



Editor's Choice

How well do implicit solvation models represent intermolecular binding energies in organic-inorganic solutions?

Blaire A. Sorenson^a, Spencer S. Hong^b, Henry C. Herbol^a, Paulette Clancy^a^a Dept. of Chemical and Biomolecular Engineering, The Johns Hopkins University, 3400 N. Charles St, Baltimore, MD 21218, United States^b, Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, 113 Ho Plaza, Ithaca, NY 14853, United States

ARTICLE INFO

Keywords:

Perovskites

Density functional theory

Solvation models

ABSTRACT

Computational methods, including density functional theory, are proving to be powerful approaches to tame the otherwise overwhelming selection of optimal species and processing conditions to fabricate hybrid organic-inorganic perovskite (HOIP) thin films *via* solution processing. In that processing, the choice of solvents is known to play a critical role in the quality of the resulting thin film, but their inclusion in a simulation dominates the overall size of the system, and hence the computational effort. This creates an incentive to understand the minimal representation of solvent medium necessary to adequately model the interactions between HOIP building blocks. These building blocks are chiefly the lead salt and cationic species dissolved in the processing solvent, whose interactions, monitored by calculations such as binding energies, govern nucleation and growth into thin films. We show that a simple implicit solvent model is surprisingly effective in terms of representing the intermolecular binding energies between lead salts and cation species in solution. Use of an implicit solvent produces binding energies that are typically within 2–3 kcal/mol of more accurate “all atom” models of solvent molecules. This is an important result since implicit solvents are roughly 100 times more efficient than explicit solvent models in terms of computational resource effort. We find that it is generally sufficient to use a Generalized Gradient Approximation for the DFT calculations, rather than more accurate, but more expensive, PW6B95 models.

1. Introduction

The ability to represent solvation effects using a computational approach as a means to understand the effects of solvent choice on processing is important for a large number of chemical systems. This includes biomolecular systems for which solvation effects and binding affinities are considered to be extremely important, for example, computer-assisted drug discovery [1]. Solvation effects in biological systems are also important in protein folding, self-assembly, biomolecular recognition and virtually all biological processes [2]. Solvent choice has also played a role in affecting the kinetics, morphology, and resulting structure manifested in crystal growth [3]. Given the computational expense of explicitly modeling every atom for all solvent molecules present in the simulation cell, there has long been an incentive to simplify solvent modeling, invariably with what are called “implicit” solvent models.

Assessing the accuracy of implicit solvent models is important, *vis-a-vis* explicit but expensive “all-atom” models, especially in systems that are required by system considerations to be large in scale or complexity. Understanding the effectiveness of implicit models has been

investigated, as listed below, but no study has yet uncovered how well an implicit solvent model captures the intermolecular binding energies of two species in solution, modeled using a quantum mechanical representation. Previous studies have looked at the influence of different parameters on various solvation models and the associated speed of the calculations for small molecules, proteins and protein-ligand complexes, as well as the predictive power of implicit solvent models for solvation free energies of organic molecules in organic solvents. [4,5] A second consideration is that sampling conformational space involving solute-solvent systems is challenging if the solvent is treated explicitly. This has also driven the use of implicit solvent models, which have been shown to speed up conformational sampling significantly [6]. Similar studies to understand the effectiveness of the implicit ‘Conductor-like Screening MOdel’ (COSMO) [7] used in quantum-mechanical calculations for intermolecular binding energies between two distinct species, neither of which are the solvent itself, has been overlooked. This is a targeted study here.

One application where this omission is particularly important involves the synthesis of hybrid organic-inorganic perovskites (HOIPs) from solution, since these are promising materials for solar cells. The

E-mail address: bsorens4@jhu.edu (B.A. Sorenson).<https://doi.org/10.1016/j.commsatsci.2019.109138>

Received 17 April 2019; Received in revised form 25 June 2019; Accepted 12 July 2019

0927-0256/ © 2019 Elsevier B.V. All rights reserved.

molecular self-assembly process that creates HOIP thin films begins with the dissolution of lead salts by a processing solvent, in the presence of organic, or inorganic, cations. This dissolution (solvation) process is poorly understood and under-investigated due to experimental and computational difficulties in probing interactions in solution. The first step in the process involves lead complexation with the processing solvent, as studied by Stevenson et al. [8] In that work, they used accurate, but computationally expensive, *ab initio* calculations via density functional theory (DFT) to describe lead complexation in HOIP solutions of both solvents. DFT remains our default approach given the lack of semi-empirical alternatives, but is restricted to small-scale studies due to its expense. This puts a premium on minimizing the time needed to perform these costly calculations and provided the motivation for this study.

In this work, we modeled the following solvents, most of which are commonly used in experimental processing: Dimethyl sulfoxide (DMSO), dimethylformamide (DMF), N-methyl-2-pyrrolidone (NMP), γ -butyrolactone (GBL), as well as less common ones, like tetrahydrothiophene 1-oxide (THTO) to study their interaction with the APbX_3 monomer. THTO was predicted computationally to act as an effective potential solvent for the dissolution of lead salts, later confirmed by experiments where it was used as an additive to perovskite solutions. THTO slowed nucleation and growth processes, while gaining control over the orientation of the resulting thin films [8,9]. We also included acetone (ACE) in our study. Although acetone has not been used in experimental studies as a processing solvent, it has been used as a solvent in the post-processing of HOIPs and studied theoretically [8,10,11].

In an attempt to curtail costs, we represent these solvent molecules in the systems under study using implicit solvent models that represent the molecules as a homogeneous medium, rather than providing an *explicit* description of all the atoms that constitute the solvent molecules. Use of implicit solvent models significantly speeds up the simulation process, [12] which then expands our ability to study a more extensive set of solvents and solvent blends in conjunction with the vast number of possible combinations of lead salt, PbX_2^- ($X = \text{I}, \text{Br}$ or Cl) and A-site cations, methylammonium (MA^+), cesium (Cs^+) or formamidinium (FA^+). The question arises: How well does the use of an implicit solvent in simulations represent perovskite complexation studies of the lead salt and A-site cation? This paper aims to answer that question.

