

Supplementary Material to
 Structure, Dynamics and Hydration Free Energy of Carbon Dioxide in
 Aqueous Solution: A Quantum Mechanical/Molecular Mechanics
 Molecular Dynamics Thermodynamic Integration (QM/MM MD TI)
 Simulation Study
 Niko Prasetyo & Thomas S. Hofer

S.I Choice of the Correlated Ab Initio Level of Theory

Basis set and level of theory are crucial choices in the setup of a QM/MM MD simulation. A number of test calculations using the 6-311++G**¹ and 6-31G**² for CO₂ and H₂O as well as the cc-pVDZ and cc-pVTZ basis sets³ applied to all atoms of the system were carried out at different levels of theory, performing energy minimisations of small carbon dioxide-water cluster CO₂(H₂O)_n ($n = 1 - 3$) in gas phase using the software package Gaussian16.⁴ The results of these calculations at frozen-core (FC) as well as all-electron (AE) MP2 and CCSD level are summarised in Tab. S.1. The comparison shows that MP2 combined with 6-311++G** and 6-31G** basis sets deliver results close to the most sophisticated correlated level considered being all-electron CCSD/cc-pVTZ, the respective deviation being in the range of 1.5 kcal/mol and below. Based on the presented data, the MP2 level combined with 6-311++G**¹ and 6-31G**² for carbon dioxide and water provides an adequate compromise between computational effort and accuracy of results. Screenshots of the CO₂(H₂O)_n ($n = 1 - 3$) clusters obtained at CCSD(AE)/cc-pVTZ level are depicted in fig. S.1.

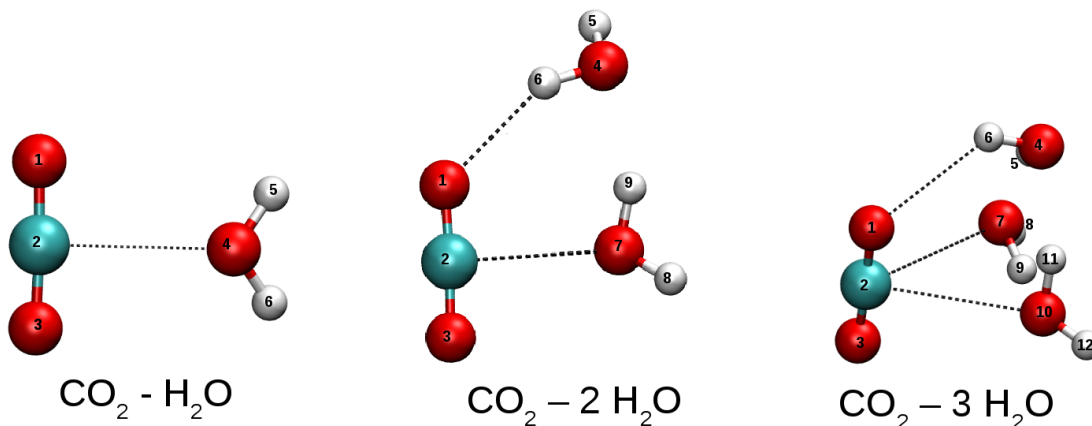


Figure S.1: Screenshots of the optimised carbon dioxide clusters CO₂(H₂O)_n ($n = 1 - 3$) at all-electron CCSD/cc-pVTZ level.

S.II Choice of Population Analysis Scheme

The reason to chose Mulliken population analysis to derive the partial charges of all atoms in the QM-region is solely based on practical considerations. According to Martin and Zipse,⁵ who investigated the charge distribution of an isolated water molecule using different combinations of theoretical level, basis set and population analysis scheme, the MP2/6-31G** level yields an overall dipole moment of 1.81 D in very good agreement with the experimental value of 1.86 D⁶ in the Mulliken case (see Tab. 3 in Ref.⁵).

