Supplementary Information: Title

Joseph Heindel[1,2], Teresa Head-Gordon[1,2,3]

[1]Kenneth S. Pitzer Theory Center and Department of Chemistry
[2]Chemical Sciences Division, Lawrence Berkeley National Laboratory
[3]Departments of Bioengineering and Chemical and Biomolecular Engineering
University of California, Berkeley, CA, USA
corresponding author: thg@berkeley.edu

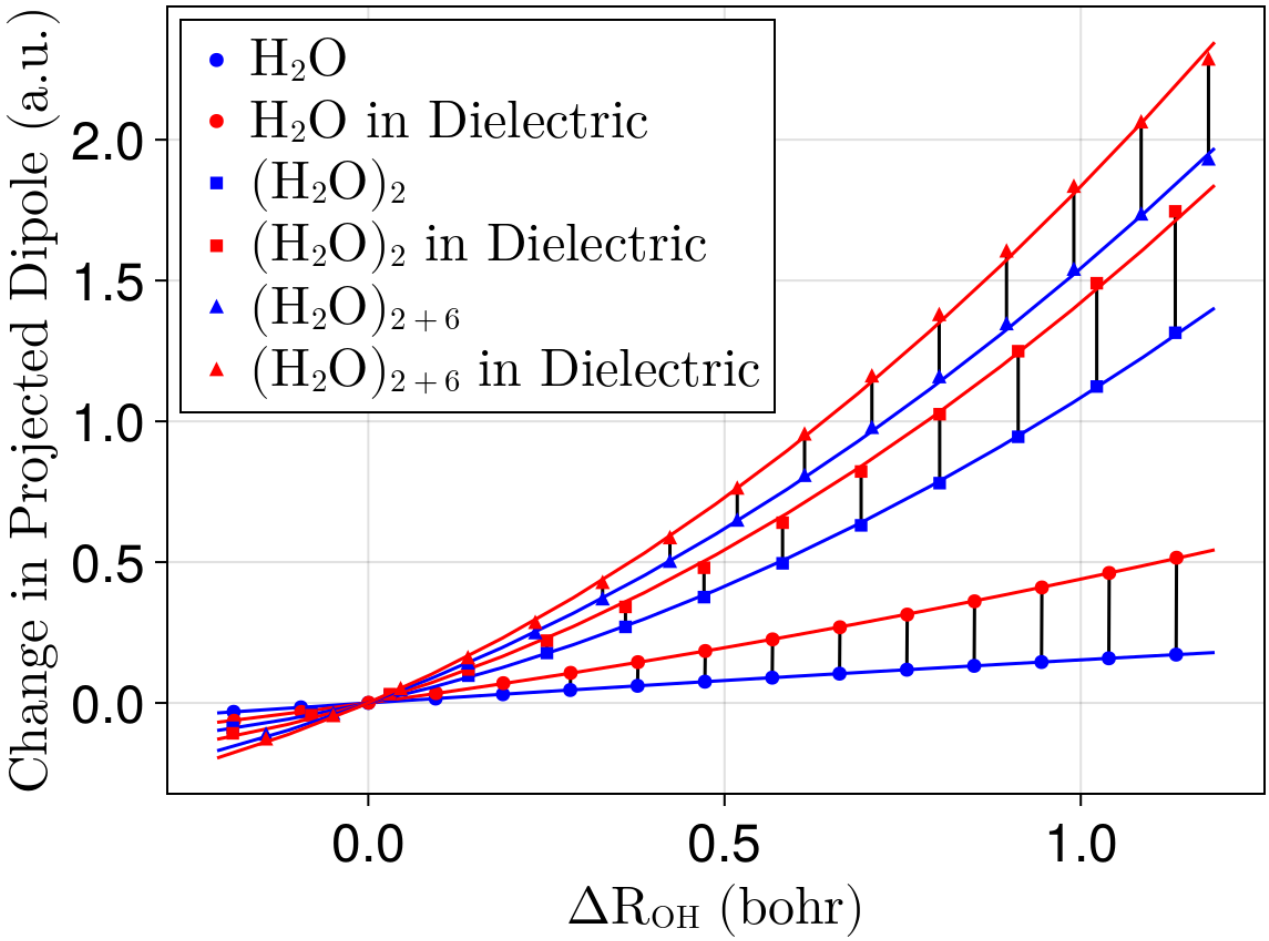
In the course of developing this model, we found that we were unable to reproduce the expected correlation between change in bond length and change in harmonic frequency for O−H stretches.[1] We ultimately determined that the failure of our model to reproduce this correlation was due to a lack of coupling between our bonding potential and the environment. To that end, we extended our morse potential to be field-dependent in a manner first described elsewhere.[1] In short, the field-dependence of the potential requires specifying the first and second dipole derivatives along the O−H stretch.

