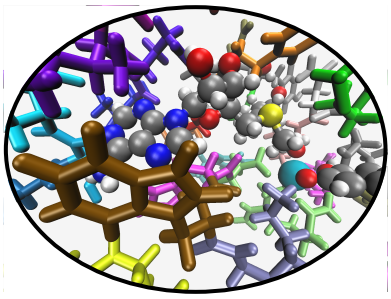
Fragment-Based Calculations of Enzymatic Thermochemistry Require Dielectric Boundary Conditions

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] *Dept. of Chemistry & Biochemistry, The Ohio State University, Columbus, Ohio 43210 USA*

Abstract

Quantum-chemical calculations of enzymatic thermochemistry require hundreds of atoms to obtain converged results, severely limiting the levels of theory that can be used. Fragment-based approaches offer a means to circumvent this problem, and we present calculations on enzyme models containing 500–600 atoms using the many-body expansion with three- and four-body terms. Results are compared to benchmarks in which the supramolecular enzyme–substrate complex is described at the same level of theory. When the amino acid fragments contain ionic side chains, the many-body expansion oscillates under vacuum boundary conditions, exaggerating the role of many-body effects. Rapid convergence is restored using low-dielectric boundary conditions. This implies that full-system calculations in the gas phase are inappropriate benchmarks for assessing errors introduced by fragment-based approximations. For calculations with dielectric boundary conditions, a three-body protocol with distance cutoffs retains sub-kcal/mol fidelity with respect to a supersystem calculation at the same level of theory, as does a two-body protocol when combined with a full-system correction at a low-cost level of theory. Both calculations dramatically reduce the cost of large-scale enzymatic thermochemistry, paving the way for application of high-level *ab initio* methods to very large systems.

TOC Graphic



Fragment-based approximations[1–6] represent an attractive way to circumvent the nonlinear scaling of computational quantum chemistry (QC), whose floating-point cost normally grows like $\mathcal{O}(N^p)$ as a function of system size (N), with exponents ranging from $p = 3$ for density functional theory (DFT) up to $p = 7$ or higher for levels of theory that provide thermochemical benchmarks. Fragmentation into $N_{\rm sub}$ separate subsystems, each of size n, reduces that cost to $N_{\rm sub} \times \mathcal{O}(n^p)$ in a manner that is amenable to distributed computing and which does not require modification to electronic structure codes. Nonlinear growth in $N_{\rm sub}$ with system size can be mitigated by means of distance- or energy-based thresholds.[6–8]

The present work presents a protocol for using fragmentation to compute enthalpy changes and activation barriers for enzyme-catalyzed reactions. Over the past decade, benchmark QM calculations have revealed that enzymatic thermochemistry does not converge until hundreds of atoms are included,[9–17] which is much larger than the QC region in typical quantum mechanics/molecular mechanics (QM/MM) calculations. Fragmentation may therefore offer an efficient route to obtain converged thermochemical calculations at benchmark levels of theory for $N > 500$ atoms, provided that errors associated with the fragmentation approximation can be controlled. The present work demonstrates that these errors can be reduced below the "thermochemical accuracy" threshold of 1 kcal/mol, yet highlights the fact that straightforward comparison of fragment-based approximations to full-system benchmarks (as a means to assess errors) is ill-posed, if the calculations are carried out with vacuum boundary conditions.

We will consider sizable models of enzyme–substrate complexes containing $N \sim 500$–600 atoms. Total energies are approximated by means of a many-body expan-

*herbert@chemistry.ohio-state.edu

sion (MBE),

$$E = \sum_{I=1} E_I + \sum_{I=1}\sum_{J>I} \Delta E_{IJ} + \sum_{I=1}\sum_{J>I}\sum_{K>J} \Delta E_{IJK} + \cdots . \quad (1)$$

Individual terms are

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \quad (2)$$

for the two-body corrections, where E_{IJ} is the energy of the dimer formed from fragments I and J, and

$$\Delta E_{IJK} = E_{IJK} - E_{IJ} - E_{IK} - E_{JK} - E_I - E_J - E_K \quad (3)$$

for the three-body corrections. Truncating eq. 1 at n-body terms, we will denote the resulting approximation as MBE(n). Electrostatic embedding of the subsystem calculations, using classical point charges derived from the fragment wave functions, is often used in an effort to hasten convergence of the MBE.[18–28] We avoid this, however, because we have found that charge embedding can lead to inconsistent convergence of the n-body expansion.[29–31] The use of self-consistent point charges also significantly complicates the formulation of analytic energy gradients.[32–36]

Fragments $I, J, K, \ldots$ are taken to be individual amino acids of the enzyme (except where stipulated otherwise, for testing purposes), with the substrate as its own fragment. Although larger fragments have sometimes been used for proteins,[37] we are able to achieve our target accuracy of 1 kcal/mol using mostly single-residue fragments, except for the substrate whose treatment is discussed below. Alternatively, overlapping fragments have sometimes been used for polypeptides and proteins,[24–28,37–42] which can be rationalized in terms of a generalized (G)MBE.[1,6,43,44] To date, most overlapping-fragment applications use a one-body approach that captures through-bond interactions but not through-space interactions. A two-body GMBE can capture both, but is relatively expensive in terms of the number of subsystems that are generated.[31,37] As such, we stick to the simple MBE(n) approach in this work.

As a first test, we consider S_N2 methyl transfer[45] catalyzed by then enzyme human catechol O-methyltransferase (COMT).[46–48] This particular enzyme has become something of a benchmark,[14,49–52] because it has a well-resolved crystal structure,[48] kinetics data,[46] and numerous known inhibitors.[14–16] A Mg^{2+} ion in the active site is essential to its function,[53] but leads to charge-transfer effects in QC calculations that can significantly alter the barrier height, depending on the size of the model system.[14,49,52] Kulik *et al.*[14] considered a sequence of COMT models with QM regions up to 940 atoms, and we selected "model 8" from Ref. 14, which contains 632 atoms and 35 fragments. The largest fragment consists of the octahedral coordination sphere around Mg^{2+}, including deprotonated catechol (2-hydroxyphenolate, $C_6H_5O_2^-$), two aspartic acid residues,

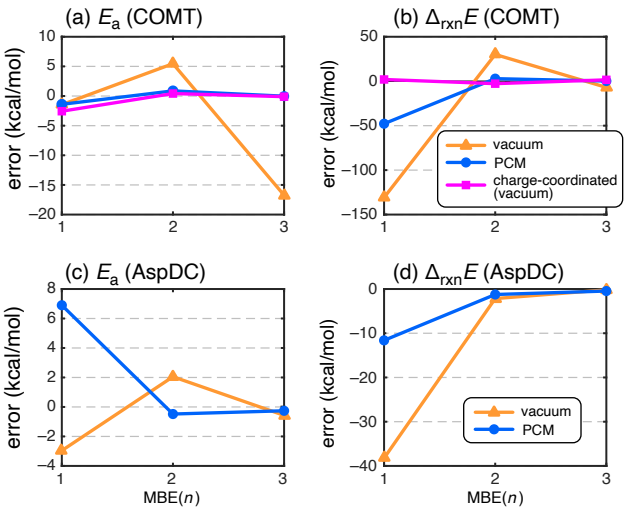


Figure 1: Errors in MBE(n) calculations at the ωB97X-D/def2-SVP level as compared to a supersystem calculation at the same level of theory. (a) E_a for COMT, (b) $\Delta_{rxn}E$ for COMT, (c) E_a for AspDC, and (d) $\Delta_{rxn}E$ for AspDC. For the COMT model, results are shown for vacuum boundary conditions ($\varepsilon = 1$) versus PCM boundary conditions ($\varepsilon = 4$), and also for a charge-coordinated model that uses larger, charge-neutral fragments. The AspDC system does not contain charged fragments.

