Supporting Information for: "Fragment-Based Calculations of Enzymatic Thermochemistry Require Dielectric Boundary Conditions"

Paige E. Bowling,[1,2], Dustin R. Broderick,[2] and John M. Herbert[1,2*]
[1] *Biophysics Graduate Program, The Ohio State University, Columbus, Ohio 43210 USA*
[2] *Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210 USA*

April 1, 2023

S1 Fragment Definition

In our code (PyFragme∩t), fragments are generally defined using residue number. To avoid severing an polar C–N peptide bond, which is not recommended,[1,2] we instead sever the C–C bond at C_α–C(=O) as shown in Fig. S1. A capping hydrogen atom cap is added at position

$$\mathbf{r}_{\rm cap} = \mathbf{r}_1 + \left(\frac{R_1 + R_{\rm H}}{R_1 + R_2} \right) (\mathbf{r}_2 - \mathbf{r}_1) \,, \tag{S1}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the positions of the two carbon atoms in the original C–C bond, and R_1, R_2, and $R_{\rm H}$ are the van der Waals radii of the atoms in question. This construction was also used in our previous work on protein fragmentation.[1]

S2 Supersystem Calculations

Tables S1 and S2 list the errors in $E_{\rm a}$ and $\Delta_{\rm rxn}E$, respectively, for COMT using several combinations of functional and basis set. Methods tested include MBE(n) with $n \leq 3$, using both vacuum and PCM ($\varepsilon = 4$) boundary conditions. Error is defined with respect to the supersystem calculation at the same level of theory. No screening or supersystem corrections are applied in these calculations.

Table S3 contains the ONIOM-style supersystem corrections to $E_{\rm a}$ and $\Delta_{\rm rxn}E$, using each of the four low-level models that were tested. These corrections were applied to the MBE(2) and MBE(3) data in Fig. 3.

S3 Convergence Tests

Figures S2 and S3 illustrate convergence of the COMT calculations (ωB97X-D/def2-SVP + PCM) as a function of the solvent dielectric constant, ε. (The data for $\varepsilon = 4$ were presented as Fig. 1a and 1b.) These plots demonstrate that vacuum boundary conditions ($\varepsilon = 1$) are an outlier. For both $E_{\rm a}$ and $\Delta_{\rm rxn}E$, results with $\varepsilon = 2$ are distinguishable from those using larger values of ε, but only barely, and results with $\varepsilon = 4$ are practically indistinguishable from those obtained using $\varepsilon = 32$, especially for MBE(3). Quantitatively,

*herbert@chemistry.ohio-state.edu

for the three-body expansion that can provide converged results without a supersystem correction, errors in both E_a and $\Delta_{rxn}E$ are smaller than 1 kcal/mol for all values $\varepsilon \geq 2$. This rapid convergence as a function of ε is consistent with results for quantum-chemical cluster models of other enzymes.[3,4]

Figure S4 presents fragmentation errors, defined as fidelity with respect to a supersystem calculation at the same level of theory, for MBE(n) calculations on COMT. Several affordable functional and basis-set combinations are tested. Whereas the levels of theory used in the main body of this work (*e.g.*, in Fig. 3 and Table 1) represent model chemistries that one might routinely use in practice for large systems, here our goal is to test convergence of MBE(n), including up to four-body terms. Because MBE(4) calculations are quite expensive in the absence of any cutoffs, more affordable levels of theory are used in these tests. Table S4 provides the MBE(3) errors at a production level of theory (ωB97X-D/def2-SVP), corresponding to the plots in Fig. 1.

The use of a cutoff threshold (R_{cut}) is tested in Fig. S5, which illustrates how errors in E_a (for COMT) change as a function of R_{cut}. This is tested for several different functionals (Fig. S5a) and for various basis sets (Fig. S5b), although the results are not strongly dependent on either choice. Figure S6 illustrates how the number of subsystems required for a MBE(3) calculation is reduced as a function of R_{cut}.

S4 Timing Data

Table S5 contains the aggregate CPU time required for a single-point energy calculation in COMT at the ωB97X-D/def2-SVP level of theory. Data are provided for both a supersystem calculation (*in vacuo* and also using PCM boundary conditions), and for MBE(n) approximations. These calculations were run on 28-core nodes (Dell Intel Xeon E5-2680 v4), as described in the caption to Fig. 4 where the same data are presented as a bar graph. Supersystem calculations were performed on a single node.

References

[1] Liu, J.; Herbert, J. M. Pair–pair approximation to the generalized many-body expansion: An efficient and accurate alternative to the four-body expansion, with applications to *ab initio* protein energetics. *J. Chem. Theory Comput.* **2016,** *12,* 572–584.

