

Computation of host-guest binding free energies with a new quantum mechanics based mining minima algorithm

Cite as: J. Chem. Phys. 154, 104122 (2021); doi: 10.1063/5.0040759

Submitted: 15 December 2020 • Accepted: 11 February 2021 •

Published Online: 12 March 2021



Peng Xu,¹ Tosaporn Sattasathuchana,¹ Emilie Guidez,² Simon P. Webb,³ Kilinoelani Montgomery,² Hussna Yasini,² Iara F. M. Pedreira,² and Mark S. Gordon^{1,a)}

AFFILIATIONS

¹Department of Chemistry, Iowa State University, Ames, Iowa 50014, USA

²Department of Chemistry, University of Colorado Denver, Denver, Colorado 80204, USA

³VeraChem LLC, 12850 Middlebrook Rd. Ste 205, Germantown, Maryland 20874-5244, USA

Note: This paper is part of the JCP Special Collection in Honor of Women in Chemical Physics and Physical Chemistry.

^{a)}Author to whom correspondence should be addressed: mark@si.msg.chem.iastate.edu

ABSTRACT

A new method called QM-VM2 is presented that efficiently combines statistical mechanics with quantum mechanical (QM) energy potentials in order to calculate noncovalent binding free energies of host-guest systems. QM-VM2 efficiently couples the use of semi-empirical QM (SEQM) energies and geometry optimizations with an underlying molecular mechanics (MM) based conformational search, to find low SEQM energy minima, and allows for processing of these minima at higher levels of *ab initio* QM theory. A progressive geometry optimization scheme is introduced as a means to increase conformational sampling efficiency. The newly implemented QM-VM2 is used to compute the binding free energies of the host molecule cucurbit[7]uril and a set of 15 guest molecules. The results are presented along with comparisons to experimentally determined binding affinities. For the full set of 15 host-guest complexes, which have a range of formal charges from +1 to +3, SEQM-VM2 based binding free energies show poor correlation with experiment, whereas for the ten +1 complexes only, a significant correlation ($R^2 = 0.8$) is achieved. SEQM-VM2 generation of conformers followed by single-point *ab initio* QM calculations at the dispersion corrected restricted Hartree-Fock-D3(BJ) and TPSS-D3(BJ) levels of theory, as post-processing corrections, yields a reasonable correlation with experiment for the full set of host-guest complexes ($R^2 = 0.6$ and $R^2 = 0.7$, respectively) and an excellent correlation for the +1 formal charge set ($R^2 = 1.0$ and $R^2 = 0.9$, respectively), as long as a sufficiently large basis set (triple-zeta quality) is employed. The importance of the inclusion of configurational entropy, even at the MM level, for the achievement of good correlation with experiment was demonstrated by comparing the calculated ΔE values with experiment and finding a considerably poorer correlation with experiment than for the calculated free energy $\Delta E - T\Delta S$. For the complete set of host-guest systems with the range of formal charges, it was observed that the deviation of the predicted binding free energy from experiment correlates somewhat with the net charge of the systems. This observation leads to a simple empirical interpolation scheme to improve the linear regression of the full set.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0040759>

I. INTRODUCTION

Ever since the Nobel Prize was awarded to Cram, Lehn, and Pedersen^{1–4} in 1987 for their seminal work on host-guest supramolecular chemistry, host-guest chemical systems have been widely studied in basic research laboratories.^{5,6} They have now also been widely adopted as a means to utilize molecular recognition

mechanisms in various applied chemistry fields, including, but not limited to, drug development,^{7,8} materials sciences,^{9,10} analytical separation sciences,^{11,12} chemical pollutant cleanup technology,^{13,14} and the agrochemical industry.¹⁵ For example, many pharmaceutical drug candidates exhibit poor solubility and, therefore, poor bioavailability (the ability to reach the site of action unaltered), but their bioavailability can be improved by the formation of

inclusion complexes with host molecules such as cyclodextrins^{16,17} and cucurbiturils.¹⁸

Both the host–guest binding affinity strength and the structural nature of host–guest complexation (i.e., the guest molecule inside, partly inside, or outside the host cavity) control the effect on the physicochemical properties of the guest molecule; therefore, key data required for the optimization of host–guest complexes for specific chemical applications are the binding free energy and the most energetically favored host–guest structures. There have been concerted efforts to develop routinely usable accurate physics-based computational methods to predict these host–guest properties. If such a capability were available, many rounds of expensive chemical synthesis and experimental measurement, usually required by chemical research and development programs, could be avoided. Furthermore, there has been recent interest in accurate computational prediction of host–guest properties in the context of providing simplified models to aid development and refinement of computational protein–ligand binding affinity methods.^{19–23}

Two main challenges arise when computing host–guest binding free energies and structures: First, the guest molecule, depending on its size and flexibility, may adopt many energetically favorable conformations and orientations within the host cavity, and even more so if the host also exhibits some flexibility. This introduces a requirement for significant conformational searching, with a goal of finding low energy structures of the system. Such studies are computationally demanding. Notably, multiple distinct thermally accessible molecular structures result in an increase in conformational entropy and are, therefore, important to account for. In addition, if, upon binding, the geometries of predominant conformations of the guest molecule change, this can also change the configurational entropy compared to that of the free guest, affecting the total binding affinity.²⁴ Second, the host–guest noncovalent interaction potential is relatively weak, but, at the same time, highly complicated to model accurately, requiring treatment of energy contributions such as electrostatics (Coulomb), polarization, exchange repulsion, dispersion, charge transfer, and solvation.

Current computational approaches fall into two camps with respect to the interaction potential: classical molecular mechanics (MM) based and quantum mechanics (QM) based approaches. The fast turnaround of MM-based methods allows for significant conformational sampling, and computations can be performed on thousands of atoms, facilitating inclusion of explicit solvent molecules. However, while MM-based methods, e.g., free energy perturbation (FEP),^{25,26} MM Poisson–Boltzmann Surface Area (MMPBSA),²⁷ and attach–pull–release (APR), a combined docking and molecular dynamics (MD) based approach,²⁸ have, for some systems, proved capable of providing relative binding free energies that correlate well with experiment, consistency as well as accurate absolute binding free energies remain a challenge.^{29–31} Furthermore, the empirical force field potentials^{32,33} that these methods rely on are unlikely to have optimal parameters for an arbitrary system of interest,³⁴ due to the presence of empirically fitted parameters that are typically tied to specific molecule types, nor do they provide adequate descriptions of complex chemical interactions involving, for example, π -stacking, polarization, and charge transfer, limiting their applicability. In contrast, QM based potentials can naturally account for these complex interactions, and, in addition, the level of QM theory (method and basis set) can, in principle, be systematically improved to provide

the required accuracy. A significant difficulty, though, with applying *ab initio* QM (AIQM) potentials to the calculation of host–guest binding affinities is that their increased computational expense can preclude adequate conformational sampling. This has resulted in a tendency to (a) rely solely on computationally cheaper semi-empirical QM (SEQM) methods, e.g., dispersion corrected PM6,^{35–37} and (b) rely on a conformational search step that uses only MM methods, sometimes leading to a situation in which QM corrections (energy and/or geometry) are unable to recover from the poor quality of the provided MM structures. This has led to somewhat mixed results regarding the accuracy of QM predicted host–guest binding affinities compared to experiment.^{38,39} In the situation where guest molecules are relatively small and rigid and the necessary seed conformations can be intuited and generated “by hand,” the application of density functional theory (DFT) with a good quality basis set has resulted in predicted absolute binding affinities that correlate highly with experiment, providing a good proof of principle.⁴⁰ Such a manual approach to the generation of host–guest conformers, however, is not generally and routinely feasible, especially when the guest and/or host molecules have significant flexibility.

This paper presents a new approach to the calculation of host–guest binding free energies, called QM-VM2, which tightly integrates SEQM and AIQM potentials with the statistical mechanics based second-generation mining minima method M2.^{41,42} This was achieved by interfacing the VeraChem LLC implementation of the second-generation mining minima method, VM2, with the QM software package GAMESS.^{43–45} The MM-only version of VM2 has already been applied to the calculation of protein–ligand and host–guest free energies.^{46–50} The new QM-VM2 approach is designed to address the problems with current approaches to the calculation of host–guest binding affinities, which were outlined above. QM-VM2 efficiently couples the use of SEQM energies and geometry optimizations with an underlying MM-based conformational search, guiding the search toward conformers that have low SEQM energies, instead of low MM energy conformers. This new scheme also allows for processing of the conformers produced at higher levels of QM theory, and it needs no manual initial placement of guests in the host, providing an automated placement mechanism to seed conformational searching. The first demonstration of QM-VM2 presented here is to compute the binding free energies of the host cucurbit[7]uril and a set of 15 guest molecules²⁰ and to compare the results with experimentally determined binding affinities.

This paper is divided into the following sections: Sec. II is a theory section, where the QM-VM2 method is described in detail, Sec. III gives the computational details of the first application of QM-VM2, Sec. IV presents the results and discussion, and Sec. V provides the conclusions and future work.

II. THEORY

A. MM-VM2

The VeraChem mining minima (VM2) algorithm is an implementation of the second-generation Mining Minima (M2) method originally developed by Gilson *et al.*^{24,42,51–53} VM2 is an end-point approach, whereby the binding free energy of a host–guest complex

is computed as the difference between the standard chemical potentials of the bound complex (HG) and the free host (H) and guest (G) at constant volume: $\Delta F^0 = \mu_{HG}^0 - \mu_H^0 - \mu_G^0$. For host–guest systems, the volume change on binding is small, so the Helmholtz binding free energy obtained is a good approximation to the Gibbs binding free energy $\Delta F^0 \approx \Delta G^0$.^{24,54}

The classical statistical mechanics based standard chemical potential of a molecule in solution can be expressed as

$$\mu^0 = -RT \ln \left(\frac{8\pi^2}{C^0} Z \right), \quad (1)$$

where

$$Z = \int e^{-E(\mathbf{r})/RT} d\mathbf{r} \quad (2)$$

and

$$E(\mathbf{r}) = U(\mathbf{r}) + W(\mathbf{r}). \quad (3)$$

Here, Z is the configuration integral over all molecular conformations, C^0 is the standard concentration, which, combined with the factor of $8\pi^2$, accounts for the positional and orientational mobility of the free molecule at standard concentration, $E(\mathbf{r})$ is the energy comprising the potential energy $U(\mathbf{r})$ plus the solvation energy $W(\mathbf{r})$ as a function of internal coordinates, R is the ideal gas constant, and T is the absolute temperature.²⁴ The VM2 method approximates Z , an integral over all space, as the sum over local configuration integrals z_i for a manageable set of M local energy wells, which correspond to the low energy minima of the system,

$$Z \approx \sum_i^M z_i, \quad (4)$$

$$z_i \equiv \int e^{-E(\mathbf{r})/RT} d\mathbf{r}, \quad (5)$$

where $E(\mathbf{r})$ is again the energy as a function of internal coordinates, but the integration is restricted to the local energy well i . The standard chemical potential may now be conveniently expressed in terms of a sum over local standard chemical potentials μ_i^0 ,

$$\mu^0 \approx -RT \ln \left(\frac{8\pi^2}{C^0} \sum_i^M z_i \right) = -RT \ln \left(\sum_i^M e^{-\mu_i^0/RT} \right), \quad (6)$$

where

$$\mu_i^0 = -RT \ln \left(\frac{8\pi^2}{C^0} z_i \right). \quad (7)$$

The two key computational requirements of the VM2 algorithm then become the determination of the low energy minima of a system and the calculation of the local standard chemical potentials of these minima. The latter is calculated using an enhanced harmonic approximation method called HAMS, i.e., harmonic approximation with mode scanning.^{55,56}

$$\mu_i^0 = -RT \ln \left(\frac{8\pi^2}{C^0} z_i^{\text{HAMS}} e^{-E_{0,i}/RT} \right), \quad (8)$$

$$= E_{0,i} - RT \ln \left(\frac{8\pi^2}{C^0} z_i^{\text{HAMS}} \right). \quad (9)$$

Here, $E_{0,i}$, the energy at the bottom of the potential energy well i , is the leading term, and z_i^{HAMS} is calculated via the matrix of the energy second derivatives (Hessian) plus numerical integration of low energy modes to correct for anharmonicity. The determination of the low energy minima of a system is achieved by the use of an aggressive torsional mode-distort-minimize algorithm, a heavily modified version of the Tork search method,⁵⁷ which generates torsional modes via diagonalization of the MM-based dihedral angle Hessian matrix and is designed to repeatedly drive molecular conformations over high energy barriers and subsequently geometry optimize them to produce lower and lower energy minima.

