



Cite this: *Phys. Chem. Chem. Phys.*, 2025, 27, 20421

# An efficient workflow for generation of conformational ensembles of density functional theory quality: dimers of polycyclic (hetero-)aromatics

Jessica J. Ortlieb,  Nathanael J. King  and Alex Brown \*

The determination of geometries and relative energetics (binding as well as free energies) of ensembles of dimers, as well as for small  $n$ -mers, is an important property in physical chemistry, connected to understanding both properties and spectroscopic measurements. In this tutorial review, a workflow for generating conformational ensembles is presented and highlighted for several homodimers. The workflow involves six steps: (i) generate an initial ensemble, using the Conformer-Rotamer Ensemble Sampling Tool (CREST), and its underlying GFN2-xTB method; (ii) reoptimize each member of the ensemble using B97-3c; (iii) discard duplicates; (iv) reoptimize the remaining conformers using  $\omega$ B97X-D4/def2-SVP; (v) if needed, discard any duplicate conformers; (vi) compute vibrational frequencies using  $\omega$ B97X-D4/def2-SVP and final single point energies using  $\omega$ B97X-V/def2-QZVPP. B97-3c was selected for step (ii) due to its performance in a screening of several composite density functional theory (DFT) methods, namely HF-3c, B97-3c, PBEh-3c,  $r^2$ SCAN-3c, and  $\omega$ B97X-3c, for the pyrene homodimer. The six-step workflow allows the generation of large DFT-quality ensembles efficiently, as demonstrated on the known pyrene dimer ensemble, and then applied to the homodimers of eight small polycyclic (hetero-)aromatic molecules related to asphaltenes: anthracene, phenanthrene, fluorenone, dibenzofuran, dibenzothiophene, dibenzothiophene oxide, *N*-methylcarbazole, and benzo[*h*]quinoline. Methods are suggested for analysis of trends in dimerization structures and energies for these monomers, including an analysis revealing a strong dependence of binding energy on the magnitude of dipole cancellation.

Received 14th March 2025,  
 Accepted 6th September 2025

DOI: 10.1039/d5cp01010a

[rsc.li/pccp](http://rsc.li/pccp)

## 1 Introduction

Non-covalent interactions (NCIs) are an important topic in chemistry, particularly in computational chemistry. One of the more unique and well-known forms of NCIs is  $\pi$ - $\pi$  stacking, which is a significant factor in phenomena as diverse as the stacking of DNA bases,<sup>1</sup> the origin of optoelectronic properties in organic electronic materials,<sup>2,3</sup> and the nanoaggregation of petroleum asphaltenes.<sup>4,5</sup> Despite the ubiquitous nature of  $\pi$ - $\pi$  interactions, there is still disagreement on the physical mechanisms behind  $\pi$ - $\pi$  stacking,<sup>6-8</sup> which leads to uncertainty when predicting the strength and geometry of interaction, particularly in cases where multiple aromatic rings are fused together.<sup>9-13</sup>

In computational studies of non-covalent complexes and other flexible systems, there are many possible geometries to consider, and it is crucial that the computations are performed using the correct geometry or geometries. In particular, finding

the lowest-energy structure or structures is of the utmost importance. If computations are done on higher-energy structures, without taking the lowest-energy structures into account, then the results will have no relation to experiments, as they represent conformations that have a very low probability of occurring at equilibrium. Thus, many algorithms have been developed to try to find as many conformations of a given system as possible,<sup>14-20</sup> with the hope that finding enough structures will ensure that the correct, low-energy structures are among those found. When working with flexible systems, best practice is to perform computations on the full ensemble of structures, and to present the results for any given system as a Boltzmann-weighted average of the results for all members of the ensemble. Unfortunately, many interesting systems contain numerous atoms and therefore a concomitant large number of degrees of freedom. Thus, there are many conformations in the ensemble and each one is expensive to compute accurately. Taken together, these factors usually make it prohibitively expensive to compute high-quality results for properties of large, flexible molecules and for non-covalent complexes composed of such molecules.

Department of Chemistry, University of Alberta, Edmonton, AB, T6G 2G2, Canada.  
 E-mail: alex.brown@ualberta.ca; Tel: +1-780-492-1854



One of the more effective methods for the generation of conformer ensembles is the Conformer-Rotamer Ensemble Sampling Tool (CREST).<sup>14,15</sup> This tool uses RMSD-biased metadynamics to sample possible geometries for a system. The generated geometries are then iteratively screened and optimized to generate a final ensemble consisting of all conformers found in a given energy window (default: all conformers in the lowest 6 kcal mol<sup>-1</sup> by energy). This algorithm has been found to give the best coverage and the best success rate at finding lowest-energy conformers,<sup>16</sup> for systems which are complex enough that systematic conformer generation algorithms such as Confab<sup>21</sup> would be too expensive. Although CREST has been highly successful, other conformational sampling tools are still being developed, such as the global optimizer algorithm (GOAT).<sup>22</sup>

CREST generates large ensembles and more reliably finds low-energy conformers than algorithms using force fields or knowledge-based algorithms,<sup>16</sup> but the generated geometries should still be reoptimized at a level of theory higher than GFN2-xTB which CREST uses before any computation of properties is done.<sup>14,15</sup> Unfortunately, for large systems, the computational cost for reoptimization of the full ensemble (often hundreds or thousands of geometries) using density functional theory (DFT) can be enormous. As a result, in our previous work,<sup>5,23</sup> we have reoptimized only the lowest-energy geometry for large systems (hundreds of atoms) or a low-lying fraction of the ensemble (often the lowest 2 kcal mol<sup>-1</sup>, instead of the full 6 kcal mol<sup>-1</sup>) for medium-sized systems (50–100 atoms). The problem with such an approach is that those geometries which are lowest in energy using the GFN2-xTB<sup>24</sup> method, upon which CREST depends, are not necessarily the same as the geometries which are lowest in energy when using DFT. Thus, choosing the structure or structures which CREST reports as being lowest in energy is no guarantee that all the true low-energy structure(s) will be found. Indeed, the original CREST paper<sup>14</sup> indicated that the ranking of conformers using GFN2-xTB often differs significantly from rankings using DFT, and the usual procedure among users of CREST is to reoptimise ensembles at some higher level of theory.<sup>14,25–29</sup> Additionally, when CREST-generated ensembles are reoptimized using DFT, many of the conformers, up to 90% of them in some cases, coalesce to become equivalent, indicating that the GFN2-xTB potential energy surface may contain many spurious minima.<sup>23</sup>

The Grimme group has very recently introduced a replacement for GFN2-xTB, which they call g-xTB,<sup>30</sup> with the g standing for “general.” It shows great promise for significantly improved results over GFN2-xTB. It will be very interesting to see how conformer searches using g-xTB perform, and whether they require the same level of refinement, but such testing will need to wait until g-xTB is finished and incorporated into codes such as CREST and ORCA. The benchmark results available so far for g-xTB, especially for NCIs,<sup>30</sup> look very promising.

In the past few years, a promising new set of methods has been introduced to computational chemistry: the 3c composite methods.<sup>31–35</sup> These use either Hartree–Fock or DFT with a very small basis set, carefully curated for maximum error

cancellation, plus two or three empirical corrections. Some recently introduced 3c methods include HF-3c,<sup>31</sup> B97-3c,<sup>32</sup> r<sup>2</sup>SCAN-3c,<sup>34</sup> PBEh-3c,<sup>33</sup> and  $\omega$ B97X-3c.<sup>35</sup> In principle, and in practice,<sup>29,36</sup> reoptimizing ensembles using one of the 3c methods will significantly reduce errors in energy and geometry present in CREST-derived ensembles, and will also allow for the elimination of most of the spurious minima. Therefore, in this work, we explore re-optimizing and re-ranking the ensembles using 3c methods before choosing low-energy structures to reoptimize using DFT. Because the 3c methods are much less computationally expensive than DFT, this should result in greatly reduced computational effort compared to reoptimizing full ensembles using DFT, while much more reliably finding the correct low-energy structures for DFT optimization.

