How Accurate are Approximate Quantum Chemical Methods at Modelling Solute-Solvent Interactions in Solvated Clusters?

Junbo Chen,[a] Bun Chan,*[b] Yihan Shao[c] and Junming Ho*[a]

[a] School of Chemistry, University of New South Wales, Sydney, NSW 2052, Australia

[b] Graduate School of Engineering, Nagasaki University, Bunkyo-machi 1-14, Nagasaki 852-8521, Japan

[c] Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019, Unites States

SUPPORTING INFORMATION

A. Basis Set Superposition Errors (BSSE) and Timing

Table S1. Average BSSE of M06-2X in various basis sets for ZGLY in 5-, 10-, 20-, and 40-H_2O, clusters over Frame 0, 9 and 19, in kJ mol^{-1}

Basis set	ZGLY-5H_2O	ZGLY-10H_2O	ZGLY-20H_2O	ZGLY-40H_2O
aug'-cc-pVDZ	9.89	17.47	21.69	23.64
aug-cc-pVDZ	8.01	15.04	22.03	27.11
aug'-cc-pVTZ	2.35	3.78	5.00	5.48
aug-cc-pVTZ	2.69	4.44	5.89	-

Table S2. BSSE and wall time of M06-2X in various basis sets, in NGLY-10H_2O clusters

Frame 0

Basis set	Number of Basis Functions	Raw interaction energy (kJ mol^{-1})	Counterpoised (CP) corrected interaction energy (kJ mol^{-1})	BSSE (kJ mol^{-1})	Wall time (hours)
aug'-cc-pVDZ	470	-82.8688	-69.9115	12.96	0.19
aug-cc-pVDZ	570	-83.0818	-71.6421	11.44	0.28
def2-TZVPPD	1016	-76.8173	-74.6507	2.17	0.85
aug'-cc-pVTZ	1040	-79.1719	-75.8506	3.32	0.94
aug-cc-pVTZ	1265	-79.514	-75.6687	3.85	2.04
def2-QZVPPD	1806	-74.9285	-74.2614	0.67	5.60
pc-3	1810	-75.1843	-74.3244	0.86	2.99
aug'-cc-pVQZ	1950	-76.7079	-75.0766	1.63	9.35

Frame 9

Basis set	Number of Basis Functions	Raw interaction energy (kJ mol^{-1})	Counterpoised (CP) corrected interaction energy (kJ mol^{-1})	BSSE (kJ mol^{-1})	Wall time (hours)
aug'-cc-pVDZ	470	-66.8838	-56.8382	10.05	0.14
aug-cc-pVDZ	570	-67.1121	-58.2432	8.87	0.24
def2-TZVPPD	1016	-62.9898	-61.2073	1.78	0.67
aug'-cc-pVTZ	1040	-65.5558	-62.9407	2.62	0.79
aug-cc-pVTZ	1265	-65.9018	-62.6271	3.27	2.15
def2-QZVPPD	1806	-61.3994	-60.9609	0.44	4.72
pc-3	1810	-61.8841	-61.1078	0.78	2.73
aug'-cc-pVQZ	1950	-63.2107	-61.8888	1.32	7.15

Frame 19

Basis set	Number of Basis Functions	Raw interaction energy (kJ mol^{-1})	Counterpoised (CP) corrected interaction energy (kJ mol^{-1})	BSSE (kJ mol^{-1})	Wall time (hours)
aug'-cc-pVDZ	470	-84.6842	-70.8645	13.82	0.15
aug-cc-pVDZ	570	-85.3492	-73.0776	12.27	0.27
def2-TZVPPD	1016	-78.027	-76.1679	1.86	0.82
aug'-cc-pVTZ	1040	-81.4384	-78.2522	3.19	0.91
aug-cc-pVTZ	1265	-82.0984	-78.2035	3.89	2.41
def2-QZVPPD	1806	-76.6156	-76.0019	0.61	5.21
pc-3	1810	-76.9536	-76.0794	0.87	3.05
aug'-cc-pVQZ	1950	-78.3198	-76.7756	1.54	7.71

Table S3. BSSE and wall time of M06-2X in various basis sets, in TS-10H_2O clusters

Frame 0

Basis set	Number of Basis Functions	Raw interaction energy (kJ mol^{-1})	Counterpoised (CP) corrected interaction energy (kJ mol^{-1})	BSSE (kJ mol^{-1})	Wall time (hours)
aug'-cc-pVDZ	470	-137.726	-124.419	13.31	0.17
aug-cc-pVDZ	570	-137.756	-125.816	11.94	0.29
def2-TZVPPD	1016	-131.616	-129.689	1.93	0.76
aug'-cc-pVTZ	1040	-135.715	-132.059	3.66	0.91
aug-cc-pVTZ	1265	-135.749	-131.757	3.99	2.36
def2-QZVPPD	1806	-130.666	-129.923	0.74	5.14
pc-3	1810	-130.643	-129.577	1.07	3.18
aug'-cc-pVQZ	1950	-132.518	-130.612	1.91	8.17

Frame 9

Basis set	Number of Basis Functions	Raw interaction energy (kJ mol^{-1})	Counterpoised (CP) corrected interaction energy (kJ mol^{-1})	BSSE (kJ mol^{-1})	Wall time (hours)
aug'-cc-pVDZ	470	-151.446	-137.902	13.54	0.16
aug-cc-pVDZ	570	-150.427	-138.202	12.22	0.26
def2-TZVPPD	1016	-143.171	-141.382	1.79	0.75
aug'-cc-pVTZ	1040	-147.087	-143.966	3.12	0.98
aug-cc-pVTZ	1265	-147.468	-143.665	3.80	2.31
def2-QZVPPD	1806	-141.914	-141.383	0.53	5.18
pc-3	1810	-142.045	-141.136	0.91	2.95
aug'-cc-pVQZ	1950	-143.508	-141.953	1.55	7.38

