Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

Crystal structure predictions for 4-amino-2,3,6-trinitrophenol using a tailor-made first-principles-based force field

Michael P. Metz,[1,2] Muhammad Shahbaz,[1,3] Hongxing Song,[4] Leslie Vogt-Maranto,[4] Mark E. Tuckerman,[4,5,6,*] and Krzysztof Szalewicz[1,†]

[1] *Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716, USA*

[2] *Applied Research Laboratories, University of Texas at Austin, TX 78758, USA*

[3] *Department of Physics, University of Punjab, Lahore, Pakistan*

[4] *Department of Chemistry, New York University, New York, New York 10003, USA*

[5] *Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA*

[6] *NYU-ECNU Center for Computational Chemistry at NYU Shanghai, 3663 Zhongshan Road North, Shanghai 200062, China*

Abstract

Predictions of crystal structures from first-principles electronic structure calculations and molecular simulations have been performed for an energetic molecule, 4-amino-2,3,6-trinitrophenol. This physics-based approach consists of a series of steps. First, a tailor-made two-body potential energy surface (PES) was constructed with recently developed software, autoPES, using symmetry-adapted perturbation theory based on density-functional theory description of monomers [SAPT(DFT)]. The fitting procedure ensures asymptotic correctness of the PES by employing a rigorous asymptotic multipole expansion which seamlessly integrates with SAPT(DFT) interaction energies. Next, crystal structure prediction (CSP) was performed by generating possible crystal structures with rigid molecules, minimizing these structures using the SAPT(DFT) force field, and running isothermal-isobaric molecular dynamics (MD) simulations with flexible molecules based on the tailor-made SAPT(DFT) intermolecular force field and a generic/SAPT(DFT) intramolecular one. This workflow led to the experimentally-observed structure being identified as one of the forms with the lowest lattice energy, demonstrating the success of a first-principles, bottom-up approach to CSP. Importantly, we argue that the accuracy of the intermolecular potential, here the SAPT(DFT)-based potential, is determinative of the crystal structure, while generic/SAPT(DFT) force fields can be used for the structure generation step. This force field approach simplifies the CSP workflow, without significantly compromising the accuracy of the prediction.

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

I. INTRODUCTION

Organic molecular crystals are materials of critical importance in numerous industries including pharmaceuticals, electronics, vector-borne disease management, and defense. In any of these areas, the performance of a given crystalline formulation depends on its crystal structure. If a given compound can crystallize into multiple forms – commonly referred to as *polymorphism* – then mapping out the landscape of possible polymorphs becomes the first step toward selecting a polymorph with desired properties. In particular, since the existence of polymorphs can cause drugs to fail[1–4], affect the performance of organic semiconductors[5], or alter the energetic properties of explosives[6], predicting these structures and ranking them thermodynamically becomes all the more critical.

Polymorph landscapes can be generated using computational techniques tailored for crystal structure prediction (CSP). CSP is an essential step in the discovery of metastable polymorphs that may be sufficiently stable for a desired application. In CSP, the energy landscape of crystalline phases is explored to find a set of thermodynamically plausible polymorphs. Starting with information only about the molecule itself, CSP generates trial crystal structures and evaluates their lattice energies. While the idea is, in principle, straightforward, the practical implementation of CSP methods faces significant challenges in both the structure search and the calculation of accurate energies. Over the years, CSP has matured[7–9], and its progress has been followed in a series of blind test competitions organized by the Cambridge Crystallographic Data Centre (CCDC)[10–15]. Strategies for structure searches include random searches[16,17], genetic algorithms[18–20], simulated annealing[21], or Monte Carlo methods[22]. For energy calculations, significant advances have been made in the development of reliable force fields[23,24], and, at the same time, the application of electronic structure calculations with periodic boundary conditions to molecular crystals is also becoming increasingly important[25–28]. While the latter can provide very accurate energetics, it is also computationally much more expensive and has been mainly used for the final ranking of polymorphs after performing the structure search with a force field. However, a separate energy ranking step can be avoided if the force field is of sufficient quality. Generic, empirical force fields are generally insufficiently accurate and one has to resort to tailor-made force fields fitted to *ab initio* calculations. The first CSPs based on first-principles potential energy surfaces (PESs) have been performed by Podeszwa *et al.* in Refs. 29–31 using a workflow including crystal packing, lattice energy minimizations, and molecular dynamics (MD) simulations. A similar approach, except without the MD step, was used by Misquitta *et al.*[32] Both approaches fitted PESs to symmetry-adapted perturbation theory[33] based on density functional theory (SAPT(DFT)) calculations for dimers[34–42]. Some of us and coauthors have previously established a CSP workflow[15,43] combining a random search together with MD and enhanced sampling[44,45] to predict low

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

free-energy crystal structures at experimentally relevant conditions[15,43,46–48]. This workflow is now being employed and adapted by other groups[49]. It has been shown, in particular, that accurate landscapes[15,24] can be predicted within this protocol using molecule-specific force fields generated using SAPT(DFT).

In this article, we employ a version of our CSP workflow using a tailor-made SAPT(DFT) force field to determine the structure of an energetic material, 4-amino-2,3,6-trinitrophenol (**1**, see Fig. 1). The key component of our workflow is an accurate intermolecular force field, with the total interaction energy determined based on the sum of pairwise interactions between molecules. Compound **1** poses a challenge for our dimer-based approach to generating the molecule-specific force field as it lacks internal symmetry, exhibits a large dipole moment, and is fairly polarizable. The outline of this paper follows our CSP workflow. First, an intermolecular PES for dimer configurations of **1** has been developed from first-principles quantum mechanical calculations, i.e., by fitting the PES to *ab initio* interaction energies computed using SAPT(DFT). This step requires generation of suitable monomer geometries and subsequent calculation of *ab initio* interaction energies for sampled dimers (Sec. II A). The automated selection of dimers and fitting of the PES is controlled by the autoPES software package[24,50] (Sec. II B). Second, possible crystal structures are generated by randomly sampling monomer orientations and locations within random unit cells of a given space group. The crystals generated in this way are first optimized with rigid molecules using an empirical force field, followed by rigid molecule optimization using the SAPT(DFT) potential. Notice that this step is different from the workflow used in the 6th CCDC blind test, where a simplified version ofthe SAPT(DFT) PES was used in the second stage lattice energy minimizations. The 0 K optimizations are followed by flexible molecule MD simulations under ambient conditions to refine the structures using the SAPT(DFT) intermolecular force field plus generic/SAPT(DFT) intramolecular terms (Sec. II C). To evaluate the accuracy of our results, the generated potentials are analyzed with respect to SAPT(DFT) calculations (Sec. III A), and monomer and dimer geometries corresponding to the minima on the PES are compared to the experimental crystal structure (Sec. III B, III C). Finally, we show that with a tailor-made intermolecular force field, the experimentally-observed form is found to be one of the lowest energy crystal structures (Sec. III D).

II. METHODS

A. First-Principles Calculations

The first step in the CSP workflow is to generate an appropriate monomer geometry. The monomer geometry can be taken from an experimental structure if available, or can

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

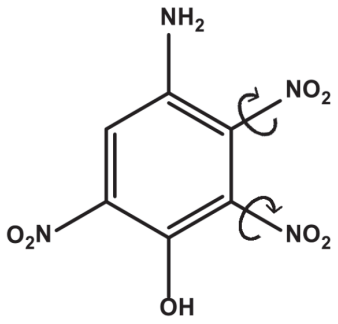

FIG. 1. The 4-amino-2,3,6-trinitrophenol (**1**) molecule. Deformations of the two easily-rotatable nitro groups (indicated by arrows) are included in the set of short-range dimer configurations used to fit the intermolecular potential.

be determined by optimizing the molecule with an electronic-structure method to obtain the so-called equilibrium (r_e) geometry. Although experimental structure is known for **1**[51], we found the r_e geometry from scratch by performing DFT optimization using the ORCA electronic structure package[52,53], the PBE0 functional[54,55], and the aug-cc-pVTZ basis set[56]. This geometry was used in PES development and the subsequent structure generation steps.

The dimer PES has been fitted to two types of *ab initio* calculations: SAPT(DFT) at close range and to an expansion in inverse powers of intermonomer separation R, built from *ab initio* computed monomer multipole moments and polarizabilities, at long range. SAPT(DFT) was used in the density-fitting version[38,57,58] coded in the SAPT2016 software package[59]. Monomer DFT calculations were performed using the ORCA electronic structure package[52,53]. The PBE functional[54] was used with the gradient-regulated asymptotic correction (GRAC)[60,61]. The ionization potential of the monomer, required for GRAC, was computed using separate PBE calculations. The aug-cc-pVDZ basis set[56] together with the corresponding correlation-energy fitted auxiliary bases from Ref. 62 were used in all cases. In SAPT calculations, the basis was in the monomer-centered 'plus' basis set (MC^+BS) form with midbond functions, see Ref. 63. Midbond function exponents and placement are as described in Ref. 64, with the midbond auxiliary basis set from Ref. 58. See Table S1 in the Supporting Information for values of some properties of **1**.

The component of the interaction energy arising from the Coulomb interaction of charge distributions of the monomers is called electrostatic energy and is given by the SAPT correction $E^{(1)}_{\rm elst}$. The antisymmetrization of the total wave functions required by the fermionic nature of electrons leads to permutations (exchanges) of electrons between monomers and when applied to zeroth-order wave functions results in the first-order exchange energy, $E^{(1)}_{\rm exch}$. The electric field due to the permanent charge distribution of monomer A (B) induces multipole moments on monomer B (A) and these multipole moments

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

interact with the permanent multipoles of monomer A (B) resulting in the induction energy, $E^{(2)}_{\rm ind}$. The long-range correlation between electrons from different monomers results in the dispersion interaction, $E^{(2)}_{\rm disp}$. The antisymmetrization of the induction and dispersion wave functions leads to the exchange-induction and exchange-dispersion energies, denoted as $E^{(2)}_{\rm exch-ind}$ and $E^{(2)}_{\rm exch-disp}$, respectively. The components $E^{(2)}_{\rm ind}$ and $E^{(2)}_{\rm disp}$ were computed from coupled Kohn-Sham (CKS) frequency-dependent density susceptibilities (FDDS) (in the former case, only at zero frequency). The fast method of Ref. 65 was used to obtain FDDSs. When this approach is used, $E^{(2)}_{\rm exch-disp}$ can be computed only in the uncoupled form. It was then scaled to approximate the CKS value as described in Ref. 41. The term $E^{(2)}_{\rm exch-ind}$ was computed from coupled amplitudes.

