This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ARTICLE TYPE

Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape

Hugh P. G. Thompson[a] **and Graeme M. Day**[b*]

The ability to anticipate the shape adopted by flexible molecules in the solid state is crucial for engineering and predicting crystal packing and, hence, properties. In this study, the conformations adopted by flexible molecules in their crystal structures are assessed in terms of their relationship to the calculated global conformational landscape. The study quantifies the limits on molecular strain that can be induced by intermolecular interactions in single-component crystal structures of molecules with no intramolecular hydrogen bonding, demonstrating that some molecules are distorted by up to 20 kJ mol^{-1} by crystal packing forces. Furthermore, we find that crystallisation often selects high energy conformers, but only when the high energy conformer is more extended than the lower energy options, allowing for greater intermolecular stabilisation. Based on these observations, we propose that the crystallisability of conformers is assessed in terms of their energies and surface areas. We formulate this as a parameterised pseudo-energy related to molecular surface area, which leads to a dramatic improvement in our ability to predict the conformations adopted by molecules in their crystal structures.

1 Introduction

Crystal engineering involves the design of crystal structures and the deliberate targeting of solid-state properties through our understanding of structure-property relationships. Successful crystal engineering relies, in part, on a knowledge of the overall shape of molecular building-blocks, as well as the relative arrangement of functional groups within a molecule that can participate in structure-directing interactions. Molecular shape can be easily predicted for rigid molecules, but becomes more challenging to anticipate as molecular flexibility is increased and the molecules of interest have a choice of conformer when self-assembling into a crystal. Conformation determines the overall molecular shape and position of functional groups, so different conformers may lead to very different crystal packing arrangements, ultimately influencing the properties of the crystal. Therefore, conformational flexibility can be seen as a potential obstacle to crystal engineering. This is particularly relevant for the engineering of pharmaceutical

materials, where solid form properties of active pharmaceutical ingredients may be manipulated, either by selection between polymorphs, or the design of multi-component crystals, such as salts or cocrystals.

Here, we address the issue of predicting which conformation will be adopted in a flexible molecule's crystal structure. We approach this problem from the point of view of computationally-guided crystal engineering and the development of methods for crystal structure prediction (CSP). CSP, as applied to organic molecules, has made substantive progress over the past decade[1–3]. The most commonly applied CSP method is based upon a global search of the lattice energy surface[4]. The main assumption is that the most likely observable crystal structures correspond to the lowest energy minima. This approach has been successfully applied to the prediction of both single-[5–8] and multi-component[6,7,9] molecular crystal structures, sometimes guiding experimental efforts towards the discovery of new solid forms.[10,11]

In terms of crystal engineering and crystal structure modelling, a rigid molecule should be defined as a molecule that has a single conformation and where that conformation is unaffected by crystal packing forces. For a rigid molecule, CSP requires an exploration of possible structures as a function of unit cell dimensions, molecular positions and orientations. Once a set of crystal structures is generated, their ranking by relative energies requires the calculation of intermolecular interactions only. Putative crystal structures are often separated

[a]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.*
[b]*School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. E-mail: G.M.Day@soton.ac.uk*

by small energy differences, and the goal of CSP has encouraged the development of accurate models for intermolecular interactions.[12]

The challenges involved with extending the capabilities of CSP methods to dealing with flexible molecules have hindered applications to pharmaceutical molecules, and are apparent in the results of the CSP blind tests[2,3], where success rates for flexible molecules are low. The two principal difficulties associated with molecular flexibility are that:

1. The flexibility of intramolecular degrees of freedom means that the molecular geometry may be distorted under the influence of intermolecular interactions in a crystal structure, to reach the optimum balance between inter- and intramolecular energies. Therefore, the impact of the crystalline environment on the molecular geometry must be modelled.

2. Conformational degrees of freedom increase the dimensionality of the energy landscape that must be explored when generating crystal structures.

