Range-dependence of two-body intermolecular interactions and their energy components in molecular crystals ©

Cite as: J. Chem. Phys. **157**, 084503 (2022); https://doi.org/10.1063/5.0103644 Submitted: 16 June 2022 • Accepted: 01 August 2022 • Accepted Manuscript Online: 02 August 2022 • Published Online: 25 August 2022

🔟 Derek P. Metcalf, 🔟 Andrew Smith, 🔟 Zachary L. Glick, et al.

COLLECTIONS

EP This paper was selected as an Editor's Pick



ARTICLES YOU MAY BE INTERESTED IN

Density-functional theory vs density-functional fits The Journal of Chemical Physics **156**, 214101 (2022); https://doi.org/10.1063/5.0091198

A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu The Journal of Chemical Physics **132**, 154104 (2010); https://doi.org/10.1063/1.3382344

Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package

The Journal of Chemical Physics 155, 084801 (2021); https://doi.org/10.1063/5.0055522









J. Chem. Phys. **157**, 084503 (2022); https://doi.org/10.1063/5.0103644 © 2022 Author(s).

Range-dependence of two-body intermolecular interactions and their energy components in molecular crystals ©

Cite as: J. Chem. Phys. 157, 084503 (2022); doi: 10.1063/5.0103644 Submitted: 16 June 2022 • Accepted: 1 August 2022 • Published Online: 25 August 2022 Derek P. Metcalf,
Andrew Smith,
Zachary L. Glick,
Andrew Sherrill^a)

AFFILIATIONS

Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry and School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, USA

^{a)}Author to whom correspondence should be addressed: sherrill@gatech.edu

ABSTRACT

Routinely assessing the stability of molecular crystals with high accuracy remains an open challenge in the computational sciences. The many-body expansion decomposes computation of the crystal lattice energy into an embarrassingly parallel collection of computations over molecular dimers, trimers, and so forth, making quantum chemistry techniques tractable for many crystals of small organic molecules. By examining the range-dependence of different types of energetic contributions to the crystal lattice energy, we can glean qualitative understanding of solid-state intermolecular interactions as well as practical, exploitable reductions in the number of computations required for accurate energies. Here, we assess the range-dependent character of two-body interactions of 24 small organic molecular crystals by using the physically interpretable components from symmetry-adapted perturbation theory (electrostatics, exchange-repulsion, induction/polarization, and London dispersion). We also examine correlations between the convergence rates of electrostatics and London dispersion terms with molecular dipole moments and polarizabilities, to provide guidance for estimating convergence rates in other molecular crystals.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0103644

I. INTRODUCTION

The different possible three-dimensional arrangements of a crystal lattice, known as polymorphs, can underlie vast changes in physical properties.^{1–3} This is especially relevant in pharmaceuticals, where different polymorphs of the same molecule can impact stability in solvent, compressibility, and even pharmacological efficacy.⁴⁻ Since the stability of polymorphs can vary by small amounts, sometimes less than a kilojoule per mole, predictive rank-ordering becomes an immense computational challenge.⁶ Therefore, consistent evaluation of crystal lattice energies (CLEs) often requires the most accurate available quantum mechanical methods. In practice, one may opt to generate a bespoke ab initio force field, parameterized to reproduce forces from quantum mechanics, to optimize the crystal structure and obtain its energy, and even propagate dynamics to estimate entropic contributions to polymorph stability. Dispersion-corrected plane-wave density functional theory (DFT) may be used for energy estimates of a particular configuration, or

to estimate the entropic contribution with the harmonic approximation, but this may still be unreliable.^{9,10} In particular, it has been shown that dispersion-corrected plane wave DFT (PBE-D3) was unable to conserve the geometry of a known polymorph of ROY.¹¹ There, the inclusion of Hartree Fock (HF) exchange alleviated the problem in gas phase, but explicit HF exchange is usually neglected in plane wave optimizations to reduce computational expense. Similarly, higher accuracy wavefunction-based methods are generally too costly to be used in this manner. Therefore, accurate and tractable approximations of the CLE remain a lucrative target for computational scientists. The many-body expansion (MBE) represents one attempt to circumvent the computational complexity of accurate CLE prediction.¹²⁻¹⁴ In the MBE, one may decompose the energy of a cluster as a sum of contributions from molecular monomers, dimers, trimers, tetramers, and so forth,

$$E = \sum_{I} \Delta E_{I} + \sum_{I < J} \Delta E_{IJ}^{(2)} + \sum_{I < J < K} \Delta E_{IJK}^{(3)} + \cdots, \qquad (1)$$

where $\Delta E^{(N)}$ denotes only the *non-additive* part of each *N*-mer energy. In the case of the dimer, this can be written with the supermolecular expression $\Delta E_{IJ}^{(2)} = E_{IJ} - E_I - E_J$, and the trimer contribution as $\Delta E_{IJK}^{(3)} = E_{IJK} - E_I - E_J - E_K - \Delta E_{IJ}^{(2)} - \Delta E_{IK}^{(2)} - \Delta E_{JK}^{(2)}$. Successive terms are constructed analogously. ΔE_I is simply the monomer energy difference from the gas and the cluster, a deformation energy. Explicitly computing this series to include all terms provides no computational benefit but is rigorously exact. The key benefit provided by the MBE is the observation that, very often, high-order contributions become very small after trimers or tetramers.¹⁵

