

# Small basis set density-functional theory methods corrected with atom-centered potentials

Viki Kumar Prasad<sup>a</sup>, Alberto Otero-de-la-Roza<sup>b,\*</sup> and Gino A. DiLabio<sup>a,\*</sup>

<sup>a</sup>. Department of Chemistry, University of British Columbia, Okanagan, 3247 University Way, Kelowna, British Columbia, Canada V1V 1V7.

<sup>b</sup>. Departamento de Química Física y Analítica, Facultad de Química, Universidad de Oviedo, MALTA Consolider Team, E-33006 Oviedo, Spain.

\* E-mail correspondence: aoterodelaroz@gmail.com and gino.dilabio@ubc.ca

## Abstract

Density-functional theory (DFT) is currently the most popular method for modeling non-covalent interactions and thermochemistry. The accurate calculation of non-covalent interaction energies, reaction energies, and barrier heights requires choosing an appropriate functional and, typically, a relatively large basis set. Deficiencies of the density-functional approximation and the use of a limited basis set are the leading sources of error in the calculation of non-covalent and thermochemical properties in molecular systems. In this article, we present three new DFT methods based on the BLYP, M062X and CAM-B3LYP functionals in combination with the 6-31G\* basis set and corrected with atom-centered potentials (ACPs). ACPs are one-electron potentials that have the same form as effective-core potentials, except they do not replace any electrons. The ACPs developed in this work are used to generate energy corrections to the underlying DFT/basis-set method such that the errors in predicted chemical properties are minimized while maintaining the low computational cost of the parent methods. ACPs were developed for the elements H, B, C, N, O, F, Si, P, S, and Cl. The ACP parameters were determined using an extensive training set of 118,655 data points, mostly of complete basis set coupled-cluster level quality. The target molecular properties for the ACP-corrected methods include non-covalent interaction energies, molecular conformational energies, reaction energies, barrier heights, and bond separation energies. The ACPs were tested first on the training set and then on a validation set of 42,567 additional data points. We show that the ACP-corrected methods can predict the target molecular properties with accuracy close to complete basis set wavefunction theory methods, but at a computational cost of double- $\zeta$  DFT methods. This makes the new BLYP/6-31G\*-ACP, M062X/6-31G\*-ACP, and CAM-B3LYP/6-31G\*-ACP methods uniquely suited to the calculation of non-covalent, thermochemical, and kinetic properties in large molecular systems.

## 1. Introduction

Over the past few decades, density functional theory (DFT) has become the leading approach for the quantum mechanical (QM) modeling of various molecular properties.<sup>1</sup> DFT's success can be attributed to its favorable balance between computational cost and accuracy, along with the existence of efficient algorithmic implementations widely available in modern software packages. One of the main sources of error in DFT is the choice of exchange-correlation functional approximation, which determines the accuracy of a given DFT method for a particular purpose. For instance, it is known that common density functional approximations are unable to accurately describe dispersion forces, which are critical when modeling non-covalent interactions and chemical reactions involving large molecules.<sup>2-4</sup> As a result, several works have focused on improving the accuracy of common density functional approximations by developing various dispersion correction techniques.<sup>5-8</sup>

It has been shown that dispersion-corrected DFT methods in combination with large basis sets can predict various molecular properties and in particular non-covalent interaction energies of small and medium-size systems with accuracy similar or slightly lower than that obtained with nearly complete basis set wavefunction theory methods.<sup>9-14</sup> However, even with recent advances in computer technology, the applicability of dispersion-corrected DFT methods with a large basis set is a challenge for systems containing more than ca. one hundred atoms. This is unfortunate because there are many interesting problems involving systems in this molecular size range: supra-molecular and (bio)chemical complexes, nanostructured materials and surfaces, enzyme active sites, and many more.<sup>15-17</sup> The reason for this limitation is the unfavorable increase in computational cost as approximately the third power in the number of basis functions for common DFT methods. Consequently, the development of computationally inexpensive DFT based methods that allow efficient and accurate modeling of large systems is an important area of research.<sup>18-21</sup>

The reduction in computational cost of dispersion-corrected DFT with the use of small double- $\zeta$  basis sets, such as 6-31G\*, allows for the modeling of large molecular systems. The primary sources of error in dispersion-corrected DFT plus 6-31G\* are the exchange-correlation functional approximation and basis set incompleteness error caused by the small size of the basis set. Methods have been recently proposed that show that these shortcomings can be mitigated efficiently.<sup>22-24</sup> For instance, Grimme and co-workers have shown that the D3<sup>25,26</sup> dispersion correction when combined with two additional semi-empirical corrections<sup>27,28</sup> designed to mitigate basis set incompleteness error (collectively known as the “3c” approach<sup>28</sup>), alleviate the deficiencies of the PBEh, HSE, and B3LYP functionals with a double- $\zeta$

basis set (yielding PBEh-3c<sup>29</sup>, HSE-3c<sup>30</sup>, and B3LYP-3c<sup>31</sup> methods). The 3c approach, which was initially proposed for a minimal basis set Hartree–Fock (HF) method (HF-3c<sup>28</sup>), has also recently been extended to triple- $\zeta$  basis set DFT methods (B97-3c<sup>32</sup> and r<sup>2</sup>SCAN-3c<sup>33</sup>). Several other methodologies based on DFT have been proposed in the literature<sup>34–40</sup>, and this underscores the interest in developing computationally inexpensive DFT based methods for modeling large molecular systems.

We have shown in earlier works that atom-centered potentials<sup>41</sup> (ACPs) offer a useful way to mitigate the underlying shortcomings of HF and DFT methods.<sup>42–55</sup> ACPs are similar to one-electron effective-core potentials<sup>56,57</sup> (ECPs) in functional form, except ACPs do not replace any electrons. Sharing a form similar to ECPs allows ACPs to be used in any software package that implements the use of ECPs. ACPs can be developed to yield energy corrections that minimize the errors in predicted properties for a target method and basis set combination by parametrization against high-level reference data. In this way, ACPs can be used to efficiently mitigate the shortcomings of double- $\zeta$  basis set DFT methods, since the use of ACPs incurs only a ca. 10% increase<sup>58</sup> in the computational cost relative to the uncorrected method.

In our recent work<sup>58</sup>, we developed four sets of ACPs for ten elements in combination with small and minimal basis set HF. These ACPs were trained against a set of 73,832 non-covalent properties (interaction energies, molecular conformational energies, and molecular deformation energies). Only non-covalent properties were used in the training set because small basis set HF, and HF in general, is limited to applications that do not involve bond breaking or formation, such as fast geometry optimizations, high-throughput conformer screening, and prediction of non-covalent interaction strengths in large systems. Small basis set DFT is a much more promising approach for modeling thermochemical, kinetic, as well as non-covalent properties since, as previously mentioned, the parent DFT methods are already reasonably successful in the calculation of these quantities. Therefore, ACPs developed for double- $\zeta$  basis set DFT methods that mitigate the errors associated with the density functional approximation and basis set incompleteness error can be used to model all the aforementioned non-covalent properties as well as reaction energies, transition state searches, and in barrier height calculations of large molecular systems.

In this work, we developed ACPs for three density functionals combined with a double- $\zeta$  basis set (6-31G\*<sup>59,60</sup>) and Grimme's D3<sup>25,26</sup> dispersion correction scheme where applicable. The functionals chosen were BLYP<sup>61,62</sup> (generalized-gradient approximation or GGA functional), M062X<sup>63</sup> (hybrid meta-GGA functional), and CAM-B3LYP<sup>64</sup> (range-separated hybrid functional), mainly due to their popularity and performance for the target properties.<sup>13,14,22,65–67</sup> The main target molecular properties are non-covalent interaction energies, molecular conformational energies, reaction energies, barrier heights, and bond

separation energies. ACPs were developed for ten elements commonly encountered in organic chemistry and biochemistry (H, C, N, O, F, P, S, Cl) plus boron and silicon. The ACP development was carried out using a training set composed of 118,655 data points calculated at a high level of theory. We used a regularized linear least-squares fitting procedure (the LASSO<sup>68-70</sup> regression method) to obtain the parameters of the ACPs, which greatly simplifies the use of such a large training set. The strengths and weaknesses of the developed ACPs are evaluated and discussed based on their performance on the training set and a validation set consisting of additional 42,567 data points.

## 2. Computational Details

The ACP development procedure employed in this article has been described in detail in our earlier works<sup>42-44,58</sup>. We summarize it here for convenience. The mathematical form of an ACP is:

$$\hat{V}_{ACP} = \sum_{\alpha} \left( V_{local}^{\alpha}(r) + \sum_{l=0}^{L-1} \sum_{m=-l}^l \delta V_l^{\alpha}(r) |Y_{lm}\rangle\langle Y_{lm}| \right) \quad (1)$$

where  $\delta V_l^{\alpha}(r) = V_l^{\alpha}(r) - V_{local}^{\alpha}(r)$ ,  $\alpha$  represents atom,  $r$  is the distance, and  $|Y_{lm}\rangle\langle Y_{lm}|$  are projection operators using real spherical harmonics based on atom  $\alpha$  with  $l$  angular momentum quantum numbers and  $m$  magnetic quantum numbers. The  $V_{local}^{\alpha}(r)$  and  $\delta V_l^{\alpha}(r)$  terms in Equation 1 are represented by  $N$  Gaussian-type functions:

$$V_l^{\alpha}(r) = \sum_{n=1}^N c_{ln}^{\alpha} \exp(-\xi_{ln}^{\alpha} r^2) \quad \text{for } l = 0, 1, 2, \dots, L \quad (2)$$

where the coefficients ( $c_{ln}^{\alpha}$ ) and exponents ( $\xi_{ln}^{\alpha}$ ) are adjustable parameters determined via a regularized least-squares fit to reference data during the ACP development.

In order to find the exponents and coefficients that best mitigate the errors in the properties predicted using the target method and basis set combination, the ACP operator (Equation 1) is first added as a perturbative correction to the Hamiltonian. To first order in the ACP perturbation, the energy correction is:

$$E_{ACP}(\{c_{ln}^{\alpha}\}, \{\xi_{ln}^{\alpha}\}) = \sum_i \langle \psi_i | \hat{V}_{ACP} | \psi_i \rangle \quad (3)$$

$$E_{ACP}(\{c_{ln}^\alpha\}, \{\xi_{ln}^\alpha\}) = \sum_{\alpha ln} c_{ln}^\alpha \sum_i \langle \psi_i | (|Y_{lm}\rangle \exp(-\xi_{ln}^\alpha r^2) \langle Y_{lm}|) | \psi_i \rangle \quad (4)$$

$$E_{ACP}(\mathbf{c}, \boldsymbol{\xi}) = \sum_{\alpha ln} c_{ln}^\alpha \Delta E_{ln}^\alpha(\xi_{ln}^\alpha) = \mathbf{c} \cdot \boldsymbol{\Delta E}(\boldsymbol{\xi})^T \quad (5)$$

where the index  $i$  runs over occupied molecular orbitals  $\psi$ .

The  $\Delta E_{ln}^\alpha(\xi_{ln}^\alpha) = \langle \psi_i | (|Y_{lm}\rangle \exp(-\xi_{ln}^\alpha r^2) \langle Y_{lm}|) | \psi_i \rangle$  integrals (Equations 4 and 5) are the ACP energy terms and they are equal to the difference between the energy when an ACP with exponent  $\xi_{ln}^\alpha$  is applied and the energy in absence of any ACP, divided by the ACP coefficient. In matrix notation, the  $\mathbf{c}$  and  $\boldsymbol{\Delta E}(\boldsymbol{\xi})^T$  are vectors of ACP coefficients and ACP energy terms, respectively. Equations 1–5 indicate that the energy corrections introduced by the ACPs decay exponentially with interatomic distances and depend on the molecular wavefunction as well as the local chemical environment of a given atom.

The ACP development process starts with the compilation of a training set of target molecular properties. The training set must be chemically diverse and composed of systems exclusively with atoms for which the ACPs are being developed. Next, a set of exponents ( $\xi_{ln}^\alpha$ ) and angular momenta ( $l$ ) on each atom are selected, and the ACP energy terms ( $\Delta E_{ln}^\alpha(\xi_{ln}^\alpha)$ ) are computed for each exponent and each entry in the training set. Once all the ACP energy terms are calculated, the optimal ACP coefficients  $c_{ln}^\alpha$  and the associated exponents are determined using a regularized least-squares fitting subject to a constraint on the sum of the absolute values of the coefficients. The ACP development process ends with the generation of the ACPs for the target method and basis set combination.

The ACPs in this work are designed for correcting the BLYP, M062X, and CAM-B3LYP functionals in combination with the 6-31G\* basis set. These ACPs are tied to these method and basis set combinations and are not transferable to other methods. The target elements chosen for ACP development are H, B, C, N, O, F, Si, P, S, and Cl; most of these atoms are common in organic chemistry and biochemistry applications. Twenty-nine exponents ( $\xi_{ln}^\alpha$ ) were considered: 0.12 to 0.30 in 0.02 steps, 0.40 to 2.00 in 0.10 steps, and 2.50 to 3.00 in 0.50 steps. Angular momenta for the ACP energy terms ( $\Delta E_{ln}^\alpha(\xi_{ln}^\alpha)$ ) were used up to the maximum angular momentum of the valence orbital basis functions present in the 6-31G\* basis set: up to  $s$  for H and  $d$  for B, C, N, O, F, Si, P, S, and Cl. The total number of ACP energy terms was 1,102. This way of generating the ACP energy terms and carrying out the fitting procedure is identical to our previous works<sup>42–44,55,58</sup>. LASSO (Least Absolute Shrinkage and Selection Operator) regression<sup>68–70</sup> is employed to solve the regularized least-squares fitting problem. The advantage of LASSO is that it automatically selects the best subset of ACP energy terms and discards the others by assigning a zero

coefficient to them. All the single-point energy calculations were performed with the *Gaussian16*<sup>71</sup> software package. The D3 parameters used are listed in the SI, along with the ACP coefficients and exponents for each method. An example of the usage of ACPs in the *Gaussian16* program is also given in the SI.

The training set (Table 1) used to parameterize the ACPs comprises data sets from the literature that represent non-covalent and covalent properties such as interaction energies, molecular conformational energies, reaction energies, barrier heights, and bond separation energies. This choice of training set properties is motivated by the potential target applications of the ACP-corrected small basis set DFT methods, namely fast geometry optimizations, conformer screening, and modeling of chemical reactions and non-covalent interaction strengths of large systems. The training set comprises 19,439 non-covalent interaction energies, 11,161 molecular conformational energies, 8,315 reaction energies, 58,197 barrier heights, and 4,502 bond separation energies. Our training set also includes 240 molecular isomerization energies, 219 total atomization energies, and 16,582 molecular deformation energies. In addition to the training set, we also assembled a validation set (Table 2), a collection of data used to test the accuracy of properties computed using ACPs for systems not included in the training set. In total, the validation set consists of 27,783 non-covalent interaction energies, 9,491 molecular conformational energies, 5,205 reaction energies, and 88 barrier heights. The structures and reference energies of all data points in the training and validation sets are provided in the SI. Most of the reference data used in both the training and validation sets were calculated with nearly complete basis set wavefunction theory methods, with the exceptions noted in Tables 1 and 2.

