PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2023, **25**, 24031

Received 22nd June 2023,
Accepted 20th July 2023

DOI: 10.1039/d3cp02903d

rsc.li/pccp

When catchers meet – a computational study on the dimerization of the Buckycatcher†

Filipe Menezes * and Grzegorz Maria Popowicz*

We study the dimerization of the buckycatcher in gas phase and in toluene. We created an extensive library of 36 different complexes, which were characterized at semi-empirical and DFT levels. Semi-empirical geometries and dimerization energies compare well against reference data or Density Functional Theory calculations we performed. Born–Oppenheimer molecular dynamics was used to understand what happens when two molecules of the buckycatcher meet, allowing us to infer on the lack of kinetic barriers when dimers form. Thermodynamically, it is possible that room temperature solutions contain dimerized buckycatcher. Using a very simple exchange model, it is shown, however, that dimerization cannot compete thermodynamically against complexation with fullerenes, which accounts for experimental observations.

Introduction

The buckycatcher is a derivative of corannulene, in which two units of the aromatic system are bound together by a tetraphenylene moiety.[1] Because of its naturally bent structure, a concave–convex surface is created, which pairs perfectly with fullerenes to form a stable adduct in solution. The existence of two corannulene units per molecule of the buckycatcher is important to yield an interaction strong enough from the enthalpy point of view. Theoretical calculations show that corannulene itself (1 unit) cannot bind fullerenes in the gas phase,[2] even with considerably strong enthalpies of binding.

Despite over a decade and a half of existence, many are still the open questions about this fascinating chemical system. To the best of our knowledge, studies focused on conformational equilibria,[1,3–5] binding to fullerenes[1,2,4–11] and to less extent the characterization of a few dimers.[3,4] The interactions of the buckycatcher with small molecules in gas and in solution have only recently been tackled by us, in another paper of this series.[12]

Dimerization remains largely undescribed. Denis and Iribarne[4] evaluated the Gibbs free energies of formation for three of a large set of possible dimers, and they concluded that the dimerization process was thermodynamically favourable. Calculated Gibbs free energies were furthermore lower than for the formation of adducts with fullerenes. This suggested that dimerization is thermodynamically favoured over complexation with fullerenes.[4] However, in solutions of the buckycatcher with fullerenes, there is barely any dimer observed.[8] In fact, there is no competition between processes, which suggests that kinetics hinder dimerization. We found, in previous work, that there is virtually no barrier for fullerenes to form adducts with the catcher.[10] This rightfully raises the question of whether kinetics or thermodynamics hinder dimerization, since the formation of catcher's dimers should be ruled by similar principles.

The current work aims primarily at contributing to a better understanding of the buckycatcher's dimerization behaviour, identifying the root for lack of dimerization after a fullerene is mixed with a solution of the catcher: is this a thermodynamic hindrance or should kinetics really be invoked? With this in view, we start by studying the thermodynamics of formation for 19 dimers of the corannulene pincer at several semi-empirical levels and the dynamics for the formation of a dimer in the dielectric of toluene. The significance of our semi-empirical calculations is inferred from density functional theory (DFT) calculations on the semi-empirical optimized structures. These results are analysed in depth to shed further lights on how to further improve currently existing semi-empirical methods. Most manuscripts dealing with the accuracy of quantum chemical methods, semi-empirical included, use large and varied datasets. Though useful for overall trends, this is blinding when and where potential problems occur. Furthermore, benefiting from large statistical analyses does not give any information on how to improve the current collection of different methods. This can only be performed with in-depth analysis of a few critical and exclusive outliers, or challenging molecular systems, like the buckycatcher. With semi-empirical and higher-quality DFT data at hands, we calculate Gibbs free energies of dimerization in gas and in toluene using 2 solvation models. Contrary to other studies in the literature

---

Institute of Structural Biology, Helmholtz Munich, Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany. E-mail: filipe.menezes@helmholtz-munich.de, grzegorz.popowicz@helmholtz-munich.de

† Electronic supplementary information (ESI) available. See DOI: **https://doi.org/10.1039/d3cp02903d**

where dimerization reactions are treated independently, we consider conformational effects in our thermodynamic calculations. Finally, the present study will serve as additional benchmarking for the methods and capabilities of ULYSSES.[13]

Results

Systems studied and nomenclature

The main conformations in which the corannulene pincer might exist and their respective nomenclature was already established in other work.[10] This is repeated in Fig. 1 for ease of reading of the present document. Here, we investigate the dimerization of the buckycatcher by being as extensive as possible. Therefore, 19 different species were optimized, involving combinations of each conformer with itself as well as with the other conformers (Fig. 1). Note that there is only experimental evidence for the existence of a few of these dimers,[3] namely species 1, 5A and 7A. A species like 8A was also observed in crystal, and the complexes 6 and 15 were considered in other theoretical work.[4] Optimized structures and their thermodynamic data will be available in **https://gitlab.com/siriius/buckycatcherrevisited.git** upon publication of this manuscript.

Choice of reference DFT method

A large study on several dimeric forms of the buckycatcher is uncharted terrain. As we are also interested in understanding limitations of current state-of-the-art semi-empirical methods for π-stacked systems, good reference points are of importance.[14] Most studies on non-covalent binding make use of extremely large basis sets or modified versions thereof.[11] This can be impractical when tackling even larger molecular systems, like those of biological interest, as it demands significant computational power. Using non-standard basis sets might be overwhelming for less experienced users. We therefore decided to take a different approach and identify instead several density functionals that can reproduce reference binding data using smaller, and more computationally affordable basis sets. Technically, this would require rerunning calculations on S30L. That would defocus the scope of the manuscript and introduce benchmarks of little relevance for the problem we are interested in studying here, the dimerization of the buckycatcher. Furthermore, the conclusions we take from our study agree quite well with the results of Sure and Grimme.[11] Given this, we benchmark density functionals exclusively against the buckycatcher-C_{60} complex.

Table 1 compiles binding energies for the buckycatcher-C_{60} system according to several density functional methods.