In this work, we test the available 3c methods for speed and accuracy, and then apply the B97-3c method to a reanalysis of an ensemble of previously reported pyrene dimers. The workflow – CREST to B97-3c to DFT – is then applied to exploring the ensembles of a series of polycyclic aromatic and heteroaromatic homodimers. Studying these particular systems provides insight into the effects which are produced by substituting heteroatoms into polycyclic aromatic systems as they relate to non-covalent interactions, particularly  $\pi$ – $\pi$  stacking. However, they also highlight to the reader the challenges in determining conformational ensembles for homodimers, a process readily extended to heterodimers or even higher order *n*-mers.

## 2 Computational methods

Initial conformer ensembles for the homodimers were generated using CREST,<sup>14</sup> in non-covalent interactions (NCI) mode. By default, CREST uses the GFN2-xTB method<sup>24</sup> for all computations. All 3c and DFT computations were done in ORCA 5.0.4.<sup>37</sup> ORCA uses the RIJCOSX<sup>38</sup> approximation for all 3c and DFT computations by default. For computations on solvated systems, where the solvent considered was toluene, the conductor-like polarizable continuum solvation model (CPCM)<sup>39</sup> was used for geometry optimizations and the solvation model based on density (SMD)<sup>40</sup> for single-point energy calculations.

Several composite methods, namely, B97-3c,<sup>32</sup> HF-3c,<sup>31</sup> PBEh-3c,<sup>33</sup> r<sup>2</sup>SCAN-3c,<sup>34</sup> and  $\omega$ B97X-3c<sup>35</sup> were used to compute the single point energies of the seven known conformers of the pyrene homodimer<sup>23</sup> in the gas phase and in toluene solution; these previous geometries were determined at the  $\omega$ B97X-D4<sup>41</sup>/def2-TZVP<sup>42,43</sup> level of theory. The best performing method was B97-3c, see Section 3.1. To gauge the differences in optimal geometry and energy between B97-3c and  $\omega$ B97X-D4, the seven conformers were reoptimized and Gibbs free energies of dimerization were determined using B97-3c. Finally, single point energies were determined using  $\omega$ B97X-D4/def2-TZVP at the B97-3c geometries. The results showed that while the differences in geometry between conformers optimized using B97-3c and those using  $\omega$ B97X-D4/def2-TZVP were small (average all-atom RMSD was 0.02 Å in the gas phase, and 0.09 Å in toluene),



the energy differences between the conformers, as evaluated using  $\omega$ B97X-V/def2-TZVP single point energies were large enough (averaging  $5 \text{ kJ mol}^{-1}$ ) to warrant further reoptimization using  $\omega$ B97X-D4/def2-SVP, rather than using the B97-3c geometries directly. See Section 3.1 for further discussion.

As a further test, the entire workflow was applied to the pyrene dimer, to test whether it could quickly and accurately reduce a large field of potential conformers to a small, accurately ordered list of optimized structures. Optimization using B97-3c and tight thresholds (ORCA keyword *TightOPT*) produced an ensemble containing several of the previously published structures, but also many spurious minima. Similarly to how CREST operates, we determined that any two conformers with energies within  $0.2 \text{ kJ mol}^{-1}$  of each other, and with all three rotational constants matching to within 1 MHz were duplicates of each other. Unfortunately, simple RMSD calculations cannot be used to identify duplicates, since symmetric monomers like pyrene can be rotated to give a rotamer which is identical to the starting conformer, but nevertheless yields a large RMSD because the individual atoms are in different positions. Duplicate structures were then removed from the ensemble, and further optimization was done using B97-3c in conjunction with very tight thresholds (the *VeryTightOPT*, *VeryTightSCF*, and *DefGrid3* ORCA keywords). Again, duplicates were removed, and final optimization was performed using  $\omega$ B97X-D4/def2-SVP. Final single points were computed using  $\omega$ B97X-V/def2-TZVP. While we use  $\omega$ B97X-V/def2-QZVPP for final single points in the rest of this work, def2-TZVP was used here to allow for comparison with the previously published results.<sup>23</sup> The results are shown in Section 3.2.

Having determined that B97-3c works well as an intermediate method for this kind of dimer, ensembles were generated and refined for a series of homodimers of tricyclic (hetero-) aromatic compounds, namely anthracene (Anth), phenanthrene (Phen), benzo-*[h]*-quinoline (BzQuin), dibenzothiophene (DBT), dibenzothiophene oxide (DBTO), dibenzofuran (DBF), fluorenone (Fluor), and *N*-methylcarbazole (MeCarb), see Fig. 1. After optimizing the geometries of the CREST-generated dimer

structures using B97-3c, any duplicates were removed from the ensemble. We then used  $\omega$ B97X-V<sup>44</sup>/def2-QZVPP<sup>45</sup>// $\omega$ B97X-D4/def2-SVP to compute high-quality geometries and single point energies for the reduced ensemble. Computations were performed both in vacuum and in SMD//CPCM toluene solution. Gibbs free energies of dimerization were computed for the final geometries using both the equations and, for cases in solution, the correction for concentration discussed previously.<sup>23,46</sup> This gives Gibbs free energies of dimerization at standard conditions – that is, in the gas phase, both monomer and dimer are at partial pressures of 1 atm, and a temperature of 298 K, and in solution, both monomer and dimer are at concentrations of 1 M, and a temperature of 298 K.

Note that, as per convention, all binding energies in this paper are reported as the energy of the monomers minus that of the dimer, to ensure that a bound complex has a positive binding energy. Gibbs free energies of dimerization are reported as the free energy of the dimer minus those of the monomers, to ensure that a spontaneously formed complex has a negative Gibbs free energy of formation.

For better comparison of geometries, the *xyz* coordinates for each conformer of each dimer were transformed into new coordinates that describe the dimers in terms of rotation and displacement along the long axis, short axis, and perpendicular axis. That is, one monomer was chosen as the reference monomer, while the other is considered the displaced monomer, and both monomer centres of mass were determined. For DBTO, the oxygen atom was omitted from centre of mass calculations, to keep the centre of mass in the plane of the aromatic system. The long and short axes are indicated by blue and red dashed lines, respectively, in Fig. 1, with the direction of positive displacement indicated by arrows, where applicable. In general, the long axis passes through the two most distant carbon atoms of the molecule, while the short axis is perpendicular to the long axis, and bisects the central bond which connects the most distant aromatic rings to each other. The normal, or *z*, axis is that line which is perpendicular to both the long and short axes (and thus to the best-fit plane of the



Fig. 1 Monomers of the homodimers under study. Long axes are indicated by blue dashed lines. Short axes are indicated by red dashed lines. Where applicable, the direction defined as positive displacement is indicated by an arrowhead.



monomer) through the point of their intersection. Separation between centres of mass was described as translation of the displaced centre of mass along the long, short, and normal axes of the reference monomer. Rotation of the displaced monomer was described by three angles: (a) the angle between the reference long axis and the projection of the displaced long axis into the plane of the reference monomer, (b) the inclination of the displaced long axis with respect to the plane of the reference monomer, and (c) the inclination of the displaced short axis with respect to the plane of the reference monomer. This gives us a coordinate system which ignores the small intramonomer distortions, and describes any dimer using only six degrees of freedom. Further details on these definitions, including on positive *vs.* negative displacements and rotations for specific monomers, are provided in the SI, Section S1. Transformed coordinates for each dimer can also be found in the SI.

Conformations are named according to their rotation and stacking pattern, with additions for short axis rotations in BzQuin and for relative oxygen atom direction in DBTO. Conformation names start with the numerical value of their rotation angle (measured in degrees), followed by one or two letters indicating lateral displacement patterns. Most conformers are simply labeled SR, for “slipped and rotated.” X denotes a conformer where the centre points (either centre of mass or centroid of central ring) of the two monomers remain superimposed, while the conformers are rotated relative to each other around this common axis. For dimers whose monomers have only six-membered rings (Anth, Phen, BzQuin, and Py), the possibility of graphitic-type stacking arises, where rings of one monomer are centred over 3-way ring junctions of the other. This motif is designated G. For BzQuin, if a conformer has a short axis inclination near 180°, then an F (for “flipped”) is appended to the conformer name. For DBTO, one of three letters, I, O, or P, is appended to describe the relative orientation of the oxygen atoms. If each oxygen atom is oriented toward the molecular plane of the opposing monomer, then an I (for “inward”) is appended to the conformer name. If each oxygen atom is oriented away from the molecular plane of the opposing monomer, then an O (for “outward”) is appended to the conformer name. If one is oriented toward the opposing monomer’s molecular plane but the other is oriented away, then a P (for “parallel”) is appended. Finally, if two or more distinct conformers are not otherwise distinguished by our naming system, we then append a number to the end of the name, with conformers with lower Gibbs free energies having lower numbers.

