THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL Letter

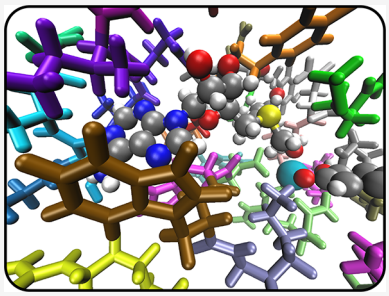
Fragment-Based Calculations of Enzymatic Thermochemistry Require Dielectric Boundary Conditions

Paige E. Bowling, Dustin R. Broderick, and John M. Herbert*

Cite This: J. Phys. Chem. Lett. 2023, 14, 3826−3834 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Electronic structure calculations on enzymes require hundreds of atoms to obtain converged results, but fragment-based approximations offer a cost-effective solution. We present calculations on enzyme models containing 500−600 atoms using the many-body expansion, comparing to benchmarks in which the entire enzyme−substrate complex is described at the same level of density functional theory. When the amino acid fragments contain ionic side chains, the many-body expansion oscillates under vacuum boundary conditions but 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 in fragment-based approximations. A three-body protocol retains sub-kilocalorie per mole fidelity with respect to a supersystem calculation, as does a two-body calculation combined with a full-system correction at a low-cost level of theory. These protocols pave the way for application of high-level quantum chemistry to large systems via rigorous, *ab initio* treatment of many-body polarization.



Fragment-based approximations[1−6] are an attractive way to circumvent nonlinear scaling of computational quantum chemistry (QC), whose floating point cost normally grows like $O(N^p)$ as a function of system size (N), with exponents (p) ranging from 3 for density functional theory (DFT) to ≥7 for levels of theory that can provide thermochemical benchmarks. Fragmentation into N_F separate subsystems (fragments), each of size n, reduces that cost to $N_F \times O(n^p)$ in a manner that is amenable to distributed computing and does not require modification to electronic structure codes. Nonlinear growth in N_F with respect to 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 reaction energies and activation barriers for enzyme-catalyzed chemical reactions. Over the past decade, benchmark calculations have revealed that enzymatic thermochemistry does not converge until hundreds of atoms are included in the QC calculation,[9−17] which is much larger than is typical in contemporary quantum mechanics/molecular mechanics (QM/MM) calculations. Fragmentation may therefore offer an efficient route for obtaining converged thermochemistry (for N > 500 atoms) at benchmark levels of theory, provided that errors associated with fragmentation can be controlled. We demonstrate that these errors can be reduced below the "thermochemical accuracy" threshold of 1 kcal/mol. At the same time, our calculations reveal that straightforward comparison of fragment-based approximations to full-system benchmarks (as a means to assess error) is ill-posed if the calculations are carried out under vacuum boundary conditions.

We will consider models of enzyme−substrate complexes containing approximately 500−600 atoms. Total energies are approximated by means of a many-body expansion (MBE),

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

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

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

for the three-body corrections. When eq 1 is truncated at n-body terms, we 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 to hasten convergence of the MBE.[18−28] We avoid this, however, having found that charge embedding can lead to inconsistent convergence of the n-body expansion.[29−31] Furthermore, the use of self-consistent point

Received: February 25, 2023
Accepted: April 6, 2023

https://doi.org/10.1021/acs.jpclett.3c00533
J. Phys. Chem. Lett. 2023, 14, 3826−3834

charges significantly complicates the formulation of analytic energy gradients.[32−36]

Fragments *I*, *J*, *K*, ..., 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 single-residue fragments. (The treatment of the substrate is discussed below.) Alternatively, overlapping fragments have sometimes been used for polypeptides and proteins.[24−28,38−45] This can be motivated in terms of a generalized (G)MBE,[1,6,46,47] but most overlapping fragment applications to date have used a one-body approach that captures through-bond interactions but not through-space interactions.[6] A two-body GMBE can capture both, but is relatively expensive in terms of the number of subsystems that are generated.[31,39] As such, we stick to the simple MBE(n) approach in this work.

