



Cite this: *Phys. Chem. Chem. Phys.*,  
2025, 27, 11343

# Prediction of solution phase association constants by mapping contact points in intermolecular complexes†

Katarzyna J. Zator  and Christopher A. Hunter \*

Atomic surface site interaction points (AIP) provide a complete description of the non-covalent interactions that one molecule can make with another. The surface site interaction model for the properties of liquids at equilibrium (SSIMPLE) algorithm can be used to calculate the free energy change associated with the pairwise interaction between two AIPs on two different molecules in any solvent. Summing these pairwise AIP interactions across an intermolecular interface that involves multiple interacting sites can be used to calculate solution phase binding free energies and association constants. A computational tool that converts the three-dimensional structure of a complex into a set of AIP contacts has been developed along with a visualisation tool to display AIP interaction maps, allowing straightforward identification of the key intermolecular contacts that contribute most to the overall binding free energy in a complex. The method successfully reproduces solution phase association constants (to within an order of magnitude) for a range of host–guest complexes involving H-bonding, aromatic and hydrophobic interactions, but performs less well for halogen-bonds and complexes involving interactions between the extended  $\pi$ -surfaces of fullerene-type compounds.

Received 29th January 2025,  
Accepted 13th May 2025

DOI: 10.1039/d5cp00398a

rsc.li/pccp

## 1. Introduction

The analysis of the nature of interactions that stabilise intermolecular complexes is central to the fields of medicinal and supramolecular chemistry.<sup>1–7</sup> The design of complementary interacting partners is usually based on the identification of specific sites that can be used for formation of attractive non-covalent interactions. Although great progress has been made in understanding the relationship between non-covalent interaction energies and chemical structure for simple functional group interactions, the prediction of solution-phase binding affinities for more complex intermolecular interfaces that involve multiple interaction sites remains a real challenge.<sup>8,9</sup> Molecular mechanics and molecular dynamics simulations can provide insights into the three-dimensional structures of such interfaces and allow identification of key non-covalent interactions.<sup>10</sup> However, conversion of structural information into the free energy change associated with binding is significantly more challenging.<sup>11,12</sup>

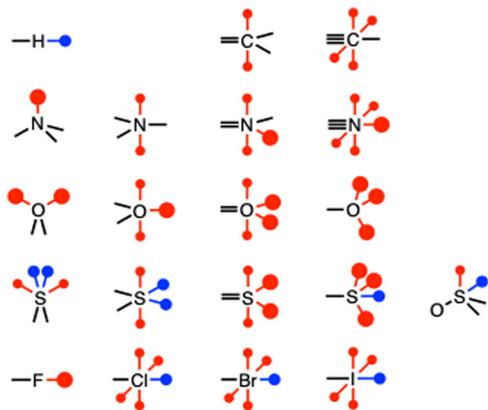
Free energy perturbation techniques show some promise, but these tools have been developed for simulation of biomolecular systems in aqueous solution, and it is not clear whether they have wider applicability.<sup>13</sup> Here we describe a different approach to the prediction of solution phase binding free energies based on summing the contributions due to pairwise functional group interactions between two molecules.

We have been developing computational tools for the analysis of binding interfaces based on the molecular surface properties of the interacting molecules.<sup>14–21</sup> Atomic surface site interaction points (AIP) are used to describe individual interaction sites on the surface of a molecule and can be calculated *ab initio* from molecular electrostatic potential surfaces (MEPS) using density functional theory (DFT).<sup>22</sup> Each AIP represents  $9 \text{ \AA}^2$  on the van der Waals surface of the molecule (*i.e.* the footprint of a H-bonding interaction)<sup>20</sup> and is assigned an interaction parameter,  $\epsilon_i$ , based on the maximum or minimum electrostatic potential for that patch of surface. This set of AIPs describes all possible interactions that the molecule can make with the surroundings, and the SSIMPLE algorithm can be used to calculate solvation energies for each AIP in any solvent or solvent mixture.<sup>17,20</sup> We have previously shown that this approach can be used to make accurate predictions of phase transfer free energies between two solvents (*e.g.* water and *n*-hexadecane) and to identify cocrystal cofomers by virtual screening.<sup>23</sup> Here, we use AIPs to predict association constants for the formation of intermolecular complexes by summing the free energy contributions due to pairwise AIP contacts in the binding interface.

Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: herchelsmith.orgchem@chem.ac.uk

† Electronic supplementary information (ESI) available: Details of the algorithm, polynomials for calculation of solvation energies in 261 different solvents, AIP contacts in host–guest complexes. The code to calculate AIPs is available at <https://github.com/k-zator/AIP> and [https://github.com/k-zator/AIP\\_MEPS](https://github.com/k-zator/AIP_MEPS). The code to identify AIP–AIP pairings has been developed in Python and has been documented and deployed at [https://github.com/k-zator/AIP\\_map](https://github.com/k-zator/AIP_map). See DOI: <https://doi.org/10.1039/d5cp00398a>





**Fig. 1** Atomic interaction point (AIP) representation of different functional groups. Tetravalent carbon, sulfur, phosphorus and silicon have no AIPs. The large red dots represent polar H-bond acceptor sites (lone pairs) assigned using the  $0.0300 \text{ e Bohr}^{-3}$  electron density isosurface. The small red dots represent non-polar H-bond acceptor sites (e.g.  $\pi$ -systems) assigned using the  $0.0020 \text{ e Bohr}^{-3}$  electron density isosurface, and these sites are treated as fractional AIPs:  $f = 1.0$  for iodine,  $0.75$  for bromine,  $0.5$  for chlorine and carbon,  $0.25$  for nitrogen and  $0.0$  for oxygen. The blue dots represent H-bond donors and  $\sigma$ -holes assigned using the  $0.0104 \text{ e Bohr}^{-3}$  electron density isosurface (sulfur  $\sigma$ -holes are fractional AIPs  $f = 0.5$ ).

**Fig. 1** shows the AIP representation of different functional groups. In order to obtain the AIP description of a molecule, three different MEPS are calculated using DFT.<sup>22</sup> Minima on the  $0.0300 \text{ e Bohr}^{-3}$  electron density isosurface identify the H-bond acceptor sites highlighted in **Fig. 1**, and the corresponding interaction parameters are assigned using eqn (1).

$$\varepsilon_i = 0.0336E_{\min} + c_{\beta} \quad (1)$$

where  $E_{\min}$  is the minimum in the MEPS in  $\text{kJ mol}^{-1}$ , and  $c_{\beta}$  is a constant that depends on the functional group (see ESI† for details).

Local minima on patches of the  $0.0020 \text{ e Bohr}^{-3}$  electron density isosurface associated with the non-polar acceptor sites highlighted in **Fig. 1** are used in eqn (2) to obtain the corresponding interaction parameters.

