The Journal of Chemical Physics | COMMUNICATION | scitation.org/journal/jcp

Self-consistent charge embedding at very low cost, with application to symmetry-adapted perturbation theory

Cite as: J. Chem. Phys. 151, 031102 (2019); doi: 10.1063/1.5111869
Submitted: 30 May 2019 • Accepted: 30 June 2019 •
Published Online: 16 July 2019

Kuan-Yu Liu, Kevin Carter-Fenk, and 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

Extended symmetry-adapted perturbation theory (XSAPT) uses a self-consistent charge embedding to capture many-body polarization, in conjunction with a pairwise-additive SAPT calculation of intermolecular interaction energies. The original implementation of XSAPT is based on charges that are fit to reproduce molecular electrostatic potentials, but this becomes a computational bottleneck in large systems. Charge embedding based on modified Hirshfeld atomic charges is reported here, which dramatically reduces the computational cost without compromising accuracy. Exemplary calculations are presented for supramolecular complexes such as $C_{60}@C_{60}H_{28}$, a DNA intercalation complex, and a 323-atom model of a drug molecule bound to an enzyme active site. The proposed charge embedding should be useful in other fragment-based quantum chemistry methods as well.

Quantum-based modeling of noncovalent interactions in sizable supramolecular assemblies has become possible, thanks to increases in computing power, but hardware improvements alone are insufficient to tackle the complexes of interest in drug discovery, which often involve binding of ligands to proteins or DNA.[1,2] A plethora of fragment-based quantum chemistry methods has emerged to address this issue by reducing the supersystem problem to a many-body problem involving relatively small fragments.[3–15] Along these lines, our group has developed extended symmetry-adapted perturbation theory (XSAPT),[16–23] an accurate and efficient monomer-based approach to compute intermolecular interaction energies that generalizes traditional SAPT-based energy decomposition analysis[24,25] to the case of more than two monomers.

XSAPT combines traditional dimer SAPT with the explicit polarization or "XPol" method,[26] using the latter to obtain the monomer wave functions.[16,17] In this way, many-body polarization is included in the unperturbed wave functions by means of self-consistent electrostatic embedding. Considering closed-shell fragments for simplicity, the XPol energy expression is

$$E = \sum_A \left[2 \sum_n^{\text{occ}} \left(\mathbf{c}_n^A\right)^\dagger \left(\mathbf{h}^A + \mathbf{J}^A - \tfrac{1}{2}\mathbf{K}^A\right)\mathbf{c}_n^A + E_{\text{nuc}}^A \right] + E_{\text{embed}}. \quad (1)$$

The expression in square brackets represents the Hartree-Fock energy for monomer A, expressed in terms of "absolutely localized" molecular orbitals (MOs),[27] $\mathbf{c}_n^A$. The final term, E_{embed}, is the sum of electrostatic embedding energies involving wave function-derived point charges.[17]

Variation of the energy expression in Eq. (1) within the absolutely localized *ansatz* affords a Fock matrix

$$\mathbf{F}^A = f_{\mu\nu}^A - \frac{1}{2}\sum_{B\neq A}\sum_{b\in B} q_b (\mathbf{I}_b)_{\mu\nu} + \sum_{a\in A} \frac{\partial E_{\text{embed}}}{\partial q_a}\frac{\partial q_a}{\partial P_{\mu\nu}} \quad (2)$$

for monomer A.[17,28] Here, $f_{\mu\nu}^A$ is the Fock matrix for the isolated monomer and

$$(\mathbf{I}_b)_{\mu\nu} = \left\langle \phi_\mu \middle| \frac{1}{\|\mathbf{r} - \mathbf{R}_b\|} \middle| \phi_\nu \right\rangle \quad (3)$$

is a one-electron integral representing the electrostatic potential generated by the function pair $\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})$ at the point $\mathbf{R}_b$, which specifies the location of the atomic embedding charge q_b.