As a first test, we consider S_N2 methyl transfer[48] catalyzed by human catechol *O*-methyltransferase (COMT).[49−51] This particular enzyme has become something of a benchmark,[14,52−55] because it has a well-resolved crystal structure,[51] kinetics data,[49] and numerous known inhibitors.[14−16] A Mg^{2+} ion in the active site is essential to COMT's function[56] but leads to charge-transfer effects in QC calculations that can significantly alter the barrier height, depending on the size of the model system.[14,52,55] Kulik et al.[14] considered a sequence of COMT models with QM regions with 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, an asparagine residue, and a water molecule (58 atoms). Reactant, product, and transition state structures for the transfer of a methyl from *S*-adenosyl-L-methionine (SAM) to catecholate were protonated and relaxed as described in 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 the MBE toward a well-defined 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 nine 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-},[57,58] and delocalization errors in DFT can exacerbate this problem.[58] To avoid artifacts, charged residues are often neutralized in fragment-based calculations on proteins.[59−61] 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 Figure 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 nearest-neighbor fragments are combined to neutralize charge; e.g., a negatively charged

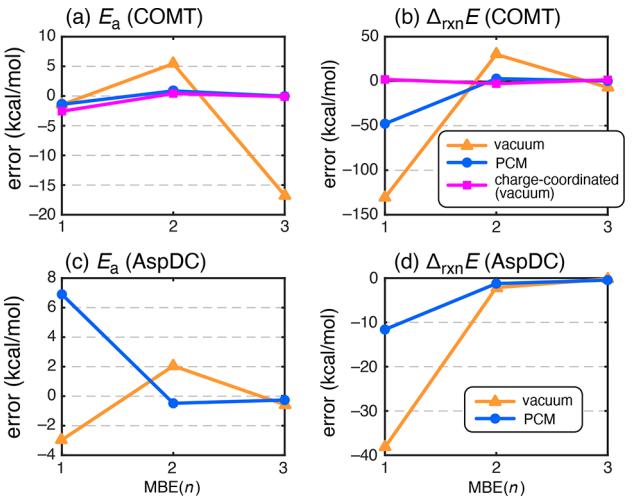


Figure 1. Errors in MBE(n) calculations at the ωB97X-D/def2-SVP level as compared to a supersystem calculation at the same level of theory: (a) E_a for COMT, (b) $\Delta_{rxn}E$ for COMT, (c) E_a for AspDC, and (d) $\Delta_{rxn}E$ for AspDC. For COMT, results are shown for vacuum boundary conditions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

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 alter 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 for both E_a and $\Delta_{rxn}E$ (see Figure 1). 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).[62] For protein electrostatics calculations based on the Poisson−Boltzmann equation, it is common to use a dielectric constant ε in the range of 2−4 to represent the hydrophobic interior of the protein,[63−68] although larger values have occasionally been suggested.[68−74] The precise value of ε may matter for pK_a calculations, but E_a and $\Delta_{rxn}E$ often converge quickly as a function of ε, such that results for ε = 4 are indistinguishable from much larger values.[75,76] That is indeed the case for COMT, as shown in Figures S2 and S3. Gas-phase energetics (ε = 1) are an outlier, but ε = 4 is nearly indistinguishable from ε = 32. Subsequent PCM calculations in this work use ε = 4.

When the subsystem calculations required for MBE(n) are performed using PCM boundary conditions with ε = 4, and results compared to a supramolecular calculation with the same boundary conditions, we observe excellent convergence of MBE(n) even in the presence of single-residue fragments with net charge. The results for several other DFT functionals and basis sets are listed in Tables S1 and S2, and we note that stable results are obtained even when the basis set contains diffuse functions. (Diffuse functions can be problematic for self-consistent charge-embedding schemes.[77−79]) In Figure S4, we extend some of these results to n = 4 in order to check convergence. Using PCM boundaries, the difference between

https://doi.org/10.1021/acs.jpclett.3c00533
J. Phys. Chem. Lett. 2023, 14, 3826−3834

the MBE(3) and MBE(4) results is $\lesssim$1 kcal/mol, while gas-phase calculations sometimes afford errors of >150 kcal/mol at the four-body level! MBE(3) calculations with low-dielectric boundary conditions consistently provide sub-kcal/mol fidelity for various functionals and basis sets, whereas MBE(3) with vacuum boundary conditions affords errors of 10−30 kcal/mol in many cases.