Frame 19

Basis set	Number of Basis Functions	Raw interaction energy (kJ mol^{-1})	Counterpoised (CP) corrected interaction energy (kJ mol^{-1})	BSSE (kJ mol^{-1})	Wall time (hours)
aug'-cc-pVDZ	470	-163.952	-150.673	13.28	0.15
aug-cc-pVDZ	570	-164.204	-151.483	12.72	0.28
def2-TZVPPD	1016	-155.9	-154.07	1.83	0.76
aug'-cc-pVTZ	1040	-159.28	-156.033	3.25	0.92
aug-cc-pVTZ	1265	-159.586	-155.811	3.77	2.01
def2-QZVPPD	1806	-154.589	-154.072	0.52	5.58
pc-3	1810	-154.595	-153.787	0.81	3.07
aug'-cc-pVQZ	1950	-156.189	-154.564	1.62	8.09

Table S4. Average BSSE and wall time of M06-2X in various basis sets, in ZGLY-10H_2O clusters

Frame 0

Basis set	Number of Basis Functions	Raw interaction energy (kJ mol^{-1})	Counterpoised (CP) corrected interaction energy (kJ mol^{-1})	BSSE (kJ mol^{-1})	Wall time (hours)
aug'-cc-pVDZ	470	-282.389	-265.706	16.68	0.20
aug-cc-pVDZ	570	-282.452	-266.884	15.57	0.31
def2-TZVPPD	1016	-274.359	-271.983	2.38	1.23
aug'-cc-pVTZ	1040	-278.23	-274.552	3.68	1.09
aug-cc-pVTZ	1265	-278.917	-274.634	4.28	2.92
def2-QZVPPD	1806	-273.556	-272.504	1.05	6.95
pc-3	1810	-273.018	-271.782	1.24	3.73
aug'-cc-pVQZ	1950	-274.764	-272.889	1.88	16.67

Frame 9

Basis set	Number of Basis Functions	Raw interaction energy (kJ mol^{-1})	Counterpoised (CP) corrected interaction energy (kJ mol^{-1})	BSSE (kJ mol^{-1})	Wall time (hours)
aug'-cc-pVDZ	470	-256.489	-239.089	17.40	0.16
aug-cc-pVDZ	570	-254.48	-240.479	14.00	0.26
def2-TZVPPD	1016	-248.063	-245.727	2.34	0.91
aug'-cc-pVTZ	1040	-250.735	-247.057	3.68	0.82
aug-cc-pVTZ	1265	-251.626	-247.206	4.42	2.79
def2-QZVPPD	1806	-247.22	-246.083	1.14	6.84
pc-3	1810	-246.735	-245.497	1.24	3.15
aug'-cc-pVQZ	1950	-248.594	-246.556	2.04	8.54

Frame 19

Basis set	Number of Basis Functions	Raw interaction energy (kJ mol^{-1})	Counterpoised (CP) corrected interaction energy (kJ mol^{-1})	BSSE (kJ mol^{-1})	Wall time (hours)
aug'-cc-pVDZ	470	-274.615	-256.281	18.33	0.16
aug-cc-pVDZ	570	-273.646	-258.096	15.55	0.26
def2-TZVPPD	1016	-264.985	-262.486	2.50	0.83
aug'-cc-pVTZ	1040	-268.454	-264.46	3.99	0.91
aug-cc-pVTZ	1265	-269.179	-264.559	4.62	2.60
def2-QZVPPD	1806	-263.885	-262.754	1.13	6.11
pc-3	1810	-263.549	-262.252	1.30	3.39
aug'-cc-pVQZ	1950	-265.289	-263.104	2.19	8.85

Table S5. Average BSSE of various DFT functionals in conjunction with aug'-cc-pVTZ, in 10-H_2O clusters over Frame 0, 9 and 19

Method	Solute = NGLY AVG. BSSE (kJ mol^{-1})	Solute = TS AVG. BSSE (kJ mol^{-1})	Solute = ZGLY AVG. BSSE (kJ mol^{-1})
ωB97M-V	3.35	3.74	4.38
ωB97X-D	3.35	3.79	4.33
M06-2X	3.04	3.34	3.78
B3LYP-D3(BJ)	3.27	3.54	3.87
B97M-rV	3.14	3.70	4.16
B97-D3(BJ)	3.30	3.85	4.09
DSD-PBEP86	16.55	18.10	23.90

Table S6. Average combined basis error (CBE = BSSE + BSIE) for M06-2X, ωB97M-V and DSD-PBEP86/aug'-cc-pVTZ interaction energies (relative to aug'-cc-pVQZ values) in 10-H_2O clusters over Frame 0, 9 and 19.

Method	Solute = NGLY AVG. CBE (kJ mol^{-1})	Solute = TS AVG. CBE (kJ mol^{-1})	Solute = ZGLY AVG. CBE (kJ mol^{-1})
M06-2X	2.64	3.29	2.92
ωB97M-V	2.66	3.12	2.91
DSD-PBEP86	3.73	3.90	4.03

As shown in Table S5, the BSSE in DSD-PBEP86 is significantly higher due to the MP2 component. Nonetheless, it appears that the combined basis set error (BSSE + BSIE), as quantified from the deviations between aug'-cc-pVTZ and CBS limit (estimated using the aug'-cc-pVQZ basis set – see Figure 3), is very similar for the different DFT methods. This suggests that the large BSSE in the DSD-PBEP86 basis set is somehow counteracted by a similarly large BSIE. Importantly, the results in Table S6 indicate that comparison between the different DFT functionals should not be affected by basis set effects.