For significantly polar monomers, the $\delta E^{\rm HF}_{\rm int,resp}$ correction[66,67] should be included in addition to the standard SAPT(DFT) corrections. The term $\delta E^{\rm HF}_{\rm int,resp}$ accounts mostly for induction and exchange-induction effects beyond second order. The inclusion of $\delta E^{\rm HF}_{\rm int,resp}$ increases the accuracy of the SAPT(DFT) interaction energies for most systems, but also increases the associated computational cost by about 60%. A criterion is automatically checked by autoPES to determine if $\delta E^{\rm HF}_{\rm int,resp}$ should be included for a given dimer based on the magnitude of the electric dipole moment of its monomers and the induction component of its interaction energy, see Ref. 24. This criterion was met, so the $\delta E^{\rm HF}_{\rm int,resp}$ correction was included in the total interaction energy for each of the close-range grid points.

The total second-order interaction energy, $E^{(2)}$, is the sum to the corrections discussed above

$$E^{(2)} = E^{(2)}_{\rm ind} + E^{(2)}_{\rm exch-ind} + E^{(2)}_{\rm disp} + E^{(2)}_{\rm exch-disp}. \quad (1)$$

We will denote the sum of $E^{(2)}$ and $\delta E^{\rm HF}_{\rm int,resp}$ as $E^{(2)} + \delta$.

In the asymptotic region, the interaction energy is computed using the multipole expansion of the interaction potential operator rather than the much more computationally expensive SAPT(DFT) method. In the original autoPES approach, the origins of this expansion are placed at the centers of mass (COM) of monomers, i.e., the standard asymptotic expansion described in Refs. 33 and 68 is used. Here, we have used instead an extension of the distributed asymptotic expansion approach of Ref. 69. The distributed expansion is a sum of expansions located on each of the atoms. This approach converges at shorter monomer separations than the COM-COM asymptotic approach, and so is appropriate for monomers of the size investigated here. The coefficients of this expansion are computed from *ab initio* distributed monomers' charge densities and FDDSs. These monomer properties are computed using the same basis set and level of theory as is used for the close-range calculations. Therefore, the resulting interaction energies connect seamlessly to those of the close-range calculations. While in Ref. 69 only the dispersion energies were computed, requiring distributed dynamic polarizabilities, we compute here also the electrostatic and induction energies, requiring also static distributed polarizabilities and

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

multipole moments. Static distributed polarizabilities were computed by us using the same constrained density-fitting of FDDSs as is used for the dynamic polarizabilities, described in Ref. 69. The distributed multipole moments were computed from distributed monomer charge densities. Whereas the density-fitting constraint of Ref. 69 cannot be used for fitting charge densities, it is possible to devise constraints appropriate in this case. However, we have performed unconstrained fitting of charge densities. The rationale for this choice is that we use the asymptotic expansion only at very large separations where essentially any kind of distributed density is adequate. This fitting represents the total monomer density $\rho(\boldsymbol{r})$ as the sums of atomic densities $\rho_a(\boldsymbol{r})$. The latter densities are used to compute a set of multipole moments on each atom, which are subsequently used to compute electrostatic energies at asymptotic separations, see Sec. II B.

We approximate the total crystal lattice energy as a sum of two-body interactions, whereas the total interaction energy includes contributions from many-body effects, primarily in the form of polarization energy. Although the pairwise approximation is usually adequate for CSPs, the quality of predictions would certainly improve if many-body effects were included, in particular since **1** has a large dipole moment (5.9 Debye) and polarizability (152 a.u.). AutoPES is capable of developing polarizable PESs that recover many-body polarization effects, but the available CSP and MD programs cannot currently accommodate such PESs.

B. PES Generation

The autoPES software[24,50] that we used fits dimer PESs assumes rigid-geometry monomers. Although a new version of this software, flex-autoPES, can include intramonomer degrees of freedom[70], the rigid-geometry approximation was used here to reduce the complexity of the fit. If the two soft nitro group rotations were included explicitly as intramonomer degrees of freedom, the resulting PES would be 10-dimensional, with two additional degrees of freedom per monomer. Such a fit would include many more free parameters, and therefore would require a larger set of training data, leading to a much more expensive calculations than in the case of the present 6-dimensional PES. However, since the effects of varying selected internal degrees of freedom can be included using autoPES in an average way, we have done so in this case, as is discussed below.

The generation of potential energy surfaces is almost entirely automated in autoPES. The procedure can be roughly divided into five parts: asymptotic calculations, generation of a grid of close-range dimer configurations, calculation of interaction energies at each such grid point, fitting an analytic functional form to the data, and finally evaluation of the quality of the fit and iterative improvement. Each step is described in detail in Ref. 24. We give a brief outline below.

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

The purpose of the asymptotic calculations is to describe the interaction energy in as large a region of the dimer configuration space as possible using only monomer properties. This reduces the number of close-range dimer configurations required for the fit, and by extension, the overall computational cost of PES generation. This approach works when the monomers are sufficiently separated that charge overlap effects can be neglected. The region where this condition is met is determined by autoPES using geometrical criteria. A grid of about 10,000 long-range dimer configurations is selected and then the corresponding interaction energies are computed using the distributed multipole expansion.

The region where charge overlap effects cannot be neglected, which extends to about 1.5 times the radial van der Waals minimum separation and is referred to as the close-range grid region, must be adequately sampled by SAPT(DFT) calculations. The close-range grid is generated using a guided Monte Carlo procedure, which samples the entire relevant region of configuration space, while placing a higher density of grid points near the most physically important regions. A guiding potential is used to determine the approximate interaction energy and locations of local minima. More grid points are placed in more energetically negative regions, and, in addition, a small percentage of points are placed specifically near each local minimum. In the initial iteration of grid generation, the OPLS-AA force field of Jorgensen and coworkers[71] is used as the guiding potential, while in later iterations the generated PES of the previous iteration is used. The interaction energy at each close-range grid point is computed as described in Sec. II A.

As discussed above, monomer **1** includes soft intramonomer degrees of freedom, i.e., nitro-group rotation angles which can change relatively significantly from their gas-phase values in the presence of other monomers in a crystal. Since in the MD stage of our CSPs the monomers are allowed to deform, we fitted rigid-monomer, six-dimensional form of PESs to grid points including some monomer deformations. For any interacting rigid monomers, each grid point is described by a set of six coordinates: the distance between the monomers' COMs R and five Euler angles describing the relative angular orientation of the monomers. However, the functional fit form used depends only on the atom-atom distances r_{ab}, where atom a (b) belongs to monomer A (B). The PESs therefore can account for internal deformations of the monomers to a limited degree, even if the training data contains only a single, fixed geometry for each monomer. In this case, by explicitly varying coordinates which specify intramonomer deformations in addition to the six intermolecular coordinates described above, configurations including such deformations are better sampled. Four additional coordinates at each intermonomer grid point were used per dimer to represent rotation of the two nitro groups indicated in Fig. 1. The rotations were sampled in the range $\pm 17^\circ$ from the equilibrium positions. (The rotation of the third nitro group was not included because it is well constrained by a hydrogen bond to the neighboring hydroxyl group.) Note that this implicit inclusion of intramonomer degrees of freedom does not change

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

the functional form of the fit: the same functional form is simply optimized using a data set which includes the intramonomer deformations. The resulting flexibilized rigid-monomer fit for the intermolecular PES was used with a generic/SAPT(DFT) intramonomer potential for MD simulations (described in more detail in Sec. II C).

After the asymptotic and close-range interaction energies are computed, the data is fit with an analytic function V of the form

$$V = \sum_{a \in A} \sum_{b \in B} u_{ab}(r_{ab}), \tag{2}$$

where a and b go over the sets of atoms in monomers A and B, respectively. The atom-atom function u_{ab} is of the form

$$u_{ab}(r_{ab}) = \left[1 + \sum_{i=1}^{2} a_i^{ab} (r_{ab})^i\right] e^{\alpha^{ab} - \beta^{ab} r_{ab}} + \frac{A_{12}^{ab}}{(r_{ab})^{12}} + \frac{q_a q_b}{r_{ab}} - \sum_{n=6,8} f_n(\delta_n^{ab}, r_{ab}) \frac{C_n^{ab}}{(r_{ab})^n}, \tag{3}$$

where f_n are the Tang-Toennies damping functions[72]

$$f_n(\delta_n^{ab}, r_{ab}) = 1 - e^{-\delta r} \sum_{m=0}^{n} \frac{(\delta r)^m}{m!}. \tag{4}$$

We will refer to the sum of the first two terms in Eq. (3) as V_{exp}, the third term as V_{elst}, and the final term as $V_{\text{asym}}^{(2)}$.

The parameters q_a and q_b of the V_{elst} term are fitted to the values of $E_{\text{elst}}^{(1)}$ computed using the distributed asymptotic multipole expansion. In this fit, the charges were constrained based on the atomic charges computed using the CHELPG method[73]. Since no damping is used in this term, V_{elst} is expected to differ from $E_{\text{elst}}^{(1)}$ at short separations. The C_n^{ab} coefficients in $V_{\text{asymp}}^{(2)}$ (constrained to be positive and in a combination rule form, see below) are fit to the sum of the asymptotic multipole expansions of $E_{\text{ind}}^{(2)}$ and $E_{\text{disp}}^{(2)}$. The damping parameters δ_n^{ab} are fit to $E^{(2)} + \delta$ computed at close range. Note that this means that the damping is used to recover not only the charge-overlap effects, but also the second-order exchange effects. Fitting the sum of $E_{\text{ind}}^{(2)}$ and $E_{\text{exch}-\text{ind}}^{(2)}$ is necessary to trim the excessive values of $E_{\text{ind}}^{(2)}$ at short separations resulting from unphysical tunneling of electrons between monomers due to the violation of Pauli's exclusion principle in Rayleigh-Schrödinger perturbation theory. This approach is not needed for the dispersion component, but we use it for consistency. The use of $\delta E_{\text{int,resp}}^{\text{HF}}$ in the fitting of the damping factors is a bit inconsistent since we neglect the asymptotic behavior of this term, but since it decays as $1/R^9$, this neglect is inconsequential. Finally, the free parameters α^{ab}, β^{ab}, a_i^{ab}, and A_{12}^{ab} of the V_{exp} term are fit to $E_{\text{int}} - V_{\text{elst}} - V_{\text{asym}}^{(2)}$. This term mainly reproduces the $E_{\text{exch}}^{(1)}$ component, but also takes care of a part of the short-range electrostatic interactions. For details of these fitting stages, see Ref. 24, Sec. VI.

