The Journal of Chemical Physics PERSPECTIVE scitation.org/journal/jcp

Fantasy versus reality in fragment-based quantum chemistry

Cite as: J. Chem. Phys. 151, 170901 (2019); doi: 10.1063/1.5126216
Submitted: 31 August 2019 • Accepted: 6 October 2019 •
Published Online: 4 November 2019

John M. Herbert[a)]

AFFILIATIONS

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA

[a)]herbert@chemistry.ohio-state.edu

ABSTRACT

Since the introduction of the fragment molecular orbital method 20 years ago, fragment-based approaches have occupied a small but growing niche in quantum chemistry. These methods decompose a large molecular system into subsystems small enough to be amenable to electronic structure calculations, following which the subsystem information is reassembled in order to approximate an otherwise intractable supersystem calculation. Fragmentation sidesteps the steep rise (with respect to system size) in the cost of *ab initio* calculations, replacing it with a distributed cost across numerous computer processors. Such methods are attractive, in part, because they are easily parallelizable and therefore readily amenable to exascale computing. As such, there has been hope that distributed computing might offer the proverbial "free lunch" in quantum chemistry, with the entrée being high-level calculations on very large systems. While fragment-based quantum chemistry can count many success stories, there also exists a seedy underbelly of rarely acknowledged problems. As these methods begin to mature, it is time to have a serious conversation about what they can and cannot be expected to accomplish in the near future. Both successes and challenges are highlighted in this Perspective.

I. INTRODUCTION

Quantum chemistry is expensive business. Even for mean-field methods, the cost of electronic structure calculations grows with system size as $\mathcal{O}(N^3)$, and typically as $\mathcal{O}(N^4)$ or worse for wave function-based methods that include electron correlation. For a method such as coupled-cluster theory with single, double, and (perturbative) triple excitations [CCSD(T)] that is capable of achieving ~1 kcal/mol accuracy in thermochemical applications, the cost is $\mathcal{O}(N^7)$. Operationally, this means that a 2× increase in system size equates to a 128× increase in computational cost. Stated differently, and recognizing that $\sqrt[7]{2} \approx 1.10$, a doubling of computer speed equates to a 10% increase in the system size that can be tackled (in a fixed amount of computer time) using CCSD(T), and this 10% increase likely translates into no more than a few additional atoms. Furthermore, even this pessimistic estimate understates the problem somewhat, since the bottleneck in CCSD(T) is often not floating-point operations but rather storage limitations (memory and disk), which grow as $\mathcal{O}(N^4)$.

At least since Pulay first discussed the "localizability of dynamic electron correlation,"[1–3] it has been recognized that the most intractable parts of the electron correlation problem are also the most short-ranged in real space. This idea was later enshrined by Kohn as the "near-sightedness of electronic matter."[4,5] In principle, it ought to be possible to design methods that treat electron correlation effects over very short length scales at a high level theory and then seamlessly transition to lower-level (eventually, classical) approximations at longer length scales. Despite significant effort and some progress,[6–16] and a few single-point CCSD(T) calculations in proteins,[17–19] one may judge by the paucity of hundred-atom CCSD(T) calculations that the problem remains a formidable one, with significant technical challenges.

An alternative (but complementary) approach to large-scale quantum chemistry is parallelization. Technical challenges associated with designing massively parallel implementations of quantum-chemical models are different from those faced in developing local correlation methods, but once again the severity of the problem can be inferred from the lack of widespread applications. To date, one of

the largest examples of a massively parallel, post-Hartree-Fock calculation is a MP2/cc-pVTZ calculation on a $(C_{150}H_{30})_2$ nanoscale graphene dimer, which runs in under an hour on 71 288 processors.[20] CCSD(T)/cc-pVTZ single-point energies have also been reported for a few isomers of $(H_2O)_N$, $N = 16–20$,[21,22] each requiring several hours' wall time on 120 000 processors.[22]

Fragment-based quantum chemistry methods,[23–39] in which a large system is decomposed into small, tractable pieces more amenable to electronic structure calculations, have sometimes been advertised as an end-run around both the nonlinear scaling of traditional quantum chemistry as well as the technical complexities of local correlation and massively parallel algorithm development. By dividing a system of size N into a number of subsystems (N_{sub}), each of size n, fragmentation reduces the formal complexity of the problem according to

$$\mathcal{O}(N^p) \rightarrow N_{sub} \times \mathcal{O}(n^p). \quad (1)$$

The exponent p ranges from $p = 3$ (for density functional theory, DFT) to $p = 7$ [for CCSD(T)], reflecting the inherent complexity of the quantum-chemical model. The fragment size (n) controls the cost of the correlated electronic structure problem. Insofar as n reflects the length scale of electron correlation, it is independent of the system size and therefore linear scaling is achieved, provided that N_{sub} does increase faster than $\mathcal{O}(N)$. (In practice, the latter requirement demands some sort of screening procedure for the subsystems.) To the extent that calculations on individual fragments are independent of one another, fragment-based methods are straightforward to parallelize at the level of a script or a driver program, without modification of the underlying electronic structure program. This makes it relatively easy to jump right into thousand-atom quantum chemistry calculations with very little start-up effort, and calculations on systems larger than 20 000 atoms (equivalent to more than 164 000 basis functions) have been reported,[40] using fragment-based approaches. This ease of entry has occasionally created a tendency to circumvent careful calibration of theoretical models in favor of proceeding directly to large-scale applications with insufficient validation.

As fragmentation methods begin to mature, it becomes time for a reckoning of what problems they can and cannot be expected to solve in the foreseeable future. The very large number of individual calculations required by the fragment-based approach imbues a computational overhead whose magnitude is seldom discussed and may not be widely appreciated, such that fragmentation methods may not always constitute the "free lunch" that they are occasionally portrayed to be. Furthermore, one needs to be clear-eyed about the sorts of chemical applications that are accessible with these methods. For example, what useful questions can be answered based on a few CCSD(T) single-point energy calculations in a large system? To be fair, the same criticism applies to local correlation methods as well, but it is worth asking nonetheless.

With this context in mind, we embark on a survey of fragment-based quantum chemistry methods that is intended to provide a sense of what is feasible and what is not, and to tie together the menagerie of fragmentation approaches by emphasizing similarities among them. This Perspective is not intended as a comprehensive review of these methods, for which the reader is referred to several overarching reviews,[24–26] books,[23,39] and overviews from individual research groups.[27–38] Instead, this article seeks to highlight some success stories in the field but also to draw attention to certain challenges that have not been widely discussed in the literature, and finally to offer opinions as to what directions this field should take.

The remainder of this work is organized as follows. Section II provides an overview of the litany of fragment-based methods in quantum chemistry. These methods are numerous, but it is my contention that they can be understood as relatively minor variations on several key ideas. The present work focuses on similarities. Section III examines the performance of these methods in selected applications to molecular clusters, molecular crystals, and liquids, i.e., to systems where fragmentation does not sever covalent bonds. By eliminating covalent fragmentation as a source of error, these applications serve to highlight other issues intrinsic to the fragmentation approach. The strengths and weaknesses revealed in noncovalent applications are then used to inform a succinct overview of macromolecular applications, in Sec. IV. Finally, Sec. V concludes with some open questions for the field.

II. OVERVIEW OF METHODS

The low entry barrier to doing fragment-based quantum chemistry by script has spawned a plethora of methods, but most of them can be understood as variations on a small number of key ideas. These main ideas are

- the many-body expansion (MBE);
- approximate functional group additivity; and
- multilayer composite approaches, in the spirit of the "ONIOM" scheme.[41–43]

Two parallel lines of argument, based on different physical ideas, have led to the development of two broad categories of fragment-based approaches. The first of these is most readily understood by considering systems in which the fragments are not covalently bonded, such as a cluster or a molecular crystal. This leads naturally to consideration of the MBE, a sequential approach in which calculations are performed on monomers, dimers, trimers, etc. Alternatively, by reflecting upon the approximate additivity of bond enthalpies in thermochemical calculations, one is naturally led to consider methods that decompose a macromolecular system into small functional groups. This leads to an ostensibly distinct category of fragment-based methods. These seemingly disparate ideas can in fact be consolidated, within the unified framework of a generalized (G)MBE.[44–46]

A. Many-body expansion

Consider a molecular liquid or a molecular crystal, in which the fragments (monomers) suggest themselves naturally and where the whole system can be envisaged as a large collection of small monomers, whose interfragment interactions are noncovalent and therefore relatively weak. Other examples might include a ligand nestled in the binding pocket of an enzyme, where each of the nearby amino acids is considered as a separate fragment, as depicted in Fig. 1. Alternatively, one might have in mind any

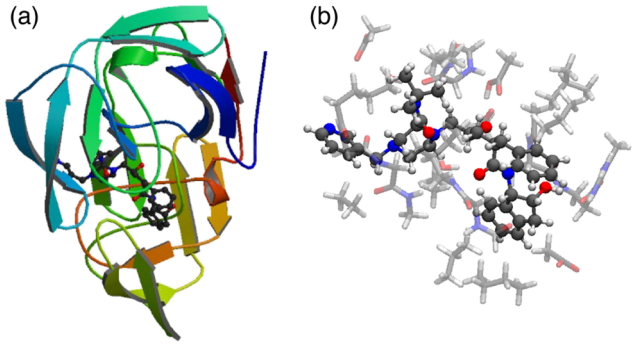

FIG. 1. (a) Illustration of a ligand (the protease inhibitor indinavir) bound to HIV-2 protease. (b) Enlarged view of the binding pocket, consisting of indinavir (ball-and-stick model) along with nearby amino acids.[47] Panel (b) is reproduced with permission from Ucisik *et al.*, J. Chem. Phys. **135**, 085101 (2011). Copyright 2011 AIP Publishing LLC.

calculation in which explicit solvent molecules are added around a given solute.

In these situations, the *many-body expansion* (MBE),

$$E = \sum_{I=1}^{N} E_I + \sum_{I=1}^{N} \sum_{J>I} \Delta E_{IJ} + \sum_{I=1}^{N} \sum_{J>I} \sum_{K>J} \Delta E_{IJK} + \cdots, \tag{2}$$

provides a context in which the description of the ligand or the solute (or simply the monomer units, more generally) can be systematically improved by sequential introduction of additional fragments. The first term in Eq. (2) represents the sum of the energies of N distinct fragments, and subsequent terms are corrections for dimers, trimers, etc. For example,

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \tag{3}$$

is a two-body correction, where E_{IJ} is the energy of dimer IJ, and

$$\Delta E_{IJK} = E_{IJK} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK} - E_I - E_J - E_K \tag{4}$$

is a three-body correction. The expansion in Eq. (2) is trivially valid insofar as each successive term subtracts out the terms that have come before it; however, this approach is only useful if higher-order n-body terms quickly become negligible, beyond $n = 3$ or perhaps $n = 4$. Truncating Eq. (2) at n-body terms forms the basis of the method that I will call MBE(n).

Neglect of terms beyond $n = 2$ affords a pairwise-additive approximation that is known to fail badly for clusters of polar monomers, despite being nearly ubiquitous in force-field development.[48] To put things quantitatively, the three-body polarization effect in isomers of $(H_2O)_6$ ranges from −9 kcal/mol to −13 kcal/mol,[49] and for representative trimers extracted from liquid water, three-body interaction energies range from −6 kcal/mol (attractive) to +1 kcal/mol (repulsive).[48] The total four-body contribution exceeds 1 kcal/mol even for water hexamer,[50,51] and for $(H_2O)_{16}$, the terms with $n \geq 5$ contribute ≈2 kcal/mol to the total interaction energy,[52] although large-basis calculations suggest that these five-body contributions may be artifacts of basis-set superposition error (BSSE).[53] Consistent with that interpretation, MBE(n) converges at $n = 4$ in calculations using a polarizable force field,[54] which contains many-body effects but not BSSE, although five-body terms are necessary to converge the forces.[54]

Fortunately, there is good numerical evidence that electron correlation effects are nearly pairwise-additive in water clusters,[53,55–58] as are counterpoise corrections.[50,59–61] Nonadditivity is dominated by classical polarization,[62] although three-body charge-transfer effects may be important in some ion–water clusters.[63,64] That electron correlation is essentially a pairwise phenomenon makes sense if one considers that the dispersion interaction, which is absent in Hartree-Fock theory and arises solely from electron correlation, also falls off very rapidly with distance, as R^{-6}. Three-body dispersion effects thus contribute $\lesssim$2%–3% of the total interaction energy in clusters of small monomers.[58] (Estimates of the three-body dispersion contribution to the lattice energy of benzene, which might be considered a worst-case scenario due to the importance of dispersion and smallness of the monomers, range from 0.6 to 1.6 kcal/mol,[65–67] representing 5%–13% of the lattice energy.) The molecular electron density decays even more rapidly (exponentially) with distance and therefore so does the exchange interaction (i.e., Pauli repulsion), and also the BSSE.

In contrast, many-body contributions to induction (i.e., polarization) cannot be neglected and become increasingly important in large systems. This is demonstrated in Fig. 2(a), which plots errors in the MBE(n) approximation to the total energy for a sequence of increasingly large water clusters.[68] Two- and three-body approximations quickly diverge from the exact result as cluster size increases.

These many-body polarization effects are present already at the Hartree-Fock level, or even at the classical level if the monomers are polarizable. In an attempt to hasten convergence of the MBE, it is therefore common to perform the subsystem electronic structure calculations embedded in some classical electrostatic representation of the rest of the system. In its simplest version, this electrostatically embedded (EE)-MBE(n) approach[68–72] might simply use atomic point charges to represent the other fragments, in the style of hybrid quantum mechanics/molecular mechanics (QM/MM) calculations. For that reason, EE-MBE(n) has been called a "spatially homogeneous" QM/MM method,[73] since the entire system is treated quantum-mechanically, one (or more) fragment(s) at a time. This is equivalent to the *generalized molecular fractionation* (EE-GMF) method of Liu and He.[74–78]

A version of the MBE that merits special attention is the *fragment molecular orbital* (FMO) method.[23,38,79–83] Historically, FMO was the first fragment-based quantum chemistry method to be identified as such,[79] even if applications of the MBE in quantum chemistry go back even further,[84–88] and FMO is especially prevalent due to its long-standing implementation in the GAMESS program.[38,81] The FMOn method uses an n-body expansion for the energy in conjunction with an electrostatic embedding that consists of full electrostatic potentials (generated from the fragment densities) at short range and point charges at longer range.[82,89] The vast majority of FMO calculations reported in the literature are actually FMO2, which is not particularly accurate for total energies, but FMO3 and FMO4 methods have been formulated and implemented as well.[90,91]

Importantly, in FMOn only the one-body subsystems are iterated to self-consistency, whereas the dimer energies [E_{IJ} in Eq. (3)], trimer energies [E_{IJK} in Eq. (4)], etc., are evaluated using frozen

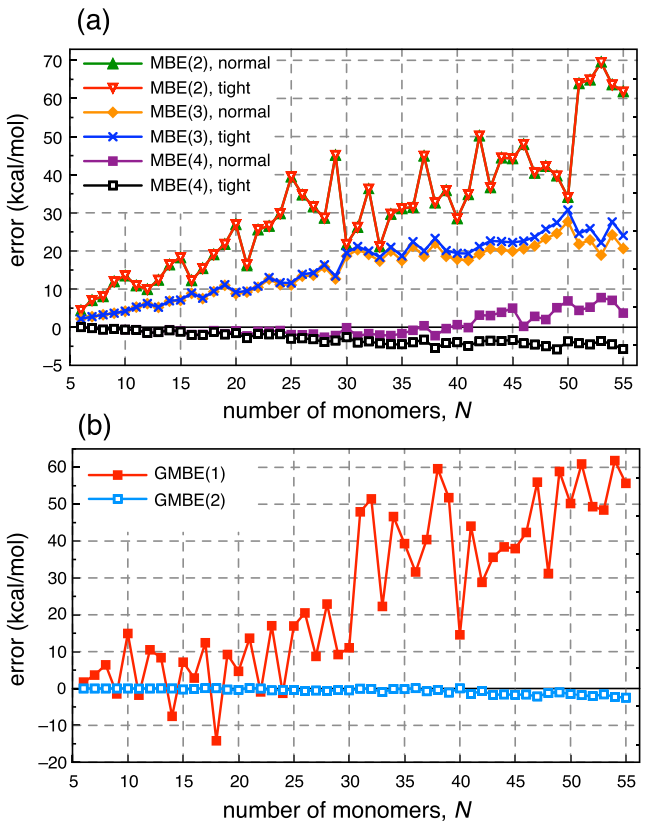

FIG. 2. Errors in MBE(n) and GMBE(n) approximations to the total interaction energy for a sequence of $(H_2O)_N$ clusters, based on calculations performed at the B3LYP/cc-pVDZ level. (a) MBE(n) results with one H_2O per fragment, using normal vs tight convergence and integral-screening thresholds. (b) GMBE(n) results with 3–4 water molecules per fragment. Adapted with permission from Lao *et al.*, J. Chem. Phys. **144**, 164105 (2016). Copyright 2016 AIP Publishing LLC.

fragment densities.[79] In contrast, most other examples of MBE(n) iterate all of the subsystems (dimers, trimers, etc.) to self-consistency. This reduces the cost and enhances the parallel efficiency of FMOn because no communication between n-body subsystems is required beyond $n = 1$, but it also has the undesirable effect of limiting FMOn to small basis sets,[92,93] as discussed in Sec. III B. The self-consistent EE-MBE(n) approach suffers no such limitation.

Although introduced here in the context of noncovalent systems, both FMOn and EE-MBE(n) can be extended to macromolecular systems by introducing "link atoms" to replace covalent bonds that are severed by fragmentation, as in standard QM/MM calculations.[94–98] Alternatively, frozen hybrid orbitals can be used to saturate the severed valencies,[92,99,100] again following QM/MM protocols.

B. Inclusion/exclusion principle

There also exist fragment-based methods whose origins are not rooted in the MBE at all but can be understood (in the context of

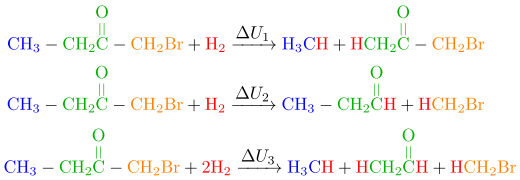

FIG. 3. Three possible H_2 addition reactions for the molecule 1-bromobutan-2-one, which has been divided into three color-coded fragments. Reaction energies ΔU are defined in Eqs. (5)–(7).

macromolecular applications) in terms of the thermochemical concept of approximate additivity of bond or functional-group energies. To illustrate this idea, consider the $CH_3CH_2C(O)CH_2Br$ molecule, divided into functional groups CH_3–, –$CH_2C(O)$–, and –CH_2Br. Figure 3 suggests three possible hydrogenation reactions involving this molecule, the first two of which involve cleavage of a single C–C bond, with the H_2 reagent serving to cap the two carbon valencies that are created when this bond is severed. The corresponding energy changes for these two reactions are

$$\Delta U_1 = E(H_3CH) + E(HCH_2COCH_2Br) - E(CH_3CH_2COCH_2Br) - E(H_2) \quad (5)$$

and

$$\Delta U_2 = E(CH_3CH_2COH) + E(HCH_2Br) - E(CH_3CH_2COCH_2Br) - E(H_2). \quad (6)$$

The third reaction in Fig. 3 cleaves both of the C–C bonds simultaneously, corresponding to a reaction energy

$$\Delta U_3 = E(H_3CH) + E(HCH_2COH) + E(HCH_2Br) - E(CH_3CH_2COCH_2Br) - 2E(H_2). \quad (7)$$

To the extent that bond energies are indeed additive, then $\Delta U_3 \approx \Delta U_1 + \Delta U_2$. Combining this approximation with Eqs. (5)–(7) affords

$$E(CH_3CH_2COCH_2Br) \approx E(CH_3CH_2COH) + E(HCH_2COCH_2Br) - E(HCH_2COH). \quad (8)$$

Notice that the molecules on the right side of Eq. (8) contain only one or two functional groups, whereas the original molecule contains three. Equation (8) therefore constitutes a fragment-based approximation to the total energy of the original molecule.

This procedure can be applied sequentially until the energy of any large molecule has been approximated in terms of the energies of small molecules consisting of a single functional group each (with hydrogen caps), although it stands to reason that the accuracy is likely to be better if the fragments are larger. In the *systematic molecular fragmentation* (SMF) approach pioneered by Collins and co-workers,[31,101–105] three levels of approximation are suggested

based on whether the fragmented bonds are separated by one, two, or three intact bonds. Denoting a single functional group as G_k, with the notation G_kG_{k+1} understood to mean two functional groups connected by a covalent bond, the SMFn schemes (for levels n = 1, 2, or 3) can be written schematically as[31,101,105]

$$G_1G_2G_3G_4G_5$$

$$\xrightarrow[1]{\text{Level}} G_1G_2 + G_2G_3 + G_3G_4 + G_4G_5 - G_2 - G_3 - G_4 \tag{9a}$$

$$\xrightarrow[2]{\text{Level}} G_1G_2G_3 + G_2G_3G_4 + G_3G_4G_5 - G_2G_3 - G_3G_4 \tag{9b}$$

$$\xrightarrow[3]{\text{Level}} G_1G_2G_3G_4 + G_2G_3G_4G_5 - G_2G_3G_4. \tag{9c}$$

The SMF3 scheme in Eq. (9c) corresponds to an energy expression

$$E(G_1G_2G_3G_4G_5) \overset{\text{SMF3}}{\approx} E(G_1G_2G_3G_4) + E(G_2G_3G_4G_5) - E(G_2G_3G_4). \tag{10}$$

Since fragment size increases with n, the SMFn framework provides something of a systematically improvable hierarchy. This is relatively straightforward for unbranched hydrocarbons such as the one considered in Fig. 3, but more complicated molecules require a more sophisticated decision tree. Multiple fragmentation strategies could be envisaged that are probably equally reasonable, at least according to the crude metric of "chemical intuition." Collins[104] has suggested systematic procedures to obtain approximate energy formulas, but these prescriptions are certainly not unique.

An important step toward systematizing the panoply of fragment-based methods was the recognition that the details of how the molecule is partitioned into fragments can be separated from the manner in which the energy formula is obtained.[44] Notice that the SMFn schemes in Eq. (9) possess a certain symmetry wherein the large molecule $G_1G_2G_3G_4G_5$ is divided into overlapping fragments, e.g., $G_1G_2G_3G_4$ and $G_2G_3G_4G_5$, and the terms that are subtracted correspond precisely to the energies of additional fragments formed from intersections. In the SMF3 energy expression, for example, there are two fragments and one intersection, the latter being

$$G_2G_3G_4 = G_1G_2G_3G_4 \cap G_2G_3G_4G_5. \tag{11}$$

This idea was discovered independently by Zhang and co-workers,[30,106–109] in the context of a method that they introduced specifically for proteins called *molecular fractionation with conjugated caps* (MFCC). As illustrated in Fig. 4, the MFCC procedure consists in fragmenting a protein across the peptide bond, introducing amino (–NH_2) and formyl (–CHO) groups to cap the severed valencies. The intersections between overlapping fragments are formamide molecules, NH_2CHO. Denoting this "conjugate" pair of functional groups as a "cap" and "cap*," the approximation that is inherent to MFCC can be expressed succinctly as

$$E(G_kG_{k+1}) \overset{\text{MFCC}}{\approx} E(G_k\text{–cap}_k) + E(\text{cap}^*_{k+1}\text{–}G_{k+1}) - E(\text{cap}_k\text{–cap}^*_{k+1}). \tag{12}$$

These principles can be generalized to encompass several other seemingly disparate methods. The generalization uses the

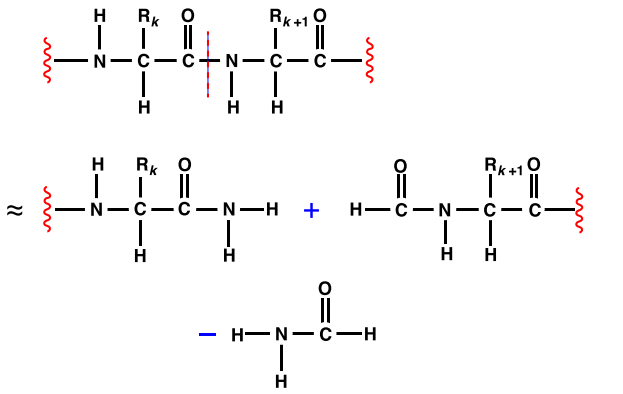

FIG. 4. Illustration of the MFCC fragmentation scheme for proteins, developed by Zhang and co-workers.[30,106–109]

set-theoretical inclusion/exclusion principle, which is a theorem about the cardinality of sets. Let $F = F_1 \cup F_2 \cup \cdots \cup F_N$ be a union of subsets F_i that need not be disjoint. Then the cardinality $|F|$ of the superset can be expressed in terms of the cardinalities $|F_i|$ of the subsets according to

$$|F| = \sum_{i=1}^{N}|F_i| - \sum_{i=1}^{N}\sum_{j>i}|F_i \cap F_j| + \sum_{i=1}^{N}\sum_{j>i}\sum_{k>j}|F_i \cap F_j \cap F_k| - \cdots + (-1)^{N-1}|F_1 \cap F_2 \cap \cdots \cap F_N|. \tag{13}$$

Setting aside (for now) the theoretical justification, let us equate the subsets F_i with fragments of a large molecule, and replace cardinality with energy. An approximation to the total energy is then obtained from Eq. (13), in the form

$$E \approx \sum_{\alpha=1}^{N}\mathcal{E}_\alpha, \tag{14}$$

where

$$\mathcal{E}_\alpha = E_\alpha - \sum_{\beta>\alpha}E_{\alpha\cap\beta} + \sum_{\beta>\alpha}\sum_{\gamma>\beta}E_{\alpha\cap\beta\cap\gamma} + \cdots. \tag{15}$$

The quantity $\mathcal{E}_\alpha$ has been called the "intersection-corrected" energy for fragment α.[44] Terms such as $E_{\alpha\cap\beta}$ represent the energy of a fragment constructed from the intersection $F_\alpha \cap F_\beta$ (with appropriate caps), and Eq. (15) terminates when the intersections vanish, i.e., when $F_\alpha \cap F_\beta \cap F_\gamma \cap \cdots = \varnothing$.

Together, Eqs. (14) and (15) form the basis of several different methods whose formal similarity was noted only after each was introduced independently.[44,45,110] In addition to the aforementioned MFCC scheme,[30,106–109] these methods include the *molecular tailoring approach* (MTA) developed by Gadre and co-workers,[29,111–118] the *generalized energy-based fragmentation* (GEBF) method of Li and co-workers,[27,37,119–121] and the *molecules-in-molecules* (MIM) method developed by Raghavachari and co-workers.[110,122–124] SMFn is very similar but was developed in a more *ad hoc* way and includes only a subset of the terms suggested by the inclusion/exclusion principle.[44]

The inclusion/exclusion principle is invoked as justification in early papers describing molecular tailoring[112] and MIM,[110]

but in other cases the methods mentioned above were introduced by way of a fragmentation strategy, from which an energy formula was subsequently deduced, consisting of fragment energies multiplied by coefficients ±1 whose values were determined by counting arguments.[105,119,120] Thus, the connection between these methods is not immediately obvious from the GEBF and SMF papers, and continues to be downplayed (by omission) in the recent literature. Each of these methods (MFCC, GEBF, MTA, MIM, and SMF*n*) differs in how fragments are selected, a choice that certainly affects the accuracy; nevertheless, it is satisfying that a shared energy formula can be found to unite these ostensibly disparate methodologies.

The leap from the inclusion/exclusion principle in Eq. (13) to the energy formula in Eqs. (14) and (15) can be formally justified by a careful rewriting of the supersystem's Hamiltonian.[44–46] Specifically, Eq. (14) derives from an exact partition

$$\hat{H} = \sum_{i=1}^{N} \hat{H}(F_i) - \sum_{i=1}^{N} \sum_{j>i} \hat{H}(F_i \cap F_j) + \cdots \tag{16}$$

that is analogous to Eq. (13). This is an exact expression for the supersystem's Hamiltonian $\hat{H}$, and its expectation value $\langle\Psi|\hat{H}|\Psi\rangle$ would afford the exact energy if $|\Psi\rangle$ were the wave function for the entire supersystem. To obtain a tractable approximation, one instead evaluates expectation values of individual terms using wave functions computed for localized subsystems, e.g.,[46]

$$\langle\Psi|\hat{H}(F_i \cap F_j)|\Psi\rangle \approx \langle\Psi_{F_i\cap F_j}|\hat{H}(F_i \cap F_j)|\Psi_{F_i\cap F_j}\rangle. \tag{17}$$

The energy expression in Eqs. (14) and (15) follows from this approximation.

Setting aside the thermochemical motivation discussed above, imagine now that the molecule $G_1G_2G_3\cdots$ in Eq. (9) is a protein or other macromolecule that is large enough so that its conformational preferences are interesting. The fragmentation strategy suggested by Eq. (9) is concerned only with primary (sequence) structure, and makes no allowance for the noncovalent interactions that control molecular conformation (secondary structure). Within the context of SMF*n*, Addicoat and Collins[103] suggest incorporating nonbonded interactions by adding interfragment dispersion corrections *a posteriori*, but a first-principles approach is desirable.

The inclusion/exclusion principle that underlies Eq. (16) is agnostic as to how the system is partitioned into fragments, and one can take advantage of this flexibility to introduce a *generalized* (G)MBE, defining GMBE(*n*) to be the method that considers interactions between *n* (possibly overlapping) fragments at a time. The procedure is:[44,46]

1. Tesselate the system into fragments $\{F_i\}$, which may or may not overlap, capping valencies wherever fragmentation severs covalent bonds.
2. Generate a new set of fragments $\{F_\alpha^{(n)}\}$ consisting of the unique *n*-mers formed from the original set of fragments $\{F_i\}$.
3. Apply the energy expression in Eq. (14), where $\mathcal{E}_\alpha$ in Eq. (15) is the intersection-corrected energy for a fragment drawn from the set $\{F_\alpha^{(n)}\}$.

If the fragments $\{F_i\}$ are disjoint (e.g., if each F_i is one H_2O molecule in a water cluster), then the instructions in the second step are to consider all unique dimers (for $n = 2$), trimers (for $n = 3$), etc. In that case, the MBE(*n*) and GMBE(*n*) methods are completely equivalent, so the latter is a true generalization of the former. If the primitive fragments $\{F_i\}$ do overlap, e.g., because we chose fragments G_1G_2, G_2G_3, G_3G_4, . . ., then the GMBE(*n*) approach is useful already starting at $n = 1$. In fact, the method that my group calls GMBE(1) uses precisely the same energy expression as the GEBF, MTA, and "MIM level 1" (MIM1) methods.[44] The SMF*n* approach is similar in spirit although GMBE(1) is formally more complete in the sense of the inclusion/exclusion principle.

