

Cite this: *Chem. Sci.*, 2021, **12**, 13193

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 22nd June 2021
Accepted 7th September 2021
DOI: 10.1039/d1sc03392a
rsc.li/chemical-science

Introduction

The solubility of an organic molecule in different chemical environments is a fundamentally important property that has implications in the petrochemical industry, in synthesis, and for the bioavailability of drug candidates. Although numerous approaches have been developed for the prediction of solubility directly from chemical structure, the accuracy and generality of these methods are still limited. The most pragmatic approaches are based on empirical parameterisation of functional group contributions, but these methods are limited to solvents where a large body of experimental data is available.^{1–4} The most sophisticated approaches are based on atomistic simulation using molecular dynamics, but these methods require significant computational resource.^{5,6} There are some promising approaches that marry the predictive power of *ab initio* calculation with empirical modelling.^{7–15}

Abraham developed a different approach by using experimental data on 1 : 1 complexation to develop summation solvation parameters to describe the total H-bond donor or acceptor capacity of a molecule.^{16–18} Here we show that the description of a molecule as a set of specific interaction points, which are associated with individual functional groups, can be used to sum interactions across the molecular surface and

accurately predict solvation properties directly from chemical structure. The approach integrates the treatment of intermolecular interactions between solutes and phase transfer equilibria, which means that diverse experimental data can be used for parameterisation. The description of molecules as a collection of the functional groups allows extrapolation to compounds for which experimental data is not available.

The generalised H-bond donor and acceptor parameters α and β can be used to describe the non-covalent interaction properties of both solvents and solutes.^{19,20} The free energy of formation ($-\Delta G^\circ$) for complexes in solution is determined by the competition between solute–solute, solute–solvent and solvent–solvent interactions, as illustrated in Fig. 1.

Given experimentally determined H-bond parameters for two solutes (α and β) and the solvent (α_s and β_s), it is possible to make a reliable quantitative estimate of the free energy of formation of a 1 : 1 complex between a H-bond donor (D) and H-bond acceptor (A) using eqn (1).

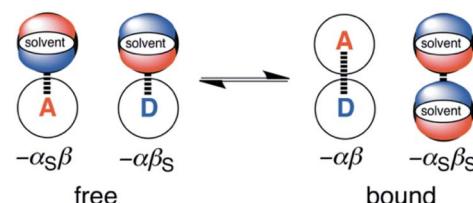


Fig. 1 The solvent competition model for the complex formed between a H-bond donor solute, D, and a H-bond acceptor solute, A. The relative stabilities of the four complexes can be estimated using the H-bond parameters, α , β , α_s and β_s , as indicated.

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

† Electronic supplementary information (ESI) available: Experimental data and references, substructure fragments, H-bond parameters, solvent parameters, surface area calculations, and spreadsheets illustrating the calculations. See DOI: 10.1039/d1sc03392a

$$\Delta G^\circ/\text{kJ mol}^{-1} = -(\alpha - \alpha_S)(\beta - \beta_S) + 6 \quad (1)$$

where the constant of 6 kJ mol^{-1} was experimentally determined in carbon tetrachloride solutions.

This surprisingly simple model is based solely on the properties of the pairwise local contacts between molecules, and it does not require any consideration of long range interactions, solvent dielectric constant, solvent structure, cavitation or entropic terms. The success of the model suggests that these terms are small relative to the free energy contributions due to local interactions, or that they cancel out, or that they are captured in some way by the constant of 6 kJ mol^{-1} , which does not vary much between solvents.

The approach illustrated in Fig. 1 has been extrapolated from single point interactions to a complete description of molecular surfaces by assigning a set of surface site interaction points (SSIP) to describe all of the non-covalent interactions that a molecule can make with the surroundings. Each SSIP is assigned an interaction parameter, which is equivalent to the empirically measured α and β parameters illustrated in Fig. 1. Fig. 2 shows the result for water, which is represented by two donor SSIPs and two acceptor SSIPs. Two approaches have been developed for obtaining the SSIP description of a molecule: manual assignment based on functional groups,²¹ and a computational method based on footprinting of *ab initio* calculated molecular electrostatic potential surfaces.^{22,23} The molecular SSIP description can be used to estimate the free energy contribution of non-covalent interactions to the stability of a solid by assuming that SSIPs are paired in a hierarchical fashion to maximise the total interaction energy, and this approach has been applied to successfully predict cocrystal formation.²⁴ The free energy of interaction of a molecule with a solvent can be calculated by using the SSIP descriptions of solute and solvent in the surface site interaction model for the properties of liquids at equilibrium (SSIMPLE), and this approach has been applied in the calculation of solution phase properties like partition.^{21,23}

In SSIMPLE, solvation free energies are obtained by calculating the equilibrium distribution of all pairwise SSIP contacts, also allowing for non-bonded states to account for the significant void space present in a liquid. There is an implicit assumption in eqn (1) that all four complexes shown in Fig. 1 are fully bound and that non-polar van der Waals interactions cancel out, so that the equilibrium is dominated by polar interactions. The introduction of non-bonded states in

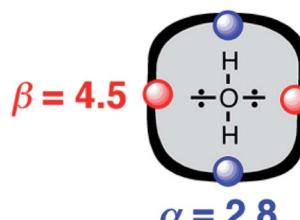


Fig. 2 Water is represented by four SSIPs that describe the two H-bond donor (blue) and two H-bond acceptor (red) sites.

SSIMPLE therefore required an additional treatment of van der Waals interactions, and calculation of solvation energies using this approach is significantly more complicated than eqn (1) would suggest. Here, we present an alternative approach, which is operationally simpler and has the advantage that empirical elements are introduced to improve the accuracy. For teaching purposes, it is also useful to have an approach where partition can be calculated using only a pocket calculator or simple spreadsheet.

Solvation free energies

As we have previously suggested, the constant of 6 kJ mol^{-1} in eqn (1) is related to the difference between the concentration of the solvent and the standard state of 1 M used to describe the chemical potentials of the solutes.^{21,25} Fig. 3 shows experimental data for formation of a 1 : 1 H-bonded complexes in various ether solvents. There is a linear dependence of the value of $-\Delta G^\circ$ on ether oxygen concentration expressed in entropy units (*i.e.* $RT \ln[O]$), and the slope is close to -1 . This result suggests that if we define $[S \cdot S]$ as the effective concentration of the solvent–solvent interactions that are made on formation of a complex between two solutes, eqn (1) can be rewritten as eqn (2). For solvents like ethers that contain a single polar functional group, the effective concentration of solvent–solvent interactions is similar to the functional group concentration, so eqn (2) provides a reasonable description of the experimental data in Fig. 3. This relationship is also consistent with the behaviour of H-bonded complexes in solvent mixtures, where the free energy change for complex formation was found to be directly proportional to the logarithm of the concentration of the more polar solvent present in the mixture.^{26–28}

$$\Delta G^\circ = -\alpha\beta + \alpha_S\beta + \alpha_S\beta_S - \alpha_S\beta_S + RT \ln[S \cdot S] \quad (2)$$

In general, the effective concentration of solvent–solvent interactions is not well-defined, so reliable values of $[S \cdot S]$ can only be deduced from experimental data. Experimentally

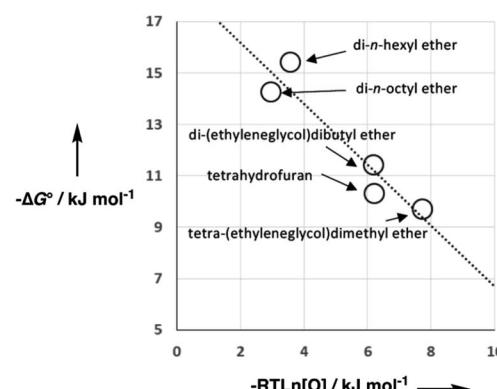


Fig. 3 Free energy of formation for 1 : 1 H-bonded complexes in ether solvents ($-\Delta G^\circ$) plotted as a function of solvent oxygen concentration in entropy units ($-RT \ln[O]$). The linear regression line has formula $-1.2x + 18.5$ with $R^2 = 0.90$.



determined association constants for formation of 1:1 H-bonded complexes in carbon tetrachloride were used to derive eqn (1), which gave a value of 10 M for [S·S] for this solvent. In this paper, we describe an approach based on partitioning of solutes between different solvents, which allows experimental determination of [S·S] for a wide range of different solvents.

The first term in eqn (2) describes the free energy of the A·D complex, and the other terms describe the free energy changes associated with desolvation of the two solutes on complexation. If we define Δg_s as the free energy of transfer of a SSIP from an external reference state into a solvent S, then eqn (2) can be rewritten as eqn (3).

$$\Delta G^\circ = -\alpha\beta - \Delta g_s(\alpha) - \Delta g_s(\beta) \quad (3)$$

where

$$\Delta g_s(\alpha) = -\alpha\beta_S - C_\beta \quad (4)$$

$$\Delta g_s(\beta) = -\alpha_S\beta - C_\alpha \quad (5)$$

The two constants C_α and C_β describe the solvent–solvent interactions that are disrupted when a solute enters the solvent. The sum of these constants should match the corresponding terms in eqn (2), which provide a composite description of the polarity and the concentration of the interactions that each solvent SSIP makes with the surrounding bulk solvent. Since these solvent–solvent terms are self-association, then in solvents where there is only one type of solvent–solvent interaction, the two constants must equal, as defined in eqn (6).

$$C_\alpha = C_\beta = \frac{1}{2}(-\alpha_S\beta_S + RT \ln[S·S]) \quad (6)$$

This reformulation of eqn (1) leads to a description of the free energy of solvation of an individual solute SSIP, and summing over a set of SSIPs that represent the surface of a molecule provides a straightforward method for calculating molecular solvation energies and hence partition. Eqn (7) shows the free energy change for transfer of a solute with N_α H-bond donor SSIPs and N_β H-bond acceptor SSIPs from the non-solvated reference state to solvent S.

$$\Delta G_S^\circ = \sum_{i=1}^{N_\alpha} \Delta g_S(\alpha_i) + \sum_{i=1}^{N_\beta} \Delta g_S(\beta_i) + C_0 \quad (7)$$