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

The coefficients A_{12}^{ab} are constrained to be positive, and are included to ensure the correct repulsive behavior of the potential at very close range. We constrain the C_n^{ab} parameters by using the geometric mean combination rule $C_n^{ab} = \sqrt{C_n^a C_n^b}$, where the parameters C_n^a and C_n^b are defined per atom, rather than per atom-pair. Similarly, the parameters α^{ab} and β^{ab} are constrained using the arithmetic sum combination rule. We have found that such constraints have only a small effect on the accuracy of the resulting PES, while substantially reducing the number of free parameters.

For all cases in which some sites of a molecule are symmetrically equivalent, autoPES constrains all parameters associated with those sites to have the same value to preserve symmetry. To reduce fitting complexity, we also applied such constraints to some sites that are not exactly equivalent but are approximately so. In particular, all nitro groups are the same, even if they have stabilizing hydrogen bonds with a neighboring group. This additional constraint sacrifices some of the ability of the functional form to fit the *ab initio* data, but results in fewer free parameters and so fewer grid points are required. All atom equivalences are shown, together with the partial charges q_x, in Fig. 2.

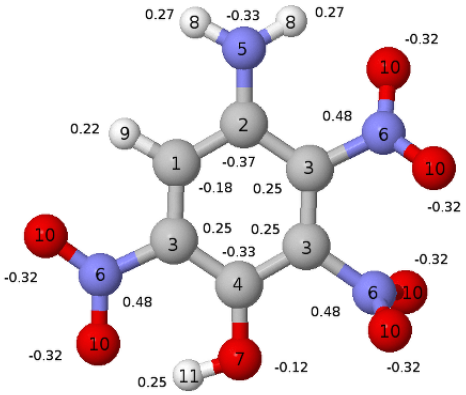

FIG. 2. Equivalent atom types and partial charges of **1**. Atom equivalences are indicated by the atom numbering.

The PES was developed in an iterative way. Following each fitting stage, the quality of the PES is evaluated, and, if convergence criteria are not met, additional iterations are performed. There are two criteria used in this evaluation. The first criterion ensures that the repulsive wall does not contain 'holes', that is, regions where the PES does not have the physically correct repulsive behavior at close range. Due to the cancellations between large positive and negative terms in the functional form used, such 'holes' can occur if even a small region of the repulsive wall of the PES is insufficiently sampled. In such cases, additional grid points are placed at the appropriate configurations, and the potential is refit. The second criterion evaluates the general accuracy of the PES by computing its root mean square error (RMSE) on a set of grid points not used in the fit. Before each fitting iteration, the total

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

set of grid points is split into a fitting set, consisting of 85% of the total number of grid points, and a test set consisting of the remaining 15%. If the RMSE of the PES evaluated on the test set is greater than 1.3 times the RMSE evaluated on the fitting set, then the PES is considered to be unconverged. In such cases, additional grid points are added using the Monte Carlo type grid generation described above. In the final fitting stage, the test set is not used and the PES is fit to all the available data. The subroutines computing the fit as well as a text file with fit parameters and other information on the fit are available in the Supplementary Information (SI).

C. Crystal Structure Predictions

The pairwise PES fit described above is used as the intermolecular potential in CSP for the solid form of **1**. The details of CSP workflow can be found in Ref. 43 and are outlined below. First, the UPACK program suite[16] is employed to generate more than 1000 random Z'=1 structures for each of the most common 13 space group found in organic molecular crystals, including: $C2$, $C2/c$, Cc, $P2_12_12_1$, $Pbca$, $Pbcn$, $P2_1$, $P2_1/c$, $Pca2_1$, Pc, $Pna2_1$, $P1$ and $P\bar{1}$. During the generation step, rigid molecules may be randomly placed with interatomic distances smaller than the repulsive wall of the SAPT(DFT) potential, so the unit cells are first optimized using a combination of the standard OPLS-AA Lennard-Jones parametrization[71] and the SAPT(DFT) fitted charges to avoid any holes in the potential. The final 0 K optimization uses the full SAPT(DFT) intermolecular potential and a modified version of UPACK[74]. These energies are used to rank the possible rigid molecule structures.

Second, the top 30 structures are expanded into a supercell and equilibrated at the target pressure and temperature using MD simulations implemented in PINY_MD package,[75] where all molecules in the simulation box are considered to be fully flexible. Note that PINY_MD cannot be applied to rigid monomers, and with flexible monomers, one needs an intramonomer force field. While the SAPT(DFT) PES described above is capable of describing the intermolecular interaction energy in the case of small internal deformations, it does not include a functional form to evaluate the intramolecular energy component. To allow for molecular flexibility, we used a generic/SAPT(DFT) force field to describe the intramolecular terms of the potential energy. While this type of force field is generally not accurate enough to produce reliable CSPs, we argue that the intramonomer component is adequate. For CSPs of the type performed here, the monomers deform only slightly from their gas-phase geometries and so the intramonomer force field needs only to describe a small region of configuration space near a single reference point. In the case of the intermolecular component of the force field, no such simplification is possible, so the CSP benefits from a tailor-made intermolecular PES.

The generic/SAPT(DFT) force field parameters used for intramolecular terms of the

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

potential energy in the MD simulations are assigned using the ACPYPE[76] wrapper for AnteChamber[77]. The resulting force field uses bonds, bends, and dihedrals from the generalized AMBER force field (GAFF)[77]. For the atom-atom interactions that are separated by three bonds, known as 1-4 interactions, OPLS-AA parameters are assigned[71], with the standard approach of scaling both Coulomb and LJ terms by 0.5. Atomic charges for all atoms, and intramolecular interactions for atoms separated by more than three bonds, use the SAPT(DFT) parameterization (see Table S2 in the SI).

The molecules in each supercell are equilibrated at 100 K using a 10 ps constant volume and temperature (NVT) MD simulation, followed by a 10 ps flexible-cell constant pressure and temperature (NPT) equilibration at 100 K and 1 bar for the full crystalline system. A subsequent 20 ps flexible-cell NPT MD simulation is used to determine the average potential energy to rank the crystalline structures. The flexible-cell NPT averaged supercells are collapsed into unit cells, and the final symmetry and space group assignments are determined by the PLATON package.[78]

For all MD runs, the cut-off distance for the intermolecular interactions is taken to be 10 Å with an integration time step of 0.2 fs. In evaluating the long range Coulomb potential, we use an Ewald screening factor of $\alpha = 0.35$ Å^{-1}, and the smooth particle mesh Ewald (SPME) summation[79] is used with the interpolation order of 10. The massive Nosé-Hoover chain (NHC) thermostat[80–82] is used for all system variables and each NHC has a length of 4 with a characteristic time scale of $\tau = 20$ fs. The NHC integrator uses the Suzuki-Yoshida[83,84] factorization scheme up to the 6-th order, or $n_{\rm sy} = 7$, and a multiple time step factor $n_c = 4$.[85] In integrating the Martyna-Tobias-Klein (MTK) equations in the flexible-cell NPT ensemble[85,86], a characteristic time scale $\tau = 0.5$ ps is used for the barostat and $\tau = 0.1$ ps for its NHC thermostat.

III. RESULTS

A. PES Analysis

The interaction energies from the PES are plotted versus *ab initio* interaction energies for all the fitting grid points in Fig. 3. The large error in the positive energy region visible in these plots is expected due to the low fitting weight given to these grid points. Because these points are located on the steep repulsive wall of the PES, even a large error corresponds only to a very small shift in the position of the wall.

The accuracy of the PES is analyzed in Table 1. The RMSE in the most important region with $E_{\rm int} < 0$, which is weighted more heavily in the PES fitting, is 0.3 kcal/mol. These are expected RMSEs for molecules of this size and fitting functions defined by Eqs. (2)-(4). The error relative to the depths of PESs is 2.4%. On one hand, this should be a

