Supplementary Material for "Understanding the many-body expansion for large systems. I. Precision considerations"

Ryan M. Richard, Ka Un Lao and John M. Herbert*

Department of Chemistry and Biochemistry,

The Ohio State University, Columbus, OH 43210

(Dated: May 12, 2014)

Reproducibility Tests

As discussed in the paper, we tested the reproducibility of a driver-based implementation of the many-body expansion (MBE) in on our previously-developed program, FRAGME∩T,[1] which interfaces with the Q-CHEM electronic structure program,[2,3] versus an implementation of the MBE that is fully self-contained within Q-CHEM. Figure S1 shows the difference in the total binding energies predicted by these two programs, comparing both a two- and a three-body expansion. The largest disagreement is for $(H_2O)_{46}$ and amounts to about 0.4 kcal/mol. Figure S2 is analogous, except that electrostatically embedded (EE) two- and three-body expansions are used. Note that in the EE-3B case the discrepancies are considerable (on the order of several kcal/mol) starting around $N = 30$–35 monomer units. This is caused by lack of precision in the driver-based Q-CHEM input files, as discussed in the papers. Repeating the calculations using all digits of precision (which requires reading Q-CHEM's binary scratch files, so is not completely driver-based), we are able to achieve a much higher level of agreement, even for the EE-3B case, as shown in Fig. S3.

Numerical Thresholds

Table II in the paper reports errors in EE-nB calculations of $(H_2O)_{40}$ as a function of two thresholds, τ_{SCF} and τ_{ints}, when Mulliken charges are used for the embedding. TIP3P charges have also been used in EE-2B and EE-3B studies of water clusters,[4] so Table S1 reports the same study of errors versus thresholds, for the case of TIP3P embedding charges. (Since the format of this table is somewhat different than that of Table II, for ease of comparison the Mulliken embedding data from

* herbert@chemistry.ohio-state.edu

Table II have been re-tabulated in Table S2 in a format analogous to that used in Table S1.) Although TIP3P embedding affords much larger errors at the EE-2B level, errors obtained at the EE-3B level are more comparable to those obtained when Mulliken embedding charges are used, and in either case the trends with respect to the two thresholds are quite similar. In particular, there seems to be no need to reduce τ_{SCF} beyond 10^{-6} hartree, and while tightening τ_{ints} beyond $\sim 10^{-10}$ a.u. does change the EE-3B(TIP3P) results by values approaching 1 kcal/mol, this is small in comparison to the overall error of $\approx$ 17–18 kcal/mol. As with the Mulliken embedding data, however, there is little evidence that the EE-4B(TIP3P) energies have converged as $\tau_{\mathrm{ints}} \to 10^{-14}$ a.u.

To obtain a more incisive understanding of the role of τ_{ints}, we have modified the Q-Chem source code in order to use two separate thresholds, τ_{SP} and τ_{CS}, for shell-pair screening and Cauchy-Schwarz integral screening,[5] respectively. (Ordinarily, both types of screening are controlled by a single parameter in Q-Chem, and we have furthermore modified the code to remove a hard-coded lower limit of $\tau_{\mathrm{SP}} = \tau_{\mathrm{CS}} = 10^{-14}$ a.u. that is present in the officially-released version of the code.) These results are presented in Table S3.

Consistent with results presented in the paper, we see no difference between EE-2B and EE-3B errors obtained using $\tau_{\mathrm{SCF}} = 10^{-6}$ versus 10^{-9} hartree, suggesting that the looser value is perfectly acceptable. Setting $\tau_{\mathrm{SP}} = 10^{-15}$ a.u. (essentially, keeping all shell pairs), we see that the value of τ_{CS} has no effect on the errors, even at the EE-4B level, owing to the relatively small sizes of the subsystems which limits the amount of Cauchy-Schwarz integral screening that is available. This reveals the shell-pair screening as the underlying source of the variation in the errors with respect to the τ_{ints} parameter discussed above.

Quadrature Grid

Finally, Table S4 presents EE-nB errors for four increasingly-large $(H_2O)_N$ clusters, computed at the Hartree-Fock/cc-pVDZ level. (As usual, "error" is defined with respect to a calculation on the entire cluster at the same level of theory.) As was observed for B3LYP/cc-pVDZ calculations, errors increase (often dramatically) as cluster size increases. The point of this exercise is to demonstrate that the DFT quadrature grid is not the fundamental reason for the size-dependent errors that are

reported in this work.

---

[1] R. M. Richard and J. M. Herbert, J. Chem. Phys. **137**, 064113:1 (2012).