A constant, C_0 , is introduced in eqn (7) to take care of any additional free energy contributions. The requirement for this constant can be demonstrated by examining experimental data for the free energies of transfer of alkanes from the gas phase into different solvents. Fig. 4 compares the gas–liquid partitioning of alkanes into *n*-hexadecane with the corresponding values for partitioning into *n*-hexane. Although the slope is very close to unity (1.01), there is an offset of 1.39 kJ mol⁻¹ in favour of transfer into *n*-hexane. The fact that the offset is a constant for different alkane solutes with very different numbers of interaction sites implies that this phenomenon is due to

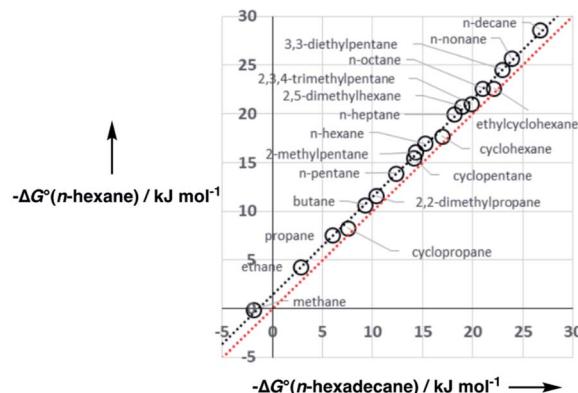


Fig. 4 Comparison of the free energies of transfer of alkanes from the gas phase into *n*-hexadecane with the corresponding values for transfer into *n*-hexane. The red line is $y = x$, and the line of best fit (black) is $y = 1.01x + 1.39 \text{ kJ mol}^{-1}$.

differences in the intrinsic properties of the two solvents. Similar behaviour is observed for pairwise comparisons of some other non-polar solvents (see ESI Section S1†). Values of the offset vary from solvent to solvent, so we define the constant C_0 relative to a reference solvent, *n*-hexadecane, to describe this intrinsic difference between different solvents.

The overall free energy of partition for a molecule is obtained by summing over all solute SSIPs using eqn (7), and the free energy of transfer between two different solvents S1 and S2 can then be obtained using eqn (8).

$$\Delta G_{S1 \rightarrow S2}^\circ = \Delta G_{S2}^\circ - \Delta G_{S1}^\circ \quad (8)$$

Many of the parameters required for the implementation of eqn (4)–(8) can be determined from the experimentally determined association constants for formation of 1:1 H-bonded complexes between polar solutes in non-polar solvents. However, H-bonded complexes are not sufficiently stable in polar solvents for characterisation of the solvation properties of a wider range of solvents, and interactions between non-polar solutes are too weak to lead to complexation in solution. Here we investigate the use of experimental partition data as a way of deducing parameters to describe these systems.

SSIP description of aliphatic hydrocarbons

A SSIP description of polar functional groups is relatively straightforward to implement using the basic principles of chemical bonding and hybridisation. For example, water is described by four SSIPs, which represent the two H-bond donor sites and the two lone pairs (Fig. 2). For non-polar functional groups, the choice is less obvious. Previously, we have used molecular surface area as the reference point to determine the number of SSIPs required to describe the non-covalent interaction properties of a molecule. The model developed here is different, because all SSIPs are assumed to be fully paired in the liquid state. If the van der Waals contribution to solvation is to cancel out when a molecule is transferred from one liquid to another, then the total concentration of SSIPs in different

liquids should be approximately the same. Liquid water provides an obvious benchmark with four SSIPs per molecule and a concentration of 55 M at 298 K. The total number of SSIPs ($N_\alpha + N_\beta$) required to describe hydrocarbons was therefore defined based on the molar concentration of the pure liquid using eqn (9).

$$[\text{SSIP}] = (N_\alpha + N_\beta)[\text{liquid}] \approx 220 \text{ M} \quad (9)$$

Solvent H-bond parameters for alkanes have been determined experimentally by applying eqn (1) to measurements of association constants for the formation of 1 : 1 H-bonded complexes.²⁹ The experimental values of $\alpha_S = 1.20$, $\beta_S = 0.60$, and $RT \ln[S \cdot S] = +6 \text{ kJ mol}^{-1}$ can be used in eqn (6) to obtain the constants required to describe alkane solvents: $C_\alpha = C_\beta = 2.64 \text{ kJ mol}^{-1}$. In contrast, eqn (1) cannot be used directly to determine the parameters required to describe the properties of alkane solutes, because they are not polar enough to form stable 1 : 1 complexes. We therefore used experimental data on phase transfer free energies to develop a SSIP description of alkane solutes.

Fig. 5 shows experimental data for the free energy of transfer of alkanes from water to *n*-hexadecane. The correlation with the number of hydrogen atoms (N_H) is significantly better than the correlation with the number of carbon atoms (N_C), because the hydrogen atoms are always exposed on the surface of the

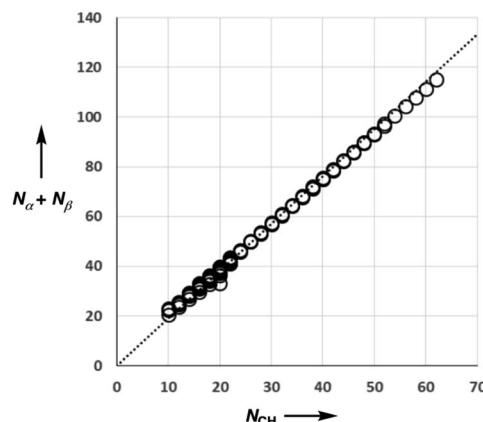


Fig. 6 Comparison of the total number of SSIPs ($N_\alpha + N_\beta$) calculated using eqn (9) with the number of CH bonds (N_{CH}) for 182 alkanes. The line of best fit through the origin is shown ($y = 1.90x$).

molecule, whereas the carbon atoms are not. For example, the quaternary carbon atom in neopentane is completely buried. The approach to construction of a SSIP description of alkane solutes is therefore based on the number of CH bonds. Application of eqn (9) to alkanes indicates that the total number of SSIPs required to describe an alkane ($N_\alpha + N_\beta$) is approximately twice the number of CH bonds (Fig. 6).³⁰ Assigning two SSIPs to each CH bond results in a total SSIP concentration that varies from 208 M for *n*-pentane to 232 M for *n*-hexadecane (cf. benchmark value of 220 M for water). Fig. 7 shows the calculated MEPS of methane. There are four regions of positive potential over the hydrogen atoms on the end of each CH bond and four regions of negative potential over the carbon atoms at the back of each CH bond. We therefore assign one α and one β for every CH bond in an alkane, and assume that the SSIP parameters determined for alkane solvents can be used to describe the non-covalent interaction properties of alkane solutes, *i.e.* $\alpha = 1.20$, $\beta = 0.60$.

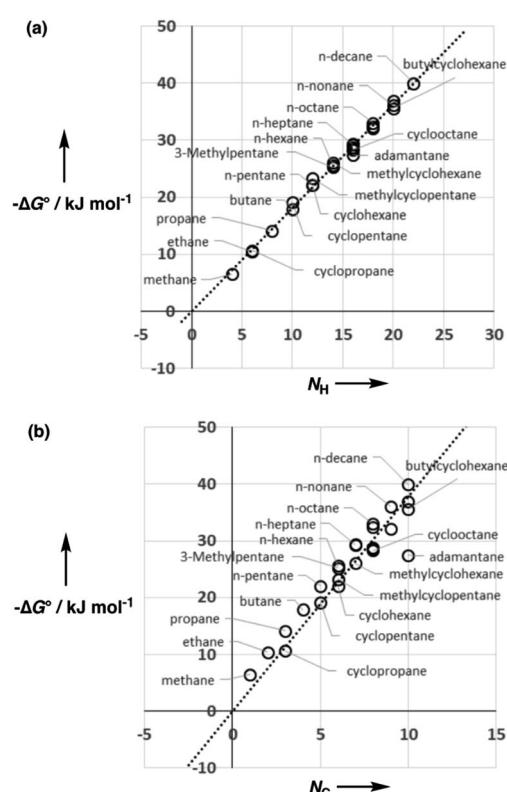


Fig. 5 Comparison of the experimental free energy of transfer of alkanes from water to *n*-hexadecane (ΔG°) (a) with the number of hydrogen atoms (N_H , $R^2 = 0.9956$) and (b) with the number of carbon atoms (N_C , $R^2 = 0.8778$) in the molecule.

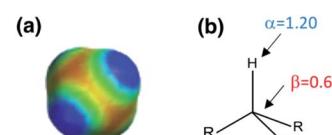


Fig. 7 (a) The molecular electrostatic potential surface of methane (red = -25 kJ mol^{-1} , blue = $+25 \text{ kJ mol}^{-1}$) calculated on the $0.002 \text{ e bohr}^{-3}$ electron density isosurface using DFT with a B3LYP 6-31G* basis set. (b) The SSIP description of a CH bond in an aliphatic hydrocarbon.



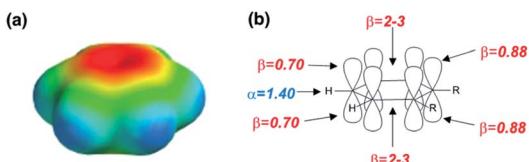


Fig. 8 (a) The molecular electrostatic potential surface of benzene (red = -80 kJ mol^{-1} , blue = $+80 \text{ kJ mol}^{-1}$) calculated on the $0.002 \text{ e bohr}^{-3}$ electron density isosurface using DFT with a B3LYP 6-31G* basis set. (b) The SSIP description of aromatic hydrocarbons (R is an alkyl substituent).

Table 1 H-bond parameters for the polar interaction sites on the π -faces of aromatic hydrocarbons

Aromatic acceptors	β
Benzene	2.00
Toluene	2.20
ortho-, meta- or para-xylene	2.40
Mesitylene	2.70
Hexamethylbenzene	3.10

one α on the end of the CH bond and two β SSIPs above and below the plane of the ring (Fig. 8b). This description gives a total of 18 SSIPs, and the MEPS shown in Fig. 8a suggests that the two additional SSIPs should be used to represent the more polar negative electrostatic potential located over the centres of the two faces of the π -system. The properties of these two SSIPs were determined by applying eqn (1) to measurements of association constants for the formation of 1 : 1 H-bonded complexes with H-bond donors (see Table 1 and Fig. S2†). The other parameters required to describe the less polar SSIPs associated with aromatic CH groups ($\alpha = 1.40$, $\beta = 0.70$) were obtained by fitting to experimental data for the free energy of transfer of aromatic hydrocarbons from water to *n*-hexadecane. Using this approach, it was also possible to obtain parameters for aromatic molecules with alkyl substituents. The aromatic carbon bearing the alkyl substituent is described by two SSIPs, one above and one below the plane of the ring, each with a value of $\beta = 0.88$ (Fig. 8b).

