DALHOUSIE UNIVERSITY

Analysis of Density-Functional Errors for Non-Covalent Interactions between Charged Molecules

A. Otero-de-la-Roza*,† and Erin R. Johnson*,‡

Departamento de Química Física y Analítica and MALTA-Consolider Team, Facultad de Química, Universidad de Oviedo, 33006 Oviedo, Spain, and Department of Chemistry, Dalhousie University, 6274 Coburg Rd, Halifax, Nova Scotia, B3H 4R2, Canada

E-mail: aoterodelaroza@gmail.com; erin.johnson@dal.ca

Abstract

The study of the structure and chemistry of biological systems with density-functional theory requires an accurate description of intermolecular interactions involving charged moieties. While dispersion-corrected functionals accurately model non-covalent interactions in neutral systems, a systematic study of the performance and errors associated with intermolecular interactions between charged fragments is missing. We undertake this study by examining the performance of a series of dispersion-corrected functionals with varying degrees of exact exchange for the side-chain protein interactions from the BioFragment Database (BFDb) of Burns et al. (the SSI set). In general, hybrid functionals with 20-30% exact exchange are accurate across the board, with the lowest mean absolute errors of 0.11 kcal/mol obtained from the 20% exact-exchange BLYP and PW86PBE hybrids coupled with the exchange-hole dipole moment (XDM) dispersion model. In addition, our analysis shows that functionals with higher exact-exchange fractions overestimate the electrostatic contributions to the binding energies, and that GGA functionals overestimate zwitterion binding energies due to delocalisation error and overestimated charge transfer. In addition, the (quite large) repulsion in the dications is systematically overestimated by all functionals and the trends for the monoanionic and dianionic dimers can be successfully explained by appealing to the ability of the underlying GGA to describe Pauli repulsion, as given by its exchange enhancement factor. Going beyond studies of biomolecules, this latter result has important implications for selecting appropriate GGA functionals for applications to ionic solids and layered materials containing anion-anion interactions.

Introduction

With steady advances in computer technology, density-functional theory (DFT) is being routinely applied to study large biochemical systems. Under the umbrella of biological applications,[1] it is now possible to apply DFT to the simulation of entire proteins, evaluate protein-ligand binding energies,[2–5] and investigate mechanistic details of enzymatic reactions.[6] A methodological advance in DFT that is crucial to these applications is the development of London dispersion models[7–14] that may be paired with popular DFT methods to provide accurate treatment of intermolecular interactions. However, as dispersion-corrected DFT methods are typically developed

*To whom correspondence should be addressed
†Universidad de Oviedo
‡Dalhousie University

and tested on small molecular dimers at or near their equilibrium geometries, their accuracy for non-covalent interactions within complex biomolecules is not guaranteed and requires further investigation.

One aspect that is often overlooked in the development and benchmarking of DFT methods is the treatment of non-covalent interactions between charged systems. However, charged fragments (e.g. charged ligands, protein zwitterions) are omnipresent in biological systems. Burns et al.[15] recently introduced the BioFragment Database (BFDb) for benchmarking computational methods regarding their ability to treat intermolecular interactions in biological systems. The largest subset of this database is the "SSI" collection of 3380 side-chain–side-chain interactions, comprising pairs of standard amino acids found within 47 proteins.[15] In the construction of this set, the protein structures were relaxed with the ff99SB force-field[16] using a generalized Born model[17] of implicit aqueous solvent. Pairs of side chains containing non-hydrogen atoms within a distance of 4Å were selected and terminated by breaking bonds to sp^3-hybridized carbon atoms and capping with a hydrogen at a fixed C-H distance of 1.1Å. The side chains were otherwise fixed at the same geometries as in the relaxed protein structures, so that the molecular dimers are generally far from their equilibrium gas-phase minima. CCSD(T) extrapolated to the complete basis-set limit was employed to obtain reference interaction energies.[15] Its design gives the SSI dataset more geometric diversity than the S22x5[18] or S66x8[19] sets, which are commonly used to develop and test dispersion-corrected DFT methods.

Importantly, the SSI set is also rare in that it involves interactions between charged monomers. Due to its construction from aqueous protein geometries in their zwitterionic state, each side chain can be neutral, cationic, or anionic. This results in six possible configurations for the side-chain dimers: neutral (0/0), zwitterionic (+1/-1), monoanions (-1), monocations (+1), dianions (-2), or dications (+2). The decomposition of the SSI set into these six categories is summarised in Table 1, which also shows the average (reference) interaction energy for each dimer class. The neutral, monoanionic, and monocationic interactions are generally weakly attractive. The ion-dipole interactions in the monoanion and monocation dimers are stronger on average (binding energies ca. 8 kcal/mol) than the interactions between two neutral fragments (ca. 1 kcal/mol), for which London dispersion is the dominant intermolecular interaction. Dispersion contributes a consistent 1–2 kcal/mol to the interaction energies for all dimer classes, with larger contributions occurring for bulkier side chains. However, dispersion only represents a significant percentage of the total energy for cases with one or more neutral side-chain fragments. Conversely, the interactions between two charged side chains are dominated by electrostatics: the zwitterions are strongly bound, while the dianion and dication interactions are strongly repulsive. The interested reader is directed to Ref. 15 for a more rigorous and detailed energy decomposition of the SSI set.

Virtually all non-covalent interaction benchmarks involve only neutral molecules and are dominated by dispersion and other van der Waals interactions, such as hydrogen bonding or halogen bonding.[18,19,22–24] While there are some exceptions for charge-transfer complexes[25,26] and halogens bonds in ionic systems,[27,28] these are well established as problem cases for conventional DFT methods.[27] In contrast, electrostatics is the dominant contribution to the interaction energies for the classes of the SSI dataset involving one or more ionic side chains. Due to the small dispersion contributions to the overall interaction energies (Table 1), it is not expected that results for these ionic dimers will be particularly sensitive to the choice of dispersion correction. However, it is not clear how various base density functionals will perform for these interaction energies. In their work, Burns et al.[15] used the SSI dataset to benchmark the accuracy of 12 density functionals, including several hybrid, double hybrid, and range-separated hybrid functionals, combined with various dispersion corrections. The conclusion of this analysis was that the performance of DFT-based methods is "unsystematic".

