

Beyond Jacob's Ladder: A Physics-driven AI Approach to Density Functional Theory

Ekaterina Pas

`katya.pas@monash.edu`

Monash University <https://orcid.org/0000-0002-2506-4890>

Peter Halat

Monash University

Florencia Stella

Monash University <https://orcid.org/0009-0005-2333-6644>

Russhil Khurana

Monash University

Sophie Crouch

Monash University

Nicole Wang

Monash University

Physical Sciences - Article

Keywords:

Posted Date: September 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10727280/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Beyond Jacob's Ladder: A Physics-driven AI Approach to Density Functional Theory

Peter Halat^{1*}, Florencia Stella¹, Russhil Khurana¹, Sophie A. Crouch,¹ Nicole Wang,¹ Ekaterina I.
Izgorodina^{1*}

¹ School of Chemistry, Monash University, Clayton, Victoria, Australia.

*Corresponding authors. E-mails: peter.halat@monash.edu, katya.pas@monash.edu

Density Functional Theory (DFT) has become a widely used approach for accurate prediction of electronic structure and thermodynamic properties of molecular systems - from single molecules to large-scale systems. Although the accuracy of DFT functionals has significantly improved over the last decade, this is typically achieved through inclusion of more complexity in the functional's description and larger basis sets. This significantly prevents applications of DFT to realistic molecular systems beyond single medium-sized molecules. The recent development of AlphaFold has demonstrated AI algorithms' ability to learn physics, whereby achieving tremendous predictive power for the structure of peptides. Here we show that the inclusion of a more complex formulation in DFT can be overcome by training a machine learning (ML) code, termed δ ML-DFT, to extract the needed information from molecule's electron density using a neural network, whereby robustly learning essential physics governing the behaviour of molecular systems. The ML code was trained to predict atomisation energies of molecules of varying functionality and intermolecular interaction energies of diverse neutral and charged molecular complexes, producing accuracy below of 1 kcal/mol and outperforming the best available DFT functionals. This remarkable outcome was achieved by constructing the loss function based on relative electronic energies to directly compare with thermodynamic properties – atomisation energy and intermolecular interaction energy. The ability of the trained ML code to predict unseen molecular systems was demonstrated with its successful application to predict reaction energies and barriers of chemical reactions including radicals and molecules with strong electron correlation. Subsequent inclusion of a small portion of these reactions in the training process has significantly improved its performance. This work is the first report identifying how electron density of molecular systems can be the only input parameter requiring for a ML code to learn physics reliably. The proven concept can be easily extended to study complex chemical reactions, whereby enabling discoveries of new chemistries, and providing an invaluable computational tool that can be used by empowering scientists in general to pioneering such discoveries.

Main text

DFT has become a widely used electronic structure approach to predict electronic energy, $E(\rho)$, and physicochemical properties of molecular systems, including thermodynamics of chemical reactions and strength of intermolecular interactions. Accuracy of individual DFT functionals relies on approximations to formulate the exchange-correlation (XC) potential (Eq. 1) whose true analytical form is yet to be identified.

$$E_{tot}(\rho) = \hat{T}_e(\rho) + J(\rho) + E_{eN}(\rho) + E_{XC} \quad (1)$$

Jacob's ladder proposed by Perdew¹ summarises several decades of development in DFT functionals presented as rungs. Apart from the common parameter – electron density (ρ) – the rungs on the ladder are characterised by inclusion of its derivatives starting from the second rung containing functionals with electron density gradient such as BLYP and PBE referred to as GGA (general gradient approximation) functionals.²⁻⁴ The fourth rung contains accurate meta-GGA functionals such as M06-2X and ω B97XD incorporating the exact Hartree-Fock (HF) exchange and kinetic energy density (τ). The improved performance of meta-GGA functionals over GGA functionals comes with increased computational cost due to the use of large basis sets and SCF (self-consistent field) convergence issues in larger molecular systems. At the dawn of the DFT field, independence of the exchange and correlation terms (Eq. 1) was adopted, mainly due to significant simplifications that it introduced in subsequent implementation of DFT functionals. From a fundamental point, the exchange and correlation terms are likely to be dependent on each other and therefore, their treatment as a combined effect has been largely unexplored in DFT.

ML models can bypass the complexity of formulating the XC potential by identifying

non-linear patterns between molecular structure parameters and the predicted property. In particular, electron density of a molecule has been used to generate robust descriptors for ML models by projecting the density onto atom-centred orbitals.^{5,6} In the work of Dick *et al.* this approach was used to successfully correct the baseline DFT electronic energies, thus significantly reducing the cost of DFT calculations.⁷ In the work by Bogojeski *et al.* electron density-derived descriptors were used to achieve chemical accuracy (1 kcal/mol) for the prediction of raw electronic energies of small organic molecules such as water and ethanol calculated with the “gold” standard benchmark method in computational chemistry, CCSD(T), a coupled cluster method including single, double, and perturbative triple excitations.⁸ Rucsko *et al.* demonstrated that self-consistent charge densities calculated with already published XC potentials could also serve as inputs for ML models to predict components of total electronic energy in semi-local DFT functionals.⁹ ML approaches have been shown to enable robust parameterisation of the already existing exchange and correlation potentials. Ryabov *et al.* showed how electron density together with its gradient and the second derivative could be used to parameterise an exchange-correlation (XC) potential using a neural network (NN) approach¹⁰ by constructing the loss function to minimise the error in total electronic energies of atoms and isomerisation reaction energies of small organic and inorganic molecules. Dick and Fernandez-Serra used NNs to create a series of functionals termed NeuralXC, whose design was aimed to provide a correction to the baseline DFT functional.¹¹ Each NeuralXC was individually trained for a given dataset of molecules (e.g. short alkanes) by constructing the loss function to predict electronic energies from CCSD(T). The trained NeuralXC functionals showed limited transferability to larger and more complex molecular systems and revealed divergence from asymptotic constraints

satisfied in conventionally designed DFT functionals. Although examples of inclusion of physical constraints in the NN-trained XC potentials^{12,13} start to emerge, these developments are yet to be applied across a broad range of chemical systems and their physicochemical properties. For example, in the work of Nagai et al.¹² the NN training was performed with respect to CCSD(T) electronic energies of small molecules such as water and ammonia. The trained model produced a mean absolute error of 3.6 kcal/mol and a standard deviation of 4.5 kcal/mol for 144 of small organic molecules,¹⁴ with benzene being the largest molecule in the dataset, further highlighting the limited transferability of the model.