These results suggest that large errors for enzymatic thermochemistry obtained using MBE(n) with vacuum boundary conditions originate not from the fragmentation approximation, nor do they arise from the simple hydrogen atom caps that we use to saturate the severed valencies. (Our approach to capping is less sophisticated than the use of "conjugated caps" that attempt to replicate amino acid moieties,[25−28,80,81] yet our results demonstrate that sub-kcal/mol accuracy can be achieved with hydrogen atom caps.) Instead, large and oscillatory errors in MBE(n) calculations under vacuum boundary conditions arise due to inconsistent charge delocalization in the n-body calculations. 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 of mimicking this polarization. In principle, one might consider the use of heterogeneous dielectric boundaries,[82−85] 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.

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.[86] 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 were placed together in a single fragment to avoid having any charged fragments. Results for E_a and $\Delta_{rxn}E$ (Figure 1) demonstrate that n-body results converge similarly for both vacuum boundary conditions (ε = 1) and PCM boundary conditions with ε = 4, 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 application of MBE(n) with vacuum boundaries to large enzyme−substrate complexes need not converge to the supermolecular result at low orders ($n \leq 4$). Oscillatory behavior results from inconsistent charge delocalization across the subsystems (monomer to dimer to trimer, etc.) when the fragments have net charge. Low-dielectric boundary conditions with $\varepsilon \approx 1.5$ have been shown to reduce the density delocalization error in isolated peptide DFT calculations,[87] and in the present context the use of ε = 4 appears to prevent oscillatory changes to corrections ΔE_{IJ}, ΔE_{IJK}, etc.

Given a two- or three-body approximation for a large enzyme model with charged side chains, one might worry about neglect of long-range interactions. We address this by assessing a multilayer fragmentation scheme 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,[88−90] and is illustrated in Figure 2 by analogy to the "ONIOM" approach to QM/MM calculations.[91] 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 neglect of long-range polarization. Raghavachari and co-workers have made extensive use of this idea for calculations in biological macromolecules,[59−61,92−95] and our two-layer procedure is equivalent to the "MIM2" strategy defined in ref 89.[6]

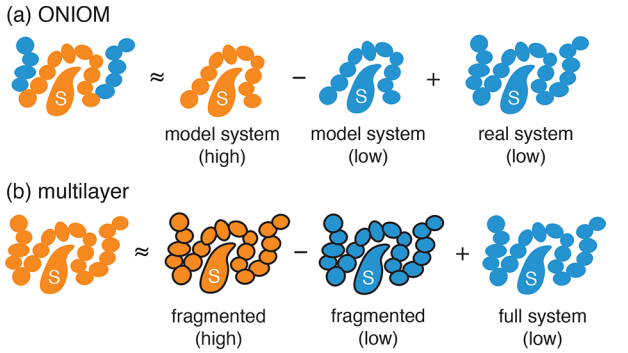


Figure 2. Multilayer techniques applied to the complex of an enzyme (depicted as a chain of amino acids) and a substrate labeled S. Colors encode the level of theory, with the higher-level method in orange and the lower-level method in blue. (a) Conventional ONIOM method, in which high-level calculations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

We tested several low-level supersystem corrections in combination with various target levels of DFT, for the activation energy in COMT. Errors with respect to the target level (applied to the entire enzyme−substrate model) are illustrated in Figure 3, and numerical values for each supersystem correction are listed in Table S3. The low-level methods that we tested include the semiempirical thrice-corrected methods HF-3c[96] and PBEh-3c,[97] which use a minimal and a double-ζ basis set, respectively. We also tested conventional Hartree−Fock (HF) theory and the density functional LRC-ωPBEh,[98,99] each with the 6-31G basis set. (Note that 6-31G is much less expensive than other double-ζ basis sets if the electronic structure software can take advantage of compound shells.[100]) For this particular 632-atom enzyme−substrate complex, all four of these supersystem corrections require a similar computational effort, which constitutes <20% overhead on top of a MBE(2) calculation.

