

Efficient Handling of Molecular Flexibility in Lattice Energy Minimization of Organic Crystals

A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, and C. C. Pantelides*

Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom

ABSTRACT: This paper presents a novel algorithm, CrystalOptimizer, for the minimization of the lattice energy of crystals formed by flexible molecules. The algorithm employs isolated-molecule quantum mechanical (QM) calculations of the intramolecular energy and conformation-dependent atomic multipoles in the course of the lattice energy minimization. The algorithm eliminates the need to perform QM calculations at each iteration of the minimization by using Local Approximate Models (LAMs), with a minimal impact on accuracy. Additional computational efficiencies are achieved by storing QM-derived components of the lattice energy model in a database and reusing them in subsequent calculations whenever possible. This makes the approach particularly well suited to applications that involve a sequence of lattice energy evaluations, such as crystal structure prediction. The algorithm is capable of handling efficiently complex systems with considerable conformational flexibility. The paper presents examples of the algorithm's application ranging from single-component crystals to cocrystals and salts of flexible molecules with tens of intramolecular degrees of freedom whose optimal values are determined by the interplay of conformational strain and packing forces. For any given molecule, the degree of flexibility to be considered can vary from a few torsional angles to relaxation of the entire set of torsion angles, bond angles, and bond lengths present in the molecule.

1. INTRODUCTION

The knowledge of the three-dimensional atomic structure of a crystal is the basis toward understanding and predicting the physical properties of the material (color, density, solubility, dissolution rate, etc.).¹ Hence, the development of computational algorithms to predict the structure and the thermodynamic stability of single and multicomponent crystals is of significant practical importance. Prediction of thermodynamic stability requires the minimization of the Gibbs free energy with respect to the unit cell dimensions and the positions of all atoms in the unit cell:

$$\min G = \min(U + PV - TS) \quad (1)$$

where U is the internal energy (which includes the zero-point energy contributions²), V the volume, and S the entropy. The PV term only becomes significant at very high pressures (typically above 1 GPa) and is usually neglected. Furthermore, the thermal, entropic, and zero-point contributions to the Gibbs free energy of flexible molecules cannot, at present, be readily and accurately computed. There is also emerging evidence that most minima on the free energy surface are also minima on the lattice energy surface.³ Consequently, most of the existing computational approaches for crystal structure prediction focus on the minimization of lattice energy.⁴ The lattice energy, E^{latt} , refers to the internal energy, U , at 0 K and 0 Pa (ignoring the zero-point energy) and can be partitioned into the intramolecular and intermolecular contributions. The intramolecular energy, ΔE^{intra} , is the energy required to deform the molecular conformation from its gas-phase geometry. The intermolecular contributions, U^{inter} , consist primarily of the repulsion–dispersion and electrostatic interactions, although intermolecular induction models are also currently being developed.⁵

The assumption underpinning crystal structure prediction is that crystal structures that occur in nature correspond to low-lying minima in the crystal energy landscape, which is usually approximated as the lattice energy. Crystal structure prediction methodologies, such as the ones that have been used in the series of blind tests organized by the Cambridge Crystallographic Data Centre (e.g., Day et al.⁴), generally consist of a structure generation step, in which many possible crystal geometries are constructed, and a refinement step, in which the most promising structures are optimized further with more elaborate models. In the most recent approaches, hundreds of thousands of structures are generated, while only hundreds to thousands of distinct structures are retained in the refinement stage. As many hypothetical crystals have very similar lattice energies,⁶ their reliable ranking requires an accurate representation of all components of the lattice energy. A prerequisite for crystal structure prediction is therefore the availability of reliable local lattice energy minimization methodologies that have a reasonable computational cost. Such an approach, the CrystalOptimizer algorithm, is presented in this paper.

Ideally, the lattice energy of a crystal structure would be evaluated by modeling the entire crystal quantum mechanically. Significant progress toward this goal has been made through the development of the GRACE method,⁷ which combines full crystal QM calculations with an empirical dispersion correction. This, however, remains computationally expensive for many molecules of interest. Instead, an accurate intermolecular potential can be obtained by modeling the electrostatic interactions with distributed multipoles^{8,9} derived directly from the isolated-molecule charge density.^{10,11} In order to avoid unphysical

Received: October 18, 2010

Published: May 09, 2011

distortions of the molecular geometry in the crystal,¹² this anisotropic intermolecular energy model needs to be coupled with an accurate and well-balanced representation of the intramolecular energy.¹³ Electronic structure calculations can generally provide the accuracy required for modeling the deformations of the molecular structure within the crystal. However, the use of quantum mechanics (QM) in the computation of the lattice energy is expensive especially for large and flexible molecules, as the intramolecular energy and atomic multipoles need to be recalculated after any conformational change. Albeit computationally demanding, this approach has been successfully embedded in several lattice energy minimization methodologies described below.

One of the first accurate algorithms to take into account molecular flexibility in lattice energy minimization is UPACK.^{2,14–16} In this approach, the intramolecular energy is calculated using a quadratic approximation constructed from the results of an *ab initio* molecular geometry optimization. The intermolecular potential is fitted to high-level quantum mechanical calculations of alkanes, alcohols, and ethers¹⁷ and involves terms for atomic multipole moments, dipole polarizabilities, and repulsion–dispersion contributions. In order to model the conformational dependence of the electrostatic model, the atomic multipole moments are defined in terms of their local-axis system and rotated with the local environment. Following a significant change in conformation during lattice energy minimization, the intramolecular potential and the electrostatic model are recalculated to maintain accuracy. A feature of this approach is that the use of Cartesian coordinates for the representation of the molecular structure during lattice energy minimization forces the user either to neglect flexibility altogether (rigid-body approach) or to account for full flexibility (atomistic representation). In the latter case, the computational cost becomes prohibitive for any molecule of nontrivial size (more than 20 atoms).

An algorithm that allows the optimization of crystal structures with user-defined flexibility is DMAFlex.¹³ However, in order to calculate the lattice energy accurately, the method incorporates a full isolated-molecule quantum mechanical molecular geometry optimization and charge density calculation at every iteration, which results in very high computational cost. The computational burden is further compounded by the use of a gradient-free (simplex) minimization algorithm¹⁸ that limits the extent of molecular flexibility that can be practically handled to a small number (fewer than 10) of torsional angles.

The CrystalOptimizer algorithm presented in this paper is a local lattice energy minimization scheme for crystal structures containing flexible molecules. It is designed to reduce the computational cost associated with quantum mechanical evaluations without compromising accuracy. It is applicable to molecules of the size, complexity, and flexibility typically encountered in pharmaceutical development. The main novelty of the approach is the use of local approximate models (LAMs) to represent the intramolecular energy and conformationally dependent charge density. These models are practically as accurate as explicit QM calculations but carry a much smaller computational burden. They are presented in section 2. This is followed by the formulation of the lattice energy minimization problem (section 3.1) and a description of the structure of the CrystalOptimizer algorithm (section 3.2). The computational performance and the accuracy of the algorithm are critically assessed by its ability to reproduce the lattice geometry and conformational degrees of freedom for a set of experimentally determined

crystal structures (sections 4.4 and 4.5). Finally, the applicability of CrystalOptimizer to the refinement of hypothetical crystal structures in crystal structure prediction is discussed (section 4.6).

2. LOCAL APPROXIMATE MODELS (LAMs)

2.1. Molecular Flexibility. The geometry of a flexible molecule can be completely defined by its Z-matrix consisting of $3N - 6$ intramolecular degrees of freedom θ (torsion angles, bond angles, and bond lengths), where N is the number of atoms in the molecule. In molecular crystals, the intermolecular forces are significantly weaker than the energy of typical covalent interactions. Consequently, only a subset of the intramolecular degrees of freedom, θ , is expected to deviate significantly from their gas-phase values. These flexible degrees of freedom, θ^f (such as torsions around single bonds), are often sufficient in capturing the effect of molecular flexibility and thus need to be explicitly modeled during lattice energy minimization. However, as the values of θ^f for a given molecule change significantly, the rest of the intramolecular degrees of freedom adjust so as to minimize the intramolecular energy. Hence, these remaining, more rigid degrees of freedom, θ^r (such as torsions in aromatic ring systems, most bond angles and bond lengths), and the intramolecular energy, ΔE^{intra} , can be approximated as functions of the flexible degrees of freedom, θ^f , in the solution of a constrained isolated-molecule quantum mechanical geometry optimization:

$$\Delta E^{\text{intra}}(\theta^f) = \min_{\theta^r} [E^{\text{intra}}(\theta^r; \theta^f)] - E^{\text{vac}} \quad (2)$$

where E^{vac} is the global (or at least a local) minimum gas-phase molecular energy used as a correction and needs to be computed only once. Because the evaluation of ΔE^{intra} requires a minimization with respect to the “rigid” degrees of freedom, θ^r , these are not truly constant. The change θ^r in response to changes in the flexible degrees of freedom, θ^f , can be seen explicitly in the following equation:

$$\theta^r(\theta^f) = \arg \min_{\theta^r} [E^{\text{intra}}(\theta^r; \theta^f)] \quad (3)$$

where “arg min” denotes the value of the optimization variables in the solution of the minimization problem shown in eq 2. These changes, however small, will have an effect on the intramolecular energy, molecular geometry, and final structure reproduction (especially for large molecules) and therefore cannot be ignored by keeping θ^r fixed at some nominal values, such as those in the gas-phase conformational minimum. The validity of this partitioning of intramolecular degrees of freedom will be examined in sections 4.4 and 4.5 by comparing the results obtained with different sets of θ^f . The main benefit of considering “rigid” degrees of freedom is a reduction in computational cost with little loss of accuracy. However, the approach presented is also valid when the vector θ^r is empty, i.e., when all intramolecular variables are treated as flexible, as shown in section 4.4.

2.2. Intramolecular Energy LAM. The intramolecular energy for a given conformation in close proximity to a reference conformation, $\theta_{\text{ref}} \equiv (\theta_{\text{ref}}^f, \theta_{\text{ref}}^r)$, can be estimated using a local approximate model (LAM) based on a quadratic Taylor

expansion:

$$\begin{aligned}\Delta E^{\text{intra}}(\theta^f, \theta^r) = & \Delta E^{\text{intra}}(\theta_{\text{ref}}) + \left[\frac{\partial E^{\text{intra}}}{\partial \theta^f} \right]_{\theta_{\text{ref}}}^T (\theta^f - \theta_{\text{ref}}^f) \\ & + \left[\frac{\partial E^{\text{intra}}}{\partial \theta^r} \right]_{\theta_{\text{ref}}}^T (\theta^r - \theta_{\text{ref}}^r) \\ & + \frac{1}{2} (\theta^f - \theta_{\text{ref}}^f)^T \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^{f2}} \right]_{\theta_{\text{ref}}} (\theta^f - \theta_{\text{ref}}^f) \\ & + (\theta^r - \theta_{\text{ref}}^r)^T \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^f \partial \theta^r} \right]_{\theta_{\text{ref}}}^T (\theta^f - \theta_{\text{ref}}^f) \\ & + \frac{1}{2} (\theta^r - \theta_{\text{ref}}^r)^T \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^{r2}} \right]_{\theta_{\text{ref}}} (\theta^r - \theta_{\text{ref}}^r) \quad (4)\end{aligned}$$

The above equation is valid for *any* reference point $\theta_{\text{ref}} \equiv (\theta_{\text{ref}}^f, \theta_{\text{ref}}^r)$ in the coordinate space. Throughout this paper, the subscript “ref” denotes the point around which the Taylor expansion is constructed. The values of θ_{ref}^r at the reference point are obtained via an isolated-molecule quantum-mechanical constrained optimization (eq 2) for the fixed values $\theta^f = \theta_{\text{ref}}^f$. This calculation also yields the minimum molecular deformation energy, $\Delta E^{\text{intra}}(\theta_{\text{ref}})$, and the first- and second-order derivatives of ΔE^{intra} with respect to all intramolecular degrees of freedom, θ_{ref} , necessary to construct the LAM. Since θ_{ref}^r is obtained by minimizing ΔE^{intra} , it must satisfy the first-order optimality condition:

$$\left[\frac{\partial E^{\text{intra}}}{\partial \theta^r} \right]_{\theta_{\text{ref}}^r, \theta_{\text{ref}}^f} = 0 \quad (5)$$

