Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2025, **16**, 23129

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

Solvation strategies for free-energy calculations in a halogen-bonded complex: implicit, explicit, and machine learning approaches

Jaroslav Vacek, [abc] Dávid Vrška, [ad] Debashree Manna, *[a] Rabindranath Lo *[a] and Pavel Hobza *[ab]

In pursuit of an efficient solvation approach for the halogen bonded complex between molecular iodine and tetramethylthiourea that reliably reproduces experimental trends, we investigated a range of solvent models, from implicit representations to periodic metadynamics simulations alongside micro-solvation and ONIOM-based methods as robust alternatives. Implicit solvent models fail to describe halogen-bonded complexes in high-polar solvents but provide surprisingly accurate estimates of binding free energies in all low to moderately polar solvents. For accurate and reliable modeling, especially in polar media, explicit solvent representations are essential. Periodic metadynamics simulations typically provide enhanced accuracy in calculating free energy differences, particularly for systems with complex solvation behavior. However, they are computationally demanding and restricted to generalized gradient approximation functionals (GGA). To overcome this limitation and improve accuracy, we employed the machine learning perturbation theory technique, enabling the estimation of free energies at levels of theory beyond the GGA.

Received 19th August 2025
Accepted 24th October 2025

DOI: 10.1039/d5sc06336a

rsc.li/chemical-science

Introduction

The solvent environment plays a critical role in modulating halogen bonding (X-bonding), significantly affecting both the strength and geometry of these noncovalent interactions. Understanding and modelling these solvent effects is crucial, as they impact a wide range of applications,[1–3] including drug design,[4–6] supramolecular self-assembly,[7] protein engineering, catalysis, molecular recognition, and chemical sensing.[8,9]

In aqueous or other polar environments, noncovalent interactions are often attenuated due to competitive solvation by the surrounding solvent molecules. However, it is evident that the effect of the solvent largely depends on the nature of the interactions between the subsystems. When electrostatic interactions are predominant, the dielectric constant of the medium plays a crucial role, as polar solvents tend to screen these interactions, thereby weakening them. Systems dominated by polarization are moderately affected by the solvent, and the solvent effect varies proportionally with the extent of polarization. Charge transfer interactions can be stabilized or destabilized depending on how well the solvent stabilizes the charge-separated resonance forms. Recently, we have shown that the stability of neutral hydrogen-bonded (H-bonded) complexes increases in a more polar solvent, provided there is substantial charge transfer between the subsystems.[10,11] This is against the conventional belief that the stability of H-bonded complexes decreases in a more polar solvent.[12–19] In the case of halogen bonds (XBs), solvent effects do not follow a simple or universal trend. Research on neutral halogen bonded (X-bonded) systems in solution is still limited and often contradictory.[20–26] For example, Carlsson *et al.* suggested that neutral XBs are slightly stabilized with an increase in solvent polarity.[27] Hunter *et al.* reported that the X-bonded complex between molecular iodine (I_2) and tetramethylthiourea (TMTU) is rather insensitive to solvent polarity.[25] Conversely, Fan-Hagenstein *et al.* found that the formation of X-bonded complexes between pyridine and perfluoroalkyl iodides is significantly influenced by the solvent, with polar solvents enhancing the stability of these complexes.[23] Theoretical investigations into neutral XBs in solution also remain relatively sparse.[28–32] For the complexes formed by a series of aromatic XB acceptors and donors, Frontera *et al.* found that the interaction energies are less favourable in solution compared to those in the gas phase.[32] However, several studies by Cabot *et al.* on perfluorohexyl iodide, and by Li *et al.* on BrCl$\cdots H_2CS$ and HOX$\cdots H_2CS$ (where X = F, Cl, and Br), indicate that solvent effects can stabilize XBs.[17,33,34] Both

[a]*Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo náměstí 542/2, 160 00 Prague, Czech Republic. E-mail: debashree.manna@uochb.cas.cz; rabindranath.lo@uochb.cas.cz; pavel.hobza@uochb.cas.cz*

[b]*IT4Innovations, VŠB-Technical University of Ostrava, 17. Listopadu 2172/15, 708 00 Ostrava-Poruba, Czech Republic*

[c]*Faculty of Science, Palacký University Olomouc, 17. Listopadu 1192/12, 779 00 Olomouc, Czech Republic*

[d]*Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University in Bratislava, Mlynská dolina, Ilkovičova 6, 842 15 Bratislava, Slovakia*

Carlsson *et al.* and Taylor *et al.* argued that a purely electrostatic model fails to adequately describe X-bonding behaviour in solution.[22,27]

Given these diverse findings, identifying an appropriate and efficient computational solvent model for X-bonded systems is both timely and necessary. To evaluate computational solvent models, we utilize the extensive experimental data available for the TMTU$\cdots$I$_2$ system investigated by Hunter and co-workers.[25] The paper presents data on two H-bonded and two X-bonded complexes in nine low to moderately polar and six highly polar solvents. The results are, at first glance, surprising since binding free energies of X-bonded complexes measured in various solvents differ by less than 2 kcal mol^{-1} and, furthermore, they are systematically larger in low to moderately polar solvents. This paper serves as a unique source of high-quality binding free energy data for X-bonded complexes across different solvents. The complexity of solvent interactions, particularly in biological contexts, makes the accurate modelling of XB behaviour in solution challenging and underscores the need for quantum mechanical (QM) methods. Implicit solvent models are widely employed across various fields of chemistry and biochemistry due to their computational efficiency and cost-effectiveness. While implicit solvent models[35–37] capture general solvent trends, they oversimplify solute–solvent interactions by treating the solvent as a uniform dielectric continuum.[38] This prevents accurate modelling of competitive binding scenarios, in which solvent molecules interact directly with donor or acceptor sites. As a result, explicit solvation models are often required to account for specific, localized solute–solvent interactions that can alter both the binding strength and geometry.

A rigorous computational modelling is crucial for designing and optimizing molecules that rely on noncovalent interactions in solutions, particularly when applied in medicinal chemistry, enzyme design, or material synthesis.

Computational details

Implicit solvent models

All geometries were initially optimized at the DFT-D[39] level using the PBE0-D3 functional[40–42] with zero damping and the def2-TZVPP basis set.[43,44] These calculations were performed with the Gaussian 16 program package.[45] To account for solvent effects, we employed the COSMO, SMD, and C-PCM continuum solvation models,[35–37] as well as periodic explicit solvent models and a combined explicit–implicit micro-solvation approach. The geometries were further refined using the range-separated hybrid meta-GGA functional ωB97M-V, which includes VV10 nonlocal correlation[46] and is known to deliver reliable results for non-covalent systems. These calculations were carried out with ORCA 6.0.[47,48] The ΔG values were computed using the rigid rotor–harmonic oscillator–ideal gas approximation at the PBE0 level, while thermodynamic corrections at the ωB97M-V level were derived using the Quasi-RRHO approach.[49]

To evaluate the total interaction energies ΔE (eqn (1)), we calculated the energy difference between the bound state (complex) and the separated state (monomers). This corresponds to the binding process TMTU + I_2 → TMTU$\cdots$$I_2$.

$$\Delta E = E_{\text{complex}} - E_{\text{monomer 1}} - E_{\text{monomer 2}} \quad (1)$$

Here, E_{complex}, $E_{\text{monomer 1}}$ and $E_{\text{monomer 2}}$ denote the electronic energies of the optimized structures of the complex, monomer 1 (TMTU), and monomer 2 (I_2), respectively.

