

Supporting Information for “A Simple Correction for Nonadditive Dispersion within Extended Symmetry-Adapted Perturbation Theory (XSAPT)”

Ka Un Lao and John M. Herbert*

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210

1 Benchmark Structures and Interaction Energies

1.1 Cluster Geometries

Geometries for $F^-(H_2O)_{n \leq 6}$ and $Cl^-(H_2O)_{n \leq 6}$ were obtained from Ref. 1, where they were optimized at the RIMP2/aug-cc-pVTZ level. Structures for $F^-(H_2O)_{10}$ also come from Ref. 1 but were optimized at the B3LYP/6-31G* level. The $(H_2O)_6$ geometries are from Ref. 2, optimized at the MP2/aug-cc-pVTZ level. Structures for $(H_2O)_{20}$ are from Ref. 3 and were optimized using the TIP4P force field.

Structures for $SO_4^{2-}(H_2O)_{n=3-6}$ taken from Ref. 4 were re-optimized at the RIMP2/aug-cc-pVDZ level. For completeness, we also optimized $SO_4^{2-}(H_2O)$ and $SO_4^{2-}(H_2O)_2$ at the same level of theory, although in the gas phase these species are unstable with respect to electron autodetachment.^{5,6} Geometries of cationic $M^+(H_2O)_{n \leq 6}$ clusters were optimized using TPSS+D3,⁷⁻⁹ with Becke-Johnson damping for the dispersion correction¹⁰ and the def2-TZVPP basis set.¹¹

Starting geometries for the clathrate hydrate building-block structures $(H_2O)_{20}$, $(H_2O)_{24}$, and $(H_2O)_{28}$ are taken from Ref. 12. Next, small guest molecules were inserted into these structures, and the geometry of the host/guest system was optimized at the B97M-V/6-31+G(d) level, but fixing the oxygen atoms of the clathrate cage in place.

Cartesian coordinates for all of these structures are available as a separate attachment.

1.2 CCSD(T)/CBS Interaction Energies

CCSD(T)/CBS benchmarks for $(H_2O)_6$, $(H_2O)_{20}$, $F^-(H_2O)_{n=1-6}$, $Cl^-(H_2O)_{n=1-6}$, and $F^-(H_2O)_{10}$ are taken from Ref. 1. For the $SO_4^{2-}(H_2O)_{n=1-6}$ clusters, the MP2 correlation energies in the CBS

*herbert@chemistry.ohio-state.edu

limit were estimated using a two-point (aTZ/aQZ) extrapolation, and this correlation energy was added to the Hartree-Fock/aQZ energy to estimate the MP2/CBS energy. Then, correction

$$\delta_{\text{CCSD(T)}} = E_{\text{CCSD(T)}} - E_{\text{MP2}} \quad (\text{S1})$$

was evaluated using the def2-TZVPP basis set, and the sum of this correction plus the MP2/CBS interaction energy is taken to be the CCSD(T)/CBS interaction energy. MP2/CBS energies for the alkali-water complexes were evaluated in a similar fashion, with the def2-TZVPP and def2-QZVPP basis sets used for the alkali metal. To evaluate $\delta_{\text{CCSD(T)}}$ for the alkali-water clusters we used the def2-TZVPP basis set for the alkali metal and a “heavy augmented” version of aug-cc-pVTZ for the remaining atoms, meaning that diffuse functions on hydrogen are removed from aug-cc-pVTZ.

Interaction energies for the clathrate hydrate structures were evaluated in a manner similar to that outlined about, but using def2-TZVPP and def2-QZVPP for the triple- and quadruple- ζ basis sets in the MP2/CBS calculation, and def2-TZVP in the $\delta_{\text{CCSD(T)}}$ calculation. In addition, the CCSD(T) energies for the clathrate structures were computed using the domain based local pair-natural orbital (DLPNO) scheme,^{13,14} with tight PNO of truncation threshold.¹⁵ This approximation has been shown to approach the canonical CCSD(T) result within 1 kJ/mol.¹⁶ The frozen core approximation was used in all post-Hartree-Fock calculations where 1s orbital has been frozen for C, O, F, and Na, and 1s2s2p orbitals have been frozen for S, Cl, and K .