We require that eq 5 also apply to the intramolecular energy LAM. Hence, if θ^f is changed by a small amount $\delta\theta^f$ from the reference value, then θ^r needs to change so that the intramolecular energy remains at a minimum. This can be enforced by ensuring that the first-order optimality conditions continue to be satisfied at $(\theta^r + \delta\theta^r, \theta^f + \delta\theta^f)$, i.e.:

$$\left[\frac{\partial E^{\text{intra}}}{\partial \theta^r} \right]_{\theta_{\text{ref}}^r + \delta\theta^r, \theta_{\text{ref}}^f + \delta\theta^f} = 0 \quad (6)$$

where $\delta\theta^r$ is the corresponding change in θ^r . Performing a first-order Taylor expansion, subtracting eq 5 from eq 6 and solving for $\delta\theta^r$ yields the approximate expression:

$$\delta\theta^r = - \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^{r2}} \right]_{\theta_{\text{ref}}}^{-1} \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^f \partial \theta^r} \right]_{\theta_{\text{ref}}}^T \delta\theta^f \quad (7)$$

which then allows us to approximate θ^r via an explicit linear function of θ^f :

$$\theta^r(\theta^f) = \theta_{\text{ref}}^r + \mathbf{A}(\theta_{\text{ref}})(\theta^f - \theta_{\text{ref}}^f) \quad (8)$$

where the matrix $\mathbf{A}(\theta_{\text{ref}})$ is defined as

$$\mathbf{A}(\theta_{\text{ref}}) \equiv \frac{\partial \theta^r}{\partial \theta^f} = - \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^{r2}} \right]_{\theta_{\text{ref}}}^{-1} \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^f \partial \theta^r} \right]_{\theta_{\text{ref}}}^T \quad (9)$$

By substituting eq 8 into eq 4 and also taking account of eq 5, an estimate for the intramolecular energy as a quadratic function solely of the flexible degrees of freedom is obtained:

$$\begin{aligned}\Delta E^{\text{intra}}(\theta^f) = & \Delta E^{\text{intra}}(\theta_{\text{ref}}^f) + \mathbf{b}(\theta_{\text{ref}})^T (\theta^f - \theta_{\text{ref}}^f) \\ & + \frac{1}{2} (\theta^f - \theta_{\text{ref}}^f)^T \mathbf{C}(\theta_{\text{ref}}) (\theta^f - \theta_{\text{ref}}^f) \quad (10)\end{aligned}$$

where the vector $\mathbf{b}(\theta_{\text{ref}})$ and matrix $\mathbf{C}(\theta_{\text{ref}})$ are defined as

$$\mathbf{b}(\theta_{\text{ref}}) \equiv \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^f} \right]_{\theta_{\text{ref}}} \quad (11)$$

$$\mathbf{C}(\theta_{\text{ref}}) \equiv \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^{f2}} \right]_{\theta_{\text{ref}}} - \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^f \partial \theta^r} \right]_{\theta_{\text{ref}}} \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^{r2}} \right]_{\theta_{\text{ref}}}^{-1} \left[\frac{\partial^2 E^{\text{intra}}}{\partial \theta^f \partial \theta^r} \right]_{\theta_{\text{ref}}}^T \quad (12)$$

Equations 8 and 10 allow the explicit and fast calculation of the values of θ^r and ΔE^{intra} for any given values of θ^f without performing new quantum mechanical calculations; the values obtained are accurate in the proximity of a reference point θ_{ref} . If a molecule is optimized atomistically, i.e., all intramolecular degrees of freedom are treated as flexible, then eq 10 reduces to the standard quadratic Taylor expansion in UPACK.¹⁶

2.3. Electrostatic Model LAM. For limited conformational changes, the conformational dependence of the intermolecular electrostatic model can be captured by rotating the multipole moments with their local environment.^{16,19} Once the distributed multipole moments (Ω) have been computed⁹ (up to the hexadecapole level) for a reference molecular conformation, θ_{ref} , each atom is assigned a local axis system using two directly connected atoms (or first and second bonded atoms for the case of terminal atoms). The calculated multipole moments are then converted to their Cartesian form and rotated to the local axis system of each atom. The locally expressed multipoles are kept constant for small conformational changes during lattice energy minimization. The conformational variability of the electrostatic model is limited to the analytical rotation (using Cartesian tensors) of the local atomic multipoles to the molecular axis system of each newly generated conformation:

$$\begin{aligned}\Omega_{k_1 k_2 \dots k_n}^i(\theta^f, \theta^r(\theta^f)) \\ \approx \sum_{k'_1} \sum_{k'_2} \dots \sum_{k'_n} \text{Rot}_{k_1 k'_1}^i(\theta^f, \theta^r(\theta^f)) \text{Rot}_{k_2 k'_2}^i(\theta^f, \theta^r(\theta^f)) \dots \text{Rot}_{k_n k'_n}^i(\theta^f, \theta^r(\theta^f)) \\ \times \Omega_{k'_1 k'_2 \dots k'_n}^i(\theta_{\text{ref}}^f, \theta_{\text{ref}}^r(\theta_{\text{ref}}^f)) \quad (13)\end{aligned}$$

where the multipole moment of rank n for atom i is calculated using the rotation matrix $\text{Rot}(\theta^f, \theta^r(\theta^f))$ that transforms the local axis system of each atom to the molecular axis system and $\theta^r(\theta^f)$ denotes the LAM-based estimate of the rigid degrees of freedom θ^r using eq 8. After each conformational rotation, the multipoles expressed in the molecular axis system are used to compute the intermolecular electrostatic contribution to lattice energy.

The conformational transferability of multipole moments varies from one molecule to another. For certain functional groups, it cannot be assumed that the localized multipole moments remain

constant even for small conformational changes.²⁰ For instance, the pyramidalization of the $-\text{NH}_2$ group has a direct influence on the position of the electron lone pair on the nitrogen atom that cannot be captured by the analytical rotation of the atomic multipole moments. In such cases, improved accuracy may be obtained by applying a linear correction to the result of eq 13 based on a first-order Taylor expansion:

$$\Omega(\theta^f) = \Omega(\theta^f, \theta^r(\theta^f)) + \left[\frac{\partial \Omega}{\partial \theta^f} \right]_{\theta_{\text{ref}}^f} (\theta^f - \theta_{\text{ref}}^f) \quad (14)$$

The partial derivatives on the right-hand side cannot be obtained in a straightforward manner from currently available QM codes, and consequently the derivatives in the current version of CrystalOptimizer are approximated using finite differences. This requires one additional QM charge density calculation for each flexible degree of freedom being perturbed if one-sided first order finite differences are used.

We note that the correction of eq 14 may not be necessary for all flexible degrees of freedom under consideration, as in many cases the computationally cheaper LAM of eq 13 will already lead to the required accuracy. More research is needed to establish the functional groups for which the application of 14 is necessary.

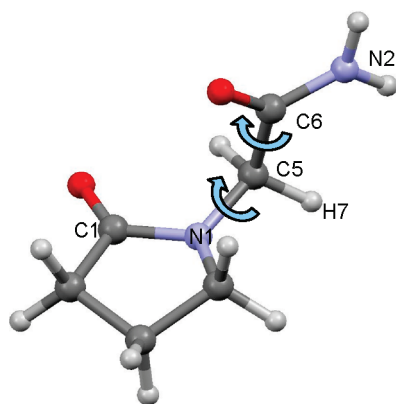


Figure 1. Molecular diagram and atom labeling for piracetam (2-oxo-pyrrolidine-acetamide).

Our preliminary analysis indicates that it is advisable to apply the correction in the case of torsional angles involving nitrogen atoms (e.g. $-\text{NH}_2$ group) or $-\text{OH}$ groups.

2.4. Range of LAM Validity. The range of validity of the proposed LAMs for the estimation of the intramolecular energy and the rigid degrees of freedom (eqs 10 and 8) has been tested against *ab initio* calculations using the GAUSSIAN²¹ suite of programs. The tests were carried out for piracetam (2-oxo-pyrrolidine-acetamide, Figure 1), a molecule comprising 20 atoms (54 intramolecular degrees of freedom). For the purpose of illustration, the two most important torsional angles, N2–C6–C5–N1 and C6–C5–N1–C1, shown as blue arrows in Figure 1, have been considered as the only flexible degrees of freedom.

The *ab initio* intramolecular energy surface, ΔE^{intra} (relative to the energy at the global conformational minimum), is shown in Figure 2a as a function of the two flexible degrees of freedom. The results shown were computed using a total of 64 points on an 8×8 grid. At each point, the two flexible degrees of freedom were fixed at the corresponding grid values, and the remaining rigid degrees of freedom were determined via a quantum mechanical, constrained, isolated-molecule geometry optimization.

As seen in Figure 2b, the intramolecular energy LAM approximates the QM surface with a maximum error of 0.15 kJ mol^{-1} over a range of 5° around the reference point at 90.8° and 155.6° for C6–C5–N1–C1 and N2–C6–C5–N1 torsions, respectively. This error is less than 3% of the 5 kJ mol^{-1} intramolecular energy variation over the conformational region considered. The maximum error is reduced to 0.07 kJ mol^{-1} within 3° of the reference point and 0.02 kJ mol^{-1} within 2° of the expansion point. Similarly, Figure 3 shows that the LAM provides an excellent approximation for the dependence of the rigid degrees of freedom on the flexible torsions. The maximum errors for the rigid degrees of freedom are less than 0.10° for the torsional angle H7–C5–N1–C1 even though this torsion changes by up to 14.5° as the values of the C6–C5–N1–C1 and N2–C6–C5–N1 torsions are modified. Correspondingly, the maximum error is less than 0.06° for bond angle C5–N1–C1 and less than 0.0006 \AA for bond length C1–N1. As expected, the absolute errors are seen to decrease for rigid degrees of freedom that are less sensitive to changes in the flexible degrees of freedom:

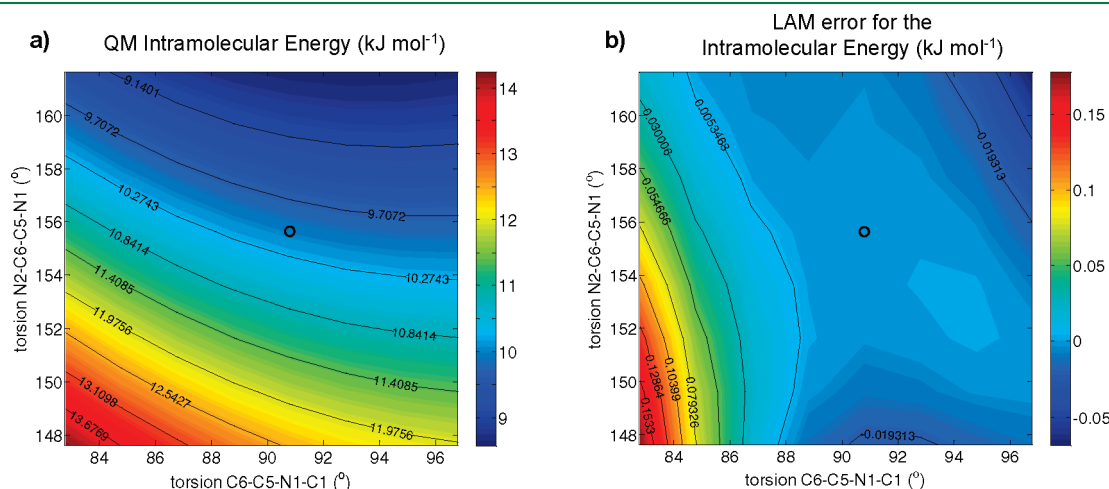


Figure 2. (a) QM intramolecular energy, $\Delta E^{\text{intra, QM}}$ at the HF/6-31G(d,p) level of theory and (b) LAM error for the intramolecular energy defined as $\Delta E^{\text{intra, QM}} - \Delta E^{\text{intra, LAM}}$ as a function of two “flexible” dihedral angles (blue arrows in Figure 1) for piracetam. Open circles correspond to the reference point for the LAM. Reproduced with permission from Adjiman, C. S.; Galindo, A. *Process Systems Engineering: Volume 6: Molecular Systems Engineering*; Wiley-VCH Verlag GmbH & Co. KGaA: New York, 2010; p 15.