With some minor modifications, the MBE may also be applied to computing a CLE. Since the CLE is defined as the energy of assembling the perfect repeating crystal from an infinitely separated non-interacting gas of monomers, CLEs may be computed on a *per-unit cell* basis. When a unit cell only contains a single unique monomer, a reference monomer I is chosen, and the expansion thereafter refers to interactions with that monomer,

$$E_{I}^{CLE} = \Delta E_{I} + \frac{1}{2} \sum_{J} \Delta E_{IJ}^{(2)} + \frac{1}{3} \sum_{J < K} \Delta E_{IJK}^{(3)} + \cdots$$
(2)

To obtain a CLE on a *per-molecule basis*, we need to divide the contribution of each *N*-mer by *N*, as in Eq. (2). The origin of this prefactor is elaborated upon in Sec. S1 of the supplementary material. This form is easily extended to unit cells with Z' > 1. Due to the regularity of crystal-packing, many *N*-mers are identical by symmetry so their contributions can be computed once and multiplied by the number of replicas. It is notable that this approach does not attempt to reduce the complexity of the underlying energy approximation method, but rather partition the computation into many smaller computations.

Unlike the supermolecular method used to compute two-body interaction energies $\Delta E_{IJ}^{(2)}$ described above, symmetry adapted perturbation theory (SAPT) is a quantum mechanical method to compute interaction energies directly.^{16,17} SAPT naturally yields energies decomposed into physically interpretable components: electrostatics, exchange-repulsion, induction/polarization, and London dispersion. These components have been used to characterize intermolecular interactions in protein-ligand systems, to enumerate interactions between amino acid sidechains, and to parameterize force fields and machine learning potentials, among many other applications.^{18–24} While previous studies have used other energy decomposition analyses to understand local interactions in the crystalline environment,^{25–28} this work seeks to probe the rangedependence of two-body interactions in the crystal and elucidate practical advice on their treatment in estimations of CLEs of new organic molecular crystals.

II. METHODS

In this study, we seek to draw generalizable conclusions on small organic molecular crystals without formal charges. The wellstudied X23 dataset contains experimental structures for 23 common small molecules encoded as Crystallographic Information Files (CIFs),^{29,30} which we additionally supplemented with one structure for ice for this study. We believe this to be a representative set for our purposes since it contains molecules of various crystallographic space groups, size, and electronic structure. Figure 1 shows a representation of the crystalline supercell in the many-body framework, where the yellow molecule is used as the reference monomer from which all dimers are formed. Color differences indicate ranges where different interaction types may be dominant. Component dominance is defined by the largest contribution to the CLE as a function of distance from the reference monomer.

To deduce range-dependent two-body effects, one must generate all pairs of monomers in the larger crystal from the simple unit cell provided by the CIF. We used the CrystaLattE program previously developed by this group to tessellate each asymmetric unit according to its crystallographic space group, copy the unit cell by translation, remove redundant particles, identify all monomers, and generate a set of symmetry-unique dimers with their corresponding replica factors.¹²

The Psi4 quantum chemistry program was used to compute SAPT components of each dimer as well as the molecular dipoles and polarizabilities.³¹ Specific computational details for the usage of CrystaLattE and Psi4 can be found below. All structures and results reside in the supplementary material.

A. CrystaLattE for dimer generation

As input, CrystaLattE requires a valid CIF for a molecular crystal containing its crystallographic space group, (fractional or Cartesian) atomic coordinates, and unit cell side lengths and angles. The user additionally must specify the desired cutoff distances, specifically the dimer cutoff for this study, which determines which dimers to include in the MBE based on closest-contact proximity. We used a large dimer cutoff of 60 Å to unambiguously determine converged two-body CLEs for all crystals at our chosen level of theory. With this information, CrystaLattE automatically determines a sufficiently large supercell of molecules and generates the set of symmetryunique dimers within the supercell in the form of Psi4 input files with the user's choice of level of theory and basis set superposition error correction. The input file includes the number of symmetryequivalent replicas of the dimer within the supercell for CLE computation and the closest-contact distance between the monomers for



FIG. 1. A cross-section of the triazine crystal with the many-body reference monomer in yellow. Approximate regimes of dominant exchange-repulsion, electrostatic, and London dispersion energies are indicated by green, red, and orange, respectively. In triazine crystal, dispersion is surprisingly prominent at ranges beyond the convergence of electrostatics.

simplified analysis. Replica identification becomes increasingly challenging as the number of monomers and the distances between them increases. We use an approach for replica identification inspired by techniques from chemical machine learning, as described in the CrystaLattE paper,¹² which we believe is an improvement over some other possibilities. Nevertheless, empirically, we found that identifying replicas of ammonia dimers at closest-contact distances greater than 30 Å is especially difficult (the eigenvalues of the Coulomb matrix become very similar between symmetry-equivalent and nonsymmetry-equivalent dimers at long range). Since even the longestrange energetic contributions for ammonia converge well before this point, we use a 30 Å cutoff for ammonia and 60 Å for the remaining 23 molecules.