[2] Thapa, B.; Beckett, D.; Jose, K. V. J.; Raghavachari, K. Assessment of fragmentation strategies for large proteins using the multilayer molecules-in-molecules approach. *J. Chem. Theory Comput.* **2018,** *14,* 1383–1394.

[3] Sevastik, R.; Himo, F. Quantum chemical modeling on enzymatic reactions: The case of 4-oxalocrotonate tautomerase. *Bioorg. Chem.* **2007,** *35,* 444–457.

[4] Dasgupta, S.; Herbert, J. M. Using atomic confining potentials for geometry optimization and vibrational frequency calculations in quantum-chemical models of enzyme active sites. *J. Phys. Chem. B* **2020,** *124,* 1137–1147.

List of Figures

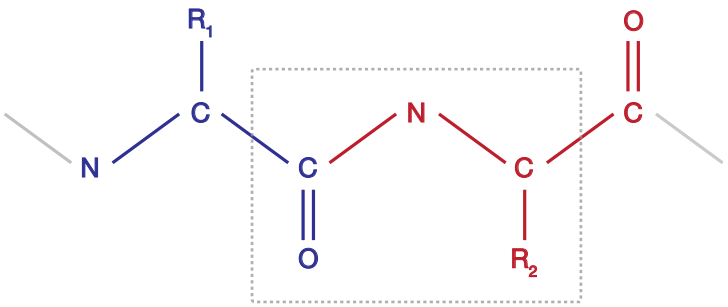

Figure S1: Fragment definition for the amino acid with side chain R_2.

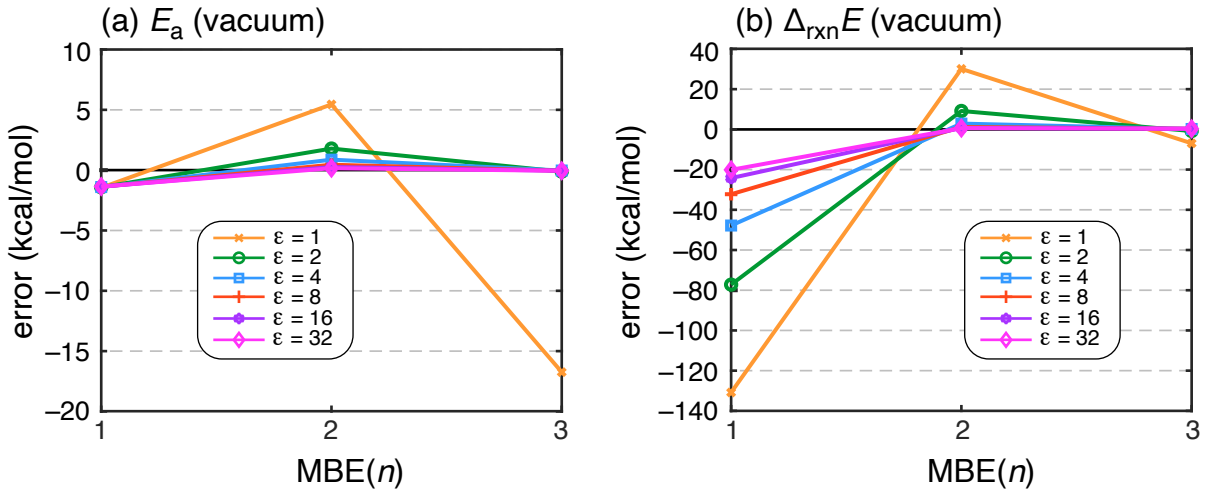

Figure S2: Errors in MBE(n) calculations of COMT at the ωB97X-D/def2-SVP level, as a function of the dielectric constant ε used for the PCM boundary conditions. The value $\varepsilon = 1$ corresponds to vacuum boundary conditions, and the $\varepsilon = 4$ data are the same as those plotted in Fig. 1a and 1b.

