A non-empirical intermolecular force-field for trinitrobenzene and its application in crystal structure prediction

ⓘD **Alex A. Aina,** ⓘD **Alston J. Misquitta and** ⓘD **Sarah L. Price**

COLLECTIONS

Paper published as part of the special topic on Special Collection in Honor of Women in Chemical Physics and Physical Chemistry

View Online   Export Citation   CrossMark

ARTICLES YOU MAY BE INTERESTED IN

A non-empirical intermolecular force-field for trinitrobenzene and its application in crystal structure prediction

View Online   Export Citation   CrossMark

Alex A. Aina,[1] (ID) **Alston J. Misquitta,**[2] (ID) **and Sarah L. Price**[1,a] (ID)

AFFILIATIONS

[1] Department of Chemistry, University College London, 20 Gordon St., London WC1H 0AJ, United Kingdom
[2] School of Physics and Astronomy and The Thomas Young Centre for Theory and Simulation of Materials at Queen Mary, University of London, London E1 4NS, United Kingdom

Note: This paper is part of the JCP Special Collection in Honor of Women in Chemical Physics and Physical Chemistry.
[a] Author to whom correspondence should be addressed: s.l.price@ucl.ac.uk

ABSTRACT

An anisotropic atom–atom distributed intermolecular force-field (DIFF) for rigid trinitrobenzene (TNB) is developed using distributed multipole moments, dipolar polarizabilities, and dispersion coefficients derived from the charge density of the isolated molecule. The short-range parameters of the force-field are fitted to first- and second-order symmetry-adapted perturbation theory dimer interaction energy calculations using the distributed density-overlap model to guide the parameterization of the short-range anisotropy. The second-order calculations are used for fitting the damping coefficients of the long-range dispersion and polarization and also for relaxing the isotropic short-range coefficients in the final model, DIFF-srL2(rel). We assess the accuracy of the unrelaxed model, DIFF-srL2(norel), and its equivalent without short-range anisotropy, DIFF-srL0(norel), as these models are easier to derive. The model potentials are contrasted with empirical models for the repulsion–dispersion fitted to organic crystal structures with multipoles of iterated stockholder atoms (ISAs), FIT(ISA,L4), and with Gaussian Distributed Analysis (GDMA) multipoles, FIT(GDMA,L4), commonly used in modeling organic crystals. The potentials are tested for their ability to model the solid state of TNB. The non-empirical models provide more reasonable relative lattice energies of the three polymorphs of TNB and propose more sensible hypothetical structures than the empirical force-field (FIT). The DIFF-srL2(rel) model successfully has the most stable structure as one of the many structures that match the coordination sphere of form III. The neglect of the conformational flexibility of the nitro-groups is a significant approximation. This methodology provides a step toward force-fields capable of representing all phases of a molecule in molecular dynamics simulations.

I. INTRODUCTION

An accurate potential energy surface (PES) for the intermolecular interaction between organic molecules is a key requirement for simulating the properties of the gas phase and the liquid, the solid, and other condensed phases including crystalline polymorphs and the transitions between them.[1] The methods of realistically evaluating many condensed phase properties require an analytical model so that huge assemblies of molecules can be simulated over long time periods. Such analytical models are generally pairwise intermolecular potentials; however, non-additive many-body contributions also need to be considered when studying condensed

phases. Furthermore, potentials for flexible molecules may depend on molecular conformation. Many molecules change conformation when packed with other molecules, which can result in changes in the molecular charge distribution that affect the intermolecular interactions. Hence, while model intermolecular potentials that are able to describe the properties in all phases of the inert gases, such as argon, have been available for decades[2] and considerable progress has been made on a highly accurate pair potential for water (see Refs. 3 and 4 and references therein), most atomistic modeling of condensed phases of organic molecules has used empirical force-fields. Force-fields have been heavily developed for biomolecular simulations, but the fitting to experimental data inevitably absorbs

system-specific errors and so limits accuracy if they are used for molecules and properties outside the scope of fitting and validation.[5] Hence, advancing the computer simulation work requires a transferable approach to deriving non-empirical potentials, based on the theory of intermolecular forces, and building on the electronic level modeling of intermolecular interactions.

Since electron correlation is the cause of the dispersion, which is important and sometimes dominant to the long-range attractive force between organic molecules, even calculating dimer interaction energies requires high-level electronic structure treatments.[6,7] As the charge densities of closed shell organic molecules are only slightly perturbed by intermolecular interactions, their intermolecular forces are well described by perturbation methods. Indeed, perturbational methods such as symmetry-adapted perturbation theory (SAPT),[4,8,9] and its variant based on density-functional theory [SAPT(DFT)],[6,7] are ideally suited for this problem as they are not only accurate, and also, SAPT(DFT) is computationally efficient. Additionally, perturbation theory provides a rigorous match between the short-range energies and the long-range analytic expansions using multipoles, polarizabilities, and dispersion coefficients. The size of organic molecules means that a density-functional charge distribution of each molecule is the most practical basis for developing a model intermolecular potential using SAPT(DFT)[6,7] to provide the essential link between electronic and atomistic level modeling of the intermolecular forces.

In SAPT(DFT),[10,11] the intermolecular interaction energy between a pair of rigid molecules is defined through the sum of energy components,

$$E_{int} = E_{elst}^{(1)} + E_{exch}^{(1)} + E_{IND}^{(2)} + E_{DISP}^{(2)} + \delta_{int}^{HF}. \tag{1}$$

Here, $E_{elst}^{(1)}$ and $E_{exch}^{(1)}$ are the first order electrostatic and exchange–repulsion energies, respectively. $E_{IND}^{(2)}$ is the second-order induction energy, which includes the exchange-induction energy, and $E_{DISP}^{(2)}$ is the total second-order dispersion energy, which includes the exchange-dispersion energy,

$$E_{IND}^{(2)} = E_{ind}^{(2)} + E_{exch-ind}^{(2)} \text{ and } E_{DISP}^{(2)} = E_{disp}^{(2)} + E_{exch-disp}^{(2)}. \tag{2}$$

The term δ_{int}^{HF} estimates the main error in truncating the expansion to second-order[12] and is dominated by the higher order induction energy contributions. For condensed phases, the usual summation over the intermolecular pair potential between all pairs of molecules in the system is an approximation (the pairwise additive approximation). The induction energy includes contributions from quantum charge-delocalization and classical polarization effects.[13] While both terms are likely to be non-additive, it is the classical polarization contribution that is dominant in the condensed phase as it is long-ranged. This term is difficult to evaluate and include in simulations as it requires numerical iteration to self-consistently account for the mutual polarization of all components in the lattice.[14]

SAPT(DFT) and related supermolecular partitioning methods have been the basis of various methods of deriving non-empirical intermolecular potentials (see Refs. 15 and 16). Non-empirical potentials are particularly valuable for modeling energetic materials, where the key functional groups, such as NO_2, have rather different charge distributions from the N and O atoms in well-parameterized

biomolecular force-fields. The difficulties in handling explosives naturally leads to a lack of experimental data for empirical parameterization, and this has led to a desire for reliable computational prediction of condensed phase properties. Pioneering studies of the organic solid-state using analytical fits to SAPT(DFT) interaction energy calculations have concentrated on small energetic molecules,[17] such as RDX[18–20] and FOX-7[21] where predictive modeling requires reliable extrapolation into the repulsive region that is not sampled in fitting empirical potentials. These force-fields are in an isotropic atom–atom form for ease of use in molecular dynamics simulation programs.

This paper seeks a non-empirical model potential for use in Crystal Structure Prediction (CSP) studies of energetic molecules. The vision behind one of the earliest CSP methods was to predict whether an energetic molecule would crystallize in a sufficiently dense crystal to be a promising material prior to synthesis,[22] although CSP studies can also be used to help solve the crystal structures.[23,24] The anisotropy of the atomic charge density arising from features such as lone pairs and π electron densities has proved to be important in realistic CSP studies,[25,26] with one of the leading family of methodologies being based on using a distributed multipole representation of the molecular charge density.[27] While most CSP studies have used empirical isotropic *exp*-6 repulsion-dispersion potentials in conjunction with an anisotropic atom–atom electrostatic model, an early non-empirical potential that additionally included anisotropy in the repulsion[28] for $C_6Br_2FClH_2$ was successful in an international blind test.[29] This potential completely neglected the induction contribution to the lattice energy.[28] A model for the use of atomic dipolar polarizabilities has since been implemented in the organic solid state modeling code DMACRYS,[14] although the numerical iteration required means that the polarization energy cannot be included in structure optimization. A fully anisotropic non-empirical atom–atom model intermolecular potential for pyridine has been derived using the program CamCASP[30] and used within DMACRYS in CSP studies to study the polymorphs of pyridine.[16] The non-empirical potential was successful in identifying the structure of a high-pressure polymorph[31] unlike the empirical repulsion–dispersion potential FIT[14] that is often used in CSP studies.[16] Thus, this paper seeks to extend this approach to predicting the crystal structures of energetic molecules. The use of an atomistic intermolecular potential is far more computationally efficient than using periodic dispersion-corrected density functional (DFT-D) calculations on hundreds of low energy crystal structures, as it is being increasingly used in CSP studies.[27] Thus, further examples of evaluating the use of non-empirical intermolecular pair potentials in CSP are urgently needed.

The energetic molecule chosen for this study was the high-explosive trinitrobenzene (TNB), which is extremely volatile in its dehydrated powder form.[32] It has a melting point of ~400 K,[32,33] and although it was first discovered in 1883,[34] a crystal structure was not determined until 1972.[35] The polymorphism of TNB (Fig. 1) was first reported in 2004 after the addition of trisindane in an attempted cocrystallisation[36] that resulted in the serendipitous crystallization of forms II ($Z' = 2$) and III ($Z' = 1$). The experimental data on forms II and III are limited to the melting points and crystal structures, with form III having the highest melting point. This, and a force-field lattice energy estimate, suggests[36] that form III is more thermodynamically stable than kinetically favored form I. Form III
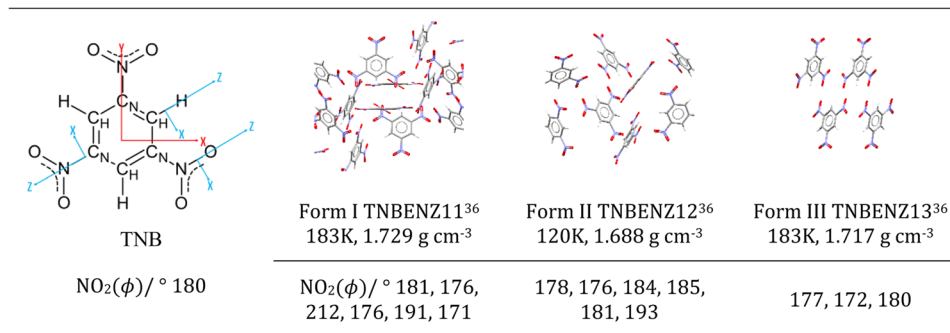
FIG. 1. The molecular and crystal structures of the energetic polymorphs of trinitrobenzene (TNB).[36] The form, Cambridge Structural Database reference code, temperature of the structure determination at ambient pressure, density, and NO_2 torsion angles (ϕ) are given below the visual representation of each experimental structure constructed using Mercury 3.6. The molecular diagram defines the atomic types and the molecule-fixed axes (red) and local atom-fixed axes (blue) used for the DIFF models, with the local axes for C_N and N and C_H and H being parallel.

