a'_{n/m}, b'_{n/m}$ and $c'_{n/m}$ are known. Furthermore, from the definitions of the parameters a, b and c in Eq. (111) (see section 4.2) follows, that Eq. (113) can be rearranged to

$$E_{AB} = a - b Z_A Z_B + c'_n Z_A^{7/3} + c'_m Z_B^{7/3} \quad (114)$$

where $a = (a'_n + a'_m)/2$ and $b = (b'_n + b'_m)/2$. Exemplarily, BDEs for C-Si and C-Cl are calculated with Eq. (114) and parameters $a'_{n/m}, b'_{n/m}$ and $c'_{n/m}$ from Table 2. The BDEs predicted by Eq. (114) are more than 200 (C-Si) and 450 kcal/mol (C-Cl) higher than predictions with DFT (PBE0). In contrast, the errors with respect to DFT are 5 (C-Si) and 2 kcal/mol (C-Cl) with Pauling’s model. The reason for the large error of our model is that the expression has the large stabilizing $c Z_A^{7/3} + c Z_B^{7/3}$ and the large destabilizing term $-b Z_A Z_B$. If $Z_A \approx Z_B$ both terms are approximately quadratic in $Z_{A/B}$ but with opposite sign such that the terms largely cancel. However, if $Z_A \ll Z_B$, the destabilizing term is much smaller than the term stabilizing term leading to the overestimation of the energy. In Pauling’s model the covalent term $(E_{AA} + E_{BB})/2$ is a lower bound to the energy to which the stabilizing ionic term $23(\chi_A - \chi_B)^2$ is added. In contrast to nuclear charges, electronegativities are of similar magnitude across different rows of the periodic table. This prevents the overestimation of the BDEs for A and B coming from different rows. Predictions with our model could likely be improved if effective core charges were used instead of the full nuclear charge because the effective core charges are of similar size across different rows. However, such modifications are out of the scope of this work.

The value of simple models, such as Pauling’s or Eq. (111), arguably lies mainly in the conceptual insights they can offer about trends in chemical compound space. Pauling’s model explains for example, that heteronuclear bonds A-B are more stable than homonuclear bonds A-A due to the stabilizing ionic term. In addition, his model defines electronegativities that have been used to rationalize phenomena beyond bond stability, like for example inductive effects. Our model might be useful in other applications beyond the prediction of BDEs. In particular, the relation of the model to Hammett’s is investigated in the following section.

4.5 Comparison to Hammett’s equation

Hammett proposed a simple equation to describe an equilibrium or rate constant K with respect to a reference constant K_0 for reactions involving benzene derivatives as

$$\log\left(\frac{K}{K_0}\right) \approx \zeta \cdot \sigma, \quad (115)$$

where ζ accounted for different reactions conditions, e.g. solvent, temperature or reaction type, and σ described the effect of different substituents of the benzene ring.¹⁵⁹ Subsequently, the model has been applied to describe trends in chemical compound space for other properties, such as orbital energies of metal organic complexes¹⁶⁰, activation energies¹⁶¹, or dipole moments.¹⁶²

The model has been a useful tool to describe property trends in chemical compound space but is based on empirical observations.¹⁶³ Our model was constructed from the physics-based alchemical energy decomposition scheme and can be used to formulate a Hammett type relationship. Thus, it might be helpful to get a better understanding under which conditions Hammett’s empirical ansatz can be used to quantify property-structure trends. In particular, the change in E_{AB} with respect to the nuclear charge of binding partner B

$$\frac{\partial E_{AB}}{\partial Z_B} = -bZ_A + \frac{7}{3}cZ_B^{4/3}. \quad (116)$$

can be approximated by a product of two variables $\zeta\sigma$ for a fixed Z_A . This is shown in Figure 16, where $\frac{\partial E_{AB}}{\partial Z_B} + \Delta$ with $\Delta = bZ_A$ is plotted against the nuclear valence charge of binding partner B so that

$$\begin{aligned} \frac{\partial E_{AB}}{\partial Z_B} + bZ_A &= \frac{7}{3}cZ_B^{4/3} \\ &\approx \underbrace{\frac{7}{3}c}_{=\zeta} \underbrace{Z_{V,B}}_{=\sigma}. \end{aligned} \quad (117)$$

The slope of the curves, $\frac{7}{3}c$, can be identified as ζ and the effect of the binding partner is modeled via $\sigma = Z_{V,B} \approx Z_B^{4/3}$. The slope ζ becomes flatter for increasing row number in the periodic table. This can be interpreted as a lower sensitivity of $\frac{\partial E_{AB}}{\partial Z_B}$ on the binding partner B. The decrease of the slope is particularly large from row 3 to row 4, which might be due to the addition of the 10 d -block electrons in the fourth row.

Due to the nearly quadratic ($Z_B^{7/3}$) dependence of E_{AB} on Z_B , the derivative $\partial_{Z_B} E_{AB}$ is approximately linear in Z_B and consequently σ can be modeled by the number of valence electrons which correlates linearly with Z_B .

The Hammett-like relationship for $\frac{\partial E_{AB}}{\partial Z_B} + \Delta$ is obtained from the BDE expression Eq. (111), which was constructed from the linear model for atomic energies, Eq. (106). Thus, the decomposition of a property into atomic contributions could be a general strategy to discover Hammett-type relations by revealing simple linear relationships between decomposed property and structure.

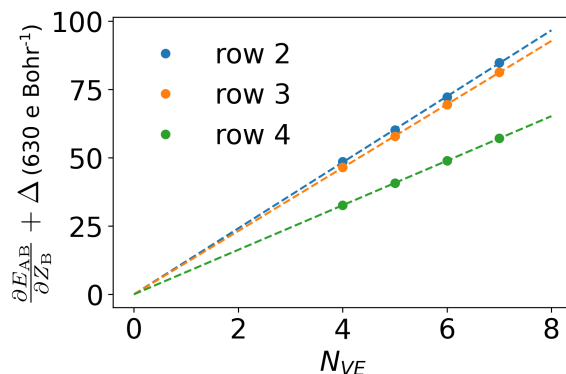


Figure 16: A Hammett-like relation between the derivative of the bond dissociation energy with respect to the nuclear charge $\frac{\partial E_{AB}}{\partial Z_B}$ and the number of valence electrons N_{VE} in the binding partner B. The curves are shifted by a constant Δ such that they intersect the origin. Adapted with permission from *J. Am. Chem. Soc.* 2023, 145, 5899-5908. Copyright 2023 American Chemical Society

4.6 Limitations to covalent bonds

The performance of our model for binding between *s* and *p*-block elements is explored in Figure 17. Z_A is equal to either 6, 7, 8 or 9 and fixed, while Z_B is varied from 3 to 9. As discussed before, our model agrees well with DFT

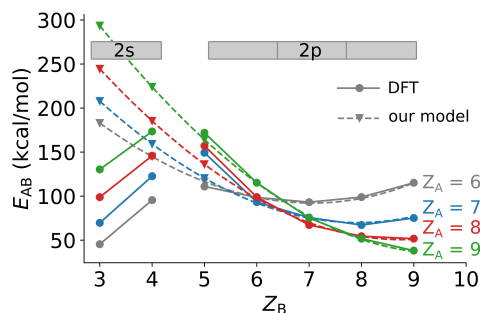


Figure 17: Limitations of applicability: BDEs between atom A and B corresponding to C/N/O/F or Li/Be/B/C/N/O/F, respectively. For second quantum number differing between A and B, our model [Eq. (111)] breaks down. Reprinted with permission from *J. Am. Chem. Soc.* 2023, 145, 5899-5908. Copyright 2023 American Chemical Society

for Z_B in the range from 6 to 9. The accuracy slightly deteriorates for boron, e.g. $Z_B = 5$, but yields still a reasonable prediction. However, as Z_B is reduced further to the *s*-block elements beryllium and lithium, our model predicts an increase in BDE while DFT predicts a decrease. The worse performance of the model for the binding between *s* and *p*-block elements could be associated with the change of the bonding character from covalent for A and B being both *p*-block elements to the more ionic character if one partner is from the *s*-block. Our model was developed based on trends for predominantly covalent bonding. Thus, its problems in describing ionic bonding should not be surprising.

Furthermore, our model has the lowest accuracy for binding among *p*-block elements if one of the partners is boron. In this case, the BDE is systematically

underestimated. The reason might be that bonds involving boron have a bond order of 1.4 while the other bonds are of order 1.^{164–166} Our model was calibrated to systems with bond order 1, which might explain the observed underestimation of the BDEs involving boron.

4.7 Bond order

Our model seems to predict BDEs with lower accuracy if the bond order deviates significantly from 1. However, is it possible to accurately predict BDEs for a different bond order after recalibration to such systems? This question is investigated by refitting our model to double bonded molecules $H_xA=BH_y$. A and B are either the elements C, N, O or Si, P, S or Ge, As, Se. x and y are chosen such that all remaining valencies are saturated. Six unique compounds $H_xA=BH_y$ can be generated for each set of elements. A notable difference to the single bonded systems H_xA-BH_y is that for the double bonds the spin in the ground state is not the same for all compounds. O_2 , S_2 and Se_2 have triplet ground states while the other systems are singlets. Thus, for consistency the BDE was calculated for every molecule $H_xA=BH_y$ in the singlet state and all fragments H_xA and BH_y in the triplet state. The BDEs were denoted as ΔE^{TS} to highlight that fragments are in triplet and molecules in singlet state. After recalibration of our model to ΔE^{TS} , energy trends were reproduced accurately with MAEs of 0.9, 1.3 and 2.0 kcal/mol for the compounds in either second, third or fourth row, respectively (see A.2.2 Figure S1 and Table S1).

4.8 Electronic Configuration

The ability of our model to describe BDEs in different spin states was also studied for the compounds $H_xA=BH_y$. For this purpose, the model was recalibrated to energy differences between fragments and compounds being either both in their energetically lowest triplet state ΔE^{TT} or in their energetically lowest singlet state ΔE^{SS} . Furthermore, the model was fitted to the bond dissociation energy E_{AB} that is calculated from the energetically lowest electronic state for each compound and fragment. In contrast to the energies ΔE^{TS} , ΔE^{TT} and ΔE^{SS} , for the energy E_{AB} , the compounds are not all in the same electronic state because O_2 , S_2 and Se_2 are triplets in the ground state while the other compounds are singlets. The model shows a worse performance for E_{AB} (MAE 7.4 kcal/mol) than for ΔE^{TS} for the compounds in the second row. The reason for the higher error is the worse prediction of E_{AB} for O_2 , the compound with another spin state (see A.2.2 Figure S2A; note that in A.2.2 the energy E_{AB} is denoted as ΔE^{GS} , where GS stands for ground state). The lower prediction accuracy if the spin state changes is inline with the observed worse performance if the bond order or bonding type (covalent vs ionic) varies within the dataset (see section 4.6). However, the MAE of E_{AB} for third and fourth row (1.6, 1.9 kcal/mol) is similar to the one of ΔE^{TS} although mixed spin states have to be modelled (see A.2.2 Figure S2A). This indicates that inconsistencies in the spin state throughout the considered chemical space do not necessarily lead to a deterioration in accuracy.

The prediction errors for ΔE^{TT} and ΔE^{SS} are in the order of 7-8 kcal/mol for the second row and 10-14 kcal/mol for the third and fourth row (see A.2.2 Figure S2B and C and Tables S3 and S4). Since our model was designed based on trends for the ground state of single bonds, the lower prediction accuracy for different spin states and bond order should be expected. Nevertheless, the trends can be reproduced reasonably well.

The overall good accuracy might be due to the relatively low complexity of the dataset with two variables (Z_A, Z_B) and only six datapoints. It might be that the performance of our model is due to its flexibility instead of its capability to describe the underlying physics. Thus, the model was also fitted to points drawn from a random distribution spanning the same range as the energies in the different spin states. The fit to the random data yields significantly higher errors than for the BDEs (see A.2.2 Figure S3) indicating that the performance of our model can be indeed associated with its ability to describe the relevant physics of energy-structure relationships.

4.9 Trends and dependence on period

It was also attempted to generalize our model to describe BDEs in either second, third or fourth row with a single set of parameters a, b, c . This could for example be achieved by expressing the parameters as a function of the principal quantum number. However, such a function would need to be very accurate because our model has two large terms of opposite sign, $-bZ_AZ_B$ and $c(Z_A^{7/3} + Z_B^{7/3})$. Hence, already small inaccuracies in b or c could lead to large errors in the BDEs.

Differences in BDEs, $\Delta E = E_{AB} - E_{AC}$, are less sensitive to the exact choice of the parameters because the attractive respectively repulsive terms in A-B and A-C cancel to a large degree. It was found that ΔE can be approximated as

$$\Delta E \approx \Delta Z \left[28(n-1) + (8.5 + (n-3)^2)Z_A \right] + 6(Z_B^2 - Z_C^2), \quad (118)$$

where $\Delta Z = Z_C - Z_B$ and n is the principal quantum number of the elements A, B and C (see A.2.2 p. S-7 for the derivation).

In each of the considered rows of the periodic table, there are 24 unique combinations of A, B and C with $Z_B \leq Z_C$. The MAE over all 72 compounds is just 4.2 kcal/mol, only slightly higher than for Eq. (111) (see Figure 18).

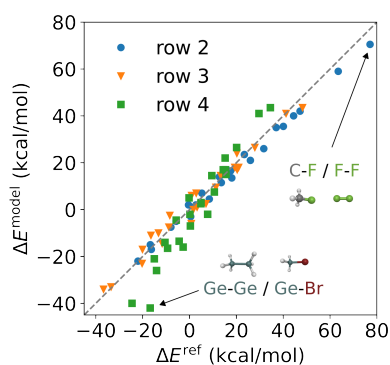


Figure 18: Trends in covalent binding: Predicted (Eq. (118)) vs. Truth (corresponding to PBE0/def2-TZVP). 72 bond dissociation energy differences between bonds A-B and A-C (ΔE), both within either the second, third, or fourth row of the periodic table, $\Delta E = E_{\text{AB}} - E_{\text{AC}}$, where $Z_{\text{B}} \leq Z_{\text{C}}$. Reprinted with permission from *J. Am. Chem. Soc.* 2023, 145, 5899-5908. Copyright 2023 American Chemical Society

The model developed in the previous chapter showed promising results for the description of bond stability trends of simple molecules containing two *p*-block elements from the same row. In this chapter, the work will be extended to a larger variety of bonding patterns. For 1325 organic molecules, the contribution of the atoms to the atomisation energy will be studied.

5.1 Calculation of the atomic energies

The subset of molecules in QM9¹⁹ with 38 valence electrons was selected for the investigation. This subset consists in total of 1329 molecules from which four were excluded due to convergence issues resulting in atomic energies for 1325 molecules. These molecules have 6-8 heavy atoms of element type carbon, nitrogen, oxygen or fluorine and the Lewis structures of all compounds are given in A.1.2. Only molecules with the same number of valence electrons were selected because these systems are isoelectronic in pseudopotential calculations. Hence, an identical UEG reference for all molecules can be generated by choosing always the same cell size in the calculations with periodic boundary conditions. This facilitates the comparison of atomic energies across different molecules.

The atomic energies were calculated with Eq. (92). Derivatives $\partial_\lambda E_I$ were evaluated with the perturbative ansatz Eq. (94) for six equally spaced points along the transformation path ($\lambda = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$) and atomic energies ΔE_I were obtained by numerical integration of $\partial_\lambda E_I$ with Simpson's rule. In a previous study, alchemical derivatives were calculated by only rescaling the local pseudopotential parameters but not the non-local parameter h_{ij} , and the same approach was adopted here.¹⁶⁷ This does not have a significant effect on the atomic energies (see A.1.2 Figure 2). The atomic energies were rescaled to match the exact ΔE with Eq. (95). The rescaled atomic energies are given with respect to the UEG, while conventionally molecular energies are reported as atomisation energies, e.g. with respect to the energy of the free atoms. Thus, the rescaled atomic energies were shifted such that their sum is equal to the atomisation energy. For this purpose, atomic energies with respect to the vacuum were calculated first by partitioning the energy of the UEG into atomic contributions as

$$E_I^{\text{UEG}} = \frac{Z_{I,V}}{\sum_J Z_{J,V}} E^{\text{UEG}}, \quad (119)$$

where $Z_{I,V}$ is the nuclear valence charge of atom *I* and E^{UEG} is the energy of the UEG. Since all compounds are isoelectronic, $\sum_J Z_{J,V}$ is identical for all molecules. Hence, E_I^{UEG} depends only on the element but not the local

environment of an atom within a molecule. The energy of a molecule with respect to the vacuum can then be expressed as

$$\begin{aligned} E^{\text{mol}} &= \sum_I \Delta \tilde{E}_I + E^{\text{UEG}} \\ &= \sum_I \left(\Delta \tilde{E}_I + \frac{Z_{I,V}}{\sum_J Z_{J,V}} E^{\text{UEG}} \right), \end{aligned} \quad (120)$$

where $\tilde{E}_I$ denotes a rescaled atomic energy and the partitioning of E^{UEG} from Eq. (119) is used in the second line.

Atomic contributions to the atomisation energy can be calculated by subtracting the energies of the free atoms, E_I^{free} , as

$$\Delta E_I^{\text{at}} = \Delta \tilde{E}_I + \frac{Z_{I,V}}{\sum_J Z_{J,V}} E^{\text{UEG}} - E_I^{\text{free}}. \quad (121)$$

The sum of the atomic atomisation energies $\sum_I \Delta E_I^{\text{at}}$ is exactly equal to the atomisation energy of the respective molecule due to the use of the rescaled atomic energies $\Delta \tilde{E}_I$.

5.2 Alchemical atomic energies of organic molecules

The atomic atomisation energies split by element are shown in Figure 19. In

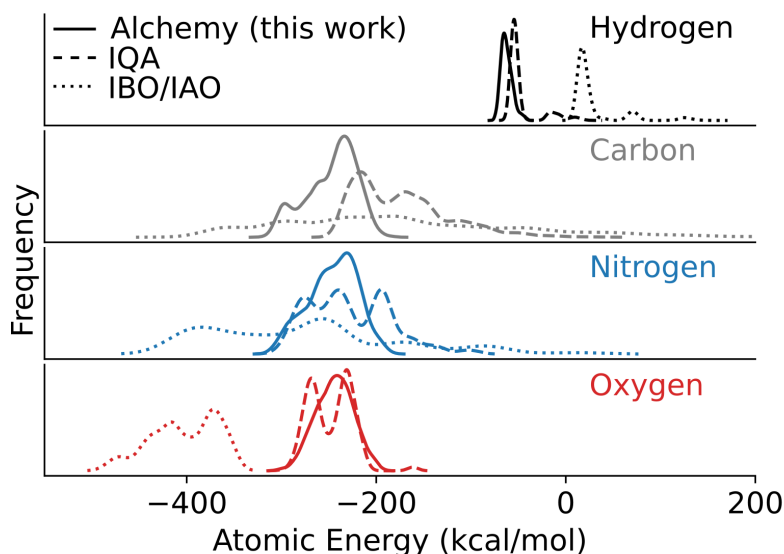


Figure 19: The distribution of atomic contributions to the atomisation energy in different local environments for 1325 drug-like molecules (~ 18000 atoms) with 6-8 heavy atoms split by element from the alchemical decomposition scheme, the interacting quantum atom (IQA) and intrinsic bond orbitals and intrinsic atomic orbitals (IBO/IAO). Reprinted from *J. Chem. Phys.* 160, 054106 (2024), with the permission of AIP Publishing.

total there are 8644 hydrogen, 6323 carbon, 1661 nitrogen and 1341 oxygen atoms. The dataset contains also 9 fluorine atoms that are omitted in the figure. The atomic energies of carbon, nitrogen and oxygen are of similar magnitude (mean values are -247 , -246 and -244 kcal/mol, respectively),

while the energies of hydrogen are shifted to more positive values (on average -62 kcal/mol). It is also noteworthy that atomic contributions to the total energy, i.e. before the energy of the free atom is subtracted, are for all atoms in the order of the energy of the respective free atom. The atomic energies from the alchemical decomposition can thus be interpreted as perturbations of the free atom energy due to the interaction with the other atoms in the molecule. This observation is also consistent with the model for BDEs Eq. (111). Eq. (111) contains the stabilizing terms $c(Z_{A/B}^{7/3})$, which are approximately proportional to the energy of the free atoms A and B, as well. If the atomic energies are interpreted as perturbed free atom energies, the shape and width of the respective distributions in Figure 19 should be connected to the different local environments of the atoms. A first indication that the width is linked to the local environment is the fact that the range of the atomic energy distribution of hydrogen (37 kcal/mol) is much narrower than for carbon, nitrogen and oxygen (135, 135 and 110 kcal/mol, respectively) because hydrogen can have only three different binding partners (carbon, nitrogen and oxygen) while for the heavy elements, a much greater variety of binding partner combinations and bond orders exists in the dataset.

The relationship between atomic energy and local structure is explored in more detail by splitting the atomic energy distributions of each element into partial distributions in which each atom has the same binding partners. For hydrogen, only three different sub-distributions exist (as mentioned above), while carbon, nitrogen and oxygen have 34, 14, and 6 unique sets of binding partners, respectively. A selection of these distributions is shown for carbon in Figure 20A. For each distribution, the covalent binding partners are on display, as well. For example, the distribution with the lowest energy (on the left in Figure 20A) corresponds to the atomic energies of a carbon atom which has three hydrogens and one carbon atom as binding partners. Similar plots for the other elements can be found in A.1.2 Figure 3. The distributions

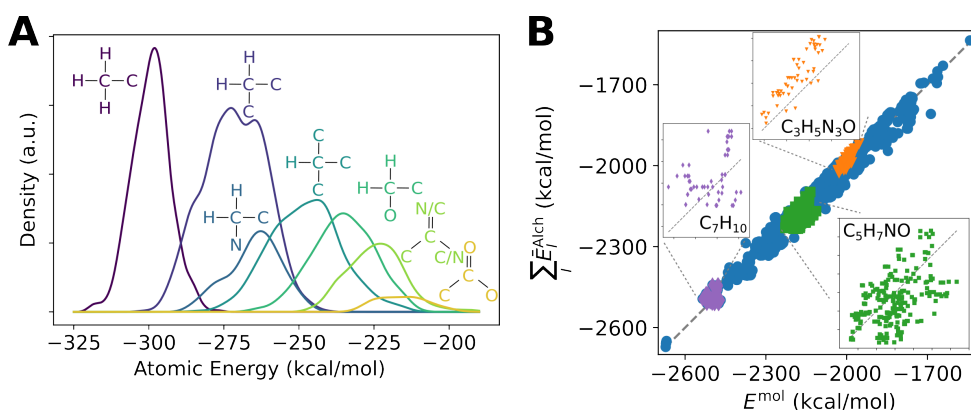


Figure 20: Modelling the atomisation energy from atomic contributions obtained via the alchemical energy decomposition. Panel A: Atomic energy distributions for carbon in different local environments. Panel B: Atomisation energies estimated as the sum of average atomic energies for the local environments present in the molecule. Insets show predictions for selected stoichiometries. Reprinted from *J. Chem. Phys.* 160, 054106 (2024), with the permission of AIP Publishing.

for the individual local environments are narrower than the distribution for all carbon atoms together. They span energy ranges of 50 kcal/mol or less

compared to 135 kcal/mol for all carbons together. Thus, the atomic energies from the alchemical decomposition scheme should be transferable across different molecules, e.g. atoms in the same local environment have similar atomic energies. This is also true for the other elements (see A.1.2 Figure 3).

The transferability of the atomic energies can be exploited to estimate the atomisation energy of a molecule. For each local environment, the mean value $\bar{E}(Z_I, P_I)$ of the corresponding atomic energy distribution is calculated. Z_I is the nuclear charge and P_I identifies the set of binding partners. The energy of a given molecule can then be evaluated by summing up these mean energies for all local environments in the molecule

$$E^{\text{mol}} = \sum_I \bar{E}(Z_I, P_I). \quad (122)$$

The molecular atomisation energies of the 1325 molecules in our dataset can be estimated this way with an error of 29 ± 23 kcal/mol. Inspection of the errors reveals that energies are overestimated if the exact atomisation energy is low and underestimated if the exact atomisation energy is high. This systematic error can be corrected by combining Eq. (122) with a linear regression model as

$$E^{\text{mol}} = m \cdot \left(\sum_I \bar{E}(Z_I, P_I) \right) + b \quad (123)$$

with slope m and intercept b obtained by minimizing

$$L = \sum_J \left(m \sum_I \sum_{I(J)} \bar{E}(Z_I, P_I) + b - E_J^{\text{mol}} \right)^2, \quad (124)$$

where $\sum_J$ is a sum over all molecules J in the dataset. The removal of the systematic error by Eq. (123) reduces the MAE to 23 ± 16 kcal/mol for our dataset. This error is one order above chemical accuracy but our model can still capture the stability trend across the whole dataset (see Figure 20B). Thus, it might be useful as a baseline for machine learning models.

Furthermore, we compared the performance of our estimate to other approaches that are used as a baseline for machine learning models. For example, the atomisation energies of the molecules in the QM7-dataset were estimated with bond-counting with a MAE of 71 kcal/mol using bond energies from literature.²⁷ The error might be higher than for Eq. (123) because QM7 is a larger and likely more diverse dataset. Nevertheless, this indicates that the alchemical atomic energies can yield competitive energy estimates. The dressed atom was introduced as another, statistical approach for the estimation of atomisation energies.²⁹ In this model, the energy of a molecule is approximated as

$$E^{\text{mol}} \approx \sum_{\text{elements}} N^{\text{el}} \times E^{\text{mean,el}}, \quad (125)$$

where $E^{\text{mean,el}}$ is an element-dependent atomic energy, N^{el} is the number of atoms of this element type in the molecule and the sum runs over all elements

contained in the molecule. The atomic energies $E^{\text{mean,el}}$ are determined by minimizing the loss function

$$L = \sum_J \left(E_J^{\text{mol}} - \sum_{\text{elements}} N_J^{\text{el}} \times E^{\text{mean,el}} \right)^2, \quad (126)$$

where E_J^{mol} is the atomisation energy of molecule J and $\sum_J$ is a sum over all molecules in the training set. After fitting Eq. (125) to the atomisation energies of the 1325 molecules in our dataset, the atomisation energies can be reproduced with an error of 20 ± 14 kcal/mol, which is somewhat lower than with Eq. (123) but of similar magnitude. However, an advantage of our expression is that the energy of molecules with the same stoichiometry can be in principle distinguished, while Eq. (125) yields in this case always identical energies. The energy estimates with Eq. (123) for molecules of the same stoichiometry are highlighted for the compositions C_7H_{10} , $\text{C}_5\text{H}_7\text{NO}$ and $\text{C}_3\text{H}_5\text{N}_3\text{O}$ in Figure 20B. The relative ordering of the atomisation energies can be captured quite well for $\text{C}_3\text{H}_5\text{N}_3\text{O}$ and $\text{C}_5\text{H}_7\text{NO}$ but not for C_7H_{10} . This shows that Eq. (123) can indeed resolve energy differences between molecules of identical stoichiometry to some extent.

In addition, the mean atomic energies $\bar{E}(Z_I, P_I)$ are obtained from the physics-based alchemical decomposition scheme, while the dressed atom energies are purely statistical. Hence, Eq. (123) is likely more reliable for the energy prediction of compounds that differ strongly from the ones in the training set.

A disadvantage is however, that the energy of a molecule can only be predicted if all local environments in this molecule are contained in the training data. To address this limitation, the variation in atomic energy due to a change in the local environment is investigated in more detail in the next section.

5.3 The relation between atomic energy and nuclear electrostatic repulsion potential

The atomic energy distributions in Figure 20A are shifted to more positive values as the nuclear charges of the binding partners increase. For example, for carbon bonded to one carbon and three hydrogen atoms ($\text{C-H}_3\text{C}$), the mean atomic energy is -299 kcal/mol. If one hydrogen is substituted by a carbon atom ($\text{C-H}_2\text{C}_2$) the energy increases to -271 kcal/mol and as another hydrogen is replaced by a carbon atom (C-HC_3) the energy increases further to -248 kcal/mol. Similar trends can also be observed for the other elements (see A.1.2 Figure 3).

A descriptor quantifying this trend could be the average electrostatic potential $\bar{V}_I$ due to the nuclear valence charges of the binding partners at the position $\mathbf{R}_I$ of the respective atom I . This potential is given by

$$V_I = \frac{1}{\mathcal{P}^N} \sum_{J=1}^{\mathcal{P}^N} \frac{Z_{J,V}}{|\mathbf{R}_J - \mathbf{R}_I|} \quad (127)$$

with $\mathcal{P}^N$ being the number of binding partners, $\mathbf{R}_J$ their positions and $Z_{J,V}$ their nuclear valence charges. Nuclear valence charges are chosen instead of

the total nuclear charges because the atomic energies were calculated with pseudopotentials. If the local environments would contain elements of the same main group with identical nuclear valence charges, e.g. oxygen and sulfur, effective core charges could be used instead. An average potential $\bar{V}_I(Z_I, P_I)$ for each local environment can be obtained in analogy to $\bar{E}(Z_I, P_I)$ as the average of the potentials V_I for all atoms of same element-type and with the same local environment.

A plot of $\bar{E}(Z_I, P_I)$ vs $\bar{V}_I(Z_I, P_I)$ is shown in Figure 21 for the elements H, C, N and O. The two quantities are approximately linearly correlated for all

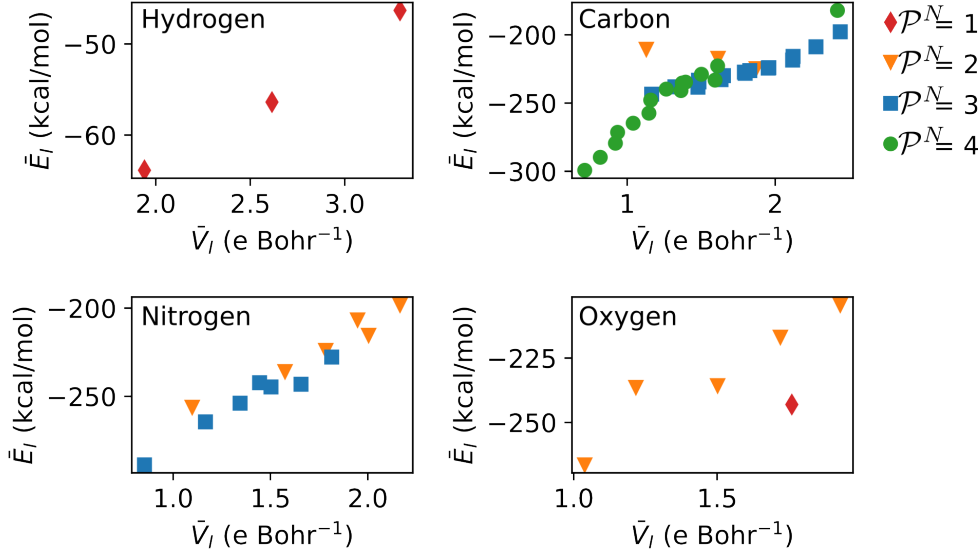


Figure 21: The influence of the local environment, i.e. the unique set of binding partners, on the respective mean atomic energy $\bar{E}_I$ (Eq. (123)) quantified via the relation between the average electrostatic potential of the binding partners $\bar{V}_I$ (Eq. (127)) at the position of atom I for H, C, N and O and different number of binding partners $\mathcal{P}^N$, which restrict bond orders. Reprinted from *J. Chem. Phys.* 160, 054106 (2024), with the permission of AIP Publishing.

elements. Thus, $\bar{E}_I$ can be estimated as

$$\bar{E}_I \approx m\bar{V}_I + b \quad (128)$$

with m and b determined from a linear fit to the respective data for $\bar{E}_I$ and $\bar{V}_I$.

V_I can be easily calculated for any compound as long as a 3D structure exists. Even if no 3D structure but only a graph representation, e.g. a Lewis structure, is available, V_I could be estimated using tabulated bond lengths for the distances $|\mathbf{R}_I - \mathbf{R}_J|$. Therefore, Eq. (123) could also be used to predict atomisation energies of molecules where $\bar{E}(Z_I, P_I)$ is not known for all local environments by estimating the unknown $\bar{E}(Z_I, P_I)$ via Eq. (128).

The ability of the nuclear electrostatic potential Eq (127) to capture atomic energy trends can be explained by the strong dependence of the atomic energy on short-range electrostatic interactions. This dependence can be understood from the atomic energy definition

$$\Delta E_I = \Delta Z_I \int_0^1 d\lambda \int d\mathbf{r} \frac{\rho(\lambda, \mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} + \frac{1}{2} Z_I \sum_{J \neq I} \frac{Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \quad (129)$$

that was derived in Ref 86 (see also Eq. (68) in this thesis). The second term in the expression is proportional to the electrostatic potential due to all nuclei and because of its inverse relation to the distance $|\mathbf{R}_I - \mathbf{R}_J|$, the binding partners should have the largest contribution to this term. Also in the first term, contributions of the density $\rho(\lambda, \mathbf{r})$ to the integral decay with inverse distance $|\mathbf{r} - \mathbf{R}_I|^{-1}$. Furthermore, the density has maxima at the positions of the nuclei and could be approximated by effective point charge placed at the positions $\mathbf{R}_J$ of the other nuclei J . Note that Eq. (129) neglects the interaction with periodic images and the background potential but these effects can be ignored if we want to model only the impact of the local environment.

The influence of the bond order on atomic energy trends is visualized in Figure 21 as well by plotting $\bar{E}_I$ versus $\bar{V}_I$ with different symbols and colors for different numbers of binding partners $\mathcal{P}^N$. Hydrogen has always only a single binding partner $\mathcal{P}^N = 1$ with bond order equal to 1. In contrast, carbon can have from one to four binding partners. If $\mathcal{P}^N = 4$, carbon has four single bonds, and if $\mathcal{P}^N = 3$, carbon has a double bond to one atom and two single bonds to the other atoms. In case of $\mathcal{P}^N = 2$, there are two possible combinations of bonds. Either carbon has two double bonds or one single and one triple bond. The slopes of $\bar{E}_I$ vs $\bar{V}_I$ change for different values of $\mathcal{P}^N$ indicating that the bond order influences the atomic energy trend. A reason could be the different electron density topology around the central atom for different bond orders which could affect the atomic energy (see Eq. (129)). Also for oxygen the atomic energy for $\mathcal{P}^N = 1$ seems to deviate from the trend for $\mathcal{P}^N = 2$. However, in this case, there is not enough data to make a definite conclusion. In the case of nitrogen, the bond order seems to have no influence on the atomic energy trends since the slopes for $\bar{E}_I$ vs $\bar{V}_I$ are similar for $\mathcal{P}^N = 2$ and $\mathcal{P}^N = 3$. In summary, the effect of the bond order does not seem to be consistent across different elements. For carbon, the bond order affects the change in energy $\bar{E}_I$ with respect to $\bar{V}_I$ while this is not the case for nitrogen. Further investigations would be needed to understand this element-specific behaviour but this is out of the scope of this work.

The relationship between the energies of single atoms and molecules and the electrostatic potential resulting from an alchemical transformation has been studied before.^{40,79–85} In particular Politzer and Murray proposed to express the energy of a molecule as a sum over atomic terms

$$E^{\text{mol}} = a \sum_A Z_A V_{A,0} + b, \quad (130)$$

where Z_A is the nuclear charge of atom A and $V_{A,0}$ is the electrostatic potentials at its position due to the other nuclei and the electron density. a and b are two parameters obtained from a fit to reference molecular energies.^{79,80}

The electrostatic interaction between nuclei is also used as a descriptor in representations for machine learning models such as the Coulomb matrix²⁷ or the bag of bonds representation,²⁹ and has been applied to model quantum properties, including energies, throughout chemical space.^{27–30,32}

5.4 Comparison to other decomposition methods

The alchemical energy decomposition scheme was also compared to two other atomic energy decompositions; a mean-field decomposition technique⁶² using

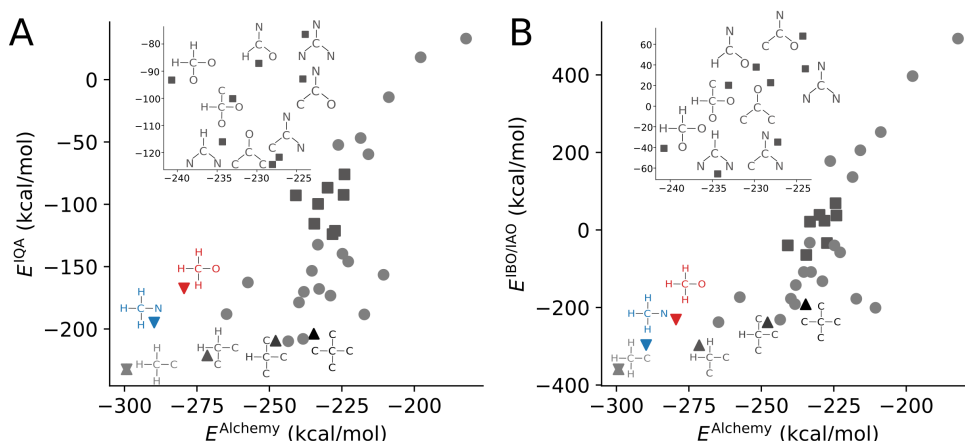


Figure 22: Comparison of the interacting quantum atom (IQA, panel A) and intrinsic bond orbitals and intrinsic atomic orbitals (IBO/IAO, panel B) partitioning schemes to the alchemical partitioning in varying environments of the carbon atom. Similarities and differences are highlighted for selected series of environments. Up-pointing triangles correspond to a series of environments where hydrogens are substituted by carbons, down-pointing triangles show trends if a single heavy atom neighbor X in H_3C-X is varied, cubes correspond to a set of environments generating similar nuclear electrostatic potentials at the position of the central carbon atom and circles denote the remaining local environments in the dataset. Reprinted from *J. Chem. Phys.* 160, 054106 (2024), with the permission of AIP Publishing.

intrinsic bond orbitals and intrinsic atomic orbitals (see section 2.6.1) and the interacting quantum atom^{60,61} (see section 2.6.2). The atomic energy distributions for these schemes are shown in Figure 19. For the interacting quantum atom and intrinsic bond and intrinsic atomic orbitals, the energies of the hydrogen atoms are more positive than those of the other elements, like for the alchemical energy decomposition. However, the energy distributions span a wider range than for alchemy. For example, the difference between the lowest and the highest atomic energy of carbon is 138 kcal/mol for alchemy, 277 kcal/mol for the interacting quantum atom and 863 kcal/mol for intrinsic bond and intrinsic atomic orbitals. As a consequence of the broader energy range, energy estimates with Eq. (123) yield slightly higher errors for the interacting quantum atom (MAE = 25 kcal/mol) and significantly higher errors for intrinsic bond and intrinsic atomic orbitals (MAE = 50 kcal/mol). Thus, the alchemical energy decomposition or the interacting quantum atom should be the preferred choice for atomisation energy estimates via mean atomic energies with Eq. (123). However, intrinsic bond and intrinsic atomic orbitals might be better suited as a fingerprint for local environments. Since the atomic energies span a wider energy range, the different environments of an atom in a molecule could be resolved in more detail.