We focus on the intermolecular binding energy of the organic and inorganic A-site cation to the lead salt PbX_2^- in the both explicit and implicit solvation models using DFT simulations. Our tasks are as follows: (1) First, we need to determine the species that constitute the smallest and most energetically favorable nuclei to study. (2) We will need to compare our results for the intermolecular binding energy using explicit solvent models to those using the COSMO implicit solvent model. Our goal is to determine just how accurately we need to model the solvents to obtain accurate binding energetics of the A-site cation to the lead salt. Our results will show that an implicit representation is accurate enough to represent the role of the solvent, which saves a factor of 100 or more CPU [7,13]. Finally, we test the effect of using two separate levels of theory, a pure generalized gradient approximation (GGA) B97 [14,15], and a more computationally expensive hybrid PW6B95 functional [16]. This test will allow us to report on the impact of functional choice on the intermolecular binding energies of these systems. Given the expense of DFT calculations, the results of this test are also important regarding computational cost-savings which might otherwise limit the scope of future studies.

2. Identifying energetically favorable nuclei in solution

Since our goal is to study perovskite nucleation in solution, we began by investigating the nature of structural building blocks; task (1) outlined above. Since the chemical composition of the most stable

monomer species is unknown, we built clusters of the following potential structures: APbX_3 , APbX_2^+ and APbX^{2+} , for all combinations of halide choices (allowing X to be Cl , Br , or I) and three A-site cations, MA^+ , FA^+ , or Cs^+ . All these clusters were created in Avogadro [17] and their preferred geometry was optimized using the quantum mechanical DFT package Orca [18].

Initially, these structures were modeled in a vacuum (*i.e.*, no solvent is included at this point). For each combination of lead salt and cation, we generated a set of samples that differed in the initial positions of the cation relative to the lead salt. This provided an ensemble of structural options for the DFT calculations to explore. We optimized the geometry of the samples for all these structures (APbX_3 , APbX_2^+ and APbX^{2+}) and hence determined the resulting lowest energy conformation across all samples for each halide and cation combination. For equity, these comparisons considered the same number of atoms and atom types in each system.

The result of this analysis showed that the neutral species APbX_3 was the most energetically stable for all combinations lead salt and cation choice, compared to APbX and APbX_2 . As a result, the rest of this paper will *only consider the APbX_3 complex* as representative of the smallest and most energetically stable building block of the perovskite solution, which we define as a monomer. Although we focus on the neutral APbX_3 , other building blocks, such as the charged species APbX_2^+ , APbX^{2+} , or others, almost certainly exist in solution. Indeed, APbX_3 precursor solutions exhibit diverse lead coordination environments as a function of precursor solution chemistry. Sharenko *et al.* showed that, as excess MAI is added to the solution, iodine ions displace coordinated dimethylformamide (DMF) molecules to form higher-order iodoplumbate complexes [19]. Their work does not, however, elucidate the energetics of species formed as a function of the molar ratio of processing species. We are currently investigating the formation of iodoplumbates and their interaction with the cationic species; but a study of these larger, charged systems is outside the scope of this paper.

Now that we have identified the APbX_3 monomer as the lowest energy complex in solution, there are nine possible combinations for homogeneous halide compositions, given that there are three choices of halide and three choices of A-site cation. Studying each of these nine combinations in the six pure solvents amounts to 54 combinations, all of which will be studied in this paper. We chose four different solvation models for this study that are discussed in Section 3. Together, this amounts to 216 combinations to consider.

3. Organization of this paper

The rest of this paper will be organized as follows: (1) Section 4 will cover methods of the DFT calculations used in this study. (2) Cost is also governed by the complexity of the solvent's representation for the explicitly modeled solvent systems. Here, we used molecular dynamics (MD) to generate random samples of the solvent around the solute. These simulations were used to determine the number of solvent molecules that are actually bound to the lead salt (Section 5). (3) Although our goal is to *minimize* the solvent representation, we are also interested to determine what resources would be needed to calculate the enthalpy of solvation directly. This would involve consideration of the entire shell of solvents surrounding the salt. However, what constitutes a full solvent shell is unknown. Section 6 determines what would be needed and, as a result, rules out enthalpy of solvation as a viable route to screen for solvent quality in the absence of unlimited resources. (4) With these important underpinnings in place, Section 7 shows how the intermolecular binding energies between lead salt and solvent are calculated. Section 8 provides the results of these studies. Our recommendations for the minimal required representation of the solvent are given in Section 9.

4. *Ab initio* simulation details

Density functional theory geometry optimizations were conducted in Orca [18] employing either (i) the pure GGA B97 functional [20] or (ii) the more accurate (and expensive) hybrid functional PW6B95 [16] and the polarized basis set TZVPP [21] with dispersion corrections, as recommended by Grimme [22,15]. The goal here was to determine how accurately DFT calculations have to be performed in order to be adequate for complexation studies. We optimized each monomer in an *implicit* solvent, using either B97 or PW6B95 functionals. The only input for the implicit solvation model is the relative dielectric of the solvent, which generates an electric field experienced by the APbX_3 species.

To reduce computational effort, the usually highly contracted and chemically inert core basis functions were eliminated by employing Effective Core Potentials (ECPs) [23,21]. ECP calculations comprise a “valence-only” basis and thus are subject to the “frozen core” approximation. An Effective Core Potential was applied to Pb, Cs, I, and Br. The Geometrical Counterpoise Correction (GCP) [24] was added to remove artificial over-binding effects from basis set superposition errors (BSSE). This correction uses atomic corrections and thus also yields the ability to correct for intramolecular BSSE. All systems were given a Tight SCF and slow convergence criteria with a grid size of 7. Upon convergence, each system was re-optimized using the same inputs described above, but this time including the density-based solvation model, SMD [25,26].

The SMD solvation model is based on the quantum mechanical charge density of a solute molecule interacting with a continuum description of the solvent with surface tension at the solute-solvent boundary. In SMD, the full solute electron density is used without defining partial atomic charges. The solvent is not represented explicitly in these calculations, but rather, the system is passed a dielectric value to represent the solvent molecules. We determined the number of coordinated solvent molecules to each monomer according to the number of solvents that formed bonds with the monomer. From each “bin” that this created, we identified the complex with the lowest energy and marked it for further study to uncover the nature of the complexation and geometry, as well as its electronic structure.