We use the "SAPT0" energy formula,[25] which describes the interaction energy through second order in the intermolecular perturbation,

$$E_{\text{int}} = E^{(1)}_{\text{elst}} + E^{(1)}_{\text{exch}} + E^{(2)}_{\text{ind}} + E^{(2)}_{\text{exch-ind}} + E^{(2)}_{\text{disp}} + E^{(2)}_{\text{exch-disp}} . \quad (4)$$

In order to incorporate *intra*molecular electron correlation in an efficient fashion, we adopt the SAPT(KS) variant of this theory,[29–32] where "KS" indicates that the MOs are obtained from Kohn-Sham density functional theory (DFT). This approach should not be confused with SAPT(DFT),[33] in which the dispersion energy is computed using density susceptibilities obtained from Kohn-Sham response theory, as an alternative to the second-order expressions in Eq. (4). Second-order dispersion energies are especially sensitive to problems with the asymptotic behavior of the exchange-correlation (XC) functional,[17,32] but it is possible to obtain dispersion energies that are no worse than their SAPT0 counterparts (based on Hartree-Fock theory), by using range-separated hybrid functionals that are tuned for each monomer.[20,21,32] The other SAPT(KS) energy components are improved relative to SAPT0.[32]

XSAPT approximates the total interaction energy by pairwise application of Eq. (4), with non-pairwise-additive polarization effects included implicitly in the unperturbed XPol wave functions.[16,19] The SAPT0 formula can also be corrected for nonadditive dispersion.[22] The electrostatics, exchange, induction, and exchange-induction contributions to Eq. (4) can be evaluated at $\mathcal{O}(N^3)$ cost, but the dispersion and exchange-dispersion terms scale as $\mathcal{O}(N^4)$ and $\mathcal{O}(N^5)$, respectively. These are also the least accurate parts of a SAPT0 or SAPT(KS) calculation,[19,32] so we usually replace them with either *ab initio* atom–atom dispersion potentials ("+*ai*D")[18–21] or self-consistently screened many-body dispersion (MBD).[23,34]

Construction of the XPol Fock matrix in Eq. (2) requires a prescription for how the embedding charges are to be derived from the monomer wave functions. For this, we have used "ChElPG" charges[17,35,36] that are fit to reproduce the molecular electrostatic potential, evaluated on a real-space grid in regions beyond the atomic van der Waals radii. ChElPG charges are physically appealing, numerically stable,[37] and afford good accuracy when used in XSAPT.[16–21,23] However, the equations for the charge derivatives $\partial q_a/\partial P_{\mu\nu}$ that appear in the monomer Fock matrices are quite complicated,[17,37,38] and evaluation of these derivatives is costly.[21,37] In fact, this quickly becomes the computational bottleneck for XSAPT calculations involving large monomers,[21] e.g., the buckycatcher/fullerene (C_{60}@$C_{60}H_{28}$) complex that is depicted in Fig. 1. In an XSAPT calculation for this complex, 35% of the computational time is spent in evaluating ChElPG charge derivatives.[21]

In view of this, we sought an alternative way to perform the charge embedding and settled on "Charge Model 5" (CM5),[39] which amounts to an empirical modification to Hirshfeld population analysis.[40] Hirshfeld atomic charges are obtained from the molecular density by using a superposition of isolated-atom densities $\{\widetilde{\rho}_a\}$ to define a weight function

$$W_a(\mathbf{r}) = \frac{\widetilde{\rho}_a(\mathbf{r})}{\sum_b \widetilde{\rho}_b(\mathbf{r})} \quad (5)$$

for atom a. The weight functions are then used to partition the molecular density $\rho(\mathbf{r})$ into atomic contributions, with the Hirshfeld charge on atom a defined as

$$q_a^{\text{Hirsh}} = Z_a - \int W_a(\mathbf{r})\, \rho(\mathbf{r})\, d\mathbf{r}. \quad (6)$$

