



Cite this: *Phys. Chem. Chem. Phys.*,  
2025, 27, 23336

# Performance of semiempirical DFT methods for the supramolecular assembly of Janus-face cyclohexanes

Bruno A. Piscelli, <sup>ab</sup> Tiger Swithenbank-Michel, <sup>b</sup> Rodrigo A. Cormanich, <sup>a</sup>  
David O'Hagan <sup>b</sup> and Michael Bühl <sup>\*b</sup>

A series of GFN-xTB methods were benchmarked against high-level DFT and *ab initio* thermodynamic data for a set of conformational equilibria and driving forces for the formation of non-covalent complexes involving the Janus-face fluorocyclohexanes based on the all-*syn*-C<sub>6</sub>F<sub>n</sub>R<sub>12-n</sub> motif (*n* = 3, 5, 6). When used alone, GFN methods showed moderate performance, with mean absolute errors (MAEs) from the high-level benchmarks of approximately 2.5 kcal mol<sup>-1</sup> for conformational equilibria and ~5.0 kcal mol<sup>-1</sup> for molecular complexes. However, applying DFT-level single-point energy corrections on GFN-optimised geometries significantly improved the accuracy, reducing MAEs to ~0.2 and ~1.0 kcal mol<sup>-1</sup> for the same systems. This hybrid approach achieves DFT-D3-level accuracy while maintaining a low computational cost, offering up to a 50-fold reduction in computational time. As such, it provides a new cost-efficient and accurate tool for the computational modeling of Janus-face systems. An illustrative application to a flexible system, C<sub>6</sub>F<sub>5</sub>H<sub>6</sub>O<sub>2</sub>C(CH<sub>2</sub>)<sub>3</sub>NHCOC<sub>6</sub>H<sub>2</sub>(OR)<sub>3</sub>, is reported (R = alkyl), highlighting the relative stabilities of folded and extended forms and their supramolecular assembly into helical stacks.

Received 28th July 2025,  
Accepted 6th October 2025

DOI: 10.1039/d5cp02879e

rsc.li/pccp

## Introduction

Because fluorine forms very strong covalent bonds with carbon (~105 kcal mol<sup>-1</sup>, on average), the development of new organofluorine materials is flourishing as they are chemically stable through wide temperature ranges and resist interactions with hydrocarbons and water.<sup>1-5</sup> These per- and polyfluoroalkyl substances, commonly referred to as PFAS, exhibit high thermal stability and surface repellent properties, both arising from the strength and low reactivity of C-F bonds, making them ubiquitous in the manufacture of modern electronics, textiles, and the development of paints and coatings.<sup>6-11</sup>

However, as a class, PFAS are non-biodegradable and are often referred to as 'forever chemicals' due to their long half-lives.<sup>12-16</sup> They are also highly resistant to metabolic degradation, often breaking down into other stable PFAS rather than innocuous products.<sup>17,18</sup> As a result, their widespread use has raised concerns due to their high persistence in the environment,<sup>19-21</sup> propelling the search for new classes of organo-fluorine motifs as an alternative to the existing PFAS. In recent years, selectively fluorinated all-*syn* cyclohexanes with

all of the fluorine atoms on one face of the cyclohexane ring have emerged as extraordinarily polar aliphatics, with all-*syn*-1,2,3,4,5,6-hexafluorocyclohexane **1** (Fig. 1(A)), first prepared by O'Hagan *et al.*, being a prominent representative of this new class of compounds.<sup>22</sup> In this molecule, three C-F bonds are co-aligned in an axial orientation. This results in a large dipole moment (calculated at 6.2 D), which polarizes the molecule. Also, the strong electron-withdrawing nature of the fluorine atoms polarizes the hydrogens, rendering them highly electropositive and leading to an unprecedented electrostatic profile for an aliphatic molecule, with cyclohexane **1** exhibiting a negatively charged fluorine face and a positively charged hydrogen face, as illustrated in Fig. 1. These aspects result in high thermal stability with decomposition, rather than melting, occurring above 200 °C.

Santschi and Gilmour aptly coined the term 'Janus Face' cyclohexanes to describe this dual characteristic, referencing the Roman god Janus, known for his two opposing faces.<sup>23</sup> This unique Janus characteristic induces highly organized supramolecular assembly around this motif, with molecular stacking of such Janus rings enabled by intermolecular interactions of the fluorine faces with hydrogen faces.<sup>24-26</sup>

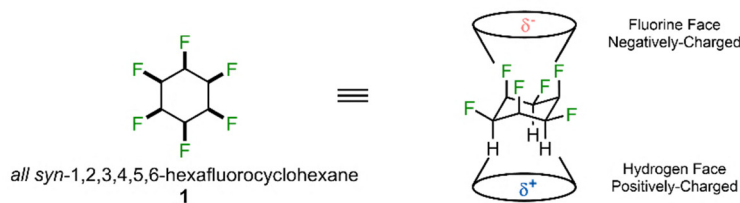
Despite their unique properties and potential applicability, to our knowledge only one report, by Pavan and Delius *et al.* in 2021, has used the Janus characteristic of these cyclohexanes to induce and control dynamic supramolecular assembly.<sup>27</sup> In our

<sup>a</sup> UNICAMP, Universidade Estadual de Campinas, Instituto de Química,  
Rua Monteiro Lobato 270, 13083-862, Campinas, São Paulo, Brazil

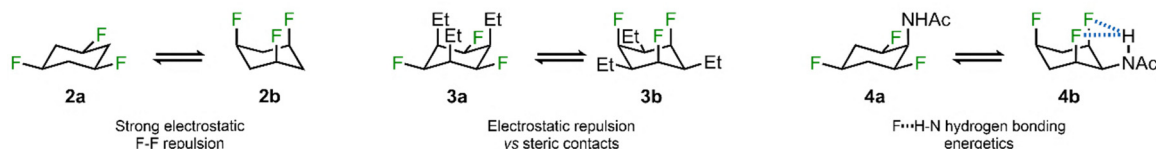
<sup>b</sup> EaStCHEM, School of Chemistry, University of St Andrews, St Andrews, Fife,  
KY16 9ST, UK. E-mail: buehl@st-andrews.ac.uk



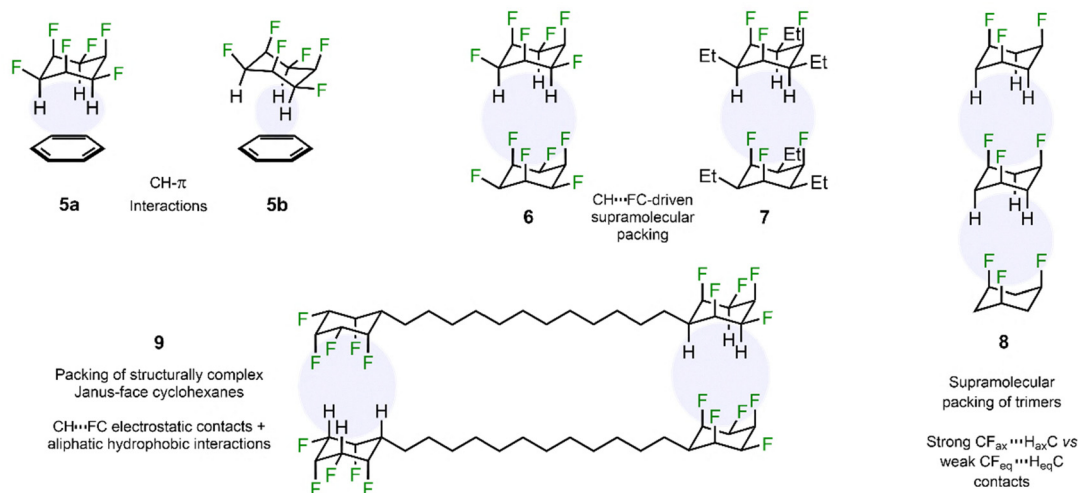
## A) The Highly Polarized Janus-Hexafluorocyclohexane



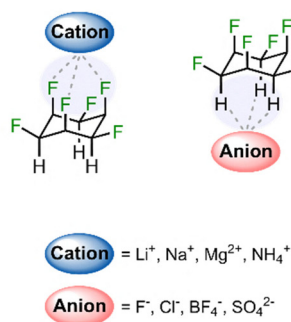
## B) Group 1 - Conformation



## C) Group 2 - Non-Covalent Interactions



## D) Group 3 - Complexation of Ions



## E) Group 4 - Application in Supramolecular Chemistry

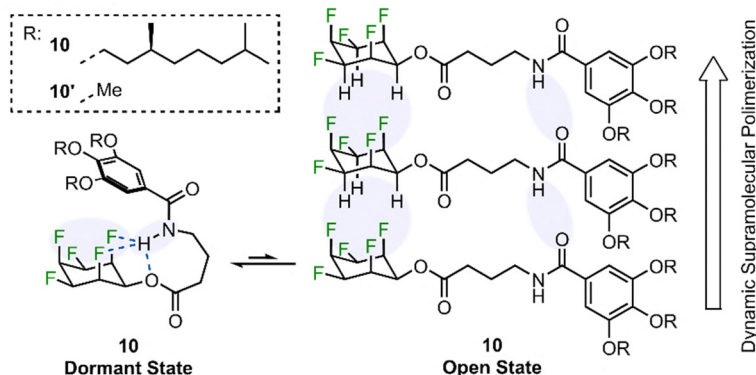


Fig. 1 (A) The parent hexafluorocyclohexane, (B)–(E) Systems proposed for benchmarking xTB methods.

opinion, a computational method able to deal with the large systems used in supramolecular applications (in the order of thousands of atoms), while maintaining a good compromise between computational time and accuracy, would help facilitate the use of these fluorinated rings in supramolecular chemistry. Therefore, we propose a set of systems divided into

4 groups (Fig. 1) to assess the accuracy of cheap semiempirical DFT variants, specifically tight-binding GFN1-xTB and GFN2-xTB methods<sup>28,29</sup> as well as xTB's universal forcefield GFN-FF<sup>30</sup> on predicting: (1) conformational preferences; (2) non-covalent assemblies; and (3) ion complexation thermodynamics. The low computational cost and large applicability of the xTB



methods make them good options for this case. Finally, the molecules used by Pavan and Delius *et al.*<sup>27</sup> are used as a proof of concept for a 'real' supramolecular application of these methods (Group 4).