As an example, for a conformer of the Phen<sub>2</sub> dimer where the monomers are rotated 180° relative to each other, and then shifted along the long axis until ring centres are superimposed over ring junctions, the conformer name would be Phen-180G, while a BzQuin<sub>2</sub> dimer where the monomers are rotated 105°, the short axis is inclined 175°, and the monomers are displaced laterally along both the long and short axes would be named BzQuin-105SR-F. As another example, a DBTO<sub>2</sub> dimer where the oxygens each point away from the opposing monomer and the

monomers are rotated 72° around the axis passing through the centre points of the 5-membered rings would be named DBTO-72X-O. We found only one conformer where the short axis inclination was not within 10° of either 0° or 180°, and this conformer is named DBTO-180T.

## 3 Results and discussion

### 3.1 Pyrene homodimer reoptimizations

When using a 3c method to refine an ensemble before reoptimization using DFT, it would be helpful for the binding energies and geometries given by the 3c method to be close to those given by DFT. Equally importantly, the energetic ordering should remain consistent, such that conformers which are low in energy according to the 3c method are also low in energy when using DFT. In other words, the shape of the potential energy surface should be similar in both the 3c method and in full DFT. Thus, prospective methods were evaluated both for absolute error in binding energies and for relative binding energy ordering, as compared to the published  $\omega$ B97X-V/def2-TZVP values,<sup>23</sup> at the published geometries. Thirdly, efficient methods are highly desirable for screening, so the methods were compared in terms of computer time required, measured in core-seconds. All computations in this section were performed on a single core, for ease of comparison.

In the gas phase, all of the 3c methods gave similar energy orderings to  $\omega$ B97X-D4/def2-TZVP. PBEh-3c gave exactly the same ordering, while B97-3c, r<sup>2</sup>SCAN-3c, and  $\omega$ B97X-3c moved one conformer lower in the ranking but were otherwise the same, and HF-3c inverted the order of several energetically similar conformers, see Table 1. In SMD-modeled toluene solution, all of the methods changed the ordering slightly compared to  $\omega$ B97X-D4/def2-TZVP, with HF-3c showing larger changes than the other methods, see Table 2.

While all of the 3c methods produced fairly accurate energy orderings of the conformers, the differences in binding energies were more pronounced. In the gas phase, B97-3c showed

**Table 1** Gas phase pyrene dimer binding energies (kJ mol<sup>-1</sup>) using different 3c methods at previously reported optimized geometries, including the published values<sup>23</sup> ( $\omega$ B97X-D4/def2-TZVP). In the last row, average computation time for a single point energy for each method on a single processor is given in seconds

| Conformer     | $\omega$ B97X-D4 | HF-3c | B97-3c | PBEh-3c | r <sup>2</sup> SCAN-3c | $\omega$ B97X-3c |
|---------------|------------------|-------|--------|---------|------------------------|------------------|
| 0PD-L         | 53.9             | 59.0  | 53.1   | 47.8    | 47.8                   | 49.7             |
| 60G2          | 53.4             | 59.8  | 53.5   | 47.4    | 48.0                   | 49.5             |
| 60G1          | 53.2             | 57.1  | 52.6   | 46.4    | 47.5                   | 50.2             |
| 0G            | 52.5             | 57.5  | 52.1   | 45.4    | 47.2                   | 49.2             |
| 71qG          | 49.8             | 53.8  | 49.5   | 43.9    | 44.1                   | 47.3             |
| 90X           | 48.5             | 56.5  | 48.9   | 40.9    | 43.3                   | 45.4             |
| MAE           | —                | 5.4   | 0.4    | 6.6     | 5.6                    | 3.3              |
| Avg. time (s) | 7415             | 52    | 357    | 1143    | 586                    | 6816             |



**Table 2** Pyrene dimer binding energies ( $\text{kJ mol}^{-1}$ ) in toluene solution, using different 3c methods at previously reported optimized geometries, including the published values<sup>23</sup> ( $\omega$ B97X-D4/def2-TZVP/CPCM(toluene)). In the last row, average computation time for a single point energy for each method on a single processor is given in seconds

| Conformer     | $\omega$ B97X-D4 | HF-3c | B97-3c | PBEh-3c | $r^2$ SCAN-3c | $\omega$ B97X-3c |
|---------------|------------------|-------|--------|---------|---------------|------------------|
| OPD-L         | 33.5             | 41.4  | 32.4   | 26.6    | 28.5          | 28.6             |
| 60G2          | 33.0             | 42.0  | 32.8   | 26.2    | 28.5          | 28.1             |
| 60G1          | 33.0             | 39.5  | 32.1   | 25.2    | 28.2          | 29.0             |
| 0G            | 32.5             | 32.5  | 28.1   | 18.4    | 26.9          | 27.0             |
| 67qG          | 30.3             | 30.1  | 27.2   | 18.1    | 25.9          | 26.3             |
| 90X           | 29.3             | 39.7  | 29.4   | 21.6    | 25.4          | 25.5             |
| MAE           | —                | 5.7   | 1.6    | 9.3     | 4.7           | 4.5              |
| Avg. time (s) | 7747             | 69    | 368    | 774     | 457           | 5175             |

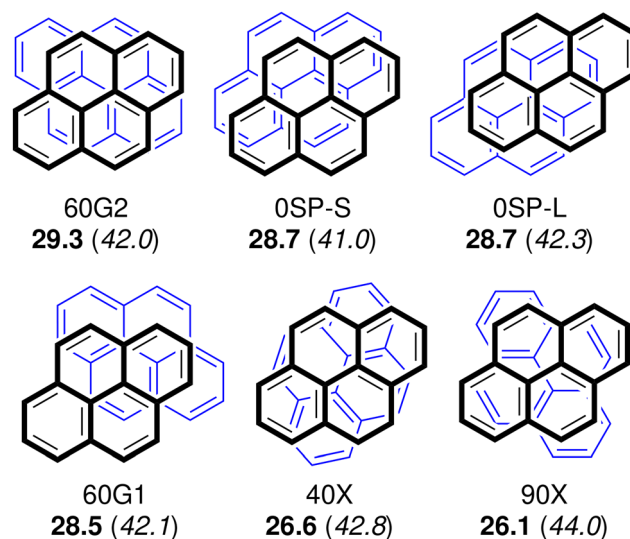
astonishingly close agreement with the  $\omega$ B97X-D4/def2-TZVP binding energies, with most energies within  $1 \text{ kJ mol}^{-1}$  (MAE  $0.5 \text{ kJ mol}^{-1}$ ) while the other 3c methods differed by  $3 \text{ kJ mol}^{-1}$  or more (MAEs range from  $3.4$  to  $6.6 \text{ kJ mol}^{-1}$ ). B97-3c was also the second fastest method, taking an average of just 357 core-seconds (six core-minutes) on a single core to compute a single-point energy for a pyrene dimer. Only HF-3c was faster.

In solution, the agreement was not as good, but B97-3c still maintained an MAE of  $1.6 \text{ kJ mol}^{-1}$ , while the others ranged from  $4.5$  to  $9.3 \text{ kJ mol}^{-1}$ . B97-3c also maintained its speed advantage in solution, with an average of 368 core-seconds per single point energy.

Having determined that B97-3c gives good energies and energy ordering at these previous DFT-optimized geometries, another important metric is its performance for geometry optimization, as typically, one would be using it to refine the geometries determined using semi-empirical (xTB) approaches. To test this performance, the optimized geometries of the six conformers of the pyrene dimer reported in our previous paper were reoptimized using B97-3c. Differences in geometry were quantified as the pairwise RMSD of atomic coordinates for each conformer before and after reoptimization. These values were quite small, averaging only  $0.020 \text{ \AA}$  in the gas phase and  $0.087 \text{ \AA}$  in solution, which indicates that B97-3c does a good job of reproducing the geometries found by  $\omega$ B97X-D4/def2-TZVP. However, when single point energies for the B97-3c geometries were computed using  $\omega$ B97X-D4/def2-TZVP at the B97-3c geometries, the computed energies were an average of  $5.2 \text{ kJ mol}^{-1}$  higher in energy in the gas phase, and  $4.6 \text{ kJ mol}^{-1}$  higher in solution. This indicates that while the B97-3c geometries are much better than those provided by GFN2-xTB, the energies are sensitive enough to geometry that a final optimization using DFT is required to obtain accurate binding energies.