There is no unique prescription for choosing fragments in the first step of the procedure, and various methods differ on this point. In the context of the GEBF approach, Li and co-workers[119,125–127] have suggested an automated procedure based on a single distance cutoff parameter, R_{cut}. The primitive fragments $\{F_i\}$ are generated by looping through all of the nonhydrogen atoms, grouping into fragments any atoms that lie within R_{cut} of one another, with hydrogen atoms assigned based on distance or covalency requirements. This procedure generates a relatively large number of fragments but can be used to provide extremely accurate results at the $n = 2$ level, for both noncovalent clusters[46,68] and macromolecules.[128] This is demonstrated in Fig. 2(b) for the same set of water clusters used to evaluate the efficacy of MBE(*n*), only this time with 3–4 H_2O molecules per fragment F_i, corresponding to $R_{\text{cut}} = 3.0$ Å. Absolute errors in the GMBE(2) approximation remain quite small even in clusters as large as $(H_2O)_{55}$, while requiring electronic structure calculations on subsystems $F_i \cup F_j$ that are no larger than $(H_2O)_8$. For applications to macromolecules, the GMBE(2) approach includes both "through-bond" interactions (due to the use of overlapping fragments) and "through-space" interactions (from the dimers of fragments), even if the fragments G_1G_2, G_2G_3, G_3G_4, . . . are selected strictly based on the primary amino acid sequence of a protein. The "through-space" terms are not included at the level of GMBE(1), which is why methods such as SMF*n* must incorporate noncovalent interactions in a more *ad hoc* way.

C. Composite approaches

As discussed in Sec. II A, electrostatic embedding of the subsystem calculations is designed to capture many-body induction without the need for high-order *n*-body electronic structure calculations. Even so, for large water clusters it is found that EE-MBE(4) is necessary to achieve a fidelity of ~1 kcal/mol between the supersystem calculation and its fragment-based approximation.[53,68] At the four-body level, however, the number of distinct subsystems increases in a catastrophic way with respect to system size. A system with $N = 50$ fragments has 230 300 distinct tetramers, for example, as compared to 19 600 trimers. Partly due to concerns over finite precision,[33,72] which are discussed in Sec. III A, the latter number is probably tractable, but the former number may not be.

Higher-order polarization effects are essentially classical, however, so there is no physical reason why three- and four-body terms need to be described at a high level of theory. Composite approaches take advantage of this physics by performing low-order *n*-body

calculations (perhaps only for $n \leq 2$) at the highest feasible level of quantum theory, then using a lower-level method to describe higher-order n-body interactions. In practice, this means that the higher-level method is operational at short length scales and the lower-level method at longer length scales. A straightforward example of this idea is to describe the many-body interactions ($n \geq 3$) using a polarizable force field. This forms the basis of the *hybrid many-body interactions* (HMBI) method developed by Beran and co-workers.[36,66,73,129–132] The force field can be parameterized on-the-fly using atomic polarizabilities derived from fragment electronic structure calculations.[66]

Force-field calculations are essentially zero-cost in this context, hence there is little reason to truncate the n-body expansion if the many-body terms are described classically. Instead, one performs the force-field calculation on the entire (super)system and uses the result to sum the n-body interactions from $n = 3$ to $n = N$ at the classical level of theory. This is the same idea used in the "ONIOM" method,[41–43] which was originally developed to combine different levels of theory, e.g., for QM/MM calculations. In the traditional ONIOM approach, a small model system is carved out of a larger system of interest, the latter of which is known as the "real system" in ONIOM terminology.[43] A low level of theory is selected that is affordable enough to be applied to the real system, while the higher level of theory is applied only to the smaller model system. The energy of the whole system is then approximated as

$$E \overset{\text{ONIOM}}{\approx} E^{\text{high}}(\text{model}) - E^{\text{low}}(\text{model}) + E^{\text{low}}(\text{real}). \quad (18)$$

Figure 5(a) provides a pictorial illustration.

To connect this idea to the HMBI approach, one could imagine the polarizable force field serving as the low level of theory with some QM model as the higher-level theory, as in a traditional QM/MM application of ONIOM. However, the spatially homogeneous nature of the fragmentation approximation allows the QM method to be applied to the entire real system (albeit over short length scales), by means of a low-order GMBE(n) approximation. A lower-level method, applied to the entire supersystem, corrects for errors introduced by fragmentation, including higher-order induction effects missing from the truncated n-body expansion.

The idea of an ONIOM-style composite with MBE(n) used in the high-level component was originally put forward by Tschumper and co-workers,[133–138] who called it the "n-body:many-body" method. This approach is illustrated in Fig. 5(b) and corresponds to the approximation

$$E^{\text{high}}(\text{super}) \overset{\text{MIM2}}{\approx} E^{\text{high}}(\text{GMBE}) - E^{\text{low}}(\text{GMBE}) + E^{\text{low}}(\text{super}). \quad (19)$$

Here, E(super) indicates a supersystem calculation, whereas E(GMBE) denotes a GMBE(n) approximation of some sort. As with the original ONIOM approach, this could be systematically improved by inserting intermediate levels of theory applied to larger fragments,[110] as in the three-layer approach that is illustrated in Fig. 5(c). A version based on overlapping fragments has also been suggested.[139–141]

As indicated by the notation in Eq. (19), the n-body:many-body approach is equivalent to a two-layer molecules-in-molecules

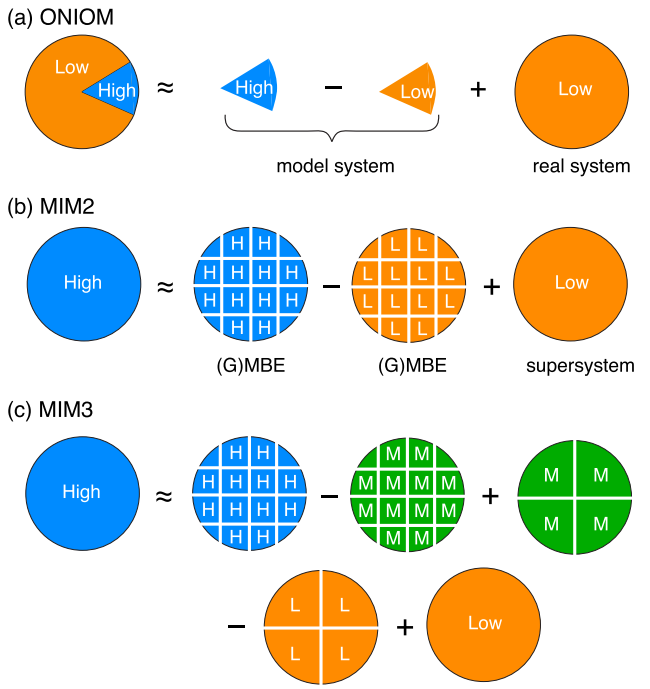

FIG. 5. Schematic illustrations of (a) the ONIOM composite approach;[41–43] (b) its fragment-based analog, the two-layer MIM2 method;[110] and (c) the three-layer MIM3 extension.[110] Orange, green, and blue represent low (L), medium (M), and high (H) levels of theory, respectively. In traditional ONIOM, the high level of theory is confined to a small "model" system, whereas the low level of theory is applied to the entire ("real") system. The multilayer MIM schemes employ one or more fragmentation methods to apply higher levels of theory to the entire system in a homogeneous way, in conjunction with a low-level calculation applied to the entire supersystem. This supersystem calculation captures many-body induction effects and corrects for other errors introduced by fragmentation.

or MIM2 method.[110,122] The MIM1 approximation is essentially GMBE(1) but has seldom been used on its own. Instead, it constitutes the fragment-based component of a multilayer MIM2 [Fig. 5(b)] or MIM3 [Fig. 5(c)] approach. The basic strategy can be understood as a sort of telescoping sum, in which each lower-level layer attempts to correct errors introduced by the fragmentation approximation at the level above, and which ultimately terminates with a calculation performed on the entire supersystem. In order to target large systems, Raghavachari and co-workers tend to perform the supersystem calculation using Hartree-Fock theory or DFT with a small basis set (e.g., M06-2X/6-31+G**),[122,124,142] or else a semiempirical method such as PM6-D3.[143,144] In contrast, Tschumper and co-workers have focused on achieving high accuracy in clusters,[133–138] and their calculations typically consist of MBE(2) at the CCSD(T) level combined with a supersystem calculation at the MP2 level.

A multilayer FMO method has also been proposed wherein two- and three-body terms are computed at different levels of theory.[145] Despite similar semantics, however, this is actually quite different from the multilayer MIMn strategy that relies on a supersystem calculation and therefore abrogates any chance at $\mathcal{O}(N)$ scaling.

Nevertheless, methods such as PM6 or even Hartree-Fock/6-31G* can nowadays be applied quite readily to systems with hundreds of atoms. If the goal is to use a very accurate (and therefore expensive) method for the fragmentation part of the calculation, then a low-level supersystem calculation may be worth the effort, to clean up residual errors introduced by fragmentation.

One variant of MIM2 is a method in which the high- and low-level methods differ only in the use of a large or a small basis set, respectively. This strategy was developed by Gadre and co-workers[29,117,118,146–150] (who call it "grafting") as a correction for their own GMBE(1)-type method. The procedure, which can be considered a special case of Eq. (19), is captured by the formula[118]

$$E^{\text{large}}(\text{super}) \approx E^{\text{large}}(\text{GMBE}) + \delta^{\text{frag}}, \tag{20}$$

where

$$\delta^{\text{frag}} = E^{\text{small}}(\text{super}) - E^{\text{small}}(\text{GMBE}) \tag{21}$$

is a correction for errors introduced by fragmentation, evaluated using a small-basis calculation on the entire supersystem, $E^{\text{small}}(\text{super})$. This strategy has most often been deployed for MP2-level geometry optimizations and vibrational frequency calculations,[29,147–149] using a combination of large-basis Hartree-Fock and small-basis MP2. In that context, the relevant approximation is

$$E^{\text{large}}(\text{super}) \approx E^{\text{large}}_{\text{HF}}(\text{super}) + E^{\text{large}}_{\text{corr}}(\text{GMBE}) + \delta^{\text{frag}}_{\text{corr}}, \tag{22}$$

where $E_{\text{corr}} = E_{\text{MP2}} - E_{\text{HF}}$ is the correlation energy and the correction term is

$$\delta^{\text{frag}}_{\text{corr}} = E^{\text{small}}_{\text{corr}}(\text{super}) - E^{\text{small}}_{\text{corr}}(\text{GMBE}). \tag{23}$$

The grafting idea brings to mind the notion of a MBE(n) approximation for the correlation energy only, applied on top of a Hartree-Fock calculation for the entire system.[151] This is somewhat similar to Stoll's *method of increments*.[86–88,152,153] The latter is an "incremental" expansion for the correlation energy consisting, for example, of two-body corrections

$$\Delta\varepsilon_{ij} = \varepsilon_{ij} - \varepsilon_i - \varepsilon_j, \tag{24}$$

where i and j are molecular orbital indices. This idea is an old one,[84] but it has been revived in recent years as a means to compute near-exact energies for small molecules,[154–164] approaching full configuration interaction. It also forms the basis of a more general "cluster-in-molecule" approach to local correlation.[12,165–169] This is conceptually apart, however, from the manner in which the MBE is used in the fragmentation methods that are the focus of this Perspective. The incremental correlation approach is not discussed any further here.

III. LARGE COLLECTIONS OF SMALL MOLECULES

Molecular crystals and liquids such as water practically cry out for a fragment-based treatment, as does any large noncovalent cluster composed of relatively small monomers. For this reason, the survey of applications that is presented herein focuses first and foremost on noncovalent systems. While highlighting some successful applications, especially in the context of molecular crystals, we will also use noncovalent systems as a platform to expose certain underlying issues that limit how broadly applicable fragmentation methods can be made, and at what cost. These lessons will then be carried into the more complex realm of macromolecular fragmentation, applications of which are surveyed in Sec. IV.

This work focuses exclusively on ground-state energies and properties. For excited states, there have been only preliminary applications of MBE(n),[78,170] although FMOn has been formulated for excited states[171–174] and an overlapping-fragment approach has been described as well.[175,176] Clusters-in-molecules methods have also been applied to excited states,[177–180] as have exciton models that couple together excited-state calculations on more than one fragment, using these as a basis for describing collective (multifragment) excitations.[178–183] Each of these methods is far less mature as compared to ground-state fragmentation, and the ground state suffices to illustrate the primary themes of this Perspective.

A. Accuracy and precision issues

Clusters of water molecules are the quintessential test systems for fragmentation methods, due to the general importance of water, the amenability of clusters to a fragmentation approach, and the fact that many-body induction is significant in water clusters.[48–54,62] Examining the errors engendered by the MBE(n) approximation, i.e., comparing the supersystem energy to the n-body result computed at the same level of theory, one finds that the errors are size-extensive.[33,68] This is evident from the data in Fig. 2(a). Even at the four-body level, the total error is several kcal/mol for $(H_2O)_N$ clusters of size $N \gtrsim 45$. Part of this error is likely attributable to BSSE,[53,59,60,68,184] and indeed, the many-body convergence errors are smaller in calculations using polarizable force fields, e.g., an MBE(4) error of only 1.2 kcal/mol for $(H_2O)_{216}$.[54]

Despite these inherent errors, Paesani and co-workers[48,185,186] have parameterized a very accurate force field for water based on dimer and trimer calculations performed at the CCSD(T) level in the complete basis set (CBS) limit. The resolution of this apparent paradox comes in noting that this "MB-pol" potential[186] is polarizable, therefore many-body induction beyond the three-body level is treated classically but is not truncated, in what is essentially an example of the multilayer approach described in Sec. II C. (The low-level supersystem method is a classical, polarizable force field.) For ion–water clusters such as $H_3O^+(H_2O)_5$, four- and five-body interactions remain on the order of ≈0.5 kcal/mol.[187]

More generally, embedding charges have been suggested as a means to reduce the importance of higher-order n-body terms. Figure 6 examines the same sequence of $(H_2O)_N$ clusters as in Fig. 2, this time using EE-MBE(3) and EE-MBE(4) approximations with various flavors of point charges derived from the one-body wave functions. At the four-body level, these charges do systematically reduce the error (as compared to a supersystem calculation at the same level of theory), but the same is *not* true at the three-body level. For EE-MBE(3), differences between various charge schemes are insignificant in comparison with the overall error in any one of these approximations, which increases rapidly as a function of cluster size. These large, size-extensive errors are sometimes masked by reporting errors in per-monomer terms.

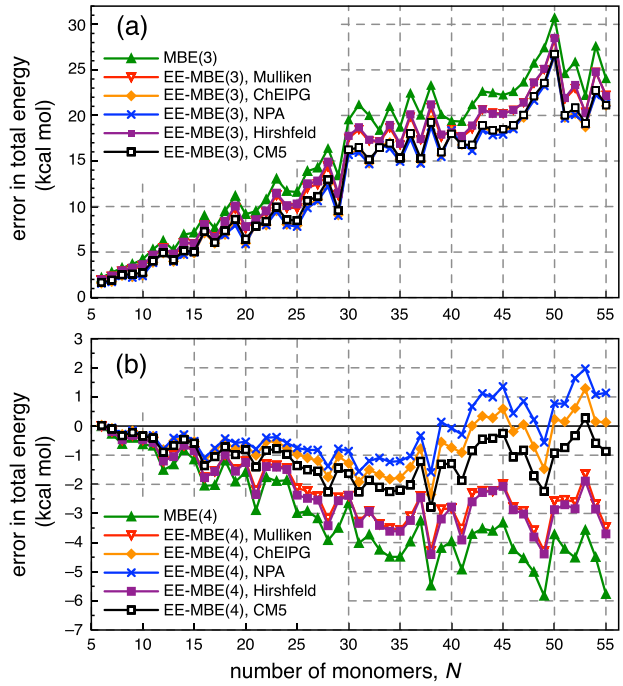

FIG. 6. Errors in (a) EE-MBE(3) and (b) EE-MBE(4) approximations for a sequence of $(H_2O)_N$ clusters, using various types of atomic partial charges. All calculations were performed at the B3LYP/cc-pVDZ level, and the error is measured relative to a supersystem calculation at the same level of theory. Adapted with permission from Lao *et al.*, J. Chem. Phys. **144**, 164105 (2016). Copyright 2016 AIP Publishing LLC.

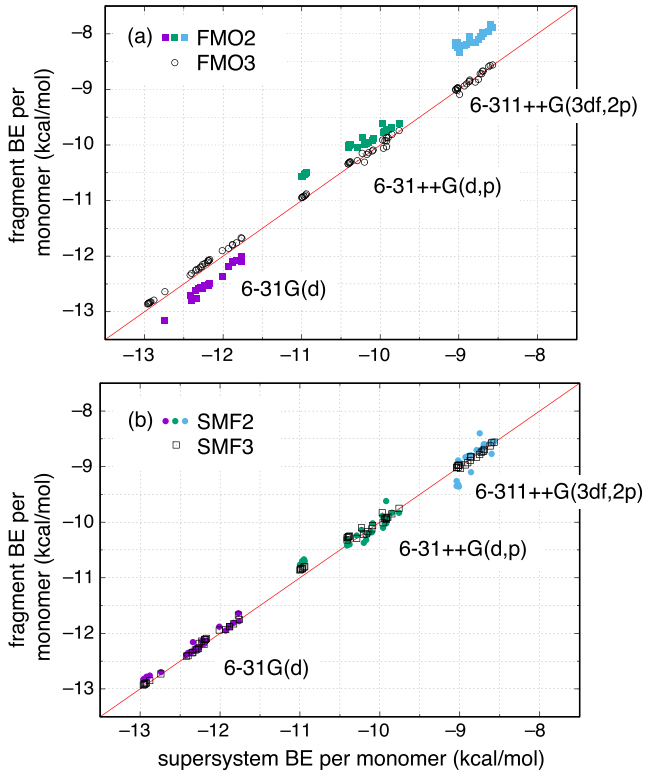

FIG. 7. Binding energies (BEs) per monomer for isomers of $(H_2O)_N$ (N = 16, 20, 32, and 64), computed at the MP2 level in three different basis sets. The horizontal axis provides the supersystem MP2 result, and the vertical axis affords the result obtained from either (a) FMO*n* calculations or (b) SMF*n* calculations. The FMO*n* calculations use one H_2O monomer per fragment, whereas the SMF2 calculations use fragments that range in size from 2 to 6 monomers (average size ≈3.2) and the SMF3 calculations use fragments ranging from 4 to 11 monomers (average size ≈6.4). Data are from Ref. 188.

As compared to EE-MBE(n), the FMOn approach uses a more sophisticated, hierarchical embedding that includes density-based electrostatic embedding at short range.[82,89] The accuracy of FMO2 and FMO3 approximations is considered in Fig. 7(a) for a different set of water clusters.[188] These data are plotted in units of binding energy per water monomer, but the FMO3 results are accurate to within ≈3 kcal/mol (on an absolute scale) for N = 64. The FMO2 data, on the other hand, are not accurate enough to be useful, with absolute errors of ≈28 kcal/mol for N = 64.

Figure 7(b) plots binding energies for the same data set obtained using SMF2 and SMF3,[103,104] both of which are tantamount to GMBE(1) but augmented with *ad hoc* intermolecular polarization and dispersion potentials obtained from calculations performed on the one-body wave functions.[103,188] Fragments are selected based on a distance criterion and include an average of ≈6 monomers per fragment in the case of the more accurate SMF3 approach.[188] This means that a limited set of six-body interactions are included in these calculations. Li and co-workers[119,125–127] have reported accurate energies for $(H_2O)_N$ clusters with only point-charge embedding, using the EE-GEBF approach [equivalent to EE-GMBE(1)] in conjunction with fragments of similar size.

Prediction of absolute binding energies is not the same as accurate prediction of relative energies for different cluster isomers, however, and it is unclear from Fig. 7 whether relative energies are preserved with good fidelity by fragmentation. In fact, the relative energy problem proves to be a very challenging one for fragment-based methods. It can be surmounted using EE-MBE(4) or GMBE(2),[68,184] or GMBE(1) if the fragments are large enough. GMBE(1) with 3–4 water molecules per fragment is clearly *not* sufficient,[68] as demonstrated by the large errors in Fig. 2(b), although the accuracy of the GMBE(2) results in the same figure suggests that 6–8 water molecules per fragment is likely sufficient. Indeed, using EE-GMBE(1) with fragments no larger than $(H_2O)_7$, Li and co-workers[127] demonstrated errors that were consistently <1 kcal/mol for isomers of $(H_2O)_{32}$ at the MP2/cc-pVTZ level, although sub-kcal/mol accuracy for $(H_2O)_{64}$ required fragments as large as $(H_2O)_{10}$. The same fragmentation strategy was then used to compute otherwise unobtainable CCSD(T)/CBS benchmarks for $(H_2O)_{64}$, and these benchmarks were used to evaluate the accuracy of various density functionals.[127]

Methods based on smaller fragments have difficulty with relative energies, however. An example for isomers of $(H_2O)_{16}$ is shown in Fig. 8, where FMO2 and FMO3 are applied at the MP2 level.[189] [Results are also shown using the *effective fragment molecular orbital* (EFMO) method, which will be discussed in Sec. III G.] Oddly, the

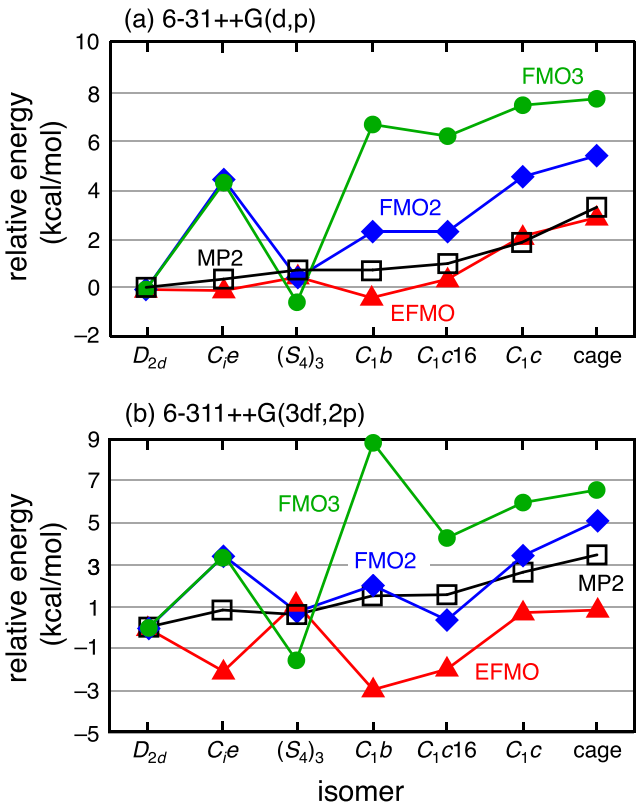

FIG. 8. Relative energies for isomers of $(H_2O)_{16}$ computed with FMO2 and FMO3 at the MP2 level in (a) the 6-31++G(d, p) basis set and (b) the 6-311++G(3df, 2p) basis set. Also shown are results from the EFMO approximation to FMO2. The data labeled "MP2" (in black) are the supersystem results. Data are from Ref. 189.

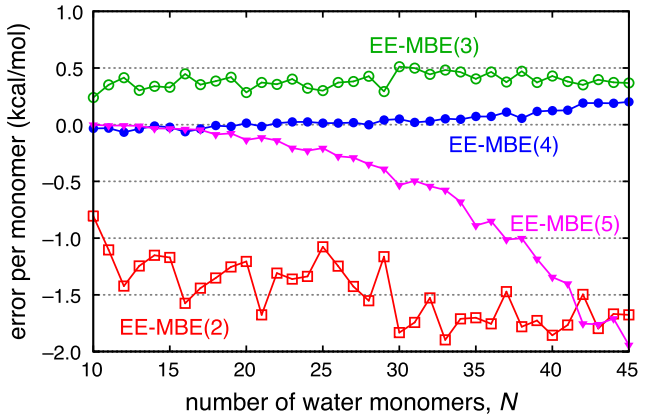

FIG. 9. Error per monomer in various EE-MBE(n) approximations to the total interaction energy, for a sequence of water clusters computed at the B3LYP/cc-pVDZ level using TIP3P embedding charges. These calculations were performed with the same thresholds as the "normal" threshold results in Fig. 2(a). Adapted with permission from R. M. Richard, K. U. Lao, and J. M. Herbert, Acc. Chem. Res. **47**, 2828–2836 (2014). Copyright 2014 AIP Publishing LLC.

FMO2 results are actually more faithful to the supersystem MP2 calculation than are the FMO3 results, the latter of which are quite erratic. In contrast, MP2 itself predicts that all isomers lie within ≈3 kcal/mol of one another. Both FMO2 and FMO3 fail to predict the correct energetic ordering of the isomers.

Figure 9 recasts the size-dependent EE-MBE(n) errors for water clusters from Fig. 6 in per-monomer terms. This demonstrates convincingly that the errors are size-extensive,[33] or in other words that a given n-body approximation can reduce the error only to a roughly constant value per fragment. One might expect this intrinsic error to be reduced as n increases, but the data in Fig. 9 suggest that this is true only up to a point. Five-body calculations have seldom been reported for systems with this many fragments, because nearly 1.4×10^6 separate subsystem calculations are required for $(H_2O)_{45}$, but EE-MBE(5) results up to $N = 45$ are reported in Fig. 9. At this level of approximation, the errors increase rapidly and dramatically with system size, in contrast to results obtained using lower-order n-body expansions.

The origin of these divergent errors can be traced to finite-precision problems leading to error accumulation as the number of subsystem calculations grows.[33,68,72] One consequence is that MBE(n) calculations are far more sensitive to the values of various thresholds than is a supersystem electronic structure calculation for the same system.[68,72,190] This is especially true for the self-consistent field (SCF) convergence threshold (τ_{SCF}) and the integral screening threshold (τ_{ints}), as seen by comparing MBE(n) results using "normal" vs "tight" thresholds [Fig. 2(a)]. The impact of τ_{SCF} and τ_{ints} is detectable but small for two- and three-body approximations, but for MBE(4) the tight-threshold results diverge from the normal-threshold ones as the number of fragments increases.

Finite-precision problems are even more pronounced when embedding charges are employed. In the case that these charges are computed on-the-fly from the fragment wave functions, EE-MBE(n) results may be unreliable unless the charges are obtained from the electronic structure program in full double precision,[72] whereas the text-based output of most quantum chemistry programs rounds the atomic charges to a few digits of decimal precision. To guard against catastrophic loss-of-precision error, the software that manages the subsystem calculations should be tightly integrated with the electronic structure program itself, so that data can be passed between the two in full double precision.[72]

Avoidance of this problem is a key advantage of GMBE(1) over EE-MBE(n). In the former approach, the fragments are generally larger but also overlapping, meaning that far fewer fragments are needed in order to obtain reasonable accuracy. For example, the SMF3 calculations on $(H_2O)_{64}$ that are shown in Fig. 7(b) used only 626 fragments in total,[188] a number that is small enough to avoid serious problems with roundoff error. In contrast, $N = 64$ corresponds to 635 376 unique tetramers required for MBE(4), assuming one H_2O monomer per fragment.

B. Basis sets and basis-set superposition error

One important use for fragment-based approaches is to provide high-accuracy benchmark data in large systems, which can

then be used to evaluate the accuracy of more affordable methods. For benchmark-quality results with correlated wave functions, large basis sets are required. A minimalist procedure to obtain CCSD(T)/CBS benchmarks would be to perform MP2 calculations in triple- and quadruple-ζ basis sets, from which the MP2/CBS result could be extrapolated and then combined with a correction

$$\delta_{\mathrm{CCSD(T)}} = E_{\mathrm{CCSD(T)}} - E_{\mathrm{MP2}}, \tag{25}$$

evaluated using a smaller basis set. EE-(G)MBE(n) approaches have been used to approximate the energies in these calculations and thereby to obtain CCSD(T)/CBS-caliber benchmark interaction energies.[127,150,184,191]

FMO calculations, in contrast, have proven problematic in large basis sets, especially those containing diffuse functions.[92,93] Errors are considerably larger as compared to small-basis results, and in some cases the embedded-fragment SCF calculation simply fails to converge in large basis sets.[92] These difficulties have been traced to the use of electrostatic embedding based on the fragment densities themselves,[92] and is likely exacerbated by the fact that only the one-body calculations are iterated to self-consistency in the FMO approach. Because the molecular orbitals on different fragments are not orthogonal to one another, there is no interfragment Pauli repulsion to confine the fragment wave functions, and given a sufficiently flexible basis set these wave functions will artificially delocalize. One suggestion to avoid this problem is to use separate basis sets for the FMOn energy calculations as compared to what is used in the embedded-fragment SCF calculation, e.g., 6-31++G** for the former and 6-31G* for the latter,[93] although neither of these is a benchmark-quality basis set. Lack of stability in large basis sets renders FMOn effectively unusable for high-accuracy, correlated wave function calculations. It may offer a decent approximation to something such as the MP2/6-31G* energy of a large system, and this might have some spectroscopic applications, but such a calculation is of questionable utility for absolute or relative energies.

One might wonder whether similar problems can afflict the use of embedding charges that are derived from the fragment wave functions, and indeed, convergence problems have sometimes been noted for Mulliken or Löwdin charges.[192,193] Problems are avoided if the embedding charges are defined in a way that connects more directly to the electron density. Examples include the "ChElPG" charges that are derived from the molecular electrostatic potential[192–195] and also "natural" charges obtained from natural bond orbital population analysis.[196]

Another issue that bears on basis-set selection is that of BSSE, which can lead to serious overestimation of interaction energies not only for noncovalent clusters but also in the context of conformational preferences of flexible molecules,[197–200] where compact structures are artificially stabilized by BSSE relative to more extended structures. This leads to an interesting conundrum because dispersion interactions, which can be challenging to describe in electronic structure calculations,[201–204] also stabilize more compact structures. This means that BSSE may masquerade as stabilization by dispersion.

Bettens and co-workers[205] have noted the sometimes oscillatory nature of the convergence of MBE(n) as $n \to N$, a limit in which the supersystem energy ought to be recovered exactly. In calculations on $(H_2O)_{16}$, they identified cases where 10- and 11-body terms were as large as ~1 mE_h in small basis sets, with the problem becoming less severe in larger basis sets and disappearing entirely when the full cluster basis was used for all of the subsystem calculations.[205] These high-order n-body terms turn out to be artifacts of BSSE.[53,59] "Errors" in the n-body expansion that persist to high orders n originate in an imbalance between how BSSE manifests in supersystem vs subsystem calculations, which makes the definition of "error" somewhat questionable when evaluated order-by-order with MBE(n).[60,184] A more robust definition would first remove the BSSE from both calculations, but this requires a BSSE correction that is compatible with MBE(n).