Even without the supersystem correction, results in Figure 3a indicate that a two-body expansion can achieve ∼1 kcal/mol accuracy for E_a using various density functionals. Low-cost supersystem corrections decrease this to ∼0.5 kcal/mol. MBE(3) is an order of magnitude more accurate than MBE(2), achieving ∼0.1 kcal/mol fidelity 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, the HF-3c supersystem correction performs just as well as HF/6-31G, despite using only a minimal basis set ("MINIX").[96] For the 632-atom enzyme−substrate

https://doi.org/10.1021/acs.jpclett.3c00533
J. Phys. Chem. Lett. 2023, 14, 3826−3834

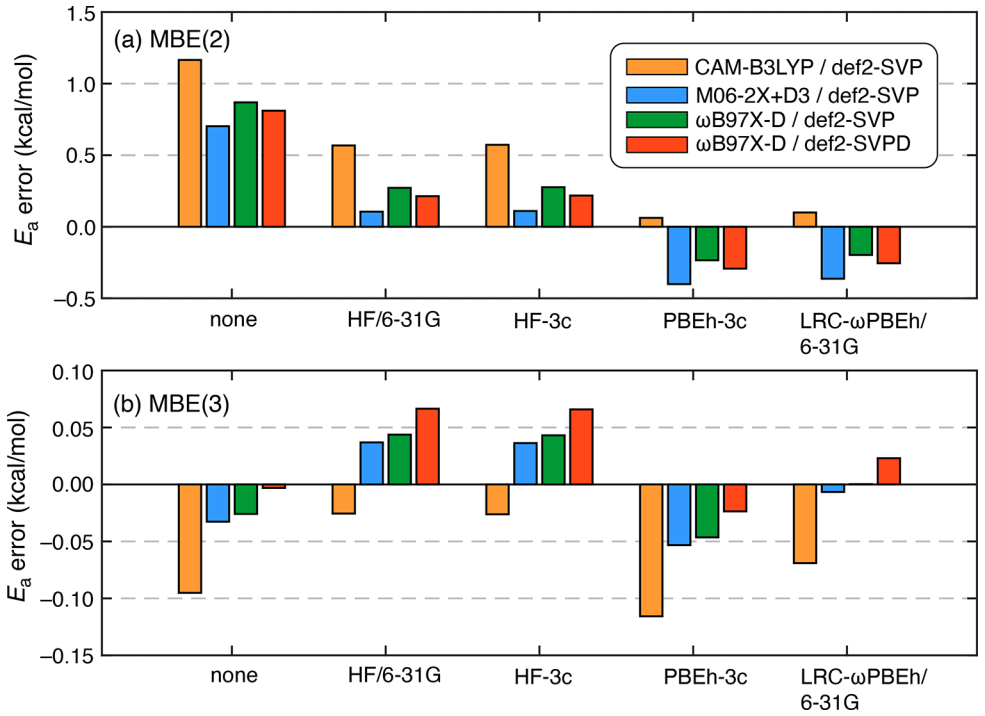


Figure 3. Errors in E_a for COMT, computed using two-layer fragmentation methods at the (a) MBE(2) or (b) MBE(3) level. Target levels of theory (used for the fragments) are indicated by the colored bars, and the error is assessed with respect to a supersystem calculation at that level of theory. Several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

complex used to model COMT, this means 1944 basis functions for HF-3c/MINIX versus 3510 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 as compared to two-electron integrals. The PCM cost could be reduced by using a less dense surface discretization.