In the MM-only based implementation of the VM2 algorithm (MM-VM2), the energy $E(\mathbf{r})$ is calculated only with classical methods. The potential energy $U(\mathbf{r})$ is obtained from empirical force fields such as the CHARMM General Force Field (CGenFF)³³ or a general Amber force field (GAFF)³² or CHARMM,⁵⁸ which contain bond-stretch, bond-angle, torsion, van der Waals, and Coulombic terms. During geometry optimizations and Hessian matrix involved steps, such as torsional mode generation, the solvation energy and energy derivative terms are calculated using the Generalized Born (GB) continuum model,^{59,60} and the more accurate Poisson–Boltzmann Surface Area (PBSA) method⁶¹ is applied as a final solvation energy correction $W(\mathbf{r})$ to the minima found.

Given the forgoing discussion, the basic MM-VM2 algorithm proceeds by searching for low energy conformations/minima of the system, and any repeat conformations found during the conformational search are discarded by a symmetry aware structural RMSD method.⁶² The local configuration integral [Eq. (8)] is calculated for the remaining minima, and then, the standard chemical potential is calculated according to Eqs. (1) and (4). In practice, this process is repeated iteratively until no new minima are found and the chemical potential converges within a given tolerance.

Upon convergence of a VM2 calculation, a probability p_i can be assigned to each of the local wells (or conformations), assuming a Boltzmann distribution,

$$p_i = \frac{e^{-\mu_i^0/RT}}{\sum_i^M e^{-\mu_i^0/RT}}. \quad (10)$$

In addition, again applying a harmonic oscillator (H.O.) approximation to the wells, an average energy of each well can be obtained using the equipartition theorem,

$$\langle E_i \rangle = E_{0,i} + N_{\text{int}} RT/2, \quad (11)$$

where N_{int} is the number of internal degrees of freedom. Equations (10) and (11) lead to an expression for the energy averaged over all wells,

$$\langle E \rangle = \sum_i^M p_i E_i = \sum_i^M p_i E_{0,i} + N_{\text{int}} RT/2. \quad (12)$$

This, in turn, allows a useful decomposition of the total chemical potential, providing expressions for the total configurational entropy as well as the entropy of each local well,⁶³

$$-TS^0 = \mu^0 - \langle E \rangle, \quad (13)$$

$$-TS_i = \mu_i^0 - \langle E_i \rangle. \quad (14)$$

The quantities in Eqs. (13) and (14) will be discussed further in Sec. IV D.

B. QM-VM2

The VM2 algorithm has now been interfaced with quantum mechanical (QM) methods, producing a new mining minima free energy method QM-VM2. In the QM-VM2 scheme [Fig. 1(a)], the semi-empirical QM (SEQM) based free energy is iterated until self-consistency, that is, until the change of the SEQM free energy falls within a predefined threshold. Once the SEQM free energy is converged, single point AIQM calculations may be performed to obtain a more accurate total free energy. As will be shown in Sec. IV, while the converged SEQM-VM2 calculation can alone yield reasonably good correlation with experiments, this post-processing step with *ab initio* QM methods can produce absolute binding free energies in excellent agreement with experiment, provided adequate basis sets are used.

The QM-VM2 conformational search [Fig. 1(b)], like that for MM-VM2, occurs through a torsional mode-distort-minimize process, with the torsional modes calculated via the MM-based dihedral angle Hessian matrix, and the distortion steps along these modes and initial minimization again using the MM-based potential. However, once an MM-based conformer is produced, it is then passed through an interface to the necessary GAMESS quantum chemistry package routines for SEQM energy or geometry optimization (note that the interface is actually general and not limited to SEQM, but rather can access any type of QM method implemented in GAMESS), and it is these SEQM energies and structures, when passed back from GAMESS through the interface to the VM2 drivers, that are used in energy cutoff decisions and to seed, if they are low in energy, the next round of conformational searches.

Once the search is complete, as in the case of MM-VM2, duplicate conformers are identified and discarded.⁶² The local chemical

potentials are then calculated for the remaining set, and from these, the total chemical potential is determined. The implementation of QM-VM2 allows the use of the MM potential to calculate the Hessian based and mode scanning terms in the HAMS based local configuration integrals, see Eq. (9),^{56,57} but with the QM based energy, e.g., $E_{0,i}^{SEQM}$, used as the leading term, i.e., used to adjust the bottom of the well,

$$\mu_i^0 = E_{0,i}^{SEQM} - RT \ln \left(\frac{8\pi^2}{C^0} z_i^{HAMS} \right). \quad (15)$$

Similarly, the QM based energies, e.g., $E_{0,i}^{SEQM}$, can be used in the average energy expressions, Eqs. (11) and (12), leading to a means of expressing the total MM-based configurational entropy and local well entropy for SEQM and QM adjusted energy wells.

The QM-VM2 implementation also allows for the use of an AIQM or SEQM Hessian based rigid rotor harmonic oscillator (RRHO) approximation, with the conformers produced by the search passed to the GAMESS Hessian and thermodynamic analysis routines to provide the required terms by the non-classical statistical mechanics RRHO based expression [i.e., zero point energy (ZPE), rotational, translational, and vibrational enthalpy (H) and entropy (S) terms] with an additional term to adjust from 1 atm pressure (P) to 1M standard concentration,^{38,64,65}

$$\mu_i^0 = E_{0,i}^{SEQM} + ZPE_i + H_i^{SEQM} + S_i^{SEQM} - RT \ln(P/RT), \quad (16)$$

$$\mu^0 \approx -RT \ln \left(\sum_i e^{-\mu_i^0/RT} \right). \quad (17)$$

The use of AIQM or SEQM Hessian based RRHO will usually be too computationally demanding to include inside the VM2

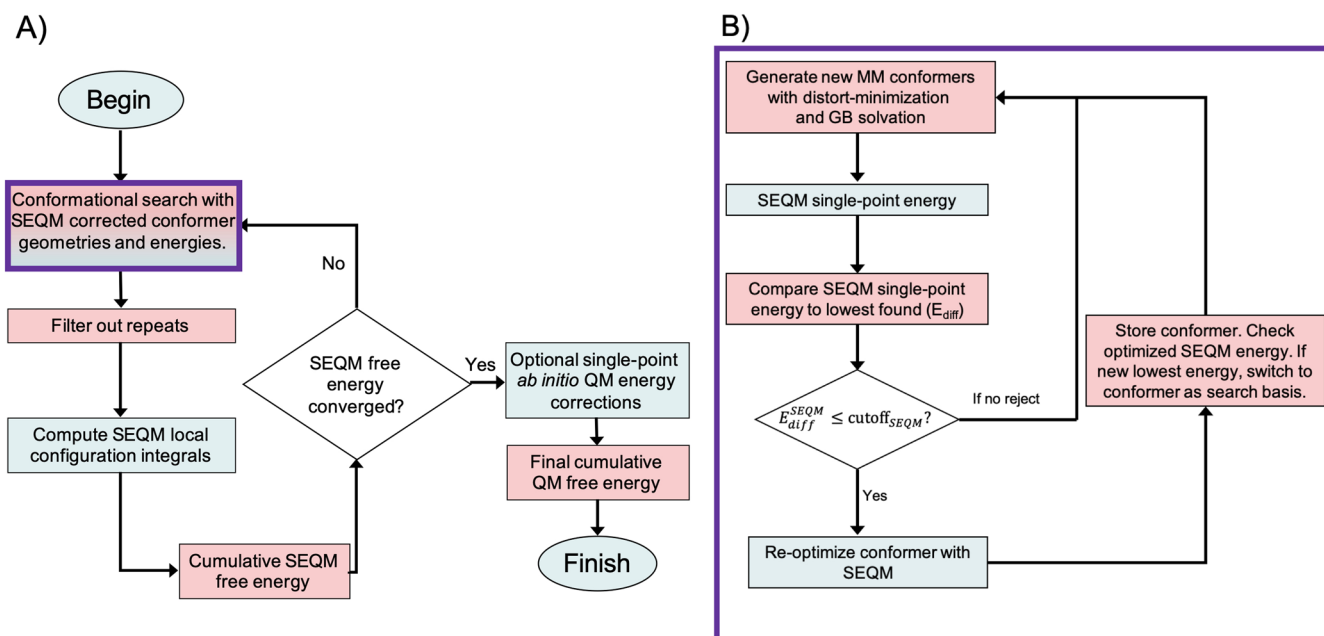


FIG. 1. QM-VM2 method. (a) Mining minima method. (b) Generation of SEQM corrected conformers [first step in (a), purple border]. Blue indicates QM calculations that are performed using GAMESS. Pink indicates calculations carried out with the VM2 software.

iterations and is more likely to be used for the post-processing of a limited set of conformers produced by a converged VM2 calculation, in an approach similar to the AIQM single-point energy correction post-processing described above. Note that while the classical formulation of VM2 results in Helmholtz free energies as described above, the QM thermodynamic analysis from GAMESS includes enthalpy terms and so provides the Gibbs free energy.

C. Progressive scheme (PGSS)

The most computationally expensive step of the conformational search procedure in QM-VM2 calculations is the SEQM geometry optimization of the MM-generated conformers. The MM-based mode-distort-minimize procedure, with subsequent PBSA solvation energy correction, takes on the order of several seconds or less per conformer. In contrast, the SEQM geometry optimization of a host-guest conformer can, for the systems presented in this study (see Fig. 3), take between 3 min and 20 min on a single compute processor (CPU) core, with a total of 800 minimizations attempted per VM2 iteration by default – see Sec. III. This is due, in part, to their significant size: the number of atoms for the set of host-guest complexes in this study ranges between 146 and 157 atoms, with corresponding basis function counts for SEQM of 416 and 442, respectively. It is also because the relatively flat potential energy surface encountered for such noncovalently bound systems results in the frequent need for more than one hundred steps to converge the geometry. Note that, in the conformational search scheme shown in Fig. 1(b), single-point SEQM energies of MM-generated conformers are used as a metric to decide whether to discard a conformer or further process it through SEQM geometry optimization, with the goal of reducing the number of geometry optimizations carried out. However, this is far from an ideal solution because some MM-generated conformers may have a high initial SEQM single-point energy, but then, after a SEQM geometry optimization, may become competitively low in energy. Therefore, the discarding of these conformers could slow down the overall convergence of the QM-VM2 calculation and even prevent the lowest energy conformers of the system from being found.

In order to address these problems, a partial optimization scheme, called progressive scheme (PGSS), has been developed to generate SEQM corrected conformers (Fig. 2). The key idea of the progressive scheme is straightforward: large gradient vector components indicate steep descent along the trajectory down to the local minima, more likely leading to a low-energy potential well. Therefore, apart from the energy, the energy gradient is also used as a means to determine if a geometry optimization calculation should be stopped and the conformer should be discarded or continued until completion. In this scheme, four new control parameters have been introduced for the partial SEQM optimization of MM-generated conformers (Table I).

The first parameter, *npgstep*, the number of PGSS partial optimization steps allowed, is typically set to 5. If the geometry has converged within *npgstep* optimization steps, the energy difference (Ediff) between the current conformer and the lowest energy conformer found so far is examined. The current conformer is kept when Ediff is within a predefined threshold *ecutpg* (the default is 5.0 kcal/mol); otherwise, it is discarded. If the geometry has not converged but Ediff is less than the *ecutpg* value, suggesting that a

potential low-lying local minimum is nearby, further optimization is carried out until completion. If Ediff is greater than the predefined threshold, the energy gradient is analyzed. The PGSS process is repeated if the largest gradient component is larger than a threshold value (*grad_{cut}*); otherwise, the process is stopped, and the partially optimized structure is discarded. The default value for the number of PGSS iterations, *npgopt*, is 3. In essence, this scheme introduces another mechanism for selecting low-energy minima contributing to the configuration integral with an early checking and intervention capability that should avoid unnecessary time-consuming SEQM geometry optimizations.

