Accurate and Robust Molecular Crystal Modeling Using Fragment-Based Electronic Structure Methods

Gregory J.O. Beran, Shuhao Wen, Kaushik Nanda, Yuanhang Huang, and Yonaton Heit

Abstract Accurately modeling molecular crystal polymorphism requires careful treatment of diverse intra- and intermolecular interactions which can be difficult to achieve without the use of high-level ab initio electronic structure techniques. Fragment-based methods like the hybrid many-body interaction QM/MM technique enable the application of accurate electronic structure models to chemically interesting molecular crystals. The theoretical underpinnings of this approach and the practical requirements for the QM and MM contributions are discussed. Benchmark results and representative applications to aspirin and oxalyl dihydrazide crystals are presented.

Keywords Fragment methods · Molecular crystals · Polymorph prediction

Contents

G.J.O. Beran (✉), S. Wen, K. Nanda, Y. Huang, and Y. Heit
Department of Chemistry, University of California, Riverside, CA 92521, USA
e-mail: gregory.beran@ucr.edu

1   Introduction

Organic molecular crystals often exhibit a variety of different packing motifs, or polymorphs. These different crystal packing motifs can have diverse physical properties, making crystal structure critically important in a wide range of fields. Polymorphism plays a major role in the pharmaceutical industry, for example, where a substantial fraction of drugs including aspirin, acetaminophen, Lipitor, and Zantac have known polymorphs. Polymorphs of a given pharmaceutical can have drastically different solubilities and bioavailabilities, making the understanding of polymorphism critical for the drug industry.

Pharmaceutical polymorphism has led to several major drug recalls or withdrawals in recent years. For instance, the HIV drug ritonavir was temporarily removed from the market in 1998 when a new, insoluble polymorph appeared in production facilities, leading to shortages of this desperately needed medicine and costing its maker hundreds of millions of dollars in lost sales [1, 2]. Polymorphism is also believed to be behind multiple recalls of the anti-seizure drug carbamazepine, which exhibits several low-solubility polymorphs [3]. In 2008, Neupro brand skin patches for the Parkinson's disease drug rotigotine were withdrawn from the market when a less-effective crystal form appeared visibly on the patches as dendritic structures [4]. Pharmaceutical polymorphism has also been the subject of many legal battles arising from the fact that unique crystal forms are patentable. Major examples include the ulcer/heartburn medication Zantac and the antibiotic cefadroxil [5].

Crystal packing is also important for foods such as chocolate. Solid cocoa butter form V is desired to achieve chocolate with a shiny appearance, a $34°C$ melting point that causes it to melt pleasingly on the tongue, and other favorable characteristics. However, form VI cocoa butter, which leads to chocolate that is dull, soft, grainy tasting, and has a melting point a few degrees higher, is thermodynamically more stable [6]. At room temperature the transition from form V to form VI occurs on a timescale of months, and it occurs even faster at elevated temperatures. The chocolate industry expends considerable effort to produce and maintain chocolate in the proper form to ensure a high-quality product with a reasonable shelf life.

Many other areas of chemistry and materials science must cope with molecular crystal polymorphism as well. Crystal packing influences the stability, sensitivity, and detonation characteristics of energetic materials, for instance [7, 8]. It also can have drastic effects on organic semiconductor materials. Solid rubrene currently holds the record for the highest-known carrier mobility in an organic molecular crystal. However, a rubrene derivative with a different crystal packing motif exhibits no measurable carrier mobility [9].

Predicting molecular crystal structure from first principles is extremely challenging. The problem involves (1) a search over many (millions or more) possible crystal packings, (2) the accurate evaluation of the lattice energy of the possible structures (at 0 K), (3) calculation of the finite-temperature thermodynamic

contributions, and (4) an understanding of the competition between thermodynamically and kinetically preferred packing arrangements.

Significant progress on the "search" problem has been made in recent years. For example, the Price group uses a hierarchical series of ever-improving theoretical models to screen out unlikely structures and eventually identify the most stable forms [10, 11]. The initial screening might consider some ~10^7 randomly generated structures with different space group symmetries and numbers of molecules in the asymmetric unit cell using a very simple force field. This simple force field allows one to rule out a significant fraction of the energetically uncompetitive structures. Subsequent rounds improve the quality of the force field, winnowing down the potential structures toward a final prediction. Neumann and co-workers use a mixture of density functional theory (DFT) with dispersion corrections and a force-field fitted to match the DFT results to search over possible structures [12–15]. In their work the force field identifies a subset of likely structures which are then refined with DFT. Both strategies proved effective in the two most recent blind tests of crystal structure prediction [16, 17]. Recent progress in crystal structure global optimization algorithms may also help solve the search problem [18].

Once a relatively small number of candidate crystal structures have been identified, discrimination among them requires predicting the lattice energies or relative energies very accurately. This necessitates using a theoretical approach that can handle the subtle balances between intra- and intermolecular interactions that characterize conformational polymorphism [19]. One must treat the diverse non-covalent interactions – hydrogen bonding, electrostatics, induction (a.k.a. polarization), and van der Waals dispersion – with high and uniform accuracy to avoid biasing the predictions toward certain classes of structures (e.g., hydrogen-bonded vs π-stacking motifs). Overall, the energy differences between experimentally observed polymorphs are typically less than 10 kJ/mol, and they are often closer to ~1 kJ/mol.

After obtaining reliable predictions at 0 K, one can start to think about finite temperature effects. The computation of finite-temperature enthalpies, entropies, and free energies is much less mature. The entropic contributions to relative polymorph stabilities are commonly assumed to be smaller than the enthalpic ones [20], but of course there will be many exceptions. Thermal effects are typically estimated using simple (quasi-)harmonic approaches (e.g., [21]) though sampling-based free energy methods are starting to be explored more actively in this context (e.g., [22, 23]).

Finally, real-world crystallization is often driven by kinetics, not thermodynamics. Understanding the kinetics of crystallization is probably even more difficult than the thermodynamics, since it requires a dynamical understanding of the nucleation and crystal growth processes with sufficient accuracy to differentiate correctly among the different packing motifs. Progress in this direction is also being made, but it will likely be quite some time before one can reliably predict which crystal structures will form kinetically under a given set of crystallization conditions.

Nevertheless, the ability to predict the thermodynamically preferred structures reliably would be very useful for predicting temperature and pressure regions of phase stability, to determine whether a given pharmaceutical polymorph is stable or metastable, or to predict a variety of crystal properties. Toward this end, this chapter focuses on computing accurate lattice energies and relative energies at 0 K.

2 Theory

The theoretical treatment of molecular crystals has traditionally relied on either molecular mechanics (MM) force fields or electronic structure methods with periodic boundary conditions. Force field modeling of molecular crystals has improved dramatically over the past decade [19, 24–27], as demonstrated by the major successes in the recent blind tests of crystal structure prediction [11, 12, 15–17, 28]. Much of this success arises from the inclusion of increasing amounts of quantum mechanical information into the force fields, ranging from the parameterization to the determination of intramolecular conformation.

In that vein, one should be able to achieve even better accuracy by treating the systems fully quantum mechanically. This requires a careful balance between accuracy and computational expense. Most quantum mechanical (QM) calculations on molecular crystals are performed with periodic DFT. Widely used semi-local density functionals generally do not describe van der Waals dispersion interactions. However, there has been substantial progress toward including van der Waals dispersion interactions either self-consistently or as an a posteriori correction using a variety of empirical and non-empirical strategies [29–35]. Despite the tremendous progress in this area, it can sometimes be difficult to identify when DFT methods are performing well enough or to interpret the results when different density functionals make contradictory predictions.

Wavefunction methods offer the potential to improve the quality of the predictions systematically by improving the wavefunction. The simplest useful wavefunction technique for molecular crystals is second-order Møller–Plesset perturbation theory (MP2). A number of periodic MP2 implementations exist, including efficient ones based on local-correlation ideas [36–47]. These provide a nice alternative to DFT, but they remain relatively computationally expensive. Furthermore, MP2 correlation itself is often insufficient, as will be discussed below. Unfortunately, more accurate periodic coupled cluster implementations are too expensive to be applied to most chemically interesting molecular crystals [48–51].

2.1 *Fragment-Based Methods in Electronic Structure Theory*

Fragment-based methods provide a lower-cost alternative to traditional periodic boundary condition electronic structure methods. These techniques partition the

system in some fashion, perform quantum mechanical calculations on each individual fragment, and piece them together to obtain information about the system as a whole. While many individual fragment calculations are needed in a single crystal, each is relatively inexpensive compared to a full periodic crystal calculation. This allows one to utilize higher-level electronic structure methods at much lower computational cost.

Early fragment techniques include the fragment molecular orbital method [52, 53], divide-and-conquer techniques [54], and incremental schemes [55, 56], but there has been an explosion of fragment methods in recent years, as discussed in two reviews [57, 58] and in a thematic issue of *Physical Chemistry Chemical Physics* [59]. A couple of groups have also identified a common framework that unifies most or all of these fragment methods [60, 61].

One very common fragment strategy decomposes the total energy of a set of interacting molecules, whether in a cluster or a crystal, according to a many-body expansion:

$$E_{\text{total}} = E_{\text{1-body}} + E_{\text{2-body}} + E_{\text{3-body}} + \cdots. \tag{1}$$

The expansion is formally exact, but any computationally useful application of this expansion requires approximation of the higher-order terms in some fashion. For a typical molecular crystal, the three-body and higher terms account for 10–20% of the total interaction energy, making those terms necessary for accurate crystal modeling.

Approximations to the many-body expansion typically fall into two categories. The first category uses electrostatic embedding to incorporate polarization effects into the lower-order terms, thereby reducing the importance of the higher-order terms. The many-body expansion can then be truncated after two-body or three-body terms. Examples of such embedding approaches include the electrostatically-embedded many-body expansion approach [62–64], binary interaction [65–67], and the exactly embedded density functional many-body expansion [68].

The electrostatic embedding methods are very successful, but they suffer from two potential disadvantages. First, while it is often true that induction effects dominate the many-body contributions, many-body dispersion has also proved important in molecular crystal systems [30, 58, 69, 70], and those effects are not captured via electrostatic embedding. Second, embedding complicates the calculation of nuclear gradients and Hessians of the energy. When embedding a particular monomer or dimer in a potential arising from the other molecules, the embedding potential depends on the positions of all other atoms in the system. Therefore, the energy gradient of that monomer or dimer now depends on all 3N atomic coordinates, instead of just the coordinates of the monomer or dimer in question. These additional gradient contributions arising from the embedding-potential are often neglected, but they can sometimes be significant [65].

The methods in the second category of many-body expansion approximations make no truncation in the many-body expansion. Instead, the higher-order terms are approximated at some lower level of theory. The incremental method and

other related schemes treat the important low-order terms with accurate ab initio methods and approximate the higher-order terms using Hartree-Fock (HF), DFT, or any higher-level QM technique. Alternatively, the hybrid many-body interaction (HMBI) approach (and a nearly identical model simultaneously and independently proposed by Manby and co-workers [71]) approximates the higher-order terms using a polarizable force field. The advantage of this approach is that polarizable force fields are much less expensive to evaluate than electronic structure methods. Compared to embedding techniques, these approaches can avoid the need to make a priori assumptions about where to truncate the many-body expansion and provide the flexibility to build in whatever physical terms are needed in the system, including the aforementioned many-body dispersion effects.