an asparagine residue, and a water molecule (58 atoms). Reactant, product, and transition state structures for methyl transfer from *S*-adenosyl-L-methionine (SAM) to catecholate were protonated and relaxed as described in the Computational Details. All calculations were performed at DFT levels of theory, so that we may obtain energies for the full enzyme–substrate complex at the same level of theory and thereby examine convergence of MBE(n) towards a well-defined supersystem target. As such, the errors discussed below are defined with respect to a supersystem calculation at the same level of theory.

The overall charge on this QM model is -1 but the system contains 9 fragments with non-zero charge. Small anions in the gas phase are sometimes inherently unstable (or metastable), as in the case of SO_4^{2-},[54,55] and delocalization errors in DFT can exacerbate this problem.[55] To avoid artifacts, charged residues are often neutralized in fragment-based calculations on proteins.[56–58] This is not always a viable or realistic option, however, as charged side chains may be directly involved in stabilizing the protein structure or binding to a ligand (as in the present example), or may be vital to a reaction mechanism. A general procedure for enzymatic thermochemistry must admit the possibility of fragments with non-zero charge.

When we naively apply MBE(n) to a large COMT model with charged residues, however, we find that convergence is erratic. This is shown for the barrier height (activation energy E_a) in Fig. 1a, where MBE(2) overestimates the barrier by 5.4 kcal/mol but MBE(3) underestimates it by 16.7 kcal/mol. To verify that charged

residues are the problem, we prepared a second model of COMT in which fragments are combined to neutralize charge, *e.g.*, a negatively-charged aspartic acid residue is combined with a positively-charged ligand, forming a single fragment. This increases the largest fragment size from 58 to 124 atoms but does not change any protonation states. Using this "charge-coordinated" model of COMT, we observe rapid convergence of MBE(n), such that two- and three-body calculations afford essentially identical values of both E_a (Fig. 1a) and $\Delta_\mathrm{rxn}E$ (Fig. 1b). Even one-body calculations perform reasonably well for the charge-coordinated model, due to the larger fragment size, but enlarging the fragments is not an attractive strategy for levels of theory beyond DFT.

Each of the calculations described above was performed using vacuum boundary conditions. As an alternative, we introduce low-dielectric boundary conditions using a polarizable continuum model (PCM).[59] For protein electrostatics calculations based on the Poisson-Boltzmann equation, it is common to use a dielectric constant in the range $\varepsilon = 2$–4 to represent the hydrophobic interior of the protein,[60–65] although larger values have occassionally been suggested.[65–71] The precise value of ε may matter for pK_a calculations, but reaction barrier heights converge quickly as a function of ε and results for $\varepsilon = 2$ are often indistinguishable from much larger values,[72,73] although different from gas-phase ($\varepsilon = 1$) values.

When the fragment calculations required for MBE(n) are performed using PCM boundary conditions, and results compared to a supramolecular calculation with the same boundary conditions, we recover good convergence of MBE(n) even for single-residue fragments having net charge. Results for several other DFT functionals and basis sets are provided in Tables S1 and S2, and in Fig. S2 we extend some of these results to $n = 4$ in order to check convergence. Using PCM boundaries, the difference between the MBE(3) and MBE(4) results is $\lesssim 1$ kcal/mol, while gas-phase calculations sometimes afford errors $>$ 150 kcal/mol at the four-body level! MBE(3) calculations with low-dielectric boundary conditions consistently provide sub-kcal/mol accuracy for various functionals and basis sets, whereas MBE(3) with vacuum boundary conditions affords errors of 10–30 kcal/mol in many cases. Notably, we obtain stable results even when the basis set contains diffuse functions. These can be problematic when self-consistent charge-embedding schemes are used.[6,74–76]

These results suggest that large errors for enzymatic thermochemistry obtained using MBE(n) with vacuum boundary conditions originate not from the fragmentation approximation itself, or from the simple hydrogen atom caps that we use to saturate the severed valencies upon fragmentation. (This is less sophisticated as compared to "conjugated caps" that try to replicate amino acid moieties,[25–28,77,78] but our results demonstrate that sub-kcal/mol accuracy is achievable even with hydrogen atom caps.) Instead, errors arise due to inconsistent charge (de)localization in the n-body calculations for charged fragments. To obtain a polarization environment that is comparable to that of the supersystem, high-order n-body calculations are required, beyond $n = 4$. Alternatively, dielectric boundary conditions provide a simple and low-cost means to mimic this polarization. In principle, one might consider the use of heterogeneous dielectric boundaries,[79–81] such that hydrophobic parts of the protein are treated differently from solvent-exposed portions. This has not been pursued in the present work, where we simply aim to demonstrate that convergence of the MBE *in vacuo* is not well-defined.

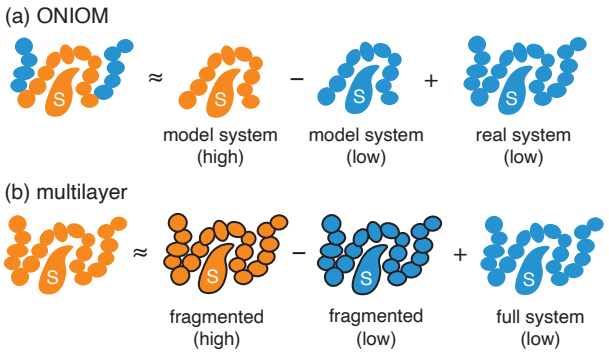


Figure 2: Multilayer techniques applied to the complex of an enzyme (depicted as a chain of amino acids) and a substrate (labeled "S"). Colors encode the level of theory, with the higher-level method in orange and the lower-level method in blue. (a) Conventional ONIOM method, in which high-level calculations are applied only to the model system (b) Multilayer fragmentation method, where the high-level method is applied to the entire system by means of fragmentation.