Many of the reported ML models using electron density as input have aimed at minimizing errors in raw electronic energies of molecules calculated with CCSD(T), widely accepted as the “gold” standard method for electronic structure property prediction. It must be noted that CCSD(T) is a non-variational method in its formulation due to the perturbative triple excitation contribution and therefore, electronic energies may appear randomly above and below the exact electronic energy. Therefore, ML training of the XC potential with respect to CCSD(T) electronic energies could result in inconsistent outcomes, which may explain why the reported ML models to-date have limited transferability to more complex molecular systems.

The most recent attempt to use deep-learning approaches to learn the exchange-correlation potential by utilising electron density and its two components – gradient and kinetic energy density – is SKALA developed by Microsoft Research and Microsoft Quantum.¹⁵ SKALA’s architecture still uses a generic formula for the exchange-correlation potential incorporating basic physical constraints. The approach uses the well accepted enhancement factor to train the local component of electronic density using ML. Additionally, non-local representations of

electronic structure are introduced directly from data via a two-stage training protocol. The training relies on the availability of a reasonable approximation of the ground state density, which is taken from B3LYP calculations. For these reasons, the computational cost of SKALA falls in between a GGA and a meta-GGA hybrid DFT functional. For the GMTKN 55 benchmark database,¹⁶ SKALA only slightly outperforms a state-of-the-art DFT functional, ω B97M-V, by 0.43 kcal/mol on average as measured through weighted total mean absolute deviations (WTMADs).

In this work we have showcased how to use electron density and its features – gradient and kinetic energy density – to teach needed physics to the ML code to design a robust XC potential that can achieve chemical accuracy for electronic structure properties, competing in its performance with the best DFT functionals. The hypothesis of the study stemmed from the second Hohenberg-Kohn theorem identifying the one-on-one correspondence between the electron density and the electronic energy, which can be further extrapolated suggesting that *electron density has all the information needed about molecular systems to predict their electronic energy*. In this regard, we show that even electron density at the lowest level of the DFT functional formulation such as Local Density Approximation (LDA) is enough to train the XC potential.

There are three important distinctions that separate our proposed approach from the published ones. Firstly, the XC potential was not biased by any preconceived formulation and was treated as one entity during the training process, allowing the code to learn the interrelationship between the exchange and correlation components. Secondly, no atom’s identity (number of electrons and charge of nucleus) or its spatial position were passed to the neural network and the ML code learned physics by connecting the points on the electron density surface to the XC potential and finally

to the total electronic energy. Thirdly, the loss function was constructed with respect to the two selected electronic structure properties – AE and EINT – to facilitate the learning process. The proposed approach is agnostic to atomic identities and their connectivity, making it universal so long as all elements from the Periodic Table are represented in the training dataset.

ML Model Training. In our work we adopted a two-tiered approach in training of the XC potential, whereby achieving the robust prediction of electronic structure properties. The workflow of the adopted approach is illustrated in Figure 1. In the first tier, LDA was used to generate four input ingredients such as electron density, its gradient, kinetic energy density and spin polarisation that were subsequently converted into dimensionless quantities to satisfy a meta-GGA approximation (for more detail see Methods). The input into the ML model was a matrix consisting of four columns, representing these four dimensionless quantities (electron density, its gradient, kinetic energy density and spin polarisation), with respect to Cartesian coordinates on the electron

density (Figure 1). Depending on the size of the molecule, the number of points represented in each matrix varied between 5,000 and 365,000, with the median value of 107,000. One-electron energies (kinetic energy of non-interacting electrons, Coulomb energy of electrons and potential energy of electrons interacting with nuclei) were calculated using *PySCF* and remained unchanged throughout the training process. Unlike in the conventional formulation of DFT functionals, the XC energy was treated as one entity without any preconceived formulation. The input matrix was passed to a neural network consisting of three layers containing 35, 37 and 196 nodes (obtained through hyperpametrised) to predict the XC energy for each electron density point in the input matrix. Total electronic energy of a molecule or an atom was then calculated as the sum of the XC energy for all electron density points and augmented with the one-electron energies (Figure 1). In the second tier, the predicted total electronic energies of molecules and atoms were used to calculate two electronic structure properties - atomisation energies (AEs) and interaction energies (EINTs). The

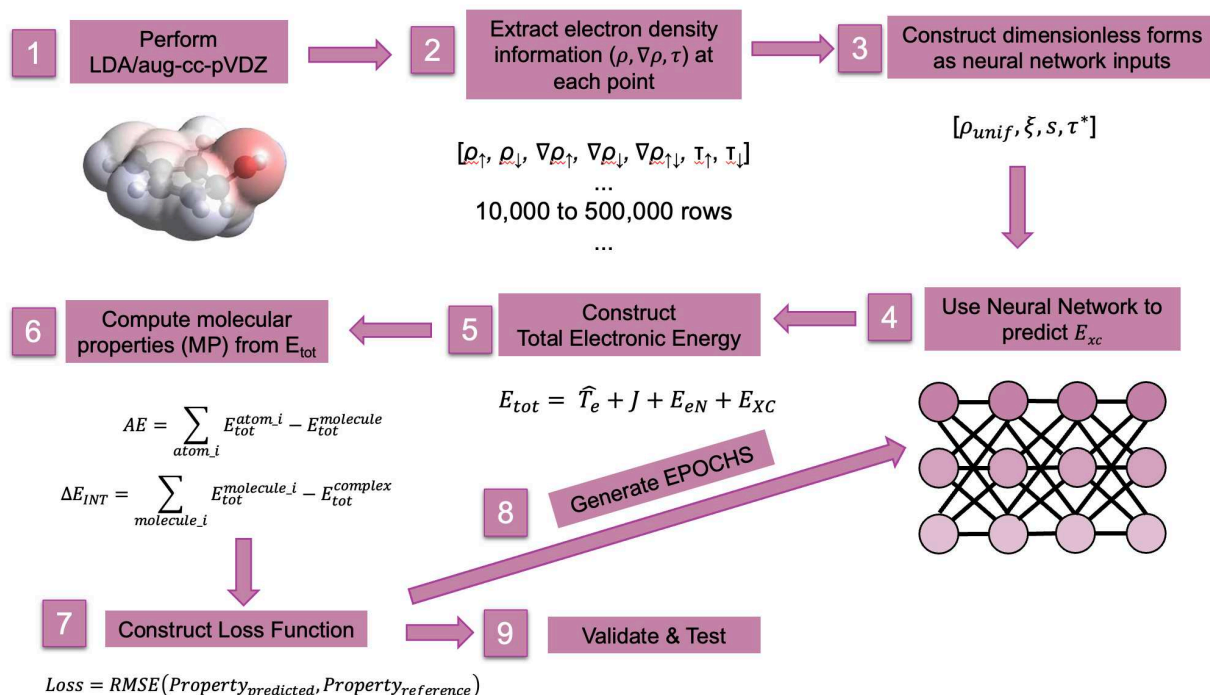


Figure 1. Schematic of the proposed δ ML-DFT approach.

loss function was constructed to minimise absolute errors in these predicted properties with respect to corresponding reference energies.