The CCSD(T)/CBS interaction energies obtained in this way are listed in Tables S1–S4. The MP2/CBS interaction energies obtained in this way are listed in Tables S1–S3 and S5.

1.3 Benchmark Energy Components for S66

The SAPT2+(3) is defined as follows:¹⁷

$$\begin{aligned} E_{\text{SAPT2+(3)}} = & E_{\text{elst}}^{(10)} + E_{\text{elst,resp}}^{(12)} + E_{\text{elst,resp}}^{(13)} + E_{\text{exch}}^{(10)} + E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)} \\ & + E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + {}^tE_{\text{ind}}^{(22)} + {}^tE_{\text{exch-ind}}^{(22)} + \delta E_{\text{int,resp}}^{\text{HF}} \\ & + E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)} + E_{\text{disp}}^{(30)} . \end{aligned} \quad (\text{S2})$$

(Consult Ref. 17 for an explanation of the notation, which is standard.) SAPT2+(3) interaction energies are comparable (in cost and quality) to CCSD(T) values, and we have used SAPT2+(3) in previous work to provide benchmarks for individual energy components. To define the energy

components, we group the terms in Eq. (S2) as suggested by Sherrill and co-workers:¹⁸

$$E_{\text{electrostatic}} = E_{\text{elst}}^{(10)} + E_{\text{elst,resp}}^{(12)} + E_{\text{elst,resp}}^{(13)} \quad (\text{S3a})$$

$$E_{\text{exchange}} = E_{\text{exch}}^{(10)} + E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)} \quad (\text{S3b})$$

$$E_{\text{induction}} = E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + {}^tE_{\text{ind}}^{(22)} + {}^tE_{\text{exch-ind}}^{(22)} + \delta E_{\text{int,resp}}^{\text{HF}} \quad (\text{S3c})$$

$$E_{\text{dispersion}} = E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)} + E_{\text{disp}}^{(30)} . \quad (\text{S3d})$$

Using these definitions, we computed energy components for the S66 data set¹⁹ at the SAPT2+(3)/aug-cc-pVTZ level. These are listed in Tables S7, S8, and S9 for hydrogen-bonded dimers, dispersion-bound dimers, and dimers of mixed influence, respectively.

1.4 Interaction Energies of clathrate hydrate host-guest complexes

CCSD(T)/CBS, MP2/CBS, DFT, and XSAPT binding energies for $\text{CH}_4@(\text{H}_2\text{O})_{20}$, $\text{CH}_4@(\text{H}_2\text{O})_{24}$, and $\text{THF}@(\text{H}_2\text{O})_{28}$ are listed in Table S6.

1.5 Software and Other Computational Details

The TPSS+D3 geometry optimizations for alkali-water clusters, DLPNO-CCSD(T) and MP2 calculations for clathrate hydrates were performed using the ORCA program,²⁰ v. 3.0.3, with ‘‘TightSCF’’ convergence criteria. In the TPSS+D3 calculations, the RI approximation²¹ for the Coulomb integrals was employed, using the matching (default) auxiliary basis set.²² The numerical quadrature grid denoted in ORCA as ‘‘Grid4’’ was used. All SAPT2+(3) calculations, along with single-point MP2 and CCSD(T) calculations for $\text{SO}_4^{2-}(\text{H}_2\text{O})_n$ and alkali-water clusters were performed using the ‘‘Beta5’’ version of the PSI4 program.²³ The RI-JK algorithm²⁴ for the SCF calculations and the RI approximation for correlation parts were employed in all MP2 and CCSD(T) calculations in this work. Second-order (X)SAPT calculations were performed using a locally-modified version of Q-CHEM.²⁵

For the DFT and XSAPT(KS) calculations of intermolecular interaction energies, the SCF convergence criterion is set to 10^{-7} a.u. and the the integral screening threshold is set to 10^{-12} a.u. A fairly dense Euler-Maclaurin-Lebedev quadrature grid ($N_{\text{radial}} = 75$, $N_{\text{angular}} = 302$) is employed for the supersystem DFT calculations. All of the supersystem calculations are counterpoise-corrected.²⁶

All XSAPT(KS) calculations reported here use ‘‘ChEIPG’’ charges²⁷ derived from the electrostatic potential for the XPol embedding, using the smooth ChEIPG algorithm in Ref. 28. These calculations furthermore use what Jacobson and Herbert have called ‘‘projected’’ dimer basis sets²⁹ (pseudocanonicalized,^{30,31} but monomer-centered). ChEIPG charges were computed using Lebedev grids³² using a ‘‘head space’’ of 3.0 Å, with radial shells spaced $\Delta r = 0.25$ Å apart. Each Lebedev

shell contains 590 points.