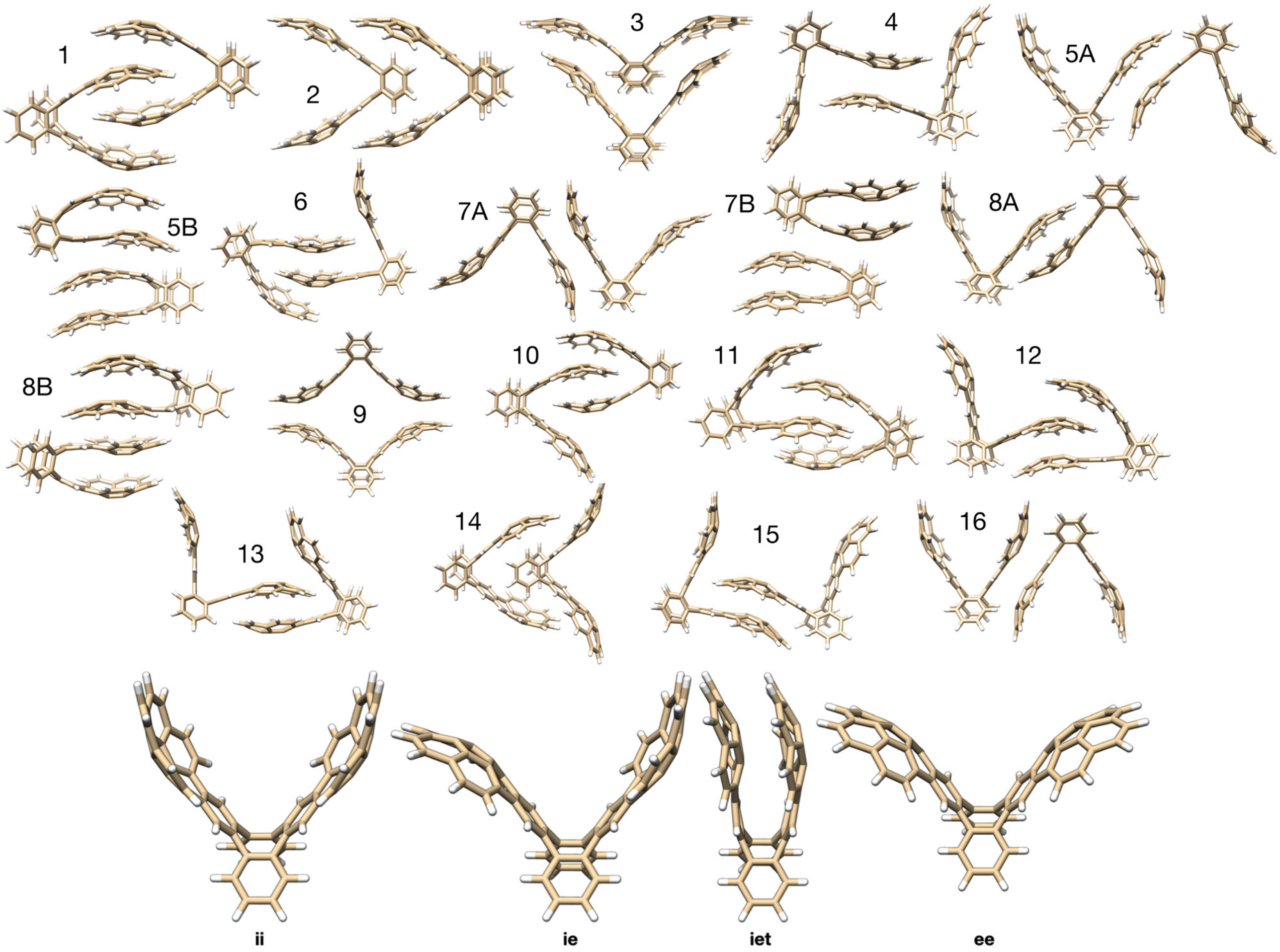

Fig. 1 Dimers of the corannulene pincers considered in this work. Below we present the monomers and the naming convention followed.

Table 1 Binding energies for the buckycatcher-fullerene according to several DFT/basis set combinations. Energies in kcal mol^{-1}. Geometries from S30L.[11] Values in italic are methods within 1.0 kcal mol^{-1} of the reference binding energy. Values in bold are predictions off by at least 5.0 kcal mol^{-1}

	6-31G*	def2-SVP	def2-TZVP	cc-pVTZ
r^2SCAN-3c	**35.4**	**39.4**	30.8	30.9
PBE-D3BJ	**38.4**	**39.6**	31.1	31.2
PBE-D3BJ (ATM)	**35.1**	**36.3**	*27.8*	*27.9*
PW6B95-D3BJ	**43.4**	**46.4**	**35.7**	**35.5**
PW6B95-D3BJ (ATM)	**40.1**	**43.1**	32.4	32.2
M06-2X	31.7	**36.9**	*27.4*	*28.5*
Ref. 11	28.4 ± 0.6			

We chose methods known to be robust for this type of system when combined with large basis sets,[6,11] and we decided to test also r^2SCAN-3c. Note that ultimately we wish to understand the thermodynamic properties of the dimerization of the buckycatcher, not to benchmark density functionals. Hence our choice of methods we knew *a priori* to be robust for these systems. In this section, however, we wish to determine a good and simple combination between density functional and basis set. Therefore, a single conformer is used for the calculations.

Use of double-zeta basis sets (6-31G*, def2-SVP) leads to binding energies significantly far from the 28.4 ± 0.6 kcal mol^{-1} given as reference by Sure and Grimme (see ESI,† Section S2 for discussion on choice of reference).[11] The only method coming reasonably close is M06-2X with the Pople basis set. It is well described in the literature that this functional lacks long-range attractive interactions.[15] The fact that M06-2X is the best performing density functional with double-zeta basis sets indicates lack of repulsion in the double-zeta data. The situation changes drastically with triple-zeta basis sets. Here, PBE with full dispersion corrections, *i.e.*, with the 3-body or Axilrod–Teller–Muto (ATM) terms included, and M06-2X are spot-on with respect to the reference data. We are particularly fond of using M06-2X to verify other data because this functional includes dispersion in its parametrization. In fact, it is recommended not to use dispersion corrections with M06-2X, since it worsens predictions.[16,17] Including dispersion interactions *via* parametrization has the disadvantage of not allowing the method to correctly reproduce long-distance behavior.[17] Still, in studies involving extremely large molecular systems, M06-2X predicted binding energies within 10% of high-level reference data, remaining one of the best-performing density functional methods.[17] We also believe that M06-2X serves as a good orthogonal check, because it is the only method used here that does not make use of Grimme's dispersion corrections.[18–20] This allows us (1) to estimate the magnitude of deviations in the calculations and (2) infer on which might be the faults exhibited by other methods. Despite M06-2X being the best performing method from the data in Table 1, due to its known limitations,[15,17] we choose PBE-D3BJ as our reference method for further calculations. This density functional, together with other basis sets,[11] provided also good results for systems dominated by π dispersion, validating the choice.

r^2SCAN-3c is also quite performant for this system, though deviations are slightly larger. We note that 3-body dispersion corrections should not be considered additionally because the method uses D4.[20] Comparing with other functionals, it appears that r^2SCAN-3c lacks repulsion with triple-zeta basis sets (the 3c correction includes corrections to the basis set superposition error). Nevertheless, the speed with which we performed all r^2SCAN-3c calculations renders this a very attractive alternative. Differences in binding energies using PM6-D3H4 geometries are minimal, especially when used in conjugation with r^2SCAN-3c. PBE-D3BJ (ATM) data shows slightly larger deviations (0.7 kcal mol^{-1}), which should result from fine differences in the binding pose (Fig. 2). With GFN2-xTB geometries there are significant deviations though. Binding is impaired by up to 8 kcal mol^{-1}. Fig. 2 compares the binding pose for the C_{60}@catcher complex according to different methods. The GFN2-xTB geometry has the corannulene units closer than in any other case and the fullerene sits deeper in the catcher's cavity. This means that the equilibrium geometry has catcher and C_{60} closer than in other method, evidencing lack of repulsion. This agrees with the analysis of binding energies. It also agrees with the fact that differences in the structure are quite fine (overlap-repulsion should exhibit an exponential decay with respect to interatomic distance).[21]

To estimate errors in the semi-empirical data we reran DFT calculations using the semi-empirical optimized structures instead of S30L ones. Table 2 summarizes the respective results.

We also performed calculations using quadruple-zeta basis sets, primarily to investigate how sensitive the results are with respect to basis set size. For PW6B95+D3BJ/cc-pvQZ a difference of 0.2 kcal mol^{-1} was recorded while for r^2SCAN-3c/cc-pvQZ 0.5 kcal mol^{-1} the binding energy was 0.5 kcal mol^{-1} lower. When possible, we also compared our triple-zeta data against the results of Sure and Grimme.[11] Similar considerations apply. Since

Table 2 Binding energies for the buckycatcher-fullerene according to some density functionals using the def2-TZVP basis set. Energies in kcal mol^{-1}. Geometries obtained by semi-empirical optimization

Geometry	PBE-D3BJ	PBE-D3BJ (ATM)	r^2SCAN-3c
GFN2-xTB	24.0	20.4	23.7
PM6-D3H4	31.7	28.5	30.9

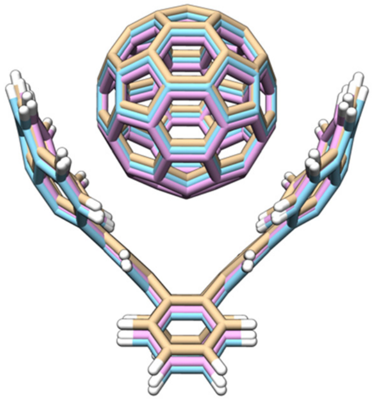

Fig. 2 Overlay of structures of the buckycatcher-C_{60} complex: brown S30L; blue PM6-D3H4; pink GFN2-xTB.

basis set superposition errors are quite reduced at quadruple-zeta level[11] and our triple-zeta calculations are identical to the former, we refrained from estimating basis set superposition errors.