Rather than treating these as fitting parameters, we computed the dipole derivatives from electronic structure as follows. First, we scan along an O−H stretch and compute the total dipole moment, μ, at each point along the scan. This dipole moment is then projected onto the O−H bond vector, $\mu_{\mathrm{OH}} = \frac{R_{\mathrm{OH}} \cdot \mu}{|R_{\mathrm{OH}}|}$. This allows us to isolate how the dipole changes as the O−H vibrates in a particular environment.

We considered three different environments to see how the dipole derivatives depend on the local environment and ultimately chose a set of parameters we deemed appropriate for polarization and one to represent the full system. Remember that in EDA, the polarization energy represents the response of a molecule to the rest of the environment plus mutual polarization. To that end, we computed an O−H stretch of an isolated H_2O in a polarizable continuum described by the COSMO model.[2] This resulted in the dipole surface shown with circles in Fig. 1. The lines are fits with a second order polynomial, $\Delta\mu_{\mathrm{OH}} = \mu^{(1)}_{\mathrm{OH}} \Delta R_{\mathrm{OH}} + \mu^{(2)}_{\mathrm{OH}} \Delta R^2_{\mathrm{OH}}$. For H_2O in a dielectric, we find $\mu^{(1)}_{\mathrm{OH}} = 0.345$ and $\mu^{(2)}_{\mathrm{OH}} = 0.095$. This is to be compared to the same values for an isolated H_2O of $\mu^{(1)}_{\mathrm{OH}} = 0.165$ $\mu^{(2)}_{\mathrm{OH}} = -0.012$. The effect of an environment is, therefore, to promote dissociation to ionic products rather than radicals which is what the change in sign of $\mu^{(2)}_{\mathrm{OH}}$ implies and to make it easier to elongate the O−H bond which is what the magnitude of $\mu^{(1)}_{\mathrm{OH}}$ means.

Interestingly, when the environment is represented by an actual molecule, as with $(H_2O)_2$, the second dipole derivative increases by a lot and increases even further when when a dielectric continuum is used. This is seen in the points marked with squares in Fig. 1. We attribute this large non-linearity in the dipole surface to the role of charge delocalization, which makes these parameters suitable for the full potential which includes charge transfer.

A water dimer, however, is not a representative environment for liquid water, or even large water clusters. To remedy this, we constructed a tetrahedrally coordinated water dimer, $(H_2O)_{2+6}$. The notation $(H_2O)_{2+6}$ indicates that this is a water dimer solvated by six other water molecules. The cartesian coordinates are available with the manuscript. When this larger solvation environment is used, we see a further increase in the dipole derivatives beyond just a water dimer. The dipole derivatives for the central O−H stretch in $(H_2O)_{2+6}$ are $\mu^{(1)}_{\mathrm{OH}} = 1.082$ and $\mu^{(2)}_{\mathrm{OH}} = 1.54$. When these values are used in the model, the $\Delta\omega$ vs ΔR_e slope is ≈ -17.1 cm^{-1}/.001 over the same clusters.



Supplementary Figure 1: *Projected dipole moments along various O-H stretches.* The dipole moment of (H_2O), $(H_2O)_2$, and $(H_2O)_{2+6}$ are computed with ωB97X-V/def2-QZVPPD as a function of the O−H stretch distance. All other degrees of freedom are fixed. The dipole moment is projected along the O−H stretch unit vector. $(H_2O)_{2+6}$ is a water dimer with six additional water molecules placed so that the water dimer is tetrahedrally coordinated. The blue points and lines are gas-phase systems while the red points and lines include a polarizable continuum. Each point is connected vertically to make it clear they are the same structure but in the presence of a polarizable continuum. Circles are used for a single water molecule, squares for a water dimer, and triangles for the tetrahedrally solvated water dimer. See text for more details and values of computed dipole derivatives.

In principle, the curves in Figure 1 should eventually be equal when sufficient explicit solvent is included in the calculation. Since the dipole derivatives computed from $(H_2O)_{2+6}$ are not sufficient to reproduce the Badger correlation, we increase the dipole derivatives until the appropriate correlation is reproduced. This results in values of $\mu^{(1)}_{OH} = 1.19$ and $\mu^{(2)}_{OH} = 3.69$.

We have already shown in Figure 5 that our model produces accurate three-body contributions to polarization and charge transfer for water trimers extracted from ion-water clusters. It is interesting to see, however, the 2-body and many-body contributions to each component of the energy for whole water clusters. To that end, we computed the 2-body and many-body contributions for a subset of the reference structures used in Tables 3 and 4. A comparison of these quantities as

computed with ωB97X-V/def2-QZVPPD and FQCT is shown in Table 1.