B. Symmetry-adapted perturbation theory (SAPT)

For the analysis in this work, we used the lowest truncation of SAPT, SAPT0, which employs zeroth-order intramonomer correlation (i.e., none) and second-order intermonomer interaction, based on a reference obtained by using Hartree–Fock for separated monomers. Previous studies have shown that SAPT0 paired with the augmented correlation-consistent polarized valence double zeta (aug-cc-pVDZ)³² basis set achieves low errors relative to reference coupled cluster computations, at least in part due to cancelation between basis set incompleteness error and the lack of more extensive treatment of electron correlation.³³ We do expect SAPT0 to yield reasonable component energies, which will enable an interpretable range-dependent analysis of the two-body contribution to the CLE. We specifically examine the electrostatic, exchange, induction, and dispersion energies from SAPT0. All computations employ the density fitting approximation, and the self-consistent field routine is tightly converged in energy to 10^{-10} E_h and density to 10^{-10} .

C. Monomer properties

We computed molecular dipoles and static polarizabilities for use in predicting convergence behavior. For molecular dipoles, we used the Hartree–Fock wavefunction in the same aug-cc-pVDZ

TABLE I. A summary of molecular and two-body crystalline characteristics of X23 and ice.

				Cumulative two-body energy (kJ mol ^{-1})			
Molecule	Dimers	Unique dimers	Elst.	Exch.	Ind.	Disp.	Tot.
Trioxane	10 380	1 740	-39.2	53.9	-10.8	-65.2	-61.3
Ethyl carbamate	8 797	6 594	-72.1	73.3	-22.3	-65.0	-86.1
Oxalic acid α	13 326	4 167	-124.5	122.5	-38.3	-85.0	-125.3
Uracil	9 2 8 4	5 800	-120.2	108.7	-38.0	-99.8	-149.4
Pyrazine	10 280	2 070	-41.0	54.2	-9.6	-80.1	-76.5
Triazine	10616	1 246	-36.8	38.8	-6.8	-65.3	-70.2
Oxalic acid β	13 302	4 987	-155.9	168.0	-67.3	-89.9	-144.9
Imidazole	12 001	7 495	-99.0	122.2	-44.3	-85.1	-106.2
Adamantane	5714	756	-9.1	29.3	-2.1	-79.7	-61.5
Cyanamide	19 077	10733	-72.9	66.0	-25.4	-51.5	-83.9
Pyrazole	11 923	8 940	-66.3	80.2	-19.4	-72.2	-77.7
Benzene	9 100	2841	-30.6	59.0	-8.4	-90.9	-70.9
Naphthalene	6 668	2 504	-39.5	79.2	-10.8	-140.8	-111.9
Formamide	17 835	11 145	-81.4	68.4	-27.1	-44.5	-84.6
Ice	28 943	11 153	-69.3	69.2	-16.9	-18.3	-35.3
Ammonia ^a	3 7 5 0	627	-39.9	39.0	-9.7	-24.0	-34.6
Carbon dioxide	21 528	1 775	-23.5	23.5	-3.5	-30.9	-34.5
Succinic acid	9 2 2 8	3 461	-156.2	182.1	-64.4	-111.4	-149.8
Hexamine	6 470	206	-48.2	66.5	-10.4	-101.6	-93.7
Cytosine	9 2 4 8	4 624	-160.5	150.8	-62.7	-118.9	-191.4
Urea	14112	1 988	-126.7	111.7	-46.3	-65.9	-127.2
1,4-Cyclohexanedione	7 852	3 926	-72.9	85.9	-25.0	-102.2	-114.1
Acetic acid	13 190	6 595	-91.8	101.1	-39.8	-55.0	-85.4
Anthracene	5 272	1 973	-50.1	101.2	-14.3	-190.7	-153.9
Total	277 896	107 346					

^aUsing a reference convergence distance of 30 Å, not 60 Å.

basis set to remain consistent with SAPT0 electrostatics. For our analysis, we used only the magnitudes of these vectors as to isolate the monomeric property dependence from the relative dipole orientations observed in the larger crystal. Polarizabilities were computed with coupled cluster singles and doubles (CCSD) in the same aug-cc-pVDZ basis.