To confirm this explanation, we also examined a different enzymatic reaction that does not involve charged moieties near the active site. For this example we chose the decarboxylation of L-apartate by the enzyme L-aspartate α-decarboxylase (AspDC), which has also been studied using QC models of varying size.[82] Here, we consider only the C–C cleavage step, using a model consisting of 30 monomers (511 atoms), corresponding to a 5 Å radial cutoff around the active site of the relaxed crystal structure. This system has zero net charge but two ionic amino acids, which we placed together in a single fragment in order to avoid having any charged fragments. Results for E_a (Fig. 1c) and for $\Delta_\mathrm{rxn}E$ (Fig. 1d) demonstrate that n-body results with either vacuum ($\varepsilon = 1$) or PCM ($\varepsilon = 4$) boundary conditions converge similarly, although the PCM-based error is smaller at the $n = 2$ level. Unlike the charge-coordinated results for COMT, where the fragments are large and thus many-body effects are small, here the $n = 1$ results are unacceptable but two-body results with low-dielectric boundary conditions are rather good.

Together, these results demonstrate that the application of MBE(n) with vacuum boundaries, to an enzyme–substrate model extracted from a crystal structure, need

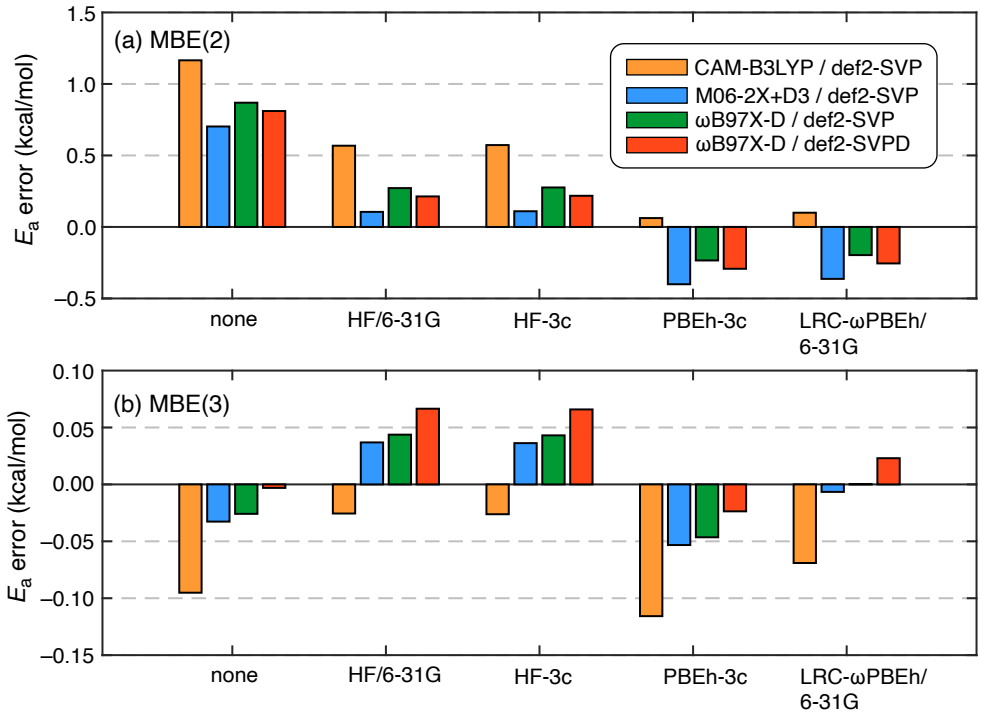


Figure 3: Errors in E_a for COMT, computed using two-layer fragmentation methods at the (a) MBE(2) or (b) MBE(3) level. Target levels of theory (used for the fragments) are indicated by the colored bars, and error is assessed with respect to a supersystem calculation at that level of theory. Several low-cost supersystem corrections are evaluated, as indicated along the horizontal axis. All calculations use PCM boundary conditions with $\varepsilon = 4$ without any distance-based cutoff applied the MBE(n) calculations.

not converge to the supermolecular result at low orders, $n \leq 4$. This behavior results from charge (de)localization that may vary greatly from monomer to dimer to trimer, etc., when the subsystems contain fragments with net charge. Elsewhere, low-dielectric boundary conditions ($\varepsilon \approx 1.5$) have been shown to reduce density delocalization error in isolated-peptide DFT calculations,[83] and in the present context the use of $\varepsilon = 4$ appears to prevent oscillatory changes the corrections $\Delta E_{IJ}, \Delta E_{IJK}, \ldots$.

Given a two- or three-body approximation for a large enzyme model possessing charged side chains, one might worry about neglect of long-range interactions. We address this by assessing a multilayer fragmentation scheme[6] in which a low-level calculation on the entire system is used to correct for errors introduced by fragmentation, while the subsystems are described at a higher level of theory. This strategy has been suggested by others under various names,[84–86] and is illustrated in Fig. 2 by analogy to the "ONIOM" approach for QM/MM calculations.[87] Both the subsystems and the supersystem are computed at the lower level of theory and the difference between low-level supersystem and low-level MBE(n) calculations provides a correction for the effects of fragmentation, including the possible neglect of long-range polarization. Raghavachari and co-workers have made extensive use of this idea for calculations in proteins,[56–58,88–90] and our two-layer procedure is equivalent to the "MIM2" strategy defined in Ref. 85.

We tested several low-level supersystem corrections in combination with four different target levels of DFT, for the activation energy in COMT. Errors with respect to the target level of DFT (applied to the entire enzyme–substrate model) are illustrated in Fig. 3 and numerical valuess for each supersystem correction can be found in Table S3. The low-level methods that we tested include the semi-empirical thrice-corrected methods HF-3c[91] and PBEh-3c,[92] which use a minimal and a double-ζ basis set, respectively. We also tested Hartree-Fock (HF) theory and the functional LRC-ωPBEh,[93] both with the 6-31G basis set. Note that 6-31G is much less expensive than other double-ζ basis sets,[94] if the electronic structure software can take advantage of compound *sp* shells. For this particular 632-atom enzyme–substrate complex, all four of these supersystem corrections require similar computational time, which constitutes less than 20% overhead on top of a MBE(2) calculation.

Even without the supersystem correction, results in Fig. 3a indicate that a two-body expansion can achieve ~ 1 kcal/mol accuracy for E_a using various density functionals. Low-cost supersystem corrections reduce this to ~ 0.5 kcal/mol. MBE(3) is an order-of-magnitude more accurate than MBE(2) and achieves ~ 0.1 kcal/mol accu-

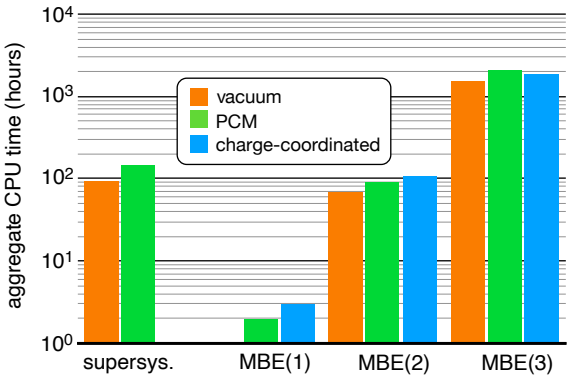


Figure 4: Total aggregate CPU time (on a logarithmic scale) for a single-point calculation on the COMT enzyme–substrate complex, at the ωB97X-D/def2-SVP level. The supersystem calculation contains 6,042 basis functions and was performed on a single 28-core node (Dell Intel Xeon E5-2680 v4). Fragment calculations were performed on the same hardware with 7 worker processes per node, each using 4 cores.

racy even without the supersystem correction. MBE(3) seems to represent something of an accuracy limit, as low-cost supersystem corrections no longer improve the results.