2.2   The Hybrid Many-Body Interaction (HMBI) Method

In the HMBI model, the intramolecular (one-body) and short-range (SR) pairwise intermolecular (two-body) interactions are treated with electronic structure theory, while the long-range (LR) two-body and the many-body terms are approximated with the polarizable force field:

$$E_{\mathrm{HMBI}} = E_{\mathrm{1\text{-}body}}^{\mathrm{QM}} + E_{\mathrm{SR2\text{-}body}}^{\mathrm{QM}} + E_{\mathrm{LR2\text{-}body}}^{\mathrm{MM}} + E_{\mathrm{many\text{-}body}}^{\mathrm{MM}}, \qquad (2)$$

where the "many-body" term includes all three-body and higher interactions (Fig. 1). To evaluate the energy in practice, one exploits the fact that one can write a many-body expansion purely in terms of MM contributions:

$$E_{\mathrm{total}}^{\mathrm{MM}} = E_{\mathrm{1\text{-}body}}^{\mathrm{MM}} + E_{\mathrm{SR2\text{-}body}}^{\mathrm{MM}} + E_{\mathrm{LR2\text{-}body}}^{\mathrm{MM}} + E_{\mathrm{many\text{-}body}}^{\mathrm{MM}}, \qquad (3)$$

which can be rearranged as

$$E_{\mathrm{LR2\text{-}body}}^{\mathrm{MM}} + E_{\mathrm{many\text{-}body}}^{\mathrm{MM}} = E_{\mathrm{total}}^{\mathrm{MM}} - E_{\mathrm{1\text{-}body}}^{\mathrm{MM}} - E_{\mathrm{SR2\text{-}body}}^{\mathrm{MM}}. \qquad (4)$$

Substituting this expression into (2) leads to the following working HMBI energy expression:

$$E_{\mathrm{total}}^{\mathrm{HMBI}} = E_{\mathrm{total}}^{\mathrm{MM}} + \sum_i \left(E_i^{\mathrm{QM}} - E_i^{\mathrm{MM}} \right) + \sum_{ij} d_{ij}(R) \left(\Delta^2 E_{ij}^{\mathrm{QM}} - \Delta^2 E_{ij}^{\mathrm{MM}} \right), \qquad (5)$$

where E_i corresponds to the energy of monomer i, and $\Delta^2 E_{ij} = E_{ij} - E_i - E_j$ is the interaction energy between monomers i and j evaluated as the difference between the total energy of the dimer E_{ij} and the individual monomer energies E_i and E_j. Note that the monomer energies in these expressions are evaluated at the same
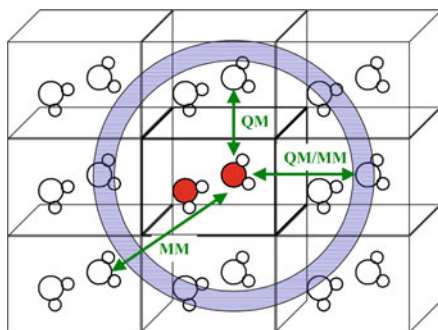
Fig. 1 Schematic of the HMBI method. Each individual molecule in the unit cell and its short-range pairwise interactions are treated with QM, while longer-range interactions and interactions involving more than two molecules are treated with MM. The *shaded region* indicates the smooth transition QM and MM via interpolation. Reprinted with permission from [115]. Copyright 2010 American Chemical Society

geometry as in the dimer (and in the full cluster or crystal). The $d_{ij}(R)$ term ensures no discontinuities arise in the potential energy surface as the model transitions from the short-range QM to the long-range MM two-body interaction regimes. This function decays from 1 to 0 as a function of the intermolecular distance R (defined here as the shortest distance between the two molecules) over a user-defined interval governed by two parameters, r_1 and r_0 [72]:

$$d_{ij}(R) = \frac{1}{1 + e^{2\left|r_1 - r_0\right|/(r_1 - R) - \left|r_1 - r_0\right|/(R - r_0)}}. \tag{6}$$

By default, we conservatively transition from QM to MM between $r_1 = 9.0$ Å and $r_0 = 10.0$ Å, but more aggressive cutoffs can often be used. In water/ice, for instance, using the ab initio force field described in the next section, one can transition from QM to MM between 4.5 and 5.5 Å with virtually no loss in accuracy (see Sect. 2.3).

The energy expression in (5) resembles the expressions used in a two-layer ONIOM-style QM/MM model. One computes the energy of the entire system at a low-level of theory, and then corrects it with smaller calculations at a higher level of theory. The physics described by the two types of models is quite different, however. In ONIOM QM/MM, one partitions the system into distinct QM and MM regions. In the fragment-based HMBI QM/MM approach, one instead partitions based on the nature of the interaction. There are no specific QM and MM regions. Each molecule has both QM and MM interactions in the HMBI model. The important interactions are treated with QM, while the less important ones are treated with MM. In this sense, the HMBI model is spatially homogeneous, as is appropriate for modeling a molecular crystal.

For systems with periodic boundary conditions like crystals, the HMBI energy expression in (5) must also include pairwise interactions between central unit cell molecules and their periodic images:

$$
\begin{aligned}
E_{\text{total}}^{\text{HMBI}} = E_{\text{total}}^{\text{MM}} &+ \sum_i \left(E_i^{\text{QM}} - E_i^{\text{MM}} \right) + \sum_{ij} d_{ij(0)} \left(\Delta^2 E_{ij(0)}^{\text{QM}} - \Delta^2 E_{ij(0)}^{\text{MM}} \right) \\
&+ \frac{1}{2} \sum_i \sum_{j(\mathbf{n})}^{\text{images}} d_{ij(\mathbf{n})} \left(\Delta^2 E_{ij(\mathbf{n})}^{\text{QM}} - \Delta^2 E_{ij(\mathbf{n})}^{\text{MM}} \right).
\end{aligned}
\tag{7}
$$

In this expression, i runs over molecules in the central unit cell, while j can either be in the central unit cell ($j(0)$) or in the periodic image cell $\mathbf{n}$ ($j(\mathbf{n})$). Unit cell $\mathbf{n}$ is defined as the cell whose origin lies at vector $\mathbf{n} = n_{v_1} \mathbf{v}_1 + n_{v_2} \mathbf{v}_2 + n_{v_3} \mathbf{v}_3$ in terms of the three lattice vectors, $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. In practice, thanks to the function $d_{ij}(R)$, the sum over $j(\mathbf{n})$ runs only over molecules within a distance r_0 of the current central unit cell molecule i.

The computational bottleneck in HMBI is the evaluation of the QM pairwise interaction energies $\Delta^2 E_{ij(0)}$ and $\Delta^2 E_{ij(\mathbf{n})}$. Because only short-range pairwise interactions are treated with QM, the HMBI model scales linearly with the number of molecules in the system (non-periodic) or unit-cell (periodic). More precisely, the MM terms do not scale linearly, but their cost is so much smaller than that of the QM terms for any practical system that linear scaling behavior is observed. For non-periodic systems, the onset of linear scaling occurs once the system becomes larger than the outer extent of the QM-to-MM transition region, r_0. In periodic systems, which are formally infinite, linear-scaling behavior is observed for unit cells of any size. This linear-scaling behavior is a key advantage of fragment methods compared to fully QM methods like periodic DFT or MP2. Large unit cells or supercells of the sort that might be used to perform lattice dynamics or to examine localized/defect behavior can be much cheaper with a fragment method than with a traditional periodic QM methods.

2.2.1   Nuclear Gradients and Hessians

The HMBI energy expression contains only additive energy contributions and does not use any sort of embedding, so derivatives of the energy can be computed straightforwardly. For instance, if q_l corresponds to the Cartesian x, y, or z coordinates (not its fractional coordinate) of the l-th atom in the central unit cell, then the gradient of the energy with respect to q_l is given by [73]

$$\frac{\partial E_{\text{total}}^{\text{HMBI}}}{\partial q_l} = \frac{\partial E_{\text{total}}^{\text{MM}}}{\partial q_l} + \sum_i \left(\frac{\partial E_i^{\text{QM}}}{\partial q_l} - \frac{\partial E_i^{\text{MM}}}{\partial q_l} \right) + \sum_{ij(\mathbf{0})} \frac{\partial d_{ij(0)}}{\partial q_l} \left(\Delta^2 E_{ij(0)}^{\text{QM}} - \Delta^2 E_{ij(0)}^{\text{MM}} \right)$$

$$+ \sum_{ij(\mathbf{0})} d_{ij(0)} \left(\frac{\partial \Delta^2 E_{ij(0)}^{\text{QM}}}{\partial q_l} - \frac{\partial \Delta^2 E_{ij(0)}^{\text{MM}}}{\partial q_l} \right)$$

$$+ \frac{1}{2} \sum_i \sum_{j(\mathbf{n})}^{\text{images}} \frac{\partial d_{ij(\mathbf{n})}}{\partial q_l} \left(\Delta^2 E_{ij(\mathbf{n})}^{\text{QM}} - \Delta^2 E_{ij(\mathbf{n})}^{\text{MM}} \right)$$

$$+ \frac{1}{2} \sum_i \sum_{j(\mathbf{n})}^{\text{images}} d_{ij(\mathbf{n})} \left(\frac{\partial \Delta^2 E_{ij(\mathbf{n})}^{\text{QM}}}{\partial q_l} - \frac{\partial \Delta^2 E_{ij(\mathbf{n})}^{\text{MM}}}{\partial q_l} \right).$$

$$(8)$$

The individual one-body and two-body energy gradient terms in (8) are obtained readily from the monomer and dimer gradients computed in standard electronic structure or MM software packages. The expression for the gradient of the QM-to-MM smoothing function d_{ij} has been provided previously [73].

To optimize the size and shape of the unit cell, one also needs the gradient with respect to the lattice vectors. When working in Cartesian coordinates instead of fractional coordinates, changing the lattice vectors does not affect the one-body or two-body terms within the central unit cell, but it does affect the two-body contributions due to interactions between molecules in the central unit cell and those in periodic image cells and the MM energy of the entire crystal. The resulting gradient with respect to the qth coordinate (x, y, or z) of the ϵth lattice vector ($\mathbf{v}_1$, $\mathbf{v}_2$, or $\mathbf{v}_3$) is given by

$$\frac{\partial E}{\partial v_{\epsilon q}} = \frac{\partial E_{\text{total}}^{\text{MM}}}{\partial v_{\epsilon q}} + \frac{1}{2} \sum_i \sum_{j(n_v)} n_v \sum_k \left\{ \frac{\partial d_{ij(n_v)}}{\partial q_k} \left(\Delta^2 E_{ij(n_v)}^{\text{QM}} - \Delta^2 E_{ij(n_v)}^{\text{MM}} \right) \right.$$

$$\left. + d_{ij(n_v)} \left(\left(\frac{\partial E_{ij(n_v)}^{\text{QM}}}{\partial q_k} - \frac{\partial E_{j(n_v)}^{\text{QM}}}{\partial q_k} \right) - \left(\frac{\partial E_{ij(n_v)}^{\text{MM}}}{\partial q_k} - \frac{\partial E_{j(n_v)}^{\text{MM}}}{\partial q_k} \right) \right) \right\}.$$

$$(9)$$

where k sums over the atoms in periodic image monomer j. Often, one expresses the unit cell in terms of three lattice constants (a, b, and c) and three angles (α, β, and γ). The expressions for the nine components of the lattice vector gradients in (9) can be transformed into expressions for these six lattice parameters [73].

Note that, except for the term $\partial E_{\text{total}}^{\text{MM}}/\partial v_{\epsilon q}$, all the terms needed to compute the lattice vector gradients in (9) are already available from the gradients with respect to the atomic positions (8). The $\partial E_{\text{total}}^{\text{MM}}/\partial v_{\epsilon q}$ term can be computed inexpensively, since it is at the MM level. Therefore, the calculation of the lattice vector gradient typically requires minimal additional work once the nuclear gradients with respect to atomic position have been obtained. The evaluation of the one-body and

two-body QM gradients forms the computational bottleneck in evaluating the full
nuclear gradient.

The nuclear Hessian can be computed similarly by differentiating (8) with
respect to a second nuclear coordinate, $q_{l'}$:

$$
\begin{aligned}
\frac{\partial^2 E_{\text{total}}^{\text{HMBI}}}{\partial q_l \partial q_{l'}} = {} & \frac{\partial E_{\text{PBC}}^{\text{MM}}}{\partial q_l \partial q_{l'}} + \sum_i \left(\frac{\partial^2 E_i^{\text{QM}}}{\partial q_l \partial q_{l'}} - \frac{\partial^2 E_i^{\text{MM}}}{\partial q_l \partial q_{l'}} \right) \\
& + \sum_{ij(\mathbf{n})} \zeta_{ij} d_{ij(\mathbf{n})} \left(\frac{\partial^2 \Delta^2 E_{ij(\mathbf{n})}^{\text{QM}}}{\partial q_l \partial q_{l'}} - \frac{\partial^2 \Delta^2 E_{ij(\mathbf{n})}^{\text{MM}}}{\partial q_l \partial q_{l'}} \right) \\
& + \sum_{ij(\mathbf{n})} \zeta_{ij} \frac{\partial d_{ij(\mathbf{n})}}{\partial q_l} \left(\frac{\partial \Delta^2 E_{ij(\mathbf{n})}^{\text{QM}}}{\partial q_{l'}} - \frac{\partial \Delta^2 E_{ij(\mathbf{n})}^{\text{MM}}}{\partial q_{l'}} \right) \\
& + \sum_{ij(\mathbf{n})} \zeta_{ij} \frac{\partial d_{ij(\mathbf{n})}}{\partial q_{l'}} \left(\frac{\partial \Delta^2 E_{ij(\mathbf{n})}^{\text{QM}}}{\partial q_l} - \frac{\partial \Delta^2 E_{ij(\mathbf{n})}^{\text{MM}}}{\partial q_l} \right) \\
& + \sum_{ij(\mathbf{n})} \zeta_{ij} \frac{\partial^2 d_{ij(\mathbf{n})}}{\partial q_l \partial q_{l'}} \left(\Delta^2 E_{ij(\mathbf{n})}^{\text{QM}} - \Delta^2 E_{ij(\mathbf{n})}^{\text{MM}} \right),
\end{aligned}
\tag{10}
$$

where the sums over $j(\mathbf{n})$ run over molecules in both the central unit cell and in
periodic image cells, $\zeta_{ij} = 1$ for dimers where both monomers lie within the central
unit cell ($\mathbf{n} = 0$), and $\zeta_{ij} = 1/2$ for dimers where the second monomer lies outside
the central unit cell. Evaluation of the Hessian requires gradients and Hessians for
each individual monomer and dimer. Once the Hessian has been computed, one can
compute harmonic vibrational frequencies, lattice dynamics, statistical thermody-
namic partition functions, etc.