5. Molecular dynamics studies to determine explicit solvent binding motifs

All the explicit solvation models in this study were obtained using a combination of MD and DFT simulations. The MD simulations, done using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software [27], are used to determine the solvation of the monomer at room temperature. The resulting configuration is then used as a good starting point for a more accurate *ab initio* DFT calculation to produce an equilibrated structure for the solvated monomer complex. First, for the MD studies, we created initial geometries for the coordination complexes by equilibrating the monomer species in solvent using the semi-empirical OPLS force field [28]. We used the similarly sized Ba ion in place of Pb for the MD simulations since there is no OPLS model for Pb systems. This generated an approximate packing of solvent molecules around the solute. The MD simulation studied a single monomer of ABaX_3 exposed to abundant solvent; here, a box containing 100 solvent molecules.

The complexes have many degrees of freedom, thus no optimization of a single posited structure will reliably approach the global optimum geometry. Accordingly, we created ten different sample geometries of the $[\text{ABaX}_3 + \text{solvent}]$ complex using packmol [29]. For each sample, we used a different pseudo-random seed to ensure that the starting configurations of the solvent molecules around the solute (the ABaX_3 monomer) had different initial coordinates. These ten simulations were equilibrated for 0.1 ns at 300 K, until the energy fluctuations were less than 0.1 kcal/mol.

To determine how many solvents coordinated with the monomer,

we used two criteria, the bond length between the lead ion and the most polar atom in the solvent molecule (invariably oxygen), and the Mayer Bond Order [8] which measures the electron density between those same atoms. The cut-offs for these two “bonding” criteria were a bond length of 3.0 Å, and an MBO of 0.1. The MD-generated complexes containing Ba were replaced with Pb then further optimized using DFT to obtain electronic and structural information. Using these metrics, the number of solvent molecules coordinated to the lead atom on the monomer *never exceeded three solvent molecules* for all the solvents we studied. This allowed us to limit our study of explicit models to the consideration of *no more than three* solvent molecules around the lead salt, which we refer to in later sections as the $N_S = 1, 2$ or 3 solvation model.

While the purpose of this work is to determine how well solvents need to be represented to achieve accurate intermolecular binding energies between PbX_3^- and the A-site cation, A^+ , we further probed the expense of obtaining a full solvation shell of explicitly modeled solvents around the solute, including solvents that are not bound to the monomer complex. Once a full shell is achieved, we can then calculate the enthalpy of solvation for this system (using Eq. 1, Section 7) as we have done for the intermolecular binding energy. The subtle difference in the two calculations lies only in the number of solvents in the included calculations. Such calculations are highly desired among both computationalists and experimentalists for a variety of reasons. We elucidate the investment of computational time need to acquire such a value in the following section.

6. Determination of the expense of enthalpy of solvation calculations

Given unlimited computing resources, the enthalpy of solvation (rather than ersatz metrics such as the Mayer Bond Order [8] or the Gutmann donor number [30]) would be the calculation of choice for a chosen combination of lead ion, halide and A-site cation in the presence of a full solvation shell of solvent molecules. But how many molecules does that “full shell” imply for a lead ion? And how does the enthalpy of solvation deviate from this limiting case as you consider fewer and fewer solvent molecules to be part of that shell?

To answer these questions and determine the computational expense of capturing a representative solvation shell of a monomer, we calculated the intermolecular binding energy between N solvents and a single ABX_3 , in this case, CsPbCl_3 in γ -butyrolactone. Energies were calculated using DFT with the B97-D3 functional [14,15], split valence basis set, def2-TZVP effective core potentials [21], and the COSMO solvation model [25]. We observe the expected result that the solvation energy in the bulk asymptotically begins to flatten as the number of solvent molecules increases, up to the limit of a full shell of solvent molecules. We fitted this curve to a function, $f(x) = A - Be^{-Cx}$, where A, B, and C are fitting constants, and x is the number of solvents.

To capture how many solvent molecules are potentially required to “converge” to a full shell value, we determined the point at which the slope of the curve in Fig. 1 drops below that of thermal fluctuations (1 kT at 300 K; 0.6 kcal/mol). This point, a converged enthalpy of solvation, occurs at a shell size containing around 23 solvent molecules (an extrapolation of the data shown in Fig. 1). The blue curve in Fig. 1 illustrates the computational time needed to complete one data point on this graph. A simulation that considers 16 solvent molecules took approximately 388 CPU hours of computing time (approximately 97 h per core) when run in parallel on a fast CPU for one choice of halide, cation and solvent. Furthermore, these calculations were done for a simplified level of theory in which only a split valence basis set was used, and the convergence criteria was set loosely. Clearly, these calculations are too slow to allow us to screen a large set of candidates using the enthalpy of solvation as a metric. It provides a clear motivation to look for other less expensive models and metrics. Our choice for such a metric was to determine the binding energy between the lead salt and an A-site cation