[2] Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, et al., Phys. Chem. Chem. Phys. **8**, 3172 (2006).

[3] A. I. Krylov and P. M. W. Gill, WIREs Comput. Mol. Sci. **3**, 317 (2013).

[4] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 46 (2007).

[5] M. Häser and R. Ahlrichs, J. Comput. Chem. **10**, 104 (1989).

[6] P. M. W. Gill, B. G. Johnson, and J. A. Pople, Chem. Phys. Lett. **209**, 506 (1993).

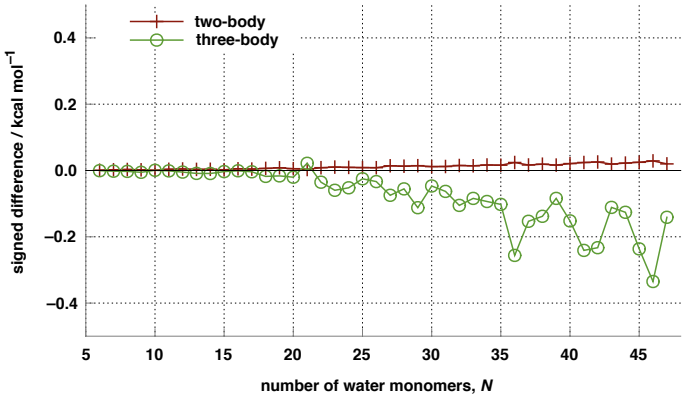

FIG. S1: Differences in non-embedded two- and three-body approximations to the total binding energies of water clusters, comparing the driver-based FRAGME∩T implementation to the fully integrated Q-CHEM implementation. Calculations were performed at the B3LYP/cc-pVDZ level using the SG-1 quadrature grid,[6] with numerical thresholds $\tau_{\mathrm{SCF}} = 10^{-5}$ a.u. and $\tau_{\mathrm{ints}} = 10^{-9}$ a.u.

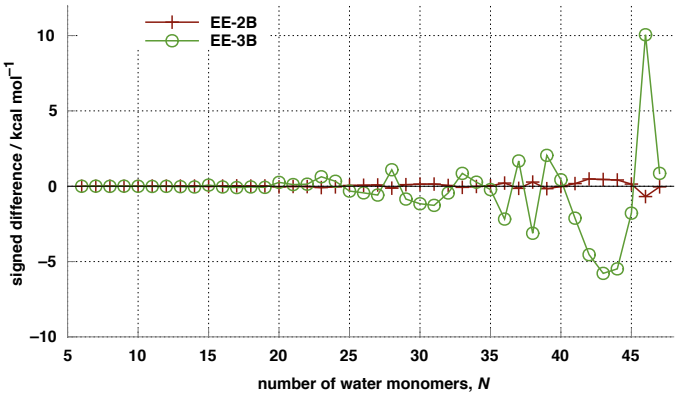

FIG. S2: Differences in EE-2B and EE-3B total binding energies for water clusters computed with FRAGME∩T versus Q-CHEM. Calculations were performed at the B3LYP/cc-pVDZ level using the SG-1 quadrature grid,[6] with numerical thresholds $\tau_{\mathrm{SCF}} = 10^{-5}$ a.u. and $\tau_{\mathrm{ints}} = 10^{-9}$ a.u. Mulliken charges were used for the electrostatic embedding but were rounded off at six decimal digits in atomic units.

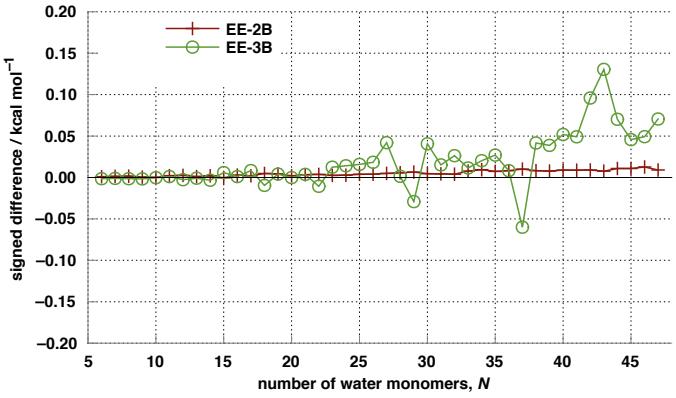

FIG. S3: Differences in EE-2B and EE-3B total binding energies for water clusters computed with FRAGME∩T versus Q-CHEM, using all digits of precision in preparing Q-CHEM input files. Calculations were performed at the B3LYP/cc-pVDZ level using the SG-1 quadrature grid,[6] with numerical thresholds $\tau_{\mathrm{SCF}} = 10^{-5}$ a.u. and $\tau_{\mathrm{ints}} = 10^{-9}$ a.u.