B. Mean absolute deviation against W1X-2 for NGLY and TS

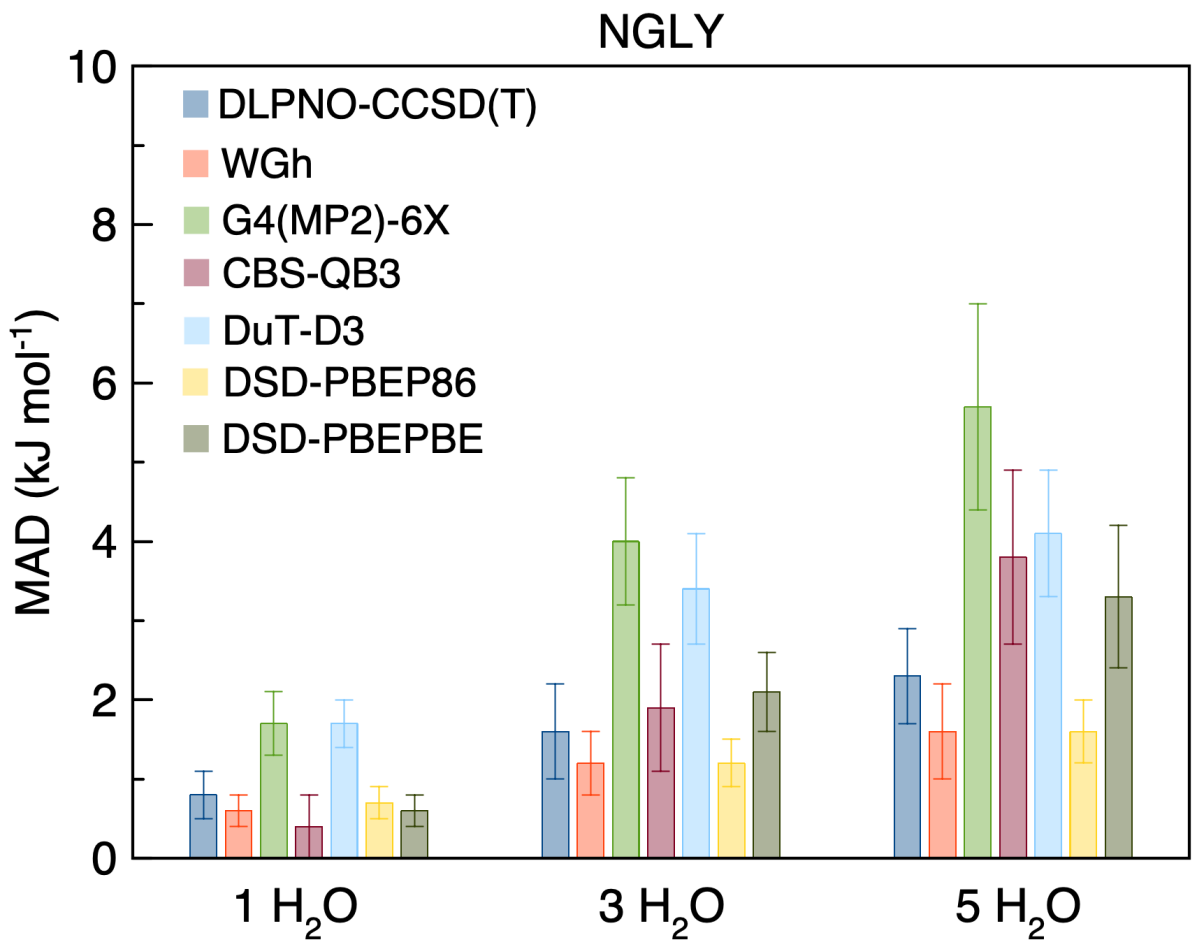

Figure S1. The mean absolute deviation in calculated interaction energies relative to W1X-2 values for clusters containing neutral glycine (NGLY) and 1, 3 and 5 solvent molecules. The error bars represent the standard deviation over 20 frames.