is estimated to be more stable than form I by 0.6 kJ mol^{-1} by periodic PBE-D3(BJ) lattice energy calculations.[37]

The crystal structures of TNB (Fig. 1) show that the crystalline conformations do not differ much in the NO_2 torsion angles, and consequently, a rigid-molecule intermolecular potential is worth developing as a first approximation. TNB is highly symmetrical in its gas-phase optimized geometry so that there are only five distinct atomic types. This type of rigid-molecule approximation was often reasonably successful for substituted aromatic molecules in early CSP studies.[38,39] Incorporating flexibility into the force-field so that it accurately balances the inter and intramolecular forces is highly challenging, as shown by a contemporary study of α-RDX[20] as well as CSP studies on flexible pharmaceuticals.[40,41] Indeed, larger changes in the NO_2 torsion angles would affect the charge distribution of the nitro group to an extent that would require reparameterization of the model potential for accurate lattice energies.[42]

We can develop non-empirical derived intermolecular force-fields (DIFFs) for TNB with an atomic multipolar expansion of long-range components V_{LR} and an exponential model for the short-range terms V_{SR}. The functional form[1] for the interaction between molecules M and N containing atoms i and k of types ι and κ (types are C_H, C_N, N, O, or H) at a separation R_{ik} and orientation Ω_{ik} can be written as

$$V_{int} = V_{M,N}(R_{MN}, \Omega_{MN}) = V_{SR} + V_{elst} + V_{disp} + V_{pol}$$
$$= \sum_{i \in M, k \in N} U_{INT}^{ik}(R_{ik}, \Omega_{ik}) = G \exp\left[-\alpha^{\iota\kappa}\left(R_{ik} - \rho^{\iota\kappa}(\Omega_{ik})\right)\right]$$
$$+ Q_t^i T_{tu}^{ik} Q_u^k + \sum_{n=0}^{2} f_{2n+6}\left(\beta_{disp}, R_{ik}\right)\frac{C_{2n+6}^{\iota\kappa}}{R_{ik}^{2n+6}}$$
$$+ f_n\left(\beta_{pol}, R_{ik}\right)\left(\Delta Q_t^i T_{tu}^{ik} Q_u^k + \Delta Q_t^k T_{tu}^{ki} Q_u^{ik}\right). \tag{3}$$

The electrostatic and induction terms are implicitly summed over the permanent atomic multipoles Q_t^i (where t = lk; l is the total angular momentum, k is the z-component of angular momentum, and $l \leq 4$), induced dipoles ΔQ_t^i and the interaction tensors T_{tu}^{ik} are defined using the axes (Fig. 1) on the two molecules.[1] The induced moments and the isotropic dispersion coefficients, $C_6^{\iota\kappa}$, $C_8^{\iota\kappa}$, and $C_{10}^{\iota\kappa}$, are calculated from atomic polarizabilities that are derived from the

molecular charge distribution using linear-response Kohn–Sham theory together with suitable distribution techniques. The dispersion and polarization expansions are damped using the Tang–Toennies[43] damping function f_n. All the short-range terms are modeled by an exponential term where the atomic repulsion anisotropy $\rho(\Omega_{ik})$ is derived specifically for the molecule by fitting to SAPT(DFT) calculations, with guidance from the distributed density overlap model.[16] The partitioning of the interaction energy and molecular properties into atomic contributions is based on the basis-space version[44] of the iterated stockholder atoms approach of Lillestolen and Wheatley.[45,46]

The DIFF force-field functional form is complex. However, the complication of the anisotropy is likely to be important for many molecules as there is a significant difference in the reported quality of fits to SAPT(DFT) dimer intermolecular potential energy surfaces from using isotropic or anisotropic models. One automated generation of isotropic model potentials reported a typical error of about 0.8 kJ mol^{-1} in the negative energy region,[47] whereas a different approach that included anisotropy exhibits average errors of 0.2 kJ mol^{-1} in the attractive region for a different wide range of interacting dimers.[48] Given that lattice energy differences between polymorphs are often less than 2 kJ mol^{-1}, with about 80% calculated to be less than 4 kJ mol^{-1},[49] the modeling of the organic solid state requires the best possible accuracy. The solid state also samples the potential energy surface at a much wider set of relative separations and orientations in the known and CSP generated crystal structures than those that are tested in dimer calculations. Thus, the ability to reproduce crystal structures and the relative energies of known and hypothetical structures generated in a CSP study is a severe test of the accuracy of a force-field.

In this paper, a set of DIFF potentials was developed using a combination of CamCASP,[30] Psi4,[50] NWChem,[51] and ORIENT.[52] These potentials have the same long-range terms but increase the cost of development, first by the inclusion of anisotropy in the short range term and then allowing the isotropic short range coefficients to be relaxed by fitting to the total interaction energy, calculated up to second-order SAPT(DFT) interaction energies.[30,52] The size of TNB meant that we needed to make some changes in the DIFF potential development method compared to that used for pyridine[16] and

water;[31] however, the philosophy of the approach remains the same. The various DIFF model potentials were tested in a CSP ($Z' = 1$) study to assess whether form III was found and also to study the relative stability of all observed polymorphs. The aim is to advance our ability to model the condensed phases of energetic molecules such as TNB by developing a workflow for deriving a sufficiently realistic non-empirical potential and implementing it in a CSP study.

II. METHODS

The non-empirical potential derivation begins from the isolated molecular structure of TNB obtained by optimizing the isolated molecular conformation using the Gaussian09 program,[53] with the PBE0 functional[54–56] and the d-aug-cc-pVTZ Dunning basis.[57] This molecular conformation was held rigid for the development of the force-fields and in all simulations, except those used to assess this approximation. The three non-empirical DIFF potentials were constructed using the workflow in Fig. 2. These had the same long-range terms but differed in whether the short-range terms were isotropic, DIFF-srL0, or anisotropic with terms up to L = 2, DIFF-srL2. Additionally, the model parameters determined using first-order SAPT(DFT) energies were either allowed to relax against the total SAPT(DFT) intermolecular energies, indicated by "(rel)," or kept at their un-relaxed values, indicated by "(norel)." The comparisons were made with the non-empirical isotropic exp-6 model potential (FIT)[14] whose parameters were fitted to organic crystal structures and have widely been used in CSP studies. The FIT model was either combined with multipoles computed using the Gaussian-Distributed Multipole Analysis (GDMA)[58] method with the PBE0/6-31G(d,p) charge density, denoted as FIT(GDMA,L4), or the same ISA distributed multipoles used in the DIFF potentials [FIT(ISA,L4)]. Here, "L4" denotes that these multipoles contained terms of maximum rank of 4 on all sites.

A. Numerical details for electronic structure calculations

The molecular properties of TNB were computed using the modified basis-space iterated stockholder atoms (BS-ISA)[44,59] and the ISA-Pol[59] algorithms implemented in CamCASP.[30] The Kohn–Sham orbitals and orbital energies needed for these calculations were calculated using the NWChem. 6.6[60] code. We used the CS00 (Casida-Salahub) asymptotically corrected[61] PBE0 functional and the augA-Sadlej basis set,[62] a Sadlej-pVTZ basis that has been augmented with additional diffuse functions and functions of f-symmetry. This basis set was chosen as a cost-efficient yet accurate alternative to the larger d-aug-cc-VTZ basis set (see Fig. 1 of the supplementary material). The asymptotic shift (Δ) for TNB was calculated using the experimental ionization potential of 0.4028 a.u.[63] and the energy of the highest occupied molecular orbital, also computed with PBE0/augA-Sadlej, to give $\Delta = 0.0539$ a.u. The linear-response density functional calculations were performed using the hybrid ALDA+CHF kernel in CamCASP 6.0.[30]

The SAPT(DFT) dimer interaction energy calculations used to determine the induction energies and regularized induction energies for the polarization models were computed using CamCASP

6.0 with the above numerical details except that we used the aug-cc-pVTZ basis in the MC+BS format using a 3s2p1d1f midbond basis from the CamCASP basis library. Note that in CamCASP 6.0, the second-order exchange-induction energy is computed without the single-exchange (S^2) approximation using a closed-shell implementation of the general energy expression derived by Schäeffer and Jansen.[64] The first-order SAPT(DFT) interaction energies were computed using the above basis but with CamCASP 5.9. However, for reasons of computational efficiency, the interaction energies computed up to second-order were performed using the CamCASP with DFT orbitals and orbital energies calculated with the Psi4 1.1 code. In this case, the asymptotic correction was changed to be the GRAC correction[65] as Psi4 does not support the CS00 scheme. The two asymptotic correction schemes use the same long-range form but differ in the manner in which the long-range functional is spliced with the PBE0 functional. Consequently, some differences are to be expected, but these are likely smaller than the overall errors of the approach

B. Long-range terms

The molecular charge density was partitioned into atomic charge densities to obtain the permanent distributed multipole moments Q_t^i of rank $l \leq 4$ using the most recent iterated stockholder atoms algorithm[44] implemented in using CamCASP 6.0, the ISA-A algorithm.[59] The ISA-A method used the augA-Sadlej main basis with PBE0/CS00. The additional basis sets needed for the ISA calculation included the aug-cc-pVTZ[66] basis for the density-fitting and the aug-cc-pVQZ-RI basis modified with the ISA Set2 s-functions from the CamCASP basis library. The permanent atomic multipoles Q_t^i of rank up to hexadecapole were obtained for each atom.[59] The ISA-Pol[59] algorithm as implemented in CamCASP was used to derive the anisotropic dipolar (L1) polarizabilities $\alpha_{tt'}^{ik}$ for use in modeling the polarization energy and the isotropic L3 atom–atom dispersion model,[67] i.e., the C_6^{iK}, C_8^{iK}, and C_{10}^{iK}.[14,67] The long-range multipolar properties used in the DIFF potentials are given in the supplementary material.

C. Short-range potential parameters

The parameters in the short-range potential include the terms in the exponential part of the Born–Mayer potential and also the damping parameters for the dispersion and polarization models. The latter were determined by direct fitting to selected SAPT(DFT) second-order energies. However, for the Born–Mayer repulsive terms, we adopted the multi-stage fitting technique, which has been extensively described by Misquitta and Stone.[16] In this approach, the initial parameter estimates are obtained via the distributed density-overlap parameterized against a dense set of first-order SAPT(DFT) energies. This provides the potential DIFF-srL2(norel) and a model with only isotropic repulsion DIFF-srL0(norel).