Several generalizations of the Boys-Bernardi "function counterpoise" (CP) correction,[206,207] which are compatible order-by-order with MBE(n), have been suggested.[59,61,191,208,209] To understand these generalizations, note first that the Boys-Bernardi correction for dimers can be generalized to a cluster of monomers by defining a CP-corrected interaction energy,[202]

$$E_{\mathrm{int}}^{\mathrm{SSFC}} = E_{IJK\cdots} - \sum_{i=1}^{N} E_i^{IJK\cdots}. \tag{26}$$

Here, $E_i^{IJK\cdots}$ indicates the energy of monomer i computed in the full cluster ($IJK\cdots$) basis set. The correction in Eq. (26) is sometimes called the "site-site function counterpoise" (SSFC) correction,[210–213] although it is a natural generalization of the original Boys-Bernardi procedure wherein both monomer energies are computed using the dimer basis,

$$E_{\mathrm{int}}^{\mathrm{CP}} = E_{IJ} - E_I^{IJ} - E_J^{IJ}. \tag{27}$$

A many-body counterpoise (MBCP) procedure is obtained by applying the MBE(n) approximation to each individual energy in Eq. (26).[191] This procedure is called MBCP(n),[191] and a version for use with the GMBE has also been reported, called GMBCP(n).[68] Other alternatives have been suggested,[59,61,208,209] but these tend to be equivalent or nearly equivalent in leading order, which is the only significant contribution.[60,61,212–214] Similar relative energies are obtained from these various CP corrections, even if absolute energies differ.[61]

The Boys-Bernardi procedure has sometimes been criticized as an "overcorrection" (see Ref. 207 for a discussion and a refutation) and also for failing to offer a systematic improvement to the uncorrected results.[215] Much of this supposed conventional wisdom, however, seems to trace back to older literature using basis sets that (by modern standards) do not seem adequate for benchmark purposes. Results are more consistent when aug-cc-pVXZ basis sets are employed, as shown in Fig. 10 for MP2 calculations on $(H_2O)_6$.[191] Both the CP-corrected and uncorrected MP2 energies converge smoothly to the CBS limit, albeit from different directions. Moreover, EE-MBCP(3) provides an approximation to the CP-corrected MP2 result that is consistently accurate across this sequence of basis sets, just as EE-MBE(3) accurately approximates the uncorrected MP2 energy. The fact that both energies can be approximated in a consistent way is especially important for larger systems where extrapolations are unreliable due to basis-set limitations. As illustrated by the aug-cc-pVDZ results in Fig. 10, CP correction in a small basis set may indeed overcorrect the result, but

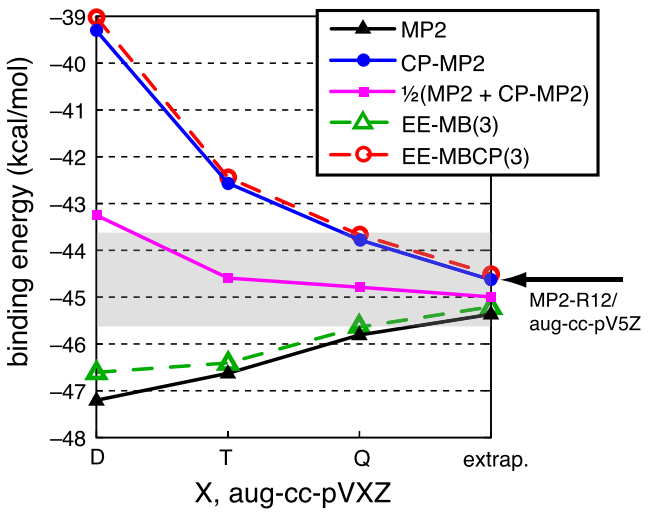

FIG. 10. Convergence of MP2/aug-cc-pVXZ energies for an isomer of $(H_2O)_6$, using EE-MBE(3) to approximate the MP2 total energy and EE-MBCP(3) to approximate the CP-corrected energy. The shaded region delineates ±1 kcal/mol of the MP2/CBS result from Ref. 216. Adapted with permission from R. M. Richard, K. U. Lao, and J. M. Herbert, J. Phys. Chem. Lett. **4**, 2674–2680 (2013). Copyright 2013 American Chemical Society.

averaging the corrected and uncorrected results achieves much better balance. For aug-cc-pVTZ, this average is quite close to the CBS limit.

This may prove useful in applications to macromolecules, in order to address cases where certain conformations might be artificially stabilized by BSSE.[197–200] As an alternative, it has been suggested in the context of FMO2 that one could simply apply a standard Boys-Bernardi correction [Eq. (27)] to each ΔE_{IJ} in Eq. (3).[217] This approach has not yet been widely tested.

C. Cluster spectroscopy

The discussion up to now has focused on total energies, which are important as benchmarks but often not the most chemically or physically relevant quantities. (An exception is that accurate prediction and interpretation of intermolecular interaction energies occupies an important niche that is discussed in Sec. III E.) Prediction of spectroscopic observables, on the other hand, affords a direct point of contact with experiment and in this section we discuss one success story and one illustrative failure, each involving the application of fragment-based quantum chemistry to cluster spectroscopy.

Numerous spectroscopic observables can be formulated as energy derivatives,[218–220] so it is worth noting explicitly that GMBE(n) energy formulas can be differentiated term-by-term to afford fragment-based approximations to various energy gradients. These include the nuclear gradients dE/dx that are needed for geometry optimizations and *ab initio* molecular dynamics (MD) simulations, but also electric-field ($\mathcal{F}$) derivatives such as $d^2E/d\mathcal{F}_\alpha d\mathcal{F}_\beta$, which define polarizabilities. More exotic derivatives include $d^2E/d\mu_\alpha dB_\beta$, where $\boldsymbol{\mu}$ is the magnetic moment for a given nucleus and $\mathbf{B}$ is an applied magnetic field. These derivatives define the magnetic shielding tensor needed to compute nuclear magnetic resonance (NMR) chemical shifts.[218–220]

Considering the MBE, one might be tempted to express the energy gradient in the form

$$\frac{dE}{dx} = \sum_{I=1}^{N} \frac{dE_I}{dx} + \sum_{I=1}^{N} \sum_{J>I} \left(\frac{dE_{IJ}}{dx} - \frac{dE_I}{dx} - \frac{dE_J}{dx} \right) + \cdots, \tag{28}$$

which looks as if it can be evaluated by means of separate gradient calculations for monomers (dE_I/dx), dimers (dE_{IJ}/dx), etc. This is strictly correct only in the absence of any electrostatic embedding, however. Otherwise, there are additional terms in the analytic gradient that describe how the embedding potential on fragment I changes when fragment J is perturbed.[221,222] This significantly complicates the formulation of analytic gradients for most EE-GMBE(n) approaches, a fact that has not always been recognized. The present author has recently introduced a variational formulation that avoids this complexity,[222] which is discussed in Sec. III F. This issue does not affect the examples presented in this section, which eschew the use of electrostatic embedding.

In the context of vibrational spectroscopy of clusters, there have been numerous efforts to extend the reach of high-level *ab initio* methods by focusing on two-layer approaches that require a supersystem calculation but apply the most expensive level of theory using either the GMBE(1) approximation,[29,113,148,149,223,224] or else MBE(2).[137,138] A successful example is depicted in Fig. 11, illustrating harmonic vibrational spectra for one isomer of $(H_2O)_{16}$, computed at the MP2/aug-cc-pVDZ level using the MTA method developed by Gadre and co-workers.[148] The spectrum in red is the result of a traditional supersystem calculation, whereas the spectrum overlaid in

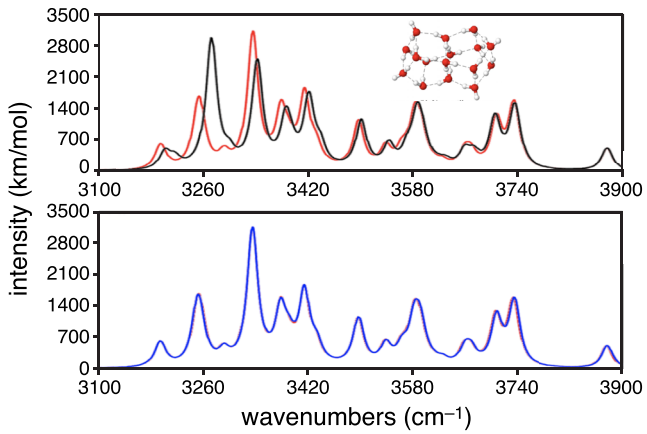

FIG. 11. Harmonic vibrational spectra (using 10 cm^{-1} broadening) for an isomer of $(H_2O)_{16}$ computed at the MP2/aug-cc-pVDZ level of theory. The supersystem result is shown in red in both panels, whereas the spectrum in black (upper panel) is obtained using the MTA approximation at the same level of theory. The blue spectrum (lower panel) is computed by "grafting" MP2/aug-cc-pVDZ onto MP2/6-31G according to Eq. (22), and is essentially indistinguishable from the supersystem MP2/aug-cc-pVDZ spectrum in red. Adapted with permission from N. Sahu and S. R. Gadre, J. Chem. Phys. **142**, 014107 (2015). Copyright 2015 AIP Publishing LLC.

black is computed using MTA [equivalent to GMBE(1)], with 11 primary fragments ranging in size from $(H_2O)_6$ to $(H_2O)_8$, along with 22 intersections that are smaller still. The MTA calculation thus consists in computing a Hessian for 33 different subsystems, none larger than $(H_2O)_8$. The spectrum obtained in this way is in good agreement with that obtained from the full-cluster Hessian computed at the same level of theory, except for an error in the relative intensities of a pair of peaks around 3300 cm^{-1}.

Even this small error disappears upon application of the "grafting" correction in Eq. (22), using 6-31G as the small basis set; see Fig. 11(b). Differentiating Eq. (22) twice with respect to nuclear coordinates, one obtains an approximation for the correlated Hessian matrix in the large basis set, namely,

$$\mathbf{H}^{\text{large}}(\text{super}) \approx \mathbf{H}^{\text{large}}_{\text{HF}}(\text{super}) + \mathbf{H}^{\text{large}}_{\text{corr}}(\text{GMBE}) + \mathbf{H}^{\text{small}}_{\text{corr}}(\text{super}) - \mathbf{H}^{\text{small}}_{\text{corr}}(\text{GMBE}), \quad (29)$$

where $\mathbf{H}_{\text{corr}}$ represents the difference between the MP2 and Hartree-Fock Hessians. (A similar multilevel scheme is used for the dipole moment derivatives needed to compute infrared intensities.[148]) Application of Eq. (29) does require the Hessian for the entire cluster, computed at the MP2/6-31G level in this particular case, but the resulting spectrum is indistinguishable from the full-system MP2 result computed in the larger basis set.

Even for strongly interacting systems such as the hydrated bisulfate cluster $HSO_4^-(H_2O)_{12}$, MP2 frequencies obtained in this manner differ from supersystem results by no more than 1–2 cm^{-1}.[148] The grafting procedure affords essentially exact infrared spectra for a variety of other cluster systems,[149,225] and has also been used as a basis-set correction for DFT calculations, e.g., with 6-311++G* as the large basis and 6-31G as the small basis in Eqs. (20) and (21).[149]

As a counterbalance to this highly successful application, we next discuss an unsuccessful attempt to use MBE(n) to model chiroptical spectra of solution-phase molecules,[226,227] the failure of which poses some interesting questions for the field. The scientific issue of interest is theoretical prediction of the specific optical rotation, $[\alpha]_\omega$, for chiral molecules such as methyloxirane (C_3H_6O) and methylthiirane (C_3H_6S). In both of these molecules, the sign of the specific rotation is wavelength-dependent in the gas phase and solvent-dependent in solution.[228] These are examples in which the solute imposes a "chiral imprint" on its solvation environment,[229,230] an effect that is not reproduced by continuum solvation models.[231] The sign of $[\alpha]_\omega$ can be sensitive to the level of quantum theory that is employed, and theoretical calculations of electronic circular dichroism spectra are generally found to be sensitive to high-level electron correlation effects, basis-set effects, and solvation effects.[228] As such, there is incentive to employ the very highest (and costliest) levels of *ab initio* theory, but at the same time these calculations must include some explicit solvent molecules.

This seems like a perfect situation in which to use MBE(n) in order to introduce explicit solvent effects in a tractable way, but in practice it is found that convergence of the MBE(n) sequence of approximations is dramatically worse for specific rotation as compared to properties such as the total energy or the static dipole moment, or even a response property such as the frequency-dependent polarizability, $\alpha(\omega)$.[226,227] Each of the latter properties converges to <1% error (as compared to the supersystem result) by $n = 3$ in the MBE,[226] whereas the specific rotation exhibits wildly oscillatory behavior, as seen in Fig. 12(a) for $C_3H_6O(H_2O)_{13}$. Convergence is not achieved until $n = 10$. (The $n = 14$ result is exact by definition.)

It is worth noting that specific rotation can be written as a derivative of the time-averaged quasi-energy,[227] which satisfies both a variational principle and a Hellmann-Feynman theorem,[219,220] meaning that application of MBE(n) to this property is theoretically justified, in principle. Evidently, however, not all molecular properties are similar in their many-body convergence properties. That said, the rotatory strength (mixed electric-dipole/magnetic-dipole polarizability) tensor that controls chiroptical properties is known to be more sensitive, as compared to other properties, to molecular vibrations, solvation effects, and other small changes in electronic structure.[220]

Slow convergence for specific rotation has been traced to a BSSE effect.[227] As discussed in Sec. III B, the MBE(n) sequence of approximations is oscillatory even for total energies[205] although the energy

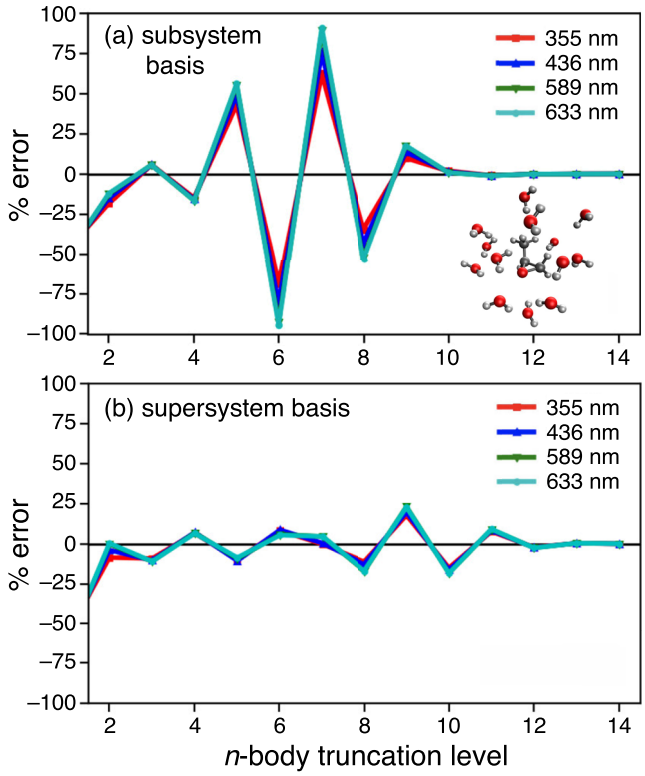

FIG. 12. Convergence of the MBE(n) sequence of approximations as applied to compute the specific rotation $[\alpha]_\omega$ at four different excitation wavelengths, for the $C_3H_6O(H_2O)_{13}$ cluster that is shown. (a) Normal MBE(n) approach, using the subsystem basis set for each subsystem calculation. (b) Results using the full cluster basis set for each subsystem calculation. All calculations were performed at the CAM-B3LYP/aug-cc-pVDZ level. Adapted with permission from B. G. Peyton and T. D. Crawford, J. Phys. Chem. A **123**, 4500–4511 (2019). Copyright 2019 American Chemical Society.

oscillations are quite small in percentage terms, as compared to what is plotted in Fig. 12(a) for $[\alpha]_\omega$, and may therefore go unnoticed. In contrast, oscillations in $[\alpha]_\omega$ are dramatic and unmistakable. These oscillations are significantly damped, however, if the subsystem calculations are each performed using the full cluster basis set, as shown in Fig. 12(b). The n-body sequence continues to oscillate, reflecting the inherent sensitivity of $[\alpha]_\omega$ to small changes in the electronic structure, but the magnitude of the oscillations is much reduced so that semiquantitative results might be obtainable with a low-order n-body approximation. It remains unclear whether the counterpoise corrections discussed in Sec. III B, and/or a GMBE(n) approximation with overlapping fragments, might be enough to tame the highly oscillatory nature of the many-body contributions to specific rotation.

D. Molecular crystals

Perhaps even more so than molecular clusters, molecular crystals cry out for a fragment-based approach. In contrast to solid-state inorganic materials, *molecular* crystals naturally decompose into weakly interacting monomer units and are essentially nothing more complicated than periodically replicated noncovalent molecular clusters. If the monomers are small, then high-accuracy quantum chemistry can be deployed rather easily despite the extended nature of the system, and there is a growing list of examples in which cohesive energies of crystals have been computed using the very best levels of theory [e.g., CCSD(T)/CBS] for the one-, two-, and (sometimes) many-body interactions.[65,129,152,232–239] One example is a theoretical prediction of the lattice energy of benzene that is precise enough (at −55.90 ± 0.76 kJ/mol) to warrant revisiting assumptions used to extrapolate the experimental data to 0 K.[236]

To obtain this level of quantitative agreement requires QM calculations at the four-body level, which remains challenging with highly correlated wave function approaches. However, significant progress on other properties can be made by combining multiscale frameworks with two-body electronic structure. This is exemplified by the HMBI scheme developed by Beran and co-workers,[36,129,131,239–241] which involves one- and two-body QM calculations under periodic boundary conditions in conjunction with either a polarizable force field or else a periodic Hartree-Fock or DFT calculation, in order to capture higher-order induction. Closely related is the *binary interaction method* (BIM) of Hirata *et al.*,[242,243] which is a periodic version of EE-MBE(2). Both approaches take advantage of the fact that periodic boundary conditions are relatively straightforward to incorporate into a monomer-based treatment of a molecular crystal,[244] especially if the far-field interactions (between distant subsystems) are already described using classical multipoles. The energy expression for periodic EE-MBE(2) is[243]

$$E_{\mathrm{BIM}} = E_{\mathrm{LR}} + \sum_I E_{I(\mathbf{0})} + \frac{1}{2}\sum_{I,J}\sum_{\boldsymbol{k}}\left[E_{I(\mathbf{0})J(\boldsymbol{k})} - E_{I(\mathbf{0})} - E_{J(\boldsymbol{k})}\right], \quad (30)$$

where $\boldsymbol{k}$ indexes a sum over lattice vectors. The quantity $E_{I(\mathbf{0})J(\boldsymbol{k})}$ is the energy of a dimer in which monomer I resides in the unit cell ($\mathbf{0}$) and monomer J is in replica cell $\boldsymbol{k}$. The quantity E_{LR} is an Ewald-type long-range electrostatic correction.

As noted by Hirata *et al.*,[243] the fragment-based approach to crystals is only weakly dependent on the periodic boundary conditions as compared to plane-wave methods. As a result, a reasonable approximation to the force constants needed in the Hessian, and in the dynamical force-constant matrix that is used to compute the phonon dispersion curve, is

$$\frac{\partial^2 E}{\partial x_{I(\mathbf{0})}\partial y_{J(\boldsymbol{k})}} \approx \frac{\partial^2 E_{I(\mathbf{0})J(\boldsymbol{k})}}{\partial x_{I(\mathbf{0})}\partial y_{J(\boldsymbol{k})}}. \quad (31)$$

This greatly facilitates calculation of phonon densities of states and thus thermal properties of crystals.[243] It is worth noting that if the embedding potential E_{LR} in Eq. (30) is derived on-the-fly from the monomer wave functions (e.g., in the form of atomic partial charges), then the gradient $\partial E_{\mathrm{LR}}/\partial x$ is nontrivial;[222] see Sec. III F.

Fragmentation methods can be used to predict finite-temperature properties of molecular crystals by performing geometry optimizations on either the Helmholtz (A) or Gibbs (G) free energy surface,[243]

$$A = E_{\mathrm{elec}} + U_{\mathrm{vib}} - TS_{\mathrm{vib}}, \quad (32a)$$

$$G = E_{\mathrm{elec}} + PV + U_{\mathrm{vib}} - TS_{\mathrm{vib}}. \quad (32b)$$

The electronic energy $E_{\mathrm{elec}}(V)$ depends on the volume of the unit cell, and the vibrational energy (U_{vib}) and entropy (S_{vib}) depend on both volume and temperature. These vibrational quantities can be estimated from harmonic partition functions, or better yet within a quasi-harmonic approximation (QHA).[245–248]

Even at 0 K, vibrational zero-point corrections have a significant effect on the structure of the solid, as shown for ice I_h in Fig. 13. Minimizing $E_{\mathrm{elec}}(V)$ without consideration of vibrations affords $\bar{V} \approx 18.3$ cm^3/mol (the "no QHA" result in Fig. 13), when the one- and two-body terms are described at the MP2/CBS level within the

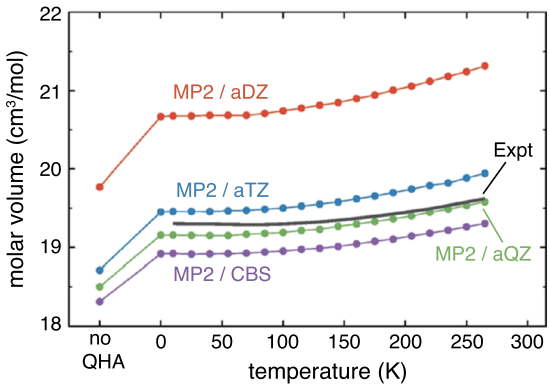

FIG. 13. Molar volume $\bar{V}(T)$ for ice I_h, computed using the HMBI formalism with one- and two-body terms described at the indicated levels of theory.[246] The molar volume is obtained by minimizing the free energy $A(V, T)$ with respect to V, using a quasi-harmonic approximation (QHA) for the vibrations. The "no QHA" result minimizes $E_{\mathrm{elec}}(V)$ at $T = 0$ K. Experimental data are from Ref. 249. Adapted with permission from Y. N. Heit and G. J. O. Beran, Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. **72**, 514–529 (2016). Copyright 2016 International Union of Crystallography.

composite HMBI scheme.[246] Staying at absolute zero but including zero-point energy [U_{vib} in Eq. (32)] increases the MP2/CBS molar volume to 18.9 cm^3/mol, which is closer to experiment (19.3 cm^3/mol at T = 0 K[249]) but not in quantitative agreement. Thermal expansion effects are smaller than zero-point effects and are described reasonably well within the QHA. Results for CO_2(s)[246] and CH_3OH(s)[239] show similar trends: MP2/aug-cc-pVQZ results are in good agreement with experiment, but the extrapolated MP2/CBS volume is systematically too small. CCSD(T)/CBS results are available but these offer little additional correction in the case of CO_2,[246] and move the results slightly farther from experiment in the case of CH_3OH.[239]

Hirata and co-workers[243,250–252] have used the periodic EE-MBE(2) approach to explore the finite-temperature properties of ice I_h. Figure 14(a) shows the unusual phenomenon of negative thermal expansion.[252] Upon warming from absolute zero, ice I_h initially undergoes a volume contraction, with the more typical thermal expansion observed only for $T \gtrsim 70$ K.[249] The calculations are in good agreement with experiment up to temperatures high enough to see this compression-to-expansion turnover. The calculations also provide a molecular-level explanation for the observed anomaly: bending modes of the hydrogen-bond network, once thermally populated, function to collapse the hexagonal cagelike voids in the structure of ice I_h, leading to an initial contraction of the volume with respect to temperature. This collapsing motion leads to a decrease in the frequencies of these modes upon thermal excitation, unlike most of the other hydrogen-bonding modes whose frequencies shift in the opposite direction.[243]

Figure 14(b) compares the inelastic neutron scattering spectrum of ice I_h to a calculated phonon density of states.[243,250] In the original experimental work,[253,255] the existence of two peaks in the translational density of states, labeled "3" and "4" in Fig. 14(b), was interpreted as evidence of two distinct forms of hydrogen bonds, one strong and one weak. This interpretation proved to be controversial.[256,257] Fortunately, the good agreement between theory and experiment in Fig. 14(b) lends credence to the molecular-level interpretation of the phonon spectrum obtained from the calculations, which can put the hypothesis to the test. EE-MBE(2) calculations suggest that there is no difference between the hydrogen bonds; rather, in proton-disordered ice, the hydrogen-bond stretching frequencies along the crystallographic axes of ice I_h are simply not equivalent.[243] Periodic DFT calculations support the same conclusion.[258]

As compared to water ice, CO_2(s) seems like a much simpler crystal, but one whose high-temperature and high-pressure phase diagram are nevertheless relevant to planetary geochemistry. In fact, our understanding of the extreme regions of CO_2's phase diagram continues to evolve, with several (putative) new phases having been reported since the late 1990s,[259] including polymeric or "nonmolecular" phases.[260] However, experiments under extreme thermodynamic conditions (such as Raman spectroscopy in a diamond anvil cell) sometimes provide only indirect evidence for phase transitions, and theoretical calculations may help to reduce the ambiguity associated with such measurements. There have been several attempts to compute the phase diagram for CO_2 using fragment-based quantum chemistry,[261–263] and one example is shown in Fig. 15.[261] The pressure-induced transition between phase I (ambient dry ice) and phases II, III, and VII is reproduced quantitatively by calculations at the MP2/aug-cc-pVDZ level, using the periodic EE-MBE(2) methodology.

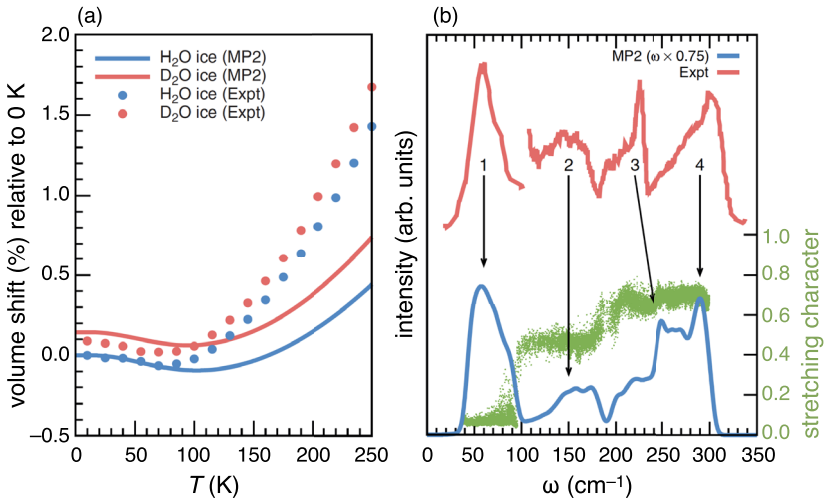

FIG. 14. Theory vs experiment for ice I_h, with calculations performed using periodic EE-MBE(2) at the MP2/aug-cc-pVDZ level. (a) Vibrationally corrected volume change at finite temperature, relative to $\bar{V}$ at 0 K. Experimental data[249] are shown in comparison to quasiharmonic calculations.[252] (b) Experimental inelastic neutron scattering spectrum (in red, from Ref. 253) compared to the MP2 phonon density of states (in blue, from Ref. 243, computed for a fully proton-disordered structure). Bond-stretching parameters (in green, from Ref. 254 and to be read from the axis on the right) measure the amount of stretching character at a given frequency. Panel (a) is adapted with permission from M. A. Salim, S. Y. Willow, and S. Hirata, J. Chem. Phys. **144**, 204503 (2016). Copyright 2016 AIP Publishing LLC. Panel (b) is from Hirata *et al.*, *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, Ch. 9, pp. 245–296. Copyright 2017 John Wiley & Sons.

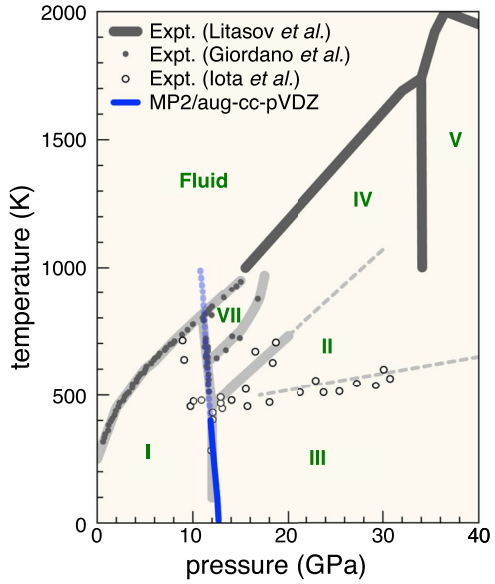

FIG. 15. Phase diagram for CO_2 based on three sets of experimental measurements.[259,264–267] The putative coexistence boundary between phases II and III is disputed,[263] and the measurements are kinetic (based on Raman spectroscopy[268]) rather than thermodynamic.[259] Also shown is a coexistence boundary computed at the MP2/aug-cc-pVDZ level using the EE-MBE(2) approach.[261] Adapted with permission from Li *et al.*, Nat. Commun. **4**, 2647 (2013). Copyright 2013 Springer Nature Publishing.

Periodic fragment-based methods can also be used for spectroscopic applications. NMR spectroscopy in particular can be useful to discriminate between crystal polymorphs,[36,132] and composite strategies based on DFT exhibit good fidelity with respect to periodic (plane-wave) DFT calculations.[269] Perhaps more importantly, the fragment-based approach facilitates the use of hybrid functionals at negligible additional cost, assuming that the method is interfaced with a Gaussian orbital electronic structure code, whereas the use of hybrid functionals in plane-wave DFT is often prohibitively expensive. Furthermore, MP2 theory is known to outperform even hybrid DFT for NMR chemical shifts,[270] but has not seen widespread use in this capacity. MP2-based NMR calculations would be straightforward to implement for molecular crystals using a fragment-based approach.

Raman spectra of CO_2(s) have been reported based on periodic EE-MBE(2) calculations.[243,263,271,272] At the level of CP-corrected MP2/aug-cc-pVDZ, these calculations afford a semiquantitative description of the pressure dependence of various Raman bands,[243,271] which provides the data necessary to determine pressure-induced splitting of the $\nu_1 + 2\nu_2$ Fermi dyad.[243,271,272] (This pressure dependence is used as a geophysical barometer.[243]) Experimentally, Raman spectroscopy of CO_2(s) was used to establish the "kinetic line" delineating phases II and III (see Fig. 15). Theoretical calculations of these same spectra, using the HMBI approach with MP2/CBS energetics, present a compelling case that the experiments have not detected a genuine phase boundary and that phases III and VII of CO_2(s) have precisely the same structure.[263]

This success is encouraging but of course CO_2(s) is the very simplest of molecular crystals, with intermolecular interactions limited to dispersion and quadrupole–quadrupole electrostatics. Its monomers have little internal structure except possibly in the most extreme regions of the phase diagram.[259,260] In contrast, CH_3OH(s) presents the additional complexity of flexible internal degrees of freedom and an intermolecular hydrogen bond. Figure 16(a) shows the phase diagram for methanol computed using two-body calculations at the CCSD(T)/CBS level, in conjunction with a periodic Hartree-Fock calculation to capture higher-order induction. This represents the current state-of-the-art in terms of what is routinely feasible from first principles and affords a semiquantitative picture of the phase diagram below the melting transition.

Figure 16(b) shows a close-up view of the α/β coexistence curve for CH_3OH(s), comparing MP2 and CCSD(T) results. This comparison demonstrates that post-MP2 correlation effects are relatively small but that basis-set effects are sizable. In the author's view, these results also suggest that the most likely path toward better accuracy involves a better description of three-body interactions, or perhaps improvements to the embedding scheme more generally. That said, results for molecular crystals are extremely promising, especially in comparison with the somewhat pessimistic view of the accuracy of MBE-based methods that was presented in Sec. III A.

To reconcile these two views, note that the periodic solids discussed here have rather small unit cells, meaning that the number of individual n-body subsystems is much smaller than in typical cluster or liquid calculations. The use of composite approaches furthermore allows the n-body expansion to be truncated at low order, while preserving sufficient accuracy to make useful predictions. Predictive accuracy does require the use of correlated wave functions and large basis sets, which demonstrates the need for fragment-based approaches that can handle the latter. This, in turn, appears to require the n-body subsystems to be iterated to self-consistency in the presence of interfragment exchange interactions. If these interactions are absent, as in the FMO approach, then problems arise with spurious delocalization of the fragment wave functions in large basis sets.[92,93] Methods that cannot handle large basis sets are of limited utility when correlated wave function calculations are desired or required.