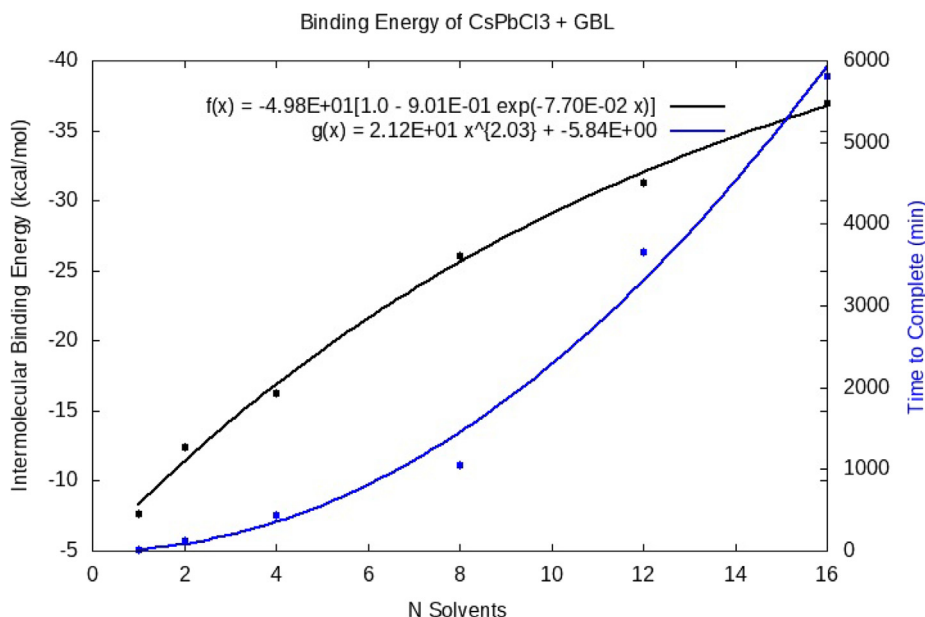


Fig. 1. Enthalpy of solvation as a function of the number of neighboring solvent molecules explores how many solvent molecules are necessary to calculate a converged solvation energy shell around CsPbCl_3 in a γ -butyrolactone solvent via DFT. The black line represents a best fit of the enthalpy of solvation data to a fitting function, $f(x)$. The blue line represents a best fit of the time taken for each calculation to converge.

in solvent.

7. Calculation of cation binding energies

From the simulations described above, we were able to identify the most accurate geometry and lowest energy sample for each monomer-solvent combination, and the number of solvents that bind to it. This allowed us to calculate the intermolecular binding energy between two species: the lead salt, PbX_3^- , and the A-site cation. Further, it also allows us to see how the intermolecular binding energy changes as we vary the number of solvent molecules bound to the monomer. While our simulations in Section 5 suggest that no more than three solvents from bonds to the complex, from a computational effort point of view, the fewer the number of solvents we consider, the better, as will be considered in Section 8.1.

The binding energy, E_b , is defined as:

$$E_b = E_{\text{APbX}_3} - E_{\text{PbX}_3} - E_A \quad (1)$$

where E_{APbX_3} is the energy of a single monomer containing the lead salt and A-site cation, with a chosen number of coordinated solvents (1, 2, or 3). E_{PbX_3} is the energy of an isolated lead salt with coordinated solvents, and E_A is the energy of the isolated A-site cation (see Figs. 2 and 3). The binding energy between the lead salt and A-site cation is closely related to the Coulombic force, F_C , and hence the relative dielectric of the medium in which the two species reside:

$$F_C = \frac{q_1 q_2}{r^2 4\pi \epsilon_0 \epsilon_r} \quad (2)$$

To compare these solvated systems to that of a monomer in a solid state, we calculated the intermolecular binding energy of the MAPbI_3 monomer in a vacuum (no solvent) to be -101 kcal/mol. This value falls within the range of binding energy values for the MAPbI_3 system found computationally by Varadwaj et al. Their work showed the binding energy dependence in MAPbI_3 as a function of the geometry of the structure, in both the monomer unit and a unit cell structure [31]. This work has important implications in the area of device-related properties, but there has been no relationship that has shown to what extent the processing solvent affects the final geometry of the monomer, or the position that the cation assumes once incorporated into a thin film.

8. Results and discussion

8.1. Comparison of Solvation Models

As mentioned in Section 5, we considered three explicit solvation models in this study, identified by the number of solvent molecules, N_S , that form bonds to the lead atom in the APbX_3 system ($N_S = 1, 2$ or 3 solvents). We were able to limit N_S to 3, since no more than three solvent molecules bound to APbX_3 .

Using the largest and most computationally expensive system, $N_S = 3$, we looked at the range of potential energies from DFT, produced from ten different samples for MAPbI_3 and PbI_3^- in three different solvents: THTO ($\epsilon = 42.84$) DMF ($\epsilon = 36.7$) and NMP ($\epsilon = 32.2$). We found that the difference in the resulting potential energies, corresponding to slightly different geometries for these randomly generated samples, all fell within 2 kcal/mol, and the standard deviation of the sample population showed that all samples are very close to the average. Values for the range and standard deviations are reported in Table 1.

We gauged the accuracy of the intermolecular binding energy between lead salt to the A-site cation of the other solvation models ($N_S = \text{implicit}, 1, 2$) to the most accurate model we considered, $N_S = 3$. For MAPbX_3 systems, the three explicit solvation models together with the COSMO implicit model, differ, at most, by around 2 kcal/mol, close to the intrinsic uncertainty in the DFT calculations [32] (ESI Tables 1 and 2). The largest relative difference in the intermolecular binding energy between PbX_3 and the MA cation is shown to occur when modeling the solvent NMP using the implicit solvation model and the PW6B95 functional. For the CsPbX_3 systems, differences in intermolecular binding energy of PbX_3 to the cesium cation, exhibit a larger range where the largest difference was 3.5 kcal/mol (ESI Tables 5 and 6). However, there were some notable differences when modeling FAPbX_3 systems. These systems were much less consistent in capturing the binding energy from model to model (ESI Tables 5 and 6). This is reflected in the largest relative difference in energies that compare the $N_S = 3$ to the $N_S = 1$ system which occurs in FAPbX_3 , where the energy difference is 4–6 kcal/mol.

Figs. 2 and 3 show the overall trend in the intermolecular binding energy between lead salt and A-site cation as a function of relative dielectric of the solvent for all choices of halides and A-site cations, four