$-\log_{10}(\tau/\text{a.u.})$		error / kcal mol^{-1}			$-\log_{10}(\tau/\text{a.u.})$		error / kcal mol^{-1}		
τ_{ints}	τ_{SCF}	EE-2B	EE-3B	EE-4B	τ_{ints}	τ_{SCF}	EE-2B	EE-3B	EE-4B
9	5	−74.63	16.80	4.63	9	5	−74.63	16.80	4.63
9	6	−74.63	16.80	3.42	10	5	−74.60	17.58	1.77
9	7	−74.63	16.80	3.41	11	5	−74.63	17.85	1.98
9	8	−74.63	16.80	3.41	12	5	−74.60	18.00	1.75
10	5	−74.60	17.58	1.77	13	5	−74.62	18.22	1.16
10	6	−74.60	17.58	0.57	14	5	−74.63	18.36	0.68
10	7	−74.60	17.58	0.56	9	6	−74.63	16.80	3.42
10	8	−74.60	17.58	0.56	10	6	−74.60	17.58	0.57
11	5	−74.63	17.85	1.98	11	6	−74.63	17.85	0.77
11	6	−74.63	17.85	0.77	12	6	−74.60	18.01	0.55
11	7	−74.63	17.85	0.76	13	6	−74.62	18.22	−0.04
11	8	−74.63	17.85	0.76	14	6	−74.63	18.36	−0.52
12	5	−74.60	18.00	1.75	9	7	−74.63	16.80	3.41
12	6	−74.60	18.01	0.55	10	7	−74.60	17.58	0.56
12	7	−74.60	18.01	0.54	11	7	−74.63	17.85	0.76
12	8	−74.60	18.01	0.54	12	7	−74.60	18.01	0.54
13	5	−74.62	18.22	1.16	13	7	−74.62	18.22	−0.05
13	6	−74.62	18.22	−0.04	14	7	−74.63	18.36	−0.53
13	7	−74.62	18.22	−0.05	9	8	−74.63	16.80	3.41
13	8	−74.62	18.22	−0.05	10	8	−74.60	17.58	0.56
14	5	−74.63	18.36	0.68	11	8	−74.63	17.85	0.76
14	6	−74.63	18.36	−0.52	12	8	−74.60	18.01	0.54
14	7	−74.63	18.36	−0.53	13	8	−74.62	18.22	−0.05
14	8	−74.63	18.36	−0.53	14	8	−74.63	18.36	−0.53

TABLE S1: Errors in EE-nB calculations on $(H_2O)_{40}$ as a function of the thresholds τ_{ints} and τ_{SCF}, computed at the B3LYP/cc-pVDZ level (SG-1 quadrature grid[6]) and compared to a supersystem benchmark computed using $\tau_{\text{ints}} = 10^{-14}$ a.u. and $\tau_{\text{SCF}} = 10^{-9}$ a.u. TIP3P embedding charges were used in all EE-nB calculations, which were performed using the fully-integrated Q-Chem implementation of the MBE. As an aid to the eye in gauging the convergence with respect to either threshold, the same data set is tabulated twice, in two different orders.