Much like Garcia and Szalewicz observed in their calculations,[22] our study shows that deformation energies have a small contribution for binding that must be accounted for. It is also noteworthy to put their dispersion corrected Hartree Fock method in perspective with our calculations.

Dimerization energies

Fig. 3 shows binding or dimerization energies ($E_{\text{dimer}} = E_{\text{bind}} = E_{\text{monomerA}} + E_{\text{monomerB}} - E_{\text{dimer}}$) for the buckycatcher's dimers according to several methods. We optimized the dimers using several semi-empirical methods and for the case of PM6-D3H4 and GFN2-xTB geometries, energies were re-evaluated using several DFT methods. Binding energies are also given in the ESI.† It may be easily observed that even if the corannulene pincer deforms to dimerize, from the energy point of view, the process is always favourable.

To great extent, the PM6-D3H4 curve for dimerization energies is parallel to all DFT calculations evaluated on the same geometries (DFT-P). Exceptions are dimers 2 and 3, for which the semi-empirical method expects similar dimerization energies, though the latter is clearly favoured by all density functionals tested. Also, the relative placement of dimers 4, 5A, and 5B shows some disagreement. Here, PM6-D3H4 follows the behaviour of M06-2X, which contrasts that of PBE and r^2SCAN. Lastly, contrary to all DFT calculations, the formation of 8A is overly favoured. Despite these 3 incidents, the agreement with the DFT curves is remarkably good, with the standard deviation of deviations in dimerization energy (σ) around 2.0–3.0 kcal mol^{-1} (Table 3). Removing the largest outlier, 8A, reduces σ to 1.5–2.5 kcal mol^{-1}. Means Deviations (MD) in dimerization energy improve by similar factors. This strengthens the parallelism between PM6-D3H4 and DFT curves. Overall, we conclude that, with very few exceptions, PM6-D3H4 replicates the DFT dimerization energies. Note that the standard deviation thus far discussed, σ, reflects the error in relative binding energies, or the ability of PM6-D3H4 to predict the correct ordering of dimerization energies. The average deviation, MD, reflects the overall error. The overall error of PM6-D3H4 with respect to PBE-D3BJ (ATM) lies at about 3.5 kcal mol^{-1} (or $\pm$1.8 kcal mol^{-1}), while the ability to predict relative dimerization energies is *ca.* 2.5 kcal mol^{-1}.

The situation seems significantly less favourable for GFN2-xTB, and indeed the statistics reveal larger MD and σ (Table 2). Compared against DFT energies evaluated on the GFN2-xTB geometries (DFT-G), MDs may be as large as 13 kcal mol^{-1}. σs go about 4–5 kcal mol^{-1}. This places however GFN2-xTB at significant disadvantage, as we compare against DFT energies evaluated on geometries lacking repulsion (see Benchmark Studies). Comparing GFN2-xTB binding energies against DFT-P ones leads to significantly better agreement (GFN2(*) set in Table 2). Considering that the whole curve is shifted, three main outliers arise, namely dimers 2, 8B, and 14. Removing these from the statistics leads the conclusion that GFN2-xTB is also able to predict the relative dimerization energies, though less systematically than PM6-D3H4. In fact, the statistical evaluator σ with respect to PBE-D3BJ (ATM) without the 3 outliers becomes lower than 2 kcal mol^{-1}. Still, the overall accuracy is, in this case, quite poor, as energies are predicted within 7.7 kcal mol^{-1}. The ESI,† shows further data related to this study. Though unsuitable for evaluating binding energies on dimers of the buckycatcher, GFN2-xTB can replicate the energy surface. Based on the outliers identified, potential problems may arise in structures involving T-stacks. These are expected to be less stabilized, thus with shorter lifetimes/reduced populations.

We also calculated dimerization energies using other variants of the selected semi-empirical methods. In agreement with previous studies of ours,[10,12] PM6-D3H +'s lack of repulsive interactions leads to even further accentuated overbinding. Exceptions are dimers 7B and 9. Including the Axilrod–Teller–Muto (ATM) contribution to the PM6-D3H+ data (with $s_9 = 1$) would decrease the first dimerization energy by 6.5 kcal mol^{-1}. The resulting value, 39.3 kcal mol^{-1}, is the closest to PBE-D3BJ (ATM) data. Note that there are fine differences in the PM6-D3H4 and PM6-D3H+ geometries, which indicates that including the ATM terms during geometry optimization could lead to further decreasing of the respective dimerization energy. Similar conclusions apply for dimers 6 (corrected to 27.8 kcal mol^{-1}) and 15 (corrected to 22.4 kcal mol^{-1}). Including 3-body dispersion brings PM6-D3H+ data closer to values in the literature,[4] as well as our double-zeta results.

Calculations were also performed for GFN1-xTB and GFN0-xTB. From the plot in Fig. 3D, we read indirectly that shifts, MD, are larger for these two methods than they are for GFN2-xTB. Interestingly however, GFN1-xTB is the best performing method in this family with respect to reproducing the energy curve of the catcher's dimerization at the PBE-D3BJ (ATM) level: $\sigma_{\text{GFN1-xTB}} = 3.0$; $\sigma_{\text{GFN0-xTB}} = 4.1$. Here, dimer 1 is the largest outlier, with dimerization energy of 47.4 kcal mol^{-1}. This reduces to 40.2 when considering a shift of the magnitude of MD (7.2 kcal mol^{-1}).

At this point it is important to stress that our goal with the previous analysis is not to propose using shifts on calculated energies in semi-empirical methods. This would defy the very nature of semi-empirical method development. We wish however to evaluate the applicability and validity of these methods in several situations. While, from our data, we feel that PM6-D3H4 offers a great choice in replicating the energy surface and, within certain errors, in predicting dimerization energies, our study indicates that the xTB family of methods can only yield reasonable relative values ($\Delta\Delta E_{\text{bind}}$) for this specific system. Thus, though not accurate enough for calculating thermodynamic data, xTB methods should be suitable for enhanced sampling studies. This is particularly attractive, since then we benefit from using the ALPB solvation model.[23] We note that similar observations on semi-empirical methods have been made previously by Acevedo and Jorgensen.[24,25]

Thermodynamics in gas and in solution

Calculated contributions from dimerization entropies ($T\Delta S_{\text{bind}}$) are all between -15 and -19 kcal mol^{-1}, which are quite systematic for all complexes (*cf.* ESI,† for a plot of this data).