Furthermore, it was explored whether the different decomposition schemes describe changes in atomic energy due to a change in local environment in a similar way. This was done by plotting the mean atomic energies $\bar{E}_I(Z_I, P_I)$ from alchemy against the corresponding atomic energies obtained from the other two decomposition schemes (see Figure 22).

The overall trend is the same. Carbon bonded to one carbon and three hydrogens has always the lowest atomic energy. The atomic energy increases as the binding partners become more electronegative and carbon bonded to

one carbon and three fluorines has always the highest atomic energy (see Figure 22A and B).

However, the sensitivity of the atomic energy to a change in the local environment is different. In the region from -300 kcal/mol to -250 kcal/mol, the slope of the interacting quantum atom energies, E^{IQA} , vs E^{Alchemy} is significantly flatter than in the region from -250 kcal/mol to -200 kcal/mol and a similar behaviour can be observed for energies from intrinsic bond and intrinsic atomic orbitals, $E^{\text{IBO/IAO}}$, vs E^{Alchemy} . This shape of the plots seems to originate from a higher sensitivity of the alchemical energies to a change in the electrostatics compared to the other decomposition schemes. The series of local environments where carbon is bonded to one carbon and three hydrogens and where the hydrogens are subsequently replaced by carbon is located in the low energy range from -300 to -250 kcal/mol. The atomic energies for these local environments are plotted with upwards-pointing triangles in Figure 22). In this series, the electrostatic potential V_I due to the nuclei (see Eq. (127)) changes strongly because hydrogen is replaced by carbon. This causes a large change of the alchemical atomic energies due to the previously discussed dependence of the atomic energies on the electrostatic potential. In contrast, atomic energies from the interacting quantum atom and intrinsic bond and intrinsic atomic orbitals seem to be less sensitive to a substitution of carbon with hydrogen. Instead, these decompositions depend more strongly on a variation in the electronegativity of the binding partners than the alchemical one does. For example, the slope of the series where carbon is bonded to three hydrogens and one heavy atom and the heavy atom is varied from carbon over nitrogen to oxygen (indicated by downwards pointing triangles in Figure 22) has a steeper slope than the series where hydrogen is substituted by carbon.

Furthermore, the electrostatic potential does not vary a lot for the local environments in the subpanels of Figure 22. Consequently, the atomic energies from alchemy are very similar (they span a range of 15 kcal/mol) while the atomic energies from the interacting quantum atom or intrinsic bond and intrinsic atomic orbitals vary much more (they span ranges of 40 and 120 kcal/mol, respectively). Similar relationships between the decomposition schemes can be observed for the other elements (see A.1.2 Figure 4).

This analysis suggests, that the alchemical energy decomposition scheme is more sensitive to changes in the electrostatic potential of the environment than the other decomposition schemes. Consequently, it might be particularly useful to study processes with strong variations in the electrostatic potential, like the diffusion of ions in solids. The alchemical atomic energy variations due to e.g. diffusion might be explainable by simple electrostatic models. In contrast, the relationship between structure and atomic energy might be more complicated for the other decomposition schemes.

CONCLUSION & OUTLOOK

In this thesis, an alchemical atomic energy decomposition was extended and utilized to advance efforts in the computational design of compounds. In particular, physically motivated models of stability trends throughout the chemical compound space were developed based on the alchemical decomposition.

For this purpose, first an existing decomposition scheme was modified, in which atomic energies of a compound are obtained via an alchemical transformation from the uniform electron gas into the compound. The modifications were (1) the consideration of the uniform external potential of the uniform electron gas and (2) the inclusion of periodic boundary conditions in the atomic energy expression. The modified and the previously proposed (see Ref 86) decompositions were compared for a test set of 15 small molecules. The mean absolute deviation of the sum of atomic energies from the total undecomposed energy was significantly reduced from 8 Ha with the scheme proposed in Ref 86 to 0.004 Ha with the modified scheme developed in this thesis. Furthermore, the sensitivity of the atomic energies on different approximations, e.g. numerical integration over the transformation path, was investigated. As a result, an approximate and affordable, yet sufficiently accurate approach for the calculation of atomic energies was proposed.

Next, the alchemical energy decomposition was used to study chemical bonding trends throughout the chemical compound space. The relationship between atomic contributions to the bond dissociation energy and the nuclear charge of the binding partner was analysed for single bonds A-B with A and B being *p*-block elements from the second row. Based on this analysis, a simple model of bond dissociation energies just depending on the nuclear charges of A and B as variables and three parameters was constructed. The model can describe covalent single and double bonds as long as the binding partners share the same first and second quantum number. This was demonstrated by accurately reproducing bond dissociation energies for A and B being *p*-block elements from either second, third or fourth row of the periodic table. Furthermore, the model was compared to Pauling's electronegativity based model for binding energies and its consistency with Hammett's equation was discussed. This study on bond dissociation energy trends highlights two distinct features of the alchemical energy decomposition. Firstly, the atomic energies depend strongly on the local environment, allowing an accurate description of bonding by considering only the binding partners. Secondly, binding energies can be expressed as simple functions of the nuclear charges. In the case of free atoms, it had been previously shown, that such an energy expansion in nuclear charge can be related to the integrated electrostatic potential along the alchemical transformation path.⁸² This thesis has provided numerical evidence that such an expansion could also be used to model stability trends among molecules. Hence, the aforementioned two features of the alchemical decomposition scheme allow the construction of simple, physically motivated models for bonding throughout the chemical compound space. While the model presented here is limited to a rather small set of

molecules, it could inspire more sophisticated models, applicable throughout larger parts of the chemical compound space. It might for example be possible to account for different local environments by adapting the parameters of the model. Another approach to extend the model to more diverse chemistries could be the use of environment dependent partial charges instead of the fixed nuclear charges.

Modelling of stability trends for a more diverse set of compounds was also investigated in a second application by calculating atomic energies for 1325 organic molecules. The analysis of this dataset revealed that atoms have similar atomic energies across different molecules if they have the same binding partners – confirming the strong dependence on the local environment. This transferability was exploited to estimate atomization energies from atomic contributions with a MAE of 23 kcal/mol. Furthermore, the dependence of alchemical atomic energies on the electrostatic potential motivated the use of the nuclear electrostatic potential of the binding partners as simple physically meaningful descriptor of atomic energy trends for varying local environment.

While the prediction error was quite high compared to chemical accuracy (1 kcal/mol), the estimates might still be useful to assess relative stabilities in large and diverse chemical spaces for two reasons. The approach is remarkably efficient (80 times faster than the semi-empirical xTB method)^{168,169} due to its low computational complexity and predictions should be robust due to their connection to the physics based alchemical energy decomposition scheme. Furthermore, it might serve as baseline for machine learning models, also known as Δ -machine learning,¹⁷⁰ if higher accuracy is needed or it could be possible to improve estimates by considering the influence of atoms beyond the direct binding partners. An interesting approach in this context might be the partitioning of the energy based on higher order derivatives $\partial_{Z_I}^n E$, which is in principle possible via an alchemical energy decomposition.

The alchemical energy decomposition was also compared to two other atomic decomposition schemes, the interacting quantum atom and a basis set based decomposition utilizing intrinsic bond and intrinsic atomic orbitals. Although differently defined, for all three decomposition schemes a consistent, chemically intuitive relationship can be observed: lower atomic energies are associated with lower electronegativity of the binding partners. This result indicates, that physics based decompositions should yield similar atomic energy relationships and hence *ab initio* decomposition, such as the alchemical one, could be useful to validate and interpret atomic energies from machine learning models.

While the overall trend between the *ab initio* decompositions is similar, the alchemical decomposition is more sensitive to changes in the electrostatic potential. Thus, it might be specifically useful to resolve and model systems and processes where the electrostatic potential varies substantially, such as defect formation or diffusion in solids, extending on previous work in this area with other decomposition schemes.^{171–174}

Overall, this thesis has shown the potential of an alchemical atomic energy decomposition for computational design by providing simple physically motivated models for the prediction of bond dissociation and atomization energies. The success can be assigned to two inherent features of the method. The atomic energy depends strongly on the local environment and the impact of this environment can be modelled by simple descriptors such as nuclear

charges or nuclear electrostatic potential of binding partners. In future work, energy predictions could likely be improved by including interactions beyond the direct binding partners. Furthermore, the results of this thesis indicate, that alchemical atomic energies might be specifically useful to study systems with large changes in electrostatics, such as defects or diffusion in materials and to improve machine learning models by validating their atomic energies.

BIBLIOGRAPHY

- [1] J. W. ERISMAN, M. A. SUTTON, J. GALLOWAY, Z. KLIMONT, W. WINIWARTER: How a century of ammonia synthesis changed the world, *Nature Geoscience*, **1**, 636 (2008).
- [2] W. A. ADEDEJI: The treasure called antibiotics, *Annals of Ibadan Postgraduate Medicine*, **14**, 56 (2016).
- [3] I. DOYTCHINOVA: Drug design—past, present, future, *Molecules*, **27**, 1496 (2022).
- [4] S. MYERS, A. BAKER: Drug discovery—an operating model for a new era, *Nature Biotechnology*, **19**, 727 (2001).
- [5] S. Y. TAN, Y. TATSUMURA: Alexander Fleming (1881–1955): Discoverer of penicillin, *Singapore Medical Journal*, **56**, 366 (2015).
- [6] I. DOYTCHINOVA, M. ATANASOVA, I. VALKOVA, G. STAVRAKOV, I. PHILIPOVA, Z. ZHIVKOVA, D. ZHELEVA-DIMITROVA, S. KONSTANTINOV, I. DIMITROV: Novel hits for acetylcholinesterase inhibition derived by docking-based screening on zinc database, *Journal of Enzyme Inhibition and Medicinal Chemistry*, **33**, 768 (2018).
- [7] M. SEIFRID, R. POLLICE, A. AGUILAR-GRANDA, Z. MORGAN CHAN, K. HOTTA, C. T. SER, J. VESTFRID, T. C. WU, A. ASPURU-GUZIĆ: Autonomous chemical experiments: Challenges and perspectives on establishing a self-driving lab, *Accounts of Chemical Research*, **55**, 2454 (2022).
- [8] P. A. M. DIRAC, R. H. FOWLER: Quantum mechanics of many-electron systems, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **123**, 714 (1929).
- [9] P. HOHENBERG, W. KOHN: Inhomogeneous electron gas, *Physical Review*, **136**, B864 (1964).
- [10] W. KOHN, L. J. SHAM: Self-consistent equations including exchange and correlation effects, *Physical Review*, **140**, A1133 (1965).
- [11] W. KOCH, M. C. HOLTHAUSEN: *A Chemist's Guide to Density Functional Theory*, John Wiley & Sons, Ltd (2001).
- [12] A. REMMEL: What exascale computing could mean for chemistry, *C&EN*, **100** (2022).
- [13] Powering catalysis with supercomputers, *Nature Catalysis*, **1**, 633 (2018).
- [14] G.-J. CHENG, X. ZHANG, L. W. CHUNG, L. XU, Y.-D. WU: Computational organic chemistry: Bridging theory and experiment in establishing the mechanisms of chemical reactions, *Journal of the American Chemical Society*, **137**, 1706 (2015).
- [15] L. GONZÁLEZ, D. ESCUDERO, L. SERRANO-ANDRÉS: Progress and challenges in the calculation of electronic excited states, *ChemPhysChem*, **13**, 28 (2012).
- [16] O. A. VON LILIENFELD: Towards the Computational Design of Compounds from First Principles, in V. BACH, L. D. SITE (editors), *Many-Electron Approaches in Physics, Chemistry and Mathematics: A Multidisciplinary View*, 169–189, Springer Cham (2014).
- [17] S. CURTAROLO, G. L. W. HART, M. B. NARDELLI, N. MINGO, S. SANVITO, O. LEVY: The high-throughput highway to computational materials design, *Nature Materials*, **12**, 191 (2013).

- [18] N. OKIMOTO, N. FUTATSUGI, H. FUJI, A. SUENAGA, G. MORIMOTO, R. YANAI, Y. OHNO, T. NARUMI, M. TAJI: High-performance drug discovery: Computational screening by combining docking and molecular dynamics simulations, *PLOS Computational Biology*, **5**, 1 (2009).
- [19] R. RAMAKRISHNAN, P. O. DRAL, M. RUPP, O. A. V. LILIENFELD: Quantum chemistry structures and properties of 134 kilo molecules, *Scientific Data*, **1**, 1 (2014).
- [20] S. LEE, B. KIM, H. CHO, H. LEE, S. Y. LEE, E. S. CHO, J. KIM: Computational screening of trillions of metal–organic frameworks for high-performance methane storage, *ACS Applied Materials & Interfaces*, **13**, 23647 (2021).
- [21] P. BORLIDO, J. SCHMIDT, H.-C. WANG, S. BOTTI, M. A. L. MARQUES: Computational screening of materials with extreme gap deformation potentials, *npj Computational Materials*, **8**, 156 (2022).
- [22] B. HUANG, O. A. VON LILIENFELD: Ab Initio Machine Learning in Chemical Compound Space, *Chemical Reviews*, **121**, 10001 (2021).
- [23] R. S. BOHACEK, C. McMARTIN, W. C. GUIDA: The art and practice of structure-based drug design: A molecular modeling perspective, *Medicinal Research Reviews*, **16**, 3 (1996).
- [24] D. W. DAVIES, K. T. BUTLER, A. J. JACKSON, A. MORRIS, J. M. FROST, J. M. SKELTON, A. WALSH: Computational screening of all stoichiometric inorganic materials, *Chem*, **1**, 617 (2016).
- [25] A. FRANCESCHETTI, A. ZUNGER: The inverse band-structure problem of finding an atomic configuration with given electronic properties, *Nature*, **402**, 60 (1999).
- [26] K. KARANDASHEV, J. WEINREICH, S. HEINEN, D. J. ARISMENDI ARRIETA, G. F. VON RUDORFF, K. HERMANSSON, O. A. VON LILIENFELD: Evolutionary Monte Carlo of QM Properties in Chemical Space: Electrolyte Design, *Journal of Chemical Theory and Computation*, **19**, 8861 (2023).
- [27] M. RUPP, A. TKATCHENKO, K. R. MÜLLER, O. A. V. LILIENFELD: Fast and accurate modeling of molecular atomization energies with machine learning, *Physical Review Letters*, **108**, 1 (2012).
- [28] G. MONTAVON, M. RUPP, V. GOBRE, A. VAZQUEZ-MAYAGOITIA, K. HANSEN, A. TKATCHENKO, K. R. MÜLLER, O. A. V. LILIENFELD: Machine learning of molecular electronic properties in chemical compound space, *New Journal of Physics*, **15**, 095003 (2013).
- [29] K. HANSEN, F. BIEGLER, R. RAMAKRISHNAN, W. PRONOBIS, O. A. V. LILIENFELD, K. R. MÜLLER, A. TKATCHENKO: Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space, *The Journal of Physical Chemistry Letters*, **6**, 2326 (2015).
- [30] F. A. FABER, L. HUTCHISON, B. HUANG, J. GILMER, S. S. SCHOENHOLZ, G. E. DAHL, O. VINYALS, S. KEARNES, P. F. RILEY, O. A. VON LILIENFELD: Prediction errors of molecular machine learning models lower than hybrid dft error, *Journal of Chemical Theory and Computation*, **13**, 5255 (2017).
- [31] B. HUANG, O. A. VON LILIENFELD: Quantum machine learning using atom-in-molecule-based fragments selected on the fly, *Nature Chemistry*, **12**, 945 (2020).
- [32] M. RUPP, R. RAMAKRISHNAN, O. A. V. LILIENFELD: Machine learning for quantum mechanical properties of atoms in molecules, *Journal of Physical Chemistry Letters*, **6**, 3309 (2015).

-
- [33] F. M. PARUZZO, A. HOFSTETTER, F. MUSIL, S. DE, M. CERIOTTI, L. EMSLEY: Chemical shifts in molecular solids by machine learning, *Nature Communications*, **9**, 4501 (2018).
- [34] B. HUANG, O. A. VON LILIENFELD: Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity, *The Journal of Chemical Physics*, **145**, 161102 (2016).
- [35] E. A. HYLLERAAS: Über den Grundterm der Zweielektronenprobleme von H^- , He, Li^+ , Be^{++} usw., *Zeitschrift für Physik*, **65**, 209 (1930).
- [36] W. KOHN: Two Applications of the Variational Method to Quantum Mechanics, *Physical Review*, **71**, 635 (1947).
- [37] L. L. FOLDY: A Note on Atomic Binding Energies, *Physical Review*, **83**, 397 (1951).
- [38] P.-O. LÖWDIN: Scaling problem, virial theorem, and connected relations in quantum mechanics, *Journal of Molecular Spectroscopy*, **3**, 46 (1959).
- [39] J. WILSON, E. BRIGHT: Four-Dimensional Electron Density Function, *The Journal of Chemical Physics*, **36**, 2232 (2004).
- [40] P. POLITZER, R. G. PARR: Some new energy formulas for atoms and molecules, *The Journal of Chemical Physics*, **61**, 4258 (1974).
- [41] P. G. MEZEY: Electronic energy inequalities for isoelectronic molecular systems, *Theoretica Chimica Acta*, **59**, 321 (1980).
- [42] P. G. MEZEY: Level set topology of the nuclear charge space and the electronic energy functional, *International Journal of Quantum Chemistry*, **22**, 101 (1982).
- [43] P. G. MEZEY: A simple relation between nuclear charges and potential surfaces, *Journal of the American Chemical Society*, **107**, 3100 (1985).
- [44] O. A. V. LILIENFELD, M. E. TUCKERMAN: Molecular grand-canonical ensemble density functional theory and exploration of chemical space, *The Journal of Chemical Physics*, **125**, 154104 (2006).
- [45] O. A. V. LILIENFELD: First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties, *International Journal of Quantum Chemistry*, **113**, 1676 (2013).
- [46] O. A. V. LILIENFELD, R. D. LINS, U. ROTHISBERGER: Variational particle number approach for rational compound design, *Physical Review Letters*, **95**, 1 (2005).
- [47] R. BALAWENDER, M. A. WELEAREGAY, M. LESIUK, F. DE PROFT, P. GEERLINGS: Exploring chemical space with the alchemical derivatives, *Journal of Chemical Theory and Computation*, **9**, 5327 (2013).
- [48] T. P. STRAATSMAN, J. A. MCCAMMON: Computational alchemy, *Annual Review of Physical Chemistry*, **43**, 407 (1992).
- [49] M. LESIUK, R. BALAWENDER, J. ZACHARA: Higher order alchemical derivatives from coupled perturbed self-consistent field theory, *The Journal of Chemical Physics*, **136**, 034104 (2012).
- [50] G. DOMENICHINI, O. A. VON LILIENFELD: Alchemical geometry relaxation, *The Journal of Chemical Physics*, **156**, 184801 (2022).
- [51] G. F. VON RUDORFF, O. A. VON LILIENFELD: Alchemical perturbation density functional theory, *Physical Review Research*, **2**, 23220 (2020).

- [52] R. BALAWENDER, M. A. WELEAREGAY, M. LESIUK, F. D. PROFT, P. GEERLINGS: Exploring chemical space with the alchemical derivatives, *Journal of Chemical Theory and Computation*, **9**, 5327 (2013).
- [53] G. F. VON RUDORFF, O. A. VON LILIENFELD: Rapid and accurate molecular deprotonation energies from quantum alchemy, *Physical Chemistry Chemical Physics*, **22**, 10519 (2020).
- [54] M. MUÑOZ, A. ROBLES-NAVARRO, P. FUENTEALBA, C. CÁRDENAS: Predicting deprotonation sites using alchemical derivatives, *The Journal of Physical Chemistry A*, **124**, 3754 (2020).
- [55] K. SARAVANAN, J. R. KITCHIN, O. A. V. LILIENFELD, J. A. KEITH: Alchemical predictions for computational catalysis: Potential and limitations, *Journal of Physical Chemistry Letters*, **8**, 5002 (2017).
- [56] J. G. FREEZE, H. R. KELLY, V. S. BATISTA: Search for catalysts by inverse design: Artificial intelligence, mountain climbers, and alchemists, *Chemical Reviews*, **119**, 6595 (2019).
- [57] C. D. GRIEGO, J. R. KITCHIN, J. A. KEITH: Acceleration of catalyst discovery with easy, fast, and reproducible computational alchemy, *International Journal of Quantum Chemistry*, **121**, e26380 (2021).
- [58] N. MARZARI, S. DE GIRONCOLI, S. BARONI: Structure and phase stability of $\text{Ga}_x\text{In}_{1-x}\text{P}$ solid solutions from computational alchemy, *Physical Review Letters*, **72**, 4001 (1994).
- [59] K. Y. CHANG, O. A. V. LILIENFELD: Alx gal-xas crystals with direct 2 ev band gaps from computational alchemy, *Physical Review Materials*, **2**, 1 (2018).
- [60] J. M. GUEVARA-VELA, E. FRANCISCO, T. ROCHA-RINZA, A. M. PENDÁS: Interacting Quantum Atoms - A Review, *Molecules*, **25**, 1 (2020).
- [61] M. A. BLANCO, A. MARTÍN PENDÁS, E. FRANCISCO: Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme Based on the Quantum Theory of Atoms in Molecules, *Journal of Chemical Theory and Computation*, **1**, 1096 (2005).
- [62] J. J. ERIKSEN: Mean-field density matrix decompositions, *The Journal of Chemical Physics*, **153**, 214109 (2020).
- [63] A. MARTÍN PENDÁS, E. FRANCISCO, M. A. BLANCO: Binding Energies of First Row Diatomics in the Light of the Interacting Quantum Atoms Approach, *The Journal of Physical Chemistry A*, **110**, 12864 (2006).
- [64] M. G. DARLEY, P. L. POPELIER: Role of short-range electrostatics in torsional potentials, *The Journal of Physical Chemistry A*, **112**, 12954 (2008).
- [65] I. ALKORTA, J. C. R. THACKER, P. L. A. POPELIER: An interacting quantum atom study of model $\text{S}_\text{N}2$ reactions ($\text{X}^-\cdots\text{CH}_3\text{X}$, $\text{X} = \text{F}, \text{Cl}, \text{Br}, \text{and I}$), *Journal of Computational Chemistry*, **39**, 546 (2018).
- [66] J. JARA-CORTÉS, J. HERNÁNDEZ-TRUJILLO: Energetic analysis of conjugated hydrocarbons using the interacting quantum atoms method, *Journal of Computational Chemistry*, **39**, 1103 (2018).
- [67] P. L. POPELIER: QCTFF: On the construction of a novel protein force field, *International Journal of Quantum Chemistry*, **115**, 1005 (2015).
- [68] P. L. A. POPELIER: Molecular simulation by knowledgeable quantum atoms, *Physica Scripta*, **91**, 033007 (2016).

- [69] M. J. BURN, P. L. POPELIER: Creating Gaussian process regression models for molecular simulations using adaptive sampling, *The Journal of Chemical Physics*, **153** (2020).
- [70] B. C. B. SYMONS, M. K. BANE, P. L. A. POPELIER: DL_FFLUX: A Parallel, Quantum Chemical Topology Force Field, *Journal of Chemical Theory and Computation*, **17**, 7043 (2021).
- [71] M. J. BURN, P. L. A. POPELIER: Gaussian process regression models for predicting atomic energies and multipole moments, *Journal of Chemical Theory and Computation*, **19**, 1370 (2023).
- [72] F. ZIELINSKI, P. I. MAXWELL, T. L. FLETCHER, S. J. DAVIE, N. DI PASQUALE, S. CARDAMONE, M. J. L. MILLS, P. L. A. POPELIER: Geometry optimization with machine trained topological atoms, *Scientific Reports*, **7**, 12817 (2017).
- [73] M. L. BROWN, J. M. SKELTON, P. L. A. POPELIER: Construction of a gaussian process regression model of formamide for use in molecular simulations, *The Journal of Physical Chemistry A*, **127**, 1702 (2023).
- [74] M. L. BROWN, J. M. SKELTON, P. L. A. POPELIER: Application of the FFLUX Force Field to Molecular Crystals: A Study of Formamide, *Journal of Chemical Theory and Computation*, **19**, 7946 (2023).
- [75] B. C. B. SYMONS, P. L. A. POPELIER: Application of Quantum Chemical Topology Force Field FFLUX to Condensed Matter Simulations: Liquid Water, *Journal of Chemical Theory and Computation*, **18**, 5577 (2022).
- [76] F. Ø. KJELDAL, J. J. ERIKSEN: Properties of Local Electronic Structures, *Journal of Chemical Theory and Computation*, **19**, 9228 (2023).
- [77] J. J. ERIKSEN: Electronic excitations through the prism of mean-field decomposition techniques, *The Journal of Chemical Physics*, **156**, 061101 (2022).
- [78] F. Ø. KJELDAL, J. J. ERIKSEN: Decomposing Chemical Space: Applications to the Machine Learning of Atomic Energies, *Journal of Chemical Theory and Computation*, **19**, 2029 (2023).
- [79] P. POLITZER, J. S. MURRAY: The fundamental nature and role of the electrostatic potential in atoms and molecules, *Theoretical Chemistry Accounts*, **108**, 134 (2002).
- [80] P. POLITZER, J. S. MURRAY: Electrostatic potentials at the nuclei of atoms and molecules, *Theoretical Chemistry Accounts*, **140**, 7 (2021).
- [81] M. LEVY, Y. TAL, S. C. CLEMENT: A discontinuous energy-density functional, *The Journal of Chemical Physics*, **77**, 3140 (1982).
- [82] H. TERUYA, T. ANNO: An energy-density equation for isoelectronic changes in atoms as revealed from the z dependence of the electrostatic potential at the nucleus for isoelectronic series of atoms, *The Journal of Chemical Physics*, **79**, 6162 (1983).
- [83] J. L. GÁZQUEZ, A. VELA: Binding energies of atoms and ions: The Z^{-1} perturbation expansion and the Thomas-Fermi limit, *Physical Review A*, **38**, 3264 (1988).
- [84] Y. TAL, M. LEVY: Recursion theory for nonrelativistic ground-state atomic energies and expectation values of r^{-1} , *Physical Review A*, **25**, 1838 (1982).
- [85] K. D. SEN, J. M. SEMINARIO, P. POLITZER: Z-transition state calculations of energy changes and electrostatic potentials in isoelectronic atoms and molecules, *The Journal of Chemical Physics*, **90**, 4373 (1989).

- [86] G. F. VON RUDORFF, O. A. VON LILIENFELD: Atoms in Molecules from Alchemical Perturbation Density Functional Theory, *The Journal of Physical Chemistry B*, **123**, 10073 (2019).
- [87] A. SZABO, N. S. OSTLUND: *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications, Inc., Mineola, first edition (1996).
- [88] M. BORN, R. OPPENHEIMER: Zur Quantentheorie der Molekeln, *Annalen der Physik*, **389**, 457 (1927).
- [89] S. MATSIKA: Electronic structure methods for the description of nonadiabatic effects and conical intersections, *Chemical Reviews*, **121**, 9407 (2021).
- [90] V. FOCK: Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems, *Zeitschrift für Physik*, **61**, 126 (1930).
- [91] J. C. SLATER: Note on Hartree's Method, *Physical Review*, **35**, 210 (1930).
- [92] D. R. HARTREE, W. HARTREE: Self-consistent field, with exchange, for beryllium, *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences*, **150**, 9 (1935).
- [93] C. DAVID SHERRILL, H. F. SCHAEFER: The configuration interaction method: Advances in highly correlated approaches, volume 34 of *Advances in Quantum Chemistry*, 143–269, Academic Press (1999).
- [94] J. P. PERDEW, K. SCHMIDT: Jacob's ladder of density functional approximations for the exchange-correlation energy, *AIP Conference Proceedings*, **577**, 1 (2001).
- [95] W. KOCH, M. C. HOLTHAUSEN: *A chemist's guide to density functional theory*, p. 19, Wiley-VCH (2001).
- [96] W. KOCH, M. C. HOLTHAUSEN: *A chemist's guide to density functional theory*, p. 33–34, Wiley-VCH (2001).
- [97] W. KOCH, M. C. HOLTHAUSEN: *A chemist's guide to density functional theory*, p. 41, Wiley-VCH (2001).
- [98] L. H. THOMAS: The calculation of atomic fields, *Mathematical Proceedings of the Cambridge Philosophical Society*, **23**, 542–548 (1927).
- [99] E. FERMI: Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente, *Zeitschrift für Physik*, **48**, 73 (1928).
- [100] K. BURKE: The ABC of DFT, *Flexo*, **40**, 22 (2015).
- [101] M. DÄNE, A. GONIS: On the v-Representability Problem in Density Functional Theory: Application to Non-Interacting Systems, *Computation*, **4**, 24 (2016).
- [102] W. KOCH, M. C. HOLTHAUSEN: *A chemist's guide to density functional theory*, p. 41–43, Wiley-VCH (2001).
- [103] P. ROBERT G., W. YANG: *Density-Functional Theory of Atoms and Molecules.*, number Vol. 16 in International Series of Monographs on Chemistry, Oxford University Press (1989).
- [104] A. D. BECKE: Perspective: Fifty years of density-functional theory in chemical physics, *The Journal of Chemical Physics*, **140**, 18A301 (2014).
- [105] P.-F. LOOS, P. M. W. GILL: The uniform electron gas, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **6**, 410 (2016).

- [106] W. KOCH, M. C. HOLTHAUSEN: *A chemist's guide to density functional theory*, pp. 70, Wiley-VCH (2001).
- [107] F. BLOCH: Bemerkung zur Elektronentheorie des Ferromagnetismus und der elektrischen Leitfähigkeit, *Zeitschrift für Physik*, **57**, 545 (1929).
- [108] P. A. M. DIRAC: Note on Exchange Phenomena in the Thomas Atom, *Mathematical Proceedings of the Cambridge Philosophical Society*, **26**, 376–385 (1930).
- [109] J. C. SLATER: A Simplification of the Hartree-Fock Method, *Physical Review*, **81**, 385 (1951).
- [110] D. M. CEPERLEY, B. J. ALDER: Ground State of the Electron Gas by a Stochastic Method, *Physical Review Letters*, **45**, 566 (1980).
- [111] S. H. VOSKO, L. WILK, M. NUSAIR: Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, *Canadian Journal of Physics*, **58**, 1200 (1980).
- [112] J. P. PERDEW, Y. WANG: Accurate and simple analytic representation of the electron-gas correlation energy, *Physical Review B*, **45**, 13244 (1992).
- [113] W. KOCH, M. C. HOLTHAUSEN: *A chemist's guide to density functional theory*, pp. 75, Wiley-VCH (2001).
- [114] W. KOCH, M. C. HOLTHAUSEN: *A chemist's guide to density functional theory*, pp. 20, Wiley-VCH (2001).
- [115] A. PRIBRAM-JONES, D. A. GROSS, K. BURKE: DFT: A Theory Full of Holes?, *Annual Review of Physical Chemistry*, **66**, 283 (2015).
- [116] A. D. BECKE: Density functional calculations of molecular bond energies, *The Journal of Chemical Physics*, **84**, 4524 (1986).
- [117] J. P. PERDEW, K. BURKE, M. ERNZERHOF: Generalized gradient approximation made simple, *Physical Review Letters*, **77**, 3865 (1996).
- [118] F. ZAHARIEV, S. S. LEANG, M. S. GORDON: Functional derivatives of meta-generalized gradient approximation (meta-GGA) type exchange-correlation density functionals, *The Journal of Chemical Physics*, **138**, 244108 (2013).
- [119] W. KOCH, M. C. HOLTHAUSEN: *A chemist's guide to density functional theory*, pp. 78, Wiley-VCH (2001).
- [120] A. D. BECKE: A new mixing of Hartree-Fock and local density-functional theories, *The Journal of Chemical Physics*, **98**, 1372 (1993).
- [121] P. J. STEPHENS, F. J. DEVLIN, C. F. CHABALOWSKI, M. J. FRISCH: Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields, *The Journal of Physical Chemistry*, **98**, 11623 (1994).
- [122] J. P. PERDEW, M. ERNZERHOF, K. BURKE: Rationale for mixing exact exchange with density functional approximations, *The Journal of Chemical Physics*, **105**, 9982 (1996).
- [123] C. ADAMO, V. BARONE: Toward reliable density functional methods without adjustable parameters: The PBE0 model, *The Journal of Chemical Physics*, **110**, 6158 (1999).
- [124] I. Y. ZHANG, X. XU: On the top rung of Jacob's ladder of density functional theory: Toward resolving the dilemma of SIE and NCE, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **11**, e1490 (2021).

- [125] J. P. PERDEW, K. SCHMIDT: Jacob's ladder of density functional approximations for the exchange-correlation energy, *AIP Conference Proceedings*, **577**, 1 (2001).
- [126] G. P. CHEN, V. K. VOORA, M. M. AGEE, S. G. BALASUBRAMANI, F. FURCHE: Random-Phase Approximation Methods, *Annual Review of Physical Chemistry*, **68**, 421 (2017).
- [127] S. GRIMME, F. NEESE: Double-hybrid density functional theory for excited electronic states of molecules, *The Journal of Chemical Physics*, **127**, 154116 (2007).
- [128] N. W. ASHCROFT, N. D. MERMIN: *Solid State Physics*, pp. 64–83, Holt-Saunders (1976).
- [129] N. W. ASHCROFT, N. D. MERMIN: *Solid State Physics*, pp. 86–94, Holt-Saunders (1976).
- [130] R. M. MARTIN: *Electronic Structure: Basic Theory and Practical Methods*, pp. 262–264, Cambridge University Press (2004).
- [131] R. M. MARTIN: *Electronic Structure: Basic Theory and Practical Methods*, pp. 238–241, Cambridge University Press (2004).
- [132] D. R. HAMANN, M. SCHLÜTER, C. CHIANG: Norm-conserving pseudopotentials, *Physical Review Letters*, **43**, 1494 (1979).
- [133] S. GOEDECKER, K. MASCHKE: Transferability of pseudopotentials, *Physical Review A*, **45**, 88 (1992).
- [134] C. HARTWIGSEN, S. GOEDECKER, J. HUTTER: Relativistic separable dual-space Gaussian pseudopotentials from H to Rn, *Physical Review B*, **58**, 3641 (1998).
- [135] M. KRACK: Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals, *Theoretical Chemistry Accounts*, **114**, 145 (2005).
- [136] H. HANS: *Einführung in die Quantenchemie*, Deuticke (1937).
- [137] R. P. FEYNMAN: Forces in molecules, *Physical Review*, **56**, 340 (1939).
- [138] G. KNIZIA: Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts, *Journal of Chemical Theory and Computation*, **9**, 4834 (2013).
- [139] F. L. HIRSHFELD: Bonded-atom fragments for describing molecular charge densities, *Theoretica chimica acta*, **44**, 129 (1977).
- [140] R. F. W. BADER: *Atoms in Molecules*, Oxford University Press (1990).
- [141] A. D. BECKE: A multicenter numerical integration scheme for polyatomic molecules, *The Journal of Chemical Physics*, **88**, 2547 (1988).
- [142] R. S. MULLIKEN: Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I, *The Journal of Chemical Physics*, **23**, 1833 (1955).
- [143] J. J. ERIKSEN: Erratum: “Electronic excitations through the prism of mean-field decomposition techniques” [J. Chem. Phys. 156, 061101 (2022)], *The Journal of Chemical Physics*, **158**, 029901 (2023).
- [144] A. MARTÍN PENDÁS, E. FRANCISCO, D. SUÁREZ, A. COSTALES, N. DÍAZ, J. MUNÁRRIZ, T. ROCHA-RINZA, J. M. GUEVARA-VELA: Atoms in molecules in real space: a fertile field for chemical bonding, *Physical Chemistry Chemical Physics*, **25**, 10231 (2023).

- [145] J. IHM, A. ZUNGER, M. L. COHEN: Momentum-space formalism for the total energy of solids, *Journal of Physics C: Solid State Physics*, **12**, 4409 (1979).
- [146] R. M. MARTIN: *Electronic Structure: Basic Theory and Practical Methods*, pp. 267, Cambridge University Press (2004).
- [147] R. M. MARTIN: *Electronic Structure: Basic Theory and Practical Methods*, p. 284, Cambridge University Press (2004).
- [148] R. M. MARTIN: *Electronic Structure: Basic Theory and Practical Methods*, pp. 607, Cambridge University Press (2004).
- [149] COPYRIGHT IBM CORP. 1990-2023 AND COPYRIGHT 1994-2001 BY MAX PLANCK INSTITUTE, STUTTGART.: CPMD 4.3, <https://github.com/CPMD-code/CPMD>. (date of access: 13:42 30.01.2022).
- [150] C. D. GRIEGO, K. SARAVANAN, J. A. KEITH: Benchmarking Computational Alchemy for Carbide, Nitride, and Oxide Catalysts, *Advanced Theory and Simulations*, **2**, 1800142 (2019).
- [151] M. G. KENDALL: A new measure of rank correlation, *Biometrika*, **30**, 81 (1938).
- [152] J. SCOTT: LXXXII. The binding energy of the Thomas-Fermi Atom, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **43**, 859 (1952).
- [153] J. SCHWINGER: Thomas-Fermi model: The leading correction, *Physical Review A*, **22**, 1827 (1980).
- [154] A. KARTON, N. SYLVETSKY, J. M. MARTIN: W4-17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods, *Journal of Computational Chemistry*, **38**, 2063 (2017).
- [155] M. RUPP: Machine learning for quantum mechanics in a nutshell, *International Journal of Quantum Chemistry*, **115**, 1058 (2015).
- [156] J. STEWART: Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters., *Journal of Molecular Modelling*, **19**, 1 (2013).
- [157] R. HAUNSCHILD, W. KLOPPER: New accurate reference energies for the G2/97 test set, *The Journal of Chemical Physics*, **136**, 164102 (2012).
- [158] L. PAULING: *The Nature of the Chemical Bond*, 79–96, Cornell University Press (1960).
- [159] L. P. HAMMETT: The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives, *Journal of the American Chemical Society*, **59**, 96 (1937).
- [160] A. M. CHANG, J. G. FREEZE, V. S. BATISTA: Hammett neural networks: prediction of frontier orbital energies of tungsten-benzylidyne photoredox complexes, *Chemical Science*, **10**, 6844 (2019).
- [161] M. BRAGATO, G. F. VON RUDORFF, O. A. VON LILIENFELD: Data enhanced Hammett-equation: Reaction barriers in chemical space, *Chemical Science*, **11**, 11859 (2020).
- [162] L. K. H. VAN BEEK: A relationship between dipole moments and the Hammett equation, *Recueil des Travaux Chimiques des Pays-Bas*, **76**, 729 (1957).
- [163] I. FERNÁNDEZ, G. FRENKING: Correlation between Hammett Substituent Constants and Directly Calculated π -Conjugation Strength, *The Journal of Organic Chemistry*, **71**, 2251 (2006).

