Atom-Centered Potentials with Dispersion-Corrected Minimal Basis Set Hartree-Fock: An Efficient and Accurate Computational Approach for Large Molecular Systems

Viki Kumar Prasad, [a,1] Alberto Otero-de-la-Roza,[a,2] and Gino A. DiLabio [a,b,3]

[a)] *Department of Chemistry, University of British Columbia, 3247 University Way, Kelowna, British Columbia, Canada V1V 1V7*
[b)] *Faculty of Management, University of British Columbia, 1137 Alumni Avenue, Kelowna, British Columbia, Canada V1V 1V7*

(Dated: December 11, 2017)

We present a computational methodology based on atom-centered potentials (ACPs) for the efficient and accurate structural modeling of large molecular systems. ACPs are atom-centered one-electron potentials that have the same functional form as effective-core potentials. In recent works, we showed that ACPs can be used to produce a correction to the ground-state wavefunction and electronic energy to alleviate shortcomings in the underlying model chemistry. In this work, we present ACPs for H, C, N, and O that are specifically designed to produce accurate non-covalent binding energies and inter- and intra-molecular geometries when combined with dispersion-corrected Hartree-Fock (HF-D3) and a minimal basis-set (scaled MINI or MINIs). For example, the combined HF-D3/MINIs-ACP method demonstrates excellent performance, with a mean absolute error of 0.36 and 0.28 kcal/mol for the S22x5 and S66x8 benchmark sets, respectively, relative to highly-correlated complete-basis-set data. The application of ACPs results in a significant decrease in error compared to uncorrected HF-D3/MINIs for all benchmark sets examined. In addition, HF-D3/MINIs-ACP, has a cost only slightly higher than a minimal-basis-set HF calculation and can be used with any electronic structure program for molecular quantum chemistry that uses Gaussian basis-sets and effective-core potentials.

[1] orcid.org/0000-0003-0982-3129
[2] orcid.org/0000-0002-4866-5816
[3] orcid.org/0000-0002-3778-3892; Email: gino.dilabio@ubc.ca

1. Introduction

Recent interest in the modeling of large supramolecular systems[1-3] and molecular crystals[4-6] with density-functional theory (DFT) has caused a resurgence of low-cost computational approaches for intermolecular interactions[7–9]. These "cheap methods" offer the option of taking a calculated trade-off between accuracy and cost in the spirit of force-field approaches but preserving the generality of an electronic structure calculation, particularly regarding their ability to model chemical reactions. In fact, a persistent challenge in this area is to find a way to accurately model covalent and non-covalent interactions simultaneously, since both play important roles in determining the structure of supramolecular systems. While good relative accuracy for covalent bond breaking and formation can be obtained with DFT and a modest basis-set (B3LYP/6-31G*, for instance, is widely used to successfully elucidate many organic reaction mechanisms), modeling non-covalent interactions requires either high-level wavefunction theory (CCSD(T) with complete-basis-set extrapolation) or dispersion-corrected DFT methods with a very large basis-set.[10-12] Although recent developments have reduced the asymptotic scaling of these techniques[13-16], computationally inexpensive methods are still in high demand[17-19] particularly if they can be combined with the new reduced-scaling techniques, as is the case with methods based on atom-centered potentials (ACPs)[20-25].

DFT-based methods suffer from additional problems that make them inefficient in practice for large systems. In addition to inaccuracies caused by the numerical integration of the exchange-correlation energy, generalized-gradient-approximation (GGA) functionals predict erroneous charge-transfer between molecules or within a single molecule due to delocalization error, which may lead to significantly overestimated binding energies and reaction barriers, and spuriously small band gaps.[26–30] In addition to these inaccuracies, a practical difficulty is that many generalized-gradient-approximation (GGA) functionals (e.g. BLYP[31–33], PBE[34]) and exchange functionals with a small fraction of exact exchange (e.g. B3LYP[35,36]) have considerable difficulties in arriving at converged solutions of the self-consistent field (SCF) equations for charged systems such as zwitterions, which are essential in the description of proteins. This practical hurdle greatly hinders the application of DFT methods to large supramolecular systems.

Semi-empirical approaches (e.g. AM1[37], PMx[38-40]) based on a minimal-basis-set HF approximation have been used in the simulation of supramolecular systems extensively thanks to their reduced cost. Until recently, these methods were unable to model intermolecular interactions adequately, particularly when dispersion effects are dominant.[9,41-44] The HF-3c method, recently proposed by Sure and Grimme is similar in spirit to traditional semi-empirical methods.[45] Instead of discarding or parametrizing certain two-electron integrals, HF-3c uses minimal-basis-set Hartree-Fock (HF) combined with three ad hoc formulas to account for dispersion (D3-BJ) [46-48], basis-set superposition error (gCP) [49], and short-range covalent over-binding (SRB)[45]: The last two errors are caused by basis-set incompleteness. Although more computationally expensive than a semi-empirical method,

HF-3c is substantially cheaper than any electronic structure method that uses more complete-basis-sets, and circumvents the self-consistent field (SCF) convergence problems of GGA and hybrid density functionals in large systems. While its performance for small molecular systems is superior to semi-empirical methods,[45] HF-3c has demonstrated suboptimal performance in comparison to higher-level methods in the ranking of lattice energies in molecular crystals for molecular crystal structure prediction in the sixth CCDC blind test[50], which indicates that there is still room for improvement.

In previous works, we showed that atom-centered potentials (ACPs) represent a simple and effective means to improve the accuracy of DFT-based methods. By fitting to high-level *ab initio* wavefunction data, ACPs can be developed to correct for missing dispersion physics in conventional DFT functionals (in this context, ACPs were termed dispersion-correcting potentials, DCPs).[20-22] These DCPs also implicitly correct inherent deficiencies in the underlying functional, such as delocalization and basis-set incompleteness error.[23,24] Recently, we proposed a newer and more systematic way to develop ACPs, and showed that they can be used as a computationally inexpensive means of mitigating the effects of extreme basis-set incompleteness. These basis-set incompleteness potentials (BSIPs) allow the use of minimal or small double-zeta basis-sets with conventional DFT functionals to obtain almost complete-basis-set quality molecular properties.[25] Because ACPs have the same form as conventional effective-core potentials (ECPs)[51-53], they can be used in most computational chemistry programs that allow for the use of ECPs without additional changes to the software.

In this work, we develop a fast and accurate HF-based minimal-basis-set method for the calculation of molecular properties in large molecular systems. Grimme's D3 correction with Becke-Johnson (BJ) damping is used to account for the absence of dispersion in HF, and ACPs are used to correct for the remaining errors, the leading contribution to which is severe basis-set incompleteness error from the minimal-basis-set. The method uses the minimal-basis-set of Huzinaga with scaled exponents (MINIs).[54] ACPs are proposed for the H, C, N, and O atoms and their parameters are obtained by fitting to an extensive set of molecular properties determined using either highly-correlated wavefunction methods at the complete-basis-set (CBS) limit or LC-ωPBE-XDM/aug-cc-pVTZ[55] where obtaining reference data using wavefunction theory is not feasible. The HF-D3/MINIs-ACP approach is computationally efficient and shows excellent performance in the modeling of intermolecular and intramolecular energies and geometries, and serves as an initial proof-of-concept result using the H, C, N, and O atoms that indicate that ACPs can be used successfully to develop computationally inexpensive techniques based on minimal-basis-set electronic structure calculations. One negative aspect of ACPs is that for general use they must be developed for each atom of the periodic table. However, in our (unsystematic) experience, HF-D3/MINIs-ACP gives good results provided ACPs are applied to the majority of atoms in the system.

2. Computational Methodology

2.1 Theoretical background

The method employed to develop the ACPs in this work has been described in detail elsewhere[25], and is only briefly reviewed here. ACPs are one-electron potentials with the same mathematical form as effective-core potentials (ECPs)[51-53]:

$$V^{ACP}(\boldsymbol{r}) = \sum_A U_{L_A}^A(r_A) + \sum_{l=0}^{L_A-1} \sum_{m=-l}^{l} |Alm\rangle\, U_l^A(r_A)\, \langle Alm| \tag{1}$$

with,

$$U_l^A(r) = \sum_k c_{lk}^A\, e^{-\xi_{lk}^A {r_A}^2} \tag{2}$$

where A represents the atoms on which the potentials are centered (H, C, N, and O), r_A is the distance from nucleus, and $|Alm\rangle$ are spherical harmonics centered on atom A. We will refer to the L_A term as the "local" angular-momentum channel. The coefficients c_{lk}^A and exponents ξ_{lk}^A in Eq. 2 are adjustable parameters that are determined via least-squares fitting, as described below. The sum in Eq. 2 runs over the number of Gaussian terms defined for atom A and angular-momentum channel l. Unlike ECPs, ACPs do not replace any electrons of the atoms to which they are applied.

The energy contribution that arises from the application of each ACP is obtained from:

$$E^{ACP}(\boldsymbol{c}, \boldsymbol{\xi}) = \sum_{Alk} c_{lk}^A\, \Delta E_{lk}^A(\xi_{lk}^A) = \boldsymbol{c} \cdot \boldsymbol{\Delta E}(\boldsymbol{\xi})^{\mathsf{T}} \tag{3}$$

where the ACP terms ΔE_{lk}^A are independent of the coefficients to first order in the perturbation induced by the ACP,

$$\Delta E_{lk}^A(\xi_{lk}^A) = \sum_i \langle\psi_i| \left(\sum_{m=-l}^{l} |Alm\rangle\, \exp(-\xi_{lk}^A\, r_A^2)\, \langle Alm|\right) |\psi_i\rangle\,, \tag{4}$$

The $\boldsymbol{c}$ and $\boldsymbol{\Delta E}(\boldsymbol{\xi})^{\mathsf{T}}$ are the corresponding ACP coefficients (in energy units) and energy term vectors (dimensionless), respectively, and ψ_i's are the self-consistent ground-state Kohn-Sham orbitals.