## Experimental

### Computational details

Calculations at the force-field level (GFN-FF)<sup>30</sup> and semi-empirical tight-binding methods (GFN1<sup>28</sup> and GFN2<sup>29</sup>) were performed using xTB 6.0.2 software.<sup>31</sup> Density functional theory (DFT) calculations employed the widely used B3LYP functional<sup>32</sup> with Ahlrichs' triple- $\zeta$  def2-TZVP basis set<sup>33</sup> (def2-TZVPD for anionic species<sup>34</sup>) and Grimme's empirical dispersion correction D3<sup>35</sup> with the original damping. This approach, referred to here as DFT-D3, was chosen due to its popularity and strong performance in previous studies on Janus-face cyclohexane systems.<sup>24–26</sup> For non-covalent complexes, the effect of Counterpoise (CP) corrections for basis set superposition error (BSSE) was assessed, using a 2-fragment scheme for dimers and a 3-fragment scheme for trimeric structures. When included, this correction is indicated by the superscript "CP" (*i.e.*, DFT-D3<sup>CP</sup>). DFT calculations were carried out using Gaussian 16 Rev C.01 software.<sup>36</sup> Geometry optimisation in both xTB and Gaussian software was carried out using internal coordinates and the default thresholds for convergence criteria. Additionally, the impact of computing single-point electronic energies at a higher level than the geometry optimisation was examined. This is denoted as "method1//method2," where method2 corresponds to the geometry optimisation and thermodynamic correction (obtained from frequency calculations within the perfect gas, rigid-rotor, and harmonic oscillator approximations) calculations, while method1 refers to the higher-level single-point energy calculation. Relative energies are discussed in terms of Gibbs free energy differences ( $\Delta G$ ) evaluated at 298.15 K. Unless otherwise specified, reference values are taken from previously reported DFT or *ab initio* calculations.

For the conformational analysis of compound **10**, a conformational search was carried out using the iterative-static metadynamics algorithm as implemented in the CREST 2.11.2 software package.<sup>37,38</sup> The GFN-FF, GFN1, and GFN2 methods were employed both for the conformational sampling and subsequent reoptimisation steps. Preliminary calculations revealed that the majority of conformational flexibility in **10** arose from the large alkyl chains appended to the aromatic core. Consequently, most of the conformers identified during the sampling process reflected variations in the orientation of these chains, with minimal changes to the central scaffold. Given that the primary objective was to assess the geometry of the molecular core, the long alkyl substituents were truncated and replaced with methyl groups (**10'**, R = Me in Fig. 1). Under these conditions, each xTB method yielded approximately 1000 conformers. These structures were then ranked using a principal component analysis (PCA) and *k*-means approach, also implemented in CREST, to identify the 100 most representative

conformers per method. The 4 conformers qualitatively resembling the relevant structures I–IV of **10** determined through well-tempered metadynamics calculations, as reported by Pavan and Delius *et al.*,<sup>27</sup> were selected by visual inspection and subsequently reoptimised at the corresponding theoretical level. Finally, harmonic frequency calculations were performed at the xTB levels to obtain the relative Gibbs free energies for each conformer.

## Results and discussion

### Group 1: conformational equilibria

The first group of molecules includes the conformational pairs of molecules **2**, **3**, and **4** (Fig. 1(B)), which were selected to represent increasing complexity. Compound **2** is an *all-syn* 1,3,5-fluorocyclohexane, where conformer **2a** places all fluorine atoms in equatorial positions, while in **2b**, they adopt axial positions, resulting in destabilizing F<sub>ax</sub>–F<sub>ax</sub> electrostatic interactions, resulting in a preference for the **2a** conformer. The complexity of the system increases in compound **3**, an *all-syn* 1,3,5-triethyl-2,4,6-trifluorocyclohexane, where **3a** exhibits strong Et<sub>ax</sub>–Et<sub>ax</sub> steric clashes, and **3b** displays the same F<sub>ax</sub>–F<sub>ax</sub> electrostatic interactions as in **2b**. In compound **3**, however, the tri-axial arrangement of the fluorines in **3b** is energetically favoured. Compound **4** is structurally similar to **2** but contains a carbamate group at position 2 of the ring, *syn* to the fluorine atoms, enabling in **4b** a F<sub>ax</sub>–HN hydrogen bond in addition to F<sub>ax</sub>–F<sub>ax</sub> interactions, again resulting in a preference for **4b**. We calculated the Gibbs free energy differences ( $\Delta G$ ) between each conformational pair at a variety of theoretical levels, namely fully optimised at the GFN and DFT-D3 levels, as well as selected GFN and DFT-D3 single points, to probe how well the results from full optimisations at the higher level are approximated through single points on lower-level geometries. As a benchmark, we use the DLPNO-CCSD(T)/CBS//B3LYP-D3/def2-TZVP level, which affords relative free energies of  $\Delta G$  = 3.53, –9.52, and –2.03 kcal mol<sup>–1</sup> for the **2b**–**2a**, **3b**–**3a** and **4b**–**4a** pairs, respectively (where a negative sign indicates that the conformer with axial F atoms is favoured, in agreement with our previous findings to analogous systems<sup>25</sup>). The results are summarized graphically in Fig. 2.

Overall, force-field calculations using GFN-FF, which was originally designed to describe intermolecular (and not intramolecular) interactions in supramolecular assemblies and large molecules,<sup>30</sup> do not provide reasonable results compared to reference DFT calculations, exhibiting a mean absolute error (MAE) near 3 kcal mol<sup>–1</sup> across compounds **2**, **3**, and **4**. The only exception is the GFN1//GFN-FF composite method, which achieves an MAE of 0.94 kcal mol<sup>–1</sup>. Although single-point energy calculations at the DFT-D3 level (DFT-D3//GFN-FF) should theoretically yield better results than those at GFN1//GFN-FF, this is not the case. Additionally, the second-generation semi-empirical xTB method, GFN2, also exhibits poor performance, with an MAE of 2.51 kcal mol<sup>–1</sup> when associated with GFN-FF geometry optimisations (GFN2//GFN-FF). Thus, the relatively good accuracy of the GFN1//GFN-FF



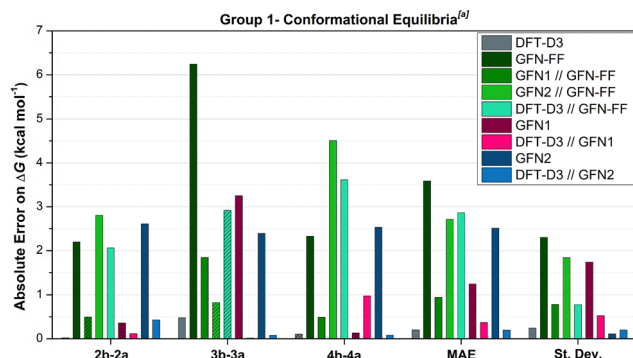


Fig. 2 Absolute (AE) and mean (MAE) errors and standard deviation of absolute errors (st. dev.) on  $\Delta G$  of different computational methods for compounds **2**, **3** and **4** (Group 1), in  $\text{kcal mol}^{-1}$ . Hashed and solid bars indicate negative and positive errors, respectively. <sup>a</sup> Reference theoretical level: DLPNO-CCSD(T)/CBS//B3LYP-D3/def2-TZVP.

approach likely arises from error cancellation between the two methods, rather than true chemical accuracy. Importantly, the F...F contact distances predicted by GFN-FF are consistently shorter (2b and 4b) or longer (3b) compared to those obtained with GFN1 and GFN2, which in turn provide very similar geometries (Fig. S1). This suggests that GFN1 treats F...F contacts in a less distance-dependent manner, such that the energetic penalties associated with overly short or long contacts in GFN-FF geometries are mitigated when single-point energies are evaluated at the GFN1 level. In contrast, for the DFT-D3//GFN-FF and GFN2//GFN-FF composite methods, the computed energies appear to be more sensitive to the precise F...F distances. Since the fluorine atomic charges predicted by GFN-FF and GFN1 are generally similar, the observed error cancellation most likely originates from the dispersive contribution included in GFN1 rather than from electrostatics. Moreover, the poor performance of the GFN1//GFN-FF method for the conformational equilibrium of **3** suggests that this approach may not adequately capture the steric destabilization associated with hydrophobic alkyl-alkyl contacts and may be sensitive to the system studied, highlighted by the relatively high standard deviation across the series.

In contrast, the GFN1 and GFN2 methods were benchmarked on largely the same test sets for conformational equilibria in their original papers,<sup>28,29</sup> and both display strong overall performance, reflecting their reliable description of intramolecular interactions. Moreover, both methods deliver geometries of sufficiently high quality that the corresponding composite methods, DFT-D3//GFN1 and DFT-D3//GFN2, perform remarkably well, with MAEs of 0.37 and 0.20  $\text{kcal mol}^{-1}$ , respectively. Notably, the DFT-D3//GFN2 approach even matched full DFT-D3 calculations for the studied systems (MAE = 0.20  $\text{kcal mol}^{-1}$ ), highlighting that these hybrid strategies are cost-effective options for computing relative Gibbs free energies of Janus cyclohexane conformational equilibria without compromising chemical accuracy.