- [164] I. MAYER: Charge, bond order and valence in the AB initio SCF theory, *Chemical Physics Letters*, **97**, 270 (1983).
- [165] I. MAYER: Bond order and valence: Relations to Mulliken's population analysis, *International Journal of Quantum Chemistry*, **26**, 151 (1984).
- [166] A. J. BRIDGEMAN, G. CAVIGLIASSO, L. R. IRELAND, J. ROTHERY: The Mayer bond order as a tool in inorganic chemistry, *Journal of the Chemical Society, Dalton Transactions*, 2095–2108 (2001).
- [167] O. ANATOLE VON LILIENFELD: Accurate ab initio energy gradients in chemical compound space, *The Journal of Chemical Physics*, **131**, 164102 (2009).
- [168] C. BANNWARTH, S. EHLERT, S. GRIMME: GFN2-xTB-An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions, *Journal of Chemical Theory and Computation*, **15**, 1652 (2019).
- [169] C. BANNWARTH, E. CALDEWEYHER, S. EHLERT, A. HANSEN, P. PRACHT, J. SEIBERT, S. SPICHER, S. GRIMME: Extended tight-binding quantum chemistry methods, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **11**, e1493 (2021).
- [170] R. RAMAKRISHNAN, P. O. DRAL, M. RUPP, O. A. V. LILIENFELD: Big data meets quantum chemistry approximations: The Δ -machine learning approach, *Journal of Chemical Theory and Computation*, **11**, 2087 (2015).
- [171] Y. DAN, D. R. TRINKLE: First-principles core energies of isolated basal and prism screw dislocations in magnesium, *Materials Research Letters*, **10**, 360 (2022).
- [172] Z. XU, S. TANAKA, M. KOHYAMA: Atomic configurations and energies of mg symmetric tilt grain boundaries: ab initio local analysis, *Modelling and Simulation in Materials Science and Engineering*, **29**, 085010 (2021).
- [173] H. S. OH, S. J. KIM, K. ODBADRAKH, W. H. RYU, K. N. YOON, S. MU, F. KÖRMANN, Y. IKEDA, C. C. TASAN, D. RAABE, T. EGAMI, E. S. PARK: Engineering atomic-level complexity in high-entropy and complex concentrated alloys, *Nature Communications*, **10**, 2090 (2019).
- [174] C. LI, J. YIN, K. ODBADRAKH, B. C. SALES, S. J. ZINKLE, G. M. STOCKS, B. D. WIRTH: First principle study of magnetism and vacancy energetics in a near equimolar NiFeMnCr high entropy alloy, *Journal of Applied Physics*, **125**, 155103 (2019).

APPENDIX: REPRINTED PUBLICATIONS

The papers on which this thesis is based on are reported in this appendix. The results on the extension of the method in chapter 3 and the application of the decomposition to a large set of organic molecules in chapter 5 are presented in appendix A.1. The results on the bond dissociation energy model in chapter 4 are reported in appendix A.2.

APPENDIX A.1 TRANSFERABILITY OF ATOMIC ENERGIES FROM ALCHEMICAL DECOMPOSITION

APPENDIX A.1.1

Transferability of atomic energies from alchemical decomposition

MICHAEL J. SAHRE, GUIDO FALK VON RUDORFF, PHILIPP MARQUETAND AND O.
ANATOLE VON LILIENFELD

J. Chem. Phys. 160, 054106 (2024).
<https://doi.org/10.1063/5.0187298>

Contributions:

MICHAEL J. SAHRE contributed to the conceptualization of the project, developed the methodology, performed calculation and analysis, wrote the initial draft and contributed to the review and editing of the final manuscript.

GUIDO FALK VON RUDORFF contributed to the analysis and method implementation, and to the review and editing of the final manuscript.

PHILIPP MARQUETAND envisioned and supervised the project, and contributed to the review and editing of the final manuscript.

ANATOLE VON LILIENFELD supervised the project and contributed to the analysis and to the review and editing of the final manuscript.

Reprinted from *J. Chem. Phys.* 160, 054106 (2024), with the permission of AIP Publishing.

RESEARCH ARTICLE | FEBRUARY 06 2024

Transferability of atomic energies from alchemical decomposition

Michael J. Sahre  ; Guido Falk von Rudorff  ; Philipp Marquetand  ; O. Anatole von Lilienfeld  



J. Chem. Phys. 160, 054106 (2024)

<https://doi.org/10.1063/5.0187298>



CrossMark



Lock-in Amplifier



Boxcar Averager



Zurich Instruments

Find out more

Boost Your Optics and Photonics Measurements

Transferability of atomic energies from alchemical decomposition

Cite as: J. Chem. Phys. 160, 054106 (2024); doi: 10.1063/5.0187298

Submitted: 13 November 2023 • Accepted: 9 January 2024 •

Published Online: 6 February 2024



View Online



Export Citation



CrossMark

Michael J. Sahre,¹  Guido Falk von Rudorff,^{2,3}  Philipp Marquetand,⁴ 
and O. Anatole von Lilienfeld^{5,6,7,8,9,10,11,a)} 

AFFILIATIONS

¹Vienna Doctoral School in Chemistry (DoSChem) and Institute of Theoretical Chemistry and Faculty of Physics, University of Vienna, 1090 Vienna, Austria

²Department of Chemistry, University Kassel, Heinrich-Plett-Str.40, 34132 Kassel, Germany

³Center for Interdisciplinary Nanostructure Science and Technology (CINaT), Heinrich-Plett-Straße 40, 34132 Kassel, Germany

⁴Faculty of Chemistry, Institute of Theoretical Chemistry, University of Vienna, Währinger Str. 17, 1090 Vienna, Austria

⁵Chemical Physics Theory Group, Department of Chemistry, University of Toronto, St. George Campus, Toronto, M5S 3H6 Ontario, Canada

⁶Department of Materials Science and Engineering, University of Toronto, St. George Campus, Toronto, M5S 3E4 Ontario, Canada

⁷Vector Institute for Artificial Intelligence, Toronto, M5S 1M1 Ontario, Canada

⁸ML Group, Technische Universität Berlin and Institute for the Foundations of Learning and Data, 10587 Berlin, Germany

⁹Berlin Institute for the Foundations of Learning and Data, 10587 Berlin, Germany

¹⁰Department of Physics, University of Toronto, St. George Campus, Toronto, M5S 1A7 Ontario, Canada

¹¹Acceleration Consortium, University of Toronto, Toronto, M5R 0A3 Ontario, Canada

^{a)} Author to whom correspondence should be addressed: anatole.vonlilienfeld@utoronto.ca

ABSTRACT

We study alchemical atomic energy partitioning as a method to estimate atomization energies from atomic contributions, which are defined in physically rigorous and general ways through the use of the uniform electron gas as a joint reference. We analyze quantitatively the relation between atomic energies and their local environment using a dataset of 1325 organic molecules. The atomic energies are transferable across various molecules, enabling the prediction of atomization energies with a mean absolute error of 23 kcal/mol, comparable to simple statistical estimates but potentially more robust given their grounding in the physics-based decomposition scheme. A comparative analysis with other decomposition methods highlights its sensitivity to electrostatic variations, underlining its potential as a representation of the environment as well as in studying processes like diffusion in solids characterized by significant electrostatic shifts.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0187298>

I. INTRODUCTION

Understanding structure–property relationships is essential for the rational design of chemical compounds. In this regard, the decomposition of molecules into smaller fragments has emerged as a valuable concept in chemistry. The impact of a specific fragment on the molecular properties is often transferable,¹ enabling the estimation of molecular properties from fragment-based coarse-grained models. Due to their reduced dimensionality compared to models

explicitly considering the interactions between all electrons and nuclei, such fragment-based approaches can, for example, be used to study very large molecules.^{2,3} Furthermore, the lower complexity can be helpful in constructing simplified models that encompass only the essential molecular characteristics for explaining trends among various molecules. This is for example exploited in the Hammett model, which assumes the effect of a specific functional group on a molecular property to be a scalar value that is transferable across different molecules.^{4,5} More recently, the decomposition of

molecules into smaller fragments has also been used in combination with machine learning.^{6,7} For instance, so-called atom models infer the energy of a query compound based on a linear combination of reference results for its constituting building blocks,⁸ and many artificial neural networks predict total energies as a sum of atomic energies.^{9–11} These models decompose properties based on statistics and for a given set of training molecules but accurate out-of-sample predictions are generally not guaranteed.

Molecular properties can also be partitioned via rigorous physics-based decomposition schemes. One of the earliest definitions of atoms in molecules dates back to Politzer and Parr who showed that the energy of a molecule can be expressed exactly as a function of the electrostatic potentials $V_{A,0}$ at the positions of the nuclei as

$$E^{\text{mol}} = \frac{1}{2} \sum_A Z_A V_{A,0} - \frac{1}{2} \sum_A \int_0^{Z_A} \left[Z'_A \left(\frac{\partial V_{A,0}}{\partial Z'_A} \right) - V_{A,0} \right] dZ'_A, \quad (1)$$

with the electrostatic potential

$$V_{A,0} = \sum_{B \neq A} \frac{Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} - \int d\mathbf{r} \frac{\rho(\mathbf{r}, \{Z'_A\})}{|\mathbf{r} - \mathbf{R}_A|}, \quad (2)$$

and the nuclear charges and positions $Z_{A/B}$, $\mathbf{R}_{A/B}$ for all nuclei in the molecule.^{12–14} In the second term of Eq. (1), the nuclear charges $\{Z'_A\}$ are scaled up linearly from 0 to their respective value in the molecule. The electron density $\rho(\mathbf{r}, \{Z'_A\})$ is a function of the varying nuclear charges and transforms into the electron density of the molecule along the integration path starting from an infinitely spread out charge distribution. This idea was built on work by Wilson who employed the Hellman–Feynman theorem to demonstrate that the energy of a molecule can be expressed as a function of only the electron density and the external potential by scaling the nuclear charges as described above.¹⁵ Similar expressions for the energy of single atoms were derived even earlier^{16–18} and can be traced back to Hylleraas who calculated energies of Helium-like atoms by treating the energy dependence on the nuclear charge with perturbation theory.¹⁹ Since the scaling of nuclear charges corresponds to the generation or transformation of elements, the term quantum alchemy was introduced to describe these processes, which have been utilized to investigate changes in energy due to changes in elemental composition.^{20–27}

Equation (1) is an exact decomposition of the energy into atomic contributions but the calculation of these atomic energies is challenging since the isoelectronic density must be known as a function of the variable nuclear charges along the scaling path. In addition, it was demonstrated that atomic energies are path-dependent such that a meaningful comparison of atomic energies requires a common reference and scaling path.²⁸ To address these issues, an infinitely extended uniform electron gas (UEG) was proposed as a reference from which arbitrary isoelectronic compounds can be generated by linear up-scaling of the nuclear charges in the respective compound.^{28,29} We note that this is not the only option. Another reference could, for example, be a UEG inside a sphere³⁰ in combination with the “exact Coulomb cutoff technique”³¹ inside the same sphere.

In the case of the infinitely extended uniform electron gas, a positive background charge density equal to the negative uniform

electron density is present to ensure electric charge neutrality. As the positive background charge density approaches zero, the reference becomes equal to the infinitely spread-out electron density originally proposed by Wilson. For this limiting case, an alternative atomic energy expression to Eq. (1), decomposing the energy of an arbitrary compound E^{mol} with respect to an isoelectronic uniform electron gas reference E^{UEG} , was derived by application of the Hellmann–Feynman theorem and the chain rule as

$$E^{\text{mol}} - E^{\text{UEG}} = \sum_I Z_I \underbrace{\left(- \int d\mathbf{r} \frac{\int_0^1 d\lambda \rho(\lambda, \mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} + \frac{1}{2} \sum_{J \neq I} \frac{Z_J}{|\mathbf{R}_J - \mathbf{R}_I|} \right)}_{:= \Delta E_I}, \quad (3)$$

where Z_I and $\mathbf{R}_I$ indicate the nuclear charge and position, respectively. λ is a coupling variable linearly connecting the Hamiltonian of the uniform electron gas at $\lambda = 0$ with the Hamiltonian of the molecule at $\lambda = 1$ and $\rho(\lambda, \mathbf{r})$ is the electron density as a function of spatial coordinate $\mathbf{r}$ and λ . The equation was originally derived for the external potential in both the reference and the target compound due to a set of point charges. If the external potential of the reference is set to zero, Eq. (3) is obtained. It was proposed to use this model also for a UEG as a reference where the positive background charge density, and thus, the external potential approaches zero.

This is not the only possibility to define atomic energies. Another approach is, for instance, the interacting quantum atom (IQA),^{32,33} which is based on the quantum atom in molecule introduced by Bader.³⁴ This scheme relies on the decomposition of the electron density into well-defined fragments based on the topology of the density. Furthermore, it was recently proposed to calculate atomic energies by partitioning the energy among localized molecular orbitals, combined with an atomic charge partitioning scheme (in the following abbreviated as MO/AC).³⁵

Due to their well-defined connection to the underlying physics, such decomposition schemes can be useful to gain insights into the nature of chemical bonding. The interacting quantum atom approach has, for example, been employed to examine binding energy trends of first row diatomics,³⁶ the role of electrostatics in torsional energy potentials,³⁷ or the development of machine learning force fields.^{38,39} MO/AC was used to classify different types of electronic excitations⁴⁰ or the development of machine learning models predicting atomization energies.⁴¹

Recently, the alchemical partitioning scheme in Eq. (3) was also utilized to construct a simple expression for binding energies between *p*-block elements with only the nuclear charges of the binding partners as variables.⁴² For such practical applications of Eq. (3) with the UEG as a reference, the λ -dependent electron density $\rho(\lambda)$ must be calculated. In principle, this can be done with density functional theory and periodic boundary conditions (pbc), where the pbc are used to model the infinitely extended uniform potential of the UEG. However, in Eq. (3), it is assumed that the external potential of the reference is zero, which is only valid in the limit of the background charge density of the UEG approaching zero. This would require an infinitely large supercell, which cannot be realized in practice. If the electron density $\rho(\lambda)$ obtained for a finite supercell is used in Eq. (3) instead, the sum of atomic energies $\sum_I \Delta E_I$ does

not add up to $E^{\text{mol}} - E^{\text{UEG}}$ because the external potential of the UEG is non-zero for a finite cell size.

Thus, we introduce and benchmark here an alternative approach to Eq. (3) where the uniform background potential of the UEG and periodic boundary conditions are included. The modified alchemical decomposition scheme is then applied to a diverse set of 1325 organic isoelectronic molecules with six to eight heavy atoms, extracted from the QM9 dataset. We analyze the relationship between the local environment and alchemical atomic energies and develop a model to estimate the atomization energies of the respective molecules. The alchemical decomposition scheme is also compared to the IQA and MO/AC methods by calculating the respective atomic energies for the same dataset. By analyzing the different relations between local structure and atomic energies in these methods, we evaluate which kind of problems, e.g., estimation of energy or fingerprint for molecules in machine learning and which kind of systems, e.g., molecules or materials, might be treated best by which decomposition scheme.

II. METHODS

A. Atomic energies with periodic boundary conditions

We calculated atomic energies with Eq. (3) for 12 molecules of structure $\text{CH}_3\text{-YH}_y$ with $\text{Y} = \text{C, N, O, F, Si, P, S, Cl, Ge, As, Se, Br}$, and y such that all valencies of Y are saturated and the molecules HCN , ethene, and ethyne (15 molecules in total).

This was done by first calculating the electron density for selected λ -values along the transformation path from the UEG to the respective compound. The calculations were carried out with DFT and periodic boundary conditions by placing the compounds in a cubic supercell with a length of 15–18 Å. The electron densities for different λ -values were then used to numerically evaluate the integrals over λ and $\mathbf{r}$ in Eq. (3) with the integral over $\mathbf{r}$ carried out over a single supercell (see Sec. V).

According to Eq. (3), the sum of atomic energies for any compound should add up to the energy difference between the energy of the compound and the uniform electron gas. However, for the 15 molecules, the sum of atomic energies was on average about 8 Ha lower than the exact energy difference as illustrated in Fig. 1(a).

The reason for this deviation is that the Hamiltonian of the uniform electron gas contains a positive, uniform background potential. This potential is modeled via periodic boundary conditions during the evaluation of the electron density $\rho(\lambda)$ with density functional theory (DFT). However, the background potential and the periodic boundary conditions are neglected in Eq. (3). Furthermore, the external potential is not Coulombic [as assumed in Eq. (3)] if pseudopotentials are used. In the following, we derive an alternative expression to Eq. (3) which includes the background potential and periodic boundary conditions.

If the Hamiltonians of the uniform electron gas H^{UEG} and an isoelectronic compound H^{mol} are linearly coupled as

$$H(\lambda) = H^{\text{mol}}\lambda + (1 - \lambda)H^{\text{UEG}}, \quad (4)$$

with λ being a dimensionless variable defined over the interval $[0, 1]$, the periodic external potential is given by

$$V_{\text{ext}}^{\text{pbc}}(\mathbf{r}, \lambda) = \sum_{\mathbf{T}} \sum_I V_I^{\text{PP}}(\lambda, \mathbf{r} - \mathbf{R}_I - \mathbf{T}) + \rho^{\text{bg}}(\lambda) V^{\text{bg}}(\mathbf{r} - \mathbf{T}), \quad (5)$$

where $\sum_{\mathbf{T}}$ is a sum over the lattice vectors of the periodic system, V_I^{PP} is the pseudopotential of the I -th nucleus, ρ^{bg} is the uniform background charge, and V^{bg} is the associated potential. The background charge must be introduced to obtain a neutral system for $\lambda < 1$ since the energy would diverge otherwise. In this work, we use pseudopotentials from Goedecker, Teter, and Hutter,^{43,44} which can be expressed analytically as

$$V_I^{\text{PP}}(\lambda, r) = V_{I,\text{loc}}(\lambda, r) + V_{I,\text{nl}}(\lambda, \mathbf{r}, \mathbf{r}'), \quad (6)$$

with

$$V_{I,\text{loc}}(\lambda, r) = -\frac{Z_V(\lambda)}{r} \text{erf}(\alpha^{\text{PP}} r) + \sum_i C_i^{\text{PP}}(\lambda) \left(\sqrt{2} \alpha^{\text{PP}} r \right) \times \exp \left\{ -(\alpha^{\text{PP}} r)^2 \right\} \quad (7)$$

and

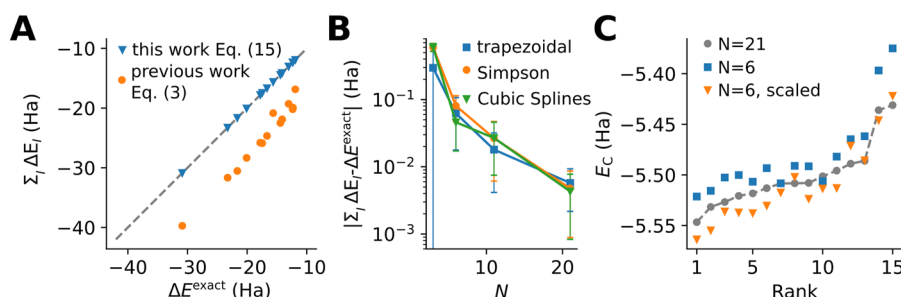


FIG. 1. (a) The sum of atomic energies for 15 different molecules vs the exact molecular energy with respect to the uniform electron gas ΔE^{exact} computed including periodic boundary conditions (pbc) and background potential via Eq. (15) or without pbc via Eq. (3). (b) MAE of $\sum_i \Delta E_i$ vs ΔE for the 15 molecules for an increasing number of integration points along the λ -path using different integration methods. (c) Comparison of atomic energy rankings for carbon atoms in 15 different molecular environments with varying number of integration points N and rescaling [Eq. (17)].

$$V_{I,\text{nl}}(\lambda, \mathbf{r}, \mathbf{r}') = \sum_{lm} \sum_{ij} \langle \mathbf{r} | p_i^{lm} \rangle h_{ij}^l(\lambda) \langle p_j^{lm} | \mathbf{r}' \rangle. \quad (8)$$

The λ -dependence of the parameters follows from the linear coupling of the Hamiltonians in Eq. (4) as $Z_V(\lambda) = \tilde{Z}_V \lambda$, $C_i^{\text{PP}}(\lambda) = \tilde{C}_i^{\text{PP}} \lambda$, and $h_{ij}^l(\lambda) = \tilde{h}_{ij}^l \lambda$. $\tilde{Z}_V$, $\tilde{C}_i^{\text{PP}}$ and $\tilde{h}_{ij}^l$ denote the optimized parameters reported in Refs. 43 and 44 such that the pseudopotential is zero for $\lambda = 0$ and equal to the potentials in Refs. 43 and 44 at $\lambda = 1$. A detailed description of the other parameters can be found in Ref. 44.

The uniform background charge density is given by

$$\rho^{\text{bg}}(\lambda) = \frac{N_e - \sum_I Z_{I,V}(\lambda)}{\Omega_{\text{cell}}}, \quad (9)$$

where N_e is the number of valence electrons, $Z_{I,V}(\lambda)$ is the valence charge of nucleus I at a given λ -value, and Ω_{cell} is the volume of the supercell. The uniform background potential is defined as

$$V^{\text{bg}}(\mathbf{r}) = - \int_{\Omega_{\text{cell}}} d\mathbf{r}' \frac{1}{|\mathbf{r} - \mathbf{r}'|}, \quad (10)$$

where $\int_{\Omega_{\text{cell}}}$ indicates integration over a single supercell.

Furthermore, the nuclear repulsion energy E^{nuc} depends on the interaction between the nuclei E^{NN} , between nuclei and background potential E^{Nbg} , and the self-interaction of the background E^{bg} such that

$$E^{\text{nuc}} = E^{\text{NN}} + E^{\text{Nbg}} + E^{\text{bg}}, \quad (11)$$

with

$$E^{\text{NN}} = \frac{1}{2} \sum_{I,J'}' \sum_{\mathbf{T}} \frac{Z_{I,V}(\lambda) Z_{J,V}(\lambda)}{|\mathbf{R}_I - \mathbf{R}_{J'} - \mathbf{T}|}, \quad (12)$$

where the $'$ indicates the exclusion of the divergent terms for $\mathbf{R}_I = \mathbf{R}_J$ at $\mathbf{T} = 0$,

$$E^{\text{Nbg}} = \sum_{I,\mathbf{T}} -Z_{I,V}(\lambda) \rho^{\text{bg}}(\lambda) V^{\text{bg}}(\mathbf{R}_I - \mathbf{T}) \quad (13)$$

and

$$E^{\text{bg}} = \frac{1}{2} \rho^{\text{bg}}(\lambda) \sum_{\mathbf{T}} - \int_{\Omega_{\text{cell}}} d\mathbf{r} V^{\text{bg}}(\mathbf{r} - \mathbf{r}' - \mathbf{T}). \quad (14)$$

In analogy to Ref. 28 (see also the supplementary material, Sec. I), the atomic energy decomposition can be derived by application of the Hellmann–Feynman theorem⁴⁵ and chain rule as

$$E^{\text{mol}} - E^{\text{UEG}} = \sum_I \int_0^1 d\lambda \underbrace{\frac{\partial E^{\text{nuc}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} + \int d\mathbf{r} \Delta V_I \rho_\lambda(\mathbf{r})}_{:= \frac{\partial E_I}{\partial \lambda}}, \quad (15)$$

with $\Delta V_I = \sum_{\mathbf{T}} V_I^{\text{PP}}(\lambda = 1, \mathbf{r} - \mathbf{R}_I - \mathbf{T}) - \tilde{Z}_{I,V}/\Omega_{\text{cell}} V^{\text{bg}}(\mathbf{r} - \mathbf{T})$ and $\tilde{Z}_{I,V}$ being the nuclear valence charge of the I -th nucleus at $\lambda = 1$.

The largest contribution of the background interaction to the atomic energies can already be included by evaluating Eq. (15) only at $\mathbf{T} = 0$, that is, by including the interaction of the electron density with the background and of the background with itself within the same supercell. This reduces the deviation of $\sum_I \Delta E_I$ from ΔE^{exact} from 8 Ha [if Eq. (3) is used] to 1.3 Ha. The interaction with the periodic images can be considered by evaluating the sum over all lattice vectors $\mathbf{T}$.

Here, we achieve this by approximating $\partial_\lambda E_I$ by central finite differences as

$$\frac{\partial E_I}{\partial \lambda} \approx \frac{E[V_I(\lambda + \Delta\lambda)] - E[V_I(\lambda - \Delta\lambda)]}{2\Delta\lambda}. \quad (16)$$

$E[V_I(\lambda \pm \Delta\lambda)]$ is the energy of a system where the parameters Z_V , $\{C_i\}$ and $\{h_{ij}\}$, of all pseudopotentials except for the I -th one are scaled by λ and the same parameters for V_I are scaled by $\lambda \pm \Delta\lambda$ with $\Delta\lambda$ being a small value (in this work, $\Delta\lambda = 5 \times 10^{-5}$). The periodic boundary conditions and the background potential are then consistently considered in the evaluation of $E[V_I(\lambda \pm \Delta\lambda)]$ by the respective electronic structure code. This improves the agreement between the sum of atomic energies and the exact energy difference between compound and uniform electron gas, ΔE^{exact} , by three orders of magnitude (MAE = 4 mHa) compared to Eq. (3) (MAE = 8 Ha) for the compounds mentioned above [see Fig. 1(a)].

B. Technical aspects

The calculation of atomic energies with Eq. (15) is performed by numerical integration of $\partial_\lambda E_I$ along the λ -path. Inspection of the derivatives for different atoms $\partial_\lambda E_I$ as a function of λ shows that the derivatives are zero at $\lambda = 0$ and increase monotonously with increasing λ (Fig. S1) in line with previous reports.^{25,46} Thus, the derivative at $\partial_\lambda E_I(\lambda = 0)$ was not calculated explicitly but always set to 0. The error of $\sum_I \Delta E_I$ with respect to ΔE^{exact} for the test molecules mentioned above is shown in Fig. 1(b) for different numerical integration methods and increasing number of equally spaced integration points along the λ -path. The MAE over all molecules decreases with an increasing number of grid points. For 21 equally spaced integration points, the errors are 5.9, 4.5, and 4.4 mHa for integration with the trapezoidal rule, Simpson's rule, and cubic splines, respectively, which is still somewhat above chemical accuracy (1.6 mHa). Unfortunately, the use of an even denser integration grid is not feasible, particularly if the decomposition scheme is applied to a larger set of molecules because from our experience the SCF-calculations for $\lambda < 1$ often need more than 100 SCF cycles to converge.

While accurate absolute energies require a dense integration grid, qualitative trends in atomic energies can already be recovered for a coarse grid. This is illustrated in Fig. 1(c), which shows atomic energies for carbon in 15 different molecular environments sorted by increasing energy value as predicted with 6 or 21 equally spaced integration points along λ . The overall trend is already reproduced if atomic energies are calculated via integration with Simpson's rule over six points along the λ -path as evident from Kendall's τ coefficient of 0.77.⁴⁷

Furthermore, the atomic energies can be rescaled to yield the exact energy difference as

$$\Delta \tilde{E}_I^N = \Delta E^{\text{exact}} \frac{\Delta E_I^N}{\sum_j \Delta E_j^N}, \quad (17)$$

where ΔE^{exact} is the exact energy difference between energy of the whole molecule and the uniform electron gas, ΔE_I^N is the atomic energy obtained after integration over N points along λ , and $\sum_j \Delta E_j^N$ is the sum of all atomic energies in the molecule numerically integrated over N points. Rescaled atomic energies can be used if exact values are required, for example, for the training of machine learning models. For $N = 6$ and our test set, rescaling even improves the ordering of atomic energies (Kendall's $\tau = 0.85$) with respect to the dense integration grid with $N = 21$. Thus, we expect to recover atomic energy trends with varying local environment at least qualitatively at this grid size.

Furthermore, finite size effects can impact atomic energies. For example, the symmetrical hydrogen atoms in ethane have different interactions with periodic images because the molecular symmetry differs from the cell symmetry. As a consequence, their atomic energies differ by up to 6 mHa. It might be possible to account for this effect via finite size corrections as used in the calculation of point defect energies.⁴⁸ However, in this study, we simply average over the energies of symmetrical atoms because changes in atomic energy due to varying local environment tend to be larger than the error due to finite size effects. For example, in the case of the symmetrical hydrogen atoms in ethane, the error due to the finite size effects (6 mHa) is significantly smaller than the change in atomic energy due to the different local environment of hydrogen in the amino group in CH_3NH_2 (15 mHa).

The computation of derivatives via central finite differences [Eq. (16)] entails an additional 2 MN self-consistent field (SCF) calculations, where M is the number of atoms in the system and N is the number of integration points along the λ coordinate. Although these SCF calculations converge rapidly, they significantly increase the computational cost of this method. According to the Hellmann–Feynman theorem, a perturbative ansatz can be employed instead,^{49,50} where the energy $E[V(\lambda \pm \Delta\lambda), \rho_\lambda]$ is evaluated for the density ρ_λ at λ instead of the self-consistent density $\rho_{\lambda \pm \Delta\lambda}$ as

$$\frac{\partial E_I}{\partial \lambda} = \frac{E[V(\lambda + \Delta\lambda), \rho_\lambda] - E[V(\lambda - \Delta\lambda), \rho_\lambda]}{2\Delta\lambda}. \quad (18)$$

Such a perturbative approach has already been used previously for the calculation of alchemical derivatives with periodic boundary conditions.^{51,52} As shown in Sec. II in the supplementary material, Eq. (18) should recover the exact derivative $\partial_\lambda E_I$ independent of the chosen $\Delta\lambda$. In practice, the agreement between Eq. (18) and the finite difference approach Eq. (16) worsens with increasing values of $\Delta\lambda$. This might be due to the practical implementation of the energy expression with a uniform background potential in the used electronic structure code (CPMD⁵³). Nevertheless, perturbative atomic energies are nearly identical to those obtained via self-consistent calculations for a small value of $\Delta\lambda = 5 \times 10^{-5}$ with a mean absolute deviation of 0.03 mHa for the 91 atoms in our test set of

15 small molecules. We assessed the same error for 361 fragments in 55 molecules of the QM9 database. In this case, the deviation was 0.02 mHa. Thus, atomic energies for the molecules in QM9 were evaluated via Eq. (18).

Furthermore, the scaling of non-local parameters h_{ij} of the pseudopotentials does not have a significant impact on the atomic energies (see Fig. S2) as has been reported in a previous study.⁴⁹ Therefore, only the parameters of V_{loc} were scaled for the calculations of the QM9 molecules but rescaling of the h_{ij} could be included straightforwardly if necessary.

The comparison of absolute atomic energies from the alchemical energy decomposition scheme with other decompositions is not meaningful because the alchemical energies are given with respect to the uniform electron gas and also differ significantly from energies obtained from an all-electron calculation. Thus, we define atomic contributions to the atomization energy as

$$E_I^{\text{at}} = \Delta E_I + \frac{Z_{I,V}}{N_e} E^{\text{UEG}} - E_I^{\text{free}}, \quad (19)$$

with $\Delta E_I = \int_0^1 d\lambda \partial_\lambda E_I$ as defined in Eq. (15), the nuclear valence charge $Z_{I,V}$, the total number of valence electrons N_e , the energy of the uniform electron gas E^{UEG} , and the energy of the free pseudo atom (with nuclear charge $Z_{I,V}$) E_I^{free} . The second term removes the shift in energy due to the choice of the uniform electron gas as the reference. The sum of atomic atomization energies $\sum_I E_I^{\text{at}}$ is then equal to the atomization energy of the respective molecule.

III. RESULTS AND DISCUSSION

A. Alchemical atomic energies of organic molecules

The alchemical atomic energy decomposition scheme was applied to the subset of molecules in QM9 with 38 valence electrons (in total 1325 molecules). Lewis graphs of the corresponding molecular structures (coordinates directly taken from QM9⁵⁴) of the respective molecules are shown in the supplementary material (Figs. S5–S17). All alchemical atomic energies presented in this section correspond to atomization energies as defined in Eq. (19). They were obtained by first calculating atomic energies with respect to the uniform electron gas by numerical integration of Eq. (15) for each atom in a molecule. The required derivatives were evaluated with the perturbative approach Eq. (18). These atomic energies were further rescaled according to Eq. (17), and finally, atomic atomization energies were obtained from Eq. (19).

Figure 2 shows the atomic energy distributions for hydrogen (8644 atoms), carbon (6323 atoms), nitrogen (1661 atoms), and oxygen (1341 atoms). The dataset contains also nine fluorine atoms for which atomic energies are omitted in the figure.

The magnitude of the atomic energies for carbon, nitrogen, and oxygen is similar (mean values are -247 , -246 , and -244 kcal/mol, respectively) and the distributions span similar ranges (135 kcal/mol for carbon and nitrogen and 110 kcal/mol for oxygen). The average atomic energy of hydrogen is -62 kcal/mol and the respective distribution spans a range of 37 kcal/mol. The narrower range of atomic energies for hydrogen could be explained by a lower diversity of molecular environments. Hydrogen can have only three different binding partners (carbon, nitrogen, or oxygen). In contrast, in our

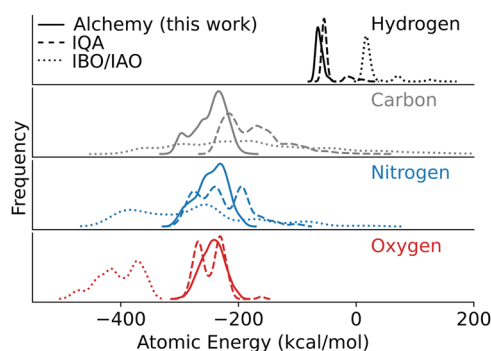


FIG. 2. The frequency of atomic contributions to the atomization energy in different local environments for 1325 drug-like molecules (~18 000 atoms) with six to eight heavy atoms split by element from different decomposition schemes.

dataset, oxygen has 6, nitrogen 14, and carbon 34 different sets of binding partners.

We explore the impact of the local environment on the atomic energy in more detail by grouping the atomic energies by binding partners. Exemplarily, selected atomic energy distributions for carbon with different binding partners are shown in Fig. 3(a). The atomic energies for the same set of binding partners span energy ranges of 50 kcal/mol or less indicating that the atomic energy is sensitive to the local environment. A similar relation can be observed for the elements H, N, and O (Fig. S3).

This can be exploited to estimate atomization energies by adding up the mean atomic energy for each local environment in a molecule as

$$E^{\text{mol}} \approx \sum_I \bar{E}(Z_I, P_I), \quad (20)$$

where I is a sum over all atoms in the molecule, Z_I is the nuclear charge of I , P_I describes the set of binding partners of I , and $\bar{E}(Z_I, P_I)$ is the atomic energy averaged over all atoms with nuclear charge Z_I and the binding partners P_I in our dataset. The atomization energies of the 1325 molecules can be estimated with an MAE of

29 ± 23 kcal/mol by Eq. (20). Furthermore, energies are overestimated for molecules with a low atomization energy and underestimated for molecules with a high atomization energy. The MAE can be reduced to 23 ± 16 kcal/mol by accounting for the systematic error by combining Eq. (20) with a linear fit as

$$E^{\text{mol}} \approx m \sum_I \bar{E}(Z_I, P_I) + b, \quad (21)$$

with slope m and intercept b chosen such that $\sum_I (m \sum_I \bar{E}(Z_I, P_I) + b - E_j^{\text{mol}})^2$ is minimal, where I runs over all molecules in the dataset.

As displayed in Fig. 3(b), Eq. (21) can predict the stability trend for the molecules in our dataset. While the overall error is one order above chemical accuracy, the model might still be useful as a baseline for machine learning models of the atomization energy.

To set the accuracy of our baseline into perspective, we compare it to other simple estimates of atomization energies. For example, bond-counting yields an MAE of 71 kcal/mol for the QM7-dataset.⁵⁵ While QM7 is a larger and likely more diverse dataset than the subset of 1325 QM9 molecules used in this study, this result indicates that the alchemical decomposition scheme can be competitive to conventional energy estimates. The dressed atom is another, statistical estimate of atomization energies.⁵⁶ In the dressed atom model, the atomization energy is approximated as

$$E^{\text{mol}} \approx \sum_{\text{elements}} N^{\text{el}} \times E^{\text{mean,el}}, \quad (22)$$

where an average energy $E^{\text{mean,el}}$ is defined for each element and N^{el} is the number of atoms of a given element type in the molecule. The atomic energies $E^{\text{mean,el}}$ are chosen such that the loss function $\sum_j (E_j^{\text{mol}} - \sum_{\text{elements}} N_j^{\text{el}} \times E^{\text{mean,el}})^2$ over all atomization energies E_j^{mol} in the training set is minimized. The prediction error of the dressed atom approach is 20 ± 14 kcal/mol, which is somewhat lower than for atomic energies from quantum alchemy. However, since alchemical atomic energies depend on the local environment, stability trends can in some cases also be reproduced for the same stoichiometry [e.g., $\text{C}_3\text{H}_5\text{N}_3\text{O}$, see Fig. 3(b)] while molecules

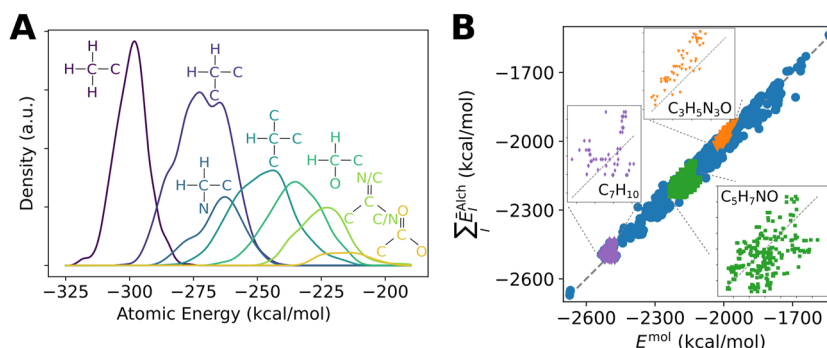


FIG. 3. Modeling the atomization energy from atomic contributions obtained via the alchemical energy decomposition. (a) Atomic energy distributions for carbon in different local environments. (b) Estimated atomization energies from the sum of average atomic energies for the local environments in a molecule. Insets show predictions for selected stoichiometries.

with the same stoichiometry have always the same energy with the dressed atom approach. In addition, the dressed atom approach is a purely statistical model while the energies $\bar{E}(Z_I, P_I)$ originate from a physics-based decomposition scheme. Thus, they are less sensitive to the choice of the training set and should be more transferable.