Having established that we can 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.[6,31,39,101] Timing data for single-point energy calculations on COMT are provided in Figure 4, with the corresponding numerical data in Table S5. In the absence of any supersystem correction, MBE(2) with PCM boundary conditions is ∼60% of the cost of a supersystem calculation at the same level of theory (ωB97X-D/def2-SVP), whereas MBE(3) is ∼14 times more expensive than the supersystem calculation, although this cost can be distributed across a large number of compute nodes. Despite using larger fragments, the charge-coordinated MBE(3) calculation is actually ∼10% cheaper than MBE(3) with single-residue fragments, because the former calculation decreases the number of unique subsystems from 7175 to 3581. This balance would likely shift in favor of the single-residue calculation if a method more expensive than DFT were used.

The number of subsystems required for MBE(n) increases as N^n for a protein with N residues, which imposes a severe computational bottleneck even for $n = 3$.[31] In what follows, we screen the dimers and trimers on the basis of distance, removing them from the calculation if the minimum interatomic distance between any two fragments exceeds a

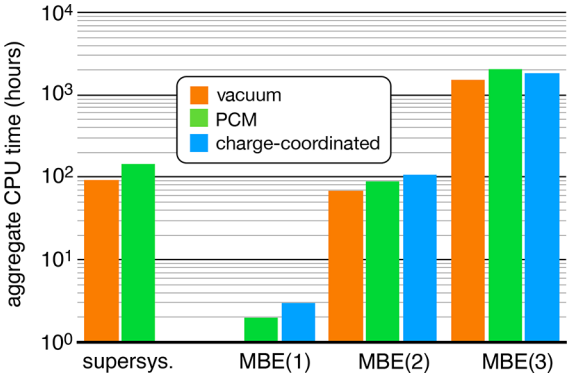


Figure 4. Aggregate computational time (on a logarithmic scale) for a single-point calculation on the 632-atom COMT enzyme−substrate complex, at the ωB97X-D/def2-SVP level of theory. The supersystem calculation contains 6042 basis functions and was performed on a single 28-core node (Dell Intel Xeon E5-2680 v4). Fragment calculations were performed on the same hardware with seven worker processes per node, each using four cores.

specified threshold, R_{cut}. We then recompute E_a and $\Delta_{rxn}E$ 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 (Figure S5) demonstrate that the predicted value of E_a changes by <0.1 kcal/mol as R_{cut} is decreased from 25 to 8 Å. Setting R_{cut} = 8 Å decreases the number of subsystems from 7175 to 1499 (as shown in Figure S6), with a negligible effect on accuracy. Using R_{cut} = 8 Å, the computational effort for ωB97X-D/def2-SVP is reduced from 2025 h (the value shown in Figure 4) to 657 h. The latter

https://doi.org/10.1021/acs.jpclett.3c00533
J. Phys. Chem. Lett. 2023, 14, 3826−3834

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

method	error in E_a (kcal/mol) vs same method[b]	error in E_a (kcal/mol) vs triple ζ[c]	CPU time (h)[d]
MBE(3)	−0.00	0.76	11253
MBE(3) + HF/6-31G[e]	0.04	0.83	11656
MBE(3) + 8 Å[f]	−0.05	0.71	4170
MBE(3) + HF/6-31G[e] + 8 Å[f]	0.04	0.80	4300
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 of theory. [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 four cores. [e]HF/6-31G as a supersystem correction. [f]Screening threshold $R_{cut} = 8$ Å.

figure is still 5 times greater than the cost of the corresponding supersystem calculation, however.

We include diffuse functions in our next set of tests (Table 1), because a method that is intended for general application to enzymatic thermochemistry must be able to accommodate diffuse functions to describe anionic side chains, yet these functions often prove to be problematic for self-consistent charge schemes.[77,78] Even if the electrostatic embedding charges are taken from a force field and held fixed, QM/MM-style, the use of diffuse functions may lead to over-polarization of the QM system by the MM charges.[102]

Errors in E_a for COMT, computed using MBE(2) and MBE(3) at the ωB97X-D/def2-SVPD level, are summarized in Table 1, which includes results using a HF/6-31G supersystem correction and/or distance-based screening using $R_{cut} = 8$ Å. We have also tabulated errors with respect to a ωB97X-D/def2-TZVP supersystem calculation, providing a measure of the basis set incompleteness error when the smaller def2-SVPD basis set is used. The two- and three-body approximations afford sub-kcal/mol errors 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) provides converged results without the need for a supersystem correction, which scarcely alters the results, whereas such a correction affords a small but noticeable improvement to MBE(2).