Importantly, HF-3c performs just as well as HF/6-31G as a supersystem correction despite using only a minimal basis set ("MINIX"[91]). For the 632-atom COMT enzyme–substrate complex, this means 1,944 basis functions for HF-3c/MINIX versus 3,510 functions for HF/6-31G. On a single 28-core node, these supersystem calculations can be completed in 0.6 h (HF-3c/MINIX) and 1.0 h (HF/6-31G), with 80–90% of that time spent in the PCM solver, which is less well-parallelized than the two-electron integrals. (The PCM cost could be reduced by using a less dense surface discretization.)

Having established that we can consistently obtain converged results, we next turn to computational efficiency. The cost of fragmentation methods is not always discussed honestly, and should be measured in aggregate computer time rather than wall time.[6,37,95] Timing data for single-point energy calculations on COMT are provided in Fig. 4, with the corresponding numerical data in Table S5. In the absence of any supersystem correction, MBE(2) with PCM boundary conditions costs about 60% as much a supersystem calculation at the same level of theory (ωB97X-D/def2-SVP), whereas MBE(3) is about 14× more expensive than the supersystem calculation. Despite using larger fragments, the charge-coordinated MBE(3) calculation is actually about 10% cheaper than MBE(3) with single-residue fragments, because the former calculation reduces the number of unique subsystems from 7,175 to 3,581. This balance would likely shift in favor of the single-residue calculation if a method more expensive than DFT were used, provided that good fidelity is maintained.

The number of subsystems required for MBE(n) grows as N^n for a protein with N residues, and this combinatorial growth imposes a severe computational bottleneck, even for $n = 3$.[31] In what follows, we screen the dimers and trimers based on distance, removing them from the calculation if the minimum interatomic distance between any two fragments exceeds a specified threshold, R_{cut}. We then recompute E_{a} and $\Delta_{\text{rxn}}E$ for COMT, with the caveat that we are careful to ensure that the same residues are included in the reactant, product, and transition state models. Tests of a distance-screened MBE(3) approximation (Fig. S3) demonstrate that the predicted value of E_{a} for COMT changes by < 0.1 kcal/mol as R_{cut} is reduced from 25 Å to 8 Å. Setting $R_{\text{cut}} = 8$ Å reduces the number of subsystems from 7,175 to 1,499 (as shown in Fig. S4), yet has negligible effect on accuracy. Setting $R_{\text{cut}} = 8$ Å, the computational effort is reduced from 2,025 h (which is the value shown in Fig. 4) to 657 h. This figure is still 5× greater than the cost of the corresponding supersystem calculation, however.

We include diffuse functions in our next set of tests (ωB97X-D/def2-SVPD), because a method that is intended for general application to enzymatic thermochemistry must be able to accommodate diffuse functions, in order to describe anionic side chains, yet these can be quite problematic for self-consistent charge schemes.[74,75] Even if electrostatic embedding charges are fixed (say, from a force field), the use of diffuse functions can lead to overpolarization of the QM system by the MM charges.[96] Errors in E_{a} for COMT, computed using MBE(2) and MBE(3) approximations, are provided in Table 1. This includes results both with and without a HF/6-31G supersystem correction, and also with and without distance-based screening using $R_{\text{cut}} = 8$ Å. We have also tabulated errors with respect to a ωB97X-D/def2-TZVP calculation, which provides a measure of the basis-set incompleteness error when the smaller def2-SVPD basis set is used.

Both the MBE(2) and MBE(3) approximations at the ωB97X-D/def2-SVPD level afford sub-kcal/mol error with respect to supersystem results using the larger def2-TZVP basis set, suggesting that the basis-set incompleteness error is < 1 kcal/mol. MBE(3) achieves this feat without a supersystem correction, inclusion of which scarcely alters the results, whereas the supersystem correction does afford a small but noticeable improvement to MBE(2). It is worth noting that the supersystem ωB97X-D/def2-TZVP calculation on this 632-atom model consists of 11,767 basis functions and requires an aggregate computation time of 17,546 h running on a single 40-core node.

These results once again demonstrate that consistent, sub-kcal/mol accuracy is achievable is two ways: MBE(3) alone, or MBE(2) with a supersystem correction. Distance cutoffs with $R_{\text{cut}} = 8$ Å can safely be applied in either case. This consistency indicates that the supersystem correction (which is performed at the HF/6-31G level for the calculations reported in Table 1) primarily accounts for three-body polarization, and that four-

Table 1: Errors in E_a for COMT, Computed at the ωB97X-D/def2-SVPD Level.[a]

Method	Error in E_a (kcal/mol)		CPU Time
	vs. same method[b]	vs. triple-ζ[c]	(hours)[d]
MBE(3)	−0.00	0.76	11,253
MBE(3) + HF/6-31G[e]	0.04	0.83	11,656
MBE(3) + 8 Å[f]	−0.05	0.71	4,170
MBE(3) + HF/6-31G[e] + 8 Å[f]	0.04	0.80	4,300
MBE(2)	0.81	1.57	460
MBE(2) + HF/6-31G[e]	0.20	0.96	491
MBE(2) + 8 Å[f]	0.83	1.59	354
MBE(2) + HF/6-31G[e] + 8 Å[f]	0.20	0.96	379

[a]All calculations were performed using a PCM with $\varepsilon = 4$. [b]Error with respect to a supersystem calculation at the same level of theory. [c]Error with respect to a supersystem calculation at the ωB97X-D/def2-TZVP level. [d]Aggregate computer time for one single-point energy calculation, using a single 48-core node (Intel Xeon Platinum 8268). Fragment calculations employ 12 worker processes, each running on 4 cores. [e]HF/6-31G as a supersystem correction. [f]$R_{cut} = 8$ Å screening threshold.

body terms make a negligible (sub-kcal/mol) contribution when PCM boundaries are applied. (See Fig. S2.) Of these two high-fidelity fragment-based procedures, MBE(2) with cutoffs and a supersystem correction is more affordable, by 11× as compared to MBE(3) with cutoffs and no supersystem correction. Although the best measure of real-world cost is total (aggregate) time across all processors, if one wants to use throughput as the figure of merit then it is worth noting that the 379 h required for the supersystem-corrected MBE(2) calculation corresponds to 329 distinct subsystems that can be distributed across compute nodes.

In summary, we find that low-dielectric boundary conditions lead to rapid convergence of the many-body expansion, which otherwise suffers from oscillatory behavior in the presence of charged fragments. Larger, charge-neutral fragments can be used as an alternative strategy to avoid these oscillations, but this will significantly increase the cost if a correlated wave function method is used for the two-body interactions. At the same time, ionic residues must be anticipated in general, and this makes dielectric boundary conditions effectively mandatory for QC calculations of enzymatic thermochemistry. These observations furthermore suggest that the use of gas-phase supersystem calculations to benchmark fragmentation approximations distorts the performance of those approximations. Where charged fragments are involved, comparison to a gas-phase benchmark may exaggerate the role of higher-order n-body terms.