A limitation of Eq. (21) is that the atomization energy of a molecule can only be calculated if all of its atomic energies $\bar{E}(Z_I, P_I)$ are known. This restriction can be addressed by constructing a descriptor to calculate $\bar{E}(Z_I, P_I)$ for unknown local environments. Thus, we analyzed how changes in atomic energy are related to respective changes in the local environment.

B. The relation between atomic energy and nuclear electrostatic repulsion potential

Figure 3(a) shows that the atomic energy distributions are shifted to higher energies as the nuclear charges of the binding partners increase. For example, for a carbon atom bonded to three hydrogens and one carbon atom (C-H₃C), the mean atomic energy is -299 kcal/mol. By replacing one hydrogen with a carbon atom (C-H₂C₂), the atomic energy increases to -271 kcal/mol and exchanging another hydrogen for a carbon atom (C-HC₃) yields an atomic energy of -248 kcal/mol. Similar trends can be observed for the other elements as well (Fig. S3).

The structural change can be quantified by the average electrostatic potential V_I generated by the nuclei of the binding partners at the position $\mathbf{R}_I$ of atom I as

$$V_I = \frac{1}{N} \sum_{j=1}^N \frac{Z_{j,V}}{|\mathbf{R}_j - \mathbf{R}_I|}, \quad (23)$$

with N being the number of covalent binding partners and $Z_{j,V}$ and $\mathbf{R}_j$ being their respective nuclear valence charges and positions. By averaging over the potentials V_I for all atoms in the dataset with the same nuclear charge and the same binding partners, a potential

$\bar{V}(Z_I, P_I)$ can be defined analogous to $\bar{E}(Z_I, P_I)$. As illustrated in Fig. 4, $\bar{E}(Z_I, P_I)$ and $\bar{V}(Z_I, P_I)$ correlate for the elements H, C, N, and O [with $\bar{E}_I$ corresponding to $\bar{E}(Z_I, P_I)$ and $\bar{V}_I$ corresponding to $\bar{V}(Z_I, P_I)$]. Since V_I can easily be calculated for a given 3D structure, it could serve as a descriptor to obtain $\bar{E}(Z_I, P_I)$ for an unknown local environment. Furthermore, the correlation can be understood from Eq. (3). Equation (3) has an attractive and a repulsive electrostatic term that partially cancels motivating the description of the effects of the local environment by the effective potential given in Eq. (23). We note that Eq. (3) does not include periodic boundary conditions as discussed in Sec. II but it should be sufficient to explain the effect of the local environment where interactions with periodic images can be neglected. Furthermore, we choose the nuclear valence charge instead of the total nuclear charge because atomic energies are calculated using pseudopotentials. If the approach was extended to compounds containing elements like phosphorus or sulfur, with the same nuclear valence charge as nitrogen and oxygen, an effective nuclear charge⁵⁷ would be a more suitable choice to distinguish binding partners from the same group in the periodic table.

The close relationship between electrostatic potential and atomic energy emerging from the alchemical transformation has been noted before by Politzer and Murray and motivated the following expression for the molecular energy:

$$E^{\text{mol}} = a \sum_A Z_A V_{A,0} + b, \quad (24)$$

with the nuclear charges Z_A , the electrostatic potential at the position of Z_A due to the other nuclei and the electron density [see Eq. (2)], and a and b are fitted to reference molecular energies.¹⁴ Note that the fundamental importance of these relationships also motivated the definition of the Coulomb matrix and the bag of bonds representation of molecules and materials for training machine learning models of quantum properties, including energies, throughout chemical compound space.^{55,56,58}

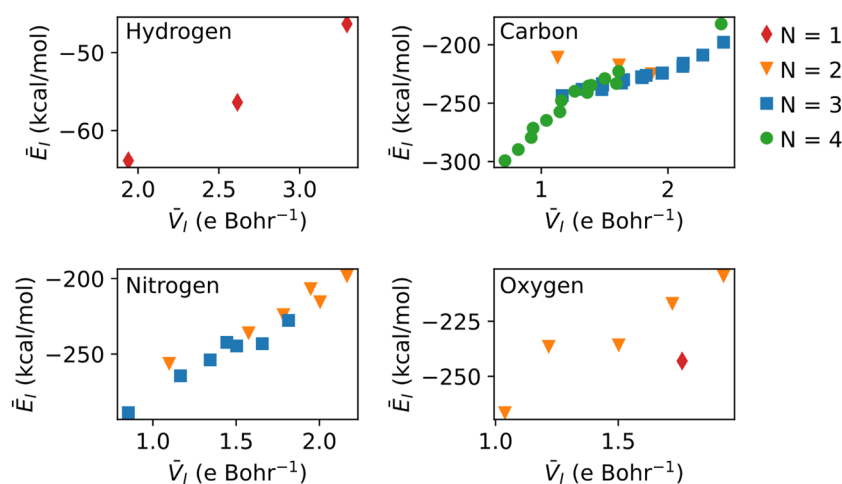


FIG. 4. The influence of the local environment on the respective mean atomic energy $\bar{E}_I$ [Eq. (20)] quantified via the relation between the average electrostatic potential of the binding partners $\bar{V}_I$ [Eq. (23)] at the position of atom I for H, C, N, and O and different number of binding partners N that restrict bond orders.

We also investigated the influence of the bond order on the atomic energy by inspecting the energy–potential relationships for different numbers of binding partners N for C, N, and O. For example, carbon can have a single and a triple bond or two double bonds ($N = 2$), two single bonds and one double bond ($N = 3$), or four single bonds ($N = 4$). The relationship between energy and potential is approximately linear for fixed N (Fig. 4). However, the slope changes with varying N indicating that the bond order impacts the sensitivity to changes in the local environment. This should be expected as different bond orders lead to different electron density profiles in the local environment. Due to the dependence of the atomic energy on the electron density according to Eq. (3), this should result in different atomic energy trends. Oxygen seems to behave similarly with the atomic energy for $N = 1$ (a carbonyl group) deviating from the linear trends observed for $N = 2$; however, in this case, there is not enough data for a definite conclusion. In contrast, the relation between $\bar{V}_I$ and E_I is linear for nitrogen with similar slopes for $N = 2$ and $N = 3$, showing that the influence of the bond order is not the same for all elements.

In summary, this analysis reveals a linear correlation between the average nuclear electrostatic potential of the local environment [Eq. (23)] and the respective atomic energy (in the case of carbon for a fixed number of binding partners). This correlation can be rationalized by the dependence of the atomic energy on the electrostatic interactions between nuclei and electrons and could be used to estimate atomic energies for local environments not contained in the training data.

C. Comparison to other decomposition methods

We also calculated atomic energies with the interacting quantum atom approach³³ and a MO/AC partitioning method. In the MO/AC decomposition, different types of molecular orbitals and atomic charges can be chosen. We select intrinsic bond orbitals⁵⁹ and charges from intrinsic atomic orbitals,⁵⁹ e.g., the IBO/IAO partitioning, because it suffers less from basis set artifacts compared to other MO/AC partitionings.^{35,41} The respective atomic energy distributions are displayed in Fig. 2. We note that atomic energies

of 224 molecules were excluded for the interacting quantum atom approach because they contained non-nuclear attractors, likely due to basis set artifacts,⁶⁰ which prevents the partitioning of the energy into atomic contributions.

As for the alchemical decomposition, atomic energies for hydrogen are higher than for the other elements. However, the atomic energies from IQA and IBO/IAO span wider energy ranges than energies from quantum alchemy. For example, the difference between the lowest and highest atomic energies of carbon is 138 kcal/mol for alchemy, 277 kcal/mol for IQA, and 863 kcal/mol for IBO/IAO. The broader energy range is also correlated with increased error in predicting atomization energy using Eq. (21) (MAE of 25 kcal/mol for IQA and 50 kcal/mol for IBO/IAO). Consequently, the alchemical decomposition method is the preferred choice for estimating energies solely based on average atomic energies with a specific set of binding partners. However, atomic energies derived from the IBO/IAO decomposition scheme might be more suitable as molecular fingerprints, e.g., for machine learning applications. This preference arises from their ability to provide finer-grained differentiation between distinct local environments since the respective atomic energy distributions are more separated from each other.

Furthermore, the atomic energy distributions displayed in Fig. 2 have different shapes indicating a different relation between the local environment and respective atomic energy. We explore these differences by plotting the mean atomic energies $\bar{E}(Z_I, P_I)$ in different local environments from either IQA or IBO/IAO against the respective energies from the alchemical decomposition (Fig. 5). This analysis reveals an overall similar dependence of the atomic energy on the local environment for the three decomposition methods. Carbon in the $\text{H}_3\text{C}-\text{C}$ fragment has always the lowest energy and the energy increases when a binding partner is replaced by a more electronegative one up to $\text{F}_3\text{C}-\text{C}$, which has always the highest atomic energy.

Nevertheless, there are also distinct differences in the relative change in atomic energy from one environment to another. Alchemical energies change strongly if a hydrogen atom is substituted

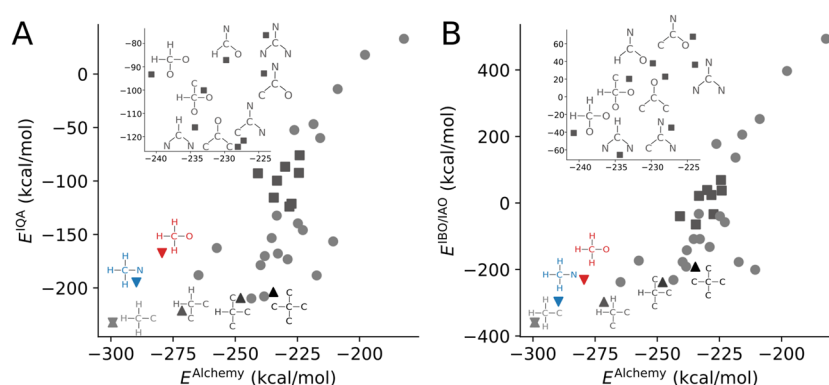


FIG. 5. Comparison for IQA (a) and IBO/IAO (b) partitioning schemes to alchemical partitioning in varying environments of the carbon atom. Similarities and differences are highlighted for selected series of environments. Up-pointing triangles correspond to a series of environments where hydrogens are substituted by carbons, down-pointing triangles show trends if a single heavy atom neighbor X in $\text{H}_3\text{C}-X$ is varied, cubes correspond to a set of environments generating similar nuclear electrostatic potentials at the position of the central carbon atom, and circles denote the remaining local environments in the dataset.

by the second row element, e.g., for the series $\text{H}_3\text{C}-\text{C}$, $\text{H}_2\text{C}-\text{C}_2$, $\text{HC}-\text{C}_3$, and $\text{C}-\text{C}_4$, because the electrostatic potential of the environment varies strongly. If the heavy atom is varied, e.g., for the fragments $\text{H}_3\text{C}-\text{C}$, $\text{H}_3\text{C}-\text{N}$, and $\text{H}_3\text{C}-\text{O}$ (see Fig. 5), the change in atomic energies with respect to the change in the local environment is smaller due to the smaller variation in the electrostatic potentials V_I [Eq. (23)]. IQA and IBO/IQA show the opposite behavior with a stronger change in atomic energy for the series $\text{H}_3\text{C}-\text{C}$, $\text{H}_3\text{C}-\text{N}$, and $\text{H}_3\text{C}-\text{O}$ compared to the series $\text{H}_3\text{C}-\text{C}$, $\text{H}_2\text{C}-\text{C}_2$, $\text{HC}-\text{C}_3$, and $\text{C}-\text{C}_4$.

Furthermore, the local environments in the insets in Fig. 5 have similar potentials V_I leading only to a small variation of the respective alchemical atomic energies, while these energies vary more for IQA and IBO/IAO. These differences in the dependence on the local environment are responsible for the non-linear relation between E^{Alchemy} and E^{IQA} respectively $E^{\text{IBO/IAO}}$. Similar correlations between the decomposition schemes were also found for the other elements H, N, and O (see Fig. S4).

The main difference between the alchemical and the other decomposition schemes in describing the relation between atomic energy and environment is the stronger dependence of the alchemical energies on the electrostatic potential of the local environment. Thus, the alchemical decomposition might be particularly well-suited to resolve energy variations for systems with substantial electrostatic fluctuations, such as atomic diffusion in solids. Furthermore, the atomic energies derived from the alchemical decomposition could potentially be efficiently approximated using only a simple electrostatic model, while the relation between atomic energy and local environment might be more complicated for the other decomposition schemes.

IV. CONCLUSION

An atomic energy partitioning based on an alchemical transformation was proposed already in 1974,¹² but practical applications were scarce due to arbitrary coupling and reference choices. An important step forward, necessary to obtain the results of this work, was the introduction of the uniform electron gas as a universal reference system, which can be transformed consistently into arbitrary isoelectronic compounds through linear coupling to the respective Hamiltonians.²⁸ However, in this case, periodic boundary conditions must be considered during atomic energy computation. Here, we addressed this problem by evaluating changes in energy due to a perturbation of the individual nuclear pseudopotentials consistent with periodic boundary conditions.

This enables the systematic and quantitative investigation of the relationship between local structure and atomic energy, yielding interesting trends hinting to a general transferability of the atomic energies. This refined decomposition approach has been employed to compute atomic energies for a subset of 1325 molecules drawn from the QM9 dataset. Analysis of the numerical results indicates a consistent and not surprising pattern: atoms sharing the same local environment exhibit similar atomic energies across distinct molecules. Consequently, the summation of averaged atomic energies corresponding to various environments within a molecule can approximate atomization energies. For the molecules studied in this work, we find an associated mean absolute error of 23 ± 16 kcal/mol. This level of accuracy aligns closely with

statistically determined atomic energies (utilizing the dressed atom approach),⁵⁶ while also potentially offering enhanced transferability. Naturally, this approach is remarkably efficient in terms of computational complexity, being ~ 80 times faster than the semi-empirical xTB method,^{61,62} indicating that it might be an attractive alternative for tasks such as rapid filtering of stabilities in vast chemical libraries encoding diverse chemistries. Such speed in energy estimation could accelerate research efforts that require swift assessments of molecular or material stability. Our averaging decomposition model might also prove useful as a meaningful base-line for machine learning models with respect to which one could predict atomization energies, also known as Δ -machine learning⁶³—in analogy to recently proposed applications of Hammett's equation.^{64,65}

Moreover, we have also demonstrated on a fundamental level that the connection between the atomic energy and local environment can be quantified using just the electrostatic potential produced by the nuclei of the binding partners at the atom's location. This enables the explanation of atomic energy trends across diverse local environments by considering just the respective nuclear charges and positions.

We also compared alchemical atomic energies and those obtained by two other decomposition methods: the interacting quantum atom (IQA)³² and a partitioning technique employing intrinsic bond orbitals and charge distribution via intrinsic atomic orbitals (IBO/IAO).³⁵ Across these three decomposition methods, a consistent relationship emerges between the atomic energy and local environment: The less electronegative the binding partners, the lower the atomic energy contribution. By contrast to IQA and IBO/IAO, the alchemical decomposition method displays a heightened sensitivity to substantial alterations in electrostatics within the local environment. Consequently, the alchemical decomposition approach appears well-suited for discerning subtle energy changes tied to significant shifts in electrostatics. This capability could be particularly valuable in the modeling of scenarios such as defect formation or diffusion in solids characterized by significant variations in electrostatics between lattice and interstitial sites. For example, the alchemical atomic energies offer the potential to assess the stability of metal or high entropy alloys, contributing to investigations in this domain based on prior studies using other decomposition schemes^{66–69} or could be used to explore compounds with elevated ionic conductivity, a crucial consideration in the advancement of solid-state electrolyte design.^{70,71}

V. COMPUTATIONAL DETAILS

Initial geometries for the 15 molecules used in the Sec. II were generated through the LERULI API.^{61,62,72,73} The geometries of CH_3Y (with $\text{Y} = \text{SiH}_3$, PH_2 , SH , GeH_3 , AsH_2 , and SeH) and HCN , ethene, and ethyne were further optimized with the CPMD code⁵³ making use of gnu parallel⁷⁴ with the PBE functional,⁷⁵ Goedecker Teter Hutter pseudopotentials,^{43,76} a cutoff of 200 Ryd, and a boxsize of $(N_{\text{VE}}/8 \times 15^3)^{1/3}$ Å with N_{VE} being the number of valence electrons in the molecule. For the optimized geometries, 21 single point density functional theory (DFT) calculations were carried out along the λ path in equally spaced intervals (step size = 0.05) from $\lambda = 0$ to $\lambda = 1$ by scaling the parameters Z_V , $\{C_i^{\text{pp}}\}$, and $\{h_{ij}\}$ in the pseudopotentials by the respective λ -value. Derivatives with respect to the variation in the external potential at a specific atomic site

were calculated according to Eq. (16) with $\Delta\lambda = 5 \times 10^{-5}$. For the perturbative approach, the SCF calculations were stopped after a single SCF iteration. For the 15 molecules mentioned in Sec. II and one compound from QM9, we verified that the variation in energy with respect to the consecutive iterations is small ($<10^{-7}$ Ha) until self-consistency is reached. For the calculation of atomic energies with Eq. (3), the valence electron density was generated with CPMD2CUBE.⁷⁷ The numerical integration with respect to λ was performed with the help of scipy⁷⁸ and numpy.⁷⁹

In total, there are 1329 molecules with 38 valence electrons in QM9. Four molecules (dsgdb9nsd_000226, dsgdb9nsd_000229, dsgdb9nsd_003330, and dsgdb9nsd_003492) were excluded due to conversion problems and atomic energies were calculated for the remaining 1325 molecules. Geometries were extracted from the QM9 dataset⁵⁴ and single points were calculated with the LDA functional, Goedecker Teter Hutter pseudopotentials, a cutoff of 200 Ryd, and a boxsize of 20 Å for $\lambda = 8/38, 15/38, 23/38, 30/38, 1$ with the parameters Z_V and $\{C_i^{PP}\}$ scaled by the respective λ -value. Derivatives with respect to the variation in the external potential at a specific atomic site were calculated perturbatively via Eq. (18). The numerical integration with respect to λ was performed with Simpson's rule using scipy and numpy. The atomic energies were rescaled using Eq. (17) and atomic atomization energies were calculated via Eq. (19). The free atom energies were calculated with the same parameters as for the molecules.

The atomic energies from the interacting quantum atom approach were calculated using the AIMAll code.⁸⁰ The required wavefunction files were generated with Gaussian⁸¹ from single point DFT calculations with the svwn-functional and the 6-31G(2df,p) basis set, utilizing the molecular geometries reported in the QM9 dataset. Free atom energies were calculated with the same settings with unrestricted DFT.

The atomic energies from the IBO/IAO method were calculated using the decondense code³⁵ with the ibo-4 option and pyscf.⁸² Required single-point DFT calculations were performed on the geometries reported in the QM9 dataset with the LDA functional and the 6-31G(2df,p) basis set. Free atom energies were calculated with the same settings using unrestricted DFT.

Average timings for the energy evaluation with Eq. (20) or a single point xTB^{61,62} calculation were generated by evaluating the energy for the compounds with the QM9 IDs dsgdb9nsd_000227, dsgdb9nsd_001768, dsgdb9nsd_002277, dsgdb9nsd_002839, and dsgdb9nsd_008977.

SUPPLEMENTARY MATERIAL

This section includes the detailed derivation of Eqs. (15) and (18), additional data on the properties of the alchemical derivatives $\partial_\lambda E_I$, the relation between local structure and alchemical atomic energies, comparison of atomic energies for the different decomposition schemes, and Lewis structures of the selected QM9 compounds.

ACKNOWLEDGMENTS

M.J.S. acknowledged discussions with Giorgio Domenichini, Danish Khan, Jan Weinreich, and Dominik Lemm. ChatGPT-3.5

and GPT4 were used for rephrasing parts of the paper. O.A.v.L. received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement No. 772834). This research was part of the University of Toronto's Acceleration Consortium, which received funding from the Canada First Research Excellence Fund (CFREF). O.A.v.L. received support as the Ed Clark Chair of Advanced Materials and as a Canada CIFAR AI Chair.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Michael J. Sahre: Conceptualization (equal); Data curation (lead); Methodology (lead); Software (lead); Writing – original draft (lead); Writing – review & editing (equal). **Guido Falk von Rudorff:** Conceptualization (equal); Methodology (supporting); Software (supporting); Writing – review & editing (equal). **Philipp Marquetand:** Conceptualization (equal); Supervision (lead); Writing – review & editing (equal). **O. Anatole von Lilienfeld:** Conceptualization (supporting); Supervision (supporting); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material. Additionally, the atomic energies of the QM9 compounds calculated with the different decomposition schemes (alchemy, IQA and IBO/IAO) and code to generate the input-file and the rescaled pseudopotential files for the DFT calculations with the CPMD program and to calculate alchemical atomic energies from the CPMD output are openly available in Zenodo at 10.5281/zenodo.10041083.⁸³ The latest version of the code can be found at https://github.com/michasahre/Alchemy_PP.

REFERENCES

- ¹S. Fias, F. Heidar-Zadeh, P. Geerlings, and P. W. Ayers, "Chemical transferability of functional groups follows from the nearsightedness of electronic matter," *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11633–11638 (2017).
- ²J. M. Herbert, "Fantasy versus reality in fragment-based quantum chemistry," *J. Chem. Phys.* **151**, 170901 (2019).
- ³J. Liu and X. He, "Fragment-based quantum mechanical approach to biomolecules, molecular clusters, molecular crystals and liquids," *Phys. Chem. Chem. Phys.* **22**, 12341–12367 (2020).
- ⁴L. P. Hammett, "Some relations between reaction rates and equilibrium constants," *Chem. Rev.* **17**, 125–136 (1935).
- ⁵L. P. Hammett, "The effect of structure upon the reactions of organic compounds. Benzene derivatives," *J. Am. Chem. Soc.* **59**, 96–103 (1937).
- ⁶O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, "Exploring chemical compound space with quantum-based machine learning," *Nat. Rev. Chem.* **4**, 347–358 (2020).
- ⁷F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, "Physics-inspired structural representations for molecules and materials," *Chem. Rev.* **121**, 9759–9815 (2021).

- ⁸B. Huang and O. A. von Lilienfeld, "Quantum machine learning using atom-in-molecule-based fragments selected on the fly," *Nat. Chem.* **12**, 945–951 (2020).
- ⁹J. Behler, "Four generations of high-dimensional neural network potentials," *Chem. Rev.* **121**, 10037–10072 (2021).
- ¹⁰O. T. Unke and M. Meuwly, "A reactive, scalable, and transferable model for molecular energies from a neural network approach based on local information," *J. Chem. Phys.* **148**, 241708 (2018).
- ¹¹K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, "Quantum-chemical insights from deep tensor neural networks," *Nat. Commun.* **8**, 13890 (2017).
- ¹²P. Politzer and R. G. Parr, "Some new energy formulas for atoms and molecules," *J. Chem. Phys.* **61**, 4258–4262 (1974).
- ¹³P. Politzer and J. S. Murray, "The fundamental nature and role of the electrostatic potential in atoms and molecules," *Theor. Chem. Acc.* **108**, 134 (2002).
- ¹⁴P. Politzer and J. S. Murray, "Electrostatic potentials at the nuclei of atoms and molecules," *Theor. Chem. Acc.* **140**, 7 (2021).
- ¹⁵E. B. Wilson, "Four-dimensional electron density function," *J. Chem. Phys.* **36**, 2232–2233 (1962).
- ¹⁶W. Kohn, "Two applications of the variational method to quantum mechanics," *Phys. Rev.* **71**, 635–637 (1947).
- ¹⁷L. L. Foldy, "A note on atomic binding energies," *Phys. Rev.* **83**, 397–399 (1951).
- ¹⁸P.-O. Löwdin, "Scaling problem, virial theorem, and connected relations in quantum mechanics," *J. Mol. Spectrosc.* **3**, 46–66 (1959).
- ¹⁹E. A. Hylleraas, "Über den grundterm der zweielektronenprobleme von, H⁺, He, Li⁺, Be⁺⁺ usw.," *Z. Phys.* **65**, 209–225 (1930).
- ²⁰N. Marzari, S. de Gironcoli, and S. Baroni, "Structure and phase stability of Ga_xIn_{1-x}P solid solutions from computational alchemy," *Phys. Rev. Lett.* **72**, 4001–4004 (1994).
- ²¹O. A. von Lilienfeld and M. E. Tuckerman, "Molecular grand-canonical ensemble density functional theory and exploration of chemical space," *J. Chem. Phys.* **125**, 154104 (2006).
- ²²R. Balawender, M. A. Welearegay, M. Lesiuk, F. De Proft, and P. Geerlings, "Exploring chemical space with the alchemical derivatives," *J. Chem. Theory Comput.* **9**, 5327–5340 (2013).
- ²³M. Muñoz and C. Cárdenas, "How predictive could alchemical derivatives be?," *Phys. Chem. Chem. Phys.* **19**, 16003–16012 (2017).
- ²⁴P. Geerlings, S. Fias, T. Stuyver, P. Ayers, R. Balawender, and F. D. Proft, "New insights and horizons from the linear response function in conceptual DFT," in *Density Functional Theory*, edited by D. Glossman-Mitnik (IntechOpen, Rijeka, 2018), Chap. 1.
- ²⁵R. Balawender, M. Lesiuk, F. De Proft, C. Van Alsenoy, and P. Geerlings, "Exploring chemical space with alchemical derivatives: Alchemical transformations of H through Ar and their ions as a proof of concept," *Phys. Chem. Chem. Phys.* **21**, 23865–23879 (2019).
- ²⁶M. Muñoz, A. Robles-Navarro, P. Fuentealba, and C. Cárdenas, "Predicting deprotonation sites using alchemical derivatives," *J. Phys. Chem. A* **124**, 3754–3760 (2020).
- ²⁷G. Domenichini, G. F. von Rudorff, and O. A. von Lilienfeld, "Effects of perturbation order and basis set on alchemical predictions," *J. Chem. Phys.* **153**, 144118 (2020).
- ²⁸G. F. von Rudorff and O. A. von Lilienfeld, "Atoms in molecules from alchemical perturbation density functional theory," *J. Phys. Chem. B* **123**, 10073–10082 (2019).
- ²⁹P. F. Loos and P. M. Gill, "The uniform electron gas," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **6**, 410 (2016).
- ³⁰P.-F. Loos, C. J. Ball, and P. M. W. Gill, "Uniform electron gases. II. The generalized local density approximation in one dimension," *J. Chem. Phys.* **140**, 18A524 (2014).
- ³¹C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, "Exact Coulomb cutoff technique for supercell calculations," *Phys. Rev. B* **73**, 205119 (2006).
- ³²J. M. Guevara-Vela, E. Francisco, T. Rocha-Rinza, and A. M. Pendás, "Interacting quantum atoms—A review," *Molecules* **25**, 4028 (2020).
- ³³E. Francisco, A. Martín Pendás, and M. A. Blanco, "A molecular energy decomposition scheme for atoms in molecules," *J. Chem. Theory Comput.* **2**, 90–102 (2006).
- ³⁴R. F. W. Bader, "A quantum theory of molecular structure and its applications," *Chem. Rev.* **91**, 893–928 (1991).
- ³⁵J. J. Eriksen, "Mean-field density matrix decompositions," *J. Chem. Phys.* **153**, 214109 (2020).
- ³⁶A. Martín Pendás, E. Francisco, and M. A. Blanco, "Binding energies of first row diatomics in the light of the interacting quantum atoms approach," *J. Phys. Chem. A* **110**, 12864–12869 (2006).
- ³⁷M. G. Darley and P. L. Popelier, "Role of short-range electrostatics in torsional potentials," *J. Phys. Chem. A* **112**, 12954–12965 (2008).
- ³⁸P. L. Popelier, "QCTFF: On the construction of a novel protein force field," *Int. J. Quantum Chem.* **115**, 1005–1011 (2015).
- ³⁹M. J. Burn and P. L. Popelier, "Creating Gaussian process regression models for molecular simulations using adaptive sampling," *J. Chem. Phys.* **153**, 054111 (2020).
- ⁴⁰J. J. Eriksen, "Electronic excitations through the prism of mean-field decomposition techniques," *J. Chem. Phys.* **156**, 061101 (2022).
- ⁴¹F. Ø. Kjeldal and J. J. Eriksen, "Decomposing chemical space: Applications to the machine learning of atomic energies," *J. Chem. Theory Comput.* **19**, 2029 (2023).
- ⁴²M. J. Sahre, G. F. von Rudorff, and O. A. von Lilienfeld, "Quantum alchemy based bonding trends and their link to Hammett's equation and Pauling's electronegativity model," *J. Am. Chem. Soc.* **145**, 5899–5908 (2023).
- ⁴³C. Hartwigsen, S. Goedecker, and J. Hutter, "Relativistic separable dual-space Gaussian pseudopotentials from H to Rn," *Phys. Rev. B* **58**, 3641–3662 (1998).
- ⁴⁴M. Krack, "Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals," *Theor. Chem. Acc.* **114**, 145–152 (2005).
- ⁴⁵R. P. Feynman, "Forces in molecules," *Phys. Rev.* **56**, 340–343 (1939).
- ⁴⁶R. Balawender, A. Holas, F. De Proft, C. Van Alsenoy, and P. Geerlings, "Alchemical derivatives of atoms: A walk through the periodic table," in *Many-Body Approaches at Different Scales: A Tribute to Norman H. March on the Occasion of His 90th Birthday*, edited by G. Angilella and C. Amovilli (Springer International Publishing, Cham, 2018), pp. 227–251.
- ⁴⁷M. G. Kendall, "A new measure of rank correlation," *Biometrika* **30**, 81–93 (1938).
- ⁴⁸Y. Kumagai and F. Oba, "Electrostatics-based finite-size corrections for first-principles point defect calculations," *Phys. Rev. B* **89**, 195205 (2014).
- ⁴⁹O. Anatole von Lilienfeld, "Accurate *ab initio* energy gradients in chemical compound space," *J. Chem. Phys.* **131**, 164102 (2009).
- ⁵⁰G. F. von Rudorff and O. A. von Lilienfeld, "Alchemical perturbation density functional theory," *Phys. Rev. Res.* **2**, 023220 (2020).
- ⁵¹O. A. V. Lilienfeld, R. D. Lins, and U. Rothlisberger, "Variational particle number approach for rational compound design," *Phys. Rev. Lett.* **95**, 153002 (2005).
- ⁵²K. Y. Chang and O. A. V. Lilienfeld, "Al_xGa_{1-x}As crystals with direct 2 eV band gaps from computational alchemy," *Phys. Rev. Mater.* **2**, 073802 (2018).
- ⁵³Copyright IBM Corp. 1990–2023 and Copyright 1994–2001 by Max Planck Institute, Stuttgart, CPMD 4.3, <https://github.com/CPMD-code/CPMD>; accessed on 13:42, 30 January 2023.
- ⁵⁴R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, "Quantum chemistry structures and properties of 134 kilo molecules," *Sci. Data* **1**, 140022 (2014).
- ⁵⁵M. Rupp, A. Tkatchenko, K. R. Müller, and O. A. von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," *Phys. Rev. Lett.* **108**, 058301 (2012).
- ⁵⁶K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K. R. Müller, and A. Tkatchenko, "Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space," *J. Phys. Chem. Lett.* **6**, 2326–2331 (2015).
- ⁵⁷J. C. Slater, "Atomic shielding constants," *Phys. Rev.* **36**, 57–64 (1930).
- ⁵⁸G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. Anatole von Lilienfeld, "Machine learning of

molecular electronic properties in chemical compound space,” *New J. Phys.* **15**, 095003 (2013).

⁵⁹G. Knizia, “Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts,” *J. Chem. Theory Comput.* **9**, 4834–4843 (2013).

⁶⁰P. Salvador, M. Duran, and I. Mayer, “One- and two-center energy components in the atoms in molecules theory,” *J. Chem. Phys.* **115**, 1153–1157 (2001).

⁶¹C. Bannwarth, S. Ehlert, and S. Grimme, “GFN_{2-x}TB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions,” *J. Chem. Theory Comput.* **15**, 1652–1671 (2019).

⁶²C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and S. Grimme, “Extended tight-binding quantum chemistry methods,” *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **11**, e1493 (2021).

⁶³R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, “Big data meets quantum chemistry approximations: The Δ -machine learning approach,” *J. Chem. Theory Comput.* **11**, 2087–2096 (2015).

⁶⁴M. Bragato, G. F. Von Rudorff, and O. A. Von Lilienfeld, “Data enhanced Hammett-equation: Reaction barriers in chemical space,” *Chem. Sci.* **11**, 11859–11868 (2020).

⁶⁵M. Bragato, G. F. von Rudorff, and O. A. von Lilienfeld, “Occam’s razor for AI: Coarse-graining Hammett inspired product ansatz,” *Chem. Phys.*, [arXiv:2305.07010](https://arxiv.org/abs/2305.07010) (2023).

⁶⁶Y. Dan and D. R. Trinkle, “First-principles core energies of isolated basal and prism screw dislocations in magnesium,” *Mater. Res. Lett.* **10**, 360–368 (2022).

⁶⁷Z. Xu, S. Tanaka, and M. Kohyama, “Atomic configurations and energies of Mg symmetric tilt grain boundaries: *Ab initio* local analysis,” *Modell. Simul. Mater. Sci. Eng.* **29**, 085010 (2021).

⁶⁸H. S. Oh, S. J. Kim, K. Odbadrakh, W. H. Ryu, K. N. Yoon, S. Mu, F. Körmann, Y. Ikeda, C. C. Tasan, D. Raabe, T. Egami, and E. S. Park, “Engineering atomic-level complexity in high-entropy and complex concentrated alloys,” *Nat. Commun.* **10**, 2090 (2019).

⁶⁹C. Li, J. Yin, K. Odbadrakh, B. C. Sales, S. J. Zinkle, G. M. Stocks, and B. D. Wirth, “First principle study of magnetism and vacancy energetics in a near equimolar NiFeMnCr high entropy alloy,” *J. Appl. Phys.* **125**, 155103 (2019).

⁷⁰Y. Zheng, Y. Yao, J. Ou, M. Li, D. Luo, H. Dou, Z. Li, K. Amine, A. Yu, and Z. Chen, “A review of composite solid-state electrolytes for lithium batteries: Fundamentals, key materials and advanced structures,” *Chem. Soc. Rev.* **49**, 8790–8839 (2020).

⁷¹S. Xia, X. Wu, Z. Zhang, Y. Cui, and W. Liu, “Practical challenges and future perspectives of all-solid-state lithium-metal batteries,” *Chem* **5**, 753–785 (2019).

⁷²D. Lemm, G. F. von Rudorff, and A. von Lilienfeld, “LERULL.com, online molecular property predictions in real time and for free,” 2021, www.leruli.com, accessed on 13:46, 30 January 2022.

⁷³G. Landrum, “RDKit,” 2010, accessed on 13:45, 30 January 2022.

⁷⁴O. Tange, “GNU parallel—The command-line power tool,” *USENIX Mag.* **36**, 42–47 (2011).

⁷⁵J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized gradient approximation made simple,” *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

⁷⁶S. Goedecker, M. Teter, and J. Hutter, “Separable dual-space Gaussian pseudopotentials,” *Phys. Rev. B* **54**, 1703–1710 (1996).

⁷⁷W. Silvestri, A. P. Seitsonen, M. Boer, and A. Kohlmeier, “CPMD2CUBE,” <https://github.com/CPMD-code/Addsions>; accessed on 13:39, 30 January 2022.

⁷⁸P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental algorithms for scientific computing in Python,” *Nat. Methods* **17**, 261–272 (2020).

⁷⁹C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, “Array programming with NumPy,” *Nature* **585**, 357–362 (2020).

⁸⁰T. A. Keith, “AIMAll (version 19.10.12),” aim.tkgristmill.com, TK Gristmill Software, Overland Park, KS, 2019.

⁸¹M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09 Revision D.01, 2009.

⁸²Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, “Recent developments in the PySCF program package,” *J. Chem. Phys.* **153**, 024109 (2020).