2.2.2   Crystal Symmetry

Molecular crystals often exhibit high symmetry, both translational (due to the
periodic boundary conditions) and space group symmetry, and significant compu-
tational savings can be reaped by exploiting this symmetry in the monomer and
dimer calculations. The straightforward approach would identify the space group of
the crystal and use the symmetry operations of that group to determine which
monomers and dimers are symmetry-equivalent. One need only compute the
energies and forces for the symmetry-unique monomers and dimers and then
scale their contributions based on the number of symmetry-equivalent monomers
or dimers.

Alternatively, one can simply rotate the monomers and dimers to a common
reference frame (e.g., aligned along the principle axes of inertia) and simply test
which monomer/dimer geometries are identical within some numerical threshold.

This approach avoids the need (1) to specify the space group and (2) to program the symmetry operations for all 230 space groups.

A sizable fraction of organic molecular crystals exhibit $P2_1/c$ symmetry, for which roughly fourfold computational savings can be obtained. The savings can be even larger for other space groups. For acetamide crystals in the R3c space group, one obtains 18-fold speed-ups by exploiting symmetry. So while the details are system-dependent, for a crystal with one molecule in the asymmetric unit cell and a conservative QM to MM transition distance, one might typically need to perform ~50 symmetry-unique dimer calculations.

2.3 Accurate Force Fields for Long-Range and Many-Body Interactions

The success of the HMBI approach depends critically on the quality of the polarizable force field used for the long-range two-body and the many-body terms; c.f. (2). This means that it needs to capture two-body electrostatics, two-body van der Waals dispersion, self-consistent long-range and many-body induction, and many-body dispersion interactions (approximated here with only the leading three-body Axilrod–Teller term):

$$E^{\mathrm{MM}} = E_{\text{2-body es}} + E_{\text{2-body disp}} + E_{\text{induction}} + E_{\text{3-body disp}}. \qquad (11)$$

In our early work [74] we used the Amoeba force field [75, 76], which includes all of these contributions except for the many-body dispersion. It works fairly well in this context, but even better results are achieved by constructing an ab initio force field (AIFF) "on the fly" based on QM calculations for each individual monomer [77, 78].

The key idea behind the AIFF is to parameterize the force field in terms of atom-centered distributed multipole moments [79–81], distributed polarizabilities [82, 83], and distributed dispersion coefficients [84]. These are obtained from the molecular electron density, the static polarizabilities, and the frequency-dependent polarizabilities, respectively. The form of the force field is well-justified at long ranges from intermolecular perturbation theory, and empirical short-range damping functions help avoid serious problems for shorter-range interactions [85, 86].

This AIFF model mimics the much more expensive QM treatment very well, as shown in Figs. 2, 3, and 4. The use of multipolar expansions and the lack of exchange terms lead to some problems at short range, but the long-range interactions are modeled very accurately. Figure 2 shows that the predicted lattice energy of ice Ih is virtually invariant to the distance at which HMBI transitions from a QM to an MM description of the long-range pairwise interactions once the two molecules are ~4.5 Å apart. For other systems, the corresponding transition distance may

Fig. 2 Performance of the AIFF for long-range pairwise interactions: Change in the lattice energy of ice Ih as a function of the QM to MM transition distance (r_0 in (6), and $r_1 = r_0 - 1$ Å). Starting around ~4.5 Å, the AIFF reproduces the QM to within 0.1 kJ/mol. Neglecting the MM terms entirely leads to much larger errors [58]. Adapted with permission of the PCCP Owner Societies
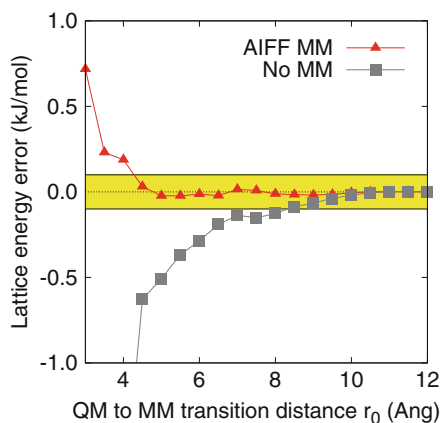
Fig. 3 Performance of the AIFF for many-body induction: error distributions for the AIFF and Amoeba force field many-body induction relative to MP2 for 101 (formamide)$_8$ geometries. The AIFF errors are smaller than the Amoeba ones and are nearly centered on zero. Adapted with permission from [77]. Copyright 2010 American Chemical Society
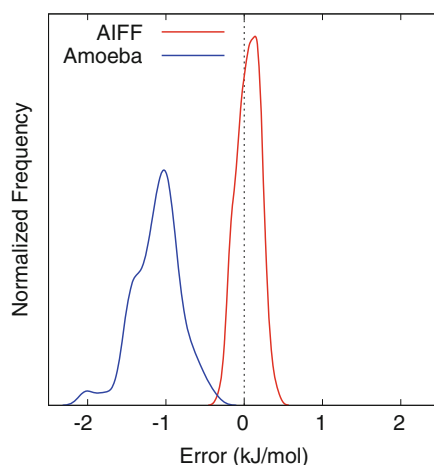
Fig. 4 Performance of the AIFF for three-body dispersion: Correlation between the HMBI three-body dispersion and the same contribution evaluated from symmetry adapted perturbation theory (SAPT) for 78 trimers taken from the formamide crystal [58]. Adapted with permission of the PCCP Owner Societies
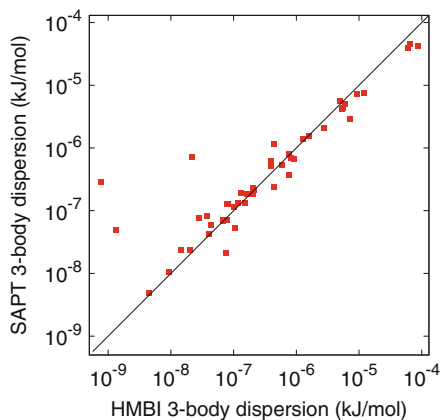
need to be longer (e.g., ~7 Å in formamide), but the AIFF always behaves well at sufficiently long distances [58].

The AIFF also reproduces the QM many-body induction effects accurately [87]. Figure 3 shows the errors in the many-body induction in a set of 101 (formamide)$_8$ geometries for the Amoeba force field and AIFF relative to RI-MP2. The Amoeba force field performs fairly well, but it systematically underestimates the many-body induction interactions by up to ~2 kJ/mol. The AIFF performs significantly better, with errors of roughly ±0.5 kJ/mol and a mean error very close to zero. Compared to Amoeba, the success of the AIFF stems from (1) its use of higher-rank multipoles (up to hexadecapole), (2) its use of higher-rank polarizabilities (up to quadrupole–quadrupole), and (3) the fact that these parameters are computed for each molecule in its current geometry rather than frozen at the values for some averaged/equilibrium geometry.

Finally, Fig. 4 demonstrates that the three-body dispersion model performs well compared to the three-body dispersion term in symmetry-adapted perturbation theory (SAPT). The simple, isotropic coupled Kohn–Sham dispersion coefficient representation provides a good approximation to the more complete SAPT calculation [58].

Overall, the AIFF does an excellent job of reproducing the QM interactions it replaces at much lower cost. HMBI predictions for the lattice energies of ammonia and carbon dioxide crystals differ from full periodic MP2 by only 1–2 kJ/mol, for instance [58].

2.3.1   AIFF Implementation

The long-range two-body electrostatics in the AIFF are implemented in standard fashion, with electrostatic interaction energy between two atoms A and B being given by

$$E_{\text{es}} \leftarrow \sum_{tu} Q_t^{\text{A}} T_{tu} Q_u^{\text{B}}, \tag{12}$$

where the T_{tu} matrix includes the distance- and orientation-dependent contributions for the interaction of two different spherical-tensor multipole moment components Q_t and Q_u. To evaluate the induction contributions, one first finds the induced multipole moments according to

$$\Delta Q_t^{\text{A}} = -\sum_{t'u} \alpha_{tt'}^{\text{A}} T_{t'u} \left(Q_u^{\text{B}} + \Delta Q_u^{\text{B}} \right), \tag{13}$$

where $\alpha_{tt'}^{\text{A}}$ is the static polarizability tensor on atom A and ΔQ is an induced multipole moment. Clearly the induced multipole moment on atom A depends on the induced multipole moment on atom B, so this process is done self-consistently until the induced multipoles reach convergence. Once the induced multipoles are

known, one can compute the induction energy contribution for the pair of atoms A and B according to

$$E_{\text{ind}} \leftarrow \sum_{tu} \Delta Q_t^A T_{tu} Q_u^B + Q_t^A T_{tu} \Delta Q_u^B. \tag{14}$$

The two-body dispersion between atoms A and B is evaluated via Casimir–Polder integration over the frequency-dependent polarizabilities:

$$E_{\text{2-body disp}} \leftarrow \sum_{tu} \sum_{t'u'} T_{tu} T_{t'u'} \int_0^\infty \alpha_{tt'}^A(i\omega) \alpha_{uu'}^B(i\omega) d\omega. \tag{15}$$

The integration is typically evaluated via numerical quadrature at ten frequencies ω. The expression in (15) corresponds to an anisotropic model for atom-atom dispersion. However, to a fairly good approximation, one can approximate this with a simple isotropic dispersion model (i.e., one which averages over the diagonal dipole–dipole and quadrupole–quadrupole elements of the frequency-dependent polarizability and neglects the off-diagonal elements). In that case, the dispersion model reduces to the standard C_6, C_8, etc., terms divided by the interatomic distance R to the corresponding power,

$$E_{\text{2-body disp}} \leftarrow \frac{C_6^{AB}}{R_{AB}^6} + \frac{C_8^{AB}}{R_{AB}^8} + \cdots. \tag{16}$$

The isotropic dispersion coefficients C_n are obtained from the Casimir–Polder integration over the appropriate elements of the isotropic frequency-dependent polarizabilities $\overline{\alpha}$:

$$C_n^{AB} \propto \int_0^\infty \overline{\alpha}^A(i\omega) \overline{\alpha}^B(i\omega) d\omega. \tag{17}$$

The many-body dispersion is approximated using the leading Axilrod–Teller three-body dispersion contribution. In the isotropic case, the three-body dispersion between atoms A, B, and C is computed according to

$$E_{\text{3-body disp}} \leftarrow C_9^{ABC} \frac{\left(1 + 3\cos\hat{A}\cos\hat{B}\cos\hat{C}\right)}{R_{AB}^3 R_{BC}^3 R_{AC}^3}, \tag{18}$$

where the R_{IJ} correspond to the distances between pairs of atoms and the cosines involve the interior angles of the triangle formed by the three atoms. The C_9 dispersion coefficient is obtained via Casimir–Polder integration over the frequency-dependent dipole–dipole polarizabilities on all three atoms:

$$C_9^{ABC} \propto \int_0^\infty \overline{\alpha}^A(i\omega) \overline{\alpha}^B(i\omega) \overline{\alpha}^C(i\omega) d\omega. \tag{19}$$

Note that most of the interaction terms described here also involve various empirical damping functions at short range to avoid divergences and the "polarization catastrophe," [78] but those are omitted here for simplicity.

The most difficult aspect of the force field implementation involves the evaluation of the $E_{\text{total}}^{\text{MM}}$ term in (7). The periodic crystal's induced multipoles are first computed self-consistently in a large, finite cluster consisting of the central unit cell and some number of periodic images (e.g., those within 25 Å). The algorithm works by inducing multipoles on the central unit cell molecules, replicating those induced moments on the periodic image molecules, and repeating the process until self-consistency is achieved. This procedure minimizes the edge effects arising from the finite cluster. Next, the permanent and induced multipole moments are combined and the overall interaction is evaluated via a multipolar Ewald sum [88]. Finally, the two-body and three-body dispersion contributions are evaluated via explicit lattice summation with large cutoffs. The evaluation of the AIFF interactions is cheap compared to the QM calculations, so one can afford to evaluate all of these contributions with tight cutoffs, thereby minimizing any errors due to truncating the lattice sums. See [78] for more details on the AIFF implementation and its performance for molecular crystals.

2.4 Electronic Structure Treatment of the Intermolecular Interactions

As the results described in Sect. 1 will demonstrate, high quality electronic structure methods need to be used for the QM one-body and two-body terms in HMBI. The electronic structure treatment must be capable of balancing the different types of intramolecular and intermolecular interactions to discriminate properly among different packing motifs. Ideally, one would employ high-level coupled cluster methods, like coupled cluster singles, doubles, and perturbative triples (CCSD(T)), with large basis sets, but this is usually impractical due to its N^7 scaling with system size N.