SSIP description of water

The parameters required to describe the properties of water as a solute have been determined experimentally by applying eqn (1) to measurements of association constants for the formation of 1 : 1 H-bonded complexes: $\alpha = 2.80$ and $\beta = 4.50$.¹⁹ In contrast, eqn (1) cannot be used directly to determine the parameters required to describe the properties of water as a solvent, because it is too polar to allow formation of singly H-bonded 1 : 1 complexes. We therefore used experimental data on phase transfer free energies to develop solvent parameters for water.

Water is a unique solvent in that the concentration of solvent-solvent interactions is known, because the H-bonds are almost fully bound in liquid water at 298 K.³¹ We can therefore set the value of $[S \cdot S]$ equal to 110 M in eqn (6). If we assume that the H-bond parameters that have been experimentally

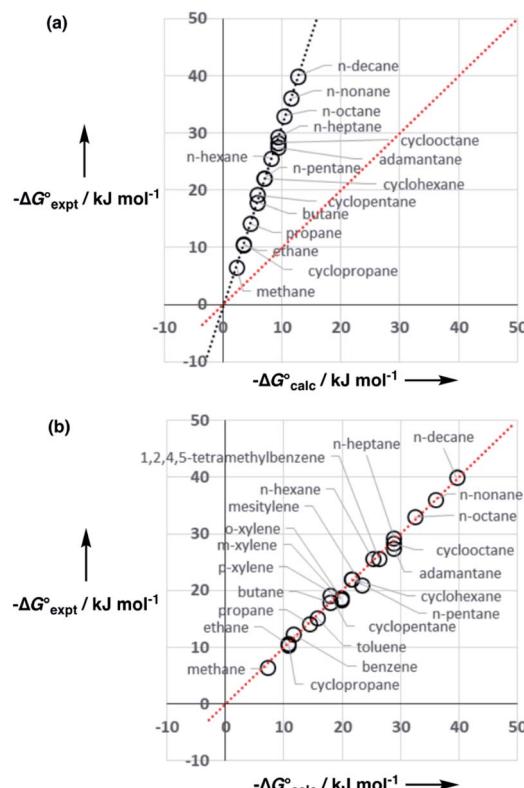


Fig. 9 (a) Comparison of the experimental free energy of transfer of alkanes from water to *n*-hexadecane ($-\Delta G^\circ_{\text{expt}}$) with the values calculated using solvent parameters for water of $\alpha_S = 2.80$, $\beta_S = 4.50$ and $C_\alpha = C_\beta = -0.47 \text{ kJ mol}^{-1}$ ($-\Delta G^\circ_{\text{calc}}$). The red line is $y = x$, and the line of best fit is black. (b) Comparison of the experimental free energy of transfer of aliphatic and aromatic hydrocarbons from water to *n*-hexadecane ($-\Delta G^\circ_{\text{expt}}$) with the values calculated using solvent parameters for water of $\alpha_S = 3.80$, $\beta_S = 3.47$ and $C_\alpha = C_\beta = -0.76 \text{ kJ mol}^{-1}$ ($-\Delta G^\circ_{\text{calc}}$). The red line is $y = x$.

measured for water as a solute can be used to describe water as a solvent, *i.e.* $\alpha_S = 2.80$ and $\beta_S = 4.50$, eqn (6) gives the constants required to describe water as $C_\alpha = C_\beta = -0.47 \text{ kJ mol}^{-1}$. However, calculation of the partition of alkanes between water and *n*-hexadecane using these parameters failed to reproduce the experimentally measured values (Fig. 9a). The preference for alkanes to partition into *n*-hexadecane is overestimated, which suggests that the solute parameters for water cannot be used to describe the properties of water as a solvent in this model. By using experimental data for the free energy of transfer of aliphatic and aromatic hydrocarbon solutes from water to *n*-hexadecane, it was possible to optimise the solvent parameters for water: $\alpha_S = 3.80$ and $\beta_S = 3.47$. Fig. 9b shows that the resulting parameters provide an excellent description of partitioning of a variety of different types of hydrocarbon into water.

SSIP description of polar solutes

The SSIP models for water and hydrocarbons described above allow accurate calculation of the partition of hydrocarbon solutes between water and *n*-hexadecane as shown in Fig. 9b. The same approach was then used to assign SSIP parameters for



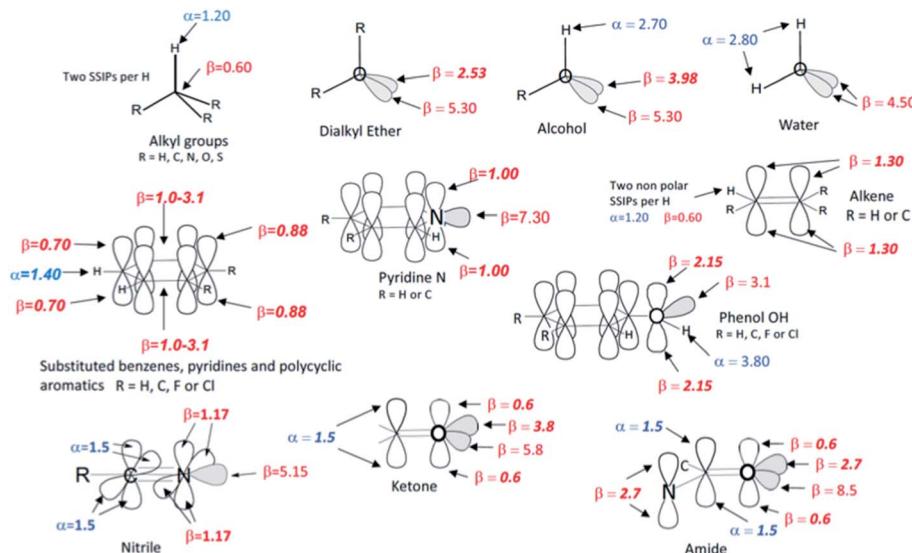


Fig. 10 The SSIP description of functional groups. The values shown as bold italic were optimised in order to minimise the rmsd between calculated and experimental free energies for the partition models listed in Table 2. The other values are based on previously measured experimental parameters.

an expanded solute set of 134 molecules that included alkenes, polycyclic aromatics, phenols, pyridines, water, alcohols and ethers. An initial set of SSIP parameters was derived from partition between water and hydrocarbons and then further refined using additional data from partition into other non-polar solvents (see below). The SSIP parameters that describe the functional groups present in these compounds are summarised in Fig. 10.

For some functional groups, experimental values of α and β have previously been determined by applying eqn (1) to measurements of association constants for the formation of 1 : 1 H-bonded complexes, and these values were used unmodified and assigned to the relevant SSIPs. However, these experimental parameters only provide information on the most polar site present in a solute, so when more than one type of α or β is required to describe a functional group, optimisation of the parameter describing the second site was often required. For example, an alcohol is described by one α and two β SSIPs (Fig. 10), but using the experimental value of 5.30 for both β SSIPs overestimated the polarity of this functional group. An accurate description of the partition coefficients of alcohols was obtained by reducing one of the two β parameters from 5.30 to 3.98. Alkyl groups are well-described by the same α and β parameters developed for alkanes above and to simplify the assignment problem the effects of electronegative substituents (O, N or S) on these parameters were assumed to be negligible. Similarly, the parameters developed for aromatic hydrocarbons generally provide a good description of substituted aromatic rings, with the exception that the value of β for the two H-bond acceptor sites over the centre of the ring had to be varied depending on the electronic properties of the ring substituents. The same approach starting from experimental functional group H-bond parameters and atom type hybridisation models was applied to ketones, nitriles, amines, amides, thioethers, alkyl fluorides, alkyl

chlorides, fluoro and chloro substituted benzenes and phenols in order to expand the range of solute functional groups (see ESI Section S3† for full details). The SSIP parameters for these functional groups were optimised using data from partition into both non-polar and polar solvents (see below).

Parameters for non-polar solvents

It is possible to determine experimental values of α_s and β_s for simple non-polar solvents by applying eqn (1) to measurements of association constants for the formation of 1 : 1 H-bonded complexes. For solvents that have only one type of donor and one type of acceptor SSIP, like aliphatic hydrocarbons, the values of C_α and C_β are equal and can be estimated using eqn (6). This approach was used to obtain parameters for carbon tetrachloride: $\alpha_s = 1.40$, $\beta_s = 0.60$, $C_\alpha = C_\beta = 2.58 \text{ kJ mol}^{-1}$. However, most organic solvents have more than one type of α or β site, so the experimentally determined values of α_s and β_s represent a weighted average of different interactions between solvent and solute, and the values of C_α and C_β are unlikely to be equal. Provided the parameters are optimised using experimental partition data, it turns out that only one α_s value and one β_s value can be used to accurately describe the properties of non-polar solvents. The experimental values of α_s and β_s for 10 non-polar solvents were used as a starting point for the development of a self-consistent set of solvent parameters by optimising against experimental partition data. The resulting values of α_s , β_s , C_α and C_β are listed in Table 2, and Fig. 11a illustrates the quality of the description of transfer free energies for dichloromethane.