The fact that the first-order perturbation energy term arising from the application of $V^{ACP}(\boldsymbol{r})$ (Eq. 1) is linear in the coefficients is key to our ACP development method.[25] For all molecules in the training set, the ACP terms ($\Delta E_{lk}^A(\xi_{lk}^A)$) are computed for a pre-determined set of exponents designed to affect the region of space relevant to correct the desired molecular properties. The coefficients are determined through a least-squares procedure based on a least-absolute-shrinkage-and-selection-operator (LASSO) method by Tibshirani[56] in which we perform variable selection to determine the optimal exponents to use in our ACPs. At the same time, the magnitude of the coefficients are constrained to ensure that

the contribution of second- and higher-order perturbation terms are negligible. We use the name "non-linearity error" to described the combined contribution of the second- and higher-order perturbation terms.[25]

Other important features of the ACPs are: (i) the use of angular projection operators in the potential allows for the introduction of local anisotropic corrections to respond to changes in the chemical environment of a given atom; (ii) the exponential decrease in the ACP as a function of distance from the nucleus ensures that the effect of a given ACP is localized in the vicinity of the atom; and (iii) the cost of the one-electron integrals associated with the ACPs is negligible compared to SCF calculation, provided an efficient implementation is used. All calculations in this work use the Gaussian 09 program[57].

2.2 Training Data Set

For the purpose of fitting our ACPs, a training set composed of several benchmark sets from the literature was assembled. Only molecules containing H, C, N, and O atoms are used in the set. The molecular properties targeted by the training set include non-covalent binding energies, conformational energies of amino acid dimers and trimers, and molecular deformation energies in small organic molecules. A detailed list of the subsets in our training set is given in Table 1. In total, the training set comprises 9814 data points, including 3235 non-covalent binding energies (S22x5[58,59], S66x8[43,60-62], ACHC[63,64], BBI[63,65], and SSI[63,65] sets), 1599 conformational energies (DIPEPCONF and P26[66] sets), and 4980 covalent molecular deformation energies (MOLdef set).

Table 1. Subsets of the training set used for the ACP fit. The classes are: non-covalent binding energies (NCI), conformational energies (CONF), and DEF (molecular deformation energies). The "Num." column indicates number of data points present in the set, and "Ref. Level" is the calculation level of the reference data.

Data Set	Class	Num.	Description	Ref. level	Ref.
S22x5	NCI	110	Potential energy curves of small non-covalently interacting dimers.	CCSD(T)/CBS	58, 59
S66x8	NCI	528	Potential energy curves of small non-covalent dimers.	CCSD(T)/CBS	43, 60-62
ACHC	NCI	54	Interaction energies of adenine-cytosine nucleobase stacking configurations.	DW-CCSD(T**)-F12/aug-cc-pVDZ	63, 64
BBI [a]	NCI	94	Interaction energies of dipeptide backbone-backbone complexes.	DW-CCSD(T)-F12/aug-cc-pV(D+d)z	63, 65

SSI [a,b]	NCI	2449	Interaction energies of amino acid side chain-side chain complexes.	DW-CCSD(T)-F12/aug-cc-pV(D+d)z	63, 65
P26 [c]	CONF	69	Conformational energies of five isolated small peptides containing aromatic side chains.	CCSD(T)/CBS	66
DIPEPCONF [a]	CONF	1530	Relative energies of ten conformers for each of 153 dipeptide combinations.	LC-ωPBE-XDM/aug-cc-pVTZ	This work
MOLdef	DEF	4980	Molecular deformation energies relative to the equilibrium geometry of 49 small molecules.	LC-ωPBE-XDM/aug-cc-pVTZ	This work

[a] Excluding the dimers containing methionine and cysteine, which contain sulfur. [b] Excluding the dimers with charged monomers. [c] Excluding WGG10 and WGG12 conformers because of a possible error in the data provided in the supporting information of the cited reference.

Most of the subsets in Table 1 are from the literature. For the present work, two new sets were developed called DIPEPCONF and MOLdef. In both sets, LC-ωPBE-XDM/aug-cc-pVTZ was used as the reference method. The rationale for this choice is that given the size of these sets, running wavefunction theory calculations would be too computationally expensive, but LC-ωPBE-XDM/aug-cc-pVTZ is expected to have a much higher accuracy than HF-D3/MINIS or any ACP-corrected version for the systems in these sets.[55] To further justify its use, we checked the performance of LC-ωPBE-XDM/aug-cc-pVTZ on several benchmark sets for conformational energies from the literature. The mean absolute errors (MAEs) are: ACONF[67,68] (alkane conformations), 0.12 kcal/mol; PCONF[63,67,69] (peptide conformations), 0.61 kcal/mol; SCONF[67,70] (sugar conformational energies), 0.24 kcal/mol. PCONF is a subset of the previously proposed P26 set[66], for which LC-ωPBE-XDM/aug-cc-pVTZ has an MAE of 0.52 kcal/mol. For comparison, LC-ωPBE-XDM/aug-cc-pVTZ gives MAEs of 0.27 kcal/mol and 0.18 kcal/mol for the S22[58] and S66[60] sets, and 0.23 kcal/mol and 0.15 kcal/mol for the S22x5[59] and S66x8[43,61,62] sets of non-covalent binding energies, respectively.

DIPEPCONF is a new dataset for dipeptide conformational energies proposed in this work. The dipeptides in the DIPEPCONF set contain only neutral side chains and are capped with acetyl (N-terminal, ACE) and primary amide (C-terminal, NHE) groups. The initial geometries of the dipeptides were generated using the *tleap* tool in *Amber16*[71]. Molecular dynamics (MD) simulations for each amino acid dimer were carried out in the gas-phase using the *ff14SB* force field[72,73], with a heating step of 200 picoseconds followed by a production run of 4200 picoseconds, from which structures were extracted at uniform time intervals to generate a total of 4000 conformers. Each conformer was then subjected to

energy minimization using the same force field. The 4000 conformers for each dipeptide were separated into energy bins and eleven conformers were selected spanning a range of energies. These conformers were then subjected to a single point calculation at the LC-ωPBE-XDM/aug-cc-pVTZ level of theory in order to generate the conformational energy reference data. In total, the DIPEPCONF dataset contains 10 conformational energies (11 conformations) for each of the 153 dipeptide sequences considered. Future work is in progress to extend this set, and this will be published elsewhere.

The MOLdef set contains reference data for molecular deformation energies. For 49 small molecules containing only H, C, N, and O, the equilibrium geometries and vibrational frequencies and normal modes were determined using LC-ωPBE-XDM/aug-cc-pVTZ. The geometry of these molecules was then deformed along each calculated normal mode. For each deformation, the reference energy is the LC-ωPBE-XDM/aug-cc-pVTZ energy difference between the deformed and the equilibrium structures. The intent behind fitting our ACPs to this set is to correct for the erroneous intramolecular geometries that arise from using HF combined with a minimal-basis-set, particularly the spuriously short covalent bonds. Three deformations on each side of the equilibrium geometry were considered along each normal mode, such that the relative energy of the distorted structure never exceeded a few dozen kcal/mol. In total, the MOLdef set consists of 4980 relative energies. The Cartesian coordinates along with the reference energies can be found in the Supporting Information (SI).

Each subset in the training set is assigned a weight for the ACP least-squares fit in order to account for the variable magnitude of the numerical values and number of data points. The weight of each subset in Table 1 is calculated using the formula:

$$w_i = \frac{1}{M_i \times N_i} \tag{5}$$

where M_i is the mean absolute value of the reference energies and N_i is the number of data points for subset i. The weights for all subsets are normalized, and the weighted root-mean-square (wRMS) is defined as:

$$\mathrm{wRMS} = \sqrt{\frac{\sum_{\mathrm{i}}^{\mathrm{subsets}} \sum_{\mathrm{j}}^{\mathrm{N_i}} \mathrm{w_j} \left(\mathrm{y_{ref,j}} - \mathrm{y_{HF\text{-}D3/MINIs\text{-}ACP,j}}\right)^2}{\sum_{\mathrm{i}}^{\mathrm{subsets}} \mathrm{N_i}}} \tag{6}$$

with $y_{\mathrm{ref},j}$ the reference energy and $y_{\mathrm{HF\text{-}D3/MINIs\text{-}ACP},j}$ the energy given by the HF-D3/MINIs-ACP method. The wRMS is minimized as a function of the ACP coefficients c_{lk}^{A} in our least-squares fitting procedure.

