

Accurate Density Functional Theory for Non-Covalent Interactions in Charged Systems

Heng Zhao¹, Balázs Lőrincz², Tobias Henkes³, Dénes Berta², Péter Nagy², Alexandre Tkatchenko³, Stefan Vuckovic¹

1. University of Fribourg

2. Budapest University of Technology and Economics

3. University of Luxembourg

Abstract

Accurately modeling non-covalent interactions (NCIs) involving charged systems remains an outstanding challenge in Density Functional Theory (DFT), with implications across natural and life sciences, engineering, e.g., in biochemistry, catalysis, and materials science. For these interactions, the interplay between electrostatics, polarization, and dispersion leads to systematic errors of up to tens of kcal/mol in standard dispersion-enhanced DFT methods. We solve this problem by introducing (r2SCAN+MBD)@HF, a DFT method without empirically fitted parameters that combines the r2SCAN functional and many-body dispersion, both evaluated on Hartree-Fock densities. We show that the unique synergy of these three components enables balanced treatment of short- and long-range correlation, which is crucial for accurate description of NCIs involving charged systems. Evaluations on standard benchmarks and a new Metal Ion Protein Clusters dataset introduced here show that (r2SCAN+MBD)@HF significantly improves accuracy for NCIs involving charged systems while maintaining robust performance for neutral systems. Given the ubiquity of such interactions, (r2SCAN+MBD)@HF is broadly applicable from biochemistry and materials science, including for generating high-quality data to train machine-learning force fields.

Accurate Density Functional Theory for Non-Covalent Interactions in Charged Systems

Heng Zhao^{1*}, Balázs D. Lőrincz^{2,3,4}, Tobias Henkes⁵, Dénes Berta^{2,3,4}

Péter R. Nagy^{2,3,4}, Alexandre Tkatchenko⁵, Stefan Vuckovic¹

¹Department of Chemistry, University of Fribourg, Fribourg & 1700, Switzerland.

²Department of Physical Chemistry and Materials Science, Faculty of Chemical Technology and Biotechnology,
Budapest University of Technology and Economics, Budapest & H-1111, Hungary.

³HUN-REN-BME Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary.

⁴MTA-BME Lendület Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary.

⁵Department of Physics and Materials Science,
University of Luxembourg, Luxembourg City & L-1511, Luxembourg.

*Corresponding author. Email: stefan.vuckovic@unifr.ch

Accurately modeling non-covalent interactions (NCIs) involving charged systems remains an outstanding challenge in Density Functional Theory (DFT), with implications across natural and life sciences, engineering, e.g., in biochemistry, catalysis, and materials science. For these interactions, the interplay between electrostatics, polarization, and dispersion leads to systematic errors of up to ten of kcal/mol in standard dispersion-enhanced DFT methods. We solve this problem by introducing (r²SCAN+MBD)@HF, a DFT method without empirically fitted parameters that combines the r²SCAN functional and many-body dispersion, both evaluated on Hartree-Fock densities. We show that the unique synergy of these three components enables balanced treatment of short- and long-range correla-

tion, which is crucial for accurate description of NCIs involving charged systems. Evaluations on standard benchmarks show that (r²SCAN+MBD)@HF significantly improves accuracy for NCIs involving charged systems while maintaining robust performance for neutral systems. Further tests on the Metal Ion Protein Clusters dataset introduced here demonstrate its improved accuracy for metal–protein interactions, including cases where the metal ion is surrounded by negatively charged ligands for which standard DFT can completely break down. Given the ubiquity of such interactions, (r²SCAN+MBD)@HF is broadly applicable from biochemistry and materials science, including for generating high-quality data to train machine-learning force fields.

Non-covalent interactions (NCIs) involving charged systems (charged NCIs for simplicity) are ubiquitous across scientific disciplines, such as in acid-base, electro-, redox-, coordination chemistry, ionic crystals and liquids, ionization and electron attachment processes, charge conduction and transfer, etc. For example, in biology, charged NCIs stabilize enzymatic transition states (1), mediate protein folding via salt bridges between charged amino acid side chains (2), and regulate ion transport (3). In catalysis, ionic complex stability dictates reaction pathways and selectivity, (e.g., in Brønsted and Lewis acid catalysis (4)). In materials science, they enable gas separation in metal-organic frameworks (MOFs) (5) and play a crucial role in the design of batteries by governing alkali metal intercalation in electrode materials (6). Thus, accurate computational models of charged NCIs are essential, spanning applications from ionic liquids (7) to electrochemical processes in energy storage systems (8).

When it comes to computational simulations of NCIs, including those involving charged species, benchmark-level accuracy is typically achieved using high-level wavefunction-based methods, such as Coupled Cluster with Single, Double, and perturbative Triple excitations [CCSD(T)] (9) and Quantum Diffusion Monte Carlo (DMC) (10). Despite advances in cost-reduction (11–14), these wavefunction methods remain impractical for biological or condensed-phase systems exceeding hundreds of atoms.

The unparalleled trade-off between accuracy and cost has made Density Functional Theory (DFT) the workhorse for NCI simulations (15–18). While DFT initially struggled with dispersion interactions, this issue has been largely mitigated, especially for neutral cases, by various dispersion

53 methods (19–27). However, the growing interest in charged NCI datasets (28–30), including the
54 extensive DES15K benchmark set (31), has revealed the major limitations of dispersion-enhanced
55 DFT for these interactions (17). Namely, while dispersion-enhanced DFT typically performs well
56 for neutral complexes, with interaction energy errors around 0.5 kcal/mol, its errors for charged
57 species can be up to 10 times larger, regardless of the chosen method (“D3” (22, 23), “D4” (32, 33),
58 “XDM” (19, 20), “TS” (21), or “MBD” (26, 27)). This issue is systematic, as recently demonstrated
59 by Johnson’s study of ~15k (31) NCI complexes (17).

60 Computational simulations of charged NCIs fundamentally differ from neutral ones, especially
61 for interactions between metallic cations and neutral molecules. A prototypical example is the
62 Li^+ –benzene complex (17), where the strong inhomogeneous electric field from Li^+ ’s localized
63 positive charge distorts the benzene electron cloud, inducing partial electron transfer. From the per-
64 spective of the theory of intermolecular interactions, describing such interactions in inhomogeneous
65 electric fields is notoriously challenging, as electrostatics, polarization, and dispersion become in-
66 herently coupled (34). This coupling critically influences a broad range of (bio)chemical processes,
67 such as selectivity in biological ion channels (see Ref. (3)). Further complexity arises because
68 external charges substantially alter dispersion interactions, stabilizing or destabilizing molecular
69 binding depending on the charge sign (34).

70 Within DFT, accurately describing strongly polarized complexes under electric fields is par-
71 ticularly challenging due to delocalization errors—often (35, 36) termed “the greatest outstanding
72 challenge in DFT” (? , 35). Another critical issue is accurately capturing electrostatics, polarization,
73 and dispersion and their coupling, requiring careful balancing of short- and long-range correlation
74 without double-counting effects (37, 38). Along these lines, dispersion methods in DFT compensate
75 for the long-range correlation missed by (semi-local) DFT approximations (15–17). However, dis-
76 persion methods must be balanced with base DFT to avoid double-counting of correlation effects
77 (see Refs. (37, 38)). Although general, this issue is particularly severe for charged NCIs, due to the
78 interplay of polarization, density errors, and poor transferability of dispersion methods trained on
79 neutral NCIs. Simply put, DFT typically underbinds neutral NCIs, so dispersion methods improve
80 their interaction energies. In contrast, for charged NCIs, where DFT already overbinds, adding an
81 attractive dispersion energy only worsens the overbinding. Thus, developing a dispersion-enhanced
82 DFT approximation that describes both neutral and charged NCIs remains a major challenge.

Resolving this challenge with charged NCIs is crucial, especially today as machine-learning force fields (MLFFs) trained on DFT data expand the reach of DFT quality modeling to larger systems and longer dynamics simulations (39). However, any errors in DFT propagate to MLFFs, reducing their reliability and predictive power (40–42). Given the critical role of charged NCIs across many applications and the growing use of MLFFs to extend DFT’s reach (43), it is imperative to address DFT’s deficiencies in describing these interactions.

In this work, we solve the DFT problem of charged NCIs by ensuring a balanced description of correlation effects in both neutral and charged systems through the integrative development of (r²SCAN+MBD)@HF. In (r²SCAN+MBD)@HF, the r²SCAN functional (44) and many-body dispersion (MBD) are evaluated on the Hartree-Fock (HF) orbitals – a combination essential for ensuring a consistent and accurate treatment of both neutral and charged NCIs. The r²SCAN functional was developed by Perdew, Sun, and co-workers to overcome the numerical issues present in its predecessor, SCAN (45). On the other hand, evaluating approximate DFT functionals on HF orbitals has a long history in quantum chemistry (46, 47), and more recently this procedure has been widely employed in the context of density-corrected DFT, a formalism developed by Burke, Sim, and co-workers to isolate density-driven errors in approximate DFT energies (47–50). Over the past decade, the same authors have identified many classes of systems for which evaluating approximate density functionals on HF orbitals yields substantially improved accuracy over self-consistent DFT (18).

Going back to (r²SCAN+MBD)@HF, beyond solving the problem of charged NCIs, our integrative design of (r²SCAN+MBD)@HF yields several key advantages over existing dispersion-enhanced DFT methods. **(1)** It achieves a balanced treatment of short- and long-range correlations across diverse NCIs, which is missed by other dispersion-enhanced DFT approximations. **(2)** By using a non-empirical approach (the only MBD parameter is set to unity and is largely system-insensitive), we avoid extensive empirical fitting of dispersion methods, which can be highly sensitive to the training set (51–53) and limit method transferability (54). **(3)** Despite being designed to address the DFT deficiencies for charged NCIs, our method retains robust accuracy for neutral systems, matching or surpassing leading semilocal and hybrid functionals on main-group benchmarks. Furthermore, using HF densities within the “density- and dispersion-corrected DFT” framework of Sim, Burke, and co-workers (51–53) serves a dual purpose in (r²SCAN+MBD)@HF: while HF

densities typically improve upon approximate DFT densities for charged NCIs, their primary role here is to maintain balance between baseline DFT and MBD across different NCI types. Crucially, this balance is lost if DFT's self-consistent density is used (e.g., if r^2 SCAN+MBD is employed in place of $(r^2$ SCAN+MBD)@HF). More broadly, altering any of its three components disrupts the consistent description of neutral and charged NCIs, typically leading to severe overbinding of charged complexes.

By addressing DFT's deficiencies for charged NCIs while retaining accuracy for neutral systems, $(r^2$ SCAN+MBD)@HF provides a robust framework for quantum simulations of complex NCIs. The most challenging examples from our benchmark sets that include dications, ranging from small complexes to clusters extracted from our Metal Ion Protein Clusters dataset (see Results), are shown in Fig. 1 and illustrate major and consistent improvements of $(r^2$ SCAN+MBD)@HF over the widely used PBE0+MBD methods. As demonstrated in Results through the consistent success of $(r^2$ SCAN+MBD)@HF for charged NCIs, our method is ideally suited for computational studies in fields where charged NCIs are critical, including biomolecular interactions, adsorption in MOFs, and electrode materials in batteries.

$(r^2$ SCAN+MBD)@HF Resolves Charged NCI Issues: DES15K insights

Figure 2 highlights a central finding of this work, demonstrating how $(r^2$ SCAN+MBD)@HF significantly improves dispersion-enhanced DFT accuracy for charged NCIs. We assess $(r^2$ SCAN+MBD)@HF on the DES15K dataset (31), which comprises 3,052 unique dimers (distinct monomer pairs) and 14,651 dimer geometries, spanning both charged (approximately one third) and neutral complexes (approximately two thirds); see Tab. S1 for the distribution across the interaction-type subsets. Each unique dimer in DES15K contains compressed, near-equilibrium, and slightly expanded geometries, sampled from potential energy surface scans and molecular dynamics simulations (31). Specifically, Fig. 2 presents error distribution histograms for interaction energies in different DES15K categories, comparing $(r^2$ SCAN+MBD)@HF to the widely used PBE0+MBD. While many other DFT+dispersion combinations could be considered, it is crucial to note that Johnson and co-workers have shown that dispersion methods systematically fail for charged NCIs (17), leading to large errors regardless of the method used. For this reason, we focus on comparing

(r^2 SCAN+MBD)@HF with PBE0+MBD, which is not only representative of the performance of dispersion-enhanced DFT for charged NCIs (17) but also a widely adopted in recent molecular dataset generation efforts (55, 56).