Experimental data on association constants for the formation of 1 : 1 H-bonded complexes provide an independent validation of these solvent parameters. Combining eqn (3)–(5) gives eqn (10), which can be used to calculate values of ΔG° for formation of 1 : 1 complexes. For each solvent, the values of α_s ,

Table 2 Parameters for water and non-polar organic solvents and comparison of calculated free energies with experimental data. The values shown as bold italic were optimised in order to minimise the rmsd between calculated and experimental free energies for the partition models listed. The other values are based on previously measured parameters from experimental 1 : 1 complexation data

Solvent	Solvent descriptors					Solvent/water partition		1 : 1 complexation	
	α_s	C_α	β_s	C_β	C_0	n	rmsd/kJ mol ⁻¹	n	rmsd/kJ mol ⁻¹
Water	3.80	-0.76	3.47	-0.76	0				
Hexadecane	1.20	2.64	0.60	2.64	0	219	1.2		
Benzene	1.40	2.50	2.00	1.09	0	37	1.2	108	1.7
Toluene	1.40	2.50	2.00	1.09	0	34	1.1	46	1.7
Hexane	1.20	2.62	0.60	2.62	1.38	87	1.1	6	0.7
Cyclohexane	1.20	2.61	0.60	2.61	1.33	51	1.2	109	1.2
Carbon tetrachloride	1.40	2.58	0.60	2.58	1.34	86	1.2	475	0.8
Dichloromethane	1.80	2.16	1.40	1.76	1.73	28	1.3	42	1.1
Chloroform	2.10	1.78	1.30	2.11	0.60	71	1.6	13	1.2
1,2-Dichloroethane	1.70	2.23	1.60	1.41	1.93	58	1.6	34	1.0
Chlorobenzene	1.40	2.51	1.40	1.71	1.48	57	1.2	22	1.5
Perfluoroalkane	1.2	2.41	0.60	2.41	0.81	27	1.85	15	1.7

β_s , C_α and C_β in Table 2 were used in conjunction with previously determined experimental values of solute α and β parameters to calculate the free energy change for complexes formed by a variety of H-bond acceptors with H-bond donors. In all cases, the rmsd between the calculated and experimental values is less than 1.7 kJ mol⁻¹ (see ESI Section S4†) confirming

the reliability of the solvent parameters. Fig. 11b illustrates the quality of the description of 1 : 1 complexation data for dichloromethane.

$$\Delta G^\circ = -\alpha\beta + \alpha\beta_s + \alpha_s\beta + C_\alpha + C_\beta \quad (10)$$

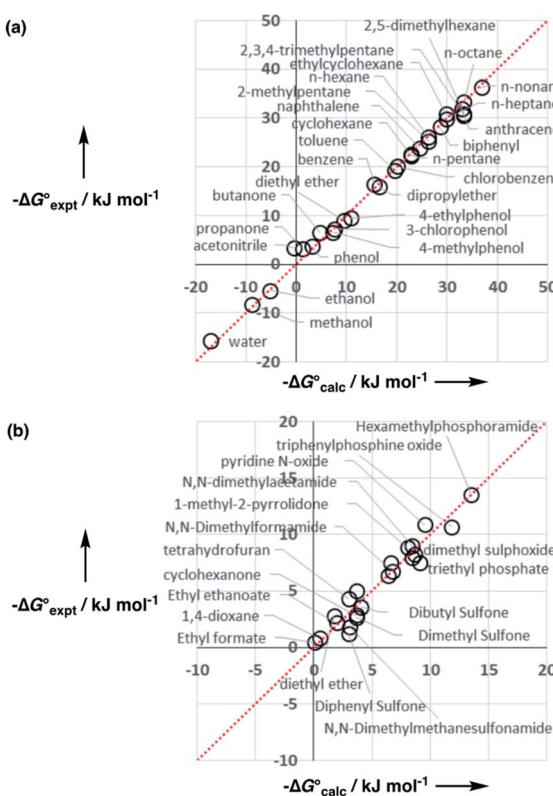


Fig. 11 Comparison of calculated ($-\Delta G^\circ_{\text{calc}}$) and experimental free energies ($-\Delta G^\circ_{\text{expt}}$) for (a) transfer of various solutes from water to dichloromethane (rmsd = 1.0 kJ mol⁻¹), and (b) formation of 1 : 1 complexes between 4-fluorophenol and various H-bond acceptors in dichloromethane (rmsd = 0.95 kJ mol⁻¹). The red lines are $y = x$.

Parameters for polar organic solvents

Organic solvents which contain polar functional groups must be treated differently, because it is no longer possible to use only one α_s value and one β_s value to describe a weighted average of solvation states. The reason for the failure of this simplistic description is that the non-polar regions of the solvent preferentially solvate non-polar solutes, and the polar groups on the solvent preferentially solvate polar solutes. In this case, both the polarities and the concentrations of the different sites on the solvent become important.

We will describe development of the model for ethers to illustrate the approach used to parameterise polar organic solvents. Ethers are the simplest class of polar organic solvent, because they have two types of acceptor SSIPs, which describe the non-polar hydrocarbon region (β_{s1}) and the polar oxygen sites (β_{s2}), but only one type of CH donor SSIP (α_{s1}). Fig. 12 shows that the free energy of transfer of alkanes from the gas phase into alkyl ethers is very similar to the value for transfer into *n*-hexadecane. This result suggests that the hydrocarbon component of an ether has similar properties to an alkane, so the solvation of non-polar solute SSIPs can be described using the same parameters used to describe alkane solvents, *i.e.* $\alpha_{s1} = 1.2$ and $\beta_{s1} = 0.6$. The parameter for the polar oxygen acceptor SSIP can be estimated from the solute values in Fig. 10, which gives $\beta_{s2} = 5.3$.

The solvation of solute H-bond acceptor SSIPs is described in a straightforward manner by eqn (5) using $\alpha_s = \alpha_{s1} = 1.2$ for the solvent H-bond donor SSIPs. Solvation of solute H-bond donors is more complicated, because they interact with two different acceptor sites. We assume that the solvent is present in a large

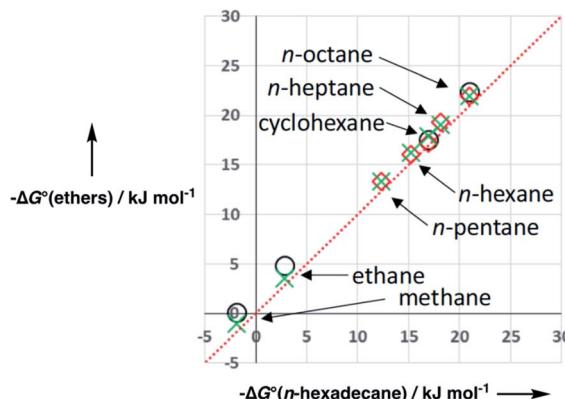


Fig. 12 Free energy of transfer of alkane solutes from the gas phase into ether solvents (circles tetrahydrofuran, crosses diethyl ether, diamonds di-*n*-butyl ether) compared to transfer from the gas phase into *n*-hexadecane. The red line is $y = x$.

excess relative to the solute, so that each solute SSIP is solvated according to an effective equilibrium constant for interaction with each type of solvent SSIP. The equilibrium constants for the interaction of a solute H-bond donor SSIP α with the two different solvent H-bond acceptor SSIPs β_{S1} and β_{S2} are given by eqn (11) and (12).

$$K_1 = e^{\frac{\alpha\beta_{S1} + C_{\beta_1}}{RT}} \quad (11)$$

$$K_2 = e^{\frac{\alpha\beta_{S2} + C_{\beta_2}}{RT}} \quad (12)$$

where the constants C_{β_1} and C_{β_2} describe the solvent–solvent interactions that are disrupted on formation of an interaction between solute and solvent (cf. eqn (4)).

The total free energy contribution due to solvation of the solute H-bond donor SSIP α is given by the sum of these equilibrium constants weighted by the fraction of interactions made with the relevant solvent SSIP, as shown in eqn (13).

$$\Delta g_s(\alpha) = -RT \ln \left(\frac{K_1}{K_1 + K_2} K_1 + \frac{K_2}{K_1 + K_2} K_2 \right) \quad (13)$$

The constants C_{β_1} and C_{β_2} are determined by the polarity and concentration of solvent–solvent interactions and are likely to vary from solvent to solvent. However, when values of transfer free energies from water to different ether solvents were used to independently optimise values of $C_{\alpha 1}$, $C_{\beta 1}$, $C_{\beta 2}$ and C_0 for tetrahydrofuran, diethyl ether and di-*n*-butyl ether, the variation in the values of the constants between solvents was rather small (see ESI Section S5†). Fig. 13 shows that using a single generic value of each constant for ethers (see Table 3) gives a good description of the experimental data for all three solvents.

For alcohols, there are two different types of acceptor SSIP, the polar oxygen and non-polar hydrocarbon, so eqn (11)–(13) can be used to describe the solvation of solute donor SSIPs. There are also two different types of donor SSIP, the polar OH and the nonpolar CH groups, so a similar set of eqn (14)–(16) are required to describe the equilibria involved in the solvation of solute acceptor SSIPs.

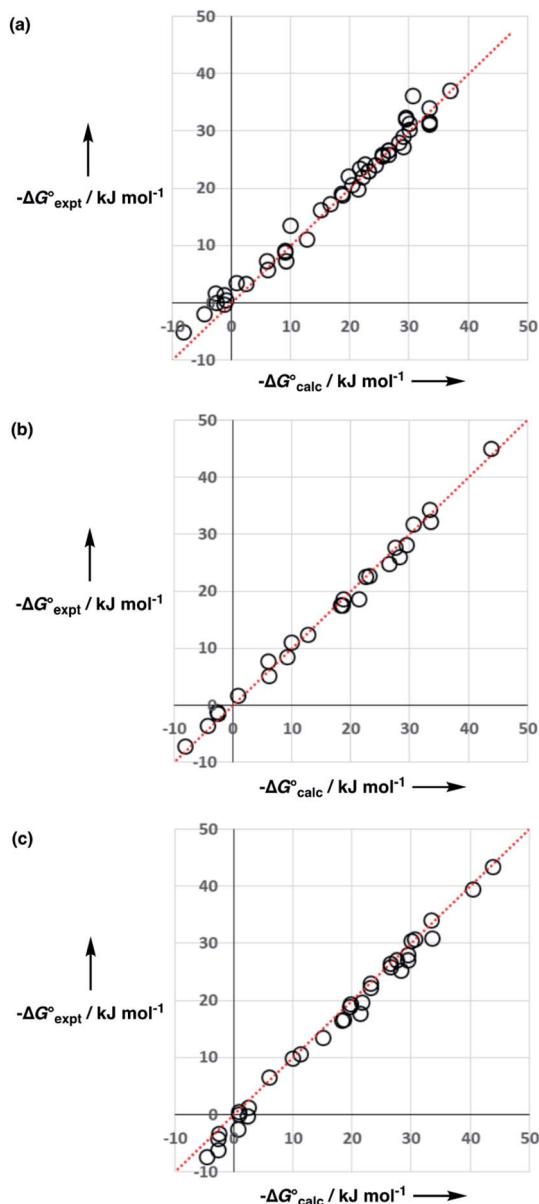


Fig. 13 Comparison of calculated ($-\Delta G^\circ_{\text{calc}}$) and experimental free energies ($-\Delta G^\circ_{\text{expt}}$) for transfer from water to (a) tetrahydrofuran (rmsd = 2.0 kJ mol⁻¹, $n = 49$), (b) diethyl ether (rmsd = 1.2 kJ mol⁻¹, $n = 25$), and (c) di-*n*-butyl ether (rmsd = 1.7 kJ mol⁻¹, $n = 35$). The red lines are $y = x$.