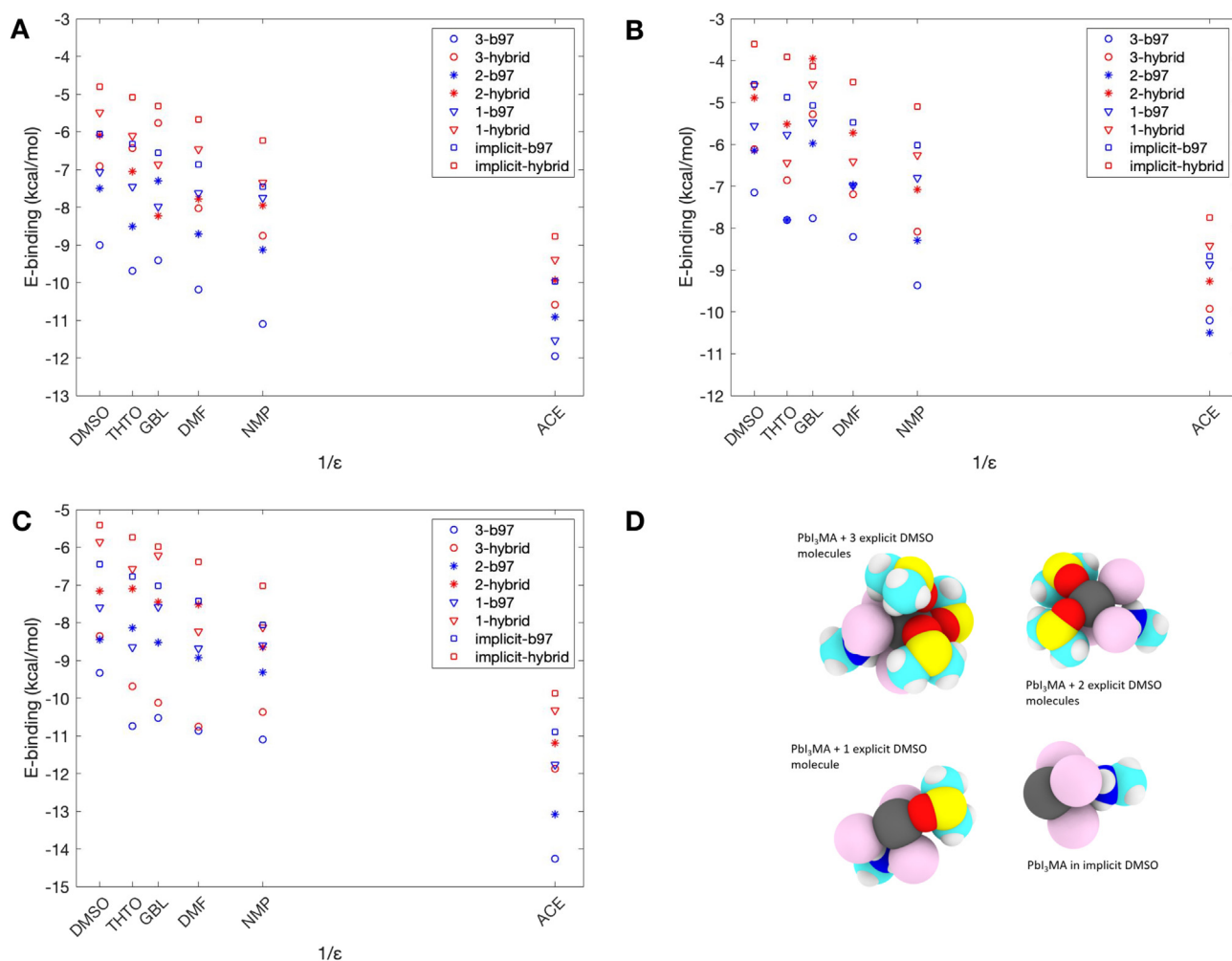


Fig. 2. Intermolecular binding energies between PbX₃ and MA for: (A) PbI₃MA, (B) PbBr₃MA, and (C) PbCl₃MA for all four solvation models, plotted against the dielectric constant of each solvent given as 1/ε. For each solvent, the difference between each solvation model shows a difference of less than 2 kcal/mol. Across the different solvents there is a clear trend for the cation binding energy to increase as a function of decreasing relative dielectric. This is true whether the model is atomically explicit or features an implicit model. The same trend is shown for both the hybrid PW6B95 and pure GGA B97 functionals. (D) PbI₃MA in DMSO for four different solvation models, N_s = 3, N_s = 2, N_s = 1, N_s = Implicit.

solvation models and two different levels of theory. As mentioned above, the binding energy between the two species, the lead salt and the A-cation, is related to the dielectric constant, as shown in Eq. 2, making it a natural metric for solvent comparison. The trend that we observe in intermolecular binding energy between the lead salt and cation using the more expensive explicit solvation models of N_s = 3, 2 or 1 solvents shows a linear relationship with the dielectric of the solvent for all monomer combinations.

For each cation in this study, the difference in the calculated intermolecular binding energy to lead salt shows a consistent variation as you reduce the number of explicit solvents in the model from N_s = 3 to N_s = 2 to N_s = 1 to N_s = implicit (ESI Tables 1–6). The intermolecular binding energy varies between 0.5 and 2 kcal/mol for all systems in this study, within the uncertainty of the DFT calculations. This is true for both the B97 and PW6B95 functionals, albeit the difference is slightly smaller using the PW6B95 functional. Importantly, a linear trend is preserved for all models and systems: as the dielectric decreases, the intermolecular binding energy increases. The significance of this result is that, if all that is needed is a trend that shows how the solvent affects the intermolecular binding energy between two species, then the implicit model is more than sufficiently accurate. To illustrate this trend, Fig. 4 shows results only for the implicit model for the PW6B95 level of theory for all solvents and monomer combinations in this study.

8.2. Effect of the level of DFT theory

Calculations for the intermolecular binding energy between PbX₃ and the A-site cation as a function of solvent were completed for each model for two levels of DFT theory, B97 and PW6B95. These energies are reported in Tables 1–6 in the ESI. Figs. 2a, b, c and 3 show that B97 calculations follow the same trend as the more expensive PW6B95 level of theory.

For the same system using two different functionals, we observed consistent differences in the binding energies for each cation system. For MAPbX₃ systems, we observed a 2 kcal/mol difference in which the B97 level calculations exhibit the higher (more negative value) of the two. This was also observed for FAPbX₃ systems, albeit the difference in the intermolecular binding energy is less consistent from model to model. The largest difference we observed was found in the N_s = 1 system, where the change in binding energy from B97 to PW6B95 was 5 kcal/mol. For CsPbX₃, the difference between theory levels is larger than that for different organic cations, an average difference of around 5 kcal/mol, but this difference is consistent across all models.