The combined dataset of 7,351 data points for AEs and 1,324 data points for EINTs was divided into three subsets: training, validation, and testing (Figure 2a), with both properties being balanced in each subset with regards to the total number of the molecular systems for each property. Training was repeated until the best (*i.e.* minimum) score, defined as the diagonal between the model's performance for the training and validation subsets (Figure 2b), was achieved with no significant improvement after 200 epochs. In general, the model's performance was found to stabilise over 2500 epochs, with 3000 epochs representing the upper bound to observe the best score falling within chemical accuracy of 1 kcal/mol (Figure 2b).

The ML model achieved the best performance when the combined dataset was randomly divided into 65% of training, 17.5% of validation and 17.5% of testing, ensuring the balanced presentation of both AE and EINT properties in each subset (Figure 2a). Mean absolute errors (MAEs), root mean square

errors (RMSEs) and maximum errors for the training, validation and test subsets are presented in the supplementary information (Table S1).

Model's performance for Atomisation Energies. The model performed remarkably well for the QM7b database¹⁷ with an RMSE of 0.61 kcal/mol (Figure 2c). This is coupled with the fact that a maximum error of 4.4 kcal/mol was recorded for a complex molecule containing a C≡C bond and N=N bond. These results are particularly reassuring confirming that when the element is abundantly present in the training set, our model can learn physics without being told the element's identity and its environment beforehand.

The performance of the model was slightly reduced with respect to the TAE140 dataset,¹⁸ with a wider error distribution as shown in Figure 3, exhibiting an increased RMSE to 3.92 kcal/mol. Apart from the atoms such as C, N, O, S and H commonly observed in organic molecules, the dataset contains molecules consisting of elements such as Be, B, Al, Si *etc.* from the second and third row elements of the Periodic Table usually observed in inorganic molecules. The uniqueness of the dataset comes from the fact that such atoms are only

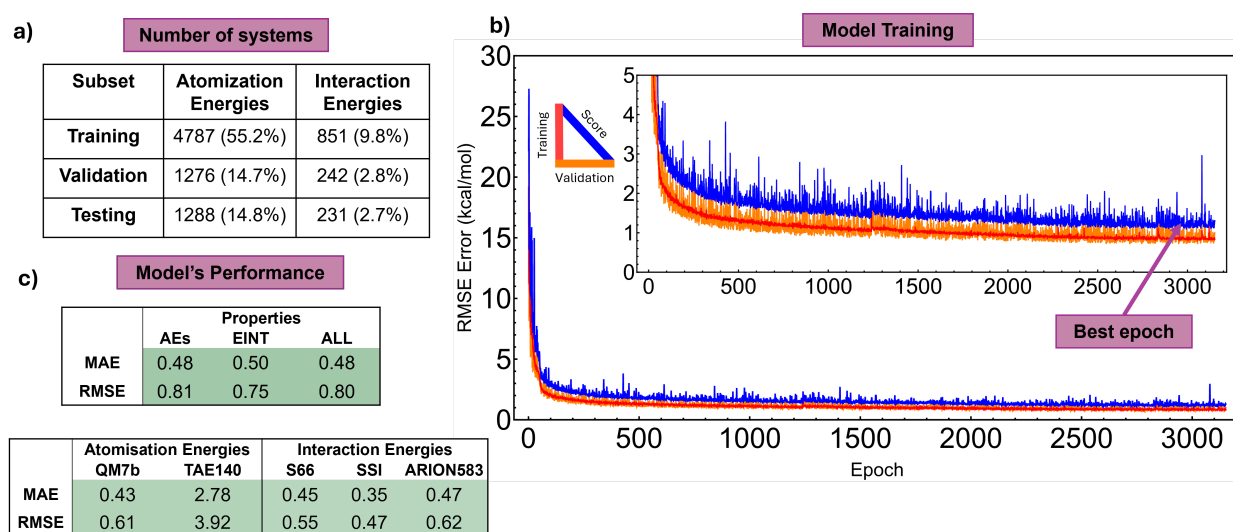


Figure 2. a) Number of systems included in the training, validation, and test sets for the two predicted properties – AEs and EINTs; b) an example of the model's training over 3000 epochs. The inset further elucidates how the model improves after 2500 epochs; c) Model's performance for AEs and EINTs overall and select datasets. All errors are given in kcal/mol.

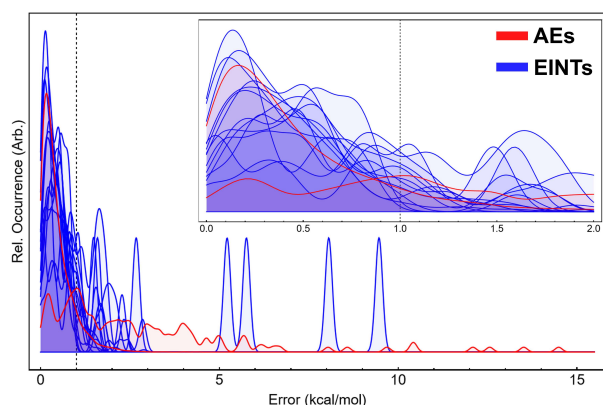


Figure 3. Distribution of absolute errors (in kcal/mol) in the used datasets. Distributions are built from combined training, validation, and test dataset results.

represented in two molecules on average with a very low percentage $<0.1\%$ of occurrence across the entire datasets used (Figure S1). Therefore, it is not surprising that our model does not manage to learn physics adequately for such elements. None of the molecular systems were treated separately by dividing them into single- and multi-reference systems as is regularly performed for assessment of DFT functionals.¹⁹ The largest errors of 14.5 kcal/mol and 13.5 kcal/mol came from SiF_4 and S_4 , respectively, with the latter being a multi-reference (MR) system. Other largest deviations came from a MR system such as BN (12.6 kcal/mol) and ClO_2 (12 kcal/mol). For MR systems whose atoms had a larger occurrence in the dataset the errors were significantly lower. For example, for MR systems such as C_2 , and F_2O , absolute errors were lower at 6.5 and 2.4 kcal/mol, respectively. This finding can be explained by the varied environment of the C, F and O atoms included in the training subset, allowing the code to learn physics associated with these elements' behaviour in molecular systems, regardless of their multireference character. The TAE140 dataset was purposefully included in this work to demonstrate the need to strike a balance between the element occurrence in molecular systems included in the training.