**Table 1.** List of data sets used for training the ACPs.

| Data set(s)                                                                                                                                                                                                 | Data points | Range of reference data (in kcal/mol) | Description of data points                                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---------------------------------------|--------------------------------------------------------------|
| <i>Non-covalent interaction energies of molecular complexes<sup>a</sup>:</i>                                                                                                                                |             |                                       |                                                              |
| HBC6 <sup>77,78</sup> , MiriyalaHB104 <sup>79,80</sup> , IonicHB <sup>81</sup> , HB375x10 <sup>82</sup> , IHB100x10 <sup>82</sup> , HB300SPXx10 <sup>83</sup> , CARBH12 <sup>14</sup>                       | 6,409       | -37.01 to +16.30                      | hydrogen bonding interactions                                |
| S22x5 <sup>78,84,85</sup> , S66x8 <sup>86-88</sup> , S66a8 <sup>87</sup> , A21x12 <sup>89-91</sup> , NBC10ext <sup>78,92-95</sup> , 3B-69-DIM <sup>96</sup> , 3B-69-TRIM <sup>96</sup> , HW30 <sup>97</sup> | 1,895       | -35.76 to +9.34                       | mixed character non-covalent interactions                    |
| B-set <sup>b,64</sup> , F-set <sup>b,64</sup> , Si-set <sup>b,64</sup> , P-set <sup>b,64</sup> , S-set <sup>b,64</sup> , Cl-set <sup>b,64</sup> , Sulfurx8 <sup>98</sup>                                    | 1,000       | -68.05 to +21.57                      | monomers containing at least one B, F, Si, P, S, and Cl atom |
| Pisub <sup>b,99,100</sup> , Pi29n <sup>101</sup> , BzDC215 <sup>102</sup> , C2H4NT <sup>95</sup>                                                                                                            | 379         | -18.30 to +10.33                      | non-stacked and stacked $\pi$ - $\pi$ interactions           |
| Hill18 <sup>103</sup> , X40x10 <sup>104</sup>                                                                                                                                                               | 238         | -14.14 to +11.95                      | halogen bonding interactions                                 |
| PNICO23 <sup>14,105</sup>                                                                                                                                                                                   | 23          | -10.97 to -0.64                       | pnictogen bonding interactions                               |
| ADIM6 <sup>14,25,106</sup> , HC12 <sup>107</sup>                                                                                                                                                            | 18          | -5.60 to -1.30                        | hydrophobic interactions                                     |
| BBI <sup>108</sup> , SSI <sup>108</sup> , NucTAA <sup>b,c,109-112</sup> , CarbyhdBz <sup>113</sup> , CarbyhdNaph <sup>114</sup> , CarbyhdAroAA <sup>b,115</sup> , CarbyhdAro <sup>b,116</sup> ,             | 4,756       | -100.86 to +64.19                     | non-covalent interactions present in various biomolecules    |

|                                                                                                                                                                                                                                                                                   |       |                    |                                                                                                                                               |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| WatAA <sup>b,117</sup> , HSG <sup>78,118</sup> , PLF547 <sup>119</sup> , JSCH <sup>84</sup> , DNASTack <sup>120</sup> , DNA2body <sup>120</sup> , ACHC <sup>121</sup> , BDNA <sup>122</sup> , NucBTrimer <sup>b,123</sup>                                                         |       |                    |                                                                                                                                               |
| Water38 <sup>124</sup> , Water1888 <sup>95,125-127</sup> , Water-2body <sup>d,54</sup>                                                                                                                                                                                            | 2,336 | -92.89 to +5.10    | hydrogen-bonded water dimers and (H <sub>2</sub> O) <sub>n</sub> clusters where n=3–10                                                        |
| CH4PAH <sup>128,129</sup> , CO2MOF <sup>130</sup> , CO2PAH <sup>131</sup> , CO2NPHAC <sup>132</sup> , BzGas <sup>133</sup>                                                                                                                                                        | 876   | -6.02 to +12.17    | non-covalent interactions between gas and substrate molecules                                                                                 |
| SSI-anionic <sup>c,108</sup> , WatAA-anionic <sup>b,c,117</sup> , HSG-anionic <sup>c,78,118</sup> , PLF547-anionic <sup>c,119</sup> , IonicHB-anionic <sup>c,81</sup> , IHB100x10-anionic <sup>c,82</sup> , Ionic43-anionic <sup>c,134</sup>                                      | 1,509 | -135.11 to +88.94  | anionic interactions                                                                                                                          |
| <b>Molecular conformational energies<sup>c</sup>:</b>                                                                                                                                                                                                                             |       |                    |                                                                                                                                               |
| 37Conf8 <sup>135</sup> , DCONF <sup>136</sup> , ICONF <sup>14</sup> , MCONF <sup>137</sup> , Torsion21 <sup>138</sup> , MolCONF <sup>139</sup>                                                                                                                                    | 8,280 | +0.0005 to +25.06  | various molecules representing pharmaceuticals, catalysts, synthetic precursors, industrial chemicals, and organic compounds                  |
| PEPCONF-Dipeptide <sup>b,140</sup> , TPCONF <sup>141</sup> , P76 <sup>142</sup> , YMPJ <sup>143</sup> , SPS <sup>f,144</sup> , rSPS <sup>f,145</sup> , UpU46 <sup>f,146</sup> , SCONF <sup>14,147</sup> , DSCONF <sup>148</sup> , SacchCONF <sup>149</sup> , CCONF <sup>150</sup> | 2,082 | -4.09 to +19.74    | molecules representative of proteins, DNA, RNA, and carbohydrates                                                                             |
| ACONF <sup>151</sup> , BCONF <sup>152</sup> , PentCONF <sup>153</sup>                                                                                                                                                                                                             | 421   | +0.14 to +16.66    | hydrocarbon-like molecules                                                                                                                    |
| Undecamer125 <sup>154</sup>                                                                                                                                                                                                                                                       | 124   | +0.06 to +1.87     | (H <sub>2</sub> O) <sub>11</sub> clusters                                                                                                     |
| PEPCONF-Dipeptide-anionic <sup>b,f,140</sup> , MolCONF-anionic <sup>f,139</sup>                                                                                                                                                                                                   | 254   | -0.47 to +10.96    | negatively charged molecules                                                                                                                  |
| <b>Reaction energies<sup>g</sup>:</b>                                                                                                                                                                                                                                             |       |                    |                                                                                                                                               |
| MN-RE <sup>155</sup>                                                                                                                                                                                                                                                              | 7,555 | -217.97 to +242.47 | automatically generated reactions using molecules from Minnesota Database2015B <sup>156</sup>                                                 |
| BH9-RE <sup>74</sup>                                                                                                                                                                                                                                                              | 449   | -89.85 to +116.88  | from BH9 set comprising chemical reactions belonging to nine types common in organic chemistry and biochemistry                               |
| DIE60 <sup>157</sup>                                                                                                                                                                                                                                                              | 60    | -6.14 to +8.60     | double-bond migration reactions in conjugated dienes                                                                                          |
| FH51 <sup>158,159</sup>                                                                                                                                                                                                                                                           | 51    | -150.81 to -0.18   | reactions involving various organic and inorganic molecules                                                                                   |
| BSR36 <sup>160,161</sup>                                                                                                                                                                                                                                                          | 36    | +2.24 to +49.82    | hydrocarbon bond separation reactions                                                                                                         |
| BH76RC <sup>162-164</sup>                                                                                                                                                                                                                                                         | 30    | -103.91 to +5.60   | hydrogen and non-hydrogen atom transfer reactions of small molecules                                                                          |
| G2RC <sup>14,164,165</sup>                                                                                                                                                                                                                                                        | 23    | -154.04 to -2.18   | reactions whose reactants and products had been taken from the G2/97 set                                                                      |
| RC21 <sup>14</sup>                                                                                                                                                                                                                                                                | 21    | -6.72 to +126.56   | organic radical fragmentation and rearrangement reactions                                                                                     |
| CR20 <sup>166</sup>                                                                                                                                                                                                                                                               | 20    | -35.70 to -7.66    | cyclo-reversion reactions                                                                                                                     |
| PlatonicHD6 <sup>167</sup> , PlatonicID6 <sup>167</sup> , PlatonicIG6 <sup>167</sup>                                                                                                                                                                                              | 18    | -43.64 to +501.85  | homodesmotic,isodesmic, and isogyric reactions involving platonic hydrocarbon cages, C <sub>n</sub> H <sub>n</sub> (where n = 4,6,8,10,12,20) |
| AlkIsod14 <sup>168</sup>                                                                                                                                                                                                                                                          | 14    | +2.20 to +15.40    | isodesmic reactions involving C <sub>n</sub> H <sub>2n+2</sub> alkanes (where n=3-8)                                                          |
| DARC <sup>14,164,169</sup>                                                                                                                                                                                                                                                        | 14    | -60.80 to -14.00   | Diels-Alder reactions                                                                                                                         |
| DC13 <sup>14,63,178,179,170-177</sup>                                                                                                                                                                                                                                             | 12    | -106.00 to +152.60 | reactions that were known to be difficult for DFT methods                                                                                     |
| WCPT6 <sup>180</sup>                                                                                                                                                                                                                                                              | 6     | -0.86 to +11.14    | tautomeric water-catalyzed proton transfer reactions                                                                                          |
| NBPRC <sup>161,164,181</sup>                                                                                                                                                                                                                                                      | 6     | -31.20 to +40.40   | reactions involving NH <sub>3</sub> /BH <sub>3</sub> and PH <sub>3</sub> /BH <sub>3</sub>                                                     |
| <b>Barrier height energies<sup>h</sup>:</b>                                                                                                                                                                                                                                       |       |                    |                                                                                                                                               |

|                                                                                                                                                                                                                                                                                                                                                            |        |                      |                                                                                                                                                                |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Grambow2020-B97D3 <sup>ij,182</sup>                                                                                                                                                                                                                                                                                                                        | 32,722 | -44.42 to +221.34    | reactions involving H, C, N, and O generated using automated potential energy surface exploration                                                              |
| Grambow2020-ωB97XD3 <sup>ij,182</sup>                                                                                                                                                                                                                                                                                                                      | 23,922 | -15.70 to +201.33    | reactions involving H, C, N, and O generated using automated potential energy surface exploration                                                              |
| BH9 <sup>k,74</sup>                                                                                                                                                                                                                                                                                                                                        | 898    | -96.26 to +144.39    | chemical reactions belonging to nine types common in organic chemistry and biochemistry                                                                        |
| E2SN2 <sup>k,183</sup>                                                                                                                                                                                                                                                                                                                                     | 418    | -15.54 to +50.78     | competing E2 and S <sub>N</sub> 2 reactions                                                                                                                    |
| HTBH38 <sup>k,163</sup>                                                                                                                                                                                                                                                                                                                                    | 38     | +1.70 to +38.40      | hydrogen atom transfer reactions of small molecules                                                                                                            |
| NHTBH38 <sup>k,162</sup>                                                                                                                                                                                                                                                                                                                                   | 38     | -12.54 to +106.18    | non-hydrogen atom transfer reactions of small molecules                                                                                                        |
| WCPT27 <sup>k,180</sup>                                                                                                                                                                                                                                                                                                                                    | 27     | -6.38 to +81.24      | water-catalyzed proton-transfer reactions                                                                                                                      |
| BHROT27 <sup>k,14</sup>                                                                                                                                                                                                                                                                                                                                    | 27     | +1.01 to +17.24      | rotation around single bonds                                                                                                                                   |
| BHPERI26 <sup>k,164,184</sup>                                                                                                                                                                                                                                                                                                                              | 26     | +0.50 to +39.70      | pericyclic reactions                                                                                                                                           |
| DBH24 <sup>k,185,186</sup>                                                                                                                                                                                                                                                                                                                                 | 24     | -2.40 to +82.14      | diverse reactions involving small molecules                                                                                                                    |
| INV24 <sup>k,187</sup>                                                                                                                                                                                                                                                                                                                                     | 24     | +4.10 to +79.70      | inversion and racemization reactions                                                                                                                           |
| CRBH20 <sup>k,188</sup>                                                                                                                                                                                                                                                                                                                                    | 20     | +33.71 to +52.42     | cyclo-reversion reactions of heterocyclic rings                                                                                                                |
| PX13 <sup>k,189,190</sup>                                                                                                                                                                                                                                                                                                                                  | 13     | -29.97 to +56.19     | proton exchange reactions in small clusters of H <sub>2</sub> O, NH <sub>3</sub> , and HF                                                                      |
| <b>Bond separation energies<sup>l</sup>:</b>                                                                                                                                                                                                                                                                                                               |        |                      |                                                                                                                                                                |
| BSE49 <sup>76</sup>                                                                                                                                                                                                                                                                                                                                        | 4,502  | 9.38 to 177.24       | breaking of 49 unique X-Y type single bonds (except H-H, H-F, and H-Cl) into corresponding radical fragments, where X and Y are H, B, C, N, O, F, Si, P, S, Cl |
| <b>Others<sup>m</sup>:</b>                                                                                                                                                                                                                                                                                                                                 |        |                      |                                                                                                                                                                |
| MOLdef <sup>b,43</sup> , MOLdef-H2O <sup>d,191,192</sup> , MOLdef-TS <sup>n,o,74</sup>                                                                                                                                                                                                                                                                     | 16,582 | -98.43 to +49.38     | molecular deformation energies of various molecules deformed along their normal modes                                                                          |
| ISO34 <sup>14,193</sup> , ISOL24 <sup>14,194</sup> , IDISP <sup>14,161,164,193,195,196</sup> , EIE22 <sup>197</sup> , PAre1 <sup>14</sup> , AlkIsomer1 <sup>168</sup> , PAH6 <sup>198</sup> , Styrene45 <sup>170</sup> , TAUT15 <sup>14</sup> , H2O16Rel5 <sup>199</sup> , H2O20Rel10 <sup>200</sup> , SW49Rel6 <sup>201</sup> , SW49Rel345 <sup>201</sup> | 240    | -60.28 to +124.46    | isomerization energies                                                                                                                                         |
| W4-17 <sup>202</sup> , PlatonicTAE6 <sup>167</sup> , AlkAtom19 <sup>168</sup>                                                                                                                                                                                                                                                                              | 219    | +2484.26 to +4621.46 | total atomization energies                                                                                                                                     |

<sup>a)</sup> defined as the difference between the energy of the complex and the sum of the monomer energies. A negative interaction energy indicates the complex is more stable than the separated monomers.

<sup>b)</sup> the reference data was recalculated using DLPNO-CCSD(T)/CBS (see Reference 58).

<sup>c)</sup> comprises non-covalently bound dimer complexes where at least one of the monomers is negatively charged.

<sup>d)</sup> the reference data was calculated at the CCSD(T)/CBS level using the same extrapolation method as in Reference 124.

<sup>e)</sup> defined as the difference between the energy of a particular conformer and a lower-energy conformer of the same molecule.

<sup>f)</sup> comprises negatively charged conformers.

<sup>g)</sup> defined as the difference between the sum of energies of reactants minus that of products.

<sup>h)</sup> forward barrier height is the energy difference between transition state and reactant(s) or pre-reaction complex; reverse barrier height is the energy difference between transition state and product(s) or post-reaction complex.

<sup>i)</sup> the reference data was recalculated using DLPNO-CCSD(T)/CBS with the same extrapolation method as in Reference 74.

<sup>j)</sup> the barrier heights are relative to the pre- or post-reaction complexes for forward and reverse barriers, respectively.

<sup>k)</sup> the barrier heights are relative to the isolated reactant(s) or product(s) for forward and reverse barriers, respectively.

<sup>l)</sup> defined as the difference between the energy of a molecule and its radical fragments formed by cleavage of a particular bond.



<sup>m)</sup> includes molecular deformation energies, isomerization energies, and total atomization energies. Molecular deformation energy is the difference between the energy of a molecule deformed along a particular normal mode and the energy of the same molecule at equilibrium. Isomerization energy is the energy difference between a molecule and one of its isomers. Total atomization energy is the energy difference between a molecule and the sum of the energies of all its constituent atoms.

<sup>n)</sup> includes deformations along the imaginary mode of transition state structures from Reference 74. The reference data was calculated at DLPNO-CCSD(T)/CBS level using the same extrapolation method as used in Reference 58.

<sup>o)</sup> contains reference data which is negative in magnitude due to deformation along the imaginary normal mode, indicating that the deformed molecule is more stable than the transition state structure.