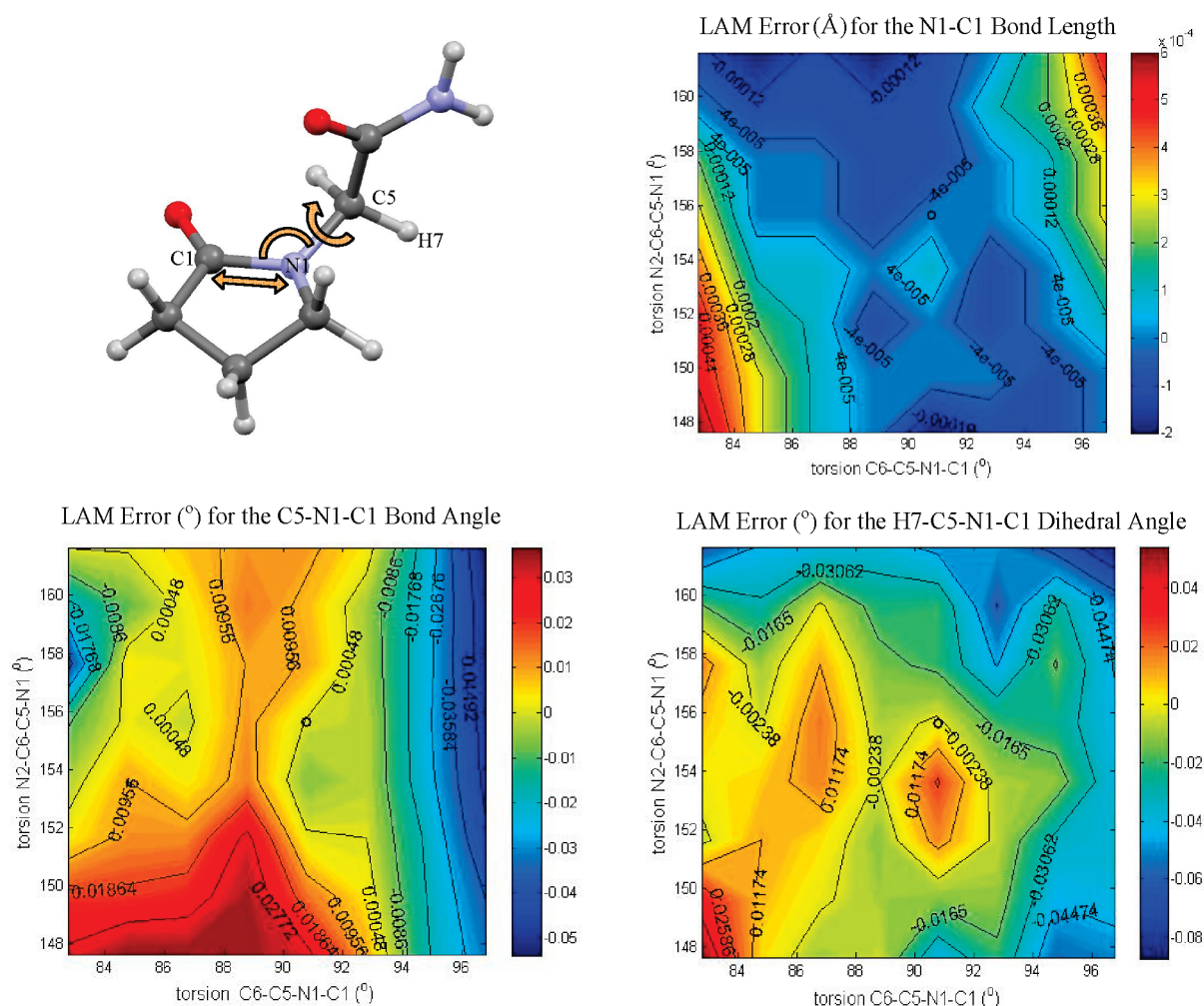


Figure 3. Difference between the quantum mechanical and estimated LAM values for selected “rigid” degrees of freedom (orange arrows, top left) as a function of two “flexible” dihedral angles (blue arrows in Figure 1) for piracetam (2-oxo-pyrrolidine-acetamide). QM calculations performed at the HF/6-31G(d,p) level of theory. Open circles correspond to the reference point for the LAM. Reproduced with permission from Adjiman, C. S.; Galindo, A. *Process Systems Engineering: Volume 6: Molecular Systems Engineering*; Wiley-VCH Verlag GmbH & Co. KGaA: New York, 2010; p 16.

bond angle C5–N1–C1 varies by 1.41° over the region considered and bond length C1–N1 by 0.0094 Å.

Figure 4 is concerned with the accuracy of the multipole LAM (eq 13) in modeling the electrostatic potential energy of xylitol (1,2,3,4,5-pentapentanol) as a selected torsional angle (H1–O1–C1–C2) deviates from the LAM's reference point. The latter corresponds to the molecular conformation at the minimized experimental crystal structure²² using nine major torsional angles (all angles involving hydroxyl groups, and four selected backbone torsions, as shown in Table 1). The quantum mechanical electrostatic potential energy of the reference conformation is shown at the center of Figure 4. For the perturbed conformations, the selected hydroxyl torsion (indicated by a blue arrow on the molecule in Figure 4) was varied by up to $\pm 10^\circ$ from its reference value while the other eight flexible degrees of freedom were held constant at their reference values. For each perturbed molecular geometry, the rigid degrees of freedom were computed using a LAM (eq 8) constructed at the PBE/6-31G(d,p) level of theory. The atomic multipole moments and the electrostatic potential were consequently evaluated using both the LAM (eq 13) and explicit quantum mechanical isolated-molecule

calculations at the PBE/6-31G(d,p) level of theory. The maximum error in the electrostatic potential energy increases the further the conformation moves away from the reference molecular geometry but does not exceed 0.03 eV over the entire range $\pm 10^\circ$ of the flexible torsion angle considered. This error is approximately 2% of the electrostatic potential energy range for the reference conformation. The error is reduced to less than 0.02 eV when the LAM is used to model the electrostatic potential energy within $\pm 5^\circ$ of the reference geometry. Of course, what is important for the purposes of crystal structure prediction is not the electrostatic potential energy *per se* but the intermolecular electrostatic contributions to the lattice energy. In section 4.3, we shall return to consider in more detail the accuracy of the multipole LAMs given by eqs 13 and 14.

In conclusion, the error inherent in any LAM increases as one moves further away from the reference point around which the LAM was constructed. It is therefore necessary to reconstruct the LAM after a significant change in the flexible degrees of freedom. What constitutes a significant change is molecule dependent, but in our experience, ranges of $\pm 5^\circ$ for torsional angles, $\pm 5^\circ$ for bond angles, and ± 0.1 Å for bond lengths leads to reliable results

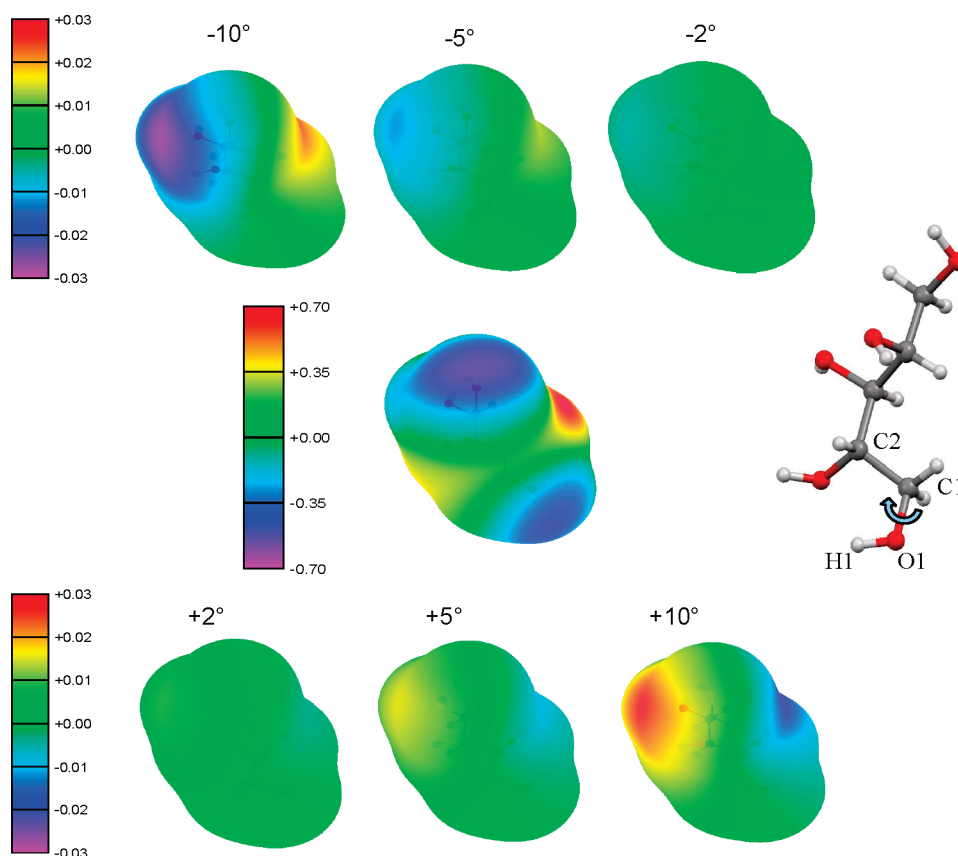


Figure 4. Error in the electrostatic potential energy (in eV; $1 \text{ eV} = 96.5 \text{ kJ mol}^{-1}$) on twice the van der Waals radii surface of xylitol (1,2,3,4,5-pentapentanol) as a function of the H1–O1–C1–C2 torsional angle. LAM used to estimate the rigid degrees of freedom and rotate the atomic multipole moments with their local environment. The quantum mechanical electrostatic potential energy in eV for the reference molecular conformation is also shown for comparison (center). All electrostatic potential energy surfaces were computed with atomic multipoles up to the hexadecapole level, at the PBE/6-31G(d,p) level of theory, using ORIENT.³⁹ The van der Waals radius for hydroxyl hydrogen atoms was set to 1 au; the radii for other atoms were taken from the work of Bondi.⁴⁰ Reproduced with permission from Adjiman, C. S.; Galindo, A. *Process Systems Engineering: Volume 6: Molecular Systems Engineering*; Wiley-VCH Verlag GmbH & Co. KGaA: New York, 2010; p 18.

in the majority of systems. These ranges are comparable to the conformational changes observed during the lattice energy minimization of flexible molecules. Hence, it should be possible to perform such calculations with only a few explicit quantum mechanical optimizations, which provides the motivation for the algorithm presented in this paper.

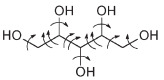
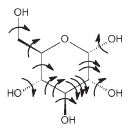
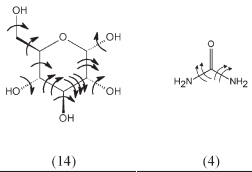
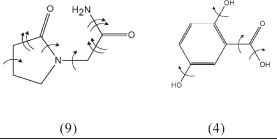
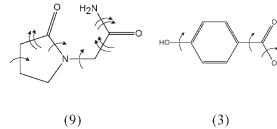
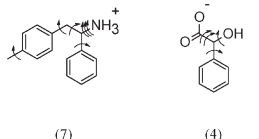
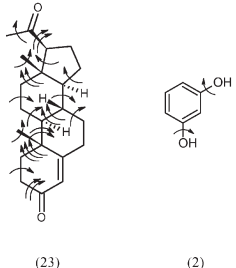
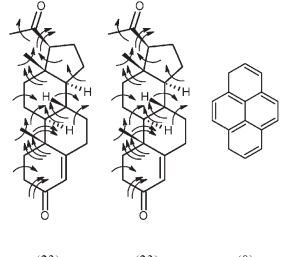
2.5. Reusability of Results of QM Calculations and LAM/QM Databases. For a given molecular species, the local approximate models described above are derived completely from the results of isolated-molecule QM calculations performed at a certain reference point θ_{ref} with a certain level of QM theory. These QM results are independent of the crystalline environment and the physical conditions (e.g., pressure) under which lattice energy minimizations took place. Consequently, they can be reused in repeated calculations involving the same molecule. This is particularly useful in the case of *ab initio* crystal structure prediction studies where it is necessary to perform lattice energy minimizations starting from a large set (possibly hundreds) of candidate structures of a given molecule.

To take advantage of this reusability, all QM-computed quantities (reference conformation, the intramolecular energy, first- and second-intramolecular energy derivatives with respect to all intramolecular degrees of freedom, the localized atomic multipole moments, and their gradients with respect to the flexible degrees of

freedom) used to construct a LAM at any point during a calculation are stored in a database. Torsional angles are allowed to take any value in the course of lattice energy minimization, but are stored only in the $[-180, +180^\circ]$ range in the database. For molecules exhibiting enantiomerism, an entry for the enantiomer of the molecule is also automatically generated from the same *ab initio* calculation, by inverting the values of the torsional angles, intramolecular energy derivatives with respect to the torsional angles, and the relevant components of the multipole moments.