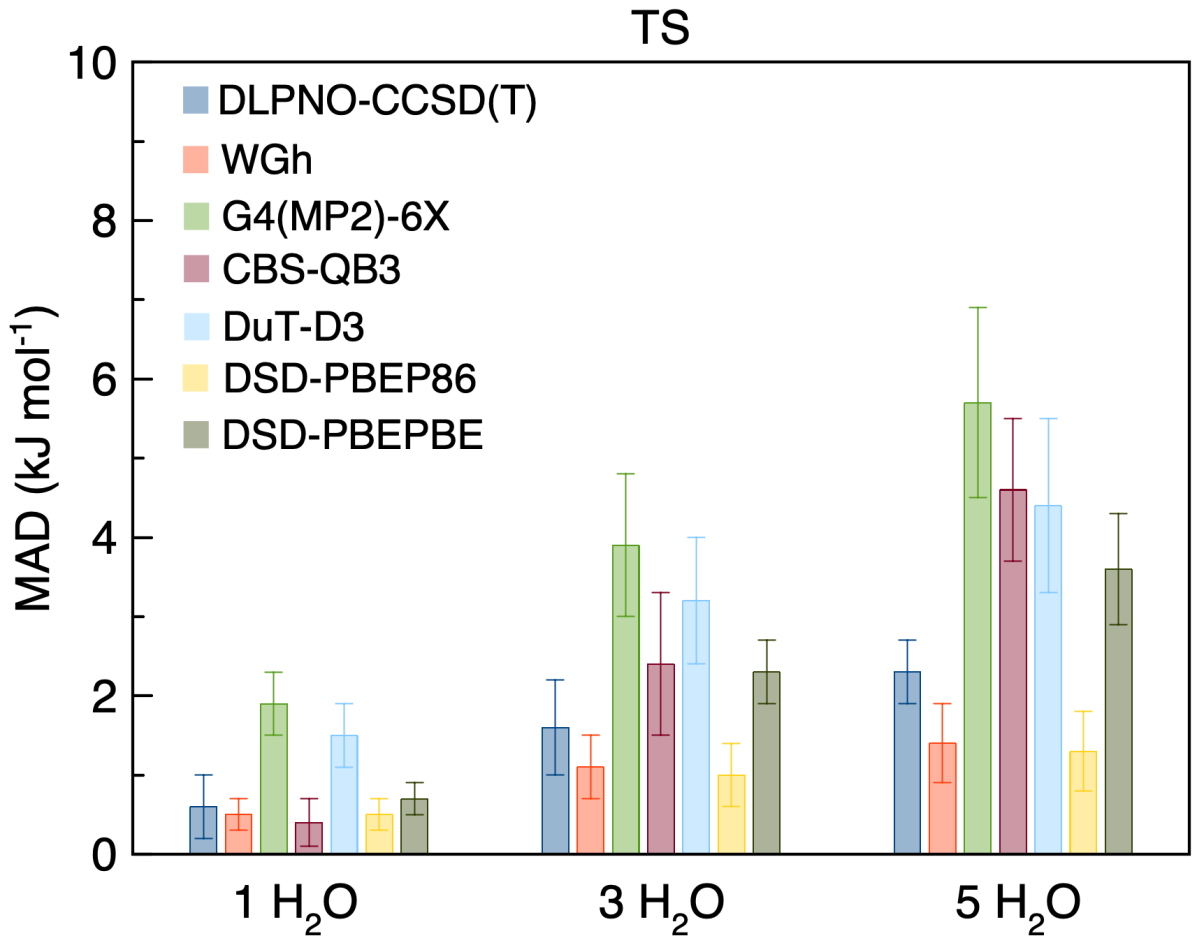

Figure S2. The mean absolute deviation in calculated interaction energies relative to W1X-2 values for clusters containing proton transfer transition state (TS) and 1, 3 and 5 solvent molecules. The error bars represent the standard deviation over 20 frames.

C. DLPNO-CCSD(T)/CBS and DSD-PBEP86/aug'-cc-pVTZ interaction energies

Table S7. DLPNO-CCSD(T)/aug-CBS(2,3) and DSD-PBEP86/aug'-cc-pVTZ interaction energies (in kJ mol^{-1}) in **1**-H_2O water clusters. The mean absolute deviation (MAD), mean signed error (MSE) and standard deviation (SD) are also shown.

	NGLY		TS		ZGLY	
Frame #	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86
0	-7.1	-7.1	-14.6	-14.2	-21.5	-20.8
1	-6.0	-5.8	-24.8	-25.0	-28.0	-27.2
2	-15.1	-15.0	-7.3	-7.1	-38.0	-38.3
3	-11.6	-11.7	-15.8	-15.5	-49.4	-48.3
4	-13.4	-13.3	-24.5	-24.4	-10.9	-9.9
5	-17.2	-17.4	-21.5	-21.7	-29.5	-28.4
6	-9.7	-9.5	-24.6	-24.9	-26.1	-25.7
7	7.5	8.4	-16.6	-16.5	-22.7	-21.8
8	-14.2	-13.8	-11.1	-10.5	-21.1	-20.0
9	-11.7	-11.5	-11.8	-10.9	-23.6	-22.9
10	3.8	4.5	-11.0	-10.9	-11.7	-11.1
11	-8.8	-8.3	-14.7	-14.1	-36.7	-36.4
12	-12.2	-12.3	-25.3	-25.6	-32.0	-31.8
13	-1.7	-1.4	-4.2	-3.6	-32.7	-31.9
14	-18.0	-17.5	-21.2	-21.3	-29.5	-29.0
15	-20.8	-20.8	-15.4	-15.6	-18.4	-17.8
16	-15.2	-15.6	-21.9	-21.4	-38.3	-37.7
17	-10.1	-9.7	-29.8	-30.3	-24.4	-23.3
18	-21.4	-21.6	-21.7	-21.6	-32.4	-31.9
19	19.4	-19.4	-31.6	-31.9	-37.0	-37.0
MAD	0.3		0.3		0.7	
MSE	0.2		0.1		0.6	
SD	0.3		0.4		0.4	

Table S8. DLPNO-CCSD(T)/aug-CBS(2,3) and DSD-PBEP86/aug'-cc-pVTZ interaction energies (in kJ mol^{-1}) in **3**-H_2O water clusters. The mean absolute deviation (MAD), mean signed error (MSE) and standard deviation (SD) are also shown.

	NGLY		TS		ZGLY	
Frame #	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86
0	-40.8	-39.8	-44.0	-42.6	-76.4	-75.1
1	-21.2	-19.9	-72.5	-72.4	-94.2	-92.5
2	-19.5	-18.7	-48.5	-48.2	-115.4	-113.7
3	-42.1	-42.3	-47.4	-46.1	-101.8	-99.7
4	-33.7	-33.2	-73.0	-72.5	-70.6	-68.6
5	-43.3	-43.4	-73.1	-73.2	-83.5	-81.8
6	-25.9	-25.7	-53.8	-53.4	-95.5	-93.9
7	-21.9	-20.7	-63.0	-62.8	-69.9	-67.7
8	-26.4	-25.6	-40.2	-38.5	-94.3	-93.2
9	-49.1	-48.9	-58.9	-58.1	-86.5	-85.8
10	-28.9	-28.4	-38.4	-37.3	-51.9	-49.8
11	-38.4	-38.0	-59.4	-58.9	-78.1	-76.5
12	-34.2	-34.2	-54.3	-53.1	-71.7	-69.8
13	-30.9	-30.2	-48.3	-47.3	-108.7	-107.8
14	-38.5	-37.6	-50.3	-49.5	-73.4	-72.2
15	-50.5	-50.4	-55.7	-54.5	-60.4	-59.2
16	-33.4	-33.7	-44.6	-43.8	-98.2	-97.6
17	-31.5	-30.8	-52.8	-52.4	-86.1	-85.3
18	-46.6	-46.3	-31.6	-32.6	-100.0	-98.8
19	-34.1	-34.8	-75.4	-74.9	-76.4	-75.2
MAD	0.5		0.8		1.4	
MSE	0.4		0.7		1.4	
SD	0.5		0.6		0.5	

Table S9. DLPNO-CCSD(T)/aug-CBS(2,3) and DSD-PBEP86/aug'-cc-pVTZ interaction energies (in kJ mol^{-1}) in **5**-H_2O water clusters. The mean absolute deviation (MAD), mean signed error (MSE) and standard deviation (SD) are also shown

