Supporting Information For

QM/MM simulations with the Gaussian Electrostatic Model, a density–based potential

H. Gökcan[a], **E. Kratz**[b], **T. A. Darden**[c], **J.-P. Piquemal**[d], **G. A. Cisneros**[*,a]

[a]Department of Chemistry, University of North Texas, Denton, TX, 76201

[b]Department of Chemistry, Wayne State University, Detroit, MI, 48202, USA

[c]OpenEye Scientific Software, Santa Fe, NM, USA

[d]Laboratoire De Chimie Théorique, UPMC Univ. Paris 06, Paris, France

e-mail: andres@unt.edu

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

ABBREVIATIONS

EDA : Energy Decomposion Analysis
MM : Molecular Mechanics
QM : Quantum Mechanics

1. Theory

The total energy of a QM/MM system may be seperated into three components;

$$E_{Tot} = E^{QM} + E^{MM} + E^{QM/MM} \tag{1}$$

where E_{QM} corresponds to the energy of the QM subsystem, E_{MM} is the energy of the MM subsystem, and $E_{QM/MM}$ component corresponds to the interaction energy between the quantum and classical subsystems.

For our current implementation the MM subsystem is represtented by GEM and thus all MM and QM/MM interactions are calculated using the GEM interaction terms. In the present case, the MM subsystem energy (and forces) do not include non–bonded interactions since we are only considering water at this stage, similar to our first QM/GEM implementation where only the Coulomb interaction was considered [3], and thus $E^{MM} = 0.0$. In the case of $E^{QM/GEM}$, four terms have been considered, Coulomb, Exchange–repulsion, polarization and dispersion:$E^{QM/GEM}_{Coul} + E^{QM/GEM}_{exch} + E^{QM/GEM}_{pol} + E^{QM/GEM}_{disp}$.

The Coulomb and exchange–repulsion terms involve 3 center integrals between the QM density the fitted GEM density:

$$E^{QM/GEM}_{Coul} = \int\int \frac{\rho(r_1)\tilde{\rho}(r_2)}{r_{12}} dr_1 dr_2 \tag{2}$$

and

$$E^{QM/GEM}_{exch} = K_{exch} \int\int \rho(r_1)\tilde{\rho}(r_2) dr_1 dr_2 \tag{3}$$

where K_{exch} is a proportionality constant. [8, 6, 2] The polarization interaction is calculated by using Mulliken point charges to approximate the QM wavefunction for the polarization of the MM environment in the calculation of the inducible point dipoles as previously described [4]. The dispersion term is approximated by the multipolar expansion taking the 6, 8 and 12 terms into consideration. The exchange proportionality parameter, as well as the coefficients for the dispersion term have been parametrized by linear least squares to match the SAPT2+3/aug–cc–pVTZ components for the ten water dimers [7]

In the present implementation we have considered two approaches for the calculation of the total interaction depending on the calculation of the QM/GEM exchange–repulsion term. In both cases the QM/GEM polarization and dispersion terms are added *a posteriori* to the total energy (and forces). The Coulomb term is caluclated by include the frozen GEM density in the core Hamiltonian in both alternative approaches. Thus, the difference in the two approaches involves only the exchange–repulsion component. In one approach, the exchange–repulsion inter–molecular interaction is calculated after the SCF has completed, i.e., by employing the relaxed one–electron density matrix for the calculation of the overlap by means of Eq. 3. The second approach involves the inclusion of the exchange–repulsion term explicitly in the SCF. This is achieved by including the GEM density in the core Hamiltonian:

$$H_{eff} = H_{core} + V_{GEM} \tag{4}$$

where V_{GEM} involves the introduction of the 3–center integrals of the QM MOs with the GEM density in the core Hamiltonian by:

$$V_{GEM} = \sum_{l} x_l \sum_{\mu\nu} \langle \mu\nu \parallel l \rangle + K'_{exch} \sum_{l} x_l \sum_{\mu\nu} \langle \mu\nu \mid l \rangle \tag{5}$$

where the first term corresponds to the Coulomb integrals and the second term to the overlap integrals multiplied by the exchange–repulsion proportionality constant. In this second case the exchange–repulsion proportionality constant is different than the former case since the QM wavefunction experiences a different external potential (see below). In all cases the required integrals have been programmed into a modified version of Psi4 as described in the main text [5].

2. Water Dimers Dataset: Subset Selection

The chosen molecules for the present study involve a selected subset of the water dimer potential energy surface (PES) reported by Babin *et al.* [1]. The complete PES includes a large number of dimers with internal geometries that deviate significantly from the equilibrium geometry, as well as dimers with very short inter–molecular distances. Given the fact that GEM employs frozen densities, the accuracy of the inter–molecular interaction suffers for severly distorted intra–molecular geometries as shown in Figure S.1.