Importantly, the methods GFN-FF, GFN2//GFN-FF, DFT-D3//GFN-FF, and GFN2 failed to correctly predict the preferred

tri-axial conformation in **4b**, underscoring significant limitations in their ability to accurately capture conformational energy differences in this case (see Table S1 for further details).

## Group 2: intermolecular non-covalent complexes

The second group of molecules includes the noncovalently bonded complexes **5**–**9** (Fig. 1(C)), where we calculated the Gibbs free energy of association ( $\Delta G$ ) relative to the separated (fully optimised) molecular fragments. We used calculations at DLPNO-CCSD(T)/CBS//B3LYP-D3<sup>CP</sup>/def2-TZVP and B3LYP-D3<sup>CP</sup>/def2-TZVP as reference values (see Fig. 3 for details), affording target  $\Delta G$  values of 4.62, 1.55,  $-16.53$ , 8.94 and  $-1.86$   $\text{kcal mol}^{-1}$  for complexes **5a**, **6**, **7**, **8** and **9**, respectively. Importantly, although complexes **5a**, **6**, and **8** display positive Gibbs free energies of association, all complexation processes are predicted to be exothermic in terms of enthalpy (Table S2). In these cases, the positive  $\Delta G$  values arise from the treatment of translational degrees of freedom in the gas phase, which imposes substantial entropic penalties on the reactions involving a reduction in the number of particles. This well-known effect can be mitigated by including solvation effects or, more recently, by performing elevated-pressure gas-phase calculations that mimic an ideal gas at liquid-like density.<sup>39</sup> Still, because the goal of this study is an internal validation of methods under the same conditions, none of these corrections were applied here. Counterpoise corrections played an important role in the performance of the tested methods and led to general improvements in accuracy, thus only CP-corrected DFT-D3<sup>CP</sup> calculations are reported (DFT-D3 results are displayed in the SI).

Molecular complex **5a** was chosen to study how the xTB methods deal with hydrophobic CH- $\pi$  interactions between hexafluorocyclohexane **1** and benzene.<sup>40</sup> To our great surprise, electronic energy corrections at DFT-D3<sup>CP</sup> on GFN structures continues to perform well compared to the pure DFT-D3<sup>CP</sup> calculations in Group 2, in light of their good performance in Group 1. DFT-D3<sup>CP</sup>, DFT-D3<sup>CP</sup>//GFN1 and DFT-D3<sup>CP</sup>//GFN2

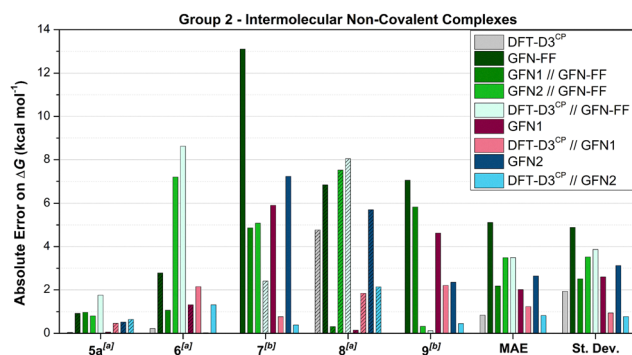


Fig. 3 Absolute (AE) and mean (MAE) errors and standard deviation of absolute errors (st. dev.) on  $\Delta G$  of different computational methods for the complexation energies of **5a**, **6**, **7**, **8** and **9** (Group 2), in  $\text{kcal mol}^{-1}$ . Hashed and solid bars indicate negative and positive errors, respectively. Reference theoretical levels: <sup>a</sup>DLPNO-CCSD(T)/CBS//B3LYP-D3/def2-TZVP and <sup>b</sup>B3LYP-D3<sup>CP</sup>/def2-TZVP.





methods exhibit AEs of 0.05, 0.46 and 0.65 kcal mol<sup>-1</sup>, respectively, on predicted binding  $\Delta G$ 's, a strong performance compared to the high-level wave function reference values (DLPNO-CCSD(T)/CBS//B3LYP-D3<sup>CP</sup>/def2-TZVP). Pure xTB methods also perform well for the binding Gibbs free energy of complex **5a**, especially GFN1, with an AE of 0.06 kcal mol<sup>-1</sup>, and GFN-FF and GFN2 maintaining strong performance with AEs of 0.93 and 0.52 kcal mol<sup>-1</sup>, respectively. However, when the electronic energy is calculated at the semi-empirical or DFT level over GFN-FF geometries, the absolute errors progressively move away from the reference values when we go from GFN1, GFN2 and DFT-D3<sup>CP</sup>, in all cases with AEs over 1 kcal mol<sup>-1</sup>.

It is noteworthy that the global minimum conformer for the 1-benzene complex is not C<sub>3v</sub>-symmetric **5a**, but the "slipped-sandwich" complex **5b** instead, which is more stable than **5a** by 0.47 kcal mol<sup>-1</sup> according to B3LYP-D3/def2-TZVP calculations in terms of  $\Delta G$ , or by 2.30 kcal mol<sup>-1</sup> when CP corrections are applied. This sheds light on an important limitation of the GFN methods, which may be insensitive to distinct conformers which are very close in energy and separated by small energy wells in the potential energy surface (PES). Considering that geometry optimisations in GFN-FF, GFN1 and GFN2 all led to the c<sub>3v</sub> complex, even when the starting structure was an already DFT pre-optimised slipped sandwich geometry, most probably the latter is not even an energy minimum on the PES of the xTB methods. However, considering both interaction modes are very close in energy, this limitation is not critical in this case and should not imply major concerns when one chooses one over another method.

The dimeric arrangement of hexafluorocyclohexane **1**, complex **6**, was chosen to assess the ability of the computational methods to estimate the energy of the strong CF<sub>ax</sub>-H<sub>ax</sub>C contacts that are electrostatic in nature and drive supramolecular arrangements in Janus-face cyclohexanes. In this case, the absolute errors of DFT-D3<sup>CP</sup>, DFT-D3<sup>CP</sup>/GFN1 and DFT-D3<sup>CP</sup>/GFN2 methods are at 0.22, 2.15 and 1.32 kcal mol<sup>-1</sup>. Pure GFN-FF and GFN1 calculations exhibit moderate AEs of 2.78 and 1.32 kcal mol<sup>-1</sup>, respectively, while GFN2 performs exceptionally better, with an AE of only 0.03 kcal mol<sup>-1</sup>. Semi-empirical single-point energy corrections on GFN-FF geometries greatly improve chemical accuracy in GFN1/GFN-FF to an AE of 1.07 kcal mol<sup>-1</sup>, best performance of the series, while single-point energies at DFT-D3<sup>CP</sup> and GFN2 provide the worst performing methods among the composite methods with GFN-FF, with absolute errors higher than 7 kcal mol<sup>-1</sup>.

Molecular complex **7** is similar to **6**, with an additional three ethyl groups that increase molecular complexity and the conformational degrees of freedom of the system. In this case, the DFT-D3<sup>CP</sup> results were taken as a reference, and DFT-D3<sup>CP</sup> single points on GFN-FF structures yield an absolute error of 2.40 kcal mol<sup>-1</sup>. The pure GFN methods exhibit varied performance, with GFN2 yielding the highest AE of 7.23 kcal mol<sup>-1</sup>, whereas GFN1 performs slightly better with an AE of 5.91 mol<sup>-1</sup>. Hybrid approaches involving DFT-D3<sup>CP</sup> corrections on GFN geometries perform considerably better, with DFT-D3<sup>CP</sup>/GFN1 and DFT-D3<sup>CP</sup>/GFN2 achieving AEs of 0.77

and 0.39 kcal mol<sup>-1</sup>, respectively. Pure GFN-FF calculations perform considerably worse than GFN2, with an AE of 13.10 kcal mol<sup>-1</sup>, while composite methods utilizing GFN/GFN-FF geometries show improvements, both for GFN1/GFN-FF (AE = 4.87 kcal mol<sup>-1</sup>) and GFN2/GFN-FF (AE = 5.08 kcal mol<sup>-1</sup>).

In order to study possible cooperativity effects, we studied the trimeric arrangement of all-*syn* 1,3,5-trifluorocyclohexane **8**, where fluorine atoms occupy the axial positions of the cyclohexane and engage in strong CF<sub>ax</sub>-H<sub>ax</sub>C electrostatic interactions, leading to supramolecular assembly. The reference Gibbs free energy of trimerization is  $\Delta G = 8.94$  kcal mol<sup>-1</sup> relative to three separated monomers **2a** at the DLPNO-CCSD(T)/CBS//B3LYP-D3<sup>CP</sup>/def2-TZVP level. The results reveal that DFT-D3<sup>CP</sup> calculations yield an AE of 4.77 kcal mol<sup>-1</sup>, while pure GFN methods exhibit a broad range of performances, with GFN1 yielding an impressive AE of 0.15 kcal mol<sup>-1</sup>, whereas GFN2 shows a higher error of 5.70 kcal mol<sup>-1</sup>, indicating worse agreement with high-level reference calculations. Hybrid approaches incorporating DFT-D3<sup>CP</sup> corrections over GFN structures also demonstrate strong performance, particularly for DFT-D3<sup>CP</sup>/GFN1 and DFT-D3<sup>CP</sup>/GFN2, which achieve AEs of 1.84 and 2.14 kcal mol<sup>-1</sup>, respectively. Interestingly, GFN-FF alone produces a high AE of 6.83 kcal mol<sup>-1</sup>, while the composite methods relying on GFN/GFN-FF geometries display contrasting trends. While GFN1/GFN-FF yields an impressively low AE of 0.31 kcal mol<sup>-1</sup>, GFN2/GFN-FF performs considerably worse with an AE of 7.52 kcal mol<sup>-1</sup>, indicating strong method-dependent variations. Corrections at the DFT level over GFN-FF structures result in an AE of 8.04 kcal mol<sup>-1</sup> for DFT-D3<sup>CP</sup>/GFN-FF, once again suggesting that the great accuracy of GFN1/GFN-FF is most probably due to the cancellation of errors between both methods.