**Table 2.** List of data sets used for validating the ACPs.

| Data set(s)                                                                                                                                                                                                                                                                                                                                                                      | Data points | Range of reference data (in kcal/mol) | Description of data points                                                                                                   |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| <i>Non-covalent interaction energies of molecular complexes<sup>a)</sup>:</i>                                                                                                                                                                                                                                                                                                    |             |                                       |                                                                                                                              |
| BlindNCI <sup>203</sup> , DES15K <sup>204</sup> , NENCI-2021 <sup>205</sup>                                                                                                                                                                                                                                                                                                      | 17,413      | -33.78 to +186.83                     | mixed character non-covalent interactions                                                                                    |
| CE20 <sup>189,190</sup> , WaterOrg <sup>206</sup>                                                                                                                                                                                                                                                                                                                                | 2,396       | -46.58 to -10.76                      | hydrogen bonding interactions                                                                                                |
| R160x6 <sup>207</sup> , R739x5 <sup>208</sup>                                                                                                                                                                                                                                                                                                                                    | 5,290       | -12.02 to +6.79                       | close contact interactions                                                                                                   |
| CHAL336 <sup>209</sup>                                                                                                                                                                                                                                                                                                                                                           | 48          | -30.85 to -1.57                       | chalcogen bonding interactions                                                                                               |
| XB45 <sup>210</sup>                                                                                                                                                                                                                                                                                                                                                              | 33          | -13.11 to -0.89                       | halogen bonding interactions                                                                                                 |
| L7 <sup>211,212</sup> , S12L <sup>9,11,212</sup> , S30L <sup>213</sup> , Ni2021 <sup>214</sup>                                                                                                                                                                                                                                                                                   | 54          | -416.08 to -1.68                      | large molecules relevant in supramolecular chemistry and biochemistry                                                        |
| C60dimer <sup>215</sup>                                                                                                                                                                                                                                                                                                                                                          | 14          | -6.88 to +12.07                       | C <sub>60</sub> dimers                                                                                                       |
| H2O20Bind10 <sup>200</sup>                                                                                                                                                                                                                                                                                                                                                       | 10          | -200.54 to -196.59                    | (H <sub>2</sub> O) <sub>20</sub> clusters                                                                                    |
| HW6Cl <sup>b,200,216</sup> , HW6F <sup>b,200,216</sup> , FmH2O10 <sup>b,200,216</sup> , SW49Bind345 <sup>b,201</sup> , SW49Bind6 <sup>b,201</sup> , Anionpi <sup>b,217</sup> , IL236 <sup>b,218</sup> , DES15K-anionic <sup>b,204</sup> , NENCI-2021-anionic <sup>b,205</sup> , CHAL336-anionic <sup>b,209</sup> , XB45-anionic <sup>b,210</sup> , S30L-anionic <sup>b,213</sup> | 2,525       | -171.42 to +66.15                     | anionic interactions                                                                                                         |
| <i>Molecular conformational energies<sup>c)</sup>:</i>                                                                                                                                                                                                                                                                                                                           |             |                                       |                                                                                                                              |
| SafroleCONF <sup>219</sup> , AlcoholCONF <sup>220</sup> , BeranCONF <sup>221</sup> , Torsion30 <sup>d,222</sup> , ANI1ccxCONF <sup>e,f,223</sup>                                                                                                                                                                                                                                 | 7,447       | +1E-3 to +49.96                       | Safrole or 5-(2-propenyl)-1,3-benzodioxol, small alcohol molecules, biaryl drug-like molecules, and small organic molecules, |
| MPCONF196 <sup>g,224</sup> , PEPCONF-Tripeptide <sup>h,140</sup> , PEPCONF-Disulfide <sup>140</sup> , PEPCONF-Cyclic <sup>140</sup> , PEPCONF-Bioactive <sup>140</sup>                                                                                                                                                                                                           | 1,874       | -0.47 to +81.00                       | peptide-like molecules                                                                                                       |
| PEPCONF-Disulfide-anionic <sup>i,140</sup> , PEPCONF-Bioactive-anionic <sup>i,140</sup>                                                                                                                                                                                                                                                                                          | 170         | +0.17 to +33.79                       | negatively charged molecules                                                                                                 |
| <i>Reaction energies<sup>l)</sup>:</i>                                                                                                                                                                                                                                                                                                                                           |             |                                       |                                                                                                                              |
| W4-17-RE <sup>f,155</sup>                                                                                                                                                                                                                                                                                                                                                        | 5,205       | -380.97 to +364.92                    | automatically generated reactions using molecules from the W4-17 <sup>202</sup> set                                          |
| <i>Barrier height energies<sup>k)</sup>:</i>                                                                                                                                                                                                                                                                                                                                     |             |                                       |                                                                                                                              |
| WaterOrgBH <sup>1,m)</sup>                                                                                                                                                                                                                                                                                                                                                       | 88          | +12.81 to +61.50                      | pericyclic reactions in absence and presence of water clusters                                                               |

<sup>a)</sup> defined as the difference between the energy of the complex and the sum of energy of the monomers. A negative interaction energy indicates the complex is more stable than the separated monomers.

<sup>b)</sup> comprises non-covalently bound complexes with at least one negatively charged monomer.

<sup>c)</sup> defined as the difference between the energy of a particular conformer and a lower-energy conformer of the same molecule.

<sup>d)</sup> only 30 systems used; we could not find the rest of the systems mentioned in the supporting information of Reference 222.

<sup>e)</sup> contains mostly conformational energies but also some molecular deformation energies.

<sup>f)</sup> only a subset of the actual data used.

<sup>g)</sup> only macrocyclic peptides used.

---

<sup>b)</sup> only a subset from the PEPCONF<sup>140</sup> database for which reference data was recalculated at the DLPNO-CCSD(T)/CBS level of theory (see Reference 58 for more details).

<sup>i)</sup> comprises negatively charged conformers.

<sup>j)</sup> defined as the difference between the sum of energies of reactants minus that of products.

<sup>k)</sup> forward barrier height is the energy difference between transition state and reactant(s) or pre-reaction complex; reverse barrier height is the energy difference between transition state and product(s) or post-reaction complex.

<sup>l)</sup> the barrier heights are relative to the pre- or post-reaction complexes for forward and reverse barriers, respectively.

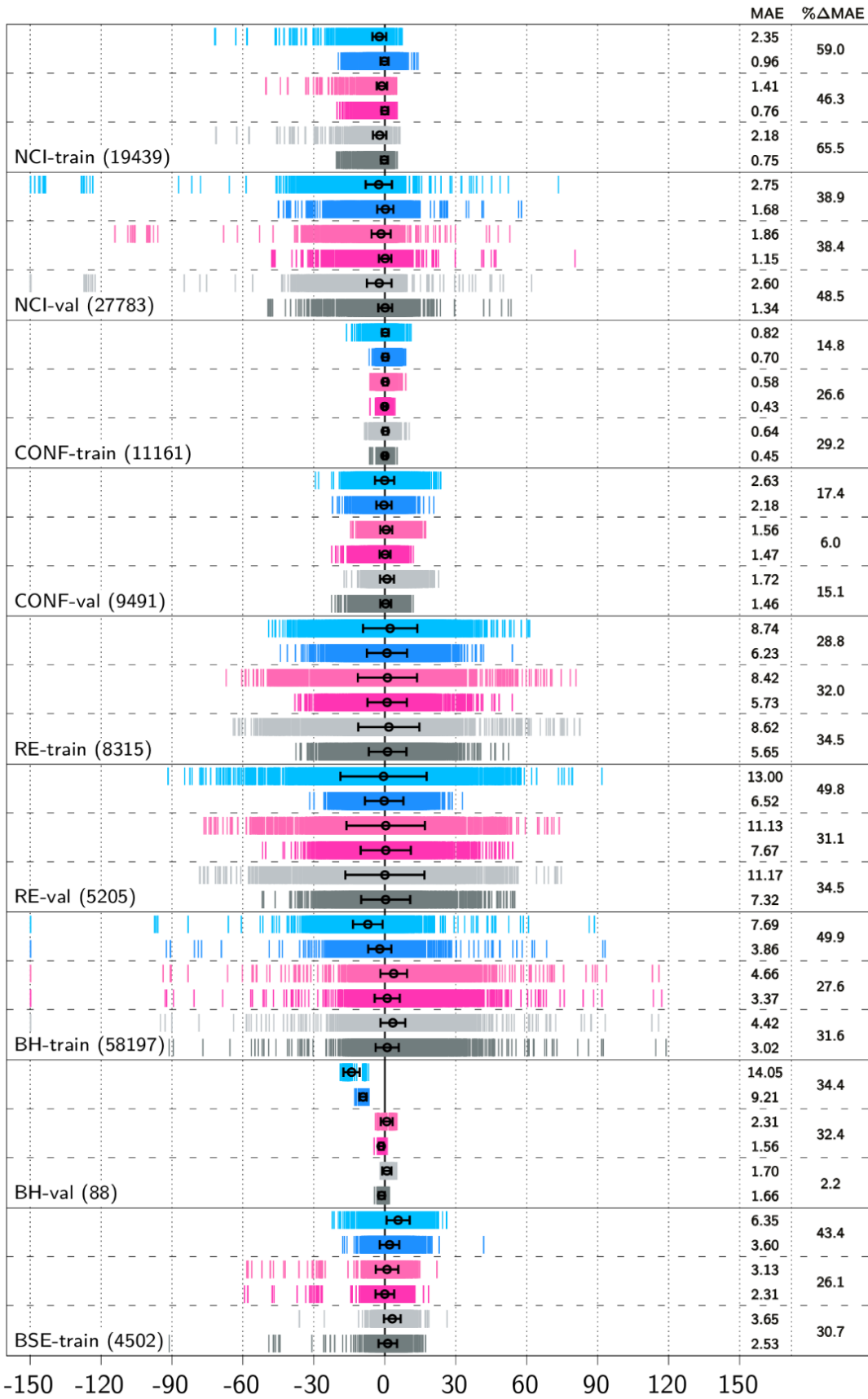
<sup>m)</sup> contains unpublished data generated in an ongoing project, which will be published elsewhere.

---

### 3. Results and Discussion

Results obtained using the target methodology (BLYP-D3/6-31G\*, M062X/6-31G\*, and CAM-B3LYP-D3/6-31G\*) with and without the proposed ACPs compared against the reference data are shown in Figure 1. The figure depicts the signed error distribution as vertical lines along with the mean signed errors (MSEs) (open circles) and the standard deviations (SDs) of the errors (horizontal black lines). The mean absolute errors (MAEs) of each method and the percentage changes in mean absolute error ( $\% \Delta \text{MAE}$ ) on application of the ACPs are also given on the right. A more detailed comparison can be found in the supporting information (Tables S1 and S2 of the SI). A detailed breakdown of the error analysis of each method by subset can be found in Figures S1 and S2 and Tables S3 and S4 of the supporting information. In the following, the results obtained from the application of ACPs are discussed for the different molecular properties in the training and validation sets.

**Figure 1.** Error distribution (in kcal/mol) associated with the uncorrected and ACP-corrected double- $\zeta$  DFT methods. The various molecular properties represented are: “NCI” or non-covalent interaction energies, “CONF” or molecular conformational energies, “RE” or reaction energies, “BH” or barrier height energies, and “BSE” or bond separation energies. Suffixes “train” and “val” are short for training and validation, respectively (see Table 1 and 2). Methods shown include BLYP-D3/6-31G\* (light blue), BLYP-D3/6-31G\*-ACP (blue), M06-2X/6-31G\* (light pink), M06-2X/6-31G\*-ACP (pink), CAM-B3LYP-D3/6-31G\* (light grey), and CAM-B3LYP-D3/6-31G\*-ACP (grey). The black circles represent the mean signed errors (MSEs, kcal/mol) and the black error bars are the standard deviations of the error (SDs, kcal/mol). The numbers on the right-hand side of each panel are the mean absolute errors (MAEs, kcal/mol) and the percentage change in MAEs upon the application of ACPs ( $\% \Delta \text{MAE}$ ) for each method.  $\% \Delta \text{MAE}$  is defined as  $[\text{MAE}(\text{base method}) - \text{MAE}(\text{ACP-corrected method})] / \text{MAE}(\text{base method}) \times 100\%$ . The X-axis has been capped at -150 (left) and +150 kcal/mol (right) for clarity.



## (i) Non-covalent interaction energies

Regarding non-covalent interaction energies (“NCI”), the application of ACPs to the double- $\zeta$  basis set DFT methods decreases the overall MAE by 46–65%, indicating a substantial improvement in the description of non-covalent interactions. This is reasonable given that ACPs have been shown to be adequate at mitigating the basis set incompleteness error which is known to greatly affect the calculation of non-covalent interaction energies.<sup>43,44,72</sup> The best methods based on the overall performance for non-covalent interactions are M062X/6-31G\*-ACP and CAM-B3LYP-D3/6-31G\*-ACP, both with an MAE of about 0.75 kcal/mol. For comparison, BLYP-D3/6-31G\*-ACP yields an overall MAE of 0.96 kcal/mol. A closer look into individual NCI data sets (Figure S1 of SI) reveals that the ACPs applied to CAM-B3LYP-D3/6-31G\* result in a more uniform reduction in the MAEs of the uncorrected method compared to M062X/6-31G\*. BLYP-D3/6-31G\*-ACP also performs uniformly better than the parent uncorrected method, although the individual data set MAEs are slightly higher than CAM-B3LYP-D3/6-31G\*-ACP.

We now take a more detailed view at the performance of CAM-B3LYP-D3/6-31G\*-ACP on the various non-covalent interaction types in the training set. Application of the ACPs to this functional result in MAE reductions for the hydrogen bonding and mixed-character non-covalent interactions greater than 50%. The description of halogen bonding and pnictogen bonding is also improved with MAE reductions in the range of about 19–57%, even though the number of data points for these interaction types in the training set is comparatively smaller than the others. Stacked and non-stacked  $\pi$ - $\pi$  interactions are also described better with the ACPs, with MAEs reduced by about 17–34%. In addition to the various interaction types, ACPs were also trained on typical systems relevant in organic chemistry and biochemistry, where these non-covalent interactions operate co-operatively. Examples include interacting nucleotides or proteins interacting with carbohydrates, nucleotides, drugs, water, and other proteins. For CAM-B3LYP-D3/6-31G\*, the MAE reduction caused by the ACPs in the data sets of biochemical significance range between 29% and 94%, indicating that ACPs significantly enhance the performance of CAM-B3LYP-D3/6-31G\* for modeling non-covalent interactions in biomolecular systems. ACPs for CAM-B3LYP-D3/6-31G\* also lead to a better description of other non-covalently interacting systems in the training set, including gas-substrate and water-water complexes, with MAE reductions in the 24–91% range.

Even though CAM-B3LYP-D3/6-31G\*-ACP improves on the base uncorrected method for almost all non-covalent interaction types, some outliers with relatively high error exist, which is expected given the enormous size of the training set. There are only two NCI subsets in the training set (ADIM6 and

HC12) where CAM-B3LYP-D3/6-31G\*-ACP yields higher MAEs than uncorrected CAM-B3LYP-D3/6-31G\*, both featuring mainly hydrocarbon interactions and contributing only 18 data points to the training set. The increase in MAE upon application of ACPs is likely the result of the relative scarcity of pure hydrophobic type interactions in the training set. Besides hydrophobic contacts, anionic interactions are the other set of interaction types (SSI-anionic, WatAA-anionic, HSG-anionic, PLF547-anionic, IonicHB-anionic, IHB100x10-anionic, and Ionic43-anionic) where there is room for improvement. In this case, ACPs lead to an improved description compared to the uncorrected double- $\zeta$  basis set DFT methods. However, the remaining relatively high errors are probably due to the fact that the 6-31G\* basis set lacks diffuse basis functions, which are known to be required for modeling anionic systems.<sup>73</sup>

Given the overall good performance of ACPs for non-covalent interactions in the training set, we now examine their performance for systems outside the training set. Figure 1 shows that ACPs successfully bring down the overall MAE of the double- $\zeta$  basis set DFT methods by about 38–48% for the NCI validation subset (“NCI-val”). A more detailed look into the results (Figure S3 of SI) shows that application of ACPs leads to MAE reductions (% $\Delta$ MAE) in the range of about 27–67% for the NCI validation subsets containing complexes featuring a mix of common interactions found in large molecular systems (BlindNCI, DES15K, and NENCI-2021 data sets with a total of 17,413 data points). This range of % $\Delta$ MAE for the subsets used in the validation resembles the % $\Delta$ MAE obtained for data sets in training set with mixed character interactions such as S22x5, S66x8, and S66a8. Regarding hydrogen bonding interactions, the CE20, WaterOrg, and H2O20Bind10 subsets of the validation set feature these types of interactions. The MAEs for these three subsets are improved significantly on the application of ACPs by about 62–92%, depending on the method. The large reduction in error observed for mixed character and hydrogen bonding interactions in the validation are probably a consequence of the fact that the training set contains more data points of these two kinds than any other interaction type. In any case, the similarity between the error reduction in the validation and training set suggests that the proposed ACPs are fairly robust regarding these interactions, i.e., they can be applied to similar systems outside the training set.