Among the four control parameters for the progressive scheme summarized in Table I, *npgstep*, *ecutpg*, *npgopt*, and *grad_{cut}*, the first three are single-valued parameters that can be altered in the VM2 input by users. The gradient threshold, *grad_{cut}*, on the other hand, is not a user-specified input parameter; rather, it changes as a function of the PGSS iteration number *n*, $(0.5)^n \times 10^{-3}$ hartree/bohr. This is because the magnitude of the gradient vector gradually decreases when falling toward the bottom of the potential well, assuming a harmonic shaped potential well. By the third progressive iteration, the *grad_{cut}* value becomes 0.25×10^{-3} , which is not too far from the typical gradient cutoff value (0.0001) in electronic structure codes. Hence, the default value for *npgopt* is set to 3, and if users set *npgopt* larger than 4, from the fourth progressive iteration onwards, the formula for *grad_{cut}* will be disregarded and the value 0.0001 will be used.

D. Coarse-grained parallelism for QM-VM2

Like the MM-VM2 implementation, QM-VM2 uses coarse-grained parallelization of the conformational search, based on the Message Passing Interface (MPI) library, to speed up the turnaround of calculations. A conformational search occurs every VM2 iteration, see Figs. 1 and 2, and each conformational search initiates hundreds of mode-distort-minimize calculations, i.e., by default, 400 single torsional mode-distort-minimize calculations followed by 400 random combinations of mode-pairs-distort-minimize calculations. In the current work, these mode-distort-minimize calculations were distributed across all MPI processes available to each QM-VM2 calculation (for most of the calculations presented here, this was 24 MPI processes), with each MPI process itself carrying out multiple “serial” executions of the procedure outlined in Fig. 2. The parallel algorithm alternates between two schemes, one in which all MPI processes are seeded with the same initial conformer, and, through global communication, all MPI processes are periodically reseeded with the current lowest energy conformation found so far. The other scheme seeds each MPI process with a different conformation, each one taken from the full set of conformations produced so far in the VM2 calculation, and carries out independent rounds of the procedure shown in Fig. 2. This approach is designed to introduce structural diversity, which, on the basis of an extensive experience with host-guest and protein-ligand conformational searches at the VM2-MM-only level, helps the search avoid becoming trapped and stalling in local energy wells, before the lowest energy conformers of a system are found.

As a single unified QM-VM2 executable is built by linking the GAMESS and VM2 compiled object files, the VM2 drivers can access and utilize the generalized distributed data interface (GDDI)⁶⁶ in

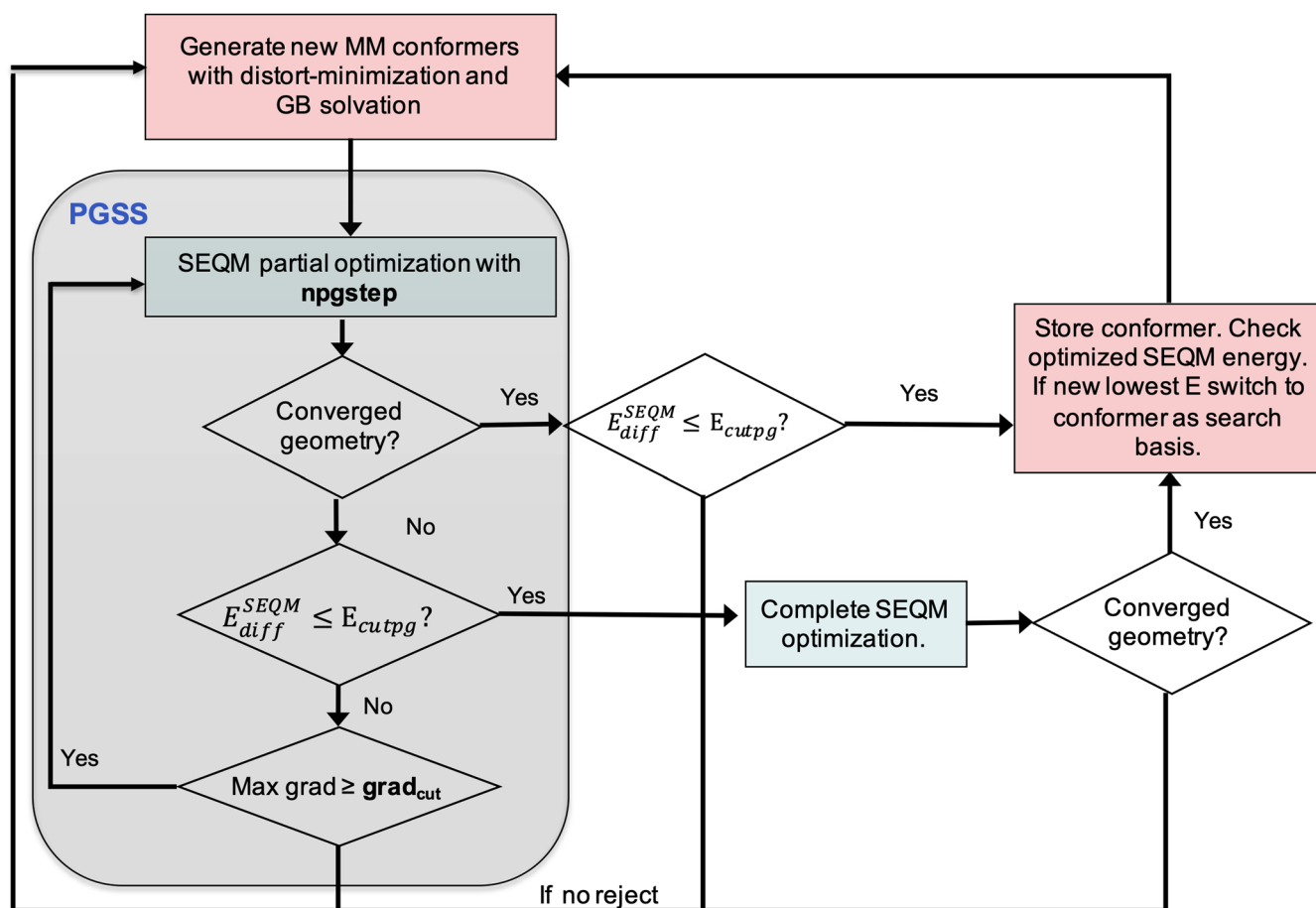


FIG. 2. Progressive optimization scheme (PGSS). This scheme can replace the standard scheme used in Fig. 1(b) to generate SEQM corrected conformers. Blue indicates QM calculations performed using GAMESS. Pink indicates calculations performed with the VM2 software.

GAMESS, which is built on top of MPI. This allows not only the use of a team of processes, where each individual process carries out a “serial” execution of mode-distort-minimize calculations, as described above, but also the use of multiple teams, each with multiple processes, providing for a combined coarse grained-fine grained multi-level parallel approach; i.e., parallelized SEQM or

AIQM energy or energy-gradient calculations are carried out by the already distributed (across teams) mode-distort-minimize calculations. While not used in the current work, the latter multi-level parallelism has been applied in other projects and allows for faster turnaround if the computational resources are available; it also allows for application to larger molecular systems.

TABLE I. A summary of the new control variables for PGSS.

PGSS control parameters	Default	Function
<i>npgstep</i>	Five steps	The number of PGSS optimization steps
<i>ecutpg</i>	5 kcal/mol	Energy cutoff when comparing with the lowest-found local minimum
<i>npgopt</i>	Three iterations	The maximum number of PGSS iterations
<i>grad_{cut}</i>	$(0.5)^n \times 10^{-3}$ hartree/bohr, where <i>n</i> is the progressive iteration number	Threshold value with which the largest gradient vector component is compared

III. COMPUTATIONAL DETAILS

The Statistical Assessment of the Modeling of Proteins and Ligands (SAMPLs) challenges were founded to provide prospective validation for computational tools in rational drug design.^{19–23} Participants in these blinded challenges are tasked with predicting molecular properties, such as binding affinities, for given chemical systems. The participants then submit their predictions to the organizers, who then compare the results of each submission to as yet unpublished high quality experimental data, generated especially for each challenge. There have been seven challenges to date, and SAMPL3 through SAMPL7 have all included a host–guest binding affinity component,^{19–23} with the rationale that these smaller simplified systems provide for faster computational turnaround than the larger protein–ligand systems, but still provide a means of testing and validating many components of the computational models employed. The datasets for completed SAMPLn challenges are also useful for retrospective studies, given that they comprise the curated sets of molecular system coordinate files supplied to the challenge participants, along with the corresponding high-quality experimental data, such as binding affinities. In this study, the host molecule cucurbit[7]uril (CB7) and associated guest molecules from the SAMPL4 challenge²⁰ were used for an initial application of the QM-VM2 method and the proposed progressive scheme (Fig. 3), and the QM-VM2 calculated binding free energies of these host–guest systems are compared against the published SAMPL4 experimental binding affinities.

Initial starting structures of the host–guest complexes were generated automatically by translation of the center of geometry (COG) of the guest molecules to the COG of the host, followed by

the removal of any resulting steric clashes via an initial MM-based geometry optimization that damps very large energy-gradient values [the COG is calculated the same way as the center of mass (COM), but with all the masses set to one].

For the MM-based parts of the QM-VM2 calculation, the parameters (bond, angle, torsion, van der Waals, etc.) for the potential energy were assigned according to the CHARMM force field,⁵⁸ using the Discovery Studio Visualizer (Biovia), and atomic partial charges were assigned using the VCharge software (Ver-aChem LLC).⁶⁷ For the MM-based mode-distort-minimize procedure, the MM solvation energy was included using the generalized Born (GB) continuum model,^{59,61} and the Poisson–Boltzmann Surface Area (PBSA) method⁶² was applied to provide a more accurate MM-based single-point solvation energy correction. The torsional mode-distort-minimize based conformational search carried out during each VM2 iteration included 400 single-mode distortion based searches and 400 searches in which random combinations of pairs of modes were used to generate distortions. At least, four VM2 iterations were carried out in all mining minima calculations, and typically, most of the free energy lowering occurs within these iterations. All final VM2 free energies were converged to an energy difference <0.3 kcal/mol compared to the previous VM2 iteration.

Both the SEQM and post-processing AIQM calculations were carried out with the electronic structure package, GAMESS.^{42–44} The SEQM calculations employed the third order density functional tight binding method, DFTB3,⁶⁸ in combination with the D3 Grimme dispersion correction⁶⁹ modified with the Becke–Johnson (BJ) damping [DFTB3-D3(BJ)].^{70–73} The set of interatomic interaction parameters used (3OB)^{74,75} was specifically designed for DFTB3,

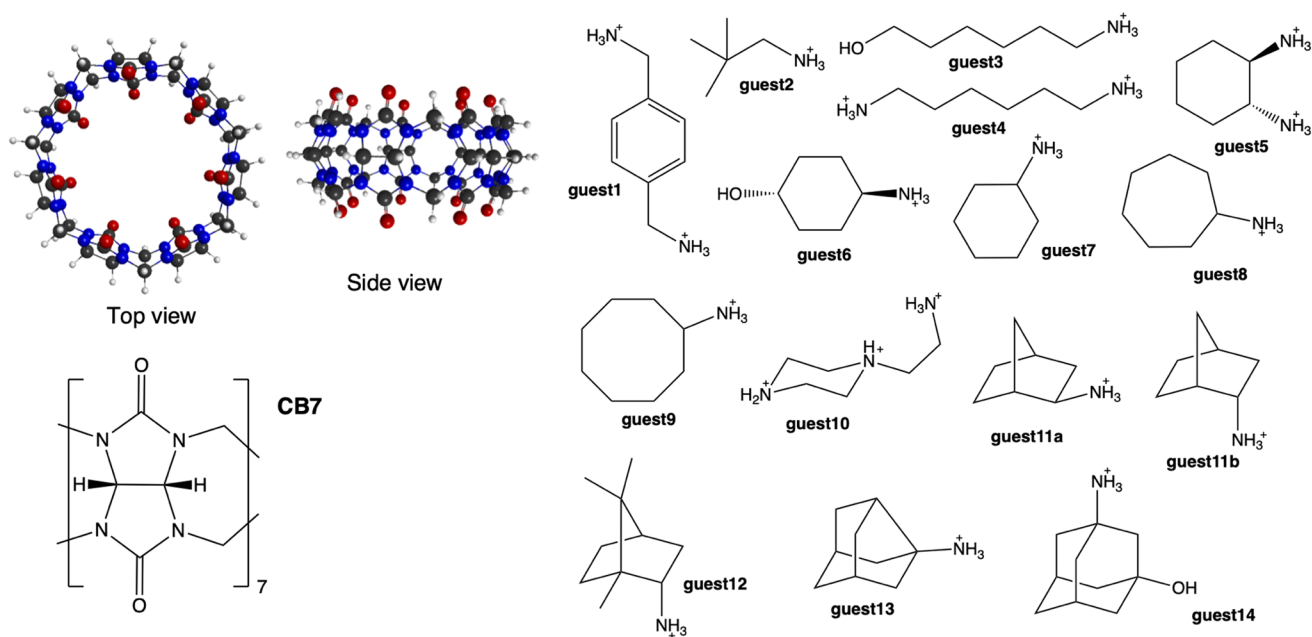


FIG. 3. Structure of host molecule cucurbit[7]uril (CB7) and the guest molecules that comprise the host–guest systems in this study. Dark gray: carbon, blue: nitrogen, red: oxygen, and white: hydrogen.

with improvements mainly in non-covalent bonding. SEQM geometry optimizations were considered converged when the largest component of the gradient was less than 0.0001 hartree/bohr, and, in addition, the root-mean-square gradient was less than a third of this maximum component tolerance.