A more pragmatic approach will primarily use techniques that scale no more than N^5, like MP2. It is well known, however, that while MP2 describes van der Waals dispersion, it does so poorly [89–91]. Comparisons with intermolecular perturbation theory reveal the reasons for this poor performance: MP2 treats intermolecular dispersion interactions at the uncoupled HF (UCHF) level [92, 93]. In this view, the intermolecular dispersion interaction involves matrix elements of the intermolecular interaction between the ground state and excited state wavefunctions for molecules A and B divided by an energy denominator that depends on the excitation energies from ground to excited states:

$$E_{\text{disp}} = -\sum_{ab} \frac{<\psi_A^0 \psi_B^0 | \hat{V} | \psi_A^a \psi_B^b>}{E_A^a - E_A^0 + E_B^b - E_B^0}. \tag{20}$$

In UCHF (and MP2) these excited states for ψ^a and excitation energies $E^a - E^0$ are approximated using unrelaxed, ground-state HF orbitals and Fock matrix orbital eigenvalue energy differences.

Better results can be obtained by replacing the UCHF treatment of intermolecular dispersion with a coupled HF (CHF) [94] or coupled Kohn–Sham (CKS) [95, 96] one in which the excited states and excitation energies are computed with time-dependent HF or time-dependent DFT. Or one can obtain improved dispersion coefficients through other means [97]. The CKS variant has proved particularly successful, and is known as MP2C:

$$E_{\mathrm{MP2C}} = E_{\mathrm{MP2}} - E_{\mathrm{disp}}^{\mathrm{UCHF}} + E_{\mathrm{disp}}^{\mathrm{CKS}}. \qquad (21)$$

MP2C works very well for a variety of intermolecular interaction types across a wide spectrum of intermolecular separations and orientations [89, 91, 96, 98–102]. The dispersion correction scales N^4, so MP2C retains the overall fifth-order scaling of MP2. However, the prefactor for the dispersion correction is relatively large, and it consumes a non-trivial amount of computer time. For example, computing the HMBI RI-MP2C/aug-cc-pVTZ single-point energy of the aspirin crystal requires ~2,450 h for the MP2 and ~390 h for the dispersion correction. So while the underlying MP2 calculation clearly dominates, accelerating the MP2C dispersion correction would be beneficial.

MP2C appears to work very well for molecular crystal problems [70, 103]. Of course, other alternatives for accurately describing intermolecular interactions with similar computational cost exist, including spin-component-scaled MP2 methods [104–106], dispersion-weighted MP2 [107], and van der Waals-corrected density functionals. Recent improvements in the random-phase approximation (RPA) are also promising and may prove useful in the near future [108–111].

2.4.1 Faster MP2C for Molecular Crystals

Evaluating the MP2C dispersion correction requires computing UCHF and CKS density–density-response functions χ for each monomer at a series of frequencies ω,

$$[\chi_0(\omega)]_{PQ} = -4\sum_{ia} \frac{(P|ia)\epsilon_{ia}(ia|Q)}{\epsilon_{ia}^2 + \omega^2}, \qquad (22)$$

and then computing the dispersion energy via Casimir–Polder integration [96]:

$$E_{\mathrm{disp}} = -\frac{1}{2\pi} \int_0^\infty d\omega \widetilde{\chi}^{\mathrm{A}}(\omega) \mathbf{J}^{\mathrm{AB}} \left(\widetilde{\chi}^{\mathrm{B}}(\omega)\right)^T \left(\mathbf{J}^{\mathrm{AB}}\right)^T. \qquad (23)$$

In these expressions, i and a are occupied and virtual molecular orbitals, P and Q are auxiliary basis functions, ϵ_{ia} is the difference between the HF orbital energies for orbitals i and a, $J^{\mathrm{AB}} = (P^{\mathrm{A}}|r_{12}^{-1}|Q^{\mathrm{B}})$, $\widetilde{\chi} = \mathbf{S}^{-1}\chi\mathbf{S}^{-1}$, and $S = (P|r_{12}^{-1}|Q)$. For CKS, one must obtain the coupled density response functions from the uncoupled one in (22) by solving the Dyson equation [96].

Equations (22) and (23) are typically computed using a so-called dimer-centered (DC) basis, in which the calculations on the isolated monomers are carried out in the presence of ghost basis functions on the other monomer it is interacting with. This has two disadvantages for fragment-based calculations in molecular crystals. First, the ghost basis functions substantially increase the basis set size and therefore the computational cost. Second, when computing the many different pairwise intermolecular interaction energies, one must recompute each monomer's response function repeatedly for each different dimer interaction in which it is involved.

However, while the calculation of accurate CKS or UCHF dispersion energies requires large basis sets [112], it turns out that the energy difference between them is much less sensitive to the basis set [103]. In fact, one obtains nearly identical results when using a comparable monomer-centered (MC) basis (i.e., with no ghost basis functions). This indicates that the basis-set dependence of UCHF and CKS dispersion are similar, and the higher-order basis-set effects are well described in the DC UCHF dispersion that is inherent in the supermolecular MP2 calculation. It is unnecessary to replace those contributions with CKS ones. Figure 5 demonstrates the excellent agreement between the MP2C results obtained in monomer-centered and dimer-centered basis sets, nearly independent of whether one uses a double-zeta basis or extrapolated complete-basis-set limit.

In practice, switching to an MC basis accelerates the calculation of the dispersion correction for a single dimer roughly fivefold. Much more dramatic savings can be obtained in a molecular crystal, however. Thanks to space group symmetry and the use of periodic boundary conditions, typical molecular crystal unit cells contain no more than a handful of symmetry-unique monomers. Therefore, one needs only calculate the monomer density–density response functions in (22) for those few unique monomers. Then one can compute the dispersion interaction for each dimer according to (23) with trivial effort. For instance, using this approach reduces the computational time for evaluating the MP2C/aug-cc-pVTZ dispersion correction in crystalline aspirin by two orders of magnitude, from 390 CPU hours to only 2.8 CPU hours. At the same time, the MC MP2C approach affects the relative energies of the two polymorphs by no more than a few tenths of a kJ/mol relative to the DC MP2C values (e.g., Fig. 6). For all practical purposes, this approach makes computing the MP2C correction "free" for molecular crystals. With its high accuracy and low cost, MP2C makes an excellent choice for describing the non-covalent interactions in molecular crystals.

Of course, the aforementioned speed-ups only affect the evaluation of the dispersion correction. As noted previously, the underlying RI-MP2/aug-cc-pVTZ calculations in aspirin require ~2,450 CPU hours. Further computational savings must come from the MP2 part such as through local MP2 or other similar methods. One must be careful, however, that the truncation schemes do not hinder the ability to describe the non-covalent interactions at intermediate distances (i.e., beyond nearest-neighbor interactions but not yet far enough apart to be treated by the force field in HMBI).
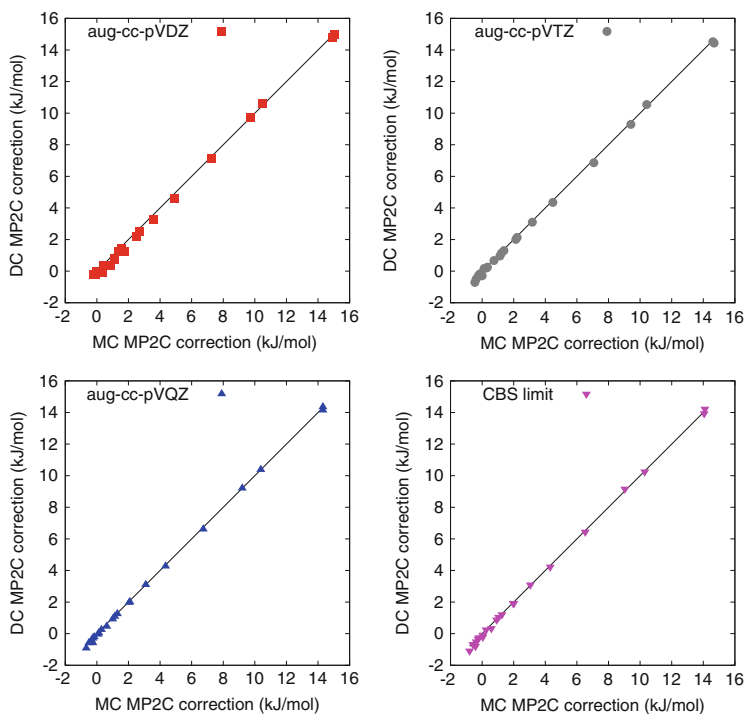
Fig. 5 For the dimers in the S22 test set [139], the MP2C dispersion correction calculated in an MC basis is nearly identical to that computed in a DC basis, even for a small aug-cc-pVDZ basis. Adapted with permission from [103]. Copyright 2013, American Institute of Physics
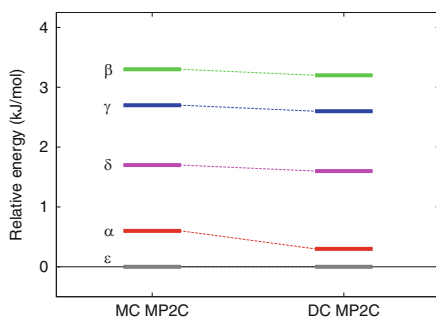

Fig. 6 Comparison between MC MP2C and DC MP2C for the relative energies of the five experimentally known polymorphs of oxalyl dihydrazide. The relative polymorph energies differ by 0.3 kJ/mol or less. Adapted with permission from [103]. Copyright 2013, American Institute of Physics

2.4.2 Basis Sets

It is well know that describing intermolecular interactions with correlated wave function methods requires large basis sets, so it is no surprise that molecular crystal energetics can often be very sensitive to the basis sets. This is particularly true for cases of conformational polymorphism, where the crystal packing motifs differ in both intramolecular conformation and intermolecular packing. Counterpoise corrections can help correct for basis set superposition error (BSSE). Different conformational polymorphs, however, may exhibit varying degrees of intramolecular BSSE, which is much harder to correct. In the oxalyl dihydrazide example discussed in Sect. 3.3, for example, the qualitative ordering of the experimental polymorphs differs dramatically as a function of basis set for exactly this reason. The intramolecular conformations of certain functional groups, such as the pyramidalization of nitrogen groups, can also be sensitive to basis sets.

In general, one should take care to ensure that molecular crystal calculations are converged with respect to basis set. Basis set extrapolation toward the complete basis set (CBS) limit, where feasible, can be very useful. One can also employ explicitly correlated techniques like MP2-F12 to achieve large-basis accuracy at reasonable computational cost [113, 114].

3 Performance and Applications of HMBI

To demonstrate the capabilities of the HMBI method, this section discusses molecular crystal benchmarks for predicting crystal geometries and lattice energies (Sect. 3.1) along with applications to polymorphic aspirin (Sect. 3.2) and oxalyl dihydrazide (Sect. 3.3) crystals.

The ability to predict crystal lattice energies accurately provides a demanding test for any theoretical treatment of molecular crystals. Lattice energies measure the energy required to dissociate the crystal to isolated gas-phase molecules. Whereas relative polymorph energies allow for some degree of error cancellation between the treatments of two different sets of non-covalent interactions, lattice energies expose any errors in calculating the strength of those interactions (at least for small, rigid molecules for which the intramolecular geometry is similar in the gas and crystal phases). The situation is analogous to the differences between computing reaction energies vs atomization energies, with the former being much easier due to cancellation of errors. In molecular crystals, the degree of error cancellation in the relative energies between polymorphs will depend on the differences in intramolecular configurations and intermolecular packing. In cases like aspirin, where the two polymorphs exhibit similar structures, substantial error cancellation occurs. On the other hand, cases like oxalyl dihydrazide exhibit more diverse interactions and packing modes, leading to less error cancellation and making it much more difficult to obtain accurate polymorphic energies.
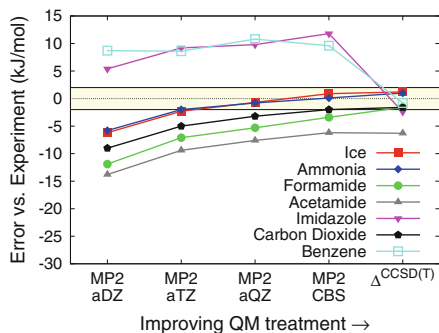
Fig. 7 Benchmark lattice energy predictions compared to experiment (in kJ/mol). Improving the QM treatment systematically improves the lattice energy predictions. The *shaded region* highlights ±2 kJ/mol to indicate the excellent agreement with the nominal experimental values. See [78] for complete details. Adapted with permission from [78]. Copyright 2011 American Chemical Society

3.1 Predicting Molecular Crystal Lattice Energies and Geometries

3.1.1 Lattice Energies

Benchmark lattice energy predictions have been performed on a series of seven small-molecule crystals: ammonia, acetamide, benzene, carbon dioxide, formamide, ice Ih, and imidazole. These crystals were chosen because they span a diverse range of intermolecular interactions, ranging from hydrogen-bonded to van der Waals dispersion bonded. The calculations start in a relatively small basis (aug-cc-pVDZ) RI-MP2, then increase the basis set toward the TQ-extrapolated basis set limit, and then finally examine higher-order correlation effects evaluated with CCSD(T) in a modest basis. Initially the Amoeba force field was used for the MM terms [115], but even better results are obtained when the AIFF (in the Sadlej basis set [116, 117]) described in Sect. 2.3 is used [78].