2.3 ACP Development

Angular momentum channels for all ACPs are considered up to the maximum angular momentum primitive in the MINIs basis-set, including the local (i.e. local and s for H; and local, s and p for C, N, and O). The thirty chosen ACP exponents are 0.01, from 0.02 to 0.28 in 0.02 steps, from 0.40 to 2.00 in 0.20 steps, and from 2.50 to 5.00 in 0.50 steps. The ACP terms (ΔE_{lk}^{A} (ξ_{lk}^{A})) corresponding to each atom, angular momentum channel, and exponent were determined for each entry in the training set using the Hartree-Fock (HF) method with the MINIs minimal-basis-set and the D3 dispersion correction. The D3 parameters correspond to those for the HF/aug-cc-pVTZ method and with Becke-Johnson damping: $s_6 = 1.0$, $s_8 = 0.9171$, $a_1(\mathrm{BJ}) = 0.3385$, $a_2(\mathrm{BJ}) = 2.8830$ Å. Note that the D3 parameters used in this work are close the ones reported in Ref. 45 for the HF-3c method. To find the optimal ACP, we need to determine the subset of all 330 calculated ACP terms that gives good performance as measured by the magnitude of the wRMS, and at the same time yields coefficients that are small enough that the non-linearity error is small. In this way, we ensure that the statistics resulting from our least-squares fit are a faithful representation of the actual performance of the ACP, and that non-linearity error will not be a large contribution to the method's performance in actual applications.

In our previous work on ACPs for basis-set incompleteness error (BSIPs, Ref. 25), we used an iterative procedure by which all combinations of ACP terms for each atom were explored in turn. Although this method resulted in ACPs with good performance, it is also time consuming and limited by the maximum number of ACP terms in each atom, since the number of combinations increases factorially with this value. In this work, we used an alternative procedure based on the least-absolute-shrinkage-and-selection-operator (LASSO) method by Tibshirani.[56] In LASSO, the least-squares function (in our case, the wRMS in Eq. 6) is minimized subject to the condition that l_1-norm of the ACP coefficients does not exceed a certain bound chosen beforehand. The l_1-norm is given as,

$$\|c\|_1 = \sum_i |c_i| \tag{7}$$

This allows us to constrain the ACP fits to give coefficients as small as we choose. In addition, LASSO also performs variable selection, i.e., for the given constraint on the l_1-norm of the coefficients, LASSO automatically selects the best subset of ACP terms and assigns zero coefficients to the rest. The ACPs determined using the LASSO method have lower wRMS than using the method in Ref. 25. Perhaps more importantly, the fit takes minutes, instead of days, which enables the use of much larger training sets, and the ACPs are not limited to have a certain number of terms per atom (for the ACPs listed in Table 3, LASSO selected 4-9 terms per channel). In this work, we used the local linearization plus active set method proposed by Osborne et al.[74] and implemented in octave/MATLAB by Mark Schmidt[75,76]. After some exploration, we determined that a l_1-norm bound of 5.0 a.u. on the coefficients is a good compromise between accuracy and non-linearity error (see

Figure S1 in SI).

2.4 Validation Data Sets

In addition to the training set, we use several other benchmark sets from the literature to validate the performance of the developed ACPs. The subsets of this validation set have been selected to test non-covalent binding energies and relative conformational energies, and are shown in Table 2.

Table 2. Subsets of the validation set. The "Num." column indicates number of data points, and "Ref. level" is the calculation level of the reference data.

Group	Subset	Num.	Description	Ref. level	Ref.
Non-covalent Interactions					
HYDROCARBONS	HC12	12	Interaction energies of saturated and unsaturated hydrocarbon dimers.	CCSD(T)/CBS	77
	ADIM6	6	Interaction energies of n-alkane dimers.	CCSD(T)/CBS	67, 78
	CH_4 · PAH	382 [b]	Interaction energies of methane with polycyclic aromatic hydrocarbons.	CCSD(T)/CBS	63, 79, 80
	C_2H_4 · NT	75	Interaction energies of ethene with coronene.	CCSD(T)/CBS	63
CO_2-CAPTURE	CO_2 · PAH	249	Interaction energies of CO_2 with polycyclic aromatic hydrocarbons.	CCSD(T**)-F12avg/CBS	63, 81
	CO_2 · NPHAC	96	Interaction energies of CO_2 with nitrogen-doped poly heterocyclic aromatic compounds.	CCSD(T)/CBS	63, 82
LARGE-SYSTEMS	S12L [a]	10 23	Interaction energies of large host-guest	corrected	1, 2,

	S30L [a]		supramolecular motifs.	expt.	83-86
WATER	SHIELDS38	38	Interaction energies of water clusters, $(H_2O)_n$, with n=2-10.	CCSD(T)/CBS	87
CHARGED	IONICHB	120	Dissociation curves of small, charged, hydrogen-bonded complexes.	CCSD(T)/CBS	43
	SSI (charged) [a]	766	Dimers of amino acid side chain-side chain complexes having charged monomers only.	DW-CCSD(T)-F12/aug-cc-pV(D+d)z	63, 65
BIOMOLECULES	A24 [a]	19	Interaction energies of small non-covalently bound complexes.	CCSD(T)/CBS	12
	HSG	21	Model protein-ligand interaction energies.	CCSD(T)/CBS	88, 89
	HBC6	118	Dissociation curves of doubly hydrogen-bonded complexes.	CCSD(T)/CBS	88, 90
Relative Conformational Energies					
CONFORMERS	ACONF	15	Conformational energies of n-alkanes.	W1h-val	67, 68
	BCONF	64	Conformational energies of butane-1,4-diol	CCSD(T)-F12b/cc-pVTZ-F12	91
	MCONF	51	Conformational energies of melatonin.	CCSD(T)/CBS	92

	PCONF	10	Conformational energies of Phenyl-Glycyl-Glycine tripeptide.	CCSD(T**)-F12a/CBS	63, 67, 69
	SCONF	17	Conformational energies of two carbohydrates.	CCSD(T)/CBS	67, 70
	TRCONF	8	Conformational energies of two tetrapeptides	CCSD(T)/CBS	93

[a] Only molecules containing H, C, N and O atoms. [b] 23 out of 405 geometries were missing from the supporting information of Ref. 63.

3. Results

3.1 Optimized ACPs for HF-D3/MINIs

The ACP exponents and coefficients resulting from the fit are listed in Table 3. In general, ACP terms with higher exponents have higher coefficients in absolute value because the corresponding potential term decays faster as it reaches farther away from the atom, and therefore gives a smaller energy contribution. Unlike the transferable BSIPs presented in Ref. 25 which were optimized for the BLYP functional with different combination of basis sets, the ACPs given in Table 3 were optimized specifically for HF-D3/MINIs against high-level reference data, so they cannot be used with other methods or basis-sets because they correct not only for basis-set-incompleteness but also for deficiencies in HF-D3.

The constraint for the LASSO fit was chosen to give an ACP that minimizes the wRMS in a self-consistent calculation over the training set. The number of terms per atom in this ACP is automatically determined by LASSO (in general, 17-21 terms per atom), and therefore the optimized ACPs in Table 3 contain many more terms than those we have developed previously. A sample input file demonstrating the use of the ACPs in the Gaussian program is provided in the SI.

Table 3. HF-D3/MINIs atom-centered potentials for the H, C, N, and O atoms. ξ_{lj}^{A} and c_{lj}^{A} indicate the ACP exponents and coefficients, respectively.

Atom (A)	Function type (l)	ξ_{lj}^{A}	c_{lj}^{A}
H	local	0.01	-0.00001938089
		0.02	-0.00003366414
		0.04	0.001323631399
		0.06	-0.00300563983
		0.10	0.00090072893
		0.12	0.00141564539
		0.22	0.00858410759
		0.40	-0.025198082700
		1.00	0.03040038924
	s	0.01	0.00897184727
		0.02	-0.03953828144
		0.04	0.04867972215
		0.06	0.03449941507
		0.10	-0.01590602911
		0.14	-0.13674945686
		0.40	0.40214230125
		2.50	-1.03353005814
C	local	0.01	0.00001861609
		0.02	0.00021332142
		0.04	-0.00214933468
		0.06	0.00516749173
		0.10	-0.01555553001
		0.16	0.05344004796
		0.26	-0.13758669254
		0.60	0.23120921343
		1.40	-0.80389740259
	s	0.01	-0.01552134192
		0.02	0.00045549650
		0.16	0.02692938642
		0.24	0.01943925488
		0.26	0.03167389276
	p	0.02	0.01111334155
		0.08	-0.01209320774
		0.18	0.00852089081
		0.20	0.05995111066
		0.22	0.01316335258
		0.24	0.12664813920

		1.20	-0.23218368640
N	local	0.01	0.00005935026
		0.02	0.00014788866
		0.04	-0.00030466045
		0.06	0.00062390375
		0.10	-0.00803928317
		0.16	0.01653732266
		0.60	-0.14314848707
	s	0.01	0.00262789783
		0.04	-0.03994239966
		0.06	-0.07041918366
		0.08	-0.03080252469
		0.40	0.07565676804
	p	0.01	-0.01376692773
		0.02	0.01612955962
		0.04	0.00621017916
		0.06	0.06263892557
		0.22	0.04296706278
		1.00	-0.01832723217
		1.20	-0.24431696314
O	local	0.01	-0.00016818935
		0.02	0.00060565334
		0.04	-0.00330007999
		0.06	0.01108579450
		0.08	-0.00827068380
		0.10	-0.00526053319
		0.28	0.00491351734
		0.80	-0.27851873745
		1.00	-0.03196194558
	s	0.01	0.02609885310
		0.02	0.00355349168
		0.04	0.03976597732
		0.10	0.07826701426
	p	0.02	-0.02356374593
		0.04	-0.00042034334
		0.14	-0.00601202155
		0.20	0.03760145141

	0.26	0.02986288464
	0.28	0.02034075728

The first step in the validation of our ACPs is to apply the resulting HF-D3/MINIs-ACP method to the training set to make sure that our least-squares fit is representative of results that would be obtained from self-consistent field (SCF) calculations in which that ACPs are applied and that non-linearity error is not detrimental to the performance of the ACP.[25] By comparing the statistics from the fit and from a self-consistent HF-D3/MINIs-ACP calculation on the training set, we make sure that the contribution from the second- and higher-order terms in Eq. 4 are not significant. Table 4 compares the mean absolute errors (MAEs) obtained from the least-squares fit and from the validation calculations. For comparison, the results obtained using HF-D3/MINIs, HF-D3/aug-cc-pVTZ (abbreviated aTZ), and with the HF-3c method are also given in the table. Note that the HF-3c parameters for the D3 and gCP corrections were obtained using a fit against the S66 and S66x8 sets, respectively, as reported in Ref. 45. The MAEs obtained from the ACP fit deviate by 0.15 kcal/mol or less from the results of the corresponding self-consistent calculations, indicating that the l_1-norm constraint in the LASSO fit was successful in preventing excessive non-linearity error. The MAEs obtained using the uncorrected method range from 0.84 to 3.42 kcal/mol, and application of our ACPs reduce the MAEs by up to a factor of five.