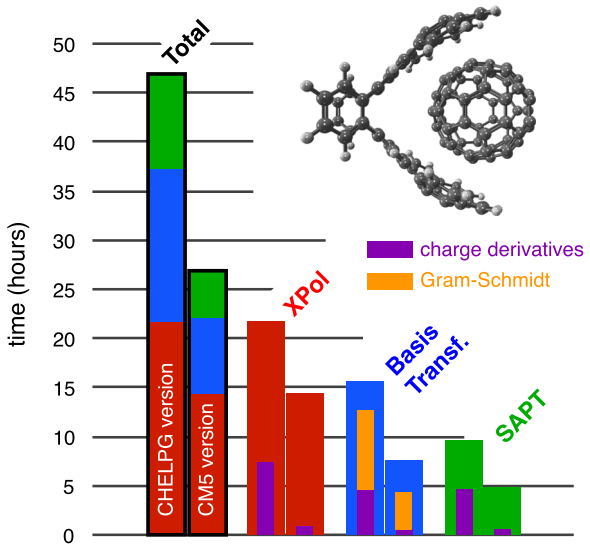

FIG. 1. Timing data for XSAPT(KS)+*ai*D/hp-TZVPP calculations on C_{60}@$C_{60}H_{28}$ (4592 basis functions), using either the original ChElPG implementation of XSAPT (left-hand bar in each pair, using data from Ref. 21) or the new CM5 implementation reported here (right-hand bars). Both calculations were parallelized across all 28 cores of a single compute node. The total calculation time is broken down into three color-coded steps: the XPol self-consistent field procedure (in red); pseudo-canonicalization to transform the monomer MOs to a dimer basis (in blue); and, finally, the SAPT calculation (in green). Charge derivatives are required in all three steps, and these timing data are separated out and depicted in purple. Orange bars represent the Gram-Schmidt orthogonalization contribution to the pseudo-canonicalization step, where the multithreading has been improved in the present implementation as compared to the one reported in Ref. 21.

These charges are sometimes considered to be too small,[41] in the sense that the dipole moment obtained from them is smaller than the true dipole moment obtained from $\rho(\mathbf{r})$, and CM5 introduces an empirical modification to correct for this.[39] We express this as

$$q_a^{\text{CM5}}[\rho(\mathbf{r})] = q_a^{\text{Hirsh}}[\rho(\mathbf{r})] + q'_a, \quad (7)$$

where the empirical correction q'_a depends on the atomic number Z_a and the molecular geometry but is independent of $\rho(\mathbf{r})$.

The charge derivatives needed in Eq. (2) are

$$\frac{\partial q_a^{\text{CM5}}}{\partial P_{\mu\nu}} = -\int W_a(\mathbf{r})\, \phi_\mu(\mathbf{r})\, \phi_\nu(\mathbf{r})\, d\mathbf{r}. \quad (8)$$

These integrals can be evaluated using the same numerical quadrature used to compute the XC contribution to the Fock matrix, but a naïve implementation proves to be costly. Introducing a molecular quadrature grid consisting of points $\{\mathbf{r}_i\}$ and weights $\{w_i\}$, we have

$$\frac{\partial q_a^{\text{CM5}}}{\partial P_{\mu\nu}} = -\sum_i w_i\, W_a(\mathbf{r}_i)\, \phi_\mu(\mathbf{r}_i)\, \phi_\nu(\mathbf{r}_i). \quad (9)$$

The cost of this implementation scales with the number of atoms (N_{atoms}) and basis functions (N_{basis}) as $\mathcal{O}(N_{\text{atoms}} \times N_{\text{basis}}^2 \times N_{\text{mol-Leb}})$, where $N_{\text{mol-Leb}}$ represents the number of Lebedev grid points that is required for accurate integration of $\rho(\mathbf{r})$. For typical quadrature

grids, this number ranges from ≈3800 points per atom (low-quality, SG-1) to 15 000–19 000 points per atom (high quality, SG-3).[42] In contrast, the cost to evaluate ChElPG charge derivatives scales as $\mathcal{O}(N_{\text{atoms}} \times N_{\text{basis}}^2 \times N_{\text{ESP-grid}})$,[37] where $N_{\text{ESP-grid}}$ represents the number of electrostatic potential grid points. Because the ChElPG procedure fits only to the long-range, slowly-varying parts of the electrostatic potential, it is possible to make $N_{\text{ESP-grid}} \ll N_{\text{mol-Leb}}$, especially if atom-centered Lebedev grids are used in the ChElPG algorithm.[37,38] (For example, $N_{\text{ESP-grid}} = 3044$ for the $C_{60}@C_{60}H_{28}$ calculations reported in Fig. 1, using the Lebedev-based ChElPG algorithm described in Ref. 37.) In this case, no actual cost savings are realized by replacing ChElPG charges with CM5 charges.

That said, the cost to implement Eq. (9) can be reduced dramatically by recognizing that the integrand vanishes as $\mathbf{r}$ moves away from $\mathbf{R}_a$, the position of nucleus a. This occurs both because the free-atom density $\widetilde{\rho}_a(\mathbf{r})$ vanishes in this limit, and also because the basis functions $\phi_\mu(\mathbf{r})$ and $\phi_\nu(\mathbf{r})$ must both be associated with atom a, else $\partial q_a^{\text{CM5}}/P_{\mu\nu} = 0$. This implies that both $\phi_\mu(\mathbf{r})$ and $\phi_\nu(\mathbf{r})$ decay to zero as r moves away from $\mathbf{R}_a$. As such, the integral required to compute this derivative can be evaluated accurately and efficiently using just the atom-centered grid for a, not the entire molecular grid. In effect, we restrict the summation in Eq. (9) to just those grid points $i \in a$. This reduces the cost of the CM5 charge derivatives to $\mathcal{O}(N_{\text{atoms}} \times N_{\text{basis}}^2 \times N_{\text{atom-Leb}})$, which is essentially the same as the cost of the XC quadrature step in a DFT calculation.