The primarily hydrogen-bonded crystals exhibit generally consistent improvement toward the experimental lattice energies as the electronic structure treatment of the one-body and short-range two-body terms is improved, as shown in Fig. 7. For those cases, the higher-order CCSD(T) correlation contributions are generally small. However, for the benzene and imidazole crystals, where the π-electron van der Waals dispersion interactions are significant, MP2 over-binds the crystals by ~10–12 kJ/mol (~10–20%). This sort of error is particularly problematic in the context of molecular crystal polymorphism, since MP2 will be biased toward such dispersion-bound packing motifs over hydrogen-bonded ones.

The CCSD(T) results correct the MP2 overestimate of the dispersion interactions in these two crystals. Overall, for six of the seven crystals, the estimated CBS-limit CCSD(T) results lie within 1–2 kJ/mol of the nominal experimental values, which

Table 1 Molecular crystal lattice energies in kJ/mol. MP2C/CBS-limit performs much better than MP2/CBS-limit relative to both the estimated complete-basis-set CCSD(T) results and experiment

Crystal	MP2[a]	MP2C[b]	CCSD(T)[a]	Experiment[c]
Ice (Ih)	59.9	60.3	60.2	59
Ammonia	39.3	40.5	40.2	39
Formamide	78.6	78.7	80.4	82
Acetamide	79.8	79.8	79.7	86
Carbon dioxide	29.1	26.3	29.5	31
Imidazole	102.8	93.1	88.6	91
Benzene		48.6	61.6	52

[a]Wen and Beran [78]
[b]Huang et al. [103]
[c]Beran and Nanda [115]

is likely within the experimental errors. The reason for the larger errors in acet-amide is unclear. Of course, CCSD(T) calculations will often be cost-prohibitive in molecular crystals. MP2C can provide a lower-cost alternative that substantially improves upon MP2. Table 1 demonstrates that including the MP2C dispersion correction has relatively small effects on the hydrogen-bonded crystals, but it reduces the large MP2 errors for the imidazole and benzene crystals from ~10–12 to ~2–3 kJ/mol, which is only slightly worse than the CCSD(T) results; see [103] for further details.

3.1.2 Crystal Structures

The HMBI model also provides accurate crystal geometries. Fully relaxed geom-etry optimizations at the HMBI counterpoise-corrected RI-MP2/aug-cc-pVDZ and Amoeba MM level for several of the benchmark crystals described above reproduced the experimental structures fairly accurately, including the space group symmetry [73]. Root-mean-square errors in the unit cell lattice parameters are only 1.6% relative to low-temperature crystal structures (100 K or below). For comparison, the same errors with B3LYP-D* [118] are 3.4% in the 6-31G** basis and 2.0% in the TZP basis. The HMBI RI-MP2 calculations also perform well for the root-mean-square deviations in the atomic positions (Table 2). Figure 8 shows overlays of the experimental and HMBI RI-MP2 optimized structures, highlighting the good agreement between them.

Overall, the RI-MP2/aug-cc-pVDZ structures are clearly better than dispersion-corrected B3LYP-D*/6-31G* [118]. They are slightly worse than those obtained with B3LYP-D* in a triple-zeta basis (TZP), though the MP2/aug-cc-pVDZ results exhibit slightly more uniform errors. In principle, larger-basis MP2 geometries ought to be even better, but optimizing in those larger basis sets becomes even more expensive. One must also worry about the effects of the MP2 treatment of dispersion on the geometries.

Given the sorts of practical calculations that are currently feasible, it is not obvious that fragment methods provide a significant advantage relative to DFT for molecular crystal structure optimization, especially since the basis set requirements

Table 2 Root-mean-square deviations (rmsd$_{15}$ [140] in Å) between the theoretically optimized and experimental geometries for clusters of 15 molecules taken from each crystal. The temperatures at which the experimental structures was obtained is also listed. See [73] for details

Crystal	B3LYP-D*/ 6-31G**	HMBI RI-MP2/ aug-cc-pVDZ	B3LYP-D*/ TZP	Experimental temperature (K)
Ice (Ih)	0.13	0.10	0.04	15
Formamide	0.29	0.16	0.22	90
Acetamide	0.16	0.08	0.08	23
Imidazole	0.20	0.12	0.14	103
Benzene	0.09	0.06	0.02	4

for correlated wavefunction methods like MP2 are steeper than those for DFT. For this reason, we often optimize the geometry using DFT and perform high-level single-point energies using HMBI. Of course, further improvements in MP2-like algorithms and continuing advances in computer hardware may change this.

3.2 Aspirin Polymorphism

The crystal structure of aspirin has been known since the 1960s [119], but suspicion of a second crystal form lingered for many years until 2004, when the structure of aspirin form II was predicted by the Price group [120]. Experimental confirmation for form II came a year later [121], though this report was followed by several more years of controversy in the crystallographic community [122, 123]. Only in the past couple years has the existence and structure of form II aspirin been firmly established [124–126].

The controversy arose from the fact that the two aspirin polymorphs exhibit very similar structures, and it proved very difficult to obtain pure crystals of form II. Rather, one often obtains mixed crystals containing domains of both forms I and II. The overall crystal packing of the two polymorphs is very similar except for the nature of the interlayer hydrogen bonding (Fig. 9). Whereas form I exhibits dimers, with each acetyl group hydrogen bonding to one other aspirin molecule in the adjacent layer, form II exhibits a catemeric structure, where each acetyl group hydrogen bonds to two adjacent molecules. This catemeric structure produces long chains of hydrogen bonds in form II.

Interestingly, earlier dispersion-corrected DFT studies suggested that form II was ~2–2.5 kJ/mol more stable than form I [127, 128], which would be surprising for two crystals that appear to grow together so readily. Other examples of crystals which form intergrowths are often separated by less than 1 kJ/mol in energy [129–131].

We investigated this system [132] using HMBI after optimizing the crystal structures for each form using B3LYP-D*/TZP. We performed a variety of calculations, including MP2, SCS(MI)-MP2 [104], and MP2C in both aug-cc-pVDZ and
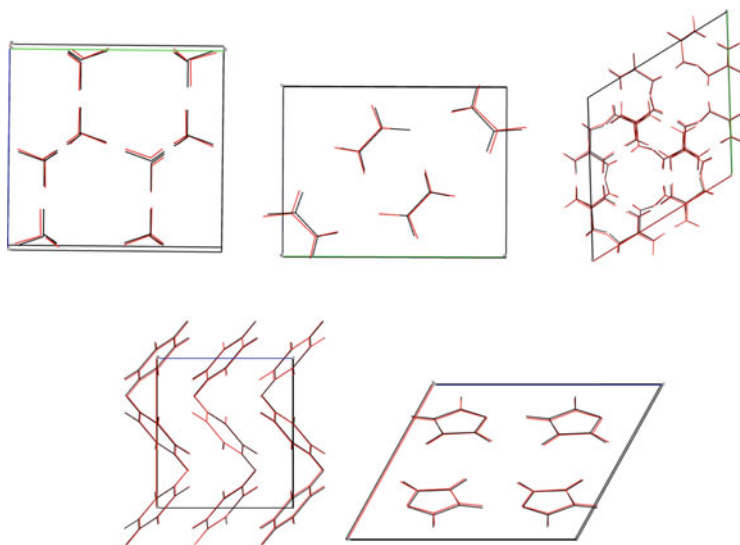
Fig. 8 Overlays of the experimental (*black*) and HMBI MP2 structures (*red*) for (clockwise from *top left*) ice Ih, formamide, acetamide, imidazole, and benzene. The experimental unit cell boundaries are drawn. Adapted with permission from [73]. Copyright 2012, American Institute of Physics
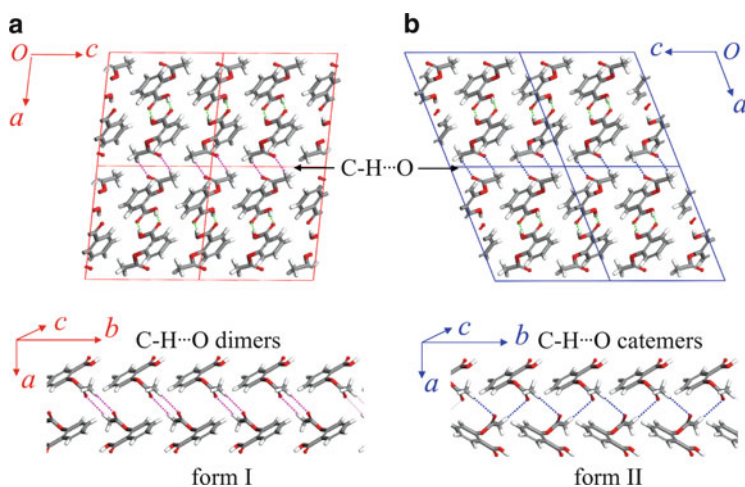

Fig. 9 Aspirin forms I and II exhibit very similar crystal packing. The key difference lies in whether the interlayer hydrogen bonding occurs as dimers (form I) or catemers (form II). Reprinted with permission from [132]. Copyright 2012 American Chemical Society

aug-cc-pVTZ basis sets. In all cases, the energy differences between the two polymorphs were only a few tenths of a kJ/mol, confirming the expected near-degeneracy of the two forms (Table 3). Interestingly, both MP2 and SCS(MI)-MP2

Table 3 Lattice energies and relative energy differences for the two polymorphs of aspirin (in kJ/mol). The two polymorphs are virtually degenerate across a range of model chemistries

	MP2[a]	SCS(MI) MP2[a]	MP2C[b]
aug-cc-pVDZ			
Form I	113.7	132.5	–
Form II	113.5	132.3	–
$\Delta E_{I \rightarrow II}$[c]	0.2	0.1	–
aug-cc-pVTZ			
Form I	132.1	135.6	116.1
Form II	132.0	135.5	116.3
$\Delta E_{I \rightarrow II}$[c]	0.1	0.0	−0.1

[a]Wen and Beran [132]
[b]Huang et al. [103]
[c]The apparent discrepancies between the lattice energies and the ΔE values arise from rounding

appear to overestimate significantly the ~115 kJ/mol experimental lattice energy [128, 133], predicting aug-cc-pVTZ values of 132 and 136 kJ/mol, respectively. In contrast, MP2C predicts a lattice energy of 116 kJ/mol, in much better agreement with the experimental value [103].

In addition to predicting the virtual degeneracy of the two aspirin polymorphs, physical insight into the nature of the polymorphism was obtained by decomposing the relative energies of the two polymorphs according to the different contributions in the HMBI many-body expansion. In particular, we found that while the intramolecular interactions favor form I, the intermolecular interactions favor form II (Fig. 10). The key differences arise from the nature of the interlayer hydrogen bonding. In form I, the acetyl group hydrogen bonds to only one adjacent aspirin molecule, which allows the acetyl group to adopt a slightly more favorable intramolecular conformation. In contrast, the catemeric hydrogen bonding in form II forces the acetyl group to adopt a conformation that is slightly less favorable, but in doing so it forms much better hydrogen bonds and achieves hydrogen bond cooperativity through the extended hydrogen-bond networks. It turns out that the energy differences in each case amount to about 1.5 kJ/mol and cancel one another almost perfectly, making the two polymorphs virtually and "accidentally" degenerate.

From a theoretical perspective, one other key feature emerged from this study: the strong similarity between the crystal packing in both polymorphs leads to excellent cancellation of errors in predicting the relative polymorph energetics. Unfortunately, such thorough error cancellation is probably the exception rather than the rule, as demonstrated by the case of oxalyl dihydrazide which is discussed in the next section.

3.3 Oxalyl Dihydrazide Polymorphism

Oxalyl dihydrazide provides another interesting example of molecular crystal polymorphism. It exhibits five experimentally known polymorphs, denoted $\alpha - \epsilon$, that differ in their degree of intramolecular and intermolecular hydrogen bonding
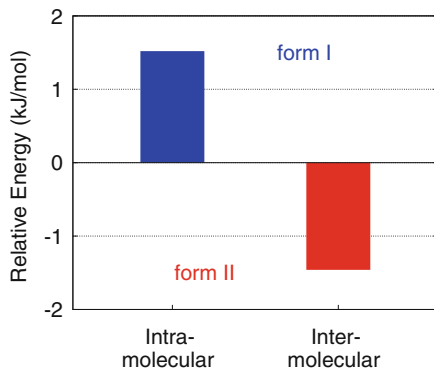
Fig. 10 Aspirin provides a classic example of conformational polymorphism: form I adopts a slightly more favorable intramolecular conformation, while form II exhibits stronger intermolecular interactions. The two effects have nearly identical energies, and the two polymorphs are virtually degenerate. Reprinted with permission from [132]. Copyright 2012 American Chemical Society

(Fig. 11). The relative stabilities of the five polymorphs is not known experimentally, but it is believed that the trend follows $\alpha, \delta, \epsilon < \gamma < \beta$ [134]. That is, the β form is the least stable, followed by the γ form. The other three are more stable, though their energetic ordering is unknown.