Figs. 2(a) and 2(b) illustrate the most striking improvements of (r^2 SCAN+MBD)@HF for charged NCIs. For alkaline earth dication–neutral complexes (Fig. 2a), PBE0+MBD severely overbinds (MAE = 4.41 kcal/mol), while (r^2 SCAN+MBD)@HF reduces this error by a factor of four and yielding much less skewed error distribution. Similarly, for alkali cation–neutral complexes (Fig. 2b), (r^2 SCAN+MBD)@HF reduces the MAE compared to PBE0+MBD (1.91 kcal/mol) by a factor of three, again resulting in a significantly less skewed error distribution. These results highlight (r^2 SCAN+MBD)@HF’s robustness in correcting DFT failures for NCIs with cations.

Fig. 2(c) includes less polarizing non-metal cations with more delocalized charges interacting with neutral molecules. Consequently, PBE0+MBD yields smaller errors (MAE = 1.08 kcal/mol) than in Fig. 2(a,b). Yet, (r^2 SCAN+MBD)@HF still significantly improves accuracy, reducing the MAE nearly by half (to 0.60 kcal/mol), further demonstrating its robustness for charged NCIs.

We now turn to another DES15K subset: anion–neutral complexes. Fig. 2(d) shows that both methods perform similarly for non-metal anion–neutral pairs, with PBE0+MBD slightly outperforming (r^2 SCAN+MBD)@HF (MAE = 0.6 vs. 0.8 kcal/mol). However, (r^2 SCAN+MBD)@HF is generally more robust for anion-containing NCIs because (i) semilocal DFT often yields unphysical results for anions with positive HOMO energies that imply unbound states (48), hence the red flag next to PBE0+MBD in Fig. 2(d); and (ii) for more challenging cases like the B30 dataset (57), (r^2 SCAN+MBD)@HF reduces the MAE by half compared to PBE0+MBD (Tab. S10).

For cation–anion pairs (Fig. 2e), dominated by strong electrostatics, PBE0+MBD again significantly overbinds (MAE = 1.39 kcal/mol). (r^2 SCAN+MBD)@HF substantially improves accuracy (MAE = 0.76 kcal/mol), with most errors within ± 2 kcal/mol. For neutral complexes (Fig. 2f), where PBE0+MBD already performs well (17), (r^2 SCAN+MBD)@HF remains comparably accurate (MAE $\sim$ 0.5 kcal/mol).

Focusing on the large errors in DES15K interaction energies ($|\text{error}| \geq 2$ kcal/mol) and using symmetry-adapted perturbation theory (SAPT), which decomposes these interactions into electrostatic, induction, and dispersion components, we find another clear difference between the performance of the two methods. (r^2 SCAN+MBD)@HF exhibits only 559 such outliers (3.8%

of DES15K), and importantly these outliers are not only fewer in number but also confined to a much narrower physical regime, roughly matching the region where all the three SAPT components contribute comparably (see the SAPT ternary plots in Fig. S1). In contrast, PBE0+MBD yields 1072 such cases with $|\text{error}| \geq 2$ kcal/mol (7.3% of DES15K), spanning a broader physical regime wherever electrostatics or induction dominate.

These results demonstrates the robustness of $(r^2\text{SCAN}+\text{MBD})@\text{HF}$, highlighting its success in addressing dispersion-corrected DFT deficiencies for charged NCIs. Additionally, despite being free from empirical parameters fitted to data, $(r^2\text{SCAN}+\text{MBD})@\text{HF}$ matches the performance of modern dispersion-corrected semi-local and hybrid functionals on the diverse organic GMTKN55 dataset (Tab. S11) (58), further emphasizing its robustness. In the following sections, we explore practical applications to metal-containing protein interactions and analyze the principles underlying $(r^2\text{SCAN}+\text{MBD})@\text{HF}$'s improved accuracy.

$(r^2\text{SCAN}+\text{MBD})@\text{HF}$ at work for metal-protein interactions

Having established the accuracy of $(r^2\text{SCAN}+\text{MBD})@\text{HF}$ for DES15K, we now assess it on biologically relevant metal–protein interactions. To this end, we have curated the Metal Ion Protein Clusters (MIPC) dataset, consisting of 45 biologically relevant metal–ligand complexes extracted from high-resolution (≤ 2.5 Å) Protein Data Bank (PDB) crystal structures with metal cations coordinated by both neutral and charged amino acid residues. The ≤ 2.5 Å resolution cutoff was applied because lower resolutions would risk poorly resolved ion-binding sites.

In our MIPC, we distinguish a subset with negatively charged ligands [Fig. 3(a)] and a subset with neutral ligands [Fig. 3(b)]. The former MIPC subset [Fig. 3(a)] includes important mono- (Na^+ and K^+) and divalent (Mg^{2+} , Ca^{2+} , and Zn^{2+}) ions in biochemistry. Mg^{2+} has a well-established catalytic role in phosphate chemistry (59). The most common superfamily of such enzymes are the P-loop containing nucleoside triphosphate hydrolases, so we included the highest resolution representatives containing Mg^{2+} with GTP (60) and GDP separately (61). Zn^{2+} ions form zinc-fingers, an important motif for nucleic acid recognition (62). Typical coordinating residue are cysteins and histidines, we represented these by including Zn^{2+} sites selected for Zn-specific force-field development (63). We also included examples of Ca^{2+} clusters, as Ca^{2+} is a very common

cation in protein structures (64, 65), with tightly regulated cellular concentrations and an important regulatory role in kinases (66).

The MIPC subset with neutral ligands [Fig. 3(b)] include catalytic ion sites, e.g., the secondary Ca^{2+} site in Phospholipase A₂ (PDB ID 1POD) (67) or one of the Zn^{2+} ions of the zinc-cluster in the YcdX (PDB ID 1M65), setting up a hydroxide nucleophile for phosphoester hydrolysis (68). In other cases present in MIPC, the ion binding is linked to allosteric regulation of substrate binding (PDB IDs 2C1U, 7O20) (69, 70). In an inhibitor screening study for sodium-coupled transporter LeuT (PDB ID 2Q6H) (71), sodium sites are in the close proximity of binding site of the leucine and the inhibitor clomipramine. The reliable computation of such binding energies is especially important when the identification of the bound cation relies on quantum-mechanical calculations (PDB ID 2DSN) (72).

For each protein in MIPC, we construct a cluster model by selecting amino acid residues within the first coordination shell of metal ions (Li^+ , Na^+ , K^+ , Mg^{2+} , Ca^{2+} , and Zn^{2+}). We then benchmark the interaction energies between metallic cation and its surrounding amino acid residues using high-level LNO-CCSD(T) calculations (12, 13, 73, 74) and include benchmark uncertainty estimates (see Methods).

Fig. 3 compares ($r^2\text{SCAN}+\text{MBD}$)@HF and PBE0+MBD errors against LNO-CCSD(T) reference interaction energies for MIPC (including uncertainty estimates), separating (a) negatively charged and (b) neutral ligands. The labels are PDB codes, and red flags mark negatively charged systems for which the MBD part of PBE+MBD is not obtainable (see below). In Figs. 3(a) and 3(b), the systems are grouped by metal cation and labeled by PDB codes, with representative proteins (cartoon) and corresponding metal–ligand clusters (ball-and-stick) from MIPC illustrated in Fig. 1.

For clusters with negatively charged ligands [Fig. 3(a)], our method achieves a MAE of 1.5 kcal/mol, while PBE0+MBD gives twice that value (see Tab. S5 in the SI for individual cluster values). However, beyond the numerical improvements, our method is fundamentally better suited to systems involving negatively charged ligands than standard self-consistent hybrids such as PBE0. For one of the complexes in Fig. 3(a), the MBD energies of PBE0+MBD cannot even be evaluated for the ligand part of the cluster (the K^+ -based cluster from PDB ID 5AH1, marked with a red flag), which contains negatively charged COO^- groups from two Asp residues. We can again trace back this failure to the deficiencies of standard self-consistent hybrid DFT for anions, which we discussed

earlier (48, 50, 75). In the negatively charged ligand of the 5AH1 cluster, the PBE0 density errors distort the Hirshfeld volumes and produce unphysical atomic polarizabilities, which in turn lead to an unphysical MBD solution (76). Furthermore, the same type of PBE0+MBD failure observed for the 5AH1 cluster can also occur for other complexes in Fig. 3(a) when more diffuse basis sets are used (see Fig. S9 for a concrete example involving the Zn^{2+} -based cluster with PDB ID 2AP1). It is crucial to understand that these failures do not originate from the MBD model itself, but rather from erroneous PBE0 densities. Within our method, there is no problem with the MBD part for the 5AH1 cluster and the resulting $(\text{r}^2\text{SCAN}+\text{MBD})@\text{HF}$ interaction energy also lies within the reference benchmark. In this way, by avoiding the fundamental limitations of PBE0 for anionic densities that can lead to the breakdown of MBD evaluated at those densities, our $(\text{r}^2\text{SCAN}+\text{MBD})@\text{HF}$ approach is not only more accurate, but also more physically grounded for the interaction energies of the protein clusters involving negatively charged ligands.

Moving now to Fig. 3(b) containing the MIPC clusters with neutral ligands, similar trends can be observed. Overall, $(\text{r}^2\text{SCAN}+\text{MBD})@\text{HF}$ achieves a MAE of 2.3 kcal/mol, compared to 5.5 kcal/mol for PBE0+MBD (see Tab. S6 in the SI for individual cluster values). For the single Li^+ MIPC cluster (selected for therapeutic relevance (77)), the $(\text{r}^2\text{SCAN}+\text{MBD})@\text{HF}$ error lies essentially within the reference uncertainty. For the Na^+ , Mg^{2+} , and Ca^{2+} clusters, $(\text{r}^2\text{SCAN}+\text{MBD})@\text{HF}$ shows slightly larger errors, yet still substantially reduces the characteristic overbinding observed with PBE0+MBD. For the Zn^{2+} clusters in Fig. 3(b), a closed-shell transition-metal cation not present in DES15K, both $(\text{r}^2\text{SCAN}+\text{MBD})@\text{HF}$ and PBE0+MBD show small errors for most cases. However, for one Zn^{2+} cluster (PDB ID 1CAK), where PBE0+MBD exhibits pronounced overbinding, $(\text{r}^2\text{SCAN}+\text{MBD})@\text{HF}$ significantly reduces the error.

While it has been reported in the literature that ion–water interaction energies computed with PBE0 combined with a dispersion method often exhibit good performance as the number of surrounding water molecules increases, it is useful to test how these errors evolve in the more enzyme-like coordination environments as those in our MIPC set. To examine this, we have analysed the Mg^{2+} cluster (PDB ID = 2YNV) already shown in Fig. 3(b). Starting from the original cluster, we have removed one ligand at a time to generate five *subclusters*, in which the number of ligands varies from one to five (Fig. 3(c)). For each subcluster, we have benchmarked the interaction energy between Mg^{2+} and the present ligand(s), enabling us to assess how DFT errors behave as the number

of ligands increases. The results in Fig. 3(d) show that although the absolute error increases with the number of ligands for both PBE0+MBD and (r²SCAN+MBD)@HF, this increase is much less steep for (r²SCAN+MBD)@HF, indicating a more robust performance of (r²SCAN+MBD)@HF as the number of coordinating ligands around the ion increases.