MP2(FC)/GEN			MP2(AE)/GEN		CCSD(AE)/GEN	
n	E_{bind}/n	$r_{\text{C-O}}$	E_{bind}/n	$r_{\text{C-O}}$	E_{bind}/n	$r_{\text{C-O}}$
1	-3.65	2.72	-3.18	2.72	-3.26	2.72
2	-6.43	2.70	-5.97	2.70	-5.76	2.71
3	-8.89	2.90	-8.43	2.90	-7.91	2.91

MP2(FC)/cc-PVDZ			MP2(AE)/cc-PVDZ		CCSD(AE)/cc-PVDZ	
n	E_{bind}/n	$r_{\text{C-O}}$	E_{bind}/n	$r_{\text{C-O}}$	E_{bind}/n	$r_{\text{C-O}}$
1	-3.77	2.78	-4.58	2.78	-3.75	2.78
2	-7.25	2.75	-7.79	2.75	-6.89	2.75
3	-9.86	2.95	-10.23	2.95	-9.15	2.95

MP2(FC)/cc-PVTZ			MP2(AE)/cc-PVTZ		CCSD(AE)/cc-PVTZ	
n	E_{bind}/n	$r_{\text{C-O}}$	E_{bind}/n	$r_{\text{C-O}}$	E_{bind}/n	$r_{\text{C-O}}$
1	-3.17	2.78	-3.23	2.78	-3.16	2.78
2	-5.91	2.77	-6.11	2.77	-5.73	2.78
3	-8.05	3.00	-8.30	2.95	-7.60	3.01

Table S.1: Average binding energy per ligand molecule E_{bind}/n in kcal/mol and average C-O_{wat} bond distance $r_{\text{C-O}}$ in Å obtained for energy minimised CO₂(H₂O)_n clusters (n=1,2 and 3) at MP2 and CCSD(AE) level in the gas-phase. The term FC and AE denotes frozen-core and all-electron respectively, while GEN represents the assignment of the 6-311++G** basis set for CO₂ and 6-31G** for H₂O, respectively.

While atomic partial charges based on the electrostatic surface potential (*e.g.* restrained electrostatic potential charges, RESP; Merz-Kollman, MK; Charges from electrostatic potentials based on a grid, CHELPG) are very popular and commonly used to derive charges for individual molecules, it is known that these methods are less suitable in case of molecular clusters.^{7,8} This is due to the fact that the surface-to-volume ratio decreases unfavourably with increasing cluster size, and it has been discussed that the data encoded on the electrostatic potential surface does not provide sufficient information to represent the electron density in the center of the cluster.^{7,8} Thus, despite MK, RESP and CHELPG charges appearing to be a suitable alternative at the MP2/(aug)-cc-pVnZ level (Tab. 3 in Ref.⁵), the short-comings of the ESP strategy prevents an application to the QM zone of the simulation consisting of CO₂ plus several water molecules.

On the other hand partial charges obtained via natural population analysis (NPA) as well as the atom-in-molecule (AIM) approach are already too high (Tab. 3 in Ref.⁵), leading to an overestimation of the dipole moment of an isolated water molecule of 1.86 D⁶ over the whole range of basis sets considered. As the dipole moment in liquid water is even higher compared to an isolated water molecule in the gaseous state, it can be expected that the deviations in the NPA and AIM schemes are further amplified in condensed phase due to the presence of hydrogen bonds and solute-solvent interactions.

Since it is imperative in QM/MM simulations to uphold the best possible compatibility between QM-derived charges and the point charges of the MM model to avoid artefacts in the QM/MM coupling, Mulliken charges in conjunction with the MP2/6-31G** level assigned to the solvent molecules are the preferred level of theory. As the comparison of cluster energies discussed above has shown, this level provides a good compromise between accuracy of results and computational effort as well.

S.III Overall charge of CO₂

In order to evaluate whether the observed overall negative net-charge of the CO₂ molecule of $-0.23a.u$ is solely the result of Mulliken population analysis (MPA), the partial charges have been re-evaluated using alternative population analysis schemes, namely Löwdin population analysis (LPA), natural population analysis (NPA) as well as population analysis based on occupation numbers as implemented in the Turbomole package (keyword PABOON), yielding partial charges based on so-called 2-center and multicenter occupancies (referred to here as OCC-2 and OCC-M). Although Turbomole is capable of also computing ESP-based charges, this scheme has not been considered due to the short-comings discussed above.^{7,8} The AIM scheme, on the other hand, is not available in TURBOMOLE 7.0.1.