Regarding the other interaction types, an assessment of the ACP performance in the validation stage for  $\pi$ - $\pi$  stacking and pnictogen bonding interactions could not be carried out due to the scarcity of high-level reference data in the literature. This scarcity compelled us to include all available systems containing these interaction types in the training set instead of reserving them for validation. The ACPs were further validated on systems containing other interaction types such as halogen bonding, chalcogen bonding, and close contact repulsions (XB45, CHAL336, R160x6, and R739x5 data sets) that were not specifically part of the training set. The MAEs for the data sets representing halogen bonding were mostly deteriorated (by

>60%) on the application of ACPs, except for BLYP-D3/6-31G\*, where the MAE was improved by 56%. Application of ACPs to subsets representing chalcogen bonding led to a decrease in the MAEs of uncorrected methods by 6–42%, most likely due to the presence of O and S containing complexes in the training set that were not purely chalcogen bonded. On the other hand, the MAEs of the uncorrected methods for the subsets representing close contact repulsions were initially low (0.50–0.97 kcal/mol), and the application of ACPs led to either increase (by 9–30%) or decrease (by 9–33%). All these findings suggest an under-representation of halogen bonding, chalcogen bonding, and close contact repulsions in the training set compared to other interaction types. Therefore, future ACP development work will require more such systems to be included in the training set in order to increase the diversity and robustness of the resulting ACPs.

The application of ACPs to complexes containing at least one monomer with negative charge in the validation set led to mostly a reduction in the MAEs of the uncorrected methods by 19–75%. Nevertheless, applying the ACPs could only bring down the MAEs to values that were greater than 2 kcal/mol. This indicates that the performance of ACPs for anionic interactions in the validation set convey the same message as in the training set: even though errors decrease on the application of ACPs, the 6-31G\* basis set is inappropriate for modeling anionic systems.

Finally, we tested the proposed ACPs for their performance regarding non-covalent interaction energies in some more challenging complexes that are significantly different from those in the training set. In particular, we used the C60dimer set of non-covalent interaction energies between C<sub>60</sub> dimeric complexes and the L7, S12L, S30L, and Ni2021 sets containing interaction energies between relatively large supramolecular systems. These data sets provide a more stringent test for ACPs than the other validation subsets because of the large size of the systems involved (the absolute reference energies range from 25 kcal/mol to 416 kcal/mol) as well as the multiplicity of and cooperativity between the non-covalent interactions present in these systems. The application of ACPs led to an overall reduction in the MAEs of the underlying methods for these sets by 11–66%, with only a few exceptions. For example, ACPs for CAM-B3LYP-D3/6-31G\* reduce the MAEs of 4 out of 5 data sets by about 30–51%, indicating once again the robustness of the corresponding ACPs. The application of ACPs with CAM-B3LYP-D3/6-31G\* led to an increase in the MAE of the C60dimer set from 1.78 kcal/mol to 3.97 kcal/mol. Although the error increases, this result also demonstrates that in the case of failure due to the systems studied being very different from those on which the ACP were trained, the results from the ACP-corrected methods are far from being catastrophic.

## (ii) Conformational energies

Another molecular property included in the training set is molecular conformational energies (“CONF”) with 11,161 data points. The purpose of these data is to inform ACPs about how molecular motion along rotatable bonds and torsional angles involving various effects ( $\pi$ -conjugation, steric interactions, intramolecular hydrogen-bonding, and electron repulsion) influence the molecular potential energy surfaces. The CONF systems include peptides, nucleotides, carbohydrates, alcohols, hydrocarbons, (H<sub>2</sub>O)<sub>11</sub> clusters, and other molecules representing pharmaceuticals, catalysts, synthetic precursors, industrial chemicals, and organic compounds. Interestingly, the overall MAE of all uncorrected double- $\zeta$  basis set DFT methods for CONF in the training set is below 1 kcal/mol, likely due to error cancellation. Application of ACPs further reduces the MAEs by about 15–29% and brings them down to relatively low values: 0.70 (BLYP-D3/6-31G\*-ACP), 0.43 (M062X/6-31G\*-ACP), and 0.45 kcal/mol (CAM-B3LYP-D3/6-31G\*-ACP).

In order to understand the performance of ACPs for CONF in the training set, we take a closer look at the results of CAM-B3LYP-D3/6-31G\*-ACP (see Figure S1 of SI). These ACPs perform well for most subsets and are particularly suitable for peptides (PEPCONF-Dipeptide, TPCONF, P76, YMPJ), carbohydrates (SCONF, DSCONF, SacchCONF, CCONF), alcohols (BCONF), the melatonin molecule (MCONF), (H<sub>2</sub>O)<sub>11</sub> clusters (Undecamer125), and a mix of various medium-sized organic molecules (37Conf8). The reduction in MAE for these subsets ranges between 24–84%, with most of them generally showing an improvement greater than 47%. The MAE reduction for other CONF subsets, like those containing organic molecules that are drug-like or have industrial relevance (DCONF and MolCONF) range between 13–18%, probably because of the already quite low MAEs of the uncorrected method (0.51 kcal/mol for DCONF and 0.41 kcal/mol for MolCONF).

Despite the general improvement in the CAM-B3LYP-D3/6-31G\* method for conformational energies caused by ACPs, some outliers exist. The conformational energy data sets where the errors of ACP-corrected CAM-B3LYP-D3/6-31G\* are higher than the uncorrected method are mainly found in the data sets for which the uncorrected method already has very small MAEs. For example, the hydrocarbon conformer subsets (ACONF and PentCONF) have MAEs of only 0.04–0.06 kcal/mol with CAM-B3LYP-D3/6-31G\*. The MAEs of the ACP-corrected method for these data sets are higher than the uncorrected method but are still below 0.50 kcal/mol. Some other outlier subsets have MAEs for the uncorrected method in the range 0.25–0.58 kcal/mol, with the MAE from the ACP-corrected method rising to 0.42–0.74 kcal/mol.

Regarding the performance of ACPs for conformational energies outside the training set, we used 9,491 conformational energy data points in the validation set. The application of ACPs to the CONF data in the validation set (“CONF-val”) shows overall good performance. The ACPs reduce the MAEs of M062X/6-31G\* and CAM-B3LYP-D3/6-31G\* by about 5–37% for two of the large data sets containing various conformers of organic molecules, viz. MAEs for Torsion30 of 0.32 and 0.36 kcal/mol and for ANI1ccxCONF of 1.72 and 1.68 kcal/mol, with M062X/6-31G\*-ACP and CAMB3LYP-D3/6-31G\*-ACP, respectively. ACPs applied to BLYP-D3/6-31G\* also result in a reduction of MAEs (22–28%) that brings the MAEs down to 0.47 kcal/mol for Torsion30 and 2.76 kcal/mol for ANI1ccxCONF.

The performance of ACPs in the representative Torsion30 and ANI1ccxCONF validation examples shows that ACPs generally perform well for CONF data points that share similarities with those used in the training set. However, systems that are significantly different from the training set compilation result in somewhat higher but not catastrophic errors, for example in the data set containing conformational energies of peptide model systems with a disulfide-bridged bond where the ACPs fail to reduce the MAEs for the corresponding validation subset. Nevertheless, ACPs work well for other peptide conformational systems in the validation set, with MAE reductions in the range of 33–63% for CAM-B3LYP-D3/6-31G\*-ACP in the case of conformational energies of other peptide systems (tripeptide, cyclic, and those that show bioactive functionality).

### **(iii) Reaction energies**

We turn our attention now to covalent properties involving bond breaking and formation. One of such properties included in the training set are chemical reaction energies (“RE”). The training set contains 8,315 RE data points. Application of ACPs to the RE subsets in the training set shows an overall reduction in the MAEs of the double- $\zeta$  basis set DFT methods by about 29–34%. This improvement in the MAE mainly reflects the MAE decrease observed in two specific RE subsets, MN-RE (7,555 data points) and BH9-RE (449 data points). These subsets contain data points representing a variety of chemical reactions. For instance, the BH9-RE subset contains reaction energies of the nine most common elementary reactions encountered in organic and bioorganic chemistry. Besides MN-RE and BH9-RE, the ACPs also perform generally well (MAE reductions between 11–87% for M062X/6-31G\*-ACP and CAM-B3LYP-D3/6-31G\*-ACP) for other RE subsets of various types of reactions with fewer data points, with only a few exceptions. The decrease in the MAEs for reaction energies is also observed for the relatively large RE validation data set (“RE-val”), composed of a subset of the W4-17-RE set with 5,205 data points. For this data set, which was not used in the ACP training, the MAEs decrease by about 31–50% on application of



ACPs compared to the uncorrected double- $\zeta$  basis set DFT methods. This indicates that the ACPs are successful and robust for reaction energies.

#### (iv) Barrier heights

Besides reaction energies, the training set also included barrier heights of chemical reactions (“BH”) in order to make the eventual ACP-corrected methods usable for kinetic studies. The total number of BH data points contributes nearly 50% of the training set data (58,197) and is the most dominant property overall. The application of ACPs reduces the overall MAE of the parent methods for barrier heights by about 27–50%. This improvement originates mostly from the good performance of ACPs (MAE reductions by 13–52%) on the four main BH data sets: Grambow2020-B97D3 (32,722 data points), Grambow2020- $\omega$ B97XD3 (23,922 data points), BH9 (898), and E2SN2 (418). Note that the BH9 data set was designed recently<sup>74</sup> to be used for the particular purpose of developing the ACPs in this work and contains various model reactions that increase the diversity of the BH data in the training set.

For the other BH data sets in the training set besides BH9, E2SN2, and the two Grambow2020 subsets, the performance of ACPs is also quite good. ACPs mostly bring down the MAEs for the other subsets by about 22–94%. Contrary to most other BH data points in the training set, the INV24 and BHROT27 subsets feature barriers for processes that do not involve bond breaking or formation. The application of ACPs on these two data sets do not show a significant reduction in MAE compared to the uncorrected methods.

To validate the performance of ACPs on barrier height prediction, we applied the uncorrected methods and their ACP-corrected counterparts to a data set with 88 data points (WaterOrgBH) not included in the training set. This data set contains to-be-published barrier height reference data for pericyclic-type reactions in the absence and presence of water clusters, calculated at the DLPNO-CCSD(T)/CBS level of theory. The MAEs of double- $\zeta$  basis set DFT methods after application of ACPs to the WaterOrgBH data set decrease from 14.46 to 9.21 kcal/mol (BLYP-D3/6-31G\*), 2.31 to 1.56 kcal/mol (M062X/6-31G\*), and 1.70 to 1.66 kcal/mol (CAM-B3LYP-D3/6-31G\*), indicating that ACPs are likely to perform well when they are applied to the calculation of barrier heights outside the training set.

As an additional test, we also applied the proposed ACPs to the systems in the work of Bistoni *et al.*<sup>75</sup>, involving the Baeyer–Villiger reaction catalyzed by the cyclohexanone monooxygenase enzyme. The geometries used for ACP testing are the relevant stationary points of the reaction (reactant, intermediate, transition state, and product), obtained independently with three active site models of

increasing size. The reference data for the relative energies along the reaction profile were obtained with DLPNO-CCSD(T0)/def2-TZVPP. Point charges were used to model the electrostatic potential from the surrounding protein environment. The errors in the calculated reaction energies and barrier heights with and without ACPs for the three differently sized active sites are shown in Table 3. The errors of various methods relative to DLPNO-CCSD(T0)/def2-TZVPP data in Table 3 demonstrates that the ACP-corrected methods yield lower errors in the predicted relative energies than the uncorrected methods for most stationary points along the reaction profile of different active site sizes, with only a few exceptions. Upon application of ACPs for the barrier height prediction of the smallest active site model (99 atoms), the errors in the barrier heights (energy of transition state relative to reactant) drop from -22.53 to -7.22 kcal/mol (BLYP-D3/6-31G\*), -4.74 to 1.67 kcal/mol (M062X/6-31G\*), and -5.14 to -1.09 kcal/mol (CAM-B3LYP-D3/6-31G\*).

**Table 3.** Reference data (calculated using DLPNO-CCSD(T0)/def2-TZVPP) and errors relative to the reference data yielded by various methods (uncorrected and ACP-corrected double- $\zeta$  basis set DFT methods) for the relative energies along the reaction profile of Baeyer–Villiger reaction catalyzed by the cyclohexanone monooxygenase enzyme.<sup>a,b,c</sup>

| Method                    | QM region with 99 atoms |                  |         | QM region with 206 atoms |                  |         | QM region with 307 atoms |                  |         |
|---------------------------|-------------------------|------------------|---------|--------------------------|------------------|---------|--------------------------|------------------|---------|
|                           | Intermediate            | Transition state | Product | Intermediate             | Transition state | Product | Intermediate             | Transition state | Product |
| DLPNO-CCSD(T0)/def2-TZVPP | -4.50                   | 8.10             | -68.60  | -2.00                    | 10.50            | -71.10  | -1.50                    | 9.90             | -71.90  |
| BLYP-D3/6-31G*            | -9.4                    | -22.53           | 0.25    | -2.98                    | -16.2            | 1.56    | -1.08                    | -15.95           | 0.43    |
| BLYP-D3/6-31G*-ACP        | -1.4                    | -7.22            | -1.94   | 0.94                     | -3.6             | -1.67   | 2.72                     | -3.41            | -2.22   |
| M062X/6-31G*              | -11.45                  | -4.74            | -5.64   | -8.5                     | 0.03             | -5.08   | -6.22                    | 1.06             | -5.81   |
| M062X/6-31G*-ACP          | -4.36                   | 1.67             | 1.94    | -2.96                    | 4.97             | 2.54    | -0.41                    | 5.82             | 2.42    |

|                         |       |       |       |       |      |       |       |     |       |
|-------------------------|-------|-------|-------|-------|------|-------|-------|-----|-------|
| CAM-B3LYP-D3/6-31G*     | -8.36 | -5.14 | -3.65 | -3.61 | 0.47 | -2.35 | -1.49 | 1.2 | -3.24 |
| CAM-B3LYP-D3/6-31G*-ACP | -2.71 | -1.09 | 0.92  | -0.63 | 2.67 | 1.73  | 1.85  | 3.3 | 1.56  |

<sup>a)</sup> all energies were calculated in the presence of point charges with kcal/mol units, <sup>b)</sup> the energies along the reaction profile were calculated relative to the reactant, <sup>c)</sup> the geometries and point charges were taken from the work of Bistoni *et al.*<sup>75</sup>

## (v) Bond separation energies and other properties

Lastly, we analyze ACP performance for bond separation energies (“BSE”). In our training set, we included 4,502 bond separation energies from our recently developed BSE49<sup>76</sup> data set. BSE49 contains the reaction energies associated with the formation of radical species upon homolytic cleavage of 49 unique single bonds with various functional group substitutions. The BSE49 set was designed so that the 49 bonds in question represent all single bonds between unique combinations of the ten atoms for which ACPs are being developed in this work. The lack of bond separation energy data in the literature was the main motivation behind the creation of the BSE49 data set. Therefore, further validation of ACPs on data outside the training set was not possible. Application of our ACPs leads to a reduction in the MAEs compared to the parent methods in BSE49 by about 26–43%, suggesting that ACPs offer an efficient way of modeling bond separation reactions.

Other chemical properties that were included in the ACP training set were: molecular isomerization energies (“ISOM”), total atomization energies (“TAE”), and molecular deformation energies (“DEF”). A few ISOM and TAE data sets from the literature were included in the training set, mainly to ensure that ACPs do not lead to a degradation in the functional performance for these properties. The results from Figure S1 in the SI show that the application of the ACPs reduce the MAEs of ISOM and TAE with CAM-B3LYP-D3/6-31G\* by about 33% and 28%, respectively. Similarly, with M06-2X/6-31G\* the MAEs are reduced by about 34% for ISOM and 8% for TAE. For BLYP-D3/6-31G\*, the MAE is only reduced for ISOM by 42% but an increase in observed for the MAE of TAE by only 10%.