Two different implicit solvation models were considered to account for solvation effects in the SEQM and AIQM based calculations: the Solvation Model Density (SMD) method⁷⁶ and the conductor-like polarizable continuum model (C-PCM).^{77–80} The cavitation and dispersion terms were excluded from the latter model in the current work. It was found that the SMD model, both with the default van der Waals radii values⁸¹ and with a set of adjusted radii,⁸² showed improved binding free energies relative to C-PCM for a few complexes, but over the full set of complexes was worse; therefore, the C-PCM results are presented here. For C-PCM, the default density of tesserae on the cavity surface (nstall = 60) is used, as increasing this value only marginally improved the binding free energies and the overall correlation, with significantly longer computational times. The simplified united atomic (SUHF) radii⁸³ were used for the generation of the C-PCM cavity holding the solute, as van der Waals radii led to increased over-binding of the host and guests and substantially lowered the correlation with experiment (see the [supplementary material](#) for more details).

For the AIQM post-processing step [see Fig. 1(a)], the 30 conformers with the lowest DFTB3-D3(BJ)/PCM energy were processed with single-point energy calculations using TPSS-D3(BJ)⁸⁴ and HF-D3(BJ) with the def2-TZVP basis set^{85–88} and C-PCM implicit solvation. For a subset of seven of the host–guest systems, second-order perturbation theory with density fitting (RI-MP2),^{89,90} again with C-PCM and the def2-TZVP basis set, was applied to the 30 lowest DFTB3-D3(BJ)/C-PCM energy conformers. RI-MP2 was applied only to a reduced host–guest set due to the computational expense resulting from the use of the def2-TZVP basis set (~3700 basis functions) for these calculations. The reduced set comprised the systems with the most strongly and weakly bound guests experimentally (guests 3 and 13), a middle range binder (guest 7), and the four highly charged systems (guests 1, 4, 5, and 10), i.e., formal charge greater than +1.

Local configuration integrals were calculated using MM for the Hessian based and mode scanning (HAMS) terms,^{55,56} but with the SEQM and AIQM energies used as the leading term, i.e., used to adjust the bottom of the well [e.g., Eq. (15)], and the absolute temperature was set as 300 K.

Default values are used for the control parameters of the progressive scheme (Fig. 2); that is, the number of progressive iterations and the number of SEQM optimization steps were set to 3 and 5, respectively, and an energy cutoff of 5 kcal/mol was used.

In the remaining sections of the paper, the use of the C-PCM solvation model is implied. References to the SEQM-VM2 step and the post-processing step will be separated by //, and values from the SEQM method are enclosed in square brackets. For example, the notation TPSS-D3(BJ)/Def2-TZVP//[DFTB3-D3(BJ)]-VM2 indicates that the SEQM-VM2 calculation was performed at the DFTB3-D3(BJ) level of theory and the conformers were post-processed at the TPSS-D3(BJ)/Def2-TZVP level of theory.

IV. RESULTS AND DISCUSSION

This section is divided into the following subsections: (a) the SEQM-VM2 binding free energies and the SEQM-VM2 plus AIQM post-processing binding free energies are presented for the host CB7 and 15 guest molecules (Fig. 3) and compared to experimental binding affinities; (b) the performance of the PGSS method, developed to avoid unproductive geometry optimizations during the VM2 conformational search, is evaluated; (c) the effect of inclusion of MM-based entropy terms [Eq. (15)] on the accuracy of the calculated binding free energies compared to experiment is examined; and (d) an interpolation scheme is proposed to examine and correct for apparent systematic error in the current implementation of QM-VM2 when applied to highly charged systems.

While the ultimate goal of the QM-VM2 methodology, and its ongoing development, is to consistently predict, with good accuracy, the absolute binding free energies of receptor–ligand complexes, the ability to predict even just the ranking of a set of ligands with respect to how strongly they bind a particular receptor is also highly sought after by medicinal chemists and other applied scientists. Therefore, examined here are not only errors of the predicted absolute binding free energies but also errors of the predicted *relative* binding free energies and linear correlation metrics between the predicted values and experimental values.

Specific error metrics presented are mean signed error (MSE), mean absolute error (MAE), and root mean-squared error (RMSE). Two definitions of RMSE are presented: $RMSE_o$ and $RMSE_r$. The former evaluates the accuracy of the absolute binding free energies and is given as

$$RMSE_o = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[\Delta G_i^{\text{exp}} - \Delta G_i^{\text{calc}} - \frac{1}{n} \sum_{j=1}^n \left(\Delta G_j^{\text{exp}} - \Delta G_j^{\text{calc}} \right) \right]^2}, \quad (18)$$

where the second sum is the MSE. The latter, $RMSE_r$, assesses the relative binding free energies by calculating the differences among all pairs of host–guest systems subject to this study,

$$RMSE_r = \sqrt{\frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n \left[\left(\Delta G_j^{\text{calc}} - \Delta G_i^{\text{calc}} \right) - \left(\Delta G_j^{\text{exp}} - \Delta G_i^{\text{exp}} \right) \right]^2}. \quad (19)$$

In Eqs. (18) and (19), n is the number of measurements, ΔG^{calc} and ΔG^{exp} are the calculated and experimental binding affinities, respectively, and just for the purposes of these definitions, ΔG^{calc}

$\equiv \Delta F^{\text{calc}}$. These RMSE definitions are those used for the assessment of Sampl4 challenge results,²⁰ allowing direct comparison between QM-VM2 and the participating methods.

Two methods to examine linear correlation are used: one, linear regression analysis via the linear regression slope and the Pearson coefficient of determination, R^2 , where a linear regression slope of 1.0 and R^2 value of 1.0 indicate a perfect correlation, and two, the Kendall rank correlation coefficient, τ ,⁹¹ which is a measure of the strength of the association between two sets of ranked data, in this case, experimental and calculated binding free energies. Kendall τ ranges between 0 and 1, with 0 being no correlation and 1 being a perfect correlation. In addition to the metrics mentioned above, the y-intercept values of the linear regression lines are also recorded as they are used in an interpolation scheme discussed later in Sec. IV D.

A. SEQM-VM2 and SEQM-VM2 with *ab initio* QM post-processing

Table II presents the [DFTB3-D3(BJ)]-VM2 calculated binding free energies of host CB7 and the guest molecules shown in Fig. 3, along with the corresponding experimental binding free energies. A MSE of -7.2 kcal/mol for the full set of host-guest systems, Table IV, indicates substantial over-binding in the DFTB3-D3(BJ)-VM2 predicted values compared to the experimental values, with very large over-binding (MSE = -12.8 kcal/mol) for the four highly charged guests (i.e., charge $>+1$) and smaller over-binding (MSE = -5.2 kcal/mol) for the remaining guests with charge $+1$. Given the significant mismatch between MSEs for the two groups, the lack of linear correlation between the computed and experimental values ($R^2 = 0.05$) for the full set is not surprising – see Fig. 4. Separate regression plots for the charge $=+1$ and charge $>+1$ sets (see Fig. 5), however, show very good correlation for the charge $=+1$ set, $R^2 = 0.8$, and significantly worse correlation, $R^2 = 0.3$, for the charge $>+1$ set, albeit with a small sample size. The high level of correlation with experiment for the charge $=+1$ set is encouraging, given that only a SEQM level of theory was employed. Furthermore, the results

TABLE II. Binding free energies of the Sampl4 CB7 set from DFTB3-D3(BJ)-VM2 calculations. The numbering scheme is given in Fig. 3.

Complex	Positive charge	Binding free energy (kcal/mol)	
		Exp.	SEQM-VM2
1	2	−9.9	−23.43
2	1	−9.6	−14.24
3	1	−6.6	−11.87
4	2	−8.4	−18.71
5	2	−8.5	−21.74
6	1	−7.9	−12.58
7	1	−10.1	−14.22
8	1	−11.8	−15.95
9	1	−12.6	−17.66
10	3	−7.9	−21.98
11a	1	−11.1	−15.97
11b	1	−11.1	−16.18
12	1	−13.3	−19.46
13	1	−14.1	−19.08
14	1	−11.6	−19.92

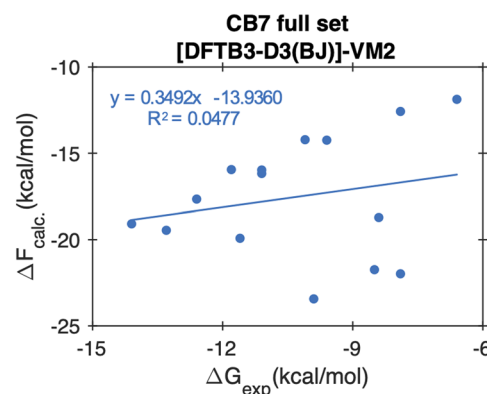


FIG. 4. Comparison of binding affinities (kcal/mol) of the Sampl4 CB7 set, experiments vs calculations obtained from [DFTB3-D3(BJ)]-VM2. Also presented is the linear regression line.

suggest that highly charged ($>+1$) systems are a particular challenge to DFTB3-D3(BJ)/PCM, with Fig. 5 indicating the possibility of systematic error that could be adjusted for. This is further examined in Sec. IV D.

Since DFTB is a SEQM method, an obvious next step toward more accurate binding free energy predictions is to apply a more sophisticated AIQM treatment. This was carried out according to the post-processing scheme indicated in Fig. 1(a); i.e., AIQM single-point energies, $E_{0,i}^{AIQM}$, at [DFTB3-D3(BJ)]-VM2 generated conformer geometries, which replace the $E_{0,i}^{SEQM}$ term in Eq. (15). The SEQM level of theory employed, DFTB3-D3(BJ)/PCM, will not necessarily agree with the AIQM level of theory as to which conformers are the lowest in energy; therefore, a significant number of [DFTB3-D3(BJ)]-VM2 conformers must be included in the AIQM based Boltzmann-averaged binding free energies. After including the top 10, 20, and 30 conformers for post-processing of the complete set, it was observed that the binding free energies converge with the inclusion of 30 conformers. Therefore, the 30 lowest energy [DFTB3-D3(BJ)]-VM2 conformers were post-processed for each host-guest system. Given this large total number of post-processing AIQM

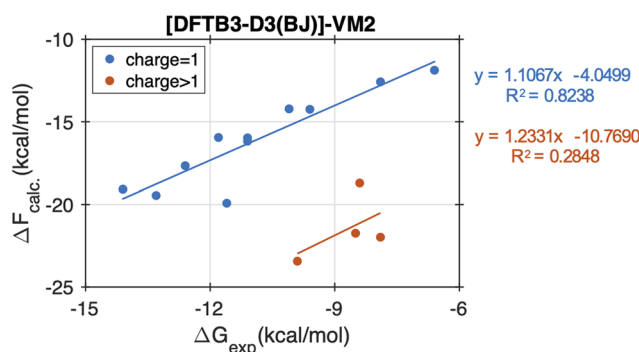


FIG. 5. Comparison of binding affinities (kcal/mol) of Sampl4 CB7 dataset experiments vs calculations obtained from [DFTB3-D3(BJ)]-VM2, separately for the set of charge $=+1$ (blue) and the set of charge $>+1$ (orange).