Table 1: The composition of the SSI set of Burns et al.,[15] showing the number of intermolecular complexes of each type. The average interaction strength from the literature reference data is also given, together with the average dispersion-binding contribution to the overall interaction energies (obtained with BHandHLYP-XDM[8,20,21] calculations), both in kcal/mol.

Complex Charge	Monomer Charges	Abbreviation	Number	Average Strength	Average Dispersion
Neutral	both neutral	0/0	2,596	-1.27	-1.67
Anion	one negatively charged	-1	391	-8.27	-1.79
Cation	one positively charged	+1	202	-7.77	-1.83
Neutral	oppositely charged	+1/-1	170	-103.4	-1.46
Dianion	negatively charged	-2	14	73.7	-2.14
Dication	positively charged	+2	7	61.1	-2.56

In this work, we aim to study and understand the performance of various DFT methods in the description of non-covalent interactions between neutral and charged protein fragments using the SSI data set. To this end, calculations were carried out using three families of hybrid functionals, based on selected generalised gradient approximations (GGAs),[21,29–31] all paired with the exchange-hole dipole moment (XDM) dispersion correction.[7,8] Our results show that the subsets of the SSI benchmark involving one or more neutral side chains, whose dimers are similar to those included in other benchmark sets for non-covalent interactions, are well described by all functionals considered. The best performance for these dimers is obtained using the exact-exchange mixing fractions normally employed for main-group thermochemistry.[32–35] Larger absolute errors are obtained for the zwitterion and diion subsets, although the performance is still excellent relative to the magnitude of the interaction energies. All methods exhibit consistent overestimation of electrostatic repulsion in the dications, which we conjecture is due to incomplete inclusion of correlation effects.[36] For the dianions, the errors can be explained by the large-gradient limit of the GGA enhancement factor of the functionals considered.[37–41] Finally, systematic overbinding is seen for the zwitterions, which may be rationalised in terms of delocalisation (or charge-transfer) error[42–46] with GGAs and overestimation of electrostatic interactions[41] with Hartree-Fock theory caused by localisation error.

Computational Methods

The geometries of all dimers and monomers in the SSI set were obtained from the work of Burns et al.[15] The electronic energies of all species were evaluated using the Gaussian 09 program[47] with the aug-cc-pVTZ basis set. Families of hybrid functionals, having an exchange-correlation energy of the form

$$E_{\mathrm{XC}} = a_{\mathrm{X}} E_{\mathrm{X}}^{\mathrm{HF}} + (1 - a_{\mathrm{X}}) E_{\mathrm{X}}^{\mathrm{GGA}} + E_{\mathrm{C}}^{\mathrm{GGA}} \quad (1)$$

were constructed based on three GGA functionals: BLYP,[21,29] PBE,[30] and PW86PBE.[30,31] The a_{X} parameter controls the fraction of exact-exchange mixing and was varied from 0 to 1 in increments of 0.1. For example, using the BLYP GGA with $a_X = 0.5$ gives the BHandHLYP[20,21] functional.

In all cases, the exchange-hole dipole moment (XDM) dispersion correction[7,8] was added to the base DFT energy computed with each hybrid functional, in a post-self-consistent-field (post-SCF) fashion. The two XDM damping-function parameters were fit for each hybrid by minimising the mean absolute percent errors (MAPEs) for the Kannemann-Becke set of 49 intermolecular complexes (KB49),[48] as in our previous work.[8,41] The parameter values and corresponding error statistics for this KB49 set, at each value of a_{X}, are given in the Supporting Information.

A small subset of the SSI, the members of which are also listed in the Supporting Information, was constructed by selecting 5 complexes

from each class in Table 1. For each of these 30 complexes, charge-transfer analysis was performed using Bader's quantum theory of atoms in molecules (QTAIM).[49] The QTAIM atomic charges were computed from the wavefunction of the intermolecular complex, using the AIMall program,[50] and summed over each molecule in the complex. The difference between these results and the formal, integer charges on each molecule determined the extent of intermolecular charge transfer.

Results and Discussion

The mean absolute errors (MAEs) and mean errors (MEs) for each subset of the SSI are shown as a function of exact-exchange mixing fraction in Figures 1 and 2, respectively. For the purposes of discussion, we will consider the neutral, monocation, and monoanion complexes together, as the error trends are similar for each of these classes. The zwitterions, dications, and dianions will then be considered individually.

Overall, the best performing methods are those with intermediate exact-exchange mixing fractions of 20-30%, as are typically employed for predition of main-group thermochemistry. In general, the binding-energy results for the side-chain dimers are excellent, with MAEs of only ca. 1% of the total interaction energies. However, the maximum errors can be substantially larger, typically ranging from 1-2 kcal/mol for the neutral dimes and 1-4 kcal/mol for the charged dimers, depending on functional (Supporting Information). The distributions of the errors for the particular case of the 20%-PW86PBE hybrid, one of the best-performing XDM-corrected functionals, are shown in Figure 3 for the neutral, zwitterion, monoanion, and monocation classes. The dianion and dication distributions are not shown due to the small sample sizes (14 and 7 complexes, respectively).