We draw a simple connection between mean static polarizabilities and the convergence rate of London dispersion in these crystal systems. An empirical dispersion estimate between all pairs of molecules (or atoms) in a crystal may be given by

$$E_{disp} = \sum_{AB} \sum_{n=6,8} \sum_{T} \frac{C_n^{AB}}{(|\mathbf{r}_{AB} + \mathbf{T}|)^n},$$
(3)

where *AB* denotes pairs of molecules (or atoms), C_n^{AB} is the isotropic dispersion coefficient of order *n*, **T** is the unit cell translation vectors, and \mathbf{r}_{AB} are the intermolecular displacement vectors. There have been several proposals for close range damping to eliminate the non-physical divergence at \mathbf{r}_{AB} , $\mathbf{T} = [0, 0, 0]$,^{34–38} which we ignore for simplicity of analysis and because our estimates are constructed for distances where damping is negligible. This energy may instead be approximated as an energy of the reference unit cell with an external homogeneous continuum of mutually polarizable electron density, with higher-order dispersion coefficients truncated,

$$E_{disp} \approx \int_{|\mathbf{r}_A|>0}^{\infty} \frac{4\pi \mathscr{C}_6^A}{|\mathbf{r}_A|^4} d\mathbf{r}_A, \tag{4}$$

where \mathscr{C}_{6}^{4} represents the continuum analog of the isotropic dispersion coefficient, and the integral is convergent for $|\mathbf{r}_{A}| > 0$ but only applicable at ranges where damping is negligible.

For homocrystals that are approximately packed with the same efficiency, one would expect a linear relationship between the monomeric mean static polarizability $\bar{\alpha}$ and the continuum dispersion coefficient: $\mathscr{C}_6^A \approx m\bar{\alpha}_A$ for a given molecule A. Therefore, fitting the single parameter *m* for some molecular crystals should provide accurate cutoff and energy information on others.^{39,40} The working equation to deduce cutoff distances for a given energy tolerance E_{disp}^{tol} is, thus, given by

$$r_A^{cut} \approx \left(E_{disp}^{tol} \right)^{-1/3} \left(\frac{4\pi}{3} \mathscr{C}_6^A \right)^{1/3},\tag{5}$$

a simple manipulation of Eq. (4) where r_A^{cut} represents *r* at which the remaining (neglected) dispersion energy is equal to E_{disp}^{tol} . We fit a value for *m* and demonstrate transferability across molecular crystals in the dataset.

III. RESULTS AND DISCUSSION

Total values of each SAPT component given in Table I illustrate the components most responsible for binding in each system. Computed molecular dipoles and polarizabilities can be found in the supplementary material. The number of unique dimers in the 60 Å cutoff sphere is typically a small fraction of the total number of dimers in the sphere. Overall, over 60% of explicit SAPT computations could be discarded in this study by identifying symmetryunique replicas. In total, this study of 24 molecular crystals entailed 107 346 SAPT computations on symmetry-unique dimers.

A. Component convergence

To illustrate the convergence of the SAPT0 interaction energy components in the crystalline setting, we determine the distance at which each component has converged to within 1 and 0.5 kJ mol⁻¹ of the asymptotic total defined by the 60 Å dimer cutoff distance. For example, Fig. 2 shows the cumulative SAPT0 component and total energies of the cyanamide and oxalic acid α crystals as the dimer cutoff increases. Different molecules can exhibit drastically different convergence character, both in magnitude and the relative contributions and rate of each component.

Applications may have variable error tolerance, so both 1.0 and 0.5 kJ mol⁻¹ are shown as our recommendations differ depending on the value of this metric. The tighter convergence criterion



FIG. 2. The cumulative two-body SAPT0 crystal lattice energies of cyanamide (top) oxalic acid α (bottom). Vertical dotted lines indicate convergence to the 60 Å reference CLE at a tolerance of 1.0 kJ mol⁻¹.