When dielectric boundaries are employed, MBE(3) provides converged results with sub-kcal/mol fidelity, without the need for electrostatic embedding, conjugated caps, or an ONIOM-style supersystem correction, and using single-residue fragments for most of the protein. This relatively simple three-body approach represents a reliable fall-back procedure for systems that are too large even for conventional DFT. That said, even for a 632-atom enzyme–substrate model, a full-system DFT calculation is far less expensive when a high-performance electronic structure code is used. A practical alternative to MBE(3) is MBE(2) with distance screening, in a double-ζ basis set, plus an ONIOM-style supersystem correction at the HF/6-31G level. This composite approach is converged below 1 kcal/mol with respect to a triple-ζ benchmark and is generally less expensive than the full-system calculation. Moreover, that cost is readily distributable across hardware.

In the end, we find that enzymatic thermochemistry can be reproduced with sub-kcal/mol fidelity using practical protocols based on fragmentation. The stage is now set to push the accuracy of these calculations beyond the DFT level, by means of hybrid methods that deploy high-level methods for the two-body interactions combined with three-body DFT to capture polarization by the environment. We are also exploring the use of fragment-based vibrational frequency calculations, as pioneered by others,[97–99] to include zero-point corrections and finite-temperature thermal corrections. (The use of smooth cutoffs in gradient calculations has already been demonstrated.[7]) Network analysis can be used to build sensible (if sizable) models of the enzyme–substrate complex,[52,100,101] and then the protocols developed here can provide converged results for any given model. Together, these developments promise to make QC modeling of enzymatic reactions more robust and systematic.

Computational Details

The crystal structure[48] of COMT (PDB code: 3BMW) was protonated using the H++ web server[102] at pH = 7.0, salinity of 0.15 M, $\varepsilon_{in} = 10$, and $\varepsilon_{out} = 80$. Ligand atoms were protonated separately using PyMOL, then validated against reactant and product structures taken from Ref. 14. As in that work, the inhibitor 3,5-dinitrocatechol in the crystal structure was re-

placed with catecholate ($C_6H_5O_2^-$). Reactant and product structures were relaxed using the GFN2-xTB semi-empirical model[103] with a generalized Born/surface area (GBSA) implicit solvent model for water.[104] GFN2-xTB with implicit solvent has been recently benchmarked for protein structure, with results that compare well to experiment.[105] To obtain the transition state, we scanned the bond length between the sulfur atom on SAM and the transferred methyl group. The system was then trimmed to obtain the 632-atom "model 8" from Ref. 14, which contains residues within a 5 Å radius of active site along with three important residues identified experimentally. This model affords converged energetics with respect to larger models.[14]

For AspDC (PDB code: 1UHE),[106] a single monomer unit can be directly downloaded from the protein database although the complete structure is an octamer. Starting from the latter, a large radial cutoff of 12 Å was used for structure relaxation using GFN2-xTB in implicit solvent. From that relaxed structure, a smaller 5 Å region was created for a scan along the bond-breaking coordinate, and from that scan a transition state and a product structure were selected. For fragmentation calculations, the negatively charged ligand and the cationic arginine residue coordinated to it were included in a single fragment, such that all fragments are uncharged.

In creating fragments, we avoid cutting the polar C–N peptide bond (following previous recommendations),[37,56] and instead create fragments by cutting the C–C bond at C_α, as indicated in Fig. S1. The severed valency is capped with a hydrogen atom that is positioned according to eq. S1, as in previous work.[37]

All QM calculations were run using a home-built interface (PyFragme∩t) to Q-Chem.[107] For all calculations, the self-consistent field convergence threshold is set to $\tau_{\mathrm{SCF}} = 10^{-8}$ Ha and the integral and shell-pair drop tolerances are set to $\tau_{\mathrm{ints}} = 10^{-12}$ a.u. We use the conductor-like PCM (C-PCM),[59] implemented with the switching/Gaussian discretization scheme.[108–111] The continuum interface is defined by a van der Waals cavity,[59] constructed using modified Bondi atomic radii[112] that are scaled by a factor of 1.2. That surface is discretized using 110 Lebedev points for hydrogen and 194 points for other atoms.[108] A conjugate gradient algorithm was used to solve the C-PCM equations for the full protein model,[111] whereas matrix inversion was used for the subsystem C-PCM calculations. Calculations with ωB97X-D and M06-2X+D3 use the SG-2 quadrature grid,[113] whereas SG-1[114] is used for other functionals.

Supporting Information

Additional data including convergence tests with various functionals and basis sets.

Notes

The authors declare the following competing financial interest(s): J.M.H. serves on the board of directors of Q-Chem Inc.

Acknowledgments

We thank Atsu Agbaglo at the University of Memphis for assistance in creating the octomer of AspDC from the crystal structure of the monomer. This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award No. DE-SC0008550, Calculations were performed at the Ohio Supercomputer Center.[115]

References

[1] Richard, R. M.; Herbert, J. M. A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory. *J. Chem. Phys.* **2012,** *137,* 064113:1–17.

[2] Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.* **2012,** *112,* 632–672.

[3] Collins, M. A.; Bettens, R. P. Energy-based molecular fragmentation methods. *Chem. Rev.* **2015,** *115,* 5607–5642.

[4] Raghavachari, K.; Saha, A. Accurate composite and fragment-based quantum chemical methods for large molecules. *Chem. Rev.* **2015,** *115,* 5643–5677.

[5] Fang, T.; Li, Y.; Li, S. Generalized energy-based fragmentation approach for modeling condensed phase systems. *WIREs Comput. Mol. Sci.* **2017,** *7,* e1297:1–13.

[6] Herbert, J. M. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.* **2019,** *151,* 170901:1–38.

[7] Liu, K.-Y.; Herbert, J. M. Understanding the many-body expansion for large systems. III. Critical role of four-body terms, counterpoise corrections, and cutoffs. *J. Chem. Phys.* **2017,** *147,* 161729:1–13.

[8] Liu, K.-Y.; Herbert, J. M. Energy-screened many-body expansion: A practical yet accurate fragmentation method for quantum chemistry. *J. Chem. Theory Comput.* **2020,** *16,* 475–487.

[9] Hu, L.; Eliasson, J.; Heimdal, J.; Ryde, U. Do quantum mechanical energies calculated for small models of protein-active sites converge? *J. Phys. Chem. A* **2009,** *113,* 11793–11800.

[10] Hu, L.; Söderhjelm, P.; Ryde, U. On the convergence of QM/MM energies. *J. Chem. Theory Comput.* **2011,** *7,* 761–777.

[11] Sumner, S.; Söderhjelm, P.; Ryde, U. Effect of geometry optimizations on QM-cluster and QM/MM studies of reaction energies in proteins. *J. Chem. Theory Comput.* **2013,** *9,* 4205–4214.