	NGLY		TS		ZGLY	
Frame #	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86
0	-53.9	-52.3	-79.0	-77.6	-142.9	-140.9
1	-58.1	-57.2	-118.0	-117.6	-138.6	-135.6
2	-50.7	-49.7	-106.1	-106.1	-182.3	-180.1
3	-70.1	-70.2	-79.6	-78.0	-155.6	-153.1
4	-63.2	-62.4	-119.6	-119.2	-131.2	-128.0
5	-65.6	-65.4	-89.1	-88.4	-157.1	-156.0
6	-41.5	-40.8	-94.6	-94.2	-169.8	-168.3
7	-64.2	-63.2	-111.1	-110.5	-127.5	-124.2
8	-48.5	-47.0	-69.6	-68.0	-146.1	-144.4
9	-45.8	-45.5	-98.7	-98.2	-162.2	-160.8
10	-56.4	-55.9	-80.8	-79.4	-109.9	-107.2
11	-46.8	-45.9	-98.2	-97.6	-136.2	-134.8
12	-61.0	-60.3	-108.8	-107.8	-107.4	-104.8
13	-49.8	-48.7	-63.0	-61.7	-155.3	-153.1
14	-62.2	-61.2	-106.9	-106.1	-123.9	-122.1
15	-74.8	-74.6	-104.0	-103.3	-113.3	-112.1
16	-35.5	-34.8	-82.1	-80.6	-144.1	-142.7
17	-65.4	-64.4	-66.8	-65.9	-143.3	-140.9
18	-68.3	-67.2	-68.3	-66.3	-156.1	-154.7
19	-56.0	-55.8	-114.1	-112.6	-152.1	-151.0
MAD	0.8		1.0		2.0	
MSE	0.8		1.0		2.0	
SD	0.5		0.5		0.7	

Table S10. DLPNO-CCSD(T)/aug-CBS(2,3) and DSD-PBEP86/aug'-cc-pVTZ interaction energies (in kJ mol^{-1}) in **10**-H_2O water clusters. The mean absolute deviation (MAD), mean signed error (MSE) and standard deviation (SD) are also shown

	NGLY		TS		ZGLY	
Frame #	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86
0	-80.6	-80.0	-133.3	-132.1	-269.4	-267.5
1	-77.5	-76.7	-144.1	-144.0	-253.5	-251.6
2	-41.4	-39.1	-100.2	-99.4	-298.7	-296.3
3	-92.4	-92.8	-123.3	-120.8	-262.2	-258.6
4	-62.8	-61.7	-168.0	-168.0	-286.6	-284.1
5	-82.5	-81.7	-167.0	-165.7	-296.6	-295.8
6	-75.8	-75.3	-116.9	-116.6	-344.0	-343.1
7	-99.6	-98.7	-140.8	-139.5	-277.3	-274.6
8	-57.1	-55.1	-120.3	-119.1	-297.0	-295.3
9	-70.0	-69.4	-149.6	-149.5	-253.9	-250.2
10	-62.4	-61.6	-106.0	-105.7	-277.8	-274.6
11	-62.8	-61.2	-125.3	-124.6	-260.7	-259.4
12	-59.3	-58.2	-179.6	-179.4	-219.3	-216.4
13	-69.6	-69.2	-132.0	-130.6	-264.6	-262.8
14	-78.5	-77.8	-166.7	-165.9	-246.7	-244.3
15	-84.3	-82.9	-110.1	-107.3	-240.3	-239.2
16	-95.0	-94.4	-109.9	-107.9	-216.3	-214.3
17	-91.9	-91.0	-151.6	-150.6	-273.5	-271.6
18	-67.3	-66.2	-126.9	-125.0	-296.6	-295.3
19	-81.5	-80.7	-155.8	-154.6	-269.4	-267.5
MAD	1.0		1.1		2.1	
MSE	0.9		1.1		2.1	
SD	0.6		0.8		0.9	

Table S11. DLPNO-CCSD(T)/aug-CBS(2,3) and DSD-PBEP86/aug'-cc-pVTZ interaction energies (in kJ mol^{-1}) in **20**-H_2O water clusters. The mean absolute deviation (MAD), mean signed error (MSE) and standard deviation (SD) are also shown

	NGLY		TS		ZGLY	
Frame #	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86
0	-93.8	-93.3	-211.1	-211.8	-314.4	-314.3
1	-123.9	-125.7	-204.2	-206.3	-308.3	-307.9
2	-105.3	-105.1	-145.7	-145.7	-323.9	-323.4
3	-151.9	-152.8	-184.2	-183.6	-303.2	-300.2
4	-85.9	-85.4	-187.9	-188.7	-348.7	-347.1
5	-128.6	-129.9	-183.1	-182.7	-345.3	-346.1
6	-73.4	-73.6	-134.2	-136.5	-426.2	-427.2
7	-141.1	-141.5	-180.7	-181.1	-259.3	-257.3
8	-99.8	-99.9	-127.8	-127.5	-403.9	-403.9
9	-60.8	-61.3	-157.8	-158.9	-409.8	-408.3
10	-95.2	-95.4	-114.4	-114.2	-318.0	-316.9
11	-92.6	-92.0	-129.8	-129.9	-365.7	-366.3
12	-118.4	-118.6	-175.0	-176.8	-327.9	-327.3
13	-98.1	-99.0	-141.1	-140.7	-403.5	-404.0
14	-128.6	-129.3	-168.5	-169.0	-296.4	-294.9
15	-123.2	-123.4	-127.3	-125.5	-289.8	-290.7
16	-114.9	-115.0	-164.4	-164.0	-349.1	-348.8
17	-90.4	-89.5	-141.1	-139.9	-439.6	-440.4
18	-102.4	-103.7	-132.5	-133.0	-361.4	-363.2
19	-107.6	-108.1	-249.7	-251.3	-379.3	-378.7
MAD	0.6		0.9		1.0	
MSE	0.3		0.3		0.3	
SD	0.7		1.1		1.2	

Table S12. DSD-PBEP86/aug'-cc-pVTZ interaction energies (in kJ mol^{-1}) in **40**-H_2O water clusters.