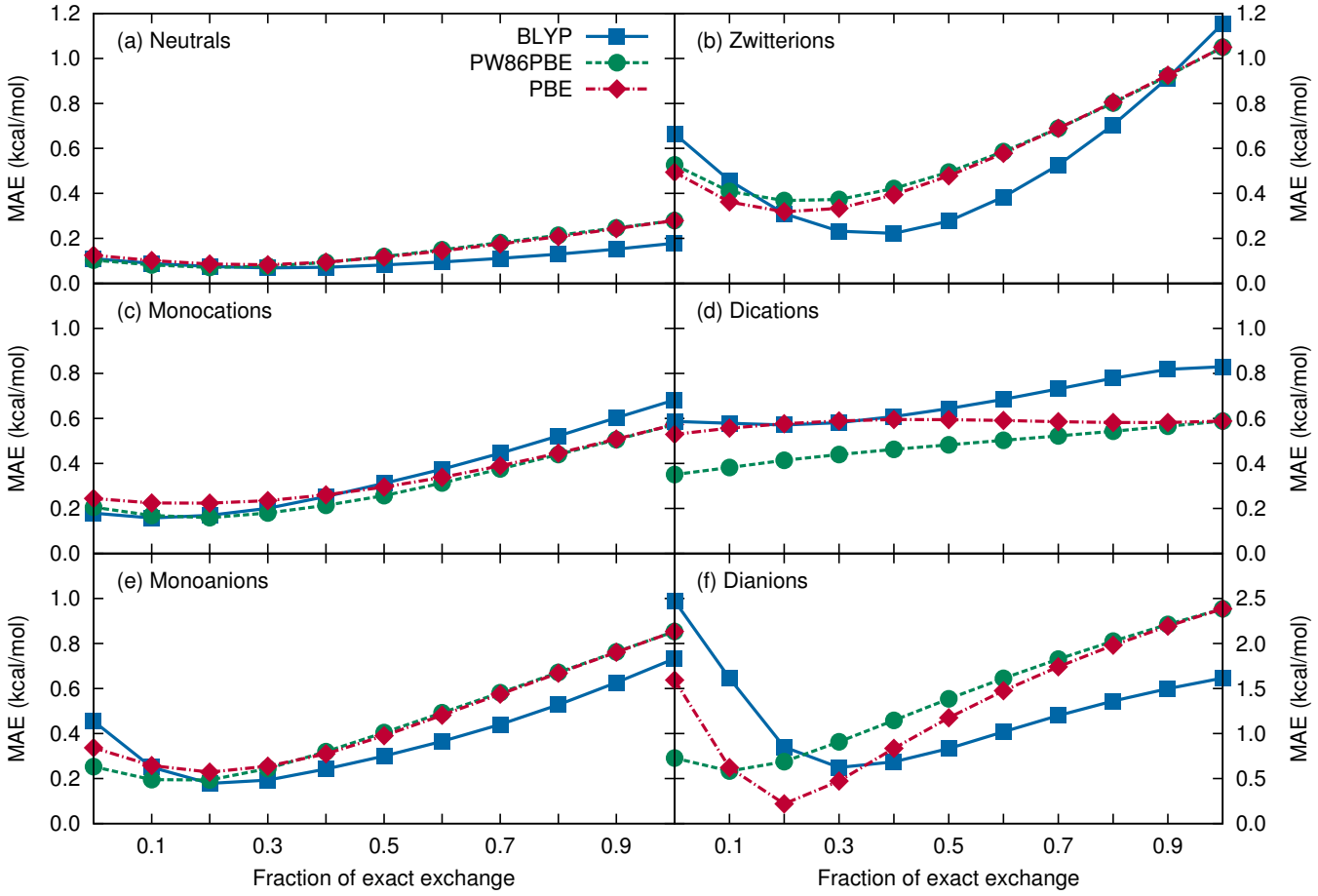
Neutral and Monoion Complexes

For the neutral complexes with neutral monomers (the 0/0 category in Table 1), the interactions are dominated by dispersion. The interaction energies are generally smaller in magnitude than for other intermolecular-interaction benchmarks, such as the KB49,[8,48] S22,[22] or S66,[23] where the interactions are a mix of dispersion, π-stacking, and hydrogen-bonding. This is primarily due to the fact that the dimers in the SSI set are frozen at their geometries in the protein, which prevents them from rearranging to maximize dispersion or form hydrogen bonds as in other benchmark sets. As a result of the smaller interaction energies, the MAEs are also typically lower than for other sets. In particular, with the full range of XDM-corrected functionals considered herein, MAEs for the KB49 set range from 0.21–0.92 kcal/mol (see Supporting Information), compared to MAEs of 0.07–0.28 kcal/mol for the 0/0 SSI set. Hence, all XDM-corrected functionals were found to perform very well at all fractions of exact-exchange mixing. From Figure 1, the lowest errors were consistently obtained with 10–40% exact exchange, in agreement with what has been seen previously for the KB49 set, as well as for other standard thermochemistry benchmarks.[8] From Figure 2, the MEs are also uniformly near zero, indicating no consistent overbinding or underbinding tendency.

The MAEs for the monocation (+1) and monoanion (-1) complexes (Figure 1) are quite similar in magnitude, ranging from 0.16 to 0.85 kcal/mol. While these errors are significantly higher than for the neutral 0/0 complexes, they are comparable to those obtained for the KB49 set. This is reasonable because both the total binding energies and the strength of the electrostatic interactions in the KB49 dimers, and other similar sets, are comparable to those in the monoanion and monocation subsets of the SSI. As for the neutral 0/0 complexes, the MAEs are minimized for 10–40% exact exchange, indicating that conventional hybrid functionals perform quite well for all three of these SSI classes.

The MEs (Figure 2) for the monocations are again uniformly near zero for all mixing fractions, indicating no systematic errors. For the monoanions, however, there is some dependence

Figure 1: Mean absolute errors for the various subsets of the SSI, as a function of exact-exchange mixing. Results are shown for (a) neutral complexes, (b) zwitterions, (c) monocationic complexes, (d) dications, (e) monoanionic complexes, and (f) dianions. Note the increased y-axis range for the dianions in plot (f).