Whenever it is necessary to create LAMs at a new point θ^f in the molecule's conformational space, the corresponding LAM/QM database is searched to identify whether any existing entry θ_{ref} in it can be used to create a LAM that would be valid at θ^f . This would be the case if the differences between the values of the elements of θ^f and the corresponding values in θ_{ref} are all within a given tolerance ε . If more than one database entry meets these validity criteria, the entry with the lowest root-mean-square deviation from θ^f is chosen. On the other hand, if no database entry satisfies the validity criteria, new QM calculations are performed to construct the required LAMs by solving the constrained minimization problem defined by eq 2 and performing a charge density calculation. The QM results are then used to create a new entry in the database with a reference point $\theta_{\text{ref}} = (\theta^f, \theta^r)$, where θ^r is given by eq 3.

Table 1. Systems Considered for CrystalOptimizer Refinement

Name	Molecular Diagram (number of flexible degrees of freedom)	Space Group	Z'
Xylitol ²²	 (12)	$P2_12_12_1$	1
α -D-glucose ²⁵	 (14)	$P2_12_12_1$	1
α -D-glucose Urea (1:1) co-crystal ²⁶	 (14) (4)	$P2_12_12_1$	1
Piracetam	Form I ²⁹	$P2_1/n$	1
	Form II ²⁷	$P\bar{1}$	1
	Form III ²⁸	$P2_1/n$	1
	Form IV ³⁰	$P2_1/c$	1
	Form V ³¹	$P\bar{1}$	1
Piracetam Gentisic acid (1:1) co-crystal ³²	 (9) (4)	$C2/c$	2
Piracetam p-Hydroxybenzoic acid (1:1) co-crystal ³²	 (9) (3)	$P2_1/n$	2
(R)-1-phenyl-2-(4-methylphenyl) ethylammonium (S)-mandelate salt ³³	 (7) (4)	$P2_12_12_1$	2
Progesterone Resorcinol (1:1) co-crystal ³⁴	 (23) (2)	$P2_12_12$	2
Progesterone Pyrene (2:1) co-crystal ³⁵	 (23) (23) (0)	$P1$	3

A LAM/QM database is specific to a particular molecule and level of QM theory. However, it can be reused and, indeed, extended during more than one calculation involving this molecule. For example, a database created during the lattice energy minimization of a given experimentally observed polymorph may be used on a later occasion for the minimization of the same experimental structure under a different pressure, or of a different polymorph, or of a cocrystal that involves the given molecule together with a different one. For this reason, LAM/QM databases are stored as persistent computer files that can be used in any calculation relating to the corresponding molecule, potentially being extended during each such calculation to contain an increasing number of points, thereby becoming more and more useful as an increasing fraction of the molecule's conformational space is covered. An illustrative example of the performance gain that can be achieved due to the use of databases in an *ab initio* crystal structure prediction study is discussed in section 4.6.

Finally, it is worth pointing out that the information being stored in a LAM/QM database entry comprises the results of a QM calculation (i.e., the values of the intramolecular energy and its partial derivatives appearing on the right-hand side of eq 4) and not the LAMs of eqs 8 and 10 (e.g., the matrices $A(\theta_{\text{ref}})$ and $C(\theta_{\text{ref}})$ and the vector $b(\theta_{\text{ref}})$) derived from them. Thus, the LAM/QM database entries are independent of the specific way in which the conformational degrees of freedom θ^r are partitioned between “flexible” θ^f and “rigid” θ^r variables. Consequently, this information may be used for different calculations pertaining to the same molecule under different degrees of flexibility, provided eq 5 still holds. In practice, this means that an existing database can be reused by subsequent calculations considering the same or higher degree of flexibility. It is worth noting that the requirement for nondecreasing flexibility does not apply to the atomic multipole moments; these can be reused across different calculations involving any degree of flexibility as long as the definition of the local axis system for each atom in the molecule remains the same.

3. COMPUTATIONAL METHODOLOGY—CRYSTALOPTIMIZER

3.1. Lattice Energy Minimization. The lattice energy minimization problem can be written as

$$\min_{\mathbf{X}, \theta} E^{\text{latt}} \equiv \min_{\mathbf{X}, \theta} [\Delta E^{\text{intra}}(\theta) + U^{\text{inter}}(\mathbf{X}; \theta, \Omega(\theta))] \quad (15)$$

where ΔE^{intra} is the energy required to deform the molecule from its most stable gas-phase conformation. Stable crystal forms can therefore be identified by minimizing the lattice energy with respect to the intramolecular degrees of freedom, θ (bond lengths, bond angles, and torsional angles), and lattice variables, \mathbf{X} , which include the unit cell geometry and the position and orientation of all crystallographically independent molecules in the lattice. Here, $\Omega(\theta)$ denotes the distributed multipole model⁸ used to represent the dominant electrostatic contributions to the intermolecular energy.

As has already been explained, the dimensionality of the above optimization problem can be reduced by the partitioning of the intramolecular degrees of freedom, θ , into flexible, θ^f , and rigid, θ^r , degrees of freedom, leading to the modified minimization

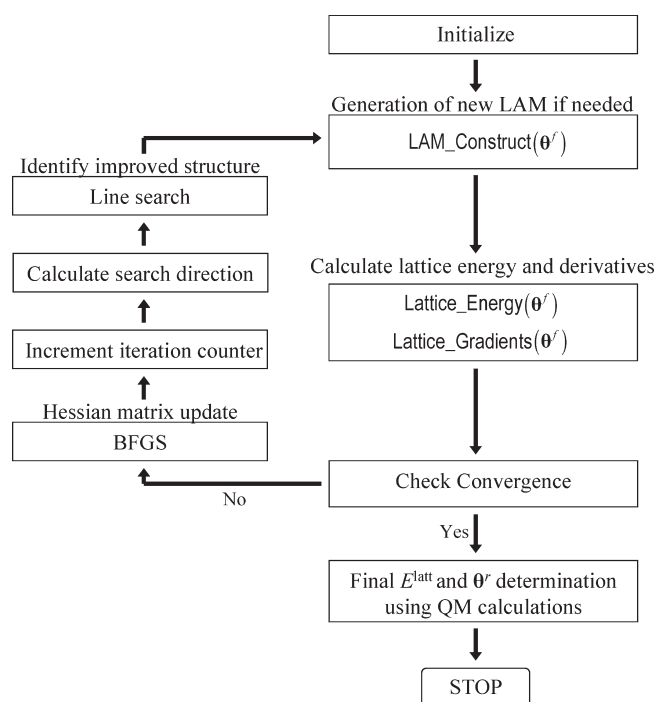


Figure 5. CrystalOptimizer local lattice energy minimization algorithm flowchart. Refer to Figure 6 for procedures for LAM construction (LAM_Construct), lattice energy (Lattice_Energy), and lattice energy derivatives (Lattice_Gradients) calculation.

problem:

$$\min_{\mathbf{x}, \theta^f} E^{\text{latt}} \equiv \min_{\mathbf{x}, \theta^f} [\Delta E^{\text{intra}}(\theta^f) + U^{\text{inter}}(\mathbf{X}; \theta^f, \theta^r(\theta^f), \Omega(\theta^f, \theta^r(\theta^f)))] \quad (16)$$

where $\Delta E^{\text{intra}}(\theta^f)$ and $\theta^r(\theta^f)$ are defined by eqs 2 and 3, respectively, and can be approximated accurately and efficiently via the LAMs shown in eqs 10 and 8, respectively. The multipoles $\Omega(\theta^f, \theta^r(\theta^f))$ can also be approximated via the LAMs defined by eqs 13 and 14.

For a given molecular conformation θ , the minimum intermolecular energy with respect to the lattice variables, \mathbf{X} , can be calculated with existing algorithms such as DMACRYS.⁵ Therefore, in order to use such codes, eq 16 is further reformulated as

$$\min_{\theta^f} E^{\text{latt}} = \min_{\theta^f} [\Delta E^{\text{intra}}(\theta^f) + \bar{U}^{\text{inter}}(\theta^f)] \quad (17)$$

where the second term on the right-hand side is given by the solution of another minimization problem:

$$\bar{U}^{\text{inter}}(\theta^f) \equiv \min_{\mathbf{X}} U^{\text{inter}}(\mathbf{X}; \theta^f, \theta^r(\theta^f), \Omega(\theta^f, \theta^r(\theta^f))) \quad (18)$$

Equations 17 and 18 define a bilevel optimization problem. The inner minimization (eq 18) determines the crystal structure for *rigid* molecular entities, whose conformation is determined by the outer minimization (eq 17) manipulating the flexible degrees of freedom θ^f . It should be noted that the DMAflex algorithm¹³ solves the same bilevel optimization problem, using DMACRYS⁵ to calculate the intermolecular energy (eq 18).

3.2. The CrystalOptimizer Algorithm. CrystalOptimizer is a local lattice energy minimization algorithm designed to solve the optimization problem given by eq 17 and making use of LAMs in

order to reduce the computational cost associated with quantum mechanical calculations. An overview of the CrystalOptimizer algorithm is shown in Figure 5 and will now be discussed in detail.

In the initialization steps, the algorithm requires the user to specify the starting crystal structure. By representing the molecular geometry in the automatically generated Z-matrix form (that avoids near-linear bond angles), the proposed algorithm allows the user to select the extent of molecular flexibility to be considered during the minimization. This can range from a few selected torsional angles to full atomistic minimization.

Before the optimization is carried out, the user is also required to specify several model parameters such as optimization convergence tolerances, the choice of quantum mechanical methods and basis sets, and the tolerance vector ε that defines the range of LAM validity for different types of flexible degrees of freedom. If required, different tolerances may be specified for the LAMs relating, respectively, to intramolecular energy and multipoles. For each flexible degree of freedom, the user also specifies whether or not the linear update to the multiple moments (eq 14) is to be used. Input information relating to DMACRYS, such as the cutoff range for Ewald summation, the type of repulsion-dispersion potential, and the values of parameters within it, must also be specified.

CrystalOptimizer uses a quasi-Newton algorithm coupled with a line-search²³ to solve the outer minimization problem. This approach ensures rapid convergence even when there are many flexible degrees of freedom by using an approximation of the Hessian matrix of the second-order derivatives of the lattice energy with respect to the flexible degrees of freedom. At the start of the lattice energy minimization, this Hessian approximation is normally initialized to the unit matrix; at each subsequent iteration, it is updated via the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method²³. This ensures that the Hessian approximation remains positive-definite, which guarantees the identification of a direction in which to change the flexible degrees of freedom θ^f , which results in a reduction of the lattice energy, thereby avoiding the location of transition states. The use of the BFGS approximation avoids the evaluation of the second-order derivatives of the lattice energy at every outer iteration, instead making use only of values of the lattice energy and its first-order gradients with respect to the flexible degrees of freedom.

As can be seen from the right-hand side of eq 17, the computation of the lattice energy for given θ^f involves two components. The first one, $\Delta E^{\text{intra}}(\theta^f)$, is computed explicitly via the LAM of eq 10. The second component is computed by solving the inner minimization problem (eq 18) using DMACRYS,⁵ with the molecular conformation and distributed multipole moments fixed at the values determined via the LAMs of eqs 8, 13, and 14.