MIPC represents a significant step toward real-life systems also in terms of ion coordination: ions in MIPC are coordinated by 4–6 ligands. This introduces several challenges for MIPC simulations. First, ion–ligand interaction errors can grow with the number of ligands, as illustrated in Fig 3(d). Second, modeling polarized ligand–ligand interactions is difficult, as these interactions often become overly repulsive due to overpolarization in standard DFT methods. Third, as noted above, if negative charges are present in the ligand, standard DFT can yield unphysical results. Importantly, (r²SCAN+MBD)@HF effectively overcomes these issue, and the results from this section demonstrate that (r²SCAN+MBD)@HF reliably describes such complex charged interactions involving ions in protein or solvent environments. For example, these are critical to biological functions such as substrate binding, catalytic activity, and allosteric regulation. Its robust accuracy across diverse charged NCIs makes (r²SCAN+MBD)@HF a powerful method for biochemical simulations.

In contrast to the clusters discussed above involving metal cations, for the interaction energies of the glycine–water complexes of Ref. (78), PBE0+MBD yields slightly smaller errors than (r²SCAN+MBD)@HF (Tab. S9 in the SI). Nevertheless, the MAEs of both methods across the complexes with different forms of glycine (neutral, proton transfer transition state and zwitterion) remain below 1 kcal/mol. At the same time, when comparing the four glycine protonation forms (negatively and positively charged, neutral, and zwitterionic) interacting with a single water molecule, (r²SCAN+MBD)@HF is slightly more accurate than PBE0+MBD (see Fig. S16 in the SI). For zwitterions, some caution is needed, as standard hybrid DFT functionals, such as PBE0, have issues for zwitterionic systems, including unphysical charges and erratic SCF behaviour. Thus, even though (r²SCAN+MBD)@HF may exhibit accuracy comparable to or slightly below that of PBE0+MBD for zwitterionic complexes, it provides a physically more robust framework for treating these systems (79).

The role of densities in (r²SCAN+MBD)@HF for cation-neutral pairs

Having demonstrated (r²SCAN+MBD)@HF's accuracy for charged NCIs, we now analyze how it outperforms standard dispersion-enhanced DFT. A natural question arises: is this success primarily due to the r²SCAN functional, the MBD term, or the use of HF densities? Analyzing the contributions of these components individually and combined, we find that (r²SCAN+MBD)@HF's accuracy relies on their synergy—altering any component disrupts this balance and compromises accuracy for charged and/or neutral NCIs.

We first examine the role of HF densities in (r²SCAN+MBD)@HF. It has been shown that evaluating functionals on the HF densities, instead of self-consistent ones, often significantly improves DFT accuracy across many systems (18, 48, 50, 52, 53). For certain systems (e.g., stretched heterodimers), HF densities are more accurate (i.e., energetically closer) to their exact counterparts than DFT ones leading to DFT evaluated on HF densities improving over self-consistent DFT *for the right reason* (80). However, for other systems, such as transition states, whether this improvement for the right reason or error cancellations has been questioned (81, 82).

Thus, to determine whether HF or DFT densities are more accurate for our systems of interest, we compare both against high-quality CCSD densities for representative cation–neutral complexes in Fig. 4. Specifically, for metallic cation-neutral molecule pairs, we analyze the spherically averaged deformation density, $\Delta\tilde{\rho}(\mathbf{r}, u)$, which isolates changes in electron distribution upon complex formation. Second, by placing $\mathbf{r}$ at the metal cation, $\Delta\tilde{\rho}(\mathbf{r}, u)$ reveals how electron density redistributes—accumulating or depleting—around the metal cation as a function of radial distance u , providing a spherically averaged view of polarization effects. To define $\Delta\tilde{\rho}(\mathbf{r}, u)$, we first introduce the spherically averaged electron density, $\tilde{\rho}(\mathbf{r}, u)$ (see Methods for a precise definition), which represents the average electron density at a distance u from a reference point $\mathbf{r}$ (83–85). Then, $\Delta\tilde{\rho}(\mathbf{r}, u)$ is defined as the difference between $\tilde{\rho}(\mathbf{r}, u)$ of the complex and the sum of those from its isolated fragments (the cation and the neutral molecule). Importantly, $\Delta\tilde{\rho}(\mathbf{r}, u)$ integrates to zero at every $\mathbf{r}$, $\int_0^\infty du 4\pi u^2 \Delta\tilde{\rho}(\mathbf{r}, u) = 0$.

Figure 4 shows $4\pi u^2 \Delta\tilde{\rho}(\mathbf{r}, u)$ for selected cation-neutral complexes computed at the CCSD, PBE0, and HF levels, with $\mathbf{r}$ placed at the cation. We also display CCSD isosurfaces of $\Delta\rho(\mathbf{r})$ (the usual deformation, i.e., interaction density) to visualize charge redistributions. For panels (a)

water - K^+ and (b) chlorobenzene - K^+ , the HF (spherically averaged deformation) densities more closely match CCSD results than PBE0, supporting their use in $(r^2\text{SCAN}+\text{MBD})@HF$. The results for more systems follow similar trends and are given in Figs. S11 and S12. However, in panel (c) 1,3-dioxane- Na^+ , we cannot clearly determine whether HF or PBE0 is more accurate. While this analysis confirms that the HF (deformation) densities are more accurate than those of PBE0 for our cation–neutral complexes, it alone does not fully explain why $(r^2\text{SCAN}+\text{MBD})@HF$ achieves its overall accuracy. Beyond improving density accuracy for charged NCIs, the HF densities also play a key role in $(r^2\text{SCAN}+\text{MBD})@HF$'s success by restoring the balance between the $r^2\text{SCAN}$ and MBD contributions to interaction energies, a critical aspect explored in the next section.

The accuracy of $(r^2\text{SCAN}+\text{MBD})@HF$ – the synergy of the three components

To demonstrate the synergy of its three components in achieving high accuracy for both neutral and charged NCIs, Fig. 5 shows dissociation curves for three prototypical NCI-bound complexes. In this figure, DFT methods are benchmarked against CCSD(T) references for: (a) Li^+ –benzene (cation–neutral interaction), (b) acetic acid dimer (hydrogen-bonded), and (c) benzene dimer (dispersion-bound π – π stacking). In Fig. 5, the top row compares $(r^2\text{SCAN}+\text{MBD})@HF$ with PBE0-based variants, while the bottom row compares it with $r^2\text{SCAN}$ -based variants, highlighting that altering any of the three components in $(r^2\text{SCAN}+\text{MBD})@HF$ compromises its accuracy.

An essential requirement for a base DFT functional is that it underbinds NCIs, allowing the dispersion term to properly compensate this underbinding (see Ref. (37) for a more formalized criterion using the concept of "dispersionless" functionals). For example, PBE0 correctly underbinds the benzene dimer (Fig. 5(c), top), allowing MBD to improve its binding. However, it already overbinds the Li^+ – benzene complex (Fig. 5(a), top), so adding MBD only exacerbates the error. Thus, PBE0 fails to meet the fundamental requirement for dispersion-enhanced DFT: a base functional that consistently underbinds complexes across diverse NCIs, enabling dispersion corrections to accurately compensate.

On the other hand, self-consistent $r^2\text{SCAN}$ exhibits the desired underbinding behavior for the complexes in Fig. 5(a, c) [bottom], but not for the hydrogen-bonded acetic acid dimer in Fig. 5(b) [bottom], where it shows excessive overbinding, a behavior previously observed in Ref. (52) for

hydrogen-bonded systems. Among the three base functionals used here, only r^2 SCAN evaluated on the HF density (r^2 SCAN) exhibits a consistent underbinding trend across the three prototypical systems for the three NCI types. The interaction energies are significantly improved once dispersion is restored through adding our MBD term to r^2 SCAN. This highlights the critical synergy among the three components (r^2 SCAN, MBD, and HF density) in (r^2 SCAN+MBD)@HF, enabling balanced accuracy across diverse NCIs—an essential feature not achieved by other combinations.

Now, we turn to the role of dispersion in (r^2 SCAN+MBD)@HF, which is more subtle. Previously, r^2 SCAN@HF combined with the D4 correction and density-corrected (DC) parameters, yielding r^2 SCAN@HF-DC4, achieved excellent general accuracy, especially for hydrogen-bonded complexes and water (52). Consistent with this, r^2 SCAN@HF-DC4 performs well for neutral complexes [Figs. 5(b,c), bottom], but notably fails for the Li^+ –benzene system [Fig. 5(a), bottom]. The reparametrized version of r^2 SCAN@HF-DC4 based on the *dual-calibration* approach (53) reduces the error at intermediate and long Li^+ –benzene separations, yet the underlying erroneous behaviour, including an unphysical jump in the dissociation curve, remains (Fig. S7). This failure arises due to the sensitivity of the r^2 SCAN@HF-DC4 results to both the density and its D4 empirical parameters (51, 52) for the cation-neutral pairs. To elaborate on this, we have to go into more technical details on the construction of the dispersion methods. Namely, both MBD and D4 rely (at least indirectly) on the density information, with the latter requiring partial atomic charges for its construction (33). By default settings, classical partial atomic charges trained on the DFT density data are used in D4 (33). Since these settings are used in both the training and application of r^2 SCAN@HF-DC4, this method still retains input from the DFT densities, even though it is designed for the HF densities. On top of the indirect DFT density input, the classical charges employed in the D4 part of r^2 SCAN@HF-DC4 are based on the so-called *electronegativity equilibration* model, which inherently suffers from artificial long-range charge transfer, especially for systems containing ions (86). In contrast to these problems present in r^2 SCAN@HF-DC4, (r^2 SCAN+MBD)@HF strictly computes both the functional and the MBD term on the HF densities without any input from the DFT densities. Moreover, (r^2 SCAN+MBD)@HF is nearly non-empirical, with its only empirical parameter β , which controls the range-separation of MBD interaction, set to unity. Crucially, when coupled with r^2 SCAN@HF, MBD@HF exhibits remarkable insensitivity to β variations across different systems, as we will show later.

Transforming r^2 SCAN@HF's initial weakness into (r^2 SCAN+MBD)@HF's key strength

In Fig. 2, we have shown that (r^2 SCAN+MBD)@HF success for charged NCIs is due to the complementarity between MBD and r^2 SCAN@HF, with MBD accurately compensating for r^2 SCAN@HF's systematic underbinding. This ensures balanced accuracy across both charged and neutral NCIs, a key advantage of (r^2 SCAN+MBD)@HF over other dispersion-enhanced methods. What enables this balance is the consistent underbinding behavior of r^2 SCAN@HF, as observed previously for prototypical complexes across different NCIs. Here we further demonstrate that this critical behavior of r^2 SCAN@HF—essential for (r^2 SCAN+MBD)@HF's success (Figs. 2, 3, 5)—is systematic, using the broader DES15K dataset. Fig. 6(a) compares signed interaction energy errors from r^2 SCAN@HF, PBE0@HF, and PBE0 across neutral (left) and metal cation–neutral complexes (right). Only r^2 SCAN@HF consistently underbinds both neutral and cation complexes, unlike PBE0-based methods. As a result, only r^2 SCAN@HF, among the three methods, can be transformed into an accurate dispersion-enhanced DFT approach by adding a dispersion term. In this way, the systematic underbinding by r^2 SCAN@HF, initially a weakness, becomes (r^2 SCAN+MBD)@HF's key strength when combined with MBD, enabling balanced accuracy for charged and neutral NCIs.

Completing the Puzzle: How MBD and r^2 SCAN Enable (r^2 SCAN+MBD)@HF's Accuracy

Now, we examine the final piece of the puzzle—MBD's role in (r^2 SCAN+MBD)@HF's success. MBD has only one empirical parameter, β , which we set to unity in (r^2 SCAN+MBD)@HF. Increasing β weakens MBD contribution to interaction energies, vanishing as $\beta \rightarrow \infty$. Figure 6(b) shows how the interaction energy for the $\text{H}_2\text{O}-\text{Ca}^{2+}$ complex depends on the MBD parameter β for r^2 SCAN@HF, PBE0, and PBE0@HF, with the reference energy indicated by the horizontal dashed line. For (r^2 SCAN+MBD)@HF (green curve), the optimal β is slightly above 1, making our choice of $\beta = 1$ nearly optimal. In contrast, since PBE0 and PBE0@HF already overbind the complex, their β -curves never cross the reference line, instead favoring large β values (ideally $\beta \rightarrow \infty$) to suppress further overbinding from MBD.