Model's performance for Interaction Energies. The model performed very well for intermolecular complexes consisting of at least two molecules interacting non-covalently. An RMSE of 0.80 kcal/mol was achieved across 14 datasets. Commonly used datasets such as S66,²⁰ SSI²¹ and the newly designed dataset ARION583²² produced an RMSE of 0.55, 0.47 and 0.62 kcal/mol. This observation is accompanied with the largest errors of 1.44 kcal/mol for the dimer of neopentane in S66, 2.88 kcal/mol for the dimer of aspartate in SSI and 2.52 kcal/mol of the complex between methoxypridine and Na^+ in ARION583. The model performed equally well for the S22TRM²³ consisting of systems of three organic molecules interacting in the same fragment. For this database, an RMSE of 1.10 kcal/mol and a maximum error of 2.16 kcal/mol for the formic acid trimer was observed. The largest errors in ARION583 came from complexes of organic molecules interacting with alkali metal ions. This is also reflected in the CHB6²⁴ dataset, where the largest errors of 9.5 kcal/mol and 8.1 kcal/mol were recorded among all 14 datasets for the intermolecular complexes between benzene and the Li^+ and Na^+ cations, respectively. In general, complexes containing alkali metal ions have diverse nature ranging from partially covalent complexes in the case of Li^+ complexes to electrostatically driven Na^+ complexes with a growing dispersion component in K^+ complexes.^{25,26} Considering a small number of complexes (only 55) containing Li^+ , Na^+ and K^+ cations in these two datasets, it is not surprising that the model does not have enough systems to accurately learn about these metals.

Among other databases that produced errors above chemical accuracy, there are three that are worth mentioning - CARBH12,¹⁶ PNICO23²⁷ and Hal59.^{28,29} These datasets contain elements in uncommon environment or underrepresented elements such as iodine. For example, most of the molecular systems considered contained carbon atoms engaged in single, double, and triple bonds, whereas CARBH12 consists of molecular complexes between model carbenes such as CH₂ and dimethylimidazole interacting with small molecule hydrogen donors such as HCl and NH₃. Despite the, the largest errors are still relatively small and are observed for HCl interacting with dimethylimidazole (2.3 kcal/mol), whereas the rest of the system produced errors below 1 kcal/mol. This finding further confirms that the code can robustly delineate between a varying carbon environment. PNICO23 consists of dimers of substituted amines and phosphines exhibiting pnictogenic bonds. Largest error of only 1.6 kcal/mol was observed for dimer PH₂F. The rest of the complexes had errors below 1 kcal/mol. This is an outstanding outcome considering that only 23 such complexes/dimers were included in the dataset. For the Hal59 dataset consisting of only 23 iodine-containing complexes, the largest errors were observed for two complexes of Br₂ interacting with trimethylamine (2.4 kcal/mol) and pyridine (2.2 kcal/mol). The errors are still

relatively small and 90% of the systems had errors below chemical accuracy.

Testing the Model for Unseen Reaction Energies and Barriers. Six datasets containing reaction energies (REs) and four datasets containing reaction barriers (RBs) were used to test the performance of the trained model. Interestingly, despite the complexity of the included reactions, the model produced an RMSE of 5.74 kcal/mol and 5.10 kcal/mol for REs and RBs, respectively. Inclusion of randomly selected systems in the training dataset – amounting to 63% of the RE systems and 65% of the RB systems – significantly improved the performance of the model, reducing the RMSEs to 1.96 kcal/mol for REs and 1.96 kcal/mol for RBs (for more detail see Table S2). For the systems in the test datasets, RMSEs of 1.94 and 2.43 kcal/mol were recorded for REs and RBs, respectively, demonstrating the model’s ability to learn diverse chemical bonding environments. The largest errors of 7.60 and 6.3 kcal/mol were recorded for complex isomerization reactions in ISOL24^{24,30} - numbers 4 and 18 in the original paper, respectively. One of the problems with more complex databases is the quality of the reference method as the “gold” standard method of CCSD(T) may not be computationally feasible. For example, for ISOL24 the DLPNO-CCSD(T) method was used to generate reference reaction energies due

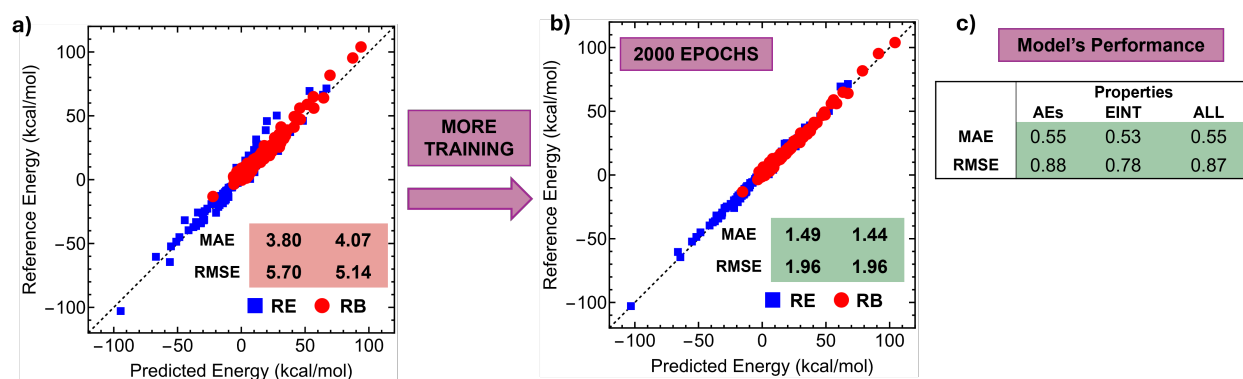


Figure 4. a) Application of the trained δ ML-DFT model to reaction energies (REs) and reaction barriers (RBs), b) Performance of the δ ML-DFT model for all RE/RB systems after retraining of the model with the addition of 63% of REs and 65% of RBs and c) Model’s performance for AEs and EINTs after the retraining. All errors are given in kcal/mol.

to the size of organic molecules, which tends to generate larger errors for systems with increased dispersion interactions.³⁰

The inclusion of more data did not change the overall performance of the retrained model, with the QM7b, S66 and SSI datasets retaining RMSEs below 1 kcal/mol of 0.66, 0.60 and 0.95 kcal/mol, respectively. Two datasets, AHB21²⁴ and CHB6, improved their performance by 0.27 and 0.43 kcal/mol for RMSEs, respectively. The largest error for the datasets with EINTs reduced from 9.5 kcal/mol to 8.2 kcal/mol. The TAE140 dataset had a slightly increased maximum error to 22.8 kcal/mol (ClO₂). Overall, the dataset retained its RMSE of 0.87 kcal/mol (combined AEs and EINTs), further confirming that addition of more systems into the training set only reinforces excellent performance of the model.