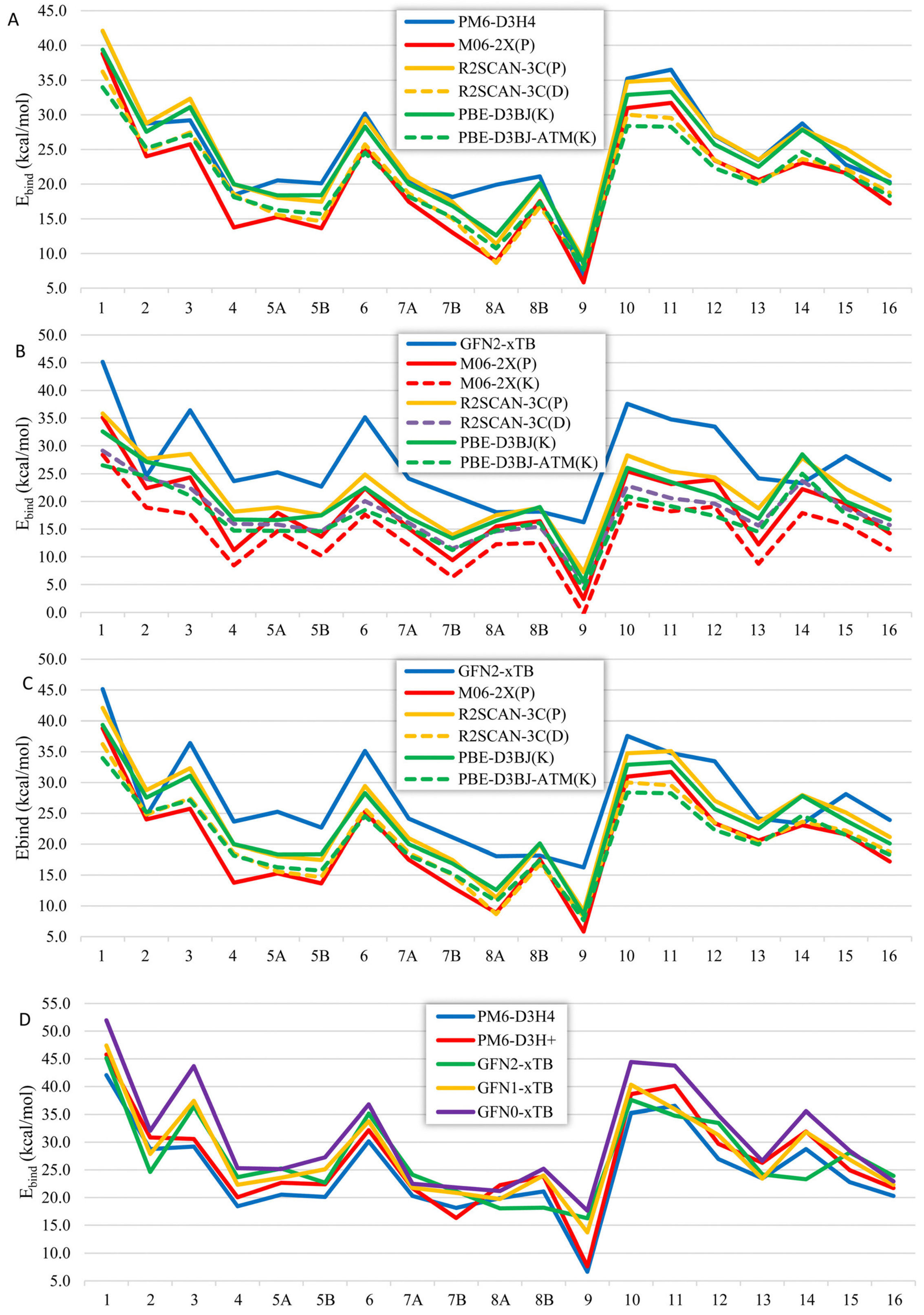

Fig. 3 Binding energies for the pincers' dimers in kcal mol^{-1}. (A) data using PM6-D3H4 optimized geometries; (B) data using GFN2-xTB optimized geometries; (C) comparison between GFN2-xTB data and DFT dimerization energies using however PM6-D3H4 geometries; (D) semi-empirical dimerization energies. P stands for Pople basis set, *i.e.*, 6-31G*. D is the Dunning basis set cc-pVTZ and K the Karlsruhe basis set def2-TZVP.

Table 3 Mean deviation (MD) and respective standard deviations (σ) between semi-empirical and DFT binding energies. Data in kcal mol^{-1} and given in the form MD ± σ. Evaluation is done on the respective semi-empirical optimized geometries. P stands for Pople basis set (6-31G*), K for Karlsruhe (def2-TZVP) and D for Dunning (cc-pVTZ). GFN2(*) compares GFN2-xTB dimerization energies against DFT ones evaluated however on the PM6-D3H4 geometries (DFT-P)

Method	Basis	PM6-D3H4	GFN2-xTB	GFN2(*)
M06-2X	P	4.3 ± 2.2	8.9 ± 4.1	6.7 ± 3.5
	K	—	13.0 ± 4.4	—
r^2SCAN-3c	P	0.4 ± 2.5	5.4 ± 4.3	2.9 ± 3.5
	D	3.7 ± 2.9	9.2 ± 4.9	6.2 ± 3.2
PBE-D3BJ	K	1.2 ± 2.1	7.0 ± 5.2	3.6 ± 3.5
PBE-D3BJ (ATM)	K	4.0 ± 2.7	10.0 ± 5.9	6.4 ± 3.6

The conformational landscape in the dimer space is consequently predominantly enthalpy driven. Gas phase Gibbs free energies follow the binding energies (Fig. 4), and we conclude, irrespective of method, that dimer 1 is the most favourable species to form in a hypothetical gas phase. Note that though it seems that the overlap between binding Gibbs free energies is lower than for the respective energies, this cannot be the case since thermodynamic corrections were systematically applied to all dimers using the semi-empirical data. In the case of DFT results, since these are evaluated on the PM6-D3H4 geometries, the thermodynamic corrections used were calculated using PM6-D3H4 on PM6-D3H4 optimized geometries. The apparent increased discrepancy arises because the statistical mechanical corrections differ from species to species. Similar considerations apply for solvation contributions, which are discussed next.

Solvation in toluene (Fig. 5) increases Gibbs free energies for the formation of complexes. In the case of COSMO, the increase is about 0.5–7 kcal mol^{-1}. ALPB is significantly more penalizing, with corrections amounting to 5–20 kcal mol^{-1}. Consequently, equilibria are less shifted to the direction of dimerization.

The COSMO corrections are indeed so soft that conclusions taken before for gas phase data remain, for practical purposes, unchanged. The respective COSMO and gas phase curves at dispersion corrected PM6 methods are quite similar, and strictly based on the COSMO model one would conclude that much does not qualitatively change with respect to the gas phase. On the other hand, quantitatively speaking, the dimerization energies of other species come closer to that for dimer 1, which indicates a slightly stronger contribution of other species to the equilibrium. The estimated effects of solvation according to ALPB are stronger than COSMO's but also not as uniform. Effectively, the range of Gibbs free energies of binding contracts by 10–15 kcal mol^{-1} when using ALPB solvation (difference between minima and maxima). Particularly interesting is the fact that most equilibria become non-spontaneous with this correction. This is in appreciable contrast to the picture built with COSMO. Compared to the DFT data, PM6-D3H4 resembles the PBE-D3BJ results (without 3-body corrections), where only 6 cases potentially show binding. Unlike the picture built with COSMO and those methods, PM6-D3H4/ALPB and PBE-D3BJ/ALPB expect more competition between different dimerization reactions, which seems more plausible from a chemical viewpoint. Also interesting are the results for PBE-D3BJ (ATM), where only one dimerization process is expected to be thermodynamically spontaneous. This clearly evidences the role of 3-body dispersion on the results.

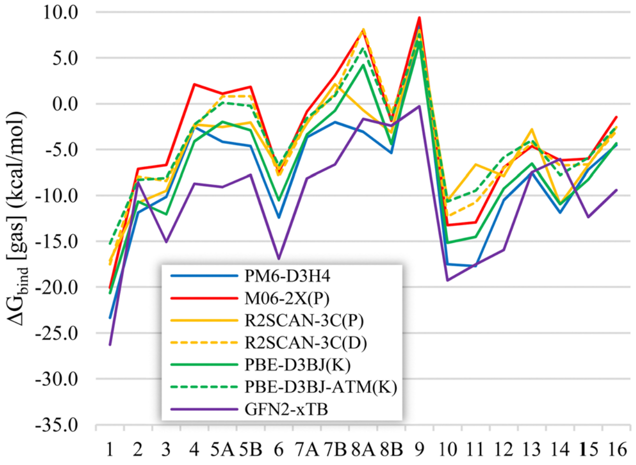

Fig. 4 Gibbs free energies for the dimerization of the buckycatcher in gas phase at 300 K. Data given in units of kcal mol^{-1}.