In addition to the value of the lattice energy for given θ^f , the outer optimization algorithm also requires values of its gradients with respect to θ^f . The gradients of its first component, $\Delta E^{\text{intra}}(\theta^f)$, can be computed in a straightforward manner by differentiating the LAM of eq 10:

$$\left. \frac{\partial \Delta E^{\text{intra}}}{\partial \theta^f} \right|_{\theta^f} = \mathbf{b}(\theta_{\text{ref}}) + \mathbf{C}(\theta_{\text{ref}})(\theta^f - \theta_{\text{ref}}^f) \quad (19)$$

The gradients of the second component, \bar{U}^{inter} , of the lattice energy with respect to θ^f cannot be obtained in a closed analytical form. In CrystalOptimizer, they are computed via a centered finite difference scheme. The gradient with respect to the k^{th}

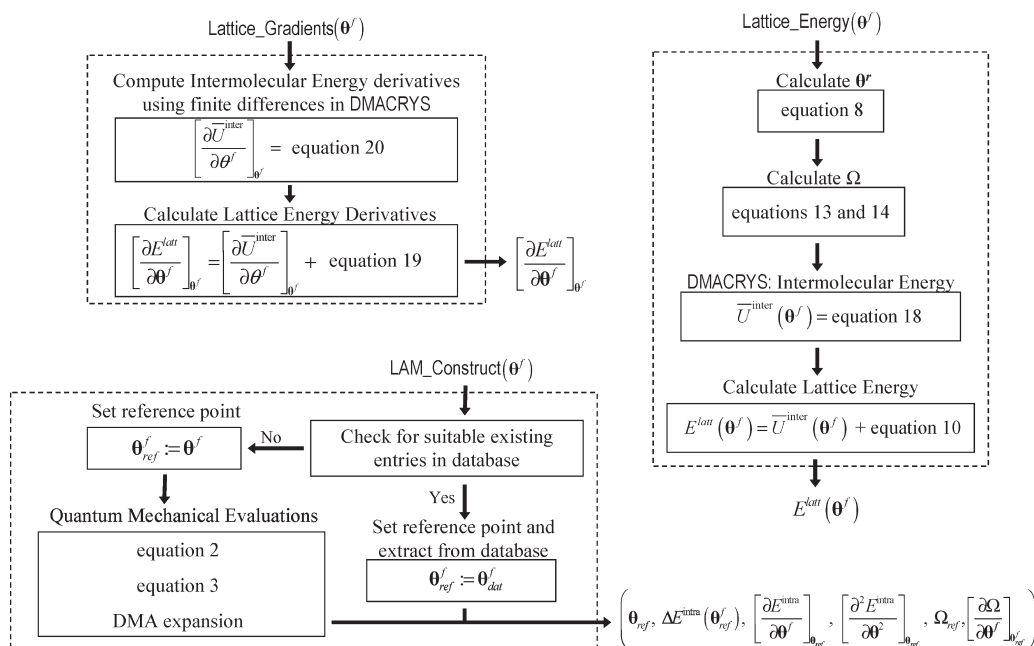


Figure 6. Procedures for lattice energy derivatives (top left), lattice energy (right) calculations, and LAM construction (bottom left) in CrystalOptimizer.

element θ_k^f of the vector θ^f requires a positive and a negative perturbation of magnitude $\delta\theta_k^f$. For each perturbed value, the LAMs of eqs 8, 13, and, where necessary, 14 are used to update the values of the rigid degrees of freedom θ^r and the multipoles Ω , before calling DMACRYS to determine \bar{U}^{inter} via the minimization described by eq 18. The required gradient is then obtained via the finite difference approximation:

$$\left.\frac{\partial \bar{U}^{\text{inter}}}{\partial \theta_k^f}\right|_{\theta^f} \approx \frac{\bar{U}_{k+}^{\text{inter}} - \bar{U}_{k-}^{\text{inter}}}{2\delta\theta_k^f} \quad (20)$$

where $\bar{U}_{k+}^{\text{inter}}$ and $\bar{U}_{k-}^{\text{inter}}$ denote the values of \bar{U}^{inter} returned by DMACRYS for the positive and negative perturbations, respectively.

The required gradients $\partial E^{\text{latt}}/\partial \theta^f$ are then simply computed as the sum of expressions 19 and 20. The procedures for calculating the lattice energy and the gradient of the lattice energy with respect to the flexible degrees of freedom are shown in Figure 6.

On the basis of the value of the lattice energy and the search direction calculated using the lattice energy gradient and the BFGS Hessian matrix approximation, the line-search procedure in the outer minimization algorithm determines a new set of flexible degrees of freedom that results in a sufficiently large reduction of the lattice energy. This involves the evaluation of lattice energy at a sequence of points along the search direction. All but the last of these points do not fulfill the criterion of sufficient reduction in the lattice energy and are immediately discarded from further consideration. As it is not generally worth performing any expensive QM calculations at such points, during the line search part of the algorithm, we relax the LAM validity tolerances ε by a factor n which is typically in the range $1 < n \leq 2$.

After the line search identifies a new improved point in conformational space θ^f , we reset the LAM validity tolerances to their original value ε , before evaluating the new lattice energy

and its gradients with respect to θ^f in the manner detailed above. The BFGS approximation is then used to provide a new estimate of the Hessian matrix, and the algorithm proceeds to the next outer iteration by calculating a new search direction and performing a line search along it.

The optimization terminates successfully when either the changes of all flexible degrees of freedom during the last step or all the lattice energy gradients with respect to θ^f are below specified tolerances. As a final step, CrystalOptimizer uses rigorous QM calculations (eqs 2 and 3) to recompute the lattice energy and molecular conformation at the values of θ^f determined by the optimization, thereby eliminating any small inconsistencies that may have arisen from the use of LAMs. It should be noted, however, that this final calculation does not eliminate any error caused by using LAMs in identifying the true minimum.

CrystalOptimizer returns a failure status if either the line search fails to identify a new set of flexible degrees of freedom that produce a sufficient reduction in the lattice energy along the search direction, or when a predefined maximum number of outer iterations is reached without either of the two convergence criteria having been satisfied.

4. ALGORITHM TESTING, RESULTS AND DISCUSSION

The main question to be answered in validating the CrystalOptimizer algorithm is whether, for a given initial point, the use of the local approximate models leads to the same local energy minimum as when using QM calculations at each minimization iteration. Once this is established, the performance of the algorithm, and particularly the reduction in computational cost for a given number of flexible degrees of freedom can be investigated. The use of additional degrees of freedom can also be studied, and the impact of flexibility on crystal structure can be assessed.

4.1. Systems Studied. The validation of the CrystalOptimizer algorithm was performed on the systems shown in Table 1. These were selected because of their scientific and/or practical

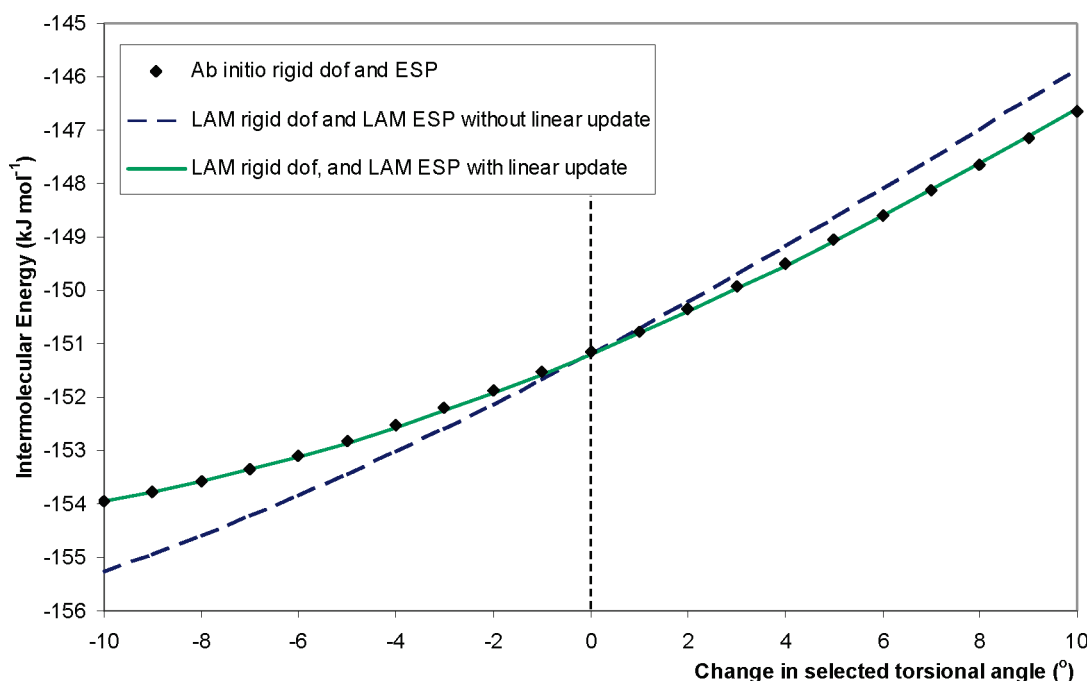


Figure 7. Intermolecular energy as a function of the H1–O1–C1–C2 torsional angle (blue arrow in Figure 4) for three different computational strategies: (1) molecular geometry and the electrostatic potential (ESP) obtained by full quantum mechanical calculations (black diamonds), (2) LAM used to update rigid degrees of freedom (dof) and multipole rotation (blue dashed line), (3) LAMs used to update the rigid degrees of freedom, multipole rotation, and a linear update of the multipole moments (continuous green line).

importance and the fact that they possess sufficient conformational flexibility to present a significant challenge for lattice energy minimization using current techniques.

Xylitol, a naturally occurring sugar alcohol, is a stereoisomer of 1,2,3,4,5-pentapentanol (pentose) for which only one crystal structure has been determined experimentally to date.²² No other organic systems containing xylitol were found in the Cambridge Crystallographic Database²⁴ (CSD).

Our second test system is glucose. We consider the structure of one of its cyclical isomers, α -D-glucose.²⁵ A cocrystal of α -D-glucose with urea²⁶ is also found in the CSD.

Piracetam (2-oxo-pyrrolidine-acetamide) is a pharmaceutical nootropic drug. There are five distinct polymorphs determined experimentally,^{27–31} two of which are observed only at high pressure. Two cocrystals³² of piracetam are also reported in the CSD.

In order to examine the applicability of CrystalOptimizer to a range of systems of practical interest, three additional crystals were considered, namely the salt (*R*)-1-phenyl-2-(4-methylphenyl)ethylammonium-(*S*)-mandelate³³ and the cocrystals of the pharmaceutically important steroid progesterone with resorcinol³⁴ and pyrene.³⁵ These were selected primarily on the basis of posing significant computational challenges in terms of both the molecular size and flexibility.

4.2. Molecular Modeling and Computational Considerations. The quantum mechanical molecular geometry optimizations and the charge density calculations were both evaluated at the PBE/6-31G(d,p) level of theory.

The LAM validity tolerances were set at $\pm 5^\circ$ for torsional and bond angles and ± 0.02 Å for bond lengths. The linear update to the multipole moments (eq 14) was used for all flexible torsional angles that involved the nitrogen atom or the OH groups. Although the extent to which the LAMs are reliable is system-dependent, the above range of validity is expected to provide a

sufficiently small error compared with the other approximations present in the computational model.

The repulsion–dispersion interactions were modeled with an empirical exp-6 potential¹⁰ and were summed in direct space up to a 30 Å cutoff.

CrystalOptimizer was used to perform local lattice energy minimizations starting from the experimental structures for each model system considered. The flexible degrees of freedom that were taken into consideration for each system are marked in the molecular diagrams shown in the second column of Table 1.

Most calculations were performed on a single Intel Xeon S150 2.66 GHz processor using 1500 MB of memory. In the case of the salt and progesterone cocrystals, the minimizations were performed on four Intel Xeon S150 2.66 GHz processors with 7 GB of shared memory. The availability of multiple processors was exploited for performing the QM computations using the GAUSSIAN²¹ code.

4.3. Use of Linear Updates for Multipole LAMs. Section 2.4 examined the range of validity of the LAMs for multipoles in terms of their ability to reproduce the electrostatic potential field surrounding an isolated molecule. As noted there, what is more important is the accuracy with which these LAMs can approximate the intermolecular electrostatic contributions to the lattice energy.

We now return to consider this question in more detail using xylitol, with the minimized experimental structure (using nine flexible hydroxyl and backbone torsions) as the reference point. The H1–O1–C1–C2 hydroxyl angle is varied by up to $\pm 10^\circ$ from the reference point while maintaining the remaining flexible degrees of freedom at their reference values; DMACRYS is used to reoptimize the crystal lattice at each different conformation under consideration, and the intermolecular energy contribution to the lattice energy at the optimal structure is recorded.

Figure 7 shows results from three different sets of calculations performed for several values of the selected torsion angle over the range mentioned above.

For each value of the flexible hydroxyl angle, the rigid degrees of freedom were computed explicitly by solving the quantum mechanical constrained optimization in eq 3. The electrostatic potential was also calculated quantum mechanically for each generated conformation. These molecular and electrostatic models were used to compute the intermolecular energy shown as black diamonds in Figure 7.

Next, the LAM of eq 8 was used to approximate the rigid degrees of freedom, and the multipoles were rotated according to eq 13. The results are shown as a blue dashed line in Figure 7. The maximum error in the intermolecular energy is 1.3 kJ mol^{-1} and occurs when the selected flexible torsion angle is 10° from its reference value. The errors for changes of up to $\pm 5^\circ$ and $\pm 2^\circ$ from the reference value are 0.6 kJ mol^{-1} and 0.3 kJ mol^{-1} , respectively.

The intermolecular energy was also calculated using the LAMs for the estimation of the rigid degrees of freedom (eq 8), multipole rotation (eq 13), and a linear update to the multipole moments (eq 14). The results are shown as a continuous green line in Figure 7. The error in the intermolecular energy is within 0.06 kJ mol^{-1} of the *ab initio* value for all values of the selected flexible degree of freedom.

