Supplementary Information (SI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2025

Supplementary material for the paper:

Solvation Strategies for Free-Energy Calculations in Halogen-Bonded Complex: Implicit, Explicit, and Machine Learning Approaches

Jaroslav Vacek,[a,b,c] Dávid Vrška,[a,d] Debashree Manna,*[a] Rabindranath Lo,*[a] Pavel Hobza*[a,b]

[a.] *Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo náměstí 542/2, 160 00 Prague, Czech Republic. Email: 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.*

Table S1 Cell volumes (a × b × c) used in molecular dynamics simulations for each investigated solvent in both bound and separated states. Values are given in Ångströms and rounded to two decimal places.

Solvent	State	a	b	c
Carbon tetrachloride	bound	15.22, 0.00, 0.00	–0.11, 15.15, 0.00	–0.06, –0.03, 15.20
Carbon tetrachloride	separated	15.93, –1.33, –1.15	–0.12, 16.47, 1.01	–0.07, –0.03, 16.37
Dichloromethane	bound	13.96, 0.00, 0.00	0.22, 14.75, 0.00	–0.24, –2.08, 14.08
Dichloromethane	separated	13.32, –0.12, –2.08	0.23, 14.18, –0.99	–0.25, –2.21, 14.96
Methanol	bound	12.37, –1.14, 0.94	–0.05, 12.12, –2.41	–0.11, –0.25, 13.22
Methanol	separated	12.83, –1.30, 0.33	–1.32, 11.95, –0.19	0.90, –2.74, 12.89

Table S2 Number of configurations (N_{data}) generated by molecular dynamics using the PBE-D3(BJ) functional, and equilibration period (τ) for each investigated solvent, for both bound and separated (in parentheses) states as determined via Mann-Kendall tests.

Solvent	N_{data}	τ
Carbon tetrachloride	60000 (86436)	5000 (35000)
Dichloromethane	80000 (60000)	30000 (10000)
Methanol	80000 (80000)	0 (0)

To minimize errors in ML predictions, we optimized two ML hyperparameters (HPs), namely σ in Gaussians representing the atomic density ρ centered at positions x_j in the atomic environment

$$\rho_{\chi_i^A}(r) = \sum_{j \in \chi_i^A} \exp\left\{-\frac{(r - x_j)^2}{2\sigma^2}\right\} \quad \text{(Eq. S1)}$$

and the regularization parameter λ used in kernel ridge regression:

$$w = (K_{train} + \lambda 1)^{-1} y_{train} \quad \text{(Eq. S2)}$$

$$y_{predict} = K_{predict} w, \quad \text{(Eq. S3)}$$

where $\boldsymbol{y_{train}}$ and $\boldsymbol{y_{predict}}$ are the vectors of labels of configurations for the training and prediction sets, $\boldsymbol{K_{train}}$ and $\boldsymbol{K_{predict}}$ denote the kernel matrices for the training and predicted configurations, $\boldsymbol{1}$ is the identity matrix and $\boldsymbol{w}$ is the vector of weights used to make predictions of the target labels $\boldsymbol{y_{predict}}$. HP optimization was performed separately for the R and P configurations, for each method and each solvent – multiple combinations of σ and λ were tested for each model independently. To search for the optimal HPs, we employed a leave-one-out cross-validation procedure.[S1]

Table S3 Hyperparameters σ and λ yielding optimal RMSE errors (see Table S4), obtained using the leave-one-out cross-validation technique for each investigated target functional in each solvent, for both bound and separated (in parentheses) states.

method	HSE06-D3(BJ)		PBE0-D3(BJ)	
solvent	σ [Å]	λ [eV]	σ [Å]	λ [eV]
Carbon tetrachloride	0.04 (0.0005)	$1e^{-13}$ ($5e^{-8}$)	0.015 (0.005)	$1e^{-13}$ ($1e^{-13}$)
Dichloromethane	0.005 (0.005)	$1e^{-13}$ ($1e^{-13}$)	0.005 (0.005)	$1e^{-13}$ ($1e^{-13}$)
Methanol	0.015 (0.001)	$1e^{-13}$ ($1e^{-13}$)	0.015 (0.001)	$1e^{-13}$ ($1e^{-13}$)

Table S4 Mean absolute errors (MAE) and root mean square errors (RMSE) (normalized per atom) of leave-one-out cross-validation technique for each investigated target functional in each solvent, for both bound and separated (in parentheses) states, obtained using the hyperparameter combinations σ and λ listed in Table S3**.**

method	HSE06-D3(BJ)		PBE0-D3(BJ)	
solvent	MAE/at. [meV]	RMSE/at. [meV]	MAE/at. [meV]	RMSE/at. [meV]
Carbon tetrachloride	0.1819 (0.2063)	0.2391 (0.2575)	0.1871 (0.2121)	0.2462 (0.2628)
Dichloromethane	0.2125 (0.2146)	0.2614 (0.2632)	0.2186 (0.2142)	0.2651 (0.2652)
Methanol	0.2517 (0.2178)	0.3046 (0.2754)	0.2536 (0.2164)	0.3066 (0.2738)