	NGLY		TS		ZGLY	
Frame #	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86	DLPNO-CCSD(T)	DSD-PBEP86
0	-139.9	-141.8	-	-282.9	-	-430.5
1	-	-168.1	-	-222.0	-	-416.4
2	-	-135.3	-	-209.2	-	-376.4
3	-	-169.0	-	-203.8	-	-394.0
4	-	-115.6	-	-270.6	-	-430.0
5	-	-152.1	-	-234.9	-	-444.5
6	-	-118.9	-	-206.1	-	-486.6
7	-	-168.8	-	-201.9	-	-337.5
8	-	-158.7	-	-180.8	-	-528.3
9	-	-97.8	-	-197.8	-	-427.3
10	-	-151.0	-	-135.2	-	-427.8
11	-	-144.0	-	-180.1	-	-470.7
12	-	-132.2	-	-236.4	-	-426.6
13	-	-140.5	-	-248.7	-	-495.8
14	-	-134.9	-	-183.8	-	-327.2
15	-	-127.4	-	-202.9	-	-365.7
16	-	-171.3	-	-208.4	-	-362.3
17	-	-135.5	-	-247.8	-	-534.6
18	-	-127.6	-	-189.2	-	-399.5
19	-	-135.9	-	-278.0	-	-407.3

D. Mean absolute percentage deviation (MAPD) against DSD-PBEP86/aug'-cc-pVTZ

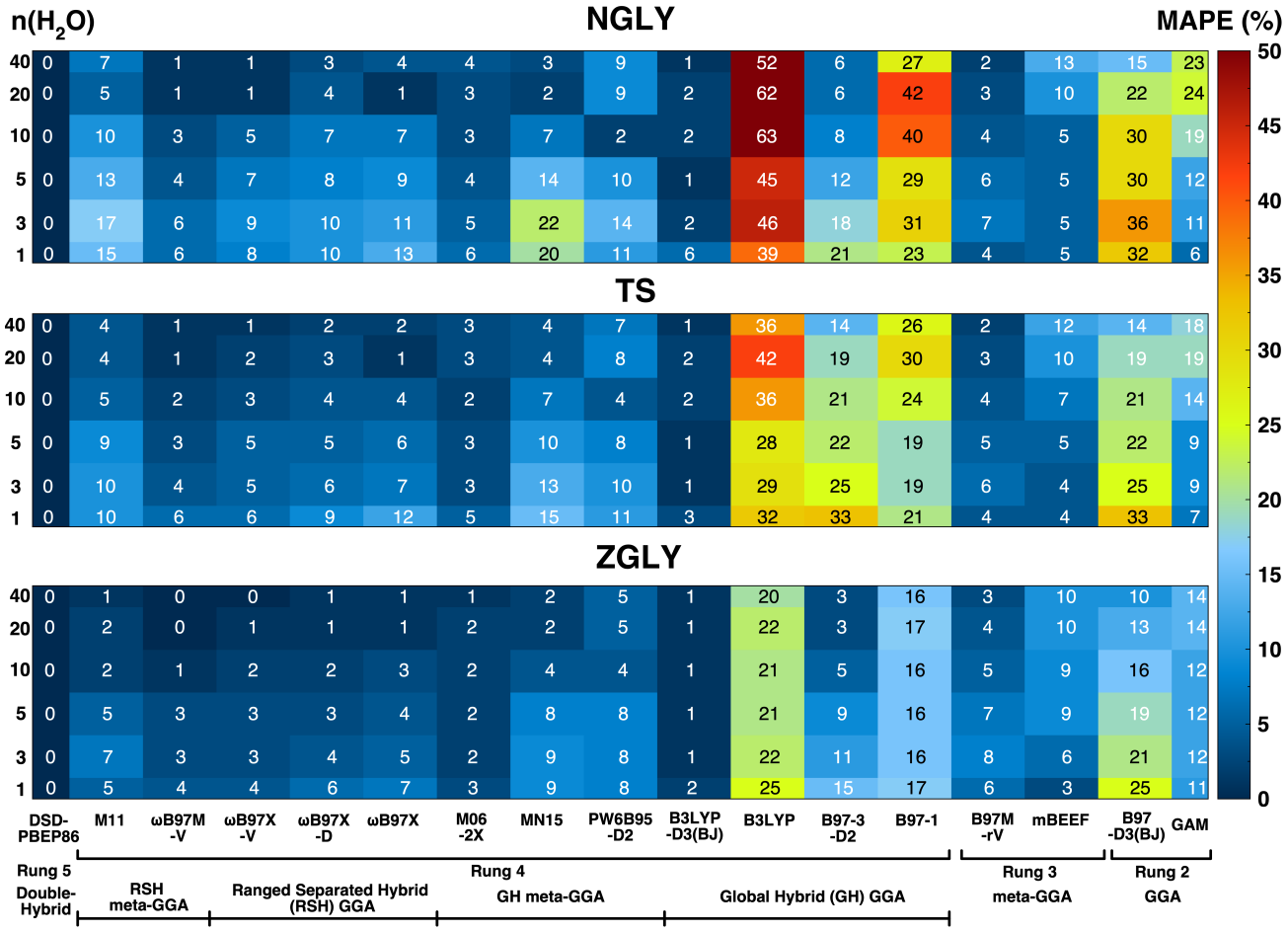

Figure S3. Mean absolute percentage deviation (MAPD) of DFT/ aug'-cc-pVTZ absolute interaction energies relative to DSD-PBEP86/aug'-cc-pVTZ values averaged over 20 different solvent configurations. n(H_2O) denotes for the number of water molecules in the glycine-water complex.