The first-order exchange, $E_{exch}^{(1)}$, and electrostatic, $E_{elst}^{(1)}$, interaction energies were computed using a combination of NWChem. and CamCASP[16] for 2000 pseudo-random dimer configurations generated using the quasi-random Sobol sequence and Shoemake's uniform random rotations algorithm,[16] implemented in CamCASP and used previously for molecules as non-spherical as pyridine and $C_6Br_2ClFH_2$.[28] The distributed electrostatic energy, $V_{elst}^{(1)}$, was then
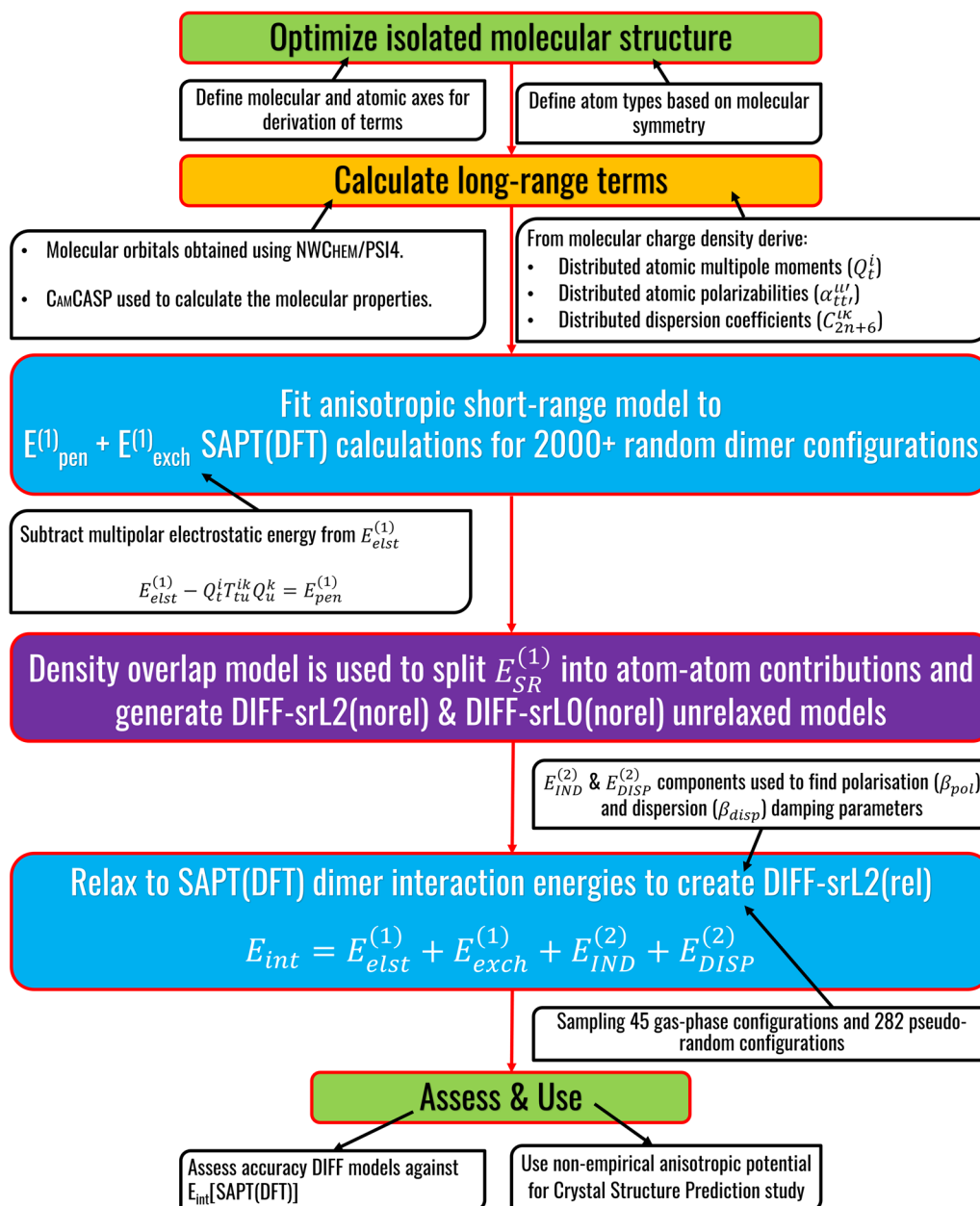
FIG. 2. A diagrammatic workflow of the stages in developing both the unrelaxed anisotropic and isotropic DIFF-srL2(norel), DIFF-srL0(norel), and the final relaxed anisotropic DIFF-srL2(rel) intermolecular potentials.

calculated for the 2000 dimer configurations using the multipoles Q_t^i and used to calculate the penetration energy ($E_{pen}^{(1)} = E_{elst}^{(1)} - V_{elst}^{(1)}$) for each dimer configuration.

The fitting of the first-order short range energies to the sum of atom–atom exponential terms in Eq. (3), $E_{exch}^{(1)} + E_{pen}^{(1)}$ ~ $\sum_{i \in M, k \in N} G \exp[-\alpha^{\iota\kappa}(R_{ik} - \rho^{\iota\kappa}(\Omega_{ik}))]$, allowing for all the

symmetry allowed shape-anisotropy terms of rank 1 or 2 (see Table VI of the supplementary material), is an ill-defined problem. Hence, the parameters are initially estimated by using the ISA atomic charge densities and assuming that $E_{SR}^{(1)}$ is proportional to the overlap of the charge distributions for each of the 15 different atom–atom interactions in TNB,

$$E_{SR}^{(1)} = E_{exch}^{(1)} + E_{pen}^{(1)} \approx \sum_{i\in M, k\in N} K^{ik} \int \rho_i(r)\rho_k(r)dr$$

$$= \sum_{i\in M, k\in N} K^{\iota\kappa} S^{ik}(\boldsymbol{R}) = \sum_{i\in M, k\in N} E_{SR}^{(1)}(ik). \quad (4)$$

The isotropic model was derived by using the fits to the distributed density overlap model to provide the initial parameters. These parameters were relaxed in stages to fit the 2000 $E_{exch}^{(1)} + E_{pen}^{(1)}$ energies using constraints to keep the relaxed parameter values close to the initial overlap model values. The anisotropic model was derived in the same way, with the fitting to the distributed density overlap model being done systematically to determine which anisotropic terms made a significant improvement on the isotropic fits. Once the choice of anisotropic terms had been made, these parameter values were tightly constrained in the relaxation stage. Further details are in 2.4 of the supplementary material.

The final potential parameters required for Eq. (3) are those required for damping the multipolar induction and dispersion expansion at a short range to allow for molecular overlap using the Tang–Toennies damping function,[43]

$$f_n(\beta_{MN}R_{ik}) = 1 - \exp(-\beta_{MN}R_{ik})\left(\sum_{m=0}^{n} \frac{(\beta_{MN}R_{ik})^m}{m!}\right). \quad (5)$$

Note that, in principle, the damping parameters are dependent on the sites involved. Indeed, this has been demonstrated for the water dimer,[4] but due to limitations of the DMACRYS[14] code, here, we are restricted to single damping parameters for the polarization and dispersion models. This results in a degradation of the quality of the resulting models and the need for compromises. A preliminary estimate of $\beta_{disp} = 1.8$ a.u. can be derived[59] from the ionization potential of TNB. This was investigated by comparing the second-order dispersion energies $E_{DISP}^{(2)} = E_{disp}^{(2)} + E_{exch-disp}^{(2)}$ and the multipolar dispersion energies $V_{DISP}^{(2)}$ of DIFF-srL2(norel) of both the configuration scan and crystal structure dimer datasets (defined below) for a range of values of β_{disp}. The estimate $\beta_{disp} = 1.8$ a.u. gave a good approximation of the dispersion energy at short-range, but a value of $\beta_{disp} = 1.65$ a.u. gave the best fit to the data (see 2.3.2 of the supplementary material) and so was used in all the DIFFs.

Defining the polarization damping parameter is more involved as this parameter is more site-dependent than the dispersion damping and also because the polarization damping needs to be determined by fitting not to the second-order induction energy, but to the second-order polarization energy defined by regularized SAPT(DFT). We compared $E_{POL}^{(2)} = E_{IND}^{(2)}(Reg)$ with the multipolar polarization energies $V_{POL}^{(2)}$ of DIFF-srL2(norel) for various values of β_{pol}. It has been suggested that $\beta_{pol} = \sqrt{\beta_{disp}} = 1.35$ a.u.,[68] but this damping is insufficient. No particular value of β_{pol} performed perfectly across all separations, particularly for repulsive configurations (see 2.2.2 of the supplementary material); however, $\beta_{pol} = 1.0$ a.u. was selected as it gave the smallest errors around the equilibrium distances of the low energy dimers and for the various two-molecule contacts in the crystal (see Figs. 3 and 4 of the supplementary material). These values of $\beta_{disp} = 1.65$ a.u. and $\beta_{pol} = 1.0$ a.u. were used in Eq. (3) to complete the parameterization of the isotropic DIFF-srL0(norel) and anisotropic DIFF-srL2(norel) model intermolecular potentials.

D. Improving DIFF by relaxing parameters to reproduce SAPT(DFT) interaction energies

1. Definition of SAPT(DFT) total (second order) interaction energies

The SAPT(DFT) energies to second-order include a charge-delocalization energy,

$$E_{CD}^{(2)} = E_{IND}^{(2)} - E_{IND}^{(2)}(Reg) = E_{IND}^{(2)} - E_{POL}^{(2)}.$$

Both the $E_{CD}^{(2)}$ and $E_{POL}^{(2)}$ include second-order exchange-induction effects.[13] $E_{CD}^{(2)}$ is very small for TNB except for highly repulsive configurations ($E_{int}^{(2)} > 150$ kJ mol^{-1}) and, in most dimers, is negligible in comparison with the polarization energy (see 2.6 of the supplementary material). Contributions to the interaction energy from third and higher-orders in the interaction operator are usually included via the δ_{int}^{HF} term; however, this term is known to overestimate these effects in complexes with little or no hydrogen bonding, as is the case for TNB. Furthermore, we do include higher-order polarization effects through the iterated polarization model. After investigation (see 2.6 of the supplementary material), we have not included the δ_{int}^{HF} term in this study.

Finally, three datasets of TNB dimer intermolecular energies, $E_{int}[SAPT(DFT)] = E_{int} = E_{elst}^{(1)} + E_{exch}^{(1)} + E_{POL}^{(2)} + E_{DISP}^{(2)}$, were calculated using the dimer configurations described below. The configuration scans and the pseudo-random dataset were used for refining the isotropic short-range parameters $\rho_{00}^{\iota\kappa}$ and $\alpha_{00}^{\iota\kappa}$ to absorb errors from the partitioning of the energies to give DIFF-srL2(rel) (see 2.5 of the supplementary material for details). A set of dimer orientations derived from the CSP-generated crystal structures was used as an independent test of the quality of all the model intermolecular potentials for TNB.

2. Generation of datasets of SAPT(DFT) dimer interaction energies

a. Gas-phase dimer orientation configuration scans. The gas phase dimers corresponding to minima on the DIFF-srL0(norel) were found using the basin-hopping algorithm, a two-part iterative global optimization treatment that employs a global stepping algorithm with local minimizations at each step[69] using ORIENT,[52] as a good way of locating minima on a bumpy potential energy surface. For TNB, an energy corresponding to kT at a temperature of 500 K determined the acceptance criterion for up to 100 000 steps with a maximum displacement $\Delta x = 0.10$ and rotation $\Delta\theta = 30°$ for each step. The CLUSTER procedure in CamCASP was used to find the unique dimer geometries orientated by their axes of inertia.