The above results indicate that, for this particular system, the multipole rotation LAM of eq 13 leads to considerable errors in intermolecular energy. However, the subsequent application of the linear correction (eq 14) produces an accurate estimate of the electrostatic potential energy, even for relatively large conformational changes.

4.4. Application of CrystalOptimizer to Single Component Crystals. We start by considering the three model systems which involve single component crystals, namely, xylitol, α -D-glucose, and piracetam form II.

In order to assess the computational performance and the effects of varying molecular flexibility on the performance of the CrystalOptimizer, these systems were studied using different optimization settings and increasingly wider sets of flexible degrees of freedom.

Initially, only the hydroxyl (amide for piracetam form II) and selected backbone dihedrals were treated as flexible degrees of freedom, while the rings in α -D-glucose and piracetam were assumed to be rigid. These simplifications result in nine, six, and four flexible degrees of freedom for xylitol, α -D-glucose, and piracetam, respectively. Although these minimal sets of flexible torsion angles are not sufficient to capture the whole molecular flexibility, they have the advantage of being within the range of applicability of earlier algorithms for lattice energy. Of particular interest in this context is the DMAFlex algorithm,¹³ which has a similar model of inter- and intramolecular interactions as CrystalOptimizer, thereby allowing some validation of the results obtained with our code, and a direct comparison of computational performance. Three different CrystalOptimizer runs were performed with the same number of degrees of freedom:

Case 1. CrystalOptimizer without LAMs or LAM/QM databases was tested. This requires quantum mechanical calculations at every iteration, as with DMAFlex, and shows the effect of using a quasi-Newton algorithm instead of a simplex algorithm.

Case 2. All features of CrystalOptimizer are used. The LAM/QM databases are initially empty and are populated

during these calculations. In fact, for the calculations reported here, no reuse of information stored in the databases has actually occurred. Therefore, only the impact of the LAMs is assessed.

Case 3. All features of CrystalOptimizer are used, including the use of the LAM/QM databases (cf. section 2.5) already populated in case 2.

Following this initial study, additional optimizations were carried out with CrystalOptimizer, by gradually increasing the degree of molecular flexibility under consideration in two further steps.

Case 4. All heavy-atom torsions, and all hydroxyl (H—O—C) and amide (H—N—C) bond angles were additionally treated as flexible, resulting in 17 flexible degrees of freedom for xylitol and 19 for α -D-glucose and piracetam.

Case 5. An atomistic representation corresponding to full molecular flexibility is considered, resulting in 60 flexible degrees of freedom for xylitol, 66 for α -D-glucose, and 54 for piracetam.

Note that in the analysis of the results reported in cases 1, 4, and 5 and also in section 4.5, the LAM/QM database (cf. section 2.5) feature of CrystalOptimizer was disabled. A more detailed analysis of the effects of using these databases is presented in section 4.6.

The detailed results from the above studies are provided in Tables 2, 3, and 4 for xylitol, α -D-glucose, and piracetam form II, respectively. In all cases, agreement between the predicted crystal structures and the experimentally observed ones is assessed on the basis of the root-mean-square deviation of the molecular conformation and the 15-molecule coordination sphere.³⁶

Overall, there is a gain in stability as we widen the set of flexible degrees of freedom considered by CrystalOptimizer, which indicates that the detailed modeling of molecular flexibility is important in order to capture the full extent of conformational distortions by the packing forces. For instance, the inclusion of the selected bond angles (case 4) stabilizes the structures of xylitol, α -D-glucose, and piracetam by roughly 1 kJ mol^{-1} compared to the value obtained by considering the flexibility of only the main torsional angles (cases 1–3). As the number of flexible degrees of freedom increases, small changes are observed in the lattice energy, but the balance between intermolecular and intramolecular energy tends to shift so that the molecule adopts a less stable conformation to reduce the energetic cost of intermolecular interactions. CrystalOptimizer makes it computationally feasible to consider a large number of degrees of freedom but it is up to the user to define the extent of flexibility required for each particular molecule or system.

More generally, explicit consideration of the flexibility of bond angles and stiff torsions, such as those in polyaromatic ring systems, is expected to be especially significant for larger systems and under high pressure. On the other hand, it can be seen by comparing cases 4 and 5 that the additional energy stabilization arising from modeling the bond lengths is more limited (unless proton transfer takes place³⁷). The energy required to perturb a bond length is usually very large, and consequently the modeled bond lengths do not change significantly during lattice energy minimization. Thus, in order to reduce computational cost, the bond lengths can normally be assumed to be rigid with little loss of predictive accuracy.

Figure 8 demonstrates the accuracy of reproduction of the experimental conformation obtained by considering full

Table 2. Lattice Energy Minimization for Xylitol

	CrystalOptimizer				
	DMAFlex ^a	no LAMs, no LAM databases	LAMs, no LAM databases	LAMs and LAM databases	more flexibility, LAMs but no LAM databases
	case 1	case 2	case 3	case 4	case 5
no. of flexible degrees of freedom ^b	9	9	9	17	60
lattice energy (kJ mol ⁻¹)	main torsions -125.32	main torsions -125.15	main torsions -125.16	all torsions, ^c OH angles -127.04	atomistic ^d -127.78
intramolecular energy (kJ mol ⁻¹)	23.31	23.55	23.43	23.18	24.47
intermolecular energy (kJ mol ⁻¹)	-148.62	-147.70	-148.58	-150.22	-152.25
rmsd ^e	0.20	0.19	0.19	0.19	0.23
CPU time (hr:min) ^f	35:32	07:30	04:01	12:20	43:55
no. of outer iterations	127	13	13	14	23
no. of QM calculations	127	4	1	5	5
molecular Hessian		3	0	4	4
charge density	127	19	1	25	25
QM molecular optimization	53.83	4.79	1.07	3.33	0.50
QM Hessian calculation		12.95	0.00	11.04	2.94
QM charge density, DMA	41.01	29.02	2.82	23.55	6.50
DMACRYs	5.13	52.96	95.60	61.75	89.58
other	0.03	0.28	0.51	0.33	0.48

^a Karamertzanis and Price.¹³ ^b Linear update to the multipole moments calculated for all torsions involving the OH group. ^c Excluding dihedrals defined as H-C-X-X. ^d All intramolecular degrees of freedom. ^e Root-mean-square deviation in 15-molecule coordination sphere compared with the experimental structure (excluding hydrogen atoms). ^f Single Intel Xeon 5150 2.66 GHz processor using 1500 MB of memory.

Table 3. Lattice Energy Minimization for α -D-Glucose

	CrystalOptimizer				
	DMAFlex ^a	no LAMs, no LAM databases	LAMs, no LAM databases	LAMs and LAM databases	more flexibility, LAMs but no LAM databases
	case 1	case 2	case 3	case 4	case 5
no. of flexible degrees of freedom ^b	6	6	6	19	66
lattice energy (kJ mol ⁻¹)	main torsions -146.08	main torsions -146.41	main torsions -146.39	all torsions, ^c OH angles -147.86	atomistic ^d -148.62
intramolecular energy (kJ mol ⁻¹)	9.60	10.08	10.05	12.54	12.55
intermolecular energy (kJ mol ⁻¹)	-155.68	-156.49	-156.44	-160.40	-161.16
rmsdis ^e	0.20	0.20	0.20	0.16	0.14
CPU time (hr:min) ^f	44:13	09:20	02:17	13:22	41:32
no. of outer iterations	118	14	11	16	22
no. of QM calculations	118	5	1	4	4
molecular optimization		4	0	3	3
molecular Hessian		25	1	19	19
charge density	118				
QM molecular optimization	55.47	8.33	2.71	4.73	0.59
QM Hessian calculation		21.69	0.00	11.20	3.57
QM charge density, DMA	41.14	41.53	6.66	21.91	7.02
DMACRYs	3.37	28.25	89.98	61.71	88.18
other	0.02	0.20	0.65	0.45	0.64

^a Karamertzanis and Price.¹³ ^b Linear update to the multipole moments calculated for all torsions involving the OH group. ^c Excluding dihedrals defined as H-C-X-X. ^d All intramolecular degrees of freedom. ^e Root-mean-square deviation in 15-molecule coordination sphere compared with the experimental structure (excluding hydrogen atoms). ^f Single Intel Xeon 5150 2.66 GHz processor using 1500 MB of memory.

Table 4. Lattice Energy Minimization for Piracetam Form II

	CrystalOptimizer				
	DMAFlex ^d		no LAMs, No LAM databases		
			case 1	case 2	case 3
no. of flexible degrees of freedom ^b	4	4	4	4	4
lattice energy (kJ mol ⁻¹)	main torsions	main torsions	main torsions	main torsions	main torsions
intramolecular energy (kJ mol ⁻¹)	-93.49	-93.31	-93.31	-93.31	-93.31
intermolecular energy (kJ mol ⁻¹)	15.27	15.47	15.47	15.47	15.47
rmsd ^f	-108.76	-108.78	-108.78	-108.78	-108.78
CPU time (hr:min) ^g	0.25	0.25	0.25	0.25	0.25
no. of outer iterations	11:42	02:11	01:28	01:28	00:17
no. of QM calculations	55	2	2	2	2
molecular optimization	55	2	2	2	1
molecular Hessian		2	1	0	0
charge density	55	10	6	6	49
QM molecular optimization	57.45	18.18	27.61	27.61	15.48
QM Hessian calculation		27.04	19.79	19.79	0.00
QM charge density, DMA	39.97	48.15	43.14	43.14	36.54
DMACRY	2.57	6.62	9.43	9.43	47.86
other	0.01	0.01	0.03	0.03	0.12
					0.13
					0.15

^a Karamertzanis and Price.¹³ ^b Linear update to the multipole moments calculated for all torsions involving the N atom. ^c Excluding dihedrals defined as H-C-X-X. ^d Excluding bond angles defined as H-C-X-X. ^e All intramolecular degrees of freedom. ^f Root-mean-square deviation in 15-molecule coordination sphere compared with the experimental structure (excluding hydrogen atoms). ^g Single Intel Xeon 5150 2.66 GHz processor using 1500 MB of memory.

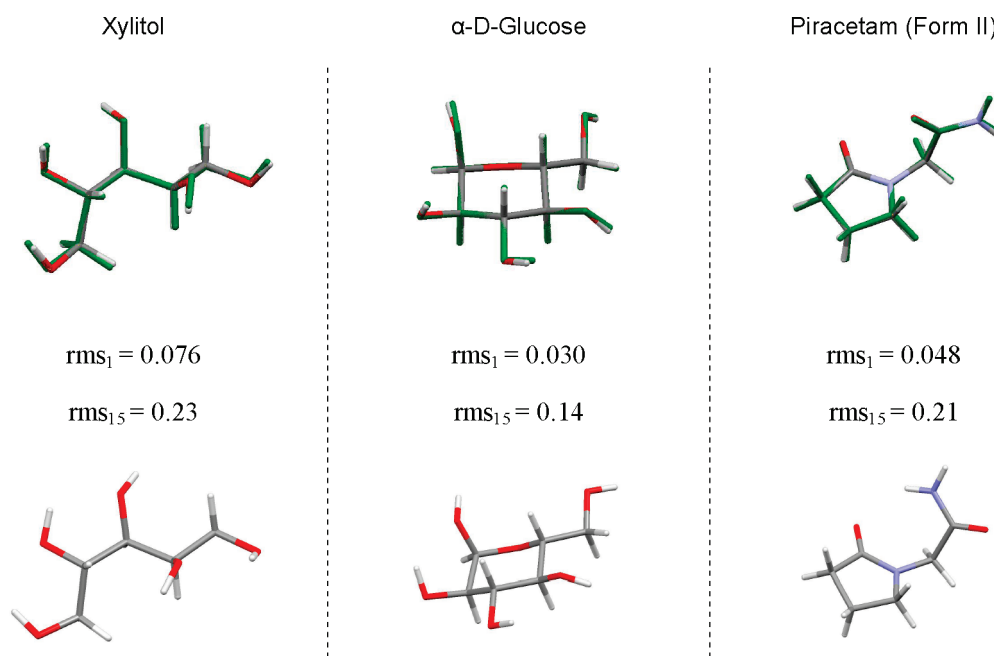


Figure 8. Top row: Overlay of the experimental (colored by element) and minimized (green) structures of xylitol, α -D-glucose, and piracetam when full molecular flexibility is allowed (atomistic representation). The root-mean-square deviations for one molecule (rms_1) and a 15-molecule coordination sphere (rms_{15})³⁶ compared to the experimental structure are shown below. Bottom row: The gas-phase molecular conformations for the three molecules. Reproduced with permission from Adjiman, C. S.; Galindo, A. *Process Systems Engineering: Volume 6: Molecular Systems Engineering*; Wiley-VCH Verlag GmbH & Co. KGaA: New York, 2010; p 36.

molecular flexibility (atomistic representation). Even with the low-quality level of theory used for the QM evaluations, the reproduction of the crystal structures with CrystalOptimizer is acceptable for crystal structure prediction. As shown in Tables 2–4, the maximum error in the root-mean-square deviation of a 15-molecule coordination sphere in all three cases is below 0.3 Å which is less than the 0.4 to 0.5 Å criterion usually taken to indicate a successful prediction in crystal structure blind tests. In general, structure reproduction improves as a larger number of flexible degrees of freedom are being considered. An exception to this trend is observed for xylitol (see the increase in rms_{15} in the last three columns of Table 2), but the overall rms_{15} variation for this molecule is too small to warrant further attention, considering that thermal effects are not taken into account.