The smallest difference we observed in the binding energy occurred for the implicit models using B97 and PW6B95. For all systems, this change in value is around 1 kcal/mol, again within DFT accuracy. This is expected as the implicit model does not have to accommodate any changes in the system or adjust its geometry upon convergence with

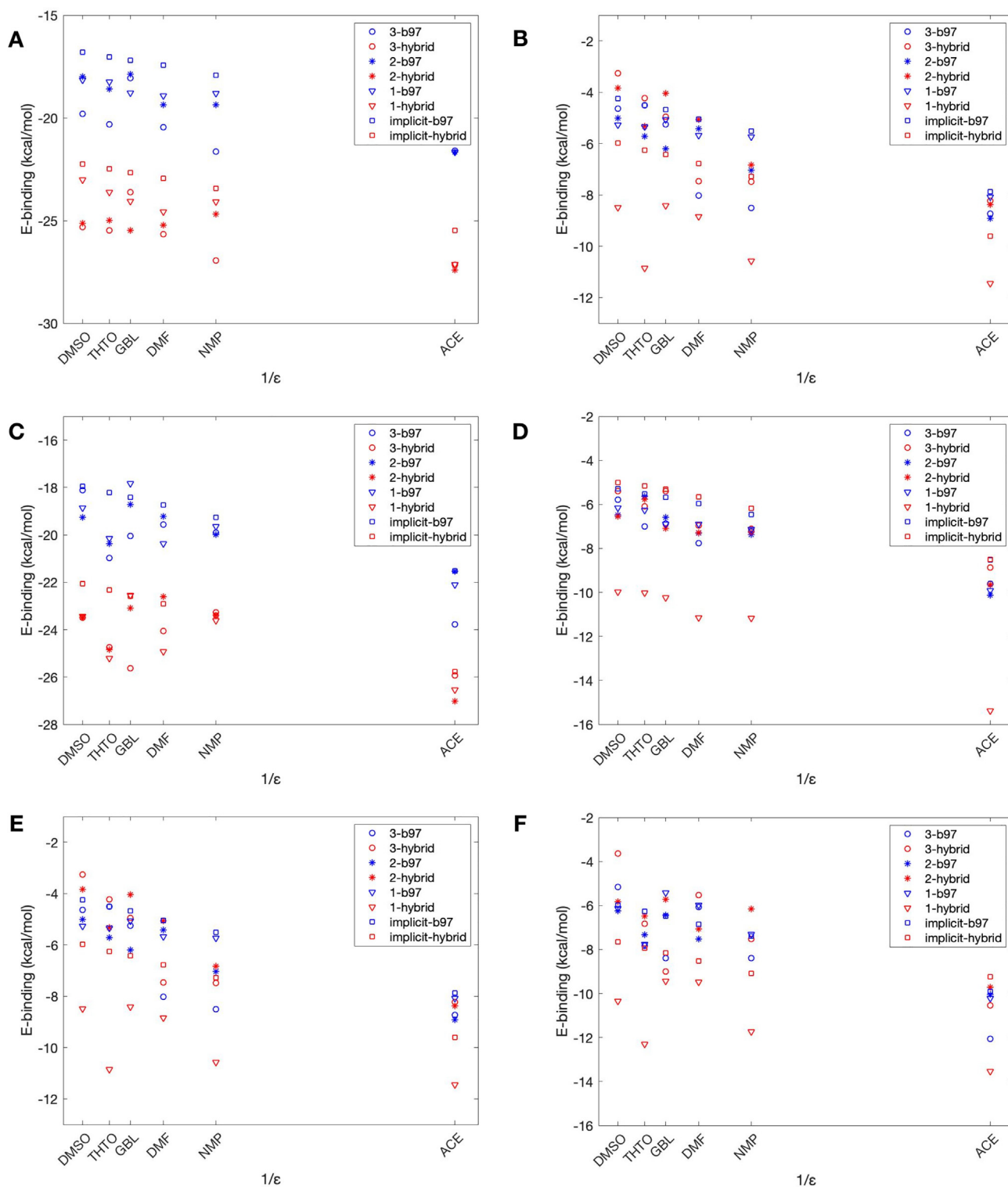


Fig. 3. Intermolecular binding energy between (A) PbI_3 and FA, (B) PbBr_3 and FA, (C) PbCl_3 and FA, (D) PbI_3 and Cs, (E) PbBr_3 and Cs, (F) PbCl_3 and Cs. A–F show results for all four solvation models, as a function of the dielectric constant of each solvent plotted as $1/\epsilon$.

explicit solvents, as would be expected in the optimization for $N_s = 1$ –3 systems.

Overall, the cheaper, pure-GGA, B97 functional is sufficiently accurate to capture *trends* in cation binding energetics of perovskite monomers in solution. But, in order to obtain *accurate values* of the binding energetics, an explicit model and more expensive PW6B95 level of theory should still be employed.

8.3. Impact of model choice on CPU time

To determine the time advantage of using an implicit solvation model, we compared its simulation time to the explicit model for the same monomer and solvent system on the same level of theory. Our work flow, as described in Sections 4 and 5, first optimized the geometry in B97, and then the resulting system was subject to further optimization on the PW6B95 functional. Therefore, we assume the total

Table 1
Range and STD of DFT energies for generated samples.

Solvent	Range (kcal/mol)	STD
MAPbI ₃		
THTO	1.295	0.0006
DMF	0.935	0.0004
NMP	1.854	0.0007
PbX ₃ ⁻		
THTO	2.051	0.0009
DMF	1.441	0.0008
NMP	1.669	0.0008

run time as the sum of B97 and PW6B95 times. We report the average run time of the B97 and PW6B95 for the MAPbI₃ system for all solvation models in this study using DFT. The cited costs do not include the cost of the MD simulations used to create the initial configurations for the DFT studies.

Fig. 5 plots the average total run time for all four models in this study. We include two bounds, the upper as the total run time for each model, and the lower bound as the shortest run time resulting from either the B97 or PW6B95 simulation. The lower bound represents the optimal scenario where structure from the B97 simulation is fully optimized and the PW6B95 is redundant. The upper bound, on the other hand, is the conservative bound where we assume the worst case scenario: the B97 geometry is incorrect and the total run time is the addition of the B97 and PW6B95 simulation times. The resulting differences in computational time and cost are substantial (ESI Table 7): The average time needed for the more accurate but *expensive* model using three explicit solvents was 32 CPU hours compared to the implicit model which took an average of 7 min for the systems under study, a factor of nearly 300 times slower.