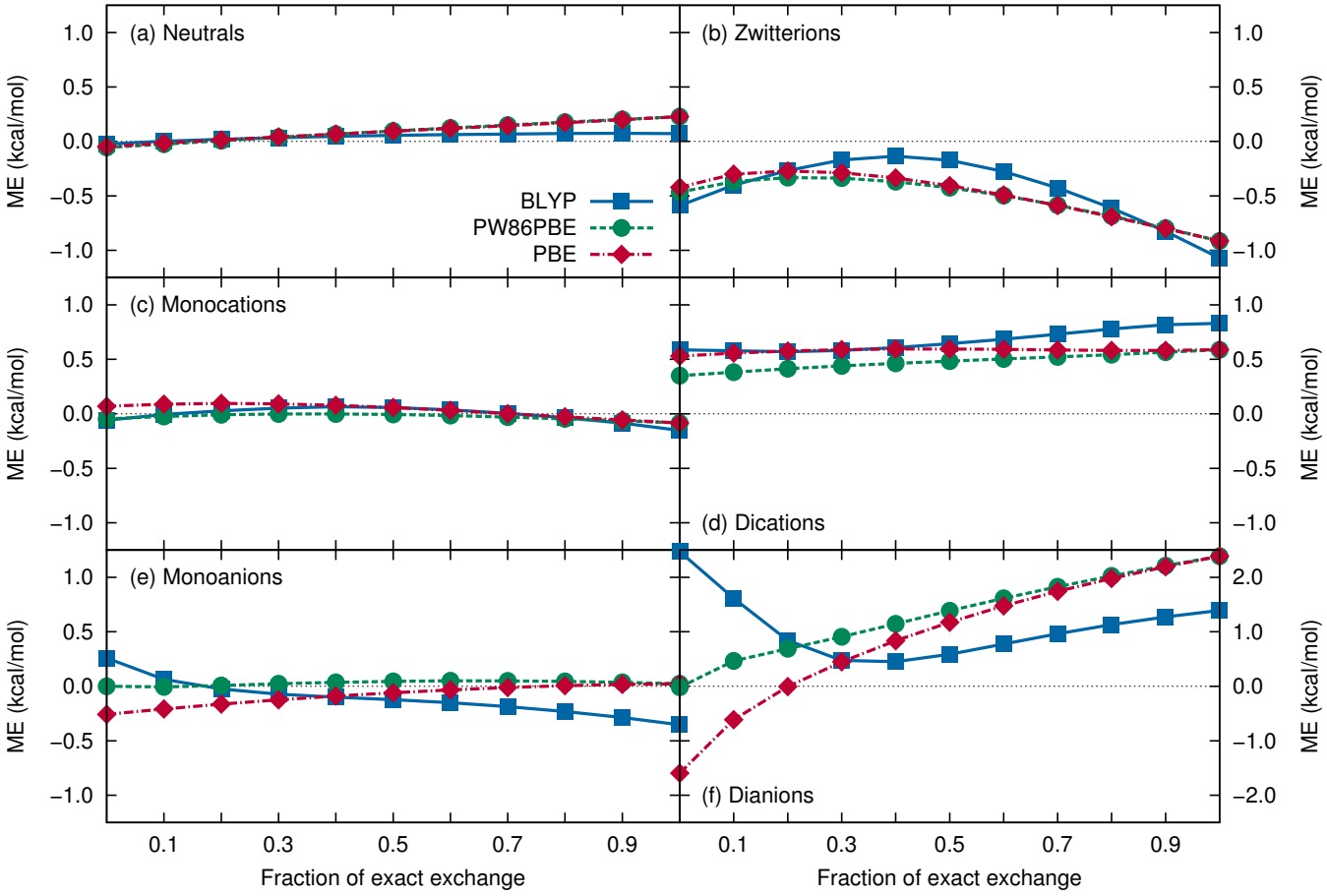
of the MEs on the choice of base functional, particularly for the pure GGAs. This effect will be observed again, but to a much larger extent, when we consider the errors for the dianions. In any case, conventional hybrid functionals, with exact-exchange mixing fractions appropriate for main-group thermochemistry and XDM dispersion corrections, give excellent performance for the neutral, monocation, and monoanion SSI binding energies.

Zwitterion Complexes

The ionic nature of the intermolecular interactions present in the zwitterion (+1/-1) complexes imparts large binding energies (Table 1), comparable to covalent bond strengths. Dispersion represents $< 2\%$ of the average interaction energies, meaning that the performance for this SSI subset will be dominated by the choice of base density functional. As shown in Figure 1, the errors for the zwitterion +1/-1 complexes are generally larger than for the neutral 0/0 complexes, ranging from 0.22 to 1.15 kcal/mol, although they are still comparable to those for the KB49 set. The lowest MAEs are again obtained with 20-40% exact exchange. However, considering the mean errors in Figure 2, the +1/-1 complexes are overbound with all functionals and a_X values considered.

Hartree-Fock theory is known over-localize electrons[51] and overestimate molecular dipoles.[52] In the +1/-1 systems, this results in overbinding of the oppositely-charged moieties due to an overestimation of the electrostatic contribution. This explains the overbinding at high exact-exchange fractions shown in Figure 2, a behaviour that is similar to what was

Figure 2: Mean errors for the various subsets of the SSI, as a function of exact-exchange mixing. Results are shown for (a) neutral complexes, (b) zwitterions, (c) monocationic complexes, (d) dications, (e) monoanionic complexes, and (f) dianions. Note the increased y-axis range for the dianions in plot (f).



seen in our previous work for hydrogen fluoride clusters.[41] Conversely, the overbinding of the +1/-1 complexes by all pure GGA functionals can be explained by density-functional delocalisation error. This error refers to the tendency of GGA functionals to overstabilise delocalised electron distributions, including those with fractional charges.[43–46,53] In the context of intermolecular interactions, delocalisation error manifests as overbinding in complexes with significant charge transfer and intermolecular delocalisation[25–27,54,55] and is consequently also known as charge-transfer error.

To illustrate this point, Figure 4 shows the intermolecular charge transfer for a 30-complex subset of the SSI database, as a function of exact-exchange mixing fraction. The values for each SSI class are averages over five representative examples of interactions within that class. The degree of charge transfer for each complex was determined as the deviation from the expected integer charges on each molecular fragment via QTAIM analysis. The +1/-1 complexes stand out in the figure as having both the largest charge transfer and the largest span of values across functionals, ranging from nearly $0.12e^-$ for BLYP to $0.06e^-$ for BHandHLYP and only $0.04e^-$ for HFLYP. This is consistent with our previous discussion: charge transfer is overestimated by GGA functionals and it decreases with increasing exact exchange fraction, indicating the appearance of (density-driven) delocalisation error from the base functional.[46,53] The excessive charge transfer predicted by the GGA functionals, shown in Figure 4, causes the observed overbinding in Figure 2.

Overall, the best performance for the +1/-

Figure 3: Error distributions for selected SSI subsets obtained from the PW86PBE-based hybrid functional with 20% exact-exchange mixing and XDM dispersion.