[12] Liao, R.-Z.; Thiel, W. Comparison of QM-only and QM/MM models for the mechanism of tungsten-dependent acetylene hydratase. *J. Chem. Theory Comput.* **2012,** *8,* 3793–3803.

[13] Liao, R.-Z.; Thiel, W. Convergence in the QM-only and QM/MM modeling of enzymatic reactions: A case study for acetylene hydratase. *J. Comput. Chem.* **2013,** *34,* 2389–2397.

[14] Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. How large should the QM region be in QM/MM calculations? The case of catechol *o*-methyltransferase. *J. Phys. Chem. B* **2016,** *120,* 11381–11394.

[15] Karelina, M.; Kulik, H. J. Systematic quantum mechanical region determination in QM/MM simulation. *J. Chem. Theory Comput.* **2017,** *13,* 563–576.

[16] Kulik, H. J. Large-scale QM/MM free energy simulations of enzyme catalysis reveal the influence of charge transfer. *Phys. Chem. Chem. Phys.* **2018,** *20,* 20650–20660.

[17] Yang, Z.; Mehmood, R.; Wang, M.; Qi, H. W.; Steeves, A. H.; Kulik, H. J. Revealing quantum mechanical effects in enzyme catalysis with large-scale electronic structure simulation. *React. Chem. Eng.* **2019,** *4,* 298–315.

[18] Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput.* **2007,** *3,* 46–53.

[19] Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body correlation energy, with applications to the calculation of accurate second-order Møller–Plesset perturbation theory energies for large water clusters. *J. Chem. Theory Comput.* **2007,** *3,* 1342–1348.

[20] Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body expansions for simulations. *J. Chem. Theory Comput.* **2008,** *4,* 1–6.

[21] Leverentz, H. R.; Truhlar, D. G. Electrostatically embedded many-body approximation for systems of water, ammonia, and sulfuric acid and the dependence of its performance on embedding charges. *J. Chem. Theory Comput.* **2009,** *5,* 1573–1584.

[22] Nagata, T.; Fedorov, D. G.; Kitaura, K. Mathematical formulation of the fragment molecular orbital method. In *Linear-Scaling Techniques in Computational Chemistry and Physics*, Vol. 13; Zaleśny, R.; Papadopoulos, M. G.; Mezey, P. G.; Leszczynski, J., Eds.; Springer: New York, 2011; Chapter 2, pages 17–64.

[23] Fedorov, D. G.; Asada, N.; Nakanishi, I.; Kitaura, K. The use of many-body expansions and geometry optimizations in fragment-based methods. *Acc. Chem. Res.* **2014,** *47,* 2846–2856.

[24] Li, S.; Li, W.; Ma, J. Generalized energy-based fragmentation approach and its applications to macromolecules and molecular aggregates. *Acc. Chem. Res.* **2014,** *47,* 2712–2720.

[25] He, X.; Zhu, T.; Wang, X. W.; Liu, J. F.; Zhang, J. Z. H. Fragment quantum mechanical calculation of proteins and its applications. *Acc. Chem. Res.* **2014,** *47,* 2748–2757.

[26] Liu, J.; Zhu, T.; He, X.; Zhang, J. Z. H. MFCC-based fragmentation methods for biomolecules. In *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*; Gordon, M. S., Ed.; Wiley: 2017; Chapter 11, pages 323–348.

[27] Jin, X.; Glover, W. J.; He, X. Fragment quantum mechanical method for excited states of proteins: Development and application to the green fluorescent protein. *J. Chem. Theory Comput.* **2020,** *16,* 5174–5188.

[28] Shen, C.; Jin, X.; Glover, W. J.; He, X. Accurate prediction of absorption spectral shifts of proteorhodopsin using a fragment-based quantum mechanical method. *Molecules* **2021,** *26,* 4486:1–19.

[29] Richard, R. M.; Lao, K. U.; Herbert, J. M. Aiming for benchmark accuracy with the many-body expansion. *Acc. Chem. Res.* **2014,** *47,* 2828–2836.

[30] Richard, R. M.; Lao, K. U.; Herbert, J. M. Understanding the many-body expansion for large systems. I. Precision considerations. *J. Chem. Phys.* **2014,** *141,* 014108:1–14.

[31] Lao, K. U.; Liu, K.-Y.; Richard, R. M.; Herbert, J. M. Understanding the many-body expansion for large systems. II. Accuracy considerations. *J. Chem. Phys.* **2016,** *144,* 164105:1–15.

[32] Parandekar, P. V.; Hratchian, H. P.; Raghavachari, K. Applications and assessment of QM:QM electronic embedding using generalized asymmetric Mulliken atomic charges. *J. Chem. Phys.* **2008,** *129,* 145101:1–10.

[33] Nagata, T.; Brorsen, K.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Fully analytic energy gradient in the fragment molecular orbital method. *J. Chem. Phys.* **2011,** *134,* 124115:1–13.

[34] Brorsen, K. R.; Zahariev, F.; Nakata, H.; Fedorov, D. G.; Gordon, M. S. Analytic gradient for density functional theory based on the fragment molecular orbital method. *J. Chem. Theory Comput.* **2014,** *10,* 5297–5307.

[35] Liu, J.; Rana, B.; Liu, K.-Y.; Herbert, J. M. Variational formulation of the generalized many-body expansion with self-consistent embedding charges: Simple and correct analytic energy gradient for fragment-based *ab initio* molecular dynamics. *J. Phys. Chem. Lett.* **2019,** *10,* 3877–3886.

[36] Holden, Z. C.; Rana, B.; Herbert, J. M. Analytic energy gradients for the QM/MM-Ewald method using atomic charges derived from the electrostatic potential: Theory, implementation, and application to *ab initio* molecular dynamics of the aqueous electron. *J. Chem. Phys.* **2019,** *150,* 144115:1–20.

[37] 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.

[38] Sahu, N.; Gadre, S. R. Vibrational infrared and Raman spectra of polypeptides: Fragments-in-fragments within molecular tailoring approach. *J. Chem. Phys.* **2016,** *144,* 114113:1–16.

[39] Xu, M.; He, X.; Zhu, T.; Zhang, J. Z. H. A fragment quantum mechanical method for metalloproteins. *J. Chem. Theory Comput.* **2019,** *15,* 1430–1439.

[40] Bozkaya, U.; Ermiş, B. Linear-scaling systematic molecular fragmentation approach for perturbation theory and coupled-cluster methods. *J. Chem. Theory Comput.* **2022,** *18,* 5349–5359.

[41] Du, J.; Liao, K.; Ma, J.; Li, W.; Li, S. Generalized energy-based fragmentation approach for the electronic emission spectra of large molecules. *J. Chem. Theory Comput.* **2022,** *18,* 7630–7638.

[42] Chen, W.-K.; Fang, W.-H.; Cui, G. Extending multilayer energy-based fragment method for excited-state calculations of large covalently bonded fragment systems. *J. Chem. Phys.* **044110:1–13,** *158,*.

[43] Richard, R. M.; Herbert, J. M. The many-body expansion with overlapping fragments: Analysis of two approaches. *J. Chem. Theory Comput.* **2013,** *9,* 1408–1416.