		Dist			
Molecule	Electrostatics	Exchange	Induction	Dispersion	Total
Trioxane	12.3 (23.0)	2.6 (2.6)	2.6 (2.6)	8.5 (11.1)	13.8 (23.1)
Ethyl carbamate	8.3 (11.4)	2.9 (2.9)	2.9 (3.5)	8.2 (10.8)	10.4 (15.9)
Oxalic acid <i>α</i>	5.8 (10.4)	3.0 (3.0)	3.0 (4.4)	8.7 (11.1)	11.1 (13.1)
Uracil	15.4 (27.6)	3.3 (3.5)	5.9 (7.4)	10.0 (13.0)	15.5 (28.9)
Pyrazine	3.5 (7.7)	3.5 (3.5)	3.5 (3.5)	9.7 (12.6)	10.8 (16.1)
Triazine	3.6 (6.4)	3.6 (3.6)	3.0 (3.6)	9.5 (12.2)	9.8 (13.6)
Oxalic acid β	8.8 (10.9)	3.4 (3.4)	3.4 (4.2)	8.7 (11.0)	10.9 (16.0)
Imidazole	12.3 (24.5)	3.4 (3.4)	4.9 (6.4)	9.3 (12.3)	13.9 (22.0)
Adamantane	2.7 (2.7)	2.7 (2.7)	2.5 (2.7)	10.1 (13.1)	12.6 (19.3)
Cyanamide	18.7 (33.1)	3.4 (3.4)	5.8 (7.7)	7.8 (9.9)	18.7 (33.1)
Pyrazole	5.0 (10.0)	3.0 (3.0)	3.0 (4.3)	9.1 (11.9)	11.2 (14.1)
Benzene	4.7 (10.0)	2.7 (2.7)	2.7 (2.7)	10.4 (13.7)	12.5 (20.8)
Naphthalene	2.9 (6.7)	2.9 (2.9)	2.9 (2.9)	12.0 (16.4)	12.3 (17.7)
Formamide	17.1 (29.5)	3.4 (3.4)	5.3 (6.7)	6.9 (9.1)	15.9 (29.5)
Ice	7.0 (13.2)	1.9 (1.9)	1.9 (3.8)	4.5 (5.7)	7.6 (13.2)
Ammonia ^a	6.7 (16.8)	2.6 (2.6)	2.6 (2.6)	5.6 (7.1)	10.5 (17.4)
Carbon dioxide	4.7 (4.7)	3.1 (3.1)	3.1 (3.1)	6.4 (8.8)	6.2 (10.1)
Succinic acid	6.4 (10.0)	3.1 (3.1)	3.1 (3.1)	9.2 (12.1)	11.3 (14.9)
Hexamine	2.9 (8.7)	2.9 (2.9)	2.9 (2.9)	11.1 (13.6)	13.1 (22.3)
Cytosine	28.6 (48.0)	3.2 (3.2)	8.0 (10.3)	10.3 (13.6)	28.9 (48.0)
Urea	26.5 (45.2)	2.5 (4.0)	5.5 (7.9)	8.1 (9.9)	24.1 (45.2)
1,4-Cyclohexanedione	10.5 (14.7)	2.7 (2.7)	3.3 (4.0)	9.7 (12.8)	11.4 (16.9)
Acetic acid	10.4 (17.4)	3.2 (3.2)	3.2 (4.1)	7.3 (9.6)	11.1 (17.9)
Anthracene	6.2 (8.5)	2.7 (2.7)	2.7 (2.7)	14.1 (18.8)	14.0 (19.3)

TABLE II. The distance to convergence of each SAPT component and the total two-body interaction energy for X23 and ice. Convergence is defined as within 1 kJ mol⁻¹ of the CLE at the 60 Å reference cutoff. Parenthetical values use a convergence criterion of 0.5 kJ mol⁻¹. Bold typeface indicates the slowest-converging component for each system.

^aComputed with respect to a reference convergence distance of 30 Å, not 60 Å.

highlights large discrepancies between molecules—the electrostatics of some molecules are well converged to under 0.5 kJ mol⁻¹ by 3 Å (essentially the first layer of monomers), while others remain unconverged at 40 Å. Following is a qualitative analysis of the component convergence phenomena, which will be supported in Sec. III B by some quantitative justification. Table II shows the large variety in convergence distances and interaction types between crystals.

Consistent with intuition, the electrostatic contribution of the SAPT0 interaction energy often exhibits convergence distances far greater than other interaction components. The range of electrostatic interactions is well-studied and understood but is especially pertinent in the context of crystals where the interactions can compound due to symmetry and the number of such interactions grows as r^3 in dimer cutoff radius. Importantly, none of the crystals in X23 are cocrystals and, therefore, cannot support total molecular charges; in fact, X23 additionally lacks zwitterions. This observation indicates that dipolar and higher-order molecular multipolar interactions, despite their more rapid decay with distance, are solely responsible for the persistence of electrostatic CLE fluctuations even at long range. Unlike other terms, electrostatics can be attractive or repulsive, sometimes causing oscillations at distances where the other terms smoothly and monotonically

approach an asymptotic result. We note that the many-body expansion about the crystalline unit cell, as opposed to the molecule, is well-understood and generally converges predictably. Because the unit cell is neutral and the cell dipole is either zero (due to symmetry) or small (for which a surface dipole term exists),⁴¹ most dipole–dipole terms are zero or small and higher-order multipoles converge absolutely. Evidently, empirical estimates of electrostatics (specifically Ewald sums to incorporate the asymptotic effect) should be employed for slowly converging systems at long distances as accurate quantum mechanical computations become intractable, while simple models should become accurate at sufficiently long range.

For all systems, exchange-repulsion is fully converged within the first one or two shells of neighboring molecules, at worst requiring a dimer cutoff of less than 5 Å. This too is consistent with intuition, as the exponential decay of the exchange energy drastically outscales the r^3 growth in number of dimers.

Induction energies are predominantly quite local for these crystals, with all crystals converging to within 1.0 kJ mol⁻¹ of the reference by 8 Å and just four systems requiring more than a 7 Å cutoff to achieve 0.5 kJ mol⁻¹ convergence. This is, again, partially a reflection of the lack of full formal charges on any of the atoms in our dataset. Of the slowly converging systems, cytosine and uracil are among the largest and most polar in the dataset. However, smaller systems can also converge slowly: cyanamide and urea, with just three and four heavy atoms, respectively, have among the most gradual two-body induction convergence. We elaborate on this disparity in Sec. III B.