Results for methanol's phase diagram also demonstrate the perniciousness of the polymorphism problem in crystal structure prediction.[247,273,274] Definitive prediction of the most stable polymorph at a given thermodynamic state point requires calculation of free energies to an accuracy of ~1 kJ/mol. This is demonstrated in vivid fashion by systematically shifting the predicted CCSD(T)/CBS free energy G_α of α-CH_3OH(s), by increments of just 0.5 kJ/mol. As shown in Fig. 16(c), this alters the α/β coexistence boundary in a qualitative way. A shift of ±0.5 kJ/mol in G_α changes the coexistence temperature by almost 150 K at $P = 1$ GPa.

E. Intermolecular energy decomposition

In the realm of noncovalent quantum chemistry, the basic theory of intermolecular interactions is fundamentally fragment-based. That theory is known as *symmetry-adapted perturbation theory* (SAPT),[203,275–277] but unlike most other methods discussed in

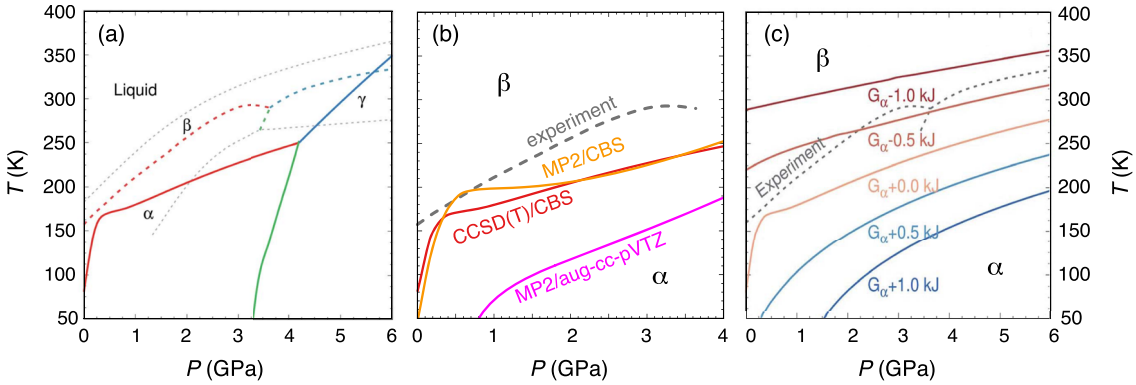

FIG. 16. (a) Experimental phase diagram for methanol (dashed curves) superimposed on a calculation of the solid/solid coexistence curves that is color-coded in the same way as the experimental data. In the calculation, two-body interactions are described at the CCSD(T)/CBS level, in a composite method that uses a periodic Hartree-Fock calculation to capture higher-order induction.[240] (b) Closer view of the α/β boundary, showing results at several levels of theory. The CCSD(T)/CBS coexistence curve is the same one shown in (a). (c) Results for the α/β coexistence curve, if the CCSD(T)/CBS free energy for the α phase is shifted by increments of 0.5 kJ/mol. The temperature scale is the same in all three panels. Reprinted with permission from C. Cervinka and G. J. O. Beran, Chem. Sci. **9**, 4622–4629 (2018). Copyright 2018 Royal Society of Chemistry.

this Perspective, which are designed as general-purpose (or at least, broadly applicable) procedures for large systems, SAPT is intended solely for the purpose of computing and analyzing intermolecular interaction energies. At its heart is a direct perturbative expansion of the interaction energy, E_{int}, meaning that SAPT is free of BSSE since there is no energy difference to compute in a potentially unbalanced basis set. Isolated-monomer wave functions serve as the unperturbed states and the perturbation consists of the intermolecular Coulomb interactions along with an operator that antisymmetrizes the monomer basis states, hence "symmetry-adapted." The terms in the perturbation series can be classified into physically meaningful components including electrostatics, exchange, induction, and dispersion. As a result, SAPT comes equipped with a natural energy decomposition analysis scheme:

$$E_{\text{int}}^{\text{SAPT0}} = E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} + E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)} + \cdots. \quad (33)$$

(Within this formalism, the charge-transfer energy is hidden in the induction energy but can be separated with additional calculations.[278–280])

Superscripts in Eq. (33) indicate orders in intermolecular perturbation theory. Intramolecular electron correlation can be introduced at low cost by using Kohn-Sham DFT calculations for the monomers, though care must be taken that the functionals exhibit correct asymptotic behavior.[281] With that proviso, the second-order "SAPT0" expression that is explicated in Eq. (33) is comparable to MP2 in both accuracy and cost. CCSD(T)-quality interaction energies can be obtained by including additional terms in the perturbation series, as in the SAPT2+ and SAPT2+(3) methods,[203,282] or by solving coupled-perturbed Kohn-Sham equations to obtain the dispersion energy from frequency-dependent density susceptibilities, as in the SAPT(DFT) approach.[275,276] Each of these methods incurs increased cost with respect to the $\mathcal{O}(N^5)$ scaling of SAPT0, however.

At second order, the dispersion and exchange-dispersion terms,

$$E_{\text{dispx}} = E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)}, \quad (34)$$

are both the least accurate components of Eq. (33) and also the most expensive, scaling as $\mathcal{O}(N^4)$ and $\mathcal{O}(N^5)$, respectively. As such, there has been much effort to replace these terms with low-cost, *ab initio* dispersion potentials of the damped atom–atom C_6/R^6 form:[283–288]

$$E_{\text{dispx}} = -\sum_{a,b}^{\text{atoms}} f_{\text{damp}}(R_{ab}) \frac{C_{6,ab}}{R_{ab}^6}. \quad (35)$$

Although similar in form to empirical dispersion corrections used in DFT,[289,290] an important distinction is that dispersion can be cleanly separated from the other energy components in SAPT, so there is no double-counting. The parameters $C_{6,ab}$ in Eq. (35) can be fit to *ab initio* dispersion data (e.g., from SAPT2+ calculations) to afford a method that Lao and Herbert have called SAPT+*ai*D.[287] An even more promising version,[291] which is closer to a first-principles approach, includes a self-consistent many-body dispersion (MBD) correction based on atoms-in-molecules polarizabilties.[292]

SAPT was originally construed to compute interaction energies of dimers. Although three-body extensions have been formulated,[293–295] they incur $\mathcal{O}(N^7)$ cost as compared to $\mathcal{O}(N^5)$ for dimer SAPT0 or $\mathcal{O}(N^3)$ for SAPT+*ai*D and SAPT+MBD. While the fragment-based nature of two-body SAPT makes it amenable to larger clusters, in keeping with the discussion in Sec. II A, pairwise application of SAPT (in any of its flavors) will omit important many-body induction. To address this, my group has introduced an extended version of SAPT called "XSAPT,"[284–288,291,296] in which the zeroth-order wave functions are obtained from a fragment-based SCF procedure using atomic embedding charges. This is essentially EE-MBE(2) with two-body interactions computed using SAPT.

For large supramolecular complexes, XSAPT is among the most accurate quantum chemistry methods for noncovalent interactions, approaching ~1 kcal/mol with respect to CCSD(T)/CBS benchmarks and rivaling the best-available DFT methods in comparison with interaction energies derived from experimental binding affinities.[287,291] At the same time, the fragment-based nature of XSAPT means that it is lower in cost even as compared to traditional supramolecular DFT. For example, an XSAPT calculation on a 157-atom DNA intercalation complex, using an augmented triple-ζ basis set (4561 basis functions), requires about 6 h on a single compute node, vs 13 h for a DFT calculation with the ωB97M-V functional.[291,296] Furthermore, no basis-set extrapolation or counterpoise correction is required in order to obtain an accurate interaction energy from XSAPT, in contrast to supramolecular calculations.

Two examples of pharmaceutical ligand binding are presented to illustrate the results. The first is the aforementioned intercalation complex consisting of the anticancer drug ellipticine ($C_{17}H_{14}N_2$) bound to a pair of dinucleotides representing double-stranded DNA. The other system consists of the antiretroviral drug indinavir ($C_{36}H_{47}N_5O_4$) bound to a model of the active site of HIV-2 protease; see Fig. 1(b). Interaction energies computed with several different approaches are listed in Table I, along with the XSAPT energy decomposition. For the ellipticine/DNA complex, the XSAPT+MBD interaction energy is similar to that obtained from the supramolecular approach, using functionals such as ωB97M-V that perform well for noncovalent interactions.[297,298] (Analogous DFT calculations for indinavir/HIV have not been attempted because this system involves 8346 basis functions.) Examining the energy decompositions, it is unsurprising to learn that the π-stacked intercalation complex would be unbound in the absence of a −71 kcal/mol dispersion energy. It is interesting to note, however, that the HIV/indinavir complex exhibits an even larger dispersion energy, −135 kcal/mol. The explanation is that dispersion is size-extensive and the HIV complex has about twice as many atoms as the DNA complex. The take-home message is that dispersion is no less important in a system having no obvious π-stacking interactions but simply a large number of electrons.

TABLE I. Interaction energies for two ligand/macromolecule complexes.[a]

	E_{int} (kcal/mol)	
Method	Ellipticine/DNA	Indinavir/HIV
B97M-V (+counterpoise)[b]	−41.3	...
ωB97M-V (+counterpoise)[b]	−43.7	...
HF-3c[c]	−41.7	−132.8
PBEh-3c[d]	−37.3	−119.1
XSAPT+MBD[e]	−41.7	−125.4
XSAPT energy decomposition		
E_{elst}	−22.2	−114.9
E_{exch}	59.2	190.0
E_{ind}	−8.0	−65.9
E_{disp}	−70.7	−134.6

[a]Data are taken from Ref. 296.
[b]def2-TZVPPD basis set.
[c]Semiempirical method of Ref. 299.
[d]Semiempirical method of Ref. 300.
[e]def2-hpTZVPP basis set.

The energy decomposition that comes with SAPT is useful as a means to construct transferrable intermolecular force fields,[301–303] by fitting individual terms in the energy expression to empirical functions with correct physical features, e.g., R^{-6} distance dependence for dispersion and exponential distance dependence for exchange repulsion.

It should also be noted that there are alternatives to SAPT when it comes to fragment-based energy decomposition analysis. The *pair energy decomposition analysis* (PIEDA) technique[304] uses the two-body nature of the FMO2 energy expression to generalize the venerated (and also frequently maligned) Kitaura-Morokuma energy decomposition scheme[305] to many-body systems. The latter is an example of a "supramolecular" energy decomposition approach, of which there exist other examples as well.[277,306–308] In principle, any of these can be applied within a fragment-based formalism provided there is a means for describing many-body polarization. An example can be found in how SAPT-style energy decomposition was extended to many-body clusters within the XSAPT formalism.[285]

F. Analytic gradients and *ab initio* molecular dynamics

The possibility of performing *ab initio* MD simulations within a fragment-based formalism is especially attractive for a system such as liquid water where the use of very small fragments facilitates simulations beyond the DFT level. Several such simulations have been reported recently using MP2 and coupled-cluster doubles (CCD) for the electronic structure, in conjunction with variants of periodic EE-MBE(2).[76,77,243,309] Radial distribution functions (RDFs) from the MP2 simulations are shown in Fig. 17(a), along with what is arguably the best available DFT simulation, using a hybrid functional and a self-consistent van der Waals correction.[310] RDFs from all three simulations compare quite well to experiment. Simulations by Hirata and co-workers using the BIM approach [Eq. (30)] were performed at T = 250 K in order to match the experimental RDFs at T = 300 K,[243] because MP2 water is known to be denser than real water,[311] thus requiring a lower temperature (at the same density) in order to maintain a stable liquid phase.[312] The self-diffusion coefficient computed at the MP2 level [Fig. 17(b)] is in much better agreement with experiment than that obtained either from the classical TIP3P potential or from Car-Parrinello DFT simulations, the latter of which are known to predict dynamics that are much too sluggish.

The basis set used in the fragment-based MP2 simulations is aug-cc-pVDZ, selected so that MP2 interaction potentials for $(H_2O)_2$ closely match CCSD(T)/CBS results, which is only the case if the basis set contains diffuse functions.[309] Although the use of diffuse functions is common in molecular calculations, and evidently necessary also for bulk water, these are ordinarily omitted in condensed-phase simulations because their presence significantly degrades the effectiveness of integral screening and is therefore cost-prohibitive. Use of diffuse functions in the condensed phase

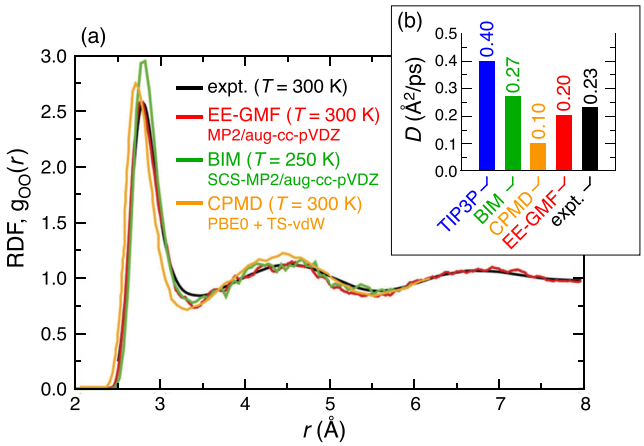

FIG. 17. (a) Oxygen–oxygen radial distribution functions and (b) self-diffusion constants for liquid water, from *ab initio* MD simulations. The EE-GMF simulations are from Ref. 76, and the BIM simulations are from Ref. 309; both methods are variants of EE-MBE(2). The Car-Parinello MD simulations are from Ref. 310. Adapted with permission from J. Liu, X. He, and J. Z. H. Zhang, Phys. Chem. Chem. Phys. **19**, 11931–11936 (2017). Copyright 2017 The PCCP Owner Societies.

also engenders problematic linear dependencies.[313] Neither of these issues is particularly troublesome in fragment-based approaches, where the linear dependency problem is no more severe than in typical small-molecule calculations.

One interesting observation from the two MP2-based simulations is that despite the high degree of similarity between the two approaches, both of which are variants of EE-MBE(2) and both of which predict similar oxygen–oxygen RDFs, these two studies predict rather different coordination environments for the oxygen atoms in liquid water. BIM simulations at the level of spin-component-scaled[314] (SCS-)MP2/aug-cc-pVDZ predict an average of 3.8 hydrogen bonds per water molecule,[243] consistent with the conventional tetrahedral view of the liquid.[315] The most probable oxygen coordination number is 4 (also consistent with the tetrahedral picture), although the distribution has a long tail and the average coordination number is 4.7.[243] The fairly large number of penta-coordinated oxygen atoms is offered to explain the slightly higher density of liquid water (as compared to ice) at 0 °C.[243]

On the other hand, EE-GMF simulations by Liu *et al.*[76,77] predict roughly equal fractions of "single-donor" and "double-donor" water molecules, at both the MP2/and also the CCD/aug-cc-pVDZ level. This brings to mind an old debate, seemingly put to rest,[316] about a possible interpretation of liquid water in terms of "rings and chains," i.e., a two-coordinate picture rather than a four-coordinate picture.[317] Additional analysis is required to resolve this discrepancy.

For applications to water, EE-MBE(n) calculations that use embedding charges have sometimes taken these charges to be fixed quantities, taken for example from classical water model. This is reasonable for neat water but for more complex and heterogeneous systems, the embedding charges should be determined on-the-fly using the fragment wave functions. This seemingly straightforward modification significantly complicates the formulation of analytic energy gradients, however. To understand why, consider the analytic gradient of the MBE(n) energy as written in Eq. (28). This simple "sum of fragment gradients" expression is valid only if the derivatives dE_I/dx and dE_{IJ}/dx include response terms that express how the embedding charges change with respect to perturbation of x.[222] Such terms do not exist within the standard machinery needed for most QM/MM calculations.[193,195] Consequently, the analytic gradient of the EE-MBE(n) energy with wave function-derived charges cannot be obtained simply by performing a sequence of subsystem gradient calculations with an off-the-shelf electronic structure program.

This fact has not always been recognized, and often the requisite response terms are simply neglected.[27,105,108,119–121,243,318] Collins[105] argues that the magnitude of the neglected contributions to the gradient is comparable to a typical stopping criterion for geometry optimizations. Other calculations suggest that the neglected response terms have little effect on geometry optimizations and introduce errors of <50 cm^{-1} in vibrational frequencies.[222] Be that as it may, energy conservation in MD simulations is highly sensitive to the quality of the forces, and neglect of charge-response contributions to the gradient is likely the cause of catastrophic energy drift that is observed in *ab initio* MD simulations of polypeptides using the EE-GEBF approach.[319] Those simulations used atomic partial charges derived from natural population analysis, without consideration of the requisite charge derivatives, and encountered energy drift in excess of 0.6 E_h over just 5 ps of dynamics!

Proper analytic gradients are now available for FMO,[320–322] along with analytic second derivatives[323,324] and other response properties.[325,326] Prior to 2011, however, a number of "approximate analytic gradients" were reported for this method.[327–332] Because the FMOn energy expression is not variational, its analytic gradient requires solution of coupled-perturbed equations for the fragments, even when the underlying electronic structure model *is* variational.[320] This adds not only to the cost but also to the formal complexity of the method, leading, for example, to incomplete implementations of the analytic gradient for FMO in combination with polarizable continuum solvation models (PCMs).[333,334] As a result, analytic and numerical Hessians for the FMO + PCM approach afford slightly different vibrational frequencies.[335]

An alternative approach is to use a variational formulation of EE-GMBE(n),[222] which sidesteps the need for coupled-perturbed equations when used with an SCF level of electronic structure theory. The variational approach is based on the "XPol" charge-embedded SCF procedure[194,336–338] and requires minimal modification to the Fock matrix and analytic gradient, although these modifications cannot be made from outside of the electronic structure program. The variational form of the charge-embedded Fock matrix for subsystem K is[194,222]

$$F^K_{\mu\nu} = f^K_{\mu\nu} - \frac{1}{2}\sum_{j\notin K}\left\langle \mu \left| \frac{q_j}{\|\mathbf{r}-\mathbf{R}_j\|} \right| \nu \right\rangle + \sum_{k\in K}\frac{\partial E_{\text{emb}}}{\partial q_k}\frac{\partial q_k}{\partial P^K_{\mu\nu}}, \qquad (36)$$

where $\mathbf{f}^K$ is the Fock matrix in isolation and E_{emb} is a classical embedding potential. Equation (36) requires expressions for the derivatives $\partial q_k/\partial P^K_{\mu\nu}$ of the embedding charges with respect to the fragment density matrices, which have so far been formulated for Mulliken,[336] Löwdin,[192] ChElPG,[194] and Hirshfeld[296] embedding charges. Unlike

FMOn, in which only the monomers are iterated to self-consistency, a variational EE-GMBE(n) calculation requires self-consistent iterations on each subsystem, up to and including n-mers of fragments. The procedure is illustrated by the flow chart in Fig. 18.

By deleting the charge-derivative term in Eq. (36), along with its companion terms in the energy gradient, it can be shown definitively that the charge-response terms are necessary to obtain energy-conserving dynamics with EE-MBE(n).[222] (The same result has been demonstrated by deliberately omitting the coupled-perturbed equations in the FMOn analytic gradient.[339]) As a matter of philosophy, it is the author's view that whatever approximations may have gone into the design and formulation of a fragment-based *ansatz* for the total energy, one should insist that the analytic energy gradient be the exact derivative of that energy expression. The variational formulation of EE-GMBE(n) provides a general framework for doing so, and one that is agnostic to the details of the fragmentation procedure.

G. Computational cost and ways to reduce it

Curiously, the cost of fragment-based quantum chemistry calculations is discussed only sporadically in the literature. More often, the parallelizability is emphasized without providing detailed timing information to illustrate what the calculations really cost in comparison with a supersystem calculation, at the same level of theory,[190] but a few glimpses are available. For example, *ab initio* MD simulations of liquid water, using EE-MBE(2) at the CCD/aug-cc-pVDZ level, took a reported 3.5 min per time step running on 30 nodes with 28 cores each.[77] With a simulation time step of 1 fs, this means that the 15.0 ps of simulation time reported in Ref. 77 represents 36.5 days of "wall time," i.e., the time-to-solution according to the clock on the wall. Multiplying by the total number of processors, this comes out to a staggering 84 central processing unit (CPU)-years of computing time.

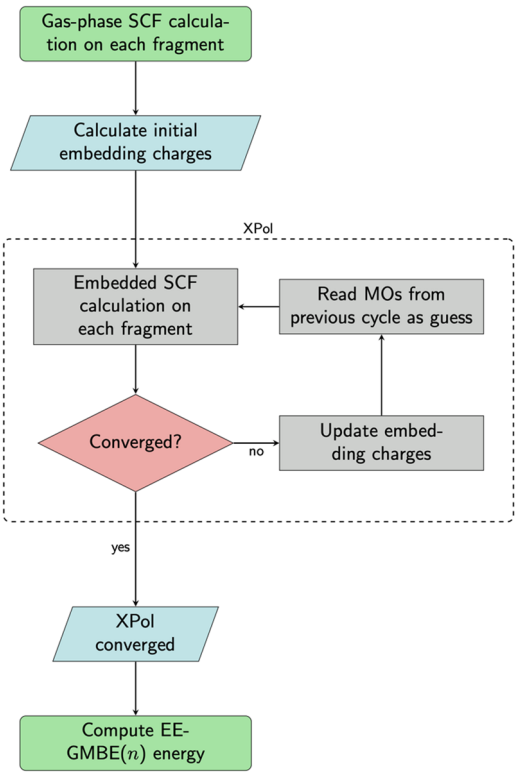

FIG. 18. Flow chart for the variational EE-GMBE(n) approach. The dashed box contains the iterative XPol procedure that determines the self-consistent embedding charges. Adapted with permission from Liu *et al.*, J. Phys. Chem. Lett. **10**, 3877–3886 (2019). Copyright 2019 American Chemical Society.

In evaluating the efficacy of fragment-based approaches, it is important to consider total CPU time and not simply wall time. Although the latter represents a particular user's time-to-solution, at a shared computing facility it is the total CPU time that is a fixed resource. Moreover, from a sustainability perspective, the best metric by which to measure the cost of a calculation may not have units of time at all but rather power consumption ("performance per watt"). It is important to note that the use of CPU throttling (i.e., dynamic frequency and/or voltage scaling) in modern processors means that power consumption is generally very low when the processor is idle, which largely negates the "might as well use them for something" mentality about computing. Efficient resource utilization often seems to be forgotten or ignored in the computational science community's march toward exascale computing.

The aforementioned *ab initio* MD simulations of liquid water used a relatively large QM system of about 140 water molecules, with an $\mathcal{O}(N^6)$ electronic structure method.[77] At present, such a simulation is only possible by means of fragmentation. Especially with lower-scaling methods, however, the crossover point at which the fragment-based approach actually becomes cheaper than the supersystem calculation that it aims to approximate occurs later than one might guess.

Figure 19 presents timing data for resolution-of-identity (RI-)MP2/aug-cc-pVTZ calculations on $(H_2O)_{20}$, from a study whose goal was to determine whether fragmentation methods could accurately predict relative energies for different cluster isomers.[68] This has proven to be a challenging problem, in part because BSSE is markedly larger in certain families of isomers than for others.[184] For this reason, timing data for various counterpoise corrections are also reported, including a full supersystem counterpoise correction [SSFC, Eq. (26)], as well as the MBCP(4) and GMBCP(2) corrections that are commensurate to MBE(4) and GMBE(2) energy calculations, respectively.

According to these data, the fragmentation approaches are significantly more expensive, in terms of aggregate CPU time, as compared to a conventional RI-MP2 calculation for the full cluster. The total CPU time for the MBE(4) calculation is about 8× the supersystem cost, and MBCP(4) is about twice as expensive as the full-system counterpoise correction. The GMBE(2) calculation is 74× more expensive than the supersystem calculation, and GMBCP(4) is 34× more expensive than SSFC. The power of fragmentation only becomes evident upon switching to a metric of cost *per subsystem*, which is shown on the right in Fig. 19. The largest subsystems in MBE(4) are water tetramers, and the cost per subsystem is reduced by more than 1000× as compared to the cost of a calculation on $(H_2O)_{20}$. The GMBE(2) calculations use 3–4 water monomers per fragment and therefore subsystems as large as $(H_2O)_8$, so the cost reduction per subsystem is smaller.

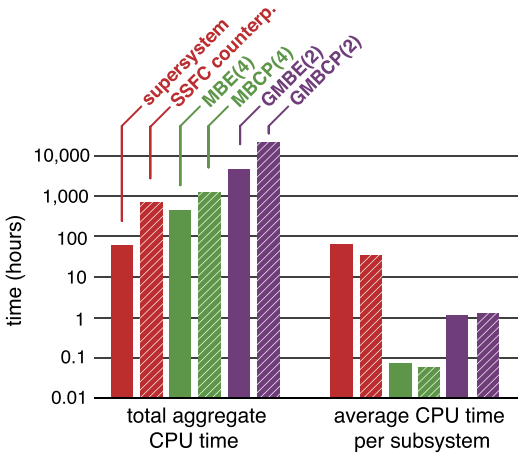

FIG. 19. Total computer time required for a single-point energy calculation on $(H_2O)_{20}$ at the RI-MP2/aug-cc-pVTZ level, showing the cost of a full supersystem calculation along with its MBE(4) and GMBE(2) approximations. Hatched bars show the time required for the counterpoise correction that is commensurate to each energy calculation. On the right, the timings are divided by the number of subsystem calculations that are required. Supersystem calculations were multithreaded across 20 cores, whereas subsystem calculations for the fragment-based methods were run in serial but distributed across processors. All calculations were performed using the Q-Chem program.[340] Data are from Ref. 68.

To understand these data, it helps to enumerate the number of subsystems that are required for each calculation, which is done in Table II for both $(H_2O)_{20}$ and $(H_2O)_{55}$. Application of MBE(4) to the smaller cluster generates 6175 distinct subsystem calculations, which is reduced to 4263 for GMBE(2), although the subsystems are somewhat larger. The corresponding counterpoise corrections engender about a fourfold increase in the number of terms although other work has concluded that only two-body counterpoise corrections are important.[59–61] These data make it clear that one should not assume that the fragment-based calculation will actually reduce the total wall time that is required, and furthermore, the total CPU time may very well increase upon fragmentation.

That said, the fact that one large calculation has been reduced to a (very large) number of smaller calculations is still useful because the fragment-based approach may be more readily restartable as compared to the supersystem calculation, and easier to perform in small segments. Segmentation is especially important when it comes to storage costs. Whereas the CPU cost of a RI-MP2 or CCSD(T) calculations scales as $\mathcal{O}(N^5)$ or $\mathcal{O}(N^7)$, respectively, both methods require $\mathcal{O}(N^4)$ storage (memory and/or disk), and it is the storage cost that is often the practical limitation in correlated wave function calculations. The use of fragmentation as a form of checkpointing in massively parallel applications represents a trivial, scalable, and fault-tolerant way to divide the computational effort over a large number of processors, possibly over a very long *calendar* time.

That said, it is worth considering ways to bring down the cost, by neglecting or approximating the smallest n-body terms. In keeping with the notion that quantum mechanics is operative only at short length scales, a natural idea is to describe all of the many-body effects classically, with only one- and two-body terms computed at a QM level. This is the approach that is taken in the HMBI method,[73] with good success for molecular crystals (Sec. III D).[36] The EE-GMF method developed by Liu *et al.*[74–77] also routinely uses screening on top of what is otherwise EE-MBE(2) or EE-MBE(3).

An alternative version of this idea is the *effective fragment molecular orbital* (EFMO) method,[32,189,341–345] originally designed as an approximation to FMO2 in which QM calculations for well-separated dimers are replaced by calculations using the effective fragment potential (EFP).[346–348] The latter is a polarizable force field that is parameterized in an automated way based on electronic structure calculations. EFP can be applied to the entire supersystem at negligible cost, thus EFMO amounts to a multilayer method in which short-range, two-body interactions (only) are described using QM calculations. Operationally, the EFMO procedure amounts to an energy expression[189]

$$E^{\mathrm{EFMO}} = \sum_I E_I^{(0)} + \sum_{\substack{I \\ R_{IJ}\le R_{\mathrm{cut}}}} \sum_{J>I} \left(\Delta E_{IJ}^{(0)} - E_{IJ}^{(\mathrm{ind})}\right) + \sum_{\substack{I \\ R_{IJ}> R_{\mathrm{cut}}}} \sum_{J>I} E_{IJ}^{\mathrm{EFP}} + E_{\mathrm{total}}^{(\mathrm{ind})}, \tag{37}$$

TABLE II. Number of subsystem calculations required for several different fragmentation methods, for two different water clusters. Adapted with permission from Lao *et al.*, J. Chem. Phys. **144**, 164105 (2016). Copyright 2016 AIP Publishing LLC.

	$(H_2O)_{20}$				$(H_2O)_{55}$			
Size	MBE(n)[a]	MBCP(n)[b]	GMBE(2)[a]	GMBCP(2)[b]	MBE(n)[a]	MBCP(n)[b]	GMBE(2)[a]	GMBCP(2)[b]
$n = 2$	190	380	...	...	1 485	2 970	...	...
$n = 3$	1140	3 420	...	...	26 235	78 705	...	...
$n = 4$	4845	19 380	...	...	341 055	1 364 220	...	...
$n = 4$–6	...	...	4113	16 040	...	...	17 883	78 650
$n = 6$–9	...	...	150	1 110	...	...	1 469	11 275
Total	6175	23 180	4263	17 150	368 775	1 445 895	19 352	89 925

[a]Calculations involving n monomers in an n-mer basis.
[b]Calculations involving one monomer in an n-mer basis.

in which $\Delta E_{IJ}^{(0)}$ is a two-body correction evaluated at the QM level, akin to that in Eq. (3) but based on *isolated* (rather than embedded) monomer energies $E_I^{(0)}$ and dimer energies $E_{IJ}^{(0)}$. The quantities E_{IJ}^{EFP} are the pairwise EFP energies, which replace the QM calculation when the intermolecular distance is large ($R_{IJ} > R_{\mathrm{cut}}$). The final term, $E_{\mathrm{total}}^{(\mathrm{ind})}$, is the total EFP induction energy for the supersystem. It serves to sum the many-body polarization to all orders, like other multilayer methods discussed in Sec. II C.

Because FMO2 is not a particularly good approximation for energies, it is not surprising that EFMO also fails in this regard, e.g., with errors for water clusters that are ~1 kcal/mol/monomer at the Hartree-Fock/6-31G* level.[341] Whereas Eq. (37) suggests an approximation to FMO2, results for isomers of $(H_2O)_{16}$ in Fig. 8 suggest that the approximation is inconsistent, especially in the larger 6-311++G(3df, 2p) basis set. In fact, EFMO is sometimes advertised as being *more* accurate than FMO2,[32,189] which may be objectively true but is a rather weak endorsement. Errors in total binding energies for isomers of $(H_2O)_{32}$ span a considerable range, from 204 kcal/mol for FMO2, to 116 kcal/mol for FMO3, to 48 kcal/mol for EFMO.[189]

Moreover, the efficacy of distance-based screening is likely to suffer in heterogeneous systems, especially if the fragments are considerably larger than H_2O. In polyalanine α-helices, for example, the secondary structure is stabilized by conformations characterized by long-range, cooperative alignment of the alanine dipole moments, and many-body effects for these geometries remain significant at much longer length scales as compared to water clusters.[349] These long-range cooperative interactions are also much more significant in α-helices as compared to β-strands,[349] so neglecting them outright will have a detrimental effect on the prediction of relative conformational energies for polypeptides and proteins. One way to account for these long-range interactions is by means of a low-level supersystem calculation, in which case aggressive distance-based cutoffs might still be acceptable for the high-level fragment calculations, which are needed only to describe short-range QM interactions while the supersystem calculation takes care of long-range classical interactions.