The distortion of molecules away from their gas phase structures has recently been discussed in the context of conformational polymorphism.[13] Off-the-shelf force field methods for modelling intramolecular interactions do not provide a sufficiently accurate description of the associated energetic cost.[14] Therefore, the high demands of CSP for flexible molecules have required either solid-state electronic structure based approaches[15,16] or hybrid energy models, where intramolecular degrees of freedom are treated quantum mechanically and atom-atom models are used for intermolecular interactions.[17–21]

The problem of sampling conformational phase space and predicting the relevant molecular geometries is less well developed. Conformational degrees of freedom can be sampled concurrently with packing parameters, but for highly flexible molecules it can be advantageous to split the high dimensional global optimisation problem into two smaller problems: a search for relevant conformers followed by a crystal packing search for each. Apart from reducing the dimensionality of the search space, this approach leaves room for additional selection criteria when assessing which molecular conformations might lead to the most stable crystal structures.[22,23]

A purpose of this work is to study the conformational energy landscapes of a large set of flexible, pharmaceutical-like molecules, to investigate the relationship between the ensemble of possible conformations of the isolated molecules and the molecular geometries seen in their observed crystal structures. These comparisons should provide insight into conformational preferences in crystal packing and inform the future development of structure prediction methods for flexible molecules. Therefore, we have carried out a series of calculations to address the following questions:

- Which conformation will a molecule adopt, and where on the energy landscape of the isolated molecule is this conformation found?

- Apart from energy, can other molecular properties be identified that predict which conformation will be observed in crystal structures?

A starting assumption in this work is that the geometry adopted by a molecule in its crystal structure will be close to that of one of the conformers of the isolated molecule. However, in comparing the energy landscape of the isolated molecule to its crystalline geometry, we also investigate to what extent molecular geometries are distorted away from the geometry of the isolated conformer by intermolecular forces. The results will help quantify the importance of considering distortions away from idealised molecular geometries when attempting to predict the crystalline arrangement of organic molecules, in a general crystal engineering context as well as for the specific goal of *ab initio* CSP.

2 Methods

The aim of this investigation is to compare the geometries that organic molecules adopt in their crystal structures to the landscape of possible conformers of the isolated molecule. For molecules whose crystal structures have been determined, the crystalline molecular geometries are obtained from the atomic coordinates in the crystal structure. However, an exhaustive set of conformers is only available from computational methods. The methods used for conformational searches and energy calculations on the observed and calculated molecular geometries are described below.
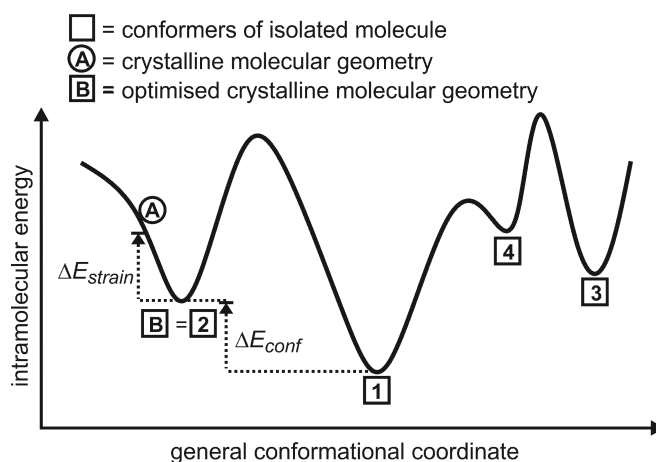
Fig. 1 Schematic of a conformational energy surface, highlighting the conformers of the isolated molecule (**1-4**), the crystalline geometry (**A**) and the conformer reached upon energy minimisation of the molecular geometry found in the crystal (**B**).

2.1 Crystal structure optimisations

The first aim is to obtain an accurate geometry and energy for all conformations of a molecule in all of its observed crystal structures. The energies of the molecular geometries observed in the known crystal structures were evaluated after lattice energy minimisation of the experimentally determined crystal structures, removing the influence of experimental errors in atomic coordinates.

For consistency, we apply the same level of theory in an atomic basis set method for crystal and molecular energy calculations: we use dispersion-corrected density functional theory (DFT-D), which has been shown to provide an accurate description of the structures and energies of molecular organic crystals[15,24,25]. All DFT-D calculations were performed using the CRYSTAL09[26,27] software, with the B3LYP hybrid functional[28,29], 6-31G* basis set, and the rescaled empirical dispersion correction suggested by Civalleri *et al.*[30] All atomic positions were relaxed during lattice energy minimisation, with unit cell dimensions constrained at experimentally determined values. All single molecule calculations were also performed in CRYSTAL09, using the same functional, basis set, DFT integration grid and optimisation thresholds as the periodic lattice energy minimisations. Full details of DFT-D calculations are provided in the Supplementary Information.