2 Parameters

2.1 Tuned Values of the Range Separation Parameter

Table S10 lists the tuned value of the range separation parameter, ω , for each of the monomers considered in this work, using both ω_{IP} tuning³³ and ω_{GDD} tuning.³⁴ These values were tuned using the LRC- ω PBE/hpTZVPP method based on monomer geometries optimized using RIMP2/aug-cc-pVDZ. (The hpTZVPP basis set corresponds to def2-TZVPP augmented using diffuse functions on non-hydrogen atoms, taken from Pople’s 6-311+G basis set.)

2.2 Empirical +*ai*D3 Dispersion Potential

As in previous work^{1,35} we use SAPT2+(3), Eq. (S2), as our benchmark level of theory for the purpose of fitting the *ai*D3 dispersion parameters. These parameters are new in this work for Li, Na, and K; see Ref. 1 for the remaining *ai*D3 parameters. Previously we used aug-cc-pVTZ for the benchmark SAPT2+(3) calculations, and continue to do so for the main-group elements, but since this basis is unavailable for the alkali metals we use def2-QZVPP for these.

The training set used to determine these parameters consists of 21 dimers optimized at the RIMP2 level, with the same basis set as for the SAPT2+(3) calculations. The training set consists of complexes $M^+ \cdots (\text{molecule})$, where $M = \text{Li, Na, or K}$ and the molecule is $\text{C}_6\text{H}_6, \text{C}_2\text{H}_2, \text{C}_2\text{H}_4, \text{CH}_3\text{NH}_2, \text{CH}_3\text{OH}, \text{H}_2\text{O},$ and NH_3 . For each complex, five different radial geometries corresponding to the same angular configuration were considered, with displacements of 0.9, 1.0, 1.2, 1.5, and 2.0 times the equilibrium separation. This affords 105 data points in total.

The *ai*D3 dispersion potential is the same one used in previous dispersion-corrected SAPT approaches,^{1,35,36}

$$E_{\text{disp}}^{\text{aiD3}} = - \sum_{i \in A} \sum_{\substack{j \in B \\ (B \neq A)}} \left[\frac{C_{6,ij}}{R_{ij}^6} f_6(\beta_{ij} R_{ij}) + \frac{C_{8,ij}}{R_{ij}^8} f_8(\beta_{ij} R_{ij}) \right], \quad (\text{S4})$$

where i and j are nuclei located on different monomers, and

$$f_n(r_{ij}) = 1 - \exp(-r_{ij}) \sum_{m=0}^n \frac{r_{ij}^m}{m!} \quad (\text{S5})$$

is the Tang-Toennies damping function.³⁷ We take

$$C_{6,ij} = \sqrt{C_{6,i} C_{6,j}} \quad (\text{S6a})$$

$$C_{8,ij} = \sqrt{C_{8,i} C_{8,j}} \quad (\text{S6b})$$

$$\beta_{ij} = \sqrt{\beta_i \beta_j} \quad (\text{S6c})$$

where $C_{6,i}$, $C_{8,i}$, and β_i are fitting parameters. For $i = \text{H, C, N, and O}$ we fix these parameters at their previously-optimized values,¹ then vary them for $i = \text{Li}^+, \text{Na}^+, \text{and K}^+$ so as to minimize

$$\chi^2 = \frac{1}{105} \sum_{n=1}^{105} \left[E_{\text{disp}}^{(n)}(\text{aiD3}) - E_{\text{disp}}^{(n)}(\text{SAPT}) \right]^2. \quad (\text{S7})$$

Here, $E_{\text{disp}}^{(n)}$ is the dispersion energy for the n th dimer, computed either using the benchmark SAPT2+(3) calculation or else the *aiD3* dispersion potential. We used a genetic algorithm followed by simplex optimization to fit the parameters, and the final deviation was $\chi = 0.1310$ kcal/mol. Optimized values of $C_{6,i}$, $C_{8,i}$, and β_i are listed in Table S11.