In an attempt to avoid the need for a supersystem calculation, energy-based screening has been suggested as an alternative to distance-based screening of MBE(n).[60,349,350] Figure 20(a) provides a proof-of-concept demonstration for water clusters of various sizes, showing that the three-body energy corrections ΔE_{IJK} obtained from QM calculations are highly correlated with classical estimates based on dipole polarizabilities. (The same is true of the four-body corrections.[349]) By using an energy cutoff rather than a distance cutoff, the overwhelming majority of three- and four-body interactions can be neglected while remaining faithful to MBE(n) energetics. This is demonstrated for $n = 3$ in Fig. 20(b), where a cutoff $E_{\mathrm{cut}} = 0.2$ kJ/mol introduces an overall error of <0.1 kJ/mol in the total energy predicted by the MBE(3) approach.

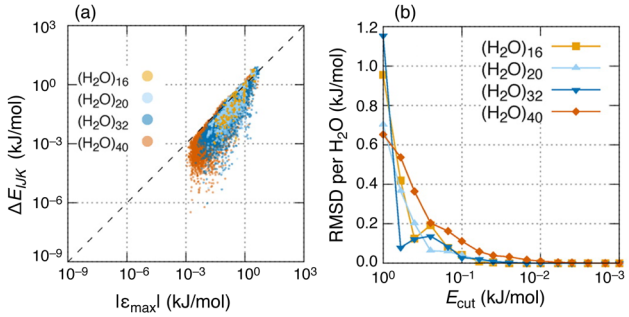

FIG. 20. Proof-of-principle demonstration of the efficacy of energy-based screening, for three-body interactions in water clusters. (a) Distribution of three-body interaction terms ΔE_{IJK} for water clusters of different sizes vs the maximum three-body polarization interaction $|\varepsilon_{\mathrm{max}}|$ estimated using classical polarization theory. (b) Error introduced when three-body terms with $|\Delta E_{IJK}| < E_{\mathrm{cut}}$ are neglected, as a function of the cutoff threshold. Adapted with permission from J. F. Ouyang and R. P. A. Bettens, J. Chem. Theory Comput. **12**, 5860–5867 (2016). Copyright 2016 American Chemical Society.

As shown in Fig. 21(a), this cutoff reduces the number of three-body terms to a value that appears to grow only linearly with respect to cluster size. The same is true for the four-body terms, and this makes energy-screened MBE(n) a genuinely $\mathcal{O}(N)$ approach, whereas in the absence of screening this method incurs a prefactor $N_{\mathrm{sub}} \propto \mathcal{O}(N^n)$ in Eq. (1). Energy-based screening also proves to be more effective and less erratic than the distance-based alternative, despite the fact that water is probably a best-case scenario for the latter. Changes in the ring structures within the data set of $(H_2O)_N$ isomers, at $N = 23$ and again at $N = 29$,[351] are

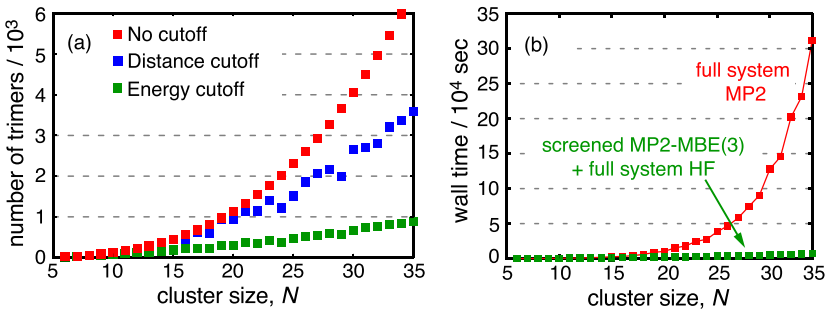

FIG. 21. (a) Total number of three-body terms before and after application of either a distance-based cutoff ($R_{\mathrm{cut}} = 7$ Å) or else an energy-based cutoff ($E_{\mathrm{cut}} = 0.25$ kJ/mol), for a sequence of clusters $(H_2O)_N$. (b) Timings (on a single 40-core node) for MP2/aug-cc-pVDZ calculations on $(H_2O)_N$, along with those for an energy-screened MBE(3) approximation at the same level of theory combined with a supersystem Hartree-Fock/aug-cc-pVDZ correction. Data are from Ref. 350.

TABLE III. Timing data and energy errors (ΔE) for GMBE(1) "molecular tailoring" calculations, including the basis-set grafting procedure of Eq. (20). Adapted with permission from S. S. Khire, L. J. Bartolotti, and S. R. Gadre, J. Chem. Phys. **149**, 064112 (2018). Copyright 2018 AIP Publishing LLC.

		aug-cc-pVDZ[a]				aug-cc-pVTZ[a]			
		N_{basis}[b]	ΔE	Wall time (min)[c]		N_{basis}[b]	ΔE	Wall time (min)[c]	
System	Method		(mE_h)	Fragment	Standard		(mE_h)	Fragment	Standard
Taxol	ωB97X-D	1885	2.38	168	361	4025	1.44	1439	4 044
α-cyclodextrin	B3LYP	2058	0.78	61	174	4416	0.28	577	2 035
γ-cyclodextrin	B3LYP	2744	0.23	110	345	5888	0.13	1107	10 054
Vancomycin	B3LYP	3006	0.74	384	517	6379	0.08	3437	10 568
$(H_2O)_{64}$	ωB97X-D	2624	0.85	173	385	5888	0.08	1324	2 844
$(H_2O)_{32}$	MP2	1312	1.02	202	675	2944	0.33	4897[d]	22 414[d]
$(H_2O)_{16}$	MP2	656	0.26	10	78	1472	0.35	202	1 178[d]

[a]Grafting procedure uses aug-cc-pVXZ as the large basis and cc-pVXZ as the small basis.
[b]Number of functions in the large basis set.
[c]Calculations were run on a single 16-core node except where otherwise indicated. The "fragment" time includes the cost of the small-basis supersystem calculation, whereas the "standard" time is the cost of the large-basis supersystem calculation.
[d]Supersystem calculations were performed on a 20-core node.

evident in the distance-based screening data in Fig. 21(a) and result in sudden changes in the efficacy of the procedure. In contrast, energy-based screening smoothly interpolates through these structural transitions.

A practical implementation of energy-based screening for MBE(n) has recently been developed using the EFP force field to make *a priori* estimates of the subsystem energies.[350] QM calculations of the n-body energy corrections $\Delta E_{IJ\ldots}$ are performed only if $|\Delta E^{\mathrm{EFP}}_{IJ\ldots}| > E_{\mathrm{cut}}$. When combined with a supersystem EFP calculation (MIM2-style), this amounts to an energy-screened MBE(n) analog of EFMO. At the three-body level, however, this method is still not accurate enough to handle the challenging problem of relative isomer energies in water clusters, even with the supersystem EFP correction.[350] Accuracy of ±1 kcal/mol with respect to supersystem MP2 can be achieved by adding a supersystem Hartree-Fock correction to the energy-screened MBE(3) energy computed at the MP2 level.[350] This composite approach remains significantly faster than supersystem MP2, as shown by the timing data in Fig. 21(b). Notably, these data represent wall times for calculations performed on a single 40-core node, meaning that the fragment-based calculation is more efficient even when measured in aggregate CPU time and does not require massive parallelization to be feasible.

As compared to MBE(3), and certainly as compared to the high-accuracy GMBE(2) and MBE(4) methods, the GMBE(1) approximation typically requires a vastly smaller number of subsystems. As a result, GMBE(1) is often cheaper than a supersystem calculation even without the need for screening. In some cases, this is true even for composite methods that require a supersystem calculation at a low level of theory, as Gadre and co-workers have consistently demonstrated.[112,113,117,118,148,149]

Some timing data from Gadre's recent work are presented in Table III, including both water clusters and macromolecules.[118] These calculations employ the "grafting" correction of Eqs. (20) and (21), with aug-cc-pVXZ as the large (target) basis set and the corresponding cc-pVXZ as the small basis set. The fragment-based calculation thus requires a full-system calculation in the smaller basis set; nevertheless, this composite approach systematically outperforms a supersystem calculation in the target basis set, even with both calculations performed on a single computer node, without resorting to large-scale parallelization to rescue the fragmentation method. This is especially true at the MP2 level; however, the fragment-based approach remains less expensive even for DFT calculations, while introducing errors $\lesssim 2 \times 10^{-3} E_h$ in the total energy. It remains to be seen whether GMBE(1) can consistently handle the challenging problem of relative energies.

IV. MACROMOLECULAR FRAGMENTATION

Molecular clusters and crystals are sufficient to highlight many of the strengths and limitations of fragment-based quantum chemistry, without introducing the complexity of fragmentation across covalent bonds. The full power of fragmentation, however, is only unleashed once these methods are brought to bear on macromolecular systems. This section briefly highlights applications along these lines, without delving into the details of how the severed bonds should be capped. This same issue is faced in the well-established context of QM/MM calculations,[94–98] so will not be discussed here. The need for link atoms,[96,97] hybrid orbital caps,[92,99,100] or "pseudobonds"[98] certainly introduces an additional layer of complexity, but issues of accuracy and cost that were documented above for noncovalent applications are equally relevant in macromolecules.

What should be clear from the noncovalent applications surveyed in Sec. III is that high accuracy is achievable (both in an absolute sense and also in the sense of high fidelity to the underlying quantum-chemical model), but at appreciable cost in many cases. In view of this, when it comes to condensed-phase and macromolecular

applications, it is worth noting that within the statistical mechanics community, there is a general feeling that when faced with a choice between an accurate-but-expensive potential energy surface and a cruder model that can be sampled effectively, the latter option is almost always preferable. (Call it the "primacy of entropy" mentality, whereas quantum chemists tend to adopt an enthalpy-centric viewpoint.) Recognizing that fragment-based approaches have not yet advanced to the point of providing good sampling, except possibly when combined with semiempirical quantum chemistry,[352–354] this section aims to discover whether there are macromolecular applications where fragmentation can yet be useful. Only a few highlights are provided, to give an appreciation for what is realistically feasible at a level of accuracy that might solve practical problems.

A. Energetics

Accurate prediction of relative conformational energies is challenging in clusters, as for example in the case of isomers of $(H_2O)_{16}$ for which FMO-based methods fail badly (Fig. 8). As compared to FMOn, MBE(n)-based methods that iterate the subsystems to self-consistency fare much better and in fact show great promise for the difficult problem of predicting relative energies of crystal polymorphs (Sec. III D), which is the solid-state analog of the conformational landscape problem. With a heavy-atom fragmentation criterion of ≈3 Å, the GMBE(2) approach has proven to be essentially exact for clusters even without resort to electrostatic embedding,[46,68] so it is natural to inquire how this method performs for macromolecules.

Proteins represent a straightforward test case, and calculations using two amino acids per fragment afford good fidelity with the macromolecular calculation, if used in conjunction with the overlapping-fragment GMBE(2) approach that can describe both "through-bond" and "through-space" interactions.[128] A sequence-based fragmentation scheme can be used, with caps that are either hydrogen link atoms or else more sophisticated "conjugated caps" that terminate the severed peptide bonds with amino acid functional groups, as in the MFCC approach of Zhang and co-workers[30,109] that is illustrated in Fig. 4.

Figure 22(a) examines the conformational energy landscape for two small proteins, comparing supersystem energies at the level of M06-2X/6-31G* with EE-GMBE(2) results at the same level of theory.[128] These calculations use a "pair-pair" algorithm to generate a set of fragments consisting of all pairs of amino acids that are either covalently bonded or else hydrogen-bonded to one another. Subsystems consist of pairs of fragments ($F_i \cup F_j$) and their mutual intersections ($F_i \cap F_j$), which are therefore no larger than four amino acids. Relative energy profiles are computed for 20 conformations of each protein, obtained by gas-phase geometry optimization of a solution-phase ensemble of structures. Relative energy profiles obtained from an EE-GMBE(2) calculation are nearly indistinguishable from those obtained by performing a DFT calculation on the entire protein.

The same is true for some larger proteins (e.g., 4DP1 with 99 residues and 1A4A with 258 residues), as shown in Fig. 22(b), where the fragment-based calculations use a form of EE-GMBE(1) in which dimers are constructed from all pairs of amino acids within 4.5 Å of one another.[355] Essentially, this approach uses the two-body energy correction in Eq. (3) in conjunction with conjugated caps and the MFCC energy formula of Eq. (12) that subtracts out the energies of the overlapping caps, GMBE(1)-style.

Both sets of calculations in Fig. 22 suggest that it is possible to approximate biomolecular quantum chemistry with extremely high fidelity using fragment-based methods. That said, the accuracy of

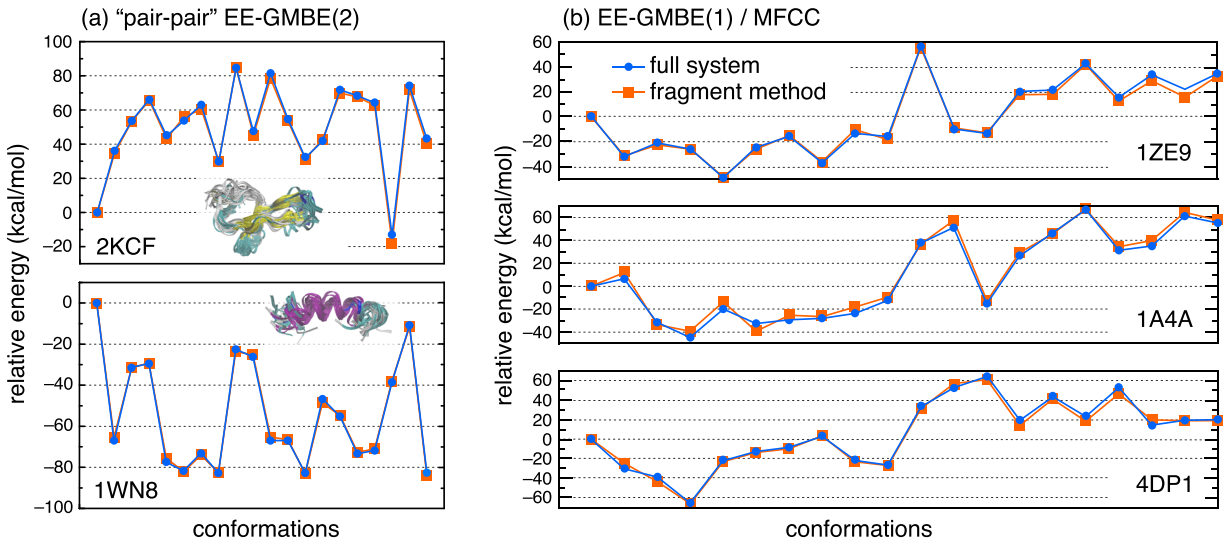

FIG. 22. Relative conformational energies of proteins, comparing supersystem calculations (in blue) to fragment-based results (in orange). Results are shown using (a) the EE-GMBE(2) method, with a "pair-pair" algorithm for generating fragments,[128] and (b) the MFCC approach,[355] which is a form of EE-GMBE(1). PDB codes for each protein are indicated, and the ensembles of conformations used in (a) are shown. All calculations were performed at the M06-2X/6-31G* level of theory. The data plotted in (a) are from Ref. 128 and those in (b) are from Ref. 355.

these calculations (with respect to the corresponding supersystem DFT calculation) raises some interesting questions. One question is how the accuracy can be so good, given that both fragmentation approximations are limited to short-range pairs of fragments. The answer is likely that all of the calculations in Fig. 22 are performed on conformations derived from the compact, native structure of the protein, such that long-range interactions are largely similar within the ensemble of structures, and errors associated with neglecting the long-range interactions largely cancel across the conformational profiles in Fig. 22. A more challenging test would be to compute the relative energies of compact structures vs elongated "random coil" structures. It is possible that distance cutoffs of 4.0–4.5 Å really are sufficient, especially once the first elements of secondary structure have begun to form, but it is unclear whether this has been rigorously tested.

A second issue is the choice of density functional. Despite its popularity in noncovalent applications, M06-2X does not include nonlocal electron correlation and therefore it lacks proper long-range dispersion,[356] which is necessary to describe the full spectrum of noncovalent interactions.[357] In the present context, this means that some of the genuine long-range interactions that are discarded due to fragmentation were already absent from the supersystem M06-2X calculations that were used as benchmarks. Benchmark calculations using functionals that include long-range dispersion would help to address this issue. These could be empirical DFT+D methods,[289,290] or else proper nonlocal correlation functionals such as ωB97M-V.[297,298]

Some DFT+D benchmarks for proteins have been reported and used to assess the MIM2 and MIM3 fragmentation schemes.[144] With the three-level MIM3 strategy [Fig. 5(c)], a semiempirical method such as PM6-D3 can be used for the supersystem component of the composite calculation, while staying within a target accuracy of 2 kcal/mol with respect to a full-protein DFT calculation.[144] Although timing data are not provided in Ref. 144, it is indicated that the supersystem PM6 calculation is the most expensive part of the composite strategy, which is therefore no more expensive than standard semiempirical methods but with potentially much better (and more controllable) accuracy.

Also useful would be benchmarks at correlated levels of wave function theory that include long-range dispersion. Such calculations generally demand larger basis sets than 6-31G*, consideration of which raises another question about the protein data in Fig. 22: why isn't BSSE a larger problem? As discussed in Sec. III B, BSSE manifests rather differently in a supersystem calculation than it does in the subsystem calculations, which can artificially hamper convergence of MBE(n) because higher-order n-body interactions are intertwined with BSSE.[184,205] In addition, BSSE can artificially stabilize compact structures of biomolecules as compared to more open ones.[197–199] Part of the reason why this imbalance does not cause a problem here is that all of the structures that are considered are relatively compact.

In general, it would be useful to have additional benchmarks to assess how faithful the fragmentation methods are for the subtle problem of energy differences in biomolecules, including ligand/protein binding energies. Chloride ion binding to some organic macrocycles has been examined recently,[142] using a two-layer strategy with CCSD(T) as the high-level method, applied GMBE(1)-style, and M06-2X-D3 as the low-level method applied to the entire supersystem. This strategy affords good agreement with supersystem CCSD(T) interaction energies.[142] For protein/ligand interaction energies, a MIM3 approach has been tested with PM6-D3 as the lowest-level method.[143] Calculated interaction energies (ΔU) correlate reasonably well with experimental free energies (ΔG), but additional testing is needed to understand where the error cancellation arises since the target level of theory (B97-D3 in Ref. 143) is of modest accuracy and the entropy change was not considered except by means of a continuum solvation model to include a desolvation penalty for the ligand. Ligand binding to macromolecules has also been considered using FMO, but these studies have tended to be more qualitative in nature,[358–363] as befitting the limitations of that method.

Both sets of protein calculations in Fig. 22 use subsystems that are 50–60 atoms in size,[128,355] meaning that these calculations should be feasible (albeit nontrivial) even in larger basis sets and with correlated wave function levels of theory. That said, it is worth considering the cost of applying fragmentation to proteins, even at modest levels of theory. Figure 23 presents timing data for GMBE(2) calculations on proteins as large as 70 amino acids (1142 atoms), at the Hartree-Fock/6-31G* level of theory.[128] Immediately evident is the fact that a Hartree-Fock calculation on the entire protein is significantly *cheaper* than the fragment-based calculation, if that cost is measured in aggregate CPU time [Fig. 23(a)]. This fact is sometimes disguised in the literature by comparing to timings from a slow quantum chemistry code, or one whose parallel efficiency is poor, as these limitations tilt the comparison toward the smaller and trivially parallelizable calculations in the fragment-based approach. The full-protein timings reported in Fig. 23, however, are parallelized across a modest 12 cores, using an efficient DFT code.[340] Zhang and co-workers[355] have also reported that fragment-based approximations can be more expensive than full-protein DFT calculations although the fragmentation approach used in Ref. 355 includes far fewer terms as compared to GMBE(2), hence the crossover point (at which the fragment-based calculation becomes lower in cost) occurs in a smaller protein.

The crossover will occur much earlier if a correlated wave function method is used instead of DFT, as demonstrated in a recent application of GMBE(1) to ionic liquid clusters containing 10 ion pairs (250–270 atoms).[75] At the M06-2X/6-31G* level of theory, the fragmentation approach is more expensive than the corresponding supersystem calculation, but this is reversed at the MP2/6-31+G* level. The latter calculations involve 2690 basis functions, and the fragmentation approximation requires only 1%–2% as much CPU time as the supersystem calculation.

The power of fragmentation thus lies mostly in its parallelizability, viz., the fact that the time-to-solution can be made almost arbitrarily small, down to the cost of a single subsystem calculation. The same is true for storage requirements, which is a significant consideration in correlated wave function calculations. Finally, for calculations that are to be distributed across a great many processors, checkpointing (in case of hardware failure) quickly becomes a necessity, and the natural divisibility of fragment-based calculations makes this easier as well. These are tremendous advantages in the context of high-performance computing, but one should not lose sight of the tremendous amount of computer time that will nevertheless be required in order to extend quantum chemistry to macromolecular systems. This consideration, in turn, must serve to

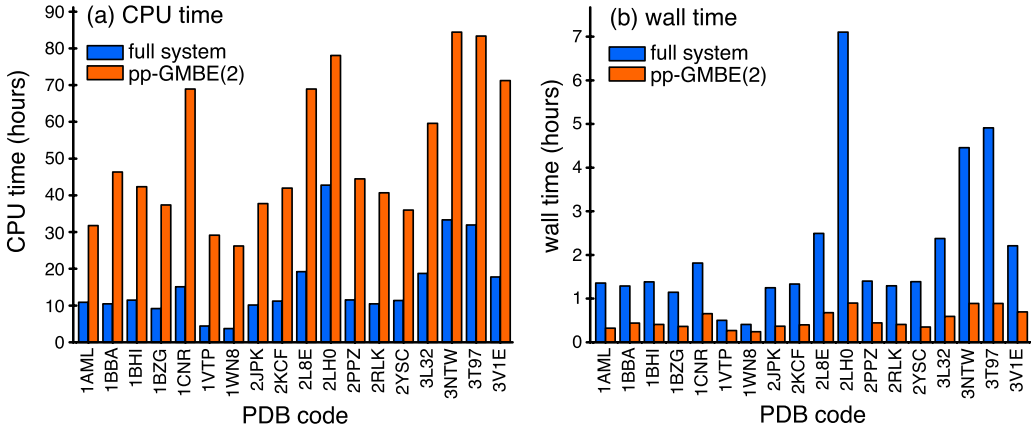

FIG. 23. Timing data for Hartree-Fock/6-31G* calculations on a set of proteins, comparing a calculation on the full protein to a "pair-pair" (pp)-GMBE(2) calculation.[128] The data are separated into (a) aggregate computer time summed across all processors vs (b) time-to-completion or "wall time." All supersystem calculations were performed on a single 12-core node, whereas the pp-GMBE(2) calculations (with two amino acids per fragment) used 10 nodes, but each subsystem calculation was performed on only one node. Adapted with permission from J. Liu and J. M. Herbert, J. Chem. Theory Comput. **12**, 572–584 (2016). Copyright 2016 American Chemical Society.

guide what applications are selected for these methods. In a system with thousands of atoms, fragment-based approaches can certainly facilitate single-point energy and gradient computations at hundreds (or possibly thousands) of geometries, but 10^6 evaluations of the gradient (as required for 1 ns of MD) is probably not feasible except using semiempirical quantum chemistry,[352–354] or possibly with multilayer GMBE(1)-style methods if these can be reduced to semiempirical cost.[144]

B. Spectroscopy

As compared to accurate calculation of energetics, spectroscopic applications may be an area where there is a clearer path forward for fragment-based approaches because direct contact with experiment can be made based on a modest number of energy and gradient evaluations, as compared to what would be required for *ab initio* MD. Spectroscopic applications might involve full or partial geometry optimization to relax the system at an *ab initio* level of theory, followed by calculation of appropriate derivatives representing molecular properties.[218–220]

For example, composite strategies have been developed for computing vibrational spectra of large molecules. One of these is the grafting-corrected GMBE(1)-style molecular tailoring method of Gadre and co-workers, which has proven very successful for cluster spectroscopy[148,149,225] (see Fig. 11) and has also been adapted for macromolecules.[148,149,224] Using fairly large fragments (≈100 atoms in some cases) and a full-system calculation of the Hessian at a low level of theory (e.g., BP86/6-31G or MP2/6-31G), essentially exact results can be obtained at a higher level of theory (BP86/TZVP or MP2/6-31G**), with the higher-level method applied only to fragments. To date, this method has not been coupled to a geometry optimization at the same level of theory, so full-system geometry optimization at the higher level of theory is required and the geometry is not quite stationary at the level of theory used to compute frequencies. In contrast, macromolecular geometry optimizations have been reported using FMO*n*,[329,334,364] and also MIM*n*[123] or its equivalent.[134,136,138]

MIM2-based approaches to vibrational spectroscopy, developed by Jose and Raghavachari,[123,124,365,366] are more promising in terms of their extensibility to large systems and their ability to optimize the geometry and compute harmonic frequencies at a consistent level of approximation. Various combinations have been considered, in which a high-level method is applied to fragments, GMBE(1)-style (in what the authors call MIM1), in combination with a low-level supersystem calculation. The high-level method is typically DFT in a triple-ζ basis set, since this is generally good enough for quantitative or semiquantitative vibrational spectroscopy within the harmonic approximation, while the low-level method might be double-ζ DFT or better yet a semiempirical method. Vibrational frequencies are generally accurate to within a few cm^{-1} as compared to supersystem benchmarks at the higher level of theory, while the low-level supersystem calculation helps to correct relative intensities.[123] These trends are illustrated in Fig. 24(a), which reports harmonic vibrational spectra for α-$(glycine)_{10}$.

The accuracy of fragment-based vibrational frequencies and intensities is consistent with the notion that the frequencies are largely dictated by local chemical structure. Typically, vibrational normal modes are delocalized only in the sense that they do not correspond to individual chemical bonds; they do tend to localize within individual chemical moieties or functional groups. There are exceptions, for example in the case of the amide I band in proteins,[369] where quasidegeneracies and transition dipole coupling amongst the carbonyl oscillators causes the excitation to delocalize over several residues.[370] However, these effects emerge naturally from diagonalization of a Hessian that is constructed from local information. The α-$(glycine)_{10}$ spectra in Fig. 24(a) that are computed from fragment-based approaches are in excellent agreement with the supersystem result, the latter of which includes the full effects of excitonic coupling. Raman spectra of

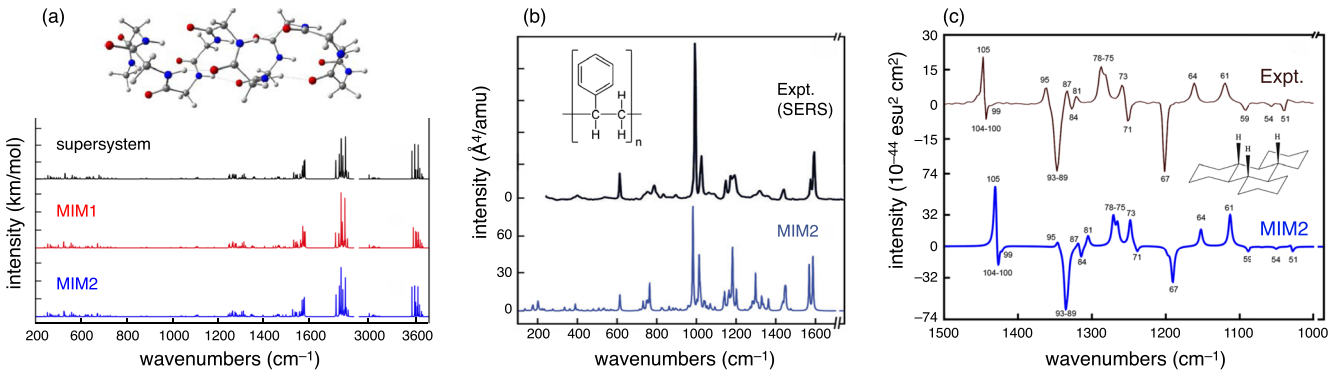

FIG. 24. Vibrational spectra from MIM*n* calculations. (a) Infrared spectra of $(\text{glycine})_{10}$ shown in an α-helix configuration, computed at the supersystem B3LYP/6-311G** level (upper panel) and its GMBE(1)-style MIM1 approximation (middle panel). The lower panel depicts the MIM2 spectrum obtained by adding a supersystem PM6 correction [Eq. (19)] to the MIM1 calculation.[123] (b) Raman spectra for polystyrene, comparing the experimental (surface-enhanced) spectrum[367] (top) to the MIM2 spectrum (bottom).[365] The latter is computed using B3LYP/6-311+G** as the high-level method and Hartree-Fock/6-31G as the low-level method. (c) Vibrational circular dichroism spectra of perhydrotriphenylene, comparing experiment[368] (top) and a MIM2 calculation (bottom, for the particular enantiomer that is shown). The calculation uses MPW1PW91/aug-cc-pVTZ-f as the high-level method and MPW1PW91/6-31+G* as the low-level method.[366] Panel (a) is adapted with permission from K. V. J. Jose and K. Raghavachari, J. Chem. Theory Comput. **11**, 950–961 (2015). Copyright 2015 American Chemical Society. Panel (b) is adapted with permission from K. V. J. Jose and K. Raghavachari, Mol. Phys. **113**, 3057–3066 (2015). Copyright 2015 Taylor & Francis. Panel (c) is adapted with permission from K. V. J. Jose, D. Beckett, and K. Raghavachari, J. Chem. Theory Comput. **11**, 4238–4247 (2015). Copyright 2015 American Chemical Society.

α-$(\text{alanine})_{10}$ computed using FMO2 are also in excellent agreement with supersystem calculations.[324,335] On the other hand, infrared and Raman intensities depend on derivatives of the dipole moment and polarizability, respectively, and these quantities are presumably sensitive to induction effects. This is why the MIM2 approach improves the intensities relative to MIM1 [see Fig. 24(a)].

In contrast to the very challenging problem of accurate energy predictions, vibrational frequencies compare well to experiment even at relatively low levels of theory such as DFT. In combination with a fragmentation approach, this allows quantum chemistry to make direct contact with experiment in complex systems, at tractable cost.[124] Examples depicted in Figs. 24(b) and 24(c) show that DFT-based MIM2 vibrational spectra compare well enough with experiment to make assignments, both for the Raman spectrum of polystyrene[365] and for the vibrational circular dichroism spectrum of perhydrotriphenylene.[366] The latter can be used to assign the absolute stereochemistry of this molecule.[368] Note that these calculations include a careful treatment of the link-atom contributions to the gradient and Hessian.[123,365,366]

In terms of cost, it is reported that the FMO2 Hessian for a molecule with 400 atoms can be evaluated in 16 h on 72 cores at the B3LYP/6-31G* level of theory, with a total memory requirement of 1.6 Gb.[335] Raman spectra on full proteins have been reported at the FMO2 level,[326] which run in a matter of several days on 228 processors with 2 Gb of memory per processor. The parallel efficiency of these calculations is as high as 90% in some cases,[324] meaning that wall-clock times could be reduced considerably by further parallelization. Additional parallelization could be obtained by computing frequencies as finite differences of analytic first derivatives, which introduces essentially no error in the fingerprint region of the spectrum,[371] and reduces the memory requirement to that of a single-point energy calculation.