A stark contrast appears in Fig. 6(c) for the neutral benzene–phenol complex. Here, PBE0 and

PBE0@HF require smaller β values (0.8–0.9) to match the reference—unlike the large β values favored previously (Fig. 6(b)). Remarkably, (r²SCAN+MBD)@HF again aligns closely with the reference at $\beta \approx 1$, highlighting its excellent consistency across charged and neutral NCIs.

Further analysis in the SI, covering additional systems (Fig. S15) and transferability tests of DFT+MBD trained on neutral NCIs and evaluated on charged NCIs and vice versa (Fig. S14) confirms the results of Figs. 6(b,c). Only when using r²SCAN@HF as a baseline does DFT+MBD exhibit consistently small sensitivity to β across different NCIs, allowing (r²SCAN+MBD)@HF to simultaneously maintain high accuracy for both neutral and charged interactions. In contrast, the greater sensitivity of other methods implies improving accuracy for one interaction type inevitably reduces it for the other. Thus, (r²SCAN+MBD)@HF's unique β -insensitivity makes it ideal for simulations of systems where different interaction types coexist, such as biomolecular complexes (e.g., Fig. 3).

To further motivate the choice of r²SCAN in our (r²SCAN+MBD)@HF, we have tested other functionals within the (DFT+MBD)@HF methodology for DES15K categories. To enable efficient comparison across many functionals, we have used a genetic algorithm (GA) to construct representative subsets of the metal-containing and neutral DES15K categories. These subsets have been selected such that they reproduce the error statistics of the full DES15K categories for the DFT methods on which we performed full-set calculations (Fig. S2 in the SI). These GA-selected subsets have enabled us to cost-effectively test a larger number of functionals for neutral and metal-containing DES15k complexes, and also to test higher-level methods (see below). Testing different functionals, such as M06 (87), TPSS (88), and B3LYP, within the (DFT+MBD)@HF methodology for both GA-selected neutral and metal-containing DES15K complexes confirm that (r²SCAN+MBD)@HF delivers the best overall performance (Fig. S3 in the SI). It is followed by other SCAN-family members [SCAN (89) and r²SCAN0 (90)], and then by (M06+MBD)@HF. Additional tests of (M06+MBD)@HF, such as those on our MIPC and on paradigmatic cases of Fig. 5, show that (r²SCAN+MBD)@HF performs substantially better across diverse NCIs than (M06+MBD)@HF (Fig. S5 and Fig. S7, respectively), despite the former being free of parameters fitted to data.

In terms of computational cost, (r²SCAN+MBD)@HF is comparable to hybrid functionals, since it combines semilocal r²SCAN with an HF calculation, resulting in an overall cost similar

to PBE0+MBD. To place our results in the broader context of DFT accuracy and cost, we have compared (r^2 SCAN+MBD)@HF with representative semilocal, hybrid, range-separated hybrid, and double-hybrid methods on three DES15K categories: the Alkaline earth subset and the GA-selected Alkali and Neutral subsets. As a semilocal (lowest-cost), we have included r^2 SCAN+D4 (33, 44), while at the hybrid level we have used PBE0+MBD and the (DFT+MBD)@HF variants mentioned earlier. At a higher cost than standard hybrids, we have selected ω B97M-V (91), the range-separated hybrid with nonlocal correlation (see Fig. S4 for a direct wall-time comparison with (r^2 SCAN+MBD)@HF on a representative MIPC cluster). Finally, we have included revDSD-PBEP86-D4 (92) as a representative double hybrid, the most computationally costly DFT class. ω B97M-V and revDSD-PBEP86-D4 were chosen because they were used in the literature to effectively model NCIs (78). Across the three DES15K categories considered (Alkaline earth, Alkali, and Neutral), (r^2 SCAN+MBD)@HF is the only method among all tested DFT approaches that achieves chemical accuracy (MAEs below 1 kcal/mol) simultaneously across all three categories, as shown in Fig. S4 in the SI. Notably, ω B97M-V fails to achieve this accuracy for the Alkaline earth subset, and revDSD-PBEP86-D4 does not achieve it for the Alkali subset (see also Fig. S6 in the SI for the corresponding error-distribution boxplots). These tests highlight the advantage of (r^2 SCAN+MBD)@HF even relative to more costly DFT methods for charged NCIs.

Discussion

Accurate yet tractable quantum-mechanical treatment of charged NCIs remains a key challenge with broad implications from biochemistry to materials. Dispersion-enhanced DFT has major difficulties in describing charged NCIs due to the interplay of density errors, coupling of polarization and dispersion, and the poor transferability of dispersion corrections trained on neutral systems. Improving charged NCIs accuracy without compromising neutral ones cannot be solved by merely adding charged systems to the dispersion method training. Instead, as we show here, it requires a foundational solution that balances correlation effects for both neutral and charged NCIs. Specifically, achieving this balance requires a dispersion method to accurately compensate for what the base DFT misses, which is easier for neutral NCIs but becomes highly nontrivial for charged ones. Restoring this balance is central to the integrative design of (r^2 SCAN+MBD)@HF for ac-

curate treatment of both charged and neutral NCIs. Crucially, altering any of its three components (r^2 SCAN, MBD, or HF densities) breaks this balance.

A distinguishing feature of (r^2 SCAN+MBD)@HF is its overall minimal empiricism, (i.e., it is free from empirical parameters fitted to data) reflected in each of its three components. Despite this, (r^2 SCAN+MBD)@HF improves accuracy for charged NCIs, even relative to higher level DFT approaches, such as ω B97M-V and revDSD-PBEP86-D4. It also maintains performance for neutral NCIs, and achieves accuracy comparable to leading semi-local functionals on main-group benchmarks (GMTKN55). This broader accuracy enables (r^2 SCAN+MBD)@HF to support applications from biochemistry to materials, where charged NCIs coexist with other interactions. For example, in biochemistry, accurate simulations of enzymatic reactions with a metallic cation in the active site require capturing both cation affinity and reaction kinetics (93, 94). These challenges often couple when cations stabilize transition states, as with Mg^{2+} in kinase-catalyzed phosphoryl transfers (93, 94). Thus, describing such systems requires a method that can capture both reaction barriers and NCIs involving the cation, making (r^2 SCAN+MBD)@HF a strong candidate for advancing enzymatic simulations. This is strongly supported by our MIPC results, which demonstrate our method's high accuracy for metal-protein interactions, including cases where the metal ion is surrounded by negatively charged ligands for which standard DFT can completely break down.

On the materials side, an example of interactions between metallic cations and neutral gases is the challenging adsorption of small molecules on open-metal-site (OMS) MOFs (95), which is another target for future applications of (r^2 SCAN+MBD)@HF. These systems are hard to simulate as classical force fields are unreliable near the OMS (96), and the recently developed MLFFs for MOFs display limitations due to the DFT training data (97). We expect that this problem can be addressed by training MLFFs on (r^2 SCAN+MBD)@HF data. More broadly, (r^2 SCAN+MBD)@HF can serve as a general-purpose DFT method for generating high-quality reference data to train MLFFs, particularly for charged systems, with (r^2 SCAN+MBD)@HF's forces implemented following Refs (98, 99).

Methods

DFT Calculations. All DFT calculations were performed using pySCF package (100), whereas ω B97M-V and double hybrid calculations were performed with ORCA 5.0.4 (101). For DES15K

calculations, we used def2-QZVPPD basis set (102), RI approximations are used with corresponding auxiliary basis sets (103) to accelerate the calculation. For MIPC calculations, we used def2-TZVPD basis set (102). The base functional r²SCAN@HF and PBE0@HF energies were evaluated on converged HF densities. The default grid level 5 in pySCF was used for all HF-DFT calculations. The MBD calculation were performed on converged HF and PBE0 densities using libMBD package (104). The empirical range-separation parameter $\beta = 1.0$ in MBD@rsSCS (27) was selected for all calculations, unless otherwise specified. The high-level reference data for Li⁺-benzene dissociation curve were reported in Ref (17). The reference data of other dissociation curves in Fig 5 are from S66x8 database (105).

Spherically averaged density that we used in Fig. 4 is defined as,

$$\tilde{\rho}(\mathbf{r}, u) = \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} \rho(\mathbf{r} + \mathbf{u}) \sin \phi \, d\theta \, d\phi,$$

and it satisfies

$$4\pi \int_0^\infty \tilde{\rho}(\mathbf{r}, u) u^2 \, du = N,$$

where N is the number of electrons.

The integrand $4\pi u^2 \tilde{\rho}(\mathbf{r}, u)$ is defined as radial density.

Spherically averaged deformation density is defined as

$$\Delta\tilde{\rho}(\mathbf{r}, u) = \tilde{\rho}_{\text{AB}}(\mathbf{r}, u) - \tilde{\rho}_{\text{A}}(\mathbf{r}, u) - \tilde{\rho}_{\text{B}}(\mathbf{r}, u),$$

and

$$4\pi \int_0^\infty \Delta\tilde{\rho}(\mathbf{r}, u) u^2 \, du = 0.$$

Clusters Creation. For the MIPC clusters in Fig. 3(a) containing negatively charged ligands, we have sourced Ca²⁺ and Mg²⁺ binding proteins from the non-redundant ccPDB 2.0 dataset (106), while the clusters including Zn²⁺, Na⁺ and K⁺ have been sourced from ion-specific classical force field development studies. (63, 107) Table S2 in the SI summarizes the dataset sources, chosen to limit redundancy and to make use of literature studies on metal-binding sites. From the initial cluster set, we excluded clusters with fewer than four ligands surrounding the metal cation. For the final MIPC subset with negatively charged ligands [Fig. 3(a)], we have used the highest-resolution representative for the most common ion-coordination profiles.

Our second MIPC cluster set of Fig. 3(b), is restricted to only contain neutral ion-binding residues, which reduces the available structures significantly. We queried the RCSB Protein Data Bank (RCSB PDB) for structures with resolution ≤ 2.5 Å as of 26 July 2024, then filtered out sites that contained charged residues within 5 Å of the metal ion. These criteria reduced the structure pool to 65 sites, further decreasing to 36 omitting those that have a coordination number less than 4. 26 non-redundant structures were manually selected to minimize structural redundancy.

The MIPC protein structures have been extracted from PDB and then completed with explicit hydrogen atoms by using CHARMM-GUI (108), relying on the HBUILD utility (109). Structures were used as is, without minimization. To create protein-ion clusters (i.e. cation with surrounding amino acid residues), we have made an initial selection of atoms within 3.5 Å of the ion. We then inspected atoms covalently bonded to this initial selection and decided, for each bond, whether the bonded atom should be included in the cluster or whether the bond should be cut. Bond cutting was allowed only at specific positions, namely, at single bonds between two sp^3 carbon atoms (and, in rare cases where this was not possible, at the C_α -N backbone bond). Any cut bond was capped with a hydrogen atom to produce the final cluster. The algorithm used for this purpose has been made publicly available on GitHub: https://github.com/bertadenes/qm_selector.