$$K'_1 = e^{\frac{\alpha\beta_{S1} + C_{\beta_1}}{RT}} \quad (14)$$

$$K'_2 = e^{\frac{\alpha\beta_{S2} + C_{\beta_2}}{RT}} \quad (15)$$

where the constants $C_{\alpha 1}$ and $C_{\alpha 2}$ describe the solvent–solvent interactions that are disrupted on formation of an interaction between solute and solvent.

$$\Delta g_s(\beta) = -RT \ln \left(\frac{K'_1}{K'_1 + K'_2} K'_1 + \frac{K'_2}{K'_1 + K'_2} K'_2 \right) \quad (16)$$



Table 3 Parameters for polar organic solvents and comparison of calculated transfer free energies with experimental data. The values shown as bold italic were optimised in order to minimise the rmsd between calculated and experimental free energies for the partition models listed. The other values are based on previously measured parameters from experimental 1 : 1 complexation data

Solvent class	Solvent descriptors								Solvent/water partition		
	α_{S1}	$C_{\alpha1}$	β_{S1}	$C_{\beta1}$	α_{S2}	$C_{\alpha2}$	β_{S2}	$C_{\beta2}$	C_0	n	rmsd/kJ mol ⁻¹
Ethers ^a	1.20	2.58	0.60	2.98	—	—	5.30	-3.65	2.26	109	1.7
Nitriles ^b	1.20	-0.79	0.60	2.69	1.50	2.67	5.15	-3.54	3.55	76	1.7
Ketones ^c	1.20	-0.78	0.60	2.82	1.50	2.71	5.80	-4.09	1.89	118	1.4
Alcohols ^d	1.20	2.75	0.60	2.95	3.50	-6.02	6.90	-6.42	0.18	604	1.5

^a Diethyl ether, di-*n*-butyl ether and tetrahydrofuran. ^b Acetonitrile, propionitrile and butyronitrile. ^c Acetone, butanone and cyclohexanone.

^d Methanol, ethanol, propan-1-ol, butan-1-ol, pentan-1-ol, hexan-1-ol, heptan-1-ol, octan-1-ol, decan-1-ol, propan-2-ol, butan-2-ol, 2-methylpropan-1-ol, 2-methylpropan-2-ol and 3-methylbutan-1-ol.

The effect of alcohols on the association constants for formation of 1 : 1 H-bonded complexes has been investigated in alcohol-alkane mixtures.³² These experiments suggest that alcohol self-association leads to an increase in the polarity of the hydroxyl group compared with monomeric alcohols in dilute solution. The effective value of the hydroxyl α increases from 2.7 to 3.5 and β increases from 5.3 to 6.9 in the self-associated bulk liquid. These parameters were therefore used as α_{S2} and β_{S2} to describe the polar SSIPs in alcohols. The alkane solvent parameters ($\alpha_{S1} = 1.2$ and $\beta_{S1} = 0.6$) were used to describe the non-polar hydrocarbon SSIPs in alcohols. Partition data were available for 14 different alcohol solvents, and these data were used to optimise a single set of constants ($C_{\alpha1}$, $C_{\alpha2}$, $C_{\beta1}$, $C_{\beta2}$ and C_0 in Table 3) that describe the solvent–solvent interactions in all of these alcohols.

Experimental values of α_s and β_s have been determined for nitrile and ketone solvents by applying eqn (1) to measurements of association constants for the formation of 1 : 1 H-bonded complexes (see ESI Section S6†). These values were therefore used to describe the polar SSIPs in these solvents, α_{S2} and β_{S2} , and the non-polar hydrocarbon SSIPs were described using the alkane parameters as above ($\alpha_{S1} = 1.2$ and $\beta_{S1} = 0.6$). As for ethers and alcohols, a generic set of constants ($C_{\alpha1}$, $C_{\alpha2}$, $C_{\beta1}$, $C_{\beta2}$ and C_0) were obtained for dialkyl ketones and for alkyl nitriles by minimising the rmsd between experimental partition data and calculated transfer free energies (Table 3).

Results

The training set used to develop the model contained 266 different solutes, 35 different solvents, and a total of 1884 independent measurements of partition free energies (see ESI Section S7† for details). These data were used to optimise 60 SSIP values used to describe the solutes and 45 parameters used to describe the solvents. The other 63 SSIP values and solvent constants used in the model were fixed using previous measurements on 1 : 1 H-bonded complexes. Thus the number of data points used in parameterisation of the model (1884) was more than 10 times the number of optimised parameters (105).

Experimental data for partition between water and *n*-hexadecane was available for 219 of the solutes in the training set. The solvent parameters in Tables 2 and 3 were used with SSIP

representations of these solutes to calculate transfer free energies between water and 34 different organic solvents (see ESI Section S9† for details). The results are illustrated in Fig. 14. The agreement is excellent with an overall rmsd of 1.4 kJ mol⁻¹ for 1713 data points, and there are no major outliers.

For 189 of the solutes in the training set, data was also available for partition between wet octanol and water, and these data were used to validate the generic alcohol solvent parameters. Comparison of calculated and experimental values was satisfactory with an rmsd of 1.6 kJ mol⁻¹. In addition, calculations were performed on a validation set of 84 solutes that were not present in any of data sets used to parameterise the model. Again good agreement with the experimental data was obtained with an overall rmsd of 1.7 kJ mol⁻¹ for octanol–water partition and 1.6 kJ mol⁻¹ for *n*-hexadecane–water partition. The original training set was limited to simple compounds, some of the validation set solutes contained more than one functional group (e.g. 1,4-dicyanobutane), and these compounds are described well. These results suggest that the model developed here is robust and has applicability beyond the systems used for parameterisation.

The performance of the model was evaluated by comparing the results with related methods. Fig. 15 compares the model with the performance of the SSIMPLEx method for the same for

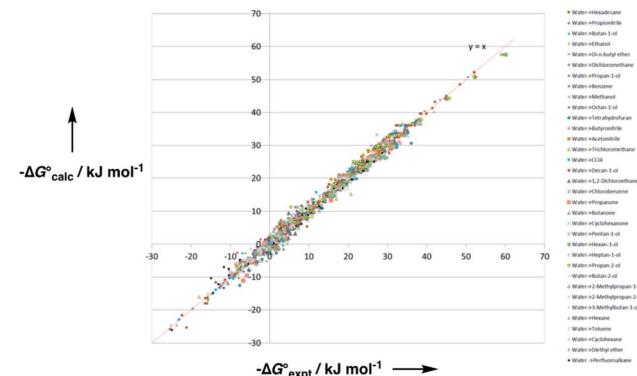


Fig. 14 Comparison of free energies calculated using the SSIP model developed here ($-\Delta G^{\circ}_{\text{calc}}$) with experimental values ($-\Delta G^{\circ}_{\text{expt}}$) for transfer of 219 different solutes from water to 35 solvents (overall rmsd is 1.4 kJ mol⁻¹ for 1713 data points). The red line is $y = x$.



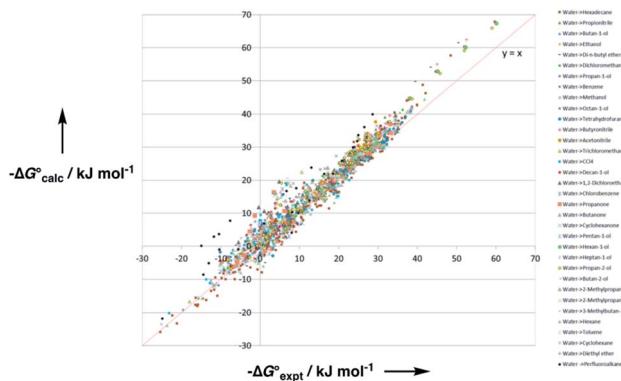


Fig. 15 Comparison of free energies calculated using SSIMPLE ($-\Delta G_{\text{calc}}^{\circ}$) with experimental values ($-\Delta G_{\text{expt}}^{\circ}$) for transfer of 219 different solutes from water to 35 solvents (overall rmsd is 3.5 kJ mol⁻¹ for 1713 data points). The red line is $y = x$.

1713 experimental data points used in Fig. 14. SSIMPLE uses *ab initio* calculations to obtain SSIPs for both solute and solvent and evaluates all pairwise SSIP interactions in the liquid phase, including van der Waals interactions.²¹ There are very few empirical parameters in SSIMPLE, and although the trend is reasonable, the performance is significantly worse than the model developed in this paper, with an overall rmsd of 3.5 kJ mol⁻¹. For octanol–water partition, a number of methods have been developed. *c log P* is a structure-based method that uses empirically derived parameters for functional group fragments, so it can be generalised to a wide range of solutes but cannot be generalised to other solvents.³³ Abraham's linear solvation energy relationship (LSER) is based on summation over solvent–solute interactions and can be applied to a wide range of solvents, but empirical parameters are required for each solute.¹⁸ For 189 wet octanol–water data points, the rmsd values obtained using *c log P* and Abraham's linear solvation energy relationship (LSER) are 0.8 kJ mol⁻¹ and 1.1 kJ mol⁻¹ respectively. Although the rmsd obtained using the model described here is slightly higher (1.6 kJ mol⁻¹), the advantage is that extrapolation to a wider range of solutes and solvents requires minimal reparameterisation.

Discussion

SSIP representation of solutes

The representation of functional groups illustrated in Fig. 10 provides a straightforward method for determining the number of SSIPs required to describe a molecule based on assignment of an integer number of SSIPs to each heavy atom according to hybridisation and number of attached hydrogen atoms. Thus each alkane and alkene CH is assigned two SSIPs, one α and one β . Each sp^2 hybridised carbon is assigned two β SSIPs that represent the two faces of the π -system. Each sp hybridised carbon is assigned four β SSIPs to represent the π -electrons. A similar approach is used for oxygen and nitrogen, but an additional β SSIP is assigned for each lone pair. The approach for halogens is slightly different as explained below.

Many of the required values of the α and β parameters were available from experimental measurements of association constants for 1 : 1 H-bonded complexes in carbon tetrachloride, and these parameters were used directly. For alkanes, which do not form H-bonds in carbon tetrachloride, experimental 1 : 1 complexation data in alkane solvents have been used to determine solvent H-bond parameters, and these parameters were used for alkane solutes.