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

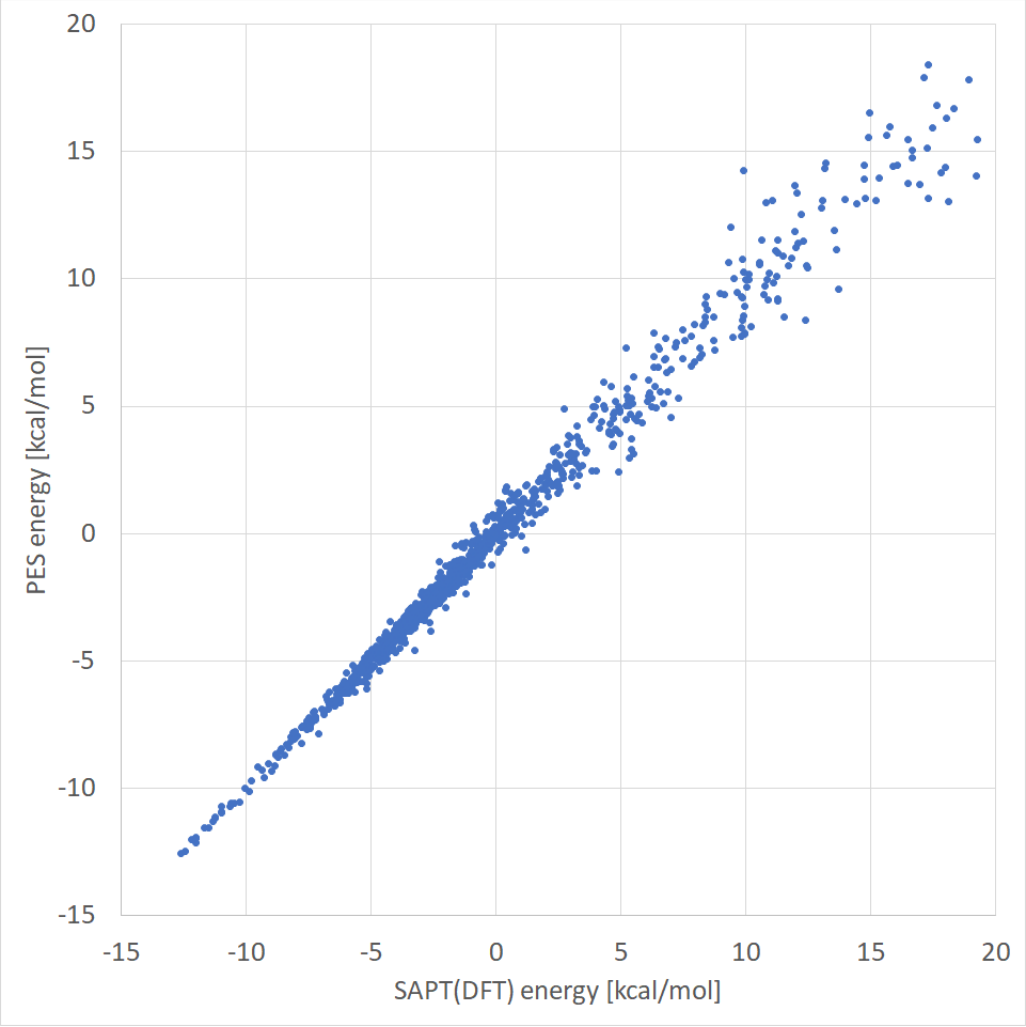

FIG. 3. SAPT(DFT) interaction energy versus PES interaction energy for all fitting points of the **1-1** dimer.

sufficient accuracy since uncertainties of current *ab initio* calculations are of the order of 5%. On the other hand, since differences in lattice energies between some polymorphs are a fraction of kcal/mol, a higher accuracy of the fit would be desirable. The simplest way to improve the accuracy of fit would be to decrease the number of approximate symmetries (the monomer contains 21 atoms, but we have only 11 atoms with distinct parameters), but this would require computing significantly more grid points at considerable computational cost. If computational expense is not a problem, one can in fact reduce the errors of autoPES fits to about 0.01 kcal/mol[87] by also including off-atomic sites. The ratio $N_{\mathrm{grid}}/N_{\mathrm{FP}}$ is 8.7. This ratio is fairly low since autoPES converged in only three iterations (the number of grid points in the first iteration is 6 times the number of free parameters and 20% of this number is added in each subsequent iteration). The total number of minima on the PES listed in Table 1, 66, is very large for this system, more than for others that have been fitted so far

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

TABLE 1. RMSEs of the potential evaluated on subsets of the close-range grid (all energies in kcal/mol). Numbers of grid points in the subsets are given in parentheses. The number of free parameters in the close-range fitting stage is denoted by N_{FP}, the number of detected local minima on the PES by N_{min}, and the total number of grid points by N_{grid}.

N_{grid}	N_{FP}	$N_{\mathrm{grid}}/N_{\mathrm{FP}}$	N_{min}	RMSE $E_{\mathrm{int}} < 0$	RMSE $E_{\mathrm{int}} < 10$
1144	132	8.7	66	0.30 (733)	0.51 (1015)

TABLE 2. Interaction energies of the 10 most negative minima on the PES and of the minima number 26, 32, 33, 54, and 58, whose geometries most closely match the dimers found in the experimental crystal (minimum 2 is also in this category).

Minimum	Energy (kcal/mol)	Minimum	Energy (kcal/mol)
1	-12.57	26	-8.36
2	-12.05	32	-7.74
3	-11.68	33	-7.68
4	-11.66	54	-6.24
5	-11.60	58	-5.38
6	-11.17		
7	-10.92		
8	-10.79		
9	-10.73		
10	-10.67		

with autoPES. The reason is the existence of approximate symmetries. For the benzene molecule, there exists a large number of symmetry-equivalent minima with exactly the same energy. When the benzene symmetry is broken by the addition of the nitro, hydroxyl, and amino groups, all these minima have different energies. The most energetically negative minima on the PES, plus the minima at dimer geometries that closely match those in the experimental crystal (discussed below), are listed in Table 2. All the minima, together with all parameters of the fit, are given in the SI. Table 2 shows that the low-lying minima are very closely spaced and correspond to variations on the slipped-parallel dimer.

The total interaction energy and its components are plotted as functions of the COM-COM distance for the orientation corresponding to the global minimum in Fig. 4. In addition to the fit and its components, we show the corresponding *ab initio* SAPT(DFT) quantities. The figure therefore shows also the quality of our fits in the region of the van der Waals

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

minimum. As one can see, this quality is excellent as the *ab initio* total interaction energies lie almost ideally on the curve from the fit. For the components, the agreement for smaller R is much worse, but this is expected. The best agreement is for the dispersion plus induction components, again as expected since the damping factors in the PES are fitted to these *ab initio* values. The discrepancies are largest for the electrostatic energy which shows the expected failure of the asymptotic expansion at small R. Note that no short-range fitting, and therefore no damping factor, was used for electrostatics. The deviations of the exponential term are a consequence of the deviations for the electrostatics since the exponential term makes up for these discrepancies. In the asymptotic regime, where the exponential term is small, both the electrostatic and induction plus dispersion components are in good agreement with SAPT values. This can be seen in Fig. 5 which shows interaction energies at the same dimer orientation as Fig. 4, but at large separations.

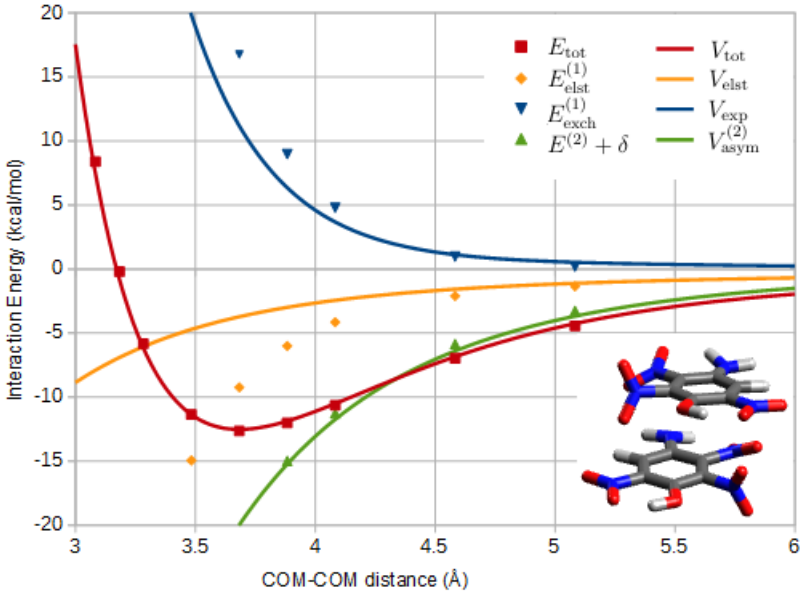

FIG. 4. Radial dependence of the interaction energy and its components for the orientation corresponding to the global minimum of the **1**-**1** dimer. The geometry of the minimum is shown in the insert.

Since molecule **1** has a large polarizability due to π-electrons in the aromatic rings, one expects the dispersion interactions to be large. However, the dipole moment is also large and the quadrupole moments, shown in Table 3, are large as well. Thus, in general one expects substantial electrostatic and induction interactions as well. This results in large interaction energy of -12.5 kcal/mol at the van der Waals minimum, compared to the corresponding benzene dimer interaction energy of only about -2.5 kcal/mol[64]. However, one cannot deduct the ratios of the attractive components from the knowledge of the monomer properties. Figure 4 shows that the dispersion plus induction energy (including

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

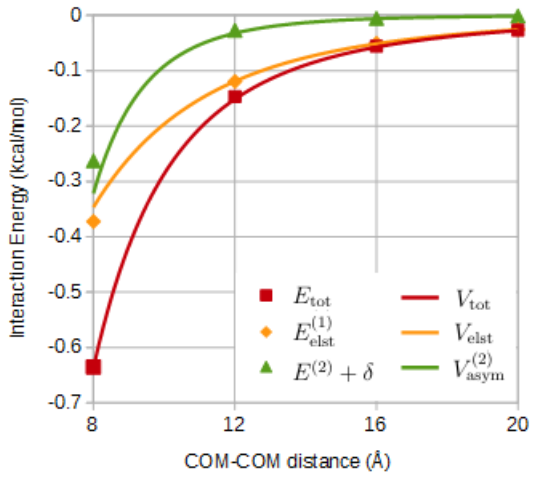

FIG. 5. Radial dependence of the interaction energy and its components in the asymptotic regime for the orientation corresponding to the global minimum of the **1**-**1** dimer.

their exchange counterparts) is about twice larger in magnitude than the electrostatic one. The quantity $E^{(2)}+\delta$ is dominated by the dispersion component (84%), so the induction component, including $\delta E^{\rm HF}_{\rm int,resp}$, is only 16%. Thus, despite the molecule being strongly polar, the dispersion energy dominates. This is a general trend for larger molecules since the electrostatic interactions between different region of monomers are of either sign and therefore have a tendency to cancel[88]. For molecules of the size of **1**, the COM multipole expansion picture does not work well at distances near van der Waals minima and smaller, so only SAPT calculations can determine precise relations between interaction energy components. At large distances shown in Fig. 5, the electrostatic interaction dominates due to its slower $(1/R^3)$ decay than the dispersion and induction components $(1/R^6)$.