Consistent, sub-kcal/mol fidelity can be achieved in two ways: MBE(3) alone or MBE(2) with a supersystem correction. Distance cutoffs with $R_{cut} = 8$ Å can safely be applied in either case. This consistency indicates that the supersystem correction primarily accounts for three-body polarization, and that four-body terms make a negligible (sub-kcal/mol) contribution when PCM boundaries are applied (Figure S4). Of these two protocols, MBE(2) with cutoffs and a supersystem correction is more affordable, by a factor of 11 as compared to MBE(3) with cutoffs and no supersystem correction. Although the best measure of real-world cost is the 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. These results are converged to within $\lesssim$1 kcal/mol of a ωB97X-D/def2-TZVP calculation that consists of 11 767 basis functions and requires an aggregate computational time of 17 546 h running on a single 40-core node.

In summary, we find that low-dielectric boundary conditions lead to rapid convergence of the MBE, 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 to describe the subsystems. Because ionic residues must be anticipated in general, 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 calculation 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, using single-residue fragments without the need for electrostatic embedding, conjugated caps, or an ONIOM-style supersystem correction. This relatively simple three-body approach represents a reliable fall-back procedure for systems that are too large for conventional DFT. A practical alternative 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 considerably less expensive than the full-system DFT calculation that it aims to approximate, with a cost that can be readily distributed 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 set to push the accuracy of these calculations beyond the DFT level, by means of a hybrid approach that deploys high-level methods for the two-body interactions combined with three-body DFT to capture polarization by the protein environment. We are also exploring the use of fragment-based vibrational frequency calculations, as pioneered by others,[103−105] 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,[55,106,107] 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.

https://doi.org/10.1021/acs.jpclett.3c00533
J. Phys. Chem. Lett. 2023, 14, 3826−3834

COMPUTATIONAL DETAILS

The crystal structure[51] of COMT [Protein Data Bank (PDB) entry 3BWM] was protonated using the H++ web server[108] at pH 7.0, a salinity of 0.15 M, ε_{in} = 10, and ε_{out} = 80. Ligand atoms were protonated separately using PyMOL and 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 replaced with catecholate ($C_6H_5O_2^-$). Reactant and product structures were relaxed using the GFN2-xTB semiempirical model,[109] with a generalized Born/surface area (GBSA) implicit solvent model for water.[110] Recent benchmarking demonstrates that GFN2-xTB with implicit solvent affords protein structures that compare favorably to experiment.[111] To obtain the transition state, we scanned the length of the bond 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 the active site along with three important residues identified experimentally. This model affords converged energetics with respect to larger models.[14]

For AspDC (PDB entry 1UHE),[112] a single monomer unit can be directly downloaded from the PDB, although the complete structure is an octamer. Starting from the latter, a large radial cutoff of 12 Å was used for structure relaxation with GFN2-xTB in implicit solvent. From that relaxed structure, a smaller 5 Å region was created for a scan along the bond-breaking coordinate. A transition state and a product structure were determined from that scan. 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),[39,59] and instead create fragments by cutting the C_α−$C_{carbonyl}$ bond as indicated in Figure S1. The severed valence is capped with a hydrogen atom that is positioned according to eq S1, as in previous work.[39]

All QM calculations were run using a home-built fragmentation code (PyFragme∩t), interfaced with Q-Chem.[113] For all calculations, the self-consistent field convergence threshold is set to $\tau_{SCF} = 10^{-8}$ Ha and the integral and shell-pair drop tolerances are set to $\tau_{ints} = 10^{-12}$ a.u. We use the conductor-like PCM (C-PCM),[62] implemented with the switching/Gaussian discretization scheme.[114−117] The continuum interface is defined by a van der Waals cavity,[62] constructed using modified Bondi atomic radii[118] 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.[114] A conjugate gradient algorithm was used to solve the C-PCM equations for the full protein model,[117] 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,[119] whereas SG-1[120] is used for other functionals.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.3c00533.