9. Conclusions

This work determined just how precisely the solvent needs to be modeled in a molecular simulation of complexation in mixed organic-inorganic solutions. The implicit solvent model came surprisingly close (generally within a couple of kcal/mol) to reproducing the intermolecular binding energy of the most complex explicit model that we studied ($N_s = 3$). This reduced the cost of the simulations by about two orders of magnitude. There is a considerable impact to this result: A simple and computationally efficient implicit model is effective at

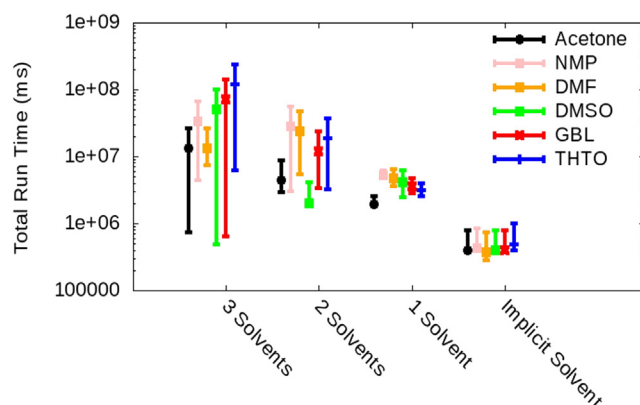


Fig. 5. Comparison of simulation run times for different PbI₃MA solvation models. The upper bound is the sum of the B97 and PW6B95 functional run times while the lower bound is the shortest run time found for the two levels of theory.

representing the major trends in intermolecular binding energies, which are important in the establishment of complexes and sub-critical nuclei in solution.

We have established several key measures of solvent representation in this exotic mix of organic and inorganic species. First, we established that around 23 solvent molecules are involved in providing a full solvation shell around a APbX₃ monomer. This quantifies the minimum number of solvents needed to obtain an accurate value for the enthalpy of solvation. The high expense of these calculations, however, is unnecessary in cases where only the intermolecular binding energy between two species is of interest.

Second, if an explicit model is desired, you can reduce the computational expense by using smaller systems, since the number of solvent molecules that *actually bind* to the lead atom in the monomer is never more than three in the cases we studied. The APbX₃ monomers showed that the intermolecular binding energy changes in a predictable fashion as you steadily reduce the number of solvent molecules in the lead salt motif.

Third, and in contrast to the mild dependence on the number of solvent molecules in the vicinity of the lead salt, we found significant differences in cation binding energies between the PW6B95 and B97 levels of DFT theory. For *accurate* lead salt-cation binding energies, an expensive PW6B95 DFT model with explicit solvent models is needed.

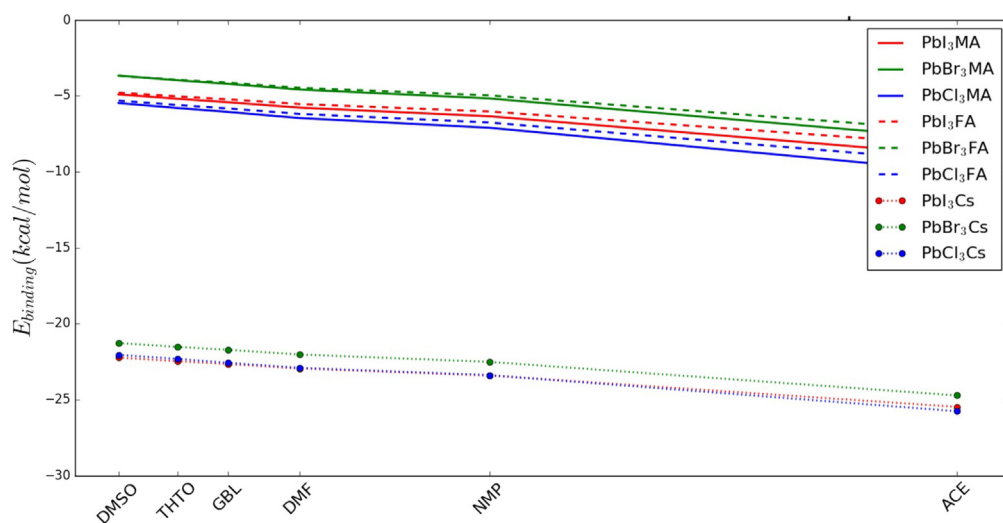


Fig. 4. Cation binding energy of APbX₃ as a function of $1/\epsilon$ showing a roughly linear relationship for a variety of halide and cation choices and both bath solvents and anti-solvents (modeled with an implicit solvent). Results for systems containing cesium show a significantly more negative (stronger) cation binding energy. Color key as in the insert.

On the other hand, if *trends* the intermolecular binding energy are all that is required, the B97 functional is generally sufficient.

Finally, and importantly for a rapid triage of solvent candidates, we uncovered an almost linear relationship between the intermolecular binding energy of a solvated lead salt to an A-site cation with the dielectric constant of the solvent, suggesting that this is a useful metric of binding strength for these systems. This combination of computational efficiency and the simple, near-linear, nature of the relationship between binding energy and dielectric constant will make it straightforward to quickly screen the effectiveness of other, as yet unexplored, potential solvents or additives to promote complexation between perovskite monomers.

Acknowledgements

This work was primarily supported by the Cornell Center for Materials Research with funding from the NSF MRSEC program (DMR-1719875). BAS thanks the Ford Foundation for support of her doctoral research, SH thanks the Hunter R. Rawlings III Cornell Presidential Research Scholars program (RCPRS) for support of his undergraduate research. The intense amount of computing resources required for this study was provided by the Cornell Institute for Computational Science and Engineering (ICSE).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.commatsci.2019.109138>.