Earlier force field and DFT calculations had significant trouble reproducing these qualitative trends. The relative polymorph ordering varied widely with the choice of density functional; see, for example, [135]. Most of those functionals lacked van der Waals dispersion corrections, however. The empirical dispersion-corrected D-PW91 functional, which has been very successful in the blind tests of crystal structure prediction and elsewhere [12, 15, 87, 136–138], does however obtain a plausible ordering for the polymorphs ($\alpha < \epsilon < \delta < \gamma < \beta$), but the overall energy range of the polymorphs is ~15 kJ/mol, which is somewhat larger than the <10 kJ/mol typically found for experimentally observable polymorphs. On the other hand, a different dispersion corrected function, B3LYP-D* [118], gives a very different ordering that is inconsistent with experiment [70]. In other words, obtaining reasonable results for this system is not simply a matter of including dispersion. Rather, the energetics are very sensitive to the specific treatment of the electronic structure and the balance between intermolecular and intramolecular interactions. Note that vibrational zero-point energy is also important in oxalyl dihydrazide, and it is included in the results shown here. Finite-temperature effects may also be significant, but no attempt has been made to estimate them.

Obtaining plausible predictions for the relative polymorph energies proved challenging, even for correlated wavefunction methods (Fig. 12) [70]. In small basis sets, MP2 predicts the α form to be the *least* stable. However, increasing the basis set size toward the CBS limit dramatically reorders the polymorphs, preferentially stabilizing the α form. The slow basis set convergence in oxalyl
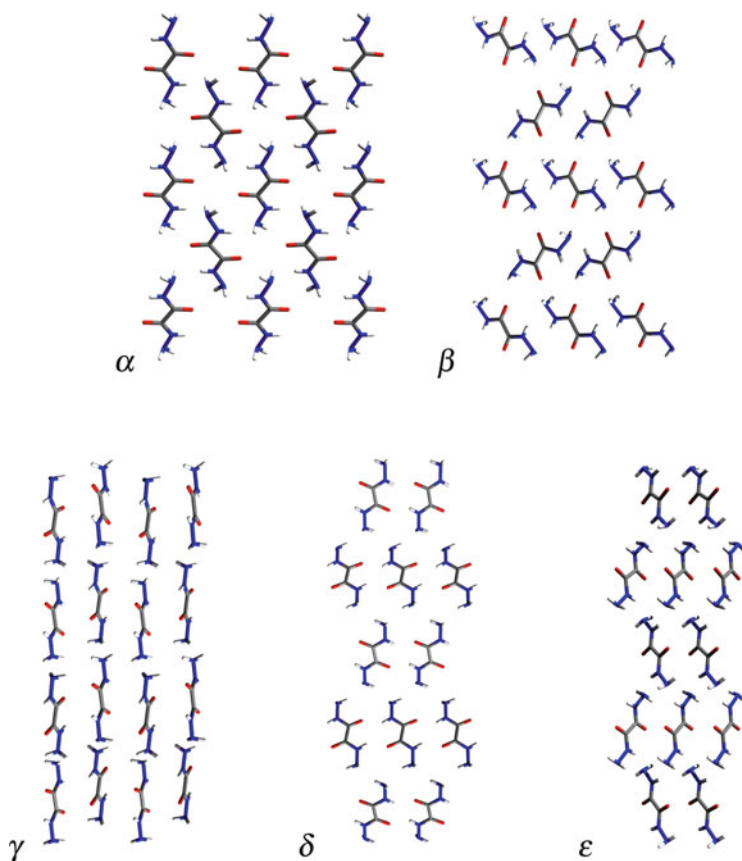
Fig. 11 The five experimentally known polymorphs of oxalyl dihydrazide

dihydrazide provides a sharp contrast to aspirin, where the relative energy differ-
ence between the two polymorphs was insensitive to the basis set.

The basis set sensitivity in oxalyl dihydrazide arises from BSSE. The α poly-
morph exhibits only intermolecular hydrogen bonds, while the other four forms
contain mixtures of intermolecular and intramolecular hydrogen bonds. Applying a
standard counterpoise correction to each QM dimer calculation helps correct the
intermolecular BSSE, but it is harder to correct intramolecular BSSE. Therefore,
the α polymorph, which has much less intramolecular BSSE, is destabilized relative
to the other four forms in small basis sets. As the basis set size increases, the
intramolecular BSSE decreases in the other four forms, and the balance between
intermolecular and intramolecular hydrogen bonding is restored.

Beyond the basis set effects, oxalyl dihydrazide exhibits some π-stacking type
van der Waals dispersion interactions, so it comes as no surprise that using MP2C
instead of MP2 also has a significant effect. The dispersion correction reorders the
β and γ polymorphs to the correct experimental ordering. It also stabilizes the α

Fig. 12 Relative polymorph energies for oxalyl dihydrazide as a function of the model chemistry. Basis set effects, improved MP2C two-body dispersion, and AIFF three-body dispersion are all important. Adapted with permission from [70]. Copyright 2012 American Chemical Society
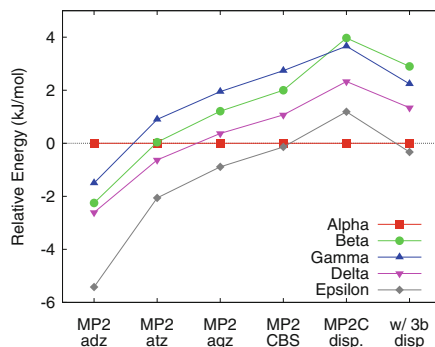
form substantially relative to the other polymorphs. The more empirical SCS(MI)-MP2 did not perform as well as MP2C in this system [70], perhaps due to the major differences in the optimal correlation energy scaling factors for intra- and intermolecular interactions.

Finally, Axilrod–Teller three-body dispersion effects prove important here. They destabilize the α form relative to the other four forms. The three-body dispersion contributions here are repulsive and die off with R^{-9}. Therefore, they are more repulsive in the dense α polymorph (1.76 g/cm^3) than in the four other, less dense polymorphs (1.59–1.66 g/cm^3). Without three-body dispersion, the calculations predict that the α polymorph is the most stable. However, including the three-body dispersion terms makes the ϵ polymorph slightly more stable than the α one, in contrast with the best dispersion-corrected DFT predictions. The predicted energetic preference for the ϵ polymorph over the α one is slight, and no experimental data currently exists to help resolve the issue. Nevertheless, it is interesting to note that the empirically dispersion-corrected DFT results in [70, 135] include only two-body dispersion and favor the α polymorph, just like the HMBI results without three-body dispersion. From this perspective it would be interesting to see what state-of-the-art many-body dispersion-corrected density functionals predict here.

4 Conclusions and Outlook

Fragment-based methods like HMBI provide a computationally viable means of achieving high-accuracy structures and lattice energies for chemically interesting molecular crystals. The systems examined here highlight the challenges inherent to modeling molecular crystals, where the subtle energetic competitions can require careful electronic structure treatments. It will not always be a priori obvious how elaborate an electronic structure treatment is needed for a given system. One can sometimes count on cancellation of errors, especially when the different crystal packing motifs are similar (like in aspirin), but much less error cancellation occurs in other cases (like oxalyl dihydrazide). Unfortunately, the reduced error-cancellation case is probably more typical for conformational polymorphs of flexible organic molecules, especially as the molecules become larger.

Periodic DFT methods can often provide acceptable accuracy for molecular crystal systems with reasonable computational costs. In many cases, however, one will wish to assess the reliability of those predictions or to obtain even more accurate predictions. From this perspective, one of the key strengths of fragment methods is their ability to improve systematically the quality of the results with respect to the underlying electronic structure treatments. The ability to demonstrate convergence of the predictions with respect to model chemistry is a powerful tool for making robust predictions. Moreover, for systems with large numbers of molecules in the unit cell, the inherently linear-scaling nature of the computationally dominant QM calculations in fragment methods will often make them cheaper than full periodic DFT calculations. Fragment methods also provide a natural decomposition scheme for understanding the nature of the energetic competitions in polymorphic crystals.

For evaluating the intermolecular interactions in molecular crystals, the dispersion-corrected MP2C method provides a useful balance between accuracy and efficiency, offering near coupled-cluster-quality results at MP2-like cost. In practice, MP2C (or its explicitly correlated variant, MP2C-F12), with an aug-cc-pVTZ basis, is practical for molecular crystals containing two to three dozen atoms per molecule and a handful of molecules in the asymmetric unit cell. In other words, these techniques are applicable to a number of interesting, small-molecule pharmaceuticals, organic semiconductors, and energetic materials (though many more such materials remain impractical for the moment!).

Thus, fragment methods like HMBI will likely play an important role in molecular crystal modeling for years to come. The next advances are likely to come from a couple of directions. First, further efficiency improvements will enable the application to larger systems. For instance, the MP2C timings described above indicate that the vast majority of the computational time is consumed on the MP2 calculations. A sizable fraction of that time comes from evaluating the long-range interactions, which are then largely discarded through the MP2C dispersion correction. New, more efficient strategies for achieving similar quality results can surely be developed. In addition, further improvements to the ab initio force field would enable one to treat fewer dimers quantum mechanically, thereby accelerating the calculations. This requires incorporating efficient treatments of the short-rate exchange interactions into the force field.

Second, one needs to consider finite-temperature entropies and free energies. This can be done through quasi-harmonic-type approximations or through dynamical free energy calculations. The latter is potentially more rigorous and capable of capturing anharmonic effects, but of course it is limited by the need for extensive computational sampling and the quality of the force fields that can be used to perform such sampling affordably. Quasiharmonic calculations are much less expensive computationally, but they involve their own severe approximations. Much more research is needed in this area to determine which techniques are useful under which circumstances. In other words, many theoretical opportunities remain in molecular crystal modeling to occupy researchers for quite some time!

References

1. Bauer J, Spanton S, Quick R, Quick J, Dziki W, Porter W, Morris J (2001) Ritonavir: an extraordinary example of conformational polymorphism. Pharm Res 18:859–866
2. Chemburkar SR, Bauer J, Deming K, Spiwek H, Patel K, Morris J, Henry R, Spanton S, Dziki W, Porter W, Quick J, Bauer P, Donaubauer J, Narayanan BA, Soldani M, Riley D, Mcfarland K (2000) Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development. Org Process Res Dev 4(5):413–417
3. Raw AS, Furness MS, Gill DS, Adams RC, Holcombe FO, Yu LX (2004) Regulatory considerations of pharmaceutical solid polymorphism in abbreviated new drug applications (ANDAs). Adv Drug Deliv Rev 56(3):397–414. doi:10.1016/j.addr.2003.10.011
4. Goldbeck G, Pidcock E, Groom C (2012) Solid form informatics for pharmaceuticals and agrochemicals: knowledge based substance development and risk assessment. http://www.ccdc.cam.ac.uk/Lists/ResourceFileList/Solid Form informatics .pdf. Accessed 28 June 2013
5. Bernstein J (2002) Polymorphism in molecular crystals. Clarendon, Oxford
6. Roth K (2005) Von Vollmilch bis Bitter, edelste Polymorphie. Chem Unserer Zeit 39:416–428
7. Politzer P, Murray JS (eds) (2003) Energetic materials: part 1. Decomposition, crystal, and molecular properties. Elsevier, Amsterdam
8. Politzer P, Murray JS (eds) (2003) Energetic materials: part 2. Detonation, combustion. Elsevier, Amsterdam
9. Haas S, Stassen AF, Schuck G, Pernstich KP, Gundlach DJ, Batlogg B, Berens U, Kirner HJ (2007) High charge-carrier mobility and low trap density in a rubrene derivative. Phys Rev B 76:115203
10. Karamertzanis PG, Kazantsev AV, Issa N, Welch GWA, Adjiman CS, Pantelides CC, Price SL (2009) Can the formation of pharmaceutical cocrystals be computationally predicted? 2 Crystal structure prediction. J Chem Theory Comput 5:1432–1448
11. Kazantsev AV, Karamertzanis PG, Adjiman CS, Pantelides CC, Price SL, Galek PTA, Day GM, Cruz-Cabeza AJ (2011) Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. Int J Pharm 418:168–178. doi:10.1016/j.ijpharm.2011.03.058
12. Kendrick J, Leusen FJJ, Neumann MA, van de Streek J (2011) Progress in crystal structure prediction. Chem Eur J 17(38):10736–10744. doi:10.1002/chem.201100689
13. Neumann MA (2008) Tailor-made force fields for crystal-structure prediction. J Phys Chem B 112(32):9810–9829. doi:10.1021/jp710575h
14. Neumann MA, Perrin MA (2005) Energy ranking of molecular crystals using density functional theory calculations and an empirical van der Waals correction. J Phys Chem B 109:15531–15541
15. Neumann MA, Leusen FJJ, Kendrick J (2008) A major advance in crystal structure prediction. Angew Chem Int Ed 47:2427–2430
16. Bardwell DA, Adjiman CS, Arnautova YA, Bartashevich E, Boerrigter SXM, Braun DE, Cruz-Cabeza AJ, Day GM, Della Valle RG, Desiraju GR, van Eijck BP, Facelli JC, Ferraro MB, Grillo D, Habgood M, Hofmann DWM, Hofmann F, Jose KVJ, Karamertzanis PG, Kazantsev AV, Kendrick J, Kuleshova LN, Leusen FJJ, Maleev AV, Misquitta AJ, Mohamed S, Needs RJ, Neumann MA, Nikylov D, Orendt AM, Pal R, Pantelides CC, Pickard CJ, Price LS, Price SL, Scheraga HA, van de Streek J, Thakur TS, Tiwari S, Venuti E, Zhitkov IK (2011) Towards

crystal structure prediction of complex organic compounds – a report on the fifth blind test. Acta Crystallogr B 67:535–551. doi:10.1107/S0108768111042868