In addition, a large set of DEF data points were also included in the training set. These data points represent energy differences between a molecule at its equilibrium geometry and the same molecule deformed along its various normal modes. Our intention when we included these data in the training set

was to improve the description of molecular potential energy surfaces to obtain reasonably accurate geometries and energy derivatives or, at least, prevent the ACP fitting procedure from deteriorating the performance of the uncorrected method for these properties. In this connection, we performed transition state searches and geometry optimizations on the stationary point structures (reactants, products, and transition states) of 100 representative chemical reactions taken from the BH9 data set. Application of ACPs resulted in geometries with average root-mean-square-deviations (RMSDs) ranging between 0.110–0.131 Å (Table S6 of SI) compared to the reference geometries. These average RMSDs are only slightly higher than the uncorrected double- $\zeta$  basis set DFT methods (0.068–0.123 Å). Note that the geometries yielded by DFT-D3 methods with a double- $\zeta$  basis set have been shown in the literature to be reasonably close to that obtained with quadruple- $\zeta$  basis sets.<sup>22</sup>

### **(vi) Overall ACP performance**

In this subsection, we briefly discuss the overall performance of ACPs. First, we consider the various biases of the base method as measured by the mean signed errors (MSE) and standard deviations (SD) shown in Figure 1. Regarding non-covalent interaction energies, the Figure shows that the uncorrected methods tend to over-estimate interaction energies leading to negative MSEs. Application of ACPs corrects for this bias and reduces both the MSE and the error spread (SD). For conformational energies, the MSEs and SDs of the base methods are reasonably low, and application of ACPs results in a very small change for these quantities. For conformational energies, ACPs bring down MSEs and SDs. Covalent properties such as reaction energies, barrier heights, and bond separation energies are mostly underestimated (positive MSEs) with a substantial error spread. In the particular case of barrier heights and the BLYP functional, this may be attributed to delocalization error<sup>74</sup>, which is partially remedied by the application of ACPs. For the covalent properties, application of the ACPs also decreases both the MSE and SDs substantially.

Lastly, we compare the performance of uncorrected and ACP-corrected double- $\zeta$  basis set DFT methods with the same DFT methods using large basis sets. We present the MAEs of the various methods in Table 4. Only a few representative data sets for which nearly complete basis set DFT results have been reported in the literature are shown in the table. For non-covalent properties (interaction and conformational energies), the MAEs of ACP-corrected methods are slightly higher than the same methods at complete basis set limit. However, a particular exception is seen in the case of large hydrogen-bonded water cluster sets, Water38 and H2O20Bind10 with CAM-B3LYP-D3/6-31G\*-ACP where the MAEs are lower than CAM-B3LYP-D3/def2-QZVPDD by almost a factor of two. In the case of covalent properties

(reaction energies and barrier heights), the MAEs of M062X/6-31G\*-ACP and CAM-B3LYP-D3/6-31G\*-ACP methods are lower than the same functionals with def2-QZVPDD basis set for a few cases like BH9-RE, BSR36, and PX13. The MAEs with CAM-B3LYP-D3/6-31G\*-ACP is also lower than CAM-B3LYP-D3/def2-QZVPDD for two other data sets, namely HTBH38 and WCPT27. Another interesting result is observed for the ACP-corrected BLYP-D3/6-31G\* method, where the MAEs of data sets containing reaction energies and barrier heights, with only a few exceptions, are generally lower than BLYP-D3/def2-QZVPDD MAEs, indicating that ACPs perform well in mitigating errors other than basis set incompleteness like those arising from the approximations in the density functionals itself. The overall comparison shown in Table 4 demonstrates that ACP-corrected methods have a performance close to nearly complete basis set DFT or sometimes even better, but at a ca. one order of magnitude lower computational time.

**Table 4.** Comparison of the mean absolute errors (MAEs) of various 6-31G\*, 6-31G\*-ACP, and nearly complete basis set DFT methods for selected data sets.

| Data set <sup>a</sup>                     | BLYP-D3/6-31G* | BLYP-D3/6-31G*-ACP | BLYP-D3/def2-QZVPDD <sup>b</sup> | M062X/6-31G* | M062X/6-31G*-ACP | M062X/def2-QZVPDD <sup>b</sup> | CAM-B3LYP-D3/6-31G* | CAM-B3LYP-D3/6-31G*-ACP | CAM-B3LYP-D3/def2-QZVPDD <sup>b</sup> |
|-------------------------------------------|----------------|--------------------|----------------------------------|--------------|------------------|--------------------------------|---------------------|-------------------------|---------------------------------------|
| <i>Non-covalent interaction energies:</i> |                |                    |                                  |              |                  |                                |                     |                         |                                       |
| S66x8 <sup>86-88</sup> (528)              | 1.67           | 0.60               | 0.15                             | 0.94         | 0.52             | 0.30                           | 1.56                | 0.52                    | 0.29                                  |
| Sulfurx8 <sup>98</sup> (104)              | 1.15           | 0.33               | 0.19                             | 0.60         | 0.34             | 0.20                           | 0.99                | 0.36                    | 0.16                                  |
| 3B-69-DIM <sup>96</sup> (207)             | 1.65           | 0.61               | 0.24                             | 0.76         | 0.48             | 0.43                           | 1.68                | 0.35                    | 0.31                                  |
| BzDC215 <sup>102</sup> (170)              | 0.94           | 0.44               | 0.14                             | 0.63         | 0.22             | 0.22                           | 0.82                | 0.33                    | 0.15                                  |
| Water38 <sup>124</sup> (38)               | 36.41          | 1.39               | 1.30                             | 26.22        | 4.81             | 1.62                           | 36.38               | 3.39                    | 5.87                                  |
| CE20 <sup>189,190</sup> (20)              | 21.46          | 2.86               | 1.02                             | 15.18        | 3.30             | 1.16                           | 21.40               | 3.81                    | 3.27                                  |
| H2O20Bind10 <sup>200</sup> (10)           | 145.88         | 24.40              | 4.50                             | 107.24       | 15.19            | 3.35                           | 154.83              | 13.15                   | 22.12                                 |
| <i>Molecular conformational energies:</i> |                |                    |                                  |              |                  |                                |                     |                         |                                       |
| YMPJ <sup>143</sup> (495)                 | 1.08           | 0.85               | 0.53                             | 0.72         | 0.62             | 0.38                           | 1.06                | 0.52                    | 0.34                                  |
| BCONF <sup>152</sup> (64)                 | 2.55           | 0.85               | 0.34                             | 1.75         | 0.45             | 0.12                           | 2.44                | 0.39                    | 0.34                                  |
| SCONF <sup>14,147</sup> (17)              | 4.03           | 0.98               | 0.41                             | 2.47         | 0.63             | 0.18                           | 3.29                | 0.75                    | 0.19                                  |
| PentCONF <sup>153</sup> (342)             | 0.30           | 0.21               | 0.36                             | 0.13         | 0.17             | 0.11                           | 0.06                | 0.42                    | 0.08                                  |
| <i>Reaction energies:</i>                 |                |                    |                                  |              |                  |                                |                     |                         |                                       |
| BH9-RE <sup>74</sup> (449)                | 5.39           | 3.12               | 7.15 <sup>e</sup>                | 3.40         | 2.02             | 2.76 <sup>d</sup>              | 3.49                | 2.43                    | 3.14 <sup>e</sup>                     |
| BH76RC <sup>162-164</sup> (30)            | 8.72           | 4.02               | 3.28                             | 7.35         | 3.98             | 0.86                           | 8.44                | 4.43                    | 1.61                                  |
| NBPRC <sup>161,164,181</sup> (6)          | 2.81           | 1.77               | 3.08                             | 2.13         | 3.34             | 1.10                           | 1.73                | 1.84                    | 2.05                                  |
| BSR36 <sup>160,161</sup> (36)             | 2.74           | 2.61               | 3.33                             | 3.62         | 0.77             | 3.51                           | 4.81                | 1.03                    | 4.26                                  |
| WCPT6 <sup>180</sup> (6)                  | 3.71           | 1.37               | 1.04                             | 2.74         | 2.42             | 0.80                           | 3.42                | 2.22                    | 0.89                                  |
| CR20 <sup>166</sup> (20)                  | 2.46           | 2.70               | 9.63                             | 3.92         | 2.61             | 1.75                           | 4.52                | 3.34                    | 2.52                                  |
| DIE60 <sup>157</sup> (60)                 | 2.39           | 1.12               | 1.43                             | 1.16         | 0.67             | 0.57                           | 1.29                | 1.05                    | 0.57                                  |
| <i>Barrier heights:</i>                   |                |                    |                                  |              |                  |                                |                     |                         |                                       |
| BH9 <sup>74</sup> (898)                   | 12.72          | 7.40               | 8.66 <sup>e</sup>                | 3.43         | 2.98             | 2.27 <sup>d</sup>              | 4.16                | 2.88                    | 2.37 <sup>e</sup>                     |

|                                  |       |      |       |       |      |      |       |      |      |
|----------------------------------|-------|------|-------|-------|------|------|-------|------|------|
| BHPERI26 <sup>164,184</sup> (26) | 7.24  | 4.47 | 3.58  | 7.24  | 4.47 | 1.35 | 2.25  | 2.60 | 2.37 |
| CRBH20 <sup>188</sup> (20)       | 13.52 | 1.34 | 16.56 | 13.52 | 1.34 | 1.32 | 2.31  | 1.61 | 1.18 |
| DBH24 <sup>185,186</sup> (24)    | 10.99 | 5.23 | 8.25  | 10.99 | 5.23 | 0.85 | 6.00  | 4.45 | 2.66 |
| HTBH38 <sup>163</sup> (38)       | 10.01 | 5.09 | 8.79  | 10.01 | 5.09 | 1.08 | 4.53  | 2.58 | 3.61 |
| NHTBH38 <sup>162</sup> (38)      | 13.59 | 5.54 | 8.95  | 13.59 | 5.54 | 1.29 | 8.14  | 5.37 | 2.71 |
| PX13 <sup>189,190</sup> (13)     | 33.76 | 2.17 | 8.88  | 33.76 | 2.17 | 6.11 | 29.49 | 3.59 | 8.06 |
| WCPT27 <sup>180</sup> (27)       | 12.81 | 3.18 | 6.53  | 12.81 | 3.18 | 2.92 | 9.66  | 2.49 | 3.43 |

<sup>a)</sup> details about the data sets can be found in Table S1 of the Supporting Information, <sup>b)</sup> from Reference 13, <sup>c)</sup> represents the value obtained with BLYP-XDM/def2-QZVPP in Reference 74, <sup>d)</sup> represents the value obtained with M062X/def2-QZVPP in Reference 74, <sup>e)</sup> represents the value obtained with CAM-B3LYP-XDM/def2-QZVPP in Reference 74.

## 4. Conclusions

The use of very accurate quantum mechanical methods, such as nearly complete basis set wavefunction theory methods, is not practical for applications involving large systems because of the steep scaling of their computational cost with system size. A low-cost alternative is the use of small basis set density-functional theory (DFT) methods, but this requires that the inherent shortcomings in these methods be mitigated. The focus of this article is to improve the performance of small basis set DFT methods without sacrificing their computational efficiency.

We developed and applied atom-centered potentials (ACPs) to mitigate the shortcomings of BLYP-D3, M062X, and CAM-B3LYP-D3 methods, in combination with the 6-31G\* basis set. We expect these shortcomings to be primarily the error from the choice of density functional approximation and basis set incompleteness from the limited size of the basis set.

The ACPs presented in this work were developed for ten elements (H, B, C, N, O, F, Si, P, S, Cl). The parametrization of the ACPs was carried out using a regularized linear least-squares fitting procedure (the LASSO regression method) using a training set of 118,655 data points calculated mostly using wavefunction theory methods extrapolated to the complete basis set limit. The main molecular properties in the training set were non-covalent interaction energies, molecular conformational energies, reaction energies, barrier heights, and bond separation energies. The performance of the proposed ACPs was assessed using the training set and an additional validation set containing 42,567 data points not used during the ACP training.

Our assessment of the new ACP-corrected methods suggests that the ACPs reduce the mean absolute errors (MAEs) of double- $\zeta$  basis set DFT methods for most subsets, and in general, lead to an improved description of all molecular properties in the training set, indicating that ACPs successfully mitigate the

deficiencies of the parent double- $\zeta$  basis set DFT methods. The best performing ACP-based method, i.e., CAM-B3LYP-D3/6-31G\*-ACP, yields mean absolute errors, relative to high-level of theory, of around 0.7 kcal/mol for non-covalent interaction energies, 0.4 kcal/mol for conformational energies, 5.6 kcal/mol for reaction energies, 3.0 kcal/mol for barrier heights, and 2.5 kcal/mol for bond separation energies.

Further analysis of the performance of ACP-corrected methods on the validation set shows that ACPs are relatively robust, i.e., they are suitable for applications in systems outside the training set. However, there is a performance penalty when the system under consideration is significantly different from our training set, although no catastrophic results were obtained at any point. This observation suggests that increasing the size of the training set and incorporating more diversity is a straightforward way to improve the ACPs. In addition, using basis sets containing diffuse functions could be a way to overcome the limitations of the present methods in the description of negatively charged systems. The proposed new ACPs correct small basis set DFT methods and improve the accuracy of their parent methods, and allow carrying out quantum mechanical calculations of large systems at a reasonably low computational cost.

## Supporting information

The supporting Information (SI) is available free of charge on the ACS publications website at DOI: <http://dx.doi.org/xx.xxxx/acs.jctc.xxxxxxxx>.

Sample Gaussian16 input file demonstrating the use of ACPs, formulas for the error measures, tables listing the details of data sets in the ACP training and validation set, figures and tables associated with the detailed error analysis of training and validation sets. (PDF)

Basis set and ACP files, detailed results, and Cartesian coordinates of all the data sets used in the study. (ZIP)

## Acknowledgements

We thank the Natural Sciences and Engineering Research Council of Canada, Canadian Foundation for Innovation, B.C. Knowledge Development Fund, and the University of British Columbia for financial support. This research was enabled by the computational resource support provided by WestGrid, Compute Canada, and the University of British Columbia's advanced research computing services. AOR thanks the Spanish Ministerio de Ciencia e Innovación and the Agencia Estatal de Investigación (AEI) for financial support (projects PGC2018-097520-A-100 and RED2018-102612-T), and the Spanish MINECO for a Ramón y Cajal fellowship (RyC-2016-20301). VKP would like to thank Eduardo Romero-Montalvo for

providing the unpublished data of WaterOrgBH set. We thank Giovanni Bistoni for kindly providing the geometries and point charges used for the results in Table 3.

## Conflicts of interest

Authors declare no conflicts of interest.