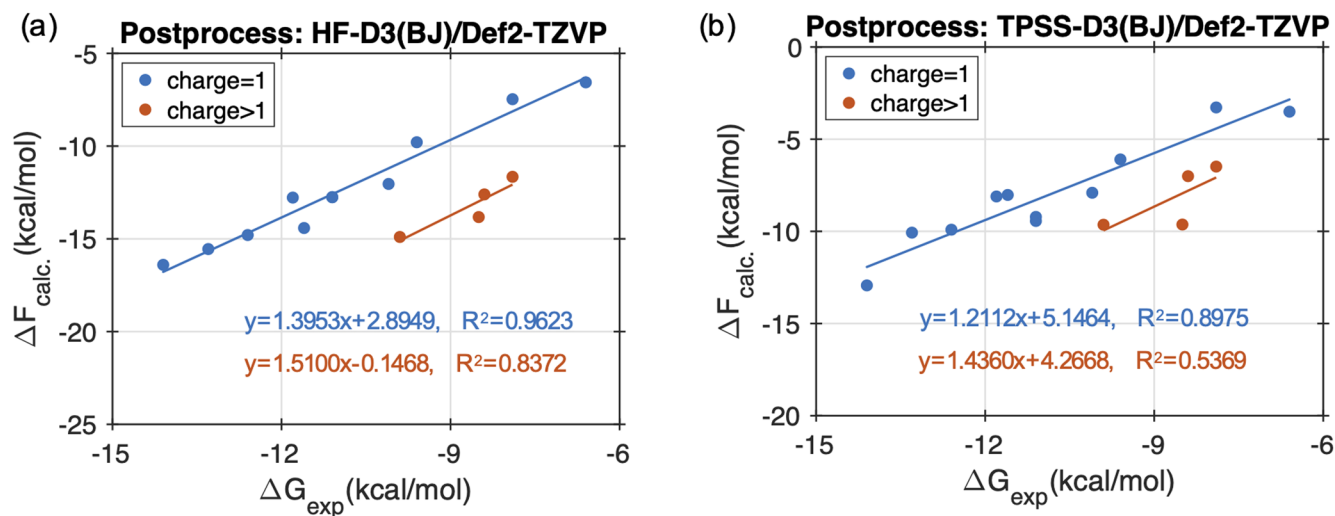
TABLE III. Binding affinities of the Samp14 CB7 set from TPSS-D3(BJ)/def2-tzvp/[DFTB3-D3(BJ)]-VM2 and HF-D3(BJ)/def2-tzvp/[DFTB3-D3(BJ)]-VM2 calculations.

Complex	Positive charge	Binding free energy (kcal/mol)			
		Experimental	HF-D3(BJ)	TPSS-D3(BJ)	RI-MP2
1	2	−9.9	−14.90	−9.64	−17.89
2	1	−9.6	−9.79	−6.10	
3	1	−6.6	−6.56	−3.51	−8.89
4	2	−8.4	−12.60	−7.01	−13.04
5	2	−8.5	−13.83	−9.63	−14.63
6	1	−7.9	−7.47	−3.28	
7	1	−10.1	−12.04	−7.91	−13.57
8	1	−11.8	−12.78	−8.11	
9	1	−12.6	−14.79	−9.91	
10	3	−7.9	−11.66	−6.49	−8.84
11a	1	−11.1	−12.76	−9.44	
11b	1	−11.1	−12.75	−9.22	
12	1	−13.3	−15.55	−10.07	
13	1	−14.1	−16.40	−12.93	−20.45
14	1	−11.6	−14.42	−8.03	

calculations, the computationally efficient HF and DFT methods were chosen, augmented with Grimme dispersion corrections (−D). Second-order perturbation theory with density fitting (RI-MP2) was also considered. Both a double-zeta + polarization basis set, 6-31G(d,p), and a triple-zeta + polarization basis set, Def2-TZVP, were explored.

The binding affinities obtained from post-processing calculations using the double-zeta basis set (see [supplementary material](#) for details) clearly demonstrate that the double-zeta + polarization basis set overestimates the binding affinities, regardless of the

choice of theory. For example, the predicted binding affinity of CB7-guest2 is −19.5 kcal/mol and −18.3 kcal/mol calculated with HF-D3(BJ)/6-31G(d,p) and TPSS-D3(BJ)/6-31G(d,p), respectively, while the experimental result is only −9.6 kcal/mol. Adding the diffuse functions on the heavy elements can significantly improve the error metrics, as illustrated by the post-process of SEQM-VM2 at HF-D3(BJ)/6-31+G(d,p) (Table S5). However, the predicted binding affinities at HF-D3(BJ)/6-31+G(d,p) are still overestimated by ~5 kcal/mol to 12 kcal/mol. On the other hand, [Table III](#) clearly shows that the triple-zeta + polarization basis set,

**FIG. 6.** Comparison of binding free energy (kcal/mol) of the Samp14 CB7 set obtained from experiments vs calculated at (a) HF-D3(BJ)/def2-TZVP/[DFTB3-D3(BJ)]-VM2 and (b) TPSS-D3(BJ)/def2-TZVP/[DFTB3-D3(BJ)]-VM2 for the training set with charge =+1 (blue) and charge >+1 (orange).

Def2-TZVP, can produce absolute binding affinities in very good agreement with experiment for both HF-D3(BJ) and TPSS-D3(BJ). It is concluded that the double-zeta + polarization basis set is inadequate for predicting binding free energies of the SAMPL4 CB7 host-guest complexes, and a triple-zeta + polarization basis set or better is necessary.

Table III, Fig. 6, and Table IV present the predicted binding free energies obtained by HF-D3(BJ) and TPSS-D3(BJ) post-processing with the triple-zeta + polarization basis set Def2-TZVP, along with linear correlation and error metrics with respect to experiment, for the complete set of host-guest complexes shown in Fig. 3. Table III also presents RI-MP2/Def2-TZVP binding free energies for a

subset of these complexes – see Sec. III for details. It may be seen from Table IV that HF-D3(BJ), with a MSE of -2.3 kcal/mol for the full set of complexes, tends to overestimate the binding free energies compared to experiment [though not nearly as much as DFTB3-D3(BJ)], whereas TPSS-D3(BJ), with a full set MSE of 2.2 kcal/mol, underestimates the binding free energies. Examining the charge $=+1$ and charge >1 sets separately, it was seen that HF-D3(BJ) yields very good results for the charge $=+1$ set, with a small over binding, MSE = -1.4 kcal/mol, but noticeably larger over binding errors for the charge >1 set (MSE = -4.6 kcal/mol). TPSS-D3(BJ), on the other hand, shows the opposite trend with clear under binding, MSE = 2.8 kcal/mol, for the charge $=+1$ set and very small errors over

TABLE IV. Error metrics of QM-VM2 binding free energy predictions compared to experiment for the SAMPL4 CB7 dataset.

Host-guest systems	Level of theory	MSE	MAE	RMSE _o	RMSE _r	τ	Slope	Intercept	R ²
Current QM-VM2 work									
Full set	[DFTB3-D3(BJ)]-VM2	-7.2	7.2	3.6	5.2	0.2	0.3	-13.9	0.0
	RHF-D3(BJ)//VM2 ^a	-2.3	2.3	1.7	2.5	0.7	1.0	-2.2	0.6
	TPSS-D3(BJ)//VM2 ^a	2.2	2.4	1.4	2.1	0.6	0.9	1.6	0.7
SAMPL4 participant results for comparison									
	OST ^{b,c}			1.9	2.8		1.4		0.8
	RRHO ^{b,d}			2.5	3.7		1.8		0.8
	Enthalpy ^{b,e}			2.7	4.0		1.6		0.7
	M2 ^{b,f}			3.4	4.5		2.0		0.7
	EES ^{b,g}			3.4	5.0		1.9		0.7
	SIE+HB ^{b,h}			1.8	2.6		0.2		0.6
	BAR ^{b,i}			2.2	3.3		1.3		0.6
	FEP ^{b,j}			3.9	5.7		1.8		0.6
	QM/M2 ^{b,k}			3.0	4.5		0.7		0.2
Current QM-VM2 work									
Charge $=+1$	[DFTB3-D3(BJ)]-VM2	-5.2	5.2	1.1	1.7	0.7	1.1	-4.0	0.8
	RHF-D3(BJ)//VM2 ^a	-1.4	1.5	1.0	1.5	1.0	1.4	2.9	1.0
	TPSS-D3(BJ)//VM2 ^a	2.8	2.8	1.0	1.4	0.8	1.2	5.1	0.9
Charge $>+1$	[DFTB3-D3(BJ)]-VM2	-12.8	12.8	1.5	2.4	0.3	1.4	-10.8	0.3
	RHF-D3(BJ)//VM2 ^a	-4.6	4.6	0.6	1.0	1.0	1.5	-0.1	0.8
	TPSS-D3(BJ)//VM2 ^a	0.5	1.0	1.0	1.7	1.0	1.4	4.3	0.5

^aResults in this work from AIQM single-point energy post-processing of the first 30 [DFTB3-D3(BJ)]-VM2 conformers. The Def2-TZVP basis set is employed for both RHF-D3(BJ) and TPSS-D3(BJ).

^bResults presented in the SAMPL4 overview article.²⁰ See references therein for additional methodological details to those shown directly below.

^cOrthogonal space tempering (OST) carried out with the GAFF/AM1-BCC energy model and a modified TIP3P water potential.⁹³

^dConformational sampling was performed manually. Final QM free energy model: PW6B95-D3/def2-QZVP(-g,-f)/COSMO-RS//TPSS-D3-cosmo/def2-TZVP/HF-3c(freq.) and includes rigid rotor harmonic oscillator (RRHO) approximation derived terms.⁴⁰

^eDirect calculation of enthalpy change from long MD simulations of end states, carried out with the GAFF/AM1-BCC energy model and TIP3P/TIP3P-Ew water potential.

^fMolecular mechanics mining minima (M2) calculations³⁹ carried out with the CHARMM force field³⁸ and Vcharge charges.⁶⁷

^gExpanded ensemble simulations (EESs) using MD with the GAFF/AM1-BCC energy model and TIP3P water potential.⁹⁴

^hSolvated interaction energy including hydrogen bonding terms (SIE+HB) using Wilma docking and the GAFF/AM1-BCC energy model with the biotechnology research institute boundary element method for solving Poisson equation (BRI BEM) continuum solvation.^{95,96}

ⁱBennett acceptance ratio (BAR)⁹⁷ using MD with the atomic multipole optimized energetics for biomolecular simulation (AMOEBA)⁹⁸ force field.

^jFree energy perturbation (FEP)²⁵ using metadynamics with the GAFF/AM1-BCC energy model and TIP3P water potential.

^kQM corrected M2 based on PM6-DH+ geometry optimizations of molecular mechanics M2 conformers⁵ and includes COSMO continuum solvation and RRHO terms.^{39,99}

and under binding, $MSE = 0.5$ kcal/mol and $MAE = 1.0$ kcal/mol, for the charge >1 set.

Both HF-D3(BJ) and TPSS-D3(BJ) yield much improved linear correlations with experiment for the full set of host–guest complexes [$R^2 = 0.6$ for HF-D3(BJ) and $R^2 = 0.7$ for TPSS-D3(BJ)], compared to a correlation of essentially zero for the SEQM-VM2 calculations for the full set, see Table IV and Figs. 4 and 6. This significant improvement in the ability to describe the whole set of host–guest systems, with their full range of charge states, $+1$ to $+3$, is likely indicative of the importance of large basis sets as well as improvement in the underlying QM method. On the other hand, if the linear correlations are again examined for the separate charge $=+1$ and charge >1 sets (see Fig. 6), the correlations are further improved: for the charge $=+1$ set, R^2 for HF-D3(BJ) and TPSS-D3(BJ) is 1.0 and 0.8, respectively; for the charge >1 set, R^2 for HF-D3(BJ) and TPSS-D3(BJ) is 0.8 and 0.5, respectively [also note the improvement of these R^2 values over the equivalent [DFTB3-D3(BJ)]-VM2 values, Fig. 5]. These results, together with the mean error data discussed above, suggest that some systematic error remains in the relative description of the charge $=+1$ set and charge >1 set. While in the long term, efforts will be made to find and address the underlying issues with the physics of the model (e.g., a possible source of this error is the use of a continuum solvation model); in the short term, simple scaling methods that can correct for these errors have been explored, as discussed in Sec. IV D.