⁸³M. J. Sahr, G. F. von Rudorff, P. Marquetand, and O. A. von Lilienfeld (2023). “Transferability of atomic energies from alchemical decomposition—Additional data,” Zenodo. <https://doi.org/10.5281/zenodo.10041083>

APPENDIX A.1.2

Supporting Information

Supplementary Material: Transferability of atomic energies from alchemical decomposition

Michael J. Sahre,¹ Guido Falk von Rudorff,^{2,3} Philipp Marquetand,⁴ and O. Anatole von Lilienfeld^{5, 6, 7, 8, 9, 10, 11}

¹*Vienna Doctoral School in Chemistry (DoSChem) and Institute of Theoretical Chemistry and Faculty of Physics, University of Vienna, 1090 Vienna, Austria*

²*University Kassel, Department of Chemistry, Heinrich-Plett-Str.40, 34132 Kassel, Germany*

³*Center for Interdisciplinary Nanostructure Science and Technology (CINSaT), Heinrich-Plett-Straße 40, 34132 Kassel*

⁴*Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger Str. 17, 1090 Vienna, Austria*

⁵*Chemical Physics Theory Group, Department of Chemistry, University of Toronto, St. George Campus, Toronto, ON, Canada*

⁶*Department of Materials Science and Engineering, University of Toronto, St. George Campus, Toronto, ON, Canada*

⁷*Vector Institute for Artificial Intelligence, Toronto, ON, M5S 1M1, Canada*

⁸*ML Group, Technische Universität Berlin and Institute for the Foundations of Learning and Data, 10587 Berlin, Germany*

⁹*Berlin Institute for the Foundations of Learning and Data, 10587 Berlin, Germany*

¹⁰*Department of Physics, University of Toronto, St. George Campus, Toronto, ON, Canada*

¹¹*Acceleration Consortium, University of Toronto, Toronto, ON*

I. DERIVATION OF EQ.(15)

According to the Hellmann-Feynman theorem, the derivative of the energy with respect to λ , $\frac{\partial E}{\partial \lambda}$, is given by

$$\frac{\partial E}{\partial \lambda} = \frac{\partial}{\partial \lambda} \langle \Psi | H(\lambda) | \Psi \rangle = \langle \Psi | \frac{\partial H}{\partial \lambda} | \Psi \rangle. \quad (1)$$

For a compound isoelectronic to the uniform electron gas, only the external potential and the nuclear repulsion depend explicitly on λ so that

$$\frac{\partial H}{\partial \lambda} = \frac{\partial V_{\text{ext}}^{\text{pb}}}{\partial \lambda} + \frac{\partial E^{\text{nuc}}}{\partial \lambda}. \quad (2)$$

From the definition of $V_{\text{ext}}^{\text{pb}}$ (see Eqs. (5)-(10) in the main text) follows by application of the chain rule

$$\frac{\partial V_{\text{ext}}^{\text{pb}}}{\partial \lambda} = \sum_I \sum_{\mathbf{T}} \left(\sum_K \frac{\partial V_I^{\text{PP}}}{\partial V_{IK}} \frac{\partial V_{IK}}{\partial \lambda} \right) + \frac{\partial \rho^{\text{bg}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} V^{\text{bg}} \quad (3)$$

where V_{IK} is the K -th λ -dependent pseudopotential parameter of nucleus I . The pseudopotentials are linearly dependent on the parameters V_{IK} (see Eqs. (7,8) in the main text). Moreover, and as defined in the main text, the parameters V_{IK} are themselves linearly dependent on λ , e.g. $V_{IK}(\lambda) = \bar{V}_{IK} \lambda$ with $\bar{V}_{IK}$ being the optimized K -th pseudopotential parameter for element I . Hence, the directional derivative $\frac{\partial V_I}{\partial \lambda}$ is equal to the respective pseudopotential at $\lambda = 1$

$$\frac{\partial V_I}{\partial \lambda} = \sum_K \frac{\partial V_I^{\text{PP}}}{\partial V_{IK}} \frac{\partial V_{IK}}{\partial \lambda} = V_I(\lambda = 1). \quad (4)$$

Furthermore from Eq. (9) in the main text follows,

$$\frac{\partial \rho^{\text{bg}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} = -\frac{\bar{Z}_{I,V}}{\Omega_{\text{cell}}} \quad (5)$$

such that

$$\frac{\partial V_{\text{ext}}^{\text{pb}}}{\partial \lambda} = \sum_I \sum_{\mathbf{T}} V_I(\lambda = 1, \mathbf{r} - \mathbf{R}_I - \mathbf{T}) - \frac{\bar{Z}_{I,V}}{\Omega_{\text{cell}}} V^{\text{bg}} = \sum_I \Delta V_I. \quad (6)$$

The derivative of the nuclear repulsion energy (see Eqs. (11)-(14) in the main text) is given by

$$\frac{\partial E^{\text{nuc}}}{\partial \lambda} = \frac{\partial E^{\text{NN}}}{\partial \lambda} + \frac{\partial E^{\text{Nbg}}}{\partial \lambda} + \frac{\partial E^{\text{bg}}}{\partial \lambda} \quad (7)$$

with

$$\frac{\partial E^{\text{NN}}}{\partial \lambda} = \sum_I \frac{\partial E^{\text{NN}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} = \sum_I \sum_{J'} \sum_{\mathbf{T}}' \frac{\bar{Z}_{I,V} Z_{J,V}(\lambda)}{|\mathbf{R}_I - \mathbf{R}_J - \mathbf{T}|} \quad (8)$$

and

$$\frac{\partial E^{\text{Nbg}}}{\partial \lambda} = \sum_I \frac{\partial E^{\text{Nbg}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} = \sum_{I,\mathbf{T}} \bar{Z}_{I,V} \left(\frac{\sum_J Z_{J,V}(\lambda)}{\Omega_{\text{cell}}} - \rho^{\text{bg}} \right) V^{\text{bg}}(\mathbf{R}_I - \mathbf{T}) \quad (9)$$

and

$$\frac{\partial E^{\text{bg}}}{\partial \lambda} = \sum_I \frac{\partial E^{\text{bg}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} = \sum_{I,\mathbf{T}} \bar{Z}_{I,V} \frac{\rho^{\text{bg}}}{\Omega_{\text{cell}}} \int_{\Omega_{\text{cell}}} d\mathbf{r} V^{\text{bg}}(\mathbf{r} - \mathbf{T}). \quad (10)$$

Consequently, $\frac{\partial E^{\text{nuc}}}{\partial \lambda}$ can be decomposed into atomic contributions as

$$\frac{\partial E^{\text{nuc}}}{\partial \lambda} = \sum_I \frac{\partial E_I^{\text{nuc}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} \quad (11)$$

with

$$\begin{aligned} \frac{\partial E_I^{\text{nuc}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} &= \sum_{\mathbf{T}} \sum_{J'} \frac{\bar{Z}_{I,V} Z_{J,V}(\lambda)}{|\mathbf{R}_I - \mathbf{R}_J - \mathbf{T}|} \\ &\quad + \sum_{\mathbf{T}} \bar{Z}_{I,V} \left(\frac{\sum_J Z_{J,V}(\lambda)}{\Omega_{\text{cell}}} - \rho^{\text{bg}} \right) V^{\text{bg}}(\mathbf{R}_I - \mathbf{T}) \\ &\quad + \sum_{\mathbf{T}} \bar{Z}_{I,V} \frac{\rho^{\text{bg}}}{\Omega_{\text{cell}}} \int_{\Omega_{\text{cell}}} d\mathbf{r} V^{\text{bg}}(\mathbf{r} - \mathbf{T}). \end{aligned} \quad (12)$$

Hence, $\frac{\partial E}{\partial \lambda}$ can be expressed as

$$\begin{aligned} \frac{\partial E}{\partial \lambda} &= \sum_I \langle \Psi | \Delta V_I + \frac{\partial E_I^{\text{nuc}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} | \Psi \rangle \\ &= \sum_I \int d\mathbf{r} \Delta V_I \rho_\lambda(\mathbf{r}) + \frac{\partial E_I^{\text{nuc}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} \end{aligned} \quad (13)$$

with the definitions of ΔV_I and $\frac{\partial E_I^{\text{nuc}}}{\partial \lambda}$ in Eq. (6) and Eq. (12), respectively. The atomic energy decomposition of the difference between uniform electron gas and compound is finally

$$E^{\text{mol}} - E^{\text{UEG}} = \sum_I \int_0^1 d\lambda \left(\frac{\partial E_I^{\text{nuc}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} + \int d\mathbf{r} \Delta V_I \rho_\lambda(\mathbf{r}) \right). \quad (14)$$

II. PERTURBATIVE DERIVATIVES

$$\begin{aligned} \langle V_{\text{ext}}^\pm \rangle &= \sum_{\mathbf{T}} \int_{\Omega_{\text{cell}}} d\mathbf{r} \left(V_I^{\text{PP}}(\lambda \pm \Delta\lambda) + \sum_{J \neq I} V_J^{\text{PP}}(\lambda) + \frac{N_e - \sum_J Z_{J,V}(\lambda) \mp \Delta\lambda \bar{Z}_{I,V}}{\Omega_{\text{cell}}} V^{\text{bg}} \right) \rho_\lambda(\mathbf{r}) \\ &= \sum_{\mathbf{T}} \int_{\Omega_{\text{cell}}} d\mathbf{r} \left(V_I^{\text{PP}}(\lambda) \pm \Delta\lambda V_I^{\text{PP}}(\lambda = 1) + \sum_{J \neq I} V_J^{\text{PP}}(\lambda) + \frac{N_e - \sum_J Z_{J,V}(\lambda) \mp \Delta\lambda \bar{Z}_{I,V}}{\Omega_{\text{cell}}} V^{\text{bg}} \right) \rho_\lambda(\mathbf{r}) \end{aligned} \quad (15)$$

$$\left\langle \frac{V_{\text{ext}}^+ - V_{\text{ext}}^-}{2\Delta\lambda} \right\rangle = \sum_{\mathbf{T}} \int_{\Omega_{\text{cell}}} d\mathbf{r} \left(V_I^{\text{PP}}(\lambda) - \frac{\bar{Z}_{I,V}}{\Omega_{\text{cell}}} V^{\text{bg}} \right) \rho_\lambda(\mathbf{r}) = \int_{\Omega_{\text{cell}}} d\mathbf{r} \Delta V_I \rho_\lambda(\mathbf{r}) \quad (16)$$

$$\begin{aligned}
E_I^{\text{NN}\pm} &= \frac{1}{2} \sum_{\mathbf{T}}' \frac{(Z_{I,V}(\lambda) \pm \Delta \bar{Z}_{I,V})^2}{|\mathbf{R}_{IT}|} + \frac{1}{2} \sum_{J \neq I} \sum_{\mathbf{T}} \frac{(Z_{I,V}(\lambda) \pm \Delta \bar{Z}_{I,V}) Z_{J,V}(\lambda)}{|\mathbf{R}_{IJT}|} \\
&+ \frac{1}{2} \sum_{K \neq I} \sum_{\mathbf{T}} \frac{Z_K(\lambda) (Z_{I,V}(\lambda) \pm \Delta \bar{Z}_{I,V})}{|\mathbf{R}_{KIT}|} + \frac{1}{2} \sum_{K \neq I} \sum_{J' \neq I} \sum_{\mathbf{T}}' \frac{Z_K(\lambda) Z_{J,V}(\lambda)}{|\mathbf{R}_{KJT}|} \\
&= \frac{1}{2} \sum_{\mathbf{T}}' \frac{(Z_{I,V}(\lambda) \pm \Delta \bar{Z}_{I,V})^2}{|\mathbf{R}_{IT}|} + \sum_{J \neq I} \sum_{\mathbf{T}} \frac{(Z_{I,V}(\lambda) \pm \Delta \bar{Z}_{I,V}) Z_{J,V}(\lambda)}{|\mathbf{R}_{IJT}|} + \frac{1}{2} \sum_{K \neq I} \sum_{J' \neq I} \sum_{\mathbf{T}}' \frac{Z_K(\lambda) Z_{J,V}(\lambda)}{|\mathbf{R}_{KJT}|}
\end{aligned} \tag{17}$$

with $\mathbf{R}_{IJT} = \mathbf{R}_I - \mathbf{R}_J - \mathbf{R}_T$.

$$\frac{E_I^{\text{NN}+} - E_I^{\text{NN}-}}{2\Delta\lambda} = \bar{Z}_{I,V} \sum_{J'} \sum_{\mathbf{T}}' \frac{Z_{J,V}(\lambda)}{|\mathbf{R}_{IJT}|} = \frac{\partial E^{\text{NN}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} \tag{18}$$

$$\begin{aligned}
E_I^{\text{Nbg}\pm} &= \sum_{\mathbf{T}} -V^{\text{bg}} \left[(Z_{I,V}(\lambda) \pm \Delta \lambda \bar{Z}_{I,V}) \left(\frac{N_e - \sum_J Z_{J,V}(\lambda) \mp \Delta \lambda \bar{Z}_{I,V}}{\Omega_{\text{cell}}} \right) \right. \\
&\quad \left. + \sum_{J \neq I} Z_{J,V}(\lambda) \left(\frac{N_e - \sum_J Z_{J,V}(\lambda) \mp \Delta \lambda \bar{Z}_{I,V}}{\Omega_{\text{cell}}} \right) \right]
\end{aligned} \tag{19}$$

$$\frac{E_I^{\text{Nbg}+} - E_I^{\text{Nbg}-}}{2\Delta\lambda} = \sum_{\mathbf{T}} \bar{Z}_{I,V} \left(\frac{\sum_J Z_{J,V}(\lambda)}{\Omega_{\text{cell}}} - \rho^{\text{bg}} \right) V^{\text{bg}}(\mathbf{R}_I - \mathbf{T}) = \frac{\partial E^{\text{Nbg}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} \tag{20}$$

$$E^{\text{bg}\pm} = \frac{1}{2} \left(\frac{N_e - \sum_K Z_K(\lambda) \mp \Delta \lambda \bar{Z}_{I,V}}{\Omega_{\text{cell}}} \right)^2 \sum_{\mathbf{T}} - \int_{\Omega_{\text{cell}}} d\mathbf{r} V^{\text{bg}}(\mathbf{r} - \mathbf{T}) \tag{21}$$

$$\frac{E^{\text{bg}+} - E^{\text{bg}-}}{2\Delta\lambda} = \bar{Z}_{I,V} \frac{\rho^{\text{bg}}}{\Omega_{\text{cell}}} \sum_{\mathbf{T}} \int_{\Omega_{\text{cell}}} d\mathbf{r} V^{\text{bg}}(\mathbf{r} - \mathbf{T}) = \frac{\partial E^{\text{bg}}}{\partial Z_{I,V}} \frac{\partial Z_{I,V}}{\partial \lambda} \tag{22}$$

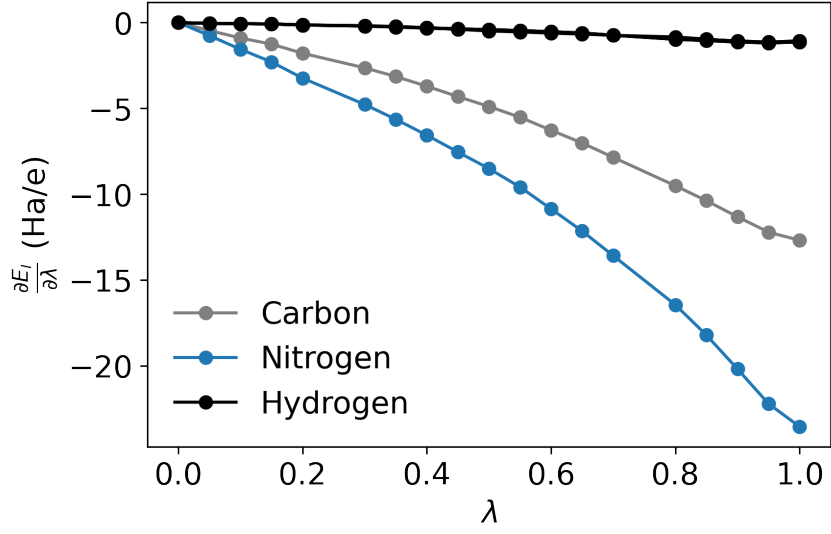


FIG. 1. The partial derivatives $\frac{\partial E_I}{\partial \lambda}$ of the energy with respect to a change in the pseudopotential of a specific atom as a function of λ for the atoms in methylamine.

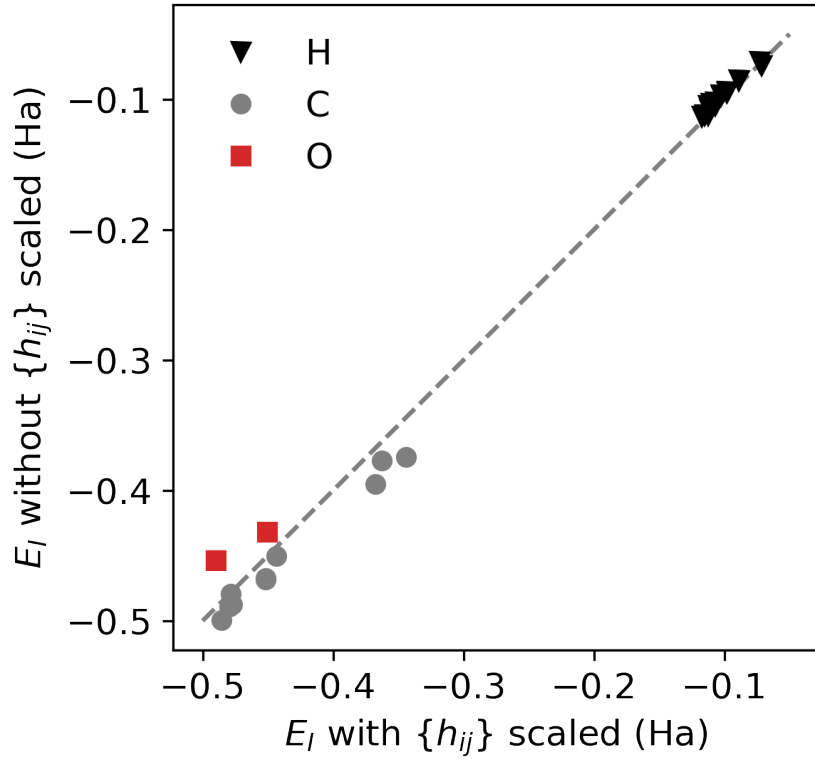


FIG. 2. Impact of the scaling of the non-local parameters of the pseudopotential $\{h_{ij}\}$ on the atomic energies exemplified for two molecules of QM9 (IDs dsgdb9nsd_000227 and dsgdb9nsd_000228).

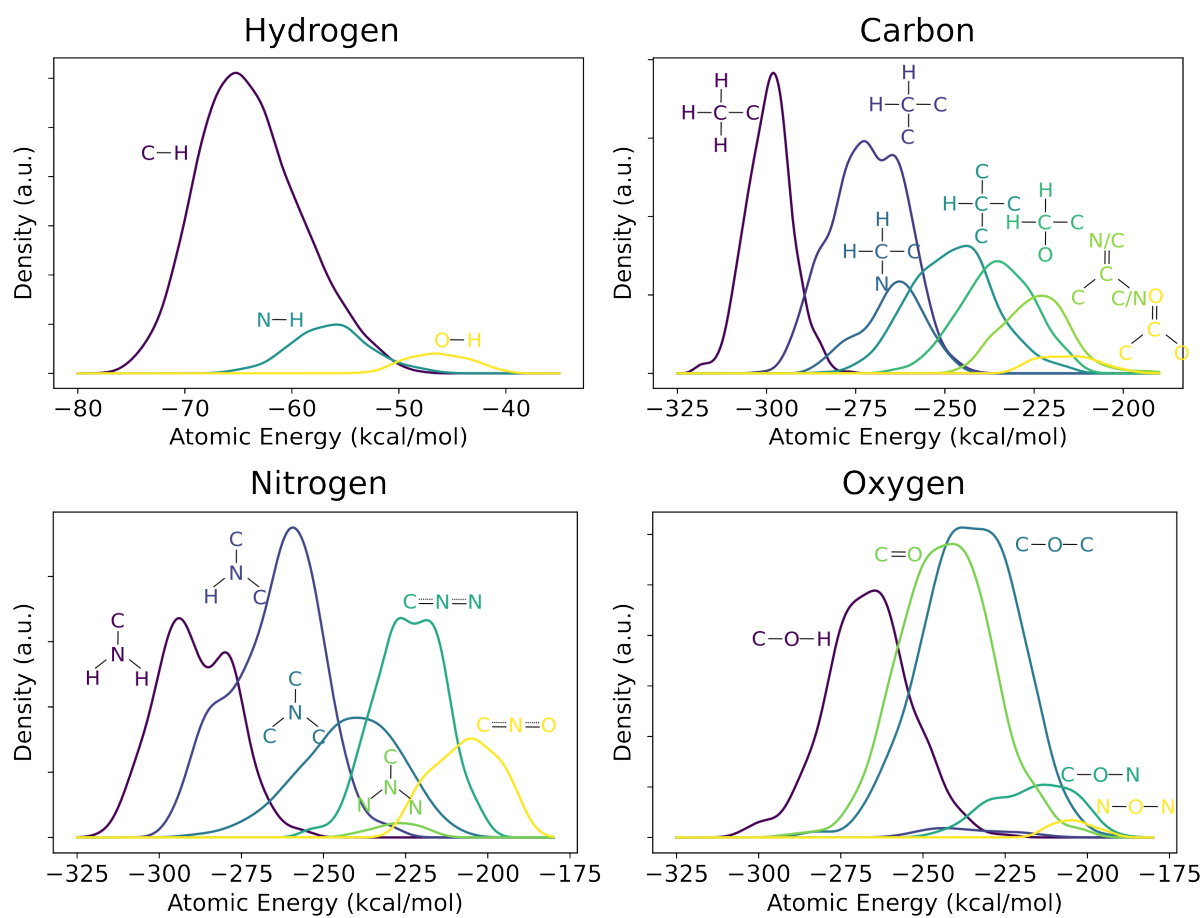


FIG. 3. Distribution of alchemical atomic energies for selected local environments for H, C, N, O.

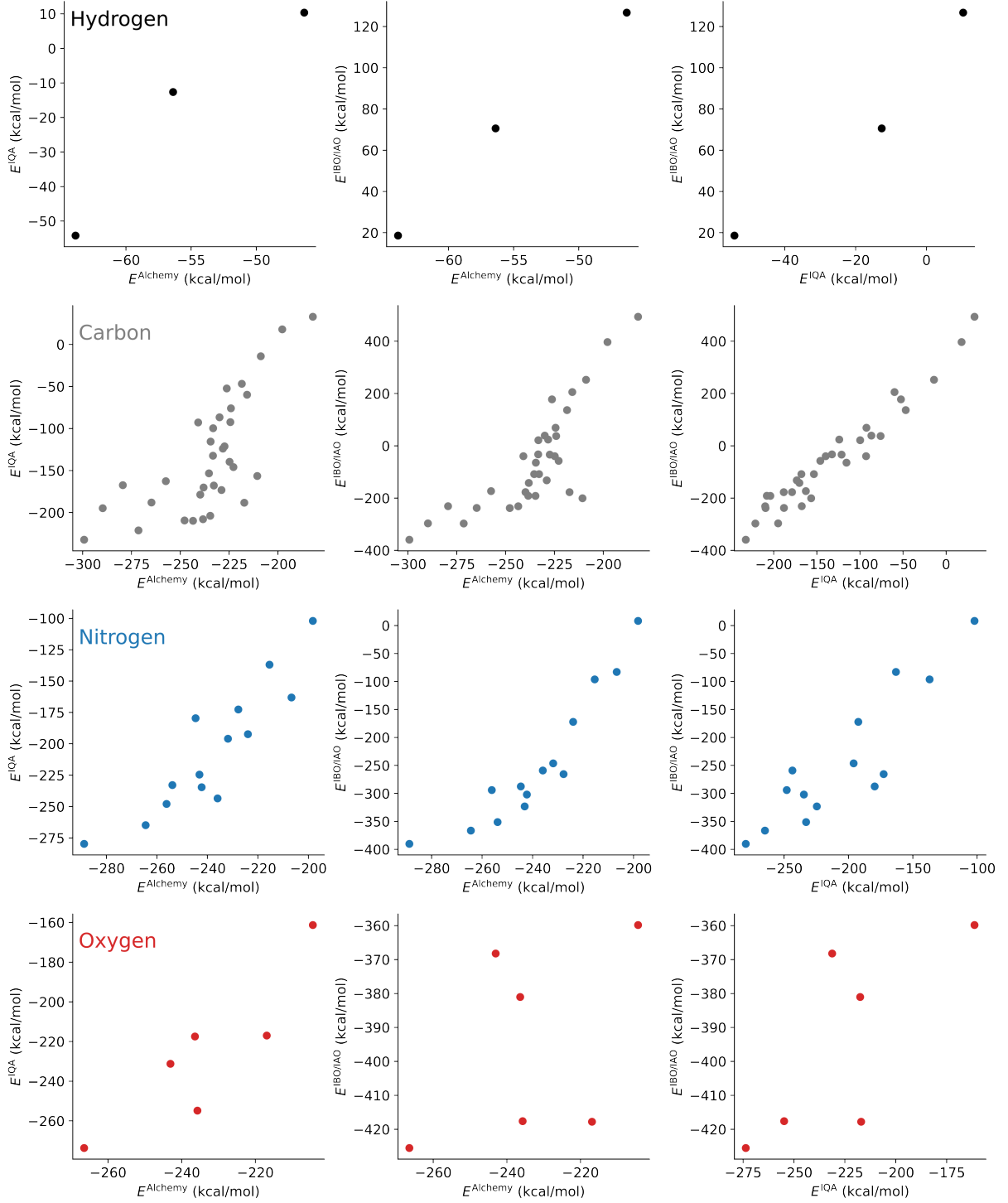


FIG. 4. The relation between atomic energy and the local environment compared for different decomposition schemes (from left to right alchemy vs IQA, alchemy vs IBO/IAO and IQA vs IBO/IAO) for the elements H, C, N and O.