The configuration scan dataset, used for fitting the damping parameters and illustrating the potential fits, comprised up to 33 points for each dimer orientation, with the center of mass separation scaled from the minimum energy structure by a factor between 0.76 and 1.4. The set of dimers in this dataset was considered too correlated to use in the relaxation process, so a truncated dataset of only 9 points per dimer orientation (45 points in all) was used in the relaxation of the short-range parameters.

b. Pseudo-random dataset. A further 1000 random configurations were added to the 2000 generated for the first order energies by restarting the quasi-random Sobol sequence. The energies of these 3000 pseudo-random dimer configurations were evaluated using DIFF-srL2(norel). $E_{int}^{(2)}[SAPT(DFT)]$ calculations using Psi4/CamCASP were initiated for the 200 most stable and 130 least stable (most repulsive) configurations. The final pseudo-random $E_{int}^{(2)}[SAPT(DFT)]$ dataset comprised 190 random dimers between −10 kJ mol^{-1} and −22 kJ mol^{-1} (the energy of the most stable

pseudo-random dimer) and 92 repulsive structures with energies ranging from 3 kJ mol^{-1} to 65 kJ mol^{-1}. The results of an initial fit to the configuration scan dataset influenced this choice of points for the pseudo-random dataset that was added to it for the fitting.

c. Test dataset of crystalline conformations. The 100 most stable CSP structures obtained after lattice energy optimization with FIT(ISA,L4) were analyzed with the Crystal Similarity tool in Mercury[70] to give the geometries of all pairs of molecules in van der
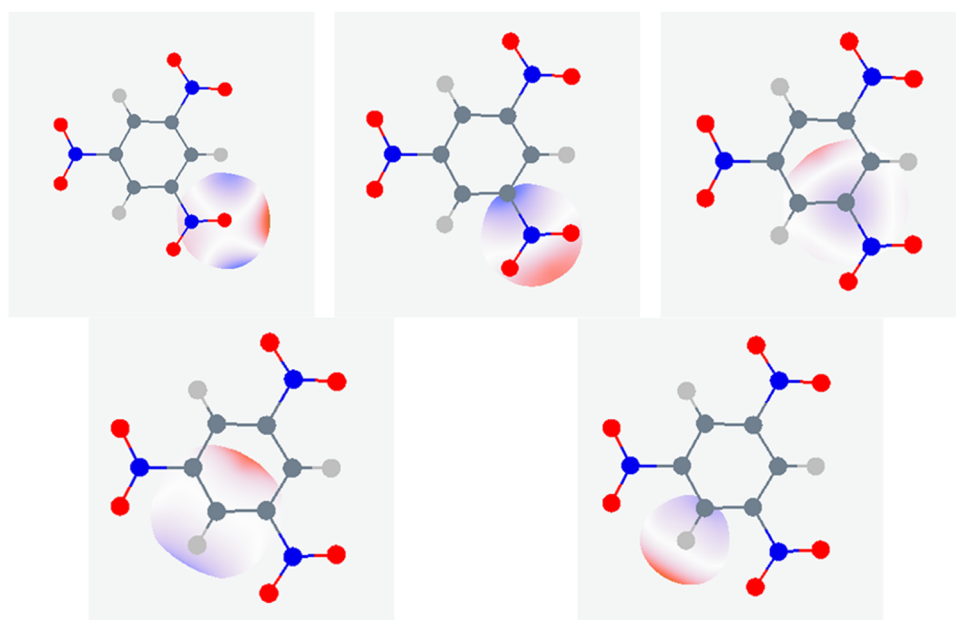

FIG. 3. The anisotropy of the charge distribution for each atomic type in TNB, as shown by the electrostatic potential (eV) arising from the cumulative effect of the ISA multipoles of ranks 1 to 4 on the iso-density surface of $10^{-3}e/bohr^3$. The scale is +0.4 eV (red) to −0.4 eV (blue).

TABLE I. The gas-phase minimum energy dimers of TNB calculated using the DIFF-srL0(norel) model together with their interaction energies computed with SAPT(DFT) and the model potentials. All energies are in kJ mol^{-1}. The energies in bold correspond to those optimized in determining the structures.

	S1	S2	S3	S4	T1
Models					
SAPT(DFT)	−21.79	−20.55	−24.43	−23.51	−19.35
DIFF-srL0 (norel)	**−26.26**	**−24.13**	**−21.01**	**−20.43**	**−18.57**
DIFF-srL2 (norel)	−25.46	−22.37	−25.13	−23.67	−18.48
DIFF-srL2 (rel)	−23.65	−20.58	−24.47	−23.86	−18.75
FIT (GDMA,L4)	−36.00	−35.35	−10.80	−24.02	−15.26
FIT (ISA,L4)	−25.58	−22.07	−22.73	−22.50	−20.58

Waals contact. $E_{int}[SAPT(DFT)]$ energies were calculated on the 66 most common distinct dimer configurations. This dataset was used to test the quality of the intermolecular potential models.

The $\rho_{00}^{l\kappa}$ and $\alpha_{00}^{l\kappa}$ short-range parameters of DIFF-srL2(norel) were relaxed to the $E_{int}[SAPT(DFT)]$ energies of the 45 of the 160 structures found in the configuration scans of the 5 gas-phase
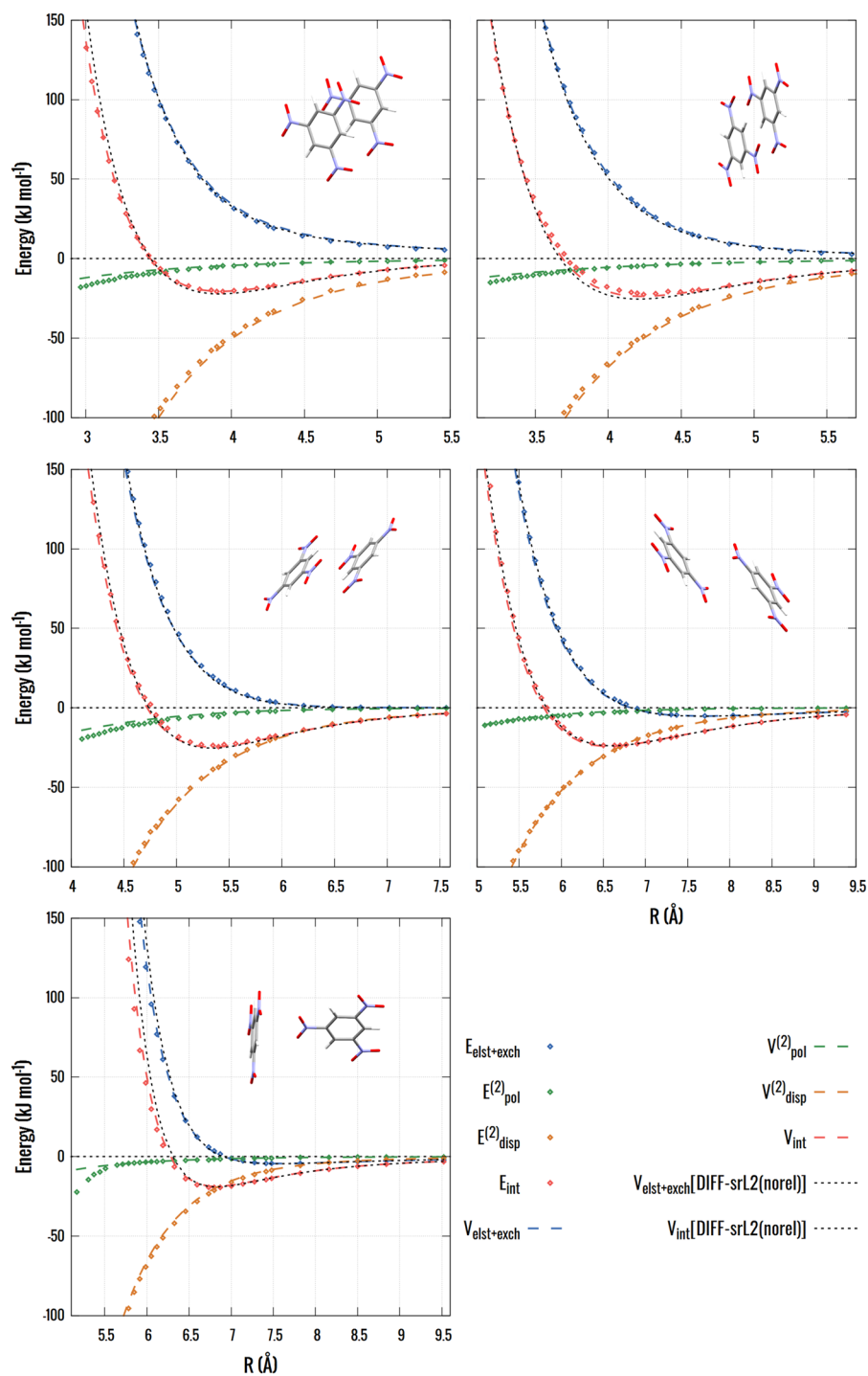

FIG. 4. The radial dependence of the $E_{int}[SAPT(DFT)]$ energy (points) for the gas-phase dimer orientations (the configuration scans) and the model potential V_{int} (dashes) for the different contributions in DIFF-srL2(rel). The total $E_{int}[SAPT(DFT)]$ and short range $E_{SR}^{(1)}$ energies are also given before relaxation, i.e., for DIFF-srL2(norel).

minima and the 282 pseudo-random structures, as described in 2.5 of the supplementary material.

E. Crystal structure prediction (CSP) and modeling for TNB

The TNB CSP study employed CrystalPredictor 2.0.1[71,72] to use the optimized isolated molecular structure of TNB to generate $Z' = 1$ crystal structures within the 59 most probable space groups[73] using the FIT potential and the point charges from the ISA multipole model [FIT(ISA,L0)]. This generated 13 956 unique structures that were lattice energy minimized using DMACRYS[14] with the three non-empirical potentials [DIFF-srL2(rel), DIFF-srL2(norel), and DIFF-srL0(norel)] and the empirical isotropic exp-6 potential FIT[14] with the two distributed multipole electrostatic models [FIT(ISA,L4) and FIT(GDMA,L4)]. Starting the lattice energy minimizations from the Crystal Predictor/FIT(ISA,L0), output structures ensured that more of the differences in the model potential energy surfaces were sampled than the usual hierarchical approach in CSP studies of minimizing from the lattice energy minima obtained with the previous model in the hierarchy of expected accuracy.

The structure optimizations were carried out without the polarization term. A Hessian check confirmed that each structure was a minimum in the lattice energy, and the structures clustered to remove duplicates. The polarization term was subsequently included as a single point energy evaluation, iterated so that the induced moments were maintained the crystal symmetry. Corresponding calculations were performed on the experimental crystal structures after the molecular conformation had been replaced by the optimized isolated molecular structure.

1. Dimer and crystal structure comparisons

The crystal structure similarity tool in Mercury 3.6[74] was used to determine the number of molecules, n, in a maximum coordination cluster (30 for crystals, 2 for dimer comparisons, and 1 for molecular conformations), which matched within a 30% distance in intermolecular atom–atom distances and 30° in interatomic intermolecular angles, and the corresponding root mean square difference, $RMSD_n$, excluding hydrogen atoms.
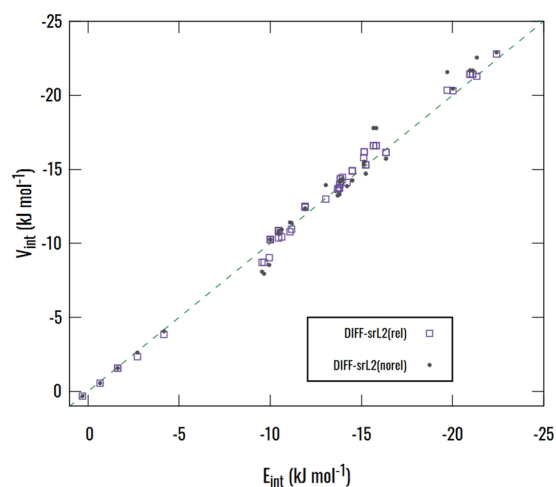

FIG. 5. Ability of the model potentials DIFF-srL2(rel) and DIFF-srL2(norel) to estimate the total intermolecular $E_{int}[SAPT(DFT)]$ energies for the crystal dimer dataset. The dashed line is $V_{int} = E_{int}[SAPT(DFT)]$.