The computational methods were further tested against a more structurally complex system, comprising the dimer of an all-*syn* pentafluorinated bis-cyclohexyl compound with a long alkyl linker (complex **9**). In this case, in addition to the strong electrostatically-driven CF<sub>ax</sub>-H<sub>ax</sub>C interactions, hydrophobic contacts between the alkyl chains may also play a role in complexation. No conformational analysis was undertaken; the alkyl chains were constructed in linear all-*trans* conformations, mimicking the X-ray diffraction data.<sup>24</sup> Due to the system size, the reference value in this case was taken as calculations at the B3LYP-D3<sup>CP</sup>/def2-TZVP theoretical level ( $\Delta G = -1.86$  kcal mol<sup>-1</sup> relative to two separated monomers). Upon optimisation, noticeable intermolecular contacts formed between the H atoms of the alkyl chains (with the H...H contact down to 2.395 Å at the DFT-D3<sup>CP</sup> level, see Fig. S2). Pure GFN methods exhibit a wide range of errors, with GFN1 yielding a relatively high AE of 4.62 kcal mol<sup>-1</sup>, whereas GFN2 performs substantially better, achieving an AE of 2.35 kcal mol<sup>-1</sup>, close to DFT-D3 accuracy. Hybrid approaches incorporating DFT-D3<sup>CP</sup> corrections on GFN geometries display varied performance. DFT-D3<sup>CP</sup>/GFN1 and DFT-D3<sup>CP</sup>/GFN2 yield AEs of 2.20 and 0.45 kcal mol<sup>-1</sup>, respectively, reinforcing the strong performance of GFN2 geometries in this case. GFN-FF alone exhibits a high AE of 7.05 kcal mol<sup>-1</sup>, making it one of the least



reliable methods for this system. However, composite methods using GFN-FF geometries display interesting trends. GFN1//GFN-FF results in an AE of  $5.83 \text{ kcal mol}^{-1}$ , while GFN2//GFN-FF delivers a remarkably low AE of just  $0.32 \text{ kcal mol}^{-1}$ . DFT-D3<sup>CP</sup> corrections on GFN-FF structures yield a very respectable AE of  $0.13 \text{ kcal mol}^{-1}$ .

In general, basis set superposition error corrections proved to be rather important in systems from Group 2 and consistently improved the results when either incorporated directly in geometry optimisation at the DFT-D3<sup>CP</sup> level or even when incorporated only in single-point energy corrections, as in the composite approaches with the GFN methods. The only exception is the binding Gibbs free energy of system **5a**, in which inclusion of CP correction led to a small overestimation (see Fig. S3 for further details). Pure GFN-FF and GFN2 calculations were the worst performers for molecules in Group 2, with MAEs of  $5.12$  and  $2.64 \text{ kcal mol}^{-1}$ , respectively. Even though single-point energy corrections obtained at higher level semi-empirical or DFT-D3 on GFN-FF geometries improved the results against the reference values in some cases, the results are not consistent and appear to arise from random error-cancellation interactions between the methods and proved to be very system-dependent. Surprisingly, the best overall performing method was achieved by composite methods based on GFN2, with an MAE of  $0.82 \text{ kcal mol}^{-1}$ , virtually the same as full DFT-D3<sup>CP</sup> calculations (MAE =  $0.84 \text{ kcal mol}^{-1}$ ). Strong performances were obtained by the hybrid approach DFT-D3<sup>CP</sup>//GFN1, which rendered a similar MAE ( $1.24 \text{ kcal mol}^{-1}$ ).

The results reported herein are consistent with the original GFN publications. GFN1, which was primarily benchmarked against noncovalent complexes stabilized by London dispersion and classical hydrogen bonding,<sup>28</sup> performs worst among the xTB methods. In contrast, GFN2 incorporates higher multipole electrostatics instead of the monopole-based treatment in GFN1,<sup>29</sup> a feature that is particularly relevant for the

electrostatically driven packing of Janus-face cyclohexanes, and consequently shows improved performance in the present set. As for GFN-FF, although noncovalent interactions are treated at the force-field level, its incorporation of flexible atomic charges and tailored lone-pair potentials is expected to bring its performance closer to that of the semiempirical GFN1 method,<sup>30</sup> an outcome that is indeed observed for the Group 2 systems. Nevertheless, none of the xTB methods were parametrized for the assembly of highly polarized aliphatic systems such as Janus cyclohexanes. Yet, their reasonable performance in this context underscores their practical usefulness.

The varying accuracies of GFN methods in predicting the binding Gibbs free energies of Group 2 non-covalent complexes prompted us to evaluate the quality of their corresponding geometries in comparison to those obtained with DFT-D3 and DFT-D3<sup>CP</sup>. As shown in Fig. 4(a), the CP correction has little effect on the predicted inter-ring distances, with both DFT-D3 and DFT-D3<sup>CP</sup> yielding similar equilibrium geometries. Notably, GFN-FF consistently predicts significantly shorter inter-ring contacts, whereas GFN1 and GFN2 geometries tend to be slightly shorter and longer, respectively, compared to DFT-D3<sup>CP</sup> across all systems studied. The PES of the inter-ring distance in dimer **6** (Fig. 4(b)) reveals that the GFN-FF equilibrium geometry corresponds to a much shorter contact and lies in a repulsive region of the DFT-D3<sup>CP</sup> PES, approximately  $4 \text{ kcal mol}^{-1}$  above the minimum. Consequently, applying higher-level single-point energy corrections to GFN-FF geometries does not consistently enhance accuracy and appears to be system-dependent, as it depends on how far the GFN-FF geometry is from the equilibrium geometry at other methods. It is worth noting that GFN-FF geometries most closely resemble those from GFN1, and thus GFN1//GFN-FF consistently outperforms other GFN-FF-based combinations. In contrast, GFN1 and GFN2 geometries, though distinct from DFT results, lie within a shallow region near the PES minimum (within

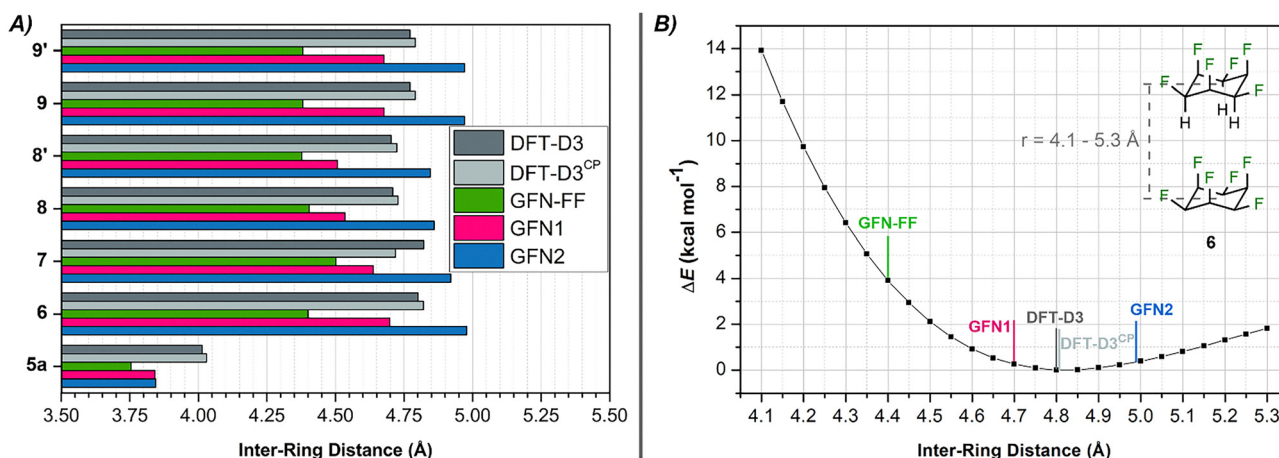


Fig. 4 (A) Distance between centres of mass of endocyclic carbon atoms on interacting cyclohexane units in the equilibrium geometries of systems **5a**, **6**, **7**, **8** and **9**, expressed in angstroms (Å). In systems **8** and **9**, where there are two pairs of interacting cyclohexanes, both distances are displayed as **8** and **8'**, and **9** and **9'**. (B) Relaxed PES, at B3LYP-D3<sup>CP</sup>/def2-TZVP, for the inter-ring distance in dimer **6**, scanned from 4.10 to 5.30 Å in increments of 0.05 Å, highlighting the equilibrium distance obtained using different computational methods. In all cases, the inter-ring distance is defined as the separation between the centers of mass of the endocyclic carbon atoms in each cyclohexane ring.



0.5 kcal mol<sup>-1</sup>). Therefore, single-point energy corrections at higher level DFT-D3 and DFT-D3<sup>CP</sup> over GFN1 and GFN2 geometries generally improve the chemical accuracy of these composite methods.