Finally, London dispersion contributes heavily to the CLE at surprising ranges: again, the competition between the decay of the component (roughly r^{-6}) and the r^3 enumeration of dimers results in important contributions to many crystals beyond 10 Å. Large delocalized systems are the most challenging to converge, with anthracene requiring a cutoff distance greater than 14 Å to reach 1.0 kJ mol⁻¹ agreement. Unlike induction, however, convergence distances are more uniformly distributed between 3 and 12 Å.

The total two-body CLE almost always converges at the same rate as the most slowly converging SAPT component. Some variability arises due to favorable cancelation of components, yielding a flat total energy profile, and some compound unfavorably, requiring slightly longer cutoffs than the slowest component to converge the CLE, often 1–2 Å longer, but occasionally more.

B. Monomer property correlations

In previous sections, we alluded to the understanding of component CLE convergence rates based on intuitive chemical concepts



FIG. 3. The distance to convergence of SAPT0 electrostatics, induction, and dispersion for each X23 crystal and ice. Dark points represent convergence to 1.0 kJ mol⁻¹ for a given system and the corresponding light point represents convergence to 0.5 kJ mol⁻¹.

and computable quantities. The electrostatic energy decay rate, for example, is closely related to the static charge distribution of a single monomer in vacuum. By representing this charge distribution in a multipole expansion, we can construct a compact (though approximate) description of the electrostatics while only considering a single monomer. Although a unit cell cannot generally carry a net charge, ionic cocrystals and zwitterions present strong charge localization within the unit cell and represent an edge case to this principle, which will be reserved for future study.

The first generally non-zero multipole moment in crystals is then the molecular dipole, which we compute and relate to two-body SAPT component convergence in the crystal. Indeed, the cutoff distance required to converge the electrostatic energy is correlated with the magnitude of the molecular dipole, as shown in the first plot of Fig. 3. Several factors make this correlation imperfect; as discussed, the dipole is an inexact representation of the density, higher-order multipole moments are ignored, and perhaps most importantly, the relative orientation of multipoles within the crystal is unknown a priori. For systems with no molecular dipole, the crystalline electrostatic energies are well converged by 20 Å, comparable to the second slowest-scaling component, dispersion. At larger dipoles, however, proportionally longer contacts must be considered. Cytosine has the largest dipole of any system considered by over 50% and is able to challenge the very generous 60 Å reference cutoff used for this study. As electrostatics is an interaction between static charge distributions, one would not anticipate a correlation between convergence distances and polarizability, an intuition verified in the second panel of Fig. 3.

Convergence of induction correlates effectively with the magnitude of molecular dipoles. Unlike electrostatics, however, molecules without a dipole are well converged by the first shell of neighbors. Cytosine requires a 10 Å shell, the largest of any in the set. Convergence rates should also be related to molecular polarizabilities,



FIG. 4. The SAPT0-predicted continuum dispersion coefficients as a function of static molecular polarizabilities from coupled-cluster singles and doubles calculations on monomers. Each line represents a different least-squares fit of the slope using a random three-point fitting set.



FIG. 5. Predicted SAPT0 dispersion convergence distances via an isotropic continuum dispersion model. 1.0 and 0.5 kJ mol⁻¹ energy tolerances are shown for each crystal in the dataset. Results are shown for each of an ensemble of 500 linear models fit to only three random training points.

but the size of the dipole shows a stronger correlation, as shown in Fig. 3; even highly polarizable molecules exhibit only local induction effects if all interacting molecules are without a dipole. Likewise, molecules that are only modestly polarizable still *do* polarize in the presence of a strong electric field like that induced by large neighboring dipoles. The upper right regions of the polarizability-induction and polarizability-electrostatics plots are notably sparse. We note that one would not generally expect these to be sparse, but rather it is likely an artifact of a small dataset in the high polarizability region. Indeed, the two most polarizable molecules in X23 are naphthalene and anthracene, which coincidentally have no dipole.

London dispersion interactions can be expressed as integrations over frequency of the products of imaginary-frequencydependent polarizabilities of each monomer, and so not surprisingly, Fig. 3 shows a striking correlation with the convergence distance for dispersion with (static) molecular polarizability. Indeed, the required distance cutoff scales approximately linearly with molecular polarizability.

We fit a single scaling parameter *m* to relate the computed static polarizability $\bar{\alpha}$ to the continuum dispersion coefficient analog \mathscr{C}_6^A in Eq. (4), which is evaluated by adding all SAPT0 contributions from 45 to 60 Å, at that point the underlying assumptions of anisotropy, homogeneity, and continuity are likely appropriate. Then, we used *m* directly for all molecules to predict the cutoff distance at which the neglected (remaining) dispersion energy is only 1.0 kJ mol⁻¹ according to Eq. (5). This correspondence is shown in Fig. 4, which illustrates the insensitivity of *m* to choice of fitting set and the accuracy with which cutoff distances are predicted, with access only to the static molecular polarizability of the molecule. We note that the predicted cutoffs are almost always larger than the actual required cutoffs, so the model tends to provide a safe estimate.