TABLE 3. Quadrupole moments of **1** computed at the same level of theory as described in Sec. II A, given in atomic units. The coordinate frame is given in the SI.

xx	yy	zz	xy	xz	yz
-15.8	11.5	4.3	6.0	-4.9	-2.0

Figure 4 shows that the minimum configuration of the **1** dimer is slipped parallel. The relative smallness of the electrostatic energy for such configuration may seem to clash with the well-known fact that for the benzene dimer the slipped parallel configuration is stabilized relative to the sandwich configuration by the electrostatic interactions: negative for the former configuration and positive for the latter (note that these relations cannot be explained using only the leading quadrupole-quadrupole term[64]). However, the benzene-dimer electrostatic interactions are actually very small in magnitude relative to the dispersion

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

interactions[64] and the importance of the electrostatic energy is significantly larger for the dimer of **1**. This is because the quadrupole-quadrupole interactions decaying as $1/R^5$ are the leading-order interactions for the benzene dimer, while for the **1** dimer the dipole-dipole and dipole-quadrupole interactions that decay as $1/R^3$ and $1/R^4$, respectively, dominate. Since the dipole moments are in parallel planes and partly rotated with respect to each other at the minimum geometry, this dipole configuration corresponds to a smaller magnitude of the electrostatic energy than one could have expected based on its polarity.

At the distances of the van der Waals minimum and smaller, the charge-penetration effects strongly increase the magnitude of the electrostatic energy. This leads to an unexpected trend in the ratio of $E^{(1)}_{\text{elst}}/E^{(2)}$ for separations shown in the close-range figure. In the multipole approximation, this ratio should increase as $R^{6-3} = R^3$. Instead, this ratio slightly decreases with increasing R. The reason is, of course, charge-penetration effects. These effects are here unusually large, the fact that to our knowledge has not been observed in literature. This observation appears to be related to the slipped parallel configuration of the dimer. The explanation of this phenomenon is as follows. The electrostatic interactions consist of electron-electron, nuclear-nuclear, and electron-nuclear interactions. When the electrons densities start to overlap significantly, the last interaction remains almost unaffected by this overlap since each nucleus of a monomer is still outside the region of substantial charge density of the interacting partner. Thus, only the positive electron-electron component is affected and it becomes reduced by the overlap compared to what it would have been if the extent of the density were significantly reduced. This makes the electrostatic energy significantly more negative than the values predicted by the asymptotic expansion. Thus, this effect cannot be recovered using the usual multiplicative damping functions which range between 0 and 1, but requires the use of purely exponentially decaying terms resulting from the bipolar expansion of the interaction potential (see Sec. V of Ref. 33). The effect is unusually large for dimers in slipped parallel configuration since charge penetration is obviously exceptionally strong.

B. Comparison of monomer minimum energy geometry with those in the experimental crystal structure

The experimental crystal structure has one molecule in the asymmetric unit cell[51], but due to the symmetry operations in the *Pbca* space group that invert the molecule, there are two monomer geometries present. These molecules differ primarily in the orientation of the nitro groups relative to the aromatic ring. A comparison of the r_e geometry with the two experimental geometries shows that the conformer used in fitting the PES is suitable to describe both forms (see Fig. 6). To quantify the differences, the experimental coordinates are aligned to the optimized geometry by minimizing the root mean square deviation

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

(RMSD) of carbon atoms. After alignment, the $RMSD_C$ is 0.019 Å and 0.014 Å, while $RMSD_{O,N}$ is 0.151 Å and 0.427 Å for the two experimental geometries, respectively. The deviations are primarily due to slight rotations of two of the nitro groups indicated in Fig. 1 (the third nitro group is in the plane of the ring due to the intramolecular hydrogen bond with the neighboring hydroxyl group). The discrepancies in the rotation angles of these two nitro groups between theory and experiment are to be expected, as the rotational potential is quite flat and therefore interactions in the crystal can easily lead to rotations relative to atomic positions in the isolated molecule.

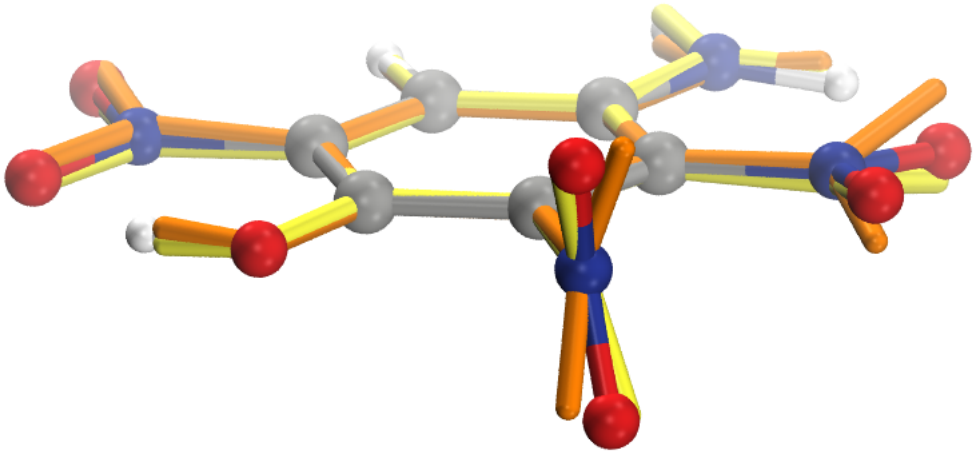

FIG. 6. Geometries of **1** from DFT optimization (spheres), compared to the geometries of the two conformers from the experimental crystal structure[51] (orange and yellow). Two of the nitro groups in the experimental geometries deviate somewhat from the optimized gas-phase equilibrium positions.

C. Comparison of dimer minimum geometries with those of dimers in the experimental crystal structure

We compare the local minima of the PES to the dimer configurations found in the experimental crystal structure. An interesting question to answer is whether close-neighbor dimers in the crystal structures are in any way related to the minima on PESs. It is generally known that in many cases the crystal dimer configurations do not correspond to global minimum geometries of the isolated dimer, although such correspondence does exist for crystals with well-defined synthons such as hydrogen bonds. The question is, however, if a crystal's dimers are matched by isolated dimers corresponding to local minima. The other

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

option is that dimer configurations in crystals are intermediate between the local minima of the isolated dimers.

We have identified the nearest-neighbor dimers from the *Pbca* experimental crystal[51] and overlapped them with all 66 dimers at the minima geometries on our PES. The matches are determined by aligning the dimers based on minimizing the RMSD for the carbon and nitrogen atoms. One of the local minimum geometries, minimum 2, is the same as a closely spaced dimer from the crystal. The five other observed dimers from the crystal are related to local minima that are higher energy, with significantly less contact between the molecular pair. We note that the first 22 minima are dimer configurations that belong to the group of slipped-parallel dimers similar to the global minimum. Therefore, it is reasonable that the next lowest energy match (minimum 26) to the experimental dimers would be so high in the energy ranking.

A section of the crystal cell highlighting the two most closely matched dimers is shown in Fig. 7. The overlaps between these dimers and the corresponding minima on the PES (minimum 2 and 54) are shown in Fig. 8 (the remaining matches are shown in the SI). The first closely matching minimum is number 2 (interaction energy of -12.05 kcal/mol, 0.52 kcal/mol above the global minimum) with an $\mathrm{RMSD}_{C,N}$ of only 0.17 Å. The other five minima, also reported in Table 2, are listed in energetic order (with $\mathrm{RMSD}_{C,N}$ to their respective dimer in parenthesis): number 26 (1.32 Å), number 32 (1.44 Å), number 33 (1.59 Å), number 54 (0.78 Å), and number 58 (1.18 Å).

We can now answer the question asked in the beginning of this section: yes, dimers from crystal structures are similar to isolated dimers representing local minima. In particular, the experimental nearest neighbour configuration is very close to the lowest-energy local minimum on the PES. Taking into account that 22 lowest-energy configurations are of slipped parallel type, the first non-parallel dimer matching experiment is also one of the lowest-energy such dimers. Thus, the knowledge of the minima on the dimer PES of a notional compound can be helpful in crystal design.

D. CSPs

Our CSP workflow ultimately generates structures that are equilibrated at experimentally-relevant temperatures and pressures. However, we can learn more about predictions by examining the structures that match the reported form at each intermediate stage of the process. Such results are presented in Table 4. The rigid molecule packing step using the OPLS force field parameters yields a match to the experimental structure with a root mean squared deviation of 20 neighboring molecules (RMSD_{20}) of 0.301 Å, but the energy is of rank 53 and it is ∼12 kJ/mol above the lowest energy. Optimizing the polymorphs with the SAPT(DFT) intermolecular potential, also with rigid molecules, significantly improves

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

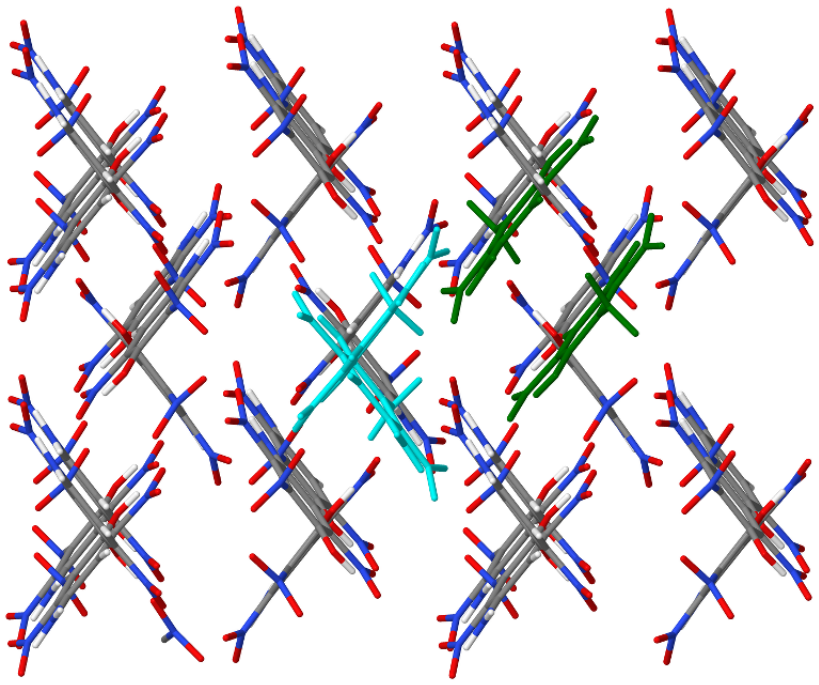

FIG. 7. Experimental crystal structure of **1**. The dimer which closely matches minima 2 is shown in green, and the dimer which closely matches minimum number 54 is shown in cyan.

the energy ranking of the experimental form. The matching structure in this case is the 4th lowest energy (see Fig. 9) and has an $RMSD_{20}$ which is slightly improved over the OPLS version, despite having an even more elongated a-axis compared to experiment.