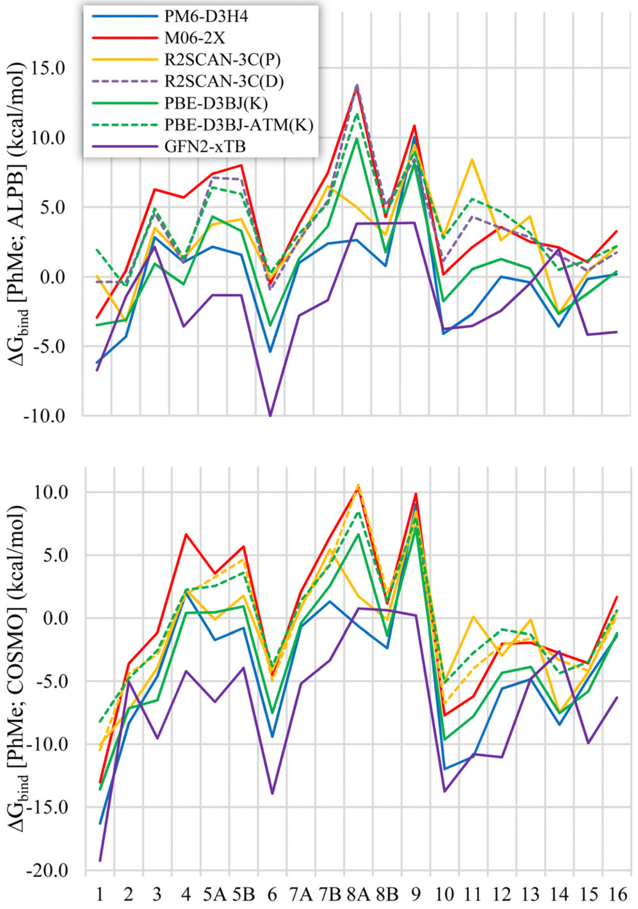

Fig. 5 Gibbs free energies for dimerization in toluene at 300 K. Data given in units of kcal mol^{-1}. Upper plot, data obtained using the ALPB contribution. Lower plot obtained using COSMO.

Table 4 Global thermodynamics of dimerization for the buckycatcher using 19 conformers on the product space. ΔG_{simple} is the Gibbs free energy in solution calculated considering only average of the chemical potentials. $T\Delta S_{\text{conf}}$ is the respective conformational terms and $\Delta G_{\text{full}} = \Delta G_{\text{simple}} - T\Delta S_{\text{conf}}$

		ΔG_{simple}	$T\Delta S_{\text{conf}}$	ΔG_{full}
PM6-D3H4		−6.1	−0.6	−5.4
M06-2X	6-31G*	−2.8	−0.8	−2.0
r^2SCAN-3c	6-31G*	−5.4	−0.5	−4.9
	cc-pVTZ	0.4	0.0	0.4
PBE-D3BJ	def2-TZVP	−3.2	−0.3	−2.9
PBE-D3BJ (ATM)	def2-TZVP	−0.4	−0.6	0.2

We note that solvation contributions with ALPB were obtained for the respective geometries. This means that the ALPB corrections to GFN2-xTB were obtained using GFN2-xTB geometries, whereas ALPB corrections for PM6-D3H4 and DFT data were obtained using PM6-D3H4 optimized geometries. Using the ALPB corrections with GFN2-xTB geometries on the DFT data, to infer on the stability of the results, showed no effective change.

Though the analysis of some specific dimerization processes showed that in toluene most processes become thermodynamically less favourable, none of the systems above studied reflects the actual experimental setup for dimerization. This is because the conformational complexity of the system was not duly considered. In actual calculations, such effects must be accounted for in order to have results comparable to experimental data. Table 4 compiles thermodynamic data obtained for this system considering

$$\Delta G_{\text{full}} = \Delta G_{\text{simple}} - T\Delta S_{\text{conf}} \tag{1}$$

In the previous expression, ΔG_{simple} is calculated as the average Gibbs free energy of all conformers, while $T\Delta S_{\text{conf}}$ is obtained from the Shannon entropy term. Weights used to calculate the last term were obtained from the relative Gibbs free energies in solution (not relative energies). Consequently, ΔG_{full} accounts for conformational effects onto the Gibbs free energy. While semi-empirical and DFT calculations with double-zeta basis sets expect dimerization to still be thermodynamically favoured, DFT data obtained using triple-zeta basis sets predict non-spontaneity. We stress the effect of entering with conformational entropy for PBE-D3BJ (ATM) results. An excel calculation file is provided as ESI,† as we believe this might be useful for other researchers.

Dynamics of dimerization

Having verified the thermodynamic (non-)spontaneity for the dimerization of the buckycatcher, we decided to investigate the dynamics for a pair of ii conformers. Starting from two such molecules at large enough distance (minimum 7.35 Å) we let them come together and interact in a toluene environment. The results may be found in Fig. 6. Note that since GFN2-xTB

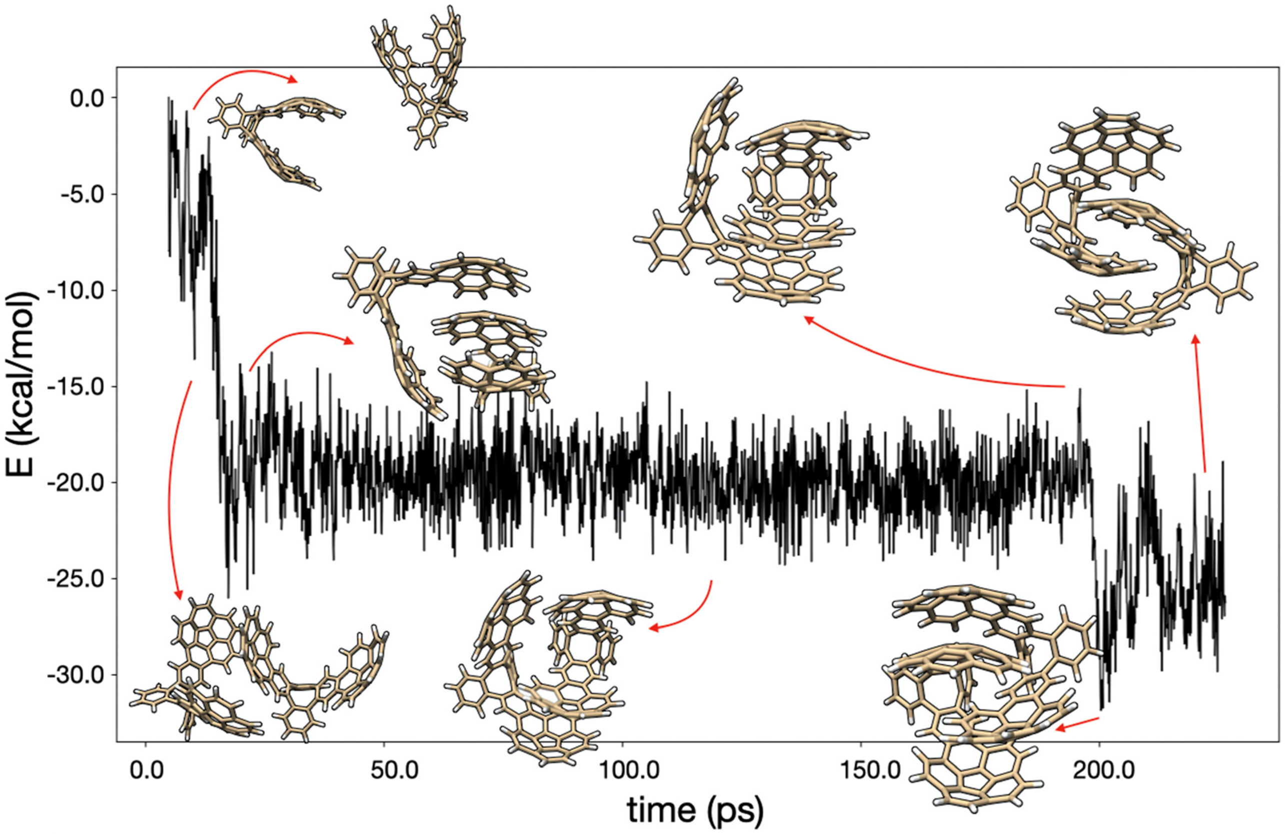

Fig. 6 Dynamics for the dimerization of two molecules of the buckycatcher in conformation ii.