17. Day GM, Cooper TG, Cruz-Cabeza AJ, Hejczyk KE, Ammon HL, Boerrigter SXM, Tan JS, Della Valle RG, Venuti E, Jose J, Gadre SR, Desiraju GR, Thakur TS, van Eijck BP, Facelli JC, Bazterra VE, Ferraro MB, Hofmann DWM, Neumann MA, Leusen FJJ, Kendrick J, Price SL, Misquitta AJ, Karamertzanis PG, Welch GWA, Scheraga HA, Arnautova YA, Schmidt MU, van de Streek J, Wolf AK, Schweizer B (2009) Significant progress in predicting the crystal structures of small organic molecules – a report on the fourth blind test. Acta Crystallogr B 65(Pt 2):107–125. doi:10.1107/S0108768109004066

18. Zhu Q, Oganov AR, Glass CW, Stokes HT (2012) Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. Acta Crystallogr B 68:215–226. doi:10.1107/S0108768112017466

19. Price SL (2008) Computational prediction of organic crystal structures and polymorphism. Int Rev Phys Chem 27(3):541–568. doi:10.1080/01442350802102387

20. Gavezzotti A, Filippini G (1995) Polymorphic forms of organic crystals at room conditions: thermodynamic and structural implications. J Am Chem Soc 117:12299–12305

21. Otero-de-la Roza A, Johnson ER (2012) A benchmark for non-covalent interactions in solids. J Chem Phys 137(5):054103. doi:10.1063/1.4738961

22. Raiteri P, Martonák R, Parrinello M (2005) Exploring polymorphism: the case of benzene. Angew Chem Int Ed 44(24):3769–3773. doi:10.1002/anie.200462760

23. Schnieders MJ, Baltrusaitis J, Shi Y, Chattree G, Zheng L, Yang W, Ren P (2012) The structure, thermodynamics and solubility of organic crystals from simulation with a polarizable force field. J Chem Theory Comput 8(5):1721–1736. doi:10.1021/ct300035u

24. Kazantsev AV, Karamertzanis PG, Adjiman CS, Pantelides CC (2011) Efficient handling of molecular flexibility in lattice energy minimization of organic crystals. J Chem Theory Comput 7:1998–2016

25. Price SL (2004) The computational prediction of pharmaceutical crystal structures and polymorphism. Adv Drug Deliv Rev 56(3):301–319. doi:10.1016/j.addr.2003.10.006

26. Price SL, Price LS (2011) Computational polymorph prediction. In: Storey R, Ymen I (eds) Solid state characterization of pharmaceuticals, 1st edn. Blackwell, London, pp 427–450

27. Price SL, Leslie M, Welch GWA, Habgood M, Price LS, Karamertzanis PG, Day GM (2010) Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. Phys Chem Chem Phys 12:8478–8490. doi:10.1039/c004055j

28. Day GM, Motherwell WDS, Ammon HL, Boerrigter SXM, Della Valle RG, Venuti E, Dzyabchenko A, Dunitz JD, Schweizer B, van Eijck BP, Erk P, Facelli JC, Bazterra VE, Ferraro MB, Hofmann DWM, Leusen FJJ, Liang C, Pantelides CC, Karamertzanis PG, Price SL, Lewis TC, Nowell H, Torrisi A, Scheraga HA, Arnautova YA, Schmidt MU, Verwer P (2005) A third blind test of crystal structure prediction. Acta Crystallogr B 61(Pt 5):511–527. doi:10.1107/S0108768105016563

29. Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Van der Waals density functional for general geometries. Phys Rev Lett 92:246401

30. DiStasio RA, von Lilienfeld OA, Tkatchenko A (2012) Collective many-body van der Waals interactions in molecular systems. Proc Natl Acad Sci U S A 109:14791–14795. doi:10.1073/pnas.1208121109

31. Grimme S (2011) Density functional theory with London dispersion corrections. WIREs Comput Mol Sci 1(2):211–228. doi:10.1002/wcms.30

32. Otero-de-la Roza A, Johnson ER (2013) Many-body dispersion interactions from the exchange-hole dipole moment model. J Chem Phys 138(5):054101. doi:10.1063/1.4789421

33. Thonhauser T, Cooper VR, Li S, Puzder A, Hyldgaard P, Langreth DC (2007) Van der Waals density functional: self-consistent potential and the nature of the van der Waals bond. Phys Rev B 76:125112

34. Tkatchenko A, Scheffler M (2009) Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. Phys Rev Lett 102(7):073005. doi:10.1103/PhysRevLett.102.073005

35. Vydrov OA, Van Voorhis T (2010) Nonlocal van der Waals density functional: the simpler the better. J Chem Phys 133(24):244103. doi:10.1063/1.3521275
36. Ayala PY, Kudin KN, Scuseria GE (2001) Atomic orbital Laplacetransformed second-order Møller–Plesset perturbation theory for periodic systems. J Chem Phys 115:9698–9707
37. Hirata S, Iwata S (1998) Analytical energy gradients in second-order Moller–Plesset perturbation theory for extended systems. J Chem Phys 109(11):4147–4155
38. Hirata S, Shimazaki T (2009) Fast second-order many-body perturbation method for extended systems. Phys Rev B 80(8):1–7. doi:10.1103/PhysRevB.80.085118
39. Izmaylov AF, Scuseria GE (2009) Resolution of the identity atomic orbital Laplace transformed second-order Møller–Plesset theory for nonconducting periodic systems. Phys Chem Chem Phys 10:3421–3429
40. Marsman M, Grueneis A, Paier J, Kresse G (2009) Second-order Møller–Plesset perturbation theory applied to extended systems. I. Within the projector-augmented-wave formalism using a plane wave basis set. J Chem Phys 130:184103
41. Maschio L, Usvyat D, Manby FR, Casassa S, Pisani C, Schutz M (2007) Fast local-MP2 method with density-fitting for crystals. I. Theory and algorithms. Phys Rev B 76:075101
42. Ohnishi YY, Hirata S (2010) Logarithm second-order many-body perturbation method for extended systems. J Chem Phys 133(3):034106. doi:10.1063/1.3455717
43. Pisani C, Maschio L, Casassa S, Halo M, Schutz M, Usvyat D (2008) Periodic local MP2 method for the study of electronic correlation in crystals: theory and preliminary applications. J Comput Chem 29:2113–2124
44. Shiozaki T, Hirata S (2010) Communications: explicitly correlated secondorder Møller–Plesset perturbation method for extended systems. J Chem Phys 132(15):151101. doi:10.1063/1.3396079
45. Suhai S (1983) Quasiparticle energy-band structures in semiconducting polymers: correlation effects on the band gap in polyacetylene. Phys Rev B 27:3506–3518
46. Sun JQ, Bartlett RJ (1996) Second-order many-body perturbation-theory calculations in extended systems. J Chem Phys 104:8553–8565
47. Usvyat D, Maschio L, Manby FR, Casassa S, Pisani C, Schutz M (2007) Fast local-MP2 method with density-fitting for crystals. II. Test calculations and applications to the carbon dioxide crystal. Phys Rev B 76:075102
48. Förner W, Knab R, Čížek J, Ladik J (1997) Numerical application of the coupled cluster theory with localized orbitals to polymers. IV. Band structure corrections in model systems and polyacetylene. J Chem Phys 106:10248–10264
49. Hirata S, Podeszwa R, Tobita M, Bartlett RJ (2004) Coupled-cluster singles and doubles for extended systems. J Chem Phys 120:2581–2592
50. Reinhardt P (2000) Dressed coupled-electron-pair-approximation methods for periodic systems. Theor Chem Acc 104:426–438
51. Yu M, Kalvoda S, Dolg M (1997) An incremental approach for correlation contributions to the structural and cohesive properties of polymers. Coupled-cluster study of trans-polyacetylene. Chem Phys 224:121–131
52. Fedorov DG, Kitaura K (2007) Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. J Phys Chem A 111:6904–6914
53. Kitaura K, Ikeo E, Asada T, Nakano T, Uebayasi M (1999) Fragment molecular orbital method: an approximate computational method for large molecules. Chem Phys Lett 313:701–706
54. Yang W (1991) Direct calculation of electron density in density functional theory. Phys Rev Lett 66(11):1438–1441
55. Paulus B (2006) The method of increments - a wavefunction-based ab initio correlation method for solids. Phys Rep 428(1):1–52. doi:10.1016/j.physrep.2006.01.003
56. Stoll H (1992) Correlation energy of diamond. Phys Rev B 46:6700–6704
57. Gordon MS, Fedorov DG, Pruitt SR, Slipchenko L (2012) Fragmentation methods: a route to accurate calculations on large systems. Chem Rev 112:632–672. doi:10.1021/cr200093j

58. Wen S, Nanda K, Huang Y, Beran GJO (2012) Practical quantum mechanics-based fragment methods for predicting molecular crystal properties. Phys Chem Chem Phys 14:7578–7590. doi:10.1039/c2cp23949c

59. Beran GJO, Hirata S (2012) Fragment and localized orbital methods in electronic structure theory. Phys Chem Chem Phys 14:7559–7561. doi:10.121/cg300358n

60. Mayhall NJ, Raghavachari K (2012) Many-overlapping-body (MOB) expansion: a generalized many body expansion for nondisjoint monomers in molecular fragmentation calculations of covalent molecules. J Chem Theory Comput 8(8):2669–2675. doi:10.1021/ct300366e

61. Richard RM, Herbert JM (2012) A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory. J Chem Phys 137(6):064113. doi:10.1063/1.4742816

62. Dahlke EE, Truhlar DG (2006) Assessment of the pairwise additive approximation and evaluation of many-body terms for water clusters. J Phys Chem B 3:10595–10601

63. Dahlke EE, Truhlar DG (2007) Electrostatically embedded many-body correlation energy, with applications to the calculation of accurate second-order Møller–Plesset perturbation theory energies for large water clusters. J Chem Theory Comput 3:1342–1348

64. Dahlke EE, Truhlar DG (2007) Electrostatically embedded many-body expansion for large systems, with applications to water clusters. J Chem Theory Comput 3:46–53

65. Hirata S (2008) Fast electron-correlation methods for molecular crystals: an application to the alpha, beta(1), and beta(2) modifications of solid formic acid. J Chem Phys 129(20):204104. doi:10.1063/1.3021077

66. Sode O, Keceli M, Hirata S, Yagi K (2009) Coupled-cluster and many-body perturbation study of energies, structures, and phonon dispersions of solid hydrogen fluoride. Int J Quantum Chem 109:1928–1939

67. Sode O, Keceli M, Hirata S, Yagi K (2009) Coupled-cluster and many-body perturbation study of energies, structures, and phonon dispersions of solid hydrogen fluoride phonon dispersions. Int J Quantum Chem 109:1928–1939. doi:10.1002/qua

68. Manby FR, Stella M, Goodpaster JD, Miller TF (2012) A simple, exact density-functional-theory embedding scheme. J Chem Theory Comput 8(8):2564–2568. doi:10.1021/ct300544e

69. Reilly AM, Tkatchenko A (2013) Seamless and accurate modeling of organic molecular materials. J Phys Chem Lett 4:1028–1033

70. Wen S, Beran GJO (2012) Crystal polymorphism in oxalyl dihydrazide: is empirical DFT-D accurate enough? J Chem Theory Comput 8:2698–2705. doi:10.1021/ct300484h

71. Neill DPO, Allan NL, Manby FR (2010) Ab initio Monte Carlo simulations of liquid water. In: Manby F (ed) Accurate quantum chemistry in the condensed phase. CRC, Boca Raton, pp 163–193

72. Subotnik JE, Sodt A, Head-Gordon M (2008) The limits of local correlation theory: electronic delocalization and chemically smooth potential energy surfaces. J Chem Phys 128:034103

73. Nanda K, Beran GJO (2012) Improved prediction of organic molecular crystal geometries from MP2-level fragment QM/MM calculations. J Chem Phys 137:174106. doi:10.1063/1.4764063

74. Beran GJO (2009) Approximating quantum many-body intermolecular interactions in molecular clusters using classical polarizable force fields. J Chem Phys 130:164115. doi:10.1063/1.3121323

75. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T (2010) Current status of the AMOEBA polarizable force field. J Phys Chem B 114 (8):2549–2564. doi:10.1021/jp910674d