TABLE 4. Cell parameters, $RMSD_{20}$, and ranks for structures matching the observed form. Experimental values are from Ref. 51.

Structure	Inter potential	Rigid?	Rank	$RMSD_{20}$ (Å)	density (g/cm^3)	a (Å)	b (Å)	c (Å)
experimental[51]					1.839	12.5540	10.3306	13.5974
0 K	OPLS	rigid	53	0.301	1.796	12.9546	10.2355	13.6158
0 K	SAPT(DFT)	rigid	4	0.272	1.754	13.2466	10.2491	13.6153
100 K	SAPT(DFT)	flexible	5	0.350	1.684	13.4277	10.4279	13.7765

To account for thermal expansion of the crystal unit cells, the predicted structures are equilibrated at 100 K and 1 bar through MD simulations with flexible molecules. After this step, the matching structure has an $RMSD_{20}$ of 0.350 Å when using an angle tolerance of 25° and is the 5th lowest energy form, as shown in Fig. 10. The overlap of the predicted and experimental crystal structures is shown in Fig. 11. The deviations from the reported structure arise from the slightly expanded a-axis for the unit cell and a slight mismatch in the

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

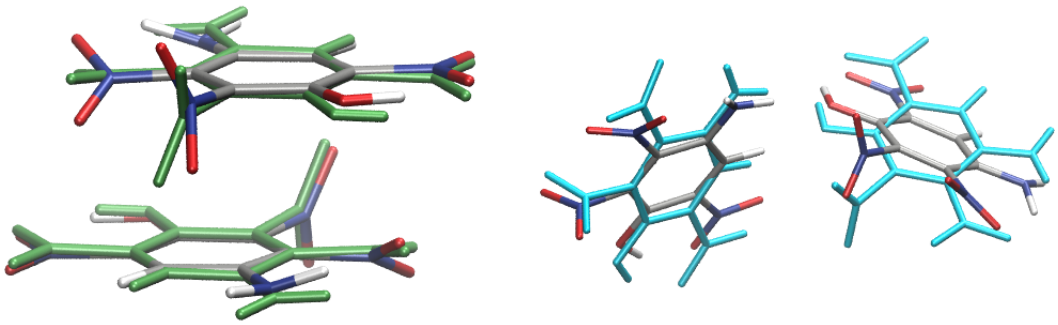

FIG. 8. Overlap of the dimers at geometries corresponding to minima number 2 (left) and 54 (right) with dimers from the experimental crystal structure. The experimental dimers are shown in green and cyan (as in Fig. 7) and the PES minima are shown colored by element. The dimers were aligned by minimizing the RMSD for all C and N atoms.

orientation of the easily rotatable nitro groups. However, the structures are clearly a good match and the hybrid force field approach using the tailor-made SAPT(DFT) intermolecular terms with a generic/SAPT(DFT) intramolecular force field is adequate for **1**. Overall, we are encouraged that the computational time required to generate a molecule-specific intermolecular force field, as compared to the use of a generic intermolecular force field, is worth the effort to refine both the structural match and energy ranking of the observed form.

IV. CONCLUSIONS

In this paper, we demonstrated the application of a bottom-up protocol for crystal structure prediction that includes zero-temperature CSP and finite-temperature MD with a tailor-made molecule-specific force field, in the prediction of the landscape of an energetic compound, 4-amino-2,3,6-trinitrophenol. The custom force field includes an accurate intermolecular component built from dimers treated at the SAPT(DFT) level of electronic structure theory with a generic/SAPT(DFT) representation of the intramolecular potential. The landscape reveals that the experimentally-observed form matches the predicted structure with an $RMSD_{20}$ of 0.35 Å, which lies well within the acceptable bounds prescribed by the CCDC blind structure prediction competition[15].

A critical finding of this work is that the quality of the molecule-specific *intermolecular*

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

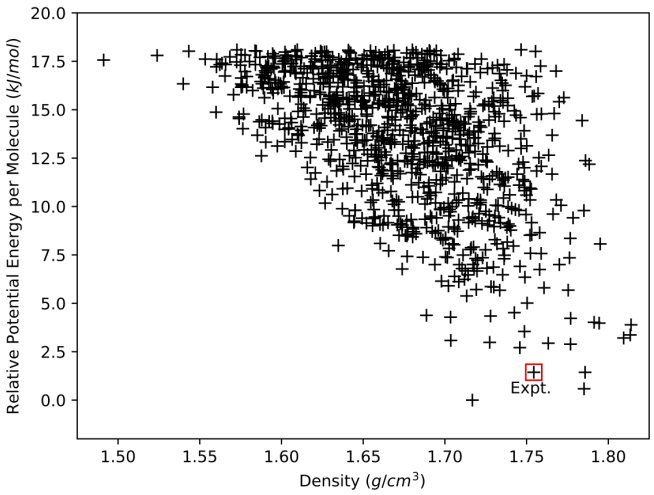

FIG. 9. CSP results for **1** evaluated for predicted crystal structures optimized using rigid molecules and the tailor-made SAPT(DFT) intermolecular potential. The observed experimental structure is indicated with a red box.

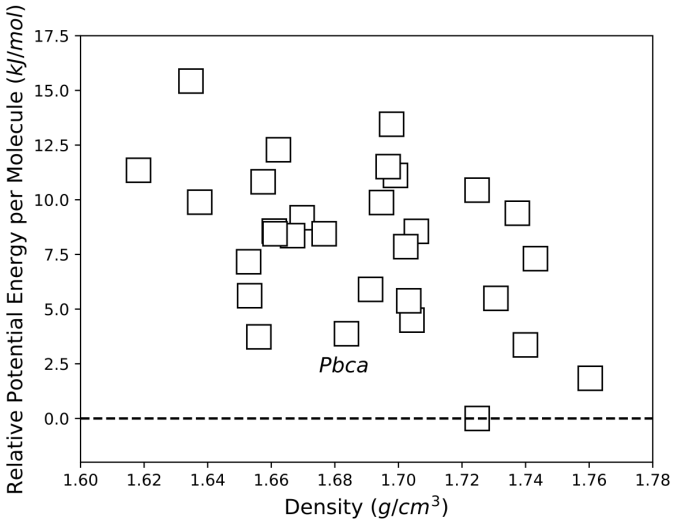

FIG. 10. CSP results for **1** when equilibrated at 100 K and 1 bar. Average energy per molecule is shown relative to the energy of the most stable predicted polymorph. The experimentally-observed *Pbca* structure has the fifth lowest average potential energy.

potential appears to be determinitive of the accuracy of predicted structures. Using a generic/SAPT(DFT) representation of the *intramolecular* potential can be employed without significantly compromising that accuracy, even when some functional groups rotate easily within the crystalline environment. If greater accuracy of the final structure is required, several improvements of the present approach are possible depending on the computational

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

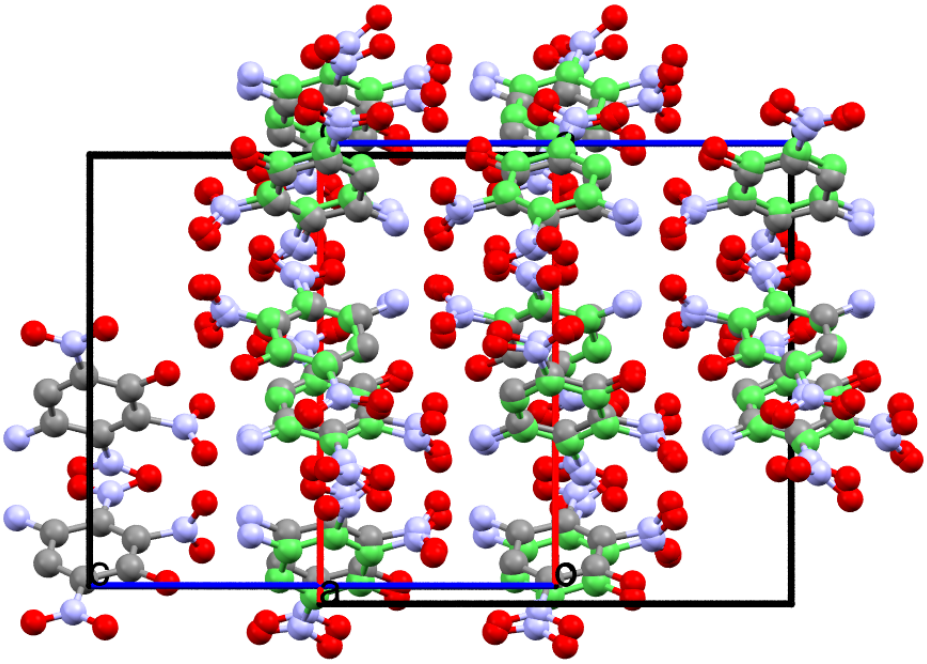

FIG. 11. Overlap of the experimentally-observed structure[51] with the predicted form equilibrated at 100 K and 1 bar using the tailor-made SAPT(DFT) force field for MD simulations. The experimental structure is shown colored by element, while the predicted structure is shown with green carbon atoms.

effort permitted: increased size of the basis set for *ab intio* calculations, increased number of parameters in the SAPT fit, a more elaborate form of the fitting function, explicit inclusion of intramonomer degrees of freedom in the SAPT PESs, an iterative refinement of SAPT PESs by including specific dimers extracted from highest-rank polymorphs within the fitting procedure, use of an intramonomer force field fitted to *ab intio* data, and by performing periodic boundary conditions dispersion-corrected DFT calculations for selected structures. However, reasonable CSP results have been obtained here for this challenging molecule without resorting to these more expensive approaches. The relative insensitivity of the landscape on the intramolecular potential, i.e., small differences between the rigid- and flexible-monomer simulations, simplifies the CSP procedure, allowing for easy application of tools such as autoPES[50] for the creation of molecule-specific force fields. Whether the idea of using approximate intramolecular potentials in combination with tailor-made intermolecular potentials applies to crystals of larger, more flexible organic molecules merits further exploration and will be the subject of future work.