Comparison with other DFT functionals. The model outperformed the most used and best performing DFT functionals such as PBE-D3,^{4,31} M06-2X,³² SCAN,³³ ω B97X-D³⁴ and ω B97M-V³⁵ for the TAE140 dataset and the 14 datasets containing EINTs (Figure 5). The model produced lower RMSEs and maximum errors for all. The performance for EINTs is particularly impressive, reaching the overall RMSE below 1 kcal/mol compared to that of 2.1 kcal/mol for ω B97X-D, the best performing DFT functional. For TAE140 the maximum error was by far the lowest among the select DFT functionals, whose maximum errors fell in the range of 28.8 kcal/mol for M06-2X and 54.4

a)	TAE140	ML-DFT	PBE-D3	SCAN	ω B97X-D	M06-2X	ω B97M-V
	RMSE	3.9	19.3	5.5	5.4	4.9	4.2
	Max.	14.5	54.4	27.7	35.7	28.8	31.2

b)	EINT	ML-DFT	PBE-D3	SCAN	ω B97X-D	M06-2X	ω B97M-V
	RMSE	0.8	2.7	2.5	2.1	2.7	2.5
	Max.	9.5	14.5	12.4	11.4	15.2	14.5

Figure 5. Comparison of the δ ML-DFT model's performance with respect to DFT functionals for a) TAE140 and b) EINTs. All errors are given in kcal/mol.

kcal/mol for PBE-D3, thus demonstrating model's versatility in capturing complex chemical environments. As the QM7b dataset uses ω B97X-D reference values, performance could not be compared.

The reported δ ML-DFT model represents the next evolutionary step in developing accurate and cost-effective DFT functionals using physics-driven ML approaches. Electronic density of molecular systems even at a low level of DFT was shown to carry enough electronic structure information. This quantity alone enabled a neural network to predict the exchange-correlation energy and achieve high accuracy for interaction energies and atomisation energies even for chemical systems with a multi-reference character. This work paves the way to how the exchange-correlation potential can be learned using ML without any bias arising from functional bias and atomic identity and connectivity. Opportunities to improve on the currently presented approach are endless, opening a new chapter in cost-effective and highly accurate methods in computational chemistry.

Methods

Local Spin Density Approximation (LSDA), specifically Slater-VWN5^{36,37}/aug-cc-pVDZ⁵ calculations were performed on each chemical system, using PySCF^{6,7} to extract molecular electronic density for training, validation and testing purposes. The 'grids.level' was set to 2, with 'RKS' and 'UKS' algorithms loaded for spin-paired and spin-unpaired systems, respectively. For the vast majority of systems, the '.run()' protocol was sufficient for SCF convergence, and '.newton()' was implemented when required.

Density matrices (defined as 'dm') were obtained via the 'make_rdm1()' function, and used to obtain the non-interacting Coulomb energy term (E_{coul}) by calling 'get_veff(dm=dm).ecoul'. The 1-electron energy term (E_{1e}) was extracted by first generating the core Hamiltonian (hcore) via 'get_hcore()' and subsequently computed through an Einstein sum across the density matrix: 'np.einsum("ij,ji->", hcore, dm)'. For spin-unpaired systems, 'dm' was replaced by 'dm[0] + dm[1]' being the sum of both spin up and spin down densities. Finally, the nuclear repulsion energy (E_{nuc}) was obtained through the 'energy_nuc()' function. The three components, E_{coul} , E_{1e} and E_{nuc} , comprise the non-interacting electronic energy term, E_s .

$$E_s = E_{coul} + E_{1e} + E_{nuc}$$

E_s was recorded for all chemical systems and electron density weights were called via the 'grids.weights' command. Before electron density could be used as input to train the proposed approach, it had to be converted to dimensionless entity as described below:

1. Atomic orbitals (defined as 'ao') were calculated using 'numint.eval_ao(mol, coords, deriv=2)', where 'mol' represented the molecular system and 'coords' represented electron density point coordinates obtained by the 'grid.coords' command. For spin paired systems, electron density, defined as ρ , and its components - electron density gradient components, defined as $\frac{\delta\rho}{\delta x}$, $\frac{\delta\rho}{\delta y}$, $\frac{\delta\rho}{\delta z}$, electron density

Laplacian, defined as $\nabla^2\rho$ and τ - were calculated for each electron density point

via 'numint.eval_rho(mol, ao, dm, xctype="mgga")'. For unpaired-spin systems, ρ , $\frac{\delta\rho}{\delta x}$, $\frac{\delta\rho}{\delta y}$, $\frac{\delta\rho}{\delta z}$, $\nabla^2\rho$ and τ were calculated from both spin up and spin down electron density matrices, using 'dm[0]' and 'dm[1]', respectively.