NMR is arguably the single most important spectroscopic technique in all of chemistry and is considered here as a final example. Calculation of chemical shielding tensors has been implemented within FMO2, and results at the Hartree-Fock/6-31G* level were shown to be in good agreement with supersystem results.[325] However, response properties have notoriously onerous basis-set demands and 6-31G* results for chemical shieldings cannot be considered reliable. In polypeptides, ^{13}C chemical shieldings computed with 6-31G* differ by up to 36 ppm with respect to converged results,[372,373] with even larger deviations for ^{17}O.[372,374] The basis-set demands make FMO*n* problematic for this particular property although some cancellation of errors may occur when chemical *shifts* are computed, by subtracting the shieldings in a reference compound such as trimethylsilane.

To obtain converged results while keeping the basis size tractable, conventional NMR calculations often rely on "locally dense" basis sets,[375] which saturate the atoms whose chemical shifts are desired while using a smaller basis set for the rest of the molecule. (Multiple separate calculations are then required in order to obtain chemical shifts for the entire molecule.) For polypeptides, reasonable agreement with supersystem calculations is obtained using as few as three amino acids, with a dense basis set only on the central one.[372] This "tripeptides-in-molecules" approach, combined with the dense psSeg-3 basis set that is specifically designed for NMR calculations,[376] achieves an accuracy of 0.20 ppm for ^{1}H and 0.76 ppm for ^{13}C shieldings in a particular decapeptide,[372,373] with respect to a supersystem calculation at the same level of theory and using the psSeg-3 basis set for the entire molecule. SMF3 calculations applied to the same decapeptide afford errors of 0.26 ppm (^{1}H) and 0.87 ppm (^{13}C).[373] Although the SMF3 results are quite good, the success of the tripeptides-in-molecules approach suggests that fragmentation may not actually be necessary for low-cost chemical shift calculations in large molecules.

In general, the correlation between fragment-based and supersystem chemical shifts is outstanding.[372,373,377] The use of locally dense basis sets may be incompatible with methods that employ overlapping fragments, and attempts to generalize this idea for use with SMF did not improve the results.[373] That said, in a fragment-based approach, the individual calculations are inherently smaller, so it may be possible to saturate the basis set on a fragment in a way that is not feasible in a macromolecule. It remains to be seen whether the subsystem approaches can reproduce subtle differences in chemical shifts in different environments, e.g., to discriminate between crystal polymorphs whose chemical shifts might differ by 1–5 ppm or even less.[36,132] To go from *ab initio* chemical shifts at a particular protein geometry to data that can be used to assign an NMR spectrum will require configurational averaging and perhaps a treatment of solvent effects.

Solvent effects have been considered in other studies where subsystem or fragmentation approaches were applied to compute NMR chemical shifts in full proteins.[378–381] Results often correlate well with experiment, even when modest basis sets are used. Examples include a pentapeptides-in-proteins QM/MM approach using either an implicit solvent approach[379] or else a combined implicit and explicit solvent approach.[380] MIM2 has also been applied to protein chemical shifts, using both implicit and explicit solvent.[381] These methods are compared for the same protein in Fig. 25, where the ^{1}H and ^{13}C chemical shifts obtained from the subsystem calculations are plotted against the experimental data set of chemical shifts. The agreement is very good, and in fact the correlations with experiment are surprisingly good even for the pentapeptides-in-proteins calculation with no solvent. Subtle differences in chemical shifts can easily hide in these correlation plots, however, and it will be interesting to see whether NMR spectra for macromolecules can be assigned using fragment-based calculations, as has been done for NMR spectra of molecular crystals.[36,132]

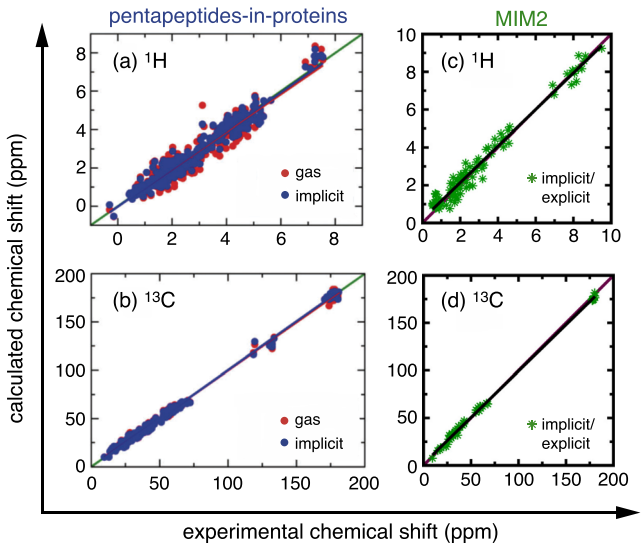

FIG. 25. Theory vs experiment for ^{1}H (top) and ^{13}C (bottom) chemical shifts in ubiquitin. The two panels on the left show results from pentapeptides-in-proteins QM/MM calculations, performed either in the gas phase (red symbols) or else in implicit solvent (blue symbols).[379] Calculations were performed at the B3LYP/6-31G** level except for the carbonyl ^{13}C atoms, for which a locally dense combination of 6-311++G** and 4-31G* was used. The two panels on the right show MIM2 results obtained with a combined implicit/explicit solvation approach.[381] The high-level method in the MIM2 calculations is MPW1PW91/6-311++G(2d, 2p), and the low-level method is MPW1PW91/6-31G. Panels (a) and (b) are adapted with permission from T. Zhu, X. He, and J. Z. H. Zhang, Phys. Chem. Chem. Phys. **14**, 7837–7845 (2012). Copyright 2012 The PCCP Owner Societies. Panels (c) and (d) are adapted with permission from K. V. J. Jose and K. Raghavachari, J. Chem. Theory Comput. **13**, 1147–1158 (2017). Copyright 2017 American Chemical Society.

V. LESSONS AND QUESTIONS

I will summarize this work with a set of lessons, which attempt to distill this Perspective into a few key points, then add some open questions that may serve as a starting point for thinking about future directions for this field.

• **Lesson #1**: *For high accuracy calculations, the subsystems must be iterated to self-consistency in the presence of exchange interactions*. In the presence of favorable electrostatic interactions with neighboring nuclei but in the absence of Pauli repulsion to prevent spurious electron delocalization, fragment-based calculations are only viable when compact basis sets confine the electrons to molecules. The noniterative procedure used for the *n*-body subsystems in FMO*n* makes this method fast and scalable,[324,341,382–384] but also limits its applicability to small basis sets.[92,93] These may not be appropriate for use with correlated wave functions, especially where high accuracy is desired. Embedded MBE(*n*) methods that iterate the *n*-body subsystems to self-consistency do not suffer this limitation, and numerous applications using MP2 or CCSD(T) with aug-cc-pVXZ basis sets (up to quadruple-ζ) have been reported.

• **Lesson #2**: *A focus on very small fragments obscures the performance for other applications*. Water clusters have been a ubiquitous test bed for MBE(*n*)-based methods, perhaps to a fault. The water monomer is small enough, yet many-body interactions in $(H_2O)_N$ are important enough, that by selecting H_2O as the fundamental fragment, one has effectively dug a very deep hole (in terms of accuracy) that requires a very tall *n*-body ladder in order to escape. A noteworthy exception is the case of molecular crystals, where the unit cell often consists of no more than a few molecules and periodic boundary conditions make the situation much more tractable.

For large but finite systems, or for liquids where sizable periodic simulation cells are required, the need for high-order *n*-body interactions has profoundly deleterious consequences for the efficacy of the method, introducing serious finite-precision and error-accumulation issues.[33,68,72] Furthermore, the combinatorics of four-body approximations quickly renders these approaches cost-prohibitive as the number of fragments grows,[68] although energy-based screening is a promising strategy to curtail the number of subsystem electronic structure calculations that is required.[349,350] Complementary to these screening approaches are ONIOM-style composite methods such as MIM*n*,[110] which obviate the need for high-order *n*-body terms by summing the many-body effects to infinite order at a low level of theory.

Using somewhat larger fragments is usually advantageous despite the higher cost per subsystem, because more of the

short-range interactions are treated exactly (at the specified level of QM theory) and classical embedding approximations are pushed to longer length scales. Overlapping fragments appear to be superior to disjoint ones, even for noncovalent clusters, and the overlapping-fragment approach is practically mandatory for macromolecules. Using overlapping fragments, good fidelity with supersystem calculations is achieved with only a relatively small numbers of fragments, using GMBE(1)-style approximations such as GEBF,[27] MFCC,[30] and molecular tailoring.[29] These methods achieve dramatic reduction in the overall cost of the calculation, even (in favorable cases) when a low-level supersystem calculation is required for accuracy.[118]

• **Lesson #3**: *Accurate energy calculations are particularly challenging*. For large water clusters, quantitative accuracy can be achieved using subsystems as large as $(H_2O)_{10}$.[46,68,127] Short of this, however, accurate prediction of relative energies remains beyond the capability of many fragment-based methods.[68,350] This problem is solvable using composite methods that combine a high-level method applied to fragments with a low-level supersystem calculation, albeit at the price of abrogating $\mathcal{O}(N)$ scaling. That said, well-chosen composite methods afford orders-of-magnitude cost reduction with respect to application of correlated wave functions to large systems, and therefore have a useful domain of applicability in "medium-size" systems that would otherwise be unreachable by high-level quantum chemistry. A focus on linear scaling at all costs obscures other potential applications that can have an impact on chemistry right now.

• **Lesson #4**: *Spectroscopic applications may be less demanding and more directly relatable to experiment*. Energy gradients may be more forgiving than absolute or relative energies, in terms of the need for high-order induction and/or electron correlation effects. In favorable cases, spectroscopic predictions may also place less stringent demands on basis sets as compared to accurate prediction of energies. As such, it may be possible to predict spectroscopic observables in large systems, to an accuracy that is useful to experimentalists, while maintaining the treatment of electron correlation and the fragmentation approximation at tractable levels. In contrast to *ab initio* MD simulations, prediction of spectra may require only tens or hundreds of energy gradient calculations, not millions or tens of millions. That said, the sampling problem remains unsolved. In other words, how should one obtain the geometries to be used in a high-level calculation of the spectroscopy? If the cost of the fragment-based component of a multilayer method can be reduced below the cost of a semiempirical calculation for the entire system, then the combination of the two could serve as a next-generation semiempirical methodology, with improved (and systematically improvable) accuracy.

Finally, let us ponder the future of fragment-based quantum chemistry by means of some existential questions for the field.

• **Question #1**: *Do we need (or want) the MP2/6-31G* answer for a large system?* That level of theory is chosen largely as a rhetorical device although it is the level used in some parallel scalability tests for FMO.[383,384] One might just as well ask, in the context of the calculations in Fig. 22, whether the M06-2X/6-31G* energy for a protein is a genuinely useful quantity. Either level of theory is relatively crude as *ab initio* quantum chemistry calculations go, so one needs to be clear-eyed about what (if any) practical information is gleaned from such a calculation. Are there genuine scientific questions that can be answered at these levels of theory? Are these questions that cannot be answered by semiempirical calculations? It is possible that the semiempirical and fragment-based *ab initio* approaches are complementary, with the latter serving to spot-check or correct the former.

• **Question #2**: *What scientific questions are answerable with these methods?* Long timescale, fragment-based *ab initio* MD simulations will not be routine any time soon; ergo, the sampling problem is not solvable at a strictly *ab initio* level of theory. The difficulty in predicting relative energies with fragment-based approaches furthermore suggests that one cannot always trust the potential energy surface generated by fragmentation to be completely faithful to the underlying quantum-chemical model. Classical force fields and semiempirical quantum chemistry are in no danger of being displaced as the workhorse methods for doing sampling, so in view of that fact, what is the role of higher-level methods in large systems? One possible answer is spectroscopy, which is only indirectly accessible (at best) via classical simulations.

Outside of molecular crystals, there have been very few fragment-based studies that can claim quantitative or predictive accuracy for thermochemical properties, i.e., for energies. This is especially true for the difficult problem of relative conformational energies, which are often dictated by subtle noncovalent interactions requiring high levels of electron correlation. The situation seems more favorable for spectroscopy. In the author's opinion, the future is promising for overlapping-fragment strategies that are applied to make spectroscopic predictions, possibly in combination with low-level supersystem embedding. Application of low-level (but *ab initio*) electronic structure methods to large systems may also have qualitative utility, e.g., in terms of energy decomposition analysis, force-field validation, or screening of candidate structures in large systems.

ACKNOWLEDGMENTS

The author thanks the current and former members of his research group who have worked on fragment-based quantum chemistry. In chronological order, they are: Dr. Leif Jacobson, Dr. Ryan Richard, Professor Ka Un Lao, Dr. Jie Liu, Dr. Adrian Morrison, Dr. Kuan-Yu Liu, Kevin Carter-Fenk, Bhaskar Rana, and Paige Bowling. 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.

REFERENCES

[1] P. Pulay, "Localizability of dynamic electron correlation," Chem. Phys. Lett. **100**, 151–154 (1983).
[2] S. Sæbø and P. Pulay, "Local configuration interaction: An efficient approach for larger molecules," Chem. Phys. Lett. **113**, 13–18 (1985).
[3] S. Saebo and P. Pulay, "Local treatment of electron correlation," Annu. Rev. Phys. Chem. **44**, 213–236 (1993).
[4] W. Kohn, "Density functional and density matrix method scaling linearly with the number of atoms," Phys. Rev. Lett. **76**, 3168–3171 (1996).
[5] E. Prodan and W. Kohn, "Nearsightedness of electronic matter," Proc. Natl. Acad. Sci. U. S. A. **102**, 11635–11638 (2005).
[6] J. E. Subotnik and M. Head-Gordon, "Exploring the accuracy of relative molecular energies with local correlation theory," J. Phys.: Condens. Matter **20**, 294211 (2008).

[7]J. Kussmann, M. Beer, and C. Ochsenfeld, "Linear-scaling self-consistent field methods for large molecules," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**, 614–636 (2014).
[8]H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, and M. Dornbach, "Scalable electron correlation methods. I. PNO-LMP2 with linear scaling in the molecular size and near-inverse-linear scaling in the number of processors," J. Chem. Theory Comput. **11**, 484–507 (2014).
[9]M. Sparta and F. Neese, "Chemical applications carried out by local pair natural orbital based coupled-cluster methods," Chem. Soc. Rev. **43**, 5032–5041 (2014).
[10]D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin, and F. Neese, "Exploring the accuracy limits of local pair natural orbital coupled-cluster theory," J. Chem. Theory Comput. **11**, 1525–1539 (2015).
[11]D. G. Liakos and F. Neese, "Is it possible to obtain coupled cluster quality energies at near density functional theory cost? Domain-based local pair natural orbital coupled cluster vs modern density functional theory," J. Chem. Theory Comput. **11**, 4054–4063 (2015).
[12]W. Li, Z. Ni, and S. Li, "Cluster-in-molecule local correlation method for post-Hartree–Fock calculations of large systems," Mol. Phys. **114**, 1447–1460 (2016).
[13]M. Schwilk, Q. Ma, C. Köppl, and H.-J. Werner, "Scalable electron correlation methods. 3. Efficient and accurate parallel local coupled cluster with pair natural orbitals (PNO-LCCSD)," J. Chem. Theory Comput. **13**, 3650–3675 (2017).
[14]H.-J. Werner, C. Köppl, Q. Ma, and M. Schwilk, "Explicity correlated local electron correlation methods," in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, edited by M. S. Gordon (Wiley, Hoboken, 2017), Chap. 1, pp. 1–80.
[15]M. Saitow, U. Becker, C. Riplinger, E. F. Valeev, and F. Neese, "A new near-linear scaling, efficient and accurate, open-shell domain-based local pair natural orbital coupled custer singles and doubles theory," J. Chem. Phys. **146**, 164105 (2017).
[16]Q. Ma and H.-J. Werner, "Scalable electron correlation methods. 5. Parallel perturbative triples correction for explicitly correlated local coupled cluster with pair natural orbitals," J. Chem. Theory Comput. **14**, 198–215 (2018).
[17]C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, "Natural triple excitations in local coupled cluster calculations with pair natural orbitals," J. Chem. Phys. **139**, 134101 (2013).
[18]Y. Guo, U. Becker, and F. Neese, "Comparison and combination of 'direct' and fragment based local correlation methods: Cluster in molecules and domain based local pair natural orbital perturbation and coupled cluster theories," J. Chem. Phys. **148**, 124117 (2018).
[19]P. R. Nagy and M. Kállay, "Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods," J. Chem. Theory Comput. **15**, 5275–5298 (2019).
[20]M. Katouda, A. Naruse, Y. Hirano, and T. Nakajima, "Massively parallel algorithm and implementation of RI-MP2 energy calculation for peta-scale and many-core supercomputers," J. Comput. Chem. **37**, 2623–2633 (2016).
[21]E. Aprà, A. P. Rendell, R. J. Harrison, V. Tipparaju, W. A. de Jong, and S. S. Xantheas, "Liquid water: Obtaining the right answer for the right reasons," in *SC '09: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* (ACM, New York, 2009), p. 66.
[22]S. Yoo, E. Aprà, X. C. Zeng, and S. S. Xantheas, "High-level *ab initio* electronic structure calculations of water clusters $(H_2O)_{16}$ and $(H_2O)_{17}$: A new global minimum for $(H_2O)_{16}$," J. Phys. Chem. Lett. **1**, 3122–3127 (2010).
[23]*The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, edited by D. G. Fedorov and K. Kitaura (CRC Press, Boca Raton, 2009).
[24]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, "Fragmentation methods: A route to accurate calculations on large systems," Chem. Rev. **112**, 632–672 (2012).
[25]M. A. Collins and R. P. Bettens, "Energy-based molecular fragmentation methods," Chem. Rev. **115**, 5607–5642 (2015).
[26]K. Raghavachari and A. Saha, "Accurate composite and fragment-based quantum chemical methods for large molecules," Chem. Rev. **115**, 5643–5677 (2015).
[27]S. Li, W. Li, and J. Ma, "Generalized energy-based fragmentation approach and its applications to macromolecules and molecular aggregates," Acc. Chem. Res. **47**, 2712–2720 (2014).
[28]S. Hirata, K. Gilliard, X. He, J. Li, and O. Sode, "*Ab initio* molecular crystal structures, spectra, and phase diagrams," Acc. Chem. Res. **47**, 2721–2730 (2014).
[29]N. Sahu and S. R. Gadre, "Molecular tailoring approach: A route for *ab initio* treatment of large clusters," Acc. Chem. Res. **47**, 2739–2747 (2014).
[30]X. He, T. Zhu, X. W. Wang, J. F. Liu, and J. Z. H. Zhang, "Fragment quantum mechanical calculation of proteins and its applications," Acc. Chem. Res. **47**, 2748–2757 (2014).
[31]M. A. Collins, M. W. Cvitkovic, and R. P. A. Bettens, "The combined fragmentation and systematic molecular fragmentation methods," Acc. Chem. Res. **47**, 2776–2785 (2014).
[32]S. R. Pruitt, C. Bertonia, K. R. Brorsen, and M. S. Gordon, "Efficient and accurate fragmentation methods," Acc. Chem. Res. **47**, 2786–2794 (2014).
[33]R. M. Richard, K. U. Lao, and J. M. Herbert, "Aiming for benchmark accuracy with the many-body expansion," Acc. Chem. Res. **47**, 2828–2836 (2014).
[34]D. G. Fedorov, N. Asada, I. Nakanishi, and K. Kitaura, "The use of many-body expansions and geometry optimizations in fragment-based methods," Acc. Chem. Res. **47**, 2846–2856 (2014).
[35]R. O. Ramabhadran and K. Raghavachari, "The successful merger of theoretical thermochemistry with fragment-based methods in quantum chemistry," Acc. Chem. Res. **47**, 3596–3604 (2014).
[36]G. J. O. Beran, J. D. Hartman, and Y. N. Heit, "Predicting molecular crystal properties from first principles: Finite-temperature thermochemistry to NMR crystallography," Acc. Chem. Res. **49**, 2501–2508 (2016).
[37]T. Fang, Y. Li, and S. Li, "Generalized energy-based fragmentation approach for modeling condensed phase systems," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **7**, e1297 (2017).
[38]D. G. Fedorov, "The fragment molecular orbital method: Theoretical development, implementation in GAMESS, and applications," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **7**, e1322 (2017).
[39]*Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, edited by M. S. Gordon (John Wiley & Sons, Hoboken, NJ, 2017).
[40]T. Ikegami, T. Ishida, D. G. Fedorov, K. Kitaura, Y. Inadomi, H. Umeda, M. Yokokawa, and S. Sekiguchi, "Full electron calculation beyond 20,000 atoms: Ground electron state of photosynthetic proteins," in *Proceedings of Supercomputing 2005* (IEEE Computer Society, 2005), pp. 1–12.
[41]T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery, K. Morokuma, and M. J. Frisch, "Combining quantum mechanics methods with molecular mechanics methods in ONIOM," J. Chem. Theory Comput. **2**, 815–826 (2006).
[42]T. Vreven and K. Morokuma, "Hybrid methods: ONIOM(QM:MM) and QM/MM," Annu. Rep. Comput. Chem. **2**, 35–51 (2006).
[43]L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. F. Ke, F. Y. Liu, H. B. Li, L. N. Ding, and K. Morokuma, "The ONIOM method and its applications," Chem. Rev. **115**, 5678–5796 (2015).
[44]R. M. Richard and J. M. Herbert, "A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory," J. Chem. Phys. **137**, 064113 (2012).
[45]R. M. Richard and J. M. Herbert, "The many-body expansion with overlapping fragments: Analysis of two approaches," J. Chem. Theory Comput. **9**, 1408–1416 (2013).
[46]L. D. Jacobson, R. M. Richard, K. U. Lao, and J. M. Herbert, "Efficient monomer-based quantum chemistry methods for molecular and ionic clusters," Annu. Rep. Comput. Chem. **9**, 25–56 (2013).
[47]M. N. Ucisik, D. S. Dashti, J. C. Faver, and K. M. Merz, Jr., "Pairwise additivity of energy components in protein-ligand binding: The HIV II protease-indinavir case," J. Chem. Phys. **135**, 085101 (2011).
[48]G. A. Cisneros, K. T. Wikfeldt, L. Ojamäe, J. Lu, Y. Xu, H. Torabifard, A. P. Bartók, G. Csányi, V. Molinero, and F. Paesani, "Modeling molecular interactions in water: From pairwise to many-body potential energy functions," Chem. Rev. **116**, 7501–7528 (2016).
[49]Y. Chen and H. Li, "Intermolecular interaction in water hexamer," J. Phys. Chem. A **114**, 11719–11724 (2010).

[50]R. A. Christie and K. D. Jordan, "*n*-body decomposition approach to the calculation of interaction energies of water clusters," in *Intermolecular Forces and Clusters II*, Structure and Bonding Vol. 116, edited by D. Wales and R. A. Christie (Springer-Verlag, Heidelberg, 2005), pp. 27–41.
[51]J. Cui, H. Liu, and K. D. Jordan, "Theoretical characterization of the $(H_2O)_{21}$ cluster: Application of an *n*-body decomposition procedure," J. Phys. Chem. B **110**, 18872–18878 (2006).
[52]F.-F. Wang, M. J. Deible, and K. D. Jordan, "Benchmark study of the interaction energy for an $(H_2O)_{16}$ cluster: Quantum Monte Carlo and complete basis set limit MP2 results," J. Phys. Chem. A **117**, 7606–7611 (2013).
[53]J. P. Heindel and S. S. Xantheas, "The many-body expansion for water clusters revisited" (submitted).
[54]O. Demerdash and T. Head-Gordon, "Convergence of the many-body expansion for energy and forces for classical polarizable models in the condensed phase," J. Chem. Theory Comput. **12**, 3884–3893 (2016).
[55]M. J. Gillan, D. Alfé, P. J. Bygrave, C. R. Taylor, and F. R. Manby, "Energy benchmarks for water clusters and ice structures from an embedded many-body expansion," J. Chem. Phys. **139**, 114101 (2013).
[56]A. Heßelmann, "Correlation effects and many-body interactions in water clusters," Beilstein J. Org. Chem. **14**, 979–991 (2018).
[57]S. S. Khire and S. R. Gadre, "Pragmatic many-body approach for economic MP2 energy estimation of molecular clusters," J. Phys. Chem. A **123**, 5005–5011 (2019).
[58]M. Alkan, P. Xu, and M. S. Gordon, "Many-body dispersion in molecular clusters," J. Phys. Chem. A **123**, 8406–8416 (2019).
[59]J. F. Ouyang and R. P. A. Bettens, "Many-body basis set superposition effect," J. Chem. Theory Comput. **11**, 5132–5143 (2015).
[60]K.-Y. Liu and J. M. Herbert, "Understanding the many-body expansion for large systems. III. Critical role of four-body terms, counterpoise corrections, and cutoffs," J. Chem. Phys. **147**, 161729 (2017).
[61]R. M. Richard, B. W. Bakr, and C. D. Sherrill, "Understanding the many-body basis set superposition error: Beyond Boys and Bernardi," J. Chem. Theory Comput. **14**, 2386–2400 (2018).
[62]W. Chen and M. S. Gordon, "Energy decomposition analyses for many-body interaction and applications to water complexes," J. Phys. Chem. **100**, 14316–14328 (1996).
[63]C. K. Egan and F. Paesani, "Assessing many-body effects of water self-ions. I: $OH^-(H_2O)_n$ clusters," J. Chem. Theory Comput. **14**, 1982–1997 (2018).
[64]C. K. Egan and F. Paesani, "Assessing many-body effects of water self-ions. II: $H_3O^+(H_2O)_n$ clusters," J. Chem. Theory Comput. **15**, 4816–4833 (2019).
[65]R. Podeszwa, B. M. Rice, and K. Szalewicz, "Predicting structure of molecular crystals from first principles," Phys. Rev. Lett. **101**, 115503 (2008).
[66]S. Wen and G. J. O. Beran, "Accurate molecular crystal lattice energies from a fragment QM/MM approach with on-the-fly *ab initio* force field parametrization," J. Chem. Theory Comput. **7**, 3733–3742 (2011).
[67]M. R. Kennedy, A. R. McDonald, A. E. DePrince III, M. S. Marshall, R. Podeszwa, and C. D. Sherrill, "Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory," J. Chem. Phys. **140**, 121104 (2014).
[68]K. U. Lao, K.-Y. Liu, R. M. Richard, and J. M. Herbert, "Understanding the many-body expansion for large systems. II. Accuracy considerations," J. Chem. Phys. **144**, 164105 (2016).
[69]E. E. Dahlke and D. G. Truhlar, "Electrostatically embedded many-body expansion for large systems, with applications to water clusters," J. Chem. Theory Comput. **3**, 46–53 (2007).
[70]E. E. Dahlke and D. G. Truhlar, "Electrostatically embedded many-body expansions for simulations," J. Chem. Theory Comput. **4**, 1–6 (2008).
[71]E. D. Speetzen, H. R. Leverentz, H. Lin, and D. G. Truhlar, "Electrostatically embedded many-body expansion for large systems," in *Accurate Condensed-Phase Electronic Structure Theory*, edited by F. R. Manby (CRC Press, Boca Rotan, FL, 2011), Chap. 5, pp. 105–127.
[72]R. M. Richard, K. U. Lao, and J. M. Herbert, "Understanding the many-body expansion for large systems. I. Precision considerations," J. Chem. Phys. **141**, 014108 (2014).
[73]G. J. O. Beran, "Approximating quantum many-body intermolecular interactions in molecular clusters using classical polarizable force fields," J. Chem. Phys. **130**, 164115 (2009).
[74]J. Liu, L. Qi, J. Z. H. Zhang, and X. He, "Fragment quantum mechanical method for large-sized ion-water clusters," J. Chem. Theory Comput. **13**, 2021–2034 (2017).
[75]J. Liu and X. He, "Accurate prediction of energetic properties of ionic liquid clusters using a fragment-based quantum mechanical method," Phys. Chem. Chem. Phys. **19**, 20657–20666 (2017).
[76]J. Liu, X. He, and J. Z. H. Zhang, "Structure of liquid water—A dynamical mixture of tetrahedral and 'ring-and-chain' like structures," Phys. Chem. Chem. Phys. **19**, 11931–11936 (2017).
[77]J. Liu, X. He, J. Z. H. Zhang, and L.-W. Qi, "Hydrogen-bond structure dynamics in bulk water: Insights from *ab initio* simulations with coupled cluster theory," Chem. Sci. **9**, 2065–2073 (2018).
[78]J. Liu, H. Sun, W. J. Glover, and X. He, "Prediction of excited-state properties of oligoacene crystals using fragment-based quantum mechanical method," J. Phys. Chem. A **123**, 5407–5417 (2019).
[79]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, "Fragment molecular orbital method: An approximate computational method for large molecules," Chem. Phys. Lett. **313**, 701–706 (1999).
[80]D. G. Fedorov and K. Kitaura, "Extending the power of quantum chemistry to large systems with the fragment molecular orbital method," J. Phys. Chem. A **111**, 6904–6914 (2007).
[81]D. G. Fedorov and K. Kitaura, "Theoretical background of the fragment molecular orbital (FMO) method and its implementation in GAMESS," in *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, edited by D. G. Fedorov and K. Kitaura (CRC Press, Boca Rotan, FL, 2009), Chap. 2, pp. 5–36.
[82]T. Nagata, D. G. Fedorov, and K. Kitaura, "Mathematical formulation of the fragment molecular orbital method," in *Linear-Scaling Techniques in Computational Chemistry and Physics*, Challenges and Advances in Computational Chemistry and Physics Vol. 13, edited by R. Zaleśny, M. G. Papadopoulos, P. G. Mezey, and J. Leszczynski (Springer, New York, 2011), Chap. 2, pp. 17–64.
[83]D. G. Fedorov, T. Nagata, and K. Kitaura, "Exploring chemistry with the fragment molecular orbital method," Phys. Chem. Chem. Phys. **14**, 7562–7577 (2012).
[84]H. Stoll and H. Preuß, "On the direct calculation of localized HF orbitals in molecule clusters, layers and solids," Theor. Chem. Acc. **46**, 11–21 (1977).
[85]J. C. White and E. R. Davidson, "An analysis of the hydrogen bond in ice," J. Chem. Phys. **93**, 8029–8035 (1990).
[86]H. Stoll, "Correlation energy of diamond," Phys. Rev. B **46**, 6700–6704 (1992).
[87]H. Stoll, "The correlation energy of crystalline silicon," Chem. Phys. Lett. **191**, 548–552 (1992).
[88]H. Stoll, "On the correlation energy of graphite," J. Chem. Phys. **97**, 8449–8454 (1992).
[89]T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, and K. Kitaura, "Fragment molecular orbital method: Use of approximate electrostatic potential," Chem. Phys. Lett. **351**, 475–480 (2002).
[90]D. G. Fedorov and K. Kitaura, "The importance of three-body terms in the fragment molecular orbital method," J. Chem. Phys. **120**, 6832–6840 (2004).
[91]T. Nakano, Y. Mochizuki, K. Yamashita, C. Watanabe, K. Fukuzawa, K. Segawa, Y. Okiyama, T. Tsukamoto, and S. Tanaka, "Development of the four-body corrected fragment molecular orbital (FMO4) method," Chem. Phys. Lett. **523**, 128–133 (2012).
[92]D. G. Fedorov, L. V. Slipchenko, and K. Kitaura, "Systematic study of the embedding potential description in the fragment molecular orbital method," J. Phys. Chem. A **114**, 8742–8753 (2010).
[93]D. G. Fedorov and K. Kitaura, "Use of an auxiliary basis set to describe the polarization in the fragment molecular orbital method," Chem. Phys. Lett. **597**, 99–105 (2014).
[94]H. Lin and D. G. Truhlar, "QM/MM: What have we learned, where are we, and where do we go from here?," Theor. Chem. Acc. **117**, 185–199 (2007).
[95]H. M. Senn and W. Thiel, "QM/MM methods for biological systems," Top. Curr. Chem. **268**, 173–290 (2007).