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

ACKNOWLEDGMENTS

The authors thank Rahul Nikhar for providing us with a modified UPACK program capable of using the SAPT(DFT) potential. This work was supported by the U.S. Army Research Laboratory and Army Research Office (Grant No. W911NF-19-1-0117) and by the NSF (Grant No. CHE-1900551).

---

* marktuckerman@nyu.edu

† szalewic@udel.edu

[1] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris, "Ritonavir: An extraordinary example of conformational polymorphism," Pharm. Res. **18**, 859 (2001).

[2] S. 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, "Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development," Org. Proc. Res. Devel. **4**, 413–417 (2000).

[3] I. B. Rietveld and R. Ceolin, "Rotigotine: Unexpected polymorphism with predictable overall monotropic behavior," J. Pharm. Sci. **104**, 4117–4122 (2015).

[4] M. Mortazavi, J. Hoja, L. Aerts, L. Quere, J. van de Streek, M. A. Neumann, and A. Tkatchenko, "Computational polymorph screening reveals late-appearing and poorly-soluble form of rotigotine," Comm. Chem. **2**, 70 (2019).

[5] O. D. Jurchescu, D. A. Mourey, S. Subramanian, S. R. Parkin, B. M. Vogel, J. E. Anthony, T. N. Jackson, and D. J. Gundlach, "Effects of polymorphism on charge transport in organic semiconductors," Phys. Rev. B **80**, 085201 (2009).

[6] S. M. Walley, J. E. Field, and M. W. Greenaway, "Crystal sensitivities of energetic materials," Mat. Sci. Tech. **22**, 402–413 (2006).

[7] S. L. Price, "Is zeroth order crystal structure prediction (CSP_0) coming to maturity? What should we aim for in an ideal crystal structure prediction code?" Faraday Discuss. **211**, 9–30 (2018).

[8] S. L. Price, "Control and prediction of the organic solid state: A challenge to theory and experiment," Proc. R. Soc. A **474**, 20180351 (2018).

[9] J. Nyman and S. M. Reutzel-Edens, "Crystal structure prediction is changing from basic science to applied technology," Faraday Discuss. **211**, 459–476 (2018).

[10] J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, W. T. M. Mooij, S. L. Price, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, and D. E. Williams, "A test of crystal structure prediction of small organic molecules," Acta Cryst. B **56**, 697–714 (2000).

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

[11] W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, J. P. M. Lommerse, W. T. M. Mooij, S. L. Price, H. Scheraga, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, and D. E. Williams, "Crystal structure prediction of small organic molecules: A second blind test," Acta Cryst. B **58**, 647–661 (2002).

[12] G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, T. Torrisi, H. A. Scheraga, Y. A. Amautova, M. U. Schmidt, and P. Verwer, "A third blind test of crystal structure prediction," Acta Cryst. B **61**, 511–527 (2005).

[13] G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Amautova, M. U. Schmidt, J. van de Streek, A. K. Wolf, and B. Schweizer, "Significant progress in predicting the crystal structures of small organic molecules – A report on the fourth blind test," Acta Cryst. B **65**, 107–125 (2009).

[14] D. A. Bardwell, C. S. Adjiman, Y. A. Amautova, E. Vartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Deskraju, B. P. Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti, and I. K. Zhitkov, "Towards crystal structure prediction of complex organic compounds – A report on the fifth blind test," Acta Cryst. B **67**, 535–551 (2011).

[15] A. 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. Cosburg, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio, 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. Hogmann, 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. Y. Li, S. Lohano, F. J. J. Leusen, A. M. Lund, J. Lv, Y. M. 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. Nees, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J.

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

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. Vazques-Mayagoitia, L. Vogt, Y. C. Want, R. E. Watson, G. A. de Wijs, J. Yang, Z. Ahu, and C. R. Groom, "Report on the sixth blind test of organic crystal structure prediction methods," Acta Cryst. **B72**, 439–459 (2016).

[16] B. P. van Eijck and J. Kroon, "Upack program package for crystal structure prediction: Force fields and crystal structure generation for small carbohydrate molecules," J. Comput. Chem. **20**, 799–812 (1999).

[17] Bouke P. van Eijck and Jan Kroon, "Structure predictions allowing more than one molecule in the asymmetric unit," Acta Crystallogr. B Struct. Sci. **56**, 535–542 (2000).

[18] Seonah Kim, Anita M. Orendt, Marta B. Ferraro, and Julio C. Facelli, "Crystal structure prediction of flexible molecules using parallel genetic algorithms with a standard force field," J. Comput. Chem. **30**, 1973–1985 (2009).

[19] Qiang Zhu, Artem R. Oganov, Colin W. Glass, and Harold T. Stokes, "Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications," Acta Crystallogr. B Struct. Sci. **68**, 215–226 (2012).

[20] Farren Curtis, Xiayue Li, Timothy Rose, Álvaro Vázquez-Mayagoitia, Saswata Bhattacharya, Luca M. Ghiringhelli, and Noa Marom, "GAtor: A First-Principles Genetic Algorithm for Molecular Crystal Structure Prediction," J. Chem. Theory Comput. **14**, 2246–2264 (2018).

[21] H.R. Karfunkel and R.J. Gdanitz, "Ab Initio prediction of possible crystal structures for general organic molecules," J. Comput. Chem. **13**, 1171–1183 (1992).

[22] J. Pillardy, Y. A. Arnautova, C. Czaplewski, K. D. Gibson, and H. A. Scheraga, "Conformation-family Monte Carlo: A new method for crystal structure prediction," Proc. Natl. Acad. Sci. **98**, 12351–12356 (2001).

[23] Jonas Nyman, Orla Sheehan Pundyke, and Graeme M. Day, "Accurate force fields and methods for modelling organic molecular crystals at finite temperatures," Phys. Chem. Chem. Phys. **18**, 15828–15837 (2016).

[24] Michael P Metz, Konrad Piszczatowski, and Krzysztof Szalewicz, "Automatic generation of intermolecular potential energy surfaces," J. Chem. Theory Comput. **12**, 5895–5919 (2016).

[25] Gregory J. O. Beran, "Modeling Polymorphic Molecular Crystals with Electronic Structure Theory," Chem. Rev. **116**, 5567–5613 (2016).

[26] J. G. Brandenburg and S. Grimme, "Organic crystal polymorphism: a benchmark for dispersion-corrected mean-field electronic structure methods," Cryst. Acta B **52**, 502–513 (2016).

[27] Johannes Hoja, Anthony M. Reilly, and Alexandre Tkatchenko, "First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure: First-principles modeling of molecular crystals," WIREs Comput. Mol. Sci. **7**, e1294 (2017).

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

[28] J. Hoja, H. Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio, and A. Tkatchenko, "Reliable and practical computational description of molecular crystal polymorphs," Sci. Adv. **5**, eaau3338 (2019).

[29] Rafał Podeszwa, Robert Bukowski, Betsy M Rice, and Krzysztof Szalewicz, "Potential energy surface for cyclotrimethylene trinitramine dimer from symmetry-adapted perturbation theory," Phys. Chem. Chem. Phys. **9**, 5561–5569 (2007).

[30] R. Podeszwa, B. M. Rice, and K. Szalewicz, "On predicting structure of molecular crystals from first principles," Phys. Rev. Lett. **101**, 115503 (2008).

[31] R. Podeszwa, B. M. Rice, and K. Szalewicz, "Crystal structure prediction for cyclotrimethylene trinitramine (RDX) from first principles," Phys. Chem. Chem. Phys. **11**, 5512–5518 (2009).

[32] A. J. Misquitta, G. W. A. Welch, A. J. Stone, and S. L. Price, "A first principles prediction of the crystal structure of $C_6Br_2ClFH_2$," Chem. Phys. Lett. **456**, 105–109 (2008).

[33] B. Jeziorski, R. Moszyński, and K. Szalewicz, "Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes," Chem. Rev. **94**, 1887–1930 (1994).

[34] H. L. Williams and C. F. Chabalowski, "Using Kohn-Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions," J. Phys. Chem. A **105**, 646–659 (2001).

[35] Alston J Misquitta and Krzysztof Szalewicz, "Intermolecular forces from asymptotically corrected density functional description of monomers," Chem. Phys. Lett. **357**, 301–306 (2002).

[36] Andreas Heßelmann and Georg Jansen, "First-order intermolecular interaction energies from kohn–sham orbitals," Chem. Phys. Lett. **357**, 464–470 (2002).

[37] Andreas Heßelmann and Georg Jansen, "Intermolecular induction and exchange-induction energies from coupled-perturbed Kohn-Sham density functional theory," Chem. Phys. Lett. **362**, 319–325 (2002).

[38] Alston J Misquitta, Bogumil Jeziorski, and Krzysztof Szalewicz, "Dispersion energy from density-functional theory description of monomers," Phys. Rev. Lett. **91**, 033201 (2003).

[39] Andreas Heßelmann and Georg Jansen, "Intermolecular dispersion energies from time-dependent density functional theory," Chem. Phys. Lett. **367**, 778–784 (2003).

[40] A. J. Misquitta and K. Szalewicz, "Symmetry-adapted perturbation theory calculations of intermolecular forces employing density functional description of monomers," J. Chem. Phys. **122**, 214109 (2005).

[41] Alston J Misquitta, Rafał Podeszwa, Bogumił Jeziorski, and Krzysztof Szalewicz, "Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations," J. Chem. Phys. **123**, 214103 (2005).

[42] A Hesselmann, G Jansen, and M Schütz, "Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

intermolecular interaction energies," J. Chem. Phys. **122**, 014103 (2005).