The combinations of σ and λ values yielding the lowest prediction errors are summarized in Table S3. The optimal λ-HP values take on the value $1e^{-13}$ eV in all cases, except in the case of separated state in CCl_4 solvent computed with HSE06-D3(BJ) functional ($5e^{-8}$ eV). The value $1e^{-13}$ eV was the lowest tested. With each decrease in the λ value, the errors were reduced by a progressively smaller amount, the difference in error between $\lambda = 1e^{-12}$ and $\lambda = 1e^{-13}$ eV was on order of $\sim 1e^{-12}$ eV/atom. This small λ value suggests that regularization is not required for these systems. The values of the σ-HP lie within a range from $\sigma = 0.0005$ Å (for the HSE06-D3(BJ) in the separated state in CCl_4 solvent to $\sigma = 0.04$ Å (HSE06-D3(BJ), bound state, CCl_4).

To quantify prediction accuracy, we monitored the statistical indicators MAE (Mean Absolute Error) and RMSE (Root Mean Square Error). The resulting MAE and RMSE values obtained from optimized HPs are reported in Table S4. The MAE/atom and RMSE/atom values are the highest for CH_3OH, followed by CH_2Cl_2 and the lowest for CCl_4, both in the bound and separated states for both hybrid functionals.

Table S5 The charge transfer values calculated from Natural bond order (NBO) charges[S2] and charge model 5 (CM5)[S3] methods of TMTU…I_2 complex in various solvents calculated with implicit solvent models at PBE0-D3/def2-TZVPP level of theory.

solvent	COSMO		SMD		CPCM	
	NBO	CM5	NBO	CM5	NBO	CM5
Carbon tetrachloride	-0.376	-0.380	-0.371	-0.374	-0.365	-0.367
Dichloromethane	-0.506	-0.514	-0.483	-0.491	-0.472	-0.479
Methanol	-0.555	-0.564	-0.525	-0.535	-0.510	-0.518

Table S6 The interaction energy, ΔE (kcal/mol), Gibbs free energy, ΔG (kcal/mol, $T = 298$ K), halogen bond length, r (Å), and charge transfer (CT_1, calculated from Hirshfeld charges; CT_2, calculated from Löwdin charges) of TMTU...I_2 complex in various solvents calculated in COSMO solvent model at ωB97M-V/def2-TZVPP level of theory.

solvent	ΔE	ΔG	r	CT_1/CT_2
Carbon tetrachloride	–11.94	–0.21	2.724	–0.292/–0.256
Dichloromethane	–12.97	–1.15	2.610	–0.521/–0.476
Methanol	–14.41	–2.14	2.574	–0.668/–0.610

Table S7 The interaction energy (ΔE), Gibbs free energy ($\Delta G_{calc.}$) (in kcal/mol, T = 298 K) and charge transfer (CT, calculated from Hirshfeld charges) of tetramethylurea…I_2 complex in various solvent media with COSMO solvent model calculated at PBE0-D3/def2-TZVPP level of theory. Experimental free energies[S4] are given for comparison.

Solvent	ε	ΔE	$\Delta G_{calc.}$	ΔG_{expt}[a]	CT
n-Octane	1.94	-9.65	0.29	-1.47	-0.193
Carbon tetrachloride	2.23	-9.59	0.37	-1.06	-0.200
Toluene	2.38	-9.56	0.40	-0.65	-0.203
Chloroform	4.81	-9.37	0.65	0	-0.234
1,1,2,2-tetrachloroethane	8.4	-9.31	0.79	~ 0	-0.232
Dichloromethane	8.93	-9.30	0.80	-0.41	-0.253
Acetone	20.7	-9.27	0.91	~ 0	-0.270
Methanol	32.61	-9.26	0.94	~ 0	-0.27
Acetonitrile	37.5	-9.26	0.94	~ 0	-0.275

[a] From ref. S4

Table S8 The relative free energy of various conformations TMTU…I_2 complex with 30 explicit solvent molecules in COSMO model calculated at PBE0-D3/def2-SVP level of theory.

Solvent		Relative free energy (kcal/mol)	
		bound state	separated state
Carbon tetrachloride	conformer 1	0.0	0.0
	conformer 2	2.07	0.92
	conformer 3	4.24	1.07
	conformer 4	5.20	5.14
	conformer 5	7.72	6.12
Dichloromethane	conformer 1	0.0	0.0
	conformer 2	0.88	1.18
	conformer 3	2.93	2.40
	conformer 4	3.42	3.68
	conformer 5	3.75	4.37
Methanol	conformer 1	0.0	0.0
	conformer 2	1.21	1.06
	conformer 3	2.14	4.99
	conformer 4	8.46	7.92
	conformer 5	10.36	9.97

Table S9 The relative free energy of various conformations TMTU…I_2 complex with 50 explicit solvent molecules calculated at PBE0-D3/def2-TZVPP:PM7 level of theory.