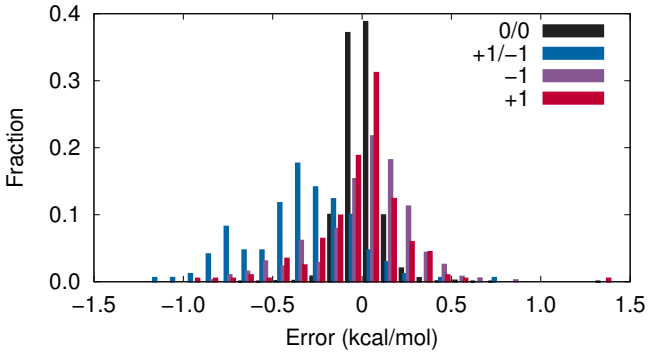
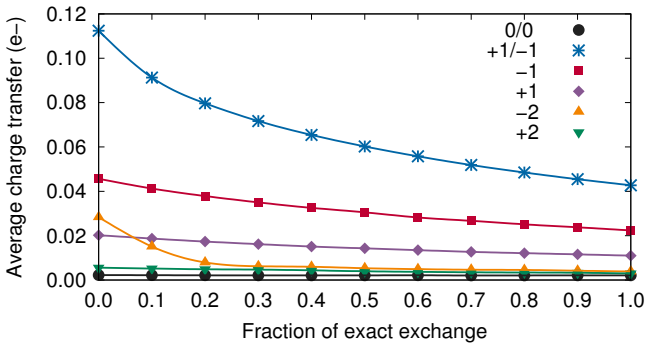


Figure 4: The average intermolecular charge transfer, as a function of exact-exchange mixing in the BLYP-based hybrid series, for each interaction class of the representative 30-complex SSI subset.



1 complexes is obtained with intermediate exact-exchange mixing fractions, which minimize delocalisation error, while also reducing the overstabilisation of electrostatic interactions from Hartree-Fock theory. The general overbinding of the zwitterions is similar to, but less dramatic than, our results for Bauzá's halogen-bonding dataset,[27,28] which includes ionic cases where halide anions act as the electron donors. For that benchmark, the complexes display even higher charge transfer, with average values of $0.22e^-$ for BLYP and $0.17e^-$ for BHandHLYP.[27] Consequently, most XDM-corrected GGA and hybrid functionals were found to shown even greater systematic overbinding.

Dication complexes

For the dication class of the SSI, the interactions are dominated by electrostatic repulsion. While dispersion contributes only ca. 4% on average to the overall interaction energy (Table 1), the absolute values of the dispersion energies are largest for this class. This is due to the large side-chain size (methylguanidinium cation) for arginine, relative to most other standard amino acids. Of the 7 dication complexes, 5 of the original amino-acid pairs are arginine dimers and two are arginine-lysine dimers.

From Figure 1, the MAEs for the dication set range from 0.35 to 0.83 kcal/mol, which represents only ca. 1% of the overall interaction energy. However, unlike for all other SSI classes, there is no minimum in the MAE for hybrid functionals with 10–40% exact exchange. Additionally, Figure 2 shows that all functionals systematically overestimate the electrostatic repulsion in these complexes. The systematic errors for the dications show minimal dependence on the extent of exact-exchange mixing. This result, taken together with the negligible charge transfer found in the dication complexes (Figure 4), indicates that the systematic deviations for the dications cannot be attributed to delocalization (or self-interaction) error. Indeed, delocalization error typically results in overbinding, rather than overestimation of nonbonded repulsion.

It is unlikely that the source of systematic error for the dications is the XDM dispersion term because an underestimation of dispersion interactions for cations would similarly affect the monocation complexes, yet this is not observed in Figure 2. It is similarly unlikely that basis-set incompleteness is responsible, as electronic energies of cations are generally less sensitive to the choice of basis set than neutral or anionic species. For the particular test case of the 20% BLYP hybrid functional, increasing the basis-set size to aug-cc-pVQZ causes a 0.07 kcal/mol increase in the repulsion energies, resulting in a slight increase in MAE from 0.57 to 0.64 kcal/mol.

A possible explanation for the results in Figure 2(d) could be a problem with the de-

scription of electron correlation for the dication interactions, which would affect semilocal GGAs and Hartree-Fock similarly. Given that cations have more compact electron densities than neutral molecules, dication complexes will possess virtually zero density overlap between monomers. It is reasonable that semilocal GGA correlation models may not be able to provide a complete treatment of dynamical electron correlation between the cations; even a small amount of additional correlation would serve to offset the excess electrostatic repulsion, improving agreement with the correlated-wavefunction referece data. It has been previously noted for dications of thiophene oligomers[36] that sophisticated treatments of electron correlation are needed to compute accurate interaction energies for dication complexes, without overestimation of electrostatic repulsion. We therefore conjecture that incomplete treatment of electron correlation effects may be the reason for the observed overestimation of electrostatic repulsion for the SSI dication complexes.

Dianion Complexes

As for the dications, the interactions in the dianion complexes are all dominated by electrostatic repulsion. The interaction energies are overly repulsive for functionals with high fraction of exact-exchange mixing, probably due to the tendency of Hartree-Fock to overestimate the electrostatic component. However, the interaction-energy errors for the dianions, which range from 0.22 to 2.47 kcal/mol (Figure 1), are generally larger in magnitude than for all other classes of complexes within the SSI set.

Figure 2 shows that the performance for the dianions exhibits the greatest dependence on the choice of base GGA functional: BLYP is strongly underbinding, while PBE is strongly overbinding, and PW86PBE has near zero mean error. This is the same qualitative trend observed for the GGAs in the treatment of the monoanions but, in this case, the MEs are 6–10 times higher in magnitude. These observations can be explained by considering the form of the enhancement factor for each of the GGA exchange functionals.

The GGA enhancement factor, $F_{\mathrm{X}}(s)$, determines the exchange-energy contribution beyond the local density approximation (LDA) by incorporating a dependence on the reduced density gradient, s:

$$E_{\mathrm{X}} = c_{\mathrm{X}} \int F_{\mathrm{X}}(s) \varepsilon_{\mathrm{X}}^{\mathrm{LDA}}(\rho) d\mathbf{r}. \qquad (2)$$

Here, c_{X} is a constant, ρ is the electron density, and $\varepsilon_{\mathrm{X}}^{\mathrm{LDA}}$ is the LDA exchange-energy density. The particular form of the enhancement factor depends on the GGA exchange functional. Because the electron density is piecewise exponential, the reduced density gradient will always have large values far from an atom or molecule. Conversely, when a van der Waals complex is formed, the reduced gradient will have small values in the intermolecular region, becoming identically zero at the bond critical point(s).[56] As all GGA functionals are constructed to reduce to the LDA for zero gradient, the difference in exchange energy upon formation of the van der Waals complex is determined by the large-gradient limit of the enhancement factor. Consequently, the limit of large reduced density gradients determines the performance of the exchange functional for Pauli repulsion in van der Waals complexes,[37–41] which is crucial in the description of intermolecular interactions.