The accuracy of the atomic-grid implementation of Eq. (9) has been tested by computing XSAPT interaction energies for the S22 data set.[43] The maximum deviation (with respect to an implementation that uses the full molecular quadrature grid) is 0.02 kcal/mol, with no systematic deviation, and the total molecular charge is conserved to within $\sim 10^{-5}e$. Figure 1 presents a timing comparison for the CM5-based implementation of XSAPT as compared to the original ChElPG-based implementation,[21] as applied to $C_{60}@C_{60}H_{28}$. The time required to compute charge derivatives $\partial q_a/\partial P_{\mu\nu}$ has been reduced from 16.7 h to 2.0 h, and even larger speedups are anticipated as the system size increases, since $N_{\text{ESP-grid}}$ increases with molecular size but $N_{\text{atom-Leb}}$ does not. A secondary cost reduction reflected in the timing data stems from improved multithreading of the repeated matrix multiplications required for the basis transformation or "pseudocanonicalization" step.[16] This transformation represents an alternative to using the dimer basis to compute the monomer wave functions,[24,44] as is often done in dimer SAPT calculations but which becomes ill-defined in XSAPT calculations involving more than two monomers.[16] Together, these improvements alter the nature of the bottleneck in large XSAPT calculations, which are now dominated by the more traditional parts of the self-consistent field calculations that comprise the XPol step.

The remainder of this work documents the accuracy of the new CM5-based implementation of XSAPT. We first consider the standard S22[43] and S66[45] data sets, which consist of dimers of charge-neutral molecules, along with the ion–molecule hydrogen bonding (IMHB) data set of Řezác and Hobza,[46] and an ion-pair data set from Lao and Herbert.[20] Error statistics for both CM5- and ChElPG-based implementations of XSAPT, as compared to the benchmark interaction energies for each data set, are listed in Table I. Both charge schemes provide comparable results for S22 and S66, but

TABLE I. Errors in XSAPT interaction energies for data sets of small dimers, as compared to benchmark values.[a]

	Error (kcal/mol)			
	Maximum		MUE	
Data set	CM5	ChElPG	CM5	ChElPG
S22[b]	−1.1	−1.2	0.4	0.4
S66[b]	−1.1	−1.1	0.3	0.3
IMHB[c]	−2.4	−5.3	1.1	1.7
Ions[c]	−3.8	−15.1	1.4	3.6

[a]Benchmarks are MP2/cc-pVTZ for IMHB and complete-basis CCSD(T) for the other data sets.
[b]SAPT(KS) calculations use LRC-ωPBE/hpTZVPP for the monomers.
[c]SAPT(KS) calculations use ωB97X-V/def2-TZVPPD for the monomers.

significant differences are observed where ions are involved. For the IMHB data set, the largest absolute deviation between CM5- and ChElPG-based XSAPT interaction energies occurs in the case of the imidazolium···methylamine complex. Here, the charge difference $|q_N - q_C|$ between the heavy atoms in methylamine is an unrealistically large $2.3e$ in the case of ChElPG charges, vs only $0.5e$ for CM5 charges. For the ion-pair data set, the largest deviation is found in the complex of Cl^- with dimethyl ethyl amine, where the ChElPG atomic charges result in bond dipoles whose positive ends point toward the nitrogen atom, whereas in the CM5 case they point away.

These observations bring to mind a frequent criticism of ChElPG charges, at least when it comes to their use in force-field parameterization, which is that the ChElPG procedure may sacrifice chemically intuitive atomic partial charges in the interest of better fitting the molecular electrostatic potential.[36] This problem becomes more severe for large molecules with "buried" atoms whose ChElPG charges may be devoid of any chemical significance whatsoever. Related to this is an unwanted conformational dependence of the ChElPG charges.[36] We have previously considered these criticisms not to be relevant in the context of XSAPT, since we have no interest in the atomic partial charges *per se*, beyond their ability to reproduce the electrostatic potential. However, it appears that for monomers with net charge, the CM5 procedure affords both more intuitive atomic charges *and* smaller errors in intermolecular interaction energies. That said, a recent study of various wave function-derived charge models concluded that CM5 charges were slightly less accurate than ChElPG charges for describing strong electrostatic interactions.[47] The reasons for this discrepancy (as compared to the present results) are unclear.