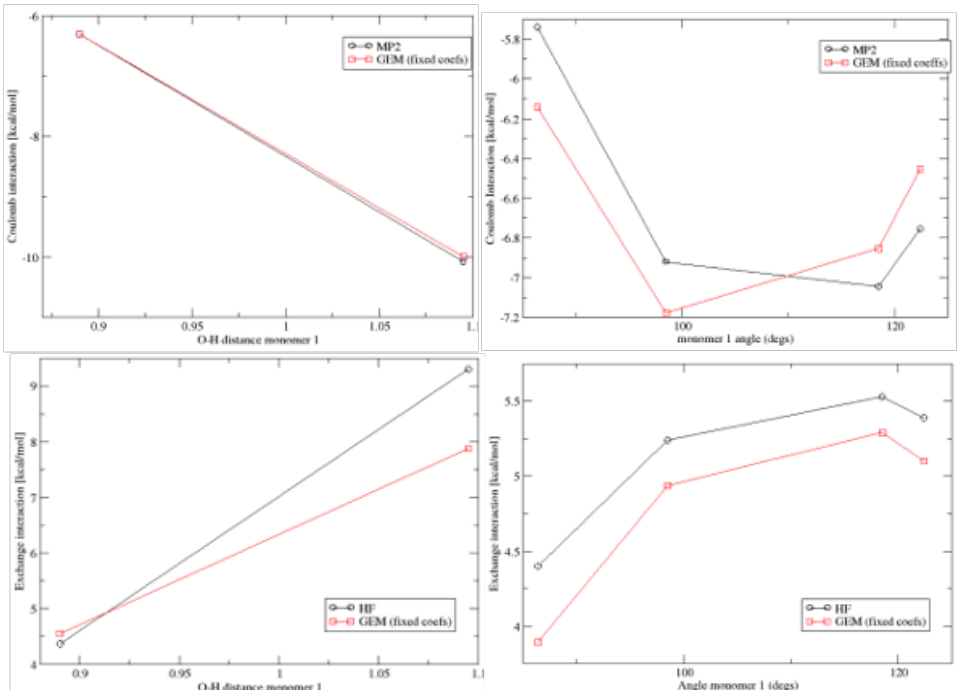

Figure S.1: Errors in Coulomb (top row) and exchange–repulsion (bottom row) inter–molecular interactions with respect to changes in internal geometry. The energies correspond to the canonical (lowest energy) water dimer with distorted internal–geometries for both monomers for either bonds or angles. Coulomb energy references are calculated at the MP2/aug–cc–pVTZ level and exchange–repulsion reference energies are calculated with RVS at the HF/aug–cc–pVTZ level.

Given the above results, we have selected a subset of the water dimers that correspond to geometries with intra–molecular geometries that do not significantly deviate from the equilibrium geometries using the geometrical parameters defined in Figure S.2:

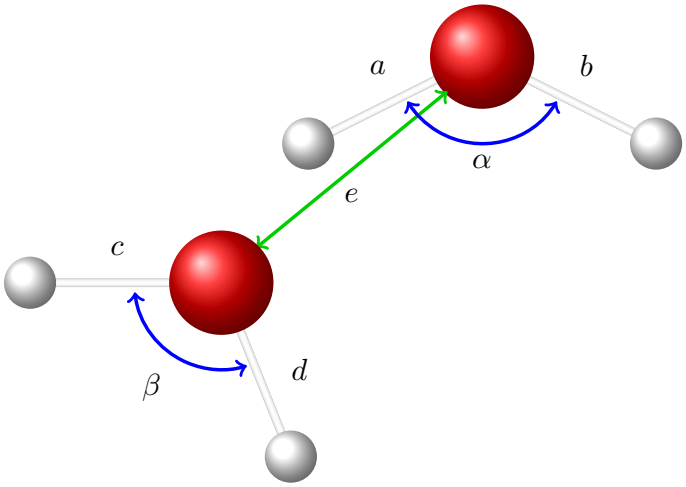

Figure S.2: Selected angles and distances in water dimers for population reduction, and their nomenclature. α and β indicate the intramolecular angles. Intramolecular distances are depicted with a, b, c, and d. The intermolecular O-O distance is denoted with e.

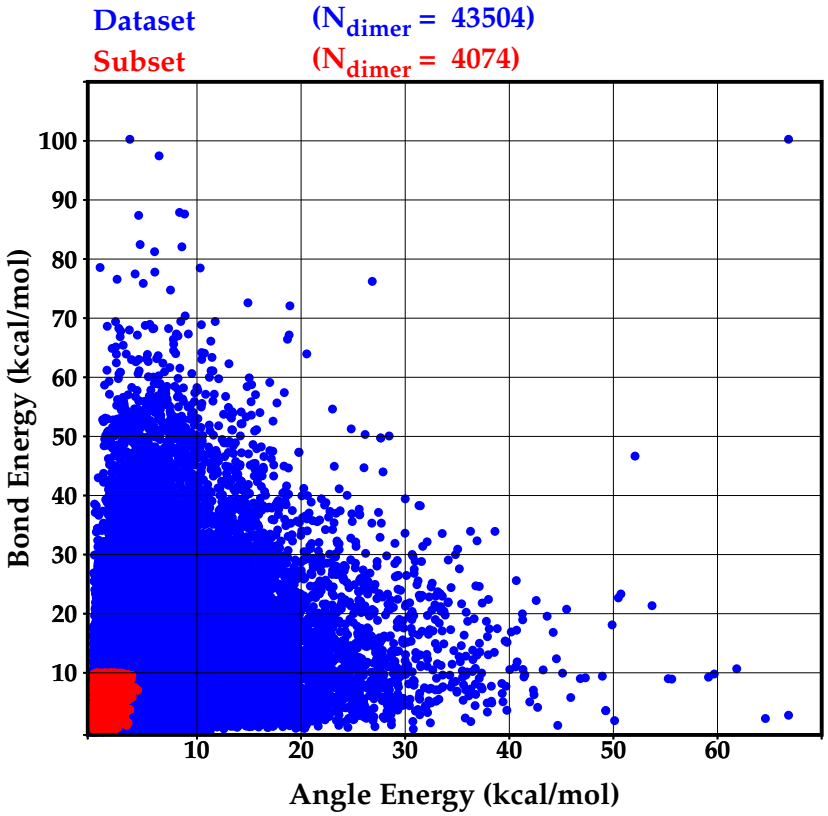

Figure S.3: Angle and bond energies of the dataset of water dimers (blue) and the subset of water dimers (red).