$-\log_{10}(\tau/\text{a.u.})$		error / kcal mol^{-1}			$-\log_{10}(\tau/\text{a.u.})$		error / kcal mol^{-1}		
τ_{ints}	τ_{SCF}	EE-2B	EE-3B	EE-4B	τ_{ints}	τ_{SCF}	EE-2B	EE-3B	EE-4B
9	5	−10.37	17.34	1.91	9	5	−10.37	17.34	1.91
9	6	−10.38	17.30	1.62	10	5	−10.35	18.19	−1.64
9	7	−10.38	17.30	1.63	11	5	−10.37	18.40	−1.16
9	8	−10.38	17.30	1.63	12	5	−10.33	18.54	−1.32
10	5	−10.35	18.19	−1.64	13	5	−10.35	18.75	−1.91
10	6	−10.35	18.15	−1.93	14	5	−10.35	18.89	−2.42
10	7	−10.35	18.14	−1.92	9	6	−10.38	17.30	1.62
10	8	−10.35	18.14	−1.92	10	6	−10.35	18.15	−1.93
11	5	−10.37	18.40	−1.16	11	6	−10.37	18.36	−1.45
11	6	−10.37	18.36	−1.45	12	6	−10.33	18.50	−1.61
11	7	−10.37	18.36	−1.44	13	6	−10.35	18.71	−2.20
11	8	−10.37	18.36	−1.44	14	6	−10.36	18.85	−2.71
12	5	−10.33	18.54	−1.32	9	7	−10.38	17.30	1.63
12	6	−10.33	18.50	−1.61	10	7	−10.35	18.14	−1.92
12	7	−10.33	18.50	−1.60	11	7	−10.37	18.36	−1.44
12	8	−10.33	18.50	−1.60	12	7	−10.33	18.50	−1.60
13	5	−10.35	18.75	−1.91	13	7	−10.35	18.71	−2.19
13	6	−10.35	18.71	−2.20	14	7	−10.36	18.85	−2.70
13	7	−10.35	18.71	−2.19	9	8	−10.38	17.30	1.63
13	8	−10.35	18.71	−2.19	10	8	−10.35	18.14	−1.92
14	5	−10.35	18.89	−2.42	11	8	−10.37	18.36	−1.44
14	6	−10.36	18.85	−2.71	12	8	−10.33	18.50	−1.60
14	7	−10.36	18.85	−2.70	13	8	−10.35	18.71	−2.19
14	8	−10.36	18.85	−2.70	14	8	−10.36	18.85	−2.70

TABLE S2: Errors in EE-nB calculations on $(H_2O)_{40}$ as a function of the thresholds τ_{ints} and τ_{SCF}, computed at the B3LYP/cc-pVDZ level (SG-1 quadrature grid[6]) and compared to a supersystem benchmark computed using $\tau_{\text{ints}} = 10^{-14}$ a.u. and $\tau_{\text{SCF}} = 10^{-9}$ a.u. Mulliken embedding charges were used in all EE-nB calculations, which were performed using the fully-integrated Q-Chem implementation of the MBE. As an aid to the eye in gauging the convergence with respect to either threshold, the same data set is tabulated twice, in two different orders.

$-\log_{10}(\tau/\text{a.u.})$			error / kcal mol^{-1}		
τ_{SP}	τ_{CS}	τ_{SCF}	EE-2B	EE-3B	EE-4B
12	10	6	−10.14	18.70	−1.42
13	11	6	−10.15	18.91	−2.01
14	12	6	−10.16	19.05	−2.51
15	13	6	−10.15	18.34	−1.74
12	10	9	−10.16	18.70	
13	11	9	−10.15	18.91	
14	12	9	−10.16	19.05	
15	13	9	−10.15	18.34	
15	9	6	−10.15	18.34	−1.74
15	10	6	−10.15	18.34	−1.74
15	11	6	−10.15	18.34	−1.74
15	12	6	−10.15	18.34	−1.74
15	13	6	−10.15	18.34	−1.74
15	14	6	−10.15	18.34	−1.74
15	15	6	−10.15	18.34	−1.74
15	10	9	−10.15	18.34	
15	11	9	−10.15	18.34	
15	12	9	−10.15	18.34	
15	13	9	−10.15	18.34	
15	14	9	−10.15	18.34	
15	15	9	−10.15	18.34	

TABLE S3: Errors in EE-nB calculations on $(H_2O)_{40}$ as a function of τ_{SP}, τ_{CS}, and τ_{SCF}. All calculations were performed at the B3LYP/cc-pVDZ level (SG-1 quadrature grid[6]) and compared to a supersystem benchmark computed using $\tau_{\text{SP}} = \tau_{\text{CS}} = 10^{-15}$ a.u. and $\tau_{\text{SCF}} = 10^{-9}$ a.u. Mulliken embedding charges were used in all EE-nB calculations, which were performed using the fully-integrated Q-Chem implementation of the MBE.

Cluster	error / kcal mol^{-1}			
	EE-2B	EE-3B	EE-4B	EE-5B
$(H_2O)_{10}$	−4.75	1.48	−0.21	0.02
$(H_2O)_{20}$	−13.94	3.60	−0.42	−0.05
$(H_2O)_{30}$	−30.73	9.90	−1.25	−0.29
$(H_2O)_{40}$	−43.40	11.81	−1.68	

TABLE S4: Errors in EE-nB total energies computed at the Hartree-Fock/cc-pVDZ level using TIP3P embedding charges, as compared to a Hartree-Fock/cc-pVDZ supersystem calculation. All calculations used $\tau_{\text{ints}} = 10^{-14}$ a.u. and $\tau_{\text{SCF}} = 10^{-6}$ a.u.