Convergence tests with various functionals and basis sets (PDF)

Coordinates for the enzyme-substrate complexes (ZIP)

Transparent Peer Review report available (PDF)

AUTHOR INFORMATION

Corresponding Author

John M. Herbert − *Biophysics Graduate Program, The Ohio State University, Columbus, Ohio 43210, United States; Department of Chemistry & Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States;* orcid.org/0000-0002-1663-2278; Email: herbert@chemistry.ohio-state.edu

Authors

Paige E. Bowling − *Biophysics Graduate Program, The Ohio State University, Columbus, Ohio 43210, United States; Department of Chemistry & Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States*

Dustin R. Broderick − *Department of Chemistry & Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.3c00533

Notes

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

ACKNOWLEDGMENTS

P.E.B. thanks Atsu Agbaglo at the University of Memphis for assistance in creating the octomer of AspDC from the crystal structure of the monomer and thanks Prof. Nate DeYonker for hosting a visit. This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences, under Grant DE-SC0008550. Calculations were performed at the Ohio Supercomputer Center.[121]

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.

(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. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *7*, No. e1297.

(6) Herbert, J. M. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.* **2019**, *151*, 170901.

(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.

(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: Hoboken, NJ, 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.
(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.
(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.
(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.
(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.
(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.
(37) Wolter, M.; von Looz, M.; Meyerhenke, H.; Jacob, C. R. Systematic partitioning of proteins for quantum-chemical fragmentation methods using graph algorithms. *J. Chem. Theory Comput.* **2021**, *17*, 1355−1367.
(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.
(39) 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.
(40) 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.
(41) Hellmers, J.; König, C. A unified and flexible formulation of molecular fragmentation schemes. *J. Chem. Phys.* **2021**, *155*, 164105.
(42) Hellmers, J.; Hedegård, E. D.; König, C. Fragmentation-based decomposition of a metalloenzyme−substrate interaction: A case study for a lytic polysaccharide monooxygenase. *J. Phys. Chem. B* **2022**, *126*, 5400−5412.
(43) 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.
(44) 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.
(45) Chen, W.-K.; Fang, W.-H.; Cui, G. Extending multi-layer energy-based fragment method for excited-state calculations of large covalently bonded fragment systems. *J. Chem. Phys.* **2023**, *158*, 044110.
(46) 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.
(47) 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.
(48) 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.
(49) 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

https://doi.org/10.1021/acs.jpclett.3c00533
J. Phys. Chem. Lett. 2023, 14, 3826−3834