Analogously, the binding free energy differences ΔG were obtained according to eqn (2a):

$$\Delta G = G_{\text{complex}} - G_{\text{monomer 1}} - G_{\text{monomer 2}} \quad (2a)$$

The following formula was used for Helmholtz free energy (ΔA):

$$\Delta A = A_{\text{complex}} - A_{\text{monomer 1}} - A_{\text{monomer 2}} \quad (2b)$$

Two different charge models are used to estimate the charge transfer: CT_1, calculated using Hirshfeld charges,[50] and CT_2, calculated using Löwdin charges.[51,52]

Explicit–implicit solvent models

Micro-solvation and ONIOM. For the micro-solvation[53–56] studies, 30 explicit solvent molecules were included in combination with the COSMO implicit solvent model at the PBE0-D3/def2-SVP level of theory. Dielectric constants (ε) of 2.23, 8.93, and 32.61 were used to represent CCl_4, CH_2Cl_2, and CH_3OH, respectively. The critical step represents the size of the solvation shell. Considering the complex, TMTU with iodine is small, as it consists of only ten heavy atoms; however, its solvation shell contains 30 solvent molecules. Evidently, for a larger complex, the number of solvent molecules in the solvation shell dramatically increases, making the respective QM calculations prohibitively expensive. One way to solve the problem is to use the ONIOM (Our own N-layered Integrated molecular Orbital and molecular Mechanics) approach.[57] The method was employed to assess the strength of the X-bonded complex in the presence of up to 50 explicit solvent molecules, treating the complex as the high layer and the solvent molecules as the low layer. The high layer was treated at the PBE0-D3/def2-TZVPP level of theory, while the low layer was described using the PM7 (ref. 58) method.

The binding free energy was computed as the energy difference,

$$G_{\text{binding}}^{\text{ONIOM/MS}} = G_{\text{close}}^{\text{ONIOM/MS}} - G_{\text{far}}^{\text{ONIOM/MS}}.$$

Here, *close* means the bound state, and for *far*, we have considered a distance of 6 Å between the monomers. For the micro-solvation method, the same formula was used for the calculation of binding free energy.

Periodic models

The TMTU$\cdots$$I_2$ complex was simulated with more sophisticated periodic descriptions of solvents. Due to the high computational cost of this approach, the number of solvent molecules

was limited to 24 per system for carbon tetrachloride and dichloromethane solvents. For methanol, 28 explicit solvent molecules were used. Including more solvent molecules would have exceeded our available computational resources.

At the beginning, the TMTU$\cdots$I$_2$ complex solvated with 24 molecules of the solvents, which were placed in a large unit cell (50 Å × 50 Å × 50 Å). Initial geometry optimizations were performed using classical force fields *via* the QuantumATK package.[59,60] Subsequently, the system was pre-equilibrated using force-field molecular dynamics (MD) in the *NpT* ensemble, employing the Martyna–Tobias–Klein (MTK) algorithm[61] keeping the pressure p and the temperature T constant. The number of particles N is constant by design.

The resulting pre-equilibrated structures were further optimized using the Vienna *Ab initio* Simulation Package (VASP),[62–64] followed by *ab initio* MD in the *NpT* ensemble. For this purpose, the Parrinello–Rahman algorithm[65,66] was used. After *NpT* equilibration, a final *ab initio* MD simulation was performed in the *NVT* ensemble – number of particles, volume, and temperature (the Nosé–Hoover thermostat[67–70] was employed) were constant, using the average unit cell volume obtained from the previous *NpT* run. Final cell volumes for each investigated system are summarized in Table S1 in the SI.

All *ab initio* MD and metadynamics simulations were carried out using the VASP at the PBE-D3 level of theory, with the Becke–Johnson damping function.[39,71,72] A plane-wave basis set was employed to describe the valence electron region, while the projector-augmented wave (PAW) method[73,74] was used to treat the core electrons. The ALGO = Fast was employed and the energy cutoff and precision mode parameters were set as follows: ENCUT = 400 eV and PREC = NORMAL. To determine the equilibration period of the MD trajectories (see Table S2 in the SI), we employed the Mann–Kendall tests.[75]

Once equilibrium was reached, the resulting structure was used as the starting point for *ab initio* metadynamics.[76,77] The S–I distance was chosen as the collective variable (CV) for the TMTU$\cdots$I$_2$ system. A very fine bias potential was applied to ensure high precision in the metadynamics simulations: the height of the Gaussian hill was set to 0.0005 eV and the width of the Gaussian hill was 0.005 Å, added after every two steps in a metadynamics run.

Since the canonical (*NVT*) ensemble is used, we work with Helmholtz free energy A instead of Gibbs free energy G. We report the free energy of reaction $\Delta A_{(\mathrm{R}\rightarrow\mathrm{P})}$ (in our case, the binding process of TMTU with the iodine dimer from separated TMTU + I$_2$ (R) to bound TMTU$\cdots$I$_2$ (P) in different solvents) defined in eqn (3):

$$\Delta A_{\mathrm{R}\rightarrow\mathrm{P}} = A_{\xi_{\mathrm{ref,R}}\rightarrow\xi_{\mathrm{ref,P}}} - k_{\mathrm{B}}T\ln\left(\frac{P(\xi_{\mathrm{ref,R}})}{P(\xi_{\mathrm{ref,P}})}\right), \quad (3)$$

where the term $\Delta A_{\xi_{\mathrm{ref,R}}\rightarrow\xi_{\mathrm{ref,P}}}$ is the reversible work needed to shift the CV, ξ, from $\xi_{\mathrm{ref,R}}$ to $\xi_{\mathrm{ref,P}}$. This contribution can be obtained *e.g.* from metadynamics simulations, where the CV (S–I distance) is gradually driven from the reactant ($\xi_{\mathrm{ref,R}}$) to the product ($\xi_{\mathrm{ref,P}}$.) state. The second term, $-k_{\mathrm{B}}T\ln\left(\frac{P(\xi_{\mathrm{ref,R}})}{P(\xi_{\mathrm{ref,P}})}\right)$, represents a correction that accounts for the normalized probabilities of observing the specific reactant $P(\xi_{\mathrm{ref,R}})$ and product $P(\xi_{\mathrm{ref,P}})$ configuration. These probabilities are obtained from molecular dynamics trajectories, for instance using histogram methods. Although the specific values $\xi_{\mathrm{ref,R}}$ and $\xi_{\mathrm{ref,P}}$ can be chosen arbitrarily, in practice we select the maximum value of the corresponding normalized histograms. Here, k_{B} denotes the Boltzmann constant and T is the thermodynamic temperature of simulation (340 K in our case). The statistical errors were determined using the block averaging method[78] with a 95% confidence interval. The overall sum of errors was calculated as follows:

$$\text{Sum of errors} = \sqrt{{s_{\mathrm{bound}}}^2 + {s_{\mathrm{separated}}}^2} \quad (4)$$

where s_{bound} and $s_{\mathrm{separated}}$ represent errors for the bound and separated state, respectively.