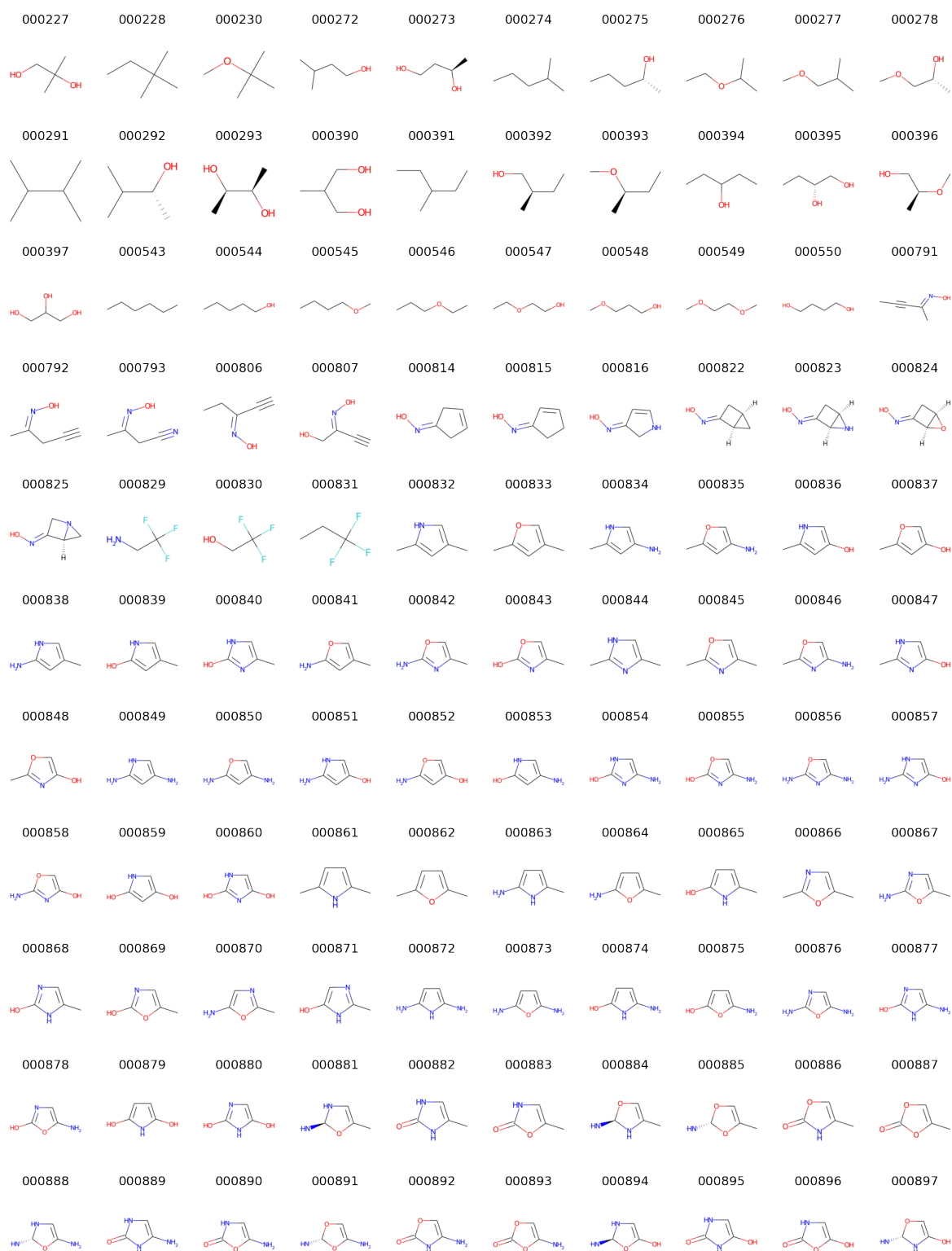


FIG. 5. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

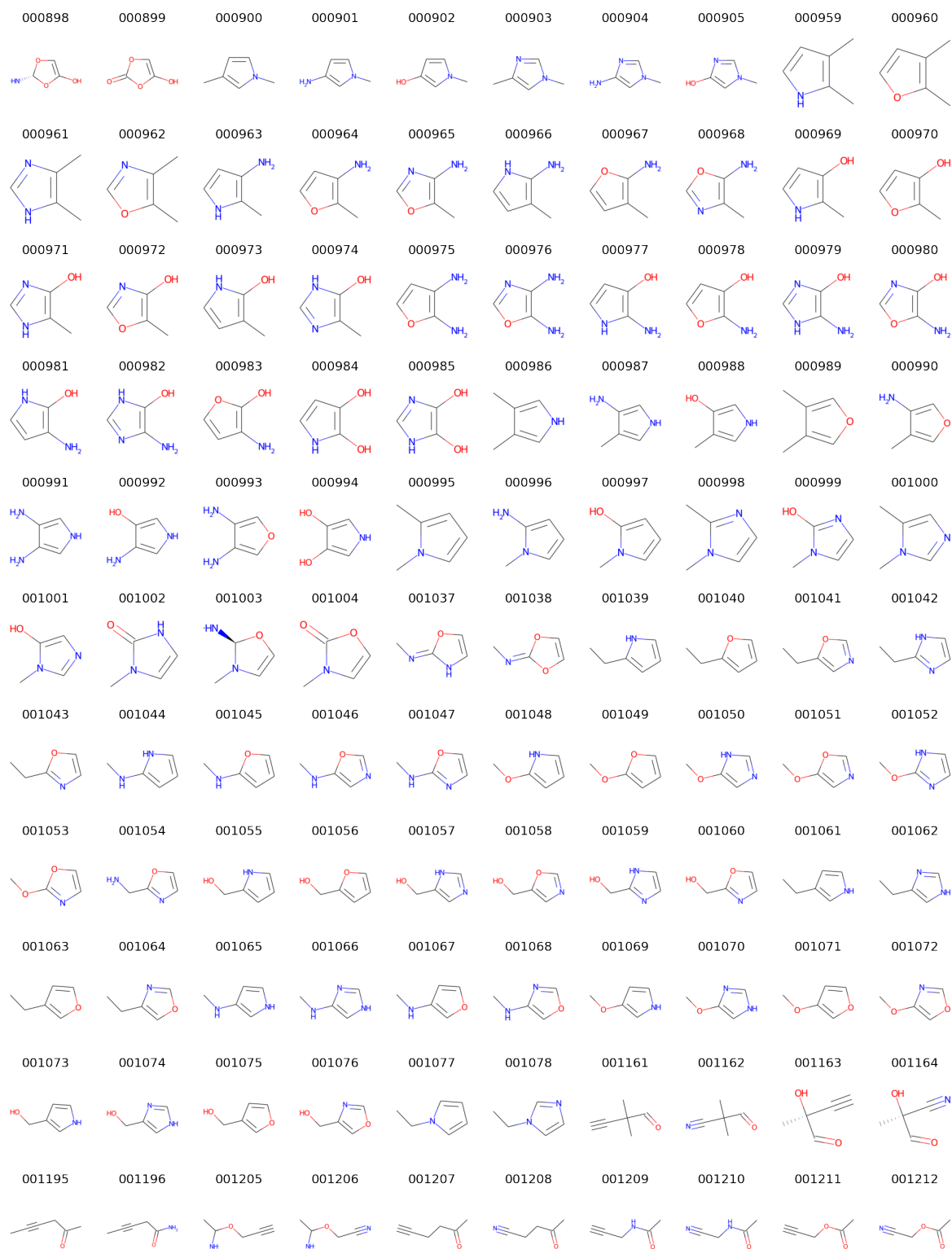


FIG. 6. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

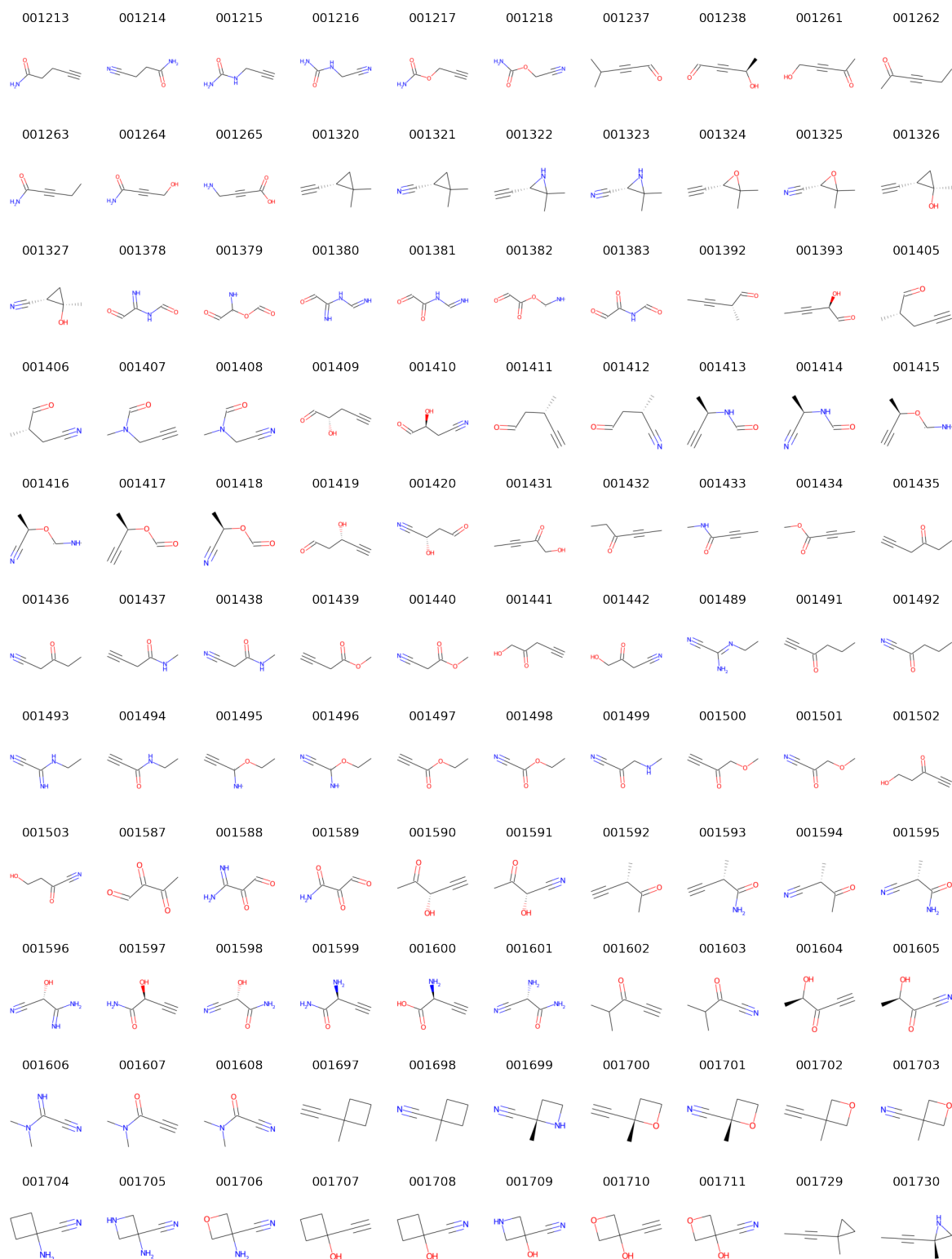


FIG. 7. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

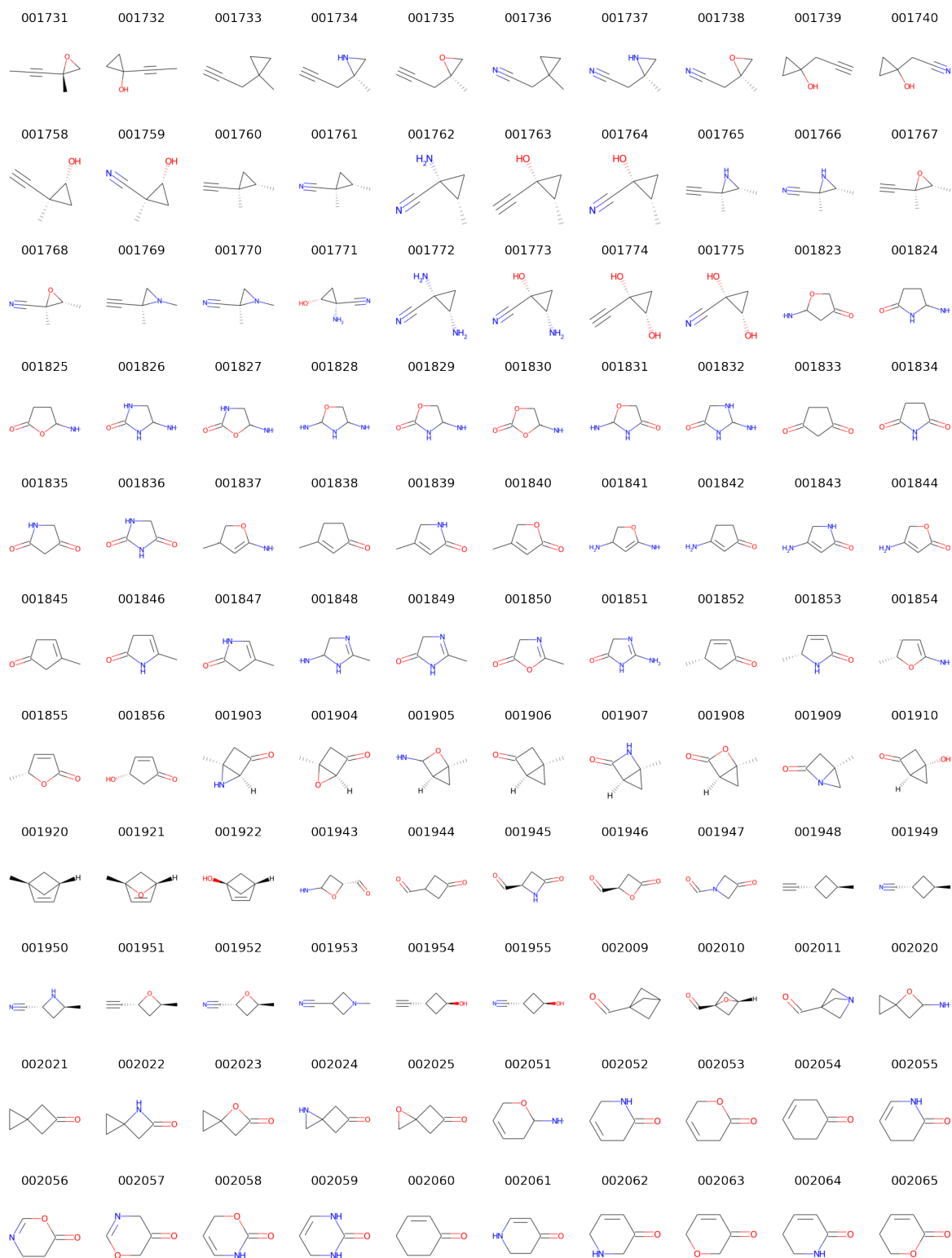


FIG. 8. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

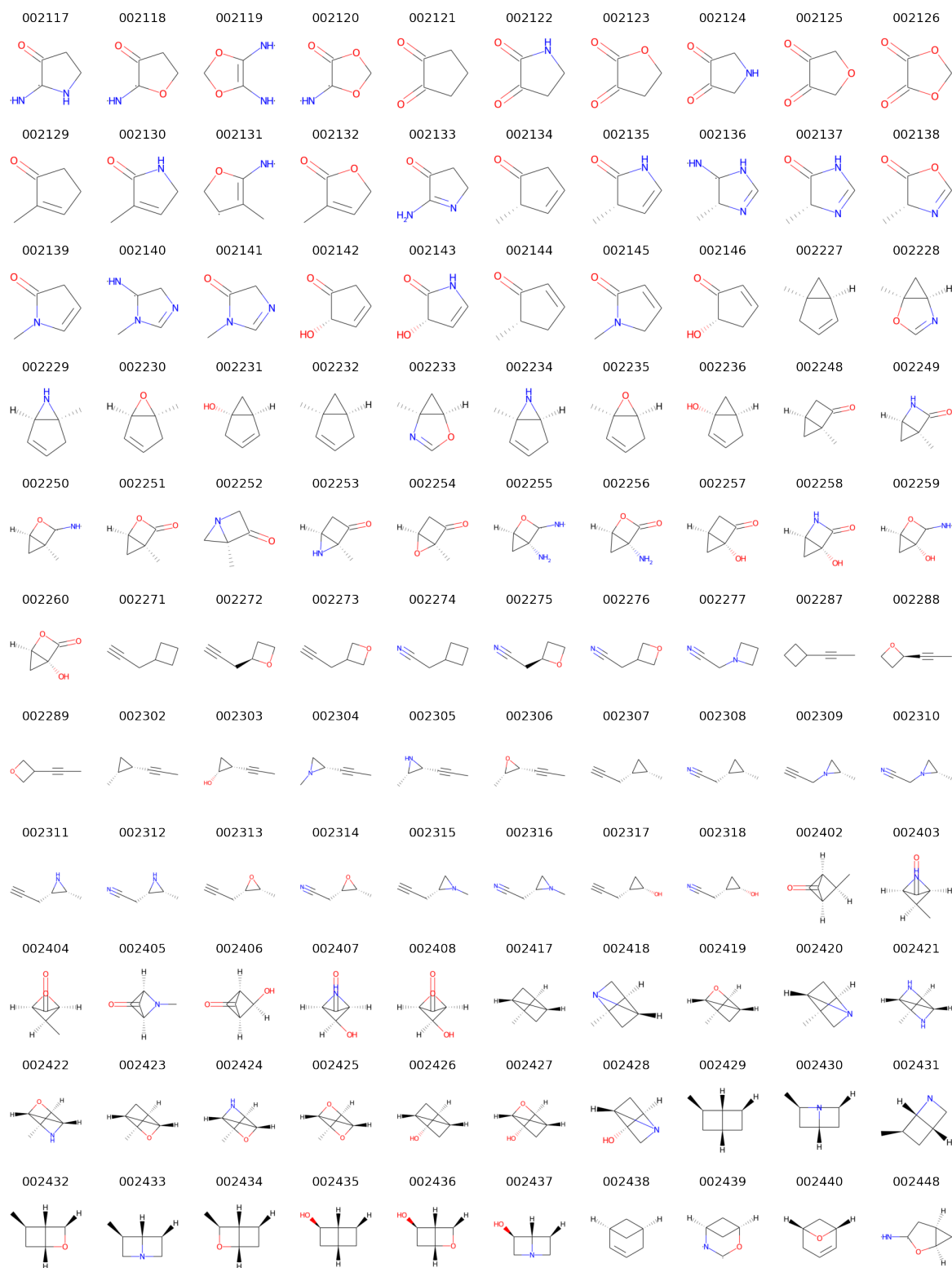


FIG. 9. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

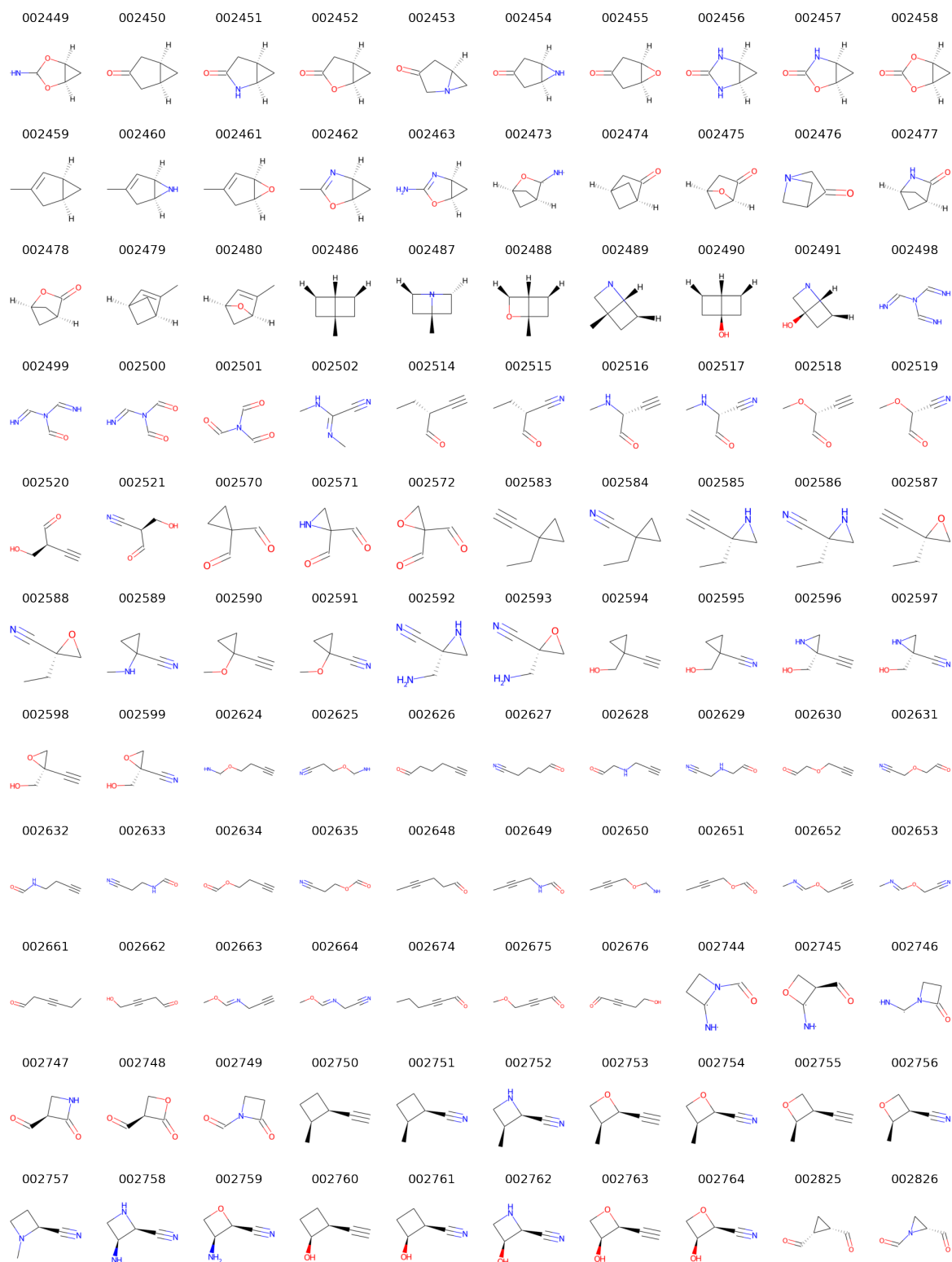


FIG. 10. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

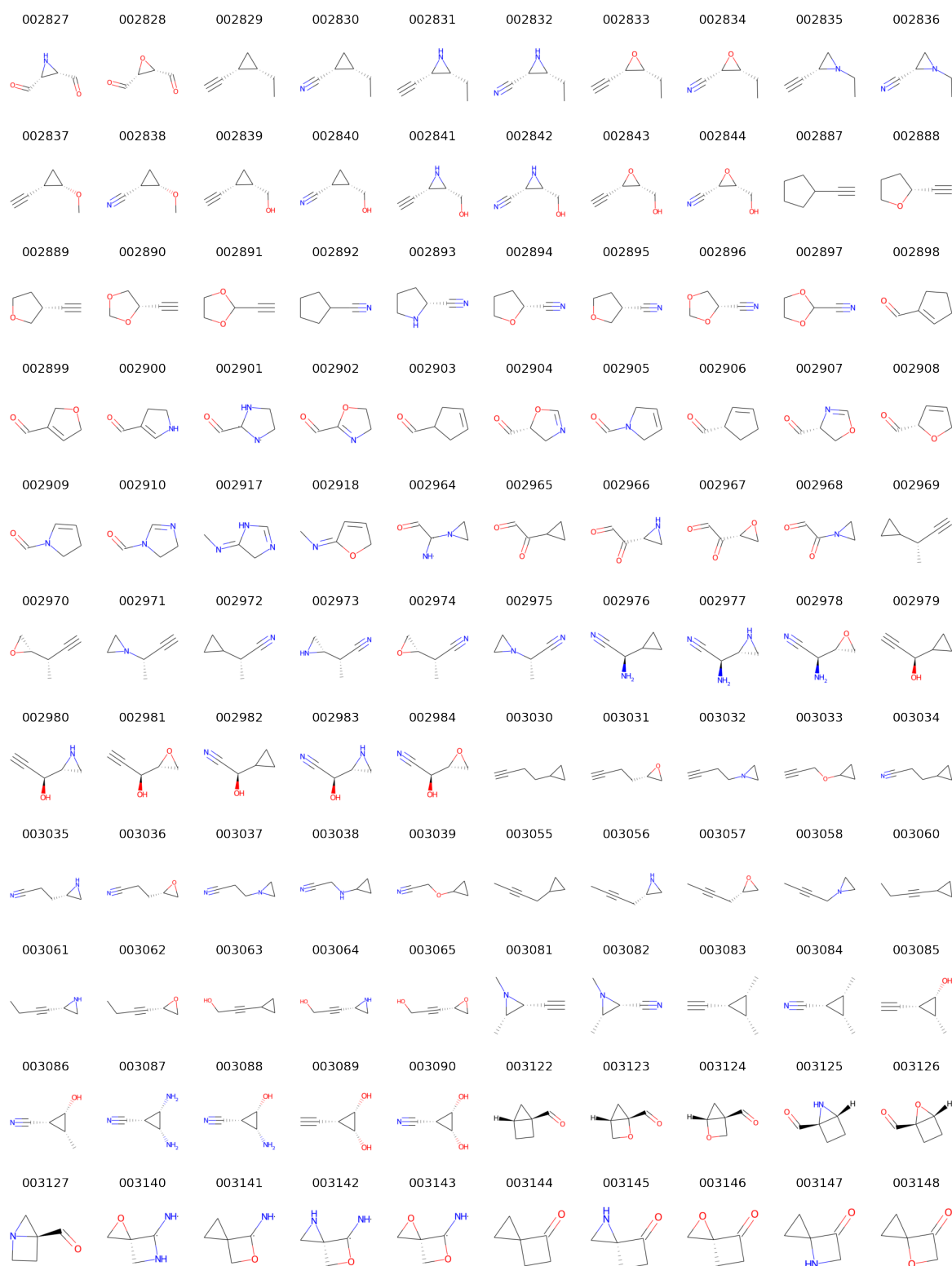


FIG. 11. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

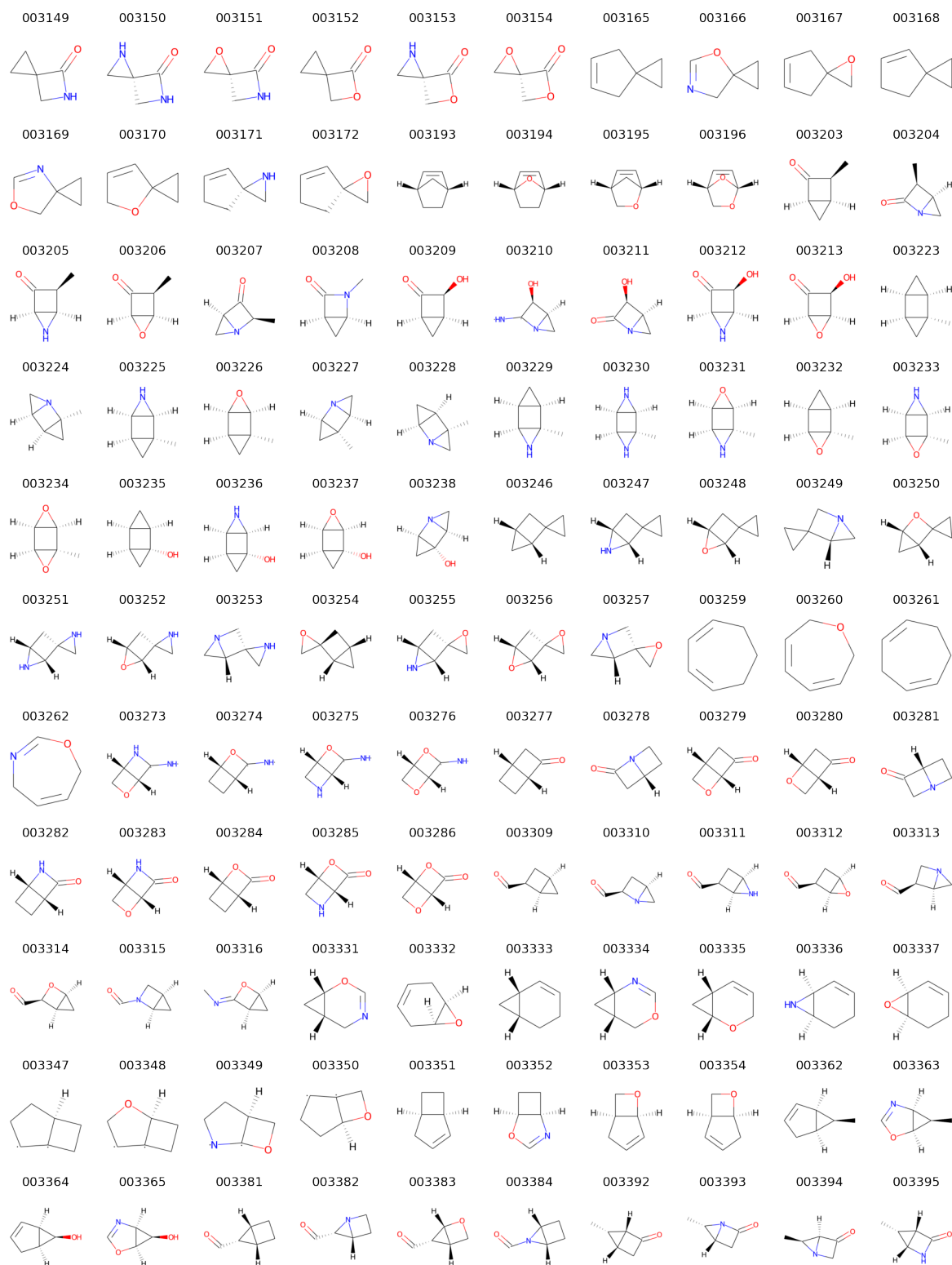


FIG. 12. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

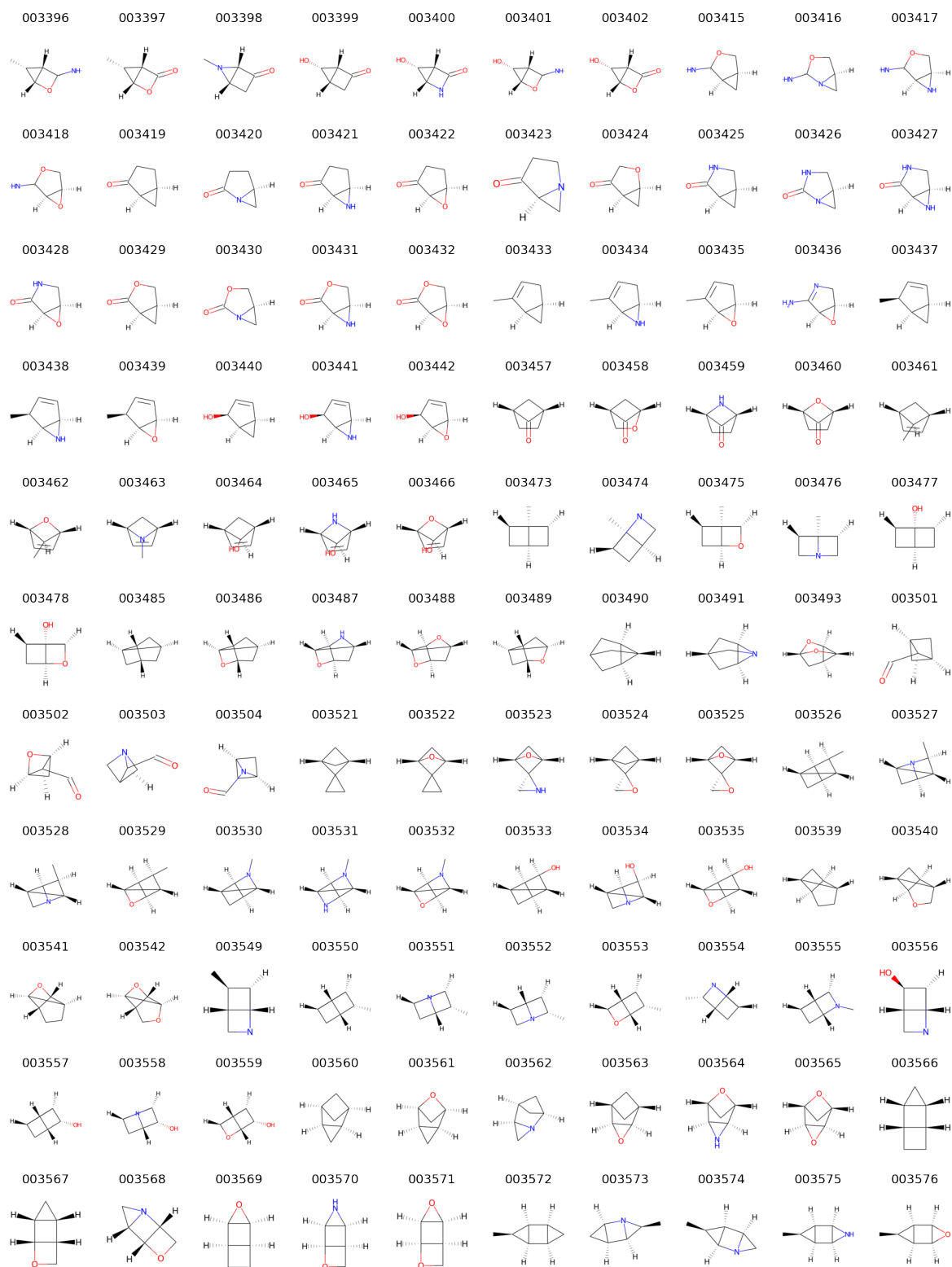


FIG. 13. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

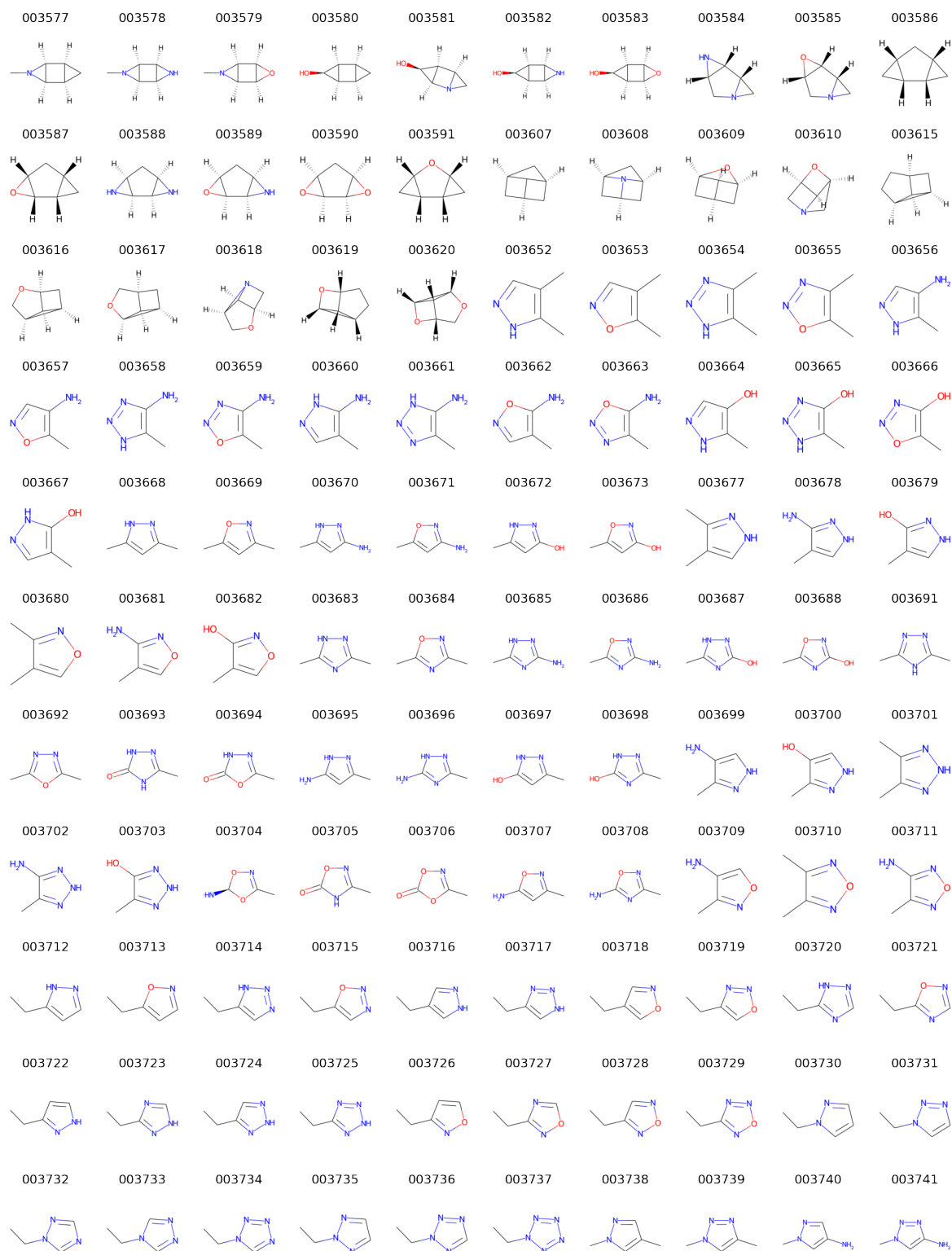


FIG. 14. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

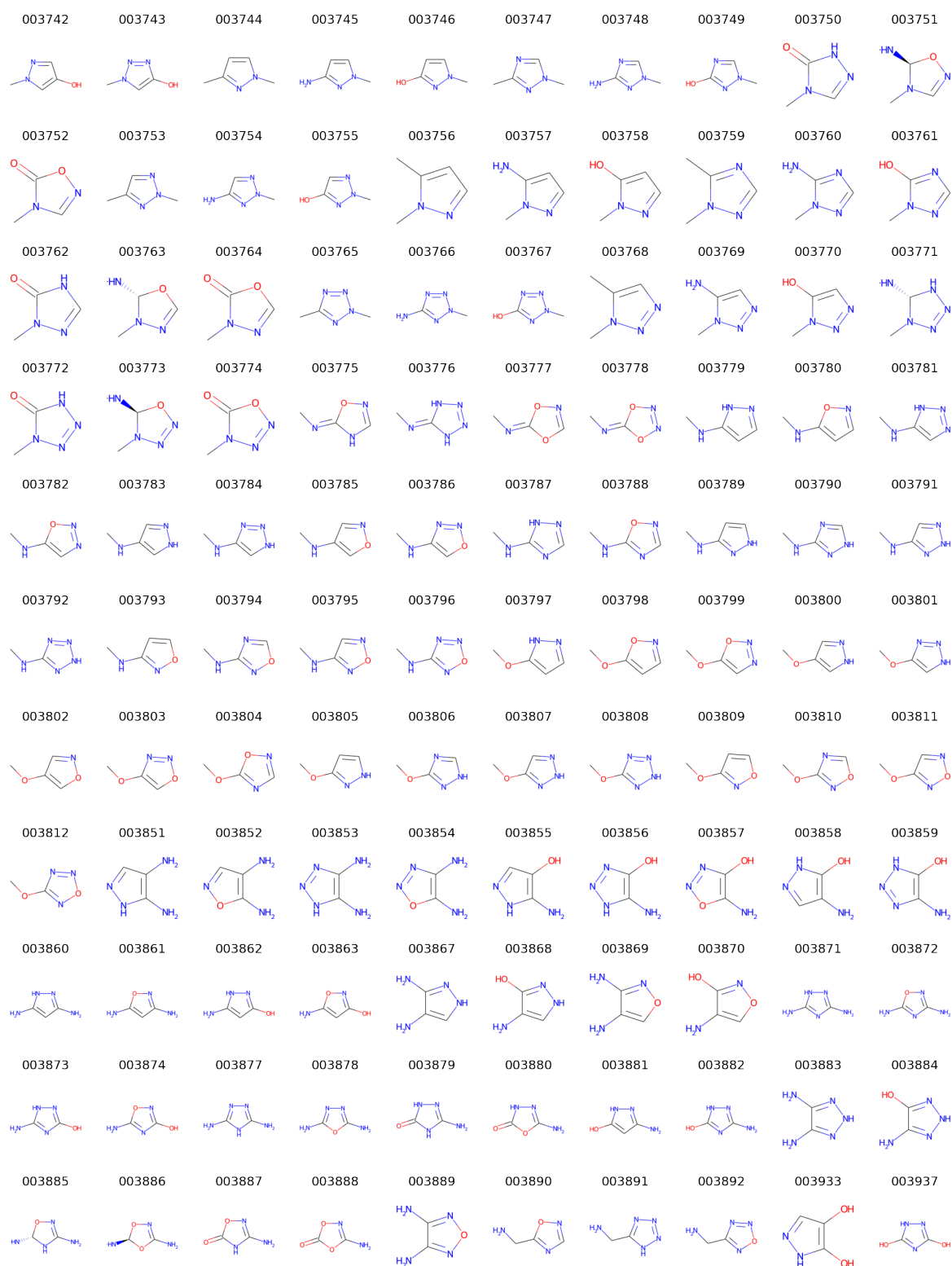


FIG. 15. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

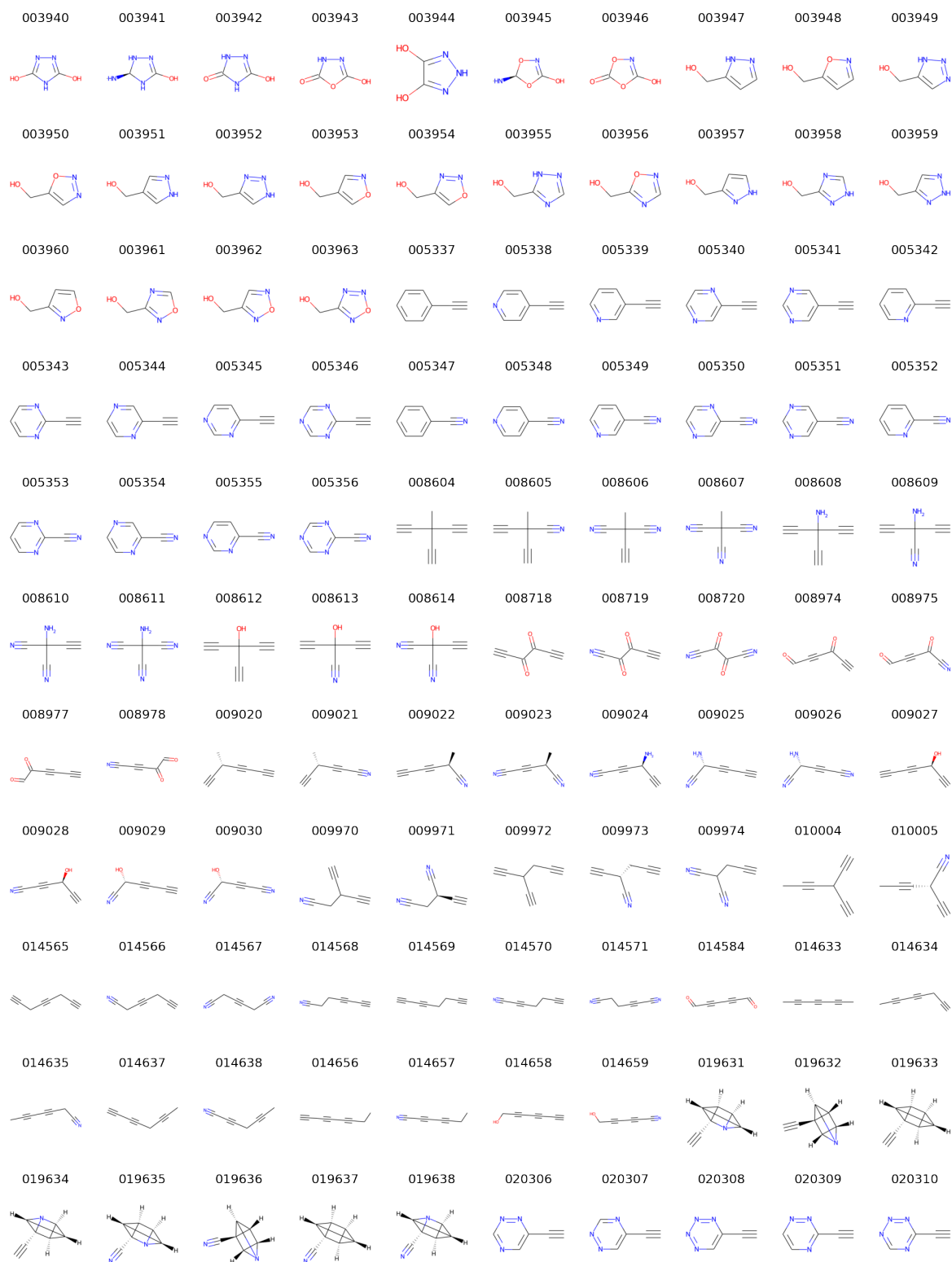


FIG. 16. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

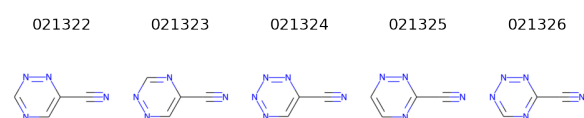


FIG. 17. The structures of the molecules for which atomic energies were calculated in this study. The numbers correspond to the respective ID in the QM9 dataset.

APPENDIX A.2 QUANTUM ALCHEMY BASED BONDING TRENDS AND THEIR
LINK TO HAMMETT'S EQUATION AND PAULING'S ELECTRONEGATIVITY MODEL

APPENDIX A.2.1

***Quantum Alchemy Based Bonding Trends and Their Link to
Hammett's Equation and Pauling's Electronegativity Model***

MICHAEL J. SAHRE, GUIDO FALK VON RUDORFF, AND O. ANATOLE VON LILIENFELD

J. Am. Chem. Soc. 2023, 145, 5899-5908.
<https://doi.org/10.1021/jacs.2c13393>

Contributions:

MICHAEL J. SAHRE performed the calculations, the analysis, and the model development, wrote the initial draft, and contributed to final version of the manuscript.

GUIDO FALK VON RUDORFF contributed to the analysis of the data and the model development, supervised the developments, and contributed to the final manuscript.

ANATOLE VON LILIENFELD conceived the scope of the manuscript, contributed to the analysis of the data and the model development, supervised the developments, and contributed to the final manuscript.

Reprinted with permission from *J. Am. Chem. Soc.* 2023, 145, 5899-5908.
Copyright 2023 American Chemical Society

Quantum Alchemy Based Bonding Trends and Their Link to Hammett's Equation and Pauling's Electronegativity Model

Michael J. Sahre, Guido Falk von Rudorff, and O. Anatole von Lilienfeld*



Cite This: *J. Am. Chem. Soc.* 2023, 145, 5899–5908



Read Online

ACCESS |



Metrics & More

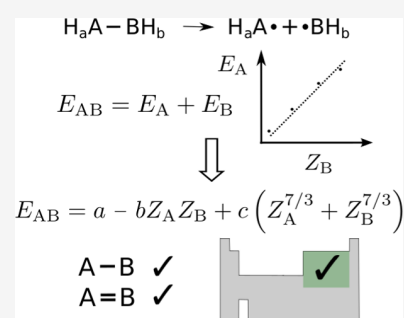


Article Recommendations



Supporting Information

ABSTRACT: We present an intuitive and general analytical approximation estimating the energy of covalent single and double bonds between participating atoms in terms of their respective nuclear charges with just three parameters, $[E_{AB} \approx a - bZ_A Z_B + c(Z_A^{7/3} + Z_B^{7/3})]$. The functional form of our expression models an alchemical atomic energy decomposition between participating atoms A and B. After calibration, reasonably accurate bond dissociation energy estimates are obtained for hydrogen-saturated diatomics composed of p-block elements coming from the same row $2 \leq n \leq 4$ in the periodic table. Corresponding changes in bond dissociation energies due to substitution of atom B by C can be obtained via simple formulas. While being of different functional form and origin, our model is as simple and accurate as Pauling's well-known electronegativity model. Analysis indicates that the model's response in covalent bonding to variation in nuclear charge is near-linear, which is consistent with Hammett's equation.



INTRODUCTION

Due to their direct link to thermodynamics stability, computational predictions of binding trends among molecules and materials have greatly improved design choices in the chemical sciences.^{1–8} Unfortunately, while numerical bonding estimates obtained from modern and computationally demanding quantum methods are very accurate and predictive, they are at the same time difficult to grasp with human intuition due to the inherent complexity of their solutions to the electronic Schrödinger equation.⁹ Already very approximate models, however, can yield intuitive descriptions of important bonding features. For example, Lewis's simple concept of binding electron pairs, in conjunction with the Aufbau principle, imposes relevant constraints from Pauli's exclusion principle,¹⁰ or the delocalization of π -electrons can be easily understood within Hückel theory (Figure 1, left).¹¹ Pauling's model for covalent bond energies is based on the idea to decompose the wave function within valence bond theory into ionic and covalent terms, leading to the chemically intuitive concept of electronegativity (Figure 1, center).¹² Because of their universality due to their foundation in quantum mechanics, these basic models are applicable across chemical compound space and have proven powerful for advancing our understanding of chemistry, despite their approximate nature. While they do not eliminate the need for experiments or accurate numerical calculations, they still offer meaningful predictions about the stability of different chemical compounds. For example, the octet rule^{13,14} postulated by Lewis in conjunction with the binding electron pair can be used to identify valid covalent bonding patterns in non-electron-deficient organic molecules. As another example,

the Hückel rule predicts whether a compound is stabilized due to aromaticity.

Less approximate and more recently, energy decomposition methods have been introduced to provide a more detailed understanding of chemical bonding. Partitioning of quantum mechanical observables onto the constituting parts of the quantum many-body system can be done in arbitrarily many ways. Some methods decompose the energy into different physical contributions;^{15–19} others partition it onto atoms.^{20–24} These methods have for example been useful to explain differences in interatomic interactions,²⁵ stability,^{26,27} or torsional energy profiles²⁸ across various molecules. Also within the framework of computational alchemical perturbation density functional theory²⁴ one can meaningfully quantify the effect of different binding partners on the atomic energy of every atom in the system.²⁹ Here, we introduce a model of covalent bonding, which we compare to molecular orbital (MO) and valence bond (VB) theory in Figure 1. The model directly emerges from quantum alchemy (QA) based atomic energy decomposition arguments. It is, to the best of our knowledge, new, yet as simple and accurate as Pauling's electronegativity model and of a distinctly different functional form and origin. We thus believe that it enables a fresh and intuitive perspective on covalent bonding. Such insights could

Received: December 15, 2022

Published: March 2, 2023



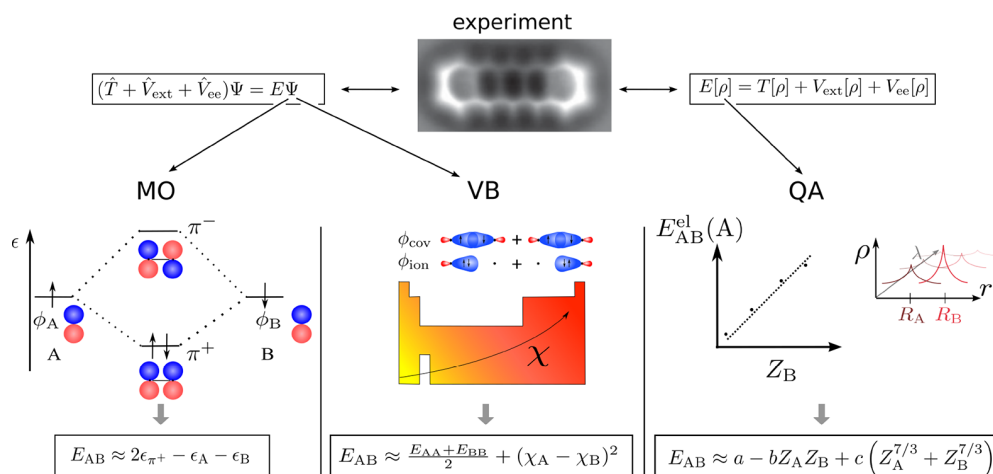


Figure 1. Experimental measurements (illustrated by an atomic force microscopy image of a pentacene molecule adapted from ref 35; reprinted with permission from AAAS) inform exact theory (Schrödinger equation and density functional theory), which informs three intuitive approximate views of the chemical bond (MO, VB, QA) in chemical compound space. MO represents molecular orbital theory accounting for bonding in energy diagrams (left). VB corresponds to valence bond theory enabling the decomposition of the wave function into covalent and ionic parts (middle). Quantum alchemy (QA) enables direct partitioning into atomic energies based on thermodynamic integration over varying nuclear charges. MO, VB, and QA can be used to account for bonding trends.