All symmetry-independent molecules were extracted from the optimised crystal structures and single-point molecular energy calculations were performed to give the energies of the molecules in their crystalline geometry (see point **A** in Figure 1). These molecular geometries were also taken as starting points for unconstrained molecular geometry optimisations, yielding the local energy minimum on the conformational energy surface associated with the crystalline molecular geometry (point **B**, Figure 1). We refer to the energy difference between the single-point and optimised molecular geometries as the intramolecular strain energy, ΔE_{strain}.

2.2 Conformational searches

We applied a low-mode conformational search (LMCS[31,32]) method, as implemented in MacroModel[33], to generate as complete and unbiased sets of conformers as possible for each molecule studied. LMCS is a mode-following algorithm - a starting molecular geometry is perturbed along one or a combination its calculated normal modes before re-minimising. Full details of the search parameters used here are available in the Supplementary Information.

All unique conformers resulting from an initial search using the OPLS2005[34] force field were re-optimised using DFT-D, applying the same computational parameters described above for the crystalline molecular geometries. The result should be a complete set of low energy conformers for each molecule (points **1-4** in Figure 1), assuming that the conformer search

was complete and that all local minima on the conformational energy surface described by DFT-D have a corresponding local minimum on the force field described energy surface. The predicted conformations were re-clustered to remove duplicates after DFT-D re-optimisation and, in most cases, additional structures were removed at this stage, suggesting that the DFT-D conformational energy surface is smoother than the OPLS surface.

Geometric comparisons were performed to locate the optimised conformations taken from the observed crystal structures in the sets of predicted conformers (e.g. point **B** = point **2** in Figure 1). We refer to the energy difference between the conformer corresponding to the crystalline geometry and the lowest energy computer-generated conformer as the relative conformational energy, ΔE_{conf} (Figure 1).
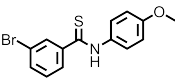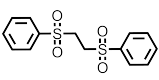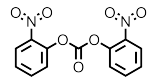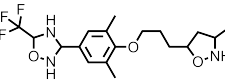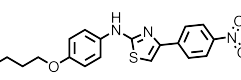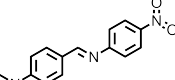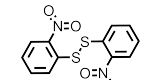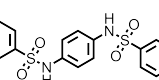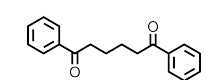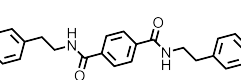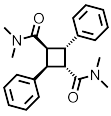
As further characterisation of the conformational energy landscape, the Connolly surface area of each DFT-D-optimised conformer was calculated using Material Studio[35] with a probe radius of 1.8 Å.

2.3 Choice of molecules

The molecules included in the study were chosen to explore a range of molecular size and flexibility. To focus on pharmaceutical-like molecules, we restricted ourselves to molecules that satisfy Lipinski's rule of five,[36] placing limits on molecular mass, number of hydrogen bond donors and acceptors and predicted octanol-water partition coefficient. Molecules with the potential for intramolecular hydrogen bonding are excluded to focus on molecules without strong, directional non-bonded intramolecular interactions. The influence of intramolecular hydrogen bonding on the conformations adopted in crystal structures is the subject of an on-going study.

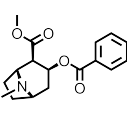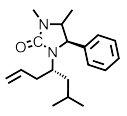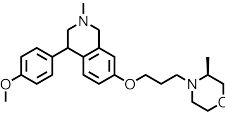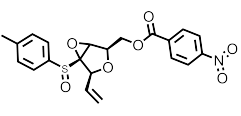Molecular flexibility was measured by the number of rotatable (exocyclic) single bonds in the molecule, excluding terminal methyl groups. Three molecules were chosen with each of 4, 5, 6, 7 and 8 such rotatable bonds, yielding a test set of 15 molecules. While their presence was not used as a criterion for the selection of molecules, the set also includes molecules with flexible ring systems, whose flexibility is sampled during the low-mode conformational search.