### 3.2 Pyrene homodimer ensemble refinement

With the 3c method of choice determined, we next revisited the refinement of the CREST-generated ensemble. All 73 members of the conformational ensemble generated by CREST from the previous paper<sup>23</sup> were reoptimized using B97-3c. In doing so, we were able to reduce the 73-conformer pyrene homodimer



**Fig. 2** The conformational ensemble of the pyrene homodimer, determined at the  $\omega$ B97X-V/def2-TZVP/SMD(toluene)// $\omega$ B97X-D4/def2-SVP/CPCM(toluene) level of theory, using the screening procedure outlined in the Methods section. Binding energies ( $E_b$ ) are listed in bold, and Gibbs free energies of dimerization ( $\Delta G_{\text{dim}}$ ) in italics.

CREST ensemble down to 10 unique structures, with the rest discarded as duplicates. The geometries of the remaining structures were optimized using  $\omega$ B97X-D4/def2-SVP, after which an additional four structures became duplicates. The final six conformers included five of the six solution-phase dimers identified in our previous paper (all but 67qG), or close approximations thereto, namely the 0SP-L, 0G, 60G1, 60G2, and 90X conformers. For some conformers, the long axes were misaligned by as much as  $12^\circ$ , relative to the previously published conformers, but the conformers were still recognizable. It is likely that at least part of the small difference in geometries is due to the smaller basis set used in the present work during optimization. Also present was the X conformer previously reported by Gonzales and Lim,<sup>47</sup> which in our naming system becomes 40X (Fig. 2).

### 3.3 Other homodimers

Having demonstrated that B97-3c gives good results for both absolute binding energies and relative energy ordering, and that the refinement approach of CREST followed by B97-3c mostly reproduces the results of previous work<sup>23</sup> on the pyrene homodimer with significant time savings, the process was applied to homodimers of additional small polycyclic aromatic molecules, with and without heteroatoms. The eight monomers, along with pyrene, are shown in Fig. 1.

Initial ensembles of the homodimers were generated with CREST and refined using B97-3c. B97-3c refinements for both gas phase and solution phase ensembles started from the gas phase CREST output. After B97-3c geometry optimization, about half of the CREST geometries became indistinguishable from other geometries, and were removed from the ensembles. Further  $\omega$ B97X-D4/def2-SVP optimization on the reduced



**Table 3** Binding energy ( $E_b$ ) and Gibbs free energy of dimerization ( $\Delta G_{\text{dim}}$ ) of the lowest energy conformer of each homodimer, in solution and in gas phase. Results were determined at the  $\omega$ B97X-V/def2-QZVPP/SMD(toluene)// $\omega$ B97X-D4/def2-SVP/CPCM(toluene) level of theory for values in solution, and at the same level of theory, but without solvation, for the gas phase values

| Dimer          | $E_{b,\text{sol}}$<br>(kJ mol <sup>-1</sup> ) | $\Delta G_{\text{dim},\text{sol}}$<br>(kJ mol <sup>-1</sup> ) | $E_{b,\text{gas}}$<br>(kJ mol <sup>-1</sup> ) | $\Delta G_{\text{dim},\text{gas}}$ |
|----------------|---|---|---|------------------------------------|
| Anth-90X       | 25.45   | 21.36   | 42.20   | 17.04                              |
| Phen-105SR     | 25.20   | 29.21   | 42.29   | 20.57                              |
| BzQuin-104SR-F | 28.30   | 26.29   | 43.93   | 18.49                              |
| DBF-96SR       | 24.00   | 29.31   | 37.87   | 23.57                              |
| DBT-92SR       | 23.58   | 30.46   | 38.85   | 23.51                              |
| DBTO-160SR-P   | 30.86   | 26.77   | 59.02   | 8.62                               |
| Fluor-135-SR   | 27.79   | 28.03   | 46.39   | 17.66                              |
| MeCarb-150SR   | 33.68   | 26.67   | 54.69   | 14.23                              |

ensembles eliminated a few more conformers, but overall B97-3c produced few false minima. Final ensembles for each dimer are reported as xyz structures in the SI.

After running high-quality single point energy ( $\omega$ B97X-V/def2-QZVPP/SMD(toluene)) and vibrational frequency computations ( $\omega$ B97X-D4/def2-SVP/CPCM(toluene)) for the conformers, some patterns became clear, see Table 3 for binding energies and free energies of the most stable conformers. The corresponding structures for these lowest-energy dimers are presented in Fig. 3. First of all, the dimers that contain nitrogen or sulfoxide tend to have larger binding energies than the molecules with sulfur, oxygen, or no heteroatoms. Second, the Gibbs free energy of dimerization ( $\Delta G_{\text{dim}}$ ) for every conformer that we found was positive. These positive  $\Delta G_{\text{dim}}$  values imply that none of these homodimers exist in significant quantities, in the gas phase or in toluene solution, at standard conditions. However, we can still gain insight into the relative energetics of binding in these and other systems by studying these trends. It should be noted that the positive Gibbs free

energies are the result of the dominance of the entropic term ( $-T\Delta S_{\text{dim}}$ ) over the enthalpic term ( $\Delta H_{\text{dim}}$ ), see Tables 4 and 5. At low temperatures, such as in a supersonic jet expansion, the entropic term would become much smaller, the Gibbs free energies of dimerization would become negative, and we would expect to find stable dimers.

## 4 Discussion

While the following sections are specific to the currently presented set of homodimers, the corresponding analyses can, in principle, be applied to other dimers, including both homodimers and heterodimers. Note that we do not discuss in any detail energy decomposition analyses (EDA), such as the local energy decomposition (LED)<sup>48</sup> available in ORCA for DLPNO-CCSD(T),<sup>49</sup> as these are beyond the scope of the present Tutorial Review. That said, LED results for a small number of dimers are presented in Table S8. Unfortunately, no clear trends emerged for these homologous dimers in the energy components, and so these specific results are not analyzed in detail.

### 4.1 Identification of homologous conformers

For all dimer conformers, the xyz coordinates were transformed into displacement coordinates, as described in the Methods section (Section 2) and SI (Section S1). All transformed coordinates are available in the SI. Two groups of monomers in our study can be considered homologous. Phen and BzQuin are homologues, with the only difference being the substitution of N for CH at one position on the ring system. Similarly, DBF, DBT, DBTO, Fluor, and MeCarb can all be considered homologues of each other, with the ring system being mostly identical and only the apical position of the 5-membered ring and the groups pendant to it varying. Since the monomers are homologous, it is possible for them to form homologous



**Fig. 3** Lowest energy conformer of each homodimer, as determined at the  $\omega$ B97X-D4/def2-QZVPP/SMD(toluene)// $\omega$ B97X-D4/def2-SVP/CPCM(toluene) level of theory. Full ensembles, as xyz coordinates, can be found in the SI.