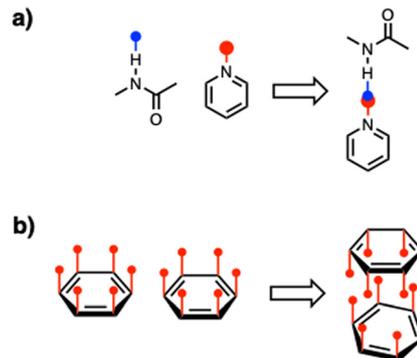
$$\varepsilon_i = 0.0232E_{\min} \quad (2)$$

Some of these non-polar H-bond acceptor AIPs are also assigned a fractional parameter  $f$  to signify that they represent a relatively small surface area, which is important for the treatment of van der Waals or dispersion interactions (see below). Local maxima on patches of the  $0.0104 \text{ e Bohr}^{-3}$  electron density isosurface associated with the H-bond donor and  $\sigma$ -hole sites highlighted in **Fig. 1** are used in eqn (3) to obtain the corresponding interaction parameters.

$$\varepsilon_i = 0.0132E_{\max} - 2.80 \quad (3)$$

For non-polar positive AIPs associated with CH and SH groups, local maxima on patches of the  $0.0104 \text{ e Bohr}^{-3}$  electron density isosurface were used in eqn (4) to obtain the interaction parameters.

$$\varepsilon_i = 0.0078E_{\max} - 0.64 \quad (4)$$



**Fig. 2** Atomic interaction point (AIP) description of (a) a H-bonding interaction between an amide and a pyridine, and (b) an aromatic stacking interaction.

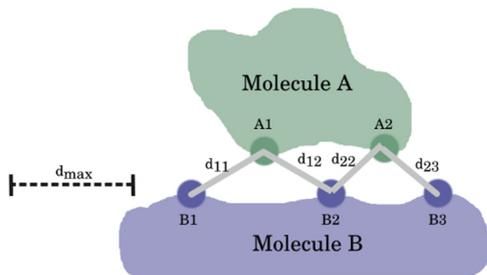
**Fig. 2** illustrates the approach for two molecules that form a 1 : 1 complex. Each molecule is represented by a set of AIPs, and the interaction between the two molecules is described by a set of pairwise interactions between AIPs on one molecule and AIPs on the other. The free energy change associated with each AIP contact can be calculated from the corresponding interaction parameters, as described previously,<sup>22,24</sup> and the total free energy change for formation of the complex can be obtained by summing pairwise AIP interactions.<sup>14</sup> For H-bonding interactions (**Fig. 2a**), when the AIPs of the two functional groups are projected onto the three-dimensional structure of the complex, the H-bond donor AIP coincides in space with the H-bond acceptor AIP, so this contact can be straightforwardly and unambiguously identified. In general, non-covalent interaction interfaces are not so simple, and the AIPs representing two interacting sites may not coincide as neatly as in **Fig. 2a**. For example, **Fig. 2b** shows an aromatic stacking interaction. In this case, when the AIPs of the two functional groups are projected onto the three-dimensional structure of the complex, multiple AIPs are close in space, and it is not obvious which AIPs should be paired to best represent the interaction.

Here we describe a general solution to the problem of identifying a unique set of pairwise AIP contacts to describe the non-covalent interactions present in an intermolecular complex. We use a collection of supramolecular host-guest complexes to show that this new computational approach can be used to analyse the three-dimensional structure of a complex in order to obtain a reliable estimate of the solution phase association constant in a wide range of different solvents.

## 2. Approach

**Fig. 3** shows a simple example of the general problem of identifying AIP contacts in an intermolecular interaction interface. Molecule A has two AIPs, and molecule B has three AIPs, so there is a total of six possible pairwise AIP contacts. The criteria used to decide which set of pairings best represents the non-covalent interactions present in the interface are based on the distances between AIPs. Firstly, a threshold is applied that requires interacting AIPs to be closer than an upper distance limit,  $d_{\max}$ . In **Fig. 3**, A1 is close to B1 and B2, and A2 is close to





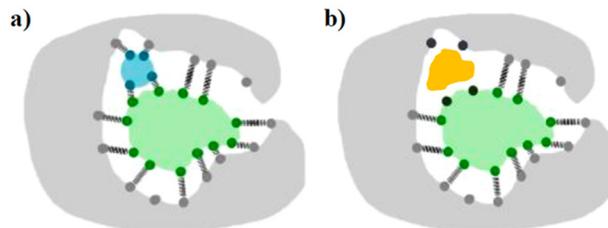
**Fig. 3** Schematic AIP representation of the interaction interface for the complex formed by molecule A and molecule B. AIPs on molecule A are labelled A1 and A2, and AIPs on molecule B are labelled B1, B2, and B3. Distances between AIPs that are closer than  $d_{\max}$  are labelled.

B2 and B3, so the  $d_{\max}$  criterion would eliminate A1-B3 and A2-B1 from the list of possible contacts. The remaining four contacts are highlighted in Fig. 3. The next criterion applied is that the number of AIP contacts should be maximised. In this case, both AIPs on molecule A make contacts with AIPs on molecule B, so there are two interactions. Three sets of pairwise AIP combinations are possible (A1-B1 and A2-B2, A1-B1 and A2-B3, or A1-B2 and A2-B3), and the criterion used to choose between them is based on the sum of the distances between pairs of interacting AIPs,  $d_{\text{total}}$ . For example, A1-B1 and A2-B2 would be selected, if  $(d_{11} + d_{22})$  is less than the two alternatives,  $(d_{11} + d_{23})$  and  $(d_{12} + d_{23})$ . As the number of AIPs in the interaction interface increases, the number of possible combinations increases rapidly, so the algorithm developed to locate AIP contacts involves a hierarchical treatment of interaction sites to reduce the scale of the problem.

### 3. Solvated interaction sites

The number of AIPs in the list of possible contacts can be reduced by eliminating all AIPs that remain in contact with the solvent in the complex. The solvent-accessible surface area (SASA) values were calculated for each AIP in a complex using an adapted version of the Shrake–Rupley algorithm implemented in mdtraj.<sup>25,26</sup> In this algorithm, each atom is described by a sphere, which has a radius equal to the van der Waals radius plus a probe radius. The solvent-accessible surface for each atom is defined by the points on the surface of this sphere that do not fall inside any of the other spheres. Each point on the atomic solvent-accessible surface was assigned to the nearest AIP in space, allowing calculation of the AIP SASA.

These SASA values were used to identify which AIPs are desolvated on binding and which remain in contact with the solvent. Parameterisation of the SASA criteria was based on X-ray crystal structures of protein–ligand complexes, which have interstitial water molecules in the binding interface. The three-dimensional structures of complexes from the CASF dataset were retrieved from the protein databank and analysed for water molecules that were resolved in the intermolecular binding interface. The AIPs that interact with the water molecules in these structures are easily identified as close contacts.