description of the thermolabile variant of the enzyme. *Biochemistry* **1995**, *34*, 4202−4210.
(50) Bonifácio, M. J.; Archer, M.; Rodrigues, M. L.; Matias, P. M.; Learmonth, 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.
(51) 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.
(52) 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.
(53) 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. U.S.A.* **2015**, *112*, 7954−7959.
(54) 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*, No. e0161868.
(55) Summers, T. J.; Cheng, Q.; Palma, M. A.; Pham, D.-T.; Kelso, D. K., III; Webster, C. E.; DeYonker, N. J. Cheminformatic quantum mechanical enzyme model design: A catechol-O-methyltransferase case study. *Biophys. J.* **2021**, *120*, 3577−3587.
(56) Axelrod, J.; Tomchick, R. Enzymatic O-methylation of epinephrine and other catechols. *J. Biol. Chem.* **1958**, *233*, 702−705.
(57) 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.
(58) 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.
(59) Thapa, B.; Beckett, D.; Jovan Jose, K. V.; Raghavachari, K. Assessment of fragmentation strategies for large proteins using the multilayer molecules-in-molecules approach. *J. Chem. Theory Comput.* **2018**, *14*, 1383−1394.
(60) 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.
(61) 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.
(62) Herbert, J. M. Dielectric continuum methods for quantum chemistry. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, No. e1519.
(63) Gilson, M. K.; Honig, B. H. The dielectric constant of a folded protein. *Biopolymers* **1986**, *25*, 2097−2119.
(64) 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.
(65) Rodgers, K. K.; Sligar, S. G. Surface electrostatics, reduction potentials, and internal dielectric constant of protiens. *J. Am. Chem. Soc.* **1991**, *113*, 9419−9421.
(66) Nakamura, H. Roles of electrostatic interaction in proteins. *Q. Rev. Biophys.* **1996**, *29*, 1−90.
(67) 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.
(68) Alexov, E.; Mehler, E. L.; Baker, N.; M. Baptista, A.; 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.
(69) King, G.; Lee, F. S.; Warshel, A. Microscopic simulations of macroscopic dielectric constants of solvated proteins. *J. Chem. Phys.* **1991**, *95*, 4366−4377.
(70) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* **1994**, *238*, 415−436.
(71) 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.
(72) 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.
(73) 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.
(74) 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.
(75) Sevastik, R.; Himo, F. Quantum chemical modeling on enzymatic reactions: The case of 4-oxalocrotonate tautomerase. *Bioorg. Chem.* **2007**, *35*, 444−457.
(76) 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.
(77) 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.
(78) 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.
(79) 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.
(80) 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.
(81) 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.
(82) 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.
(83) Si, D.; Li, H. Heterogeneous conductorlike solvation model. *J. Chem. Phys.* **2009**, *131*, 044123.
(84) 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.
(85) Stein, C. J.; Herbert, J. M.; Head-Gordon, M. The Poisson−Boltzmann model for implicit solvation of electrolyte solutions: Quantum chemical implementation and assessment via Sechenov coefficients. *J. Chem. Phys.* **2019**, *151*, 224111.
(86) 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.
(87) 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.
(88) 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.
(89) 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.

https://doi.org/10.1021/acs.jpclett.3c00533
J. Phys. Chem. Lett. 2023, 14, 3826−3834

(90) Sahu, N.; Gadre, S. R. Molecular tailoring approach: A route for *ab initio* treatment of large clusters. *Acc. Chem. Res.* **2014**, *47*, 2739−2747.
(91) 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.
(92) 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.
(93) 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.
(94) 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.
(95) Chandy, S. K.; Raghavachari, K. Accurate and cost-effective NMR chemical shift predictions for nucleic acids using a molecules-in-molecules fragmentation-based method. *J. Chem. Theory Comput.* **2023**, *19*, 544−561.
(96) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672−1685.
(97) 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.
(98) 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.
(99) Lange, A. W.; Herbert, J. M. Both intra- and interstrand charge-transfer excited states in B-DNA are present at energies comparable to, or just above, the $^1\pi\pi^*$ excitonic bright states. *J. Am. Chem. Soc.* **2009**, *131*, 3913−3922.
(100) 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.
(101) Gavini, V.; et al. Roadmap on electronic structure codes in the exascale era. *arXiv* **2022**, DOI: 10.48550/arXiv.2209.12747.
(102) 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.
(103) 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.
(104) 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.
(105) 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: Hoboken, NJ, 2017; Chapter 4, pages 141−164.
(106) 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.
(107) 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.
(108) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating p*K* prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulation. *Nucleic Acids Res.* **2012**, *40*, W537−W541.
(109) 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.
(110) 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.
(111) Řezáč, J.; Stewart, J. J. P. How well do semiempirical QM methods describe the structure of proteins? *J. Chem. Phys.* **2023**, *158*, 044118.
(112) 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.
(113) 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.
(114) 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.
(115) 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.
(116) 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.
(117) 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, FL, 2016; Chapter 11, pages 363−416.
(118) 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.
(119) 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.
(120) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. A standard grid for density-functional calculations. *Chem. Phys. Lett.* **1993**, *209*, 506−512.
(121) Ohio Supercomputer Center. http://osc.edu/ark:/19495/f5s1ph73.

https://doi.org/10.1021/acs.jpclett.3c00533
J. Phys. Chem. Lett. 2023, 14, 3826−3834