3. Water Dimers Subset: Clustering

Table S.1: The clusters of water dimers, their populations and the centroids of each cluster that are defined by intramolecular distances and angles, and intermolecular O–O distance (see Figure S.1). Clusters are obtained with kmeans clustering analysis as implemented in `sklearn.cluster` module of Scikit-learn. Maximum number of iterations for a single run is $max\_iter = 1000$, the number of runs with different centroid seeds is $n\_init = 100$, verbosity mode is $verbose = 0$, the number of seeds is $random\_state = 5000$, and the relative tolerance with regards to inertia to declare convergence is $tol = 1e - 10$.

Centroid	**Population**	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$	$\boldsymbol{d}$	$\boldsymbol{e}$
0	347	114.163	113.899	0.976	0.978	0.978	0.983	5.332
1	464	100.808	101.401	0.985	0.986	0.983	0.984	4.882
2	894	104.771	104.679	0.976	0.975	0.976	0.976	4.381
3	321	114.306	101.478	0.975	0.976	0.979	0.989	5.402
4	378	101.248	108.363	0.979	0.983	0.983	0.980	5.312
5	323	106.901	114.530	0.979	0.980	0.976	0.981	5.383
6	447	108.951	108.192	0.976	0.983	0.981	0.979	5.429
7	267	101.127	114.408	0.979	0.975	0.983	0.982	5.351
8	264	115.093	107.547	0.979	0.979	0.974	0.986	5.353
9	369	107.873	100.307	0.980	0.983	0.980	0.978	5.423

4. Fitting Results for Exchange–Repulsion & Dispersion Parameters

4.1. C_n coefficients for dispersion term

Table S.2: C_{10}, C_8, and C_6 coefficents that are obtained with least squares fitting to the 10 stable water dimers.

	H–H	**O–O**	**O–H**
$\boldsymbol{C_{10}}$	2.658583e-2	9.155587e-1	1.560157e-1
$\boldsymbol{C_8}$	6.979261e-14	9.744981e-14	8.246985e-14
$\boldsymbol{C_6}$	9.799039e-3	9.436931e-1	9.616281e-2

4.2. Proportionality parameters for exchange–repulsion component

Proportionality parameters K and K' are obtained with iterations of least squares fitting to the QM reference data of stable 10 water dimers. In case of calculations with Tychonov regularization, least squares fitting is performed with the exchange–repulsion energies that are computed with the Hermite coefficients obtained by $\lambda_1 = 0.00003$ using A2DG auxiliary set, and by $\lambda_1 = 0.01$ using A2 auxiliary set.

4.2.1 K and K' for A2DG Auxiliary Set

Table S.3: Proportionality parameter (K) obtained with 1 iteration of least squares fitting to the QM reference data of stable 10 water dimers for both of the fitting procedures. The results obtained with Tychonov regularization and Cholesky decomposition are depicted as E_{exch}^{Tyc} and E_{exch}^{Chol} respectively.

Dimer	**SAPT**	$\boldsymbol{E_{exch}^{Tyc}}$ $(\boldsymbol{K = 1.000000})$	$\boldsymbol{E_{exch}^{Tyc}}$ $(\boldsymbol{K = 5.081745})$	$\boldsymbol{E_{exch}^{Chol}}$ $(\boldsymbol{K = 1.000000})$	$\boldsymbol{E_{exch}^{Chol}}$ $(\boldsymbol{K = 5.144886})$
1	8.5381	1.6114	8.1884	1.6591	8.5366
2	7.0755	1.2349	6.2755	1.2789	6.5784
3	6.7509	1.1722	5.9577	1.2155	6.2547
4	6.3876	1.3322	6.7693	1.2695	6.5306
5	5.2836	0.9777	4.9688	0.9068	4.6640
6	4.8217	0.8547	4.3428	0.7794	4.0096
7	4.3102	1.1038	5.6101	1.0329	5.3140
8	1.2748	0.1851	0.9396	0.1751	0.8995
9	4.5257	0.7574	3.8493	0.7373	3.7939
10	2.3264	0.3031	1.5400	0.2654	1.3657

Table S.4: Proportionality parameter (K') obtained with 5 iterations of least squares fitting to the QM reference data of stable 10 water dimers for both of the fitting procedures. The results obtained with Tychonov regularization and Cholesky decomposition are depicted as E_{exch}^{Tyc} and E_{exch}^{Chol} respectively.

Dimer	SAPT	$\boldsymbol{E_{exch}^{Tyc}}$ $(\boldsymbol{K' = 1.000000})$	$\boldsymbol{E_{exch}^{Tyc}}$ $(\boldsymbol{K' = 7.075662})$	$\boldsymbol{E_{exch}^{Chol}}$ $(\boldsymbol{K' = 1.000000})$	$\boldsymbol{E_{exch}^{Chol}}$ $(\boldsymbol{K' = 7.069278})$
1	8.5381	1.5060	8.7557	1.5676	9.0918
2	7.0755	1.1923	7.1277	1.2263	7.4331
3	6.7509	1.1295	6.8289	1.1686	7.1326
4	6.3876	1.1295	5.5962	1.0740	5.1600
5	5.2836	0.8785	4.6577	0.8004	4.1570
6	4.8217	0.7530	4.3280	0.7050	3.7964
7	4.3102	0.9413	3.9612	0.8551	3.4736
8	1.2748	0.1883	0.9787	0.1648	0.9076
9	4.5257	0.6903	4.1316	0.6941	3.9861
10	2.3264	0.2510	1.5155	0.2432	1.2498

4.2.2 K and K' for A2 Auxiliary Set

Table S.5: Proportionality parameter (K) obtained with 1 iteration of least squares fitting to the QM reference data of stable 10 water dimers for both of the fitting procedures. The results obtained with Tychonov regularization and Cholesky decomposition are depicted as E_{exch}^{Tyc} and E_{exch}^{Chol} respectively.