We next examine the performance of CM5-based XSAPT in different basis sets. Table II shows mean unsigned errors (MUEs) for several different data sets containing ionic monomers.[48] These include the AHB21 and CHB6 data sets in which one monomer is an anion or a cation, respectively, and also the IL16 data set consisting of ion-pairs that are common constituents of room-temperature ionic liquids. These systems are rather small, and perhaps for that reason the XSAPT results converge already in the aug-cc-pVDZ basis set. A more detailed breakdown can be found

TABLE II. XSAPT error statistics for several data sets from Ref. 48 that contain ionic monomers.

	MUE (kcal/mol)							
	AHB21		CHB6		IL16		Overall	
Basis set	CM5	ChElPG	CM5	ChElPG	CM5	ChElPG	CM5	ChElPG
cc-pVDZ	8.3	9.0	2.4	2.7	10.7	11.6	8.7	9.4
jun-cc-pVDZ	1.2	1.3	...	...	1.4	2.9	1.3	2.0
aug-cc-pVDZ	0.9	2.9	1.2	0.8	3.0	7.2	1.8	4.3
cc-pVTZ	5.9	6.3	1.4	1.3	7.7	9.9	6.2	7.2
aug-cc-pVTZ	1.2	2.0	0.8	0.7	3.2	10.6	1.9	5.2
def2-TZVPP	3.5	3.0	0.5	0.5	1.3	2.3	2.3	2.4
def2-TZVPPD	1.2	1.1	1.0	1.1	1.9	7.7	1.4	3.6

in Tables S7–S9 of the supplementary material, and these data reveal that the difference between the CM5 and ChElPG charges is marginal for the AHB21 and CHB6 data sets but quite pronounced for IL16, where both monomers are ions. Note that ChElPG charges have occasionally been used as a metric for intermolecular charge transfer, e.g., in ionic liquids,[49,50] but this seems rather dubious in view of the unphysical values that we sometimes encounter in ionic systems.

The S30L data set[51] consists of 30 sizable host/guest complexes, including the $C_{60}@C_{60}H_{28}$ complex in Fig. 1 that we have previously used to benchmark various versions of XSAPT.[21,23] Our ChElPG-based implementation of XSAPT affords a MUE of 4.7 kcal/mol for S30L, which is competitive with the best-available quantum chemistry approaches, at reduced cost even as compared to supramolecular DFT.[23] (To give context to this MUE, note that the S30L benchmark interaction energies are back-corrected to the gas phase starting from experimental binding affinities measured in solution, and come with estimated uncertainties of 2–3 kcal/mol.[21,51]) The CM5-based version of XSAPT affords a slightly lower MUE (4.1 kcal/mol) even while it accelerates the calculation of the $C_{60}@C_{60}H_{28}$ interaction energy by nearly a factor of two!

Figure 2 displays two examples of ligands bound to macromolecules, including an intercalation complex of ellipticine with DNA,[1] and the antiretroviral drug indinavir embedded in a model of the active site of HIV-2 protease.[2] Interaction energies for these two complexes at various levels of theory are listed in Table III. (The XSAPT+MBD values are based on a slightly updated parameterization of the MBD correction, as compared to our original version in Ref. 23, and details are provided in the supplementary material.) For the smaller DNA/ellipticine complex, we can compare to counterpoise-corrected DFT/def2-TZVPPD results using two of the best-performing density functionals for noncovalent interactions,[53] namely, B97M-V and ωB97M-V, both of which afford similar interaction energies. The XSAPT+*ai*D3 method,[20] which

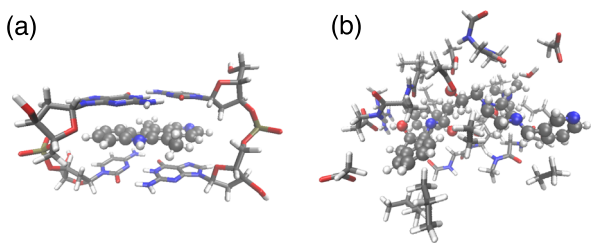

FIG. 2. Model systems for drug binding: (a) ellipticine/DNA intercalation complex (157 atoms) and (b) the indinavir molecule in a model of the HIV-2 binding pocket (323 atoms). In both cases, the drug molecule is shown in a ball-and-stick representation, while the rest of the system is depicted with a tubular representation.