All chosen molecules have experimentally determined crystal structures and the molecules were selected to include diversity in known polymorphic behaviour. For each level of flexibility, one molecule was chosen with known conformational polymorphism, one with known packing polymorphism (alternative crystal packings of the same molecular conformation) and one molecule with no reported polymorphism. To maximise the data obtained from each molecule, molecules with the most reported polymorphs at each level of flexibility were selected. For molecules with the same number of

Polymorphic	Molecular Flexibility (number of exocyclic single bonds)				
type	4	5	6	7	8
Conformational	HIBGUV(3,4,2)	FAHNOR(2,2,2)	VEMTOW(2,2,2)	HAJYUN(2,2,2)	DANQEP(3,4,3)
Packing	MABZNA(4,9,1)	ODNPDS(2,2,1)	FIBKUW(2,2,1)	GALCAX(2,2,1)	CELHIL(2,2,1)
Non-polymorphic	SIKRIN(1,1,1)	COCAIN(1,1,1)	NEQNIG(1,1,1)	SEVJAF(1,1,1)	DADNUR(1,1,1)

Table 1 Chemical diagrams and CSD REFCODEs for the studied molecules. The three numbers in parentheses following each REFCODE refer to the number of polymorphs, the number of independent molecular geometries (summed over all polymorphs) and the number of unique conformers found in all polymorphs, respectively.

known polymorphs, and for the non-polymorphic molecules, the molecule with the highest quality reported crystal structure (as measured by the R factor) was chosen. To our knowledge, no solvate structures have been reported for the molecules in our set, so the study includes only single-component crystal structures. We refer to molecules by their 6-letter Cambridge Structural Database (CSD)[37] REFCODE family names which are provided, along with chemical diagrams, in Table 1.

3   Results and Discussion

3.1   Intramolecular strain energy

Any distortion in the crystal structure away from the molecular geometry of the isolated molecule must involve an increase in intramolecular energy, ΔE_{strain}, which we calculate with reference to the nearest local minimum on the isolated molecule's energy surface (see Fig. 1).

The distribution of ΔE_{strain} is summarised in Figure 2 for the 36 distinct molecular geometries in the 29 lattice energy minimised crystal structures of the 15 molecules (see Table S1 in supplementary information), including molecular geometries from different polymorphs and independent molecules within the same crystal structure (for structures with multiple molecules in the asymmetric unit).

There is a correlation, albeit weak, between ΔE_{strain} and the
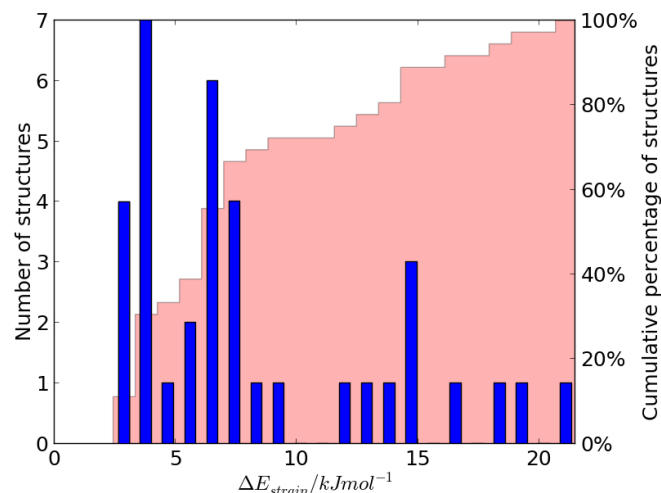
Fig. 2 Histogram (blue bars) showing the distribution of intramolecular strain energies of the 36 molecular geometries. The pink filled area shows the cumulative percentage of molecules as a function of strain energy.

geometrical distortion of the crystalline molecular geometry away from the geometry of the isolated molecule (see Figs S3 and S4 in supplementary information.). However, there are cases where a large molecular strain energy results from very small changes to stiff intramolecular degrees of freedom or

where large geometry changes come at a relatively low energetic cost. The latter typically correspond to rotation about soft dihedrals (for example, see Figure 3a).
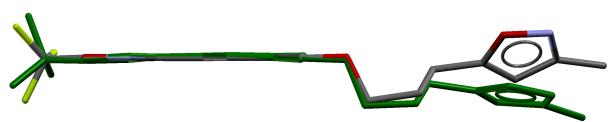
(a) HAJYUN undergoes a large distortion at low strain energy (6.7 kJ mol^{-1}), due to twisting about saturated exocyclic bonds.

(b) Large conformation changes in CELHIL enable optimised intermolecular hydrogen bonds ($\Delta E_{strain} = 19.0$ kJ mol^{-1}).

(c) Small changes in bond lengths and angles in the crystal structure GALCAX01 result in a $\Delta E_{strain} = 2.4$ kJ mol^{-1}.

Fig. 3 Overlays of crystalline molecular geometries (green) and the associated conformers of the isolated molecules (coloured by atom).