Dimer	SAPT	$\boldsymbol{E_{exch}^{Tyc}}$ $(\boldsymbol{K = 1.000000})$	$\boldsymbol{E_{exch}^{Tyc}}$ $(\boldsymbol{K = 5.158955})$	$\boldsymbol{E_{exch}^{Chol}}$ $(\boldsymbol{K = 1.000000})$	$\boldsymbol{E_{exch}^{Chol}}$ $(\boldsymbol{K = 6.165279})$
1	8.5381	1.4897	7.6845	1.2475	7.6908
2	7.0755	1.1264	5.8095	0.9118	5.6200
3	6.7509	1.0655	5.4976	0.8540	5.2654
4	6.3876	1.3824	7.1317	1.1760	7.2503
5	5.2836	1.0329	5.3276	0.8327	5.1324
6	4.8217	0.9111	4.7013	0.7122	4.3901
7	4.3102	1.1678	6.0235	1.0329	6.3686
8	1.2748	0.1544	0.7951	0.1657	1.0222
9	4.5257	0.7110	3.6672	0.6068	3.7393
10	2.3264	0.3219	1.6616	0.2811	1.7344

Table S.6: Proportionality parameter (K') obtained with 8 and 6 iterations of least squares fitting using Tychonov regularization and Cholesky decomposition respectively to the QM reference data of stable 10 water dimers for both of the fitting procedures. The results obtained with Tychonov regularization and Cholesky decomposition are depicted as E_{exch}^{Tyc} and E_{exch}^{Chol} respectively.

Dimer	SAPT	E_{exch}^{Tyc} $(K' = 1.0000)$	E_{exch}^{Tyc} $(K' = 7.5566)$	E_{exch}^{Chol} $(K' = 1.0000)$	E_{exch}^{Chol} $(K' = 10.6283)$
1	8.5381	1.3953	8.2517	1.1559	8.4010
2	7.0755	1.0724	6.6776	0.8593	6.6748
3	6.7509	1.0170	6.3806	0.8068	6.3218
4	6.3876	1.1797	6.1598	0.9846	6.1778
5	5.2836	0.9229	5.2734	0.7282	5.0469
6	4.8217	0.8349	4.9635	0.6389	4.6290
7	4.3102	0.9839	4.5176	0.8572	4.5257
8	1.2748	0.1441	0.7977	0.1558	1.1445
9	4.5257	0.6670	3.9990	0.5639	4.2549
10	2.3264	0.2994	1.7300	0.2591	1.8633

5. Effect of Different Computation Schemes in QM/GEM using A2DG Auxiliary Set

5.1. Inter–molecular energy components obtained with Cholesky decomposition using Coulomb and Exchange embedding scheme

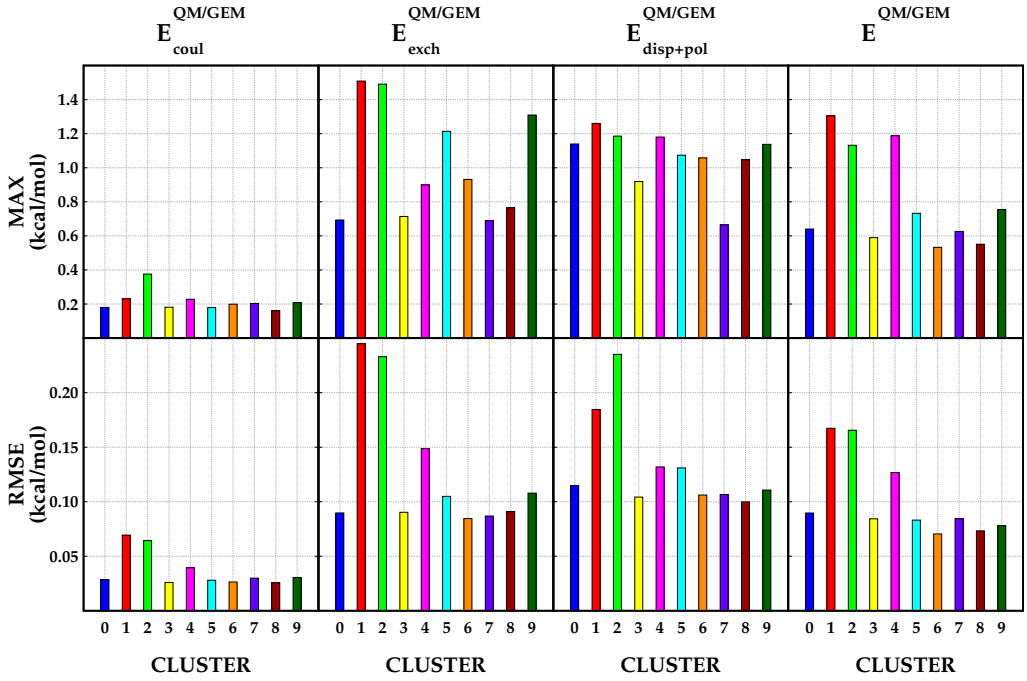

Figure S.4: Errors per cluster with respect to the SAPT calculations. Each cluster is represented with different colors of columns.