References

- [1] K. U. Lao and J. M. Herbert, *J. Phys. Chem. A* **119**, 235 (2015).
- [2] Y. Chen and H. Li, *J. Phys. Chem. A* **114**, 11719 (2010).
- [3] J. K. Kazimirski and V. Buch, *J. Phys. Chem. A* **107**, 9762 (2003).
- [4] L. C. Smeeton, J. D. Farrell, M. T. Oakley, D. J. Wales, and R. L. Johnston, *J. Chem. Theory Comput.* **11**, 2377 (2015).
- [5] A. T. Blades and P. Kebarle, *J. Am. Chem. Soc.* **116**, 10761 (1994).
- [6] X.-B. Wang, J. B. Nicholas, and L.-S. Wang, *J. Chem. Phys.* **113**, 10837 (2000).
- [7] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Phys. Rev. Lett.* **91**, 146401:1 (2003).
- [8] J. P. Perdew, J. Tao, V. N. Staroverov, and G. E. Scuseria, *J. Chem. Phys.* **120**, 6898 (2004).
- [9] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.* **132**, 154104:1 (2010).
- [10] S. Grimme, S. Ehrlich, and L. Goerigk, *J. Comput. Chem.* **32**, 1456 (2011).
- [11] F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).

- [12] E. B. Fleischer and K. C. Janda, *J. Phys. Chem. A* **117**, 4001 (2013).
- [13] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, *J. Chem. Phys.* **139**, 134101:1 (2013).
- [14] M. Sparta and F. Neese, *Chem. Soc. Rev.* **43**, 5032 (2014).
- [15] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin, and F. Neese, *J. Chem. Theory Comput.* **11**, 1525 (2015).
- [16] D. G. Liakos and F. Neese, *J. Chem. Theory Comput.* **11**, 4054 (2015).
- [17] E. G. Hohenstein and C. D. Sherrill, *WIREs Comput. Mol. Sci.* **2**, 304 (2012).
- [18] J. C. Flick, D. Kosenkov, E. G. Hohenstein, C. D. Sherrill, and L. V. Slipchenko, *J. Chem. Theory Comput.* **8**, 2835 (2012).
- [19] J. Řezáč, K. E. Riley, and P. Hobza, *J. Chem. Theory Comput.* **7**, 2427 (2011), Erratum: *ibid.* **10**, 1359–1360 (2014).
- [20] F. Neese, *WIREs Comput. Mol. Sci.* **2**, 73 (2012).
- [21] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, *Chem. Phys. Lett.* **240**, 283 (1995).
- [22] F. Weigend, *Phys. Chem. Chem. Phys.* **8**, 1057 (2006).
- [23] J. M. Turney et al., *WIREs Comput. Mol. Sci.* **2**, 556 (2012).
- [24] F. Weigend, M. Kattannek, and R. Ahlrichs, *J. Chem. Phys.* **130**, 164106:1 (2009).
- [25] Y. Shao et al., *Mol. Phys.* **113**, 184 (2015).
- [26] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, *Chem. Rev.* **94**, 1873 (1994).
- [27] C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.* **11**, 361 (1990).
- [28] J. M. Herbert, L. D. Jacobson, K. U. Lao, and M. A. Rohrdanz, *Phys. Chem. Chem. Phys.* **14**, 7679 (2012).
- [29] L. D. Jacobson and J. M. Herbert, *J. Chem. Phys.* **134**, 094118:1 (2011).
- [30] K. Wolinski and P. Pulay, *J. Chem. Phys.* **118**, 9497 (2003).
- [31] R. P. Steele, R. A. DiStasio, Jr., Y. Shao, J. Kong, and M. Head-Gordon, *J. Chem. Phys.* **125**, 074108:1 (2006).