### Group 3: complexation of ions

Group 3 consists of more strongly bound systems, where the hexafluorocyclohexane **1** interacts with atomic (Li<sup>+</sup>, Na<sup>+</sup>, Mg<sup>2+</sup>, F<sup>-</sup> and Cl<sup>-</sup>) or molecular (NH<sub>4</sub><sup>+</sup>, BF<sub>4</sub><sup>-</sup> and SO<sub>4</sub><sup>2-</sup>) ions. These are already difficult systems and are among the worst performances of the GFN methods reported in their original papers.<sup>28–30</sup> However, on top of that, the Janus-face cyclohexane **1** represents a one-of-a-kind compound, in the sense that its physical-chemical properties are not found in any similar class of compounds. Therefore, considering that the GFN calculations rely on a set of parametrizations derived from large benchmark datasets, it is expected that pure GFN calculations furnish poor electronic energies and thermodynamic parameters for the complexes of **1** with ions. In fact, calculations for molecules of Group 3 using the GFN-FF, GFN1 and GFN2 methods, as well as using the GFN1//GFN-FF and GFN2//GFN-FF composite approaches, rendered binding Gibbs free energies that were either strongly over- or underestimated compared to the reference values previously published in the literature (target  $\Delta G$  values are in the range between -62.86 and -20.00, and between -150.68 and -21.80 kcal mol<sup>-1</sup> for complexes with anions and cations, respectively, see the SI for further details). DFT-D3//GFN-FF and DFT-D3<sup>CP</sup>//GFN-FF also exhibited high absolute deviations compared to the other methods, most probably due to poor complex geometries predicted by the force-field, similarly to what was observed in Group 2. Thus, their results were omitted in the discussion and appended to the SI.

Thus, only results at DFT-D3, DFT-D3<sup>CP</sup> and their corrections to GFN1 and GFN2 geometries will be discussed in the main text. To account for the higher polarizability of anionic species, the def2-TZVPD basis set (with additional diffuse functions) was used on anions during DFT calculations.

The **1**-Li<sup>+</sup> complex, dominated by localized electrostatics, shows a moderate absolute error of 3.31 kcal mol<sup>-1</sup> with the pure DFT-D3<sup>CP</sup> method. Among the composite approaches, DFT-D3<sup>CP</sup>//GFN2 stands out with a similar AE of 4.15 kcal mol<sup>-1</sup>. DFT-D3<sup>CP</sup>//GFN1 performs slightly worse (5.80 kcal mol<sup>-1</sup>). In all cases, neglecting BSSE affords slightly lower MAEs (compare DFT-D3 and DFT-D3<sup>CP</sup> bars in Fig. 5).

The **1**-Na<sup>+</sup> complex, involving a larger monovalent ion with more diffuse interactions, shows a poor baseline performance with DFT-D3<sup>CP</sup> (7.46 kcal mol<sup>-1</sup>). Composite approaches using GFN geometries yield AEs of 11.72 kcal mol<sup>-1</sup> and 13.17 kcal mol<sup>-1</sup> (DFT-D3<sup>CP</sup>//GFN1 and DFT-D3<sup>CP</sup>//GFN2, respectively).

The **1**-Mg<sup>2+</sup> complex, featuring a highly polarizing divalent ion, challenges all methods. DFT-D3 and DFT-D3<sup>CP</sup> remain the most reliable, with AEs of 1.80 and 2.90 kcal mol<sup>-1</sup>, respectively. Note, however, that the binding Gibbs free energy of this complex is the strongest of all, with a target  $\Delta G$  of -150.68 kcal mol<sup>-1</sup>. All composite methods overshoot the

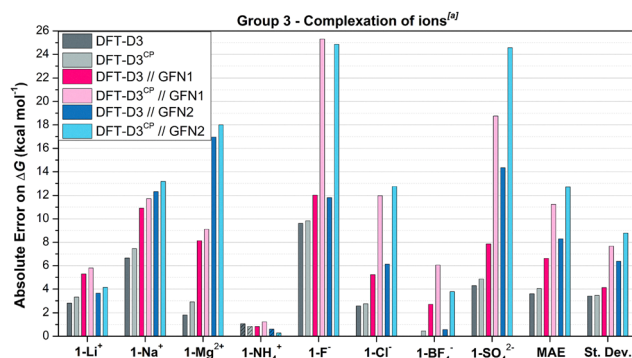


Fig. 5 Absolute (AE) and mean (MAE) errors and standard deviation of absolute errors (st. dev.) on  $\Delta G$  of different computational methods for the complexation of **1** with Li<sup>+</sup>, Na<sup>+</sup>, Mg<sup>2+</sup>, NH<sub>4</sub><sup>+</sup>, F<sup>-</sup>, Cl<sup>-</sup>, BF<sub>4</sub><sup>-</sup> and SO<sub>4</sub><sup>2-</sup>, in kcal mol<sup>-1</sup>. Hashed and solid bars indicate negative and positive errors, respectively. <sup>a</sup>Reference theoretical level: DLPNO-CCSD(T)/CBS//B3LYP-D3<sup>CP</sup>/def2-TZVP.

target significantly, affording AEs of 11.72 kcal mol<sup>-1</sup> and 13.17 kcal mol<sup>-1</sup> at DFT-D3<sup>CP</sup>//GFN1 and DFT-D3<sup>CP</sup>//GFN2, respectively.

For the **1**-NH<sub>4</sub><sup>+</sup> complex, likely governed by directional hydrogen bonding mediated by the F atoms of **1** and electrostatics, the results are generally more favourable. Standard DFT-D3 and DFT-D3<sup>CP</sup> calculations produce AEs of 1.07 and 0.84 kcal mol<sup>-1</sup>, respectively. DFT-D3<sup>CP</sup>//GFN1 and DFT-D3<sup>CP</sup>//GFN2 deliver excellent AEs of 1.22 kcal mol<sup>-1</sup> and 0.27 kcal mol<sup>-1</sup>, respectively. The presence of N-H...F hydrogen bonds likely aligns **1**-NH<sub>4</sub><sup>+</sup> more closely with systems in the training set used during the development of the xTB methods, enhancing the accuracy of force-field and semiempirical geometries and contributing to the strong overall performance of GFN approaches in this case.

As for the anionic complexes, **1**-F<sup>-</sup>, a small and highly basic anion, DFT-D3 yields a significantly absolute deviation of 9.61 kcal mol<sup>-1</sup>, which remains virtually the same at 9.83 kcal mol<sup>-1</sup> upon application of the CP correction. Unexpectedly, CP corrections have a strong worsening effect on GFN-based calculations for this system, where the hybrid methods, DFT-D3<sup>CP</sup>//GFN1 (AE 25.31 kcal mol<sup>-1</sup>) and DFT-D3<sup>CP</sup>//GFN2 (AE 24.86 kcal mol<sup>-1</sup>), show much higher deviations compared to the one non-corrected for BSSE calculations (12.03 and 11.81 kcal mol<sup>-1</sup>, respectively).

For the complex of **1** with chloride in **1**-Cl<sup>-</sup>, a larger and less basic anion compared to fluoride, a moderate deviation profile is observed. DFT-D3 gives an AE of 2.55 kcal mol<sup>-1</sup>, increased to 2.74 kcal mol<sup>-1</sup> upon CP correction, a slight increase in error in this case. Among the hybrid methods, DFT-D3//GFN1 and DFT-D3//GFN2 yield decent results (5.24 and 6.13 kcal mol<sup>-1</sup>, respectively), though CP correction leads to AEs of 11.98 and 12.74 kcal mol<sup>-1</sup> in DFT-D3<sup>CP</sup>//GFN1 and DFT-D3<sup>CP</sup>//GFN2, respectively, again highlighting an important decrease in accuracy upon correction for BSSE.

In complex **1**-BF<sub>4</sub><sup>-</sup>, where **1** interacts with the weakly coordinating anion BF<sub>4</sub><sup>-</sup> primarily through dispersion and diffuse





electrostatics, standard DFT-D3 provides a reliable AE of  $0.01 \text{ kcal mol}^{-1}$ , while DFT-D3<sup>CP</sup> brings it slightly up to  $0.44 \text{ kcal mol}^{-1}$ . Hybrid methods vary widely: DFT-D3//GFN2 delivers an exceptionally low AE of  $0.57 \text{ kcal mol}^{-1}$ , which unexpectedly increases to  $3.78 \text{ kcal mol}^{-1}$  after CP correction. A similar trend occurs with GFN1 geometries (from 2.70 to  $6.05 \text{ kcal mol}^{-1}$  upon CP correction), maintaining good accuracy.

Lastly, the  $1\text{-SO}_4^{2-}$  complex features a highly charged, polarizable anion with multiple oxygen atoms capable of forming non-conventional hydrogen bonds with the polarized C-H<sub>ax</sub> bonds of the Janus cyclohexane. DFT-D3 and DFT-D3<sup>CP</sup> show moderate deviations ( $4.29$  and  $4.85 \text{ kcal mol}^{-1}$ , respectively), with the CP correction marginally decreasing accuracy. Hybrid methods yield mixed results: DFT-D3//GFN1 produces a higher AE ( $7.86 \text{ kcal mol}^{-1}$ ), but CP correction increases the error to  $18.76 \text{ kcal mol}^{-1}$ . Similarly, GFN2-based methods jump from  $14.34$  to  $24.57 \text{ kcal mol}^{-1}$  with CP, revealing an unexpected and substantial error increase.