Machine learning perturbation theory

For free energy calculations in an improved level of theory, the Machine Learning Perturbation Theory (MLPT)[79–81] technique was employed. This approach connects Δ-ML[82] where a model is trained to predict energy differences of the target level (*i.e.* more accurate and computationally more demanding method), relative to a reference method (*i.e.* computationally low-cost, production method), with thermodynamic perturbation theory (PT).[83]

In our case, the production method is the GGA-type PBE functional supplemented by the Grimme dispersion correction and the Becke–Johnson damping function D3(BJ). As target methods, we selected two DFT functionals from the hybrid rung of Jacob's ladder,[84] both enhanced with D3(BJ) corrections: range-separated HSE06 (ref. 85) and unscreened PBE0.[40–42] For these target functionals in all solvents, we chose a time step value of TIME = 0.7 within the Damped algorithm, and we set the energy cutoff (ENCUT) of 520 eV and "Accurate" precision.

Thus, MD data were generated using the PBE-D3(BJ) functional (for more details, see Section "*Periodic models*") with a Hamiltonian $H(\boldsymbol{p},\boldsymbol{x})$. Using MLPT, simulation at the target levels described by the Hamiltonian $\tilde{H}(\boldsymbol{p},\boldsymbol{x})$ requires only a few dozen single-point calculations (we selected every (N_{data}/100)-th configuration, where N_{data} is the total number of configurations obtained from the production MD run) for training the ML model. The rest of the simulation is predicted based on this training. In this way, MLPT enables simulations that are several orders of magnitude faster without sacrificing accuracy.

Several applications of the MLPT technique in the efficient determination of various properties, such as adsorption enthalpies,[79,81,86] free activation energies[80,87] or fully anharmonic dimerization thermodynamics at challenging target levels, including CCSD(T)[88] have been found.

The formula for the target free energy difference between two stable states $\Delta\tilde{A}_{\mathrm{R}\rightarrow\mathrm{P}}$ is known (see, for instance, ref. 80 and 89):

$$\Delta\tilde{A}_{\mathrm{R}\rightarrow\mathrm{P}} = \Delta A_{\mathrm{R}\rightarrow\mathrm{P}} - k_{\mathrm{B}}T\ln\left\{\frac{\langle\exp[-\Delta V(\boldsymbol{r})/k_{\mathrm{B}}T]\rangle_{\mathrm{P}}}{\langle\exp[-\Delta V(\boldsymbol{r})/k_{\mathrm{B}}T]\rangle_{\mathrm{R}}}\right\}, \quad (5)$$

where the $\Delta A_{R\to P}$ term is the free energy of the reaction obtained using the production method (see eqn (3)). The second term represents ensemble averages over initial $\langle\cdots\rangle_R$ and final $\langle\cdots\rangle_P$ stable states. To evaluate these terms, the second-order cumulant expansion was employed (for details, see the excellent textbook by Chipot and Pohorille[90]), which assumes a Gaussian-shaped function of the potential energy distribution:

$$\left\langle \exp\left[-\frac{\Delta V(\boldsymbol{r})}{k_B T}\right]\right\rangle \approx \exp\left[-\frac{\langle \Delta V(\boldsymbol{r})\rangle}{k_B T}+\frac{\langle \Delta V^2(\boldsymbol{r})\rangle-\langle \Delta V(\boldsymbol{r})\rangle^2}{2(k_B T)^2}\right] \quad (6)$$

Note that the term $\Delta V(\boldsymbol{r})$ in eqn (5) and (6) denotes the difference in potential energy between the target and the production Hamiltonians. More generally, $\Delta H(\boldsymbol{r},\boldsymbol{p})$ represents the Hamiltonian difference, but if the two Hamiltonians differ only in their potential energy contributions, the free energy expression can be written equivalently using $\Delta V(\boldsymbol{r})$ instead of $\Delta H(\boldsymbol{r},\boldsymbol{p})$: $\tilde{H}(\boldsymbol{r},\boldsymbol{p}) - H(\boldsymbol{r},\boldsymbol{p}) = \tilde{V}(\boldsymbol{r}) - V(\boldsymbol{r}) = \Delta V(\boldsymbol{r})$.

The statistical errors were determined using the block averaging method[78] in the same way as in the case of the production method (see eqn (4)).

A kernel ridge regression[91] model trained on a small set of target-level calculations is used to predict energy differences $\Delta V(\boldsymbol{r})$ for the full ensemble, employing the REMatch kernel[92] with SOAP descriptors[93] as implemented in the DScribe[94] and scikit-learn[94,95] libraries, respectively.

Details about minimizing errors in ML predictions are presented in Tables S3 and S4 in the SI.

Results and discussion

Implicit models

Although implicit solvent models do not capture specific solvent–solute interactions as accurately as explicit models, they provide a very reasonable approximation for bulk solvation effects, like electrostatic stabilization, solvation free energy, and dielectric screening. Therefore, we begin by evaluating the performance of the implicit COSMO[35] solvation models against the experimentally observed trends in binding free energies ΔG. These experimental ΔG values are derived from association constants reported by Hunter and co-workers.[25] Table 1 presents the interaction energies (ΔE), Gibbs free energies (ΔG) and Helmholtz free energies (ΔA) at temperature $T = 298$ K, of the X-bonded TMTU$\cdots$I$_2$ complex in different solvents at the PBE0-D3/def2-TZVPP level of theory. The results indicate that the stability of the TMTU$\cdots$I$_2$ complex increases steadily with increasing solvent polarity. However, this trend is supported by experimental data only for low to moderately polar solvents (Table 1 and Fig. 1a). When a highly polar solvent is used, the correlation between polarity and stability of the complex is lost. Thus, an implicit solvent model can only be used for low to moderately polar solvents, where its performance is incredible. The correlation between calculated and experimental ΔG for six non-polar/low-polar solvents with ε between 1.9 and 8.9 is as high as 0.960 (Table 1 and Fig. 1b). A strong correlation was also observed between the experimental and calculated ΔG values and the solvent dielectric constant (Fig. 1b). Nevertheless, this relationship deteriorates markedly in the case of highly polar solvents. To elucidate the underlying reasons for this deviation, a more detailed investigation was undertaken using three representative solvents, each chosen to reflect a distinct range of dielectric properties. Carbon tetrachloride (CCl_4) was selected as a low-polar solvent ($\varepsilon = 2.23$), dichloromethane (CH_2Cl_2) as a moderately polar solvent ($\varepsilon = 8.93$), and methanol (CH_3OH) as a high-polar solvent ($\varepsilon = 32.61$). We started with evaluating the performance of three widely used implicit solvation models (COSMO, SMD, and CPCM) against the experimental ΔG trends, as summarized in Table 2. All implicit solvent models examined yielded consistent results, predicting a steady increase in complex stability with increasing solvent polarity; none captured the trend observed for more polar solvents. Tables 2 and S5 in the SI also list the S$\cdots$I bond lengths (r) and charge transfer (CT) values. To assess the functional dependency, values are also reported with the ωB97M-V functional with the COSMO solvation model in Table S6 in the SI. Tables 2 and S5 in the SI further show a significant charge transfer from I_2 to TMTU upon