The majority (75%) of molecules are strained by less than 10 kJ mol^{-1} and it is unsurprising in these cases that the increase in intramolecular energy can be compensated by improved intermolecular interactions in the crystal. However, we find several cases of surprisingly high strain energies. The highest values calculated for the molecules studied here are $\Delta E_{strain} = 18.3$ and 21.6 kJ mol^{-1} in the two polymorphs FIBKUW01 and FIBKUW02 and $\Delta E_{strain} = 19.0$ mol^{-1} in the monoclinic polymorph of N,N'-di(phenylethyl)terephthalamide (CELHIL01, see Figure 3b). One important observation is that high intramolecular strain energies are not restricted to large molecules: two of the smallest molecules in our set show large strain in their crystal structures: SIKRIN ($\Delta E_{strain} = 14.6$ kJ mol^{-1}) and HIBGUV ($\Delta E_{strain} = 14.4$ kJ mol^{-1}, in its monoclinic beta polymorph). The results that we find here are in broad agreement with those recently reported for a large set of conformational polymorphs.[13]

At the low energy end of the ΔE_{strain} distribution, we note that none of the molecules are completely unaffected by crystal packing. The smallest calculated value is $\Delta E_{strain} = 2.4$ kJ mol^{-1} for the triclinic polymorph of 1,4-dibenzoylbutane (CSD refcode GALCAX01), corresponding to an all-atom

RMS difference in atomic positions of 0.03 Å between gas and solid phase molecular geometries (Figure 3c). From a structure prediction perspective, these results demonstrate the limitations of any rigid-molecule treatment of even moderately flexible molecules, where the molecule in the crystal structure is assumed to be undistorted from its gas phase geometry. Energy contributions of 2-3 kJ mol^{-1} to the relative stability of crystal structures can be crucial in the context of CSP, where energy separations between computer-generated crystal structures are often 1 kJ mol^{-1} or less.[4]

3.2 Which conformation is adopted in a crystal structure? Energetics.

Molecules are not restricted to their lowest energy conformer during crystallisation and the best balance of inter- and intramolecular interactions sometimes involves a higher energy conformer. We compare the conformers observed in known crystal structures with the sets of predicted conformers to investigate the frequency with which higher energy conformers are adopted in the crystal structures of flexible organic molecules.

We define the conformer associated with the crystalline molecular geometry as the structure reached after DFT-D optimisation of the molecule extracted from the crystal structure. From our initial set of 36 independent molecular geometries, we find only 21 distinct conformers because some molecular geometries (e.g. of the packing polymorphs) converge to the same conformer on the gas phase molecular energy surface.

For all molecules apart from one, we find a corresponding conformer in the sets generated by the conformation search. For the remaining molecule, SEVJAF, the original LMCS search failed to return the conformation obtained from the known crystal structure. However, the observed conformer was obtained after initiating a second, longer, search from the geometrically closest conformer in the initial set. For the purposes of this study, we assume that the initial set of SEVJAF conformers is representative of the conformational surface, despite being incompletely explored, and we proceed with the original set of SEVJAF conformers supplemented by the conformer corresponding to the observed crystalline molecular geometry.

The results are summarised in Table 2 using the quantity ΔE_{conf} (Figure 1), the energy difference between the conformer corresponding to the crystalline molecular geometry and the global minimum conformer found by the conformational search. Surprisingly, only 6 of the 15 molecules adopt the global minimum conformer in one of their known crystal structures. Two of these 6 molecules display conformational polymorphism, adopting a non-global minimum conformer in one polymorph. Therefore, a non-global energy minimum conformer is observed in the crystal structures of 11 of the

15 molecules.

It is clear from these results that a range of molecular conformers must be considered when assessing which molecular conformation will be adopted in the solid state. In the majority of cases, the crystalline geometry is energetically close to the most stable conformer of the isolated molecule: of the 21 conformers in this study, the crystalline conformation is within 5.9 kJ mol^{-1} of the global minimum in 15 cases. The remaining 6 are more interesting, ranging from ΔE_{conf} = 13.9 to 25.6 kJ mol^{-1}; the highest energy observed conformer in this set of molecules is ΔE_{conf} = 25.6 kJ mol^{-1} in crystal structure DADNUR. These values suggest that, for a flexible organic molecule, conformations up to about ΔE_{conf} = 26 kJ mol^{-1} must be considered as possibilities for what will be observed in the solid state.