Solvent		Relative free energy (kcal/mol)

		bound state	separated state
Carbon tetrachloride	conformer 1	0.0	0.0
	conformer 2	1.67	3.26
	conformer 3	2.80	7.76
	conformer 4	6.60	7.76
	conformer 5	7.47	12.19
Dichloromethane	conformer 1	0.0	0.0
	conformer 2	7.71	2.93
	conformer 3	12.55	12.36
	conformer 4	12.64	15.27
	conformer 5	14.41	16.70
Methanol	conformer 1	0.0	0.0
	conformer 2	1.10	5.63
	conformer 3	1.33	6.58
	conformer 4	5.24	8.98
	conformer 5	9.20	9.88

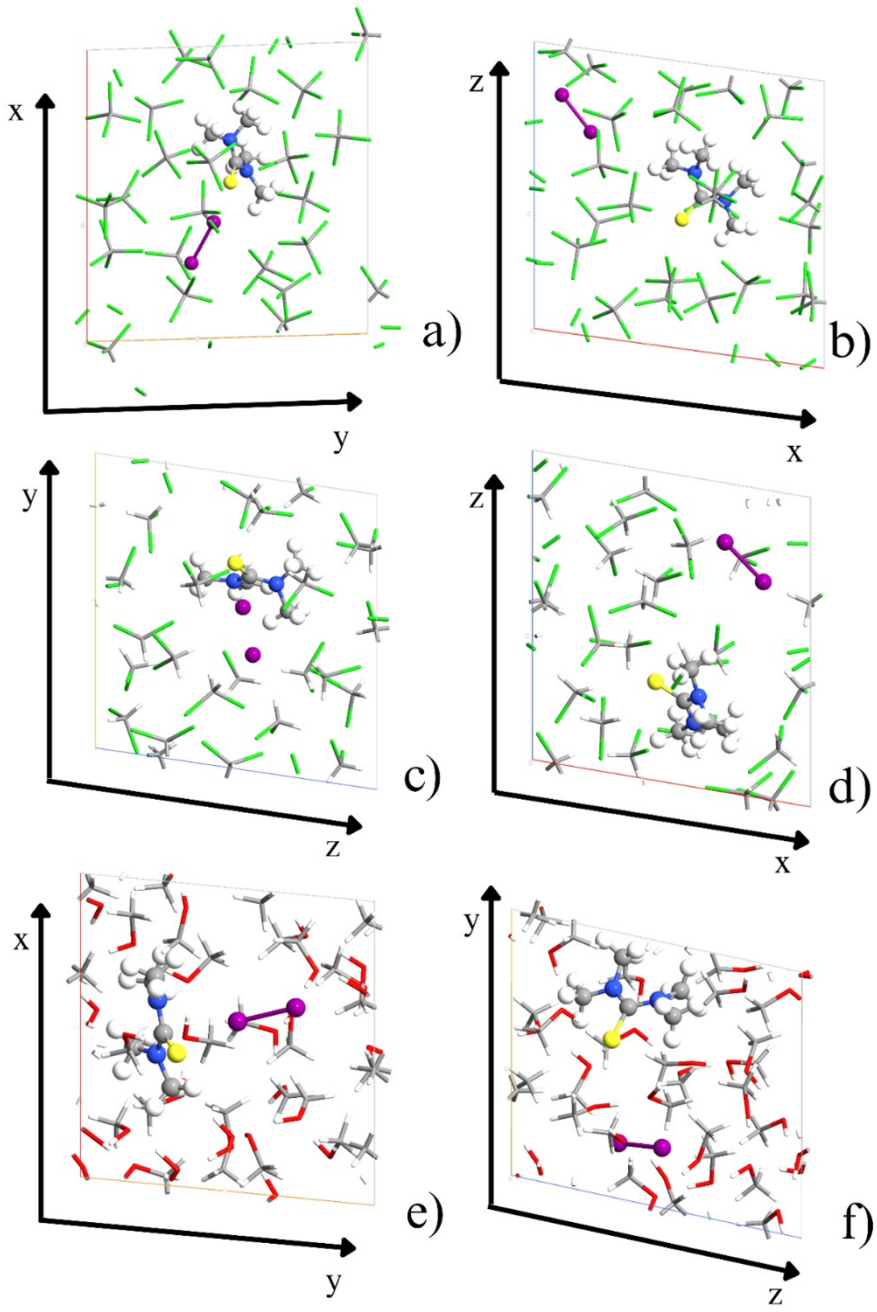

Fig. S1 Periodic models of tetramethylthiourea with an iodine molecule in CCl_4 (a,b), CH_2Cl_2 (c,d), and CH_3OH (e,f). The structures of the solvent molecules are simple. Each system was studied both as a complex (a, c, e) and as separated fragments (b, d, f). [C: grey, I: violet, N: blue, H: white, S: yellow, O: red]

References

S1 M. Stone, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 1974, **36**, 111–133.

S2 A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* 1988, **88**, 899–926.

S3 A. V. Marenich, S. V. Jerome, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* 2012, **8**, 527–541.

S2 C. C. Robertson, R. N. Perutz, L. Brammer and C. A. Hunter, *Chem Sci* 2014, **5**, 4179–4183.