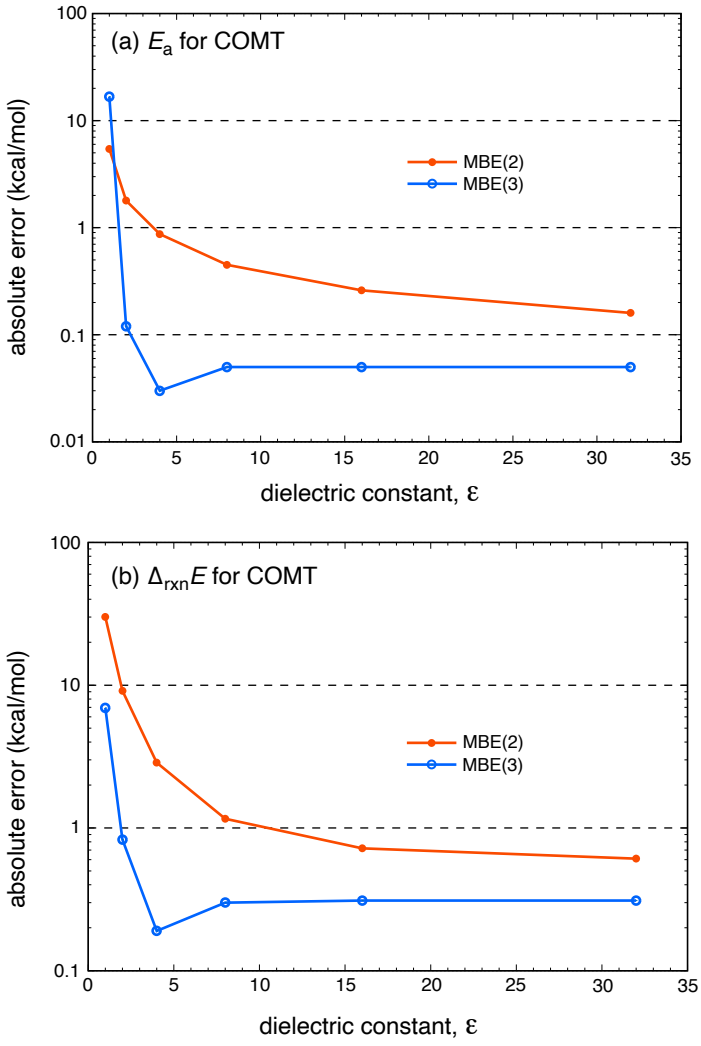

Figure S3: Absolute errors in MBE(2) and MBE(3) calculations of (a) E_a and (b) $\Delta_{rxn}E$ for COMT at the ωB97X-D/def2-SVP level, as a function of the dielectric constant ε used for the PCM boundary conditions. These are the same data as in Fig. S2, plotted in a different way.