5.2. Inter–molecular energy components obtained with Tychonov regularization ($\lambda_1 = 0.00003$) using Coulomb and Exchange embedding scheme

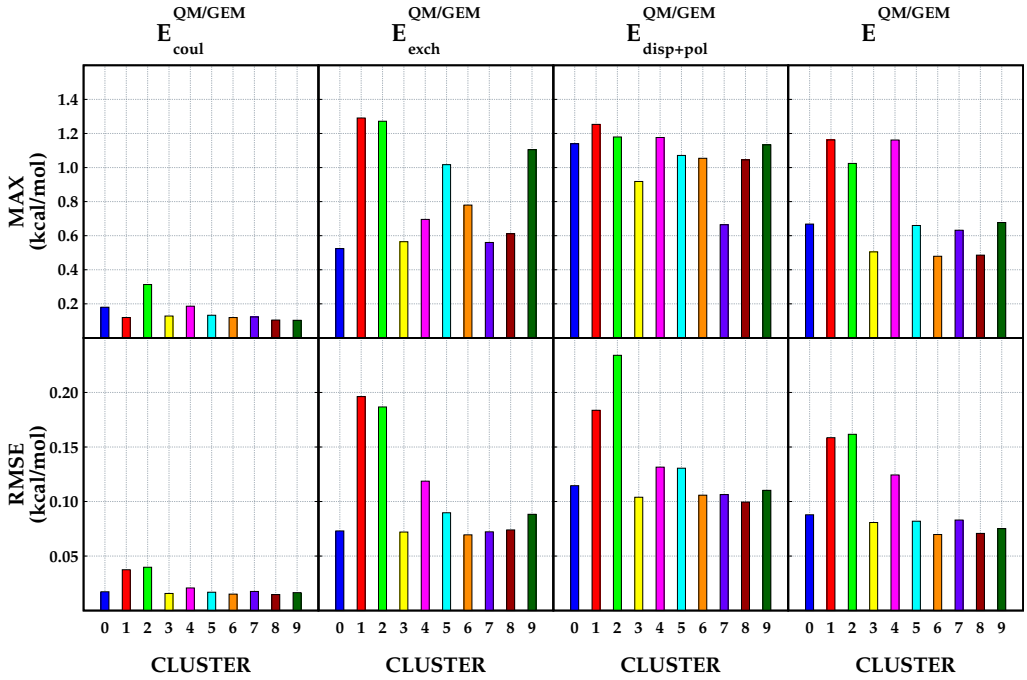

Figure S.5: Errors per cluster with respect to the SAPT calculations. Each cluster is represented with different colors of columns.

5.3. Exchange–Repulsion Component obtained with Cholesky decomposition and Tychonov regularization ($\lambda_1 = 0.00003$) using Coulomb embedding scheme

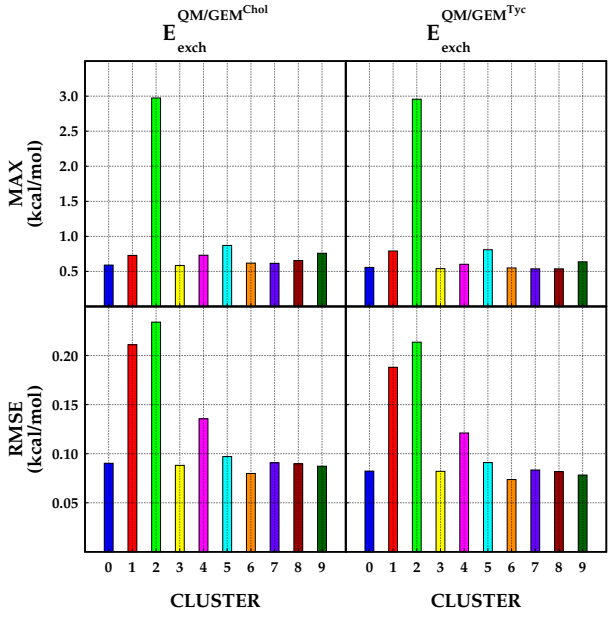

Figure S.6: Errors of $E_{exch}^{QM/GEM}$ component obtained from QM/GEM calculations with respect to the reference QM EDA data. Errors for Cholesky decomposition are depicted on the left, and for Tychonov regularization are depicted on the right.

6. Effect of Different Computation Schemes in QM/GEM using A2 Auxiliary Set

6.1. Inter–molecular energy components obtained with Tychonov regularization ($\lambda_1 = 0.01$, and $\lambda_2 = 0.012$) using Coulomb and Exchange embedding scheme

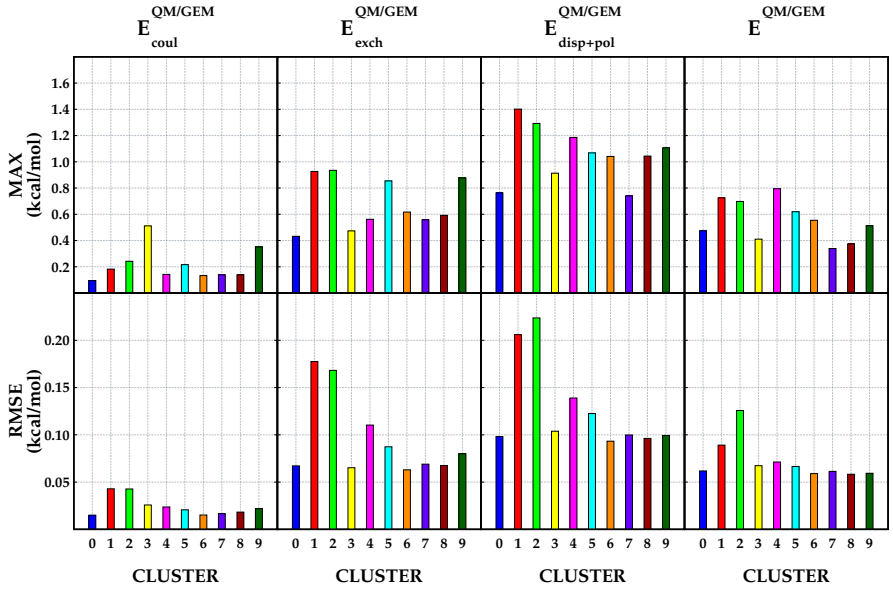

Figure S.7: Errors per cluster with respect to the SAPT calculations. Each cluster is represented with different colors of columns.