TABLE III. Interaction energies for the ligand/macromolecule complexes in Fig. 2.

		E_{int} (kcal/mol)	
Method		DNA/ellipticine	HIV/indinavir
QMC[a]		−33.6	...
B97M-V (+counterpoise)[b]		−41.3	...
ωB97M-V (+counterpoise)[b]		−43.7	...
HF-3c		−41.7	−132.8
PBEh-3c		−37.3	−119.1
XSAPT+*ai*D3 (CM5)[c,d]		−36.7	−106.2
XSAPT+*ai*D3 (ChElPG)[c,d]		−35.7	−103.9
XSAPT+MBD (CM5)[c,e]		−41.7	−125.4
XSAPT+MBD (ChElPG)[c,e]		−40.7	−123.1
XSAPT energy decomposition			
E_{elst}		−22.2	−114.9
E_{exch}		59.2	190.0
E_{ind}		−8.0	−65.9
E_{disp}	*ai*D3+ATM[d]	−65.7	−115.4
	MBD+esDQ[e]	−70.7	−134.6

[a] From Ref. 52.
[b] def2-TZVPPD basis set.
[c] def2-hpTZVPP basis set (Ref. 20).
[d] Includes Axilrod-Teller-Muto (ATM) three-body dispersion (Ref. 21).
[e] Many-body dispersion with effectively screened dipole–quadrupole dispersion (MBD+esDQ) model (Ref. 23).

describes dispersion by means of pairwise atom–atom dispersion potentials ("+*ai*D3"), underestimates the interaction energy as compared to DFT, even when Axilrod-Teller-Muto (ATM) triatomic C_9 dispersion corrections are included.[21] The recently developed XSAPT+MBD method[23] predicts an interaction energy in much better agreement with the DFT values. Most important from the standpoint of this work is the fact that both ChElPG- and CM5-based versions of XSAPT predict interaction energies within 1.0 kcal/mol of one another. For reasons that are unclear, all of these methods overestimate the interaction energy as compared to a quantum Monte Carlo (QMC) estimate from Ref. 52.

A supramolecular DFT calculation on the HIV/indinavir complex (8346 basis functions using def2-TZVPPD) would strain our computational capabilities, but XSAPT calculations are quite feasible and are reported in Table III. As a sanity check, we have computed E_{int} for both ligand/macromolecule complexes using the semiempirical HF-3c[54] and PBEh-3c[55] methods, which perform reasonably well for supramolecular complexes despite their low cost.[21] For the DNA/ellipticine complex, HF-3c predicts an interaction energy comparable to the DFT methods, while PBEh-3c underestimates this value. Assuming roughly the same behavior for the HIV/indinavir complex, we can say that XSAPT+MBD interaction energies seem reasonable while those obtained with XSAPT+*ai*D3+ATM are underestimated. Note that these two XSAPT methods differ only in the dispersion energy, and so this comparison underscores the fact that nonadditive dispersion is even more important in the larger HIV/indinavir complex than it is in the smaller intercalation complex, despite the latter being a prototypical example of π-stacking.

In summary, we have adapted CM5 charge embedding for use with the XSAPT methodology, finding that it improves both the accuracy and the efficiency of the method, as compared to our original ChElPG-based implementation. Where high-level *ab initio* benchmarks are available, interaction energies computed with CM5-based XSAPT are consistently a bit more accurate than ChElPG-based results, when the constituent molecules are charge-neutral. For systems involving ions (and especially ion pairs), however, the CM5-based version improves the accuracy considerably, mainly by removing some outliers where the ChElPG embedding charges adopt counter-intuitive values. This improvement is coupled to a dramatic reduction in the cost of CM5-based XSAPT, reducing the time required to compute charge derivatives $\partial q_a/\partial P_{\mu\nu}$ by a factor of 8.4 for the $C_{60}@C_{60}H_{28}$ complex. As a result, the charge-embedding terms no longer contribute appreciably to the cost of an XSAPT calculation, whereas previously these terms were a significant computational bottleneck.