E. Energy Decomposition Analysis (EDA) for B3LYP-D3(BJ)/aug'-cc-pVTZ

In ALMO-EDA2, interaction energy is decomposed into permanent and induced electrostatics (ELEC), Pauli repulsions (PAULI), dispersion (DISP), polarization (POL), and charge transfer (CT). Frozen potential (FRZ) is introduced as the sum from ELEC and PAULI since Pauli repulsions have the opposite sign of other contributions.

In Figure S4(a), there is a systematic trend that the FRZ component of the interaction energy decreases from 32% to 11%, which agrees with the state of the solute varying from a neutral to a charged state. In turn, the POL and CT components increase from 18% and 18% to 35% and 25% for NGLY and ZGLY, respectively. The dispersion remains unaffected by the variation of solute given that the size of the water cluster is the same. In Figure S4(b), the solute state is fixed as ZGLY, where the water cluster size is increasing from 5 to 20. As a result, the FRZ also responds to the change as the ratio decrease from 26% to 11% from $(H_2O)_5$ to $(H_2O)_{20}$, respectively. While CT component remains mostly unaffected, both POL and DISP display an increasing trend as water cluster grows. This is also observed for other states of solute (see Table S13-16). The EDA analysis has shown that for B3LYP-D3(BJ), the dispersion component continues to grow while electrostatic decreases as a result of increasing cluster size. This might have shed some light in elucidating the comparison between B3LYP and B3LYP-D3(BJ) that a proper contribution of dispersion is essential for large cluster systems.

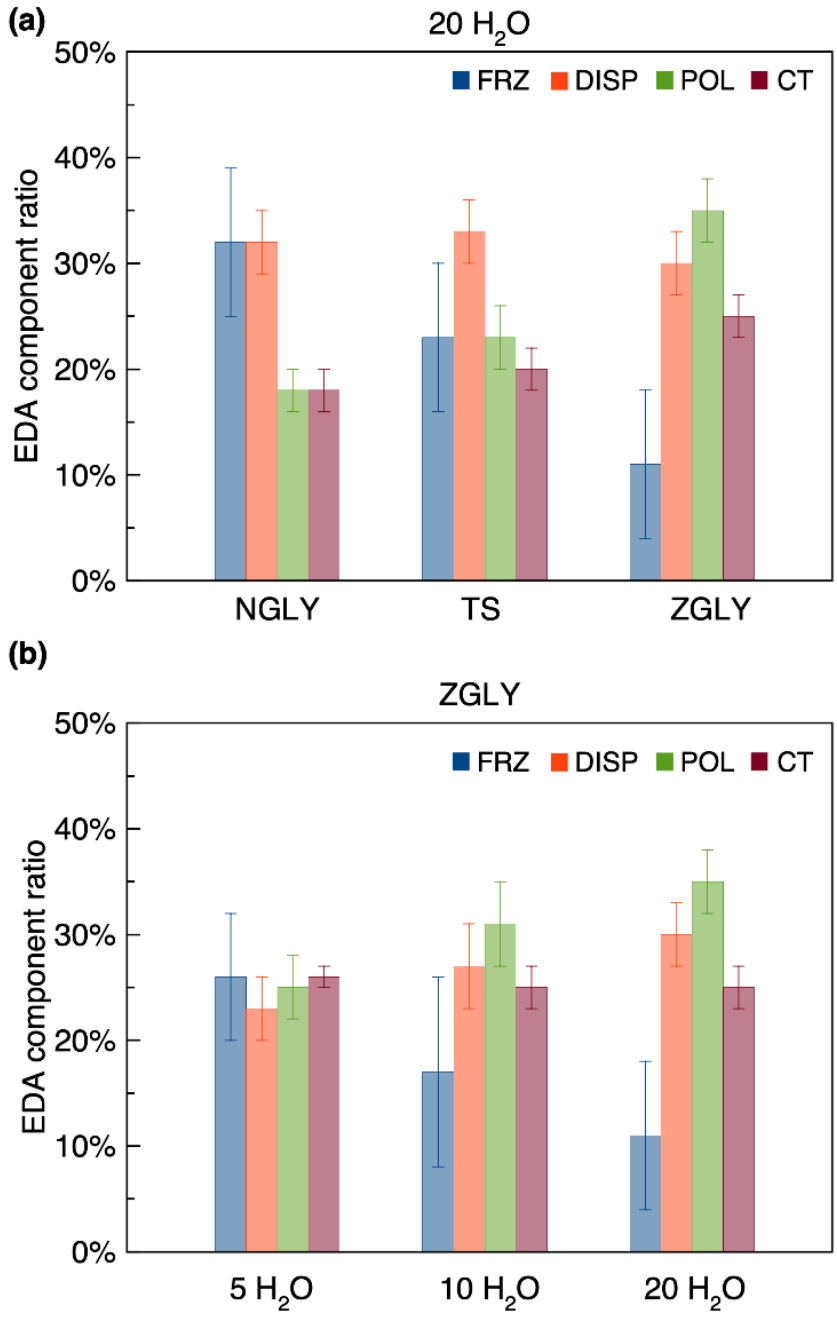

Figure S4. The EDA component ratio with standard deviation in terms of magnitude in interaction energy of the sum of all contributions, where (a) shows for $(H_2O)_{20}$ clusters the energy decomposition for each state of the solute, NGLY, TS ,and ZGLY, and (b) displays for ZGLY energy decomposition for each size of the cluster, 5, 10 ,and 20 -water clusters. All ratio is an average of 20 snapshots in B3LYP-D3(BJ)/aug'-cc-pVTZ level of theory.

Table S13. EDA of each component of the interaction energy for **ZGLY** in B3LYP-D3(BJ)/aug'-cc-pVTZ. The units are in kJ mol-1. ELEC stand for electrostatic, PAULI for Pauli repulsion, DISP for dispersion, POL for polarization, CT for charge-transfer, and TOTAL for total interaction energy. The energies reflect an average over 20 snapshots.