[96]H. M. Senn and W. Thiel, "QM/MM methods for biomolecular systems," Angew. Chem., Int. Ed. Engl. **48**, 1198–1229 (2009).
[97]E. Brunk and U. Rothlisberger, "Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states," Chem. Rev. **115**, 6217–6263 (2015).
[98]Y. Zho, S. Wang, Y. Li, and Y. Zhang, "Born–Oppenheimer *ab initio* QM/MM molecular dynamics simulations of enzyme reactions," in *Computational Approaches for Studying Enzyme Mechanism Part A*, Methods in Enzymology Vol. 577, edited by G. A. Voth (Elsevier, 2016), Chap. 5, pp. 105–118.
[99]T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, and K. Kitaura, "Fragment molecular orbital method: Application to polypeptides," Chem. Phys. Lett. **318**, 614–618 (2000).
[100]D. G. Fedorov, J. H. Jensen, R. C. Deka, and K. Kitaura, "Covalent bond fragmentation suitable to describe solids in the fragment molecular orbital method," J. Phys. Chem. A **112**, 11808–11816 (2008).
[101]V. Deev and M. A. Collins, "Approximate *ab initio* energies by systematic molecular fragmentation," J. Chem. Phys. **122**, 154102 (2005).
[102]M. A. Collins and V. A. Deev, "Accuracy and efficiency of electronic energies from systematic molecular fragmentation," J. Chem. Phys. **125**, 104104 (2006).
[103]M. A. Addicoat and M. A. Collins, "Accurate treatment of nonbonded interactions within systematic molecular fragmentation," J. Chem. Phys. **131**, 104103 (2009).
[104]M. A. Collins, "Systematic fragmentation of large molecules by annihilation," Phys. Chem. Chem. Phys. **14**, 7744–7751 (2012).
[105]M. A. Collins, "Molecular forces, geometries, and frequencies by systematic molecular fragmentation including embedding charges," J. Chem. Phys. **141**, 094108 (2014).
[106]D. W. Zhang and J. Z. H. Zhang, "Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy," J. Chem. Phys. **119**, 3599–3605 (2003).
[107]X. H. Chen and J. Z. H. Zhang, "Molecular fractionation with conjugate caps density matrix with pairwise interaction correction for protein energy calculation," J. Chem. Phys. **125**, 044903 (2006).
[108]X. Wang, J. Liu, J. Z. H. Zhang, and X. He, "Electrostatically embedded generalized molecular fractionation with conjugate caps method for full quantum mechanical calculation of protein energy," J. Phys. Chem. A **117**, 7149–7161 (2013).
[109]J. Liu, T. Zhu, X. He, and J. Z. H. Zhang, "MFCC-based fragmentation methods for biomolecules," in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, edited by M. S. Gordon (Wiley, 2017), Chap. 11, pp. 323–348.
[110]N. J. Mayhall and K. Raghavachari, "Molecules-in-molecules: An extrapolated fragment-based approach for accurate calculations on large molecules and materials," J. Chem. Theory Comput. **7**, 1336–1343 (2011).
[111]S. R. Gadre and V. Ganesh, "Molecular tailoring approach: Towards PC-based *ab initio* treatment of large molecules," J. Theor. Comput. Chem. **5**, 835–855 (2006).
[112]V. Ganesh, R. K. Dongare, P. Balanarayan, and S. R. Gadre, "Molecular tailoring approach for geometry optimization of large molecules: Energy evaluation and parallelization strategies," J. Chem. Phys. **125**, 104109 (2006).
[113]A. P. Rahalkar, V. Ganesh, and S. R. Gadre, "Enabling *ab initio* Hessian and frequency calculations of large molecules," J. Chem. Phys. **129**, 234101 (2008).
[114]S. R. Gadre, K. V. J. Jose, and A. P. Rahalkar, "Molecular tailoring approach for exploring structures, energetics and properties of clusters," J. Chem. Sci. **122**, 47–56 (2010).
[115]A. P. Rahalkar, S. D. Yeole, V. Ganesh, and S. R. Gadre, "Molecular tailoring: An art of the possible for *ab initio* treatment of large molecules and molecular clusters," in *Linear-Scaling Techniques in Computational Chemistry and Physics*, Challenges and Advances in Computational Chemistry and Physics Vol. 13, edited by R. Zaleśny, M. G. Papadopoulos, P. G. Mezey, and J. Leszczynski (Springer, New York, 2011), Chap. 10, pp. 199–226.
[116]A. P. Rahalkar and S. R. Gadre, "Tailoring approach for obtaining molecular orbitals of large systems," J. Chem. Sci. **124**, 149–158 (2012).
[117]N. Sahu, S. D. Yeole, and S. R. Gadre, "Appraisal of molecular tailoring approach for large clusters," J. Chem. Phys. **138**, 104101 (2013).
[118]S. S. Khire, L. J. Bartolotti, and S. R. Gadre, "Harmonizing accuracy and efficiency: A pragmatic approach to fragmentation of large molecules," J. Chem. Phys. **149**, 064112 (2018).
[119]W. Li, S. Li, and Y. Jiang, "Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules," J. Phys. Chem. A **111**, 2193–2199 (2007).
[120]S. Hua, W. Hua, and S. Li, "An efficient implementation of the generalized energy-based fragmentation approach for general large molecules," J. Phys. Chem. A **114**, 8126–8134 (2010).
[121]W. Li, W. Hua, T. Fang, and S. Li, "The energy-based fragmentation approach for *ab initio* calculations of large systems," in *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*, edited by J. R. Reimers (Wiley, Hoboken, NJ, 2011), Chap. 7, pp. 227–258.
[122]A. Saha and K. Raghavachari, "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. **11**, 2012–2023 (2015).
[123]K. V. J. Jose and K. Raghavachari, "Evaluation of energy gradients and infrared vibrational spectra through molecules-in-molecules fragment-based approach," J. Chem. Theory Comput. **11**, 950–961 (2015).
[124]K. V. J. Jose and K. Raghavachari, "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*, edited by M. S. Gordon (Wiley, 2017), Chap. 4, pp. 141–164.
[125]D. Yuan, X. Shen, W. Li, and S. Li, "Are fragment-based quantum chemistry methods applicable to medium-sized water clusters?," Phys. Chem. Chem. Phys. **18**, 16491–16498 (2016).
[126]L. Zhang, W. Li, T. Fang, and S. Li, "Accurate relative energies and binding energies of large ice–liquid water clusters and periodic structures," J. Phys. Chem. A **121**, 4030–4038 (2017).
[127]D. Yuan, Y. Li, Z. Ni, P. Pulay, W. Li, and S. Li, "Benchmark relative energies for large water clusters with the generalized energy-based fragmentation method," J. Chem. Theory Comput. **13**, 2696–2704 (2017).
[128]J. Liu and J. M. Herbert, "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. **12**, 572–584 (2016).
[129]G. J. O. Beran and K. Nanda, "Predicting organic crystal lattice energies with chemical accuracy," J. Phys. Chem. Lett. **1**, 3480–3487 (2010).
[130]K. D. Nanda and G. J. O. Beran, "Prediction of organic molecular crystal geometries from MP2-level fragment quantum mechanical/molecular mechanical calculations," J. Chem. Phys. **137**, 174106 (2012).
[131]S. Wen, K. Nanda, Y. Huang, and G. J. O. Beran, "Practical quantum mechanics-based fragment methods for predicting molecular crystal properties," Phys. Chem. Chem. Phys. **14**, 7578–7590 (2012).
[132]J. D. Hartman and G. J. O. Beran, "Fragment-based electronic structure approach for computing nuclear magnetic resonance chemical shifts in molecular crystals," J. Chem. Theory Comput. **10**, 4862–4872 (2014).
[133]G. S. Tschumper, "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. **427**, 185–191 (2006).
[134]A. M. Elsohly, C. L. Shaw, M. E. Guice, B. D. Smith, and G. S. Tschumper, "Analytic gradients for the multicentered QM:QM method for weakly bound clusters: Efficient and accurate 2-body:many-body geometry optimizations," Mol. Phys. **105**, 2777–2782 (2007).
[135]D. M. Bates, J. R. Smith, T. Janowski, and G. S. Tschumper, "Development of a 3-body:many-body integrated fragmentation method for weakly bound clusters and application to water clusters $(H_2O)_{n=3-10,16,17}$," J. Chem. Phys. **135**, 044123 (2011).
[136]D. M. Bates, J. R. Smith, and G. S. Tschumper, "Efficient and accurate methods for the geometry optimization of water clusters: Application of analytic

gradients for the two-body:many-body QM:QM fragmentation method to $(H_2O)_n$, n = 3–10," J. Chem. Theory Comput. **7**, 2753–2760 (2011).
[137]J. C. Howard and G. S. Tschumper, "*N*-body:Many-body QM:QM vibrational frequencies: Application to small hydrogen-bonded clusters," J. Chem. Phys. **139**, 184113 (2013).
[138]T. M. Sexton and G. S. Tschumper, "2-body:many-body QM:QM study of structures, energetics, and vibrational frequencies for microhydrated halide ions," Mol. Phys. **117**, 1413–1420 (2019).
[139]B. W. Hopkins and G. S. Tschumper, "A multicentered approach to integrated QM/QM calculations. Applications to multiply hydrogen bonded systems," J. Comput. Chem. **24**, 1563–1568 (2003).
[140]B. W. Hopkins and G. S. Tschumper, "Multicentered QM/QM methods for overlapping model systems," Mol. Phys. **103**, 309–315 (2005).
[141]B. W. Hopkins and G. S. Tschumper, "Integrated quantum mechanical approaches for extended π systems: Multicentered QM/QM studies of the cyanogen and diacetylene trimers," Chem. Phys. Lett. **407**, 362–367 (2005).
[142]S. Debnath, A. Sengupta, K. V. J. Jose, and K. Raghavachari, "Fragment-based approaches for supramolecular interaction energies: Applications to foldamers and their complexes with anions," J. Chem. Theory Comput. **14**, 6226–6239 (2018).
[143]B. Thapa, D. Beckett, J. Erickson, and K. Raghavachari, "Theoretical study of protein–ligand interactions using the molecules-in-molecules fragmentation-based method," J. Chem. Theory Comput. **14**, 5143–5155 (2018).
[144]B. Thapa, D. Beckett, K. V. J. Jose, and K. Raghavachari, "Assessment of fragmentation strategies for large proteins using the multilayer molecules-in-molecules approach," J. Chem. Theory Comput. **14**, 1383–1394 (2018).
[145]D. G. Fedorov, T. Ishida, and K. Kitaura, "Multilayer formulation of the fragment molecular orbital method (FMO)," J. Phys. Chem. A **109**, 2638–2646 (2005).
[146]A. P. Rahalkar, M. Katouda, S. R. Gadre, and S. Nagase, "Molecular tailoring approach in conjunction with MP2 and RI-MP2 codes: A comparison with fragment molecular orbital method," J. Comput. Chem. **31**, 2405–2418 (2010).
[147]J. P. Furtado, A. P. Rahalkar, S. Shanker, P. Bandyopadhyay, and S. R. Gadre, "Facilitating minima search for large water clusters at the MP2 level via molecular tailoring," J. Phys. Chem. Lett. **3**, 2253–2258 (2012).
[148]N. Sahu and S. R. Gadre, "Accurate vibrational spectra via molecular tailoring approach: A case study of water clusters at MP2 level," J. Chem. Phys. **142**, 014107 (2015).
[149]S. S. Khire, N. Sahu, and S. R. Gadre, "Harnessing desktop computers for *ab initio* calculation of vibrational IR/Raman spectra of large molecules," J. Chem. Sci. **130**, 159 (2018).
[150]N. Sahu, G. Singh, A. Nandi, and S. R. Gadre, "Toward an accurate and inexpensive estimation of CCSD(T)/CBS binding energies of large water clusters," J. Phys. Chem. A **120**, 5706–5714 (2016).
[151]E. E. Dahlke and D. G. Truhlar, "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. **3**, 1342–1348 (2007).
[152]D. Usvyat, L. Maschio, and M. Schütz, "Periodic and fragment models based on the local correlation approach," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **8**, e1357 (2018).
[153]H. Stoll, "Toward a wavefunction-based treatment of strong electron correlation in extended systems by means of incremental methods," J. Chem. Phys. **151**, 044104 (2019).
[154]L. Bytautas and K. Ruedenberg, "Correlation energy extrapolation by intrinsic scaling. I. Method and application to neon atom," J. Chem. Phys. **121**, 10905 (2004).
[155]J. Friedrich and M. Dolg, "Fully automated incremental evaluation of MP2 and CCSD(T) energies: Application to water clusters," J. Chem. Theory Comput. **5**, 287–294 (2009).
[156]L. Bytautas and K. Ruedenberg, "The range of electron correlation between localized molecular orbitals. A full configuration interaction analysis for the NCCN molecule," J. Phys. Chem. A **114**, 8601–8612 (2010).
[157]J. Friedrich and K. Walczak, "Incremental CCSD(T)(F12)/MP2-F12—A method to obtain highly accurate CCSD(T) energies for large molecules," J. Chem. Theory Comput. **9**, 408–417 (2013).
[158]J. S. Boschen, D. Theis, K. Ruedenberg, and T. L. Windus, "Correlation energy extrapolation by many-body expansion," J. Phys. Chem. A **121**, 836–844 (2017).
[159]J. J. Eriksen, F. Lipparini, and J. Gauss, "Virtual orbital many-body expansions: A possible route towards the full configuration interaction limit," J. Phys. Chem. Lett. **8**, 4633–4639 (2017).
[160]P. M. Zimmerman, "Incremental full configuration interaction," J. Chem. Phys. **146**, 104102 (2017).
[161]P. M. Zimmerman, "Strong correlation in incremental full configuration interaction," J. Chem. Phys. **146**, 224104 (2017).
[162]P. M. Zimmerman and A. E. Rask, "Evaluation of full valence correlation energies and gradients," J. Chem. Phys. **150**, 244117 (2019).
[163]J. J. Eriksen and J. Gauss, "Many-body expanded full configuration interaction: I. Weakly correlated regime," J. Chem. Theory Comput. **14**, 5180–5191 (2018).
[164]J. J. Eriksen and J. Gauss, "Many-body expanded full configuration interaction. II. Strongly correlated regime," J. Chem. Theory Comput. **15**, 4873–4884 (2019).
[165]S. Li, J. Shen, W. Li, and Y. Jiang, "An efficient implementation of the 'cluster-in-molecule' approach for local electron correlation calculations," J. Chem. Phys. **125**, 074109 (2006).
[166]Y. Guo, W. Li, and S. Li, "Improved cluster-in-molecule local correlation approach for electron correlation calculation of large systems," J. Phys. Chem. A **118**, 8996–9004 (2014).
[167]A. D. Findlater, F. Zahariev, and M. S. Gordon, "Combined fragment molecular orbital cluster in molecule approach to massively parallel electron correlation calculations for large systems," J. Phys. Chem. A **119**, 3587–3593 (2015).
[168]Z. Ni, W. Li, and S. Li, "Fully optimized implementation of the cluster-in-molecule local correlation approach for electron correlation calculations of large systems," J. Comput. Chem. **40**, 1130–1140 (2018).
[169]Y. Wang, Z. Ni, W. Li, and S. Li, "Cluster-in-molecule local correlation approach for periodic systems," J. Chem. Theory Comput. **15**, 2933–2943 (2019).
[170]S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki, and H. Sekino, "Fast electron correlation methods for molecular clusters in the ground and excited states," Mol. Phys. **130**, 2255–2265 (2005).
[171]M. Chiba, D. G. Fedorov, and K. Kitaura, "Time-dependent density functional theory with the multilayer fragment molecular orbital method," Chem. Phys. Lett. **444**, 346–350 (2007).
[172]M. Chiba, D. G. Fedorov, and K. Kitaura, "Time-dependent density functional theory based upon the fragment molecular orbital method," J. Chem. Phys. **127**, 104108 (2007).
[173]M. Chiba, D. G. Fedorov, T. Nagata, and K. Kitaura, "Excited state geometry optimizations by time-dependent density functional theory based on the fragment molecular orbital method," Chem. Phys. Lett. **474**, 227–232 (2009).
[174]M. Chiba, D. G. Fedorov, and K. Kitaura, "The fragment molecular orbital–based time-dependent density functional theory for excited states in large systems," in *The Fragment Molecular Orbital Method*, edited by D. Fedorov and K. Kitaura (Taylor & Francis, 2009), Chap. 5, pp. 91–118.
[175]Y. Ma, Y. Liu, and H. Ma, "A new fragment-based approach for calculating electronic excitation energies of large systems," J. Chem. Phys. **136**, 024113 (2012).
[176]Y. Ma and H. Ma, "Calculating excited states of molecular aggregates by the renormalized excitonic method," J. Phys. Chem. A **117**, 3655–3665 (2013).
[177]K. D. Closser, Q. Ge, Y. Mao, Y. Shao, Q. Ge, and M. Head-Gordon, "Superposition of fragment excitations for excited states of large clusters with application to helium clusters," J. Chem. Theory Comput. **11**, 5791–5803 (2015).
[178]J. Liu and J. M. Herbert, "An efficient and accurate approximation to time-dependent density functional theory for systems of weakly coupled monomers," J. Chem. Phys. **143**, 034106 (2015).
[179]J. Liu and J. M. Herbert, "Local excitation approximations to time-dependent density functional theory for excitation energies in solution," J. Chem. Theory Comput. **12**, 157–166 (2016).

[180]J. M. Herbert, X. Zhang, A. F. Morrison, and J. Liu, "Beyond time-dependent density functional theory using only single excitations: Methods for computational studies of excited states in complex systems," Acc. Chem. Res. **49**, 931–941 (2016).
[181]A. F. Morrison, Z.-Q. You, and J. M. Herbert, "*Ab initio* implementation of the Frenkel-Davydov exciton model: A naturally parallelizable approach to computing collective excitations in crystals and aggregates," J. Chem. Theory Comput. **10**, 5366–5376 (2014).
[182]A. F. Morrison and J. M. Herbert, "Low-scaling quantum chemistry approach to excited-state properties via an *ab initio* exciton model: Application to excitation energy transfer in a self-assembled nanotube," J. Phys. Chem. Lett. **6**, 4390–4396 (2015).
[183]A. F. Morrison and J. M. Herbert, "Analytic derivative couplings and first-principles exciton/phonon coupling constants for an *ab initio* Frenkel-Davydov exciton model: Theory, implementation, and application to compute triplet exciton mobility parameters for crystalline tetracene," J. Chem. Phys. **146**, 224110 (2017).
[184]R. M. Richard, K. U. Lao, and J. M. Herbert, "Approaching the complete-basis limit with a truncated many-body expansion," J. Chem. Phys. **139**, 224102 (2013).
[185]F. Paesani, "Getting the right answers for the right reasons: Toward predictive molecular simulations of water with many-body potential energy functions," Acc. Chem. Res. **49**, 1844–1851 (2016).
[186]S. K. Reddy, S. C. Straight, P. Bajaj, C. H. Pham, M. Riera, D. R. Moberg, M. A. Morales, C. Knight, A. W. Götz, and F. Paesani, "On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice," J. Chem. Phys. **145**, 194504 (2016).
[187]J. P. Heindel, Q. Yu, J. M. Bowman, and S. S. Xantheas, "Benchmark electronic structure calculations for $H_3O^+(H_2O)_n$, n = 0–5, clusters and tests of an existing 1,2,3-body potential energy surface with a new 4-body correction," J. Chem. Theory Comput. **14**, 4553–4566 (2018).
[188]S. R. Pruitt, M. A. Addicoat, M. A. Collins, and M. S. Gordon, "The fragment molecular orbital and systematic fragmentation methods applied to water clusters," Phys. Chem. Chem. Phys. **14**, 7752–7764 (2012).
[189]S. R. Pruitt, C. Steinmann, J. H. Jensen, and M. S. Gordon, "Fully integrated effective fragment molecular orbital method," J. Chem. Theory Comput. **9**, 2235–2249 (2013).
[190]In response to a query titled "Why is FMO in GAMESS not fast enough in comparison to no fragmentation with some other programs?," the following statement appears on the web page for the FMO code in GAMESS: "In addition to the obvious difference in the computational algorithms, the default accuracy (integral, SCF convergence etc) are set differently. To get a fair comparison, one should match all necessary thresholds, which can have a large impact upon timings. GAMESS has a high accuracy standard, and FMO raises that even higher, to get reliable and reproducible results for the huge total energies which large molecules have." This suggests that the authors of the FMO code have encountered some of the same precision problems that were documented by Herbert and co-workers,[33,68,72] and further suggests that they have seen some of the timing issues documented in Ref. 68 and elsewhere. (https://staff.aist.go.jp/d.g.fedorov/fmo/fmoqa.html; accessed October 5, 2019.)
[191]R. M. Richard, K. U. Lao, and J. M. Herbert, "Achieving the CCSD(T) basis-set limit in sizable molecular clusters: Counterpoise corrections for the many-body expansion," J. Phys. Chem. Lett. **4**, 2674–2680 (2013).
[192]L. D. Jacobson and J. M. Herbert, "An efficient, fragment-based electronic structure method for molecular systems: Self-consistent polarization with perturbative two-body exchange and dispersion," J. Chem. Phys. **134**, 094118 (2011).
[193]Z. C. Holden, R. M. Richard, and J. M. Herbert, "Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets," J. Chem. Phys. **139**, 244108 (2013); Erratum, **142**, 059901 (2015).
[194]J. M. Herbert, L. D. Jacobson, K. U. Lao, and M. A. Rohrdanz, "Rapid computation of intermolecular interactions in molecular and ionic clusters: Self-consistent polarization plus symmetry-adapted perturbation theory," Phys. Chem. Chem. Phys. **14**, 7679–7699 (2012).
[195]Z. C. Holden, B. Rana, and J. M. Herbert, "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. **150**, 144115 (2019).
[196]A. E. Reed, R. B. Weinstock, and F. Weinhold, "Natural population analysis," J. Chem. Phys. **83**, 735–746 (1985).
[197]A. E. Shields and T. van Mourik, "Comparison of *ab initio* and DFT electronic structure methods for peptides containing an aromatic ring: Effect of dispersion and BSSE," J. Phys. Chem. A **111**, 13272–13277 (2007).
[198]L. F. Holroyd and T. van Mourik, "Insufficient description of dispersion in B3LYP and large basis set superposition errors in MP2 calculations can hide peptide conformers," Chem. Phys. Lett. **442**, 42–46 (2007).
[199]T. van Mourik, "Determining potential energy surfaces for flexible peptides. Problems caused by intramolecular BSSE and dispersion," in *Molecular Potential Energy Surfaces in Many Dimensions*, Collaborative Computational Project on Molecular Quantum Dynamics (CCP6), edited by M. M. Law and A. Ernesti (Daresbury Laboratory, Daresbury, Warrington, UK, 2009), pp. 10–18.
[200]D. Toroz and T. van Mourik, "Structure of the gas-phase glycine tripeptide," Phys. Chem. Chem. Phys. **12**, 3463–3473 (2010).
[201]C. D. Sherrill, "Computations of noncovalent π interactions," in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and T. R. Cundari (Wiley-VCH, 2009), Vol. 26, Chap. 1, pp. 1–38.
[202]G. S. Tschumper, "Reliable electronic structure computations for weak noncovalent interactions in clusters," in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and T. R. Cundari (Wiley-VCH, 2009), Vol. 26, Chap. 2, pp. 39–90.
[203]E. G. Hohenstein and C. D. Sherrill, "Wavefunction methods for noncovalent interactions," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 304–326 (2012).
[204]M. Shahbaz and K. Szalewicz, "Do semilocal density-functional approximations recover dispersion energies at small intermonomer separations?," Phys. Rev. Lett. **121**, 113402 (2018).
[205]J. F. Ouyang, M. W. Cvitkovic, and R. P. A. Bettens, "Trouble with the many-body expansion," J. Chem. Theory Comput. **10**, 3699–3707 (2014).
[206]S. F. Boys and F. Bernardi, "The calculation of small molecular interactions by the differences of separated total energies. Some procedures with reduced errors," Mol. Phys. **19**, 553–566 (1970).
[207]F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, "State of the art in counterpoise theory," Chem. Rev. **94**, 1873–1885 (1994).
[208]M. Kamiya, S. Hirata, and M. Valiev, "Fast electron correlation methods for molecular clusters without basis set superposition errors," J. Chem. Phys. **128**, 074103 (2008).
[209]I. Mayer and I. Bakó, "Many-body energy decomposition with basis set superposition error corrections," J. Chem. Theory Comput. **13**, 1883–1886 (2017).
[210]B. H. Wells and S. Wilson, "Van der Waals interaction potentials: Many-body basis set superposition effects," Chem. Phys. Lett. **101**, 429–434 (1983).
[211]K. Mierzwicki and Z. Latajka, "Basis set superposition error in N-body clusters," Chem. Phys. Lett. **380**, 654–664 (2003).
[212]P. Salvador and M. M. Szczęśniak, "Counterpoise-corrected geometries and harmonic frequencies of N-body clusters: Application to $(HF)_n$ (n = 3, 4)," J. Chem. Phys. **118**, 537–549 (2003).
[213]K. Walczak, J. Friedrich, and M. Dolg, "On basis set superposition error corrected stabilization energies for large n-body clusters," J. Chem. Phys. **135**, 134118 (2011).
[214]I. Bakó, I. Mayer, A. Hamza, and L. Pusztai, "Two- and three-body, and relaxation energy terms in water clusters: Application of the hierarchical BSSE corrected decomposition scheme," J. Mol. Liq. **285**, 171–177 (2019).
[215]D. W. Schwenke and D. G. Truhlar, "Systematic study of basis set superposition errors in the calculated interaction energy of two HF molecules," J. Chem. Phys. **82**, 2418–2426 (1985); Erratum, **84**, 4113 (1986); Erratum, **86**, 3760 (1987).
[216]D. M. Bates and G. S. Tschumper, "CCSD(T) complete basis set limit relative energies for low-lying water hexamer structures," J. Phys. Chem. A **113**, 3555–3559 (2009).