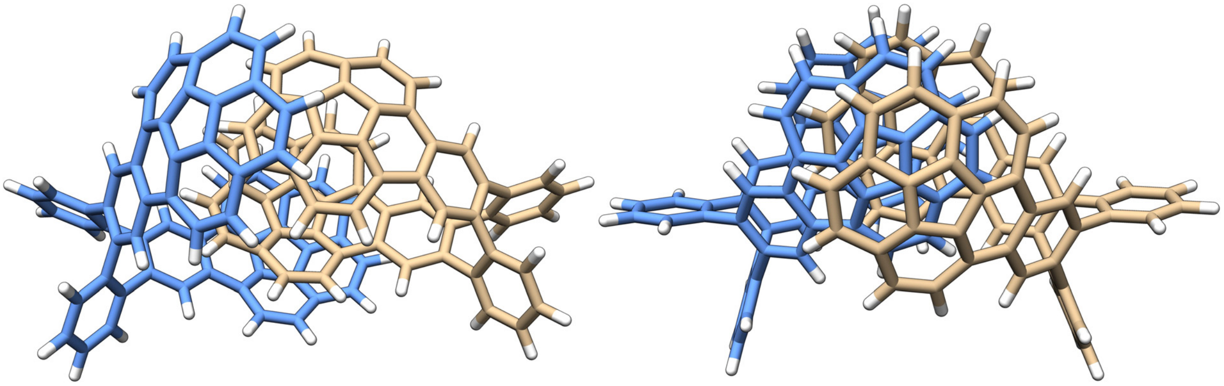

Fig. 7 Structures of the new dimers of the buckycatcher identified with the dynamical calculations. Left, the first species stabilized in the dynamics (E1), right the second and most stable one (E2).

expects dimer 1 to bind in toluene, then the method is particularly suitable to study this specific process, even though the respective binding energies are untrustworthy.

As the potential between conformers is attractive, the two species come closer together and a dimer is formed. The first species stabilized parallels dimer 2 of Fig. 1. However, the molecules interact *via* the corannulene units and not the tether (see Fig. 7, left). Following the naming convention of Fig. 1, we coined this dimer 17. After approximately 200 ps of simulation time, an interconversion takes place. The new species formed is identical to dimer 1, however, the tethers are tilted so that this new complex does not belong to the C_S point group, but to C_2. This species was named dimer 18. We optimized the respective complexes at both semi-empirical levels (PM6-D3H4 and GFN2-xTB) and calculated their relative stability with respect to dimer 1. Table 5 shows the respective data.

Species 17 is significantly less stable than dimer 1, which may be followed in the dynamical calculation. In gas phase, we observe differences in energy and thermodynamic state functions of at least 8 kcal mol^{-1}, which makes this but a short-lived intermediary in the dimerization process. Similar considerations extrapolate to other methods. Like the behaviour of all other species here studied, when placed in the environment of toluene, the chemical potential decreases, coming closer to that of dimer 1. Nevertheless, the relative chemical potential is still 5 kcal mol^{-1} higher, rendering this an irrelevant species for the chemical equilibria. On the other hand, complex 18 is, according to semi-empirical methods, slightly more stable than dimer 1 in gas. This species is favoured enthalpically and entropically. Solvation effects seem to favour the new dimer too. Contrary to the semi-empirical data, PBE calculations expect that 18 is slightly disfavoured with respect to the first dimer. This may however result from using suboptimal geometries. Nevertheless, after accounting for solvation effects, dimers 1 and 18 have the same chemical potential. Fortunately, the difference in energies is minimal, so that results on individual dimerization reactions remain valid. An additional conclusion is possible, however, that the potential energy surface around dimer 1 is quite flat, allowing a tilting motion where the tethers move concertedly. This motion is similar to water's symmetric bend. The observation agrees with the dynamics we ran on the isolated catcher (see S1, Fig. S5, ESI†). Finally, the dynamical run shows that the formation of the dimers of the buckycatcher should not be kinetically hindered.

Table 5 Stability of dimers 17 and 18 with respect to dimer 1. Data in kcal mol^{-1}. $\Delta G_{\text{RRHO-gas}}$ is the calculated correction to transform binding energies into binding Gibbs free energies in gas phase, while ΔG_{PhMe} yields the conversion to toluene. Thermodynamic data calculated at the GFN2-xTB level

	17	18
$\Delta E_{\text{GFN2-xTB}}$	8.9	−0.2
$\Delta E_{\text{PM6-D3H4}}$	9.9	−0.3
$\Delta E_{\text{PBE-D3BJ}}$	7.8	0.3
$\Delta E_{\text{PBE-D3BJ (ATM)}}$	7.4	0.3
$\Delta G_{\text{RRHO-gas}}$	−0.9	−0.1
ΔG_{PhMe}	−3.9	−0.2

Two things change however with the present additional data: (i) despite our attempt to be as exhaustive as possible, the dynamical calculations showed there are still potentially important conformers missing due to the flatness of the energy surface. Though these might not impact the enthalpy of binding for the ensemble, these can impact the conformational entropy of the product state, which cast doubt on the calculated change in conformational entropies. At least, the value we tabulated before could be biased to the reagent (monomeric catcher) state. Though in most cases conclusions are not impacted by this bias, DFT data with triple-zeta basis sets are borderline. To verify this, all previous dimers were reoptimized in tilted variants.

Re-evaluation of thermodynamics of binding

Scouting for tilted dimers allowed us to identify 15 additional species, leading to a total of 36 dimers. All the optimized species are energetically distant from the previously presented ones, such that their conformational weights are at most 0.2% at 300 K and in the dielectric of toluene. This presents no change whatsoever on data given in Table 4.

Discussion

The present manuscript presents a new dataset of similar systems for which dimerization energies are sensitive to fine binding details, like geometry or balance between relevant intermolecular forces. Despite their similarity, binding energies cover a wide range of values, offering sensitive test-grounds for computational methods. Reference dimerization energies were obtained by using a series of DFT calculations in several conditions, which included different basis sets and dispersion corrections. We also used M06-2X with sufficiently large basis sets as an orthogonal consistency check to calculations using, *e.g.*, the Grimme dispersion correction. Our results agree with other observations in the literature, for instance, on using fully dispersion-corrected PBE with large enough basis sets and on the overall accuracy of M06-2X.[11,17]

One of the goals of the present work was the identification of limitations in semi-empirical methods. It is well known that PM6 lacks repulsion,[26,27] and this result is further evidenced here in our work. Řezáč and Hobza proposed adding an empirical correction acting on hydrogen atoms.[27] Though this systematically improves the method, we feel that a deeper change involving reparameterization or even reformulation of the core-core terms might be required. Introduction of orthogonalization effects also seems to be paramount.[28] In our benchmark study, PM6-D3H4 optimized geometries were in excellent agreement with S30L ones. DFT calculations on PM6-D3H4 and S30L geometries were numerically close. The same does not apply for GFN2-xTB. Though lack of repulsion has been improving significantly with different generations of the method, there is still room for improvement. It is possible that slight changes in the repulsive terms might be necessary. Overall, we feel that from this study, and our experience in general, focus should be given on repulsive terms when developing or refining currently existing semi-empirical methods. Note that this is meaningful from a physical point of view: while other interactions decay with inverse powers of the interatomic distance, overlap-repulsion has an exponential dependence.