Five hundred different models fit to random combinations of three training points show that the learned coefficient is transferable, at least among small neutral molecules. Propagating that error to the prediction of cutoff distances produces the data in Fig. 5, which illustrates that even the worst choice of training points produces accurate predictions. Using more training points even allows reasonable predictions of the remaining dispersion energy itself, with quality growing with the cutoff distance.

The leading sources of error in this model are likely the packing efficiency and continuum assumptions, but the results indicate these assumptions are appropriate for this dataset at the distances required for 1.0 kJ mol⁻¹ convergence. Convergence of dispersion in crystals of small organic molecules lacking formal charges can therefore be predicted *a priori* via static molecular polarizabilities with near-Ångstrom fidelity. These results may not extend to dispersion energies for crystals containing formal charges or whose packing efficiency may be significantly different from those in this dataset.

IV. CONCLUSIONS

The delicate interplay of non-covalent forces between pairs of molecules comprises most of the 0 K crystal lattice energy for small molecular crystals. To understand how these forces vary by molecule and range, we computed the SAPT interaction energies of 107 K dimers across 24 molecular crystals. We show the rangedependent variation of electrostatics, exchange-repulsion, polarization, and London dispersion in the crystalline setting, with special focus on the convergence rates of each effect. These results build intuition about the nature of non-covalent interactions within crystals and can inform future range-dependent hierarchical methods for CLE prediction.

Although the decay rates of different intermolecular forces between two molecules are generally understood, how they manifest themselves in molecular crystals, and the interplay of these decay rates with the concomitant increase in the number of dimers with increasing distance r, has not been well explored. Electrostatic interactions, which in the case of static dipole interaction decay as r^{-3} , can persist in magnitude at surprising ranges in crystals since the number of such interactions *grows* as r^3 . Cytosine is one extreme example, holding the largest molecular dipole in the X23 dataset and requiring a nearly 30 Å interacting shell to converge to within 1 kJ mol⁻¹ of reference energies computed using a 60 Å cutoff. This emphasizes the need for simple empirical forms and infinite-range corrections such as Ewald summation. By contrast, small molecules without dipoles need not be treated for electrostatics beyond small 4–10 Å shells.

The rate of convergence of induction/polarization energy is also highly dependent on the magnitude of the molecular dipole, though requiring less dramatic distances even in the most extreme cases. London dispersion, nominally decaying as r^{-6} per dimer, can, nevertheless, require computing energies of dimers as far separated as 14 Å to reach convergence. By employing a continuum dispersion model, it is possible to use molecular polarizabilities to accurately estimate cutoff distances at which dispersion interactions become negligible in small organic crystals lacking formal charges.

In future work, we will investigate convergence rates of twoand three-body interactions at higher levels of theory. Additionally, we hope to further explore the usage of a range-dependent hierarchy of methods in CLE prediction to reduce computational expense.

SUPPLEMENTARY MATERIAL

The supplementary material includes all of the requisite .cif files, geometries of the generated supercells, all dimers, and the corresponding computed SAPT0 interaction energy components and replica numbers. Figures for all X23 systems and ice corresponding to Fig. 2, as well as an experiment on the impact of anisotropy on the fit in Fig. 5, are contained in the supplementary material. The CrystaLattE program used for *N*-mer generation is available at https://github.com/carlosborca/CrystaLattE.

ACKNOWLEDGMENTS

Derek P. Metcalf acknowledges support from the U.S. National Science Foundation Graduate Research Fellowship, Grant No. DGE-1650044. Andrew Smith acknowledges funding via the NSF REU program, Grant No. CHE-1852511. Additional financial support was provided by NSF Grant No. CHE-1955940. This research was supported, in part, by research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, and by the Georgia Tech Hive Cluster, funded by the NSF through Grant No. MRI-1828187.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Derek P. Metcalf: Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Supervision (supporting); Writing – original draft (lead). **Andrew Smith**: Data curation (equal); Formal analysis (supporting); Investigation (lead). **Zachary L. Glick**: Methodology (equal); Software (equal). **C. David Sherrill**: Conceptualization (lead); Formal analysis (supporting); Funding acquisition (lead); Methodology (equal); Supervision (lead); Writing – review & editing (lead).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

¹K. Raza, P. Kumar, S. Ratan, R. Malik, and S. Arora, SOJ Pharm. Pharm. Sci. 1, 10 (2014).

- ²L. Yu, Acc. Chem. Res. **43**, 1257 (2010).
- ³A. Llinàs and J. M. Goodman, Drug Discovery Today 13, 198 (2008).
- ⁴R. Censi and P. Di Martino, Molecules 20, 18759 (2015).

⁵S. R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubauer, B. A. Narayanan, M. Soldani, D. Riley, and K. McFarland, Org. Process Res. Dev. 4, 413 (2000).

⁶A. J. Cruz-Cabeza, S. M. Reutzel-Edens, and J. Bernstein, Chem. Soc. Rev. 44, 8619 (2015).