The results for the RI-MP2/Def2-TZVP calculations for a subset of seven of the host–guest systems (see Table III) show that RI-MP2 more closely matches the behavior of HF-D3(BJ) than TPSS-D3(BJ). In fact, except for complex 10, RI-MP2 exhibits over binding larger than HF-D3(BJ). This is not too surprising, given that MP2 is known for its tendency to over bind noncovalent complexes due to its incomplete treatment of electron correlation.³² Furthermore, even though the RI approximation greatly reduces the computational cost of MP2, at little cost in accuracy, it is still considerably more costly and memory-intensive than HF-D and DFT-D. Considering all of the reported error metrics in Table IV, the HF-D and DFT-D AIQM methods, with an adequately large basis set, perform similarly well over the full set of host–guest systems. In terms of computational expense, the SCF convergence of the HF-D3(BJ)/Def2-TZVP can be up to a factor of two times faster than TPSS-D3(BJ)/Def2-TZVP. An additional advantage of HF-D over DFT-D is that HF-D avoids possible double counting of dispersion via the $-D$ correction, whereas the accuracy of DFT is functional dependent. Consequentially, the HF-D3(BJ) method with a triple-zeta + polarization quality basis set is recommended for post-processing of [DFTB3-D3(BJ)]-VM2 calculations, when considering accuracy, reliability, and computational cost.

For selected SAMPL4 challenge participant methods, the error metrics $RMSE_o$, $RMSE_r$, slope, and R^2 for calculated binding affinities with respect to experimental values are also presented in Table IV to allow comparison with the current QM-VM2 work. A number of SAMPL4 participants submitted multiple entries with the same underlying methodology; in these cases, only the best performing entry is included. Furthermore, SAMPL4 methods achieving an $R^2 < 0.2$ are excluded. An overall comparison between the error metrics for the QM-VM2 approaches restricted Hartree–Fock (RHF)-D3(BJ)//VM2 and TPSS-D3(BJ)//VM2, applied to the full set of host–guest systems, and the SAMPL4 methods shows that the

QM-VM2 results are competitive with respect to R^2 and that they outperform all SAMPL4 methods with respect to $RMSE_o$, $RMSE_r$, and slope. A more detailed comparison now follows, though, for the sake of brevity only, TPSS-D3(BJ)//VM2 values are discussed, as the RHF-D3(BJ)//VM2 trends are quite similar.

Table IV shows that the best performing SAMPL4 method, OST, achieves an impressive $R^2 = 0.8$, which is slightly better than the $R^2 = 0.7$ achieved by TPSS-D3(BJ)//VM2, but its $RMSE_o$, $RMSE_r$, and slope values (1.9 kcal/mol, 2.8 kcal/mol, and 1.4, respectively) do not compare favorably with the corresponding TPSS-D3(BJ)//VM2 values of 1.4 kcal/mol, 2.1 kcal/mol, and 0.9, respectively. The SAMPL4 method labeled RRHO in Table IV also achieves $R^2 = 0.8$, but its $RMSE_o$, $RMSE_r$, and slope values of 2.5 kcal/mol, 3.7 kcal/mol, and 1.8, respectively, are worse than the OST values with respect to comparison to the TPSS-D3(BJ)//VM2 values. In addition, RRHO, a QM based approach, in contrast to the automatic initial guest molecule placement and conformational search employed by QM-VM2, relies on manual placement and conformational search. Notably, while a manual search is feasible for relatively small and rigid guests or ligand molecules, it will quickly become unmanageable and ineffective with even a modest increase in guest size and flexibility. The SAMPL4 methods achieving R^2 values of 0.7 and 0.6, with the exception of SIE+HB (see Table IV), show significantly worse $RMSE_o$, $RMSE_r$, and slope values than TPSS-D3(BJ)//VM2; for example, the method labeled enthalpy achieves $R^2 = 0.7$, $RMSE_o = 2.7$ kcal/mol, $RMSE_r = 4.0$ kcal/mol, and slope $= 1.6$. The SIE+HB method ($R^2 = 0.6$) has somewhat more competitive $RMSE_o$ and $RMSE_r$ values of 1.8 kcal/mol and 2.6 kcal/mol, but a poor slope value of 0.2.

B. Progressive scheme (PGSS)

The performance of the newly proposed PGSS method was assessed by comparison of QM-VM2 calculations using the conventional conformational search scheme and QM-VM2 calculations using the PGSS enhanced conformational search. In terms of accuracy, the differences in the binding affinities between the conventional and PGSS QM-VM2 schemes are negligibly small, less than 0.1 kcal/mol, as shown in the supplementary material. In terms of efficiency, several measures are examined in Tables V and S6 in the supplementary material. The CB7-guest10 and CB7-guest14 complexes (see Fig. 3) were chosen as representatives of the charge $>+1$ and charge $=+1$ host–guest sets, respectively, to demonstrate the performance of PGSS. All of the SEQM-VM2 calculations were carried out on an Intel® Xeon® CPU E5-2695 v2 (2.40 GHz) with 24 cores in one node. Several noteworthy observations are made. First, the total CPU time and the number of VM2 iterations may be larger for PGSS QM-VM2 runs. This is because the number of conformers explored during the conformational search step can be considerably larger for PGSS QM-VM2. On the other hand, the time spent per generated conformer is still shorter for the PGSS scheme. In addition, the number of conformers that contribute to the Boltzmann distribution is slightly larger for the PGSS scheme. In other words, within the same amount of time, the PGSS scheme can sample a larger conformational space. Thus, it is concluded that even with the current default values of PGSS control parameters, the PGSS scheme provides a powerful boost for the conformational search in

TABLE V. Comparison of timing and efficiency metrics between conventional QM-VM2 and PGSS QM-VM2 for complexes CB7-guest10 and CB7-guest14.

	CB7-guest10		CB7-guest14	
	Conventional	PGSS	Conventional	PGSS
No. of VM2 iterations	4	4	4	6
Total CPU time (s)	105 201.3	107 685.5	94 502.1	126 523.2
Time/iteration (h)	7.3	7.5	6.6	5.9
No. of Boltzmann average samples	111	132	80	82
No. of conformers	1565	1655	1814	2889
Time/conformer (s)	67.2	65.1	52.1	43.8

QM-VM2 and can be further enhanced by optimizing the PGSS control parameters.

C. Entropy effect

As mentioned in Sec. III, in the current QM-VM2 study, local configuration integrals were calculated using an MM-based enhanced harmonic approximation, but with the SEQM and AIQM energies used as the leading term, to adjust the bottom of the well – see Eq. (15). Furthermore, Sec. II B describes how the SEQM and AIQM based average energy $\langle E \rangle$ and the total configurational entropy $-TS^0$ can be backed out of the total chemical potential by the same energy adjustment to Eq. (12), followed by application of Eq. (13), which allows one to examine the importance of inclusion of configurational entropy, even at the MM level (note that the solvent entropy contribution is implicitly included in the continuum solvent model).

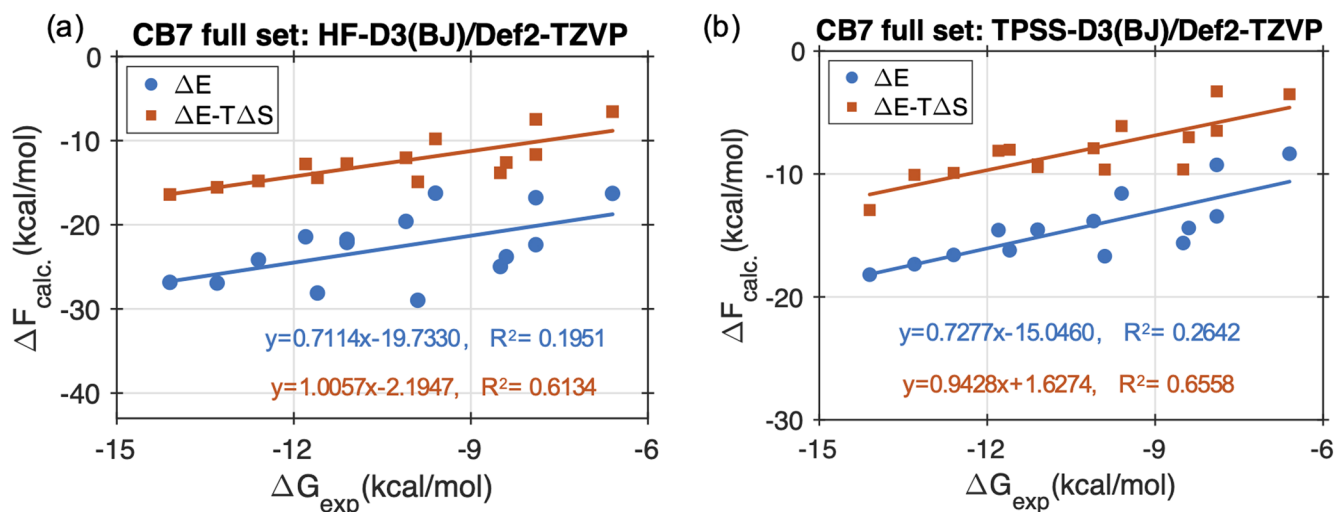
The importance of the configurational entropy to the correlation with experimental values is clearly demonstrated in Fig. 7. Using only the energy term (blue dots labeled ΔE) results in a weak

linear correlation of the computed and experimental binding affinities [$R^2 = 0.2$ for HF-D3(BJ) and $R^2 = 0.3$ for TPSS-D3(BJ) with the Def2-TZVP basis set]. Including the MM-based entropy term (orange dots labeled $\Delta E - T\Delta S$) significantly improves the linear correlation of the computed and experimental binding affinities; i.e., R^2 increases from 0.2 to 0.6 for HF-D3(BJ) and from 0.3 to 0.7 for TPSS-D3(BJ).

To assess the validity of using the MM-based entropy term, it is noted that previous MM-based mining-minima studies of host-guest and protein–ligand systems observed an approximately linear relationship between energy and entropy contributions, that is, the large negative energy contribution is canceled partly by a proportional entropy penalty.^{42,46} Such an approximately linear relationship can also be observed between the QM energy and MM entropy (see the [supplementary material](#)).

D. Interpolation

While [DFTB3-D3(BJ)]-VM2 yields good correlation with experiment for the set of charge = +1 host–guest systems in this study

**FIG. 7.** Comparison of the binding free energy (kcal/mol) obtained from experiments vs calculated with (a) HF-D3(BJ)/def2-tzvp/[DFTB3-D3(BJ)]-VM2 and (b) TPSS-D3(BJ)/def2-tzvp/[DFTB3-D3(BJ)]-VM2 with the inclusion of various components of free energy.

(see Table IV and Fig. 5, slope = 1.1, close to the ideal value of 1.0, $R^2 = 0.8$), as described in Sec. IV A, the linear regression R^2 values drop considerably when highly charged systems are included, suggesting a systematic error in the relative treatment of charge = +1 and charge >1 systems (see Figs. 4 and 5). Therefore, an interpolation method is proposed to adjust computed binding free energies of highly charged systems to account for systematic error and, thereby, improve the accuracy of predictions without additional computational cost. The scheme requires a low charge (e.g., +1) training set, for which QM-VM2 provides good correlation with known experimental binding free energies.