**Fig. 4** Schematic illustration of the approach used to determine SASA threshold parameters for identifying AIPs that are in contact with solvent in the structure of a complex. (a) The binding pocket of a protein (grey) that contains a ligand (green) and an interstitial water molecule (blue). The associated AIPs are depicted as darker dots and intermolecular AIP contacts are shown as cross-hatched lines. (b) The SASA calculated after removing the interstitial water molecule is shown in yellow, and the four navy AIPs that are known to be in contact with the interstitial water molecule fall on this surface.

For each structure, the interstitial water molecule was removed (highlighted in blue in Fig. 4a), and SASA were calculated for each AIP in the remaining pocket (highlighted in yellow in Fig. 4b). In general, the AIPs had been in contact with the water molecule in the original structure had the largest SASA values, but for seven complexes (1NC1, 1NC3, 3G2N, 3G2Z, 4K18, 4LLX, 4TWP), the SASA values were all rather small, and the discrimination between AIPs that did and did not contact the interstitial water molecule was less clear cut. These structures were therefore used to optimise the SASA criteria for identifying solvated AIPs.

The SASA probe radius was varied between 0.0 and 1.4 Å in increments of 0.05 Å, and the resulting AIP SASA values for all seven complexes were analysed for differences between the set of AIPs that contact water and the set of AIPs that do not. The 0.35 Å probe radius gave the largest difference, with 9.8 Å<sup>2</sup> providing the threshold value for assignment of a solvated site. This value can be compared with the footprint of an AIP: 9 Å<sup>2</sup> on the van der Waals surface of an oxygen atom is equivalent to 14 Å<sup>2</sup> projected onto the 0.35 Å probe radius SASA. In other words, an AIP is considered solvated, if more than 70% of the associated surface is exposed to solvent. For fractional AIPs, the threshold SASA was therefore scaled by the corresponding fractional parameter  $f$ , which represents the surface area of these sites. Although this parameterisation is based on finding a small number of sites solvated by water, the resulting analysis is based purely on the shape of the SASA of the solutes, and we assume that the parameters do not depend on the solvent or AIP type.

This SASA analysis eliminates a large number of AIPs, and only the remaining AIPs are candidates for involvement in intermolecular interactions that contribute to the overall stability of the complex.

### 4. H-bonds

H-bonds represent a special class of non-covalent interaction, because they are associated with close contacts that come well within the van der Waals (vdW) radii of the interacting atoms.<sup>27</sup> Thus H-bonds can be readily identified by simply using interatomic distances and the standard Baker–Hubbard definition as



implemented in mdtraj, *i.e.* a distance of less than 3.0 Å between the two heavy atoms.<sup>28</sup> H-bond acceptors were defined as any oxygen or nitrogen, excluding atom type N.pl3. The heavy atoms of H-bond donors were defined as any of the atom types O.3, N.3 or N.pl3 bonded to hydrogen. The distances between all H-bond donor and acceptor heavy atoms were calculated. For any donor–acceptor pair for which this distance was less than 3.0 Å, the distances between all of the AIPs on the acceptor atom and all of the AIPs on the hydrogen atom(s) of the donor were calculated. The pair of AIPs closest in space was identified as the AIP contact associated with the H-bond, provided the separation was less than 2.2 Å. The AIP distance criterion is shorter than the interatomic distance criterion for H-bonding to ensure that the two interacting AIPs lie between the two interacting atoms, *i.e.* the H-bond donor points towards the H-bond acceptor. All of the other AIPs associated with the H-bonded atoms were removed from the list of potential contacts, reducing the complexity of the AIP network and simplifying subsequent analysis.

## 5. Analysis of AIP networks

The remaining AIPs were divided into subgroups representing networks of possible pairings, and each network was considered independently of the others. For example, if two AIPs are closer than  $d_{\max}$  from one other, and both are more than  $d_{\max}$  from any other AIP, they would constitute a network of just two AIPs, and this pair would be assigned as a contact. Larger networks consist of more AIPs with multiple possible pairings within the  $d_{\max}$  threshold (*cf.* Fig. 2). These AIP contact networks were analysed with an adapted maximum bipartite pairing (MBP) algorithm, which is used to search for consensus pairings between items from two groups so that the number of pairings is maximised overall. The pairing partners are stochastically switched in an iterative process, and for each arrangement the number of pairings is evaluated until a maximum is reached. It is possible that there is more than one solution with the same maximum number of pairings. In this case, the set of contacts that minimised the sum of the distances between pairs of interacting AIPs,  $d_{\text{total}}$ , was selected.

The value of  $d_{\max}$  is the only adjustable parameter that determines the performance of the pairing algorithm. Analysis of X-ray crystal structures of protein–ligand complexes from the CASF dataset was used to test different values of  $d_{\max}$ .<sup>29</sup> Values less than 1 Å led to almost no AIP pairings, and values of greater than 2 Å led to networks that were too large for computation using the MBP algorithm. Values of  $d_{\max}$  midway between these two extremes gave AIP pairings consistent with the close contacts observed between atoms in the interaction interfaces. We use a default value of  $d_{\max}$  of 1.7 Å, which was sufficiently large to capture all of the potential AIP contacts in an intermolecular interface and small enough to keep the computational cost down (about a minute on a desktop computer for interaction surfaces involving 200 atoms).

An additional complication in the analysis of these networks arises due to the fractional AIPs that are used to represent some

functional groups. The original formulation of surface site interaction points (SSIP) used the same molecular surface area (9 Å<sup>2</sup>) for each SSIP, because this approach allows a straightforward treatment of the van der Waals (or dispersion) contribution to the interaction energy between two SSIPs.<sup>20</sup> However, AIPs that describe  $\pi$ -sites on aromatic carbon atoms are assigned a fraction parameter,  $f = 0.5$ , because they represent half the molecular surface area of a standard SSIP.<sup>22</sup> Thus an AIP with a fraction parameter  $f = 1.0$  could interact with two different AIPs with a fraction parameter  $f = 0.5$ . For example in the pairing process shown in Fig. 2, if A1 and A2 are both  $f = 0.5$  AIPs, and B2 is a  $f = 1.0$  AIP, then A1·B2 and A2·B2 are no longer mutually exclusive pairings. This situation is handled by copying instances of  $f = 1.0$  AIPs into two identical  $f = 0.5$  AIPs at the same position, allowing two different pairwise interactions with the same site. In this case, a fractional weighting of 0.5 is also applied to the distances between the pairs of interacting AIPs used to calculate  $d_{\text{total}}$ , the distance parameter used to determine the set of pairings that best represents the non-covalent interactions present in the interface.