LNO-CCSD(T) reference interaction energies. CCSD(T) reference calculations were performed with our linear-scaling local natural orbital (LNO) CCSD(T) (12, 13, 73, 74) method in the MRCC (110–112) program suite. To accelerate the convergence towards the complete basis set (CBS) limit, we combine basis set extrapolation, counterpoise correction (113) and density-based basis set correction (DBBSC) (114). The CBS and local approximation free (LAF) limit of CCSD(T) is estimated for the MIPC set as

$$E_{\text{CCSD(T)}} = E_{\text{N-T LNO}}^{\text{TZ,DBBSC}} - E_{\text{Normal LNO}}^{\text{TZ,DBBSC}} + E_{\text{Normal LNO}}^{\text{CBS(T,Q),DBBSC}}. \quad (1)$$

Here, superscripts XZ refer to the cardinal number X of the aug-cc-pV($X+d$)Z basis sets (115), while the corresponding CBS extrapolated values are denoted as CBS($X, X+1$) (116). Subscripts Normal (N) and Tight (T) denote LNO thresholds, while N-T denotes LAF extrapolation towards canonical CCSD(T) using Normal and Tight LNO settings. (12, 13) The system specific uncertainty estimates corresponding to Eq. (1), plotted in Fig. 3, are the sum of uncertainty estimates for the basis set and LNO approximations. These are obtained, respectively, as the size of the DBBSC correction at the

539 CBS(T,Q) level and via the LAF framework (12, 13) as $\pm 0.5(E_{\text{Tight LNO}}^{\text{TZ,DBBSC}} - E_{\text{Normal LNO}}^{\text{TZ,DBBSC}})$. This level
540 of convergence provides, on the average, about ± 0.5 kcal/mol uncertainty for the LNO-CCSD(T)
541 interaction energies of the MIPC set. An additional benefit of the composite Eq. (1), in addition
542 to its robust and low uncertainty estimate, is its computational cost. The required LNO-CCSD(T)
543 computations took about 1–2 days of wall time with few 10s of CPU cores and ca. 50 GBs of
544 minimal memory requirement for the complexes of 40–70 atoms. LNO-CCSD(T) for the largest,
545 90–110 atom complexes required about twice as much resource. The reliability of the CCSD(T)
546 estimate of Eq. (1) is extensively validated against even better converged LNO-CCSD(T) results
547 for 5 representative protein-ion complexes (see Supplementary Information for additional details).

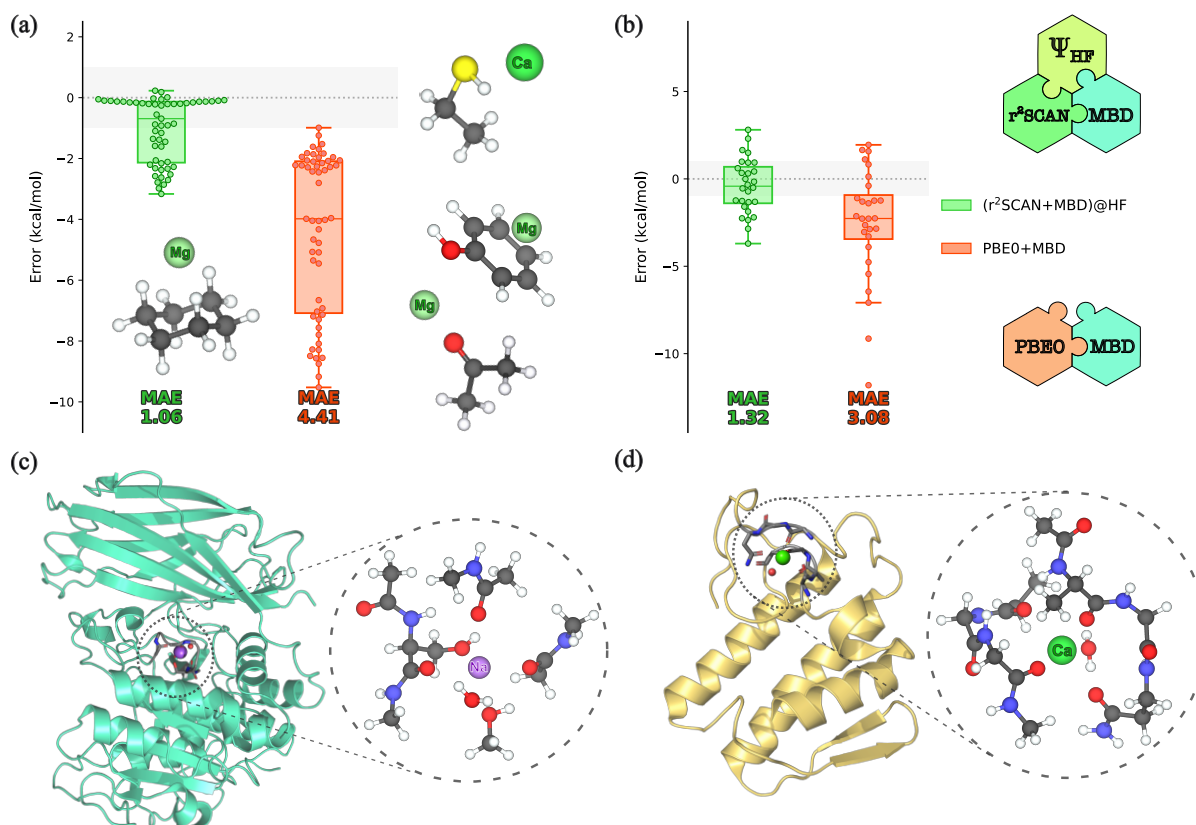


Figure 1: Beeswarm and boxplots with the interaction energy errors (kcal/mol) of $(r^2\text{SCAN}+\text{MBD})@HF$ and PBE0+MBD methods for dication complexes in (a) DES15K and (b) MIPC datasets, with numbers at the bottom denoting the mean absolute errors. Cartoon representations of proteins (c) 7O20 and (d) 1POD from the MIPC dataset, with zoomed-in views of their curated cluster structures. The labels for panels (c, d) correspond to Protein Data Bank (PDB) identifiers.

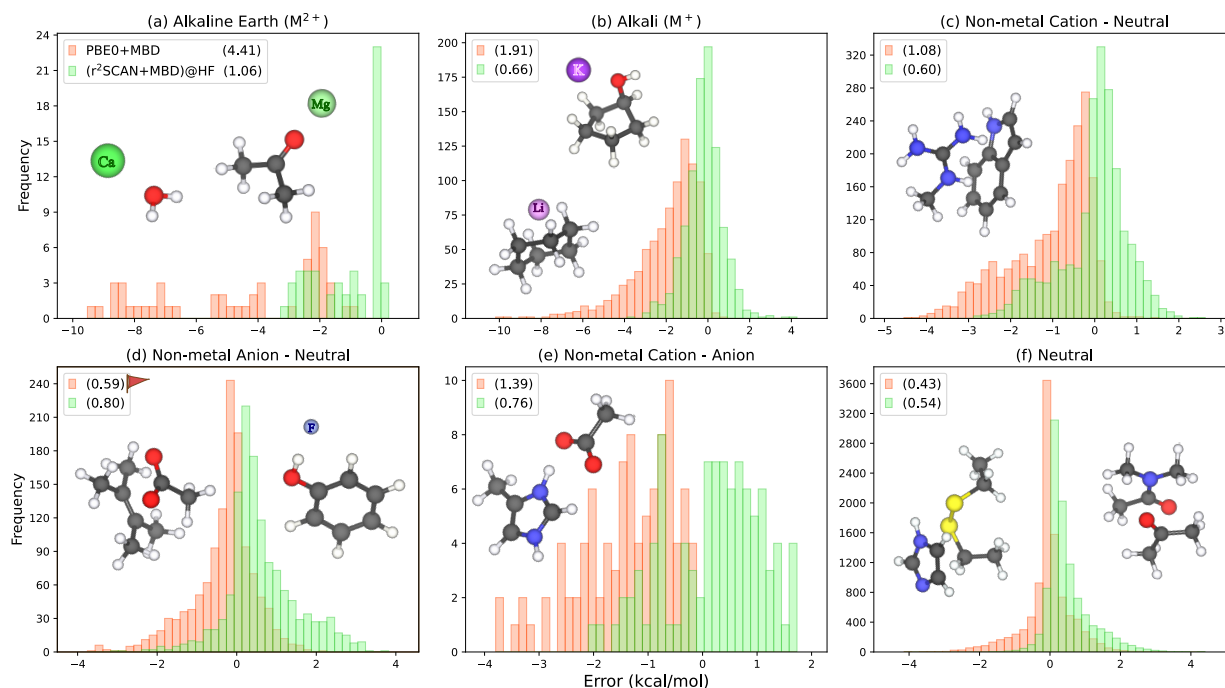


Figure 2: Histograms of error distribution for interaction energy predictions using (r²SCAN+MBD)@HF and PBE0+MBD methods for different subsets of DES15K dataset: (a) Alkaline earth dication complexes, (b) Alkali cation complexes, (c) Non-metal cation-neutral complexes, (d) Non-metal anion-neutral complexes, with the red flag indicating general and specific challenges of self-consistent DFT for anions (see text); (e) Non-metal cation-anion complexes, and (f) Neutral complexes. Numbers in parentheses represent corresponding mean absolute errors (MAEs) in kcal/mol. Negative errors indicate overbinding, while positive errors indicate underbinding.

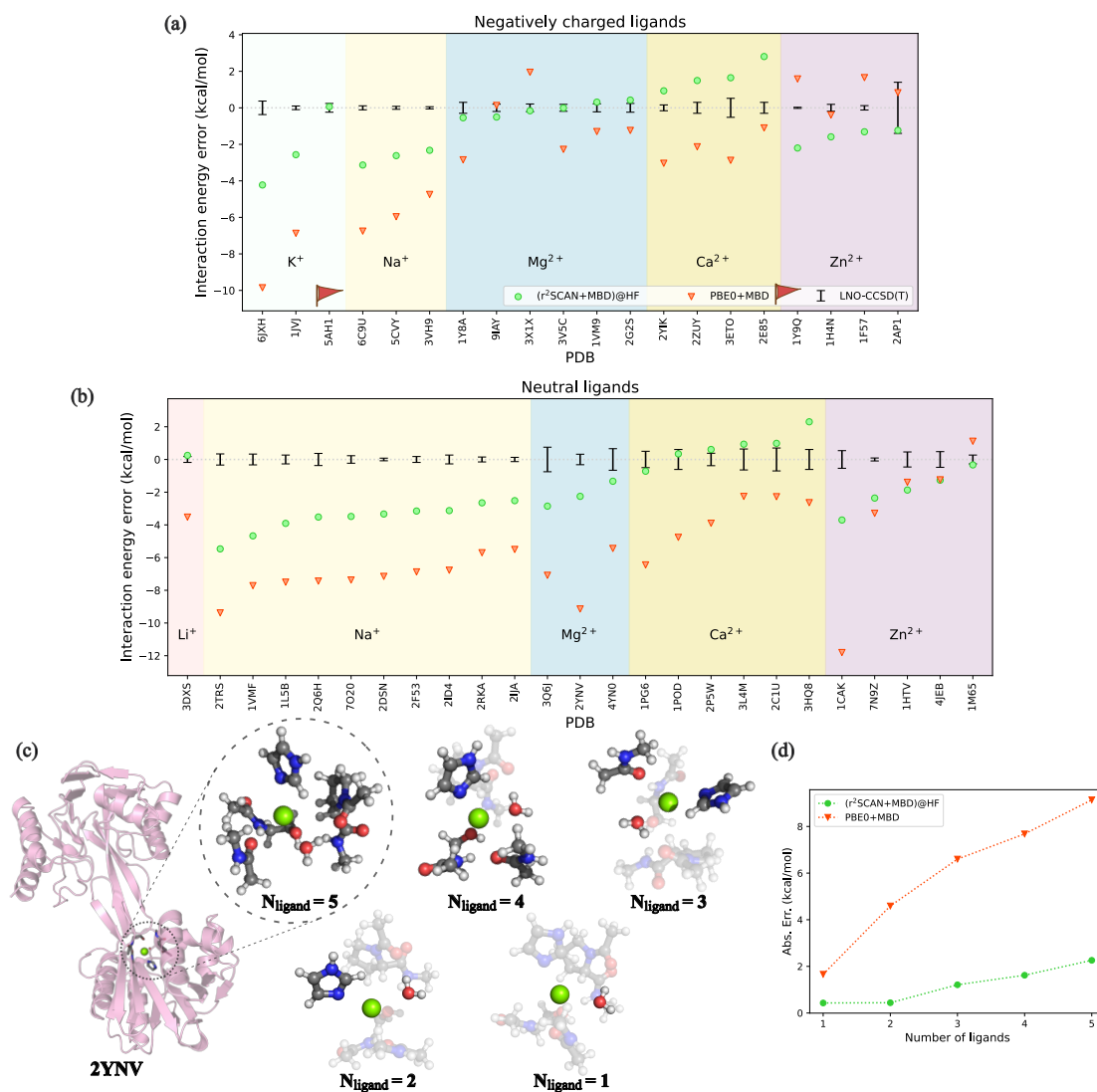


Figure 3: Comparison of the interaction energy errors of (r²SCAN+MBD)@HF and PBE0+MBD methods for the MIPC dataset, including both subset where the metallic cation is surrounded by (a) negatively charged and (b) neutral ligands, labeled by their PDB codes, with the red flag indicating the case for which the MBD part of PBE0+MBD could not be obtained (see text). Black error bars indicate uncertainty estimates of the LNO-CCSD(T) reference (see Methods). Representative protein structures (cartoon) and corresponding metal–ligand clusters (ball-and-stick) are shown in Fig. 1. (c) Subcluster generation for the Mg²⁺ cluster (PDB ID 2YNV): sequential ligand removal from the original cluster yielding subclusters with one to five coordinating ligands. (d) Absolute interaction energy errors of (r²SCAN+MBD)@HF and PBE0+MBD relative to the LNO-CCSD(T) reference for each subcluster in (c).