A total of 40 configurations distributed over the entire simulation trajectory for the QM/MM MD case with $\lambda=1.0$ have been considered (a re-evaluation of every single configuration of the trajectory was not feasible). The average partial charges for the carbon and oxygen atoms as well as the respective sum obtained from this evaluation are summarised in Tab. S.2.

Atom	LPA	MPA	OCC-2	NPA	OCC-M
C	-0.13 ± 0.04	0.29 ± 0.08	0.17 ± 0.22	0.95 ± 0.02	0.24 ± 0.19
O	-0.15 ± 0.05	-0.30 ± 0.07	-0.18 ± 0.06	-0.50 ± 0.03	-0.11 ± 0.09
O	-0.14 ± 0.04	-0.30 ± 0.07	-0.18 ± 0.06	-0.49 ± 0.01	-0.11 ± 0.09
Sum	-0.42 ± 0.10	-0.31 ± 0.08	-0.19 ± 0.14	-0.04 ± 0.01	0.02 ± 0.09

Table S.2: The average atomic charge of CO₂ in atomic units ($a.u$) and the respective standard deviations for 40 different configurations of the first hydration shell obtained using different partial charge schemes according to Löwdin (LPA), Mulliken (MPA), natural (NPA), and occupation number (OCC-2, OCC-M) population analysis at RIMP2-level combined with 6-311++G** basis set for CO₂ and 6-31G** basis set for water. It can be seen that the overall charges of CO₂ tends to be in a range from negative to neutral.

As typical for the evaluation of atomic partial charges the different schemes yield highly different values, resulting from the fact that partial charges are not physically observable and can thus only be employed for a relative comparison. (For instance two different compounds calculated using identical level/basis/population setups can be compared, however, the absolute magnitude of the partial charges should only be considered as trend.) It can be seen that the overall charge of CO₂ is found in the range from negative ($-0.42\pm0.10a.u$, Löwdin) to neutral ($0.02\pm0.09a.u$, OCC-M) using the different population analysis schemes. A pronounced positive charge for CO₂ is not observed. These results show that carbon dioxide in aqueous solution has a tendency towards an overall negative charge.

S.IV Structural properties of aqueous CO₂ at MM level

The radial distribution function (RDF) of C-O_{wat} pairs obtained from classical MD ($\lambda = 1$ in the charging step) is depicted in fig. S.2a. A weak double peak is observed with its maximum located at 4.11 Å. In analogy to the QM/MM MD results presented in the main article a conically restricted analysis of the RDF was carried out, by first performing an angular-radial distribution (ARD) analysis to identify the appropriate angle to separate the conical regions.⁹ Again, this procedure enables to decompose the shoulder in the first shell peak of the C-O_{wat} RDF from the main peak (see fig. S.2b and c, respectively). The associated conical intervals have been chosen based on the associated angular-radial distribution (dashed lines in fig. S.2d) in the