Among the various approaches, full DFT calculations remain the most reliable and consistent approach for the more challenging Group 3 ionic complexes, yielding mean absolute errors of  $3.60 \text{ kcal mol}^{-1}$  with DFT-D3 and  $4.05 \text{ kcal mol}^{-1}$  with DFT-D3<sup>CP</sup> with low standard deviations ( $3.41\text{--}3.46 \text{ kcal mol}^{-1}$ ). However, the use of GFN1 geometries in hybrid protocols delivers a cost-effective alternative to pure DFT calculations and renders results in relatively good agreement with the reference values, with MAEs of  $6.62 \text{ kcal mol}^{-1}$  for DFT-D3//GFN1 (st. dev. of  $4.13$ ) and  $11.25 \text{ kcal mol}^{-1}$  for DFT-D3<sup>CP</sup>//GFN1 (st. dev. of  $7.66$ )—closely approaching DFT-level accuracy at a significantly reduced computational cost. In contrast, geometries obtained from GFN2 exhibit greater variability and tend to perform worse than those from GFN1, with corresponding MAEs of  $8.30 \text{ kcal mol}^{-1}$  (DFT-D3//GFN2) and  $12.69 \text{ kcal mol}^{-1}$  (DFT-D3<sup>CP</sup>//GFN2).

In terms of methodological performance, the application of counterpoise correction proved beneficial for pure DFT-D3 calculations for  $1\text{-NH}_4^+$ , but consistently led to higher AEs on other systems and on hybrid approaches based on GFN geometries, which prompted us to investigate how the contact distances between the ions and cyclohexane **1** vary across equilibrium geometries obtained at different theoretical levels. Notably, contact distances between **1** and the molecular anions  $\text{BF}_4^-$  and  $\text{SO}_4^{2-}$  calculated at the GFN1 level are significantly shorter than those from DFT-optimised geometries, and slightly shorter than those predicted by GFN2 (Fig. 6). In these cases, the overly close ion–cyclohexane proximity predicted by the GFN methods amplifies the effect of the CP correction, resulting in an apparent overcorrection. This reflects an error-cancellation phenomenon, where the uncorrected values fortuitously agree better with the reference. Nevertheless, because BSSE systematically decreases with increasing basis set size, it is expected that larger basis sets in the single-point calculations would mitigate this behaviour and make CP corrections consistently beneficial.

It is worth noting that GFN1 incorporates the third generation of Grimme's empirical dispersion correction (D3),<sup>28,35</sup>

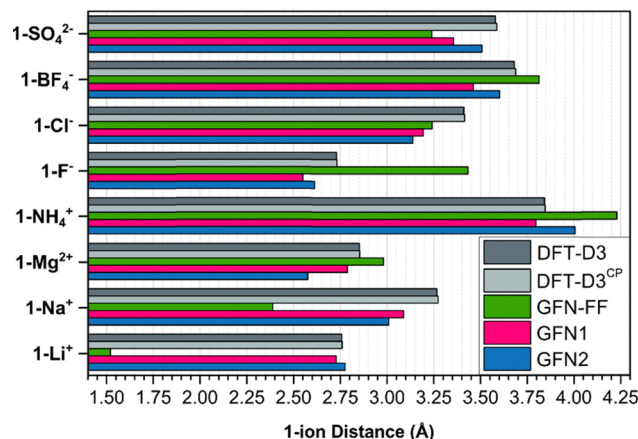


Fig. 6 Distance between the centres of mass of endocyclic carbon atoms from cyclohexane **1** and the ions in the equilibrium geometries of systems  $1\text{-Li}^+$ ,  $1\text{-Na}^+$ ,  $1\text{-Mg}^{2+}$ ,  $1\text{-NH}_4^+$ ,  $1\text{-F}^-$ ,  $1\text{-BF}_4^-$  and  $1\text{-SO}_4^{2-}$ , expressed in angstroms (Å). For the molecular ions, the distance to the central atom was considered.

whereas GFN2 employs the more advanced D4 correction.<sup>29,41</sup> Consequently, for molecular anions—where dispersion interactions are expected to play a more significant role in determining complex geometries and energetics—GFN2 generally offers comparable or improved performance over GFN1 due to its more refined treatment of dispersive contributions. Conversely, for cationic species, the trend reverses: GFN1 geometries are in better agreement with the DFT results than GFN2, even though contact distances are still shorter than those predicted by DFT, indicating that D4 dispersive corrections may overestimate binding contacts to more localized charged species.

GFN-FF geometries, on the other hand, show substantial deviations from DFT geometries. They predict contact distances up to  $\sim 1.25 \text{ Å}$  shorter, in cases such as the  $\text{Li}^+$  complex, and up to  $\sim 0.70 \text{ Å}$  longer, as seen with the hard anion  $\text{F}^-$ . While GFN-FF also includes a modified version of the D4 dispersion correction,<sup>30,41</sup> these discrepancies reflect the intrinsic limitations of force-field approaches, in which atomic ions are modeled as point charges. As a result, electrostatic interactions are often exaggerated, and stereoelectronic effects are poorly handled through parametrization that is highly system dependent. Therefore, for charged complexes, GFN-FF geometries are generally of poor quality, and applying single-point energy corrections at higher theoretical levels is insufficient to yield reliable results (see the SI for further details).

The impact of the optimiser on the resulting geometries was also evaluated. For this purpose, Gaussian16's Berny algorithm was used to optimise all Group 3 systems using the GFN-FF, GFN1, and GFN2 methods. As shown in Fig. S5, the ion–cyclohexane distances remained largely unchanged regardless of whether the optimisations were performed in Gaussian16 or with the xTB optimiser, indicating that the differences discussed above arise from the methods themselves, while the choice of optimisation algorithm has only a minimal impact on the geometry of these systems.





### Group 4: supramolecular chemistry—case study

To evaluate the performance of GFN methods on systems with increased structural complexity, we revisited compound **10**, a molecular platform proposed by Pavan and Delius *et al.*,<sup>27</sup> inspired by the large molecular dipole of all-*cis*-1,2,3,4,5,6-hexafluorocyclohexane **1**. Molecule **10** presents significant conformational complexity, and four representative conformers—**10-I**, **10-II**, **10-III**, and **10-IV**—were previously identified by the authors through a combination of experiment (NMR and DLS) and well-tempered metadynamics simulations at the GAFF level. Previously reported computational results show that conformers **10-I**, **10-II**, and **10-IV** lie within  $\sim 1.2$  and  $\sim 2.6$  kcal mol<sup>-1</sup> from the global minimum **10-III** and are considered “dormant” due to the inaccessibility of the Janus cyclohexane core, which is blocked by different types of interactions: N-H...F hydrogen bonding in **10-I**, hydrophobic interactions with long alkyl chains stabilized by C=O...H-N hydrogen bonds in **10-II**, and  $\pi$ -stacking interactions in **10-IV**. In contrast, the open-state conformer **10-III** is promoted upon the addition of well-defined seed molecules, enabling the stacking of Janus cyclohexanes and triggering kinetically controlled supramolecular aggregation.

Our first objective was to assess whether the GFN methods could reliably identify these four conformers of the core structure of **10** through unbiased conformational sampling. To this end, we performed conformational searches using the workflow detailed in the Materials and Methods section with GFN-FF, GFN1, and GFN2. To enhance sampling of the conformational space around the core, the long alkyl chains appended to the aromatic ring were substituted with methyl groups. The methyl-capped systems are referred to hereafter as **10'**. As shown in Fig. 8(A), representatives of all four conformers were qualitatively reproduced across the three theoretical levels, with the exception of conformer **10-IV** at the GFN-FF level, which predicted a non-conventional C-F... $\pi$  interaction rather than the experimentally observed C-H... $\pi$  interaction. The selected conformers taken forward for further analysis are shown in Fig. 8(A).

For a more quantitative assessment, we fully optimised the selected conformers obtained at each level and subsequently re-optimised them using DFT-D3 (including frequency calculations). We then compared both the relative  $\Delta G$  and structural parameters, namely the root-mean-square deviation (RMSD) of heavy atoms, between the GFN-derived and DFT-D3-optimised structures (Fig. 8(B)). In contrast to the behavior observed for Group 1, where single-point DFT-D3 corrections over GFN geometries improved  $\Delta G$  predictions, such corrections in the case of **10** consistently increased the absolute deviations from the DFT-D3 reference. GFN-FF performed worst among the xTB methods, with a mean absolute error (MAE) of 4.74 kcal mol<sup>-1</sup> and the highest RMSD values across nearly all conformers, averaging 0.74 Å. This structural inaccuracy carried over into composite methods involving GFN-FF, which yielded even larger MAEs: 11.46 and 7.60 kcal mol<sup>-1</sup> for GFN2//GFN-FF and DFT-D3//GFN-FF, respectively. GFN1//GFN-FF was a notable exception, with a reduced MAE of 3.16 kcal mol<sup>-1</sup>.

On the other hand, GFN1 emerged as the most accurate among the xTB methods, with an MAE of only 0.61 kcal mol<sup>-1</sup> and an average RMSD of 0.42 Å, indicating good agreement with DFT-D3 geometries. Surprisingly, the composite method DFT-D3//GFN1 performed worse, with an MAE of 3.56 kcal mol<sup>-1</sup>. GFN2, despite yielding geometries in close agreement with DFT-D3 (average RMSD of 0.41 Å), exhibited a higher MAE of 1.53 kcal mol<sup>-1</sup>. Again, DFT-D3 single-point corrections worsened the results, raising the MAE to 3.63 kcal mol<sup>-1</sup> for DFT-D3//GFN2.