Comparison of Many-Body Expansion for EDA Components (kcal/mol)

$(H_2O)_n$	Component	Cls. Elec.		Mod. Pauli		Disp.		Pol.		CT	
		FQCT	ωB97X-V	FQCT	ωB97X-V	FQCT	ωB97X-V	FQCT	ωB97X-V	FQCT	ωB97X-V
$(H_2O)_3$	2-Body	-27.28	-26.77	29.05	28.87	-6.31	-6.28	-3.38	-3.48	-6.13	-6.08
	$\geq$ 3-Body	-	-	0.0	-0.24	0.0	0.21	-1.43	-1.63	-0.69	-0.74
	Total	-27.28	-26.77	29.05	28.63	-6.31	-6.07	-4.82	-5.11	-6.82	-6.82
$(H_2O)_6$	2-Body	-79.27	-78.43	87.70	88.40	-19.65	-19.90	-10.53	-10.94	-17.86	-18.41
Prism	$\geq$ 3-Body	-	-	0.0	-0.62	0.0	0.83	-6.09	-6.40	-2.71	-2.80
	Total	-79.27	-78.43	87.70	87.78	-19.65	-19.07	-16.62	-17.34	-20.57	-21.21
$(H_2O)_6$	2-Body	-79.03	-78.67	89.12	90.10	-19.13	-19.36	-11.01	-11.25	-19.07	-19.62
Cage	$\geq$ 3-Body	-	-	0.0	-0.66	0.0	0.79	-5.99	-6.26	-2.91	-3.07
	Total	-79.03	-78.67	89.12	89.44	-19.13	-18.56	-17.00	-17.51	-21.98	-22.69
$(H_2O)_{10}$	2-Body	-163.4	-162.4	189.1	192.2	-37.89	-38.34	-24.16	-24.92	-42.01	-43.27
	$\geq$ 3-Body	-	-	0.0	-0.68	0.0	1.18	-14.24	-14.92	-7.19	-7.42
	Total	-163.4	-162.4	189.1	191.5	-37.89	-37.16	-38.40	-39.84	-49.20	-50.69
$(H_2O)_{16}$	2-Body	-280.5	-277.7	325.0	329.0	-67.23	-67.61	-41.26	-42.44	-71.59	-73.15
Antiboat	$\geq$ 3-Body	-	-	0.0	-1.06	0.0	2.17	-25.58	-27.07	-12.05	-12.37
	Total	-280.5	-277.7	325.0	327.9	-67.23	-65.44	-66.84	-69.51	-83.64	-85.52
$(H_2O)_{20}$	2-Body	-358.8	-355.4	416.5	420.9	-89.17	-89.20	-52.53	-54.26	-91.46	-92.80
ES Prism	$\geq$ 3-Body	-	-	0.0	-1.66	0.0	3.03	-32.69	-34.85	-15.04	-15.21
	Total	-358.8	-355.4	416.5	419.3	-89.17	-86.17	-85.22	-89.11	-106.5	-108.0

Table 1: Comparison of the 2-body, many-body (i.e. $\geq$ 3-Body), and total energies as predicted by FQCT and as computed with ωB97X-V/def2-QZVPPD. Both calculations are done at the *ab initio* optimized geometries. Names of isomers, when applicable, are written below the cluster size.

One of the significant advantages of using data from EDA to parameterize a force field is it allows for very fine-grained analysis of the accuracy of a model. We illustrate this in Table 1 by comparing the 2-body, many-body, and total energies as computed by our force field and ωB97X-V/def2-QZVPPD. First, we can see that the model is generally very accurate at reproducing each term, as should be expected given the excellent performance of the model in reproducing total energies and optimized strcutures. With that being said, there are a couple of apparent shortcomings. First, we currently do not include many-body dispersion which is very small but not entirely negligible, especially considering it is the only term besides Pauli repulsion which is repulsive. It is sometimes argued that many-body dispersion can be ignored since there is also a small many-body exchange effect which is attractive and hence offsets many-body dispersion. It is clear from Table 1 that this is at least approximately true in EDA. Even if the cancellation were perfect, however, that does not guarantee the forces will cancel. For that reason, we may explore adding many-body dispersion in the future, but neglect it for now since many-body charge transfer is generally much more important.

Additionally, it is apparent that our model slightly understimates two-body contributions to both polarization and charge transfer. This likely arises from a limitation of the functional forms and hence suggests improvements that can be explored in the future. Our description of many-body contributions to charge transfer, however, are excellent. Table 1 also provides information about ways in which we have gotten lucky in the parameterization of this model. For instance, in larger clusters the model's dispersion energies are slightly too attractive while our polarization model is not quite attractive enough. This cancellation of errors was not intended but likely contributes, in some small way, to the successes we have seen in Tables 4 and 5.

We are digging into such detail with the hopes that future force fields will do the same. Electronic structure only used to be able to provide coarse-grained information in the form of total energies and forces. This hampered the ability to pinpoint limitations of a force field. Now, however, we can easily identify systematic errors in a force field using the powerful combination of the many-body expansion and energy decomposition analysis.

Supplementary References

[1] Mark A Boyer, Ondrej Marsalek, Joseph P Heindel, Thomas E Markland, Anne B McCoy, and Sotiris S Xantheas. Beyond badger's rule: The origins and generality of the structure–spectra relationship of aqueous hydrogen bonds. *The Journal of Physical Chemistry Letters*, 10(5): 918–924, 2019.

[2] Andreas Klamt. The cosmo and cosmo-rs solvation models. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):699–709, 2011.