For functional groups that have more than one possible H-bond acceptor site, such as alcohols, the experimental H-bond parameter was used for the first β SSIP, and value for the second β SSIP was optimised using partition data. In these cases, the optimised value of the second SSIP was always less polar than the experimental value used for the first SSIP. These results are consistent with a range of computational and experimental studies.³⁴⁻⁴⁰ Analysis of the H-bond acceptor parameters for carbonyl groups that make an intramolecular H-bond confirms the accuracy of the SSIP values. Fig. 16 shows that the equilibrium constant K'_2 is a surrogate for K_2 , the equilibrium constant for formation of a second intermolecular H-bond with a carbonyl group. The experimental values of K'_2 correspond to β values of 3.8–4.3, which are significantly lower than the H-bond parameters measured for the first H-bond using K_1 (5.5–6.4).³⁷ These parameters agree well with the two different β values (3.8 and 5.8) used to describe the two lone pairs of a carbonyl group in Fig. 10.

Functional group parameterisation

Parameters have been derived for the most common organic functional groups in this paper, but we have not discussed halogens so far. Chloroalkanes exemplify how the general approach can be extended to other functional groups. Eqn (9) indicates that the number of SSIPs required to describe a chlorine substituent is 5, which gives $[SSIP] = 219$ M for CH_2Cl_2 , 212 M for CHCl_3 , 227 M for $\text{CH}_2\text{Cl}_2\text{CH}_2\text{Cl}_2$ and 207 M for CCl_4 . Experimental data on the association constants for 1 : 1 complexes of chloroalkanes and H-bond donors such as phenols suggest that one of these SSIPs has a β value of 2.3.²³ However chlorine substituents are non-polar, so the other four

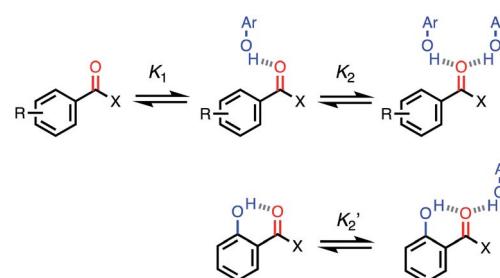


Fig. 16 Sequential equilibria for formation of a 1 : 1 and 2 : 1 complex between a carbonyl H-bond acceptor and a phenol H-bond donor ($R = H$, 2-OMe or 4-OMe, $X = H$, Me, Ph or OMe). The equilibrium constant K'_2 for formation of a 1 : 1 complex with a carbonyl H-bond acceptor that has an intramolecular H-bond can be used to estimate the value of the second β SSIP required to describe solvation of a carbonyl oxygen.

SSIPs were assigned the same parameters as for alkanes, two with $\alpha = 1.2$ and two with $\beta = 0.6$. Partition of *t*-butyl chloride between water and 10 organic solvents is well predicted using these values (rmsd = 1.1 kJ mol⁻¹), which indicates that this SSIP representation provides a reasonable description of chlorine.

However, the agreement between calculated and experimental free energies of transfer for chloroalkanes which have CH₂Cl groups was not so good (rmsd = 1.6 kJ mol⁻¹ for 10 organic solvents). We have modelled alkyl substituents in various compound classes discussed above using the same parameters as for alkanes. Polarisation of C–H bonds adjacent to heteroatoms might be expected, but reasonable results were obtained without modifying the SSIPs in most cases. As the large values of α found for dichloromethane and chloroform might suggest, this is not the case for chloroalkanes. Fig. 17 shows that increasing the value of α for the C–H groups adjacent to the chlorine from 1.2 to 1.6 significantly improved the description of these compounds (rmsd = 0.9 kJ mol⁻¹ for 10 organic solvents). Note that one of the assumptions implicit in the approach described here is that there are no changes in the net van der Waals interaction in the equilibrium illustrated in Fig. 1, so that partition can be described purely in terms of polar interactions between SSIPs. The excellent agreement between

the calculated and experimental partition data obtained for chloroalkanes suggests that this assumption holds for second row elements as well.

SSIP representation of solvents

For non-polar solvents, interactions with solutes can be described using just two types of solvent SSIP, α_S and β_S . These parameters are available for several solvents from experimental measurements of association constants for 1 : 1 H-bonded complexes, and these values were used directly. However, most polar solvents require additional SSIPs in order to account for differences in the way they interact with polar and non-polar functional groups. For these solvents, the two SSIPs used to describe alkanes were used to describe the non-polar regions of the solvent, and two SSIPs were used to describe the polar regions of the solvent. For ethers, ketones and nitriles, solvation of polar solutes can be successfully described by using the relevant functional group H-bond acceptor parameters. For alcohols, self-association increases the apparent polarity of the solvent, so the hydroxyl groups were described using experimentally determined solvent parameters. Experimental solvent parameters are not available for the other strongly self-associated solvent, water, so the values of α_S and β_S were optimised using partition data. These optimised solvent parameters suggest that water interacts more strongly with H-bond acceptors than alcohols ($\alpha_S = 3.8$ for water compared with $\alpha_{S2} = 3.5$ for alcohols) but less strongly with H-bond donors ($\alpha_S = 3.5$ for water compared with $\alpha_{S2} = 6.9$ for alcohols). The difference between the description of water and alcohols is consistent with the Taft solvent parameters determined using solvatochromic dyes: the Taft H-bond donor parameter is larger for water (1.2 compared with 0.8 for alcohols), and the Taft H-bond acceptor parameter is smaller (0.5 compared with 0.8 for alcohols).²⁰

Each solvent SSIP is also assigned a constant (C_α or C_β), which quantifies the interactions that are broken with the bulk solvent when the solvent SSIP solvates a solute. The values of these constants were optimised using partition data, and they capture information about both the polarity and the concentration of solvent–solvent interactions. For solvents with just two types of SSIP, the sum of the two constants can be directly related to the properties of the solvent–solvent interactions by eqn (17).

$$C_\alpha + C_\beta = -\alpha_S \beta_S + RT \ln[S \cdot S] \quad (17)$$

As explained above, the effective concentration of solvent–solvent interactions [S · S] is not easy to estimate for most solvents. When eqn (1) was first proposed the constant of +6 kJ mol⁻¹ was obtained from measurements of formation of 1 : 1 H-bonded complexes in carbon tetrachloride.¹⁹ This value corresponds to [S · S] = 10 M, and we have assumed that a constant of +6 kJ mol⁻¹ can also be used in eqn (1) to describe 1 : 1 complexation other organic solvents. However, the values of C_α and C_β obtained here from the partition data can be used in eqn (17) to obtain a direct experimental measurement of this parameter. The results shown in Table 4 confirm that the values of $RT \ln[S \cdot S]$ for non-polar solvents are approximately constant

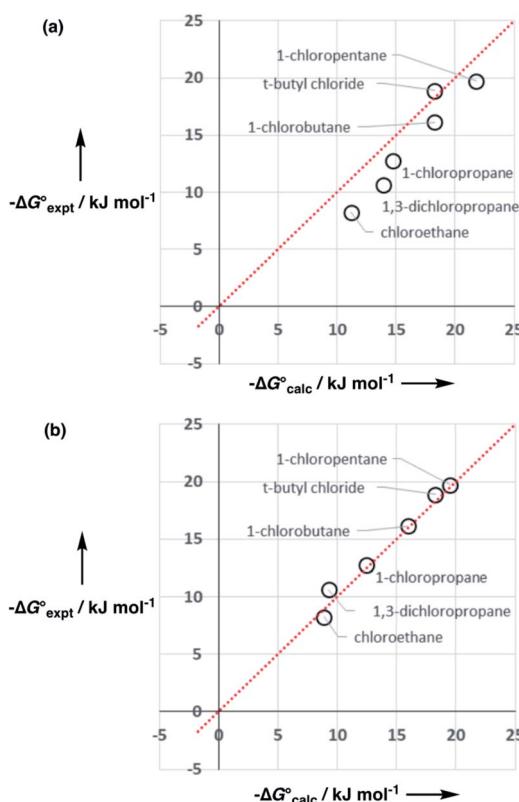


Fig. 17 Comparison of calculated ($-\Delta G^\circ_{\text{calc}}$) and experimental free energies ($-\Delta G^\circ_{\text{expt}}$) for transfer of chloroalkane solutes from *n*-hexane into water. (a) Calculated values obtained using $\alpha = 1.2$ for the C–H groups in CH₂Cl (rmsd = 2.3 kJ mol⁻¹). (b) Calculated values obtained using $\alpha = 1.6$ for the C–H groups in –CH₂Cl (rmsd = 0.7 kJ mol⁻¹). The line is $y = x$ in each case.



Table 4 Values of $RT \ln[S \cdot S]$ (kJ mol⁻¹)^a

Solvent	$RT \ln[S \cdot S]$
Hexadecane	6.0
Benzene	6.4
Toluene	6.4
Hexane	6.0
Cyclohexane	5.9
Carbon tetrachloride	6.0
Dichloromethane	6.4
Chloroform	6.6
1,2-Dichloroethane	6.4
Chlorobenzene	6.2
Perfluoroalkane	5.5

^a Calculated using eqn (17).

(+5.5 to +6.6 kJ mol⁻¹). The high value for chloroform (6.6 kJ mol⁻¹) is indicative of the greater solvating power of a more polar solvent, and the low value for perfluoroalkanes (5.5 kJ mol⁻¹) reflects the very weak intermolecular interactions present in these solvents. The results in Table 4 show why using a value of +6 kJ mol⁻¹ for the constant in eqn (1) provides an accurate description of the association constants for formation of 1 : 1 complexes in different non-polar solvents. In contrast, partition is much more sensitive to the precise values of the solvent constants, because the small differences in Table 4 are significant when they are multiplied up by the total number of SSIPs in a solute.

The behaviour of polar solvents is more complicated, because with two types of donor and two types of acceptor there are four different types of solvent–solvent interaction. However in the case of alcohols, the solvent parameters in Table 3 show that $C_{\alpha 1} \approx C_{\beta 1}$ and $C_{\alpha 2} \approx C_{\beta 2}$. Structural studies on alcohols showing clustering of the alkyl chains and separate H-bonded aggregates of the hydroxyl groups.^{41,42} In other words to a first approximation, we can think of alcohols as consisting of two independent solvating domains. The solvent parameters that describe the non-polar hydrocarbon domain in alcohols are similar to those found for alkanes. Using the values of α_{S1} , $C_{\alpha 1}$, β_{S1} and $C_{\beta 1}$ in eqn (17) gives $RT \ln[S \cdot S] = +6.4$ kJ mol⁻¹ for the hydrocarbon domain in alcohols, which is similar to the values found for non-polar solvents in Table 4. The polar hydroxyl domain is described by α_{S2} , $C_{\alpha 2}$, β_{S2} and $C_{\beta 2}$, and using these values in eqn (17) gives a value of +11.7 kJ mol⁻¹ for $RT \ln[S \cdot S]$, which is the same as the value for water in Table 4. The model that emerges from the partition data is that alcohol solvents behave as a mixture of a water-like domain and an alkane-like domain.