- [32] Z. C. Holden, R. M. Richard, and J. M. Herbert, *J. Chem. Phys.* **139**, 244108:1 (2013), Erratum: *ibid.* **142**, 059901:1–2 (2015).
- [33] R. Baer, E. Livshits, and U. Salzner, *Annu. Rev. Phys. Chem.* **61**, 85 (2010).
- [34] M. Modrzejewski, L. Rajchel, G. Chalasinski, and M. M. Szczesniak, *J. Phys. Chem. A* **117**, 11580 (2013).
- [35] K. U. Lao and J. M. Herbert, *J. Chem. Phys.* **139**, 034107:1 (2013), Erratum: *ibid.* **140**, 119901 (2014).
- [36] R. Podeszwa, K. Pernal, K. Patkowski, and K. Szalewicz, *J. Phys. Chem. Lett.* **1**, 550 (2010).
- [37] K. T. Tang and J. P. Toennies, *J. Chem. Phys.* **80**, 3726 (1984).

Isomer	Interaction Energy / kcal mol ⁻¹	
	CCSD(T)/CBS	MP2/CBS
prism	-48.31	-48.22
cage	-48.02	-48.13
book1	-47.61	-47.93
book2	-47.33	-47.65
cyclic chair	-46.52	-47.02
bag	-46.87	-47.19
cyclic boat1	-45.52	-46.04
cyclic boat2	-45.42	-45.91

Table S1: CCSD(T)/CBS and MP2/CBS interaction energies for isomers of (H₂O)₆ reported in Ref. 1.

Isomer	Interaction Energy / kcal mol ⁻¹	
	CCSD(T)/CBS	MP2/CBS
1	-169.02	-166.97
2	-169.03	-166.98
3	-168.01	-165.93
4	-169.07	-167.06
5	-169.91	-167.98
6	-171.42	-169.68
7	-169.92	-167.99
8	-166.65	-164.34
9	-163.59	-161.25
10	-168.22	-166.47

Table S2: CCSD(T)/CBS and MP2/CBS interaction energies for isomers of F⁻(H₂O)₁₀ reported in Ref. 1.

Isomer	Interaction Energy / kcal mol ⁻¹	
	CCSD(T)/CBS	MP2/CBS
1	-200.54	-197.97
2	-199.20	-196.68
3	-198.93	-196.51
4	-197.89	-195.02
5	-198.15	-195.58
6	-198.17	-195.24
7	-197.67	-194.62
8	-197.44	-194.90
9	-197.03	-194.29
10	-196.59	-193.92

Table S3: CCSD(T)/CBS and MP2/CBS interaction energies for isomers (H₂O)₂₀ reported in Ref. 1.

<i>n</i>	Interaction Energy / kcal mol ⁻¹					
	F ⁻ (H ₂ O) _{<i>n</i>}	Cl ⁻ (H ₂ O) _{<i>n</i>}	SO ₄ ²⁻ (H ₂ O) _{<i>n</i>}	Li ⁺ (H ₂ O) _{<i>n</i>}	Na ⁺ (H ₂ O) _{<i>n</i>}	K ⁺ (H ₂ O) _{<i>n</i>}
1	-32.31	-15.50	-31.99	-34.32	-23.84	-17.81
2	-52.27	-31.01	-60.09	-63.96	-44.79	-33.59
3	-70.14	-47.62	-81.06	-87.56	-63.27	-49.12
4	-85.24	-58.20	-104.75	-105.07	-78.25	-62.31
5	-101.09	-74.73	-125.97	-120.25	-92.73	-77.73
6	-116.58	-86.86	-144.08	-134.35	-106.92	-91.89

Table S4: CCSD(T)/CBS interaction energies for ion–water clusters where interaction energies for F⁻(H₂O)_{*n*} and Cl⁻(H₂O)_{*n*} were reported in Ref. 1.

<i>n</i>	Interaction Energy / kcal mol ⁻¹					
	F ⁻ (H ₂ O) _{<i>n</i>}	Cl ⁻ (H ₂ O) _{<i>n</i>}	SO ₄ ²⁻ (H ₂ O) _{<i>n</i>}	Li ⁺ (H ₂ O) _{<i>n</i>}	Na ⁺ (H ₂ O) _{<i>n</i>}	K ⁺ (H ₂ O) _{<i>n</i>}
1	-32.25	-15.75	-32.12	-34.26	-23.80	-17.94
2	-51.86	-31.30	-60.24	-63.78	-44.68	-33.81
3	-69.09	-47.74	-81.56	-87.17	-63.06	-49.43
4	-83.84	-58.40	-105.16	-104.43	-77.89	-62.65
5	-99.53	-74.94	-126.24	-119.68	-92.41	-78.43
6	-114.98	-87.16	-144.59	-133.79	-106.64	-92.63

Table S5: MP2/CBS interaction energies for ion–water clusters where interaction energies for F⁻(H₂O)_{*n*} and Cl⁻(H₂O)_{*n*} were reported in Ref. 1.