The impetus for investigating whether errors exhibited for the highly charge systems are sufficiently systematic to allow an interpolation scheme to be useful was the observation that the deviation from experiment in the calculated binding free energies for the highly charged systems in the current study is approximately a multiple of the average deviation for charge = +1 systems. The idea of the proposed empirical adjustment scheme, then, is straightforward: adjust the linear regression line for the charge = +1 system as much to the ideal as is possible, and then, for the highly charged species, apply this same correction scaled by the charge. This leads to the following equation for the interpolation of the binding affinities from DFTB3-D3(BJ)-VM2 calculations:

$$Y_{scaled} = \frac{Y_{calc} - b}{a} \times q, \quad (20)$$

where Y_{scaled} and Y_{calc} are the scaled and the computed binding affinities, respectively, a and b are the slope and the y-intercept of the regression line of the charge = +1 set, and q is the charge of the molecule. After introducing the empirical adjustment to the full set, one can observe significant improvement of the predicted binding affinities and their correlation with experiment. The R^2 value for $q > +1$ obtained for the adjusted SEQM-VM2 results increased dramatically, from 0.005 to 0.558, as illustrated in Fig. 8.

For the post-processed binding free energies, the correlation with experiment for the charge >+1 set is remarkably good—see Fig. 6 and Table V. However, the number of data points for the highly charged systems in the Sampl4 CB7 set is so limited that good correlation can be coincidental. Applying the same empirical adjustment introduced above [Eq. (20)] does not produce improved correlation. Instead of multiplying by the charges of the systems, dividing by the charges yields much better correlation,

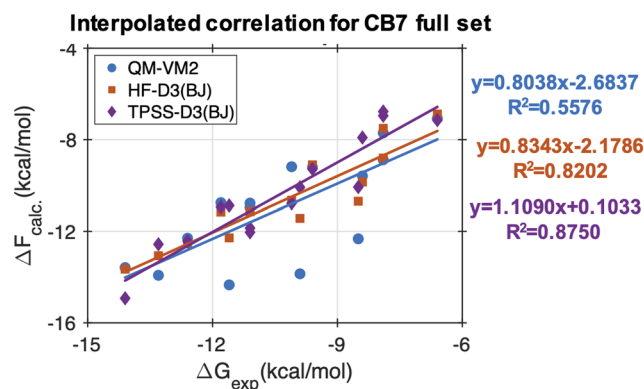


FIG. 8. Interpolated correlation for the Sampl4 CB7 full set. The binding free energies computed at various levels of theory are compared with the experiment. Blue dots, labeled QM-VM2, represent the interpolated predicted values from [DFTB3-D3(BJ)]-VM2. Those in orange and purple are the values obtained from the interpolated post-processing values from HF-D3(BJ)/Def2-TZVP and TPSS-D3(BJ)/Def2-TZVP, respectively.

$$Y_{scaled} = \frac{Y_{calc} - b}{a \times q}, \quad (21)$$

where Y_{scaled} and Y_{calc} are the scaled and the computed binding affinity at the post-processing step, respectively, and a and b are the slope and the y-intercept of the regression of charge = +1 system at the same level of theory, respectively. This change of the interpolation formulation suggests that the nature of the systematic error may have changed for the AIQM binding free energies relative to those for the SEQM binding energies. After application of the interpolation scheme, the linear correlation with experiment for both HF-D3(BJ) and TPSS-D3(BJ), for the full set of host–guest systems, improved significantly – see Table VI and Fig. 8. For HF-D3(BJ), R^2 increased from 0.6 to 0.8, and for TPSS-D3(BJ), R^2 increased from 0.7 to 0.9. Furthermore, Table VI shows that all of the error metrics improved substantially. Since this interpolation scheme is an empirical approach, it is not surprising that a significant change in the level of theory may change the nature of any systematic error and, therefore, the form of the best scaling, e.g., from multiplication by charge to dividing by charge.

TABLE VI. Error metrics of interpolated QM-VM2 binding free energy predictions compared to experiment for the SAMPL4 CB7 dataset. The pre-interpolation values are presented in parentheses.

	Level of theory	MSE	MAE	RMSE _o	RMSE _r	τ	Slope	Intercept	R^2
Interpolation of full set	[DFTB3-D3(BJ)]-VM2	−0.7 (−7.2)	1.2 (7.2)	1.6 (3.6)	2.3 (5.2)	0.6 (0.2)	0.8 (0.3)	−2.7 (−13.9)	0.6 (0.0)
	RHF-D3(BJ) ^a	−0.4 (−2.3)	0.7 (2.3)	0.8 (1.7)	1.2 (2.5)	0.8 (0.7)	0.8 (1.0)	−2.2 (−2.2)	0.8 (0.6)
	TPSS-D3(BJ) ^a	0.0 (2.2)	0.7 (2.4)	0.8 (1.4)	1.2 (2.1)	0.9 (0.6)	1.1 (0.9)	0.1 (1.6)	0.9 (0.7)

^aDef2-TZVP basis set is employed for both HF-D3(BJ) and TPSS-D3(BJ).

The proposed interpolation approach demonstrates a simple way to improve the prediction of binding affinities of a dataset when it contains both +1 and highly charged guests. The underlying assumption made here is that the set of charge =+1 guests yields reasonably good linear correlation. While the approach works well within the small dataset presented here, extensive testing of this procedure with large and varied datasets is required to establish general applicability. If found to be broadly applicable, the interpolation approach suggested here could be a practical way of achieving good predictions, even for highly charged systems that are particularly challenging for binding free energy calculations.

V. CONCLUSIONS AND FUTURE WORK

QM-VM2 is an approach that efficiently combines statistical mechanics with quantum mechanical energy potentials in order to calculate noncovalent binding free energies of receptor–ligand systems. The method efficiently couples the use of SEQM energies and geometry optimizations with an underlying MM-based conformational search and allows for processing of the conformers produced at higher levels of QM theory. A progressive scheme for conformer geometry optimizations is introduced in this work as a means to boost conformational sampling efficiency by recognizing that a steep descent on the potential energy surface implies a deep potential well and the gradient at that point can be used as a screening metric for accepting/rejecting conformers.

This first application of QM-VM2 computed the binding free energies of the host molecule cucurbit[7]uril and a set of 15 guest molecules. The results are presented here along with comparisons to experimentally determined binding affinities. SEQM-VM2 based binding free energies do not show good correlation with experiment for the full set of host–guest complexes, which includes highly charged systems (+2 and +3), whereas for just the +1 systems, a significant correlation ($R^2 = 0.8$) is achieved. SEQM-VM2 generation of conformers followed by single-point AIQM calculations, as post-processing corrections, yields good binding affinities and good correlation with experiment, even for the full set of systems, as long as a sufficiently large basis set (at least, triple-zeta quality) is employed.

The importance of the inclusion of configurational entropy, even at the MM level, to the achievement of good correlation with experiment was demonstrated by comparing ΔE values with experiment and finding considerably poorer correlation with experiment than for $\Delta E - T\Delta S$. For the complete set of host–guest systems with various charges, it was observed that the deviation of the predicted binding free energy from experiment correlates with the net charge of the systems to some extent. Thus, a simple empirical interpolation scheme was proposed to improve the linear regression of the full set.

While this work demonstrates that SEQM-VM2 with AIQM post-processing is a viable approach for predicting absolute binding free energies efficiently (with the proposed PGSS scheme) and accurately (with interpolation), there are several aspects of this method that can be further improved. For the energy model, a continuum solvation model, C-PCM, was employed due to its relatively low cost, and cavitation and dispersion solvation terms were not included. Future work could include these additional terms as a possible route to improved accuracy of the solvation treatment. In addition, to

address possible inadequacies with continuum solvation treatments of important localized interactions, especially for highly charged solutes, explicit solvent molecules represented as a polarizable model potential based on quantum mechanics, e.g., the effective fragment potential,^{100–106} could provide detailed interactions such as hydrogen bonding, without much additional computational cost. The configurational entropy contribution in the current study is obtained at the MM level. It would be interesting to see how the binding affinities and the correlation would be affected if a SEQM or AIQM based entropy term is used.

In the progressive scheme, *ecutpg*, the energy cutoff when comparing with the lowest-found local minimum, is taken to be a single-valued, user-specified parameter. An optimal value of this parameter that maximizes the benefits of a progressive scheme can be difficult to set even with extensive knowledge or experience for the system. Expressing *ecutpg* as a function of system-dependent parameters may improve the performance of the progressive scheme.

SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for preliminary calculations using various SEQM and *ab initio* QM methods for two conformers of each complex reported in Sec. I. In Sec. II, the binding free energies using two implicit solvation models are compared. Validation for using MM-based entropy contribution for the calculation of binding free energies is provided in Sec. III. Basis set effects are investigated in Sec. IV. A brief discussion of the rank 1 conformers (with the largest Boltzmann weight) is given in Sec. V. The performance of the progressive scheme is presented in Sec. VI, and the Cartesian coordinates for the rank 1 conformer of the SAMPL4 CB7 complexes are given in Sec. VII.

DEDICATION

This paper is dedicated to Professor Rosalind Franklin and Professor Patricia Thiel. Professor Franklin's research in x-ray crystallography paved the way for the determination of the structure of DNA and directly led to the discovery of the double helix. Professor Thiel, who passed away too soon in September 2020, was a leader in the field of quasicrystals, an excellent teacher and colleague, and a dear friend.

ACKNOWLEDGMENTS

This work was supported by the National Institute of General Medical Sciences of the NIH (Grant No. R44GM109679 to S.P.W.). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of General Medical Sciences. This work was also supported by the University of Colorado Denver UROP and EURECA programs. The authors thank Lawrence Stewart for automating the process of submission of 209 RI-MP2 GAMESS calculations to the AWS cloud and Michael Potter for helpful discussions regarding this work. S.P.W. declares an equity interest in VeraChem LLC.

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its [supplementary material](#).