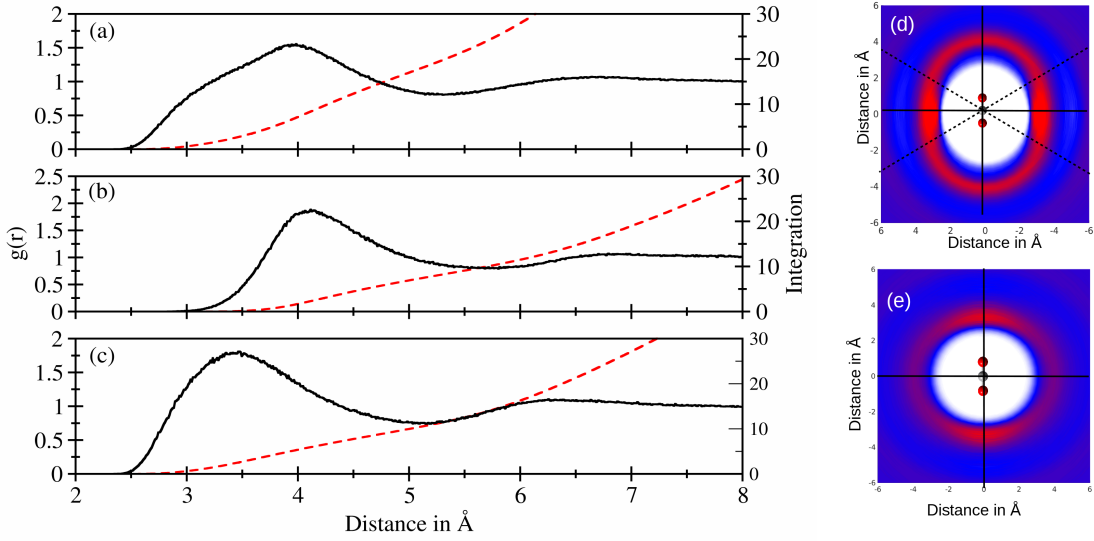


Figure S.2: C- O_{wat} pair distribution function at MM level (*i.e.* $\lambda=1.0$ of the charging path) obtained via a) a fully spherical consideration of solvent molecules and b,c) the respective decomposition employing a conically-restricted selection with respect to the molecular axis enabling a separation of the C and O_C binding regions. Angular radial distribution functions with respect to the main axis of CO_2 of d) water oxygen and e) water hydrogen atoms (Colour keys: blue = low, white = medium and red = high intensity). The respective border angle of 55° employed in the generation of conically-restricted RDFs is indicated via a dashed line in the CO_2 - O_{wat} ARD.

range of $[0^\circ, 55^\circ]$ and $[55^\circ, 90^\circ]$, respectively. The top view of the CO_2 -water ARDs (fig. S.2d and e) clearly highlight the change from the carbon-water dipole interaction to oxygen-water hydrogen bonding when moving from the normal plane of the CO_2 molecule corresponding to $y=0$ towards the molecular axis represented by $x=0$.

The chemically equivalent properties of each CO_2 oxygen atoms of the solute enabled the calculation of average $O-O_{wat}$ and $O-H_{wat}$ RDFs to characterise hydrogen bonding between solvent and solute molecules. Fig. S.3a depicts the conically restricted RDFs together with the corresponding integration number. A single peak with a maximum at $3.0 \text{ \AA}$ was observed in the $O-O_{wat}$ RDF. The hydrogen bond distance between O_C and the oxygen atoms of the solvent molecules was evaluated as $3.5 \text{ \AA}$, deduced from the maxima peak observed in $O-O_{wat}$ RDF. Based on these RDFs, the distance cutoff for searching the hydrogen bonding interaction between oxygen atoms of CO_2 and hydrogen atom of water was adopted. Integration of the $O-H_{wat}$ RDF was carried out to determine the number of bound water molecules to CO_2 , resulting in approximately 7 water molecule located in the first solvation sphere of each oxygen atoms of CO_2 .

In addition, coordination number distributions (CNDs) were calculated for the hydration shell of the entire CO_2 molecule, which is hydrated by a number of water molecules varying between 10 and 21 (average coordination number of 16.5, see Fig. S.3b). The average number of interacting hydrogen atoms within the first solvation sphere of each oxygen atoms of CO_2 was evaluated as 6.9 (Fig. S.3c). A cut-off distance of $3.0 \text{ \AA}$ was employed, according from the $g(r)$ function (see Fig. S.3a).

The broad distribution visible in the $O_C \cdots H_{wat}$ - O_{wat} angular distribution shown in Fig. S.3a