Comparisons of the DMAFlex results to those of cases 1–3, as presented in Tables 2–4, indicate that the results obtained from CrystalOptimizer and DMAFlex in terms of energy and structure reproduction are within the errors of the underlying numerical methods, and in particular of the criteria used to determine what constitutes a “converged” solution of the optimization calculations. In the case of CrystalOptimizer, some convergence inaccuracies are introduced by the use of finite differences for the evaluation of the gradients of the intermolecular energy. For DMAFlex, there are convergence issues arising from the use of the gradient-free simplex-based optimization algorithm.

CrystalOptimizer is significantly more computationally efficient than DMAFlex. With all features implemented (case 3), the CPU time is reduced by $\sim 89\%$ for xylitol, $\sim 95\%$ for glucose, and $\sim 97\%$ for piracetam when considering the same degree of molecular flexibility. One reason for this is that CrystalOptimizer’s gradient-based minimization scheme converges in an order of magnitude fewer outer iterations than the simplex algorithm in DMAFlex and scales better with the number of modeled

flexible degrees of freedom, as can be observed by comparing the DMAFlex results and case 1 columns in Tables 2–4. This accounts for an average reduction in CPU time of 63%. A second reason for the much improved computational performance of CrystalOptimizer is that the use of LAMs means that only a small fraction of the outer iterations requires full quantum mechanical calculations. It should be noted, however, that the types of QM calculations required by DMAFlex and CrystalOptimizer are not the same. Each outer iteration in DMAFlex simply requires one constrained geometry optimization and one charge density calculation, the former usually being more computationally expensive. On the other hand, the construction of the LAMs in CrystalOptimizer requires a constrained geometry optimization followed by a (usually much more expensive) analytical evaluation of the intramolecular Hessian matrix (i.e., the second-order partial derivative matrices appearing on the right-hand side of eq 4) and a charge density calculation. Moreover, if the LAM given by eq 14 is used for one or more flexible degrees of freedom, we also need one or two (depending on the finite difference method used for approximating the partial derivatives on the right-hand side of eq 14) additional charge density calculations for each of such flexible degrees of freedom. Nevertheless, despite the additional cost of constructing the LAMs in CrystalOptimizer, the frequency with which this has to be done is small (case 2, Tables 2–4), leading to a further reduction in computational cost of 46%. The poor scaling of the DMAFlex’s simplex optimization algorithm with the number of minimization variables, coupled with the significant cost of the QM calculations, indicates that this approach is unlikely to be practically applicable to systems of even moderate molecular size if they involve more than about 10 flexible degrees of freedom. We believe that any other approach that relies on explicit QM calculations at every iteration would be subject to similar limitations.

Table 5. Lattice Energy Minimization of Additional Crystal Structures Containing α -D-Glucose and Piracetam in the Asymmetric Unit (Refer to Table 1 for Molecular Diagrams)

model system		lattice energy (kJ mol ⁻¹) (pressure) ^a	rmsd ₁₅ ^b (Å)	# outer iterations	# LAM updates ^c	CPU time (hr:min) ^d
α -D-glucose urea (1:1) cocrystal		-221.62 (0.0 GPa)	0.11	13	2	19:08
	form I ^e	-91.56 (0.0 GPa)	0.21	5	1	02:58
	form II	-93.94 (0.0 GPa)	0.22	9	2	04:28
piracetam	form III	-93.97 (0.0 GPa)	0.13	13	3	07:28
	form IV	-52.51 (0.4 GPa)	0.25	9	2	05:10
	form V	291.58 (4.0 GPa)	0.30	12	3	06:04
piracetam gentisic acid (1:1) cocrystal		-188.45 (0.0 GPa)	0.26	37	2	13:32
piracetam <i>p</i> -hydroxybenzoic acid (1:1) cocrystal		-193.22 (0.0 GPa)	0.28	7	2	09:37

^a Two pressure polymorphs of piracetam, forms IV and V, were minimized at the experimental pressures of 0.4 and 4.0 GPa, respectively. ^b Root-mean-square deviation in 15-molecule coordination sphere compared with the experimental structure.³⁶ ^c For each molecule in the asymmetric unit as shown in the model system. ^d Single Intel Xeon 5150 2.66 GHz processor using 1500 MB of memory. ^e Disordered experimental form with atoms refined over two positions with occupancies 0.657:0.343. Only the most abundant conformer used for the minimization.

Table 6. Lattice Energy Minimization of Selected Crystal Structures of Varying Size and Complexity (Refer to Table 1 for Molecular Diagrams)

model system	lattice energy (kJ mol ⁻¹)	rmsd ₁₅ ^a (Å)	# outer iterations	# LAM updates ^b	CPU time (hr:min) ^c
(R)-1-phenyl-2-(4-methylphenyl)ethylammonium-(S)-mandelate salt	-596.43	1.34	36	8	45:55
progesterone resorcinol (1:1) cocrystal	-221.92	0.15	61	13	179:11
progesterone pyrene (2:1) cocrystal	-458.81	0.72	32	3	130:53

^a Root-mean-square deviation in 15-molecule coordination sphere compared with the experimental structure.³⁶ ^b For each molecule in the asymmetric unit as shown in the model system. ^c Four Intel Xeon 5150 2.66 GHz processors with 7 GB shared memory.

Case 3 in Tables 2–4 indicates the computational cost if all the information required to construct the LAMs was already available in a LAM/QM database, which essentially removes the need to perform QM calculations during the minimization. Note that one QM geometry optimization and one charge density calculation are still necessary for the explicit evaluation of the final lattice energy to remove errors associated with the use of LAMs.

On the basis of cases 2, 4, and 5 in Tables 2–4, the computational cost of the CrystalOptimizer algorithm seems to scale linearly with the number of flexible degrees of freedom and is dominated by two elements, both of them associated with the use of finite difference approximations to partial derivatives. The first element arises from the calculation of the gradients of intermolecular energy with respect to the flexible degrees of freedom (cf. eq 20) by repeated calls to the DMACRYS package; for systems involving more than about 15 flexible degrees of freedom, this calculation accounts for more than 60% of the computational time. The second element is associated with the numerical calculation of multipole derivatives with respect to the specified flexible degrees of freedom (cf. eq 14) and can account for up to 50% of the computational time in cases involving less than 15 flexible degrees of freedom.

4.5. Application of CrystalOptimizer to Cocrystals and Salts. CrystalOptimizer is directly applicable to systems with more than one species in the asymmetric unit, including cocrystals and salts, such as those listed in Table 1. It can also take account of the effects of pressure on crystal structure, via a simple extension of the intermolecular energy (eq 18) to lattice enthalpy. Table 5 presents results for cocrystals involving α -D-glucose and piracetam, and all five known forms of piracetam, two of which (forms IV and V) are studied under elevated

pressures. Table 6 presents further results for a salt and two cocrystals of progesterone.

As in the case of the systems studied in section 4.4, agreement with the experimentally observed crystal structures is generally good, with the rms₁₅ error being below 0.30 Å for all structures other than the salt and the progesterone–pyrene 2:1 cocrystal.

It is worth noting that the reproduction of these larger experimental structures is very sensitive to the selected QM model. Due to the ionic nature of the salt, it is imperative to accurately capture and reproduce the electrostatic interactions in the crystal. If the quality of the charge density calculation is improved to the MP2/6-31G(d,p) level of theory, the rms₁₅ deviation can be reduced from 1.34 Å to 0.84 Å. Similarly, if the hybrid B3LYP/6-31G(d,p) level of theory is used to represent the intramolecular interactions and the multipole model during lattice energy minimization of the progesterone–pyrene (2:1) cocrystal, the error in rms₁₅ is reduced to 0.35 Å.

The selection of the best available computational and flexibility model is therefore integral to the correct reproduction of experimental structures. However, this aspect is beyond the scope of this work.

4.6. Application of CrystalOptimizer to ab initio Crystal Structure Prediction. The main purpose for the development of an accurate local lattice energy minimization algorithm for crystal structures containing flexible molecules is not the reminimization of the experimentally determined crystals but the final refinement of a large set of stable hypothetical structures produced during a crystal structure prediction search. It is in this context that the performance gains through the use of LAM/QM databases (cf. section 2.5) become more relevant.

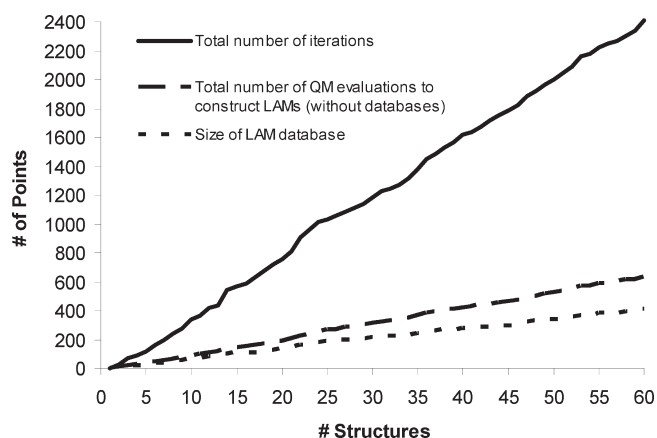


Figure 9. Total number of CrystalOptimizer minimization iterations (solid), number of LAM updates (dashed), and the actual number of QM evaluation sets necessary to construct the LAM/QM database (dotted) as a function of the number of crystal structures minimized in stage 3.

As an illustrative example, we consider the application of CrystalOptimizer to the refinement of 60 distinct crystal structures for xylitol using nine main hydroxyl and backbone torsions as the flexible degrees of freedom (cf. column 3 of Table 2). The 60 structures are typical of those that would be considered in the context of *ab initio* crystal structure prediction; in this example case, they were generated using the CrystalPredictor code³⁸ by treating xylitol as a rigid molecule.

Figure 9 shows three lines:

- The solid line shows the cumulative number of outer minimization iterations against the number of crystal structures analyzed. We can see that a total of 2412 iterations are needed, each effectively corresponding to a different molecular conformation. The computational cost would be prohibitively high if a new QM evaluation of intramolecular energy and charge density were performed at each such iteration. This effectively excludes the use of codes such as DMAFlex for this kind of application.
- The dashed line shows the number of LAMs created if no LAM/QM databases are used. This was obtained by a modified form of the CrystalOptimizer code which, like the standard form, created and used LAMs; however, if an iteration involved a molecular conformation outside the range of validity of the most recently constructed LAM, the latter was simply discarded (i.e., without any information being recorded in the LAM/QM database for possible future use) and a new one created. The total number of LAMs that need to be constructed in this case is 633. Comparing this number with the total number of iterations, we conclude that a LAM is approximately reused over a sequence of, on average, 3.8 successive iterations.
- The dotted line shows the number of LAMs created if the standard CrystalOptimizer code (including LAM/QM databases) is used. In this case, we start the refinement of the first structure using an empty LAM/QM database, and the number of entries in the database potentially increases with the consideration of each new structure. The total number of LAMs created is 408. The difference between this and the dashed line is a measure of the savings achieved via the use of a LAM/QM database.

The savings achieved via the use of a LAM/QM database would be higher if a larger number of crystal structures were refined: as the database is enlarged, there is an increasingly higher probability that a molecular conformation under consideration will be covered by one of the points that are already in the database, thereby obviating the need to perform a new set of QM calculations. This is particularly important since, in typical *ab initio* crystal structure predictions, it is not uncommon to refine many hundreds of candidate structures. These structures are often far from a local minimum, so that additional optimization iterations are needed for convergence. Thus, reducing the cost of each iteration is very important for a thorough search.

In summary, the use of LAM/QM databases is of significant benefit in the context of *ab initio* crystal structure prediction studies involving the refinement of large numbers of candidate crystal structures. It is also advantageous for the computational screening of compounds with different cocrystal formers and for crystal structure calculations under different physical conditions (e.g., a range of different pressures for identification of high-pressure polymorphs).