It has been shown that the optimal large-gradient limit for accurate description of exchange-only repulsion in noble-gas dimers[37–39] is

$$\lim_{s\to\infty} F_{\mathrm{X}}(s) \sim s^{2/5} \qquad (3)$$

and the PW86 exchange functional was designed to obey this limit. Instead, the PBE enhancement factor approaches a constant in the large-gradient limit, resulting in significant overbinding of van der Waals dimers, even without a dispersion correction. Conversely, the B88 enhancement factor scales more quickly than $s^{2/5}$, resulting in overestimated non-bonded repulsion.

Returning to the monoanions and dianions in

the SSI set, Figure 2 shows overbinding with PBE and underbinding with BLYP, which can be explained by the aforementioned differences in their enhancement factors. As PW86 has the desired large-gradient limit, PW86PBE gives near-zero mean error for both anionic and dianionic complexes. One may wonder why the enhancement factor only has a significant effect on the mean errors for the anionic complexes, and not for the neutral or cationic complexes. This can be understood through consideration of the differences in electron densities using a series of isoelectronic complexes of varying charge state.

Figure 5: The Hartree-Fock electron density (solid lines) and reduced density gradient (dashed lines), computed using the NUMOL program,[57,58] for the isoelectronic series of dimers: Na_2^{2+}, Ne_2, and F_2^{2-}.

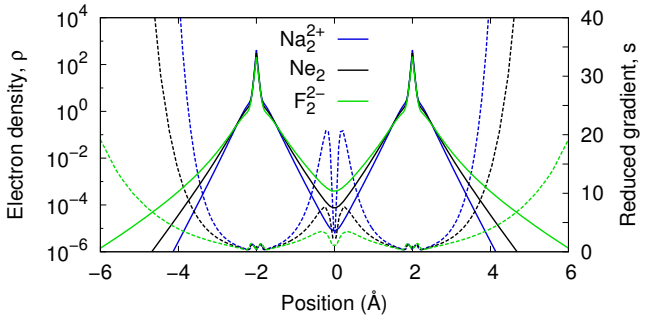


Figure 5 shows a plot of the electron densities and reduced density gradients along the "bond" axis for the F_2^{2-}, Ne_2, and Na_2^{2+} complexes, each with a 4 Å internuclear separation. The F_2^{2-} dianion has the most diffuse density distribution, as well as the lowest reduced gradient in the interatomic region. Higher density values far from the nuclei will result in a higher weighting of the gradient term for those points in the integration of the exchange energy. In turn, this will result in a larger exchange-energy difference between the dimer and separated monomers, which will cause larger variation in the predicted extent of non-bonded repulsion. Thus, the gradient term will have the largest contribution to the interaction energy for anions, followed by neutrals, and then cations, as the electron density becomes more compact. This explains why the observed error dependence with the GGA functional (Figure 2) is highest for the dianions, smaller for the monoanions, and effectively zero for the other interaction classes.

Conclusion

In this work, we have rationalized the performance of a series of dispersion-corrected hybrid functionals in the description of protein side-chain interaction energies from the SSI benchmark set. We show that, in general, excellent results are obtained with conventional hybrid density functionals paired with the XDM dispersion correction. The mean absolute errors are minimized for the entire dataset, as well as for most of its subclasses, with mixing fractions of 20-30% exact exchange. Thus, while not explicitly considered here, B3LYP-XDM[21,32] (20%) and PBE0-XDM[59] (25%) would be excellent choices for the SSI dataset. Notably, with a MAE of 0.11 kcal/mol for the entire SSI set, the 20% BLYP-XDM hybrid (essentially equivalent to B3LYP-XDM), outperforms B3LYP-D3M(BJ),[9,60] which gives a MAE of 0.17 kcal/mol with the same aug-cc-pVTZ basis set.[15] An identical MAE of 0.11 kcal/mol for the SSI set is also obtained with the PW86PBE-based 20% hybrid functional. These findings highlight the excellent performance and transferability of the XDM dispersion model to biological systems.

Our analysis has focused on trying to understand the errors for the different classes of dimers in the SSI benchmark set. The neutral, monoanion, and monocation complexes are well described in general, with all XDM-corrected functionals considered giving mean errors near zero. Although larger errors are seen for the zwitterions, dications, and dianions, the interaction strengths are comparatively larger as well, by an order of magnitude or more. Thus, the performance of the XDM-corrected functionals remains excellent for these SSI subsets, with average errors on the order of only 1% of the total interaction energies.

We show as a result of our analysis that: (i) Functionals with high exact-exchange fractions overestimate the extent of electrostatic

interactions, which is a known shortcoming of Hartree-Fock theory. (ii) Functionals with low exact-exchange fractions overstabilise the zwitterions as a consequence of density-driven delocalisation error. GGA exchange functionals predict excessive charge transfer between the interacting molecules, resulting in overstabilisation of the complex. (iii) All functionals and exact-exchange mixing fractions considered slightly overestimate the extent of electrostatic repulsion in the dications. Similar overestimation of electrostatic repulsion has been observed previously for dications,[36] where sophisticated treatments of electron correlation were needed to obtain accurate interaction energies. (iv) The performance for the dianions, and to a lesser extent for the monoanions, is sensitive to the choice of GGA exchange enhancement factor. PW86PBE gives the lowest error, as its enhancement factor obeys the large-gradient limit that ensures a reliable description of nonbonded repulsion in van der Waals complexes.