Method	binding energy / kcal mol ⁻¹			MAE (kcal mol ⁻¹)	MAPE
	CH ₄ @(H ₂ O) ₂₀	CH ₄ @(H ₂ O) ₂₄	THF@(H ₂ O) ₂₈		
Supersystem Methods					
CCSD(T)/CBS	-4.88	-4.32	-12.98	-	-
MP2/CBS	-5.01	-4.42	-13.32	0.19	2.60%
ω B97X-V	-6.47	-5.69	-15.92	1.97	29.02%
ω B97M-V	-6.04	-5.12	-14.98	1.32	19.22%
B97M-V	-5.78	-4.77	-14.39	0.92	13.29%
ω B97X-D3	-7.68	-6.43	-16.51	2.82	44.55%
M11	-5.97	-3.23	-9.61	1.85	24.53%
MN15	-9.72	-6.86	-16.36	3.59	61.40 %
PW6B95-D3	-6.35	-5.65	-14.87	1.57	25.24%
SCAN	-4.67	-3.48	-10.59	1.14	14.01%
SCAN0	-3.66	-2.72	-8.75	2.35	31.56%
Functionals Recommend Previously					
revPBE	6.60	3.50	6.10	12.79	187.82%
B3LYP	3.35	2.11	1.73	9.79	143.67%
B3LYP-D2	-5.90	-4.32	-14.10	0.72	9.91%
PBE-D2	-5.73	-4.80	-13.95	0.77	12.01%
B97-D2	-4.61	-4.35	-12.99	0.10	2.01%
BLYP-D3	-7.29	-5.35	-15.98	2.15	32.22%
M05-2X	-5.86	-4.83	-12.56	0.64	11.72%
M06-2X	-6.27	-4.10	-11.24	1.12	15.69%
ω B97X-D	-6.52	-5.10	-14.55	1.33	21.32%
XSAPT(KS) + aiD2 Methods					
XSAPT(KS)+aiD2 (ω_{IP})	-5.14	-4.85	-13.85	0.55	8.11%
XSAPT(KS)+aiD2+DC (ω_{IP})	-4.91	-4.68	-13.38	0.26	3.99%
XSAPT(KS)+aiD2 (ω_{GDD})	-5.04	-4.80	-13.79	0.49	6.97%
XSAPT(KS)+aiD2+DC (ω_{GDD})	-4.81	-4.63	-13.31	0.24	3.72%
XSAPT(KS) + aiD3 Methods					
XSAPT(KS)+aiD3 (ω_{IP})	-3.53	-3.78	-12.53	0.78	14.55%
XSAPT(KS)+aiD3+DC (ω_{IP})	-3.30	-3.60	-12.06	1.07	18.67%
XSAPT(KS)+aiD3 (ω_{GDD})	-3.44	-3.73	-12.48	0.84	15.69%
XSAPT(KS)+aiD3+DC (ω_{GDD})	-3.20	-3.56	-12.00	1.14	19.85%

Table S6: The absolute binding energies, and their MAEs and MAPEs with respect to CCSD(T)/CBS benchmarks for CH₄@(H₂O)₂₀, CH₄@(H₂O)₂₄, and THF@(H₂O)₂₈. The hpTZVPP basis set is used for the XSAPT(KS)+aiD2/aiD3 calculations and the hpTZVPP basis set is used for the non-additive dispersion corrections $E_{DC(full)}$. The def2-TZVPPD basis set is used for DFT calculations with CP corrections.