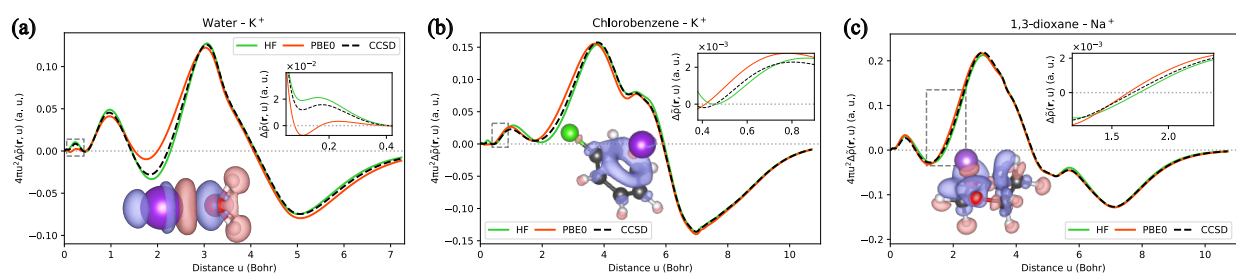


Figure 4: (a) Spherically averaged deformation density, $\Delta \tilde{\rho}(\mathbf{r}, u)$ (see Methods) for the H₂O-K⁺ complex calculated using HF, PBE0, and CCSD methods, in which $\mathbf{r}$ placed at the position of the K⁺ nucleus. Inset shows a zoom-in of this quantity in the highlighted (boxed) u region. The other inset shows the standard deformation density, $\Delta \rho(\mathbf{r})$, visualized as a 3D isosurface (isovalue 3.0×10^{-4} e/bohr³) (b, c) Same as in (a), but for chlorobenzene-K⁺ and 1,3-dioxane-Na⁺, respectively.

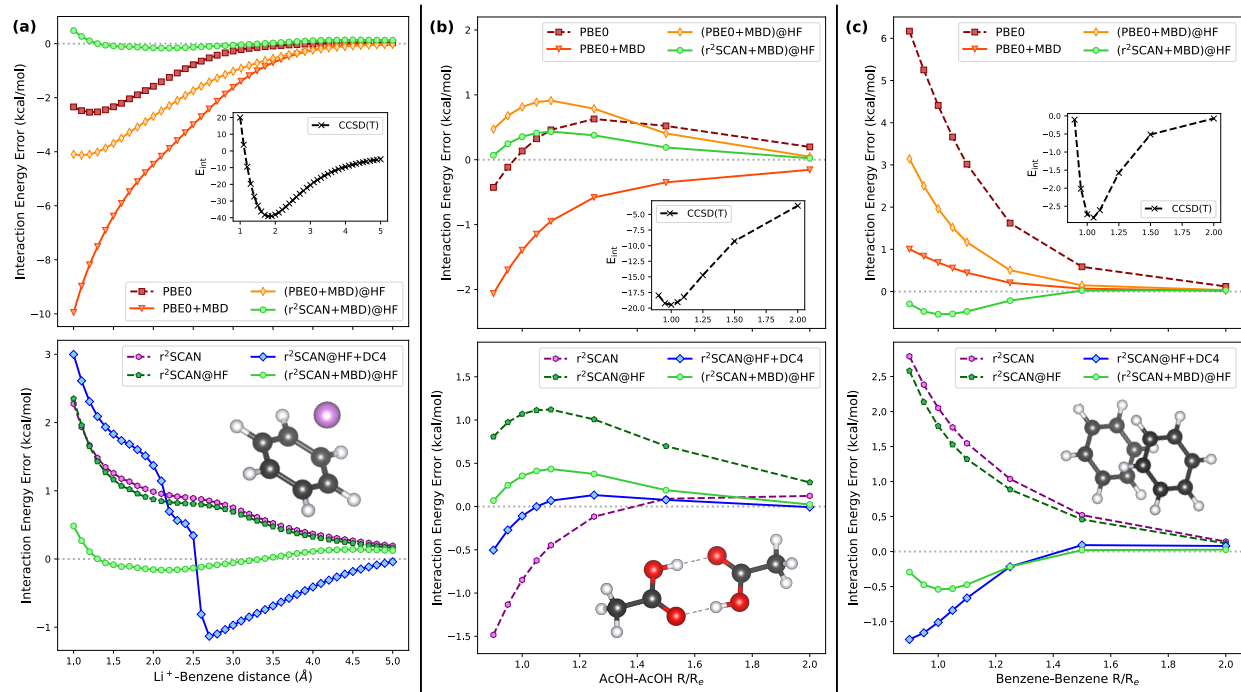


Figure 5: Errors in computed interaction energies relative to CCSD(T) reference data (interaction curves shown as insets, upper panels) for: (a) Li^+ –benzene as a function of cation–benzene distance ($\text{\AA}$); (b) acetic acid dimer; (c) benzene dimer, each as a function of separation normalized to equilibrium distance (R/R_e). Upper panels show errors progressing from bare PBE0, adding MBD (PBE0+MBD), using density-corrected version ((PBE0+MBD)@HF), to finally $(\text{r}^2\text{SCAN}+\text{MBD})@\text{HF}$. Lower panels similarly progress from bare r^2SCAN , density-corrected $\text{r}^2\text{SCAN}@\text{HF}$, previously reported $\text{r}^2\text{SCAN}@\text{HF}+\text{DC4}$, to $(\text{r}^2\text{SCAN}+\text{MBD})@\text{HF}$. As we progress along this sequence of functionals towards $(\text{r}^2\text{SCAN}+\text{MBD})@\text{HF}$, the interaction energy errors systematically decrease.

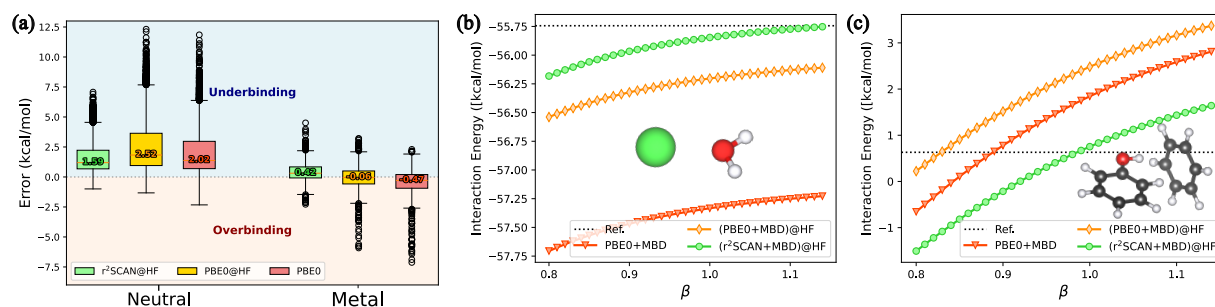


Figure 6: (a) Boxplot of interaction energy errors (kcal/mol) for r^2 SCAN@HF, PBE0@HF, and PBE0 across Neutral and Metal subsets of DES15K. Boxes show interquartile ranges, medians (lines), outliers (circles), and mean errors (numbers). (b, c) Dependence of DFT+MBD interaction energies on the MBD parameter β for (b) $\text{H}_2\text{O}-\text{Ca}^{2+}$ and (c) benzene–phenol complexes. Reference interaction energies (CCSD(T)) are indicated by dotted lines.

References and Notes

1. V. A. Adhav, K. Saikrishnan, The realm of unconventional noncovalent interactions in proteins: their significance in structure and function. *Acs Omega* **8** (25), 22268–22284 (2023).
2. R. W. Newberry, R. T. Raines, Secondary forces in protein folding. *ACS chemical biology* **14** (8), 1677–1686 (2019).
3. M. Rossi, A. Tkatchenko, S. B. Rempe, S. Varma, Role of methyl-induced polarization in ion binding. *Proceedings of the National Academy of Sciences* **110** (32), 12978–12983 (2013).
4. S. Yamada, Cation- π interactions in organic synthesis. *Chemical reviews* **118** (23), 11353–11432 (2018).
5. G. Maurin, C. Serre, A. Cooper, G. Férey, The new age of MOFs and of their porous-related solids. *Chemical Society Reviews* **46** (11), 3104–3107 (2017).
6. J.-X. Huang, G. Csányi, J.-B. Zhao, J. Cheng, V. L. Deringer, First-principles study of alkali-metal intercalation in disordered carbon anode materials. *Journal of Materials Chemistry A* **7** (32), 19070–19080 (2019).
7. K. Fumino, S. Reimann, R. Ludwig, Probing molecular interaction in ionic liquids by low frequency spectroscopy: Coulomb energy, hydrogen bonding and dispersion forces. *Physical Chemistry Chemical Physics* **16** (40), 21903–21929 (2014).
8. B. Huang, *et al.*, Non-covalent interactions in electrochemical reactions and implications in clean energy applications. *Physical Chemistry Chemical Physics* **20** (23), 15680–15686 (2018).
9. J. Rezac, P. Hobza, Describing noncovalent interactions beyond the common approximations: how accurate is the “gold standard,” CCSD (T) at the complete basis set limit? *Journal of chemical theory and computation* **9** (5), 2151–2155 (2013).
10. W. M. Foulkes, L. Mitas, R. Needs, G. Rajagopal, Quantum Monte Carlo simulations of solids. *Reviews of Modern Physics* **73** (1), 33 (2001).

11. C. Riplinger, F. Neese, An efficient and near linear scaling pair natural orbital based local coupled cluster method. *The Journal of chemical physics* **138** (3), 034106 (2013).
12. P. R. Nagy, M. Kállay, Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods. *J. Chem. Theory Comput.* **15**, 5275 (2019), doi:10.1021/acs.jctc.9b00511.
13. P. R. Nagy, State-of-the-art local correlation methods enable accurate and affordable gold standard quantum chemistry up to a few hundred atoms. *Chem. Sci.* **15**, 14556 (2024), doi:10.1039/D4SC04755A.
14. A. Zen, *et al.*, Fast and accurate quantum Monte Carlo for molecular crystals. *Proceedings of the National Academy of Sciences* **115** (8), 1724–1729 (2018).
15. S. Grimme, A. Hansen, J. G. Brandenburg, C. Bannwarth, Dispersion-corrected mean-field electronic structure methods. *Chemical reviews* **116** (9), 5105–5154 (2016).
16. P. Xu, M. Alkan, M. S. Gordon, Many-body dispersion. *Chemical Reviews* **120** (22), 12343–12356 (2020).
17. C. J. Nickerson, K. R. Bryenton, A. J. Price, E. R. Johnson, Comparison of density-functional theory dispersion corrections for the DES15K database. *The Journal of Physical Chemistry A* **127** (41), 8712–8722 (2023).
18. E. Sim, S. Song, S. Vuckovic, K. Burke, Improving results by improving densities: Density-corrected density functional theory. *Journal of the American Chemical Society* **144** (15), 6625–6639 (2022).
19. E. R. Johnson, A. D. Becke, A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *The Journal of chemical physics* **124** (17), 174104 (2006).
20. A. D. Becke, E. R. Johnson, Exchange-hole dipole moment and the dispersion interaction revisited. *The Journal of chemical physics* **127** (15), 154108 (2007).