## References

- (1) Burke, K. Perspective on Density Functional Theory. *J. Chem. Phys.* **2012**, *136* (15), 150901.
- (2) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112* (1), 289–320.
- (3) DiLabio, G. A.; Otero-de-la-Roza, A. Noncovalent Interactions in Density Functional Theory; In *Reviews in Computational Chemistry*; John Wiley & Sons Inc., 2016; pp 1–97.
- (4) Riley, K. E.; Hobza, P. Noncovalent Interactions in Biochemistry. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 3–17.
- (5) Goerigk, L. A Comprehensive Overview of the DFT-D3 London-Dispersion Correction. In *Non-Covalent Interactions in Quantum Chemistry and Physics: Theory and Applications*; Elsevier Inc., **2017**, pp 195–219.
- (6) Johnson, E. R. The Exchange-Hole Dipole Moment Dispersion Model. In *Non-Covalent Interactions in Quantum Chemistry and Physics: Theory and Applications*; Elsevier Inc., 2017; pp 169–194.
- (7) Hermann, J.; DiStasio Jr., R. A.; Tkatchenko, A. First-Principles Models for van Der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chem. Rev.* **2017**, *117* (6), 4714–4758.
- (8) Stöhr, M.; Van Voorhis, T.; Tkatchenko, A. Theory and Practice of Modeling van Der Waals Interactions in Electronic-Structure Calculations. *Chem. Soc. Rev.* **2019**, *48*, 4118–4154.
- (9) Ambrosetti, A.; Alfè, D.; DiStasio Jr., R. A.; Tkatchenko, A. Hard Numbers for Large Molecules: Toward Exact Energetics for Supramolecular Systems. **2014**, *5* (5), 849–855.
- (10) Otero-de-la-Roza, A.; Johnson, E. R. Predicting Energetics of Supramolecular Systems Using the XDM Dispersion Model. *J. Chem. Theory Comput.* **2015**, *11* (9), 4033–4040.
- (11) Risthaus, T.; Grimme, S. Benchmarking of London Dispersion-Accounting Density Functional Theory Methods on Very Large Molecular Complexes. *J. Chem. Theory Comput.* **2013**, *9* (3), 1580–1591.
- (12) Antony, J.; Sure, R.; Grimme, S. Using Dispersion-Corrected Density Functional Theory to Understand Supramolecular Binding Thermodynamics. *Chem. Commun.* **2015**, *51* (10), 1764–1774.
- (13) Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Mol. Phys.* **2017**, *115* (19), 2315–2372.
- (14) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A Look at the Density Functional Theory Zoo with the Advanced GMTKN55 Database for General Main Group Thermochemistry, Kinetics and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2017**, *19* (48), 32184–32215.
- (15) Cui, Q. Perspective: Quantum Mechanical Methods in Biochemistry and Biophysics. *J. Chem. Phys.* **2016**, *145* (14), 140901.
- (16) Ratcliff, L. E.; Mohr, S.; Huhs, G.; Deutsch, T.; Masella, M.; Genovese, L. Challenges in Large Scale Quantum Mechanical Calculations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7* (1), e1290.
- (17) Brandenburg, J. G.; Burke, K.; Civalieri, B.; Cole, D. J.; Csányi, G.; David, G.; Gidopoulos, N. I.; Gowland, D.; Helgaker, T.; Herbst, M. F.; Hourahine, B.; Irons, T. J. P.; Jacob, C. R.; Loos, P.-F.; Mehta, N.; Mulay, M. R.; Neugebauer, J.; Pernal, K.; Pribram-Jones, A.; Romaniello, P.; Ryder, M. R.; Savin, A.; Sirbu, D.; Skylaris, C.-K.; Truhlar, D. G.; Wetherell, J.; Yang, W. Challenges for Large Scale Simulation: General Discussion. *Faraday Discuss.* **2020**, *224* (0), 309–332.
- (18) Sherrill, C. D. Frontiers in Electronic Structure Theory. *J. Chem. Phys.* **2010**, *132* (11), 110902.
- (19) Hofer, T. S. From Macromolecules to Electrons—Grand Challenges in Theoretical and Computational Chemistry. *Front.*



*Chem.* **2013**, *1*, 6.

- (20) Grimme, S.; Schreiner, P. R. Computational Chemistry: The Fate of Current Methods and Future Challenges. *Angew. Chemie Int. Ed.* **2018**, *57* (16), 4170–4176.
- (21) Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017**, *50* (3), 539–543.
- (22) Hostaš, J.; Řezáč, J. Accurate DFT-D3 Calculations in a Small Basis Set. *J. Chem. Theory Comput.* **2017**, *13* (8), 3575–3585.
- (23) 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* (2), 94–109.
- (24) Kruse, H.; Goerigk, L.; Grimme, S. Why the Standard B3LYP/6-31G\* Model Chemistry Should Not Be Used in DFT Calculations of Molecular Thermochemistry: Understanding and Correcting the Problem. *J. Org. Chem.* **2012**, *77* (23), 10824–10834.
- (25) 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* (15), 154104.
- (26) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465.
- (27) 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* (15), 154101.
- (28) Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J. Comput. Chem.* **2013**, *34* (19), 1672–1685.
- (29) Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent Structures and Interactions by Density Functional Theory with Small Atomic Orbital Basis Sets. *J. Chem. Phys.* **2015**, *143* (5), 054107.
- (30) Brandenburg, J. G.; Caldeweyher, E.; Grimme, S. Screened Exchange Hybrid Density Functional for Accurate and Efficient Structures and Interaction Energies. *Phys. Chem. Chem. Phys.* **2016**, *18* (23), 15519–15523.
- (31) Pracht, P.; Grant, D. F.; Grimme, S. Comprehensive Assessment of GFN Tight-Binding and Composite Density Functional Theory Methods for Calculating Gas-Phase Infrared Spectra. *J. Chem. Theory Comput.* **2020**, *16* (11), 7044–7060.
- (32) Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S. B97-3c: A Revised Low-Cost Variant of the B97-D Density Functional Method. *J. Chem. Phys.* **2018**, *148* (6), 064104.
- (33) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r2SCAN-3c: A “Swiss Army Knife” Composite Electronic-Structure Method. *J. Chem. Phys.* **2021**, *154* (6), 064103.
- (34) Kulik, H. J.; Seelam, N.; Mar, B. D.; Martínez, T. J. Adapting DFT+ *U* for the Chemically Motivated Correction of Minimal Basis Set Incompleteness. *J. Phys. Chem. A* **2016**, *120* (29), 5939–5949.
- (35) Schneebeli, S. T.; Bochevarov, A. D.; Friesner, R. A. Parameterization of a B3LYP Specific Correction for Noncovalent Interactions and Basis Set Superposition Error on a Gigantic Data Set of CCSD(T) Quality Noncovalent Interaction Energies. *J. Chem. Theory Comput.* **2011**, *7* (3), 658–668.
- (36) García, J. S.; Brémond, É.; Campetella, M.; Ciofini, I.; Adamo, C. Small Basis Set Allowing the Recovery of Dispersion Interactions with Double-Hybrid Functionals. *J. Chem. Theory Comput.* **2019**, *15* (5), 2944–2953.
- (37) Li, W.; Miao, W.; Cui, J.; Fang, C.; Su, S.; Li, H.; Hu, L.; Lu, Y.; Chen, G. Efficient Corrections for DFT Noncovalent Interactions Based on Ensemble Learning Models. *J. Chem. Inf. Model.* **2019**, *59* (5), 1849–1857.
- (38) Mehta, N.; Goerigk, L.; Mehta, N.; Goerigk, L. Assessing the Applicability of the Geometric Counterpoise Correction in B2PLYP/Double- $\zeta$  Calculations for Thermochemistry, Kinetics, and Noncovalent Interactions. *Aust. J. Chem.* **2021**. <https://doi.org/10.1071/CH21133>
- (39) Tirri, B.; Ciofini, I.; Sancho-García, J. C.; Adamo, C.; Brémond, É. Computation of Covalent and Noncovalent Structural Parameters at Low Computational Cost: Efficiency of the DH-SVPD Method. *Int. J. Quantum Chem.* **2020**, *120* (13), e26233.
- (40) 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* (23), 234105.
- (41) DiLabio, G. A. Atom-centered Potentials for Noncovalent Interactions and Other Applications. In *Non-Covalent Interactions in Quantum Chemistry and Physics: Theory and Applications*; Elsevier Inc., **2017**, pp 221–240.
- (42) Prasad, V. K.; Otero-de-la-Roza, A.; DiLabio, G. A. Atom-Centered Potentials with Dispersion-Corrected Minimal-Basis-Set Hartree-Fock: An Efficient and Accurate Computational Approach for Large Molecular Systems. *J. Chem.*

*Theory Comput.* **2018**, *14* (2), 726–738.

- (43) Otero-de-la-Roza, A.; DiLabio, G. A. Improved Basis-Set Incompleteness Potentials for Accurate Density-Functional Theory Calculations in Large Systems. *J. Chem. Theory Comput.* **2020**, *16* (7), 4176–4191.
- (44) 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* (8), 3505–3524.
- (45) 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* (25), 6703–6713.
- (46) 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* (18), 18A542.
- (47) DiLabio, G. A.; Koleini, M.; Torres, E. Extension of the B3LYP–Dispersion-Correcting Potential Approach to the Accurate Treatment of Both Inter- and Intra-Molecular Interactions. *Theor. Chem. Acc.* **2013**, *132* (10), 1389.
- (48) Torres, E.; DiLabio, G. A. A (Nearly) Universally Applicable Method for Modeling Noncovalent Interactions Using B3LYP. *J. Phys. Chem. Lett.* **2012**, *3* (13), 1738–1744.
- (49) 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* (43), 10968–10976.
- (50) 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* (4–6), 348–353.
- (51) Mackie, I. D.; DiLabio, G. A. Accurate Dispersion Interactions from Standard Density-Functional Theory Methods with Small Basis Sets. *Phys. Chem. Chem. Phys.* **2010**, *12* (23), 6092.
- (52) Torres, E.; DiLabio, G. A. Density-Functional Theory with Dispersion-Correcting Potentials for Methane: Bridging the Efficiency and Accuracy Gap between High-Level Wave Function and Classical Molecular Mechanics Methods. *J. Chem. Theory Comput.* **2013**, *9* (8), 3342–3349.
- (53) Mackie, I. D.; DiLabio, G. A. CO<sub>2</sub> Adsorption by Nitrogen-Doped Carbon Nanotubes Predicted by Density-Functional Theory with Dispersion-Correcting Potentials. *Phys. Chem. Chem. Phys.* **2011**, *13* (7), 2780–2787.
- (54) 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* (9), 4205–4215.
- (55) Prasad, V. K.; Otero-de-la-Roza, A.; DiLabio, G. A. Performance of Small Basis Set Hartree–Fock Methods for Modeling Non-Covalent Interactions. *Electron. Struct.* **2021**, *3* (3), 034007.
- (56) Cao, X.; Dolg, M. Pseudopotentials and Modelpotentials. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (2), 200–210.
- (57) Dolg, M.; Cao, X. Relativistic Pseudopotentials: Their Development and Scope of Applications. *Chem. Rev.* **2012**, *112* (1), 403–480.
- (58) Prasad, V. K.; Otero-de-la-Roza, A.; DiLabio, G. A. Fast and Accurate Quantum Mechanical Modeling of Large Molecular Systems Using Small Basis Set Hartree–Fock Corrected with Atom-Centered Potentials. **2021**, in preparation.
- (59) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent Molecular Orbital Methods. XXIII. A Polarization-type Basis Set for Second-row Elements. *J. Chem. Phys.* **1982**, *77* (7), 3654–3665.
- (60) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta* **1973**, *28* (3), 213–222.
- (61) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100.
- (62) 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* (2), 785–789.
- (63) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120* (1–3), 215–241.
- (64) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange–Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393* (1–3), 51–57.
- (65) Density Functional Theory Poll. <https://www.marcelswart.eu/dft-poll/> (accessed 2021-10-31)

- (66) Goerigk, L.; Mehta, N. A Trip to the Density Functional Theory Zoo: Warnings and Recommendations for the User. *Aust. J. Chem.* **2019**, *72* (8), 563–573.
- (67) Witte, J.; Neaton, J. B.; Head-Gordon, M. Push It to the Limit: Characterizing the Convergence of Common Sequences of Basis Sets for Intermolecular Interactions as Described by Density Functional Theory. *J. Chem. Phys.* **2016**, *144* (19), 194306.
- (68) Tibshirani, R. Regression Shrinkage and Selection via the Lasso: A Retrospective. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2011**, *73* (3), 273–282.
- (69) Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58* (1), 267–288.
- (70) Osborne, M. R.; Presnell, B.; Turlach, B. A. On the LASSO and Its Dual. *J. Comput. Graph. Stat.* **2000**, *9* (2), 319–337.
- (71) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, 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 16, Revision B.01; Gaussian Inc.: Wallingford, CT, 2016.
- (72) Johnson, E. R.; Otero-de-la-Roza, A.; Dale, S. G.; DiLabio, G. A. Efficient Basis Sets for Non-Covalent Interactions in XDM-Corrected Density-Functional Theory. *J. Chem. Phys.* **2013**, *139* (21), 214109.
- (73) Papajak, E.; Truhlar, D. G. Efficient Diffuse Basis Sets for Density Functional Theory. *J. Chem. Theory Comput.* **2010**, *6* (3), 597–601.
- (74) Prasad, V. K.; Pei, Z.; Edelmann, S.; Otero-de-la-Roza, A.; DiLabio, G. BH9, a New Comprehensive Benchmark Dataset for Barrier Heights and Reaction Energies: Assessment of Density Functional Approximations and Basis Set Incompleteness Potentials. *J. Chem. Theory Comput.* **2021**. In press. <https://doi.org/10.1021/acs.jctc.1c00694>
- (75) Bistoni, G.; Polyak, I.; Sparta, M.; Thiel, W.; Neese, F. Toward Accurate QM/MM Reaction Barriers with Large QM Regions Using Domain Based Pair Natural Orbital Coupled Cluster Theory. *J. Chem. Theory Comput.* **2018**, *14* (7), 3524–3531.
- (76) Prasad, V. K.; Khalilian, M. H.; Otero-de-la-Roza, A.; DiLabio, G. A. BSE49, a Diverse, High-Quality Benchmark Dataset of Separation Energies of Chemical Bonds. *Sci. Data* **2021**, *8*, 300.
- (77) 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* (1), 88–96.
- (78) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis Set Convergence of the Coupled-Cluster Correction, P2CCSD(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* (19), 194102.
- (79) 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* (7), 1749–1760.
- (80) Miriyala, V. M.; Řezáč, J. Description of Non-Covalent Interactions in SCC-DFTB Methods. *J. Comput. Chem.* **2017**, *38* (10), 688–697.
- (81) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8* (1), 141–151.
- (82) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.* **2020**, *16* (4), 2355–2368.
- (83) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space. *J. Chem. Theory Comput.* **2020**, *16* (10), 6305–6316.
- (84) 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* (17), 1985–1993.
- (85) 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* (8), 2365–2376.