CM5 (or other Hirshfeld-inspired atomic charge models[47]) will likely prove beneficial in other quantum chemistry methods that employ self-consistent charge embedding. For example, our group has recently reported a variational formulation of the generalized many-body expansion that underlies most fragment-based quantum chemistry approaches,[28] which employs fragment Fock matrices identical to those in Eq. (2). Similar Fock matrices also arise in our ChElPG-based formulation of Ewald summation for quantum mechanics/molecular mechanics (QM/MM) calculations under periodic boundary conditions.[37,38] Evaluation of the charge derivatives $\partial q_a/\partial P_{\mu\nu}$ proves to be a serious bottleneck in calculation of the QM/MM-Ewald energy gradient,[38] even when the QM region is relatively small, and work is underway to implement a CM5-based version of this method.

See supplementary material for errors in E_{int} for each individual system in the data sets.

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. HF-3c and PBEh-3c calculations were performed using the Turbomole program,[56] and all other calculations were performed using a locally modified version of Q-Chem.[57] All calculations were run at the Ohio Supercomputer Center under Project No. PAA-0003.[58] J.M.H. serves on the Board of Directors of Q-Chem, Inc.

REFERENCES

[1] M. Stiborova, J. Sejbal, L. Borek-Dohalska, D. Aimova, J. Poljakova, K. Forsterova, M. Rupertova, J. Wiesner, J. Hudecek, M. Wiessler, and E. Frei, "The anticancer drug ellipticine forms covalent DNA adducts, mediated by human cytochromes P450, through metabolism to 13-hydroxyellipticine and ellipticine N^2-oxide," Cancer Res. **64**, 8374–8380 (2004).

[2] 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).

[3] K. Fukuzawa, K. Kitaura, K. Nakata, T. Kaminuma, and T. Nakano, "Fragment molecular orbital study of the binding energy of ligands to the human estrogen receptor," Pure Appl. Chem. **75**, 2405–2410 (2003).

[4] 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.

[5] 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.

[6] T. Sawada, T. Hashimoto, H. Tokiwa, T. Suzuki, H. Nakano, H. Ishida, M. Kiso, and Y. Suzuki, "How does the FMO method help in studying viruses and their binding receptors?," in *The Fragment Molecular Orbital Method*, edited by D. Fedorov and K. Kitaura (Taylor & Francis, 2009), Chap. 9, pp. 193–215.

[7] 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.

[8] E. K. Kurbanov, H. R. Leverentz, D. G. Truhlar, and E. A. Amin, "Analysis of the errors in the electrostatically embedded many-body expansion of the energy and the correlation energy of Zn and Cd coordination complexes with five and six ligands and use of the analysis to develop a generally successful fragmentation strategy," J. Chem. Theory Comput. **9**, 2617–2628 (2013).

[9] 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).

[10] R. M. Parrish, D. F. Sitkoff, D. L. Cheney, and C. D. Sherrill, "The surprising importance of peptide bond contacts in drug–protein interactions," Chem. Eur. J. **23**, 7887–7890 (2017).

[11] 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.

[12] Y. Wang, J. Liu, J. Li, and X. He, "Fragment-based quantum mechanical calculation of protein–protein binding affinities," J. Comput. Chem. **39**, 1617–1628 (2018).