Table 1 The interaction energy (ΔE), Gibbs free energy ($\Delta G_{calc.}$) and Helmholtz free energy ($\Delta A_{calc.}$) (in kcal mol^{-1}, $T = 298$ K) of the TMTU$\cdots$I$_2$ complex in various solvent media with the COSMO solvent model calculated at the PBE0-D3/def2-TZVPP level of theory. Experimental association constants[25] (K_a, M^{-1}) and free energies[25] are given for comparison

Solvent	Dielectric constant (ε)	ΔE	$\Delta G_{calc.}/\Delta A_{calc.}$	$\Delta G_{exp.}$	K_a[c]
n-Octane	1.94	−16.94	−6.52/−7.83	−5.38	8800 ± 900[a]
Carbon tetrachloride	2.23	−17.19	−6.74/−8.07	−5.27	7300 ± 600[a]
Toluene	2.38	−17.31	−6.84/−8.17	−5.51	11 000 ± 1600[a]
Chloroform	4.81	−18.50	−7.88/−9.25	−5.86	20 000 ± 1000[a]
1,1,2,2-Tetrachloroethane	8.4	−19.27	−8.62/−10.00	−6.46	55 000 ± 15 000[a]
Dichloromethane	8.93	−19.34	−8.68/−10.07	−6.50	58 000 ± 10 000[b]
Acetone	20.7	−20.02	−9.35/−10.74	−4.47	1900 ± 300[b]
Methanol	32.61	−20.24	−9.58/−10.97	−4.68	2700 ± 200[b]
Acetonitrile	37.5	−20.27	−9.61/−11.00	−4.70	2800 ± 70[b]
Water	78.4	−20.47	−9.81/−11.21	—	—

[a] From ref. 25, measured by UV/vis. [b] From ref. 25, measured by ^{1}H NMR titration. [c] From ref. 25, the errors were determined as 2× the standard deviation from multiple titration repeats.

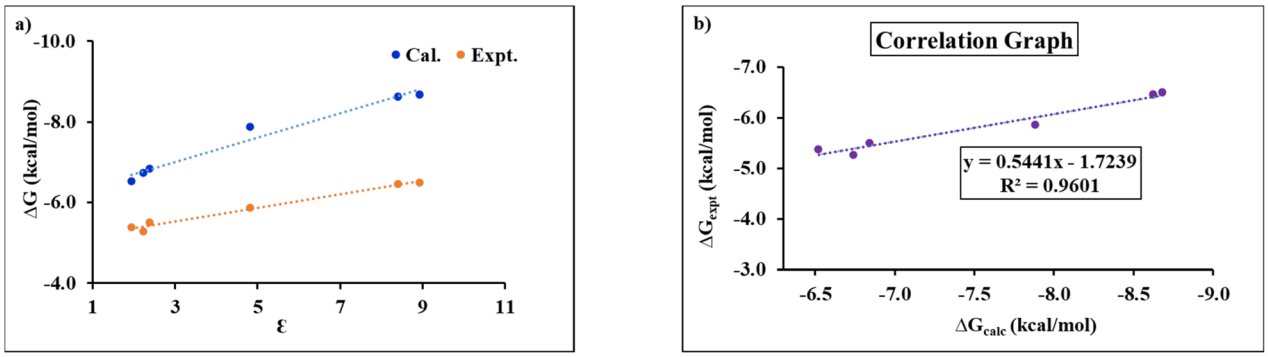

Fig. 1 Plots of (a) ΔG (kcal mol^{-1}) *vs.* ε from experimental and calculated free energy values and (b) $\Delta G_{exp.}$ (kcal mol^{-1}) *vs.* $\Delta G_{calc.}$ (kcal mol^{-1}).

Table 2 The interaction energy, ΔE (kcal mol^{-1}), Gibbs free energy, ΔG, Helmholtz free energy, ΔA (kcal mol^{-1}, $T = 298$ K), halogen bond length, r (Å), and charge transfer (CT_1, calculated from Hirshfeld charges; CT_2, calculated from Löwdin charges) of the TMTU$\cdots$I$_2$ complex in various solvents calculated with implicit solvent models at the PBE0-D3/def2-TZVPP level of theory. Experimental free energies[25] are given for comparison

	COSMO				SMD				CPCM				Exp.
Solvent	ΔE	$\Delta G/\Delta A$	r	CT_1/CT_2	ΔE	$\Delta G/\Delta A$	r	CT_1/CT_2	ΔE	$\Delta G/\Delta A$	r	CT_1/CT_2	$\Delta G_{exp.}$[a]
Carbon tetrachloride	−17.19	−6.74/ −8.07	2.724	−0.408/ −0.376	−17.86	−7.55/ −8.81	2.731	−0.416/ −0.383	−17.08	−6.79/ −8.08	2.735	−0.413/ −0.381	−5.27
Dichloromethane	−19.34	−8.68/ −10.07	2.610	−0.549/ −0.510	−20.84	−11.05/ −12.28	2.631	−0.545/ −0.507	−18.84	−8.43/ −9.76	2.637	−0.536/ −0.499	−6.50
Methanol	−20.24	−9.58/ −10.97	2.574	−0.607/ −0.564	−21.98	−13.02/ −14.26	2.599	−0.593/ −0.552	−19.52	−9.18/ −10.52	2.606	−0.580/ −0.540	−4.68

[a] From ref. 25.

complex formation. Additionally, the S$\cdots$I bond length decreases systematically with increasing solvent polarity, while the corresponding charge transfer (CT) values increase substantially. All the charge models (Hirshfeld, Löwdin, NBO and CM5) consistently indicate that the CT values for this complex are indeed significant. The strong charge-transfer nature of the interaction is further supported by the UV/Vis spectrum of the TMTU$\cdots$I$_2$ complex, which displays a distinct charge-transfer band in the 330–340 nm range.[25] Previously, we have shown that when the charge transfer (CT) between the subsystems is significant, it leads to stabilization of the X-bonded complex with increasing solvent polarity.[96] However, when the CT is low, the opposite trend may be observed. Since the CT is significant for the TMTU$\cdots$I$_2$ complex, its stabilization with increasing solvent polarity is consistent with our expectations, provided there are no substantial specific solvent–solute interactions.

In contrast, for the tetramethylurea$\cdots$I$_2$ complex, another system studied by Hunter and co-workers,[25] both experimental and computational results show destabilization with increasing solvent polarity. We have demonstrated that the CT values for this complex are significantly lower, which in turn leads to this opposite trend (Table S7, SI).

Explicit–implicit solvent models

Micro-solvation and ONIOM. The continuum solvent models characterized solely by dielectric constants have some limitations – particularly when significant specific interactions occur between the solvent molecules and the complex. These limitations can be addressed through natural refinement involving a more detailed representation of the first solvation shell. This is typically achieved by explicitly including a few solvent molecules around the complex, embedded within an implicit solvent environment – a method commonly referred to as the cluster–continuum approach or micro-solvation model.[53–56] The inclusion of the implicit environment in the micro-solvation framework allows for the necessary treatment of long-range electrostatic interactions. It is essential to have enough number of explicit solvent molecules to ensure accurate representation of the complex within the micro-solvation model for both bound and separated states.