There is an additional solvent constant C_0 , which is used to describe differences between solvents that are not accounted for by the SSIP interaction model. The values of C_0 listed in Tables 2 and 3 are all positive and small (generally less than +2 kJ mol⁻¹), and there are no obvious patterns. This result indicates that a simple model based on pairwise interactions between specific interaction sites on solvent and solute provides a rather general description of solvation phenomena and that there are no major additional factors that need to be considered.

The hydrophobic effect

Although the approach was developed from experiments on H-bonded complexes in non-polar solvents, the model provides a good description of very different phenomena, such as the hydrophobic effect in water. Interestingly, there is no need for different treatments of the water and hydrocarbon solvents or the need for a special cavitation term to describe the breaking of water structure. Free energy of transfer of a non-polar molecule from a non-polar liquid into water is a process dominated by disruption of water–water interactions and has been quantitatively linked to the molecular volume or molecular surface area of the solute.⁴³ It has been estimated that the free energy of transfer from a non-polar liquid into water is unfavourable by 0.2 kJ mol⁻¹ Å⁻² of non-polar molecular surface area, but calculation of non-polar surface area requires specialist software.⁴⁴ The total number of SSIPs that are assigned to a molecule using the relatively simple procedure outlined above is closely correlated to the molecular surface area for diverse structural type and functionality (see ESI Section S8†), so the magnitude of the hydrophobic effect can be quite accurately estimated. The advantage of this SSIP approach is that it is a simple spreadsheet calculation.

Consider for example transfer of the 9 different alkanes in Table 5 from *n*-hexadecane into water. Each of these alkanes has 16 hydrogen atoms, and each C–H groups has two SSIPs with values of $\alpha = 1.2$ and $\beta = 0.6$. The free energy of solvation for each of the SSIPs can be calculated using eqn (4) and (5) with the solvent constants listed in Table 2.

For water:

$$\Delta g_S(\alpha)/\text{kJ mol}^{-1} = -\alpha\beta_S - C_\beta = -1.20 \times 3.47 + 0.76 = -3.40$$

$$\Delta g_S(\beta)/\text{kJ mol}^{-1} = -\alpha_S\beta - C_\alpha = -3.80 \times 0.60 + 0.76 = -1.52$$

For *n*-hexadecane:

$$\Delta g_S(\alpha)/\text{kJ mol}^{-1} = -\alpha\beta_S - C_\beta = -1.20 \times 0.60 - 2.64 = -3.36$$

$$\Delta g_S(\beta)/\text{kJ mol}^{-1} = -\alpha_S\beta - C_\alpha = -0.60 \times 1.20 - 2.64 = -3.36$$

The difference between these solvation energies is $(-3.40 - 1.52) - (-3.36 - 3.36) = 1.80$ kJ mol⁻¹, which represents the free energy of transfer of one C–H group from *n*-hexadecane to water. Thus the calculated value for the free energy of transfer is the same for all of the alkanes, $16 \times 1.8 = 28.8$ kJ mol⁻¹, which agrees very well with the experimental values in Table 5 (29 ± 2 kJ mol⁻¹).

Phase change equilibria

The behaviour of alkanes with respect to liquid–liquid transfer is remarkably predictable, but this is not the case for gas–liquid and solid–liquid equilibria. Table 5 also shows experimental values for the free energy of transfer of C_xH_{16} alkanes from *n*-hexadecane into the gas phase. These values vary significantly from one alkane to another and span range of more than 10 kJ mol⁻¹. It is clear that there is an additional contribution to



Table 5 Experimental free energies of transfer for alkane solutes at 298 K

Alkane	Formula	n-Hexadecane to water	n-Hexadecane to gas	Melting point K
		$\Delta G^\circ/\text{kJ mol}^{-1}$	$\Delta G^\circ/\text{kJ mol}^{-1}$	
n-Heptane	C_7H_{16}	29.3	18.1	182
3-Methylhexane	C_7H_{16}	28.7	17.4	154
2,2-Dimethylpentane	C_7H_{16}	28.0	16.0	149
Ethylcyclohexane	C_8H_{16}	31.1	22.1	162
Propylcyclopentane	C_8H_{16}	30.6	21.7	156
cis-1,2-Dimethylcyclohexane	C_8H_{16}	28.6	21.9	223
trans-1,4-Dimethylcyclohexane	C_8H_{16}	28.7	20.8	236
Cyclooctane	C_8H_{16}	28.3	24.7	288
Adamantane	$\text{C}_{10}\text{H}_{16}$	27.5	28.1	543

the free energy of transfer of a molecule into the gas phase that is not included in the liquid phase solvation model described here. The origin of this difference is not known, but cyclic structures appear to have a dramatic effect, and in order to obtain an accurate model of gas–liquid equilibria, most methods require an empirical correction based on experimentally determined vapour pressures. Table 5 also shows that there is a variation of 400 K in the melting points of the 9 different alkanes, which highlights the dramatic difference between these phase change equilibria and liquid–liquid transfer. The equations in this paper calculate the free energy of a solute in a solvent relative to a liquid-like reference state, and additional contributions from changes in internal molecular motion are likely to be significant when a phase change is involved.

The fluorous effect

Fluorous solvents are an important class of solvents that exhibit a range of anomalous properties compared with other liquids. Perfluorocarbons are more volatile than the corresponding hydrocarbons, dissolve very high concentrations of gaseous solutes, and do not mix with either non-polar or polar solvents. The preferential dissolution of perfluorocarbon solutes in perfluorocarbon solvents is known as the fluorous effect and has been exploited in separation technology. As indicated in Table 4, the low value of $RT \ln[S \cdot S]$ provides a good description of the weak intermolecular interactions present in perfluorocarbon solvents for most solutes. However, the model breaks down for perfluorocarbon solutes. The free energy of transfer of perfluorocarbon solutes (C_5F_{12} to C_8F_{18}) from alkanes to water has a slightly higher error than for other solutes ($\text{rmsd} = 4.8 \text{ kJ mol}^{-1}$, $n = 4$), but the errors are significant for the transfer from perfluorocarbon solvents to alkanes ($\text{rmsd} = 13.8 \text{ kJ mol}^{-1}$) and to water ($\text{rmsd} = 18.6 \text{ kJ mol}^{-1}$). The strength of the interactions between perfluorocarbon solutes and perfluorocarbon solvents is consistently underestimated by the model. The origin of this special interaction between perfluorocarbons is difficult to understand, especially given the volatility of these substances, which indicates weak intermolecular interactions. Steric effects that interfere with molecular packing may contribute to this anomaly, and considerable changes in volume occur when hydrogenated and fluorinated

substances are mixed.⁴⁵ For example, when *n*-perfluorohexane is dissolved in *n*-octane at infinite dilution, the molar volume of the solute increases by 13%.^{46,47} The effect is even more pronounced when *n*-alkanes are dissolved in *n*-perfluorooctanes, where the molar volume of the solute increases by 20%.⁴⁸ It is as though layers of empty space are created around each molecule, and this behaviour clearly cannot be captured by a simple model that assumes van der Waals interactions balance and polar interactions dominate. In conclusion, the fluorous effect is outside the scope of the model developed in this paper even though the model provides a good description of the solvation of different types of solute by perfluorocarbon solvents.

Conclusions

Solvation of organic molecules is a fundamentally important property that governs intermolecular association of solutes and phase transfer equilibria. Both processes depend on non-covalent interactions between solutes and solvents, and we have shown that it is possible to quantitatively account for the free energy changes associated with intermolecular contacts by using surface site interaction points (SSIP). A molecule is described by a discrete set of SSIPs, which define the array of all possible interactions that can be made with other molecules. Association constants for the formation of 1 : 1 complexes in solution are governed by the interaction of the most polar SSIP on each solute, whereas phase transfer equilibria are governed by the sum of the interactions made by all of the solute SSIPs with the solvent. A large body of experimental data is available for both processes, and here we use this information to parameterise an integrated model for describing non-covalent interactions and solvation.

Solutes are described based on the composition of functional groups, which allows straightforward rule-based translation of chemical structure into a SSIP description without the need for the *ab initio* calculations we have used previously. Positive SSIPs (α) are assigned to hydrogen atoms, and negative SSIPs (β) are assigned to represent lone pairs and π -electron density. The total number of SSIPs used to represent a solute is fixed such that the concentration of SSIPs in a liquid is

approximately constant. The assumption is that van der Waals contributions cancel out in any association or phase transfer equilibria, so that the behaviour of the system is dominated by polar interactions between SSIPs. Equilibria are treated as a competition between pairwise interactions of solute and solvent SSIPs, which has been shown previously to provide a rather good description of experimental data on 1:1 complexation in organic solvents.

Values of the most polar SSIPs required to describe each functional group were obtained from the experimentally determined H-bond parameters α and β . The values of any additional SSIPs required to complete description of each functional group were obtained by optimisation using experimental data on phase transfer equilibria. Similarly, the values of the SSIPs required to describe many organic solvents (α_s and β_s) have been determined previously from experimental data on 1:1 complexation. The SSIP description of other solvents like water, where 1:1 complexes are not sufficiently stable for experimental study, were obtained by optimisation using experimental data on phase transfer equilibria. For solvents that contain both polar and non-polar functional groups like alcohols, two sets of solvent SSIPs were used to describe the equilibrium between the two different solvation modes. In addition, for each solvent SSIP, a constant term was used to describe the effects of solvent–solvent interactions. The resulting model is simple to implement using just a spreadsheet (see ESI Section S9†) and accurately describes the transfer of a wide range of different solutes from water to a wide range of different organic solvents (overall rmsd is 1.4 kJ mol⁻¹ for 1713 data points).

The model described above is the result of a feasibility study based on solutes with one functional group and a limited set of solvents. The purpose was to determine whether the solvent competition model originally applied to H-bond complexes in non-polar solvents could be extended to describe whole molecule solvation and partition between organic solvents and water. This simple model describes the hydrophobic effect with surprising accuracy. It has also been possible to deduce new descriptors for range of organic solvents that were not accessible by direct investigation of H-bond formation in non-polar solvents. These empirical parameters should be of value in linking the results of *ab initio* quantum mechanics calculations with the thermodynamic properties of molecules in solution.