- (86) 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* (8), 2427–2438.
- (87) 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* (11), 3466–3470.
- (88) Dilabio, G. A.; Johnson, E. R.; Otero-de-la-Roza, A. Performance of Conventional and Dispersion-Corrected Density-Functional Theory Methods for Hydrogen Bonding Interaction Energies. *Phys. Chem. Chem. Phys.* **2013**, *15* (31), 12821–12828.
- (89) Řezáč, J.; Dubecký, M.; Jurečka, P.; Hobza, P. Extensions and Applications of the A24 Data Set of Accurate Interaction Energies. *Phys. Chem. Chem. Phys.* **2015**, *17* (29), 19268–19277.
- (90) Witte, J.; Goldey, M.; Neaton, J. B.; Head-Gordon, M. Beyond Energies: Geometries of Nonbonded Molecular Complexes as Metrics for Assessing Electronic Structure Approaches. *J. Chem. Theory Comput.* **2015**, *11* (4), 1481–1492.
- (91) 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* (5), 2151–2155.
- (92) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H<sub>2</sub>S. *J. Phys. Chem. A* **2009**, *113* (38), 10146–10159.
- (93) Hohenstein, E. G.; Sherrill, C. D. Effects of Heteroatoms on Aromatic  $\pi$ - $\pi$  Interactions: Benzene-Pyridine and Pyridine Dimer. *J. Phys. Chem. A* **2009**, *113* (5), 878–886.
- (94) Takatani, T.; Sherrill, C. D. Performance of Spin-Component-Scaled Møller-Plesset Theory (SCS-MP2) for Potential Energy Curves of Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2007**, *9* (46), 6106–6114.
- (95) 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* (12), 2197–2203.
- (96) Řezáč, J.; Huang, Y.; Hobza, P.; Beran, G. J. O. Benchmark Calculations of Three-Body Intermolecular Interactions and the Performance of Low-Cost Electronic Structure Methods. *J. Chem. Theory Comput.* **2015**, *11* (7), 3065–3079.
- (97) Copeland, K. L.; Tschumper, G. S. Hydrocarbon/Water Interactions: Encouraging Energetics and Structures from Dft but Disconcerting Discrepancies for Hessian Indices. *J. Chem. Theory Comput.* **2012**, *8* (5), 1646–1656.
- (98) Mintz, B. J.; Parks, J. M. Benchmark Interaction Energies for Biologically Relevant Noncovalent Complexes Containing Divalent Sulfur. *J. Phys. Chem. A* **2012**, *116* (3), 1086–1092.
- (99) Sanders, J. M. Optimal  $\pi$ -Stacking Interaction Energies in Parallel-Displaced Aryl/Aryl Dimers Are Predicted by the Dimer Heavy Atom Count. *J. Phys. Chem. A* **2010**, *114* (34), 9205–9211.
- (100) Parrish, R. M.; Sherrill, C. D. Quantum-Mechanical Evaluation of  $\pi$ - $\pi$  Versus Substituent- $\pi$  Interactions in  $\pi$  Stacking: Direct Evidence for the Wheeler-Houk Picture. *J. Am. Chem. Soc.* **2014**, *136* (50), 17386–17389.
- (101) Steinmann, S. N.; Corminboeuf, C. Exploring the Limits of Density Functional Approximations for Interaction Energies of Molecular Precursors to Organic Electronics. *J. Chem. Theory Comput.* **2012**, *8* (11), 4305–4316.
- (102) Crittenden, D. L. A Systematic CCSD(T) Study of Long-Range and Noncovalent Interactions between Benzene and a Series of First- and Second-Row Hydrides and Rare Gas Atoms. *J. Phys. Chem. A* **2009**, *113* (8), 1663–1669.
- (103) Hill, J. G.; Legon, A. C. On the Directionality and Non-Linearity of Halogen and Hydrogen Bonds. *Phys. Chem. Chem. Phys.* **2015**, *17* (2), 858–867.
- (104) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8* (11), 4285–4292.
- (105) Setiawan, D.; Kraka, E.; Cremer, D. Strength of the Pnictogen Bond in Complexes Involving Group VA Elements N, P, and AS. *J. Phys. Chem. A* **2015**, *119* (9), 1642–1656.
- (106) 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* (11), 114304.
- (107) 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* (7), 2282–2292.
- (108) 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 (BFDdb): An Open-Data Platform for Computational Chemistry Analysis of Noncovalent Interactions. *J. Chem. Phys.* **2017**, *147* (16), 161727.
- (109) Jakubec, D.; Hostaš, J.; Laskowski, R. A.; Hobza, P.; Vondrášek, J. Large-Scale Quantitative Assessment of Binding Preferences in Protein-Nucleic Acid Complexes. *J. Chem. Theory Comput.* **2015**, *11* (4), 1939–1948.
- (110) Hostaš, J.; Jakubec, D.; Laskowski, R. A.; Gnanasekaran, R.; Řezáč, J.; Vondrášek, J.; Hobza, P. Representative Amino Acid Side-Chain Interactions in Protein-DNA Complexes: A Comparison of Highly Accurate Correlated Ab Initio Quantum Mechanical Calculations and Efficient Approaches for Applications to Large Systems. *J. Chem. Theory Comput.* **2015**, *11* (9), 4086–4092.
- (111) Jakubec, D.; Laskowski, R. A.; Vondrasek, J. Sequence-Specific Recognition of DNA by Proteins: Binding Motifs Discovered Using a Novel Statistical/Computational Analysis. *PLoS One* **2016**, *11* (7), e0158704.
- (112) Stasyuk, O. A.; Jakubec, D.; Vondrášek, J.; Hobza, P. Noncovalent Interactions in Specific Recognition Motifs of Protein-DNA Complexes. *J. Chem. Theory Comput.* **2017**, *13* (2), 877–885.
- (113) Kozmon, S.; Matuška, R.; Spiwok, V.; Koča, J. Three-Dimensional Potential Energy Surface of Selected Carbohydrates' CH/ $\pi$  Dispersion Interactions Calculated by High-Level Quantum Mechanical Methods. *Chem. - A Eur. J.* **2011**, *17* (20), 5680–5690.
- (114) Kozmon, S.; Matuška, R.; Spiwok, V.; Koča, J. Dispersion Interactions of Carbohydrates with Condensate Aromatic Moieties: Theoretical Study on the CH- $\pi$  Interaction Additive Properties. *Phys. Chem. Chem. Phys.* **2011**, *13* (31), 14215–14222.
- (115) Stanković, I. M.; Blagojević Filipović, J. P.; Zarić, S. D. Carbohydrate – Protein Aromatic Ring Interactions beyond CH/ $\pi$  Interactions: A Protein Data Bank Survey and Quantum Chemical Calculations. *Int. J. Biol. Macromol.* **2020**, *157*, 1–9.
- (116) Kumari, M.; Sunoj, R. B.; Balaji, P. V. Conformational Mapping and Energetics of Saccharide-Aromatic Residue Interactions: Implications for the Discrimination of Anomers and Epimers and in Protein Engineering. *Org. Biomol. Chem.* **2012**, *10* (21), 4186–4200.
- (117) Černý, J.; Schneider, B.; Biedermannová, L. WatAA: Atlas of Protein Hydration. Exploring Synergies between Data Mining and: Ab Initio Calculations. *Phys. Chem. Chem. Phys.* **2017**, *19* (26), 17094–17102.
- (118) Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; Sherrill, C. D.; Merz, K. M. Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein-Ligand Complexes. *J. Chem. Theory Comput.* **2011**, *7* (3), 790–797.
- (119) Kříž, K.; Řezáč, J. Benchmarking of Semiempirical Quantum-Mechanical Methods on Systems Relevant to Computer-Aided Drug Design. *J. Chem. Inf. Model.* **2020**, *60* (3), 1453–1460.
- (120) Kruse, H.; Banáš, P.; Šponer, J. Investigations of Stacked DNA Base-Pair Steps: Highly Accurate Stacking Interaction Energies, Energy Decomposition, and Many-Body Stacking Effects. *J. Chem. Theory Comput.* **2019**, *15* (1), 95–115.
- (121) 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* (9), 4197–4204.
- (122) Banáš, P.; Mládek, A.; Otyepka, M.; Zgarbová, M.; Jurečka, P.; Svozil, D.; Lankaš, F.; Šponer, J. Can We Accurately Describe the Structure of Adenine Tracts in B-DNA? Reference Quantum-Chemical Computations Reveal Overstabilization of Stacking by Molecular Mechanics. *J. Chem. Theory Comput.* **2012**, *8* (7), 2448–2460.
- (123) Kabeláč, M.; Valdes, H.; Sherer, E. C.; Cramer, C. J.; Hobza, P. Benchmark RI-MP2 Database of Nucleic Acid Base Trimers: Performance of Different Density Functional Models for Prediction of Structures and Binding Energies. *Phys. Chem. Chem. Phys.* **2007**, *9* (36), 5000–5008.
- (124) 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* (43), 12034–12046.
- (125) Mas, E. M.; Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; Wormer, P. E. S.; Van Der Avoird, A. Water Pair Potential of near Spectroscopic Accuracy. I. Analysis of Potential Surface and Virial Coefficients. *J. Chem. Phys.* **2000**, *113* (16), 6687–6701.
- (126) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; Van Der Avoird, A. Predictions of the Properties of Water from First Principles. *Science* **2007**, *315* (5816), 1249–1252.
- (127) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; Van Der Avoird, A. Polarizable Interaction Potential for Water from Coupled Cluster Calculations. I. Analysis of Dimer Potential Energy Surface. *J. Chem. Phys.* **2008**, *128* (9), 094313.
- (128) Smith, D. G. A.; Patkowski, K. Toward an Accurate Description of Methane Physisorption on Carbon Nanotubes. *J. Phys. Chem. C* **2014**, *118* (1), 544–550.

- (129) Smith, D. G. A.; Patkowski, K. Interactions between Methane and Polycyclic Aromatic Hydrocarbons: A High Accuracy Benchmark Study. *J. Chem. Theory Comput.* **2013**, *9* (1), 370–389.
- (130) Vogiatzis, K. D.; Klopper, W.; Friedrich, J. Non-Covalent Interactions of CO<sub>2</sub> with Functional Groups of Metal-Organic Frameworks from a CCSD(T) Scheme Applicable to Large Systems. *J. Chem. Theory Comput.* **2015**, *11* (4), 1574–1584.
- (131) Smith, D. G. A.; Patkowski, K. Benchmarking the CO<sub>2</sub> Adsorption Energy on Carbon Nanotubes. *J. Phys. Chem. C* **2015**, *119* (9), 4934–4948.
- (132) 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* (25), 16560–16574.
- (133) Li, W.; Grimme, S.; Krieg, H.; Möllmann, J.; Zhang, J. Accurate Computation of Gas Uptake in Microporous Organic Molecular Crystals. *J. Phys. Chem. C* **2012**, *116* (16), 8865–8871.
- (134) Lao, K. U.; Schäffer, R.; Jansen, G.; Herbert, J. M. Accurate Description of Intermolecular Interactions Involving Ions Using Symmetry-Adapted Perturbation Theory. *J. Chem. Theory Comput.* **2015**, *11* (6), 2473–2486.
- (135) Sharapa, D. I.; Genaev, A.; Cavallo, L.; Minenkov, Y. A Robust and Cost-Efficient Scheme for Accurate Conformational Energies of Organic Molecules. *ChemPhysChem* **2018**, *20* (1), 92–102.
- (136) Sellers, B. D.; James, N. C.; Gobbi, A. A Comparison of Quantum and Molecular Mechanical Methods to Estimate Strain Energy in Druglike Fragments. *J. Chem. Inf. Model.* **2017**, *57* (6), 1265–1275.
- (137) 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* (10), 2269–2277.
- (138) Tahchieva, D. N.; Bakowies, D.; Ramakrishnan, R.; Von Lilienfeld, O. A. Torsional Potentials of Glyoxal, Oxalyl Halides, and Their Thiocarbonyl Derivatives: Challenges for Popular Density Functional Approximations. *J. Chem. Theory Comput.* **2018**, *14* (9), 4806–4817.
- (139) Folmsbee, D.; Hutchison, G. Assessing Conformer Energies Using Electronic Structure and Machine Learning Methods. *Int. J. Quantum Chem.* **2021**, *121* (1), e26381.
- (140) Prasad, V. K.; Otero-de-la-Roza, A.; DiLabio, G. A. PEPCONF, A Diverse Data Set of Peptide Conformational Energies. *Sci. Data* **2019**, *6*, 180310.
- (141) 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* (19), 7028.
- (142) Valdes, H.; Pluháčková, K.; Pitonák, M.; Řezáč, 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* (19), 2747.
- (143) Kesharwani, M. K.; Karton, A.; Martin, J. M. L. Benchmark *Ab Initio* Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016**, *12* (1), 444–454.
- (144) Mládek, A.; Krepl, M.; Svozil, D.; Čech, P.; Otyepka, M.; Banáš, P.; Zgarbová, M.; Jurečka, P.; Šponer, J. Benchmark Quantum-Chemical Calculations on a Complete Set of Rotameric Families of the DNA Sugar-Phosphate Backbone and Their Comparison with Modern Density Functional Theory. *Phys. Chem. Chem. Phys.* **2013**, *15* (19), 7295–7310.
- (145) Mládek, A.; Banáš, P.; Jurečka, P.; Otyepka, M.; Zgarbová, M.; Šponer, J. Energies and 2'-Hydroxyl Group Orientations of RNA Backbone Conformations. Benchmark CCSD(T)/CBS Database, Electronic Analysis, and Assessment of DFT Methods and MD Simulations. *J. Chem. Theory Comput.* **2014**, *10* (1), 463–480.
- (146) Kruse, H.; Mladek, A.; Gkionis, K.; Hansen, A.; Grimme, S.; Šponer, J. Quantum Chemical Benchmark Study on 46 RNA Backbone Families Using a Dinucleotide Unit. *J. Chem. Theory Comput.* **2015**, *11* (10), 4972–4991.
- (147) 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* (4), 679–692.
- (148) Chan, B. Aqueous-Phase Conformations of Lactose, Maltose, and Sucrose and the Assessment of Low-Cost DFT Methods with the DSCONF Set of Conformers for the Three Disaccharides. *J. Phys. Chem. A* **2020**, *124* (3), 582–590.
- (149) Sameera, W. M. C.; Pantazis, D. A. A Hierarchy of Methods for the Energetically Accurate Modeling of Isomerism in Monosaccharides. *J. Chem. Theory Comput.* **2012**, *8* (8), 2630–2645.
- (150) Marianski, M.; Supady, A.; Ingram, T.; Schneider, M.; Baldauf, C. Assessing the Accuracy of Across-the-Scale Methods for Predicting Carbohydrate Conformational Energies for the Examples of Glucose and  $\alpha$ -Maltose. *J. Chem. Theory*

*Comput.* **2016**, *12* (12), 6157–6168.

- (151) Gruzman, D.; Karton, A.; Martin, J. M. L. Performance of Ab Initio and Density Functional Methods for Conformational Equilibria of  $C_nH_{2n+2}$  Alkane Isomers ( $n=4-8$ ). *J. Phys. Chem. A* **2009**, *113* (43), 11974–11983.
- (152) 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* (1), 293–303.
- (153) Martin, J. M. L. What Can We Learn about Dispersion from the Conformer Surface of n-Pentane? *J. Phys. Chem. A* **2013**, *117* (14), 3118–3132.
- (154) Temelso, B.; Klein, K. L.; Mabey, J. W.; Pérez, C.; Pate, B. H.; Kisiel, Z.; Shields, G. C. Exploring the Rich Potential Energy Surface of (H<sub>2</sub>O)<sub>11</sub> and Its Physical Implications. *J. Chem. Theory Comput.* **2018**, *14* (2), 1141–1153.
- (155) Morgante, P.; Peverati, R. ACCDB: A Collection of Chemistry Databases for Broad Computational Purposes. *J. Comput. Chem.* **2019**, *40* (6), 839–848.
- (156) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn–Sham Global-Hybrid Exchange–Correlation Density Functional with Broad Accuracy for Multi-Reference and Single-Reference Systems and Noncovalent Interactions. *Chem. Sci.* **2016**, *7* (8), 5032–5051.
- (157) Yu, L. J.; Karton, A. Assessment of Theoretical Procedures for a Diverse Set of Isomerization Reactions Involving Double-Bond Migration in Conjugated Dienes. *Chem. Phys.* **2014**, *441*, 166–177.
- (158) Friedrich, J.; Hänchen, J. Incremental CCSD(T)(F12\*)/MP2: A Black Box Method To Obtain Highly Accurate Reaction Energies. *J. Chem. Theory Comput.* **2013**, *9* (12), 5381–5394.
- (159) Friedrich, J. Efficient Calculation of Accurate Reaction Energies—Assessment of Different Models in Electronic Structure Theory. *J. Chem. Theory Comput.* **2015**, *11* (8), 3596–3609.
- (160) Krieg, H.; Grimme, S. Thermochemical Benchmarking of Hydrocarbon Bond Separation Reaction Energies: Jacob’s Ladder Is Not Reversed! *Mol. Phys.* **2010**, *108* (19–20), 2655–2666.
- (161) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7* (2), 291–309.
- (162) Zhao, Y.; González-Garda, N.; Truhlar, D. G. Benchmark Database of Barrier Heights for Heavy Atom Transfer, Nucleophilic Substitution, Association, and Unimolecular Reactions and Its Use to Test Theoretical Methods. *J. Phys. Chem. A* **2005**, *109* (9), 2012–2018.
- (163) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Multi-Coefficient Extrapolated Density Functional Theory for Thermochemistry and Thermochemical Kinetics. *Phys. Chem. Chem. Phys.* **2005**, *7* (1), 43–52.
- (164) Goerigk, L.; Grimme, S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions – Assessment of Common and Reparameterized (Meta-)GGA Density Functionals. *J. Chem. Theory Comput.* **2010**, *6* (1), 107–126.
- (165) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-2 and Density Functional Theories for the Computation of Enthalpies of Formation. *J. Chem. Phys.* **1997**, *106* (3), 1063–1079.
- (166) Yu, L.-J.; Sarrami, F.; O’Reilly, R. J.; Karton, A. Can DFT and Ab Initio Methods Describe All Aspects of the Potential Energy Surface of Cycloreversion Reactions? *Mol. Phys.* **2015**, *114* (1), 21–33.
- (167) Karton, A.; Schreiner, P. R.; Martin, J. M. L. Heats of Formation of Platonic Hydrocarbon Cages by Means of High-Level Thermochemical Procedures. *J. Comput. Chem.* **2016**, *37* (1), 49–58.
- (168) Karton, A.; Gruzman, D.; Martin, J. M. L. Benchmark Thermochemistry of the  $C_nH_{2n+2}$  Alkane Isomers ( $n = 2-8$ ) and Performance of DFT and Composite Ab Initio Methods for Dispersion-Driven Isomeric Equilibria. *J. Phys. Chem. A* **2009**, *113* (29), 8434–8447.
- (169) Johnson, E. R.; Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Delocalization Errors in Density Functionals and Implications for Main-Group Thermochemistry. *J. Chem. Phys.* **2008**, *129* (20), 204112.
- (170) Karton, A.; Martin, J. M. L. Explicitly Correlated Benchmark Calculations on C<sub>8</sub>H<sub>8</sub> Isomer Energy Separations: How Accurate Are DFT, Double-Hybrid, and Composite Ab Initio Procedures? *Mol. Phys.* **2012**, *110* (19–20), 2477–2491.
- (171) Manna, D.; Martin, J. M. L. What Are the Ground State Structures of C<sub>20</sub> and C<sub>24</sub>? An Explicitly Correlated Ab Initio Approach. *J. Phys. Chem. A* **2015**, *120* (1), 153–160.
- (172) Zhao, Y.; Tishchenko, O.; Gour, J. R.; Li, W.; Lutz, J. J.; Piecuch, P.; Truhlar, D. G. Thermochemical Kinetics for Multireference Systems: Addition Reactions of Ozone. *J. Phys. Chem. A* **2009**, *113* (19), 5786–5799.
- (173) Lee, J. S. Accurate Ab Initio Binding Energies of Alkaline Earth Metal Clusters. *J. Phys. Chem. A* **2005**, *109* (51), 11927–11932.