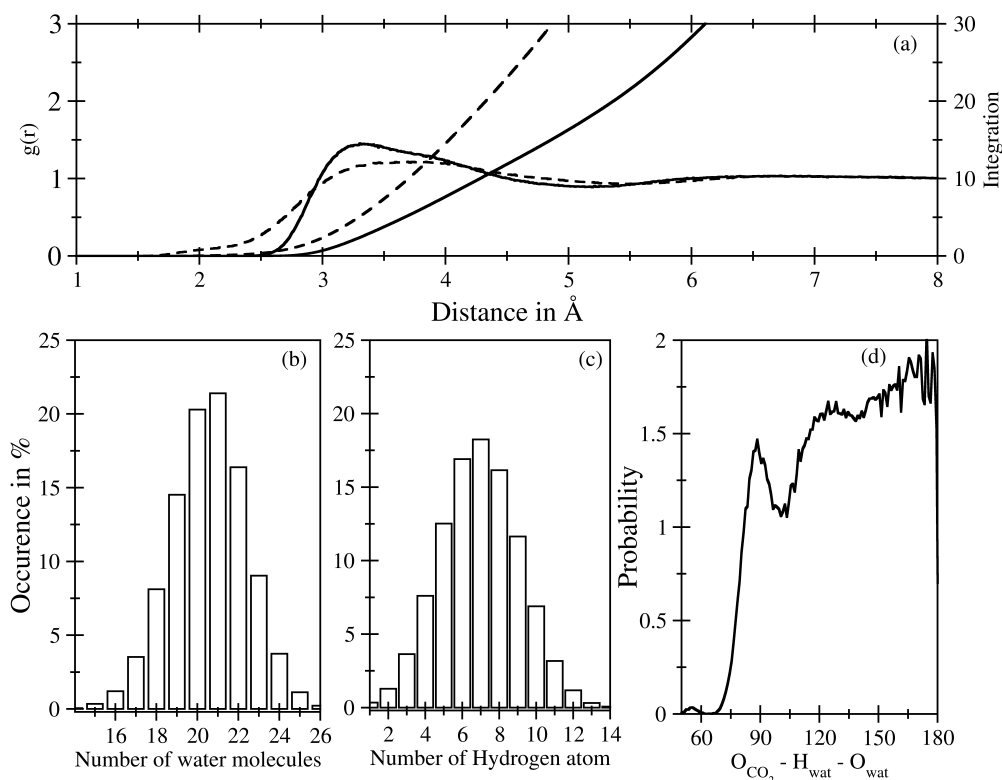


Figure S.3: Selected properties of CO₂ at MM level (*i.e.* $\lambda=1.0$ of the charging path) showing a) conically restricted O_C-water radial distribution function (RDF) for water oxygen (solid lines) and hydrogen atoms (dashed lines), respectively. Coordination number distribution of b) the entire CO₂ molecule and c) hydrogen atoms bound to each oxygen atoms of CO₂ within a cutoff radius of 3.45 Å. d) Distribution of the O_C-O_{wat}-H_{wat} hydrogen bond angle.

points towards a high degree of solvent mobility within the hydration shell of the CO₂ molecule, in line with the result from the pair and angular radial distribution analysis. The highest probability is found close to 180°, indicating a linearity of the hydrogen bonds. Nevertheless, the probability for lower angles is comparably high, pointing towards a high degree of molecular flexibility in the associated H-bonds.

The MM-based simulation shows a number of deviation from the more accurate QM/MM results, which are discussed in more detail in the main manuscript.

References

- (1) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (2) Dill, J. D.; Pople, J. A. Self-consistent molecular orbital methods. XV. Extended Gaussian-type basis sets for lithium, beryllium, and boron. *J. Chem. Phys.* **1975**, *62*, 2921–2923.
- (3) Dunning, T. Gaussian basis functions for use in molecular calculations. Contraction of (12s9p) atomic basis sets for the second row atoms. *Chem. Phys. Lett.* **1970**, *7*, 423–427.
- (4) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheese-

- man, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.
- (5) Martin, F.; Zipse, H. Charge distribution in the water molecule. A comparison of methods. *J. Comput. Chem.* **2005**, *26*, 97–105.
 - (6) Gregory, J. K.; Clary, D. C.; Liu, K.; Brown, M. G.; Saykally, R. J. The water dipole moment in water clusters. *Science* **1997**, *275*, 814–817.
 - (7) Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. Charges fit to electrostatic potentials. II. Can atomic charges be unambiguously fit to electrostatic potentials? *J. Comput. Chem.* **1996**, *17*, 367–383.
 - (8) Rigby, J.; Izgorodina, E. I. Assessment of atomic partial charge schemes for polarisation and charge transfer effects in ionic liquids. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1632–1646.
 - (9) Weiss, A. K.; Hofer, T. S. Exploiting the capabilities of quantum chemical simulations to characterise the hydration of molecular compounds. *RSC Adv.* **2013**, *3*, 1606–1635.