Table 4. Mean absolute errors (MAEs) with respect to high-level reference data for the various subsets of the training set. Uncorrected HF-D3/MINIs is compared to the results obtained from the fitting procedure ("ACP-Fit"), and to the MAE from the application of HF-D3/MINIs-ACP in self-consistent field calculations ("ACP-SCF"). For comparison, the MAEs obtained using the HF-3c and HF-D3/aTZ methods are also provided.

Set	**HF-D3/MINIs**	**ACP-Fit**	**ACP-SCF**	**HF-D3/aTZ**	**HF-3c**
S22	2.17	0.28	0.43	1.03	0.53
S22x5	1.40	0.32	0.36	0.70	0.53
S66	1.74	0.21	0.24	0.68	0.38
S66x8	1.24	0.27	0.28	0.51	0.37
ACHC	1.44	0.24	0.27	0.34	0.28
BBI	1.05	0.27	0.22	0.60	0.87
SSI	0.84	0.17	0.17	0.23	0.21
P26	2.02	0.40	0.47	0.68	1.20
DIPEPCONF	2.44	0.81	0.85	0.89	1.15
MOLdef	3.42	0.92	0.90	1.73	2.90

Table 4 also compares HF-D3/MINIs-ACP to HF-D3/aTZ, which is close to the complete-basis-set limit, and to the HF-3c method. HF-D3/MINIs-ACP outperforms HF-3c in all subsets of our training set. The improvement relative to HF-3c is quite large in the case of BBI and P26. On the other hand, the performance of the HF-D3/MINIs-ACP (0.28 kcal/mol)

and HF-3c (0.37 kcal/mol) methods on the S66x8 set, which was also used as fitting set to obtain the parameters for the gCP correction in the HF-3c approach, is quite similar. It is also interesting to compare the minimal-basis-set methods to HF-D3/aTZ, which is reasonably close to the complete-basis-set limit. The performance of HF-D3/aTZ is rather poor for all subsets, and reflects that the D3 correction provides an energy correction amounting to only about half of the total correlation energy contribution to the properties of the fitting sets. The poor performance of HF-D3/aTZ is particularly evident in the binding energies of small molecular dimers, such as the S22, for which the MAE exceeds 1 kcal/mol (c.f. B3LYP-D3, 0.36 kcal/mol[67]). The MAEs for the S22x5 and S66x8 sets for HF-D3/MINIs-ACP and HF-3c are 0.23/0.34 kcal/mol and 0.14/0.17 kcal/mol, respectively, lower than those of HF-D3/aTZ. These results (indeed all of the results shown in Table 4) indicate that the ACPs and the 3c correction rectify the errors associated with basis-set incompleteness and the partial absence of correlation. In comparison, the MAE in the S22x5 and S66x8 using B3LYP-D3(BJ)/aug-cc-pVTZ is 0.25 kcal/mol and 0.17 kcal/mol, respectively, about 0.1 kcal/mol lower than HF-D3/MINIs-ACP.[94]

Unlike HF-D3/MINIs-ACP, however, HF-3c is significantly worse than HF-D3/aTZ for conformational and, particularly, molecular deformation energies. The high MAEs for the P26, DIPEPCONF and MOLdef sets obtained with HF-3c can be directly related to the fact that the parameters for gCP and SRB corrections were not designed, in principle, for the purpose of rectifying the poor performance of HF-D3/MINIs for conformational and molecular deformation energies. To compare between the intramolecular structures predicted by different HF-based methods, all molecules in the MOLdef set were relaxed using HF-D3/MINIs, HF-D3/MINIs-ACP, and HF-3c, and the resulting geometries were compared to the LC-ωPBE-XDM/aug-cc-pVTZ equilibrium geometries using Kabsch's algorithm[95]. The average root-mean-square deviation (RMSD) of the atomic coordinates were found to be 0.0586 Å (HF-D3/MINIs), 0.0379 Å (HF-D3/MINIs-ACP), and 0.0464 Å (HF-3c). HF-3c and HF-D3/MINIs-ACP both improve HF-D3/MINIs, but the latter gave better geometries. The major difference between these three methods can be quantified using the magnitude of the mean absolute percent errors for the vibrational frequencies as calculated for the MOLdef set, which are 13.4% for uncorrected HF-D3/MINIs, 5.2% for HF-D3/MINIs-ACP, and 14.3% for HF-3c. Therefore, although the SRB term in HF-3c which was specifically fitted to reproduce higher-level intramolecular geometries of 107 small organic molecules,[45] goes a small way towards repairing the intramolecular geometries of MOLdef set and has a noticeable impact on the calculated molecular frequencies. In contrast, ACPs are very successful in correcting the significant errors in HF-D3/MINIs frequencies.

The signed errors and their averages for the four methods described above are shown in Figure 1, which display strip-charts in which all of the signed error data are plotted. In all cases, the ACPs correction is successful in reducing the spread of the errors relative to both HF-D3/MINIs and HF-D3/aTZ. The performance of HF-3c is also excellent with a small bias in all sets except BBI and MOLdef, and a spread of the error that is in general smaller than uncorrected HF-D3/MINIs. The HF-3c results show a larger spread of

errors than HF-D3/MINIs-ACP for conformational energies and molecular deformations.

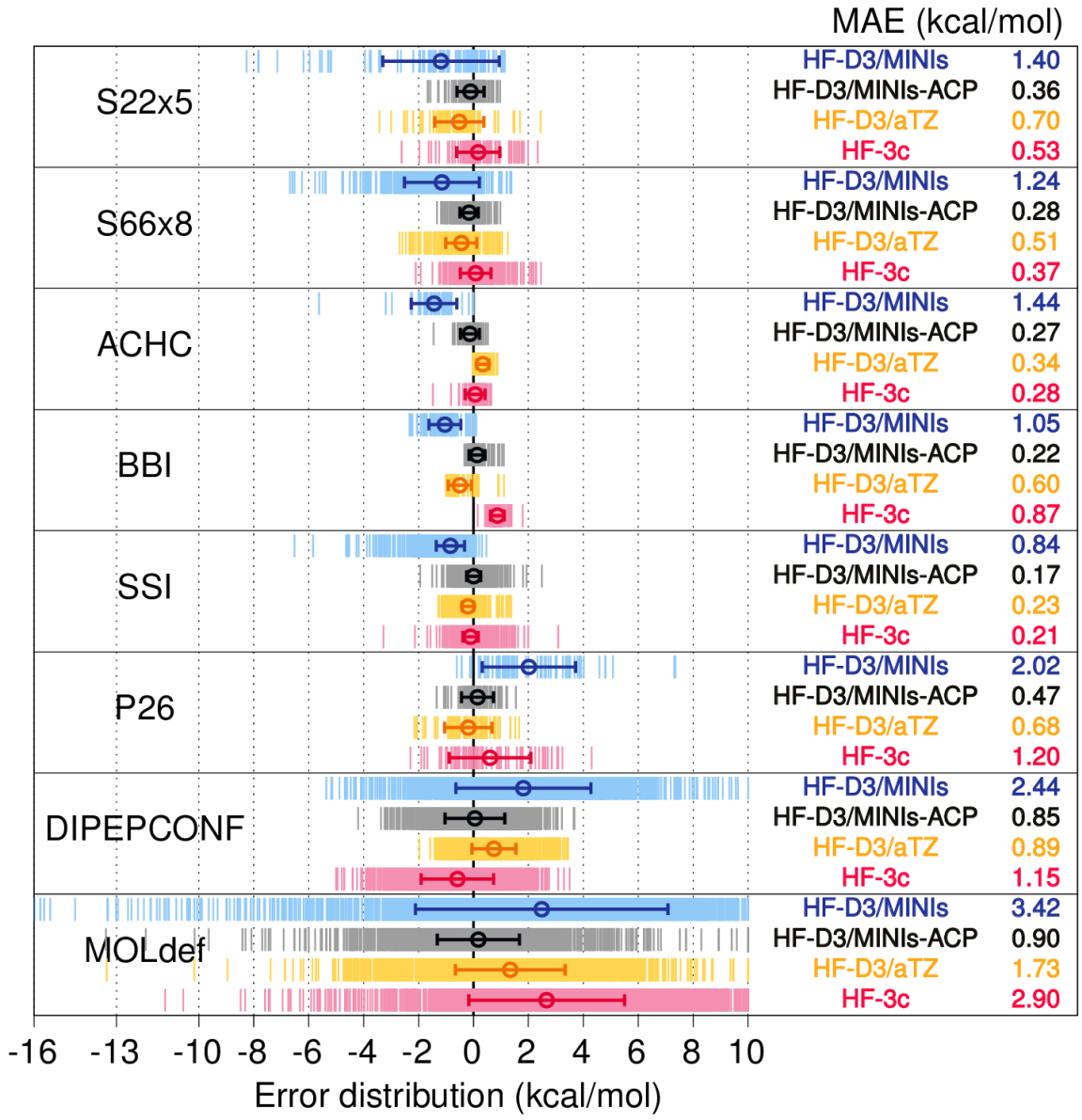

Figure 1. Signed errors associated with the subsets of the training set in Table 1 using HF-D3/MINIs corrected with ACPs (HF-D3/MINIs-ACP) and the 3c correction (HF-3c). The HF-D3/aTZ results are also given. The circle represents the mean error (ME) and the error bar is the standard deviation of the error. The numbers on the right are mean absolute error (MAE) in kcal/mol.