III. RESULTS

A. Derivation and validation of the DIFF potentials

The planar optimized structure has 5 unique atoms, all of which have anisotropic charge distributions within the ISA-A partitioning, as shown in Fig. 3. The corresponding long-range potential parameters are given in Tables I–III of the supplementary material. However, the derivation of the short-range anisotropy showed that only anisotropy on O and N made a significant difference, reducing the weighted root mean squared (rms) residuals in reproducing the $E_{exch}^{(1)} + E_{pen}^{(1)}$ energies from 1.16 kJ mol^{-1} to 0.77 kJ mol^{-1} (see Table V of supplementary material). The required repulsion and its anisotropy can be explained by considering that these are the only atoms where the repulsive wall is sampled over a range of orientations.

TABLE II. The un-weighted root mean squared deviations (RMSD) in kJ mol^{-1} of the model potentials from the SAPT(DFT) calculated energies, $\sqrt{\sum \left(E_{int}^i - V_{int}^i\right)^2/N}$, for various datasets. The 160 configuration scan dimer energies and the 282 pseudo-random dimers energies were used in the relaxation of DIFF-srL2(rel). The crystal dimers points are a validation set, not used in the fitting, that all have with some atoms in van der Waals contact. The configuration scan with the gas-phase minima orientations with R-scalings ≥ 0.85 also samples the orientations found in condensed phases and excludes the highly repulsive short-range parts of the configuration scans. The RMSDs of the various models need to be considered together with the range of intermolecular energies of the dimers given in the second row.

Model	RMSD for the 160 gas-phase dimers (all energies)	RMSD for subset of 117 points with R-scaling ≥ 0.85	RMSD for 282 pseudo-random dimers	RMSD for 59 crystal dimers	RMSD for all 458 dimers
Energy range of E_{int}	$1000 \rightarrow -25$	$125 \rightarrow -25$	$54 \rightarrow -23$	$1 \rightarrow -23$	$1000 \rightarrow -25$
FIT(GDMA,L4)	61.2	11.0	9.1	11.169	9.9
FIT(ISA,L4)	58.2	6.6	3.6	1.537	4.4
DIFF-srL0(norel)	62.0	4.7	1.6	0.873	2.7
DIFF-srL2(norel)	81.0	9.7	1.9	0.794	5.1
DIFF-srL2(rel)	25.7	3.1	1.7	0.469	2.1

The dimer structures that correspond to minima in the intermolecular pair potential are fairly insensitive to the model, with there being four structures that have the planes of the molecules roughly parallel but significantly displaced and one where there is a large angle between the molecular planes (Table I). However, the relative energies and order of the minima differ significantly with the model potential.

The relaxation of the isotropic repulsion parameters to the total second-order energies can be very dependent on the choice of the dimer configurations included if these dimers are not chosen in a balanced way. If the fitting was just done to the configuration scan dataset (with 33 dimers for each of the 5 dimer orientations), there was an improvement representing those energies, but the energies of the crystal dimer dataset became significantly worse (see 2.5 of the supplementary material). This is a reflection of the heavily correlated dimers in the fitting set and that these orientations are not important in the crystalline phase. However, relaxation to the balanced dataset of pseudo-random dimers and some configuration scan points does improve the fit to the total energies, as is just detectable in Fig. 4 for the configuration scans. Each component is well represented by the model potential, apart from the polarization energy at a short range. The real test of the model potentials is in Fig. 5 for the independent test of the crystal dimer dataset. The relaxation has somewhat improved the model potential for the crystal dimers, although it was expensive to generate sufficient total intermolecular energies $\{E_{int}[SAPT(DFT)]\}$ so that the pseudo-random dimers covered sufficient orientation space for the relaxation step to be at all worthwhile.

The overall relative success of the different model potentials in representing the second-order energies is shown in Table II. This confirms that the model potentials give a good representation of the SAPT(DFT) energies, including the relative orientations sampled in

crystal structures, as well as to the points used in the weighted relaxation process (see 2.5 of the supplementary material), many of which are significantly up the repulsive wall.

B. Modeling the solid phase: Observed and CSP generated crystal structures of TNB

The CSP lattice energy landscape for the final potential DIFF-srL2(rel) is shown in Fig. 6. The most stable and densest structures are all very similar to form III with the same 30-molecule coordination sphere. The global minimum structure is considerably more stable than the lattice energy minima reached by optimizing from the experimental form III structure, by 12 kJ mol^{-1}, but overlays 30 molecules with an RMSD$_{30}$ of 0.6 Å. It seems likely that many, if not all, structures with a 30 molecule overlay with form III are in shallow potentials energy wells and would become the same dynamic structure in a MD simulation under ambient conditions, i.e., correspond to form III if this lattice energy landscape (CSP_0) were reduced by allowing the molecular motions.[75] The lowest energy structure that does not approximate to form III (the P*bca* structure, TNB No. 55 in Fig. 6) has two layers of molecules in common with form III but differs by the third layer (see Fig. 12 of the supplementary material).

The majority of computer-generated structures are less dense than form III, although denser than form II, and more stable than form I. A selection of the most diverse structures is shown in Table III, which emphasizes that only some of the crystal structures contain a dimer related to the gas phase minima (see Fig. 11 of the supplementary material), and none contain any of the three most stable dimer configurations. Hence, the CSP with DIFF-srL2(rel) is successful, within the limitations that only Z$'$ = 1 structures were generated, and so, forms I and II could not be found in the search.
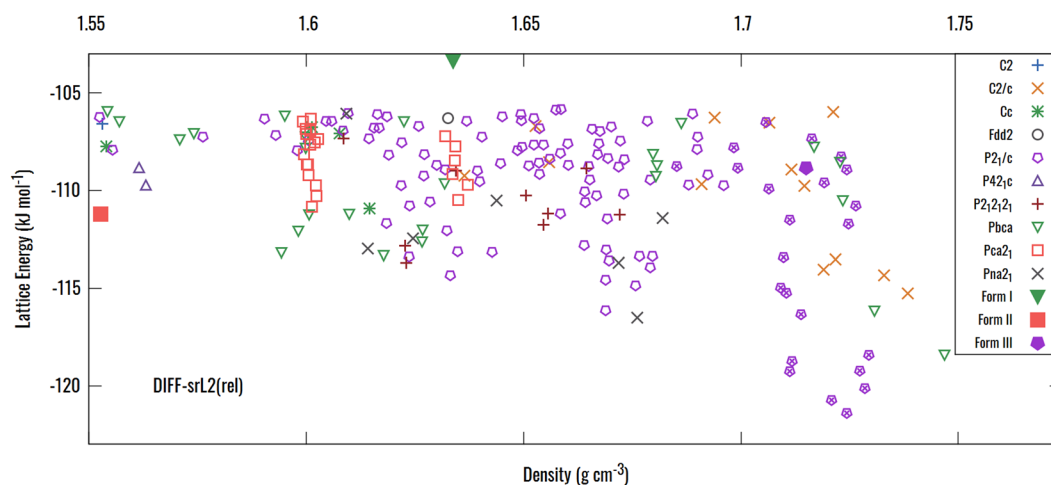

FIG. 6. The lattice energy and density of the 200 most stable CSP generated crystal structures of trinitrobenzene calculated with DIFF-srL2(rel). Each point represents a mechanically stable crystal structure classified by its space group. The lattice energy minima obtained by minimizing the experimental structures with the same rigid, planar molecular structure, and intermolecular potential are shown by the filled symbols corresponding to their space group. All CSP generated structures that had a 30/30 molecule overlay with form III are indicated by the pentagon and cross. The details for the 30 most stable structures and their structure identifier TNB number are presented in Table X of the supplementary material and all structures in .res format are available on request.

TABLE III. Solid-state energy contributions (in kJ mol^{-1}) for DIFF-srL2(rel) in the observed and the most diverse hypothetical structures. The structure labels are the order of stability after generation by CrystalPredictor, i.e., with the FIT(ISA,L0) empirical point charge model.

TNB No.	I	II	III	9955 (GM, ~III)	9	55	119	195	1612	440	5307	10 109
Space group	$Pbca$	$Pca2_1$	$P2_1/c$	$P2_1/c$	$Pna2_1$	$Pbca$	$C2/c$	Cc	$P2_12_12_1$	$Fdd2$	$P\bar{4}2_1c$	$Pca2_1$
Motifs	. . .	. . .	S4	S4	S4	S4	S4	. . .	. . .	S4	T1	S4
$V_{elst}^{(1)}$	−35	−42	−37	−42	−40	−41	−42	−40	−42	−26	−47	−45
$V_{disp}^{(2)}$	−111	−105	−125	−127	−124	−133	−132	−117	−114	−118	−110	−114
V_{sr}	58	55	66	68	68	73	73	64	63	56	63	66
V_{pol}	−14	−19	−12	−19	−21	−17	−14	−18	−20	−18	−17	−17
E_{latt}	−103	−111	−109	−121	−117	−118	−115	−111	−114	−106	−110	−111

The balance of contributions to the lattice energy varies considerably over the crystal structures (Table III), particularly the dominant dispersion energy, although this is partially offset by the variation in the short-range energies, as the denser structures are most stabilized by the dispersion. The polarization energy is significant, approaching half the electrostatic energy in some cases. Differences in the polarization energy account for a large proportion of the variation in lattice energies between the structures that approximate form III.

The relative stability of the CSP generated structures is very dependent on the potential energy model, as shown by the different crystal energy landscapes in Figs. 8–10 of the supplementary material and Fig. 7. The CSP study with the model potential commonly used for pharmaceuticals [FIT(GDMA,L4)] is almost useless, with all the known forms being unrealistically less stable than the hypothetical structures. Changing the electrostatic model to use the ISA partitioning together with a better quality charge density brings the structures within a more plausible energy range.

All the non-empirical DIFF potentials give the structures in a plausible energy range and have form I as metastable to forms II and III but differ in the relative rankings of forms II and III and the selected unobserved structures. The introduction of repulsion anisotropy introduces some reranking. The relaxation of the parameters by fitting to the second-order energies makes some difference, changing the relative stabilities of forms II and III and the structures that approximate form III from other hypothetical structures.