**Table 4** Homologous dimers – phenanthrene (Phen) and benzo[a]quinoline (BzQuin). Binding energies ( $E_b$ ), Gibbs free energies of dimerization ( $\Delta G_{\text{dim}}$ ), Enthalpy of dimerization ( $\Delta H_{\text{dim}}$ ), entropic component of dimerization ( $-T\Delta S_{\text{dim}}$ ), total dipole moments (dipole), dipole cancellation (Cancel.), contact surface area (area), and dispersion, computed at the  $\omega$ B97X-V/def2-QZVP/SMD(toluene)// $\omega$ B97X-D4/def2-SVP/CPCM(toluene) level of theory. Gas phase data is available in the SI

| Conformer       | $E_b$<br>(kJ mol <sup>-1</sup> ) | $\Delta G_{\text{dim}}^a$<br>(kJ mol <sup>-1</sup> ) | $\Delta H_{\text{dim}}$<br>(kJ mol <sup>-1</sup> ) | $-T\Delta S_{\text{dim}}$<br>(kJ mol <sup>-1</sup> ) | Dipole (D) | Cancel. (D) | Area (Å <sup>2</sup> ) | Dispersion<br>(kJ mol <sup>-1</sup> ) |
|-----------------|----------------------------------|--|--|--|------------|-------------|------------------------|---------------------------------------|
| Phen-106SR      | 25.2                             | 29.2   | -20.4  | 48.6   | 0.012      | 0.021       | 96.6                   | 64.3                                  |
| BzQuin-76SR-F   | 28.3                             | 26.3   | -23.3  | 49.5   | 0.747      | 3.882       | 86.3                   | 59.1                                  |
| BzQuin-106SR    | 27.3                             | 27.3   | -22.3  | 49.5   | 2.521      | 2.109       | 90.4                   | 62.8                                  |
| BzQuin-73SR-F   | 25.4                             | 29.0   | -20.4  | 49.4   | 3.803      | 0.827       | 94.8                   | 62.8                                  |
| Phen-44X        | 24.6                             | 29.6   | -19.8  | 49.4   | 0.047      | -0.014      | 88.9                   | 65.2                                  |
| BzQuin-43X      | 25.1                             | 29.6   | -20.1  | 49.7   | 3.395      | 0.695       | 99.7                   | 61.2                                  |
| BzQuin-145X-F   | 24.3                             | 30.1   | -19.3  | 49.5   | 3.024      | 1.606       | 98.7                   | 60.4                                  |
| Phen-58G        | 24.5                             | 28.8   | -19.6  | 48.4   | 0.139      | -0.106      | 92.3                   | 64.1                                  |
| BzQuin-116SR-F1 | 28.2                             | 24.4   | -23.0  | 47.4   | 2.133      | 2.497       | 102.5                  | 62.4                                  |
| BzQuin-68SR     | 26.4                             | 26.7   | -21.3  | 48.0   | 3.485      | 1.144       | 88.1                   | 60.8                                  |
| BzQuin-62G      | 25.9                             | 27.5   | -20.8  | 48.3   | 3.677      | 0.952       | 96.6                   | 64.3                                  |
| BzQuin-116SR-F2 | 25.0                             | 28.0   | -19.9  | 47.9   | 4.290      | 0.340       | 89.6                   | 60.6                                  |
| Phen-63G        | 24.2                             | 29.2   | -19.4  | 48.7   | 0.159      | -0.126      | 91.0                   | 62.2                                  |
| BzQuin-61G      | 25.3                             | 28.0   | -20.4  | 48.4   | 3.706      | 0.923       | 94.5                   | 61.5                                  |
| BzQuin-116G-F   | 24.2                             | 29.1   | -19.3  | 48.4   | 4.278      | 0.351       | 83.7                   | 63.1                                  |
| Phen-180G1      | 23.5                             | 29.7   | -18.9  | 48.6   | 0.000      | 0.033       | 83.0                   | 54.6                                  |
| BzQuin-180G     | 26.9                             | 25.7   | -21.9  | 47.6   | 0.001      | 4.629       | 95.2                   | 62.6                                  |
| BzQuin-4G-F     | 26.0                             | 27.9   | -20.9  | 48.8   | 1.804      | 2.825       | 87.4                   | 59.8                                  |
| Phen-151SR      | 22.3                             | 31.0   | -20.9  | 48.8   | 0.179      | -0.145      | 92.6                   | 60.4                                  |
| BzQuin-151SR    | 27.4                             | 27.0   | -22.1  | 49.1   | 0.965      | 3.665       | 83.6                   | 59.1                                  |
| BzQuin-35SR-F   | 27.0                             | 27.0   | -21.8  | 48.8   | 0.859      | 3.770       | 83.5                   | 59.9                                  |
| BzQuin-147SR    | 25.2                             | 30.2   | -20.1  | 50.3   | 1.125      | 3.505       | 75.2                   | 58.9                                  |
| BzQuin-25SR-F   | 24.9                             | 28.5   | -19.9  | 48.3   | 2.641      | 1.988       | 89.6                   | 63.4                                  |

<sup>a</sup>  $\Delta G_{\text{dim}}$  may not exactly equal  $\Delta H_{\text{dim}} - T\Delta S_{\text{dim}}$  due to rounding errors.

dimers, where the only difference between dimers is in the substituted portions of the monomers. In practice, substitutions such as N for CH or N-Me for C=O cause small distortions both within the monomers and in their relative displacement and rotation, but these are generally small enough that the conformers are recognizably homologous. Note that, for monomers of lower symmetry, such as BzQuin, multiple dimer conformers may be homologous to the same (computed or hypothetical) dimer conformer of the higher-symmetry species, in this case, phenanthrene. In such cases, the multiple conformers of the lower-symmetry dimer can also be considered to be homologous to each other. Studying the differences between homologous conformers should help to identify trends and better understand the effects of heteroatom substitution on the non-covalent interactions of polycyclic aromatic systems.

With transformed coordinates for all dimers in hand, those with homologous monomers and similar rotations (within 15°), inclinations (within 5°), and displacements (within 0.5 Å) were considered to be homologous dimer conformations. For Phen and BzQuin, six homologous arrangements were found. Six of the nine conformations of the Phen dimer, and seventeen of the twenty-two conformations of the BzQuin dimer had homologues, while the remaining conformers were more unique. For the other five homologous monomers – DBT, DBTO, DBF, Fluor, and MeCarb – the overlap in their ensembles was

substantially less, such that no arrangements were found with homologous representatives from all five dimers. However, three arrangements were found with representatives from four of the five dimers, and two with representatives from three of the five dimers. A sixth exceptional arrangement shared by only DBT and DBTO is also studied here. These six sets of homologous arrangements covered four of the eight conformations of the DBF dimer, five of the seven conformations of the DBT dimer, twelve of the twenty-two conformations of the DBTO dimer, three of the five conformations of the Fluor dimer, and two of the five conformations of the MeCarb dimer. The twelve homologous arrangements under study are illustrated in Fig. 4.

#### 4.2 Effects of substitution

Phen and BzQuin are isoelectronic, with an equal number of electrons in molecular orbitals of similar shape, which makes comparison of their homologous conformers very straightforward. The computed dipoles, binding energies, and Gibbs free energies of binding of the different homologous conformers are listed in Table 4. We see that for a given arrangement, substituting a CH unit with N on each monomer can increase the binding energy by as much as 5.09 kJ mol<sup>-1</sup>, or decrease it by as much as 3.98 kJ mol<sup>-1</sup>. In terms of Gibbs free energy of binding, the energy can decrease (become more stable) by up to 4.39 kJ mol<sup>-1</sup>, or increase by up to 3.77 kJ mol<sup>-1</sup>. It is of



**Table 5** Homologous dimers – dibenzothiophene (DBT), dibenzothiophene oxide (DBTO), dibenzofuran (DBF), fluorenone (Fluor), and *N*-methylcarbazole (MeCarb). Binding energies ( $E_b$ ), Gibbs free energies of dimerization ( $\Delta G_{\text{dim}}$ ), enthalpy of dimerization ( $\Delta H_{\text{dim}}$ ), entropic component of dimerization ( $-T\Delta S_{\text{dim}}$ ), total dipole moments (dipole), dipole cancellation (Cancel.), contact surface area (area), and dispersion, computed at the  $\omega$ B97X-V/def2-QZVPP/SMD(toluene)// $\omega$ B97X-D4/def2-SVP/CPCM(toluene) level of theory. Gas phase data is available in the SI