These observations regarding the relation between performance for intermolecular interactions in neutral and anionic systems and the large-gradient limit of the exchange enhancement factors have implications beyond biological molecules. Anion-anion interactions are important in many condensed-matter systems, including simple ionic solids where there are close contacts between anions[61] with significant anion-anion delocalization.[62] These interactions can also be used to explain, at least in part, the performance of various functionals in the description of layered materials[63–65] that have recently become relevant due to advances in the manufacturing of atomically thin films and the potential use of layered transition metal dichalcogenides in optoelectronic applications.[66]

The results in this work provide insight into the inherent errors associated with common density-functional approximations in the description of intermolecular interactions between charged fragments. This is a first step towards a more complete understanding of the performance of DFT methods for inter-residue interactions in proteins, and between charged moieties in general. However, while these interactions are definitely present in biological systems, most notably proteins, it must also be noted that they are almost always stabilized by the environment, be it solvating water molecules or other protein residues. We stress that the present results apply only to gas-phase interactions of amino-acid dimers and errors obtained with DFT methods could be larger for more complex systems, such as ligand-protein interactions. In particular, the use of continuum solvent models and neglect of conformational sampling can constitute significant limitations to application of DFT to host-guest interactions and ligand-binding affinities in biochemical systems.[67–69] Further studies involving more realistic models of biological systems would be beneficial.

Acknowledgement ERJ thanks the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support and Compute Canada (ACEnet and Westgrid) for computational resources. AOR thanks the Spanish MINECO for a Ramón y Cajal fellowship (RyC-2016-20301) and the MICINN for financial support (projects PGC2018-097520-A-100 and RED2018-102612-T), and the MALTA Consolider supercomputing centre and Compute Canada for computational resources.

Supporting Information Available: Tables of XDM damping parameters and MAEs for the KB49 fit set; MAEs, MEs, and maximum errors for the various SSI classes obtained with all XDM-corrected hybrid functionals; composition of the representative 30-molecule SSI subset. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Cole, D. J.; Hine, N. D. M. Applications of large-scale density functional theory in biology. *J. Phys. Condens. Matter* **2016**, *28*, 393001.

(2) Lukac, I.; Abdelhakim, H.; Ward, R. A.; St-Gallay, S. A.; Madden, J. C.; Leach, A. G. Predicting protein–ligand

binding affinity and correcting crystal structures with quantum mechanical calculations: lactate dehydrogenase A. *Chem. Sci.* **2019**, *10*, 2218–2227.

(3) Phipps, M. J. S.; Fox, T.; Tautermann, C. S.; Skylaris, C. K. Intuitive density functional theory-based energy decomposition analysis for protein-ligand interactions. *J. Chem. Theory Comput.* **2017**, *13*, 1837–1850.

(4) Fox, S. J.; Dziedzic, J.; Fox, T.; Tautermann, C. S.; Skylaris, C. K. Density functional theory calculations on entire proteins for free energies of binding: application to a model polar binding site. *Proteins* **2014**, *82*, 3335–3346.

(5) Antony, J.; Grimme, S. Fully ab initio protein-ligand interaction energies with dispersion corrected density functional theory. *J. Comp. Chem.* **2012**, *33*, 1730–1739.

(6) Lever, G.; Cole, D.; Lonsdale, R.; Ranaghan, K. E.; Wales, D. J.; Mulholland, A. J.; Skylaris, C.-K.; Payne, M. C. Large-scale density functional theory transition state searching in enzymes. *J. Phys. Chem. Lett.* **2014**, *5*, 3614–3619.

(7) Johnson, E. R. In *Non-covalent Interactions in Quantum Chemistry and Physics*; Otero-de-la Roza, A., DiLabio, G. A., Eds.; Elsevier, 2017; Chapter 5, pp 169–194.

(8) Otero-de-la Roza, A.; Johnson, E. R. Noncovalent interactions and thermochemistry using XDM-corrected hybrid and range-separated hybrid density functionals. *J. Chem. Phys.* **2013**, *138*, 204109.

(9) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(10) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(11) Tkatchenko, A.; Scheffler, M. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **2009**, *102*, 073005.

(12) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and efficient method for many-body van der Waals interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.

(13) Lee, K.; Murray, É. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-accuracy van der Waals density functional. *Phys. Rev. B* **2010**, *82*, 081101.

(14) Schröder, E.; Cooper, V. R.; Berland, K.; Lundqvist, B. I.; Hyldgaard, P.; Thonhauser, T. In *Non-covalent Interactions in Quantum Chemistry and Physics*; Otero-de-la-Roza, A., DiLabio, G. A., Eds.; Elsevier, 2017; Chapter 8, pp 241–274.

(15) Burns, L. A.; Faver, J. C.; Zheng, Z.; Marshall, M. S.; Smith, D. G. A.; Vanommeslaeghe, K.; MacKerell, A. D.; Merz, K. M.; Sherrill, C. D. The BioFragment Database (BFDb): An open-data platform for computational chemistry analysis of noncovalent interactions. *J. Chem. Phys.* **2017**, *147*, 161727.

(16) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.

(17) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(18) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.

(19) Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 data set: More accurate interaction energies and angular-displaced nonequilibrium geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466–3470.

(20) Becke, A. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372.

(21) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785.

(22) Jurečka, P.; Šponer, J.; Černỳ, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD (T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(23) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

(24) Kozuch, S.; Martin, J. M. L. Halogen bonds: Benchmark and theoretical analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931.

(25) Ruiz, E.; Salahub, D. R.; Vela, A. Charge-transfer complexes: Stringent tests for widely used density functionals. *J. Phys. Chem.* **1996**, *100*, 12265–12276.

(26) Steinmann, S. N.; Piemontesi, C.; Delacht, A.; Corminboeuf, C. Why are the interaction energies of charge-transfer complexes challenging for DFT? *J. Chem. Theory Comput.* **2012**, *8*, 1629–1640.

(27) Otero-de-la Roza, A.; Johnson, E. R.; DiLabio, G. A. Halogen bonding from dispersion-corrected density-functional theory: the role of delocalization error. *J. Chem. Theory Comput.* **2014**, *10*, 5436–5447.

(28) Bauzá, A.; Alkorta, I.; Frontera, A.; Elguero, J. On the reliability of pure and hybrid DFT methods for the evaluation of halogen, chalcogen, and pnicogen bonds involving anionic and neutral electron donors. *J. Chem. Theory Comput.* **2013**, *9*, 5201–5210.

(29) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(30) Perdew, J.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(31) Perdew, J.; Yue, W. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B* **1986**, *33*, 8800.

(32) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(33) Becke, A. D. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **1997**, *107*, 8554–8560.

(34) Hamprecht, F.; Cohen, A.; Tozer, D.; Handy, N. Development and assessment of new exchange-correlation functionals. *J. Chem. Phys.* **1998**, *109*, 6264.