Thus, calculations were carried out using 30 explicit solvent molecules, each embedded in a continuous solvent environment described by the COSMO implicit solvent model. The incorporation of explicit solvent molecules significantly destabilizes the TMTU$\cdots$I$_2$ complex in CH_3OH, aligning well with experimental observations (Table 3). In contrast, a slight stabilization is observed when moving from CCl_4 to CH_2Cl_2. This is again in line with the experimental trend. The differences in the S$\cdots$I bond length, with or without explicit solvation, are negligible (Table 3). We extended the study by using four more low-energy conformers from MD simulation trajectories. We used these as starting points to construct

Table 3 The Gibbs free energy (ΔG), Helmholtz free energy (ΔA) (in kcal mol^{-1}, $T = 298$ K), and halogen bond length, r (Å) of the TMTU$\cdots$I$_2$ complex in the micro-solvation approach with 30 explicit solvent molecules in the COSMO model calculated at the PBE0-D3/def2-SVP level of theory and in the ONIOM approach with 50 explicit solvent molecules calculated at the PBE0-D3/def2-TZVPP:PM7 level of theory. Experimental free energies[25] are given for comparison

	Micro-solvation		ONIOM		Exp.
Solvent	$\Delta G/\Delta A$	r	$\Delta G/\Delta A$	r	$\Delta G_{exp.}$[a]
Carbon tetrachloride	−16.10/−16.63	2.691	−12.92/−13.39	2.867	−5.27
Dichloromethane	−16.82/−17.08	2.641	−13.69/−14.18	2.461	−6.50
Methanol	−7.01/−7.44	2.587	−9.92/−10.48	2.494	−4.68

[a] From ref. 25.

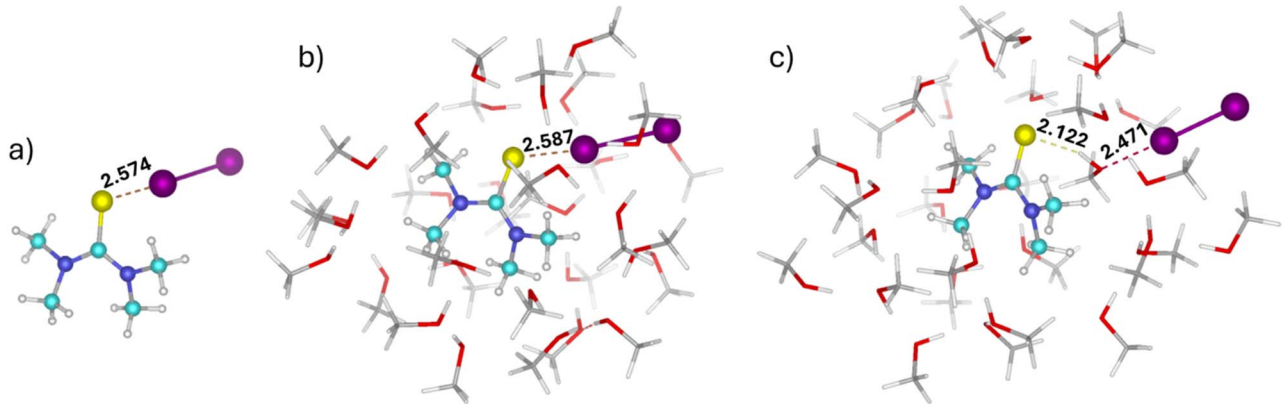

Fig. 2 The optimized geometries of the TMTU$\cdots$I$_2$ complex: (a) with the implicit COSMO solvent model in methanol medium, (b) in the complex form with 30 explicit methanol molecules included in the COSMO model, and (c) as separated systems, also with 30 explicit methanol molecules included in the COSMO model. The bond lengths are in Å [C: cyan (TMTU); grey (solvent), S: yellow, I: violet, H: white, N: blue, O: red].

complexes with 30 solvent molecules of CCl_4, CH_2Cl_2, and CH_3OH. Table S8 in the SI shows the free energy order for all the conformers, and Table 3 lists the ΔG values for the most stable conformer with 30 solvent molecules. Destabilization occurs for the TMTU$\cdots$I$_2$ complex in CH_3OH, as the separated systems are stabilized by both conventional hydrogen bonding and bridging hydrogen bonding with CH_3OH (Fig. 2c). In particular, C=S$\cdots$H and O$\cdots$I–I types of hydrogen and halogen bonds formed through the bridging CH_3OH molecule contribute to the stabilization of the separated systems. Please note that these calculations are highly computationally demanding; thus, the number of solvent molecules included in the micro-solvation model is restricted by computational cost. Additionally, the ONIOM approach was employed to evaluate the strength of the X-bonded complex in the presence of 50 explicit solvent molecules, treating the complex as the high layer (PBE0-D3/def2-TZVPP) and the solvent molecules as the low layer (PM7). For each solvent, five different conformers are considered in the ONIOM calculations (Table S9, SI), and the Gibbs free energy (ΔG) is reported for the most stable conformation (Table 3). ONIOM calculations also reproduce the experimental trends (Table 3). However, both micro-solvation and ONIOM significantly overestimate the experimental ΔG values.

Calculations of free energies of dimerization of TMTU + I_2 in explicitly defined solvents for temperature $T = 340$ K in periodic models

To increase computational accuracy, we investigated the interactions between TMTU and the iodine molecule in a periodic model. Periodic simulations are generally expected to provide improved accuracy compared to droplet models, particularly for systems exhibiting complex solvation behaviour. The periodic boundary model offers several advantages over the droplet model approach. For example, it eliminates the risk of molecules "boiling off" during a MD run. Moreover, unlike the droplet model, the periodic approach has no surface, thereby avoiding artificial interactions between the system and vacuum that could distort the results. In summary, the periodic approach enables long MD simulations while preserving stability and ensuring reliable results.

Metadynamics simulations of the TMTU$\cdots$I$_2$ molecule (Fig. 3) were performed in three different solvents: 24 explicitly defined molecules for two low to moderately polar solvents (CCl_4 and CH_2Cl_2) and 28 explicit solvent molecules for one high-polar solvent (CH_3OH), under periodic boundary conditions. We initiated the simulations from the more stable (bound) states and introduced a bias potential constructed in the space of the CV (ξ) chosen as the distance between the

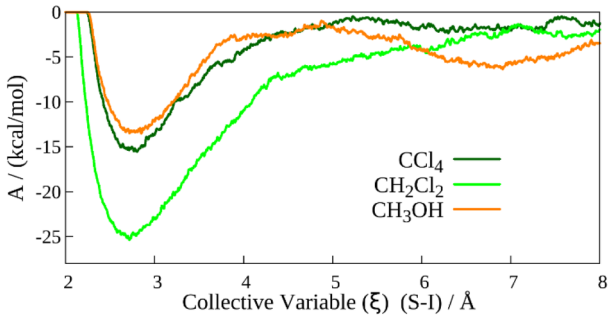

Fig. 3 The bias potential constructed in the space of the S–I distance (chosen CV) obtained from *ab initio* (PBE-D3(BJ)) metadynamics in three solvents: CCl_4, CH_2Cl_2 and CH_3OH. The temperature of metadynamics simulations is $T = 340$ K.