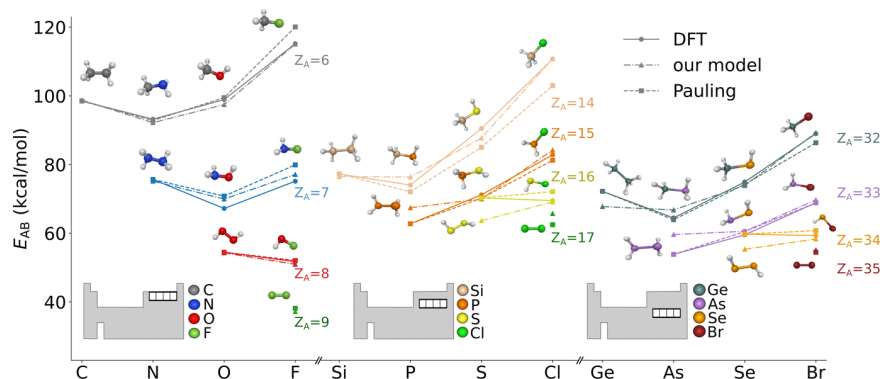


Figure 2. Calculated bond dissociation energies from density functional theory (DFT, PBE0/def2-TZVP), our quantum alchemy based chemical bond model (eq 2), and Pauling's electronegativity model in eq 4. MAEs with respect to DFT amount to 1.8 and 1.4 kcal/mol for our and Pauling's model, respectively.

be helpful for the development of more physics-based machine learning models applicable to quantum chemistry and chemical compound space. The inclusion of more physics in machine learning has been proven useful for gaining superior data efficiency and accuracy.^{30–34}

Within quantum alchemy, two systems can be interconverted by either interpolating the Hamiltonians or treating a change in the system perturbatively.^{24,36–38} We connect the energies E^{tot} and E^{UEG} of a molecule and the uniform electron gas via alchemical thermodynamic integration.²⁹ Using the Hellmann–Feynman theorem³⁹ and the chain rule affords a formally exact partitioning into a sum of atomic energies over all atoms I of the molecule as

$$E^{\text{tot}} - E^{\text{UEG}} = \sum_I Z_I \left(\underbrace{- \int d\mathbf{r} \int_0^1 d\lambda \frac{\rho(\lambda, \mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|}}_{:=E_I} + \frac{1}{2} \sum_{J \neq I} \frac{Z_J}{|\mathbf{R}_J - \mathbf{R}_I|} \right) \quad (1)$$

Z_I , Z_J and $\mathbf{R}_I$, $\mathbf{R}_J$ denote the nuclear charges and positions of nucleus I and the other nuclei J , respectively. The electron density is given as $\rho(\lambda, \mathbf{r})$ with the spatial coordinates $\mathbf{r}$ and λ as a coupling variable connecting the Hamiltonians of the molecule and the uniform electron gas.

Application of our alchemical decomposition scheme to binding energies between p-block elements reveals simple dependencies of atomic binding contributions on nuclear charges. This has motivated us to introduce the following approximate ansatz for the bond dissociation energy E_{AB} between atoms A and B,

$$E_{AB} \approx \underbrace{a}_{\text{period's offset}} - b \underbrace{Z_A Z_B}_{\propto \text{nuc. rep.}} + c \underbrace{(Z_A^{7/3} + Z_B^{7/3})}_{\propto \text{free atoms}} \quad (2)$$

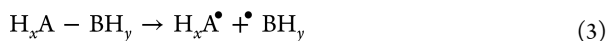
with a simple interpretation for each term and requiring just three global parameters, a , b , and c , which effectively account for interatomic distance and bond order, and the nuclear charges Z_A and Z_B as variables (see Methods section for details).

After calibration of parameters, we find that this simple model reproduces covalent binding among p-block elements of

either the second, third, or fourth row of the periodic table reasonably well. However, we note that the model is only applied to systems A–B with remaining valencies saturated with hydrogens; that is, further environmental influences are not yet being studied in this paper. In the following we also analyze and compare the model to density functional theory, quantum machine learning, semiempirical and post-Hartree–Fock quantum mechanics, Pauling’s electronegativity model, and Hammett’s equation. We subsequently discuss limitations, extensions to double bonds, and different electronic configurations and formulate a generalized expression that predicts bonding trends across multiple periods.

RESULTS AND DISCUSSION

Performance. After regression of parameters to DFT reference data, our model (eq 2) makes surprisingly accurate estimates. In particular, Figure 2 shows calculated bond dissociation energies (BDEs) for homolytic cleavage of diatomics A–B saturated with hydrogens,



Here, atoms A and B are fourth to seventh main group elements, both of either second, third, or fourth row of the periodic table. We use the term bond dissociation energy to describe the energy difference between the sum of the total energies of the two geometrically relaxed fragments and the total energy of the molecule from which the fragments are created by homolytic bond cleavage, but without the zero point energies added. BDEs for calibration were obtained from density functional theory (see Computational Details), and the row-dependent model parameters were determined from a least-squares fit to the 10 diatomics in each row (see Table 1).

Table 1. Coefficients and MAEs of Our Model Eq 2 for Different Rows n^a

n	a (kcal/mol)	b (630 Bohr $^{-1}$)	c (630 Bohr $^{-1}$ $e^{-1/3}$)	MAE (kcal/mol)
2	215.7	10.470	1.987	1.0
3	392.7	8.823	1.496	2.4
4	1109.2	6.180	0.813	2.2

^a b and c are scaled such that bZ_AZ_B and $c(Z_A^{7/3} + Z_B^{7/3})$ are given in kcal/mol if Z_A and Z_B are given in atomic units of charge e .

The residual deviation of estimated bond dissociation energies from the DFT reference amounts to an overall mean absolute error (MAE) of just 1.8 kcal/mol across all rows, falling just short of the highly coveted “chemical accuracy” threshold of ~ 1 kcal/mol. While such accuracy is extremely promising for such a simple functional form and so few parameters, we note the severe restrictions and limitations including fixed effective equilibrium geometries, participating elements, or bond orders.

To set this performance into a wider perspective, we compare calculated BDEs among other models well established in the literature. For this purpose, we selected all 10 single-bond diatomics from the second row contained within the legacy quantum chemistry W4-17 data set,⁴⁰ which provides highly accurate bond dissociation energies using explicit electron correlation methods for saturated diatomics in the second row. Figure 3 shows a scatter plot of BDEs obtained from various models. Calibrating our model using the W4-17 data yields a leave-one-out prediction error of 1.3 kcal/mol (see Computational Details). Training a chemically agnostic

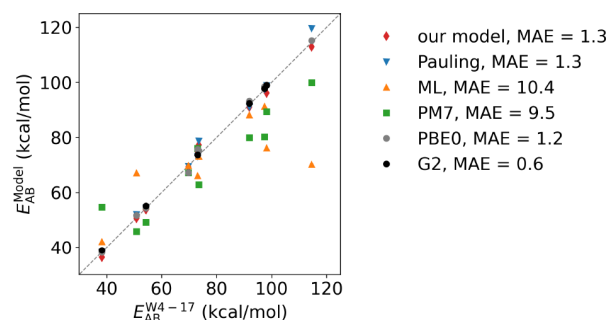


Figure 3. Calculated BDEs using our model, Pauling’s, various quantum methods (PM7, DFT (PBE0/def2-TZVP), G2), and a quantum machine learning model. Bond dissociation energies are scattered against the W4 entries⁴⁰ for 10 saturated single bond diatomics composed of carbon, nitrogen, oxygen, or fluorine.

quantum machine learning^{41–43} surrogate model (see Computational Details) results in a much higher leave-one-out prediction error of 10.4 kcal/mol. By comparison, generic QM methods such as the semiempirical PM7 method,⁴⁴ density functional theory (DFT/PBE0/def2-TZVP), and coupled cluster single double perturbative triples F12 calculations (taken from G2 ref 45) produce MAEs of 9.5, 1.2, and 0.6 kcal/mol, respectively. This indicates that our model can achieve accurate descriptions of trends in chemical compound space. We should caution, however, that our model is biased due to the calibration and is likely to perform significantly worse for other chemistries.

Furthermore, we probe the sensitivity of our model to the choice of DFT functional by comparing the leave-one-out prediction errors for the compounds in the second row fitted to BDEs computed with either PBE0 or B3LYP. The errors for PBE0 and B3LYP are 1.3 and 1.6 kcal/mol, respectively. This suggests that the accuracy of our model does not strongly depend on the DFT functional. Additionally, these findings show that the quality of the electronic structure method (DFT vs CCSDTQ5 in W4-17) does not affect the performance of the model significantly, as evident from similar errors for W4-17 (error of 1.3 kcal/mol) compared to PBE0 (1.3 kcal/mol) and B3LYP (1.6 kcal/mol).

Comparison to Pauling. Pauling’s model and our model are both simple expressions with similar overall accuracy. Pauling’s bond model¹² expresses the BDE as

$$E_{AB} = \frac{E_{AA} + E_{BB}}{2} + 23(\chi_A - \chi_B)^2 \quad (4)$$

which depends on the homolytic bond dissociation energies E_{AA} and E_{BB} and the electronegativities χ_A and χ_B . The model is based on a wave function decomposition into a covalent and an ionic part. Pauling proposed that the bond energy could accordingly be split into a covalent contribution, approximated as the mean of the homolytic bond dissociation energies, and an ionic contribution, postulated to be represented by the difference in electronegativities $(\chi_A - \chi_B)^2$. Pauling optimized electronegativities in order to reproduce bond dissociation energies as accurately as possible.¹² This model’s predictions for our test sets were discussed before (see also Figures 2, 3).

Note that Pauling’s expression requires knowledge of the homolytic bond dissociation energies and introduces electronegativities as an additional quantity, while our model relies directly on the nuclear charges. While electronegativity is useful to explain trends in chemical properties, our formulation

depending directly on nuclear charges is directly connected to the external potential in the electronic Hamiltonian and thereby more rigorously rooted in the fundamental physics governing chemistry. Furthermore, homodiatomics in the third and fourth period dominate the error of our model, while these cases are direct model parameters for Pauling. Note that our model outperforms Pauling's if the binding partners A and B have a large electronegativity difference, as observable for example for the C–F, Si–Cl, or N–F bond (Figure 2).

The reasonably accurate numerical prediction of bond dissociation energies by such simple models is intriguing. However, we believe that their main value lies rather in their conceptual power to explain bonding trends for varying chemical composition. For example, Pauling used his model to rationalize energetic trends of bonds in terms of squared electronegativity differences.¹² As obvious from visual inspection of Figure 2, our model (eq 2) provides a different explanation of such bonding trends in terms of balancing two terms with opposite sign: the linearly decaying second term versus the polynomially growing atomic energy term. Apart from the aforementioned advantage that our model does not necessitate the introduction of auxiliary concepts such as electronegativities, it is also consistent with Hammett's product ansatz, as described in the following section.

Comparison to Hammett's Equation. More than 80 years ago, Hammett proposed an empirical model^{46,47} to quantify property trends across chemical spaces as

$$P \approx \rho \cdot \sigma \quad (5)$$

Originally, $P = \log[K/K_0]$ was the equilibrium constant for various reactions of benzene derivatives normalized with respect to a reference reaction constant K_0 , σ described the effect of different substituents, and ρ accounted for the reaction type (e.g., mechanism or solvent). However, the model has been used to describe many other properties like activation energies,⁴⁸ orbital energies of metal organic complexes,⁴⁹ or dipole moments.⁵⁰ Furthermore, the model has been applied to nonbenzyl compounds.^{48,51} The relation between Hammett constants and electronegativity has also been noted in the context of nucleo- and electrophilicity relevant for mechanistic discussions in organic chemistry.⁵² While Hammett's model is very intuitive since it only requires separability of two dominating variables, its physical motivation has remained unclear.⁵³

Our model might offer a rationalization of Hammett's because it emerges from the quantum alchemy based atomic energy decomposition and accounts for the change in bond dissociation energy with respect to composition, fully consistent with Hammett's approach. In particular, the change of bond dissociation energy with respect to the nuclear charge, the partial derivative $\frac{\partial E_{AB}}{\partial Z_B}$ of our model, is simply given by

$$\frac{\partial E_{AB}}{\partial Z_B}(Z_A, Z_B) = -bZ_A + \frac{7}{3}cZ_B^{4/3} \quad (6)$$

Figure 4 displays the near-linear trend of this derivative for a fixed value of Z_A as a function of the number of valence electrons N_{VE} of Z_B for each period investigated in this study.

We have shifted each curve by a constant Δ such that it intersects the origin at $N_{VE} = 0$. The curves are approximately linear in the number of valence electrons with the slopes being proportional to $(7/3)c$ (see eq 6). Thus, $\frac{\partial E_{AB}}{\partial Z_B} + \Delta$ can be modeled by a Hammett ansatz with $\sigma = N_{VE}$ and ρ accounting

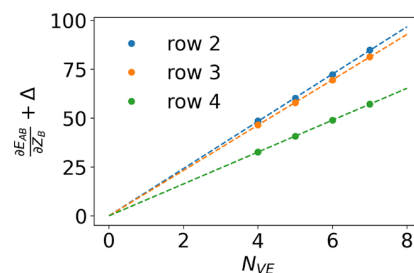


Figure 4. Sensitivity of the bond dissociation energy due to change in composition expressed by a Hammett-type relationship between the derivative of the bond dissociation energy with respect to the nuclear charge $\frac{\partial E_{AB}}{\partial Z_B}$ and the number of valence electrons N_{VE} of binding partner B for binding in different rows of the periodic table. The curves are shifted by a constant Δ such that they intersect the origin.

for different binding behavior due to a change in number of core electrons for different rows. As the row number increases, the slope becomes flatter, indicating a lower sensitivity of the bond dissociation energy to a change of the binding partner as observable in Figure 2. The drastic decrease of the slope from the third to the fourth row could be due to the additional 10 3d core electrons for elements in the fourth row.

The identification of the number of valence electrons of binding partner B as the σ -parameter is possible because the influence of the binding partner B is expressed as a function of the nuclear charge in our model. We have arrived at this expression based on the atomic energy decomposition within quantum alchemy. We believe that these findings indicate that such decomposition into atomic or fragment contributions can deepen our understanding of empirical rules such as Hammett's model. Note that this finding is also consistent with multiple other studies which found a correlation of the σ -parameters with atomic quantities like NMR shifts,^{54,55} polarizing force,⁵⁶ fragment self-similarity measures,⁵⁷ or atomic charges.^{58,59}

Limitations to Covalent Bonds. So far, we have only considered typical covalent bonding scenarios between p-block elements, i.e., elements from groups IV–VII in the periodic table. These elements all have in common that their valence electrons share the same second angular momentum quantum number. Our model has not been developed for decreased covalent character, i.e., for bonding atoms with differing second quantum numbers. Not surprisingly, for example, BDEs for single bonds shown in Figure 5 indicate a qualitatively different

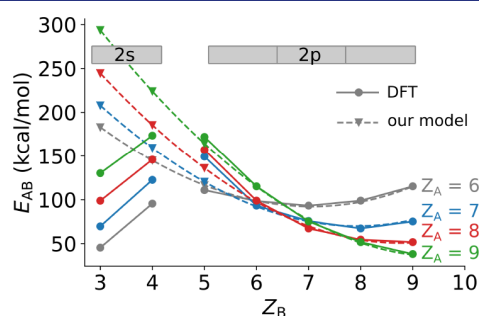


Figure 5. Limitations of applicability: BDEs between atom A and B corresponding to C/N/O/F or Li/Be/B/C/N/O/F, respectively. For second quantum number differing between A and B, our model [eq 2] breaks down.

behavior when one bonding partner, say atom B, comes from the alkaline or earth-alkaline group. According to our model calibrated for p elements only, bond dissociation energy content would monotonically increase as the nuclear charge Z_B decreases. In reality, however, bond dissociation energy content must decrease as B changes from predominantly covalent bonding among p-block elements into ionic bonding regimes with F_2 and LiF as the two opposite extremes.

Bond Order. Close inspection of results for the p-block elements shown in Figure 5 indicates that our model is the least accurate when it comes to the prediction of BDEs involving boron ($Z_B = 5$). This could be due to the parameters being optimized for bonds with bond orders being close to 1. B–N, B–O, and B–F bonds, however, are known to have a bond order of approximately 1.4.^{60–62} Consequently, one should expect our covalent single bond model to systematically underestimate the bond dissociation energy for these systems, consistent with the numerical observation.

To further explore this aspect, we have investigated whether our model can be recalibrated to predict bond dissociation energies for other bond orders. We have considered double bonds $A = B$ for all six possible combinations of elements from main groups IV, V, and VI within the same row of the periodic table, e.g., carbon, nitrogen, and oxygen for the second row. In contrast to the single-bonded molecules, the spin state is not the same for all double-bonded systems. O_2 , S_2 , and Se_2 have triplet ground states, while the other molecules are singlets. To remove the additional complexity due to a change in spin state, we used the same singlet spin state for all double-bonded molecules; that is, we refitted the parameters a , b , and c to energy differences ΔE^{TS} between fragments in a triplet state after homolytic bond cleavage and the bound molecule in a singlet state. The trend in these bond dissociation energies is very similar to the one for single bonds with a decrease in bond dissociation energies as the nuclear charges of the binding atoms grow. The resulting model's performance is equally good with prediction errors of 0.9, 1.3, and 2.0 kcal/mol for the second, third, and fourth row, respectively (see SI with Figure S1 and Table S1 for optimized parameters).

Electronic Configuration. To further investigate the impact of spin, we recalibrated our model to energy differences ΔE^{GS} between the ground states of fragments and molecules, e.g., O_2 in a triplet instead of a singlet state (Figure S2A, Table S2) and the energy differences ΔE^{TT} and ΔE^{SS} with fragments and molecules either both in triplet (Figure S2B, Table S3) or both in singlet states (Figure S2C, Table S4). The prediction of ΔE^{GS} yields an MAE of 7.4 kcal/mol for compounds in the second row. This increase in error with respect to ΔE^{TS} could be attributed to mixed spin states of the molecules. However, for the third and fourth row prediction errors for ΔE^{GS} are 1.6 and 1.9 kcal/mol, which is similar to the error for ΔE^{TS} . For these rows, the spin state changes not only for the molecules but also for the fragments. The fourth main group fragments (SiH_2 , GeH_2) have a singlet ground state, while the other fragments are triplets. Thus, changing spin states do not generally lower the accuracy.

The prediction errors for ΔE^{TT} and ΔE^{SS} are around 7–8 kcal/mol for the second row and 10–14 kcal/mol for the third and fourth row (see Table S3 and Table S4 for details). The worse performance compared to single-bonded systems is not surprising since the functional form of our model is inspired by the alchemical decomposition of bond dissociation energy

trends for single bonds in the ground state, which are significantly different from ΔE^{TT} and ΔE^{SS} .

Finally, we have also recalibrated our model (eq 2) to fit randomly drawn points from a uniform distribution in order to assess how far the performance of our model can be attributed to its mathematical flexibility or rather to its inherent capability to account for the underlying physics of the studied systems (see Figure S3). The MAE of the randomly drawn data is in all cases, except for ΔE^{TT} for the fourth row, substantially higher than for the binding energy differences. This finding corroborates the notion that our model's performance is not coincidence but rather due to its appropriate functional form approximating the relevant physics to a certain degree.

Trends and Dependence on Period. Since many questions in chemistry only require knowledge about differences in BDEs, we have investigated the applicability of our model toward the prediction of trends among bonds. More specifically, generalization of our model to deal with covalent bonding within any row from the p-block with just three parameters, a , b , and c , is impossible due to the large differences in nuclear charges with increasing row number n . We find, however, that a , b , and c vary smoothly with n when predicting changes in BDEs, i.e., $\Delta E = E_{AB} - E_{AC}$, in combination with a second-order Taylor expansion of the $Z^{7/3}$ terms. Then, parameters can be described as simple functions of the principal quantum number n (the row number) and ΔE is approximately given by

$$\Delta E \approx \Delta Z[28(n-1) + (8.5 + (n-3)^2)Z_A] + 6(Z_B^2 - Z_C^2) \quad (7)$$

with the nuclear charges Z_A , Z_B , and Z_C of the binding partners in the compounds $A-B$ and $A-C$ and the respective change in nuclear charges $\Delta Z = Z_C - Z_B$ (see SI for details).

Encouragingly, eq 7 meaningfully reproduces changes in covalent bond dissociation energies covering ranges from -40 to $+70$ kcal/mol with an MAE of just 4.2 kcal/mol with respect to density functional theory. Results are shown in Figure 6 for any single bond changes within the second, third, or fourth row of the periodic table.

Although of interest, the generalization of eq 7 to also account for atoms A, B, and C coming from different rows, or

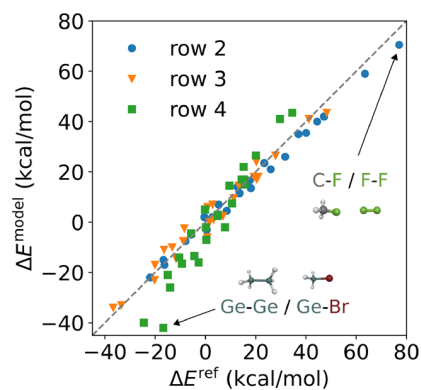


Figure 6. Trends in covalent binding: predicted (eq 7) vs truth (corresponding to PBE0/def2-TZVP). 72 bond dissociation energy differences between bonds $A-B$ and $A-C$ (ΔE), both within either the second, third, or fourth row of the periodic table, $\Delta E = E_{AB} - E_{AC}$, where $Z_B \leq Z_C$. Corresponding molecules involved are on display in Figure 2.

to involve other bond orders, has not yet been explored in this study.

CONCLUSION

We have presented a simple and, to the best of our knowledge, novel expression of covalent bonding in terms of nuclear charges. It has only three calibration parameters, which can easily be regressed to available reference data. Despite its simplicity, it is deeply rooted in the underlying physics of quantum mechanics via the computational alchemy based reasoning. We have found the model to be limited to covalent single and double bonds among atoms with the same first and second quantum number. We have presented promising numerical evidence for p-block atoms (except for rare gas elements) coming from the second, third, and fourth row. We have compared our model to Pauling's electronegativity model, and we have discussed its consistency with respect to Hammett's equation.

While only very approximate in nature and limited in scope (period and bond-order specific), we have found that our model is still sufficiently accurate to describe the trends of how bond dissociation energies respond upon modification of the chemical composition. This is reminiscent of substituent effect concepts discussed in undergraduate chemistry textbooks and used within the laboratory. As such, while limited in its numerical accuracy and scope, we believe that this model is still of fundamental importance for gaining a better intuition and understanding of chemical compound space. Furthermore, our model could enable a more efficient exploration of chemical space in design applications because first- and second-order gradients of the bond dissociation energy with respect to composition can be calculated due to the analytical form of our model.⁶³ For example, the sign of the first-order derivative would indicate whether the nuclear charge of an atomic site should be lowered or raised to increase the stability, while the magnitude would indicate how much the stability would change.

We note that the description of ionic, metallic, or van der Waals bonding through inclusion of s- and d-block elements is still outstanding and will be part of future research. This also holds for the generalization to bonds involving elements that differ in principal quantum numbers. Interactions beyond the two binding partners, e.g., the influence of immediate substituents via inductive or mesomeric effects, should be studied for further generalization. But also the inclusion of other environmental effects such as aromaticity, van der Waals interactions (hydrogen bonding, London dispersion), or electric field effects (from static multipole moments or externally) could extend the applicability of the model. Inclusion of these aspects would allow the calculation of environment-specific bond dissociation energies in molecules. This could be used as a baseline model for Δ -machine learning,^{64,65} multilevel grid combination,⁶⁶ multifidelity learning,⁶⁷ or transfer learning⁶⁸ models.

Conceptually speaking, our model relies on coarse-graining the expectation value of the electronic Hamiltonian throughout chemical space. As such, it is consistent with quantum mechanics and offers a fresh perspective on bond dissociation energies which is in line with Hammett's expression. Historically speaking, it represents an equally powerful yet possibly less empirical alternative to Pauling's electronegativity model. Future work will show to which extent this partitioning approach can be used to deepen our

understanding of chemical space with respect to other extensive properties and if it is useful for computational materials and molecular design efforts.

METHODS

Construction of the Model. Continuing from eq 1, the electronic energy $E^{\text{el}}(\text{AB})$, i.e., without nuclear repulsion, of a saturated diatomic A–B can be decomposed into

$$E^{\text{el}}(\text{AB}) = E^{\text{el}}(\text{A/AB}) + E^{\text{el}}(\text{B/AB}) + E^{\text{el}}(\text{AH/AB}) + E^{\text{el}}(\text{BH/AB}) \quad (8)$$

where $E^{\text{el}}(\text{A/AB})$ and $E^{\text{el}}(\text{B/AB})$ are the atomic energies of A and B and $E^{\text{el}}(\text{AH/AB})$ and $E^{\text{el}}(\text{BH/AB})$ are the sum of atomic energies of the hydrogens attached to A and B, respectively (see Figure 7 for an example).

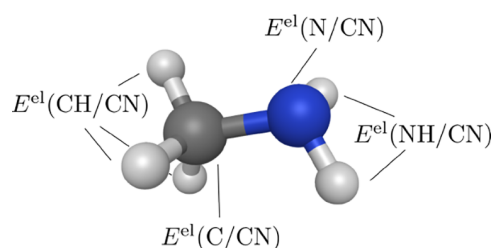


Figure 7. Atomic electronic energy decomposition notation exemplified for methylamine.

Furthermore, the electronic contribution to the binding energy $E_{\text{AB}}^{\text{el}}$ is the difference between the electronic energies $E^{\text{el}}(\text{A})$ and $E^{\text{el}}(\text{B})$ of the unrelaxed fragments A and B generated by homolytic bond cleavage and the electronic energy of compound AB.

$$E_{\text{AB}}^{\text{el}} = E^{\text{el}}(\text{A}) + E^{\text{el}}(\text{B}) - E^{\text{el}}(\text{AB}) \quad (9)$$

The electronic contribution to the *atomic* binding energy, e.g., $E_{\text{AB}}^{\text{el}}(\text{A})$ for atom A, is then defined as

$$E_{\text{AB}}^{\text{el}}(\text{A}) = E^{\text{el}}(\text{A/A}) - E^{\text{el}}(\text{A/AB}) \quad (10)$$

where $E^{\text{el}}(\text{A/A})$ and $E^{\text{el}}(\text{A/AB})$ are the atomic energies of A in fragment A and diatomic compound AB, respectively. So, the term atomic binding energy describes the contribution of a specific atom to the binding energy. The sum of the atomic binding energies of the hydrogen atoms attached to atom A is similarly defined as

$$E_{\text{AB}}^{\text{el}}(\text{AH}) = E^{\text{el}}(\text{AH/A}) - E^{\text{el}}(\text{AH/AB}) \quad (11)$$

where $E^{\text{el}}(\text{AH/A})$ and $E^{\text{el}}(\text{AH/AB})$ are the sum of atomic energies of hydrogen atoms in the fragment A and the compound AB, respectively. In analogy, the atomic binding energies $E_{\text{AB}}^{\text{el}}(\text{B})$ of atom B and $E_{\text{AB}}^{\text{el}}(\text{BH})$ of the hydrogens attached to B are defined as

$$E_{\text{AB}}^{\text{el}}(\text{B}) = E^{\text{el}}(\text{B/B}) - E^{\text{el}}(\text{B/AB}) \quad (12)$$

and

$$E_{\text{AB}}^{\text{el}}(\text{BH}) = E^{\text{el}}(\text{BH/B}) - E^{\text{el}}(\text{BH/AB}) \quad (13)$$

Using the definitions for atomic binding energies introduced in eqs 10–13 the electronic contribution to the binding energy defined in eq 9 can also be expressed as a sum of atomic terms as

$$E_{\text{AB}}^{\text{el}} = E_{\text{AB}}^{\text{el}}(\text{A}) + E_{\text{AB}}^{\text{el}}(\text{AH}) + E_{\text{AB}}^{\text{el}}(\text{B}) + E_{\text{AB}}^{\text{el}}(\text{BH}) \quad (14)$$

Calculating such atomic binding energies according to APDFT,²⁹ we have observed that $E_{\text{AB}}^{\text{el}}(\text{A})$ increases approximately linearly with the nuclear charge Z_{B} of binding partner B for saturated diatomics composed from the elements carbon, nitrogen, oxygen, and fluorine (Figure S4A). Furthermore, the atomic energy of the hydrogen atoms

remains approximately constant with varying binding partner (Figure S4B).

This enables us to express the binding energy approximately as

$$E_{AB}^{\text{el}} \approx \frac{\beta_A Z_B + \alpha_A}{=E_{AB}^{\text{el}}(A)+E_{AB}^{\text{el}}(AH)} + \frac{\beta_B Z_A + \alpha_B}{=E_{AB}^{\text{el}}(B)+E_{AB}^{\text{el}}(BH)} \quad (15)$$

The energy contributions of fragments A and B are characterized by the parameters α_A , β_A and α_B , β_B , respectively. α accounts for the constant contribution of the heavy atom (A or B) and of the hydrogens attached to it to the binding energy, while β describes the contribution of the heavy atom for varying binding partners. Furthermore, we approximate the contribution of the nuclear repulsion to the binding energy E_{AB}^{nuc} by only considering the interaction between the heavy atoms A and B at an average bond distance $\bar{d}$. This average is calculated from the bond lengths between A and B for all considered compounds A–B. Consequently, the total binding energy E_{AB} including nuclear repulsion can be expressed as

$$E_{AB} \approx \frac{\beta_A Z_B + \alpha_A}{=E_{AB}^{\text{el}}(A)+E_{AB}^{\text{el}}(AH)} + \frac{\beta_B Z_A + \alpha_B}{=E_{AB}^{\text{el}}(B)+E_{AB}^{\text{el}}(BH)} - \frac{Z_A Z_B}{\bar{d}} \quad (16)$$

The parameters α and β are determined from a least-squares fit to $E_{AB} + \frac{Z_A Z_B}{\bar{d}}$ for all 10 unique combinations A–B of carbon, nitrogen, oxygen, and fluorine. The bond dissociation energies E_{AB} are calculated with DFT using the energies of the fragments in relaxed geometries (see Computational Details). We note that the trends for atomic binding energies are based on calculations using the unrelaxed fragments. However, in the following, we fit our model to bond dissociation energies with relaxed fragments because the use of the relaxed fragments is more useful to calculate the stability of bonds. Since the trends in binding or bond dissociation energy across different compounds are similar for strained and relaxed fragments (SI Figure S7), the findings for strained fragments should be transferable to the relaxed ones.

While eq 16 is based on trends for diatomics in the second row of the periodic table, we also apply it to diatomics from the fourth to seventh main group of the third and fourth row.

The MAEs for bond dissociation energy predictions with eq 16 are 0.5, 0.4, and 0.3 kcal/mol for rows 2–4, respectively, and the optimized parameters can be found in Table S5. The model accurately reproduces the bond dissociation energy and can be applied to different rows of the periodic table. However, it is also prone to overfitting because it uses eight parameters to model bond dissociation energy for 10 compounds. Thus, we studied the relation of the parameters α and β on the nuclear charge of the respective element in an attempt to reduce the number of parameters.

The optimized parameter β_A correlates linearly with Z_A ,

$$\beta_A \approx b' Z_A \quad (17)$$

as shown in Figure S5A with b' being the slope. Thus, the terms $\beta_A Z_B$ and $\beta_B Z_A$ in eq 16 can be written as

$$\beta_A Z_B + \beta_B Z_A \approx 2b' Z_A Z_B \quad (18)$$

The β -terms account mainly for the large nuclear repulsion term $Z_A Z_B \bar{d}^{-1} \gg E_{AB}$ in the dependent variable $E_{AB} + \frac{Z_A Z_B}{\bar{d}}$. We note that the linear dependence of β_A on Z_A is also consistent with the relation

$$E_{QR}^{\text{el}} \approx E_{SR}^{\text{el}} + \frac{1}{2}(E_{QQ}^{\text{el}} - E_{SS}^{\text{el}}) \quad (19)$$

between binding energies of elements Q, R, and S with $Z_Q = Z_R - 1 = Z_S - 2$, which was derived from alchemical enantiomers.⁶⁹ Insertion of the definition for electronic binding energies from eq 15 into eq 19 yields

$$\begin{aligned} \beta_Q Z_R + \alpha_Q + \beta_R Z_Q + \alpha_R \\ = \beta_S Z_R + \alpha_S + \beta_R Z_S + \alpha_R + \beta_Q Z_Q + \alpha_Q - \beta_S Z_S - \alpha_S \\ \beta_Q Z_R + \beta_R Z_Q = \beta_S Z_R + \beta_R Z_S + \beta_Q Z_Q - \beta_S Z_S \\ \beta_R \underbrace{(Z_Q - Z_S)}_{=-2} = \beta_S \underbrace{(Z_R - Z_S)}_{=-1} + \beta_Q \underbrace{(Z_Q - Z_R)}_{=-1} \\ \beta_R = \frac{\beta_S + \beta_Q}{2} \end{aligned} \quad (20)$$

which implies a linear relation between the different values for β .

The optimized offset α_A has a nonlinear relationship with Z_A (Figure S5B). Thus, we model the dependence of α_A on Z_A as

$$\alpha_A \approx c Z_A^\gamma + a' \quad (21)$$

where the exponent γ accounts for the nonlinearity and a' is an offset.

Substitution of β and α in eq 16 with the expressions in eq 17 and eq 21 and rearrangement of the resulting equation yields

$$E_{AB} \approx a - b Z_A Z_B + c(Z_A^\gamma + Z_B^\gamma) \quad (22)$$

with $a = 2a'$ and $b = \left(\frac{1}{\bar{d}} - 2b'\right)$. Optimization of the parameters in eq 22 by a combination of nonlinear least-squares for a , b , and c and a line scan for γ leads to similar values for the optimal exponent γ_{opt} for the different rows (see Figure S6). Thus, γ is kept the same for all rows and only a , b , and c are optimized for each row independently. The optimal value for γ in this optimization procedure is $\gamma_{\text{opt}} = 2.36 \approx \frac{7}{3}$. This is an interesting result because $-0.768745 \cdot Z^{7/3}$ is the leading term in an expansion of the energy of a free atom in its nuclear charge.^{70,71} Hence, the bond dissociation energy can be expressed as

$$\begin{aligned} E_{AB} \approx a - b Z_A Z_B + c(Z_A^{7/3} + Z_B^{7/3}) \\ \approx a - b Z_A Z_B + c'(E_A^{\text{atom}} + E_B^{\text{atom}}) \end{aligned} \quad (23)$$

with the energies of the free atoms E_A^{atom} and E_B^{atom} and $c' = \frac{c}{0.768745}$. So, according to our model, the bond dissociation energies can be approximated as a sum of three terms, a , $-b Z_A Z_B$, and $c(Z_A^{7/3} + Z_B^{7/3})$. All three parameters depend on bond order and could be interpreted in the following way: a is a period-dependent energy offset reflecting the impact of the principal quantum number. $-b Z_A Z_B$ is proportional to the nuclear repulsion with b having the dimension of the inverse distance, and $c(Z_A^{7/3} + Z_B^{7/3})$ is the sum of the energies of the free atoms, scaled by c , which has dimensions of inverse distance and inverse cubic root of charge. The value of the parameter a changes much more from the third to the fourth row (≈ 700 kcal/mol) compared to the change from the second to the third row (≈ 180 kcal/mol). This difference can be explained by the addition of 18 core electrons from third to fourth in contrast to only 8 core electrons from the second to third row of the periodic table. The parameters b and c are less sensitive to a change in principal quantum number because in the respective terms $-b Z_A Z_B$ and $c(Z_A^{7/3} + Z_B^{7/3})$ mainly the nuclear charges account for changes in the number of core electrons.

COMPUTATIONAL DETAILS

The atomic energies (eq 1) of the saturated diatomics were calculated using geometries from the amons data set.⁷² The fragment structures were generated by splitting the homo-diatomics without further geometry optimization. The required electron densities were obtained following the procedure in earlier work²⁹ from calculations with the CPMD⁷³ code and the CPMD2CUBE program⁷⁴ using the PBE⁷⁵ functional in a plane wave basis with a cutoff of 200 Ry, GTH^{76,77} pseudopotentials, and a wave function gradient convergence set to 10^{-6} making partial use of gnu parallel.⁷⁸ A primitive cell

with a box length of 14.338 Å was used for saturated diatomics and of 11.380 Å for the fragments. The GTH pseudopotential parameters were scaled by $\lambda = 6/14, 8/14, 11/14$, and 1 and $\lambda = 3/7, 4/7, 6/7$, and 1 for the saturated diatomics and the fragments, respectively, to generate electron densities for different values of λ . The electron density at $\lambda = 0$ was represented as a uniform distribution. The integration with respect to $\mathbf{r}$ and λ was carried out as weighted summation over grid points and with the trapezoidal rule, respectively.

The bond dissociation energies to determine the optimal parameters in eqs 2 and 16 were calculated with PySCF^{79–82} for single-bonded systems and with GAUSSIAN⁸³ for double-bonded systems with PBE0/def2-TZVP⁸⁴ (restricted open shell for fragments). Reported bond dissociation energies are for optimized geometries of saturated diatomics and fragments. Initial guesses for the structures were generated through the LERULI API.^{85–88} The parameter optimization in eq 2, 7, 16, and 22 was performed by linear or nonlinear least-squares fitting as implemented in numpy⁸⁹ and scipy.⁹⁰ Bond dissociation energies of our model presented in Figure 3 are with respect to W4-17 data after determining the parameters via leave-one-out cross-validation. The machine learning predictions were obtained from kernel ridge regression with a Gaussian kernel and the bag of bonds⁹¹ representation in leave-one-out cross-validation. PM7 bond dissociation energies were calculated with MOPAC2016.⁹² For better comparison, enthalpy contributions and zero point energies were subtracted from the computed heat of formation. Predictions with Pauling's model eq 4 use electronegativities as reported by Pauling.¹²

■ ASSOCIATED CONTENT

Data Availability Statement

Example input files and pseudopotentials to calculate alchemical atomic energies; bond dissociation energies, total energies, and optimized structures in xyz-format calculated at the PBE0/def2-TZVP level and code used to optimize the parameters in the different versions of our model are available at 10.5281/zenodo.7684070.

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.2c13393>.

Optimized parameters and predicted bond dissociation energies for double bonds; additional data used for the derivation of our model; total energies of single- and double-bonded molecules presented in this study (PDF)

■ AUTHOR INFORMATION

Corresponding Author

O. Anatole von Lilienfeld – Vector Institute for Artificial Intelligence, Toronto M5S 1M1, Canada; Departments of Chemistry, Materials Science and Engineering, and Physics, University of Toronto, St. George Campus, Toronto M5S 0A3, Canada; Machine Learning Group, Technische Universität Berlin and Institute for the Foundations of Learning and Data, Berlin 10587, Germany; orcid.org/0000-0001-7419-0466; Email: anatole.vonlilienfeld@utoronto.ca

Authors

Michael J. Sahre – Faculty of Physics, University of Vienna, Vienna 1090, Austria; Vienna Doctoral School in Chemistry

(DoSChem), University of Vienna, Vienna 1090, Austria; orcid.org/0000-0003-3318-7484

Guido Falk von Rudorff – Department of Chemistry, University Kassel, Kassel 34132, Germany; orcid.org/0000-0001-7987-4330

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/jacs.2c13393>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge discussions with M. Meuwly, M. Bragato, and P. Marquetand, as well as support from the European Research Council (ERCCoG Grant QML). This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement #772834. O.A.v.L. has received support as the Ed Clark Chair of Advanced Materials and as a CIFAR AI chair.