All the non-empirical models were lattice energy minimized without the polarization energy, which was then added as a single-point correction. Hence, the changes in relative energies include a contribution from small changes in the structure. The exception is the comparison between the final potential DIFF-srL2(rel) before (see Fig. 8 of the supplementary material) and after the polarization term (Fig. 6). This shows that the polarization energy does make a considerable difference in spreading out the energies, including the energy range of the structures that approximate form III.
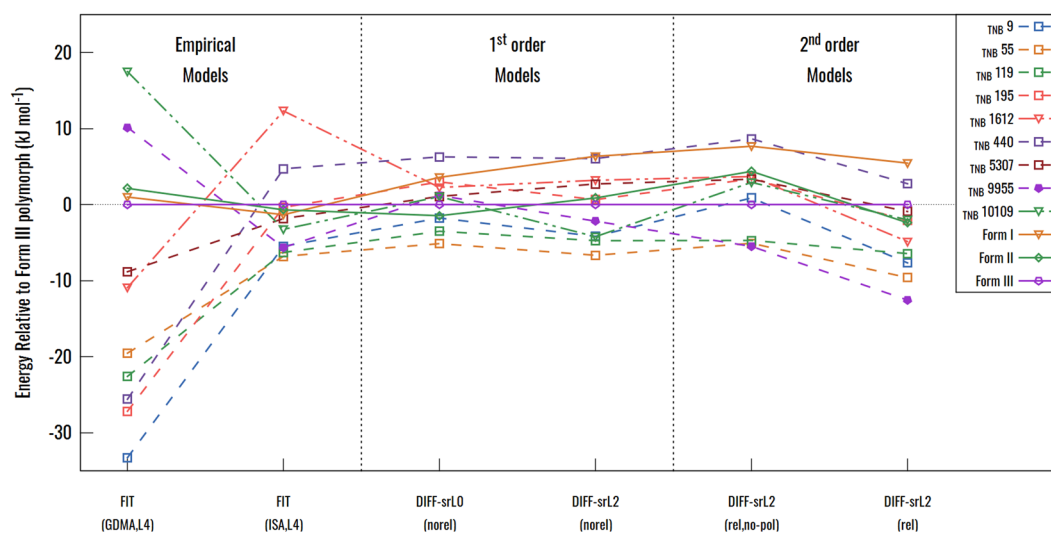
FIG. 7. Relative lattice energies of the observed and selected computer generated crystal structures of TNB in Table III relative to form III as a function of the model intermolecular potential.

However, the improvement of the electrostatic model within the model potential has the greatest effect on the CSP study.

IV. DISCUSSION

The final derived DIFF-srL2(rel) potential is highly successful in that it represents the intermolecular energies as calculated by SAPT(DFT) to a good accuracy over a wide energy range (Table II). The use of distributed multipoles, polarizabilities, and dispersion coefficients guarantees the correct long-range behavior, but the balance of the long and short-range terms is well represented in the potential energy wells for the low energy dimers (Fig. 4). This model potential has successfully translated to the solid state, providing a realistic CSP study for TNB. It is a significant improvement on a CSP study using the semi-empirical model potential [FIT(GDMA,L4)], which has been mainly used for pharmaceuticals. Indeed, early CSP studies of some chloronitrobenzenes[39] struggled to find an acceptable model potential. The improvement in the model potential was most dramatic on improving the distributed electrostatic model as the change in both the CSP (Fig. 7) and gas-phase dimer (Table I) results was greatest in going between FIT(GDMA,L4) and FIT(ISA,L4). However, the relative ranking of the minima in both phases was sensitive to the differences between the DIFF potentials.

A. The approximation of a rigid TNB molecule

This study has assumed that TNB is rigid and remains in the highly symmetrical planar conformation of the *ab initio* optimized isolated molecule upon crystallization. Analysis of the known polymorphs in Fig. 1 shows that this is an approximation, and the nitro groups twist somewhat to give non-planar structures. The most readily crystallized structure, form I, has a nitro group torsion that is more than 30° from planar in one of the two independent molecules. We estimated the effect of this conformational adjustment on crystallization by extracting the experimental molecular conformations from the crystal structures, recalculating the ISA multipoles and polarizabilities, and reoptimizing the crystal structures keeping the observed conformations rigid. This significantly improved the reproduction of form I and stabilized the lattice energies for forms II and III (see 3.1.1 of the supplementary material). However, even without including any energy penalty for the conformational change, the recently discovered forms II and III were more stable than form I. This sensitivity to the conformational change limits the conclusions that can be drawn from the CSP study of TNB. However, it seems very likely that form I is kinetically favored under most crystallization conditions, but the recently discovered forms II and III are more stable and that further polymorph screening and detailed crystallography may reveal more polymorphs or disorder in TNB. Ideally, a further CSP would be done in which the NO_2 groups of TNB could change conformation in response to the packing forces. This could be done by analytically rotating the multipoles during the structure relaxation,[42] but a final revaluation of at least the multipoles for each crystalline conformation would be needed for the final lattice energies.[42]

B. Many-body terms

A second approximation in the use of DIFF pair potentials in condensed phases is the modeling of the non-pairwise additive terms, i.e., the many-body terms that involve three or more molecules. The only non-additive term included in the DIFF potentials is the polarization energy and its exchange counterpart. The polarization contribution to the lattice energy is relatively large (Table III), several times the polarization energy per molecule for the dimers extracted from crystal structures $\left(\left| E_{POL}^{(2)}[SAPT(DFT)] \right| \right.$ < 3.5 kJ mol^{-1}); see Fig. 2 of the supplementary material}. This is in contrast to simple ionic solids, where the net electrostatic field at each ion in the crystal can be zero by symmetry so that the dipolar polarization energy is zero, although there would be a huge polarization energy for a pair of ions. The variation of the polarization energy between different crystals is significant (Table III). However, the fitting of a molecular damping parameter β_{pol} was not very satisfactory for TNB (see Fig. 2 of the supplementary material), and this could have made the polarization energy sensitive to the distance in the nearest neighbor contacts. This coupled with the uncertainty in the applicability of the δ_{int}^{HF} correction and not including the polarization energy in the lattice minimization could have led to an overestimation of the relative differences in the polarization contribution to the lattice energies of the form III type structures (Fig. 6).

The modeling of the polarization within the specific crystal structure via atomic polarizability tensors[76] is an advance on the isotropic polarizabilities within the AMEOBA force-field or the calculation of the distributed multipoles within a polarizable continuum model with a dielectric constant typical of those of organic crystals. The latter approach is often used in CSP studies,[77] as it often improves the relative energies between structures with different types of hydrogen bonding. We note that the molecular three-body Axilrod–Teller–Muto dispersion contribution was not included in modeling the crystals, and this would have provided a repulsive term that would share the sensitivity to the nearest neighbor atom–atom distances and at least partially cancel the attractive polarization term. Hence, it is important to extend our understanding of the many-molecule terms in condensed phase modeling, noting that the physics of many-body (non-additive) contributions limits the accuracy of the lattice energy evaluated from a very accurate two-body model such as the DIFF.

C. The partitioning into atoms (ISA) and effect on relaxation

This CamCASP-derived TNB DIFF model is the first time the ISA partitioning of the molecular charge density into atoms has been used for all the components of the potential for a molecule larger than water.[4] These atomic charge densities were used to derive the atomic multipoles,[44] polarizabilities,[59] and dispersion[59] parameters and, via the distributed density overlap model, to provide a first estimate of the repulsion potentials and form of anisotropy.[16] The more rigorous ISA partitioning method provides maximally spherical atoms, thus minimizing the number of important anisotropy terms. From this approach, we were able to identify the significant anisotropy on the O and N atoms in TNB (Fig. 3) and thereby construct a model with short-range anisotropy on these two atom types to significantly improve the model potential (see Table V of the supplementary material). The ISA partitioning also allows the use of better wavefunctions and hence improved polarizabilities and $C_6^{ικ}$,

$C_8^{i\kappa}$, and $C_{10}^{i\kappa}$ dispersion coefficients, which is particularly important given how the dispersion dominates the lattice energies (Table III). The use of the ISA partitioning to represent each individual component of the interaction energy is expected to reduce the errors in each contribution and hence reduce the need to relax the parameters to fit the total energies. Certainly, the unrelaxed model potential was already a good approximation for TNB and resulted in nearly the same relative energies of the crystal structures in the CSP study (Fig. 7). This is a promising finding for the development of potentials for large molecules, as the cost of deriving the DIFF-srL2(rel) potential was significantly more than the DIFF-srL2(norel). This is due to the need to calculate several hundred second-order dimer interaction energies, sampling a wide range of relative orientations, in order for the relaxation to be unbiased and hence an improvement (see 2.5 of the supplementary material). It is noteworthy that there are relatively few TNB gas-phase dimer structures (minima in the dimer potential energy surface, Table I), and so, the properties of TNB in the gas-phase would sample a very limited portion of the dimer intermolecular potential surface. The condensed phases sample a wide range of relative orientations (Table III) although the relative orientations in van der Waals contact are rarely those of the gas-phase dimers. Indeed, the non-empirical potentials have the further advantage of accurately representing the repulsive wall, which would be sampled when the crystals are compressed.[5] The use of the theory of intermolecular forces in the DIFF approach produces model pair potentials that should describe the intermolecular pair potential surface over the configurations sampled in all phases. This is a significant advantage over methods that depend critically on the sampling of configurational space, such as machine learning or fitting to total energies.[78–80]

D. Further work

An anticipated advantage of the ISA partitioning of the charge density is that the changes in the atomic potential parameters with conformation may be limited to the physical redistribution of charge with conformation[42] (i.e., the changes that go beyond the geometric changes that can be done using analytical tensor rotation). This could greatly improve the efficiency of modeling the changes in conformation that occur on crystallization. The CrystalPredictor[81,82] and CrystalOptimizer[82,83] approach to CSP models the effects of conformational changes by building up databases of conformational energies, associated forces, and distributed multipoles as a function of the torsion and bond angles that vary between polymorphs and using local approximate model interpolations. The costs of using the more accurate molecular charge densities found necessary in this study may be mitigated by the better transferability of the potential parameters with conformation. It is also possible that a DIFF-style intermolecular force-field and a separately parameterized atom–atom intramolecular force-field for the conformational energy penalty[84] could be used to efficiently generate and rank crystal structures in a CSP study. (The approximation that the same isotropic parameters can be used for intermolecular and intramolecular forces is a major limitation of traditional transferable force-fields, which has limited their usefulness in CSP studies.[84–86]) There is a great need to have force-fields that are accurate enough to give realistic relative stabilities of observable and hypothetical crystal structures generated

by CSP and can be used in molecular dynamics calculations to simulate the crystals under ambient temperature and determine the effect of varying the temperature and pressure.[75] The DIFF approach could provide such force-fields: DIFF-srL0(norel) potential could, in principle, be used in TINKER[87] after adjusting the atomic multipoles to have a maximum rank of 2.

V. CONCLUSIONS

We have derived a set of progressively more accurate and detailed non-empirical model intermolecular pair potentials for trinitrobenzene using the CamCASP and ORIENT programs. Central to these potentials is the Iterated Stockholder Atoms (ISA) partitioning of the molecular charge density into atomic charge densities using the basis-space ISA algorithm (BS-ISA) implemented in the CamCASP suite of codes. The potentials include analytic long-range parameters from the distributed atomic multipole tensors, distributed dipolar polarizability tensors, and distributed isotropic $C_6^{i\kappa}$, $C_8^{i\kappa}$, and $C_{10}^{i\kappa}$ dispersion coefficients, corresponding to a high-quality molecular charge density, and frequency-dependent molecular response functions. The long-range parameters were derived using the BS-ISA and ISA-Pol algorithms using the CamCASP code. The short-range potential parameters were determined through a hierarchical algorithm based on the distributed density-overlap model using the BS-ISA expansions for the atoms-in-a-molecule densities and SAPT(DFT) interaction energies. This algorithm involves multiple computationally well-defined steps that established a minimal set of molecule-specific atomic anisotropy terms. The initial short-range parameters in the analytical models were determined by fitting to an extensive set of computationally inexpensive SAPT(DFT) first-order exchange and penetration energies, $E_{exch}^{(1)} + E_{pen}^{(1)}$. The Tang–Toennies short-range damping of the dispersion and polarization used a molecular parameter that was fitted to second-order dispersion and polarization energies, respectively, although these parameters could have been estimated. The resulting model potential, DIFF-srL2(norel), gave a useful crystal structure prediction (CSP) study for TNB and reproduced the dimer interactions energies very well.