76. Ren P, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. J Phys Chem B 107:5933–5947

77. Sebetci A, Beran GJO (2010) Spatially homogeneous QM/MM for systems of interacting molecules with on-the-fly ab initio force-field parameterization. J Chem Theory Comput 6:155–167. doi:10.1021/ct900545v

78. Wen S, Beran GJO (2011) Accurate molecular crystal lattice energies from a fragment QM/MM approach with on-the-fly ab initio force-field parameterization. J Chem Theory Comput 7:3733–3742. doi:10.1021/ct200541h

79. Stone AJ (1981) Distributed multipole analysis, or how to describe a molecular charge distribution. Chem Phys Lett 83:233–239

80. Stone AJ (2005) Distributed multipole analysis: stability for large basis sets. J Chem Theory Comput 1:1128–1132

81. Stone AJ, Alderton M (1985) Distributed multipole analysis – methods and applications. Mol Phys 56:1047–1064

82. Misquitta AJ, Stone AJ (2008) Accurate induction energies for small organic molecules: 1. Theory. J Chem Theory Comput 4:7–18

83. Misquitta AJ, Stone AJ, Price SL (2008) Accurate induction energies for small organic molecules: 2. Development and testing of distributed polarizability models against SAPT (DFT) energies. J Chem Theory Comput 4:19–32

84. Misquitta AJ, Stone AJ (2008) Dispersion energies for small organic molecules: first row atoms. Mol Phys 106(12):1631–1643. doi:10.1080/00268970802258617

85. Stone AJ (2002) The theory of intermolecular forces. Clarendon, Oxford

86. Stone AJ, Misquitta AJ (2007) Atom-atom potentials. Int Rev Phys Chem 26:193–222

87. Neumann MA, Perrin MA (2009) Can crystal structure prediction guide experimentalists to a new polymorph of paracetamol? CrystEngComm 11(11):2475. doi:10.1039/b909819d

88. Leslie M (2008) DL MULTI – a molecular dynamics program to use distributed multipole electrostatic models to simulate the dynamics of organic crystals. Mol Phys 106 (12):1567–1578. doi:10.1080/00268970802175308

89. Grafova L, Pitonak M, Rezac J, Hobza P (2010) Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set. J Chem Theory Comput 6(8):2365–2376. doi:10.1021/ct1002253

90. Rezac J, Riley KE, Hobza P (2011) Extensions of the S66 data set: more accurate interaction energies and angular-displaced nonequilibrium geometries. J Chem Theory Comput 7:3466–3470. doi:10.1021/ct200523a

91. Riley KE, Pitonak M, Jurecka P, Hobza P (2010) Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. Chem Rev 110(9):5023–5063. doi:10.1021/cr1000173

92. Chalasinski G, Szczesniak MM (1988) On the connection between the supermolecular Møller–Plesset treatment of the interaction energy and the perturbation theory of intermolecular forces. Mol Phys 63:205–224

93. Cybulski SM, Chalasinski G, Moszynski R (1990) On decomposition of second-order Møller–Plesset supermolecular interaction energy and basis set effects. J Chem Phys 92:4357–4363

94. Cybulski SM, Lytle ML (2007) The origin of deficiency of the supermolecule second-order Møller–Plesset approach for evaluating interaction energies. J Chem Phys 127:141102

95. Hesselmann A (2008) Improved supermolecular second order Møller–Plesset intermolecular interaction energies using time-dependent density functional response theory. J Chem Phys 128(14):144112

96. Pitonak M, Hesselmann A (2010) Accurate intermolecular interaction energies from a combination of MP2 and TDDFT response theory. J Chem Theory Comput 6(1):168–178. doi:10.1021/ct9005882

97. Tkatchenko A, Distasio RA, Head-Gordon M, Scheffler M (2009) Dispersion-corrected Møller–Plesset second-order perturbation theory. J Chem Phys 131:094106

98. Granatier J, Pitonak M, Hobza P (2012) Accuracy of several wave function and density functional theory methods for description of noncovalent interaction of saturated and unsaturated hydrocarbon dimers. J Chem Theory Comput 8:2282–2292

99. Hesselmann A, Korona T (2011) On the accuracy of DFT-SAPT, MP2, SCSMP2, MP2C, and DFT + Disp methods for the interaction energies of endohedral complexes of the C(60) fullerene with a rare gas atom. Phys Chem Chem Phys 13(2):732–743. doi:10.1039/c0cp00968g

100. Hohenstein EG, Jaeger HM, Carrell EJ, Tschumper GS, Sherrill CD (2011) Accurate interaction energies for problematic dispersion-bound complexes: homogeneous dimers of NCCN, P2, and PCCP. J Chem Theory Comput 7(9):2842–2851. doi:10.1021/ct200374m
101. Jenness GR, Karalti O, Al-Saidi WA, Jordan KD (2011) Evaluation of theoretical approaches for describing the interaction of water with linear acenes. J Phys Chem A 115:5955–5964
102. Karalti O, Alfe D, Gillan MJ, Jordan KD (2012) Adsorption of a water molecule on the MgO (100) surface as described by cluster and slab models. Phys Chem Chem Phys 14(21):7846–7853. doi:10.1039/c2cp00015f
103. Huang Y, Shao Y, Beran GJO (2013) Accelerating MP2C dispersion corrections for dimers and molecular crystals. J Chem Phys 138:224112. doi:10.1063/1.4809981
104. Distasio RA, Head-Gordon M (2007) Optimized spin-component-scaled second-order Møller–Plesset perturbation theory for intermolecular interaction energies. Mol Phys 105:1073–1083
105. Gerenkamp M, Grimme S (2004) Spin-component scaled second-order Møller–Plesset perturbation theory for the calculation of molecular geometries and harmonic vibrational frequencies. Chem Phys Lett 392:229–235
106. Hill JG, Platts JA (2007) Spin-component scaling methods for weak and stacking interactions. J Chem Theory Comput 3:80
107. Marchetti O, Werner HJ (2009) Accurate calculations of intermolecular interaction energies using explicitly correlated coupled cluster wave functions and a dispersion-weighted MP2 method. J Phys Chem A 113:11580
108. Eshuis H, Bates JE, Furche F (2012) Electron correlation methods based on the random phase approximation. Theor Chem Acc 131(1):1084. doi:10.1007/s00214-011-1084-8
109. Li Y, Lu D, Nguyen HV, Galli G (2010) van der Waals interactions in molecular assemblies from first-principles calculations. J Phys Chem A 114:1944–1952
110. Lu D, Li Y, Rocca D, Galli G (2009) Ab initio calculation of van der Waals bonded molecular crystals. Phys Rev Lett 102:206411
111. Ren X, Tkatchenko A, Rinke P, Scheffler M (2011) Beyond the random-phase approximation for the electron correlation energy: the importance of single excitations. Phys Rev Lett 106 (15):153003. doi:10.1103/PhysRevLett.106.153003
112. Williams HL, Mas EM, Szalewicz K, Jeziorski B (1995) On the effectiveness of monomer-, dimer-, and bond-centered basis functions in calculations of intermolecular interaction energies. J Chem Phys 103(17):7374–7391. doi:10.1063/1.470309
113. Hättig C, Klopper W, Köhn A, Tew DP (2011) Explicitly correlated electrons in molecules. Chem Rev 112:4–74. doi:10.1021/cr200168z
114. Kong L, Bischoff FA, Valeev EF (2011) Explicitly correlated R12/F12 methods for electronic structure. Chem Rev 112:75–107. doi:10.1021/cr200204r
115. Beran GJO, Nanda K (2010) Predicting organic crystal lattice energies with chemical accuracy. J Phys Chem Lett 1:3480–3487. doi:10.1021/jz101383z
116. Sadlej AJ (1988) Medium-size polarized basis sets for high-level correlated calculations of molecular electronic properties. Collect Czech Chem Commun 53:1995–2016
117. Sadlej AJ (1991) Medium-size polarized basis sets for high-level correlated calculations of molecular electronic properties II. Second-row atoms Si–Cl. Theor Chim Acta 79:123–140
118. Civalleri B, Zicovich-Wilson CM, Valenzano L, Ugliengo P (2008) B3LYP augmented with an empirical dispersion term (B3LYP-D*) as applied to molecular crystals. CrystEngComm 10:405–410. doi:10.1039/b715018k
119. Wheatley PJ (1964) The crystal and molecular structure of aspirin. J Chem Soc 6036–6048 doi: 10.1039/JR9640006036
120. Ouvrard C, Price SL (2004) Toward crystal structure prediction for conformationally flexible molecules: the headaches illustrated by aspirin. Cryst Growth Des 4(6):1119–1127. doi:10.1021/cg049922u
121. Vishweshwar P, McMahon JA, Oliveira M, Peterson ML, Zaworotko MJ (2005) The predictably elusive form II of aspirin. J Am Chem Soc 127(48):16802–16803. doi:10.1021/ja056455b
122. Bond AD, Boese R, Desiraju GR (2007) On the polymorphism of aspirin. Angew Chem Int Ed 46(4):615–617. doi:10.1002/anie.200602378

123. Bond AD, Boese R, Desiraju GR (2007) On the polymorphism of aspirin: crystalline aspirin as intergrowths of two "polymorphic" domains. Angew Chem Int Ed 46(4):618–622. doi:10.1002/anie.200603373

124. Bauer JD, Haussuhl E, Winkler B, Arbeck D, Milman V, Robertson S (2010) Elastic properties, thermal expansion, and polymorphism of acetylsalicylic acid. Cryst Growth Des 10(7):3132–3140. doi:10.1021/cg100241c

125. Bond AD, Solanko KA, Parsons S, Redder S, Boese R (2011) Single crystals of aspirin form II: crystallisation and stability. CrystEngComm 13(2):399. doi:10.1039/c0ce00588f

126. Chan EJ, Welberry TR, Heerdege AP, Goossens DJ (2010) Diffuse scattering study of aspirin forms (I) and (II). Acta Crystallogr B 66:696–707

127. Li T (2007) Understanding the polymorphism of aspirin with electronic calculations. J Pharm Sci 96(4):755–760. doi:10.1002/jps

128. Li T, Feng S (2006) Empirically augmented density functional theory for predicting lattice energies of aspirin, acetaminophen polymorphs, and ibuprofen homochiral and racemic crystals. Pharm Res 23(10):2326–2332

129. Copley RCB, Barnett SA, Karamertzanis PG, Harris KDM, Kariuki BM, Xu M, Nickels EA, Lancaster RW, Price SL (2008) Predictable disorder versus polymorphism in the rationalization of structural diversity: a multidisciplinary study of eniluracil. Cryst Growth Des 8 (9):3474–3481. doi:10.1021/cg800517h

130. Torrisi A, Leech CK, Shankland K, David WIF, Ibberson RM, Benet-Buchholz J, Boese R, Leslie M, Catlow CRA, Price SL (2008) Solid phases of cyclopentane: combined experimental and simulation study. J Phys Chem B 112(12):3746–3758. doi:10.1021/jp710017y

131. Winkel K, Hage W, Loerting T, Price SL, Mayer E (2007) Carbonic acid: from polyamorphism to polymorphism. J Am Chem Soc 129(45):13863–13871. doi:10.1021/ja073594f

132. Wen S, Beran GJO (2012) Accidental degeneracy in crystalline aspirin: new insights from high-level ab initio calculations. Cryst Growth Des 12:2169–2172. doi:10.121/cg300358n

133. Perlovich GL, Kurkov SV, Kinchin AN, Bauer-Brandl A (2004) Solvation and hydration characteristics of ibuprofen and acetylsalicylic acid. AAPS PharmSci 6(1):22–30. doi:10.1208/ps060103

134. Ahn S, Guo F, Kariuki BM, Harris KDM (2006) Abundant polymorphism in a system with multiple hydrogen-bonding opportunities: oxalyl dihydrazide. J Am Chem Soc 128 (26):8441–8452. doi:10.1021/ja0573155

135. Karamertzanis PG, Day GM, Welch GWA, Kendrick J, Leusen FJJ, Neumann MA, Price SL (2008) Modeling the interplay of inter- and intramolecular hydrogen bonding in conformational polymorphs. J Chem Phys 128(24):244708. doi:10.1063/1.2937446

136. Perrin MA, Neumann MA, Elmaleh H, Zaske L (2009) Crystal structure determination of the elusive paracetamol form III. Chem Commun 22:3181–3183. doi:10.1039/b822882e

137. van de Streek J, Neumann MA (2010) Validation of experimental molecular crystal structures with dispersion-corrected density functional theory calculations. Acta Crystallogr B 66 (Pt 5):544–558. doi:10.1107/S0108768110031873

138. van de Streek J, Neumann MA (2011) Crystal-structure prediction of pyridine with four independent molecules. CrystEngComm 13(23):7135. doi:10.1039/c1ce05881a

139. Jurečka P, Šponer J, Černỳ J, Hobza P (2006) Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. Phys Chem Chem Phys 8:1985–1993

140. Chisholm JA, Motherwell WDS (2005) COMPACK: a program for identifying crystal structure similarity using distances. J Appl Crystall 38(1):228–231. doi:10.1107/S0021889804027074