The calculations shown in this work reveal a discrepancy in the thermodynamic feasibility of dimerization for the buckycatcher. DFT calculations with triple-zeta basis sets predicted non-spontaneity for the process (in toluene). The respective Gibbs free energies are, within the methods' accuracies, zero. All other calculations expect the opposite behaviour, *i.e.*, spontaneity. We stress that here we have an accumulation of deviations at several levels. One that is particularly important is the error associated with the solvation terms. ALPB is parametrized for GFN2-xTB, which, according to our observations from above, lacks repulsion. It is therefore extremely likely that the resulting ALPB solvation energies used in this work are over-penalizing to compensate for limitations of the tight binding method. Though not troublesome when used to estimate solvation terms for PM6 or DFT/double-zeta calculations, this bias will be incorrect for PBE-D3BJ (ATM) data, in fact, leading to over-destabilization. This would agree with previous observations of ours.[10,12] Still, we feel that ALPB solvation is far superior to COSMO's. Additionally, dispersion corrected PBE has its limitations.[11] Due to the likeliness of the remarks in this paragraph and the values obtained for Gibbs free energy of dimerization, we expect that dimerization is thermodynamically favourable, with the lowest boundary for the Gibbs free energy given by the PM6-D3H4 calculation. A precise value for the Gibbs free energy of dimerization will only be available when more refined solvation contributions will be estimated. Note that our dynamical simulations expect no significant barrier from the kinetic point of view. Consequently, we expect that room temperature solutions of the buckycatcher are composed predominantly by dimers of this species.

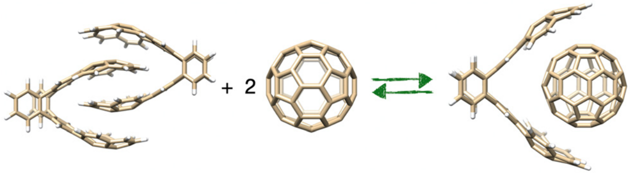

Fig. 8 Schematic for the process used to evaluate the competition between dimerization of the pincer against capturing of C_{60}.

At this point we are in position to consider the competition between the dimerization of the corannulene pincer against the encapsulation of fullerenes between the pincer's arms. Though from a theoretical viewpoint it is attractive to compare the binding to C_{60} against dimerization, this will not be reflecting the experimental conditions, where a solution of the fullerene is added to a (pre-existing) solution of the catcher. In this order of events, if dimerization is favourable, fullerene is added to a solution of the dimers, not of the isolated buckycatcher. A more adequate comparison consists thus of analysing the reaction of fullerene breaking the dimer to form C_{60}@catcher aggregates. This is represented in Fig. 8, with the specific example for dimer 1. The thermodynamically consistent procedure to perform such comparison requires accounting for conformational effects on both reactions. Note that the spontaneity of the process can be easily evaluated using thermodynamic data for the formation of the C_{60}@catcher ($\Delta G_{\mathrm{C60@catcher}}$) complexes and the dimerization of the buckycatcher ($\Delta G_{\mathrm{dimerization}}$). For dimerization to be favourable over complexation with fullerene, it is necessary that $\Delta G_{\mathrm{dimerization}} > 2\Delta G_{\mathrm{C60@catcher}}$. This is however not the case even when using semi-empirical data. We consequently conclude that dimerization does not compete thermodynamically against the formation of adducts with fullerenes.

Computational details

All calculations were performed using our newly developed C++ library, ULYSSES.[13] Geometries were minimized using the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) algorithm with the dogleg trust-region method with convergence criteria of $10^{-8}E_\mathrm{h}$ for energies and $2.5 \times 10^{-5}E_\mathrm{h}/a_0$ for gradients. The Hessian was approximated using the method of Lindh *et al.*[29]

The Hamiltonian of choice for optimizing geometries is always consistent with the method chosen for energy and Hessian evaluation. These are GFN2-xTB,[30] PM6-D3H4X,[18,27,31] and PM6-D3H+.[32,33]

Initial geometries for the pincer's dimers were obtained by direct construction from the respective optimized geometries of each conformer. Appropriate positions for each molecule in each dimer were built using Avogadro 1.2.[34,35] Figures of molecules were generated with UCSF Chimera, as developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.[36]

For the calculation of thermodynamic properties, we used the mixed free rotor/harmonic oscillator model as implemented in ULYSSES. This is based on the interpolation scheme proposed by Grimme for entropies,[6] which we extended to other thermodynamic quantities.[13] In our program, the interpolation frequency required by this model takes the value of 75 cm^{-1}.[13] Thermal corrections for obtaining thermodynamic state functions were based on the semi-empirical geometries used throughout the work. For consistency, and to minimize the impact of imaginary frequencies (and their "inversion", as suggested by Sure and Grimme[11]), anharmonic frequencies were estimated using the respective method used for geometry optimization. This means that geometries and vibrational frequencies are obtained from semi-empirical methods. The expression used to obtain the values reported in Table 4 and Fig. 4 and 5 is

$$\Delta G_{\mathrm{SOLVENT}} = \Delta G_{\mathrm{RRHO\text{-}gas}} + \delta G_{\mathrm{SOLVATION}} \qquad (2)$$

where $\Delta G_{\mathrm{SOLVENT}}$ is the quantity reported in the table and figures, $\Delta G_{\mathrm{RRHO\text{-}gas}}$ contains the gas phase Gibbs free energy, and $\delta G_{\mathrm{SOLVATION}}$ contains the solvation corrections obtained using either ALPB or COSMO with respect to gas phase data. The term $\Delta G_{\mathrm{RRHO\text{-}gas}}$ corresponds to the difference of changes in enthalpy and entropy in the gas phase ($\Delta H_{\mathrm{RRHO\text{-}gas}} - T\Delta S_{\mathrm{RRHO\text{-}gas}}$). Expressions for the partition functions used in the calculation were previously reported.[13]

Solvation Gibbs free energies at the GFN2-xTB level were estimated using the ALPB solvation model as described by Ehlert *et al.*[23] Note that ALPB is only parametrized for use with GFN2-xTB. We also used COSMO as available from MOPAC along with PM6-D3H4X.[37,38] The solvent dielectric constant used in COSMO was 7.0, larger than the experimental one (2.38). The value of 7.0 was chosen because (i) solvation free energies using $\varepsilon = 2.38$ barely differed from the gas phase ones, indicating severe underestimation of solvation effects; (ii) consistency with ALPB parametrization. Note that even using $\varepsilon = 7.0$ the results evidence underestimation of solvation effects, as noted above. For completeness, however, dimerization free energies are provided also using the real dielectric constant of the solvent (S1). Solvation Gibbs free energies were always used in conjunction with the respective data in gas phase at 300 K. In all cases, solvation effects were added to gas phase optimized geometries. This approach, used for instance in the parametrization of solvation models,[39] leads to minor deviations in the calculated quantities (see S4).

In some rare occasions we were unable to obtain adduct geometries without imaginary vibrational frequencies, which are typical for very flexible systems or also van der Waals complexes. In such situations we used instead the absolute value of the respective imaginary vibrational frequencies in the calculation of thermodynamic properties as suggested by Sure and Grimme.[11] This simplification was however only employed with less relevant species, *i.e.*, the ones for which an inaccuracy in vibrational modes would not possibly influence their thermodynamic relevance. If a dimer showed an imaginary frequency larger than 40i cm^{-1}, the structure was further reoptimized. The largest imaginary frequency persistent in the structures was 36.5i cm^{-1} (which was converted to 36.5 cm^{-1} for the statistical mechanical calculations). For two species, dimers 11 and 12, we could not avoid a second imaginary frequency. The values were themselves also of low magnitude. No frequency was disregarded in this work. If the vibrational spectrum did not conform to the standards above defined, the structure was simply further optimized. The list of imaginary frequencies for which the absolute value was taken for the calculation of thermodynamic properties is given in the ESI† (S3).