3.2 Performance of HF-D3/MINIs-ACP on the Validation Sets

The performance of HF-D3/MINIs-ACP compared to HF-3c and uncorrected HF-D3/MINIs for the validation set is shown in Table 5. The error distribution, mean errors, and standard deviation for each method and subset are shown in Figure 2.

Table 5. Mean absolute errors (MAEs, in kcal/mol) for the various subsets of the validation set using uncorrected HF-D3/MINIs, and the same method with ACPs (HF-D3/MINIs-ACP) and the 3c correction approach (HF-3c).

Group	Subset	HF-D3/MINIs	HF-D3/MINIs-ACP	HF-3c
HYDROCARBONS		**1.29**	**0.35**	**0.25**
	HC12	1.80	0.26	0.42
	ADIM6	1.90	0.21	0.47
	CH_4 · PAH	1.19	0.29	0.19
	C_2H_4 · NT	1.71	0.67	0.48
CO_2-CAPTURE		**1.72**	**0.88**	**0.57**
	CO_2 · PAH	1.64	0.87	0.55
	CO_2 · NPHAC	1.92	0.89	0.63
LARGE-SYSTEMS		**15.49**	**8.36**	**6.09**
	S12L	15.93	10.27	6.28
	S30L	15.54	7.53	6.01
WATER	SHIELDS38	**30.62**	**4.99**	**7.67**
CHARGED		**3.25**	**1.95**	**2.41**
	IONICHB	4.45	2.47	2.68
	SSI (charged)	3.07	1.87	2.38
BIOMOLECULES		**2.75**	**0.69**	**1.00**
	A24	0.73	0.32	0.44
	HSG	1.69	0.62	0.74
	HBC6	3.27	0.76	1.13
CONFORMERS		**2.27**	**0.69**	**1.05**
	ACONF	1.44	0.98	0.89
	BCONF	2.40	0.50	0.58
	MCONF	0.88	0.71	0.89
	PCONF	2.43	0.50	2.28
	SCONF	5.20	1.17	1.47
	TRCONF	5.14	0.80	3.64

Table 5 and Figure 2 show that, in all cases, the application of the ACPs improves the performance of HF-D3/MINIs. The aggregate MAEs for the binding energies (BEs) and conformational energies are reduced by a factor of 1.9 (LARGE-SYSTEMS) to 6.1 (WATER). This level of performance is similar to HF-3c (improvement factors of 1.3 - 5.2), however the two methods differ in the kinds of systems whose properties are most accurately predicted.

The HF-D3/MINIs-ACP approach improves the BEs in hydrocarbons group (HYDROCARBONS) on average by a factor of about 4, but the performance for the interactions between ethylene and carbon nanotubes (C_2H_4 · NT) and to a lesser extent

between methane and polycyclic aromatic hydrocarbons (CH_4 · PAH) is comparatively worse than the other two subsets of the HYDROCARBONS group. In these subsets, the error is dominated by the non-equilibrium dimers, particularly those in which the two monomers are at shorter distances than at the equilibrium geometry. The HF-D3/MINIs-ACP approach also shows a more modest improvement of a factor of about 2 over uncorrected HF-D3/MINIs for more specialized sets like CO_2-CAPTURE, which consists of systems relevant in carbon dioxide capture studies. This is encouraging because carbon dioxide model systems were not a part of the training set.

Similar observations can be made about the water clusters in the SHIELDS38 set, composed of $(H_2O)_n$ clusters with n=2-10. The ACPs were fitted only to water dimers, but the statistics show a generalized improvement of binding energies for larger water clusters, with the MAE reduced by a factor of 5 compared to uncorrected HF-D3/MINIs. Figure 2 shows that for this set in particular there is a strong over-binding tendency in HF-D3/MINIs caused by basis-set incompleteness error. This effect is corrected by the application of the ACPs, although the over-binding bias is not completely removed. While the ACPs presented in this work are meant to be general, specialized ACPs can also be fitted for particular systems of interest, in the spirit of force field development techniques. For instance, in a recent article,[96] we showed that ACPs fitted to the SHIELDS38 set and a collection of 2-body, 3-body, and 4-body contributions to the binding energy can correct the shortcomings of the BH&HLYP-XDM/aug-cc-pVTZ method to such a degree that the performance of the ACP-corrected method greatly exceeds wavefunction theory results. This is an interesting feature of ACPs that can be useful when extreme accuracy and computational efficiency for a single system is required, such as in molecular dynamics studies of homogeneous substances or crystal structure prediction.

The reduction of the MAE resulting by the application of our ACPs to the large host-guest complexes of S12L and S30L sets (LARGE-SYSTEMS group) and to the charged systems (CHARGED group) is more modest. The S12L, which is contained in the S30L but does not feature any hydrogen-bonded systems, is special in that the performance of various dispersion-corrected functionals show an unusual dependence on the base functional.[3] For instance, while BLYP-XDM, B3LYP-XDM, or LC-ωPBE-XDM routinely outperform PBE-XDM in the description of binding energies in small molecular dimers, it is the latter that performs best for the S12L set, with an MAE of 1.5 kcal/mol[3] (c.f. BLYP, 4.2; B3LYP, 4.0; LC-ωPBE-XDM, 6.8 kcal/mol). The reason for this dependence is unknown at present, but candidates for an explanation are a favorable error cancellation in PBE (hydrogen-bonded systems, for which PBE is notoriously bad, are absent from the S12L) or errors from the methods used in the back-correction of the experimental results from which the reference data were derived. HF-D3/MINIs for the S12L has an MAE of 15.9 kcal/mol, which is reduced to only 10.27 kcal/mol upon application of the ACPs (6.28 in the case of HF-3c). In the S30L, ACPs reduce the MAE from 15.5 kcal/mol to 7.5 kcal/mol (c.f. 6.0 kcal/mol with HF-3c), indicating that they are more successful in representing the hydrogen bonded systems in the S30L not present in the S12L. For comparison, Brandenburg *et al.* reported a MAE of 6.6 kcal/mol

using PBE-D3 at an estimated complete-basis-set limit for the S30L.[8]

The MAE for the charged systems is reduced from 3.25 kcal/mol for HF-D3/MINIs to 1.95 kcal/mol upon application of the ACPs (c.f. 2.41 for HF-3c). The charged systems in the SSI were purposefully left out of the training set, since it is clear that a minimal-basis-set does not have enough flexibility to describe anionic systems. It is also important to note that charged systems are present in the S12L, S30L, and HSG, and the errors for those systems are significantly higher than for the rest of the dimers in these sets. For instance, the MAE for HSG using HF-D3/MINIs-ACP drops from 0.62 kcal/mol to 0.29 kcal/mol when the charged systems are removed from the set. For comparison, Burns *et al.* reported a MAE of 0.48 kcal/mol for the whole HSG set using B3LYP-D3/aug-cc-pVTZ[97] and Torres and DiLabio reported an MAE of 0.15 kcal/mol using B3LYP-DCP/6-31+G(2d,2p)[98]. In spite of this, the application of ACPs is still beneficial, even for the subsets composed solely of charged systems—the MAE is decreased by a factor slightly smaller than 2 both in the IONICHB and the charged systems of the SSI set.

The application of HF-D3/MINIs-ACP to the molecular dimers with importance in biological systems (BIOMOLECULES group) improved the BEs on average by a factor of about 4. The MAE for the HBC6 subset of BIOMOLECULES obtained with HF-D3/MINIs-ACP is the largest amongst all of the subsets at 0.76 kcal/mol. Nevertheless, this level of performance is quite good, suggesting that HF-D3/MINIs-ACP approach may offer a faster alternative to accurately model non-covalent interactions in larger sized molecules of biological significance. For comparison, Burns *et al.* reported MAEs for HBC6 of 0.55 and 1.12 kcal/mol for B3LYP-D3/aug-cc-pVTZ and PBE0-D3/aug-cc-pVTZ, respectively.[97] HF-3c performs somewhat worse on the BIOMOLECULES set, viz., factor of 2.8 improvement over uncorrected HF-D3/MINIs.