| Conformer    | $E_b$<br>(kJ mol <sup>-1</sup> ) | $\Delta G_{\text{dim}}^a$<br>(kJ mol <sup>-1</sup> ) | $\Delta H_{\text{dim}}$<br>(kJ mol <sup>-1</sup> ) | $-T\Delta S_{\text{dim}}$<br>(kJ mol <sup>-1</sup> ) | Dipole (D) | Cancellation (D) | Contact area (Å <sup>2</sup> ) | Dispersion<br>(kJ mol <sup>-1</sup> ) |
|--------------|----------------------------------|--|--|--|------------|------------------|--------------------------------|---------------------------------------|
| DBT-92SR     | 23.6                             | 30.5   | -18.1  | 48.5   | 1.203      | 1.051            | 121.1                          | 62.3                                  |
| DBTO-95SR-P  | 23.3                             | 31.6   | -17.9  | 49.6   | 7.433      | 3.336            | 136.1                          | 64.9                                  |
| DBTO-92SR-I  | 21.6                             | 35.4   | -16.1  | 51.5   | 6.765      | 4.005            | 132.8                          | 62.4                                  |
| DBF-90SR     | 24.0                             | 29.3   | -18.4  | 47.7   | 0.941      | 0.786            | 86.7                           | 49.7                                  |
| DBT-157SR    | 23.4                             | 30.5   | -18.4  | 48.9   | 0.408      | 1.845            | 108.6                          | 61.2                                  |
| DBTO-159SR-P | 30.9                             | 26.8   | -25.5  | 52.2   | 5.552      | 5.217            | 122.9                          | 59.4                                  |
| DBTO-20SR-I  | 30.9                             | 27.5   | -25.4  | 52.9   | 1.857      | 8.912            | 116.4                          | 58.5                                  |
| DBTO-22SR-O  | 26.6                             | 28.2   | -21.5  | 49.6   | 1.989      | 8.781            | 124.2                          | 61.4                                  |
| DBT-13SR     | 23.0                             | 30.8   | -17.9  | 48.9   | 2.087      | 0.166            | 89.5                           | 54.6                                  |
| DBTO-164SR-O | 23.6                             | 29.4   | -18.5  | 47.9   | 9.460      | 1.309            | 149.4                          | 67.1                                  |
| DBTO-8SR-P   | 20.0                             | 36.3   | -15.0  | 51.2   | 10.170     | 0.599            | 130.6                          | 64.2                                  |
| DBF-18SR     | 21.7                             | 28.9   | -16.6  | 45.5   | 1.605      | 0.123            | 93.4                           | 53.9                                  |
| DBT-62SR     | 22.5                             | 31.5   | -17.6  | 49.1   | 1.555      | 0.699            | 104.5                          | 59.7                                  |
| DBTO-114SR-O | 25.0                             | 30.7   | -19.8  | 50.5   | 7.597      | 3.172            | 107.4                          | 59.6                                  |
| DBTO-66SR-P  | 23.0                             | 32.9   | -17.9  | 50.8   | 8.665      | 2.104            | 127.1                          | 64.9                                  |
| Fluor-73SR   | 24.6                             | 30.4   | -19.5  | 49.9   | 5.943      | 2.267            | 118.6                          | 60.5                                  |
| MeCarb-76SR  | 29.5                             | 27.3   | -24.1  | 51.4   | 3.372      | 1.205            | 76.8                           | 64.0                                  |
| DBT-41X      | 21.8                             | 31.7   | -16.8  | 48.5   | 1.764      | 0.489            | 112.4                          | 61.8                                  |
| DBTO-139X-O  | 25.2                             | 30.7   | -20.1  | 50.7   | 8.668      | 2.101            | 125.3                          | 64.4                                  |
| DBTO-40X-P   | 20.4                             | 36.7   | -15.3  | 52.0   | 9.534      | 1.236            | 144.8                          | 67.6                                  |
| DBF-44X      | 22.2                             | 33.5   | -16.6  | 45.5   | 1.433      | 0.295            | 90.3                           | 54.4                                  |
| Fluor-45X    | 25.9                             | 29.5   | -20.9  | 50.4   | 7.084      | 1.125            | 134.8                          | 64.1                                  |
| DBT-104X     | 21.6                             | 32.8   | -16.9  | 49.6   | 1.096      | 1.158            | 111.8                          | 60.9                                  |
| DBTO-73X-O   | 25.3                             | 31.4   | -20.2  | 51.7   | 5.395      | 5.374            | 113.5                          | 47.6                                  |
| DBTO-105X-P  | 18.5                             | 37.7   | -13.8  | 51.4   | 6.629      | 4.140            | 111.0                          | 62.8                                  |
| Fluor-106X   | 24.2                             | 31.9   | -19.3  | 51.2   | 4.344      | 3.866            | 118.2                          | 63.8                                  |
| MeCarb-107SR | 30.6                             | 27.1   | -25.0  | 52.1   | 2.467      | 2.110            | 80.2                           | 63.6                                  |

<sup>a</sup>  $\Delta G_{\text{dim}}$  may not exactly equal  $\Delta H_{\text{dim}} - T\Delta S_{\text{dim}}$  due to rounding errors.

interest to know whether these changes in binding energies can be correlated with other physical observables of the dimers.

One possibility would be to look for a correlation between binding energy and dipole moment. Certainly, dimers with a larger dipole moment will interact more strongly with solvent, which may increase the binding energy. At first glance, this would seem to fit well with the observation that heteroatoms, which increase the overall dipole moment, also tend to lead to larger binding energies. However, a large dipole moment for the dimer also implies that the dipoles of the monomers are aligned, which should carry an energy penalty. A more nuanced descriptor for dimers of polar monomers is dipole cancellation, *i.e.*, the extent to which the dipoles of the monomers are opposed and cancel each other out. Dipole cancellation can be calculated by subtracting the dipole of the dimer from the sum of the dipoles of the monomers.

A further concept that should affect the binding energy is the extent of  $\pi$ - $\pi$  stacking. Two simple approaches to quantifying this stacking would be by taking the energy contribution from dispersion, or the contact surface area, which are both easily extracted from standard quantum chemistry output files. The contact surface area is the area of each molecule in the

dimer in contact with the other monomer. A simple way to extract this value is to take the solvent-accessible surface area (SASA) of the monomers, subtract the SASA of the dimer, and then divide by two.

For the Phen and BzQuin dimers,  $E_b$  correlates reasonably well with dipole cancellation: linear regression gives an  $R^2$  of 0.517, where the fitting parameters are  $a = 0.8916$  kJ mol<sup>-1</sup> D<sup>-1</sup> and  $b = 23.877$  kJ mol<sup>-1</sup>. The correlation plot of  $E_b$  vs. dipole cancellation for all Phen and BzQuin dimers can be found in Fig. 5. Correlation is better yet within homologous series, where the differences in sterics and contact area are small enough to be ignored. For the three homologous series with at least four dimers, the 106SR series has an  $R^2$  of 0.9527, the 58SR series has an  $R^2$  of 0.9922, and the 151SR series has an  $R^2$  of 0.8609. The scatter plot for these three homologous series can be found in Fig. S2 in the SI. Total dipole moment, contact surface area, and dispersion were all found to correlate poorly with binding energy, but are reported for the interested reader. Numerical results for all conformers are documented in Tables 4, 5 and Table S7.

For the remaining five homologous monomers, comparison of homologous dimers is less straightforward, but still



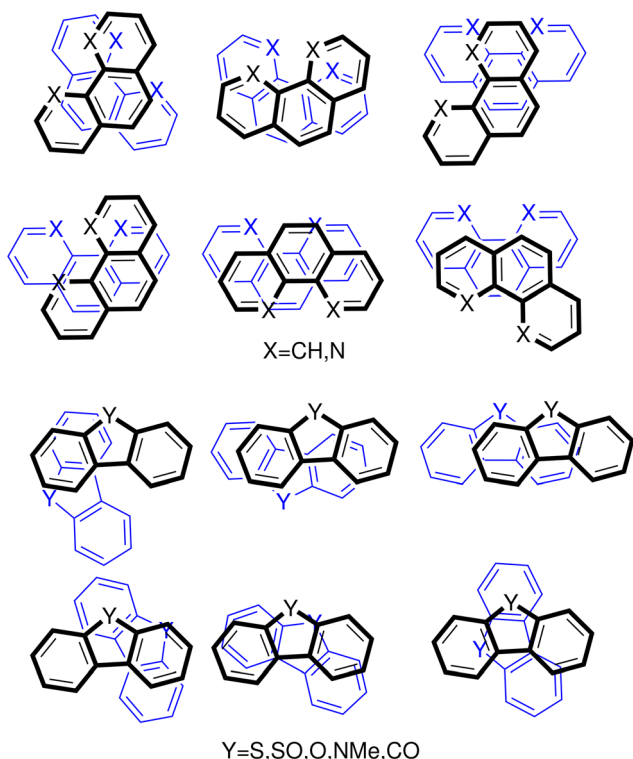


Fig. 4 Homologous arrangements of dimers under study, with variable portions indicated by X or Y.

insightful. The homologous dimers, along with their binding energies, free energies of dimerization, dipole moments, dipole cancellation, SASA, and dispersion contributions, are listed in