## 6. Solution-phase free energy changes of complexation

Once all AIP contacts have been determined for a complex, the AIP interaction parameters can be used to calculate the associated free energy contribution of each contact to the overall binding free energy in any solvent. The free energy associated with the interaction of AIP  $i$  with AIP  $j$  is given by the difference between the energies of the AIPs in the bound state,  $\Delta G_{\text{B}}$ , and the solvation energies in the free state, as described previously (eqn (5)).<sup>17</sup>

$$\Delta G(i, j) = \Delta G_{\text{B}}(i) + \Delta G_{\text{B}}(j) - \Delta G_{\text{S}}(i) - \Delta G_{\text{S}}(j) \quad (5)$$

The free energy of each AIP in the bound state is given by eqn (6).

$$\Delta G_{\text{B}}(i) = \Delta G_{\text{B}}(j) = RT \ln \left( \frac{\sqrt{1 + 8\theta} - 1}{4\theta} \right) + RT \ln \frac{\sqrt{1 + 4\theta(K_{ij} + K_{\text{vdW}})} - 1}{2\theta(K_{ij} + K_{\text{vdW}})} \quad (6)$$

where  $\theta$  is the total AIP density of the solvent, and  $K_{\text{vdW}}$  and  $K_{ij}$  are defined in eqn (7) and (8).

$$K_{\text{vdW}} = \frac{1}{2} e^{-\frac{E_{\text{vdW}}}{RT}} \quad (7)$$

$$K_{ij} = \frac{1}{2} e^{-\frac{(\epsilon_i \epsilon_j + E_{\text{vdW}})}{RT}} \quad (8)$$

where  $E_{\text{vdW}}$  is  $-5.6 \text{ kJ mol}^{-1}$ , and for repulsive interactions, where  $\epsilon_i \epsilon_j$  is positive, this value is set to zero, because the SSIMPLE formulation assumes that repulsive interactions can be avoided by reorientation of dipoles.



The solvation free energy of each AIP in the free state can be calculated for any solvent using SSIMPLE.<sup>20</sup> In SSIMPLE, all pairwise interactions between solvent and solute AIPs in the liquid phase are described by eqn (8), which allows calculation of the Boltzmann distribution of AIP contacts and hence the solvation energy for an individual solute AIP,  $\Delta G_S(i)$ . We have previously shown that for a specific solvent the relationship between  $\Delta G_S(i)$  and the AIP interaction parameter  $\varepsilon_i$  can be described accurately using a simple polynomial (eqn (9)).<sup>24,30</sup>

$$\Delta G_S(i) = \sum_{k=0}^8 a_k \varepsilon_i^k \quad (9)$$

A list of polynomial coefficients  $a_k$  and values of  $\theta$  used for 261 different solvents is provided in the ESI.†

The overall free energy change for complexation is simply the sum of the individual contributions from each AIP contact scaled by the fractional parameter,  $f$ , for interactions involving fractional AIPs (eqn (10)).

$$\Delta G^\circ = \sum_{i,j} f \Delta \Delta G(i,j) \quad (10)$$

In principle, additional terms should be included to account for the effective molarities that describe the relationship between different interaction sites and the entropic penalties associated with bimolecular interactions and conformational restriction in the complex, but it turns out that eqn (10) provides a reasonable description of the behaviour of a wide range of different complexes (see below).

Since each AIP represents  $9 \text{ \AA}^2$  of molecular surface area, the AIP contact analysis of even relatively small intermolecular interfaces turns out to be rather complicated, because a large

number of interactions are involved. A graphical method has therefore been developed to visualise the most important AIP contacts that contribute to the overall stability of a complex. These AIP interaction maps provide a three-dimensional display of all of the AIP contacts that contribute more than a user-specified threshold to the free energy change of complexation in a given solvent. The default setting is that attractive interactions worth more than  $5 \text{ kJ mol}^{-1}$  are plotted as dark green spheres, attractive interactions worth more than  $0.5 \text{ kJ mol}^{-1}$  as light green spheres, repulsive interactions worth more than  $0.5 \text{ kJ mol}^{-1}$  as orange spheres, and all weaker contacts are plotted as small grey spheres, so that they do not confuse the image unnecessarily.

## 7. Results and discussion

The field of supramolecular chemistry has created numerous examples of intermolecular complexes with complicated binding interfaces featuring different types of non-covalent interaction. High resolution X-ray crystal structure data and solution-phase measurements of association constants in a variety of different solvents are often available, so host-guest complexes provide the ideal testbed for the method described above. We used the S30L dataset of experimentally-characterised complexes compiled by Sure and Grimme as a basis for testing the AIP approach, and a number of additional complexes were added to supplement this dataset.<sup>31–46</sup> One limitation of the AIP method is that it is not possible to calculate an AIP description of a charged compound, because the MEPS is dominated by the overall molecular charge, so the complexes selected for investigation in Table 1 all involve interactions between two neutral species. The complexes are organised according to the type of non-covalent interaction

**Table 1** Free energy changes for formation of host-guest complexes in various solvents (see Fig. 5–10 for structures)

Complex	Major NCI	Solvent	$\Delta G_{\text{expt}}^\circ$ (kJ mol <sup>-1</sup> )	$\Delta G_{\text{calc}}^\circ$ (kJ mol <sup>-1</sup> )	Error (kJ mol <sup>-1</sup> )	Ref.
1	Aromatic	Chloroform	-17.6	-15.6	-2.0	32
2	Aromatic	Chloroform	-5.9	-2.6	-3.3	32
3	Aromatic	Dichloromethane	-6.3	2.0	-8.3	33
4	Aromatic	Dichloromethane	-7.5	0.6	-6.9	33
5	Aromatic	Chloroform	-21.7	-9.0	-12.7	34
6	Aromatic	Chloroform	-19.2	-9.2	-10.0	34
7	H-bond	Chloroform	-34.7	-19.8	-14.9	35
8	H-bond	Chloroform	-13.8	-13.7	-0.1	35
9	H-bond	Chloroform	-48.9	-44.0	-4.9	36
10	H-bond	Dichloromethane	-26.3	-30.9	4.6	37
11	H-bond	1,1,2,2-Tetrachloroethane	-11.8	-11.6	-0.2	38
12	H-bond	1,1,2,2-Tetrachloroethane	-15.0	-14.5	-0.5	38
13	H-bond	1,1,2,2-Tetrachloroethane	-11.1	-7.6	-3.5	39
14	Desolvated cage	Mesitylene	-41.8	-39.8	-2.0	40
15	Desolvated cage	Mesitylene	-37.6	-44.9	7.3	40
16	Hydrophobic	Water	-12.5	-12.0	-0.5	41
17	Hydrophobic	Water	-20.5	-22.8	2.3	41
18	Hydrophobic	Water	-58.9	-58.3	-0.3	42
19	Halogen bond	Cyclohexane	-2.9	-30.9	28.0	43
20	Halogen bond	Cyclohexane	-21.3	-48.8	27.5	43
21	Extended $\pi$	Chloroform	-23.0	17.7	-40.7	44
22	Extended $\pi$	Chloroform	-9.2	22.9	-32.1	44
23	Extended $\pi$	Toluene	-22.2	2.8	-25.0	45
24	Extended $\pi$	Toluene	-21.3	3.5	-24.8	45
25	Extended $\pi$	Carbon disulfide	-18.4	0.0	-18.4	46
26	Extended $\pi$	Carbon disulfide	-17.6	-0.3	-18.9	46



(NCI) responsible for binding: aromatic interactions, H-bonding, hydrophobically driven binding in water, halogen bonds. There are two additional categories listed: interactions of extended  $\pi$ -surfaces, like fullerenes, which behave quite differently from small contact surface area aromatic interactions; and cages that show anomalously high binding affinities in mesitylene, because the solvent is too large to fit inside the binding pocket, so the free host is effectively desolvated.