In addition to the AE analysis, we evaluated whether the tested methods could correctly identify conformer **10'-IV** as the global minimum, as reported for conformer **10-IV** in the work of Pavan and Delius *et al.*<sup>27</sup> Another key aspect is the relative stability of all conformers, which was previously reported to lie within approximately 2.6 kcal mol<sup>-1</sup> of the global minimum, with **10-I** and **10-II** being nearly isoenergetic at around 1.2 kcal mol<sup>-1</sup> (Fig. 7). However, these relative Gibbs free energies were obtained for the full system **10** (not the truncated **10'**) and at the GAFF force field level, which may not accurately capture the energetics of this system.<sup>42</sup> Therefore, the discussion here focuses on the general trends in relative conformational stability. As shown in Fig. 9, GFN-FF correctly predicts **10'-IV** as the most stable conformer. However, it severely overestimates the energy of the “open-state” **10'-III**, placing it  $\sim 25$  kcal mol<sup>-1</sup> above the minimum, while **10'-I** and **10'-II** are predicted to be comparably stable, lying within 1.5 kcal mol<sup>-1</sup> of the global minimum. When higher-level single-point energies are computed over GFN-FF geometries, the energetic ordering changes: **10'-IV** is no longer the global minimum, providing evidence that the C-F... $\pi$  interactions predicted by GFN-FF are not as stabilizing at the semi-empirical or DFT-D3 levels. Instead, **10'-I** becomes the most stable conformer in GFN1//GFN-FF, while the “open-state” **10'-III** is the lowest in both GFN2//GFN-FF and DFT-D3//GFN-FF, indicating the poor quality of GFN-FF geometries for this system. GFN1 offers improved predictions, placing **10'-IV** as the global minimum and locating **10'-I** and **10'-II** at 2.8 and 2.3 kcal mol<sup>-1</sup>, respectively. In this case, **10'-III** is the least stable at 7.0 kcal mol<sup>-1</sup>. DFT-D3 single-point calculations over GFN1 geometries (DFT-D3//GFN1) preserve the general energetic trends but increase the relative energy differences. GFN2 nearly reproduces the DFT-D3//GFN1 results, while

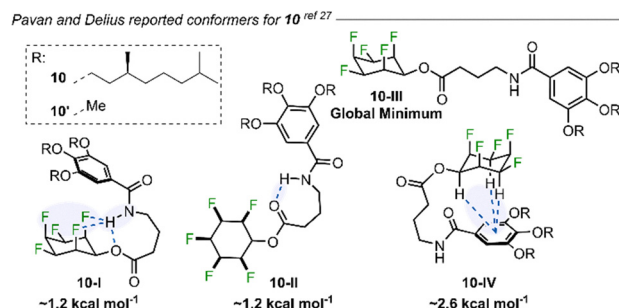
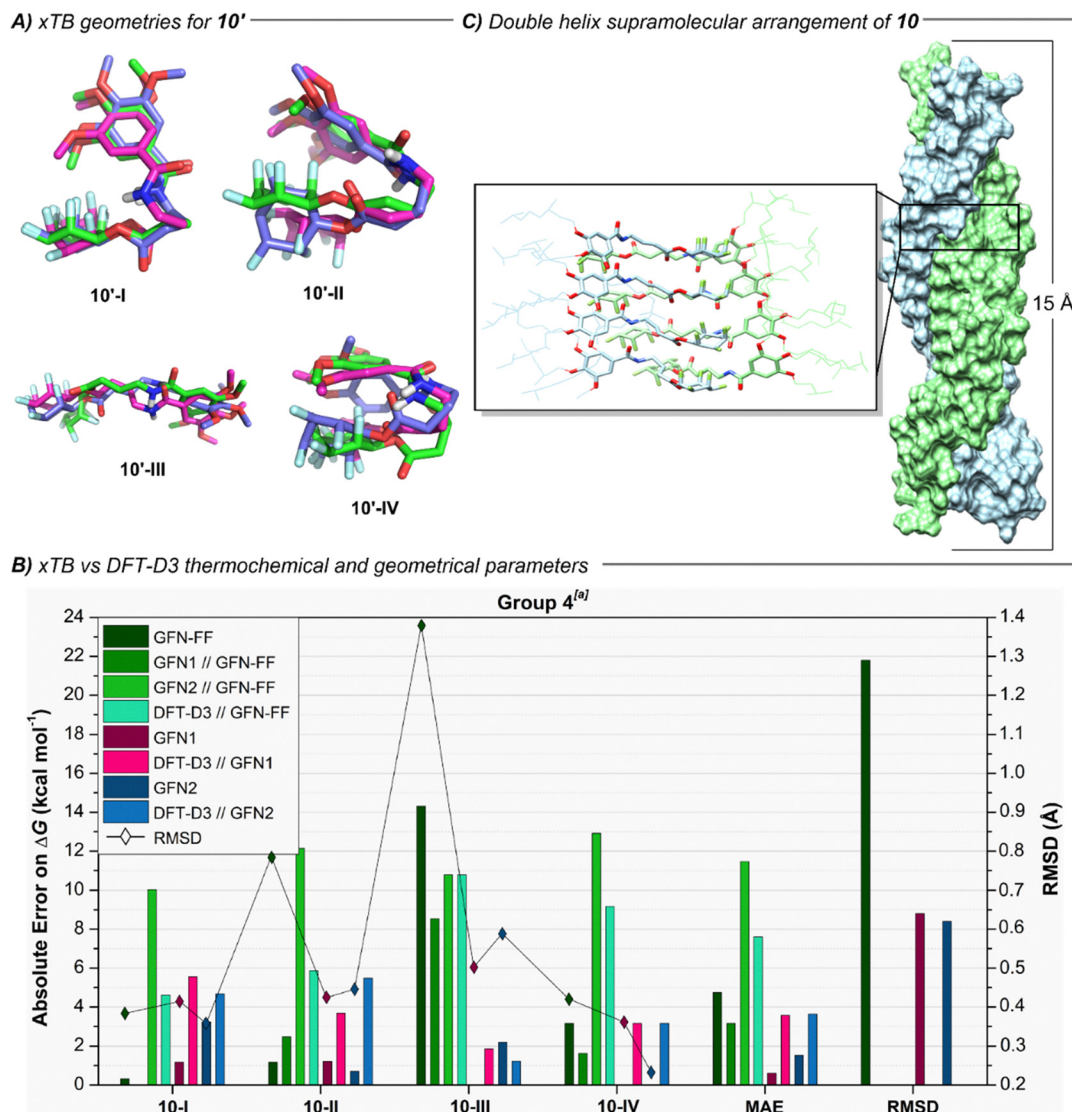


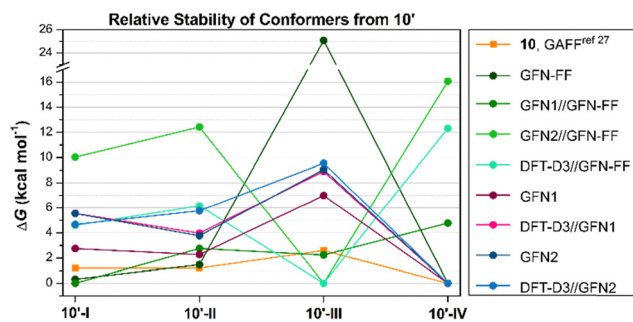
Fig. 7 Representative conformers of **10** and their relative  $\Delta G$ s found by Pavan and Delius *et al.*<sup>27</sup>





**Fig. 8** (A) Superimposed representative conformers of **10** at the GFN-FF (green), GFN1 (pink) and GFN2 (blue) theoretical levels. (B) Absolute (AE) and mean (MAE) errors on  $\Delta G$  and structural parameters (individual RMSD on heavy atoms and average RMSD) of different computational methods for conformers **10'-I**, **10'-II**, **10'-III** and **10'-IV** (Group 4), in kcal mol<sup>-1</sup>. (C) Depiction of the double-helix supramolecular arrangement of the polymer of **10**. <sup>a</sup> Calculations at the B3LYP-D3/def2-TZVP theoretical level were taken as reference for both thermodynamic and structural parameters.

DFT-D3//GFN2 slightly increases the energy gaps further and inverts the order of **10'-I** and **10'-II**.



**Fig. 9** Relative conformational stabilities of conformers **10'-I**, **10'-II**, **10'-III** and **10'-IV** obtained using different theoretical methods, with Pavan and Delius *et al.*<sup>27</sup> results as references.

Pavan and Delius *et al.*<sup>27</sup> also reported the supramolecular arrangement of polymeric **10**, which forms a double-stranded, non-covalently bound complex stabilized by electrostatic attraction between the opposing faces of the Janus cyclohexanes (Fig. 8(C)). As a final stress test, we used a double-stranded helical structure of an oligomeric **10** with a  $\sim 10$  Å pitch consisting of 40 monomeric units (in extended conformations resembling **10'-III**), equilibrated after a 1  $\mu$ s MD simulation at the GAFF level, extracted from the work of Pavel and Delius *et al.*<sup>27</sup> as the starting point for full geometry optimisations using GFN-FF, GFN1, and GFN2. Note that we employed the full system **10** with the bulky side chains for this purpose. Remarkably, all methods produced structures with relatively low RMSDs compared to the MD-equilibrated reference: 2.91, 3.13, and 2.71 Å for GFN-FF, GFN1, and GFN2, respectively. These results suggest that all tested xTB methods are capable of



handling large and complex supramolecular systems, as demonstrated by their reasonable performance on a non-covalent assembly containing 5480 atoms.