2. The electron density components were reshaped into 7 term vectors, $[\rho_\uparrow, \rho_\downarrow, \sigma_\uparrow, \sigma_\downarrow, \sigma_{\uparrow\downarrow}, \tau_\uparrow, \tau_\downarrow]$, which were defined separately for spin-paired and spin-unpaired systems, where $\uparrow$ and $\downarrow$ represented up and down spin, respectively. For spin-paired systems the terms are defined as:

$$\rho_\uparrow = \rho_\downarrow = \frac{1}{2}\rho$$

$$\sigma_\uparrow = \sigma_\downarrow = \sigma_{\uparrow\downarrow} = \frac{1}{4} \left[\left(\frac{\delta\rho}{\delta x} \right)^2 + \left(\frac{\delta\rho}{\delta y} \right)^2 + \left(\frac{\delta\rho}{\delta z} \right)^2 \right] + 10^{-7}$$

$$\tau_\uparrow = \tau_\downarrow = \frac{1}{2}\tau$$

For spin-unpaired systems, $\rho_\uparrow$, $\rho_\downarrow$, $\tau_\uparrow$, and $\tau_\downarrow$ are left unchanged, and the electron density gradient terms are defined as:

$$\sigma_\uparrow = \left[\left(\frac{\delta\rho_\uparrow}{\delta x} \right)^2 + \left(\frac{\delta\rho_\uparrow}{\delta y} \right)^2 + \left(\frac{\delta\rho_\uparrow}{\delta z} \right)^2 \right] + 10^{-7}$$

$$\sigma_\downarrow = \left[\left(\frac{\delta\rho_\downarrow}{\delta x} \right)^2 + \left(\frac{\delta\rho_\downarrow}{\delta y} \right)^2 + \left(\frac{\delta\rho_\downarrow}{\delta z} \right)^2 \right] + 10^{-7}$$

$$\sigma_{\uparrow\downarrow} = \left[\frac{\delta\rho_\uparrow}{\delta x} \cdot \frac{\delta\rho_\downarrow}{\delta x} + \frac{\delta\rho_\uparrow}{\delta y} \cdot \frac{\delta\rho_\downarrow}{\delta y} + \frac{\delta\rho_\uparrow}{\delta z} \cdot \frac{\delta\rho_\downarrow}{\delta z} \right] + 10^{-7}$$

10^{-7} is added to density gradients for numerical stability, to avoid values of zero being fed into logarithmic functions.¹³ An example PySCF input script is provided in the GitHub link (<https://github.com/PHalat/DeltaML-DFT>).

3. A separate Python script was written to perform machine learning training on the recorded electron density vectors. The electron density and its components were transformed into dimensionless representations $[\rho_{unif}, \xi, s, \tau^*]$, derived from as follows:^{12,38}

$$\begin{aligned}
\rho_{unif} &= (\rho_{\uparrow} + \rho_{\downarrow} + 10^{-7})^{\frac{1}{3}} \\
\chi &= \frac{\rho_{\uparrow} - \rho_{\downarrow}}{\rho_{\uparrow} + \rho_{\downarrow} + 10^{-7}} \\
\xi &= \frac{1}{2} \left[(1 + \chi)^{\frac{4}{3}} + (1 - \chi)^{\frac{4}{3}} \right] \\
s &= \frac{(\sigma_{\uparrow} + \sigma_{\downarrow} + 2\sigma_{\uparrow\downarrow})^{\frac{1}{2}} + 10^{-7}}{(\rho_{unif})^4} \\
\xi_{\tau} &= \frac{1}{2} \left[(1 + \chi)^{\frac{5}{3}} + (1 - \chi)^{\frac{5}{3}} \right] \\
\tau^* &= \frac{\tau_{\uparrow} + \tau_{\downarrow} + 10^{-7}}{(\rho_{unif})^5 \xi_{\tau}}
\end{aligned}$$

where the 10^{-7} terms are again added for numerical stability, particularly to avoid near-zero denominators, in addition to avoiding zero values for logarithmic functions.¹³ For each electron density point, the dimensionless vectors act as inputs to a neural network, designed to output the exchange-correlation energy at each electron density point.

- The neural network was constructed to have alternating layers of linear nodes, and softplus activation functions ($\ln(1 + e^x)$) before outputting a single scalar value, which was called $NN([\rho_{unif}, \xi, s, \tau^*])$. The number of layers and nodes within the neural network was optimised through hyperparameterisation, searching between 2 and 7 layers, and 1 and 200 nodes per layer. The learning rate was also optimised, scanning between 10^{-3} and 10^{-1} , and the Adam³⁹ optimiser was used for the δ ML-DFT model training. The scalar output was passed through another softplus function to enforce a negative value, and multiplied by ρ_{unif} and ξ to give the exchange correlation contribution of that point, NN_{xc} :

$$\begin{aligned}
NN_{xc}([\rho_{unif}, \xi, s, \tau^*]) &= \\
&-\text{softplus}(NN([\rho_{unif}, \xi, s, \tau^*])) * \rho_{unif} * \xi
\end{aligned}$$

- The δ ML-DFT model's output was multiplied by the electron density at each point, and the point's weight, w , to give the resulting exchange-correlation energy of the point. Taking the sum across all points of a chemical system

produced the exchange-correlation energy of the chemical system:

$$E_{xc}^{pred} = \sum_{i \in \text{points}} NN_{xc}(i) * (\rho_{\uparrow}(i) + \rho_{\downarrow}(i)) * w(i)$$

- Adding E_{xc} to E_s gave the total predicted electronic energy of the chemical system, which was then used to calculate electronic structure properties such as atomization energy, interaction energy, reaction energy and reaction barrier where required. The loss function, or *Score*, given to the neural network in its training was constructed from these electronic structure properties using the Euclidean norm of training and validation dataset losses, L :

$$\begin{aligned}
\text{Score} &= \\
&\sqrt{(L(\text{training}))^2 + (L(\text{validation}))^2}
\end{aligned}$$

where the loss functions, L , of training and validation datasets were the mean of root mean squared deviations of batches of reference values. The batch size was set to 10 in all training runs performed.

- Training was performed until no improvement in the *Score* was observed for at least 200 epochs. For the results presented in the manuscript, a training proportion constituted 65% of the combined AEs and EINTs datasets, leaving 17.5% to be allocated each into validation and test datasets. Training proportions between 10% and 80% were tested, with results provided in the Supplementary Information. Training, validation and test subsets were determined randomly, across the entire combined AEs and EINTs datasets. This was performed through `sklearn.model_selection.train_test_split`.

DFT Comparison. Comparative calculations were performed in TAE140 and all interaction energy databases featured in Table 1 with select DFT functionals such as PBE-D3,^{4,31} M06-2X,³² SCAN,³³ ω B97X-D³⁴ and ω B97M-

V³⁵. All calculations were performed using ORCA 6.0.1,^{40,41} with both aug-cc-pVTZ⁵ and def2-TZVPP⁴² basis sets. As aug-cc-pVTZ isn't available for potassium, bromine and iodine, 6-31++G**⁴³ and aug-cc-pVTZ-PP basis sets were substituted for these alkali metals and halogens, respectively

Datasets

Table 1 outlines the datasets featured in this work, as well as their reference levels of theory with citations. Duplicate systems between S22 and S66 were removed from S66.²⁰ A subset of 299 systems of SSI were included, representing an even mix of dispersion-driven, electrostatically-driven and mixed-interaction systems, as determined by SAPT calculations.