System	Energy / kcal mol ⁻¹				
	Electrostatics	Exchange- Repulsion	Induction	Dispersion	Total
H ₂ O···H ₂ O	-7.77	7.57	-2.32	-2.32	-4.84
H ₂ O···MeOH	-8.79	9.28	-2.88	-3.07	-5.46
H ₂ O···MeNH ₂	-12.09	13.03	-4.32	-3.51	-6.90
H ₂ O···peptide	-12.37	13.16	-4.44	-4.53	-8.19
MeOH···MeOH	-8.90	9.81	-3.05	-3.50	-5.65
MeOH···MeNH ₂	-12.72	14.61	-4.83	-4.63	-7.57
MeOH···peptide	-12.59	14.24	-4.74	-5.31	-8.39
MeOH···H ₂ O	-7.75	7.89	-2.45	-2.64	-4.95
MeNH ₂ ···MeOH	-4.11	4.92	-1.11	-2.75	-3.05
MeNH ₂ ···MeNH ₂	-5.83	7.15	-1.66	-3.83	-4.17
MeNH ₂ ···peptide	-6.88	8.47	-1.98	-5.20	-5.59
MeNH ₂ ···H ₂ O	-12.22	13.43	-4.46	-3.97	-7.23
peptide···MeOH	-7.87	8.27	-2.41	-4.24	-6.26
peptide···MeNH ₂	-10.84	12.12	-3.75	-5.08	-7.54
peptide···peptide	-11.21	12.27	-3.91	-6.10	-8.95
peptide···H ₂ O	-6.67	6.07	-1.85	-2.75	-5.21
uracil···uracil ^a	-25.98	29.77	-12.50	-9.38	-18.09
H ₂ O···pyridine	-11.21	12.18	-4.13	-3.74	-6.91
MeOH···pyridine	-11.85	13.56	-4.60	-4.63	-7.51
AcOH···AcOH	-32.36	40.13	-18.06	-10.06	-20.36
AcNH ₂ ···AcNH ₂	-25.54	27.98	-11.11	-8.23	-16.91
AcOH···uracil	-30.57	35.30	-15.64	-9.70	-20.61
AcNH ₂ ···uracil	-28.86	31.11	-13.28	-9.07	-20.10

^aWatson-Crick base pair structure.

Table S7: Energy components and total interaction energies computed at the SAPT2+(3)/aug-cc-pVTZ level for the hydrogen-bonded subset of S66.

System	Energy / kcal mol ⁻¹				Total
	Electrostatics	Exchange- Repulsion	Induction	Dispersion	
benzene···benzene ^a	-1.47	6.23	-0.68	-6.94	-2.86
pyridine···pyridine ^a	-2.93	7.29	-0.82	-7.53	-3.99
uracil···uracil ^a	-8.93	12.61	-1.85	-12.59	-10.76
benzene···pyridine ^a	-2.33	6.92	-0.77	-7.33	-3.51
benzene···uracil ^a	-4.75	10.10	-1.18	-10.25	-6.08
pyridine···uracil ^a	-5.79	10.04	-1.26	-10.31	-7.33
benzene···ethene	-0.79	4.22	-0.49	-4.13	-1.19
uracil···ethene	-3.12	5.74	-0.59	-5.34	-3.31
uracil···ethyne	-3.91	5.56	-0.63	-4.88	-3.86
pyridine···ethene	-1.47	4.78	-0.54	-4.39	-1.62
pentane···pentane	-1.80	6.07	-0.51	-7.57	-3.80
neopentane···pentane	-1.30	4.30	-0.39	-5.26	-2.65
neopentane···neopentane	-0.75	2.92	-0.29	-3.71	-1.83
cyclopentane···neopentane	-1.29	4.29	-0.40	-5.07	-2.47
cyclopentane···cyclopentane	-1.47	4.82	-0.44	-5.98	-3.06
benzene···cyclopentane	-2.24	6.15	-0.66	-6.88	-3.63
benzene···neopentane	-1.75	4.72	-0.55	-5.40	-2.97
uracil···pentane	-2.75	7.55	-0.89	-9.08	-5.18
uracil···cyclopentane	-2.21	6.43	-0.66	-7.97	-4.41
uracil···neopentane	-2.44	5.49	-0.55	-6.46	-3.97
ethene···pentane	-1.05	3.41	-0.35	-3.97	-1.97
ethyne···pentane	-1.13	3.03	-0.34	-3.26	-1.71
peptide···pentane	-2.56	6.86	-0.99	-7.79	-4.48

^a π -stacked configuration.