The SAPT(DFT) interaction energies up to second-order were then considered, with the charge-delocalization term found to be negligible and the δ_{int}^{HF} term found to be insignificant and therefore neglected. A dataset of SAPT(DFT) dimer energies, $E_{int} = E_{elst}^{(1)} + E_{exch}^{(1)} + E_{IND}^{(2)} + E_{DISP}^{(2)}$, was used to relax the isotropic short range parameters in order to absorb the errors in modeling the separate contributions and to also absorb the charge-delocalization energy. This final model, DIFF-srL2(rel), further improved the ability to model an independent dataset of total interaction energies of crystalline dimer configurations (not used in the potential derivation) with an RMSD of 0.5 kJ mol^{-1}. The improvement was sufficiently small to suggest that using ISA partitioning and SAPT(DFT) first-order calculations to produce a model for each significant contribution to the intermolecular potential could be extended to producing an accurate anisotropic atom–atom model intermolecular potential for much larger molecules than TNB.

The non-empirical DIFF intermolecular potentials were used in successful crystal structure prediction (CSP) studies of TNB, despite sampling a very different set of relative orientations of the molecules

than the minima in the dimer potential energy surface. The relative energies of the observed and computer-generated crystal structures of TNB were reasonable, unlike those generated using an empirical model potential frequently used in CSP for pharmaceuticals. The study supports other evidence that the recently discovered form III of TNB is the most stable polymorph.

The chief weaknesses of the intermolecular potential model, from both theory and the results, were in the treatment of the non-pairwise additive (many-molecule) terms and in the assumption that the molecule was rigid. We were unable to include non-additivity in the dispersion model, and the use of a single damping parameter in the polarization model may have resulted in the many-body polarization energy being too sensitive to the crystal structure. However, the rigid molecule approximation was probably the more important limitation for modeling the crystal structures of the known polymorphs and the CSP study, as small changes in the NO_2 torsion angles can significantly improve the packing of the molecules in some crystal structures.

This study makes significant progress toward a methodology for deriving non-empirical anisotropic atom–atom force-fields for organic molecules that would be capable of realistic modeling of all phases in molecular dynamics simulations.

SUPPLEMENTARY MATERIAL

See the supplementary material for the justification for choice of basis set for SAPT(DFT), the TNB atomic properties used in the DIFF potentials, the definition of polarization energy and damping the multipolar polarization energy, the distributed dispersion model, the derivation of the repulsion anisotropy, the refinement of the potential parameters, and discussion of the omitted terms and for further details of the crystal structure modeling of TNB, reproduction of experimental polymorphs, effect of conformational flexibility, sensitivity to the force-field, CSP results with different force-fields, and the CSP generated structures for DIFF-srL2(rel).

DEDICATION

This paper honors Ruth Lynden-Bell FRS, an outstanding theoretical chemist, who has mentored and inspired many students, including two of the authors, to pursue careers in academia during her career at the University of Cambridge and Queen's University Belfast.

ACKNOWLEDGMENTS

DATA AVAILABILITY

The model potentials are available as input files for ORIENT and DMACRYs from https://app.ph.qmul.ac.uk/wiki/ajm:potentials:start. The CSP generated crystal structures are available from the corresponding author on reasonable request.

REFERENCES

[1] A. J. Stone, *The Theory of Intermolecular Forces* (Oxford University Press, Oxford, 2013), Vol. 2.

[2] G. C. Maitland, M. Rigby, E. B. Smith, and W. A. Wakeham, *Intermolecular Forces: Their Origin and Determination* (Clarendon Press, Oxford, 1981).

[3] G. A. Cisneros, K. T. Wikfeldt, L. Ojamäe, J. Lu, Y. Xu, H. Torabifard, A. P. Bartók, G. Csányi, V. Molinero, and F. Paesani, "Modeling molecular interactions in water: From pairwise to many body potential energy functions," Chem. Rev. **116**(13), 7501–7528 (2016).

[4] R. A. J. Gilmore, M. T. Dove, and A. J. Misquitta, "First-principles many-body nonadditive polarization energies from monomer and dimer calculations only: A case study on water," J. Chem. Theory Comput. **16**(1), 224–242 (2020).

[5] A. A. Aina, A. J. Misquitta, and S. L. Price, "From dimers to the solid-state: Distributed intermolecular force-fields for pyridine," J. Chem. Phys. **147**(16), 161722 (2017).

[6] G. J. O. Beran, "Modeling polymorphic molecular crystals with electronic structure theory," Chem. Rev. **116**(9), 5567–5613 (2016).

[7] Y. S. Al-Hamdani and A. Tkatchenko, "Understanding non-covalent interactions in larger molecular complexes from first principles," J. Chem. Phys. **150**(1), 010901 (2019).

[8] A. J. Misquitta and K. Szalewicz, "Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers," J. Chem. Phys. **122**(21), 214103 (2005).

[9] K. Szalewicz, "Symmetry-adapted perturbation theory of intermolecular forces," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**(2), 254–272 (2012).

[10] A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz, "Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations," J. Chem. Phys. **123**(21), 214103 (2005).

[11] A. Hesselmann, G. Jansen, and M. Schutz, "Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies," J. Chem. Phys. **122**(1), 014103 (2005).

[12] M. Jeziorska, B. Jeziorski, and J. Čížek, "Direct calculation of the Hartree-Fock interaction energy via exchange perturbation expansion—The He = He interaction," Int. J. Quantum Chem. **32**(2), 149–164 (1987).

[13] A. J. Misquitta, "Charge transfer from regularized symmetry-adapted perturbation theory," J. Chem. Theory Comput. **9**(12), 5313–5326 (2013).

[14] S. L. Price, M. Leslie, G. W. A. Welch, M. Habgood, L. S. Price, P. G. Karamertzanis, and G. M. Day, "Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials," Phys. Chem. Chem. Phys. **12**(30), 8478–8490 (2010).

[15] M. J. Van Vleet, A. J. Misquitta, and J. R. Schmidt, "New Angles on Standard Force Fields: Toward a General Approach for Treating Atomic-Level Anisotropy," J. Chem. Theory Comput. **14**(2), 739–758 (2018).

[16] A. J. Misquitta and A. J. Stone, "*Ab initio* atom-atom potentials using CAM-CASP: Theory and application to many-body models for the pyridine dimer," J. Chem. Theory Comput. **12**(9), 4184–4208 (2016).

[17] K. Szalewicz, "Determination of structure and properties of molecular crystals from first principles," Acc. Chem. Res. **47**(11), 3266–3274 (2014).

[18] R. Podeszwa, R. Bukowski, B. M. Rice, and K. Szalewicz, "Potential energy surface for cyclotrimethylene trinitramine dimer from symmetry-adapted perturbation theory," Phys. Chem. Chem. Phys. **9**(41), 5561–5569 (2007).

[19] R. Podeszwa, B. M. Rice, and K. Szalewicz, "Crystal structure prediction for cyclotrimethylene trinitramine (RDX) from first principles," Phys. Chem. Chem. Phys. **11**(26), 5512–5518 (2009).

[20] H.-J. Song, Y.-G. Zhang, H. Li, T. Zhou, and F.-L. Huang, "All-atom, non-empirical, and tailor-made force field for alpha-RDX from first principles," RSC Adv. **4**(76), 40518–40533 (2014).

[21] D. E. Taylor, F. Rob, B. M. Rice, R. Podeszwa, and K. Szalewicz, "A molecular dynamics study of 1,1-diamino-2,2-dinitroethylene (FOX-7) crystal using a symmetry adapted perturbation theory-based intermolecular force field," Phys. Chem. Chem. Phys. **13**(37), 16629–16636 (2011).

[22] J. R. Holden, Z. Du, and H. L. Ammon, "Prediction of possible crystal-structures for C-, H-, N-, O- and F-containing organic compounds," J. Comput. Chem. **14**(4), 422–437 (1993).

[23] H. L. Ammon, Z. Du, J. R. Holden, and L. A. Paquette, "Structure of 1,1,5,5-tetranitro-[4]peristylane. Structure solution from molecular packing analysis," Acta Crystallogr., Sect. B **50**(2), 216–220 (1994).

[24] H. L. Ammon, Z. Du, R. D. Gilardi, P. R. Dave, F. Forohar, and T. Axenrod, "Structure of 1,3,5,7-tetranitro-3,7-diazabicyclo[3.3.0]octane. Structure solution from molecular packing analysis," Acta Crystallogr., Sect. B **52**(2), 352–356 (1996).

[25] 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).

[26] S. L. Price, "Control and prediction of the organic solid state: A challenge to theory and experiment," Proc. R. Soc. A **474**(2217), 20180351 (2018).

[27] 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. 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, "Report on the sixth blind test of organic crystal structure prediction methods," Acta Crystallogr., Sect. B **72**(4), 439–459 (2016).

[28] A. J. Misquitta, G. W. A. Welch, A. J. Stone, and S. L. Price, "A first principles solution of the crystal structure of $C_6Br_2ClFH_2$," Chem. Phys. Lett. **456**(1-3), 105–109 (2008).

[29] 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. Arnautova, 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 Crystallogr., Sect. B **65**(2), 107–125 (2009).

[30] A. J. Misquitta and A. J. Stone, CamCASP: A program for studying intermolecular interactions and for the calculation of molecular properties in distributed form, University of Cambridge, 2018, http://www-stone.ch.cam.ac.uk/programs.html#CamCASP.

[31] N. Giordano, C. M. Beavers, B. J. Campbell, V. Eigner, E. Gregoryanz, W. G. Marshall, M. Peña-Álvarez, S. J. Teat, C. E. Vennari, and S. Parsons, "High-pressure polymorphism in pyridine," IUCrJ **7**, 58–70 (2020).

[32] P. W. Cooper and S. R. Kurowski, *Introduction to the Technology of Explosives* (John Wiley & Sons, Inc., New York, 1996).

[33] M. Radomska and R. Radomski, "Calorimetric studies of binary systems of 1,3,5-trinitrobenzene with naphthalene, anthracene and carbazole. I. Phase transitions and heat capacities of the pure components and charge-transfer complexes," Thermochim. Acta **40**(3), 405–414 (1980).

[34] A. Claus and H. Becker, "Ueber trinitrotoluol und das flüssige dinitrotoluol," Ber. Dtsch. Chem. Ges. **16**(2), 1596–1598 (1883).

[35] C. S. Choi and J. E. Abel, "The crystal structure of 1,3,5-trinitrobenzene by neutron diffraction," Acta Crystallogr., Sect. B **28**(1), 193–201 (1972).

[36] P. K. Thallapally, R. K. R. Jetti, A. K. Katz, H. L. Carrell, K. Singh, K. Lahiri, S. Kotha, R. Boese, and G. R. Desiraju, "Polymorphism of 1,3,5-trinitrobenzene induced by a trisindane additive," Angew. Chem., Int. Ed. **43**(9), 1149–1155 (2004).

[37] J. C. Bennion, L. Vogt, M. E. Tuckerman, and A. J. Matzger, "Isostructural cocrystals of 1,3,5-trinitrobenzene assembled by halogen bonding," Cryst. Growth Des. **16**(8), 4688–4693 (2016).