Table 5. There is greater geometric variation between the monomers, and there are differences in the number of electrons, leading to greater differences in energy upon substitution. DBT is the highest-symmetry monomer with representation in all six homologous series, so its energies are used as the reference energies. Substituting S with O decreases binding energy by between  $0.6 \text{ kJ mol}^{-1}$  and  $3.3 \text{ kJ mol}^{-1}$ . Substitution of S with S=O can reduce the binding energy by as much as  $0.8 \text{ kJ mol}^{-1}$ , or increase it by as much as  $20.2 \text{ kJ mol}^{-1}$ . Substituting S with C=O improves binding energy by  $2.4 \text{ kJ mol}^{-1}$  to  $6.4 \text{ kJ mol}^{-1}$ . Substituting S with N-Me improves binding energy by  $11.5 \text{ kJ mol}^{-1}$  to  $13.9 \text{ kJ mol}^{-1}$ . These monomers exhibit substantially more variation than do Phen and BzQuin, and no two monomers are isoelectronic. However, the ring system is largely unchanged, allowing for some comparisons to be made. Unlike the Phen and BzQuin dimers, there is little correlation between  $E_b$  and dipole cancellation when all the dimers are considered, as  $R^2$  is only 0.0965. However, six of the DBTO dimers have significant departures from co-planarity induced by the geometry of the S=O side chains, which begin to disrupt the  $\pi$ - $\pi$  stacking. For our purposes, we consider  $\pi$ - $\pi$  stacking to be partially disrupted when the best-fit planes of the two monomers are inclined by at least  $8^\circ$  relative to each other. If these six dimers are removed from consideration, the  $R^2$  improves to 0.4288; see Fig. 6 showing  $E_b$  vs. dipole cancellation for all the DBT, DBTO, DBF, Fluor, and MeCarb dimers. The correlation is still poorer than that for the Phen and BzQuin dimers, but is large enough to indicate that dipole cancellation is a significant factor in determining the binding energy for a given dimer conformation. The inferior correlation is likely the result of the monomers being

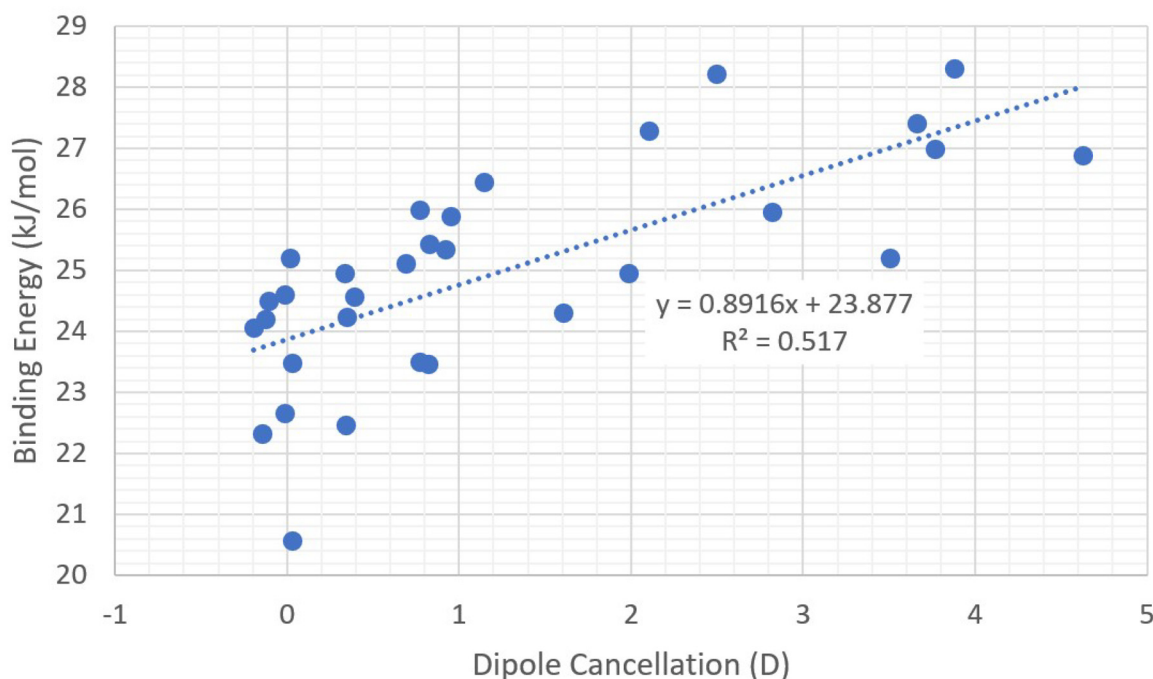


Fig. 5 Scatterplot of  $E_b$  vs. dipole cancellation for all homodimers of Phen and BzQuin, computed at the  $\omega$ B97X-V/def2-QZVPP/SMD(toluene)// $\omega$ B97X-D4/def2-SVP/CPCM(toluene) level of theory. Note that Phen homodimers are capable of negative dipole cancellation when charge transfer between monomers is greater than the original (small) dipoles.





Fig. 6 Scatterplot of  $E_b$  vs. dipole cancellation for all homodimers of DBT, DBTO, DBF, Fluor, and MeCarb, computed at the  $\omega$ B97X-V/def2-QZVPP/SMD(toluene)// $\omega$ B97X-D4/def2-SVP/CPCM(toluene) level of theory. The six DBTO homodimers with partially disrupted  $\pi$ - $\pi$  stacking are indicated with open circles, and do not contribute to the linear regression.

less similar to each other than the similarity between Phen and BzQuin. In dimers of DBT, DBTO, DBF, Fluor, and MeCarb, there is greater variation in contact surface area, shape, and polarizability, which will complicate the results, and make correlations harder to find. If the correlation is only considered within homologous series, then  $R^2$  ranges from 0.1135 to 0.7014; see Fig. S3 in the SI.

## 5 Conclusions

In this Tutorial Review, we have outlined a useful computational workflow for determining conformational ensembles. The underlying approach has been showcased by examining the pyrene dimer and then further demonstrated by generating ensembles for eight new polycyclic (hetero-)aromatic molecules.

We have shown for the pyrene dimer that, of the available 3c methods, B97-3c gives the best accuracy for computational cost, when compared to  $\omega$ B97X-D4/def2-TZVP results. Screening CREST ensembles using B97-3c reduces the size of the ensembles by filtering out conformers that are much higher in energy, and by refining the geometries such that many conformers become indistinguishable from each other. In addition, the conformer geometries produced by B97-3c which are not eliminated from consideration are much closer to those given by  $\omega$ B97X-D4/def2-SVP than those given by the GFN2-xTB method, upon which CREST depends. This multi-step approach saves a great deal of optimization time per conformer, as the computationally expensive  $\omega$ B97X-D4/def2-SVP method takes many fewer steps to optimize than if the optimization was started directly from the CREST-generated conformer. All these effects combine to mean that screening of CREST ensembles using B97-3c before DFT refinement saves a great deal of computational resources, and thus will allow for the generation of large,

high quality ensembles of larger complexes than would otherwise be feasible, similar to what others have found.<sup>29</sup>

We note that we have not applied the counterpoise correction at any point in this work. While it does, in principle, correct for basis set superposition error, the effects are negligible for DFT when using large basis sets such as def2-QZVPP, especially with dispersion-corrected functionals.<sup>50</sup> The counterpoise correction has also been found to be insignificant to improving DFT-optimized geometries,<sup>51</sup> and so we have saved computational overhead by neglecting it.

Several physically motivated analyses, outside of EDA, demonstrate useful insight that may be gained when considering dimer structures contributing to the conformational ensemble. For example, our results with the polycyclic (hetero-)aromatic dimers give valuable insight into the impact of heteroatom substitution on binding energy in these types of systems. We have shown that a significant amount of the variation in binding energy between dimers can be explained by dipole cancellation. This is why dimers of more polar compounds have greater potential for large binding energies, but not all conformers benefit from this extra binding energy: if the dipoles are not opposed, there is no significant dipole cancellation, and thus no significant benefit from the extra polarity. Investigations correlating other observable properties such as the total dipole, contact surface area, and distance between heteroatoms did not provide useful insights into binding energy changes. More work is needed to fully understand the factors that determine favourable association between polycyclic heteroaromatic compounds, including those larger than the 3-ring systems considered here, but dipole cancellation clearly plays a large role.