Figure 2 shows that for all sets that comprise non-covalent binding energies, HF-D3/MINIs shows a strong bias towards over-binding, which is successfully corrected by ACPs and by the 3c correction. The standard deviation of the errors is also greatly reduced, except for S12L. The same cannot be said about the subsets composed of conformational energies (CONFORMERS group), for which uncorrected HF-D3/MINIs shows errors on both sides of the zero-average error line. The performance of HF-D3/MINIs-ACP for conformational energies is excellent, as demonstrated by the substantial decrease in MAE to 0.69 kcal/mol from 2.27 kcal/mol for uncorrected HF-D3/MINIs. By comparison, the MAE for the CONFORMERS group is 1.05 kcal/mol using HF-3c (Table 5). The results for the tripeptides (PCONF) and tetrapeptides (TRCONF) are particularly good, which is not surprising since our training set is dominated by peptide-peptide interactions (for instance, P26 contains PCONF). The decrease in MAE is also substantial for SCONF (sugar conformations, 5.20 to 1.17 kcal/mol) and BCONF (butane-1,4-diol, 2.40 to 0.50 kcal/mol), but smaller for MCONF (melatonin, 0.88 to 0.71 kcal/mol) and ACONF (hydrocarbons, 1.44 to 0.98 kcal/mol). In all CONFORMER subsets, the bias and the spread of the errors is reduced by the application of the ACPs and the 3c corrections.

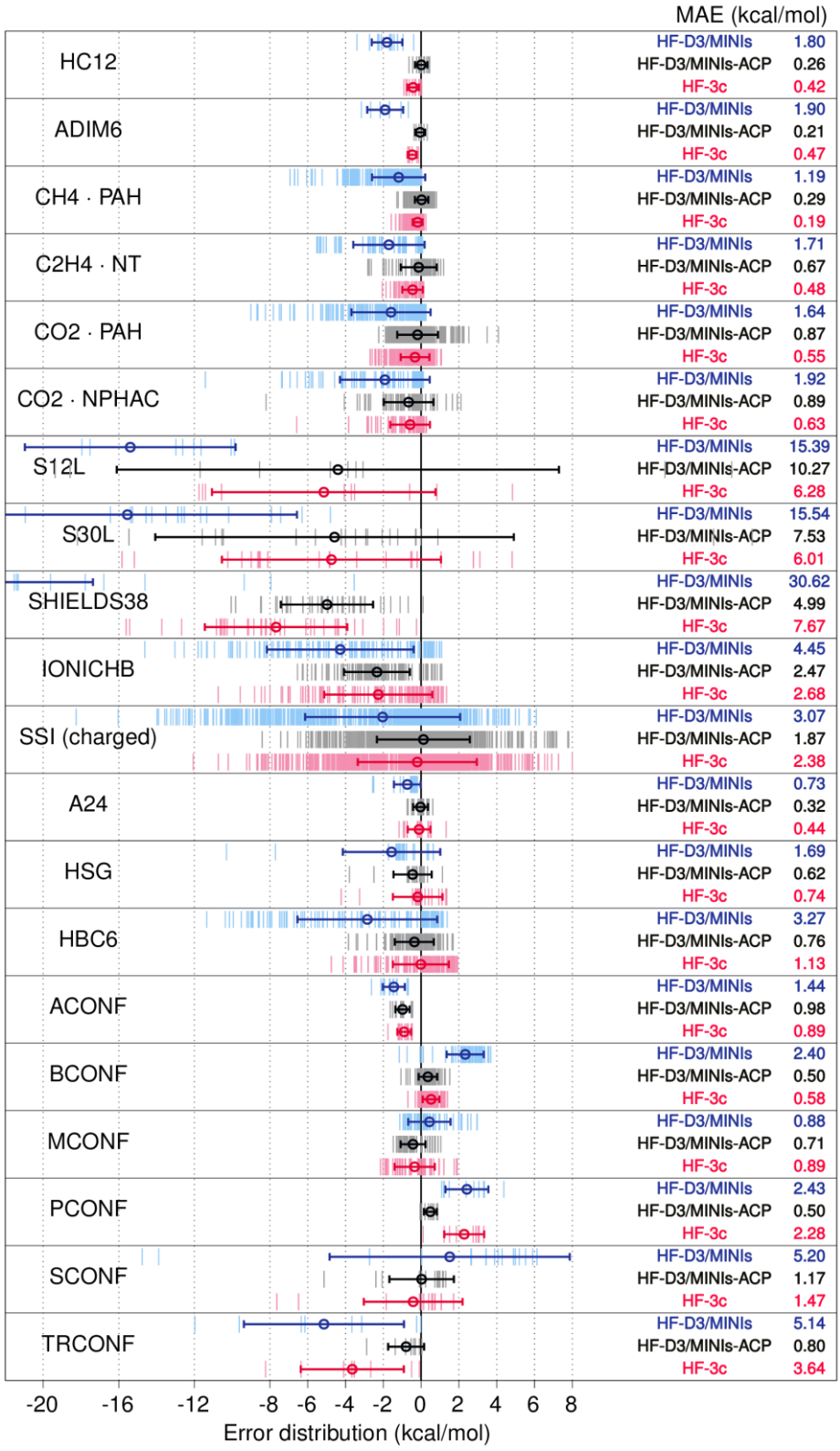

Figure 2. Signed errors associated with the subsets of the validation set in Table 2 using HF-

D3/MINIs with and without ACPs (HF-D3/MINIs-ACP) and the 3c correction approach (HF-3c). The circle represents the mean error (ME) and the error bar is the standard deviation of the error. The numbers on the right are mean absolute error (MAE) in kcal/mol.

4. Discussion and Outlook

The combined analysis of the HF-D3/MINIs-ACP performance on the training (Table 4 and Figure 1) and validation sets (Table 5 and Figure 2) offers some insight into the feasibility of using ACPs for developing a computationally inexpensive method based on minimal-basis-set quantum mechanical calculation. The overall performance is, in general, worse than conventional dispersion-corrected density functionals at the complete-basis-set limit and similar to the previously proposed HF-3c method. The performance of HF-D3/MINIs-ACP indicates that it is particularly suitable for biomolecules, and significantly better than uncorrected HF-D3/MINIs but with a similar computational cost. It is also clear that there is a certain degree of generality to the ACPs, since the MAEs for subsets of the validation set that have very little resemblance to the systems in the training set (e.g. the CO_2-CAPTURE or HYDROCARBONS groups) are consistently improved by the application of our ACPs. Nevertheless, the training set we utilized is dominated by systems derived from biological molecules, particularly proteins, and this is also reflected in the validation set. For instance, the errors in the TRCONF and PCONF conformational energies are much smaller than the other subsets. There is an essential limitation in the description of charged systems, however, caused by the very poor description of anions using a minimal-basis-set, which may have a negative impact on the calculation of zwitterionic species. For the same reason, strongly hydrogen-bonded systems (e.g. double hydrogen bonds in carboxylic acid dimers) are also difficult to model with a minimal-basis-set. As an additional analysis, we checked the performance of HF-D3/MINIs-ACP approach in prediction of the molecular dipole moment of few polar molecules. The reference structure and data of the molecular dipoles containing the four selected atoms were taken from the dipole set reported in Ref. 25. The ACP approach yields an MAE of 0.30 Debye with respect to the LC-wPBE/aug-cc-pVTZ reference, which is indicative that our proposed ACPs compensate to some extent the overall deterioration of the system's electron density description even on use of a minimal basis set like MINIs.

Another limitation of ACPs is that, in principle, they need to be developed for every atom in the system under study. However, work is under way to extend the training set to cover most atoms that usually appear in organic molecules, particularly P and S, which would enable the complete description of DNA and proteins using ACPs. Even if ACPs are only applied to a subset of the atoms in the system, their effect seems to reduce the error from HF-D3/MINIs-ACP, which is not surprising since it is dominated by the extreme basis-set incompleteness of the basis-set. For instance, we applied HF-D3/MINIs to the X40 set[99], comprising non-covalent binding energies of halogenated dimers. (Only the subset of molecules without Br and I was calculated, since there are no MINIs basis functions for

those atoms.) Using HF-D3/MINIs, the MAE is 1.50 kcal/mol, which is reduced to 1.08 kcal/mol upon application of the ACPs, an MAE similar although slightly higher than HF-3c (0.94 kcal/mol). The error is reduced, even though all the molecules in X40 contain at least one halogen atom, for which no ACPs are available.