21. A. Tkatchenko, M. Scheffler, Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density; and Free-Atom Reference Data. *Physical review letters* **102** (7), 073005 (2009).
22. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of chemical physics* **132** (15), 154104 (2010).
23. S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory. *Journal of computational chemistry* **32** (7), 1456–1465 (2011).
24. J. Klimeš, D. R. Bowler, A. Michaelides, Chemical accuracy for the van der Waals density functional. *Journal of Physics: Condensed Matter* **22** (2), 022201 (2009).
25. O. A. Vydrov, T. Van Voorhis, Nonlocal van der Waals density functional: The simpler the better. *The Journal of chemical physics* **133** (24), 244103 (2010).
26. A. Tkatchenko, R. A. DiStasio Jr, R. Car, M. Scheffler, Accurate and efficient method for many-body van der Waals interactions. *Physical review letters* **108** (23), 236402 (2012).
27. A. Ambrosetti, A. M. Reilly, R. A. DiStasio, A. Tkatchenko, Long-range correlation energy calculated from coupled atomic response functions. *The Journal of chemical physics* **140** (18), 18A508 (2014).
28. S. Spicher, E. Caldeweyher, A. Hansen, S. Grimme, Benchmarking London dispersion corrected density functional theory for noncovalent ion– π interactions. *Physical Chemistry Chemical Physics* **23** (20), 11635–11648 (2021).
29. V. Wineman-Fisher, Y. Al-Hamdani, P. R. Nagy, A. Tkatchenko, S. Varma, Improved description of ligand polarization enhances transferability of ion–ligand interactions. *The Journal of Chemical Physics* **153** (9), 094115 (2020).
30. I. Sandler, S. Sharma, B. Chan, J. Ho, Accurate quantum chemical prediction of gas-phase anion binding affinities and their structure-binding relationships. *The Journal of Physical Chemistry A* **125** (45), 9838–9851 (2021).

31. A. G. Donchev, *et al.*, Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Scientific data* **8** (1), 55 (2021).
32. E. Caldeweyher, *et al.*, A generally applicable atomic-charge dependent London dispersion correction. *The Journal of chemical physics* **150** (15), 154122 (2019).
33. E. Caldeweyher, J.-M. Mewes, S. Ehlert, S. Grimme, Extension and evaluation of the D4 London-dispersion model for periodic systems. *Physical Chemistry Chemical Physics* **22** (16), 8499–8512 (2020), doi:10.1039/d0cp00502a, <http://dx.doi.org/10.1039/d0cp00502a>.
34. A. Kleshchonok, A. Tkatchenko, Tailoring van der Waals dispersion interactions with external electric charges. *Nature communications* **9** (1), 3017 (2018).
35. K. R. Bryenton, A. A. Adeleke, S. G. Dale, E. R. Johnson, Delocalization error: The greatest outstanding challenge in density-functional theory. *WIREs Computational Molecular Science* **13** (2), e1631 (2023), doi:10.1002/wcms.1631, <https://wires.onlinelibrary.wiley.com/doi/10.1002/wcms.1631>.
36. E. Palos, A. Caruso, F. Paesani, Consistent density functional theory-based description of ion hydration through density-corrected many-body representations. *The Journal of Chemical Physics* **159** (18) (2023), doi:10.1063/5.0174577, <http://dx.doi.org/10.1063/5.0174577>.
37. K. Pernal, R. Podeszwa, K. Patkowski, K. Szalewicz, Dispersionless density functional theory. *Physical review letters* **103** (26), 263201 (2009).
38. A. J. Price, K. R. Bryenton, E. R. Johnson, Requirements for an accurate dispersion-corrected density functional. *The Journal of Chemical Physics* **154** (23), 230902 (2021).
39. O. T. Unke, *et al.*, Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments. *Science Advances* **10** (14), eadn4397 (2024).
40. O. T. Unke, *et al.*, Machine learning force fields. *Chemical Reviews* **121** (16), 10142–10186 (2021).

41. P. Montero de Hijos, C. Dellago, R. Jinnouchi, G. Kresse, Density isobar of water and melting temperature of ice: Assessing common density functionals. *The Journal of Chemical Physics* **161** (13), 131102 (2024).
42. E. Palos, *et al.*, Assessing the interplay between functional-driven and density-driven errors in DFT models of water. *Journal of Chemical Theory and Computation* **18** (6), 3410–3426 (2022).
43. T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nature communications* **12** (1), 398 (2021).
44. J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, J. Sun, Accurate and Numerically Efficient r2SCAN Meta-Generalized Gradient Approximation. *The Journal of Physical Chemistry Letters* **11** (19), 8208–8215 (2020), doi:10.1021/acs.jpclett.0c02405, <http://dx.doi.org/10.1021/acs.jpclett.0c02405>.
45. J. Sun, A. Ruzsinszky, J. P. Perdew, Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Physical Review Letters* **115** (3) (2015), doi:10.1103/physrevlett.115.036402, <http://dx.doi.org/10.1103/PhysRevLett.115.036402>.
46. P. M. Gill, B. G. Johnson, J. A. Pople, M. J. Frisch, The performance of the Becke—Lee—Yang—Parr (B—LYP) density functional theory with various basis sets. *Chemical Physics Letters* **197** (4-5), 499–505 (1992).
47. E. Sim, S. Song, K. Burke, Quantifying density errors in DFT. *The journal of physical chemistry letters* **9** (22), 6385–6392 (2018).
48. M.-C. Kim, E. Sim, K. Burke, Understanding and reducing errors in density functional calculations. *Physical review letters* **111** (7), 073003 (2013).
49. S. Vuckovic, S. Song, J. Kozłowski, E. Sim, K. Burke, Density functional analysis: The theory of density-corrected DFT. *Journal of chemical theory and computation* **15** (12), 6636–6646 (2019).

50. S. Song, S. Vuckovic, E. Sim, K. Burke, Density-corrected DFT explained: Questions and answers. *Journal of chemical theory and computation* **18** (2), 817–827 (2022).
51. S. Song, S. Vuckovic, E. Sim, K. Burke, Density sensitivity of empirical functionals. *The journal of physical chemistry letters* **12** (2), 800–807 (2021).
52. S. Song, *et al.*, Extending density functional theory with near chemical accuracy beyond pure water. *Nature Communications* **14** (1), 799 (2023), doi:10.1038/s41467-023-36094-y, <https://www.nature.com/articles/s41467-023-36094-y>.
53. M. Lee, *et al.*, Correcting dispersion corrections with density-corrected DFT. *Journal of Chemical Theory and Computation* **20** (16), 7155–7167 (2024).
54. T. Gould, B. Chan, S. G. Dale, S. Vuckovic, Identifying and embedding transferability in data-driven representations of chemical space. *Chemical Science* (2024), doi:10.1039/D4SC02358G, <https://xlink.rsc.org/?DOI=D4SC02358G>.
55. J. Hoja, *et al.*, QM7-X, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules. *Scientific data* **8** (1), 43 (2021).
56. L. Medrano Sandonas, *et al.*, Dataset for quantum-mechanical exploration of conformers and solvent effects in large drug-like molecules. *Scientific Data* **11** (1), 742 (2024).
57. A. Bauza, I. Alkorta, A. Frontera, J. Elguero, On the reliability of pure and hybrid DFT methods for the evaluation of halogen, chalcogen, and pnictogen bonds involving anionic and neutral electron donors. *Journal of chemical theory and computation* **9** (11), 5201–5210 (2013).
58. L. Goerigk, *et al.*, A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Physical Chemistry Chemical Physics* **19** (48), 32184–32215 (2017).
59. N. H. Williams, Magnesium ion catalyzed ATP hydrolysis. *Journal of the American Chemical Society* **122** (48), 12023–12024 (2000).

60. H. Noguchi, *et al.*, The structure and conformational switching of Rap1B. *Biochemical and Biophysical Research Communications* **462** (1), 46–51 (2015).
61. J. Bröker, *et al.*, Discovery of BI-2493, a pan-KRAS inhibitor showing in vivo efficacy. *Journal of Medicinal Chemistry* **68** (15), 15649–15668 (2025).
62. J. H. Laity, B. M. Lee, P. E. Wright, Zinc finger proteins: new insights into structural and functional diversity. *Current opinion in structural biology* **11** (1), 39–46 (2001).
63. Z. Yu, P. Li, K. M. Merz, Extended Zinc AMBER Force Field (EZAFF). *Journal of Chemical Theory and Computation* **14** (1), 242–254 (2017), doi:10.1021/acs.jctc.7b00773, <http://dx.doi.org/10.1021/acs.jctc.7b00773>.
64. I. Dokmanić, M. Šikić, S. Tomić, Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination. *Acta Crystallographica Section D Biological Crystallography* **64** (3), 257–263 (2008), doi:10.1107/s090744490706595x, <http://dx.doi.org/10.1107/S090744490706595X>.
65. V. Putignano, A. Rosato, L. Banci, C. Andreini, MetalPDB in 2018: a database of metal sites in biological macromolecular structures. *Nucleic Acids Research* **46** (D1), D459–D464 (2017), doi:10.1093/nar/gkx989, <http://dx.doi.org/10.1093/nar/gkx989>.
66. T. R. Soderling, J. T. Stull, Structure and Regulation of Calcium/Calmodulin-Dependent Protein Kinases. *Chemical Reviews* **101** (8), 2341–2352 (2001), doi:10.1021/cr0002386, <http://dx.doi.org/10.1021/cr0002386>.
67. D. L. Scott, *et al.*, Structures of free and inhibited human secretory phospholipase A2 from inflammatory exudate. *Science* **254** (5034), 1007–1010 (1991).
68. A. Teplyakov, *et al.*, Crystal structure of the Escherichia coli YcdX protein reveals a trinuclear zinc active site. *Proteins: Structure, Function, and Bioinformatics* **51** (2), 315–318 (2003).
69. A. Echaliier, C. F. Goodhew, G. W. Pettigrew, V. Fülöp, Activation and catalysis of the di-heme cytochrome c peroxidase from Paracoccus pantotrophus. *Structure* **14** (1), 107–117 (2006).

70. S. O. Dahms, T. Haider, G. Klebe, T. Steinmetzer, H. Brandstetter, OFF-state-specific inhibition of the proprotein convertase furin. *ACS Chemical Biology* **16** (9), 1692–1700 (2021).
71. S. K. Singh, A. Yamashita, E. Gouaux, Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. *Nature* **448** (7156), 952–956 (2007).
72. H. Matsumura, *et al.*, Novel cation- π interaction revealed by crystal structure of thermoalkalophilic lipase. *Proteins: Structure, Function, and Bioinformatics* **70** (2), 592–598 (2008).
73. P. R. Nagy, M. Kállay, Optimization of the linear-scaling local natural orbital CCSD(T) method: Redundancy-free triples correction using Laplace transform. *J. Chem. Phys.* **146**, 214106 (2017).
74. P. R. Nagy, G. Samu, M. Kállay, Optimization of the linear-scaling local natural orbital CCSD(T) method: Improved algorithm and benchmark applications. *J. Chem. Theory Comput.* **14**, 4193 (2018).
75. A. Mirtschink, C. J. Umrigar, J. D. Morgan, P. Gori-Giorgi, Energy density functionals from the strong-coupling limit applied to the anions of the He isoelectronic series. *The Journal of Chemical Physics* **140** (18) (2014), doi:10.1063/1.4871018, <http://dx.doi.org/10.1063/1.4871018>.
76. J. Hermann, *et al.*, libMBD: A general-purpose package for scalable quantum many-body dispersion calculations. *The Journal of Chemical Physics* **159** (17), 174802 (2023).
77. V. Wineman-Fisher, *et al.*, Transferable interactions of Li⁺ and Mg²⁺ ions in polarizable models. *The Journal of Chemical Physics* **153** (10), 104113 (2020).
78. J. Chen, B. Chan, Y. Shao, J. Ho, How accurate are approximate quantum chemical methods at modelling solute–solvent interactions in solvated clusters? *Physical Chemistry Chemical Physics* **22** (7), 3855–3866 (2020), doi:10.1039/c9cp06792b, <http://dx.doi.org/10.1039/C9CP06792B>.
79. S. A. Slattery, J. C. Yon, E. F. Valeev, Revisiting Artifacts of Kohn–Sham Density Functionals for Biosimulation. *Journal of Chemical Theory and Computation* **20** (15), 6652–6660 (2024), doi:10.1021/acs.jctc.4c00712, <http://dx.doi.org/10.1021/acs.jctc.4c00712>.