[38] S. A. Barnett, A. T. Hulme, N. Issa, T. C. Lewis, L. S. Price, D. A. Tocher, and S. L. Price, "The observed and energetically feasible crystal structures of 5-substituted uracils," New J. Chem. **32**(10), 1761–1775 (2008).

[39] S. A. Barnett; A. Johnston, A. J. Florence, S. L. Price, and D. A. Tocher, "A systematic experimental and theoretical study of the crystalline state of six chloronitrobenzenes," Cryst. Growth Des. **8**(1), 24–36 (2008).

[40] C. C. Pantelides, C. S. Adjiman, and A. V. Kazantsev, "General computational algorithms for ab initio crystal structure prediction for organic molecules," Top. Curr. Chem. **345**, 25–58 (2014).

[41] M. A. Neumann, "Tailor-made force fields for crystal-structure prediction," J. Phys. Chem. B **112**(32), 9810–9829 (2008).

[42] A. A. Aina, A. J. Misquitta, M. J. S. Phipps, and S. L. Price, "Charge distributions of nitro groups within organic explosive crystals: Effects on sensitivity and modeling," ACS Omega **4**(5), 8614–8625 (2019).

[43] K. T. Tang and J. P. Toennies, "An improved simple-model for the van der Waals potential based on universal damping functions for the dispersion coefficients," J. Chem. Phys. **80**(8), 3726–3741 (1984).

[44] A. J. Misquitta, A. J. Stone, and F. Fazeli, "Distributed multipoles from a robust basis-space implementation of the iterated stockholder atoms procedure," J. Chem. Theory Comput. **10**(12), 5405–5418 (2014).

[45] T. C. Lillestolen and R. J. Wheatley, "Atomic charge densities generated using an iterative stockholder procedure," J. Chem. Phys. **131**(14), 144101 (2009).

[46] T. C. Lillestolen and R. J. Wheatley, "Redefining the atom: Atomic charge densities produced by an iterative stockholder approach," Chem. Commun. **44**(45), 5909–5911 (2008).

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

[48] M. J. Van Vleet, A. J. Misquitta, A. J. Stone, and J. R. Schmidt, "Beyond Born-Mayer: Improved models for short-range repulsion in ab initio force fields," J. Chem. Theory Comput. **12**(8), 3851–3870 (2016).

[49] A. J. Cruz-Cabeza, S. M. Reutzel-Edens, and J. Bernstein, "Facts and fictions about polymorphism," Chem. Soc. Rev. **44**, 8619–8635 (2015).

[50] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, "PSI4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability," J. Chem. Theory Comput. **13**(7), 3185–3197 (2017).

[51] E. Apra, E. Bylaska, W. de Jong, N. Govind, K. Kowalski, T. Straatsma, M. Valiev, H. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. Aquino, R. Atta-Fynn, J. Autschbach, N. Bauman, J. Becca, D. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauet, Y. Chen, G. Chuev, C. Cramer, J. Daily, M. Deegan, T. Dunning, M. Dupuis, K. Dyall, G. Fann, S. Fischer, A. Fonari, H. Fruchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. Gotz, J. Hammond, V. Helms, E. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. Johnson, H. Jonsson, R. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. Lins, R. Littlefield, A. Logsdail, K. Lopata, W. Ma, A. Marenich, J. del Campo, D. Mejia-Rodriguez, J. Moore, J. Mullin, T. Nakajima, D. Nascimento, J. Nichols, P. Nichols, J. Nieplocha, A. Otero-de-la-Roza, B. Palmer, A. Panyala, T. Piro-jsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. Richard, P. Sadayappan, G. Schatz, W. Shelton, D. Silverstein, D. Smith, T. Soares, D. Song, M. Swart, H. Taylor, G. Thomas, V. Tipparaju, D. Truhlar, K. Tsemekhman, T. Van Voorhis, A. Vazquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. Vogiatzis, D. Wang,

J. Weare, M. Williamson, T. Windus, K. Wolinski, A. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao, and R. Harrison, "NWChem: Past, present, and future," J. Chem. Phys. **52**(18), 184102 (2020).

[52] A. J. Stone, A. Dullweber, O. Engkvist, E. Fraschini, M. P. Hodges, A. W. Meredith, D. R. Nutt, P. L. A. Popelier, and D. J. Wales, *ORIENT: A Program for Studying Interactions between Molecules, 4.8.29* (University of Cambridge, 2015).

[53] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09, Gaussian, Inc., Wallingford, CT, 2009.

[54] C. Adamo and V. Barone, "Toward reliable density functional methods without adjustable parameters: The PBE0 model," J. Chem. Phys. **110**(13), 6158–6170 (1999).

[55] J. P. Perdew, M. Ernzerhof, and K. Burke, "Rationale for mixing exact exchange with density functional approximations," J. Chem. Phys. **105**(22), 9982–9985 (1996).

[56] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**(18), 3865–3868 (1996).

[57] T. H. Dunning, "Gaussian-basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen," J. Chem. Phys. **90**(2), 1007–1023 (1989).

[58] A. J. Stone, *DMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System, 2.2* (University of Cambridge, Cambridge, United Kingdom, 2010).

[59] A. J. Misquitta and A. J. Stone, "ISA-POL: Distributed polarizabilities and dispersion models from a basis-space implementation of the iterated stockholder atoms procedure," Theor. Chem. Acc. **137**(11), 153 (2018).

[60] H. J. J. van Dam, W. A. de Jong, E. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, and M. Valiev, "NWChem: Scalable parallel computational chemistry," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **1**(6), 888–894 (2011).

[61] M. E. Casida and D. R. Salahub, "Asymptotic correction approach to improving approximate exchange-correlation potentials: Time-dependent density-functional theory calculations of molecular excitation spectra," J. Chem. Phys. **113**(20), 8918–8935 (2000).

[62] A. J. Sadlej, "Medium-size polarized basis-sets for high-level correlated calculations of molecular electric properties," Collect. Czech. Chem. Commun. **53**(9), 1995–2016 (1988).

[63] See http://webbook.nist.gov/chemistry/ for NIST Chemistry WebBook; accessed on 16 June 2016.

[64] R. Schäffer and G. Jansen, "Single-determinant-based symmetry-adapted perturbation theory without single-exchange approximation," Mol. Phys. **111**(16-17), 2570–2584 (2013).

[65] M. Grüning, O. V. Gritsenko, S. J. A. van Gisbergen, and E. J. Baerends, "Shape corrections to exchange-correlation potentials by gradient-regulated seamless connection of model potentials for inner and outer region," J. Chem. Phys. **114**(2), 652–660 (2000).

[66] F. Weigend, A. Köhn, and C. Hättig, "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations," J. Chem. Phys. **116**(8), 3175–3183 (2002).

[67] A. J. Misquitta and A. J. Stone, "Dispersion energies for small organic molecules: First row atoms," Mol. Phys. **106**(12-13), 1631–1643 (2008).

[68] C. Millot, J.-C. Soetens, M. T. C. Martins Costa, M. P. Hodges, and A. J. Stone, "Revised anisotropic site potentials for the water dimer and calculated properties," J. Phys. Chem. A **102**(4), 754–770 (1998).

[69] D. J. Wales and H. A. Scheraga, "Global optimization of clusters, crystals, and biomolecules," Science **285**(5432), 1368–1372 (1999).

[70] C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek, and P. A. Wood, "Mercury CSD 2.0—New features for the visualization and investigation of crystal structures," J. Appl. Crystallogr. **41**, 466–470 (2008).

[71] P. G. Karamertzanis and C. C. Pantelides, "*Ab initio* crystal structure prediction—I. Rigid molecules," J. Comput. Chem. **26**(3), 304–324 (2005).

[72] P. G. Karamertzanis and C. C. Pantelides, "*Ab initio* crystal structure prediction. II. Flexible molecules," Mol. Phys. **105**(2-3), 273–291 (2007).

[73] J. Bernstein, "Crystal polymorphism," in *Engineering of Crystalline Materials Properties: State of the Art in Modeling, Design and Applications*, edited by J. J. Novoa, D. Braga, and L. Addadi (Springer, 2008), pp. 87–109.

[74] C. F. Macrae, P. R. Edgington, P. McCabe, E. Pidcock, G. P. Shields, R. Taylor, M. Towler, and J. van de Streek, "Mercury: Visualization and analysis of crystal structures," J. Appl. Crystallogr. **39**, 453–457 (2006).

[75] N. F. Francia, L. S. Price, J. Nyman, S. L. Price, and M. Salvalaglio, "Systematic finite-temperature reduction of crystal energy landscapes," Cryst. Growth Des. **20**, 6847–6862 (2020).

[76] G. W. A. Welch, P. G. Karamertzanis, A. J. Misquitta, A. J. Stone, and S. L. Price, "Is the induction energy important for modeling organic crystals?," J. Chem. Theory Comput. **4**(3), 522–532 (2008).

[77] T. G. Cooper, K. E. Hejczyk, W. Jones, and G. M. Day, "Molecular polarization effects on the relative energies of the real and putative crystal structures of valine," J. Chem. Theory Comput. **4**(10), 1795–1805 (2008).

[78] O. Egorova, R. Hafizi, D. C. Woods, and G. M. Day, "Multifidelity statistical machine learning for molecular crystal structure prediction," J. Phys. Chem. A **124**(39), 8065–8078 (2020).

[79] D. McDonagh, C.-K. Skylaris, and G. M. Day, "Machine-learned fragment-based energies for crystal structure prediction," J. Chem. Theory Comput. **15**(4), 2743–2758 (2019).

[80] F. Musil, S. De, J. Yang, J. E. Campbell, G. M. Day, and M. Ceriotti, "Machine learning for the structure-energy-property landscapes of molecular crystals," Chem. Sci. **9**(5), 1289–1300 (2018).

[81] M. Habgood, I. J. Sugden, A. V. Kazantsev, C. S. Adjiman, and C. C. Pantelides, "Efficient handling of molecular flexibility in ab initio generation of crystal structures," J. Chem. Theory Comput. **11**(4), 1957–1969 (2015).

[82] I. Sugden, C. S. Adjiman, and C. C. Pantelides, "Accurate and efficient representation of intramolecular energy in *ab initio* generation of crystal structures. I. Adaptive local approximate models," Acta Crystallogr., Sect. B **72**, 864–874 (2016).

[83] I. J. Sugden, C. S. Adjiman, and C. C. Pantelides, "Accurate and efficient representation of intramolecular energy in *ab initio* generation of crystal structures. II. Smoothed intramolecular potentials," Acta Crystallogr., Sect. B **75**, 423–433 (2019).

[84] O. G. Uzoh, P. T. A. Galek, and S. L. Price, "Analysis of the conformational profiles of fenamates shows route towards novel, higher accuracy, force-fields for pharmaceuticals," Phys. Chem. Chem. Phys. **17**, 7936–7948 (2015).

[85] M. D. Gourlay, J. Kendrick, and F. J. J. Leusen, "Rationalization of racemate resolution: Predicting spontaneous resolution through crystal structure prediction," Cryst. Growth Des. **7**(1), 56–63 (2007).

[86] S. Brodersen, S. Wilke, F. J. J. Leusen, and G. Engel, "A study of different approaches to the electrostatic interaction in force field methods for organic crystals," Phys. Chem. Chem. Phys. **5**(21), 4923–4931 (2003).

[87] J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren, and J. W. Ponder, "Tinker 8: Software tools for molecular design," J. Chem. Theory Comput. **14**(10), 5273–5289 (2018).