Our current training set is also somewhat skewed towards peptide-peptide interactions, and this is likely detrimental to the accurate modeling of other types of non-covalent forces. On the other hand, extending the training set is a relatively simple matter. A dispersion-corrected density functional (such as LC-ωPBE-XDM) and a relatively large basis-set are good enough to generate reference data for our fits, since we have shown that these methods have in general a much higher accuracy than what we can obtain using an ACP-corrected minimal-basis-set HF calculation. The use of the LASSO fitting technique also allows training sets with hundreds of thousands or even millions of data points, which would not have been possible with the fitting procedure described in our previous work.[25]

Although extending the present work to create general-purpose ACPs is valuable, another positive feature of the current methodology is that it can be applied to develop ACPs for specific purposes. An example is our recently developed ACPs for water.[96] In addition, any property that is a linear mapping of the electronic energy can be targeted by the ACP, not just the total energy. This was nicely exemplified in our water ACPs[96], which indirectly brought the molecular dipole in water using BH&HLYP-D3/aug-cc-pVTZ to agreement with the experimental value to four significant digits. The ACPs presented in this work are inherently valuable as a general-purpose, computationally inexpensive exploratory tool, particularly for the purpose of modeling peptide-peptide interactions, which is very interesting in the field of quantum mechanical refinement of protein structures.[100,101]

Supporting information

The supporting Information (SI) is available free of charge via the Internet at http://pubs.acs.org and DOI: http://dx.doi.org/xx.xxxx/acs.jctc.xxxxxxxx

Sample Gaussian input file demonstrating the use of ACPs, formulas for the error evaluations, error analysis of fitting and validation sets, combined plot of wRMS vs. l_1-norm and the total number of ACP terms selected by LASSO vs. l_1-norm. (PDF)

Detailed results and Cartesian coordinates of all the data sets used in the study. (ZIP)

Acknowledgements

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada, the Canadian Foundation for Innovation and the University of British Columbia for financial support, and Compute Canada/Westgrid for generous allocation of computing resources. VKP acknowledges the Office of Research Services of the Okanagan Campus of The University of British Columbia for supporting the presentation of this work at the 100th

Canadian Chemistry Conference and Exhibition. The authors express their gratitude towards Lori Burns and C. David Sherill (Georgia Institute of Technology, Atlanta), for providing the BBI and SSI data sets. We thank Professor Jason Loeppky and Dr. Irene Vrbik (UBC) for helpful discussions on the LASSO approach.

Notes

The authors declare no competing financial interest.

References

(1) Risthaus, T.; Grimme, S. Benchmarking of London Dispersion-Accounting Density Functional Theory Methods on Very Large Molecular Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 1580-1591.

(2) Ambrosetti, A.; Alfè, D.; DiStasio, Jr., R. A.; Tkatchenko, A. Hard Numbers for Large Molecules: Toward Exact Energetics for Supramolecular Systems. *J. Phys. Chem. Lett.* **2014**, *5*, 849−855.

(3) Otero-de-la-Roza, A.; Johnson, E. R. Predicting Energetics of Supramolecular Systems Using the XDM Dispersion Model. *J. Chem. Theory Comput.* **2015**, *11*, 4033–4040.

(4) Brandenburg, J. G.; Grimme, S. Dispersion Corrected Hartree–Fock and Density Functional Theory for Organic Crystal Structure Prediction. *Top Curr. Chem.* **2014**, *345*, 1-24.

(5) Whittleton, S. R.; Otero-de-la-Roza, A.; Johnson, E. R. The Exchange-Hole Dipole Dispersion Model for Accurate Energy Ranking in Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2017**, *13*, 441-450.

(6) Whittleton, S. R.; Otero-de-la-Roza, A.; Johnson, E. R. Exchange-Hole Dipole Dispersion Model for Accurate Energy Ranking in Molecular Crystal Structure Prediction II: Nonplanar Molecules. *J. Chem. Theory Comput.* **2017**, *13*, 5332-5342.

(7) Thiel, W. Semiempirical Quantum-Chemical Methods. *WIREs Comput. Mol. Sci.* **2013**, *4*, 145–157.

(8) Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-Cost Quantum Chemical Methods for Noncovalent Interactions. *J. Phys. Chem. Lett.* **2014**, *5*, 4275–4284.

(9) Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116*, 5301–5337.

(10) Sure, R.; Brandenburg, J. G.; Grimme, S. Small Atomic Orbital Basis Set First-Principles Quantum Chemical Methods for Large Molecular and Periodic Systems: A Critical Analysis of Error Sources. *ChemistryOpen* **2016**, *5*, 94–109.

(11) Hohenstein, E. G.; David Sherrill, C. Wavefunction Methods for Noncovalent Interactions. *WIREs Comput. Mol. Sci.* **2011**, *2*, 304–326.

(12) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions Beyond the Common

Approximations: How Accurate Is the “Gold Standard,” CCSD(T) At the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.

(13) Riplinger, C.; Neese, F. An Efficient and Near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.* **2013**, *138*, 034106.

(14) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.

(15) Khaliullin, R. Z.; VandeVondele, J.; Hutter, J. Efficient Linear-Scaling Density Functional Theory for Molecular Systems. *J. Chem. Theory Comput.* **2013**, *9*, 4421-4427.

(16) Gale, J. D. SIESTA: A Linear-Scaling Method for Density Functional Calculations. In *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*; Reimers, J. R., Ed.; John Wiley & Sons, Inc: Hoboken, New Jersey, 2011; pp 45-75.

(17) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1-86). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

(18) Witte, J.; Neaton, J. B.; Head-Gordon, M.; Effective Empirical Corrections for Basis Set Superposition Error in the def2-SVPD Basis: gCP and DFT-C. *J. Chem. Phys.* **2017**, *146*, 234105.

(19) Krishnapriyan, A.; Yang, P.; Niklasson, A. M.N.; Cawkwell, M. J. Numerical Optimization of Density Functional Tight Binding Models: Application to Molecules Containing Carbon, Hydrogen, Nitrogen, and Oxygen. *J. Chem. Theory Comput.* **2017**, Article ASAP. (DOI: 10.1021/acs.jctc.7b00762)

(20) DiLabio, G. A. Accurate Treatment of van Der Waals Interactions Using Standard Density Functional Theory Methods with Effective Core-Type Potentials: Application to Carbon-Containing Dimers. *Chem. Phys. Lett.* **2008**, *455*, 348–353.

(21) Mackie, I. D.; DiLabio, G. A. Interactions in Large, Polyaromatic Hydrocarbon Dimers: Application of Density Functional Theory with Dispersion Corrections. *J. Phys. Chem. A* **2008**, *112*, 10968–10976.

(22) Torres, E.; DiLabio, G. A. A (Nearly) Universally Applicable Method for Modeling Noncovalent Interactions Using B3LYP. *J. Phys. Chem. Lett.* **2012**, *3*, 1738–1744.

(23) DiLabio, G. A.; Koleini, M. Dispersion-Correcting Potentials Can Significantly Improve the Bond Dissociation Enthalpies and Noncovalent Binding Energies Predicted by Density-Functional Theory. *J. Chem. Phys.* **2014**, *140*, 18A542.

(24) van Santen, J. A.; DiLabio, G. A. Dispersion Corrections Improve the Accuracy of Both Noncovalent and Covalent Interactions Energies Predicted by a Density-Functional Theory Approximation. *J. Phys. Chem. A* **2015**, *119*, 6703–6713.

(25) Otero-de-la-Roza, A.; DiLabio, G. A. Transferable Atom-Centered Potentials for the Correction of Basis Set Incompleteness Errors in Density-Functional Theory. *J. Chem.*

Theory Comput. **2017**, *13*, 3505-3524.
(26) Perdew, J. P.; Levy, M. Physical Content of the Exact Kohn-Sham Orbital Energies: Band Gaps and Derivative Discontinuities. *Phys. Rev. Lett.* **1983**, *51*, 1884–1887.
(27) Godby, R. W.; Schlüter, M.; Sham, L. J. Self-Energy Operators and Exchange-Correlation Potentials in Semiconductors. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 10159–10175.
(28) Seidl, A.; Görling, A.; Vogl, P.; Majewski, J. A.; Levy, M. Generalized Kohn-Sham Schemes and the Band-Gap Problem. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *53*, 3764–3774.
(29) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792–794.
(30) 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.
(31) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A* **1988**, *38*, 3098.
(32) 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.
(33) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results Obtained with the Correlation-Energy Density Functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200-206.
(34) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(35) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648.
(36) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
(37) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
(38) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209-220.
(39) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model* **2007**, *13*, 1173-1123.
(40) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. *J. Mol. Model* **2013**, *19*, 1-32.
(41) McNamara, J. P.; Hillier, I. H. Semi-Empirical Molecular Orbital Methods Including Dispersion Corrections for the Accurate Prediction of the Full Range of

Intermolecular Interactions in Biomolecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362-2370.

(42) Tuttle, T.; Thiel, W. OMx-D: Semiempirical Methods with Orthogonalization and Dispersion Corrections. Implementation and Biochemical Application. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159–2166.

(43) Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza P. Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *J. Chem. Theory Comput.* **2009**, *5*, 1749–1760.

(44) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141–151.

(45) Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J. Comput. Chem.* **2013**, *34*, 1672–1685.

(46) 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.

(47) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(48) Johnson, E. R.; Becke, A. D. A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections. *J. Chem. Phys.* **2006**, *124*, 174104.

(49) Kruse, H.; Grimme, S. A Geometrical Correction for the Inter- and Intra-Molecular Basis Set Superposition Error in Hartree-Fock and Density Functional Theory Calculations for Large Systems. *J. Chem. Phys.* **2012**, *136*, 154101.