■ REFERENCES

- (1) von Lilienfeld, O. A. In *Many-Electron Approaches in Physics, Chemistry and Mathematics: A Multidisciplinary View*; Bach, V., Site, L. D., Eds.; Springer: Cham, 2014; pp 169–189.
- (2) Honarparvar, B.; Govender, T.; Maguire, G. E.; Soliman, M. E.; Kruger, H. G. Integrated approach to structure-based enzymatic drug design: Molecular modeling, spectroscopy, and experimental bioactivity. *Chem. Rev.* **2014**, *114*, 493–537.
- (3) Kumalo, H. M.; Bhakat, S.; Soliman, M. E. Theory and applications of covalent docking in drug discovery: Merits and pitfalls. *Molecules* **2015**, *20*, 1984–2000.
- (4) London, N.; Miller, R. M.; Krishnan, S.; Uchida, K.; Irwin, J. J.; Eidam, O.; Gibold, L.; Cimermančič, P.; Bonnet, R.; Shoichet, B. K.; Taunton, J. Covalent docking of large libraries for the discovery of chemical probes. *Nat. Chem. Biol.* **2014**, *10*, 1066–1072.
- (5) Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H. Towards the computational design of solid catalysts. *Nat. Chem.* **2009**, *1*, 37–46.
- (6) Curtarolo, S.; Hart, G. L.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The high-throughput highway to computational materials design. *Nat. Mater.* **2013**, *12*, 191–201.
- (7) Hannagan, R. T.; Giannakakis, G.; Réocreux, R.; Schumann, J.; Finzel, J.; Wang, Y.; Michaelides, A.; Deshlahra, P.; Christopher, P.; Flytzani-Stephanopoulos, M.; Stamatakis, M.; Sykes, E. C. H. First-principles design of a single-atom-alloy propane dehydrogenation catalyst. *Science* **2021**, *372*, 1444–1447.
- (8) Schwalbe-Koda, D.; Kwon, S.; Paris, C.; Bello-Jurado, E.; Jensen, Z.; Olivetti, E.; Willhammar, T.; Corma, A.; Román-Leshkov, Y.; Moliner, M.; Gómez-Bombarelli, R. A priori control of zeolite phase competition and intergrowth with high-throughput simulations. *Science* **2021**, *374*, 308–315.
- (9) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley & Sons, Ltd, 2000.
- (10) Zhao, L.; Schwarz, W. H.; Frenking, G. The Lewis electron-pair bonding model: the physical background, one century later. *Nature Reviews Chemistry* **2019**, *3*, 35–47.
- (11) Hückel, E. Quantentheoretische Beiträge zum Benzolproblem. *Zeitschrift für Physik* **1931**, *70*, 204–286.
- (12) Pauling, L. *The Nature of the Chemical Bond*; Cornell University Press, 1960; pp 79–96.
- (13) Lewis, G. N. The Atom and the Molecule. *J. Am. Chem. Soc.* **1916**, *38*, 762–785.
- (14) Langmuir, I. The Arrangement of Electrons in Atoms and Molecules. *J. Am. Chem. Soc.* **1919**, *41*, 868–934.

- (15) Morokuma, K. Molecular Orbital Studies of Hydrogen Bonds. III. C = O-H-O Hydrogen Bond in H₂CO-H₂O and H₂CO-2H₂O. *J. Chem. Phys.* **1971**, *55*, 1236–1244.
- (16) Ziegler, T.; Rauk, A. On the calculation of bonding energies by the Hartree Fock Slater method - I. The transition state method. *Theoretica Chimica Acta* **1977**, *46*, 1–10.
- (17) Frenking, G.; Matthias Bickelhaupt, F. *The Chemical Bond*; John Wiley & Sons, Ltd, 2014; Chapter 4, pp 121–157.
- (18) Rahm, M.; Hoffmann, R. Toward an experimental quantum chemistry: Exploring a new energy partitioning. *J. Am. Chem. Soc.* **2015**, *137*, 10282–10291.
- (19) Zulueta, B.; Tulyani, S. V.; Westmoreland, P. R.; Frisch, M. J.; Petersson, E. J.; Petersson, G. A.; Keith, J. A.; Bond-Energy, A. A Bond-Energy/Bond-Order and Populations Relationship. *J. Chem. Theory Comput.* **2022**, *18*, 4774–4794.
- (20) Guevara-Vela, J. M.; Francisco, E.; Rocha-Rinza, T.; Martín Pendás, A. Interacting Quantum Atoms-A Review. *Molecules* **2020**, *25*, 4028.
- (21) Yu, M.; Trinkle, D. R.; Martin, R. M. Energy density in density functional theory: Application to crystalline defects and surfaces. *Phys. Rev. B* **2011**, *83*, 115113.
- (22) Barone, V.; Fliszár, S. Theoretical energies of representative carbon-carbon bonds. *Int. J. Quantum Chem.* **1995**, *55*, 469–476.
- (23) Eriksen, J. J. Mean-field density matrix decompositions. *J. Chem. Phys.* **2020**, *153*, 214109.
- (24) von Rudorff, G. F.; von Lilienfeld, O. A. Alchemical perturbation density functional theory. *Physical Review Research* **2020**, *2*, 23220.
- (25) Zhao, L.; Pan, S.; Frenking, G. The nature of the polar covalent bond. *J. Chem. Phys.* **2022**, *157*, 034105.
- (26) Blokker, E.; Sun, X.; Poater, J.; van der Schuur, J. M.; Hamlin, T. A.; Bickelhaupt, F. M. The Chemical Bond: When Atom Size Instead of Electronegativity Difference Determines Trend in Bond Strength. *Chem.—Eur. J.* **2021**, *3*, 1–8.
- (27) Hansen, T.; Vermeeren, P.; Bickelhaupt, F. M.; Hamlin, T. A. Stability of alkyl carbocations. *Chem. Commun.* **2022**, *58*, 12050–12053.
- (28) Darley, M. G.; Popelier, P. L. Role of short-range electrostatics in torsional potentials. *J. Phys. Chem. A* **2008**, *112*, 12954–12965.
- (29) Von Rudorff, G. F.; Von Lilienfeld, O. A. Atoms in Molecules from Alchemical Perturbation Density Functional Theory. *J. Phys. Chem. B* **2019**, *123*, 10073–10082.
- (30) Huang, B.; von Lilienfeld, O. A. Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity. *J. Chem. Phys.* **2016**, *145*, 161102.
- (31) Christensen, A. S.; Faber, F. A.; von Lilienfeld, O. A. Operators in quantum machine learning: Response properties in chemical space. *J. Chem. Phys.* **2019**, *150*, 064105.
- (32) Faber, F. A.; Christensen, A. S.; Lilienfeld, O. A. v. In *Machine Learning Meets Quantum Physics*; Schütt, K. T., Chmiela, S., von Lilienfeld, O. A., Tkatchenko, A., Tsuda, K., Müller, K.-R., Eds.; Springer International Publishing: Cham, 2020; pp 155–169.
- (33) Huang, B.; von Lilienfeld, O. A. Ab Initio Machine Learning in Chemical Compound Space. *Chem. Rev.* **2021**, *121*, 10001–10036.
- (34) Musil, F.; Grisafi, A.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M. Physics-Inspired Structural Representations for Molecules and Materials. *Chem. Rev.* **2021**, *121*, 9759–9815.
- (35) Gross, L.; Mohn, F.; Moll, N.; Liljeroth, P.; Meyer, G. The Chemical Structure of a Molecule Resolved by Atomic Force Microscopy. *Science* **2009**, *325*, 1110–1114.
- (36) Hückel, E. Quantentheoretische Beiträge zum Benzolproblem. *Zeitschrift für Physik* **1931**, *72*, 310–337.
- (37) Wilson, E. B. Four-dimensional electron density function. *J. Chem. Phys.* **1962**, *36*, 2232–2233.
- (38) Von Lilienfeld, O. A. First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties. *Int. J. Quantum Chem.* **2013**, *113*, 1676–1689.
- (39) Feynman, R. P. Forces in Molecules. *Phys. Rev.* **1939**, *56*, 340–343.
- (40) Karton, A.; Sylvetsky, N.; Martin, J. M. W4–17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods. *J. Comput. Chem.* **2017**, *38*, 2063–2075.
- (41) Huang, B.; Von Lilienfeld, O. A. Ab Initio Machine Learning in Chemical Compound Space. *Chem. Rev.* **2021**, *121*, 10001–10036.
- (42) Tkatchenko, A. Machine learning for chemical discovery. *Nat. Commun.* **2020**, *11*, 8–11.
- (43) von Lilienfeld, O. A.; Müller, K. R.; Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nature Reviews Chemistry* **2020**, *4*, 347–358.
- (44) Stewart, J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1–32.
- (45) Haunschild, R.; Klopper, W. New accurate reference energies for the G2/97 test set. *J. Chem. Phys.* **2012**, *136*, 164102.
- (46) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.
- (47) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.
- (48) Bragato, M.; Von Rudorff, G. F.; Von Lilienfeld, O. A. Data enhanced Hammett-equation: Reaction barriers in chemical space. *Chemical Science* **2020**, *11*, 11859–11868.
- (49) Chang, A. M.; Freeze, J. G.; Batista, V. S. Hammett neural networks: prediction of frontier orbital energies of tungsten-benzylidyne photoredox complexes. *Chem. Sci.* **2019**, *10*, 6844–6854.
- (50) van Beek, L. K. H. A relationship between dipole moments and the Hammett equation. *Recueil des Travaux Chimiques des Pays-Bas* **1957**, *76*, 729–732.
- (51) Ayoubi-Chianeh, M.; Kassaei, M. Z. Toward triplet disilavinylidenes: A Hammett electronic survey for substituent effects on singlet-triplet energy gaps of silylenes by DFT. *J. Phys. Org. Chem.* **2019**, *32*, e3988–e3988.
- (52) Liu, L.; Fu, Y.; Liu, R.; Li, R.-Q.; Guo, Q.-X. Hammett Equation and Generalized Pauling's Electronegativity Equation. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 652–657.
- (53) Fernández, I.; Frenking, G. Correlation between Hammett Substituent Constants and Directly Calculated π -Conjugation Strength. *Journal of Organic Chemistry* **2006**, *71*, 2251–2256.
- (54) Yoder, C. H.; Tuck, R. H.; Hess, R. E. Nuclear magnetic resonance studies of the bonding in aromatic systems. Correlation of Hammett sigma constants with methyl ¹³C-H coupling constants and chemical shifts. *J. Am. Chem. Soc.* **1969**, *91*, 539–543.
- (55) Axenrod, T.; Pregosin, P. S.; Wieder, M. J.; Milne, G. W. A. Nitrogen-15 magnetic resonance spectroscopy. Correlation of the ¹⁵N-H coupling constants in aniline derivatives with Hammett σ constants. *J. Am. Chem. Soc.* **1969**, *91*, 3681–3682.
- (56) Price, C. C. Substitution and Orientation in the Benzene Ring. *Chem. Rev.* **1941**, *29*, 37–67.
- (57) Gironés, X.; Ponc, R. Molecular Quantum Similarity Measures from Fermi Hole Densities: Modeling Hammett Sigma Constants. *J. Chem. Inf. Model.* **2006**, *46*, 1388–1393.
- (58) Genix, P.; Jullien, H.; Le Goas, R. Estimation of Hammett sigma constants from calculated atomic charges using partial least squares regression. *Journal of Chemometrics* **1996**, *10*, 631–636.
- (59) Ertl, P. Simple Quantum Chemical Parameters as an Alternative to the Hammett Sigma Constants in QSAR Studies. *Quantitative Structure-Activity Relationships* **1997**, *16*, 377–382.
- (60) Mayer, I. Charge, bond order and valence in the AB initio SCF theory. *Chem. Phys. Lett.* **1983**, *97*, 270–274.
- (61) Mayer, I. Bond order and valence: Relations to Mulliken's population analysis. *Int. J. Quantum Chem.* **1984**, *26*, 151–154.
- (62) Bridgeman, A. J.; Cavigliasso, G.; Ireland, L. R.; Rothery, J. The Mayer bond order as a tool in inorganic chemistry. *J. Chem. Soc., Dalton Trans.* **2001**, 2095–2108.

- (63) Saravanan, K.; Kitchin, J. R.; Von Lilienfeld, O. A.; Keith, J. A. Alchemical predictions for computational catalysis: Potential and limitations. *J. Phys. Chem. Lett.* **2017**, *8*, S002–S007.
- (64) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ -Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.
- (65) Huang, B.; von Lilienfeld, O. A.; Krogel, J. T.; Benali, A. Towards DMC accuracy across chemical space with scalable Δ -QML. *arXiv (Chemical Physics)*. 2022 October 12, DOI: DOI: 10.48550/ARXIV.2210.06430 (accessed 2023–02–08).
- (66) Zaspel, P.; Huang, B.; Harbrecht, H.; von Lilienfeld, O. A. Boosting Quantum Machine Learning Models with a Multilevel Combination Technique: Pople Diagrams Revisited. *J. Chem. Theory Comput.* **2019**, *15*, 1546–1559.
- (67) Batra, R.; Pilania, G.; Uberuaga, B. P.; Ramprasad, R. Multifidelity Information Fusion with Machine Learning: A Case Study of Dopant Formation Energies in Hafnia. *ACS Appl. Mater. Interfaces* **2019**, *11*, 24906–24918.
- (68) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **2019**, *10*, 2903.
- (69) Von Rudorff, G. F.; Anatole von Lilienfeld, O. Simplifying inverse materials design problems for fixed lattices with alchemical chirality. *Science Advances* **2021**, *7*, 1–10.
- (70) Scott, J. LXXXII. The binding energy of the Thomas-Fermi Atom. The London, Edinburgh, and Dublin. *Philosophical Magazine and Journal of Science* **1952**, *43*, 859–867.
- (71) Schwinger, J. Thomas-Fermi model: The leading correction. *Phys. Rev. A* **1980**, *22*, 1827–1832.
- (72) Huang, B.; von Lilienfeld, O. A. Quantum machine learning using atom-in-molecule-based fragments selected on the fly. *Nat. Chem.* **2020**, *12*, 945–951.
- (73) Copyright IBM Corp. 1990–2023 and Copyright 1994–2001 by Max Planck Institute, Stuttgart., CPMD 4.3. <https://github.com/CPMD-code/CPMD> (accessed 30–01–2022).
- (74) Silvestri, W.; Seitsonen, A. P.; Boero, M.; Kohlmeier, A. CPMD2CUBE. <https://github.com/CPMD-code/Addons>, 2022 (accessed 30–01–2022).
- (75) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (76) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.
- (77) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641–3662.
- (78) Tange, O. GNU Parallel - The Command-Line Power Tool. *login: The USENIX Magazine* **2011**, *36*, 42–47.
- (79) Sun, Q.; et al. Recent developments in the PySCF program package. *J. Chem. Phys.* **2020**, *153*, 024109.
- (80) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework. *WIREs Computational Molecular Science* **2018**, *8*, e1340.
- (81) Sun, Q. Libcint: An efficient general integral library for Gaussian basis functions. *J. Comput. Chem.* **2015**, *36*, 1664–1671.
- (82) Wang, L.-P.; Song, C. Geometry optimization made simple with translation and rotation coordinates. *J. Chem. Phys.* **2016**, *144*, 214108.
- (83) Frisch, M. J.; et al. *Gaussian 09 Revision D.01*; Gaussian Inc., 2009.
- (84) 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.

- (85) Lemm, D.; von Rudorff, G. F.; von Lilienfeld, A. LERULI.com. *Online molecular property predictions in real time and for free.* www.leruli.com, 2021 (accessed 30–01–2022).
- (86) Landrum, G. RDKit. <https://www.rdkit.org/> (accessed 30–01–2022).
- (87) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB-An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
- (88) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *WIREs Computational Molecular Science* **2021**, *11*, e1493.
- (89) Harris, C. R.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362.
- (90) Virtanen, P.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272.
- (91) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; Von Lilienfeld, O. A.; Müller, K. R.; Tkatchenko, A. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326–2331.
- (92) Stewart, J. J. P. MOPAC2016, <http://openmopac.net> (accessed 30–01–2022).

Recommended by ACS

Extracting Quantitative Information at Quantum Mechanical Level from Noncovalent Interaction Index Analyses

Erna K. Wieduwilt, Alessandro Genoni, et al.

JANUARY 19, 2023

JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ 

Reducing the Quantum Many-Electron Problem to Two Electrons with Machine Learning

LeeAnn M. Sager-Smith and David A. Mazziotti

OCTOBER 04, 2022

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

READ 

Hydrogen-Atom Electronic Basis Sets for Multicomponent Quantum Chemistry

Irina Samsonova, Kurt R. Brorsen, et al.

JANUARY 24, 2023

ACS OMEGA

READ 

Quantum Tunneling in Peroxide O–O Bond Breaking Reaction

Yangyu Zhou, Mingfei Zhou, et al.

APRIL 18, 2023

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

READ 

Get More Suggestions >

Supporting Information

Supporting Information:

**Quantum alchemy based bonding trends and
their link to Hammett's equation and Pauling's
electronegativity model**

Michael J. Sahre,^{†,‡} Guido Falk von Rudorff,[¶] and O. Anatole von Lilienfeld^{*,§,||,⊥}

[†]Faculty of Physics, University of Vienna, Vienna, 1090, Austria

*[‡]University of Vienna, Vienna Doctoral School in Chemistry (DoSChem), Vienna, 1090,
Austria.*

[¶]University Kassel, Department of Chemistry, Kassel, 34132, Germany

[§]Vector Institute for Artificial Intelligence, Toronto, M5S 1M1, Canada

*^{||}Departments of Chemistry, Materials Science and Engineering, and Physics, University of
Toronto, St. George Campus, Toronto, M5R 0A3, Canada*

*[⊥]Machine Learning Group, Technische Universität Berlin and Institute for the
Foundations of Learning and Data, Berlin, 10587, Germany*

E-mail: anatole.vonlilienfeld@utoronto.ca

Bond Order

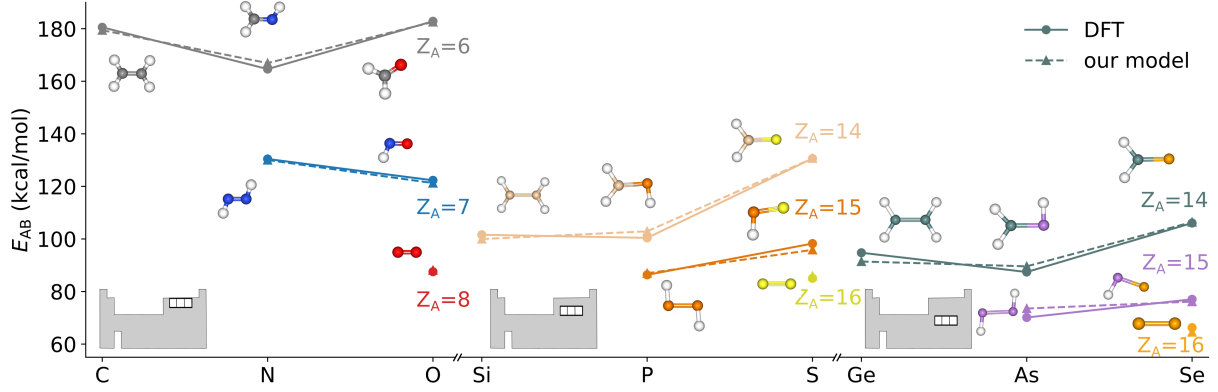


Figure S1: Calculated bond dissociation energies of double bonds from density functional theory (PBE0/def2-TZVP) and our quantum alchemy based chemical bond model (Eq. (2)). MAEs with respect to DFT amount to 0.9, 1.3 and 2.0 kcal/mol for the second, third and fourth row, respectively.

Table S1: Coefficients and MAEs of our model Eq. (2) for different rows n after refitting to ΔE^{TS} . b and c are scaled such that $bZ_A Z_B$ and $c(Z_A^{7/3} + Z_B^{7/3})$ are given in kcal/mol if Z_A and Z_B are given in atomic units of charge e .

n	a (kcal/mol)	b (630 Bohr ⁻¹)	c (630 Bohr ⁻¹ $e^{-1/3}$)	MAE (kcal/mol)
2	439.13	24.3984	4.7284	0.9
3	740.44	18.8836	3.2396	1.3
4	2475.12	14.2104	1.8714	2.0

Electronic Configuration

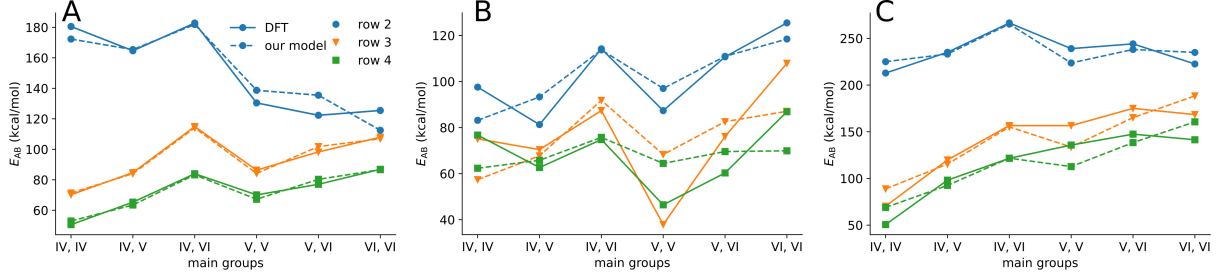


Figure S2: Energy differences between double bonded molecules $A=B$ and their fragments A , B with A and B from main groups IV, V, VI in the same row of the periodic table for various electronic states. Predictions are obtained from DFT (PBE0/def2-TZVP) and our model (Eq. (2)). All molecules and fragments are either in their electronic ground state (Panel A), the energetically lowest triplet state (Panel B) or the lowest singlet state (Panel C).

Table S2: Coefficients and MAEs of our model Eq. (2) for different rows n after refitting to ΔE^{GS} . b and c are scaled such that $bZ_A Z_B$ and $c(Z_A^{7/3} + Z_B^{7/3})$ are given in kcal/mol if Z_A and Z_B are given in atomic units of charge e .

n	a (kcal/mol)	b (630 Bohr $^{-1}$)	c (630 Bohr $^{-1}e^{-1/3}$)	MAE (kcal/mol)
2	367.73	19.7386	3.9376	7.4
3	371.53	12.4697	2.2692	1.6
4	874.76	6.7190	0.9318	1.9

Table S3: Coefficients and MAEs of our model Eq. (2) for different rows n after refitting to ΔE^{TT} . b and c are scaled such that $bZ_A Z_B$ and $c(Z_A^{7/3} + Z_B^{7/3})$ are given in kcal/mol if Z_A and Z_B are given in atomic units of charge e .

n	a (kcal/mol)	b (630 Bohr $^{-1}$)	c (630 Bohr $^{-1}e^{-1/3}$)	MAE (kcal/mol)
2	89.83	6.4716	1.7297	7.3
3	288.38	9.7959	1.7877	13.8
4	758.64	4.8042	0.6495	10.5

Table S4: Coefficients and MAEs of our model Eq. (2) for different rows n after refitting to ΔE^{SS} . b and c are scaled such that $bZ_{\text{A}}Z_{\text{B}}$ and $c(Z_{\text{A}}^{7/3} + Z_{\text{B}}^{7/3})$ are given in kcal/mol if Z_{A} and Z_{B} are given in atomic units of charge e .

n	a (kcal/mol)	b (630 Bohr $^{-1}$)	c (630 Bohr $^{-1}e^{-1/3}$)	MAE (kcal/mol)
2	333.59	17.6638	4.0305	8.2
3	80.28	8.2754	1.7258	12.9
4	0.00	3.4398	0.5523	12.6

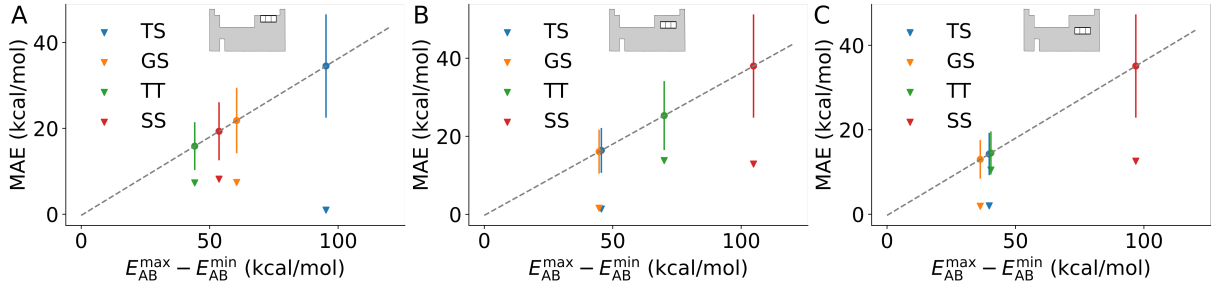


Figure S3: The performance of our model after re-calibration to binding energy differences of various electronic states in comparison to a fit to random data points from a uniform distribution. Triangles represent MAEs as a function of binding energy range after fitting our model to energy differences between fragments and double bonded molecules as triplets and singlets (TS), in their groundstate (GS), in the lowest triplet states (TT) and the lowest singlet states (SS). The circles with error bars show MAEs and standard deviation after fitting to random data points in the same energy ranges. Panels A, B, C show results for combinations of elements from main groups IV, V and VI for the second, third and fourth row of the periodic table, respectively.

Methods - Construction of the Model

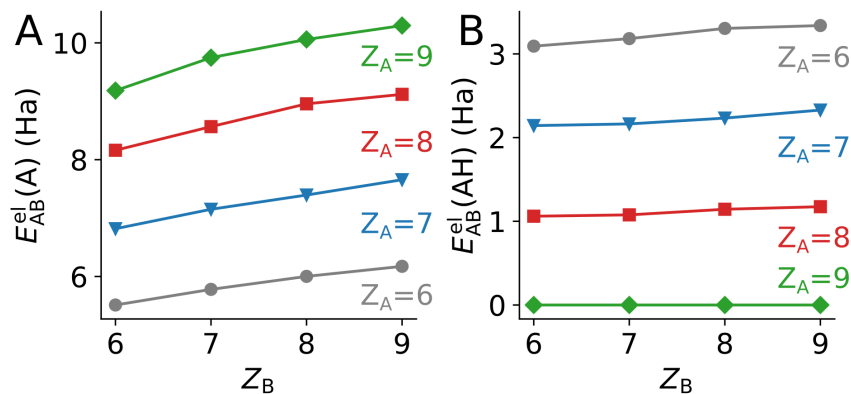


Figure S4: Atomic electronic binding energies (see Eq. (10)) of atom A as a function of the nuclear charge Z_B of the binding partner for carbon, nitrogen, oxygen and fluorine (Panel A) and the cumulated atomic binding energies of the hydrogens attached to the different atoms A (Panel B).

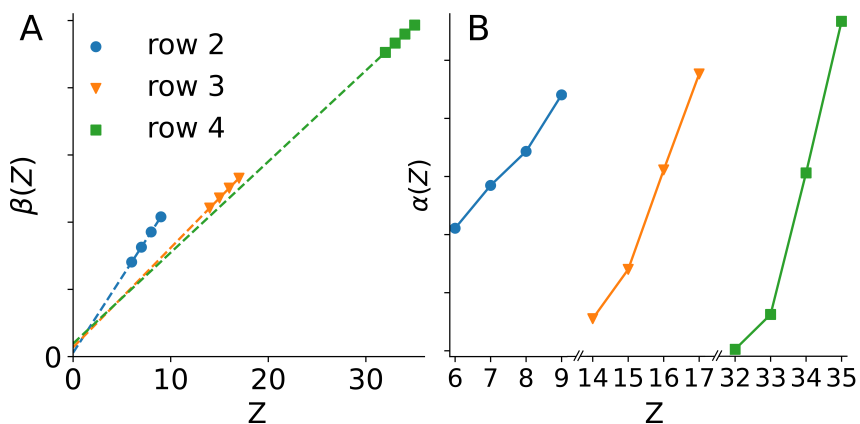


Figure S5: Optimized parameters β (Panel A) and α (Panel B) of Eq. (16) as a function of Z after a fit to bond dissociation energies calculated with DFT (PBE0/def2-TZVP).

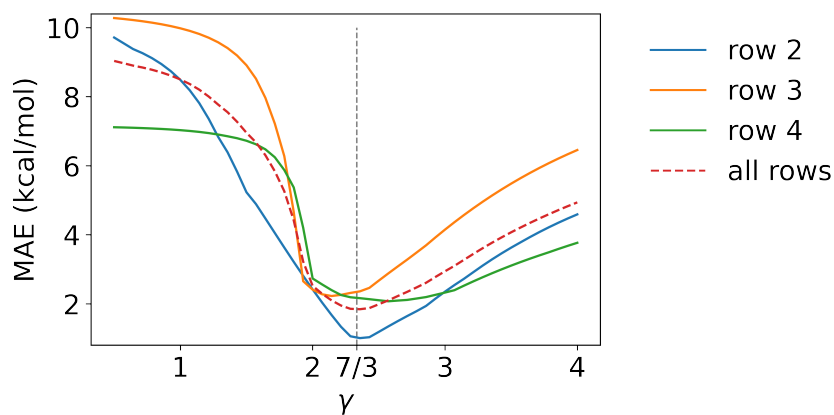


Figure S6: The mean absolute error for different values of the exponent γ in Eq. (22). The curves labeled row 2, row 3 and row 4 show the MAE if the parameters in Eq. (22) are optimized for each row individually. The dotted red curve shows the MAE for all three rows combined if γ is kept the same for all rows.

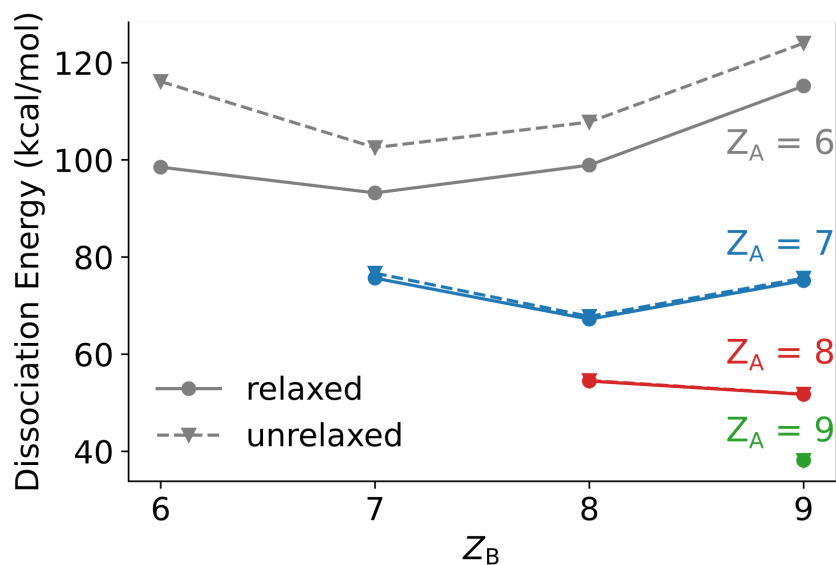


Figure S7: Calculated dissociation energies for homolytic bond cleavage of compounds H_xA-BH_y from density functional theory (PBE0/def2-TZVP). Solid lines depict dissociation energies with respect to geometrically relaxed fragments and dashed lines dissociation energies with respect to unrelaxed fragments.

Table S5: The average bond length $\bar{d}$ of A-B (in Bohr) and the optimized parameters (α, β) of Eq. (16) for the elements from main group IV to VII for different rows n of the periodic table. The unit of α is kcal/mol and β is given as $630 \cdot e\text{Bohr}^{-1}$ such that the unit of $\beta_A Z_A$ is also kcal/mol if Z_A is provided in atomic units.

n	$\bar{d}(\text{Bohr})$	IV	V	VI	VII
2	2.69	(55.39, 702.85)	(92.26, 813.36)	(121.54, 926.70)	(170.16, 1039.03)
3	4.01	(-22.70, 1102.97)	(19.71, 1177.84)	(105.29, 1251.12)	(187.89, 1324.77)
4	4.46	(-49.11, 2261.20)	(-19.08, 2330.51)	(102.73, 2397.54)	(233.19, 2464.38)

Trends and Dependence on Period - Eq.(7)

The $Z^{7/3}$ terms in our model can be approximated by a second order Taylor expansion as

$$Z^{7/3} \approx Z_0^{7/3} + \frac{7}{3}Z_0^{4/3}(Z - Z_0) + \frac{14}{9}Z_0^{1/3}(Z - Z_0)^2 \quad (1)$$

with Z_0 being a suitable reference nuclear charge. Insertion of this expression into our model leads to

$$\begin{aligned} E_{AB} &= a - bZ_A Z_B + c \left(\frac{4}{9}Z_0^{7/3} - \frac{7}{9}Z_0^{4/3}(Z_A + Z_B) + \frac{14}{9}Z_0^{1/3}(Z_A^2 + Z_B^2) \right) \\ &= \underbrace{a + \frac{4}{9}Z_0^{7/3}c}_{=\xi} - \underbrace{\frac{7}{9}Z_0^{4/3}c(Z_A + Z_B)}_{=\eta} - bZ_A Z_B + \underbrace{\frac{14}{9}Z_0^{1/3}c(Z_A^2 + Z_B^2)}_{=\mu}. \end{aligned} \quad (2)$$

The quadratic terms can be written as

$$\begin{aligned} Z_A^2 + Z_B^2 &= Z_A^2 + Z_B^2 + 2Z_A Z_B - 2Z_A Z_B \\ &= (Z_A - Z_B)^2 + 2Z_A Z_B \end{aligned} \quad (3)$$

leading to

$$\begin{aligned} E_{AB} &= \xi - \eta(Z_A + Z_B) - bZ_A Z_B + \mu([Z_A - Z_B]^2 + 2Z_A Z_B) \\ &= \xi - \eta(Z_A + Z_B) + (2\mu - b)Z_A Z_B + \mu(Z_A - Z_B)^2. \end{aligned} \quad (4)$$

The energy differences $\Delta E = E_{AB} - E_{AC}$ between A-B and A-C is then

$$\Delta E = \eta (Z_C - Z_B) - \kappa Z_A (Z_C - Z_B) + \mu ((Z_A - Z_B)^2 - (Z_A - Z_C)^2) \quad (5)$$

with $\kappa = (2\mu - b)$.

To account for the dependence of η, κ and μ on the row or principal quantum number n we make the ansatz $\eta = \tilde{\eta}(n-1)$, $-\kappa = (n-3)^2 - \tilde{\kappa}$, while we choose μ to be independent from the principal quantum number. Non-linear least squares optimization of $\tilde{\eta}$, $\tilde{\kappa}$ and μ with respect to ΔE in the resulting expression

$$\Delta E = \tilde{\eta}(n-1) (Z_C - Z_B) + [(n-3)^2 - \tilde{\kappa}] Z_A (Z_C - Z_B) + \mu ((Z_A - Z_B)^2 - (Z_A - Z_C)^2). \quad (6)$$

gives $\tilde{\eta} = 28$, $\tilde{\kappa} = 3.5$ and $\mu = 6$, so that

$$\Delta E = 28(n-1) (Z_C - Z_B) + [(n-3)^2 - 3.5] Z_A (Z_C - Z_B) + 6 ((Z_A - Z_B)^2 - (Z_A - Z_C)^2) \quad (7)$$

or alternatively

$$\Delta E = 28(n-1)\Delta Z + (8.5 + (n-3)^2) Z_A \Delta Z + 6(Z_B^2 - Z_C^2) \quad (8)$$

with $\Delta Z = Z_C - Z_B$.

Total Energies

Table S6: Total energies (in Hartree) of single bonded molecules and respective fragments after homolytic bond cleavage optimized with PBE0/def2-TZVP.

compound	energy	compound	energy	compound	energy
C ₂ H ₆	-79.750204	Cl ₂	-920.089409	BeHCH ₃	-55.172389
CH ₃ NH ₂	-95.777291	Ge ₂ H ₆	-4157.072524	BeHNH ₂	-71.250843
CH ₃ OH	-115.637262	GeH ₃ AsH ₂	-4315.341268	BeHOH	-91.138621
CH ₃ F	-139.653810	GeH ₃ SeH	-4480.425039	BeHF	-115.173176
N ₂ H ₄	-111.784922	GeH ₃ Br	-4652.430337	Li	-7.467048
NH ₂ OH	-131.622414	As ₂ H ₄	-4473.605476	BeH	-15.223816
NH ₂ F	-155.625694	AsH ₂ SeH	-4638.681646	BH ₂	-25.893672
H ₂ O ₂	-151.453106	AsH ₂ Br	-4810.679452	CH ₃	-39.796931
OHF	-175.439419	H ₂ Se ₂	-4803.749098	NH ₂	-55.832413
F ₂	-199.408535	SeHBr	-4975.731476	OH	-75.683320
SiH ₃ SiH ₃	-582.338252	Br ₂	-5147.706716	F	-99.673999
SiH ₃ PH ₂	-633.603800	BH ₂ CH ₃	-65.866689	SiH ₃	-291.107812
SiH ₃ SH	-689.861429	BH ₂ NH ₂	-81.963033	PH ₂	-342.378482
SiH ₃ Cl	-751.278691	BH ₂ OH	-101.826384	SH	-398.609973
P ₂ H ₄	-684.856580	BH ₂ F	-125.840622	Cl	-459.995117
PH ₂ SH	-741.101414	LiCH ₃	-47.336002	GeH ₃	-2078.478952
PH ₂ Cl	-802.505076	LiNH ₂	-63.410044	AsH ₂	-2236.760002
H ₂ S ₂	-797.331489	LiOH	-83.307236	SeH	-2401.827158
SHCl	-858.715428	LiF	-107.347917	Br	-2573.810130

Table S7: Total energies (in Hartree) of double bonded molecules and respective fragments after homolytic bond cleavage optimized with PBE0/def2-TZVP.

compound	energy singlet	energy triplet	compound	energy singlet	energy triplet
C ₂ H ₄	-78.512205	-78.380516	GeH ₂ Se	-4479.211917	-4479.162320
CH ₂ NH	-94.550022	-94.417719	H ₂ As ₂	-4472.399893	-4472.362319
CH ₂ O	-114.418666	-114.309740	AsHSe	-4637.469705	-4637.443125
H ₂ N ₂	-110.558763	-110.490383	Se ₂	-4802.511423	-4802.544133
NHO	-130.385650	-130.367261	CH ₂	-39.087164	-39.112821
O ₂	-150.170231	-150.230448	NH	-55.089690	-55.175865
Si ₂ H ₄	-581.083609	-581.041531	O	-74.908473	-75.015640
SiH ₂ PH	-632.371719	-632.324131	SiH ₂	-290.486074	-290.461159
SiH ₂ S	-688.639863	-688.571246	PH	-341.695472	-341.751186
H ₂ P ₂	-683.639421	-683.562517	S	-397.905241	-397.971393
PHS	-739.878544	-739.843432	GeH ₂	-2077.875509	-2077.840458
S ₂	-796.077755	-796.114103	AsH	-2236.092207	-2236.144296
Ge ₂ H ₄	-4155.831310	-4155.802722	Se	-2401.143488	-2401.203114
GeH ₂ AsH	-4314.123511	-4314.084186			