The dynamics were run using ULYSSES BOMD module, which uses a semi-empirical quantum chemical method for energy and gradient (force) evaluation. The time step was set to 1 fs and geometries for the construction of trajectories were saved every 10 fs. Integration of the equations of motion was done using the leapfrog algorithm.[40] Dynamical simulations were run always with GFN2-xTB and the ALPB solvation model for each substrate.

M06-2X,[41] r^2SCAN-3c,[42] PBE,[43] and PW6B95[44] calculations were run with ORCA 5.0.[45–48] The D3[18] correction was calculated with Becke–Johnson damping[19] directly from ORCA. 3-Body terms were calculated with the stand-alone program dftd3.[49] GFN1-xTB[50] and GFN0-xTB[51] calculations were conducted with xTB.[52]

Conclusions

In this work we studied the dimerization of the buckycatcher. New benchmark data was provided for the buckycatcher-fullerene system, which allowed us to provide robust reference points for the errors and deviations in state-of-the-art semi-empirical methods. We performed a very exhaustive description of the dimeric forms of the corannulene pincer and, overall, semi-empirical and DFT results are in good agreement. To very good extent, dimerization energies obtained with the two groups of methods are parallel.

Dynamical simulations here conducted allowed us to conclude that the dimerization process is not kinetically hindered. Thus, if the process is thermodynamically favoured, room temperature solutions of the buckycatcher should present the species in a dimerized form. As discussed above, ALPB seems to be over-penalizing in the calculation of dimerization Gibbs free energies in solution. Still, this thermodynamic state function was estimated to be borderline close to zero at sufficiently high levels of theory. This means that, given the values we calculated, it is possible that dimerization is spontaneous, however, only with a small in magnitude equilibrium constant. A reformulation of the competition between dimerization of the buckycatcher and complexation with fullerenes shows however clearly the thermodynamic preference for the latter process, accounting for experimental observations.

Author contributions

F. M.: conceptualization, data curation, formal analysis, methodology, software, validation, visualization, writing, project administration. G. M. P.: funding acquisition, project administration, resources, supervision.

Conflicts of interest

Authors have no conflicts of interest to declare.

Acknowledgements

Authors wish to thank funding by the Bundesministerium fuer Wirtschaft und Klimaschutz (BMWi) *via* ZIM (grant number: KK 5197901TS0) as well as the Bundesministerium fuer Bildung und Forschung (BMBF), project SUPREME (number: 031L0268).

Notes and references

1 A. Sygula, F. R. Fronczek, R. Sygula, P. W. Rabideau and M. M. Olmstead, *J. Am. Chem. Soc.*, 2007, **129**, 3842.
2 Y. Zhao and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2813.
3 A. Sygula, M. Yanney, W. P. Henry, F. R. Fronczek, A. V. Zabula and M. A. Petrukhina, *Cryst. Growth Des.*, 2014, **14**, 2633.
4 P. A. Denis and F. Iribarne, *Int. J. Quantum Chem.*, 2015, **115**, 1668.
5 C. Mück-Lichtenfeld, S. Grimme, L. Kobryn and A. Sygula, *Phys. Chem. Chem. Phys.*, 2010, **26**, 7091.
6 S. Grimme, *Chem. – Eur. J.*, 2012, **18**, 9955.
7 C. Villot, F. Ballesteros, D. Wang and K. U. Lao, *J. Phys. Chem. A*, 2022, **126**, 4326.
8 V. H. Le, M. Yanney, M. McGuire, A. Sygula and E. A. Lewis, *J. Phys. Chem. B*, 2014, **118**, 11956.
9 A. Tkatchenko, D. Alfè and K. S. Kim, *J. Chem. Theory Comput.*, 2012, **8**, 4317.
10 F. Menezes and G. M. Popowicz, *Molecules*, 2022, **27**, 3838.
11 R. Sure and S. Grimme, *J. Chem. Theory Comput.*, 2015, **11**, 3785.
12 F. Menezes and G. M. Popowicz, *Molecules*, 2023, **28**, 2841.
13 F. Menezes and G. M. Popowicz, *J. Chem. Inf. Model.*, 2022, **62**, 3685.
14 Y. S. Al-Hamdani and A. Tkatchenko, *J. Chem. Phys.*, 2019, **150**, 010901.
15 E. G. Hohenstein, S. T. Chill and C. D. Sherril, *J. Chem. Theory Comput.*, 2008, **4**, 1996.
16 L. Goerigk, *J. Phys. Chem. Lett.*, 2015, **6**, 3891.
17 D. Wu and D. G. Truhlar, *J. Chem. Theory Comput.*, 2021, **17**, 3967.
18 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
19 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456.
20 E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2019, **150**, 154122.
21 I. Hayes and A. J. Stone, *Mol. Phys.*, 1984, **53**, 83.
22 J. Garcia and K. Szalewicz, *J. Phys. Chem. A*, 2020, **124**, 1196.
23 S. Ehlert, M. Stahn, S. Spicher and S. Grimme, *J. Chem. Theory Comput.*, 2021, **17**, 4250.
24 O. Acevedo and W. L. Jorgensen, *Acc. Chem. Res.*, 2010, **43**, 142.
25 O. Acevedo and W. L. Jorgensen, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 422.
26 A. S. Christensen, T. Kubař, Q. Cui and M. Elstner, *Chem. Rev.*, 2016, **116**, 5301.
27 J. Řezáč and P. Hobza, *J. Chem. Theory Comput.*, 2012, **8**, 141.
28 W. Weber and W. Thiel, *Theor. Chem. Acc.*, 2000, **103**, 495.
29 R. Lindh, A. Bernhardsson, G. Karlström and P.-Å. Malmqvist, *Chem. Phys. Lett.*, 1995, **241**, 423.
30 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652.
31 J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173.
32 M. Korth, *J. Chem. Theory Comput.*, 2010, **6**, 3808.
33 J. C. Kromann, A. S. Christensen, C. Steinmann, M. Korth and J. H. Jensen, *Peer J.*, 2014, **2**, 449.
34 Avogadro 1.20: an open-source molecular builder and visualization tool. Available online: **http://avogadro.cc/**(accessed 23.12.2022).
35 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminf.*, 2012, **4**, 17.
36 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605.
37 A. Klamt and G. Schüürmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, 799.
38 J. J. P. Stewart, *MOPAC2016, Stewart Computational Chemistry*, Colorado Springs, Co, USA. Available online:**http://OpenMOPAC.net** (accessed 23.12.2022).
39 A. V. Marenich, R. M. Olson, C. P. Kelly, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 2011.
40 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford, United Kingdom, 2017.
41 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215.
42 S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes, *J. Chem. Phys.*, 2021, **154**, 064103.
43 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
44 Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 5656.
45 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**, e1606.
46 E. F. Valeev, *Libint: A library for the evaluation of molecular integrals of many-body operators over Gaussian functions, Version 2.0*. **http://libint.valeyev.net/**(accessed Dec 30 2022).
47 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297.
48 S. Lehtola, C. Steigemann, M. J. T. Oliveira and M. A. L. Marques, *Software X*, 2018, **7**, 1.
49 **https://www.chemie.uni-bonn.de/grimme/de/software/dft-d3** (accessed 19.04.2023).
50 S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989.
51 P. Pracht, E. Caldeweyher, S. Ehlert and S. Grimme, ChemRxiv, 2019.
52 C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2020, **11**, e01493.