[44] Jacobson, L. D.; Richard, R. M.; Lao, K. U.; Herbert, J. M. Efficient monomer-based quantum chemistry methods for molecular and ionic clusters. *Annu. Rep. Comput. Chem.* **2013,** *9,* 25–58.

[45] Hegazi, M. F.; Borchardt, R. T.; Schowen, R. L. S_N2-like transition state for methyl transfer catalyzed by catechol-*O*-methyltransferase. *J. Am. Chem. Soc.* **1976,** *98,* 3048–3049.

[46] Lotta, T.; Vidgren, J.; Tilgmann, C.; Ulmanen, I.; Melen, K.; Julkunen, I.; Taskinen, J. Kinetics of human soluble and membrane-bound catechol *O*-methyltransferase: A revised mechanism and description of the thermolabile variant of the enzyme. *Biochemistry* **1995,** *34,* 4202–4210.

[47] Bonifácio, M. J.; Archer, M.; Rodrigues, M. L.; Matias, P. M.; Learmouth, D. A.; Carrondo, M. A.; Soares-da-Silva, P. Kinetics and crystal structure of catechol *O*-methyltransferase complex with co-substrate and a novel inhibitor with potential therapeutic application. *Mol. Pharmacol.* **2002,** *62,* 795–805.

[48] Rutherford, K.; Le Trong, I.; Stenkamp, R. E.; Parson, W. W. Crystal structures of human 108V and 108M catechol *O*-methyltransferase. *J. Mol. Biol.* **2008,** *380,* 120–130.

[49] Jindal, G.; Warshel, A. Exploring the dependence of QM/MM calculations of enzyme catalysis on the size of the QM region. *J. Phys. Chem. B* **2016,** *120,* 9913–9921.

[50] Zhang, J.; Kulik, H. J.; Martinez, T. J.; Klinman, J. P. Mediation of donor–acceptor distance in an enzymatic methyl transfer reaction. *Proc. Natl. Acad. Sci. USA* **2015,** *112,* 7954–7959.

[51] Patra, N.; Ioannidis, E. I.; Kulik, H. J. Computational investigation of the interplay of substrate positioning and reactivity in catechol o-methyltransferase. *PLoS ONE* **2016,** *11,* e0161868:1–23.

[52] Summers, T. J.; Cheng, Q.; Palma, M. A.; Pham, D.-T.; Kelso III, D. K.; Webster, C. E.; DeYonker, N. J. Cheminformatic quantum mechanical enzyme model design: A catechol-O-methyltransferase case study. *Biophys. J.* **2021,** *120,* 3577–3587.

[53] Axelrod, J.; Tomchick, R. Enzymatic O-methylation of epinephrine and other catechols. *J. Biol. Chem.* **1958,** *233,* 702–705.

[54] Whitehead, A.; Barrios, R.; Simons, J. Stabilization calculation of the energy and lifetime of metastable SO_4^{2-}. *J. Chem. Phys.* **2002,** *116,* 2848–2851.

[55] Herbert, J. M. The quantum chemistry of loosely-bound electrons. In *Reviews in Computational Chemistry*, Vol. 28; Parill, A. L.; Lipkowitz, K., Eds.; Wiley-VCH: Hoboken, NJ, 2015; Chapter 8, pages 391–517.

[56] 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.

[57] Thapa, B.; Beckett, D.; Erickson, J.; Raghavachari, K. Theoretical study of protein–ligand interactions using the molecules-in-molecules fragmentation-based method. *J. Chem. Theory Comput.* **2018,** *14,* 5143–5155.

[58] Thapa, B.; Raghavachari, K. Energy decomposition analysis of protein–ligand interactions using molecules-in-molecules fragmentation-based method. *J. Chem. Inf. Model.* **2019,** *59,* 3474–3484.

[59] Herbert, J. M. Dielectric continuum methods for quantum chemistry. *WIREs Comput. Mol. Sci.* **2021,** *11,* e1519:1–73.

[60] Gilson, M. K.; Honig, B. H. The dielectric constant of a folded protein. *Biopolymers* **1986,** *25,* 2097–2119.

[61] Gilson, M. K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins* **1988,** *4,* 7–18.

[62] Rodgers, K. K.; Silgar, S. G. Surface electrostatics, reduction potentials, and internal dielectric constant of protiens. *J. Am. Chem. Soc.* **1991,** *113,* 9419–9421.

[63] Nakamura, H. Roles of electrostatic interaction in proteins. *Q. Rev. Biophys.* **1996,** *29,* 1–90.

[64] Grochowski, P.; Trylska, J. Continuum molecular electrostatics, salt effects, and counterion binding—A review of the Poisson–Boltzmann theory and its modifications. *Biopolymers* **2008,** *89,* 93–113.

[65] Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Nielsen, J. E.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. Progress in the prediction of pK_a values in proteins. *Proteins* **2011,** *79,* 3260–3275.

[66] King, G.; Lee, F. S.; Warshel, A. Microscopic simulations of macroscopic dielectric constants of solvated proteins. *J. Chem. Phys.* **1991,** *95,* 4366–4377.

[67] Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* **1994,** *238,* 415–436.

[68] Demchuk, E.; Wade, R. C. Improving the continuum dielectric approach to calculating pK_as of ionizable groups in proteins. *J. Phys. Chem.* **1996,** *100,* 17373–17387.

[69] Grycuk, T. Revision of the model system concept for the prediction of pK_a's in proteins. *J. Phys. Chem. B* **2002,** *106,* 1434–1445.

[70] Truchon, J.-F.; Nicholls, A.; Roux, B.; Iftimie, R. I.; Bayly, C. I. Integrated continuum dielectric approaches to treat molecular polarizability and the condensed phase: Refractive index and implicit solvation. *J. Chem. Theory Comput.* **2009,** *5,* 1785–1802.

[71] Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the dielectric "constant" of proteins: Smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J. Chem. Theory Comput.* **2013,** *9,* 2126–2136.

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

[73] 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.

[74] Fedorov, D. G.; Slipchenko, L. V.; Kitaura, K. Systematic study of the embedding potential description in the fragment molecular orbital method. *J. Phys. Chem. A* **2010,** *114,* 8742–8753.

[75] Fedorov, D. G.; Kitaura, K. Use of an auxiliary basis set to describe the polarization in the fragment molecular orbital method. *Chem. Phys. Lett.* **2014,** *597,* 99–105.

[76] Holden, Z. C.; Richard, R. M.; Herbert, J. M. Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets. *J. Chem. Phys.* **2013,** *139,* 244108:1–13 Erratum: *ibid.* **142,**

059901:1–2 (2015).

[77] Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy. *J. Chem. Phys.* **2003,** *119,* 3599–3605.

[78] Jiang, N.; Ma, J.; Jiang, Y. Electrostatic field-adapted molecular fractionation with conjugated caps for energy calculations of charged biomolecules. *J. Chem. Phys.* **2006,** *124,* 114112:1–9.