5. CONCLUSIONS

For the majority of organic crystal structures of practical interest, detailed modeling of molecular flexibility is essential for correct minimization of the lattice energy. In order to obtain balanced models for the intra- and intermolecular energy contributions, it is necessary to use quantum mechanical molecular deformation energies and charge densities, making lattice energy minimization computationally expensive.

An algorithm has been presented that makes it possible to take molecular flexibility into account when identifying crystal structures that minimize the lattice energy. This algorithm, CrystalOptimizer, is a local optimization method that overcomes the large computational cost of treating flexibility through the use of local approximate models (LAMs). This leads to a substantial reduction in the number of quantum mechanical calculations required. In addition, the algorithm is based on a quasi-Newton scheme that ensures rapid convergence even with systems involving many flexible degrees of freedom.

The results presented for a variety of model systems indicate that the implementation of local approximate models can significantly improve the computational efficiency without adversely affecting accuracy. In practice, during the course of an optimization, most iterations rely on previously constructed LAMs (rather than new QM calculations) for the computation of the values of the rigid degrees of freedom, the intramolecular energy and the electrostatic multipole model. Moreover, the use of LAM/QM databases further reduces the computational cost associated with QM calculations allowing the accurate energy minimizations performed by CrystalOptimizer to be coupled directly within the crystal structure prediction techniques. The low-energy structures computed with CrystalOptimizer can also be used as input to the study of kinetic effects⁴¹ and crystal growth,⁴² which play an important role in crystal formation.

Overall, with the CrystalOptimizer code, the computational cost of QM calculations is no longer the limiting factor in accurate lattice energy minimization, and this makes it possible to achieve even higher accuracy by using larger basis sets and/or post-HF levels of theory. The computational cost of the algorithm seems to scale linearly with the number of flexible degrees of freedom and is dominated by the calculation of the multipole

derivatives and of the intermolecular energy derivatives with respect to the flexible degrees of freedom. Both currently require the use of finite difference approximations to partial derivatives. Removing these computational bottlenecks will extend the practical applicability of CrystalOptimizer to even larger molecules.

On a more fundamental level, an important aspect of CrystalOptimizer is its ability to handle systems with more than one molecule or ion in the asymmetric unit, which makes it applicable to the prediction of crystal structures of cocrystals and salts. Our approach to date has been to treat each of these molecules or ions independently for the purposes of QM calculations and then to rely on intermolecular descriptions of the interactions between them. Whether or not this is the most appropriate way of handling such systems, or related ones such as hydrates, requires further investigation.

AUTHOR INFORMATION

Corresponding Author

*E-mail: c.pantelides@imperial.ac.uk.

ACKNOWLEDGMENT

We thank Professor S. L. Price for providing us access to the DMACRYS software. The financial support from the Engineering and Physical Sciences Research Council (EPSRC) under the Molecular Systems Engineering grant (EP/E016340) is gratefully acknowledged. Calculations were performed on the High Performance Computing Cluster at Imperial College London. An earlier version of the algorithm presented in this paper was outlined in Kazantsev et al.,⁴³ we are grateful to Wiley VCH for permission to reproduce some of the material from that publication.

REFERENCES

- (1) Bernstein, J. *Polymorphism in Molecular Crystals*; Clarendon Press: Oxford, U. K., 2002.
- (2) van Eijck, B. P. Ab initio crystal structure predictions for flexible hydrogen-bonded molecules. Part III. Effect of lattice vibrations. *J. Comput. Chem.* **2001**, *22* (8), 816–826.
- (3) Karamertzanis, P. G.; Raiteri, P.; Parrinello, M.; Leslie, M.; Price, S. L. The Thermal Stability of Lattice-Energy Minima of 5-Fluorouracil: Metadynamics as an Aid to Polymorph Prediction. *J. Phys. Chem. B* **2008**, *112*, 4298–4308.
- (4) Day, G. M.; Cooper, T. G.; Cruz Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A.; Schweizer, B. Significant progress in predicting the crystal structures of small organic molecules - a report on the fourth blind test. *Acta Crystallogr., Sect. B* **2009**, *65*, 107–125.
- (5) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modeling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478–8490.
- (6) Price, S. L. From crystal structure prediction to polymorph prediction: interpreting the crystal energy landscape. *Phys. Chem. Chem. Phys.* **2008**, *10* (15), 1996–2009.
- (7) Neumann, M. A.; Perrin, M. A. Energy ranking of molecular crystals using density functional theory calculations and an empirical van der Waals correction. *J. Phys. Chem. B* **2005**, *109* (32), 15531–15541.
- (8) Stone, A. J.; Alderton, M. Distributed Multipole Analysis - Methods and Applications. *Mol. Phys.* **1985**, *56* (5), 1047–1064.
- (9) Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005**, *1* (6), 1128–1132.
- (10) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *J. Phys. Chem.* **1996**, *100* (18), 7352–7360.
- (11) Day, G. M.; Motherwell, W. D. S.; Jones, W. Beyond the isotropic atom model in crystal structure prediction of rigid molecules: Atomic multipoles versus point charges. *Cryst. Growth Des.* **2005**, *5* (3), 1023–1033.
- (12) Brodersen, S.; Wilke, S.; Leusen, F. J. J.; Engel, G. A study of different approaches to the electrostatic interaction in force field methods for organic crystals. *Phys. Chem. Chem. Phys.* **2003**, *5* (21), 4923–4931.
- (13) Karamertzanis, P. G.; Price, S. L. Energy Minimization of Crystal Structures Containing Flexible Molecules. *J. Chem. Theory Comput.* **2006**, *2* (4), 1184–1199.
- (14) van Eijck, B. P.; Mooij, W. T. M.; Kroon, J. Attempted Prediction of the Crystal-Structures of 6 Monosaccharides. *Acta Crystallogr., Sect. B* **1995**, *51*, 99–103.
- (15) Mooij, W. T. M.; van Eijck, B. P.; Kroon, J. Ab initio crystal structure predictions for flexible hydrogen-bonded molecules. *J. Am. Chem. Soc.* **2000**, *122* (14), 3500–3505.
- (16) van Eijck, B. P.; Mooij, W. T. M.; Kroon, J. Ab initio crystal structure predictions for flexible hydrogen-bonded molecules. Part II. Accurate energy minimization. *J. Comput. Chem.* **2001**, *22* (8), 805–815.
- (17) Mooij, W. T. M.; van Eijck, B. P.; Kroon, J. Transferable ab initio intermolecular potentials. 2. Validation and application to crystal structure prediction. *J. Phys. Chem. A* **1999**, *103* (48), 9883–9890.
- (18) Nelder, J. A.; Mead, R. A SIMPLEX-method for function minimization. *Comput. J.* **1965**, *7* (4), 308–313.
- (19) Karamertzanis, P. G.; Kazantsev, A. V.; Issa, N.; Welch, G. W. A.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L. Can the Formation of Pharmaceutical Cocrystals Be Computationally Predicted? II. Crystal Structure Prediction. *J. Chem. Theory Comput.* **2009**, *5* (5), 1432–1448.
- (20) Polito, M.; D'Oria, E.; Maini, L.; Karamertzanis, P. G.; Grepioni, F.; Braga, D.; Price, S. L. The Crystal Structures of Chloro and Methyl Ortho-Benzic Acids and Their Co-crystal: Rationalizing Similarities and Differences. *CrystEngComm* **2008**, *10*, 1848–1854.
- (21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
- (22) Kim, H. S.; Jeffrey, G. A. The crystal structure of xylitol. *Acta Crystallogr., Sect. B* **1969**, *25*, 2607–2613.
- (23) Dennis, J. E.; Schnabel, R. B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*; Prentice-Hall: London, 1983.
- (24) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B* **2002**, *58* (3), 380–388.
- (25) Brown, G. M.; Levy, H. A. alpha-D-Glucose: further refinement based on neutron-diffraction data. *Acta Crystallogr., Sect. B* **1979**, *35*, 656–659.

- (26) Snyder, R. L.; Rosenstein, R. D. The crystal and molecular structure of the 1:1 hydrogen bond complex between [α]-D-glucose and urea. *Acta Crystallogr., Sect. B* **1971**, 27 (10), 1969–1975.
- (27) Admiraal, G.; Eikelenboom, J. C.; Vos, A. Structures of the triclinic and monoclinic modifications of (2-oxo-1-pyrrolidinyl)acetamide. *Acta Crystallogr., Sect. B* **1982**, 38 (10), 2600–2605.
- (28) Galdecki, Z.; Glowka, M. L. Crystal-Structure of Nootropic Agent, Piracetam-2-Oxopyrrolidin-1-Ylacetamide. *Pol. J. Chem.* **1983**, 57 (10–1), 1307–1312.
- (29) Louer, D.; Louer, M.; Dzyabchenko, A. V.; Agafonov, V.; Ceolin, R. Structure of a Metastable Phase of Piracetam From X-Ray-Powder Diffraction Using Atom-Atom Potential Method. *Acta Crystallogr., Sect. B* **1995**, 51 (Pt2), 182–187.
- (30) Fabbiani, F. P. A.; Allan, D. R.; Parsons, S.; Pulham, C. R. An exploration of the polymorphism of piracetam using high pressure. *CrystEngComm* **2005**, 7 (29), 179–186.
- (31) Fabbiani, F. P. A.; Allan, D. R.; David, W. I. F.; Davidson, A. J.; Lennie, A. R.; Parsons, S.; Pulham, C. R.; Warren, J. E. High-pressure studies of pharmaceuticals: An exploration of the behavior of piracetam. *Cryst. Growth Des.* **2007**, 7 (6), 1115–1124.
- (32) Vishweshwar, P.; McMahon, J. A.; Peterson, M. L.; Hickey, M. B.; Shattock, T. R.; Zaworotko, M. J. Crystal engineering of pharmaceutical co-crystals from polymorphic active pharmaceutical ingredients. *Chem. Commun.* **2005**, 36, 4601–4603.
- (33) Sakai, K.; Sakurai, K.; Nohira, H.; Tanaka, R.; Hirayama, N. Practical resolution of 1-phenyl-2-(4-methylphenyl)ethylamine using a single resolving agent controlled by the dielectric constant of the solvent. *Tetrahedron Asymmetr.* **2004**, 15, 3495–3500.
- (34) Dideberg, O.; Dupont, L.; Campsteyn, H. Structure cristalline et moléculaire du complexe 1:1 progestérone-résorcinol. *Acta Crystallogr., Sect. B* **1975**, 31 (3), 637–640.
- (35) Friscic, T.; Lancaster, R. W.; Laszlo, F.; Karamertzanis, P. G. Tunable recognition of the steroid a-face by adjacent p-electron density. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, 107 (30), 13216–13221.
- (36) Chisholm, J. A.; Motherwell, S. COMPACT: a program for identifying crystal structure similarity using distances. *J. Appl. Crystallogr.* **2005**, 38, 228–231.
- (37) Mohamed, S.; Tocher, D. A.; Vickers, M.; Karamertzanis, P. G.; Price, S. L. Salt or Cocrystal? A New Series of Crystal Structures Formed from Simple Pyridines and Carboxylic Acids. *Cryst. Growth Des.* **2009**, 9 (6), 2881–2889.
- (38) Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction - I. Rigid molecules. *J. Comput. Chem.* **2005**, 26 (3), 304–324.
- (39) Stone, A. J.; Dullweber, A.; Engkvist, O.; Fraschini, E.; Hodges, M. P.; Meredith, A. W.; Nutt, D. R.; Popelier, P. L. A.; Wales, D. J. *ORIENT*; University of Cambridge: Cambridge, U. K., 2006.
- (40) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, 68 (3), 441–451.
- (41) Santiso, E. E.; Trout, B. L. A general set of order parameters for molecular crystals. *J. Chem. Phys.* **2011**, 134, 064109.
- (42) Snyder, R. C.; Doherty, M. J. Predicting crystal growth by spiral motion. *Proc. R. Soc. London, Ser. A* **2009**, 465, 1145–1171.
- (43) Kazantsev, A. V.; Karamertzanis, P. G.; Pantelides, C. C.; Adjiman, C. S. CrystalOptimizer: An Efficient Algorithm for Lattice Energy Minimization of Organic Crystals Using Isolated-Molecule Quantum Mechanical Calculations. In *Process Systems Engineering: Vol. 6: Molecular Systems Engineering*; Adjiman, C. S., Galindo, A., Eds.; Wiley-VCH: Hamburg, Germany, 2010; pp 1–42.