(50) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar, L.; Gatsiou, C. A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H. Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J.; Zhu, Q.; Groom, C. R. Report on the Sixth Blind Test of Organic Crystal Structure

Prediction Methods. Acta Cryst. **2016**, *B72*, 439−459.
(51) Kahn, L. R.; Baybutt, P.; Truhlar, D. G. Ab Initio Effective Core Potentials: Reduction of All-electron Molecular Structure Calculations to Calculations Involving Only Valence Electrons. *J. Chem. Phys.* **1976**, *65*, 3826–3853.
(52) Christiansen, P. A.; Lee, Y. S.; Pitzer, K. S. Improved Ab Initio Effective Core Potentials for Molecular Calculations. In *World Scientific Series in 20th Century Chemistry, Molecular Structure and Statistical Thermodynamics: Selected Papers of Kenneth S Pitzer;* Pitzer, K. S., Ed.; WORLD SCIENTIFC PUB CO INC, **1993**; Vol. 1; pp 147–152.
(53) Christiansen, P. A.; Lee, Y. S.; Pitzer, K. S. Improved Ab Initio Effective Core Potentials for Molecular Calculations. *J. Chem. Phys.* **1979**, *71*, 4445–4450.
(54) Tatewaki, H.; Huzinaga, S. A Systematic Preparation of New Contracted Gaussian-Type Orbital Sets. III. Second-Row Atoms from Li through Ne. *J. Comput. Chem.* **1980**, *1*, 205–228.
(55) Otero-de-la-Roza, A.; Johnson, E. R. Non-Covalent Interactions and Thermochemistry Using XDM-Corrected Hybrid and Range-Separated Hybrid Density Functionals. *J. Chem. Phys.* **2013**, *138*, 204109.
(56) Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Statist. Soc. B* **1996**, *58*, 267-288.
(57) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; calmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision D.01; Gaussian Inc.: Wallingford, CT, 2009.
(58) Jurečka, P.; Šponer, J.; Černý, 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.
(59) 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.
(60) Rezáč, 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.

(61) 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.

(62) Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. L. The S66x8 Benchmark for Noncovalent Interactions Revisited: Explicitly Correlated Ab Initio Methods and Density Functional Theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905–20925.

(63) 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.

(64) Parker, T. M.; Sherrill, C. D. Assessment of Empirical Models versus High-Accuracy Ab Initio Methods for Nucleobase Stacking: Evaluating the Importance of Charge Penetration. *J. Chem. Theory Comput.* **2015**, *11*, 4197–4204.

(65) 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.

(66) Valdes, H.; Pluháčková, K.; Pitonák, M.; Rezác, J.; Hobza, P. Benchmark Database on Isolated Small Peptides Containing an Aromatic Side Chain: Comparison between Wave Function and Density Functional Theory Methods and Empirical Force Field. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747–2757.

(67) Goerigk, L.; Grimme, S. A Thorough Benchmark of Density Functional Methods for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.

(68) Gruzman, D.; Karton, A.; Martin, J. M. L. Performance of Ab Initio and Density Functional Methods for Conformational Equilibria ofC_nH_{2n+2} Alkane Isomers (n=4−8). *J. Phys. Chem. A* **2009**, *113*, 11974–11983.

(69) Reha, D.; Valdés, H.; Vondrásek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S. Structure and IR Spectrum of Phenylalanyl-Glycyl-Glycine Tripeptide in the Gas-Phase: IR/UV Experiments, Ab Initio Quantum Chemical Calculations, and Molecular Dynamic Simulations. *Chem. Eur. J.* **2005**, *11*, 6803–6817.

(70) Csonka, G. I.; French, A. D.; Johnson, G. P.; Stortz, C. A. Evaluation of Density Functionals and Basis Sets for Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 679–692.

(71) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A.

AMBER 2016, University of California, San Francisco, 2016.
(72) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
(73) 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.
(74) Osborne, M. R.; Presnell, B.; Turlach, B. A. A New Approach to Variable Selection in Least Squares Problems. *IMA Journal of Numerical Analysis* **2000**, *20*, 389-403.
(75) Schmidt, M. Graphical Model Structure Learning with l_1-Regularization. Ph.D. Thesis, The University of British Columbia, Vancouver, August 2010.
(76) Schmidt, M.; Fung, G.; Rosales, R. *Optimization Methods for l_1-Regularization*; Technical Report for The University of British Columbia (TR-2009-19), August 2009.
(77) Granatier, J.; Pitoňák, M.; Hobza, P. Accuracy of Several Wave Function and Density Functional Theory Methods for Description of Noncovalent Interaction of Saturated and Unsaturated Hydrocarbon Dimers. *J. Chem. Theory Comput.* **2012**, *8*, 2282–2292.
(78) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. Estimated MP2 and CCSD(T) Interaction Energies of N-Alkane Dimers at the Basis Set Limit: Comparison of the Methods of Helgaker *et al.* and Feller. *J. Chem. Phys.* **2006**, *124*, 114304.
(79) Smith, D. G. A.; Patkowski, K. Toward an Accurate Description of Methane Physisorption on Carbon Nanotubes. *J. Phys. Chem. C* **2014**, *118*, 544–550.
(80) Smith, D. G. A.; Patkowski, K. Interactions between Methane and Polycyclic Aromatic Hydrocarbons: A High Accuracy Benchmark Study. *J. Chem. Theory Comput.* **2013**, *9*, 370–389.
(81) Smith, D. G. A.; Patkowski, K. Benchmarking the CO_2 Adsorption Energy on Carbon Nanotubes. *J. Phys. Chem. C* **2015**, *119*, 4934–4948.
(82) Li, S.; Smith, D. G. A.; Patkowski, K. An Accurate Benchmark Description of the Interactions between Carbon Dioxide and Polyheterocyclic Aromatic Compounds Containing Nitrogen. *Phys. Chem. Chem. Phys.* **2015**, *17*, 16560–16574.
(83) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem. Eur. J.* **2012**, *18*, 9955–9964.
(84) Antony, J.; Sure, R.; Grimme, S. Using Dispersion-Corrected Density Functional Theory to Understand Supramolecular Binding Thermodynamics. *Chem. Commun.* **2015**, *51*, 1764-1774.
(85) Sure, R.; Grimme, S. Correction to Comprehensive Benchmark of Association (Free) Energies of Realistic Host–Guest Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 5990–5990.
(86) Sure, R.; Grimme, S. Comprehensive Benchmark of Association (Free) Energies of Realistic Host-Guest Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801.
(87) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark Structures and Binding Energies

of Small Water Clusters with Anharmonicity Corrections. *J. Phys. Chem. A* **2011**, *115*, 12034–12046.

(88) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis Set Convergence of the Coupled-Cluster Correction, δ(MP2)(CCSD(T)): Best Practices for Benchmarking Non-Covalent Interactions and the Attendant Revision of the S22, NBC10, HBC6, and HSG Databases. *J. Chem. Phys.* **2011**, *135*, 194102.

(89) Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; David Sherrill, C.; Merz, K. M. Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein−Ligand Complexes. *J. Chem. Theory Comput.* **2011**, *7*, 790–797.

(90) Thanthiriwatte, K. S.; Hohenstein, E. G.; Burns, L. A.; Sherrill, C. D. Assessment of the Performance of DFT and DFT-D Methods for Describing Distance Dependence of Hydrogen-Bonded Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 88–96.

(91) Kozuch, S.; Bachrach, S. M.; Martin, J. M. L. Conformational Equilibria in Butane-1,4-diol: A Benchmark of a Prototypical System with Strong Intramolecular H-bonds. *J. Phys. Chem. A* **2014**, *118*, 293−303.

(92) Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. L. The Melatonin Conformer Space: Benchmark and Assessment of Wave Function and DFT Methods for a Paradigmatic Biological and Pharmacological Molecule. *J. Phys. Chem. A* **2013**, *117*, 2269–2277.

(93) Goerigk, L.; Karton, A.; Martin, J. M. L.; Radom, L. Accurate Quantum Chemical Energies for Tetrapeptide Conformations: Why MP2 Data with an Insufficient Basis Set Should be Handled with Caution. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7028-7031.

(94) Li. A.; Muddana, H. S.; Gilson, M. K. Quantum Mechanical Calculation of Noncovalent Interactions: A Large-Scale Evaluation of PMx, DFT, and SAPT Approaches. *J. Chem. Theory Comput.* **2014**, *10*, 1563−1575.

(95) Kabsch, W. A Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Cryst.* **1976**, A32, 922-923.

(96) Holmes, J. D.; Otero-de-la-Roza, A.; DiLabio, G. A. Accurate Modeling of Water Clusters with Density-Functional Theory Using Atom-Centered Potentials. *J. Chem. Theory Comput.* **2017**, *13*, 4205-4215.

(97) Burns, L. A.; Vázquez-Mayagoitia, A.; Sumpter, B. G.; Sherrill, C. D. Density-Functional Approaches to Noncovalent Interactions: A Comparison of Dispersion Corrections (DFT-D), Exchange-Hole Dipole Moment (XDM) Theory, and Specialized Functionals. *J. Chem. Phys.* **2011**, *134*, 084107.

(98) DiLabio, G. A.; Otero-de-la-Roza, A. Noncovalent Interactions in Density Functional Theory. In *Reviews in Computational Chemistry*; Parill, A. L., Lipkowitz, K. B., Eds.; John Wiley & Sons, Inc: Hoboken, New Jersey, 2016; Vol. 29; pp 1–97.

(99) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.*, **2012**, *8*, 4285-4292.

(100) Goerigk, L.; Reimers, J. R. Efficient Methods for the Quantum Chemical

Treatment of Protein Structures: The Effects of London-Dispersion and Basis-Set Incompleteness on Peptide and Water-Cluster Geometries. *J. Chem. Theory Comput.* **2013**, *9*, 3240–3251.

(101) Goerigk, L.; Collyer, C. A.; Reimers, J. R. Recommending Hartree−Fock Theory with London-Dispersion and Basis-Set-Superposition Corrections for the Optimization or Quantum Refinement of Protein Structures. *J. Phys. Chem. B* **2014**, *118*, 14612−14626.