80. S. Nam, S. Song, E. Sim, K. Burke, Measuring density-driven errors using Kohn–Sham inversion. *Journal of chemical theory and computation* **16** (8), 5014–5023 (2020).
81. A. D. Kaplan, C. Shahi, P. Bhetwal, R. K. Sah, J. P. Perdew, Understanding density-driven errors for reaction barrier heights. *Journal of Chemical Theory and Computation* **19** (2), 532–543 (2023).
82. B. Kanungo, A. D. Kaplan, C. Shahi, V. Gavini, J. P. Perdew, Unconventional error cancellation explains the success of Hartree–Fock density functional theory for barrier heights. *The Journal of Physical Chemistry Letters* **15** (1), 323–328 (2024).
83. H. Bahmann, Y. Zhou, M. Ernzerhof, The shell model for the exchange-correlation hole in the strong-correlation limit. *The Journal of Chemical Physics* **145**, 124104 (2016), doi:10.1063/1.4962738, <https://pubs.aip.org/jcp/article/145/12/124104/290849/The-shell-model-for-the-exchange-correlation-hole>.
84. S. Vuckovic, P. Gori-Giorgi, Simple Fully Nonlocal Density Functionals for Electronic Repulsion Energy. *The Journal of Physical Chemistry Letters* **8**, 2799–2805 (2017), doi:10.1021/acs.jpclett.7b01113, <https://pubs.acs.org/doi/10.1021/acs.jpclett.7b01113>.
85. S. Vuckovic, H. Bahmann, Nonlocal Functionals Inspired by the Strongly Interacting Limit of DFT: Exact Constraints and Implementation. *Journal of Chemical Theory and Computation* **19**, 6172–6184 (2023), doi:10.1021/acs.jctc.3c00437, <https://pubs.acs.org/doi/10.1021/acs.jctc.3c00437>.
86. T. Froitzheim, M. Müller, A. Hansen, S. Grimme, The Bond Capacity Electronegativity Equilibration Charge Model (EEQBC) for the Elements Z= 1–103 (2025).
87. Y. Zhao, D. G. Truhlar, 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. *Theoretical chemistry accounts* **120** (1), 215–241 (2008).

88. J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Physical review letters* **91** (14), 146401 (2003).
89. J. Sun, A. Ruzsinszky, J. P. Perdew, Strongly constrained and appropriately normed semilocal density functional. *Physical review letters* **115** (3), 036402 (2015).
90. M. Bursch, H. Neugebauer, S. Ehlert, S. Grimme, Dispersion corrected r2SCAN based global hybrid functionals: r2SCANh, r2SCAN0, and r2SCAN50. *The Journal of Chemical Physics* **156** (13) (2022).
91. N. Mardirossian, M. Head-Gordon, ω B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *The Journal of chemical physics* **144** (21), 214110 (2016).
92. G. Santra, N. Sylvetsky, J. M. Martin, Minimally empirical double-hybrid functionals trained against the GMTKN55 database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4. *The Journal of Physical Chemistry A* **123** (24), 5129–5143 (2019).
93. R. Recabarren, K. Zinovjev, I. Tuñón, J. Alzate-Morales, How a Second Mg²⁺ Ion Affects the Phosphoryl-Transfer Mechanism in a Protein Kinase: A Computational Study. *ACS Catalysis* **11** (1), 169–183 (2020), doi:10.1021/acscatal.0c03304.
94. K. Nam, *et al.*, Magnesium induced structural reorganization in the active site of adenylate kinase. *Science Advances* **10** (32), eado5504 (2024), doi:10.1126/sciadv.ado5504.
95. U. Kökçam-Demir, *et al.*, Coordinatively unsaturated metal sites (open metal sites) in metal-organic frameworks: design and applications. *Chemical Society Reviews* **49** (9), 2751–2798 (2020), doi:10.1039/c9cs00609e.
96. A. L. Dzubak, *et al.*, Ab initio carbon capture in open-site metal-organic frameworks. *Nature Chemistry* **4** (10), 810–816 (2012), doi:10.1038/nchem.1432.
97. R. Goeminne, L. Vanduyfhuys, V. Van Speybroeck, T. Verstraelen, DFT-Quality adsorption simulations in metal-organic frameworks enabled by machine learning Potentials. *Journal of Chemical Theory and Computation* **19** (18), 6313–6325 (2023).

98. B. Rana, M. P. Coons, J. M. Herbert, Detection and Correction of Delocalization Errors for Electron and Hole Polarons Using Density-Corrected DFT. *The Journal of Physical Chemistry Letters* **13** (23), 5275–5284 (2022), doi:10.1021/acs.jpcllett.2c01187.
99. F. Belleflamme, J. Hutter, Radicals in aqueous solution: assessment of density-corrected SCAN functional. *Physical Chemistry Chemical Physics* **25** (31), 20817–20836 (2023), doi:10.1039/d3cp02517a.
100. Q. Sun, *et al.*, Recent developments in the PySCF program package. *The Journal of chemical physics* **153** (2), 024109 (2020).
101. F. Neese, Software update: the ORCA program system, version 5.0. *WIREs Comput. Molec. Sci.* **12** (1), e1606 (2022), doi:10.1002/wcms.1606.
102. F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **7** (18), 3297–3305 (2005).
103. F. Weigend, Hartree–Fock Exchange Fitting Basis Sets for H to Rn. *J. Comput. Chem.* **29**, 167 (2008).
104. J. Hermann, R. A. DiStasio Jr, A. Tkatchenko, First-principles models for van der Waals interactions in molecules and materials: Concepts, theory, and applications. *Chemical Reviews* **117** (6), 4714–4758 (2017).
105. B. Brauer, M. K. Kesharwani, S. Kozuch, J. M. Martin, The S66x8 benchmark for noncovalent interactions revisited: explicitly correlated ab initio methods and density functional theory. *Physical Chemistry Chemical Physics* **18** (31), 20905–20925 (2016).
106. P. Agrawal, *et al.*, ccPDB 2.0: an updated version of datasets created and compiled from Protein Data Bank. *Database* **2019** (2019), doi:10.1093/database/bay142, <http://dx.doi.org/10.1093/database/bay142>.
107. J. A. Delgado, V. Wineman-Fisher, S. Pandit, S. Varma, Inclusion of High-Field Target Data in AMOEBA's Calibration Improves Predictions of Protein–Ion Interactions. *Journal*

of Chemical Information and Modeling **62** (19), 4713–4726 (2022), doi:10.1021/acs.jcim.2c00758, <http://dx.doi.org/10.1021/acs.jcim.2c00758>.

108. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry* **29** (11), 1859–1865 (2008).

109. A. T. Brünger, M. Karplus, Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison. *Proteins: Structure, Function, and Bioinformatics* **4** (2), 148–156 (1988), doi:10.1002/prot.340040208, <http://dx.doi.org/10.1002/prot.340040208>.

110. M. Kállay, *et al.*, The MRCC program system: Accurate quantum chemistry from water to proteins. *J. Chem. Phys.* **152**, 074107 (2020).

111. D. Mester, *et al.*, An overview of developments in the MRCC program system. *J. Phys. Chem. A* **129**, 2086 (2025).

112. M. Kállay, *et al.* MRCC, a quantum chemical program suite. See <https://www.mrcc.hu/> (Accessed August 1, 2023).

113. S. F. Boys, F. Bernardi, The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **19**, 553 (1970).

114. D. Mester, P. R. Nagy, M. Kállay, Basis-set limit CCSD(T) energies for large molecules with local natural orbitals and reduced-scaling basis-set corrections. *J. Chem. Theory Comput.* **20**, 7453 (2024).

115. T. H. Dunning Jr., K. A. Peterson, A. K. Wilson, Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *J. Chem. Phys.* **114**, 9244 (2001).

116. T. Helgaker, W. Klopper, H. Koch, J. Noga, Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **106**, 9639 (1997).

117. P. R. Nagy, G. Samu, M. Kállay, An integral-direct linear-scaling second-order Møller–Plesset approach. *J. Chem. Theory Comput.* **12**, 4897 (2016).

118. B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning, A. K. Wilson, Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg. *Theor. Chem. Acc.* **128**, 69 (2011).
119. J. G. Hill, K. A. Peterson, Gaussian basis sets for use in correlated molecular calculations. XI. Pseudopotential-based and all-electron relativistic basis sets for alkali metal (K-Fr) and alkaline earth (Ca-Ra) elements. *J. Chem. Phys.* **147**, 244106 (2017).
120. D. Figgen, G. Rauhut, M. Dolg, H. Stoll, Energy-consistent pseudopotentials for group 11 and 12 atoms: adjustment to multi-configuration Dirac–Hartree–Fock data. *Chem. Phys.* **311**, 227 (2005).
121. A. Karton, J. M. L. Martin, Comment on: “Estimating the Hartree–Fock limit from finite basis set calculations”. *Theor. Chem. Acc.* **115**, 330 (2006).
122. E. Giner, *et al.*, Curing basis-set convergence of wave-function theory using density-functional theory: A systematically improvable approach. *J. Chem. Phys.* **149**, 194301 (2018).
123. F. Weigend, A. Köhn, C. Hättig, Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **116**, 3175 (2002).
124. C. Hättig, Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core-valence and quintuple- ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.* **7**, 59 (2005).
125. J. G. Hill, Auxiliary basis sets for density-fitting second-order Møller–Plesset perturbation theory: Weighted core-valence correlation consistent basis sets for the 4d elements Y-Pd. *J. Comput. Chem.* **34**, 2168 (2013), doi:10.1002/jcc.23372.

Acknowledgments

Funding: S.V. and H.Z. acknowledge funding from the SNSF Starting Grant project (TMSGI2.211246), as well as computational resources from the Swiss National Supercomputing Centre (CSCS; project ID s1306). B.D.L. and P.R.N. acknowledge funding of the János Bolyai Research Scholarship of the

Hungarian Academy of Sciences and the ERC Starting Grant No. 101076972, “aCCuracy”. D.B. and P.R.N. acknowledge support by the National Research, Development, and Innovation (NRDI) Office (Starting 149434, FK142489). A.T. acknowledges the ERC Advanced Grant “FITMOL”. T.H. acknowledges an AFR grant from the Luxembourg National Research Fund. The Hungarian Governmental Information-Technology Development Agency is acknowledged for awarding access to the Komondor and the LEONARDO supercomputer, owned by the EuroHPC Joint Undertaking, hosted by CINECA (Italy) and the LEONARDO consortium.

Author contributions: H.Z. and S.V. formulated the project. H.Z. performed all DFT calculations. B.D.L. and P.R.N. performed the LNO-CCSD(T) calculations on the MIPC dataset. T.H. and D.B. curated the the MIPC dataset. H.Z., B.D.L., T.H., D.B., P.R.N., A.T., and S.V. wrote the manuscript with input from all authors. S.V., P.R.N., and A.T. supervised the project.

Competing interests: There are no competing interests to declare.

Supplementary materials

Materials and Methods

Figs. S1 to S13

Tables S1 to S4

References (93-125)