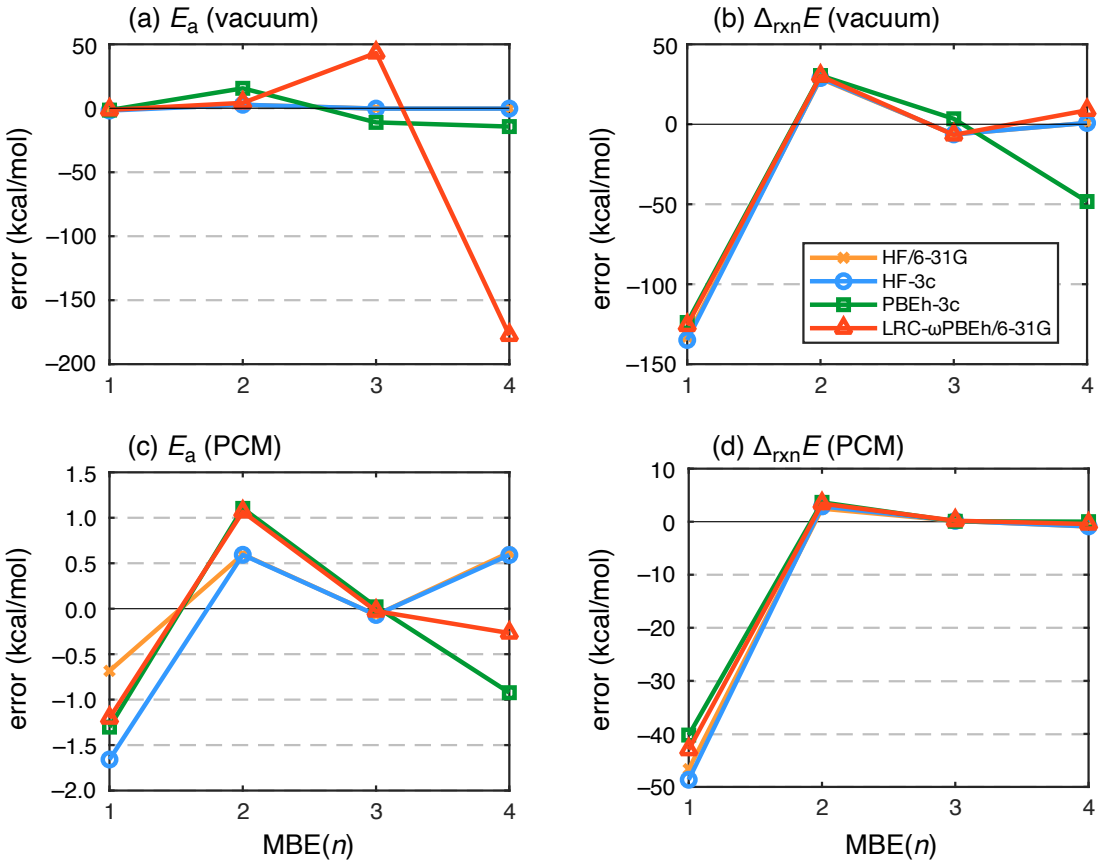

Figure S4: Errors in MBE(n) calculations for COMT, as compared to a supersystem benchmark at the same level of theory: (a) E_a with vacuum boundary conditions, (b) $\Delta_{rxn}E$ with vacuum boundary conditions, (c) E_a with PCM ($\varepsilon = 4$) boundary conditions, and (d) $\Delta_{rxn}E$ with $\varepsilon = 4$ boundary conditions.

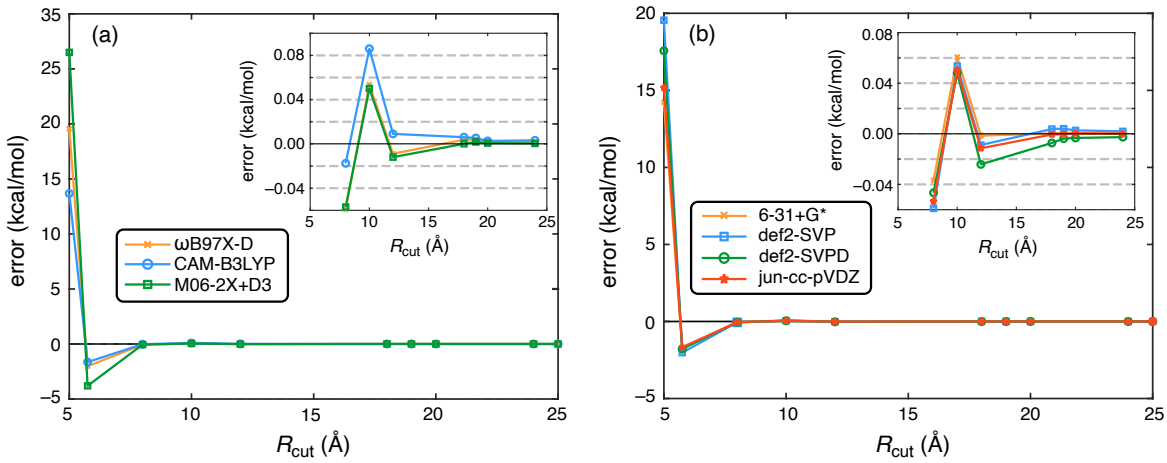

Figure S5: Error introduced in a MBE(3) calculation of E_a for COMT, as a function of the distance screening threshold R_{cut}. (a) Errors using different functionals in conjunction with the def2-SVP basis set. (b) Errors using ωB97X-D with various basis sets. In both cases, error is defined with respect to a MBE(3) calculation that does not employ cutoffs.

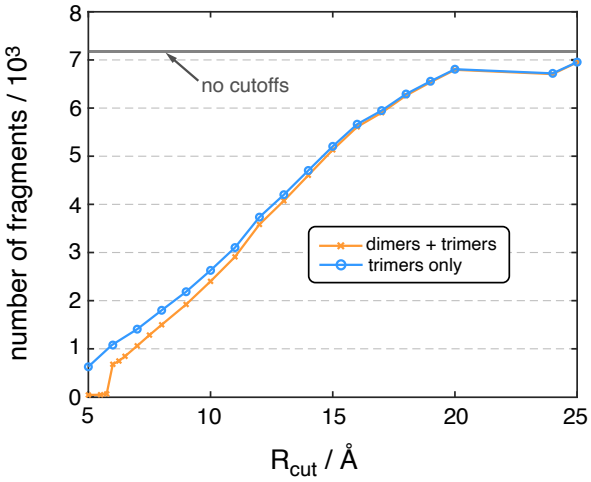

Figure S6: Number of fragments in a MBE(3) calculation of COMT, as a function of the distance cutoff threshold R_{cut}. Results are shown when the cutoff is applied to both dimers and trimers, and when only the trimers are subjected to a cutoff. In the absence of a cutoff, MBE(3) requires 7,175 subsystem calculations.

List of Tables

Table S1: Errors in E_a for COMT (in kcal/mol) using various MBE(n) methods.