6.2. Inter–molecular energy components obtained with Cholesky decomposition using Coulomb and Exchange embedding scheme

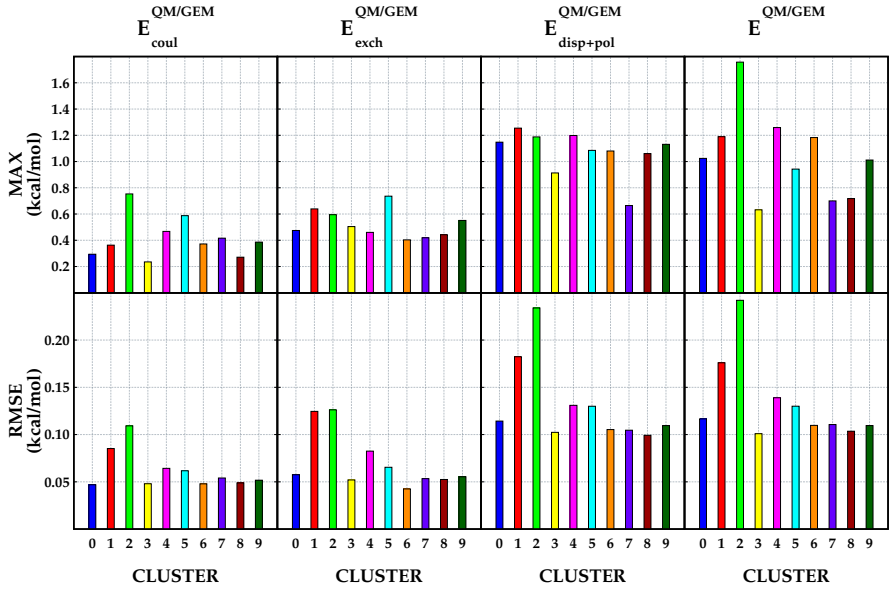

Figure S.8: Errors per cluster with respect to the SAPT calculations. Each cluster is represented with different colors of columns.

6.3. Inter–molecular energy components obtained with Tychonov regularization ($\lambda_1 = 0.01$) using Coulomb and Exchange embedding scheme

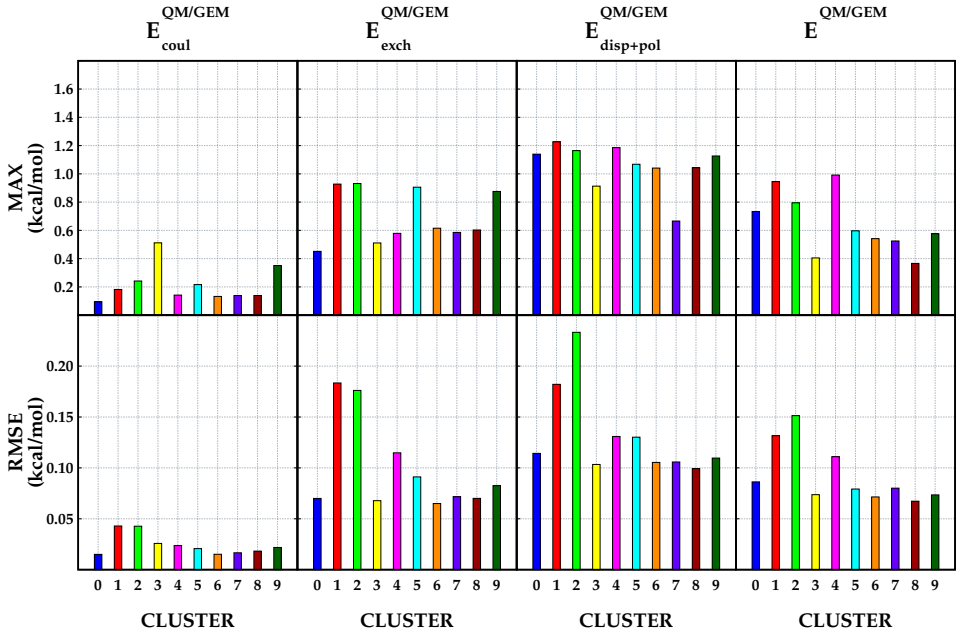

Figure S.9: Errors per cluster with respect to the SAPT calculations. Each cluster is represented with different colors of columns.