The X-ray crystal structure of the complex was used as the starting point for the calculation, and the coordinates of the two components of the complex were extracted as separate files. For each molecule, the MEPS was calculated without any geometry optimisation at the 0.0300, 0.0104 and 0.0020 e Bohr<sup>-3</sup> electron density isosurface using NWChem 7.0.228 and DFT with the B3LYP functional and a 6-31G\* basis set (or 6-31G\*\* for iodine). Footprinting software described previously was used to convert the MEPS to AIPs, and then the AIPs of the two components were superimposed on the X-ray crystal structure of the complex. The AIP pairing analysis described above was then used to identify contacts and calculate the total solution phase interaction energy.

There is an assumption implicit in this approach that the three-dimensional structure observed in the crystal does not change in solution. Since the analysis is based on assignment of AIP contacts, precise interatomic distances are not critical in determining the interaction energies. There is an inherent fuzziness in the pairing of AIPs based on the distance criteria, so small variations in the geometry of the complex will not have a significant effect on the calculated interaction energy. However, a gross rearrangement of the structure of the complex in solution would render the analysis invalid.

Fig. 5–10 shows the AIP interaction maps calculated for each of the complexes in Table 1, and the corresponding free energy

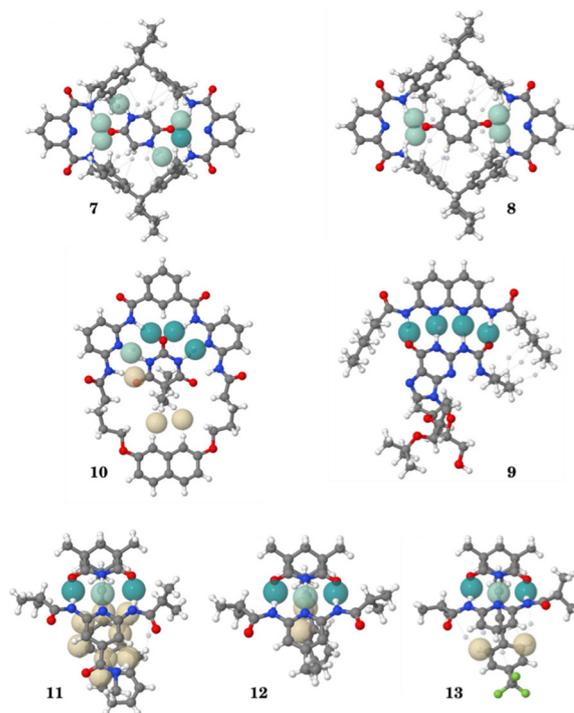
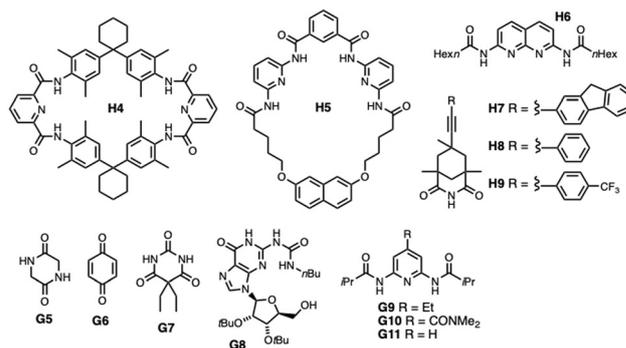


Fig. 6 AIP interaction maps for complexes where H-bonding is the major non-covalent interaction. Chemical structures of the components are shown. 7 is H4-G5. 8 is H4-G6. 9 is H5-G7. 10 is H6-G8. 11 is H7-G9. 12 is H8-G10. 13 is H9-G11.

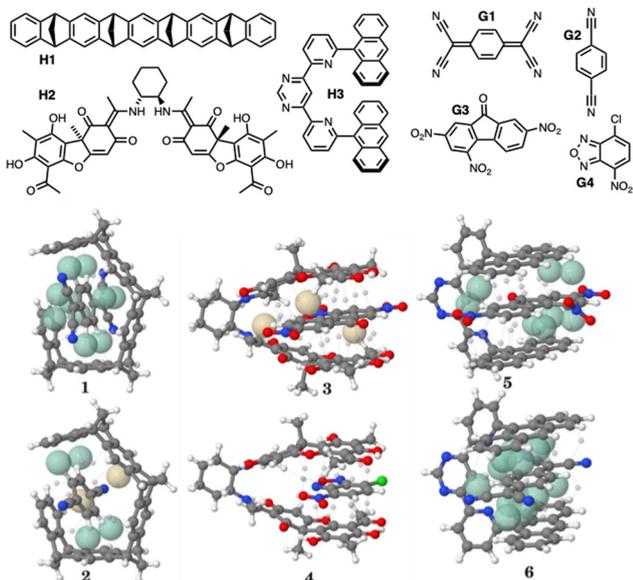


Fig. 5 AIP interaction maps for complexes where aromatic interactions are the major non-covalent interaction. Chemical structures of the components are shown. 1 is H1-G1. 2 is H1-G2. 3 is H2-G3. 4 is H2-G4. 5 is H3-G3. 6 is H3-G1.