In the end, this Tutorial Review should prove useful for those interested in using computations to study dimerization,



or even nanoaggregation. It highlights the computational tools, as well as practicalities and pitfalls in generating robust conformational ensembles. These ensembles can be useful for studying both physical and spectroscopic properties.

## Author contributions

JJO performed most of the computations and wrote much of the original draft. NJK participated in conceptualization, performed some of the computations, undertook formal analysis, and participated in writing the original draft, reviewing, and editing. AB participated in conceptualization, supervised the project, and reviewed and edited the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

Much of the data supporting this article, including full xyz coordinates for all refined ensembles, have been included as part of the SI. Supplementary information: Transformation of coordinates, screening of 3c methods against ωB97X-V in solution, B97-3c free energy computations in both gas phase and solution, DLPNO-CCSD(T) binding energies plus local energy decomposition (LED) analysis for select dimers, and xyz structures of the complete ensembles of all nine dimers. See DOI: <https://doi.org/10.1039/d5cp01010a>.

For CREST and ORCA output files, please contact the authors.

## Acknowledgements

The authors thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding (NSERC Discovery Grant RGPIN-2020-04347 to AB and NSERC-USRA to JJO) and the Digital Research Alliance of Canada (alliance-can.ca) for computational resources.

## Notes and references

- 1 P. Jurečka, J. Šponer, J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.
- 2 X. Feng, J.-Y. Hu, C. Redshaw and T. Yamato, *Chem. – Eur. J.*, 2016, **22**, 11898–11916.
- 3 S. M. Elbert, M. Reinschmidt, K. Baumgärtner, F. Rominger and M. Mastalerz, *Eur. J. Org. Chem.*, 2018, 532–536.
- 4 H. S. Silva, A. C. R. Sodero, B. Bouyssiére, H. Carrier, J.-P. Korb, A. Alfara, G. Vallverdu, D. Bégué and I. Baraille, *Energy Fuels*, 2016, **30**, 5656–5664.
- 5 N. J. King and A. Brown, *Energy Fuels*, 2023, **37**, 12796–12810.
- 6 U. Ahmed, D. Sundholm and M. P. Johansson, *Phys. Chem. Chem. Phys.*, 2024, **26**, 27431–27438.
- 7 B. Schramm, M. Gray and J. M. Herbert, *J. Am. Chem. Soc.*, 2025, **147**, 3243–3260.
- 8 J. M. Herbert, *J. Phys. Chem. A*, 2021, **125**, 7125–7137.
- 9 N. J. Silva, F. B. C. Machado, H. Lischka and A. J. A. Aquino, *Phys. Chem. Chem. Phys.*, 2016, **18**, 22300–22310.
- 10 T. Janowski and P. Pulay, *J. Am. Chem. Soc.*, 2012, **134**, 17520–17525.
- 11 S. Grimme, *Angew. Chem., Int. Ed.*, 2008, **47**, 3430–3434.
- 12 J. W. G. Bloom and S. E. Wheeler, *Angew. Chem., Int. Ed.*, 2011, **50**, 7847–7849.
- 13 E. M. Cabaleiro-Lago and J. Rodríguez-Otero, *ACS Omega*, 2018, **3**, 9348–9359.
- 14 P. Pracht, F. Bohle and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 7169–7192.
- 15 P. Pracht, S. Grimme, C. Bannwarth, F. Bohle, S. Ehlert, G. Feldmann, J. Gorges, M. Müller, T. Neudecker, C. Plett, S. Spicher, P. Steinbach, P. A. Wesolowski and F. Zeller, *J. Chem. Phys.*, 2024, **160**, 114110.
- 16 I. Iribarren and C. Trujillo, *J. Chem. Inf. Model.*, 2022, **62**, 5568–5580.
- 17 S. Riniker and G. A. Landrum, *J. Chem. Inf. Model.*, 2015, **55**, 2562–2574.
- 18 N.-O. Friedrich, F. Flachsenberg, A. Meyder, K. Sommer, J. Kirchmair and M. Rarey, *J. Chem. Inf. Model.*, 2019, **59**, 731–742.
- 19 O. Sperandio, M. Souaille, F. Delfaud, M. A. Miteva and B. O. Villoutreix, *Eur. J. Med. Chem.*, 2009, **44**, 1405–1409.
- 20 H. Tsujishita and S. Hirono, *J. Comput.-Aided Mol. Des.*, 1997, **11**, 305–315.
- 21 N. M. O'Boyle, T. Vandermeersch, C. J. Flynn, A. R. Maguire and G. R. Hutchison, *J. Cheminf.*, 2011, **3**, 8.
- 22 B. de Souza, *Angew. Chem., Int. Ed.*, 2025, **64**, e202500393.
- 23 N. J. King and A. Brown, *J. Phys. Chem. A*, 2022, **126**, 4931–4940.
- 24 S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 2847–2862.
- 25 M. Alshalalfeh, N. Sun, A. H. Moraes, A. P. A. Utani and Y. Xu, *Molecules*, 2023, **28**, 4013.
- 26 A. S. Perera, C. D. Carlson, J. Cheramy and Y. Xu, *Chirality*, 2023, **35**, 718–731.
- 27 T. P. Curran, A. Marrone, L. M. Davidson, N. Pokharel, J. F. Frempong, I. Tolbatov, M. L. Phillip, C. B. Gober, H. Yang and J. Stewart, *Pept. Sci.*, 2022, **114**, e24286.
- 28 T. Gensch, G. dos Passos Gomes, P. Friederich, E. Peters, T. Gaudin, R. Pollice, K. Jorner, A. Nigam, M. Lindner-D'Addario, M. S. Sigman and A. Aspuru-Guzik, *J. Am. Chem. Soc.*, 2022, **144**, 1205–1217.
- 29 S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher and M. Stahn, *J. Phys. Chem. A*, 2021, **125**, 4039–4054.
- 30 T. Froitzheim, M. Müller, A. Hansen and S. Grimme, *ChemRxiv*, 2025, preprint, DOI: [10.26434/chemrxiv-2025-bjxvt](https://doi.org/10.26434/chemrxiv-2025-bjxvt).
- 31 R. Sure and S. Grimme, *J. Comput. Chem.*, 2013, **34**, 1672–1685.
- 32 J. G. Brandenburg, C. Bannwarth, A. Hansen and S. Grimme, *J. Chem. Phys.*, 2018, **148**, 064104.
- 33 S. Grimme, J. G. Brandenburg, C. Bannwarth and A. Hansen, *J. Chem. Phys.*, 2015, **143**, 054107.



- 34 S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes, *J. Chem. Phys.*, 2021, **154**, 064103.
- 35 M. Müller, A. Hansen and S. Grimme, *J. Chem. Phys.*, 2023, **158**, 014103.
- 36 B. B. Mészáros, K. Kubicskó, D. D. Németh and J. Daru, *J. Chem. Theory Comput.*, 2024, **20**, 7385–7392.
- 37 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**, e1606.
- 38 B. Helmich-Paris, B. de Souza, F. Neese and R. Izsák, *J. Chem. Phys.*, 2021, **155**, 104109.
- 39 V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.
- 40 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
- 41 A. Najibi and L. Goerigk, *J. Comput. Chem.*, 2020, **41**, 2562–2572.
- 42 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- 43 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.
- 44 N. Mardirossian and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9904–9924.
- 45 F. Weigend, F. Furche and R. Ahlrichs, *J. Chem. Phys.*, 2003, **119**, 12753–12762.
- 46 V. S. Bryantsev, M. S. Diallo and W. A. Goddard III, *J. Phys. Chem. B*, 2008, **112**, 9709–9719.
- 47 C. Gonzalez and E. C. Lim, *J. Phys. Chem. A*, 2003, **107**, 10105–10110.
- 48 W. B. Schneider, G. Bistoni, M. Sparta, M. Saitow, C. Riplinger, A. A. Auer and F. Neese, *J. Chem. Theory Comput.*, 2016, **12**, 4778–4792.
- 49 Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo and F. Neese, *J. Chem. Phys.*, 2018, **148**, 011101.
- 50 A. Malloum and J. Conradie, *J. Phys. Chem. A*, 2024, **128**, 10775–10784.
- 51 R. Fodil, M. Sekkal-Rahal and A. Sayede, *J. Mol. Model.*, 2017, **23**, 31.