Table 1: An overview of datasets used for training, validation and attesting of the δ ML-DFT model. Levels of theory marked with an asterisk denote values calculated in this work. Abbreviated references for each database is given as a footnote.

Property	Database	Description	Level of Theory
Atomization Energy (AE)	QM7b ^a	Small-medium neutral organic molecules	ω B97X-D3/aug-cc-pVQZ*
	TAE140 ^b	Small neutral organic and inorganic molecules	W4
Interaction Energy (EINT)	ADIM6 ^c	Neutral alkane dimers	W1-F12
	AHB21 ^d	Small anion-neutral dimers	CCSD(T)/CBS CCSD(T)-F12/CBS (HCOO ⁻)
	ARION583 ^e	Medium aromatic neutral-ion dimers	SRS-MP2/cc-pVTZ
	CARBHB12 ^f	Small carbene hydrogen bonding complexes	W2-F12
	CHB6 ^d	Small cation-neutral dimers	CCSD(T)/CBS
	HAL59 ^{g,h}	Halogen bonded systems and dimers	CCSD(T)/CBS
	IL16 ^d	Small anion-cation dimers	CCSD(T)-F12/CBS
	IL174 ⁱ	Medium ionic liquid anion-cation dimers	CCSD(T)/CBS
	PNICO23 ^j	Small pnictogen dimers	W1-F12 & W2-F12
	RG18 ^f	Rare/Noble gas dimers	CCSD(T)/CBS (counterpoise corrected)
	S22 ^{k,l}	Small-medium neutral organic dimers	CCSD(T)/CBS
	S22TRM ^m	Small-medium neutral organic trimers	CCSD(T)/CBS
	S66 ⁿ	Small-medium neutral organic dimers	CCSD(T)/CBS
	SSI ^o	Small-medium neutral and charge amino acid side chain dimers	DW-CCSD(T**)-F12
Reaction Energy (RE)	ISOL24 ^{d,p}	Large neutral organic molecule isomerizations	DLPNO-CCSD(T)/CBS
	BH76RC ^{q,r}	Reaction energies of small organic molecules	W2-F12
	BSR36 ^{s,t}	Alkane bond separation reaction energies	DLPNO-CCSD(T)/CBS
	DARC ^{q,u}	Diels-Alder reaction energies	W1-F12
	RSE43 ^{q,v}	Methyl radical exchange reaction energies of various organic radicals	W1-F12
	ISO34 ^{q,w}	Small-medium organic isomerisation reaction energies	W1-F12
Reaction Barrier (RB)	BH76 ^{q,x,y}	Reaction barriers of miscellaneous organic molecule reactions	W2-F12
	BHPERI ^{q,z,A,B,C}	Reaction barriers of pericyclic reactions of organic molecules	W1-F12 & W2-F12

	BHDIV10 ^q	Miscellaneous small molecule reaction barriers	W1-F12 & W2-F12
	BHROT27 ^q	Single bond rotation reaction barriers of small-medium organic molecules	W1-F12 & W2-F12

a) Rupp *et al.* Phys. Rev. Lett. **108**, 058301 (2012) *b)* Karton *et al.* Chem. Phys. Lett. **510**, 165-178 (2011) *c)* Tsuzuki *et al.* J. Chem. Phys. **124**, (2006) *d)* Lao *et al.* J. Chem. Theory. Comput. **11**, 2473-2486 (2015) *e)* Stella *et al.* ChemRxiv *f)* Goerigk *et al.* Phys. Chem. Chem. Phys. **19** 32184-32215 (2017) *g)* Rezáč *et al.* J. Chem. Theory. Comput. **8**, 4285-4292 (2012) *h)* Kozuch *et al.* J. Chem. Theory. Comput. **9**, 1918-1931 (2013) *i)* Rigby *et al.* J. Chem. Theory. Comput. **10**, 3111-3122 (2014) *j)* Setiawan *et al.* J. Phys. Chem. **119**, 1642-1656 (2015) *k)* Marshall *et al.* J. Chem. Phys. **135**, 194102 (2011) *l)* Jurečka *et al.* Phys. Chem. Chem. Phys. **8**, 1985-1993 *m)* Alkan *et al.* J. Phys. Chem. **123**, 8406-8416 (2019) *n)* Rezáč *et al.* J. Chem. Theory. Comput. **11**, 2473-2486 (2015) *o)* Burns *et al.* J. Chem. Phys. **147** (2017) *p)* Heunerbein *et al.* Phys. Chem. Chem. Phys. **12**, 6940-6948 *q)* Lao *et al.* J. Chem. Theory. Comput. **11**, 2473-2486 (2015) *r)* Goerigk *et al.* J. Chem. Theory. Comput. **6** 107-126 (2010) *s)* Steinmann *et al.* J. Chem. Theory. Comput. **6**, 107-126 (2010) *t)* Krieg *et al.* Mol. Phys. **108**, 2655-2666 (2010) *u)* Johnson *et al.* J. Chem. Phys. **129** (2008) *v)* Neese *et al.* J. Chem. Theory. Comput. **5**, 3060-3073 *w)* Grimme *et al.* J. Org. Chem. **72**, 2118-2126 (2007) *x)* Zhao *et al.* Phys. Chem. Chem. Phys. **7**, 43-52 (2005) *y)* Zhao *et al.* J. Phys. Chem. A **109**, 2012-2018 (2005) *z)* Guner *et al.* J. Phys. Chem. A **107**, 11445-11459 (2003) *A)* Dinadayalane *et al.* J. Phys. Chem. A **106**, 1627-1633 *B)* Ess *et al.* J. Phys. Chem. A **109**, 9542-9553 (2005) *C)* Karton *et al.* J. Comp. Chem. **36**, 622-632 (2015)

Acknowledgements. The authors acknowledge a generous allocation of computational resources from the Monash eResearch Centre and the National Computational Infrastructure through the Merit Allocation Scheme and the Monash-NCI partnership. This work was supported by the Australian Research Council through the Discovery Project Scheme.

Author contributions.