## Conclusion

In this study, we demonstrated that the xTB methods, namely the force field GFN-FF and the semiempirical methods GFN1 and GFN2, are valuable tools for investigating conformational preferences and non-covalent interactions in Janus-face fluorinated cyclohexanes. While not always as accurate as full DFT-D3 calculations, their key advantage lies in their significantly lower computational cost, offering up to a 50-fold increase in speed for DFT-D3//GFN methods and up to a 10 000-fold reduction in computational time for pure GFN methods (see Table S10), making them suitable for systems where DFT is computationally prohibitive, while still maintaining a good compromise between chemical accuracy and computational cost across most of the systems studied.

Across different levels of molecular complexity, the performance of the xTB methods varied, as outlined below.

### Group 1 (conformational equilibria)

GFN1 outperformed its successor GFN2 in reproducing relative conformer energies.

However, the best accuracy was achieved using DFT-D3 single-point corrections over GFN-optimised geometries, with DFT-D3//GFN2 matching full DFT-D3 calculations in accuracy and surpassing it in computational efficiency for the tested set.

### Group 2 (non-covalent dimers)

GFN1 and GFN2 emerged as cost-effective alternatives to DFT-D3<sup>CP</sup>. Importantly, applying DFT-D3<sup>CP</sup> single-point corrections in this case greatly improved the results, indicating that hybrid approaches based on GFN1 and GFN2 provide the best balance of accuracy and efficiency for such non-covalent systems.

### Group 3 (complexation of 1 with ions)

The pure GFN methods performed less effectively here, especially in terms of geometries, with closer contacts to ions compared to DFT. Thus, in this case, CP corrections exerted a negative effect on accuracy, with DFT-D3//GFN1 and DFT-D3//GFN2 offering cost-effective alternatives to full DFT-D3 calculations.

### Group 4 (compound 10 and 10')

GFN1 again delivered the most accurate results in both thermodynamic parameters and structural fidelity relative to DFT-D3. Moreover, in the case of the large double-stranded supramolecular polymer of **10** (5480 atoms), all xTB methods produced reasonable geometries, demonstrating their capability in handling highly complex, non-covalent systems.

In summary, xTB methods, especially the semiempirical GFN1 and GFN2, offer a powerful combination of speed and accuracy for the study of Janus cyclohexanes and their

supramolecular assemblies. Their reliability across a range of systems highlights their importance as a tool for exploring new applications and guiding the design of advanced materials based on Janus motifs.

We hope that the insights provided in this work will stimulate further development and application of Janus cyclohexane platforms in supramolecular chemistry.

## Author contributions

Bruno A. Piscelli: investigation and writing – original draft; Tiger Swithenbank-Michel: preliminary investigation; Rodrigo A. Cormanich: funding acquisition, writing – review and editing; David O'Hagan: writing – review and editing; Michael Bühl: conceptualization, supervision, and writing – review and editing.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The research data supporting this publication can be accessed from Data underpinning “Performance of Semiempirical DFT Methods for the Supramolecular Assembly of Janus-Face Cyclohexanes”, University of St Andrews Research Portal, <https://doi.org/10.17630/2b659cbe-0b5c-4731-80b7-3b015398da66>.

Supplementary information: additional computational details, supporting results and optimised cartesian coordinates. See DOI: <https://doi.org/10.1039/d5cp02879e>.

## Acknowledgements

B. A. P. and R. A. C. acknowledge the São Paulo Research Foundation (FAPESP) for a scholarship (#2022/10156-7 and #2023/14064-2) and a young research award (#2018/03910-1), respectively. Calculations were performed on local HPC clusters at St Andrews maintained by Dr. H. Früchtl.

## Notes and references

- 1 S. J. Blanksby and G. B. Ellison, *Acc. Chem. Res.*, 2003, **36**, 255–263.
- 2 D. M. Lemal, *J. Org. Chem.*, 2004, **69**, 1–11.
- 3 D. O'Hagan, *Chem. Soc. Rev.*, 2008, **37**, 308–319.
- 4 D. O'Hagan, *Chem. Rec.*, 2023, **23**, e202300027.
- 5 B. Ameduri, *Chem. – Eur. J.*, 2018, **24**, 18830–18841.
- 6 D. Zhang, Y. Chen, X. Zheng, P. Liu, L. Miao, Y. Lv, Z. Song, L. Gan and M. Liu, *Angew. Chem., Int. Ed.*, 2025, **64**, e202500380.
- 7 J. Zhou, C. Chen, J. Sun, T. R. Fielitz, W. Zhou, D. G. Cahill and P. V. Braun, *Angew. Chem., Int. Ed.*, 2025, **64**, e202503497.



- 8 Y. Zhang, M. Eveno, F. Gallier, A. Lattuati-Derieux, N. Lubin-Germain, M. Camaiti and A. Salvini, *Prog. Org. Coat.*, 2025, **207**, 109375.
- 9 Y. Chen, X. Hua, Y. Xu, B. Gao and X. Li, *Mater. Today Chem.*, 2024, **42**, 102360.
- 10 D. Kimmer, M. Kovarova, M. Yasir, L. Lovecka, J. Cisar, L. Musilova, J. Osicka and V. Sedlářík, *Polym. Adv. Technol.*, 2025, **36**, e70203.
- 11 N. Meng, Y. Hu, Y. Zhang, N. Cheng, Y. Lin, C. Ding, Q. Chen, S. Fu, Z. Li, X. Wang, J. Yu and B. Ding, *Nano-Micro Lett.*, 2025, **17**, 218–233.
- 12 J. W. Washington, T. M. Jenkins, K. Rankin and J. E. Naile, *Environ. Sci. Technol.*, 2015, **49**, 915–923.
- 13 D. Dias, J. Bons, A. Kumar, M. H. Kabir and H. Liang, *Lubricants*, 2024, **12**, 114–154.
- 14 K. R. Miner, H. Clifford, T. Taruscio, M. Potocki, G. Solomon, M. Ritari, I. E. Napper, A. P. Gajurel and P. A. Mayewski, *Sci. Total Environ.*, 2021, **759**, 144421.
- 15 M. H. Russell, W. R. Berti, B. Szostek and R. C. Buck, *Environ. Sci. Technol.*, 2008, **42**, 800–807.
- 16 D. Renfrew and T. W. Pearson, *Environ. Soc.: Adv. Res.*, 2021, **12**, 146–163.
- 17 Z. Wang, I. T. Cousins, M. Scherlinger, R. C. Buck and K. Hungerbühler, *Environ. Int.*, 2014, **70**, 62–75.
- 18 I. T. Cousins, J. C. Dewitt, J. Glüge, G. Goldenman, D. Herzke, R. Lohmann, C. A. Ng, M. Scherlinger and Z. Wang, *Environ. Sci. Process. Impacts*, 2020, **22**, 2307–2312.
- 19 H. Brunn, G. Arnold, W. Körner, G. Rippen, K. G. Steinhäuser and I. Valentin, *Environ. Sci. Eur.*, 2023, **35**, 20–69.
- 20 R. C. Buck, J. Franklin, U. Berger, J. M. Conder, I. T. Cousins, P. De Voogt, A. A. Jensen, K. Kannan, S. A. Mabury and S. P. J. van Leeuwen, *Integr. Environ. Assess. Manage.*, 2011, **7**, 513–541.
- 21 S. Brendel, É. Fetter, C. Staude, L. Vierke and A. Biegel-Engler, *Environ. Sci. Eur.*, 2018, **30**, 9–19.
- 22 N. S. Keddie, A. M. Z. Slawin, T. Lebl, D. Philp and D. O'Hagan, *Nat. Chem.*, 2015, **7**, 483–488.
- 23 N. Santschi and R. Gilmour, *Nat. Chem.*, 2015, **7**, 467–468.
- 24 J. L. Clark, A. Taylor, A. Geddis, R. M. Neyyappadath, B. A. Piscelli, C. Yu, D. B. Cordes, A. M. Z. Slawin, R. A. Cormanich, S. Guldin and D. O'Hagan, *Chem. Sci.*, 2021, **12**, 9712–9719.
- 25 C. Yu, B. A. Piscelli, N. Al Maharik, D. B. Cordes, A. M. Z. Slawin, R. A. Cormanich and D. O'Hagan, *Chem. Commun.*, 2022, **58**, 12855–12858.
- 26 T. J. Poskin, B. A. Piscelli, K. Yoshida, D. B. Cordes, A. M. Z. Slawin, R. A. Cormanich, S. Yamada and D. O'Hagan, *Chem. Commun.*, 2022, **58**, 7968–7971.
- 27 O. Shyshov, S. V. Haridas, L. Pesce, H. Qi, A. Gardin, D. Bochicchio, U. Kaiser, G. M. Pavan and M. von Delius, *Nat. Commun.*, 2021, **12**, 3134.
- 28 S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.
- 29 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
- 30 S. Spicher and S. Grimme, *Angew. Chem., Int. Ed.*, 2020, **59**, 15665–15673.
- 31 C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, e1493.
- 32 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- 33 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- 34 D. Rappoport and F. Furche, *J. Chem. Phys.*, 2010, **133**, 134105.
- 35 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 36 M. J. Frisch *et al.*, *Gaussian 16, Revision C.01*, Gaussian, Inc., Wallingford CT, 2016. Full reference appended to SI.
- 37 S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 2847–2862.
- 38 P. Pracht, F. Bohle and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 7169–7192.
- 39 S. Gallarati, P. Dingwall, J. A. Fuentes, M. Bühl and M. L. Clarke, *Organomet.*, 2020, **39**, 4544–4556.
- 40 R. A. Cormanich, N. S. Keddie, R. Rittner, D. O'Hagan and M. Bühl, *Phys. Chem. Chem. Phys.*, 2015, **17**, 29475–29478.
- 41 E. Caldeweyher, J. M. Mewes, S. Ehlert and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 8499–8512.
- 42 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