Table 4 The PBE-D3(BJ) free energy differences of the TMTU$\cdots I_2$ system in the three solvents as defined by eqn (3). Experimental free energies[25] are given for comparison

Solvents	$\Delta A_{(R\rightarrow P)}$/kcal mol^{-1}	$\Delta G_{exp.}$[a]
Carbon tetrachloride	−11.93 ± 1.09	−5.27
Dichloromethane	−20.82 ± 1.31	−6.50
Methanol	−9.59 ± 1.32	−4.68

[a] From ref. 25.

sulphur and iodine atoms (Fig. 3). This allowed us to reach a second stable state corresponding to the separated states. It should be noted that although we considered a binding reaction (*i.e.* from the separated state to a bound state), the direction of the CV change is irrelevant in metadynamics.

After taking into account the probabilities of the occurrence of the concrete ξ values obtained from MD trajectories of both states, we were able to determine the free energy differences as functions of CV, $\Delta A(\xi)$, according to eqn (3). This methodology successfully reproduces the experimentally observed stability trends, with the corresponding values summarized in Table 4.

While the approach captures the correct qualitative trend, it tends to significantly overestimate both absolute (ΔA) and relative ($\Delta\Delta A$) free energy differences compared with the experimental reference, similar to micro-solvation and ONIOM-based methods. Moreover, the periodic metadynamics method is computationally more expensive than micro-solvation or ONIOM approaches and can accommodate comparatively fewer solvent molecules. Due to computational cost, we restricted this approach to the PBE-D3(BJ) level of theory.

Free energy finite-T calculations beyond GGA functionals

MLPT. Table 5 presents the target-level free energy differences for the binding process of TMTU + I_2 in three different solvents. When compared with experimental data (last column in Table 5), both HSE06-D3(BJ) and PBE0-D3(BJ) functionals reproduce the experimental trend: in absolute values, ΔA is the smallest for CH_3OH, followed by CCl_4, and the largest for CH_2Cl_2.

Table 5 The MLPT-assisted high-level DFT free energy differences ($\Delta\tilde{A}_{method}$) (in kcal mol^{-1}) for the binding process of TMTU–I_2 in different solvents. The temperature is 340 K. Values in bold indicate the DFT results closest to the experimental reference ΔG_{exp}. Experimental free energies[25] are given for comparison

Solvent	$\Delta\tilde{A}_{HSE06\text{-}D3(BJ)}$	$\Delta\tilde{A}_{PBE0\text{-}D3(BJ)}$	$\Delta G_{exp.}$[a]
Carbon tetrachloride	**−10.34 ±1.35**	−13.98 ± 1.28	−5.27
Dichloromethane	−19.38 ± 1.37	**−18.66 ±1.37**	−6.50
Methanol	−7.53 ± 1.09	**−7.14 ±1.09**	−4.68

[a] From ref. 25.

The MLPT-corrected free energy differences, $\Delta\tilde{A}_{HSE06\text{-}D3(BJ)}$, yield values closer to the experimental references compared to the $\Delta A_{PBE\text{-}D3(BJ)}$ baseline (see Table 4) across all three solvents. By contrast, the ML-correction with the PBE0-D3(BJ) functional, $\Delta\tilde{A}_{PBE0\text{-}D3(BJ)}$, exhibits a larger deviation from the experimental reference in CCl_4 than the corresponding baseline value.

This inverse trend arises because the difference between the target and production methods, calculated *via* the cumulant expansion (CE), is evaluated separately for the reactant (separated state) and the product (bound state). When the CE value (*i.e.*, the energy difference between target and production methods) is more negative for the reactant than for the product, the resulting correction – defined as the difference between these two CE values (see the second term of eqn (5)) – is positive. Adding this positive correction to the (already negative) $\Delta A_{PBE\text{-}D3(BJ)}$ reduces the absolute value of the free energy. This scenario applies to CH_3OH and CH_2Cl_2 for both the functionals, as well as to CCl_4 with HSE06-D3(BJ). Conversely, when the CE value is lower for the product, as observed with PBE0-D3(BJ) in CCl_4, the correction is negative, making the overall free energy even more negative. We attribute this to an artifact arising from the combination of the unscreened functionals and the solvent, possibly due to a preference of a (significantly) different density, *i.e.*, a different unit cell volume for the reactant and/or product.

Apart from this outlier, all other ML-corrected values are closer to the experimental reference than the PBE-D3(BJ) baseline. The closest agreements are observed as follows: in CCl_4 with HSE06-D3(BJ), the error is 5.07 ± 1.35 kcal mol^{-1}; in CH_2Cl_2 with PBE0-D3(BJ), 12.16 ± 1.37 kcal mol^{-1}; and in CH_3OH with PBE0-D3(BJ), 2.46 ± 1.09 kcal mol^{-1}. These values are highlighted in bold in Table 5.

Relative free energy differences between different solvents ($\Delta\Delta A$) are overestimated compared with experimental data, similar to production PBE-D3(BJ) level results. Among the functionals, the $\Delta\Delta A$ differences between the two low to moderately polar solvents are as follows: $\Delta\Delta A_{PBE0\text{-}D3(BJ)}$ = 4.68 kcal mol^{-1} and $\Delta\Delta A_{HSE06\text{-}D3(BJ)}$ = 9.04 kcal mol^{-1}. The differences between the high-polar CH_3OH and moderately polar CH_2Cl_2 remain nearly constant, 11.52 kcal mol^{-1} for PBE0-D3(BJ) and 11.85 kcal mol^{-1} for HSE06-D3(BJ). For CH_3OH and the low-polarity solvent, CCl_4, the spread is slightly wider: 2.81 kcal mol^{-1} with HSE06-D3(BJ) and 6.84 kcal mol^{-1} with PBE0-D3(BJ).

Critical evaluation of different techniques for estimation of binding free energies of X-bonded complexes in a solvent provides the following take:

(i) Continuous solvent cannot be used for both apolar and polar solvents. The model, however, provides very good estimates of ΔG for low to moderately polar solvents. COSMO and CPCM provide not only qualitative but even quantitative estimates of ΔG.

(ii) The micro-solvation and ONIOM models, contrary to the continuous models, provide reasonable qualitative agreement of ΔG for all three solvents considered, though their absolute values are strongly overestimated.

(iii) Theoretically justified periodic calculations performed at the PBE level also provide qualitative estimates of ΔG for all solvents considered, but their quantitative values are overestimated even more than in the previous case.

(iv) Upon considering the ML-corrected values, the situation is changed only slightly. HSE06 and PBE0 methods qualitatively agree with experimental values, though, again, the respective absolute values are strongly overestimated.

Conclusions

Solvent effects play a critical role in modulating the XB between molecular iodine and tetramethylthiourea in solution and their theoretical estimates are tedious and represent a challenge to current computational chemistry. The complex exhibits significant charge-transfer character and are notably stabilized in low-polarity environments (dielectric constant $\varepsilon \leq 10$) with moderate increases in solvent polarity. However, in highly polar solvents, the XB weakens due to the stabilization of charge-separated species *via* competitive solvation of the individual components. While implicit solvent models are valuable for identifying general trends and for initial screening of molecular systems, they inherently oversimplify solute–solvent interactions by neglecting specific, directional, and dynamic solvent effects. The method cannot be used for high-polar solvents, but for low to moderately polar solvents, it provides good estimates of binding free energies. To address these shortcomings, explicit solvent models are essential. By providing a detailed, dynamic representation of solvation, explicit models enable more reliable predictions of free energy trends in solution.