[43] Elia Schneider, Leslie Vogt, and Mark E. Tuckerman, "Exploring polymorphism of benzene and naphthalene with free energy based enhanced molecular dynamics," Acta Cryst. B **B72**, 542–550 (2016).

[44] Tang-Qing Yu and Mark E. Tuckerman, "Temperature-Accelerated Method for Exploring Polymorphism in Molecular Crystals Based on Free Energy," Phys. Rev. Lett. **107**, 015701 (2011).

[45] T. Q. Yu, P. Y. Chen, M. Chen, A. Samanta, E. Vanden-Eijnden, and M. E. Tuckerman, "Order-parameter-aided temperature-accelerated sampling for the exploration of crystal polymorphism and solid-liquid phase transitions," J. Chem. Phys. **140**, 214109 (2014).

[46] A.G. Shtukenberg, Q. Zhu, D. J. Carter, L. Vogt, J. Hoja, E. Schneider, H. Song, B. Pokroy, I. Polishchuk, A. Tkatchenko, A. R. Oganov, A. L. Rohl, M. E. Tuckerman, and B. Kahr, "Powder diffraction and crystal structure prediction identify four new coumarin polymorphs," Chem. Sci. **8**, 4926–4940 (2017).

[47] H. Song, L. Vogt-Maranto, R. Wiscons, A. J. Matzger, and M. E. Tuckerman, "Generating cocrystal polymorphs with information entropy driven by molecular dynamics-based enhanced sampling," J. Phys. Chem. Lett. **11**, 9751 (2020).

[48] N. Fellah, A. G. Shtukenberg, E. J. Chan, L. Vogt-Maranto, W. Xu, C. Li, M. E. Tuckerman, B. Kahr, and M. D. Ward, "Disorderly conduct of benzamide IV: Crystallogrphic and computational analysis of high entropy polymorphs of small molecules," Cryst. Growth & Design **20**, 2670 (2020).

[49] N. F. Francia, L. S. Price, J. Nyman, S. L. Price, and M. Salvalaglio, "Systematic finite-temperature reduction of crystal energy landscapes," Cryst. Growth & Design **20**, 6847 (2020).

[50] M. P. Metz, K. Piszczatowski, and K. Szalewicz, "*autoPES: Automatic Intermolecular Potential Energy Surface Generation Software*," `http://www.physics.udel.edu/~szalewic/SAPT/index.html` (2016).

[51] Thomas M Klapötke, Andreas Preimesser, and Jörg Stierstorfer, "Synthesis and energetic properties of 4-diazo-2, 6-dinitrophenol and 6-diazo-3-hydroxy-2, 4-dinitrophenol," Europ. J. Org. Chem. **2015**, 4311–4315 (2015).

[52] Frank Neese, "The orca program system," Wiley Interdis. Rev.: Comp. Mol. Sci. **2**, 73–78 (2012).

[53] Frank Neese, "An improvement of the resolution of the identity approximation for the formation of the coulomb matrix," J. Comp. Chem. **24**, 1740–1747 (2003).

[54] John P Perdew, Kieron Burke, and Matthias Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865 (1996).

[55] Carlo Adamo and Vincenzo Barone, "Toward reliable density functional methods without adjustable parameters: The pbe0 model," J. Chem. Phys. **110**, 6158–6170 (1999).

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

[56] Rick A Kendall, Thom H Dunning Jr, and Robert J Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions," J. Chem. Phys. **96**, 6796–6806 (1992).

[57] Robert Bukowski, Rafał Podeszwa, and Krzysztof Szalewicz, "Efficient calculation of coupled Kohn–Sham dynamic susceptibility functions and dispersion energies with density fitting," Chem. Phys. Lett. **414**, 111–116 (2005).

[58] Rafal Podeszwa, Robert Bukowski, and Krzysztof Szalewicz, "Density-fitting method in symmetry-adapted perturbation theory based on kohn-sham description of monomers," J. Chem. Theo. Comp. **2**, 400–412 (2006).

[59] R. Bukowski, W. Cencek, P. Jankowski, M. Jeziorska, B. Jeziorski, M. P. Metz, R. Moszynski, K. Patkowski, R. Podeszwa, F. Rob, S. Rybak, K. Szalewicz, H. L. Williams, R. J. Wheatley, P. E. S. Wormer, and P. Zuchowski, "SAPT2016: An *ab initio* program for symmetry-adapted perturbation theory," `http://www.physics.udel.edu/~szalewic/SAPT/index.html` (2016).

[60] M Grüning, OV Gritsenko, SJA Van Gisbergen, and EJ Baerends, "Shape corrections to exchange-correlation potentials by gradient-regulated seamless connection of model potentials for inner and outer region," J. Chem. Phys. **114**, 652–660 (2001).

[61] W. Cencek and K. Szalewicz, "On asymptotic behavior of density functional theory," J. Chem. Phys. **139**, 024104–(1:27) (2013), Erratum: **140**, 149902 (2014).

[62] Florian Weigend, Andreas Köhn, and Christof Hättig, "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations," J. Chem. Phys. **116**, 3175–3183 (2002).

[63] Hayes L Williams, Eric M Mas, Krzysztof Szalewicz, and Bogumił Jeziorski, "On the effectiveness of monomer-, dimer-, and bond-centered basis functions in calculations of intermolecular interaction energies," J. Chem. Phys. **103**, 7374–7391 (1995).

[64] Rafal Podeszwa, Robert Bukowski, and Krzysztof Szalewicz, "Potential energy surface for the benzene dimer and perturbational analysis of π-π interactions," J. Phys. Chem. A **110**, 10345–10354 (2006).

[65] Rafał Podeszwa, Wojciech Cencek, and Krzysztof Szalewicz, "Efficient calculations of dispersion energies for nanoscale systems from coupled density response functions," J. Chem. Theo. Comp. **8**, 1963–1969 (2012).

[66] M. Jeziorska, B. Jeziorski, and J. Cizek, "Direct calculation of the Hartree-Fock interaction energy via exchange perturbation expansion - the He–He interaction," Int. J. Quantum Chem. **32**, 149–164 (1987).

[67] Konrad Patkowski, Krzysztof Szalewicz, and Bogumił Jeziorski, "Third-order interactions in symmetry-adapted perturbation theory," J. Chem. Phys. **125**, 154107 (2006).

[68] A. J. Stone, *The Theory of Intermolecular Forces*, 2nd ed. (Clarendon Press, Oxford, 2013).

[69] Fazle Rob and Krzysztof Szalewicz, "Distributed molecular polarisabilities and asymptotic intermolecular interaction energies," Mol. Phys. **111**, 1430–1455 (2013).

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

[70] M. P. Metz and K. Szalewicz, "Automatic generation of flexible-monomer intermolecular potential energy surfaces," J. Chem. Theory Comput. **16**, 2317–2339 (2020).

[71] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," J. Am. Chem. Soc. **118**, 11225–11236 (1996).

[72] K. T. Tang and J Peter Toennies, "An improved simple model for the van der waals potential based on universal damping functions for the dispersion coefficients," J. Chem. Phys. **80**, 3726–3741 (1984).

[73] Curt M Breneman and Kenneth B Wiberg, "Determining atom-centered monopoles from molecular electrostatic potentials. the need for high sampling density in formamide conformational analysis," J. Comp. Chem. **11**, 361–373 (1990).

[74] R. Nikhar and K. Szalewicz, (2021), manuscript in preparation.

[75] Mark E. Tuckerman, D.A. Yarne, Shane O. Samuelson, Adam L. Hughes, and Glenn J. Martyna, "Exploiting multiple levels of parallelism in molecular dynamics based calculations via modern techniques and software paradigms on distributed memory computers," Comput. Phys. Commun. **128**, 333–376 (2000).

[76] Alan W. Sousa da Silva and Wim F. Vranken, "Acpype–antechamber python parser interface," BMC Research Notes **5**, 367 (2012).

[77] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case, "Development and testing of a general amber force field," J. Comput. Chem. **25**, 1157–1174 (2004).

[78] A. L. Spek, "Single-crystal structure validation with the program platon," J. Appl. Cryst. **36**, 7–13 (2003).

[79] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen, "A smooth particle mesh ewald method," J. Chem. Phys. **103**, 8577–8593 (1995).

[80] Shuichi Nosé, "A unified formulation of the constant temperature molecular dynamics methods," J. Chem. Phys. **81**, 511–519 (1984).

[81] William G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," Phys. Rev. A **31**, 1695–1697 (1985).

[82] Glenn J. Martyna, Michael L. Klein, and Mark E. Tuckerman, "Nosé–hoover chains: The canonical ensemble via continuous dynamics," J. Chem. Phys. **97**, 2635–2643 (1992).

[83] Masuo Suzuki, "General theory of fractal path integrals with applications to many-body theories and statistical physics," J. Math. Phys. **32**, 400–407 (1991).

[84] Haruo Yoshida, "Construction of higher order symplectic integrators," Phys. Lett. A **150**, 262–268 (1990).

[85] Mark Tuckerman, José Alejandre, Roberto López-Rendón, Andrea L. Jochim, and Glenn J. Martyna, "A liouville-operator derived measure-preserving integrator for molecular dynamics

Accepted Manuscript
Version of record at: https://doi.org/10.1021/acs.cgd.1c01117

simulations in the isothermal-isobaric ensemble," J. Phys. A: Math. Gen. **39**, 5629–5651 (2006).

[86] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein, "Constant pressure molecular dynamics algorithms," J. Chem. Phys. **101**, 4177–4189 (1994).

[87] M. P. Metz, K. Szalewicz, J. Sarka, R. Tóbiás, A. G. Császár, and E. Mátyus, "Molecular dimers of methane clathrates: *ab initio* potential energy surfaces and variational (ro)vibrational states," Phys. Chem. Chem. Phys. **21**, 13504–13525 (2019).

[88] Robert Bukowski, Krzysztof Szalewicz, and Cary F Chabalowski, "Ab initio interaction potentials for simulations of dimethylnitramine solutions in supercritical carbon dioxide with cosolvents," J. Phys. Chem. A **103**, 7322–7340 (1999).