[79] Iozzi, M. F.; Cossi, M.; Improta, R.; Rega, N.; Barone, V. A polarizable continuum approach for the study of heterogeneous dielectric environments. *J. Chem. Phys.* **2006,** *124,* 184103.

[80] Si, D.; Li, H. Heterogeneous conductorlike solvation model. *J. Chem. Phys.* **2009,** *131,* 044123:1–8.

[81] Coons, M. P.; Herbert, J. M. Quantum chemistry in arbitrary dielectric environments: Theory and implementation of nonequilibrium Poisson boundary conditions and application to compute vertical ionization energies at the air/water interface. *J. Chem. Phys.* **2018,** *148,* 222834:1–21 Erratum: *J. Chem. Phys.* **151**, 189901:1–2 (2019).

[82] Liao, R. Z.; Yu, J. G.; Himo, F. Quantum chemical modeling of enzymatic reactions: The case of decarboxylation. *J. Chem. Theory Comput.* **2011,** *7,* 1494–1501.

[83] Ren, F.; Liu, F. Impacts of polarizable continuum models on the SCF convergence and DFT delocalization error of large molecules. *J. Chem. Phys.* **2022,** *157,* 184106:1–11.

[84] Tschumper, G. S. Multicentered integrated QM:QM methods for weakly bound clusters: An efficient and accurate 2-body:many-body treatment of hydrogen bonding and van der Waals interactions. *Chem. Phys. Lett.* **2006,** *427,* 185–191.

[85] Mayhall, N. J.; Raghavachari, K. Molecules-in-molecules: An extrapolated fragment-based approach for accurate calculations on large molecules and materials. *J. Chem. Theory Comput.* **2011,** *7,* 1336–1343.

[86] Sahu, N.; Gadre, S. R. Molecular tailoring approach: A route for *ab initio* treatment of large clusters. *Acc. Chem. Res.* **2014,** *47,* 2739–2747.

[87] Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM method and its applications. *Chem. Rev.* **2015,** *115,* 5678–5796.

[88] Saha, A.; Raghavachari, K. Analysis of different fragmentation strategies on a variety of large peptides: Implementation of a low level of theory in fragment-based methods can be a crucial factor. *J. Chem. Theory Comput.* **2015,** *11,* 2012–2023.

[89] Jose, K. V. J.; Raghavachari, K. Fragment-based approach for the evaluation of NMR chemical shifts for large biomolecules incorporating the effects of the solvent environment. *J. Chem. Theory Comput.* **2017,** *13,* 1147–1158.

[90] Chandy, S. K.; Thapa, B.; Raghavachari, K. Accurate and cost-effective NMR chemical shift predictions for proteins using a molecules-in-molecules fragmentation-based method. *Phys. Chem. Chem. Phys.* **2020,** *22,* 27781–27799.

[91] Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013,** *34,* 1672–1685.

[92] Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.* **2015,** *143,* 054107:1–19.

[93] Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A long-range-corrected density functional that performs well for both ground-state properties and time-dependent density functional theory excitation energies, including charge-transfer excited states. *J. Chem. Phys.* **2009,** *130,* 054112:1–8.

[94] Gray, M.; Herbert, J. M. Comprehensive basis-set testing of extended symmetry-adapted perturbation theory and assessment of mixed-basis combinations to reduce cost. *J. Chem. Theory Comput.* **2022,** *18,* 2308–2330.

[95] Gavini, V. *et al.* Roadmap on electronic structure codes in the exascale era. **2022,** (arXiv:2209.12747).

[96] Lin, H.; Truhlar, D. G. QM/MM: What have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007,** *117,* 185–199.

[97] Jose, K. V. J.; Raghavachari, K. Evaluation of energy gradients and infrared vibrational spectra through molecules-in-molecules fragment-based approach. *J. Chem. Theory Comput.* **2015,** *11,* 950–961.

[98] Jose, K. V. J.; Raghavachari, K. Molecules-in-molecules fragment-based method for the evaluation of Raman spectra of large molecules. *Mol. Phys.* **2015,** *113,* 3057–3066.

[99] Jose, K. V. J.; Raghavachari, K. Molecules-in-molecules fragment-based method for the accurate evaluation of vibrational and chiroptical spectra for large molecules. In *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*; Gordon, M. S., Ed.; Wiley: 2017; Chapter 4, pages 141–164.

[100] Summers, T. J.; Daniel, B. P.; Cheng, Q.; DeYonker, N. J. Quantifying inter-residue contact through interaction energies. *J. Chem. Inf. Model.* **2019,** *59,* 5034–5044.

[101] Cheng, Q.; DeYonker, N. J. A case study of the glycoside hydrolase enzyme mechanism using an automated QM-cluster model building toolkit. *Front. Chem.* **2022,** *10,* 854318:1–14.

[102] Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulation. *Nucl. Acids Res.* **2012,** *40,* 537–541.

[103] Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—an accurate and broadly parameterized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019,** *15,* 1652–1671.

[104] Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and efficient implicit solvation model for fast semiempirical methods. *J. Chem. Theory Comput.* **2021,** *17,* 4250–4261.

[105] Řezáč, J.; Stewart, J. J. P. How well do semiempirical QM methods describe the structure of proteins? *J. Chem. Phys.* **2023,** *158,* 044118:1–8.

[106] Lee, B. I.; Suh, S. W. Crystal structure of the Schiff basse intermediate prior to decarboxylation in the catalytic cycle of aspartate α-decarboxylase. *J. Mol. Biol.* **2004,** *340,* 1–7.

[107] Epifanovsky, E. *et al.* Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021,** *155,* 084801:1–59.

[108] Lange, A. W.; Herbert, J. M. Polarizable continuum reaction-field solvation models affording smooth potential energy surfaces. *J. Phys. Chem. Lett.* **2010,** *1,* 556–561.

[109] Lange, A. W.; Herbert, J. M. A smooth, nonsingular, and

faithful discretization scheme for polarizable continuum models: The switching/Gaussian approach. *J. Chem. Phys.* **2010,** *133,* 244111:1–18.

[110] Lange, A. W.; Herbert, J. M. Symmetric versus asymmetric discretization of the integral equations in polarizable continuum solvation models. *Chem. Phys. Lett.* **2011,** *509,* 77–87.

[111] Herbert, J. M.; Lange, A. W. Polarizable continuum models for (bio)molecular electrostatics: Basic theory and recent developments for macromolecules and simulations. In *Many-Body Effects and Electrostatics in Biomolecules*; Cui, Q.; Ren, P.; Meuwly, M., Eds.; CRC Press: Boca Raton, 2016; Chapter 11, pages 363–416.

[112] Rowland, R. S.; Taylor, R. Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der Waals radii. *J. Phys. Chem.* **1996,** *100,* 7384–7391.

[113] Dasgupta, S.; Herbert, J. M. Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3. *J. Comput. Chem.* **2017,** *38,* 869–882.

[114] Gill, P. M. W.; Johnson, B. G.; Pople, J. A. A standard grid for density-functional calculations. *Chem. Phys. Lett.* **1993,** *209,* 506–512.

[115] "Ohio Supercomputer Center", http://osc.edu/ark:/19495/f5s1ph73.