By systematically comparing implicit and explicit solvent models, incorporating periodic boundary conditions to better capture realistic solvation environments, and applying machine learning perturbation theory to balance computational cost and accuracy, we provide a robust framework for evaluating solvent effects. Although the current error margins (2.5–12 kcal mol^{-1}) are too large for reliable predictive applications in catalysis, materials design, or molecular recognition, the approach effectively captures the trends observed in experimental association constants across multiple solvents. More broadly, this study contributes to the development of general strategies for modeling noncovalent interactions in complex solvent environments, which is essential for advancing both fundamental chemistry and practical applications. We also acknowledge that achieving more quantitative agreement is necessary for the direct predictive use of the free energy data.

Author contributions

D. M. and R. L. evaluated the implicit solvent models and performed micro-solvation and ONIOM calculations. J. V. performed the metadynamics simulations. D. V. calculated the free energies using machine learning perturbation theory. D. M., D. V., R. L., J. V. and P. H. prepared the manuscript. P. H. coordinated the study.

Conflicts of interest

There are no conflicts to declare.

Data availability

Additional data are available at ZENODO (**https://doi.org/10.5281/zenodo.17452749**).

Tables S1–S9, figure, and supplementary data associated with this article are provided in the supplementary information (SI). See DOI: **https://doi.org/10.1039/d5sc06336a**.

Acknowledgements

The authors would like to thank Associate Professor Tomáš Bučko for fruitful discussions in the field of free energy calculations, which greatly helped them. This work was carried out with the financial support of the European Union under the REFRESH – Research Excellence for Region Sustainability and High-tech Industries project number CZ.10.03.01/00/22_003/0000048 *via* the Operational Programme Just Transition (P. H.). J. V. acknowledges the financial support of Palacký University through the Internal Grant Association, project IGA_PrF_2025_003.

Notes and references

1 *Halogen bonding I: Impact on materials chemistry and life sciences*, ed. P. Metrangolo and G. Resnati, Springer, Switzerland, 2015.
2 *Halogen Bonding, Fundamentals and Applications*, ed. P. Metrangolo and G. Resnati, Springer-Verlag, Berlin, Heidelberg, 2008, **https://link.springer.com/book/10.1007/978-3-540-74330-9**.
3 P. Politzer, P. Lane, M. C. Concha, Y. Ma and J. S. Murray, *J. Mol. Model.*, 2007, **13**, 305–311.
4 C. Bissantz, B. Kuhn and M. Stahl, *J. Med. Chem.*, 2010, **53**, 5061–5084.
5 Y. Lu, Y. Liu, Z. Xu, H. Li, H. Liu and W. Zhu, *Expert Opin. Drug Discovery*, 2012, **7**, 375–383.
6 M. R. Scholfield, C. M. V. Zanden, M. Carter and P. S. Ho, *Protein Sci.*, 2013, **22**(2), 139–152.
7 L. C. Gilday, S. W. Robinson, T. A. Barendt, M. J. Langton, B. R. Mullaney and P. D. Beer, *Chem. Rev.*, 2015, **115**, 7118–7195.

8 G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati and G. Terraneo, *Chem. Rev.*, 2016, **116**, 2478–2601.
9 B. Li, S.-Q. Zang, L.-Y. Wang and T. C. W. Mak, *Coord. Chem. Rev.*, 2016, **308**, 1–21.
10 D. Manna, R. Lo, J. Vacek, V. M. Miriyala, P. Bouř, T. Wu, Z. Osifová, D. Nachtigallová, M. Dračínský and P. Hobza, *Angew. Chem., Int. Ed.*, 2024, **63**, e202403218.
11 R. Lo, D. Manna, J. Vacek, P. Bouř, T. Wu, Z. Osifová, O. Socha, M. Dračínský and P. Hobza, *Angew. Chem., Int. Ed.*, 2025, **64**, e202422594.
12 M. D. Driver, M. J. Williamson, J. L. Cook and C. A. Hunter, *Chem. Sci.*, 2020, **11**, 4456–4466.
13 J. L. Cook, C. A. Hunter, C. M. R. Low, A. Perez-Velasco and J. G. Vinter, *Angew. Chem., Int. Ed.*, 2007, **46**, 3706–3709.
14 A. J. A. Aquino, D. Tunega, G. Haberhauer, M. H. Gerzabek and H. Lischka, *J. Phys. Chem. A*, 2002, **106**, 1862–1871.
15 C. C. Robertson, J. S. Wright, E. J. Carrington, R. N. Perutz, C. A. Hunter and L. Brammer, *Chem. Sci.*, 2017, **8**, 5392–5398.
16 C. A. Hunter, *Angew. Chem., Int. Ed.*, 2004, **43**, 5310–5324.
17 R. Cabot and C. A. Hunter, *Chem. Commun.*, 2009, 2005–2007.
18 N. Y. Meredith, S. Borsley, I. V. Smolyar, G. S. Nichol, C. M. Baker, K. B. Ling and S. L. Cockroft, *Angew. Chem., Int. Ed.*, 2022, **61**, e202206604.
19 R. J. Burns, I. K. Mati, K. B. Muchowska, C. Adam and S. L. Cockroft, *Angew. Chem., Int. Ed.*, 2020, **59**, 16717–16724.
20 M. Erdélyi, *Chem. Soc. Rev.*, 2012, **41**, 3547.
21 P. Klaeboe, *J. Am. Chem. Soc.*, 1967, **89**, 3667–3676.
22 M. G. Sarwar, B. Dragisic, L. J. Salsberg, C. Gouliaras and M. S. Taylor, *J. Am. Chem. Soc.*, 2010, **132**, 1646–1653.
23 B. Hawthorne, H. Fan-Hagenstein, E. Wood, J. Smith and T. Hanks, *Int. J. Spectrosc.*, 2013, **2013**, 1–10.
24 O. Dumele, D. Wu, N. Trapp, N. Goroff and F. Diederich, *Org. Lett.*, 2014, **16**, 4722–4725.
25 C. C. Robertson, R. N. Perutz, L. Brammer and C. A. Hunter, *Chem. Sci.*, 2014, **5**, 4179–4183.
26 J. Cao, X. Yan, W. He, X. Li, Z. Li, Y. Mo, M. Liu and Y.-B. Jiang, *J. Am. Chem. Soc.*, 2017, **139**, 6605–6610.
27 A.-C. C. Carlsson, A. X. Veiga and M. Erdélyi, Halogen Bonding in Solution, in *Halogen Bonding II*, ed. P. Metrangolo and G. Resnati, Topics in Current Chemistry, Springer, Cham, 2014, vol. 359, pp. 49–76.
28 Y. Lu, H. Li, X. Zhu, W. Zhu and H. Liu, *J. Phys. Chem. A*, 2011, **115**, 4467–4475.
29 Y. Lu, H. Li, X. Zhu, H. Liu and W. Zhu, *Int. J. Quantum Chem.*, 2012, **112**, 1421–1430.
30 A. Forni, S. Rendine, S. Pieraccini and M. Sironi, *J. Mol. Graphics Modell.*, 2012, **38**, 31–39.
31 K. E. Riley and K. M. Merz, *J. Phys. Chem. A*, 2007, **111**, 1688–1694.
32 A. Bauzá, D. Quiñonero, A. Frontera and P. M. Deyà, *Phys. Chem. Chem. Phys.*, 2011, **13**, 20371.
33 Q. Li, R. Li, Z. Zhou, W. Li and J. Cheng, *J. Chem. Phys.*, 2012, **136**, 014302.
34 Q.-Z. Li, B. Jing, R. Li, Z.-B. Liu, W.-Z. Li, F. Luan, J.-B. Cheng, B.-A. Gong and J.-Z. Sun, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2266.
35 A. Klamt and G. Schüürmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, **5**, 799–805.
36 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
37 M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comput. Chem.*, 2003, **24**, 669–681.
38 J. Zhang, H. Zhang, T. Wu and Q. Wang, *J. Chem. Theory Comput.*, 2017, **13**, 1034–1043.
39 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
40 J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.
41 M. Ernzerhof and G. E. Scuseria, *J. Chem. Phys.*, 1999, **110**, 5029–5036.
42 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
43 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297.
44 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057.
45 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Revision C.01*, Gaussian Inc., Wallingford CT, 2016.
46 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2016, **144**, 214110.
47 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 73–78.
48 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2025, **15**, e70019.
49 S. Grimme, *Chem.–Eur. J.*, 2012, **18**, 9955–9964.
50 F. L. Hirshfeld, *Theor. Chim. Acta*, 1977, **44**, 129–138.
51 P.-O. Löwdin, *J. Chem. Phys.*, 1950, **18**, 365–375.
52 P.-O. Löwdin, On the Nonorthogonality Problem, in *Advances in Quantum Chemistry*, Elsevier, 1970, vol. 5, pp. 185–199.
53 R. B. Sunoj and M. Anand, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12715–12736.
54 C. Hadad, E. Florez, N. Acelas, G. Merino and A. Restreppo, *Int. J. Quantum Chem.*, 2019, **119**, e25766.