References

- [1] N.A. Murugan, V. Poongavanam, U.D. Priyakumar, Recent advancements in computing reliable binding free energies in drug discovery projects, *Structural Bioinformatics: Applications in Preclinical Drug Discovery Process*, Springer, 2019, pp. 221–246.
- [2] C.G. Ricci, B. Li, L.-T. Cheng, J. Dzubiella, J.A. McCammon, 'martinizing' the variational implicit solvent method (vism): solvation free energy for coarse-grained proteins, *J. Phys. Chem. B* 121 (2017) 6538–6548.
- [3] L. Shimon, M. Vaida, L. Addadi, M. Lahav, L. Leiserowitz, Molecular recognition at the solid-solution interface: a relay mechanism for the effect of solvent on crystal growth and dissolution, *J. Am. Chem. Soc.* 112 (1990) 6215–6220.
- [4] E. Katkova, A. Onufriev, B. Aguilar, V. Sulimov, Accuracy comparison of several common implicit solvent models and their implementations in the context of protein-ligand binding, *J. Mol. Graph. Model.* 72 (2017) 70–80.
- [5] J. Zhang, H. Zhang, T. Wu, Q. Wang, D. van der Spoel, Comparison of implicit and explicit solvent models for the calculation of solvation free energy in organic solvents, *J. Chem. Theory Comput.* 13 (2017) 1034–1043.
- [6] R. Anandakrishnan, A. Drozdetski, R.C. Walker, A.V. Onufriev, Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations, *Biophys. J.* 108 (2015) 1153–1164.
- [7] A. Klamt, G. Schüürmann, Cosmo: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, *J. Chem. Soc., Perkin Trans. 2* (1993) 799–805.
- [8] J. Stevenson, B. Sorenson, V.H. Subramaniam, J. Raiford, P.P. Khlyabich, Y.-L. Loo, P. Clancy, Mayer bond order as a metric of complexation effectiveness in lead halide perovskite solutions, *Chem. Mater.* 29 (2016) 2435–2444.
- [9] B.J. Foley, J. Girard, B.A. Sorenson, A.Z. Chen, J.S. Niezgoda, M.R. Alpert, A.F. Harper, D.-M. Smilgies, P. Clancy, W.A. Saidi, et al., Controlling nucleation, growth, and orientation of metal halide perovskite thin films with rationally selected additives, *J. Mater. Chem. A* 5 (2017) 113–123.
- [10] P. Tyagi, S.M. Arveson, W.A. Tisdale, Colloidal organohalide perovskite nanoplatelets exhibiting quantum confinement, *J. Phys. Chem. Lett.* 6 (2015) 1911–1916.
- [11] Y. Wang, S. Li, P. Zhang, D. Liu, X. Gu, H. Sarvari, Z. Ye, J. Wu, Z. Wang, Z.D. Chen, Solvent annealing of pbi 2 for the high-quality crystallization of perovskite films for solar cells with efficiencies exceeding 18%, *Nanoscale* 8 (2016) 19654–19661.
- [12] D.E. Tanner, K.-Y. Chan, J.C. Phillips, K. Schulten, Parallel generalized born implicit solvent calculations with namd, *J. Chem. Theory Comput.* 7 (2011) 3635–3642.
- [13] D.M. York, M. Karplus, A smooth solvation potential based on the conductor-like screening model, *J. Phys. Chem. A* 103 (1999) 11060–11079.
- [14] S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction, *J. Comp. Chem.* 27 (2006) 1787–1799.
- [15] S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory, *J. Comput. Chem.* 32 (2011) 1456–1465.
- [16] Y. Zhao, D.G. Truhlar, Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions, *J. Phys. Chem. A* 109 (2005) 5656–5667.
- [17] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, *J. Cheminform.* 4 (2012) 17.
- [18] F. Neese, The orca program system, *WIREs Comput. Mol. Sci.* 2 (2012) 73–78.
- [19] A. Sharenko, C. Mackeen, L. Jewell, F. Bridges, M.F. Toney, Evolution of iodoplumbate complexes in methylammonium lead iodide perovskite precursor solutions, *Chem. Mater.* 29 (2017) 1315–1320.
- [20] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu, *J. Chem. Phys.* 132 (2010) 154104.
- [21] F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: design and assessment of accuracy, *PCCP* 7 (2005) 3297–3305.
- [22] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu, *J. Chem. Phys.* 132 (2010) 154104.
- [23] K.A. Peterson, D. Figgen, E. Goll, H. Stoll, M. Dolg, Systematically convergent basis sets with relativistic pseudopotentials. ii. small-core pseudopotentials and correlation consistent basis sets for the post-d group 16–18 elements, *J. Chem. Phys.* 119 (2003) 11113–11123.
- [24] H. Kruse, S. Grimme, A geometrical correction for the inter-and intra-molecular basis set superposition error in hartree-fock and density functional theory calculations for large systems, *J. Chem. Phys.* 136 (2012) 04B613.
- [25] A.V. Marenich, C.J. Cramer, D.G. Truhlar, Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions, *J. Phys. Chem. B* 113 (2009) 6378–6396.
- [26] A.V. Marenich, C.J. Cramer, D.G. Truhlar, Performance of sm6, sm8, and smd on the sampl1 test set for the prediction of small-molecule solvation free energies, *J. Phys. Chem. B* 113 (2009) 4538–4543.
- [27] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *J. Comp. Phys.* 117 (1995) 1–19.
- [28] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J. Phys. Chem. B* 105 (2001) 6474–6487.
- [29] L. Martínez, R. Andrade, E.G. Birgin, J.M. Martínez, Packmol: a package for building initial configurations for molecular dynamics simulations, *J. Comput. Chem.* 30 (2009) 2157–2164.
- [30] J.C. Hamill, J. Schwartz, Y.-L. Loo, The influence of solvent coordination on hybrid organic-inorganic perovskite formation, *ACS Energy Lett.* (2017).
- [31] A. Varadwaj, P.R. Varadwaj, K. Yamashita, Hybrid organic-inorganic ch3nh3pbi3 perovskite building blocks: Revealing ultra-strong hydrogen bonding and mulliken inner complexes and their implications in materials design, *J. Comput. Chem.* 38 (2017) 2802–2818.
- [32] L. Goerigk, S. Grimme, A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions, *PCCP* 13 (2011) 6670–6688.