(35) Boese, A. D.; Handy, N. C. A new parametrization of exchange-correlation generalized gradient approximation functionals. *J. Chem. Phys.* **2001**, *114*, 5497–5503.

(36) Scherlis, D. A.; Marzari, N. π-Stacking in charged thiophene oligomers. *J. Phys. Chem. B* **2004**, *108*, 17791–17795.

(37) Lacks, D. J.; Gordon, R. G. Pair interactions of rare-gas atoms as a test of exchange-energy-density functionals in regions of large density gradients. *Phys. Rev. A* **1993**, *47*, 4681.

(38) Zhang, Y.; Pan, W.; Yang, W. Describing van der Waals Interaction in diatomic molecules with generalized gradient approximations: The role of the exchange functional. *J. Chem. Phys.* **1997**, *107*, 7921–7925.

(39) Kannemann, F. O.; Becke, A. D. Van der Waals interactions in density-functional theory: Rare-gas diatomics. *J. Chem. Theory Comput.* **2009**, *5*, 719–727.

(40) Gillan, M. J. Many-body exchange-overlap interactions in rare gases and water. *J. Chem. Phys.* **2014**, *141*, 224106.

(41) Otero-de-la-Roza, A.; DiLabio, G. A.; Johnson, E. R. Exchange-correlation effects for noncovalent interactions in density functional theory. *J. Chem. Theory Comput.* **2016**, *12*, 3160–3175.

(42) Tozer, D. J. Relationship between long-range charge-transfer excitation energy error and integer discontinuity in Kohn–Sham theory. *J. Chem. Phys.* **2003**, *119*, 12697–12699.

(43) Zhang, Y.; W, Y. A challenge for density functionals: Self-interaction error increases for systems with a noninteger number of electrons. *J. Chem. Phys.* **1998**, *109*, 2604–2608.

(44) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E. Spurious fractional charge on dissociated atoms: Pervasive and resilient self-interaction error of common density functionals. *J. Chem. Phys.* **2006**, *125*, 194112.

(45) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into current limitations of density functional theory. *Science* **2008**, *321*, 792.

(46) Kim, M.-C.; Sim, E.; Burke, K. Understanding and reducing errors in density functional calculations. *Phys. Rev. Lett.* **2013**, *111*, 073003.

(47) Frisch, M. J. et al. Gaussian 09 Revision B.1. Gaussian Inc. Wallingford CT 2009.

(48) Kannemann, F. O.; Becke, A. D. van der Waals interactions in density-functional theory: Intermolecular complexes. *J. Chem. Theory Comput.* **2010**, *6*, 1081–1088.

(49) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Oxford University Press: Oxford, 1990.

(50) Keith, T. A. AIMAll (Version 14.06.21). TK Gristmill Software, Overland Park KS, USA, 2014 (aim.tkgristmill.com).

(51) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Localization and delocalization errors in density functional theory and implications for band-gap prediction. *Phys. Rev. Lett.* **2008**, *100*, 146401.

(52) Hickey, A. L.; Rowley, C. N. Benchmarking quantum chemical methods for the calculation of molecular dipole moments and polarizabilities. *J. Phys. Chem. A* **2014**, *118*, 3678–3687.

(53) Johnson, E. R.; Otero de la Roza, A.; Dale, S. G. Extreme density-driven delocalization error for a model solvated-electron system. *J. Chem. Phys.* **2013**, *139*, 184116.

(54) Bizzarro, B. B.; Egan, C. K.; Paesani, F. Nature of halide-water interactions: Insights from many-body representations and density functional theory. *J. Chem. Theory Comput.* **2019**, *15*, 2983–2995.

(55) Becke, A. D.; Dale, S. G.; Johnson, E. R. Correct charge transfer in CT complexes

from the Becke'05 density functional. *J. Chem. Phys.* **2018**, *148*, 211101.

(56) Johnson, E.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A.; Yang, W. Revealing noncovalent interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498–6506.

(57) Becke, A. D.; Dickson, R. M. Numerical solution of Schrödinger's equation in polyatomic molecules. *J. Chem. Phys.* **1990**, *92*, 3610.

(58) Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction. *J. Chem. Phys.* **2005**, *122*, 154104.

(59) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(60) Smith, D. G. A.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised damping parameters for the D3 dispersion correction to density functional theory. *J. Phys. Chem. Lett.* **2016**, *7*, 2197–2203.

(61) Nelyubina, Y. V.; Antipin, M. Y.; Lyssenko, K. A. Anion–anion interactions: their nature, energy and role in crystal formation. *Russ. Chem. Rev.* **2010**, *79*, 167.

(62) Otero-de-la Roza, A.; Martín Pendá, A.; Johnson, E. R. Quantitative electron delocalization in solids from maximally localized Wannier functions. *J. Chem. Theory Comput.* **2018**, *14*, 4699–4710.

(63) Björkman, T.; Gulans, A.; Krasheninnikov, A. V.; Nieminen, R. M. van der Waals bonding in layered compounds from advanced density-functional first-principles calculations. *Phys. Rev. Lett.* **2012**, *108*, 235502.

(64) Björkman, T. Testing several recent van der Waals density functionals for layered structures. *J. Chem. Phys.* **2014**, *141*, 074708.

(65) Tawfik, S. A.; Gould, T.; Stampfl, C.; Ford, M. J. Evaluation of van der Waals density functionals for layered materials. *Phys. Rev. Mater.* **2018**, *2*, 034005.

(66) Jariwala, D.; Sangwan, V. K.; Lauhon, L. J.; Marks, T. J.; Hersam, M. C. Emerging device applications for semiconducting two-dimensional transition metal dichalcogenides. *ACS nano* **2014**, *8*, 1102–1120.

(67) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. Overview of the SAMPL5 host–guest challenge: Are we doing better? *J. Comput. Aided. Mol. Des.* **2017**, *31*, 1–19.

(68) Jensen, J. H. Predicting accurate absolute binding energies in aqueous solution: thermodynamic considerations for electronic structure methods. *Phys. Chem. Chem. Phys.* **2015**, *17*, 12441–12451.

(69) Ryde, U.; Söderhjelm, P. Ligand-binding affinity estimates supported by quantum-mechanical methods. *Chem. Rev.* **2016**, *116*, 5520–5566.

Graphical TOC Entry