Experimental H-bond parameters are available for most organic functional groups, and these parameters can be used to develop the method for application to a much wider range of solutes than those used in the parameterisation described here. In addition, the availability of experimental H-bond parameters for anionic and cationic functional groups may provide a method for the treatment of ionisable compounds. One limitation is that the experimental data on 1:1 complexation and partition is generally restricted to room temperature, which means that extrapolation to different temperatures would require a new treatment. There are some challenges that will have to be addressed in development of the model to tackle more complex polyfunctional compounds. Electronic interactions between functional groups through the bonding framework can affect the H-bond parameters, and intramolecular

non-covalent interactions may change the availability of SSIPs for interaction with solvent. In addition, the current model is based purely on covalent connectivity, and a more elaborate treatment would be required to tackle the effects of conformer distribution on solvation energies.

Methods

Experimental partition data

The initial training set of neutral organic solutes included a variety of common functional groups and various structural types. The selected solutes had published partition coefficients available for several different solvents. Gas-n-hexadecane partition coefficients were available for the whole set and gas–water partition for 219 molecules. Thermodynamic values for solvent–solvent partition were calculated from the available values of the free energy of transfer from gas to solvent (see ESI Section S7† for values and references to data sources).

Analysis of chemical structures

From the SMILES string each molecule was fragmented into constituent atom types.⁴⁹ Each non-hydrogen atom was assigned a group code based on SMARTS nomenclature that described the atom and the directly bonded neighbours. Each atom code was assigned a number of donor and acceptor SSIP. In the optimisation process, the number of SSIP per atom was constrained. Only integer numbers of SSIP per atom were allowed, and a close correlation between the total number of SSIP used to describe a molecule and the molecular volume was maintained. Aromatic groups were assigned an additional code to describe a SSIP in the centre of the π -face of each 6 membered ring.

Calculation procedure

For each SSIP, the free energies of interaction with solvent were calculated using eqn (4) and (5) for simple solvents or eqn (11)–(16) for complex solvents. A lookup table was used to retrieve the free energy calculated for each code, and these values were summed as in eqn (7) to calculate the overall transfer free energy from the reference phase. Free energy of transfer between solvents was calculated with eqn (8) and compared with experimental values. The calculations were performed with Microsoft Excel, and the matrix of compounds and atom codes required for the calculations was assembled by analysis of SMILES strings using a combination of the fragmentation tools in Advanced Algorithm Builder software and the open source CDK tools accessible within Excel through the LICSS add-in.^{33,50,51}

Optimisation of parameters

Published values of functional group H-bond parameters derived from experimental data were used as a starting point.^{19,23,24} An iterative approach was used to assign new α and β values and new solvent constants by minimisation of the root mean square deviation (rmsd) between calculated and experimental values. A stepwise approach was adopted starting with a small initial training set of hydrocarbon solutes from which it



was possible to deduce descriptors for the water model based on partition data for the systems hexadecane/water and benzene/water. With parameters for water established it was a straightforward process to refine parameters for other solvents and to assign α and β values to interaction sites for which there was no reliable precedent.

Data availability

All supporting data is provided in the ESI.†

Author contributions

All authors contributed to writing the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Dr V. Grigorev for providing experimental data from the HYBOT database. We thank Marianda Twydell for the artwork used as the graphical abstract. We thank the Herchel Smith Fund for financial support.

Notes and references

- 1 D. Constantinescu and J. Gmehling, *J. Chem. Eng. Data*, 2016, **61**, 2738.
- 2 A. Fredenslund, *Vapor-liquid equilibria using UNIFAC: a group-contribution method*, Elsevier, 2012.
- 3 A. Fredenslund, R. L. Jones and J. M. Prausnitz, *AIChE J.*, 1975, **21**, 1086.
- 4 J. Lohmann, R. Joh and J. Gmehling, *Ind. Eng. Chem. Res.*, 2001, **40**, 957.
- 5 D. Qiu, P. S. Shenkin, F. P. Hollinger and W. C. Still, *J. Phys. Chem. A*, 1997, **101**, 3005.
- 6 E. M. Duffy and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2000, **122**, 2878.
- 7 C. J. Cramer and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 760.
- 8 J.-L. Rivail and D. Rinaldi, *Computational Chemistry, Review of Current Trends*, World Scientific, New York, 1996, p. 139.
- 9 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999.
- 10 A. Pomogaeva and D. M. Chipman, *J. Phys. Chem. A*, 2013, **117**(28), 5812.
- 11 O. Andreussi, I. Dabo and N. Marzari, *J. Chem. Phys.*, 2012, **136**, 064102.
- 12 C. Hille, S. Ringe, M. Deimel, C. Kunkel, W. E. Acree, K. Reuter and H. Oberhofer, *J. Chem. Phys.*, 2019, **150**, 041710.
- 13 F. Eckert and A. Klamt, *AIChE J.*, 2002, **48**, 369.
- 14 A. Klamt and F. Eckert, *Fluid Phase Equilib.*, 2000, **172**, 43.
- 15 A. Klamt, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1338.
- 16 M. H. Abraham, *Chem. Soc. Rev.*, 1993, **22**, 73.
- 17 M. H. Abraham, A. Ibrahim, A. M. Zissimos, Y. H. Zhao, J. Comer and D. P. Reynolds, *Drug Discovery Today*, 2002, **7**, 1056.
- 18 M. H. Abraham and W. E. Acree Jr, *J. Org. Chem.*, 2010, **75**, 1006.
- 19 C. A. Hunter, *Angew. Chem., Int. Ed. Engl.*, 2004, **43**, 5310.
- 20 R. Cabot and C. A. Hunter, *Chem. Soc. Rev.*, 2012, **41**, 3485.
- 21 C. A. Hunter, *Chem. Sci.*, 2013, **4**, 1687.
- 22 A. Oliver, C. A. Hunter, R. Prohens and J. L. Rosselló, *J. Comput. Chem.*, 2018, **39**, 2371.
- 23 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**, 18262.
- 24 D. Musumeci, C. A. Hunter, R. Prohens, S. Scuderi and J. F. McCabe, *Chem. Sci.*, 2011, **2**, 883.
- 25 M. D. Driver, M. J. Williamson, J. L. Cook and C. A. Hunter, *Chem. Sci.*, 2020, **11**(17), 4456.
- 26 N. J. Buurma, J. L. Cook, C. A. Hunter, C. M. R. Low and J. G. Vinter, *Chem. Sci.*, 2010, **1**, 242.
- 27 V. Amenta, J. L. Cook, C. A. Hunter, C. M. R. Low and J. G. Vinter, *Org. Biomol. Chem.*, 2011, **9**, 7571.
- 28 V. Amenta, J. L. Cook, C. A. Hunter, C. M. R. Low and J. G. Vinter, *J. Phys. Chem. B*, 2012, **116**, 14433.
- 29 R. Cabot, C. A. Hunter and L. M. Varley, *Org. Biomol. Chem.*, 2010, **8**, 1455.
- 30 S. Gomez, N. Rojas-Valencia, S. A. Gomez, C. Cappelli, G. Merino and A. Restrepo, *Chem. Sci.*, 2021, **12**, 9233.
- 31 N. G. M. Pay and M. C. R. Symons, *J. Chem. Soc., Faraday Trans.*, 1993, **89**, 2417.
- 32 S. Henkel, M. C. Misuraca, P. Troselj, J. Davidson and C. A. Hunter, *Chem. Sci.*, 2018, **9**, 88.
- 33 P. Japertas, R. Didziapetrис and A. Petrauskas, *Quant. Struct.-Act. Relat.*, 2002, **21**, 23.
- 34 M. C. R. Symons, *Pure Appl. Chem.*, 1986, **58**, 1121.
- 35 G. Eaton, A. S. Pena-Nunez and M. C. R. Symons, *J. Chem. Soc., Faraday Trans.*, 1988, **84**, 2181.
- 36 G. Eaton, M. C. R. Symons, P. P. Rastogi, C. O'Duinn and W. E. Waghorne, *J. Chem. Soc., Faraday Trans.*, 1992, **88**, 1137.
- 37 M. Berthelot, C. Laurence, D. Foucher and R. W. Taft, *J. Phys. Org. Chem.*, 1996, **9**, 255.
- 38 E. S. Feldblum and I. T. Arkin, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 4085.
- 39 J. Yang, L. A. Christianson and S. H. Gellman, *Org. Lett.*, 1999, **1**, 11.
- 40 J. Yang and S. H. Gellman, *J. Am. Chem. Soc.*, 1998, **120**, 9090.
- 41 R. Böhmer, C. Gainaru and R. Richert, *Phys. Rep.*, 2014, 545, 125.
- 42 N. P. Franks, M. H. Abraham and W. R. Lieb, *J. Pharm. Sci.*, 1993, **82**, 466.
- 43 A. Leo, C. Hansch and P. Y. Jow, *J. Med. Chem.*, 1976, **19**, 611.
- 44 C. A. Hunter, *Chem. Sci.*, 2013, **4**, 834.
- 45 P. Morgado, A. R. Garcia, L. F. G. Martins, L. M. Ilharco and E. J. M. Filipe, *Langmuir*, 2017, **33**, 11429.
- 46 P. Morgado, H. Rodrigues, F. J. Blas, C. McCabe and E. J. M. Filipe, *Fluid Phase Equilib.*, 2011, **306**, 76.



47 P. Morgado, R. Tomás, H. Zhao, M. C. dos Ramos, F. J. Blas, C. McCabe and E. J. M. Filipe, *J. Phys. Chem. C*, 2007, **111**, 15962.

48 L. Lepori, E. Matteoli, A. Spanedda, C. Duce and M. R. Tiné, *Fluid Phase Equilib.*, 2002, **201**, 119.

49 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31.

50 E. L. Willighagen, J. W. Mayfield, J. Alvarsson, A. Berg, L. Carlsson, N. Jeliazkova, S. Kuhn, T. Pluskal, M. Rojas-Chertó, O. Spjuth, G. Torrance, C. T. Evelo, R. Guha and C. Steinbeck, *J. Cheminf.*, 2017, **9**, 33.

51 K. R. Lawson and J. Lawson, *J. Cheminf.*, 2012, **4**, 3.