All authors discussed the results and contributed to the final manuscript. P. H. designed the implementation workflow of the δ ML-DFT code, carried out the training of the code, performed majority of the data analysis and interpretation, performed DFT calculations used for comparison, and revised the manuscript; F. S. carried out a preliminary study in adopting electron density in a ML code as applied to the ARION583 database; R. K.

applied the trained code to study reaction energies and barriers, S. A. C. contributed to refining benchmark calculations, performed DFT calculations used for comparison and revised the manuscript; N. W. carried out a preliminary study adopting electron density in a ML code as applied to the S22, S66 and S22TRM databases; E. I. I. conceived the project, supervised the project, produced the first draft of the manuscript and revised the manuscript.

Additional information. Supplementary Information is available for this paper. A GitHub repository accompanies this paper: <https://github.com/PHalat/DeltaML-DFT> Correspondence and requests for materials should be addressed to katya.pas@monash.edu. Reprints and permissions information is available at www.nature.com/reprints.

References:

- 1 Perdew, J. P. & Schmidt, K. in *AIP conference proceedings*. 1-20 (American Institute of Physics).
- 2 Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785 (1988).
- 3 Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098 (1988).
- 4 Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
- 5 Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007-1023 (1989).
- 6 Sun, Q. *et al.* PySCF: the Python-based simulations of chemistry framework. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1340 (2018).
- 7 Sun, Q. *et al.* Recent developments in the PySCF program package. *J. Chem. Phys.* **153** (2020).
- 8 Bogojeski, M., Vogt-Maranto, L., Tuckerman, M. E., Müller, K.-R. & Burke, K. Quantum chemical accuracy from density functional approximations via machine learning. *Nat. Commun.* **11**, 5223 (2020).
- 9 Ryczko, K., Strubbe, D. A. & Tamblyn, I. Deep learning and density-functional theory. *Phys. Rev. A* **100**, 022512 (2019).
- 10 Ryabov, A., Akhatov, I. & Zhilyaev, P. Application of two-component neural network for exchange-correlation functional interpolation. *Sci. Rep.* **12**, 14133 (2022).
- 11 Dick, S. & Fernandez-Serra, M. Machine learning accurate exchange and correlation functionals of the electronic density. *Nat. Commun.* **11**, 3509 (2020).
- 12 Nagai, R., Akashi, R. & Sugino, O. Machine-learning-based exchange correlation functional with physical asymptotic constraints. *Phys. Rev. Res.* **4**, 013106 (2022).
- 13 Nagai, R., Akashi, R. & Sugino, O. Completing density functional theory by machine learning hidden messages from molecules. *npj Comput. Mater.* **6**, 43 (2020).
- 14 Curtiss, L. A., Raghavachari, K., Redfern, P. C. & Pople, J. A. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.* **106**, 1063-1079 (1997).
- 15 Luise, G. *et al.* Accurate and scalable exchange-correlation with deep learning. *arXiv* (2025).
- 16 Goerigk, L. *et al.* A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **19**, 32184-32215 (2017).
- 17 Rupp, M., Tkatchenko, A., Müller, K.-R. & Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
- 18 Karton, A., Daon, S. & Martin, J. M. W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* **510**, 165-178 (2011).
- 19 Mardirossian, N. & Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **115**, 2315-2372 (2017).
- 20 Rezáč, J., Riley, K. E. & Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **7**, 2427-2438 (2011).
- 21 Burns, L. A. *et al.* The BioFragment Database (BFDb): An open-data

- platform for computational chemistry analysis of noncovalent interactions. *J. Chem. Phys.* **147** (2017).
- 22 Stella, F., Halat, P. & Izgorodina, E. I. Do electrostatic forces correlate with dispersion forces in charged molecular systems? *ChemRxiv* **0702** (2026).
- 23 Alkan, M., Xu, P. & Gordon, M. S. Many-body dispersion in molecular clusters. *J. Phys. Chem. A* **123**, 8406-8416 (2019).
- 24 Lao, K. U., Schäffer, R., Jansen, G. & Herbert, J. M. Accurate description of intermolecular interactions involving ions using symmetry-adapted perturbation theory. *J. Chem. Theory Comput.* **11**, 2473-2486 (2015).
- 25 Chen, S., Ishii, J., Horiuchi, S., Yoshizawa-Fujita, M. & Izgorodina, E. I. Difference in chemical bonding between lithium and sodium salts: influence of covalency on their solubility. *Phys. Chem. Chem. Phys.* **19**, 17366-17372 (2017).
- 26 Wylie, L., Flynn, M., Blair, V. L., Andrews, P. C. & Izgorodina, E. I. Isomers of Alkali Metal (Methylbenzyl) allylamides: A Theoretical Perspective. *ACS Omega* **5**, 9448-9457 (2020).
- 27 Setiawan, D., Kraka, E. & Cremer, D. Strength of the pnictogen bond in complexes involving group Va elements N, P, and As. *J. Phys. Chem. A* **119**, 1642-1656 (2015).
- 28 Rezac, J., Riley, K. E. & Hobza, P. Benchmark calculations of noncovalent interactions of halogenated molecules. *J. Chem. Theory Comput.* **8**, 4285-4292 (2012).
- 29 Kozuch, S. & Martin, J. M. Halogen bonds: Benchmarks and theoretical analysis. *J. Chem. Theory Comput.* **9**, 1918-1931 (2013).
- 30 Huenerbein, R., Schirmer, B., Moellmann, J. & Grimme, S. Effects of London dispersion on the isomerization reactions of large organic molecules: a density functional benchmark study. *Phys. Chem. Chem. Phys.* **12**, 6940-6948 (2010).
- 31 Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132** (2010).
- 32 Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215-241 (2008).
- 33 Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **115**, 036402 (2015).
- 34 Chai, J.-D. & Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **128** (2008).
- 35 Mardirossian, N. & Head-Gordon, M. ω B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **144** (2016).
- 36 Slater, I. A simplification of the Hartree-Fock method. *Phys. Rev.* **81** (1951).
- 37 Vosko, S. H., Wilk, L. & Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **58**, 1200-1211 (1980).
- 38 Tao, J., Perdew, J. P., Staroverov, V. N. & Scuseria, G. E. Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **91**, 146401 (2003).
- 39 Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv* (2014).
- 40 Neese, F. Software update: the ORCA program system—version 6.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **15**, e70019 (2025).

- 41 Neese, F. The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 73-78 (2012).
- 42 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.* **7**, 3297-3305 (2005).
- 43 Blaudeau, J.-P., McGrath, M. P., Curtiss, L. A. & Radom, L. Extension of Gaussian-2 (G2) theory to molecules containing third-row atoms K and Ca. *J. Chem. Phys.* **107**, 5016-5021 (1997).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [NatureMSDFTESI.pdf](#)