[13]J. Choi, H.-J. Kim, X. Jin, H. Lim, S. Kim, I.-S. Roh, H.-E. Kang, K. T. No, and H.-J. Sohn, "Application of the fragment molecular orbital method to discover novel natural products for prion disease," Sci. Rep. **8**, 13063 (2018).
[14]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).
[15]Y. Okiyama, C. Wantanabe, K. Fukuzawa, Y. Mochizuki, T. Nakano, and S. Tanaka, "Fragment molecular orbital calculations with implicit solvent based on the Poisson–Boltzmann equation: II. Protein and its ligand-binding system studies," J. Phys. Chem. B **123**, 957–973 (2019).
[16]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).
[17]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).
[18]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).
[19]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).
[20]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).
[21]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).
[22]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).
[23]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).
[24]K. Szalewicz, "Symmetry-adapted perturbation theory of intermolecular forces," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 254–272 (2012).
[25]E. G. Hohenstein and C. D. Sherrill, "Wavefunction methods for noncovalent interactions," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 304–326 (2012).
[26]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**, 234108-1–234108-9 (2008).
[27]R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell, and M. Head-Gordon, "Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals," J. Phys. Chem. A **111**, 8753–8765 (2007).
[28]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.
[29]H. L. Williams and C. F. Chabalowski, "Using Kohn-Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions," J. Phys. Chem. A **105**, 646–659 (2001).
[30]A. J. Misquitta and K. Szalewicz, "Intermolecular forces from asymptotically corrected density functional description of monomers," Chem. Phys. Lett. **357**, 301–306 (2002).
[31]A. J. Misquitta and K. Szalewicz, "Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers," J. Chem. Phys. **122**, 214109 (2005).
[32]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).
[33]G. Jansen, "Symmetry-adapted perturbation theory based on density functional theory for noncovalent interactions," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 127–144 (2014).
[34]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).
[35]C. M. Breneman and K. B. Wiberg, "Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis," J. Comput. Chem. **11**, 361–373 (1990).
[36]M. M. Francl and L. E. Chirlian, "The pluses and minuses of mapping atomic charges to electrostatic potentials," in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (Wiley VCH, New York, 2000), Chap. 1, Vol. 14, pp. 1–32.
[37]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).
[38]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).
[39]A. V. Marenich, S. V. Jerome, C. J. Cramer, and D. G. Truhlar, "Charge model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases," J. Chem. Theory Comput. **8**, 527–541 (2012).
[40]F. L. Hirshfeld, "Bonded-atom fragments for describing molecular charge densities," Theor. Chem. Acc. **44**, 129–138 (1977).
[41]E. R. Davidson and S. Chakravorty, "A test of the Hirshfeld definition of atomic charges and moments," Theor. Chem. Acc. **83**, 319–330 (1992).
[42]S. Dasgupta and J. M. Herbert, "Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3," J. Comput. Chem. **38**, 869–882 (2017).
[43]P. Jurečka, J. Šponer, J. Černý, and P. Hobza, "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs," Phys. Chem. Chem. Phys. **8**, 1985–1993 (2006).
[44]H. L. Williams, E. M. Mas, K. Szalewicz, and B. Jeziorski, "On the effectiveness of monomer-, dimer-, and bond-centered basis functions in calculations of intermolecular interaction energies," J. Chem. Phys. **103**, 7374–7391 (1995).
[45]J. Řezác, K. E. Riley, and P. Hobza, "S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures," J. Chem. Theory Comput. **7**, 2427–2438 (2011); Erratum, **10**, 1359–1360 (2014).
[46]J. Řezác and P. Hobza, "Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods," J. Chem. Theory Comput. **8**, 141–151 (2012).
[47]T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier, and P. W. Ayers, "Minimal basis iterative stockholder: Atoms in molecules for force-field development," J. Chem. Theory Comput. **12**, 3894–3912 (2016).
[48]K. U. Lao, R. Schäffer, G. Jansen, and J. M. Herbert, "Accurate description of intermolecular interactions involving ions using symmetry-adapted perturbation theory," J. Chem. Theory Comput. **11**, 2473–2486 (2015).
[49]R. M. Fogarty, R. Rowe, R. P. Matthews, M. T. Clough, C. R. Ashworth, A. Brandt, P. J. Corbett, R. G. Palgrave, E. F. Smith, R. A. Bourne, T. W. Chamberlain, P. B. J. Thompson, P. A. Hunt, and K. R. J. Lovelock, "Atomic charges of sulfur in ionic liquids: Experiments and calculations," Faraday Discuss. **206**, 183–201 (2018).
[50]R. M. Fogarty, R. P. Matthews, C. R. Ashworth, A. Brandt-Talbot, R. G. Palgrave, R. A. Bourne, T. V. Hoogerstraete, P. A. Hunt, and K. R. J. Lovelock, "Experimental validation of calculated atomic charges in ionic liquids," J. Chem. Phys. **148**, 193817 (2018).
[51]R. Sure and S. Grimme, "Comprehensive benchmark of association (free) energies of realistic host–guest complexes," J. Chem. Theory Comput. **11**, 3785–3801 (2015); Erratum, **11**, 5990 (2015).

[52]A. Benali, L. Shulenburger, N. A. Romero, J. Kim, and O. A. von Lilienfeld, "Application of diffusion Monte Carlo to materials dominated by van der Waals interactions," J. Chem. Theory Comput. **10**, 3417–3422 (2014).
[53]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).
[54]R. Sure and S. Grimme, "Corrected small basis set Hartree-Fock method for large systems," J. Comput. Chem. **34**, 1672–1685 (2013).
[55]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**, 054107-1–054107-19 (2015).
[56]F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigand, "Turbomole," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 91–100 (2014).
[57]Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kús, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, Jr., H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, P. A. Pieniazek, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, N. Sergueev, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, V. Vanovschi, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package," Mol. Phys. **113**, 184–215 (2015).
[58]See http://osc.edu/ark:/19495/f5s1ph73 for Ohio Supercomputer Center.