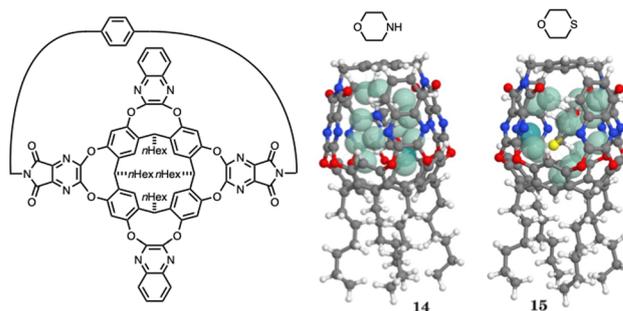


Fig. 7 AIP interaction maps for complexes where the host is a cage that is desolvated in the free state. Chemical structures of the components are shown.

changes calculated using eqn (10) are listed in Table 1 (see ESI† for a detailed list of individual AIP contacts and associated free energy contributions for each complex). It was possible to treat



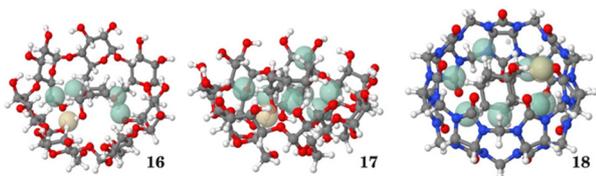


Fig. 8 AIP interaction maps for complexes where contacts between hydrophobic surfaces drive binding. **16** is cyclopentanol- $\beta$ -cyclodextrin. **17** is cyclooctanol- $\beta$ -cyclodextrin. **18** is 1-hydroxyadamantane-cucurbit[7]uril.

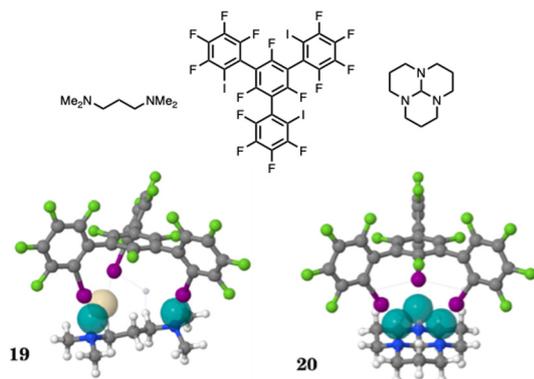


Fig. 9 AIP interaction maps for complexes where halogen bonding is the major non-covalent interaction. Chemical structures of the components are shown.

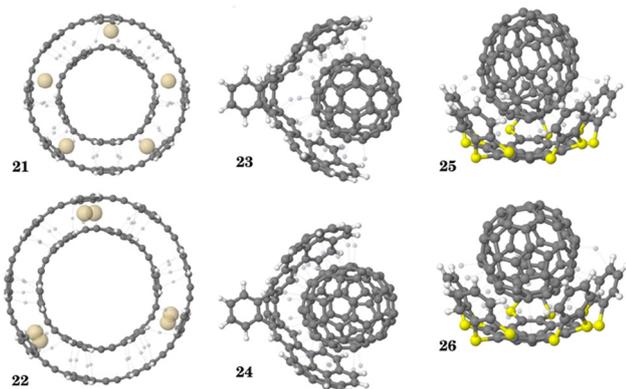


Fig. 10 AIP interaction maps for interactions between extended  $\pi$ -surfaces, like fullerenes. **21** and **22** are the cycloparaphenyleneacetylene (CPPA) complexes 5CPPA-8CPPA and 6CPPA-9CPPA. **23** and **24** are the buckycatcher complexes of  $C_{60}$  and  $C_{70}$ . **25** and **26** are the pentakis(1,4-benzodithiino)-corannulene complexes of  $C_{60}$  and  $C_{70}$ .

the desolvated cages in the same way as the other complexes by simply omitted the  $\Delta G_S$  term describing solvation of the host AIPs from eqn (5).

The aromatic interaction complexes all feature guests with electron-withdrawing groups and hosts with more electron-rich  $\pi$ -systems (Fig. 5). In complex **1**, the guest  $\pi$ -face AIPs are sufficiently positive to make attractive interactions with the negative AIPs that represent the  $\pi$ -faces of the host. There are also four attractive edge-to-face contacts leading to a large

favourable free energy change that agrees well with experiment. In complex **2**, the guest  $\pi$ -face AIPs are not positive enough to make attractive interactions with the host  $\pi$ -face AIPs, and complexation is dominated by the four edge-to-face interactions. The reduction in binding affinity compared with complex **1** agrees well with the experimental results. For complexes **3** and **4**, the AIPs are not sufficiently polar for any contacts to compete with desolvation, and the calculated free energy changes are all very small. In complexes **5** and **6**, the guest  $\pi$ -face AIPs are positive and make a large number of attractive interactions with the host  $\pi$ -face AIPs. However, the net interaction energy is somewhat less favourable than experiment in both cases. Although the AIP contact maps provide a good qualitative description of the significance of different contacts in these complexes, summing over a large number of relatively non-polar AIP interactions results in a large error in the totals for the calculated free energy changes.

In contrast, the H-bond complexes are characterised by a small number of very polar AIP interactions (Fig. 6). In this case, the accuracy of the calculated free energy changes depends on the accuracy of the AIP polar interaction parameters  $\epsilon_i$ . In all cases, the H-bonds are clearly identified as green balls in the AIP interaction maps, and the agreement with the experimental free energy changes is excellent. The one discrepancy is complex **7**, which is significantly more stable than the calculation predicts. This discrepancy appears to be due to the rather low polarity assigned to the amide AIPs involved in the H-bonding interactions.

The properties of the desolvated cage complexes are well-described by the AIP interaction maps in Fig. 7. In both cases, there are a large number of non-polar AIP contacts, which are each worth about  $-3 \text{ kJ mol}^{-1}$  due to the lack of solvent competition for van der Waals interactions with the host sites. The total calculated free energy changes agree well with experiment.

The hydrophobic complexes in Fig. 8 are also characterised by a large number of non-polar AIP contacts, each worth  $-3$  to  $-4 \text{ kJ mol}^{-1}$ . However, in this case, the driving force comes from the favourable desolvation term associated with the poor solvation of non-polar sites by water. Again the total calculated free energy changes agree well with experiment.

The halogen bond and extended  $\pi$ -surface complexes are less well-described. Although the halogen bonds are clearly identified in the AIP interaction maps (Fig. 9), the calculated free energy changes for these complexes are significantly more favourable than the experimental values.

We have previously investigated the description of halogen-bonded complexes using the empirical H-bond parameters,  $\alpha$  and  $\beta$ , on which the AIP polar interaction parameters  $\epsilon_i$  were parameterised. For many simple complexes that make a single halogen-bond, it was possible to accurately reproduce the experimentally measured association constants  $K$  using eqn (11).<sup>22,47</sup>

$$-RT \ln K / \text{kJ mol}^{-1} = -(\alpha - \alpha_S)(\beta - \beta_S) + 6 \quad (11)$$

where  $\alpha_S$  and  $\beta_S$  are the H-bond parameters of the solvent and the constant of  $6 \text{ kJ mol}^{-1}$  was experimentally determined in carbon tetrachloride solution.<sup>21</sup>

However, some halogen-bonded complexes are not well-described by eqn (11), specifically complexes of perfluoroiodocarbons with



tertiary amines and the complex formed by molecular iodine and tetramethylthiourea, where there appears to be a significant covalent contribution to the interaction energy that leads to a stabilisation of the complex.<sup>47,48</sup> In contrast, halogen-bonded complexes **19** and **20** are significantly less stable than predicted by the AIP calculations. The AIP pairing method is based on point contacts with no consideration of the precise arrangement of the interacting groups, and it is possible that there is some geometrical mismatch in these complexes that prevents optimal pairwise interactions at all of the halogen-bond sites, which results in a destabilisation that is not captured by the calculations.