⁷A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio, Jr., A. Dzyabchenko, B. P. van Eijck, D. M. Elking, J. A. van den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C.-A. Gatsiou, T. S. Gee, R. de Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. M. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. de Jong, J. Kendrick, N. J. J. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom, A. E. Masunov, P. McCabe, D. P. McMahon, H. Meekes, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szalewicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis, A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. de Wijs, J. Yang, Q. Zhu, and C. R. Groom, Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. 72, 439 (2016).

⁹J. G. Brandenburg and S. Grimme, Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. **72**, 502 (2016).

¹⁰L. Iuzzolino, P. McCabe, S. L. Price, and J. G. Brandenburg, Faraday Discuss. 211, 275 (2018).

¹¹J. Nyman, L. Yu, and S. M. Reutzel-Edens, CrystEngComm 21, 2080 (2019).

¹²C. H. Borca, B. W. Bakr, L. A. Burns, and C. D. Sherrill, J. Chem. Phys. 151, 144103 (2019).

¹³P. J. Bygrave, N. L. Allan, and F. R. Manby, J. Chem. Phys. **137**, 164102 (2012).

¹⁴S. Hirata, K. Gilliard, X. He, J. Li, and O. Sode, Acc. Chem. Res. **47**, 2721 (2014).
 ¹⁵U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, J. Chem. Phys. **135**, 224102 (2011).

¹⁶K. Szalewicz, Wiley Interdiscip. Rev.: Comput. Mol. Sci. 2, 254 (2012).

¹⁷K. Patkowski, Wiley Interdiscip. Rev.: Comput. Mol. Sci. 10, e1452 (2020).

¹⁸X.-K. Chen, B. W. Bakr, M. Auffray, Y. Tsuchiya, C. D. Sherrill, C. Adachi, and J.-L. Bredas, J. Phys. Chem. Lett. **10**, 3260 (2019).

¹⁹R. M. Parrish, D. F. Sitkoff, D. L. Cheney, and C. D. Sherrill, Chem. - Eur. J. 23, 7887 (2017).

²⁰ J. G. McDaniel and J. R. Schmidt, J. Phys. Chem. A 117, 2053 (2013).

²¹ J. G. McDaniel and J. R. Schmidt, Annu. Rev. Phys. Chem. **67**, 467 (2016).

²³D. P. Metcalf, A. Koutsoukas, S. A. Spronk, B. L. Claus, D. A. Loughney, S. R. Johnson, D. L. Cheney, and C. D. Sherrill, J. Chem. Phys. **152**, 074103 (2020).

²⁴Z. L. Glick, D. P. Metcalf, A. Koutsoukas, S. A. Spronk, D. L. Cheney, and C. D. Sherrill, J. Chem. Phys. **153**, 044112 (2020).

²⁵Q. Zeng, Y. Ma, J. Li, and C. Zhang, CrystEngComm 19, 2687 (2017).

²⁶ A. Parkin, G. Barr, W. Dong, C. J. Gilmore, D. Jayatilaka, J. J. McKinnon, M. A. Spackman, and C. C. Wilson, CrystEngComm 9, 648 (2007).

²⁷M. A. Spackman and J. J. McKinnon, CrystEngComm 4, 378 (2002).

²⁸M. A. Spackman, J. J. McKinnon, and D. Jayatilaka, CrystEngComm 10, 377 (2008).

²⁹A. Otero-De-La-Roza and E. R. Johnson, J. Chem. Phys. 137, 054103 (2012).

³⁰G. A. Dolgonos, J. Hoja, and A. D. Boese, Phys. Chem. Chem. Phys. **21**, 24333 (2019).

³¹ D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. Schaefer, A. Y. Sokolov, K. Patkowski, A. E. Deprince, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, J. Chem. Phys. **152**, 184108 (2020).

⁸G. J. O. Beran, Chem. Rev. 116, 5567 (2016).

²²M. Tafipolsky and K. Ansorg, J. Chem. Theory Comput. 12, 1267 (2016).

³² E. Papajak, J. Zheng, X. Xu, H. R. Leverentz, and D. G. Truhlar, J. Chem. Theory Comput. 7, 3027 (2011).

³³T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, and C. D. Sherrill, J. Chem. Phys. **140**, 094106 (2014).

³⁴ J. Witte, N. Mardirossian, J. B. Neaton, and M. Head-Gordon, J. Chem. Theory Comput. **13**, 2043 (2017).

³⁵D. G. A. Smith, L. A. Burns, K. Patkowski, and C. D. Sherrill, J. Phys. Chem. Lett. 7, 2197 (2016).

- ³⁶H. Schröder, A. Creon, and T. Schwabe, J. Chem. Theory Comput. **11**, 3163 (2015).
- ³⁷S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
- ³⁸S. Grimme, S. Ehrlich, and L. Goerigk, J. Comput. Chem. **32**, 1456 (2011).
- ³⁹J. Tao, J. P. Perdew, and A. Ruzsinszky, Proc. Natl. Acad. Sci. U. S. A. **109**, 18 (2012).
- ⁴⁰ J. Tao and A. M. Rappe, J. Chem. Phys. 144, 031102 (2016).
- ⁴¹V. Ballenegger, J. Chem. Phys. **140**, 161102 (2014).