6.4. Exchange–Repulsion Component obtained with Cholesky decomposition and Tychonov regularization ($\lambda_1 = 0.01$) using Coulomb embedding scheme

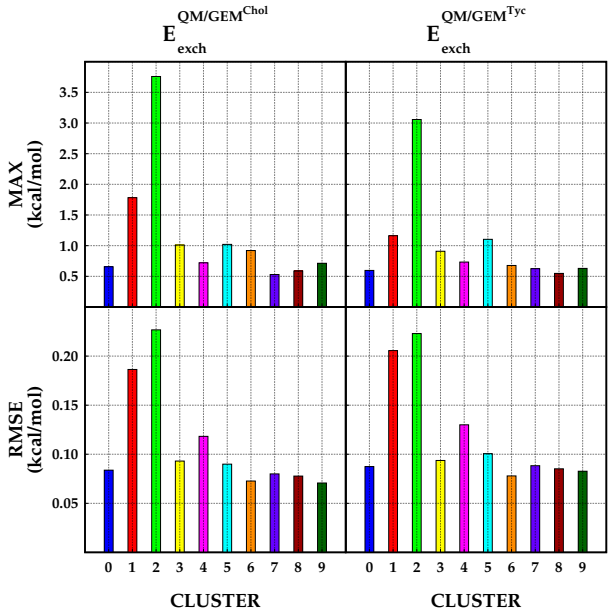

Figure S.10: Errors of $E_{exch}^{QM/GEM}$ component obtained from QM/GEM calculations with respect to the reference QM EDA data. Errors for Cholesky decomposition are depicted on the left, and for Tychonov regularization are depicted on the right.

7. Coulomb Interaction Energy Obtained with QM/AMOEBA Calculations

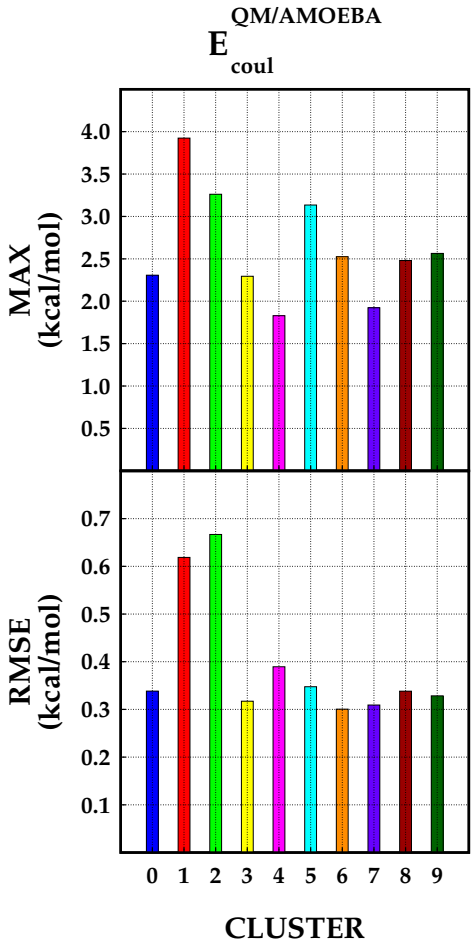

Figure S.11: Errors of $E_{coul}^{QM/AMOEBA}$ component obtained from QM/AMOEBA calculations with respect to the reference QM EDA data.

8. **Mean Signed Errors (MSE) of Inter–molecular Energy Components Obtained with Tychonov Regularization ($\lambda_1 = 0.00003$, and $\lambda_2 = 0.0001$) Using Coulomb and Exchange Embedding Scheme with A2DG Auxiliary Set**

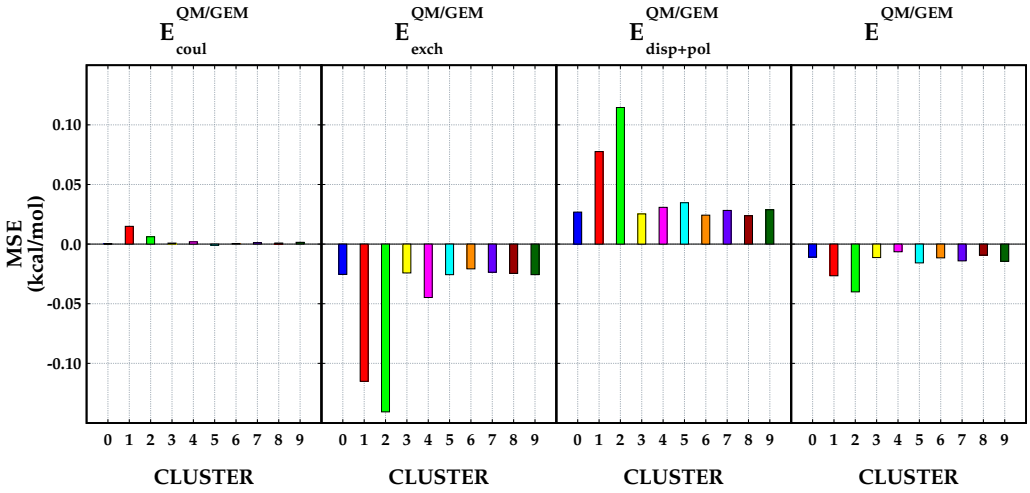

Figure S.12: Errors per cluster with respect to the SAPT calculations. Each cluster is represented with different colors of columns.