The AIP interaction maps of the extended  $\pi$  complexes are characterised by a large number of weakly unfavourable contacts (Fig. 10). For complexes **21** and **22**, contacts between negative  $\pi$ -face AIPs lead to positive calculated free energy changes. For the fullerene complexes **23–26**, most of the guest AIP values are close to zero, and the calculated free energy changes are small. These results suggest that there is an additional contribution to the stability of these complexes that is not captured by the AIP contact analysis. Since these complexes feature interactions between non-polar surfaces, the main driving force for binding is presumably dispersion. In the SSIMPLE model, dispersion is introduced through a constant van der Waals interaction energy of  $-5.6 \text{ kJ mol}^{-1}$  per AIP contact (see eqn (8)). However, dispersion interactions between solute AIPs are largely cancelled by dispersion interactions with solvent AIPs, which is consistent with experimental measurements of non-polar interactions in non-polar solvents.<sup>49,50</sup> As a result, the net contribution of dispersion to the calculated intermolecular interaction energies is small, except in the case of the desolvated cages where this cancellation does not occur. The AIP model makes reasonable predictions for the smaller aromatic interaction complexes, so the discrepancy for the extended  $\pi$  complexes appears to be related to the very large surface area of contact. One possible explanation is that there are cooperative effects across larger flat surfaces, and the AIP model, which treats individual AIP contacts as mutually independent, would not capture such effects.

The calculated free energy changes can be used to estimate association constants for the complexes ( $K = \exp(-\Delta G^\circ/RT)$ ). Fig. 11 shows the correlation between calculation and experiment for the complexes featuring conventional interaction types that are well-described by the AIP interaction maps (aromatic interactions, H-bonds, hydrophobically driven complexation, and desolvated cages). Although there are some outliers, the AIP description of these complexes is rather good, with predicted binding affinities accurate to within one order of magnitude.

## 8. Conclusions

Atomic surface site interaction points (AIP) provide a complete description of the non-covalent interactions that one molecule can make with another. The SSIMPLE algorithm can be used to obtain the free energy change associated with the pairwise interaction between AIPs on two different molecules in any

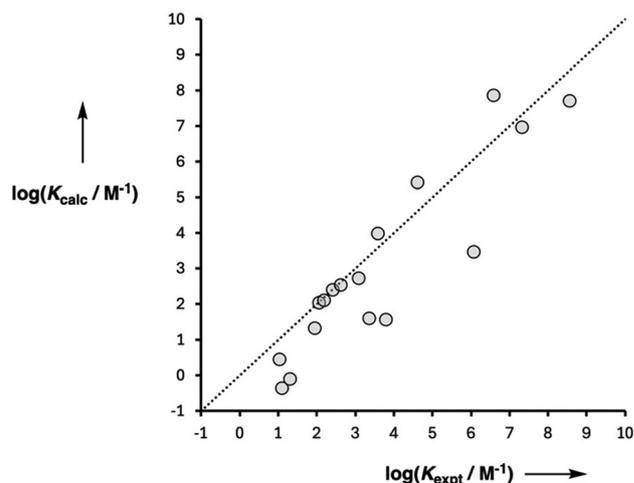


Fig. 11 Comparison of the experimentally measured association constants for formation of 1:1 complexes with the values calculated using the AIP interaction maps in Fig. 5–8 (RMSE = 1.1).

solvent. Summing these pairwise AIP interactions across an intermolecular interface that involves multiple interacting sites can be used to calculate solution phase binding free energies and association constants. For some classes of non-covalent interaction, such as H-bonding, identification of pairwise AIP interactions in the three-dimensional structure of a complex is straightforward, because the interacting AIPs are almost perfectly superimposed in space. For less polar interactions that take place over more diffuse surfaces represented by multiple AIPs, such as hydrophobic contacts, the challenge is to identify a discrete pairwise set of AIP contacts to describe the interface. This paper describes a computational tool that converts the three-dimensional structure of a complex into a set of AIP contacts that describe the interactions between the two molecules, including the free energy contributions due to desolvation of the interacting sites. A visualisation tool has also been developed to display AIP interaction maps, allowing straightforward identification of the key intermolecular contacts that contribute most to the overall binding free energy in a complex.

A range of host–guest complexes were used to investigate the utility of the methods. X-ray crystal structures of the complexes were used to identify AIP contacts and hence calculate binding free energies. Complexes that are dominated by H-bonding and aromatic interactions are described well by the AIP pairing calculations, which provide reasonably accurate predictions of the experimentally measured association constants in different solvents (*i.e.* within one order of magnitude). Similar results were obtained for complexes in water, where binding is driven by contacts between hydrophobic surfaces. The treatment of solvation can easily be adapted to describe complexes formed by cages in solvents that are too large to fit inside the host cavity. In this case, the term describing solvation of the host AIPs was simply removed from the free energy calculation to obtain accurate predictions of the association constants. In contrast, halogen-bonded complexes and complexes involving interactions between the extended  $\pi$ -surfaces were not



well-described by the AIP method. Nevertheless, AIP interaction pairs of intermolecular complexes are relatively straightforward to calculate and provide useful insight into the nature of the interactions that lead to complex formation and the role of solvent in determining observed binding affinities.

## Author contributions

The manuscript was written through contributions of all authors.

## Data availability

All supporting data is provided in the ESI.†

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The authors thank the Engineering and Physical Sciences Research Council (EP/S024220/1) for funding.

