Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics. This journal is © the Owner Societies 2019

1

SELECTION OF MOLECULES

We randomly selected 100 molecules from the QM9 database. The molecular structure is shown in Figure 1, while the geometries can be found in the original QM9 database (see main article) under the molecule IDs 477, 533, 1339, 4287, 4649, 7428, 9873, 11028, 11462, 14495, 15911, 16329, 18223, 19101, 19153, 22179, 22455, 22792, 23936, 24472, 25549, 26651, 28443, 29270, 29339, 30529, 31529, 31834, 43770, 44486, 45907, 47563, 48501, 49645, 52705, 53771, 54475, 57006, 57153, 57596, 58927, 59644, 59856, 61517, 61835, 63055, 64719, 65348, 65854, 65910, 68003, 69744, 70198, 70859, 71845, 77510, 82728, 85630, 86653, 87663, 88516, 89080, 92340, 93760, 94611, 95343, 95728, 95935, 96909, 96931, 99298, 100316, 100618, 104511, 107396, 108868, 111368, 112615, 114234, 114715, 116357, 116669, 117909, 118501, 118959, 122444, 122513, 122700, 122800, 123890, 124047, 126109, 127197, 128227, 129180, 129809, 130514, 131048, 132310, 132642.

ALCHEMICAL DERIVATIVES

The APDFT energy expression eqns 4-6 in the main text require to calculate the alchemical derivatives of the electron density. We do this by means of finite differences where the small displacement in Z space consists of changing the nuclear charges of the atoms by 0.05e. This is done with the QM/MM feature of PySCF as implemented in APDFT. To remove any numerical noise, the energy expressions are re-cast as follows:

$$\Delta E = \int_{\Omega} d\mathbf{r} \left(\sum_{I} \frac{\Delta Z_{I}}{|\mathbf{r} - \mathbf{R}_{\mathbf{I}}|} \right) \left(\sum_{i} \alpha_{i} \rho_{i}(\mathbf{r}) \right) \quad (1)$$

$$=\sum_{I}\sum_{i}\Delta Z_{I}\alpha_{i}\int_{\Omega}d\mathbf{r}\frac{\rho_{i}(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_{I}|}$$
(2)

where the weighting coefficients α_i come from the regular central finite difference stencil except for the case of mixed partial derivatives where we use

$$\frac{\partial^2 \rho}{\partial Z_I \partial Z_J} \approx \frac{1}{2h^2} \Big[\rho(Z_1, \dots, Z_I + h, \dots, Z_J + h, \dots, Z_N) \\ - \rho(Z_1, \dots, Z_I + h, \dots, Z_N) \\ - \rho(Z_1, \dots, Z_J + h, \dots, Z_N) \\ + 2\rho(Z_1, \dots, Z_N) \\ - \rho(Z_1, \dots, Z_I - h, \dots, Z_N) \\ - \rho(Z_1, \dots, Z_J - h, \dots, Z_N) \\ + \rho(Z_1, \dots, Z_I - h, \dots, Z_J - h, \dots, Z_N) \Big]$$

instead to reduce computational effort. In this form, we require the electrostatic potential at the nucleus for each proton to be removed for each point in the finite difference stencil only. The electrostatic potential at the nucleus (without any nuclear-nuclear interactions) can be calculated analytically from the density matrix. We do so using the one-electron integral code as implemented in PySCF.



FIG. 1. All molecules used for the evaluation of HF, CCSD, and APDFT. Bottom: Histogram of the deprotonation energies for single and double deprotonation in the data set considered. Data shown for HF/def2-TZVP and CCSD/6-31G*.