55 G. N. Simm, P. L. Türtscher and M. Reiher, *J. Comput. Chem.*, 2020, **41**, 1144–1145.
56 Y. Basdogan, M. C. Groenenboom, E. Henderson, S. De, S. B. Rempe and J. A. Keith, *J. Chem. Theory Comput.*, 2020, **16**, 633–642.
57 M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber and K. Morokuma, *J. Phys. Chem.*, 1996, **100**, 19357–19363.
58 J. J. P. Stewart, *J. Mol. Model.*, 2013, **19**, 1–32.
59 J. Schneider, J. Hamaekers, S. T. Chill, S. Smidstrup, J. Bulin, R. Thesen, A. Blom and K. Stokbro, *Model. Simul. Mater. Sci. Eng.*, 2017, **25**, 085007.
60 S. Smidstrup, T. Markussen, P. Vancraeyveld, J. Wellendorff, J. Schneider, T. Gunst, B. Verstichel, D. Stradi, P. A. Khomyakov, U. G. Vej-Hansen, M.-E. Lee, S. T. Chill, F. Rasmussen, G. Penazzi, F. Corsetti, A. Ojanperä, K. Jensen, M. L. N. Palsgaard, U. Martinez, A. Blom, M. Brandbyge and K. Stokbro, *J. Phys.: Condens. Matter*, 2020, **32**, 015901.
61 G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177–4189.
62 G. Kresse and J. Hafner, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1993, **47**, 558–561.
63 G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15–50.
64 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 11169–11186.
65 M. Parrinello and A. Rahman, *Phys. Rev. Lett.*, 1980, **45**, 1196–1199.
66 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
67 S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.
68 S. Nosé, *Prog. Theor. Phys. Suppl.*, 1991, **103**, 1–46.
69 W. G. Hoover, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1985, **31**, 1695–1697.
70 D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Computational science series, Academic Press, San Diego, 2nd edn, 2002.
71 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
72 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
73 P. E. Blöchl, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **50**, 17953–17979.
74 G. Kresse and D. Joubert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 1758–1775.
75 S. K. Schiferl and D. C. Wallace, *J. Chem. Phys.*, 1985, **83**, 5203–5209.
76 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12562–12566.
77 M. Iannuzzi, A. Laio and M. Parrinello, *Phys. Rev. Lett.*, 2003, **90**, 238302.
78 H. Flyvbjerg and H. Petersen, *J. Chem. Phys.*, 1989, **91**, 461–466.
79 B. Chehaibou, M. Badawi, T. Bučko, T. Bazhirov and D. Rocca, *J. Chem. Theory Comput.*, 2019, **15**, 6333–6342.
80 T. Bučko, M. Gešvandtnerová and D. Rocca, *J. Chem. Theory Comput.*, 2020, **16**, 6049–6060.
81 B. Herzog, M. Chagas da Silva, B. Casier, M. Badawi, F. Pascale, T. Bučko, S. Lebègue and D. Rocca, *J. Chem. Theory Comput.*, 2022, **18**, 1382.
82 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. von Lilienfeld, *J. Chem. Theory Comput.*, 2015, **11**, 2087–2096.
83 A. Pohorille, C. Jarzynski and C. Chipot, *J. Phys. Chem. B*, 2010, **114**, 10235–10253.
84 J. P. Perdew and K. Schmidt, *AIP Conf. Proc.*, 2001, **577**, 1.
85 A. V. Krukau, O. A. Vydrov, A. F. Izmaylov and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 224106–224111.
86 B. Herzog, A. Gallo, F. Hummel, M. Badawi, T. Bučko, S. Lebègue, A. Grüneis and D. Rocca, *npj Comput. Mater.*, 2024, **10**, 68.
87 J. Rey, A. Gomez, P. Raybaud, C. Chizallet and T. Bučko, *J. Catal.*, 2019, **373**, 361–373.
88 D. Vrška, M. Pitoňák and T. Bučko, *J. Chem. Phys.*, 2024, **160**, 174106.
89 M. Gešvandtnerová, D. Rocca and T. Bučko, *J. Catal.*, 2021, **396**, 166.
90 C. Chipot and A. Pohorille, Calculating free energy differences using perturbation theory, in *Free Energy Calculations Theory and Applications in Chemistry and Biology*, Springer, Berlin, Heidelberg, 2007.
91 M. Rupp, *Int. J. Quantum Chem.*, 2015, **115**, 1058–1073.
92 S. De, A. P. Bartók, G. Csányi and M. Ceriotti, *Phys. Chem. Chem. Phys.*, 2016, **18**, 13754.
93 A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 184115.
94 L. Himanen, M. O. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke and A. S. Foster, *Comput. Phys. Commun.*, 2020, **247**, 106949.
95 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
96 R. Lo, A. Mašínová, M. Lamanec, D. Nachtigallová and P. Hobza, *J. Comput. Chem.*, 2023, **44**, 329.