Cluster size	ELEC	PAULI	DISP	POL	CT	TOTAL
$(H_2O)_5$	-581.2	666.1	-68.1	-77.6	-78.3	-139.1
$(H_2O)_{10}$	-848.9	921.5	-108.3	-125.8	-103.0	-264.6
$(H_2O)_{20}$	-916.5	961.4	-130.2	-152.0	-108.2	-345.6

Table S14. EDA of each component of the interaction energy for **NGLY** in B3LYP-D3(BJ)/aug'-cc-pVTZ. The units are in kJ mol-1. ELEC stand for electrostatic, PAULI for Pauli repulsion, DISP for dispersion, POL for polarization, CT for charge-transfer, and TOTAL for total interaction energy. The energies reflect an average over 20 snapshots.

Cluster size	ELEC	PAULI	DISP	POL	CT	TOTAL
$(H_2O)_5$	-349.8	431.8	-53.0	-39.2	-46.1	-56.2
$(H_2O)_{10}$	-412.5	513.5	-76.0	-45.1	-52.4	-72.5
$(H_2O)_{20}$	-445.0	542.0	-94.0	-53.5	-55.1	-105.6

Table S15. EDA of each component of the interaction energy for **TS** in B3LYP-D3(BJ)/aug'-cc-pVTZ. The units are in kJ mol-1. ELEC stand for electrostatic, PAULI for Pauli repulsion, DISP for dispersion, POL for polarization, CT for charge-transfer, and TOTAL for total interaction energy. The energies reflect an average over 20 snapshots.

Cluster size	ELEC	PAULI	DISP	POL	CT	TOTAL
$(H_2O)_5$	-397.2	458.3	-54.6	-47.2	-50.9	-91.6
$(H_2O)_{10}$	-515.4	589.5	-83.3	-63.0	-61.0	-133.3
$(H_2O)_{20}$	-543.5	621.0	-102.5	-72.3	-63.9	-161.2

F. ONIOM approximation for NGLY-$(H_2O)_{40}$

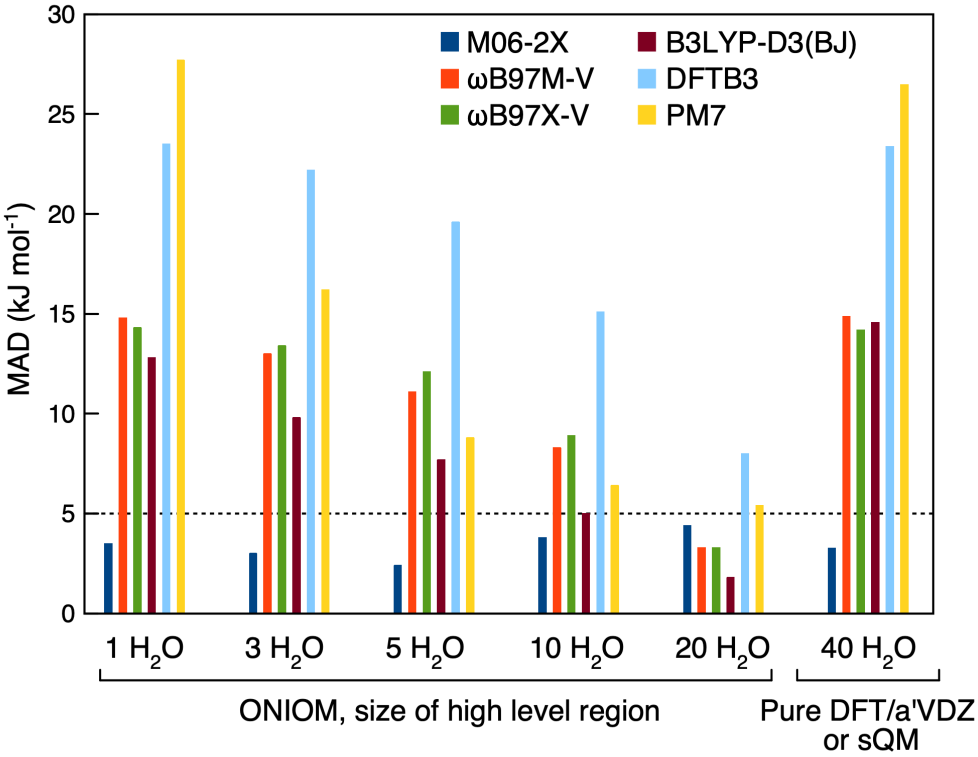

Figure S5. The MAD values for the ONIOM(DSD-PBEP86/aug'-cc-pVTZ:X/aug'-cc-pVDZ) absolute interaction energies (X=M06-2X, ωB97M-V, ωB97X-V, B3LYP-D3(BJ), DFTB3 and PM7) compared to the corresponding DSD-PBEP86/aug'-cc-pVTZ values on the **NGLY**-$(H_2O)_{40}$ cluster.

Table S16. Isodesmic reaction energies (ΔE_{iso} in Scheme 1) relative to DSD-PBEP86/aug'-cc-pVTZ values determined and various "lower-level" theories DFT/aug'-cc-pVDZ and sQM) in a 40-H_2O cluster. (N and n = 40 and 1 respectively).

LL =	DSD-PBEP86	M06-2X	ωB97M-V	ωB97X-V	B3LYP-D3(BJ)	DFTB3	PM7
NGLY (MAD)	0.0	3.5	14.9	14.4	12.8	23.4	27.8
NGLY (MSD)[a]	0.0	-1.2	-14.9	-14.4	-12.8	23.4	-27.6
NGLY (STDEV)	0.0	4.1	1.9	1.3	1.4	8.2	16.4
ZGLY (MAD)	0.0	9.4	19.8	17.3	15.6	11.8	27.9
ZGLY (MSD)[a]	0.0	-9.4	-19.8	-17.3	-15.6	9.8	-27.6
ZGLY (STDEV)	0.0	5.2	2.6	2.4	1.9	10.5	16.3

[a] MSD = mean signed deviation.