REFERENCES

- ¹D. J. Cram, *Science* **240**, 760 (1988).
- ²C. J. Pedersen, *Angew. Chem., Int. Ed. Engl.* **27**, 1021 (1988).
- ³J.-M. Lehn, *J. Inclusion Phenom.* **6**, 351 (1988).
- ⁴J.-M. Lehn, *Angew. Chem., Int. Ed. Engl.* **29**, 1304 (1990).
- ⁵M. V. Rekharsky and Y. Inoue, *Chem. Rev.* **98**, 1875 (1998).
- ⁶E. Masson, X. Ling, R. Joseph, L. Kyeremeh-Mensah, and X. Lu, *RSC Adv.* **2**, 1213 (2012).
- ⁷X. Ma and Y. Zhao, *Chem. Rev.* **115**, 7794 (2015).
- ⁸H. Cai, Y.-L. Huang, and D. Li, *Coord. Chem. Rev.* **378**, 207 (2019).
- ⁹B. V. K. J. Schmidt and C. Barner-Kowollik, *Angew. Chem., Int. Ed.* **56**, 8350 (2017).
- ¹⁰X. J. Loh, *Mater. Horiz.* **1**, 185 (2014).
- ¹¹G. K. E. Scriba, *Chromatographia* **75**, 815 (2012).
- ¹²X. Zhang, Y. Zhang, and D. W. Armstrong, in *Comprehensive Chirality*, edited by E. M. Carreira and H. Yamamoto (Elsevier, Amsterdam, 2012), pp. 177–199.
- ¹³S. Erdemir, M. Bahadir, and M. Yilmaz, *J. Hazard. Mater.* **168**, 1170 (2009).
- ¹⁴T. Aksoy, S. Erdemir, H. B. Yildiz, and M. Yilmaz, *Water, Air, Soil Pollut.* **223**, 4129 (2012).
- ¹⁵V. J. Smith, N. M. Rougier, R. H. de Rossi, M. R. Caira, E. I. Buján, M. A. Fernández, and S. A. Bourne, *Carbohydr. Res.* **344**, 2388 (2009).
- ¹⁶K. Uekama, F. Hirayama, and T. Irie, *Chem. Rev.* **98**, 2045 (1998).
- ¹⁷B. Gidwani and A. Vyas, *BioMed Res. Int.* **2015**, 198268.
- ¹⁸S. J. Barrow, S. Kaser, M. J. Rowland, J. del Barrio, and O. A. Scherman, *Chem. Rev.* **115**, 12320 (2015).
- ¹⁹H. S. Muddana, C. Daniel Varnado, C. W. Bielawski, A. R. Urbach, L. Isaacs, M. T. Geballe, and M. K. Gilson, *J. Comput.-Aided Mol. Des.* **26**, 475 (2012).
- ²⁰H. S. Muddana, A. T. Fenley, D. L. Mobley, and M. K. Gilson, *J. Comput.-Aided Mol. Des.* **28**, 305 (2014).
- ²¹J. Yin, N. M. Henriksen, D. R. Slochower, M. R. Shirts, M. W. Chiu, D. L. Mobley, and M. K. Gilson, *J. Comput.-Aided Mol. Des.* **31**, 1 (2017).
- ²²A. Rizzi, S. Murkli, J. N. McNeill, W. Yao, M. Sullivan, M. K. Gilson, M. W. Chiu, L. Isaacs, B. C. Gibb, D. L. Mobley, and J. D. Chodera, *J. Comput.-Aided Mol. Des.* **32**, 937 (2018).
- ²³M. Amezcua, L. El Khoury, and D. L. Mobley, “SAMPL₇ host–guest challenge overview: Assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations,” *J. Comput.-Aided Mol. Des.* **35**, 1 (2021).
- ²⁴M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, *Biophys. J.* **72**, 1047 (1997).
- ²⁵P. Kollman, *Chem. Rev.* **93**, 2395 (1993).
- ²⁶P. Mikulskis, D. Cioloboc, M. Andrejic, S. Khare, J. Brorsson, S. Genheden, R. A. Mata, P. Soderhjelm, and U. Ryde, *J. Comput.-Aided Mol. Des.* **28**, 375 (2014).
- ²⁷N. Homeyer and H. Gohlke, *Mol. Inf.* **31**, 114 (2012).
- ²⁸J. Yin, N. M. Henriksen, D. R. Slochower, and M. K. Gilson, *J. Comput.-Aided Mol. Des.* **31**, 133 (2017).
- ²⁹Y.-W. Hsiao and P. Söderhjelm, *J. Comput.-Aided Mol. Des.* **28**, 443 (2014).
- ³⁰S. Bhakat and P. Söderhjelm, *J. Comput.-Aided Mol. Des.* **31**, 119 (2017).
- ³¹R. G. Coleman, T. Sterling, and D. R. Weiss, *J. Comput.-Aided Mol. Des.* **28**, 201 (2014).
- ³²J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *J. Comput. Chem.* **25**, 1157 (2004).
- ³³K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. MacKerell, *J. Comput. Chem.* **31**, 671 (2010).
- ³⁴N. M. Henriksen and M. K. Gilson, *J. Chem. Theory Comput.* **13**, 4253 (2017).
- ³⁵J. Řezáč, J. Fanfrlík, D. Salahub, and P. Hobza, *J. Chem. Theory Comput.* **5**, 1749 (2009).
- ³⁶J. P. Stewart, *J. Mol. Model.* **13**, 1173 (2007).
- ³⁷J. Řezáč and P. Hobza, *Chem. Phys. Lett.* **506**, 286 (2011).
- ³⁸H. S. Muddana and M. K. Gilson, *J. Chem. Theory Comput.* **8**, 2023 (2012).
- ³⁹H. S. Muddana, J. Yin, N. V. Sapra, A. T. Fenley, and M. K. Gilson, *J. Comput.-Aided Mol. Des.* **28**, 463 (2014).
- ⁴⁰R. Sure, J. Antony, and S. Grimme, *J. Phys. Chem. B* **118**, 3431 (2014).
- ⁴¹M. S. Head, J. A. Given, and M. K. Gilson, *J. Phys. Chem. A* **101**, 1609 (1997).
- ⁴²W. Chen, C.-E. Chang, and M. K. Gilson, *Biophys. J.* **87**, 3035 (2004).
- ⁴³M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, *J. Comput. Chem.* **14**, 1347 (1993).
- ⁴⁴M. S. Gordon and M. W. Schmidt, in *Theory and Applications of Computational Chemistry*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Elsevier, Amsterdam, 2005), pp. 1167–1189.
- ⁴⁵G. M. J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, E. Guidez, T. Harville, S. Irlé, J. Ivanic, K. Kowalski, S. S. Leang, H. Li, W. Li, J. J. Lutz, I. Magoulas, J. Mato, V. Mironov, H. Nakata, B. Q. Pham, P. Piecuch, D. Poole, S. R. Pruitt, A. P. Rendell, L. B. Roskop, K. Ruedenberg, T. Sattasathuchana, M. W. Schmidt, J. Shen, L. Slipchenko, M. Sosonkina, V. Sundriyal, A. Tiwari, J. L. Galvez Vallejo, B. Westheimer, M. Włoch, P. Xu, F. Zahariev, and M. S. Gordon, *J. Chem. Phys.* **152**, 154102 (2020).
- ⁴⁶W. Chen, M. K. Gilson, S. P. Webb, and M. J. Potter, *J. Chem. Theory Comput.* **6**, 3540 (2010).
- ⁴⁷S. P. Webb, M. J. Potter, and L. E. Stewart, “Benchmarking the VM2 binding free energy software package: host–guest systems” (unpublished).
- ⁴⁸W. Chen, X. Ren, and C. A. Chang, *ChemMedChem* **14**, 107 (2019).
- ⁴⁹W. You, Y. M. Huang, S. Kizhake, A. Natarajan, and C. A. Chang, *PLoS Comput. Biol.* **12**, e1005057 (2016).
- ⁵⁰Y.-M. M. Huang, W. Chen, M. J. Potter, and C.-E. A. Chang, *Biophys. J.* **103**, 342 (2012).
- ⁵¹L. David, R. Luo, and M. K. Gilson, *J. Comput.-Aided Mol. Des.* **15**, 157 (2001).
- ⁵²V. Kairys and M. K. Gilson, *J. Comput. Chem.* **23**, 1656 (2002).
- ⁵³C.-E. Chang and M. K. Gilson, *J. Am. Chem. Soc.* **126**, 13156 (2004).
- ⁵⁴H.-X. Zhou and M. K. Gilson, *Chem. Rev.* **109**, 4092 (2009).
- ⁵⁵C.-E. Chang, M. J. Potter, and M. K. Gilson, *J. Phys. Chem. B* **107**, 1048 (2003).
- ⁵⁶M. J. Potter and M. K. Gilson, *J. Phys. Chem. A* **106**, 563 (2002).
- ⁵⁷C.-E. Chang and M. K. Gilson, *J. Comput. Chem.* **24**, 1987 (2003).
- ⁵⁸F. A. Momany and R. Rone, *J. Comput. Chem.* **13**, 888 (1992).
- ⁵⁹D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still, *J. Phys. Chem. A* **101**, 3005 (1997).
- ⁶⁰W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, *J. Am. Chem. Soc.* **112**, 6127 (1990).
- ⁶¹R. Luo, L. David, and M. K. Gilson, *J. Comput. Chem.* **23**, 1244 (2002).
- ⁶²W. Chen, J. Huang, and M. K. Gilson, *J. Chem. Inf. Comput. Sci.* **44**, 1301 (2004).
- ⁶³C.-E. A. Chang, W. Chen, and M. K. Gilson, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1534 (2007).
- ⁶⁴T. L. Hill, *An Introduction to Statistical Thermodynamics* (Dover Publications, 1986).
- ⁶⁵J. H. Jensen, *Phys. Chem. Chem. Phys.* **17**, 12441 (2015).
- ⁶⁶D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, *J. Comput. Chem.* **25**, 872 (2004).
- ⁶⁷M. K. Gilson, H. S. R. Gilson, and M. J. Potter, *J. Chem. Inf. Comput. Sci.* **43**, 1982 (2003).
- ⁶⁸M. Gaus, Q. Cui, and M. Elstner, *J. Chem. Theory Comput.* **7**, 931 (2011).
- ⁶⁹S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.* **132**, 154104 (2010).
- ⁷⁰A. D. Becke and E. R. Johnson, *J. Chem. Phys.* **123**, 154101 (2005).
- ⁷¹E. R. Johnson and A. D. Becke, *J. Chem. Phys.* **123**, 024101 (2005).
- ⁷²E. R. Johnson and A. D. Becke, *J. Chem. Phys.* **124**, 174104 (2006).

- ⁷³S. Grimme, S. Ehrlich, and L. Goerigk, *J. Comput. Chem.* **32**, 1456 (2011).
- ⁷⁴M. Gaus, A. Goez, and M. Elstner, *J. Chem. Theory Comput.* **9**, 338 (2013).
- ⁷⁵M. Kubillus, T. Kubař, M. Gaus, J. Řezáč, and M. Elstner, *J. Chem. Theory Comput.* **11**, 332 (2015).
- ⁷⁶A. V. Marenich, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. B* **113**, 6378 (2009).
- ⁷⁷H. Li and J. H. Jensen, *J. Comput. Chem.* **25**, 1449 (2004).
- ⁷⁸J. Tomasi, B. Mennucci, and R. Cammi, *Chem. Rev.* **105**, 2999 (2005).
- ⁷⁹V. Barone and M. Cossi, *J. Phys. Chem. A* **102**, 1995 (1998).
- ⁸⁰P. Su, H. Liu, and W. Wu, *J. Chem. Phys.* **137**, 034111 (2012).
- ⁸¹A. Bondi, *J. Phys. Chem.* **68**, 441 (1964).
- ⁸²J. C. Kromann, C. Steinmann, and J. H. Jensen, *J. Chem. Phys.* **149**, 104102 (2018).
- ⁸³V. Barone, M. Cossi, and J. Tomasi, *J. Chem. Phys.* **107**, 3210 (1997).
- ⁸⁴J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Phys. Rev. Lett.* **91**, 146401 (2003).
- ⁸⁵F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).
- ⁸⁶D. Feller, *J. Comput. Chem.* **17**, 1571 (1996).
- ⁸⁷K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. L. Windus, *J. Chem. Inf. Model.* **47**, 1045 (2007).
- ⁸⁸B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, and T. L. Windus, *J. Chem. Inf. Model.* **59**, 4814 (2019).
- ⁸⁹M. Katouda and S. Nagase, *Int. J. Quantum Chem.* **109**, 2121 (2009).
- ⁹⁰B. Q. Pham and M. S. Gordon, *J. Chem. Theory Comput.* **15**, 5252 (2019).
- ⁹¹M. G. Kendall, *Biometrika* **30**, 81 (1938).
- ⁹²R. T. McGibbon, A. G. Taube, A. G. Donchev, K. Siva, F. Hernández, C. Hargus, K.-H. Law, J. L. Klepeis, and D. E. Shaw, *J. Chem. Phys.* **147**, 161725 (2017).
- ⁹³L. Zheng and W. Yang, *J. Chem. Theory Comput.* **8**, 810 (2012).
- ⁹⁴J. I. Monroe and M. R. Shirts, *J. Comput.-Aided Mol. Des.* **28**, 401 (2014).
- ⁹⁵E. O. Purisima, *J. Comput. Chem.* **19**, 1494 (1998).
- ⁹⁶H. Hogues, T. Sulea, and E. O. Purisima, *J. Comput.-Aided Mol. Des.* **28**, 417 (2014).
- ⁹⁷C. H. Bennett, *J. Comput. Phys.* **22**, 245 (1976).
- ⁹⁸J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon, *J. Phys. Chem. B* **114**, 2549 (2010).
- ⁹⁹A. Klamt and G. Schüürmann, *J. Chem. Soc., Perkin Trans. 2* 799 (1993).
- ¹⁰⁰P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, and D. Cohen, *J. Chem. Phys.* **105**, 1968 (1996).
- ¹⁰¹S. P. Webb and M. S. Gordon, *J. Phys. Chem. A* **103**, 1265 (1999).
- ¹⁰²G. N. Merrill, S. P. Webb, and D. B. Bivin, *J. Phys. Chem. A* **107**, 386 (2003).
- ¹⁰³G. N. Merrill and S. P. Webb, *J. Phys. Chem. A* **107**, 7852 (2003).
- ¹⁰⁴G. N. Merrill and S. P. Webb, *J. Phys. Chem. A* **108**, 833 (2004).
- ¹⁰⁵S. Shanker and P. Bandyopadhyay, *J. Phys. Chem. A* **115**, 11866 (2011).
- ¹⁰⁶T. Sattasathuchana, P. Xu, and M. S. Gordon, *J. Phys. Chem. A* **123**, 8460 (2019).