[217]T. Ishikawa, T. Ishikura, and K. Kuwata, "Theoretical study of the prion protein based on the fragment molecular orbital method," J. Comput. Chem. **30**, 2594–2601 (2009).
[218]J. Gauss, "Molecular properties," in *Modern Methods and Algorithms of Quantum Chemistry*, NIC Series Vol. 3, 2nd ed., edited by J. Grotendorst (John von Neumann Institute for Computing, Jülich, 2000), pp. 541–592.
[219]T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, and K. Ruud, "Recent advances in wave function-based methods of molecular-property calculations," Chem. Rev. **112**, 543–631 (2012).
[220]P. Norman, K. Ruud, and T. Saue, *Principles and Practices of Molecular Properties: Theory, Modeling and Simulations* (Wiley, 2018).
[221]H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch, and T. Vreven, "QM:QM electronic embedding using Mulliken atomic charges: Energies and analytic gradients in an ONIOM framework," J. Chem. Phys. **128**, 034107 (2008).
[222]J. Liu, B. Rana, K.-Y. Liu, and J. M. Herbert, "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. **10**, 3877–3886 (2019).
[223]N. Sahu, S. S. Khire, and S. R. Gadre, "Structures, energetics and vibrational spectra of $(H_2O)_{32}$ clusters: A journey from model potentials to correlated theory," Mol. Phys. **113**, 2970–2979 (2015).
[224]N. Sahu and S. R. Gadre, "Vibrational infrared and Raman spectra of polypeptides: Fragments-in-fragments within molecular tailoring approach," J. Chem. Phys. **144**, 114113 (2016).
[225]A. P. Rahalkar, S. D. Yeole, and S. R. Gadre, "Acetylene aggregates via cluster-building algorithm and molecular tailoring approach," Theor. Chem. Acc. **131**, 1095 (2012).
[226]T. J. Mach and T. D. Crawford, "Computing optical rotation via an *N*-body approach," Theor. Chem. Acc. **133**, 1449 (2014).
[227]B. G. Peyton and T. D. Crawford, "Basis set superposition errors in the many-body expansion of molecular properties," J. Phys. Chem. A **123**, 4500–4511 (2019).
[228]T. D. Crawford, M. C. Tam, and M. L. Abrams, "The current state of *ab initio* calculations of optical rotation and electronic circular dichroism spectra," J. Phys. Chem. A **111**, 12057–12068 (2007).
[229]P. Mukhopadhyay, G. Zuber, M.-R. Goldsmith, P. Wipf, and D. N. Beratan, "Solvent effect on optical rotation: A case study of methyloxirane in water," ChemPhysChem **7**, 2483–2486 (2006).
[230]P. Mukhopadhyay, G. Zuber, P. Wipf, and D. N. Beratan, "Contributions of a solute's chiral solvent imprint to optical rotation," Angew. Chem., Int. Ed. Engl. **46**, 6450–6452 (2007).
[231]J. Kongsted, T. B. Pedersen, M. Strange, A. Osted, A. E. Hansen, K. V. Mikkelsen, F. Pawlowski, P. Jørgensen, and C. Hättig, "Coupled cluster calculations of the optical rotation of *S*-propylene oxide in gas phase and solution," Chem. Phys. Lett. **401**, 385–392 (2005).
[232]A. Hermann and P. Schwerdtfeger, "Complete basis set limit second-order Møller–Plesset calculations for the fcc lattices of neon, argon, krypton, and xenon," J. Chem. Phys. **131**, 244508 (2009).
[233]C. Müller, D. Usvyat, and H. Stoll, "Local correlation methods for solids: Comparison of incremental and periodic correlation calculations for the argon fcc crystal," Phys. Rev. B **83**, 245136 (2011).
[234]P. J. Bygrave, N. L. Allan, and F. R. Manby, "The embedded many-body expansion for energetics of molecular crystals," J. Chem. Phys. **137**, 164102 (2012).
[235]C. Müller and D. Usvyat, "Incrementally corrected periodic local MP2 calculations: I. The cohesive energy of molecular crystals," J. Chem. Theory Comput. **9**, 5590–5598 (2013).
[236]J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K.-L. Chan, "*Ab initio* determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy," Science **345**, 640–643 (2014).
[237]T. Fang, W. Li, F. Gu, and S. Li, "Accurate prediction of lattice energies and structures of molecular crystals with molecular quantum chemistry methods," J. Chem. Theory Comput. **11**, 91–98 (2015).
[238]C. Červinka, M. Fulem, and K. Růžička, "CCSD(T)/CBS fragment-based calculations of lattice energy of molecular crystals," J. Chem. Phys. **144**, 064505 (2016).
[239]C. Červinka and G. J. O. Beran, "*Ab initio* thermodynamic properties and their uncertainties for crystalline α-methanol," Phys. Chem. Chem. Phys. **19**, 29940–29953 (2017).
[240]C. Cervinkă and G. J. O. Beran, "*Ab initio* prediction of the polymorph phase diagram for crystalline methanol," Chem. Sci. **9**, 4622–4629 (2018).
[241]C. Červinka and G. J. O. Beran, "Towards reliable *ab initio* sublimation pressures for organic molecular crystals—Are we there yet?," Phys. Chem. Chem. Phys. **21**, 14799–14810 (2019).
[242]S. Hirata, "Fast electron-correlation methods for molecular crystals: An application to the α, β_1, and β_2 modifications of solid formic acid," J. Chem. Phys. **129**, 204104 (2008).
[243]S. Hirata, K. Gilliard, X. He, M. Keçeli, J. Li, M. A. Salim, O. Sode, and K. Yagi, "*Ab initio* ice, dry ice, and liquid water," in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, edited by M. S. Gordon (Wiley, 2017), Chap. 9, pp. 245–296.
[244]P. Zhang, D. G. Truhlar, and J. Gao, "Fragment-based quantum mechanical methods for periodic systems with Ewald summation and mean image charge convention for long-range electrostatic interactions," Phys. Chem. Chem. Phys. **14**, 7821–7829 (2012).
[245]P. B. Allen, "Anharmonic phonon quasiparticle theory of zero-point and thermal shifts in insulators: Heat capacity, bulk modulus, and thermal expansion," Phys. Rev. B **92**, 064106 (2015).
[246]Y. N. Heit and G. J. O. Beran, "How important is thermal expansion for predicting molecular crystal structures and thermochemistry at finite temperatures?," Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. **72**, 514–529 (2016).
[247]G. J. O. Beran, Y. N. Heit, and J. D. Hartman, "Noncovalent interactions in molecular crystals," in *Non-Covalent Interactions in Quantum Chemistry and Physics*, edited by A. O. de la Roza and G. A. DiLabio (Elsevier, 2017), Chap. 10, pp. 303–331.
[248]J. L. McKinley and G. J. O. Beran, "Identifying pragmatic quasi-harmonic electronic structure approaches for modeling molecular crystal thermal expansion," Faraday Discuss. **211**, 181–207 (2018).
[249]K. Röttger, A. Endriss, J. Ihringer, S. Doyle, and W. F. Kuhs, "Lattice constants and thermal expansion of H_2O and D_2O ice I*h* between 10 and 265 K," Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. **50**, 644–648 (1994); Erratum, **68**, 91 (2012).
[250]X. He, O. Sode, S. S. Xantheas, and S. Hirata, "Second-order many-body perturbation study of ice Ih," J. Chem. Phys. **137**, 204505 (2012).
[251]K. Gilliard, O. Sode, and S. Hirata, "Second-order many-body perturbation and coupled-cluster singles and doubles study of ice VIII," J. Chem. Phys. **140**, 174507 (2014).
[252]M. A. Salim, S. Y. Willow, and S. Hirata, "Ice Ih anomalies: Thermal contraction, anomalous volume isotope effect, and pressure-induced amorphization," J. Chem. Phys. **144**, 204503 (2016).
[253]J. C. Li and D. K. Ross, "Evidence for two kinds of hydrogen bond in ice," Nature **365**, 327–329 (1993).
[254]H. Tanaka, "Hydrogen bonds between water molecules: Thermal expansivity of ice and water," J. Mol. Liq. **90**, 323–332 (2001).
[255]J. C. Li, S. M. Bennington, and D. K. Ross, "Further evidence for the existence of two kinds of hydrogen bonds in ice Ih," Phys. Lett. A **192**, 295–300 (1994).
[256]J. S. Tse and D. D. Klug, "Comments on "Further evidence for the existence of two kinds of H-bonds in ice Ih" by Li *et al*," Phys. Lett. A **198**, 464–466 (1995).
[257]S. Klotz, T. Strässle, C. G. Salzmann, J. Philippe, and S. F. Parker, "Incoherent inelastic neutron scattering measurements on ice VII: Are there two kinds of hydrogen bonds in ice?," Europhys. Lett. **72**, 576–582 (2005).
[258]P. Zhang, L. Tian, Z. P. Zhang, G. Shao, and J. C. Li, "Investigation of the hydrogen bonding in ice Ih by first-principles density function methods," J. Chem. Phys. **137**, 044504 (2012).
[259]V. Iota, C.-S. Yoo, J.-H. Klepeis, Z. Jenei, W. Evans, and H. Cynn, "Six-fold coordinated carbon dioxide VI," Nat. Mater. **6**, 34–38 (2007).

[260]S. R. Shieh, I. Jarrige, M. Wu, N. Hiraoka, J. S. Tse, Z. Mi, L. Kaci, J.-Z. Jiang, and Y. Q. Cai, "Electronic structure of carbon dioxide under pressure and insights into the molecular-to-nonmolecular transition," Proc. Natl. Acad. Sci. U. S. A. **110**, 18402–18406 (2013).
[261]J. Li, O. Sode, G. A. Voth, and S. Hirata, "A solid–solid phase transition in carbon dioxide at high pressures and intermediate temperatures," Nat. Commun. **4**, 2647 (2013); Erratum, **6**, 8907 (2015).
[262]Y. N. Heit, K. D. Nanda, and G. J. O. Beran, "Predicting finite-temperature properties of crystalline carbon dioxide from first principles with quantitative accuracy," Chem. Sci. **7**, 246–255 (2016).
[263]W. Sontising, Y. N. Heit, J. L. McKinley, and G. J. O. Beran, "Theoretical predictions suggest carbon dioxide phases III and VII are identical," Chem. Sci. **8**, 7374–7382 (2017).
[264]K. D. Litasov, A. F. Goncharov, and R. J. Hemley, "Crossover from melting to dissociation of CO_2 under pressure: Implications for the lower mantle," Earth Planet. Sci. Lett. **309**, 318–323 (2011).
[265]V. M. Giordano, F. Datchi, and A. Dewaele, "Melting curve and fluid equation of state of carbon dioxide at high pressure and high temperature," J. Chem. Phys. **125**, 054504 (2006).
[266]V. M. Giordano and F. Datchi, "Molecular carbon dioxide at high pressure and high temperature," Europhys. Lett. **77**, 46002 (2007).
[267]V. Iota and C. S. Yoo, "Phase diagram of carbon dioxide: Evidence for a new associated phase," Phys. Rev. Lett. **86**, 5922–5925 (2001).
[268]H. Olijnyk and A. P. Jephcoat, "Vibrational studies on CO_2 up to 40 GPa by Raman spectroscopy at room temperature," Phys. Rev. B **57**, 879–888 (1998).
[269]J. D. Hartman, S. Monaco, B. Schatschneider, and G. J. O. Beran, "Fragment-based ^{13}C nuclear magnetic resonance chemical shift predictions in molecular crystals: An alternative to planewave methods," J. Chem. Phys. **143**, 102809 (2015).
[270]D. Flaig, M. Maurer, M. Hanni, K. Braunger, L. Kick, M. Thubauville, and C. Ochsenfeld, "Benchmarking hydrogen and carbon NMR chemical shifts at HF, DFT, and MP2 levels," J. Chem. Theory Comput. **10**, 572–578 (2014).
[271]O. Sode, M. Keçeli, K. Yagi, and S. Hirata, "Fermi resonance in solid CO_2 under pressure," J. Chem. Phys. **138**, 074501 (2013).
[272]S. Hirata, O. Sode, M. Keçeli, K. Yagi, and J. Li, "Response to "Comment on 'Fermi resonance in solid CO_2 under pressure'"," J. Chem. Phys. **140**, 177102 (2014).
[273]G. J. O. Beran, "Modeling polymorphic molecular crystals with electronic structure theory," Chem. Rev. **116**, 5567–5613 (2016).
[274]S. L. Price and J. G. Brandenburg, "Molecular crystal structure prediction," in *Non-Covalent Interactions in Quantum Chemistry and Physics*, edited by A. O. de la Roza and G. A. DiLabio (Elsevier, 2017), Chap. 11, pp. 333–363.
[275]K. Szalewicz, "Symmetry-adapted perturbation theory of intermolecular forces," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 254–272 (2012).
[276]G. Jansen, "Symmetry-adapted perturbation theory based on density functional theory for noncovalent interactions," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 127–144 (2014).
[277]E. Francisco and A. M. Pendás, "Energy partition analyses: Symmetry-adapted perturbation theory and other techniques," in *Non-Covalent Interactions in Quantum Chemistry and Physics*, edited by A. O. de la Roza and G. A. DiLabio (Elsevier, 2017), Chap. 2, pp. 27–64.
[278]J. Řezáč and A. de la Lande, "Robust, basis-set independent method for the evaluation of charge-transfer energy in noncovalent complexes," J. Chem. Theory Comput. **11**, 528–537 (2015).
[279]K. U. Lao and J. M. Herbert, "Energy decomposition analysis with a stable charge-transfer term for interpreting intermolecular interactions," J. Chem. Theory Comput. **12**, 2569–2582 (2016).
[280]J. Řezác and A. de la Lande, "On the role of charge transfer in halogen bonding," Phys. Chem. Chem. Phys. **19**, 791–803 (2017).
[281]K. U. Lao and J. M. Herbert, "Symmetry-adapted perturbation theory with Kohn-Sham orbitals using non-empirically tuned, long-range-corrected density functionals," J. Chem. Phys. **140**, 044108 (2014).
[282]T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, and C. D. Sherrill, "Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies," J. Chem. Phys. **140**, 094106 (2014).
[283]A. Heßelmann, "Comparison of intermolecular interaction energies from SAPT and DFT including empirical dispersion contributions," J. Phys. Chem. A **115**, 11321–11330 (2011).
[284]K. U. Lao and J. M. Herbert, "Accurate intermolecular interactions at dramatically reduced cost: XPol+SAPT with empirical dispersion," J. Phys. Chem. Lett. **3**, 3241–3248 (2012).
[285]K. U. Lao and J. M. Herbert, "An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method," J. Chem. Phys. **139**, 034107 (2013); Erratum, **140**, 119901 (2014).
[286]K. U. Lao and J. M. Herbert, "Accurate and efficient quantum chemistry calculations of noncovalent interactions in many-body systems: The XSAPT family of methods," J. Phys. Chem. A **119**, 235–253 (2015).
[287]K. U. Lao and J. M. Herbert, "Atomic orbital implementation of extended symmetry-adapted perturbation theory (XSAPT) and benchmark calculations for large supramolecular complexes," J. Chem. Theory Comput. **14**, 2955–2978 (2018).
[288]K. U. Lao and J. M. Herbert, "A simple correction for nonadditive dispersion within extended symmetry-adapted perturbation theory (XSAPT)," J. Chem. Theory Comput. **14**, 5128–5142 (2018).
[289]S. Grimme, "Density functional theory with London dispersion corrections," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **1**, 211–228 (2011).
[290]S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, "Dispersion-corrected mean-field electronic structure methods," Chem. Rev. **116**, 5105–5154 (2016).
[291]K. Carter-Fenk, K. U. Lao, K.-Y. Liu, and J. M. Herbert, "Accurate and efficient *ab initio* calculations for supramolecular complexes: Symmetry-adapted perturbation theory with many-body dispersion," J. Phys. Chem. Lett. **10**, 2706–2714 (2019).
[292]J. Hermann, R. A. DiStasio, Jr., and A. Tkatchenko, "First-principles models for van der Waals interactions in molecules and materials: Concepts, theory, and applications," Chem. Rev. **117**, 4714–4758 (2017).
[293]V. F. Lotrich and K. Szalewicz, "Symmetry-adapted perturbation theory of three-body nonadditivity of intermolecular interaction energy," J. Chem. Phys. **106**, 9668–9687 (1997).
[294]V. F. Lotrich and K. Szalewicz, "Symmetry-adapted perturbation theory of three-body nonadditivity in Ar trimer," J. Chem. Phys. **106**, 9688–9702 (1997).
[295]V. F. Lotrich and K. Szalewicz, "Perturbation theory of three-body exchange nonadditivity and application to helium trimer," J. Chem. Phys. **112**, 112–121 (2000).
[296]K.-Y. Liu, K. Carter-Fenk, and J. M. Herbert, "Self-consistent charge embedding at very low cost, with application to symmetry-adapted perturbation theory," J. Chem. Phys. **151**, 031102 (2019).
[297]N. Mardirossian and M. Head-Gordon, "ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation," J. Chem. Phys. **144**, 214110 (2016).
[298]N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals," Mol. Phys. **115**, 2315–2372 (2017).
[299]R. Sure and S. Grimme, "Corrected small basis set Hartree-Fock method for large systems," J. Comput. Chem. **34**, 1672–1685 (2013).
[300]S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen, "Consistent structures and interactions by density functional theory with small atomic orbital basis sets," J. Chem. Phys. **143**, 054107 (2015).
[301]K. Szalewicz, "Determination of structure and properties of molecular crystals from first principles," Acc. Chem. Res. **47**, 3266–3274 (2014).
[302]M. P. Metz, K. Piszczatowski, and K. Szalewicz, "Automatic generation of intermolecular potential energy surfaces," J. Chem. Theory Comput. **12**, 5895–5919 (2016).
[303]J. G. McDaniels and J. R. Schmidt, "Next-generation force fields from symmetry-adapted perturbation theory," Annu. Rev. Phys. Chem. **67**, 467–488 (2016).

[304]D. G. Fedorov and K. Kitaura, "Pair interaction energy decomposition analysis," J. Comput. Chem. **28**, 222–237 (2007).
[305]K. Kitaura and K. Morokuma, "A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation," Int. J. Quantum Chem. **10**, 325–340 (1976).
[306]P. Reinhardt, J.-P. Piquemal, and A. Savin, "Fragment-localized Kohn-Sham orbitals via a singles configuration-interaction procedure and application to local properties and intermolecular energy decomposition analysis," J. Chem. Theory Comput. **4**, 2020–2029 (2008).
[307]Q. Wu, P. W. Ayers, and Y. Zhang, "Density-based energy decomposition analysis for intermolecular interactions with variationally determined intermediate state energies," J. Chem. Phys. **131**, 164112 (2009).
[308]P. R. Horn, Y. Mao, and M. Head-Gordon, "Probing non-covalent interactions with a second generation energy decomposition analysis using absolutely localized molecular orbitals," Phys. Chem. Chem. Phys. **18**, 23067–23079 (2016).
[309]S. Y. Willow, M. A. Salim, K. S. Kim, and S. Hirata, "*Ab initio* molecular dynamics of liquid water using embedded-fragment second-order many-body perturbation theory towards its accurate property prediction," Sci. Rep. **5**, 14358 (2015).
[310]R. A. DiStasio, Jr., B. Santra, Z. Li, X. Wu, and R. Car, "The individual and collective effects of exact exchange and dispersion interactions on the *ab initio* structure of liquid water," J. Chem. Phys. **141**, 084502 (2014).
[311]M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele, "Bulk liquid water at ambient temperature and pressure from MP2 theory," J. Phys. Chem. Lett. **4**, 3753–3759 (2013); Erratum, **5**, 3066–3067 (2014).
[312]S. Y. Willow, X. C. Zeng, S. S. Xantheas, K. S. Kim, and S. Hirata, "Why is MP2-water 'cooler' and 'denser' than DFT-water?," J. Phys. Chem. Lett. **7**, 680–684 (2016).
[313]J. VandeVondele and J. Hutter, "Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases," J. Chem. Phys. **127**, 114105 (2007).
[314]S. Grimme, "Improved second-order Møller–Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies," J. Chem. Phys. **118**, 9095–9102 (2003).
[315]P. Ball, "Water—An enduring mystery," Nature **452**, 291–292 (2008).
[316]G. N. I. Clark, C. D. Cappa, J. D. Smith, R. J. Saykally, and T. Head-Gordon, "The structure of ambient water," Mol. Phys. **108**, 1415–1433 (2010).
[317]A. Nilsson and L. G. M. Pettersson, "Perspective on the structure of liquid water," Chem. Phys. **389**, 1–34 (2011).
[318]W. Hua, T. Fang, W. Li, J.-G. Yu, and S. Li, "Geometry optimizations and vibrational spectra of large molecules from a generalized energy-based fragmentation approach," J. Phys. Chem. A **112**, 10864–10872 (2008).
[319]L. Zhang, W. Li, T. Fang, and S. Li, "Ab initio molecular dynamics with intramolecular noncovalent interactions for unsolvated polypeptides," Theor. Chem. Acc. **135**, 34 (2016).
[320]T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, "Fully analytic energy gradient in the fragment molecular orbital method," J. Chem. Phys. **134**, 124115 (2011).
[321]T. Nagata, D. G. Fedorov, K. Ishimura, and K. Kitaura, "Analytic energy gradient for second-order Møller-Plesset perturbation theory based on the fragment molecular orbital method," J. Chem. Phys. **135**, 044110 (2011).
[322]T. Nagata, D. G. Fedorov, and K. Kitaura, "Analytic gradient for the embedding potential with approximations in the fragment molecular orbital method," Chem. Phys. Lett. **544**, 87–93 (2012).
[323]H. Nakata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, "Efficient vibrational analysis for unrestricted Hartree–Fock based on the fragment molecular orbital method," Chem. Phys. Lett. **603**, 67–74 (2014).
[324]H. Nakata and D. G. Fedorov, "Analytic second derivatives for the efficient electrostatic embedding in the fragment molecular orbital method," J. Comput. Chem. **39**, 2039–2050 (2018).
[325]Q. Gao, S. Yokojima, T. Kohno, T. Ishida, D. G. Fedorov, K. Kitaura, M. Fujihira, and S. Nakamura, "*Ab initio* NMR chemical shift calculations on proteins using fragment molecular orbitals with electrostatic environment," Chem. Phys. Lett. **445**, 331–339 (2007).
[326]H. Nakata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, "Simulations of Raman spectra using the fragment molecular orbital method," J. Chem. Theory Comput. **10**, 3689–3698 (2014).
[327]K. Kitaura, S.-I. Sugiki, T. Nakano, Y. Komeiji, and M. Uebayasi, "Fragment molecular orbital method: Analytical energy gradients," Chem. Phys. Lett. **336**, 163–170 (2001).
[328]D. G. Fedorov and K. Kitaura, "Second-order Møller-Plesset perturbation theory based upon the fragment molecular orbital method," J. Chem. Phys. **121**, 2483–2490 (2004).
[329]D. G. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, "The fragment molecular orbital method for geometry optimizations of polypeptides and proteins," J. Phys. Chem. A **111**, 2722–2732 (2007).
[330]T. Nagata, D. G. Fedorov, and K. Kitaura, "Derivatives of the approximated electrostatic potentials in the fragment molecular orbital method," Chem. Phys. Lett. **475**, 124–131 (2009).
[331]T. Nagata, D. G. Fedorov, and K. Kitaura, "Importance of the hybrid orbital operator derivative term for the energy gradient in the fragment molecular orbital method," Chem. Phys. Lett. **492**, 302–308 (2010).
[332]Y. Mochizuki, T. Nakano, Y. Komeiji, K. Yamashita, Y. Okiyama, H. Yoshikawa, and H. Yamataka, "Fragment molecular orbital-based molecular dynamics (FMO-MD) method with MP2 gradient," Chem. Phys. Lett. **504**, 95–99 (2011).
[333]H. Li, D. G. Fedorov, T. Nagata, K. Kitaura, J. H. Jensen, and M. S. Gordon, "Energy gradients in combined fragment molecular orbital and polarizable continuum model (FMO/PCM) calculation," J. Comput. Chem. **31**, 778–790 (2010).
[334]H. Nakata and D. G. Fedorov, "Efficient geometry optimization of large molecular systems in solution using the fragment molecular orbital method," J. Phys. Chem. A **120**, 9794–9804 (2016).
[335]H. Nakata and D. G. Fedorov, "Simulations of infrared and Raman spectra in solution using the fragment molecular orbital method," Phys. Chem. Chem. Phys. **21**, 13641–13652 (2019).
[336]W. Xie, L. Song, D. G. Truhlar, and J. Gao, "The variational explicit polarization potential and analytical first derivative of energy: Towards a next generation force field," J. Chem. Phys. **128**, 234108 (2008).
[337]J. Gao, D. G. Truhlar, Y. Wang, M. J. M. Mazack, P. Löffler, M. R. Provorse, and P. Rehak, "Explicit polarization: A quantum mechanical framework for developing next generation force fields," Acc. Chem. Res. **47**, 2837–2845 (2014).
[338]Y. Wang, M. J. M. Mazack, D. G. Truhlar, and J. Gao, "Explicit polarization theory," in *Many-Body Effects and Electrostatics in Biomolecules*, edited by Q. Cui, M. Meuwly, and P. Ren (Pan Stanford, Boca Raton, FL, 2016), Chap. 2, pp. 33–64.
[339]K. R. Brorsen, N. Minezawa, F. Xu, T. L. Windus, and M. S. Gordon, "Fragment molecular orbital molecular dynamics with the fully analytic energy gradient," J. Chem. Theory Comput. **8**, 5008–5012 (2012).
[340]A. I. Krylov and P. M. W. Gill, "Q-Chem: An engine for innovation," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**, 317–326 (2013).
[341]C. Steinmann, D. G. Fedorov, and J. H. Jensen, "Effective fragment molecular orbital method: A merger of the effective fragment potential and fragment molecular orbital methods," J. Phys. Chem. A **114**, 8705–8712 (2010).
[342]T. Nagata, D. G. Fedorov, T. Sawada, K. Kitaura, and M. S. Gordon, "A combined effective fragment potential–fragment molecular orbital method. II. Analytic gradient and application to geometry optimization of solvated tetraglycine chignolin," J. Chem. Phys. **134**, 034110 (2011).
[343]C. Steinmann, D. G. Fedorov, and J. H. Jensen, "The effective fragment molecular orbital method for fragments connected by covalent bonds," PLoS One **7**, e41117 (2012).
[344]C. Bertoni and M. S. Gordon, "Analytic gradients for the effective fragment molecular orbital method," J. Chem. Theory Comput. **12**, 4743–4767 (2016).
[345]C. Steinmann and J. H. Jensen, "Effective fragment molecular orbital method," in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, edited by M. S. Gordon (Wiley, Hoboken, 2017), Chap. 5, pp. 165–182.
[346]M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, "The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry," J. Phys. Chem. A **105**, 293–307 (2001).

[347]D. Ghosh, D. Kosenkov, V. Vanovschi, C. F. Williams, J. M. Herbert, M. S. Gordon, M. W. Schmidt, L. V. Slipchenko, and A. I. Krylov, "Noncovalent interactions in extended systems described by the effective fragment potential method: Theory and application to nucleobase oligomers," J. Phys. Chem. A **114**, 12739–12754 (2010).
[348]L. V. Slipchenko and P. K. Gurunathan, "Effective fragment potential method: Past, present, and future," in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, edited by M. S. Gordon (Wiley, Hoboken, 2017), Chap. 6, pp. 183–208.
[349]J. F. Ouyang and R. P. A. Bettens, "When are many-body effects significant?," J. Chem. Theory Comput. **12**, 5860–5867 (2016).
[350]K.-Y. Liu and J. M. Herbert, "Energy-screened many-body expansion: A practical yet accurate fragmentation method for quantum chemistry" (submitted).
[351]S. Kazachenko and A. J. Thakkar, "Water nanodroplets: Predictions of five model potentials," J. Chem. Phys. **138**, 194302 (2013).
[352]Y. Nishimoto, D. G. Fedorov, and S. Irle, "Density-functional tight-binding combined with the fragment molecular orbital method," J. Chem. Theory Comput. **10**, 4801–4812 (2014).
[353]Y. Nishimoto and D. G. Fedorov, "The fragment molecular orbital method combined with density-functional tight-binding and the polarizable continuum model," Phys. Chem. Chem. Phys. **18**, 22047–22061 (2016).
[354]V. Q. Vuong, Y. Nishimoto, D. G. Fedorov, B. G. Sumpter, T. A. Niehaus, and S. Irle, "The fragment molecular orbital method based on long-range corrected density-functional tight-binding," J. Chem. Theory Comput. **15**, 3008–3020 (2019).
[355]M. Xu, X. He, T. Zhu, and J. Z. H. Zhang, "A fragment quantum mechanical method for metalloproteins," J. Chem. Theory Comput. **15**, 1430–1439 (2019).
[356]E. G. Hohenstein, S. T. Chill, and C. D. Sherrill, "Assessment of the performance of the M05-2X and M06-2X exchange-correlation functionals for noncovalent interactions in biomolecules," J. Chem. Theory Comput. **4**, 1996–2000 (2008).
[357]N. Mardirossian and M. Head-Gordon, "How accurate are the Minnesota density functionals for noncovalent interactions, isomerization energies, thermochemistry, and barrier heights involving molecules composed of main-group elements?," J. Chem. Theory Comput. **12**, 4303–4325 (2016).
[358]K. Fukuzawa, Y. Mochizuki, T. Nakano, and S. Tanaka, "Application of the FMO method to specific molecular recognition of biomolecules," in *The Fragment Molecular Orbital Method*, edited by D. Fedorov and K. Kitaura (Taylor & Francis, 2009), Chap. 7, pp. 133–169.
[359]I. Nakanishi, D. G. Fedorov, and K. Kitaura, "Detailed electronic structure studies revealing the nature of protein–ligand binding," in *The Fragment Molecular Orbital Method*, edited by D. Fedorov and K. Kitaura (Taylor & Francis, 2009), Chap. 8, pp. 171–192.
[360]T. Ozawa, K. Okazaki, and M. Nishio, "FMO as a tool for structure-based drug design," in *The Fragment Molecular Orbital Method*, edited by D. Fedorov and K. Kitaura (Taylor & Francis, 2009), Chap. 10, pp. 217–244.
[361]T. Otsuka, N. Okimoto, and M. Taiji, "Assessment and acceleration of binding energy calculations for protein–ligand complexes by the fragment molecular orbital method," J. Comput. Chem. **36**, 2209–2218 (2015).
[362]M. P. Mazanetz, E. Chudyk, D. G. Fedorov, and Y. Alexeev, "Applications of the fragment molecular orbital method to drug research," in *Computer-Aided Drug Design*, Methods in Pharmacology and Toxicology, edited by W. Zhang (Springer Science+Business Media, New York, 2016), pp. 217–254.
[363]D. G. Fedorov and K. Kitaura, "Modeling and visualization for the fragment molecular orbital method with the graphical user interface FU, and analyses of protein–ligand binding," in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, edited by M. S. Gordon (Wiley, Hoboken, 2017), Chap. 3, pp. 119–140.
[364]D. G. Fedorov, Y. Alexeev, and K. Kitaura, "Geometry optimization of the active site of a large system with the fragment molecular orbital method," J. Phys. Chem. Lett. **2**, 282–288 (2011).
[365]K. V. J. Jose and K. Raghavachari, "Molecules-in-molecules fragment-based method for the evaluation of Raman spectra of large molecules," Mol. Phys. **113**, 3057–3066 (2015).
[366]K. V. J. Jose, D. Beckett, and K. Raghavachari, "Vibrational circular dichroism spectra for large molecules through molecules-in-molecules fragment-based approach," J. Chem. Theory Comput. **11**, 4238–4247 (2015).
[367]J. R. Anema, A. G. Brolo, A. Felten, and C. Bittencourt, "Surface-enhanced Raman scattering from polystyrene on gold clusters," J. Raman Spectrosc. **41**, 745–751 (2010).
[368]P. J. Stephens, F. J. Devlin, S. Schürch, and J. Hulliger, "Determination of the absolute configuration of chiral molecules via density functional theory calculations of vibrational circular dichroism and optical rotation: The chiral alkane D_3-anti-trans-anti-trans-anti-trans-perhydrotriphenylene," Theor. Chem. Acc. **119**, 19–28 (2008).
[369]A. Barth, "Infrared spectroscopy of proteins," Biochim. Biophys. Acta **1767**, 1073–1101 (2007).
[370]J.-H. Choi, S. Ham, and M. Cho, "Local amide I mode frequencies and coupling constants in polypeptides," J. Phys. Chem. B **107**, 9132–9138 (2003).
[371]K.-Y. Liu, J. Liu, and J. M. Herbert, "Accuracy of finite-difference harmonic frequencies in density functional theory," J. Comput. Chem. **38**, 1678–1684 (2017).
[372]R. Amos and R. Kobayashi, "Ab initio NMR chemical shift calculations using fragment molecular orbitals and locally dense basis sets," J. Phys. Chem. A **120**, 8907–8915 (2016).
[373]R. Kobayashi, R. D. Amos, D. M. Reid, and M. A. Collins, "Application of the systematic molecular fragmentation by annihilation method to *ab initio* NMR chemical shift calculations," J. Phys. Chem. A **122**, 9135–9141 (2018).
[374]D. M. Reid and M. A. Collins, "Approximating CCSD(T) nuclear magnetic shielding calculations using composite methods," J. Chem. Theory Comput. **11**, 5177–5181 (2015).
[375]D. B. Chesnut and K. D. Moore, "Locally dense basis sets for chemical shift calculations," J. Comput. Chem. **10**, 648–659 (1989).
[376]F. Jensen, "Segmented contracted basis sets optimized for nuclear magnetic shielding," J. Chem. Theory Comput. **11**, 132–138 (2015).
[377]D. M. Reid and M. A. Collins, "Calculating nuclear magnetic resonance shieldings using systematic molecular fragmentation by annihilation," Phys. Chem. Chem. Phys. **17**, 5314–5320 (2015).
[378]X. He, B. Wang, and K. M. Merz, Jr., "Protein NMR chemical shift calculations based on the automated fragmentation QM/MM approach," J. Phys. Chem. B **113**, 10380–10388 (2009).
[379]T. Zhu, X. He, and J. Z. H. Zhang, "Fragment density functional theory calculation of NMR chemical shifts for proteins with implicit solvation," Phys. Chem. Chem. Phys. **14**, 7837–7845 (2012).
[380]T. Zhu, J. Z. H. Zhang, and X. He, "Automated fragmentation QM/MM calculation of amide proton chemical shifts in proteins with explicit solvent model," J. Chem. Theory Comput. **9**, 2104–2114 (2013).
[381]K. V. J. Jose and K. Raghavachari, "Fragment-based approach for the evaluation of NMR chemical shifts for large biomolecules incorporating the effects of the solvent environment," J. Chem. Theory Comput. **13**, 1147–1158 (2017).
[382]D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, "A new hierarchical parallelization scheme: Generalized distributed data interface (GDDI), and an application to the fragment molecular orbital method (FMO)," J. Comput. Chem. **25**, 872–880 (2004).
[383]G. D. Fletcher, D. G. Fedorov, S. R. Pruitt, T. L. Windus, and M. S. Gordon, "Large-scale MP2 calculations on the Blue Gene architecture using the fragment molecular orbital method," J. Chem. Theory Comput. **8**, 75–79 (2012).
[384]B. Q. Pham and M. S. Gordon, "Hybrid distributed/shared memory model for the RI-MP2 method in the fragment molecular orbital framework," J. Chem. Theory Comput. **15**, 5252–5258 (2019).