- (174) Lepetit, C.; Chermette, H.; Gicquel, M.; Heully, J.-L.; Chauvin, R. Description of Carbo-Oxocarbons and Assessment of Exchange–Correlation Functionals for the DFT Description of Carbo-Mers. *J. Phys. Chem. A* **2006**, *111* (1), 136–149.
- (175) Schreiner, P. R.; Fokin, A. A.; Pascal Jr., R. A.; de Meijere, A. Many Density Functional Theory Approaches Fail To Give Reliable Large Hydrocarbon Isomer Energy Differences. *Org. Lett.* **2006**, *8* (17), 3635–3638.
- (176) Woodcock, H. L.; Schaefer III, H. F.; Schreiner, P. R. Problematic Energy Differences between Cumulenes and Poly-Ynes: Does This Point to a Systematic Improvement of Density Functional Theory? *J. Phys. Chem. A* **2002**, *106* (49), 11923–11931.
- (177) Piacenza, M.; Grimme, S. Systematic Quantum Chemical Study of DNA-Base Tautomers. *J. Comput. Chem.* **2004**, *25* (1), 83–99.
- (178) Grimme, S.; Mück-Lichtenfeld, C.; Würthwein, E.-U.; Ehlers, A. W.; Goumans, T. P. M.; Lammertsma, K. Consistent Theoretical Description of 1,3-Dipolar Cycloaddition Reactions. *J. Phys. Chem. A* **2006**, *110* (8), 2583–2586.
- (179) Grimme, S. Semiempirical Hybrid Density Functional with Perturbative Second-Order Correlation. *J. Chem. Phys.* **2006**, *124* (3), 034108.
- (180) Karton, A.; O’Reilly, R. J.; Radom, L. Assessment of Theoretical Procedures for Calculating Barrier Heights for a Diverse Set of Water-Catalyzed Proton-Transfer Reactions. *J. Phys. Chem. A* **2012**, *116* (16), 4211–4221.
- (181) Grimme, S.; Kruse, H.; Goerigk, L.; Erker, G. The Mechanism of Dihydrogen Activation by Frustrated Lewis Pairs Revisited. *Angew. Chemie Int. Ed.* **2010**, *49* (8), 1402–1405.
- (182) Grambow, C. A.; Pattanaik, L.; Green, W. H. Reactants, Products, and Transition States of Elementary Chemical Reactions Based on Quantum Chemistry. *Sci. Data* **2020**, *7* (1), 1–8.
- (183) Rudorff, G. F. von; Heinen, S. N.; Bragato, M.; Lilienfeld, O. A. von. Thousands of Reactants and Transition States for Competing E2 and S2 Reactions. *Mach. Learn. Sci. Technol.* **2020**, *1* (4), 045026.
- (184) Karton, A.; Goerigk, L. Accurate Reaction Barrier Heights of Pericyclic Reactions: Surprisingly Large Deviations for the CBS-QB3 Composite Method and Their Consequences in DFT Benchmark Studies. *J. Comput. Chem.* **2015**, *36* (9), 622–632.
- (185) Zheng, J.; Zhao, Y.; Truhlar, D. G. Representative Benchmark Suites for Barrier Heights of Diverse Reaction Types and Assessment of Electronic Structure Methods for Thermochemical Kinetics. **2007**, *3* (2), 569–582.
- (186) Karton, A.; Tarnopolsky, A.; Lamère, J.-F.; Schatz, G. C.; Martin, J. M. L. Highly Accurate First-Principles Benchmark Data Sets for the Parametrization and Validation of Density Functional and Other Approximate Methods. Derivation of a Robust, Generally Applicable, Double-Hybrid Functional for Thermochemistry and Thermochemical Kinetics. *J. Phys. Chem. A* **2008**, *112* (50), 12868–12886.
- (187) Goerigk, L.; Sharma, R. The INV24 Test Set: How Well Do Quantum-Chemical Methods Describe Inversion and Racemization Barriers? *Can. J. Chem.* **2016**, *94* (12), 1133–1143.
- (188) Yu, L. J.; Sarrami, F.; O’Reilly, R. J.; Karton, A. Reaction Barrier Heights for Cycloreversion of Heterocyclic Rings: An Achilles’ Heel for DFT and Standard Ab Initio Procedures. *Chem. Phys.* **2015**, *458*, 1–8.
- (189) Karton, A.; O’Reilly, R. J.; Chan, B.; Radom, L. Determination of Barrier Heights for Proton Exchange in Small Water, Ammonia, and Hydrogen Fluoride Clusters with G4(MP2)-Type, MPn, and SCS-MPn Procedures—a Caveat. *J. Chem. Theory Comput.* **2012**, *8* (9), 3128–3136.
- (190) Chan, B.; Gilbert, A. T. B.; Gill, P. M. W.; Radom, L. Performance of Density Functional Theory Procedures for the Calculation of Proton-Exchange Barriers: Unusual Behavior of M06-Type Functionals. *J. Chem. Theory Comput.* **2014**, *10* (9), 3777–3783.
- (191) Smith, B. J.; Swanton, D. J.; Pople, J. A.; Schaefer III, H. F.; Radom, L. Transition Structures for the Interchange of Hydrogen Atoms within the Water Dimer. *J. Chem. Phys.* **1990**, *92* (2), 1240–1247.
- (192) Tschumper, G. S.; Leininger, M. L.; Hoffman, B. C.; Valeev, E. F.; Schaefer III, H. F.; Quack, M. Anchoring the Water Dimer Potential Energy Surface with Explicitly Correlated Computations and Focal Point Analyses. *J. Chem. Phys.* **2002**, *116* (2), 690–701.
- (193) Grimme, S.; Steinmetz, M.; Korth, M. How to Compute Isomerization Energies of Organic Molecules with Quantum Chemical Methods. *J. Org. Chem.* **2007**, *72* (6), 2118–2126.
- (194) Huenerbein, R.; Schirmer, B.; Moellmann, J.; Grimme, S. Effects of London Dispersion on the Isomerization Reactions of Large Organic Molecules: A Density Functional Benchmark Study. *Phys. Chem. Chem. Phys.* **2010**, *12* (26), 6940–6948.
- (195) Grimme, S. Seemingly Simple Stereoelectronic Effects in Alkane Isomers and the Implications for Kohn–Sham Density Functional Theory. *Angew. Chemie Int. Ed.* **2006**, *45* (27), 4460–4464.



- (196) Schwabe, T.; Grimme, S. Double-Hybrid Density Functionals with Long-Range Dispersion Corrections: Higher Accuracy and Extended Applicability. *Phys. Chem. Chem. Phys.* **2007**, *9* (26), 3397–3406.
- (197) Yu, L.-J.; Sarrami, F.; Karton, A.; O'Reilly, R. J. An Assessment of Theoretical Procedures for  $\pi$ -Conjugation Stabilisation Energies in Enones. *Mol. Phys.* **2014**, *113* (11), 1284–1296.
- (198) Karton, A. How Reliable Is DFT in Predicting Relative Energies of Polycyclic Aromatic Hydrocarbon Isomers? Comparison of Functionals from Different Rungs of Jacob's Ladder. *J. Comput. Chem.* **2017**, *38* (6), 370–382.
- (199) Yoo, S.; Aprà, E.; Zeng, X. C.; Xantheas, S. S. High-Level Ab Initio Electronic Structure Calculations of Water Clusters (H<sub>2</sub>O)<sub>16</sub> and (H<sub>2</sub>O)<sub>17</sub>: A New Global Minimum for (H<sub>2</sub>O)<sub>16</sub>. *J. Phys. Chem. Lett.* **2010**, *1* (20), 3122–3127.
- (200) Lao, K. U.; Herbert, J. M. Accurate and Efficient Quantum Chemistry Calculations for Noncovalent Interactions in Many-Body Systems: The XSAPT Family of Methods. *J. Phys. Chem. A* **2015**, *119* (2), 235–252.
- (201) Mardirossian, N.; Lambrecht, D. S.; McCaslin, L.; Xantheas, S. S.; Head-Gordon, M. The Performance of Density Functionals for Sulfate-Water Clusters. *J. Chem. Theory Comput.* **2013**, *9* (3), 1368–1380.
- (202) Karton, A.; Sylvetsky, N.; Martin, J. M. L. W4-17: A Diverse and High-Confidence Dataset of Atomization Energies for Benchmarking High-Level Electronic Structure Methods. *J. Comput. Chem.* **2017**, *38* (24), 2063–2075.
- (203) Taylor, D. E.; Ángyán, J. G.; Galli, G.; Zhang, C.; Gygi, F.; Hirao, K.; Song, J. W.; Rahul, K.; Anatole Von Lilienfeld, O.; Podeszwa, R.; et al. Blind Test of Density-Functional-Based Methods on Intermolecular Interaction Energies. *J. Chem. Phys.* **2016**, *145* (12), 124105.
- (204) Donchev, A. G.; Taube, A. G.; Decolvenaere, E.; Hargus, C.; McGibbon, R. T.; Law, K.-H.; Gregersen, B. A.; Li, J.-L.; Palmo, K.; Siva, K.; et al. Quantum Chemical Benchmark Databases of Gold-Standard Dimer Interaction Energies. *Sci. Data* **2021**, *8* (1), 1–9.
- (205) Sparrow, Z. M.; Ernst, B. G.; Joo, P. T.; Lao, K. U.; DiStasio Jr., R. A. NENCI-2021 Part I: A Large Benchmark Database of Non-Equilibrium Non-Covalent Interactions Emphasizing Close Intermolecular Contacts. *arXiv* **2021**. (arXiv:2102.02354v1)
- (206) Romero-Montalvo, E.; DiLabio, G. A. Computational Study of Hydrogen Bond Interactions in Water Cluster-Organic Molecule Complexes. *J. Phys. Chem. A* **2021**, *125* (16), 3369–3377.
- (207) Miriyala, V. M.; Řezáč, J. Testing Semiempirical Quantum Mechanical Methods on a Data Set of Interaction Energies Mapping Repulsive Contacts in Organic Molecules. *J. Phys. Chem. A* **2018**, *122* (10), 2801–2808.
- (208) Kříž, K.; Nováček, M.; Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 3: Repulsive Contacts. *J. Chem. Theory Comput.* **2021**, *17* (3), 1548–1561.
- (209) Mehta, N.; Fellowes, T.; White, J.; Goerigk, L. The CHAL336 Benchmark Set: How Well Do Quantum-Chemical Methods Describe Chalcogen-Bonding Interactions? *J. Chem. Theory Comput.* **2021**, *17* (5), 2783–2806.
- (210) Oliveira, V.; Kraka, E.; Cremer, D. The Intrinsic Strength of the Halogen Bond: Electrostatic and Covalent Contributions Described by Coupled Cluster Theory. *Phys. Chem. Chem. Phys.* **2016**, *18* (48), 33031–33046.
- (211) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J. Chem. Theory Comput.* **2013**, *9* (8), 3364–3374.
- (212) Calbo, J.; Ortí, E.; Sancho-García, J. C.; Aragó, J. Accurate Treatment of Large Supramolecular Complexes by Double-Hybrid Density Functionals Coupled with Nonlocal van Der Waals Corrections. *J. Chem. Theory Comput.* **2015**, *11* (3), 932–939.
- (213) Sure, R.; Grimme, S. Comprehensive Benchmark of Association (Free) Energies of Realistic Host–Guest Complexes. *J. Chem. Theory Comput.* **2015**, *11* (8), 3785–3801.
- (214) Ni, Z.; Guo, Y.; Neese, F.; Li, W.; Li, S. Cluster-in-Molecule Local Correlation Method with an Accurate Distant Pair Correction for Large Systems. *J. Chem. Theory Comput.* **2021**, *17* (2), 756–766.
- (215) Sharapa, D. I.; Margraf, J. T.; Hesselmann, A.; Clark, T. Accurate Intermolecular Potential for the C60 Dimer: The Performance of Different Levels of Quantum Theory. *J. Chem. Theory Comput.* **2017**, *13* (1), 274–285.
- (216) Lao, K. U.; Herbert, J. M. An Improved Treatment of Empirical Dispersion and a Many-Body Energy Decomposition Scheme for the Explicit Polarization plus Symmetry-Adapted Perturbation Theory (XSAPT) Method. *J. Chem. Phys.* **2013**, *139* (3), 034107.
- (217) Mezei, P. D.; Csonka, G. I.; Ruzsinszky, A.; Sun, J. Accurate, Precise, and Efficient Theoretical Methods to Calculate Anion- $\pi$  Interaction Energies in Model Structures. *J. Chem. Theory Comput.* **2015**, *11* (1), 360–371.
- (218) Zahn, S.; Macfarlane, D. R.; Izgorodina, E. I. Assessment of Kohn-Sham Density Functional Theory and Møller-Plesset Perturbation Theory for Ionic Liquids. *Phys. Chem. Chem. Phys.* **2013**, *15* (32), 13664–13675.
- (219) Zhang, H.; Krupa, J.; Wierzejewska, M.; Biczysko, M. The Role of Dispersion and Anharmonic Corrections in

Conformational Analysis of Flexible Molecules: The Allyl Group Rotamerization of Matrix Isolated Safrole. *Phys. Chem. Chem. Phys.* **2019**, *21* (16), 8352–8364.

- (220) Kirschner, K. N.; Heiden, W.; Reith, D. Small Alcohols Revisited: CCSD(T) Relative Potential Energies for the Minima, First- and Second-Order Saddle Points, and Torsion-Coupled Surfaces. *ACS Omega* **2018**, *3* (1), 419–432.
- (221) Greenwell, C.; Beran, G. J. O. Inaccurate Conformational Energies Still Hinder Crystal Structure Prediction in Flexible Organic Molecules. *Cryst. Growth Des.* **2020**, *20* (8), 4875–4881.
- (222) Lahey, S. L. J.; Thien Phuc, T. N.; Rowley, C. N. Benchmarking Force Field and the ANI Neural Network Potentials for the Torsional Potential Energy Surface of Biaryl Drug Fragments. *J. Chem. Inf. Model.* **2020**, *60* (12), 6258–6268.
- (223) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x Data Sets, Coupled-Cluster and Density Functional Theory Properties for Molecules. *Sci. Data* **2020**, *7* (1), 1–10.
- (224) Řezáč, J.; Bím, D.; Gutten, O.; Rulíšek, L. Toward Accurate Conformational Energies of Smaller Peptides and Medium-Sized Macrocycles: MPCONF196 Benchmark Energy Data Set. *J. Chem. Theory Comput.* **2018**, *14* (3), 1254–1266.