9. **Total Inter–molecular Interaction Energy Differences for SAPT2+3/aug-cc-pVTZ and QM/GEM (Tychonov Regularization with $\lambda_1 = 0.00003$, and $\lambda_2 = 0.0001$ Using Coulomb and Exchange Embedding Scheme, and A2DG Auxiliary Set) with Respect to the BSSE corrected wB97X-D/aug-cc-pVTZ calculations**

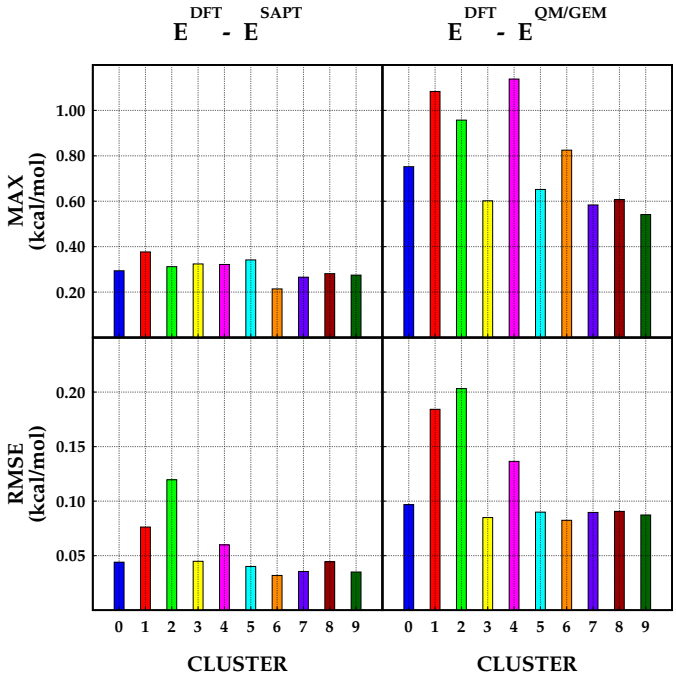

Figure S.13: Errors per cluster with respect to the BSSE corrected wB97X-D/aug-cc-pVTZ calculations. Each cluster is represented with different colors of columns.

10. Term-by-term and Total Intermolecular Interaction Energy Components

- For intermolecular energy components of the reference dataset obtained with the QM EDA calculations at the SAPT2+3/aug–cc–pVTZ level see **SAPT2+3/aug–cc–pVTZ** in **QM_EDA_and_GEM.xlsx**.

- For energy components of the subset of water dimers obtained with the QM/GEM calculations using Cholesky decomposition with A2DG auxiliary set see **GEM_Cholesky_A2DG** in **QM_EDA_and_GEM.xlsx**.

- For energy components of the subset of water dimers obtained with the QM/GEM calculations using Tychonov regularization with A2DG auxiliary set see **GEM_Tychonov_lambda1_A2DG** in **QM_EDA_and_GEM.xlsx**.

- For energy components of the subset of water dimers obtained with the QM/GEM calculations with individual fitting for exchange–repulsion energy using Tychonov regularization with A2DG auxiliary set see **GEM_Tychonov_lambda1_lambda2_A2DG** in **QM_EDA_and_GEM.xlsx**.

- For exchnage–reulsion energy of the subset of water dimers computed after SCF cycle with the QM/GEM calculations using Cholesky decomposition and Tychonov regularization with A2DG auxiliary set see **GEM_Exchange_after_SCF_A2DG** in **QM_EDA_and_GEM.xlsx**.

- For energy components of the subset of water dimers obtained with the QM/GEM calculations using Cholesky decomposition with A2 auxiliary set see **GEM_Cholesky_A2** in **QM_EDA_and_GEM.xlsx**.

- For energy components of the subset of water dimers obtained with the QM/GEM calculations using Tychonov regularization with A2 auxiliary set see **GEM_Tychonov_lambda1_A2** in **QM_EDA_and_GEM.xlsx**.

- For energy components of the subset of water dimers obtained with the QM/GEM calculations with individual fitting for exchange–repulsion energy using Tychonov regularization with A2 auxiliary set see **GEM_Tychonov_lambda1_lambda2_A2** in **QM_EDA_and_GEM.xlsx**.

- For exchnage–reulsion energy of the subset of water dimers computed after SCF cycle with the QM/GEM calculations using Cholesky decomposition and Tychonov regularization with A2 auxiliary set see **GEM_Exchange_after_SCF_A2** in **QM_EDA_and_GEM.xlsx**.

- For total inter–molecular interaction energy and approximate Coulomb interaction energy of the subset of water dimers obtained with the QM/AMOEBA calculations see **AMOEBA** in **QM_EDA_and_GEM.xlsx**.

References

[1] V. Babin, C. Leforestier, and F. Paesani. Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, vrt spectrum, and second virial coefficient. *J. Chem. Theo. Comp.*, 9(12):5395–5403, 2013.

[2] G. A. Cisneros, J.-P. Piquemal, and T. A. Darden. Generalization of the gaussian electrostatic model: Extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods. *J. Chem. Phys.*, 125(18):184101, 2006.

[3] G. A. Cisneros, J.-P. Piquemal, and T. A. Darden. Quantum mechanics/molecular mechanics electrostatic embedding with continuous and discrete functions. *J. Phys. Chem. B*, 110(28):13682–13684, 2006.

[4] E. G. Kratz, A. R. Walker, L. Lagardère, F. Lipparini, J.-P. Piquemal, and G. André Cisneros. Lichem: A qm/mm program for simulations with multipolar and polarizable force fields. *J. Comp. Chem.*, 37(11):1019–1029, 2016.

[5] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill. Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability. *J. Chem. Theo. Comp.*, 13(7):3185–3197, 2017. PMID: 28489372.

[6] J.-P. Piquemal, G. A. Cisneros, P. Reinhardt, N. Gresh, and T. A. Darden. Towards a force field based on density fitting. *J. Chem. Phys.*, 124(10):104101, 2006.

[7] G. S. Tschumper, M. L. Leininger, B. C. Hoffman, E. F. Valeev, H. F. Schaffer III, and M. Quack. Anchoring the water dimer potential energy surface with explicitly correlated computations and focal point analyses. *J. Chem. Phys.*, 116:690–701, 2002.

[8] R. J. Wheatley and S. L. Price. An overlap model for estimating the anisotropy of repulsion. *Mol. Phys.*, 69(3):507–533, 1990.