Functional	Basis Set	Gas Phase			PCM ($\varepsilon = 4$)		
		MBE(1)	MBE(2)	MBE(3)	MBE(1)	MBE(2)	MBE(3)
CAM-B3LYP	6-31+G*	−1.03	5.26	15.72	−0.95	1.04	−0.09
	def2-SVP	−1.15	11.78	−21.50	−1.12	1.17	−0.10
	def2-SVPD	−1.07	3.99	30.46	−0.92	1.07	0.18
	jun-cc-pVDZ	−1.16	3.83	51.19	−0.86	1.09	−0.04
M06-2X +D3	6-31+G*	−3.20	4.53	20.59	−3.12	0.59	−0.10
	def2-SVP	−3.20	8.63	−5.00	−3.13	0.70	−0.03
	def2-SVPD	−3.19	4.01	14.44	−3.00	0.70	−0.04
	jun-cc-pVDZ	−3.29	4.24	20.45	−3.23	0.41	−0.26
ωB97X-D	6-31+G*	−1.54	3.75	15.97	−1.47	0.79	−0.13
	def2-SVP	−1.45	5.45	−16.74	−1.40	0.87	−0.03
	def2-SVPD	−1.48	3.58	10.59	−1.35	0.81	0.00
	jun-cc-pVDZ	−1.59	3.47	15.83	−2.02	0.11	−0.74

Table S2: Errors in E_a for COMT (in kcal/mol) using various MBE(n) methods.

Functional	Basis Set	Gas Phase			PCM ($\varepsilon = 4$)		
		MBE(1)	MBE(2)	MBE(3)	MBE(1)	MBE(2)	MBE(3)
CAM-B3LYP	6-31+G*	−128.1	33.48	−3.01	−46.68	2.84	0.99
	def2-SVP	−127.0	30.07	1.00	−43.85	2.64	0.34
	def2-SVPD	−127.3	34.56	−8.73	−47.03	2.71	−0.72
	jun-cc-pVDZ	−127.5	32.92	−5.01	−46.69	2.33	0.45
M06-2X+D3	6-31+G*	−133.5	34.31	−3.18	−51.40	3.81	0.93
	def2-SVP	−131.5	30.72	−1.87	−48.05	3.33	0.18
	def2-SVPD	−132.1	34.89	−8.62	−51.28	3.39	−0.90
	jun-cc-pVDZ	−132.9	33.60	−3.52	−51.27	3.18	0.50
ωB97X-D	6-31+G*	−131.7	33.57	−8.36	−50.05	3.15	0.72
	def2-SVP	−130.8	30.08	−6.94	−47.82	2.87	0.19
	def2-SVPD	−130.8	34.59	−11.68	−50.62	2.97	−0.98
	jun-cc-pVDZ	−131.3	33.01	−8.71	−50.72	2.26	−0.03

Table S3: ONIOM-style supersystem corrections for E_a and $\Delta_\mathrm{rxn}E$, for COMT. All calculations use PCM ($\varepsilon = 4$) boundary conditions.

Quantity	Method	Correction (kcal/mol)		
		MBE(1)	MBE(2)	MBE(3)
E_a	HF/6-31G	−0.62	0.61	−0.05
E_a	HF-3c	−1.84	0.82	−0.02
E_a	PBEh-3c	−1.30	1.10	0.02
E_a	LRC-ωPBEh/6-31G	−1.20	1.07	−0.03
$\Delta_\mathrm{rxn}E$	HF/6-31G	−52.73	1.99	0.13
$\Delta_\mathrm{rxn}E$	HF-3c	−56.29	−0.17	0.06
$\Delta_\mathrm{rxn}E$	PBEh-3c	−40.23	3.65	0.08
$\Delta_\mathrm{rxn}E$	LRC-ωPBEh/6-31G	−42.98	3.41	0.19

Table S4: MBE(3) errors for ωB97X-D/def2-SVP calculations.

System	Method	Error (kcal/mol)	
		E_a	$\Delta_\mathrm{rxn}E$
COMT	vacuum	−16.7	−6.9
	PCM ($\varepsilon = 4$)	−0.0	0.2
	charge-coordinated	−0.1	1.5
AspDC	vacuum	−0.6	−0.2
	PCM ($\varepsilon = 4$)	−0.3	−0.5

Table S5: Total CPU time for a single-point energy calculation on COMT at the ωB97X-D/def2-SVP level of theory.

Method	CPU Time (hours)			
	Supersys.	MBE(1)	MBE(2)	MBE(3)
Vacuum	92	1	68	1,503
PCM	143	2	90	2,025
Charge Coordinated		3	109	1,804