Table S8: Energy components and total interaction energies computed at the SAPT2+(3)/aug-cc-pVTZ level for the dispersion-bound subset of S66.

System	Energy / kcal mol ⁻¹				
	Electrostatics	Exchange- Repulsion	Induction	Dispersion	Total
benzene···benzene ^a	-1.84	4.23	-0.62	-4.79	-3.01
pyridine···pyridine ^a	-2.78	4.91	-0.79	-5.06	-3.72
benzene···pyridine ^a	-2.32	4.45	-0.77	-4.87	-3.50
benzene···ethyne	-2.19	3.29	-0.89	-3.24	-3.03
ethyne···ethyne	-1.83	2.10	-0.54	-1.31	-1.58
benzene···AcOH ^b	-4.07	6.27	-2.07	-5.02	-4.89
benzene···AcNH ₂	-4.82	6.51	-1.68	-4.54	-4.53
benzene···H ₂ O	-2.89	3.81	-1.11	-3.12	-3.31
benzene···MeOH	-3.32	5.33	-1.32	-4.87	-4.18
benzene···MeNH ₂	-2.42	4.65	-0.75	-4.71	-3.23
benzene···peptide	-3.90	6.49	-1.47	-6.51	-5.39
pyridine···pyridine ^c	-4.75	5.81	-1.47	-4.10	-4.51
ethyne···H ₂ O	-3.91	3.41	-1.01	-1.52	-3.03
ethyne···AcOH	-7.06	8.45	-2.68	-3.77	-5.06
pentane···AcOH	-1.71	4.73	-0.59	-5.46	-3.03
pentane···AcNH ₂	-2.31	6.01	-1.06	-6.35	-3.71
benzene···AcOH	-3.12	5.87	-0.87	-5.80	-3.93
peptide···ethene	-2.69	4.64	-0.83	-4.20	-3.08
pyridine···ethyne	-5.92	6.13	-1.91	-2.62	-4.32
MeNH ₂ ···pyridine	-4.31	6.49	-1.13	-4.96	-3.91

^aT-shaped configuration. ^bO-H··· π configuration. ^cC-H···N configuration.

Table S9: Energy components and total interaction energies computed at the SAPT2+(3)/aug-cc-pVTZ level for the “mixed-influence” subset of S66.

Monomer	$\omega_{\text{IP}} / a_0^{-1}$	$\omega_{\text{GDD}} / a_0^{-1}$
adenine	0.271	0.316
2-aminopyridine	0.293	0.326
benzene	0.280	0.337
ethyne	0.397	0.428
ethene	0.359	0.420
methane	0.454	0.443
formamide	0.460	0.406
formic acid	0.412	0.429
water	0.502	0.488
HCN	0.452	0.472
indole	0.267	0.306
ammonia	0.440	0.450
phenol	0.292	0.330
pyrazine	0.367	0.350
2-pyridoxine	0.294	0.335
thymine	0.284	0.325
uracil	0.295	0.336
MeNH ₂	0.397	0.404
MeOH	0.438	0.421
AcNH ₂	0.453	0.371
AcOH	0.381	0.384
cyclopentane	0.420	0.331
neopentane	0.287	0.324
pentane	0.365	0.320
peptide	0.341	0.347
pyridine	0.316	0.349
F ⁻	0.480	0.447
Cl ⁻	0.372	0.324
SO ₄ ²⁻	0.344	0.316
Li ⁺	2.006	1.305
Na ⁺	1.049	0.839
K ⁺	0.755	0.589

Table S10: Tuned values of the range separation parameter, ω , for geometries optimized at the RIMP2/aug-cc-pVDZ level. Tuning was performed at the LRC- ω PBE/hpTZVPP level.

Element	$C_6/$ J nm ⁶ mol ⁻¹	$C_8/$ J nm ⁸ mol ⁻¹	$\beta/$ a_0^{-1}
Li	0.0002	0.0220	0.7674
Na	0.0184	0.6891	0.5271
K	0.4950	25.5348	0.4024

Table S11: Fitting parameters that define the *aiD3* dispersion potential for the alkali metals.