## Notes and references

- M. K. Gilson and H.-X. Zhou, *Annu. Rev. Biophys. Biomol. Struct.*, 2007, **36**, 21–42.
- G. Śliwowski, S. Kothiwale, J. Meiler and E. W. Lowe, *Pharmacol. Rev.*, 2014, **66**(1), 334–395.
- J. Dec and J.-M. Bollag, *Soil Sci.*, 1997, **162**(12), 858–874.
- N. Bordenave, B. R. Hamaker and M. G. Ferruzzi, *Food Funct.*, 2014, **5**(1), 18–34.
- Y. Long, J. F. Hui, P.-P. Wang, G.-L. Xiang, B. Xu, S. Hu, W.-C. Zhu, X.-Q. Lü, J. Zhuang and X. Wang, *Sci. Rep.*, 2012, **2**, 612.
- J. Hentschel, A. M. Kushner, J. Ziller and Z. Guan, *Angew. Chem., Int. Ed.*, 2012, **51**, 10561–10565.
- S. Keinan, M. A. Ratner and T. J. Marks, *Chem. Phys. Lett.*, 2004, **392**, 291–296.
- F. Jensen, *Introduction to computational chemistry*, John Wiley & Sons, 3rd edn, 2017.
- P. Pracht and S. Grimme, *Chem. Sci.*, 2021, **12**, 6551–6568.
- D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, Cambridge University Press, 2nd edn, 2004.
- H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**(7), 1198–1229.
- G. Groenhof, *Biomolecular Simulations*, Springer, 2013.
- M. Kuhn, S. Firth-Clark, P. Tosco, A. S. J. S. Mey, M. Mackey and J. Michel, *J. Chem. Inf. Model.*, 2020, **60**(6), 3120–3130.
- M. C. Storer and C. A. Hunter, *Chem. Soc. Rev.*, 2022, **51**, 10064.
- M. C. Storer and C. A. Hunter, *Phys. Chem. Chem. Phys.*, 2022, **24**(30), 18124–18132.
- M. D. Driver, M. J. Williamson, N. De Mitri, T. Nikolov and C. A. Hunter, *J. Chem. Inf. Model.*, 2021, **61**(11), 5331–5335.
- M. D. Driver, M. J. Williamson, J. L. Cook and C. A. Hunter, *Chem. Sci.*, 2020, **11**(17), 4456–4466.
- A. Oliver, C. A. Hunter, R. Prohens and J. L. Rossello, *J. Comput. Chem.*, 2018, **39**(28), 2371–2377.
- C. A. Hunter, *Chem. Sci.*, 2013, **4**(2), 834–848.
- C. A. Hunter, *Chem. Sci.*, 2013, **4**(4), 1687–1700.
- C. A. Hunter, *Angew. Chem., Int. Ed.*, 2004, **43**(40), 5310–5324.
- M. C. Storer, K. J. Zator, D. P. Reynolds and C. A. Hunter, *Chem. Sci.*, 2024, **15**(1), 160–170.
- T. Grecu, R. Prohens, J. F. McCabe, E. J. Carrington, J. S. Wright, L. Brammer and C. A. Hunter, *CrystEngComm*, 2017, **19**(26), 3592–3599.
- M. D. Driver, *The SSIP approach to solvation*, University of Cambridge, 2020.
- A. Shrake and J. A. Rupley, *J. Mol. Biol.*, 1973, **79**(3), 351–371.
- R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane and V. S. Pande, *Biophys. J.*, 2015, **109**(8), 1528–1532.
- D. A. Smith, *Modeling the hydrogen bond*, American Chemical Society, 1994.
- C. S. Calero, J. Farwer, E. J. Gardiner, C. A. Hunter, M. Mackey, S. Scuderi, S. Thompson and J. G. Vinter, *Phys. Chem. Chem. Phys.*, 2013, **15**(30), 18262–18273.
- M. Su, Q. Yang, Y. Du, G. Feng, Z. Liu, Y. Li and R. Wang, *J. Chem. Inf. Model.*, 2019, **59**, 895–913.
- R. Hubbard and E. Baker, *Prog. Biophys. Mol. Biol.*, 1984, **44**, 97–179.
- R. Sure and S. Grimme, *J. Chem. Theory Comput.*, 2015, **11**(8), 3785–3801.
- F.-G. Klärner, U. Burkert, M. Kamieth, R. Boese and J. Benet-Buchholz, *Eur. J. Org. Chem.*, 1999, (6), 1700–1707.
- J. Graton, J.-Y. Le Questel, B. Legouin, P. Uriac, P. van de Weghe and D. Jacquemin, *Chem. Phys. Lett.*, 2012, **522**(1), 11–16.
- A. Petitjean, R. G. Khoury, N. Kyritsakas and J. M. Lehn, *J. Am. Chem. Soc.*, 2004, **126**(21), 6637–6647.
- C. Allott, H. Adams, P. L. Bernad, C. A. Hunter, C. Rotger and J. A. Thomas, *Chem. Commun.*, 1998, 2449–2450.
- T. Park, S. C. Zimmerman and S. Nakashima, *J. Am. Chem. Soc.*, 2005, **127**(18), 6520–6521.
- E. Fan, A. D. Hamilton, S. K. Chang and D. Van Engen, *J. Am. Chem. Soc.*, 1991, **113**(20), 7640–7645.
- M. Harder, M. A. Carnero Corrales, N. Trapp, B. Kuhn and F. Diederich, *Chem. – Eur. J.*, 2015, **21**(23), 8455–8463.
- L. J. Riwar, N. Trapp, B. Kuhn and F. Diederich, *Angew. Chem., Int. Ed.*, 2017, **56**(37), 11252–11257.
- J. Hornung, D. Fankhauser, L. D. Shirtcliff, A. Praetorius, W. B. Schweizer and F. Diederich, *Chem. – Eur. J.*, 2011, **17**(44), 12362–12371.
- Y. Liu, E. C. Yang, Y. W. Yang, H. Y. Zhang, Z. Fan, F. Ding and R. Cao, *J. Org. Chem.*, 2004, **69**(1), 173–180.
- S. Moghaddam, C. Yang, M. Rekharsky, Y. H. Ko, K. Kim, Y. Inoue and M. K. Gilson, *J. Am. Chem. Soc.*, 2011, **133**(10), 3570–3581.
- S. H. Jungbauer, *J. Am. Chem. Soc.*, 2014, **136**(48), 16740–16743.



- 44 T. Kawase, Y. Nishiyama, T. Nakamura, T. Ebi, K. Matsumoto, H. Kurata and M. Oda, *Angew. Chem., Int. Ed.*, 2007, **46**(7), 1086–1088.
- 45 C. Muck-Lichtenfeld, S. Grimme, L. Kobryn and A. Sygula, *Phys. Chem. Chem. Phys.*, 2010, **12**(26), 7091–7097.
- 46 P. E. Georghiou, A. H. Tran, S. Mizyed, M. Bancu and L. T. Scott, *J. Org. Chem.*, 2005, **70**(16), 6158–6163.
- 47 R. Cabot and C. A. Hunter, *Chem. Commun.*, 2009, 2005–2007.
- 48 C. C. Robertson, R. N. Perutz, L. Brammer and C. A. Hunter, *Chem. Sci.*, 2014, **5**(11), 4179–4183.
- 49 L. Yang, C. Adam, G. S. Nichol and S. L. Cockroft, *Nat. Chem.*, 2013, **5**(12), 1006–1010.
- 50 C. Adam, L. Yang and S. L. Cockroft, *Angew. Chem., Int. Ed.*, 2015, **54**, 1164–1167.