Molecule	N$_{26}$	ΔE_{conf}	rank	$\Delta E_{conf,biased}$	rank
HIBGUV1	10	4.83	7	1.55	7
HIBGUV2	10	4.17	5	0.74	4
MABZNA	2	0.00	1	0.00	1
SIKRIN	1	0.00	1	0.00	1
FAHNOR1	13	4.40	5	0.00	1
FAHNOR2	13	0.00	1	0.99	2
ODNPDS	4	0.00	1	0.00	1
COCAIN	7	1.23	2	0.43	2
VEMTOW1	12	0.35	3	0.55	4
VEMTOW2	12	0.00	1	0.00	1
FIBKUW	35	5.88	10	0.00	1
NEQNIG	105	0.00	1	0.00	1
HAJYUN1	124	15.01	67	4.29	29
HAJYUN2	124	13.85	52	2.15	9
GALCAX	73	18.55	27	0.00	1
SEVJAF	623	21.25	$\geq 283^a$	3.70	$\geq 49^a$
DANQEP1	115	3.54	34	2.12	12
DANQEP2	115	2.18	18	1.82	11
DANQEP3	115	1.46	12	0.58	7
CELHIL	142	22.11	108	4.81	37
DADNUR	30	25.54	28	4.87	5

Table 2 Energies (in kJ mol^{-1}) and ranks of the conformers in the observed crystal structures, based on the calculated DFT-D energy (ΔE_{conf}) and the biased energy ($\Delta E_{conf,biased}$, see section 3.4). ΔE is the energy relative to the global minimum conformer. N$_{26}$ is the number of conformers found within 26 kJ mol^{-1} for each molecule. ^{a} Rankings and ΔE may be larger than reported for SEVJAF, due to an incomplete sampling of conformations (see text).

These results are discouraging for structure prediction and crystal engineering: such an energy window can encompass large numbers of conformers (Table 2), whose shapes and arrangement of functional groups can vary greatly. Using the molecular energy as a sole selection criterion, any conformer within this range should be considered a possibility for what will be observed in a molecule's crystal structure. This raises the question of whether non-energetic descriptors of molecular conformers can be found that highlight those conformers that are most likely to form low energy crystal structures, or filter out those that will not lead to stable crystal structures.

Molecular symmetry may be beneficial for close packing and it has been shown that molecules that adopt a centrosymmetric geometry in their crystal structure almost exclusively crystallise with molecules on crystallographic special positions[38,39]. However, it is not clear what role symmetry plays in determining the conformer that a molecule adopts. Here, we observe that, of the six molecules with potential centrosymmetry (FAHNOR, ODNPDS, FIBKUW, GALCAX, CELHIL SIKRIN), five adopt centrosymmetric or nearly centrosymmetric conformations in their crystal structures. For SIKRIN and one polymorph of FAHNOR, the observed centrosymmetric conformer is the global minimum conformer. In all other cases, the global minimum energy conformer lacks centrosymmetry. FIBKUW adopts its lowest energy centrosymmetric conformer, while CELHIL and GALCAX adopt centrosymmetric conformers that are high in energy, in favour of alternative, lower energy centrosymmetric and non-symmetric conformers. The sample is too small to judge the overall importance of centrosymmetry in conformer selection, and further studies focussed on symmetry could be interesting.

3.3 Molecular surface area as a non-energetic criterion for conformer selection

One consequence of molecular flexibility for many of the molecules studied here is the possibility of both compact and extended conformers. The lowest energy conformers tend to be relatively compact, stabilising the conformer by bringing structurally remote atoms into non-bonded contact (e.g. Figure 4a). Extended, open conformations lack this intramolecular stabilisation, while making more molecular surface area available for intermolecular interactions in the solid state. The observed conformer in crystal structures must represent a balance between these intra- and intermolecular non-bonded interactions.

(a) **(b)**
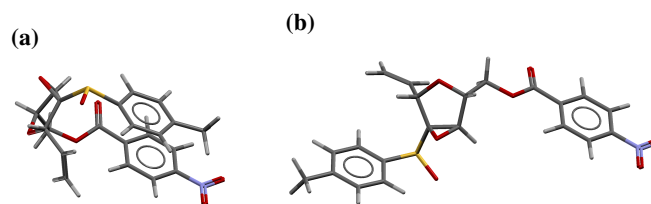
Fig. 4 The (a) lowest energy conformer and (b) conformer found in the crystal structure of DADNUR.

Indeed, in analysing the sets of observed and computer-generated conformers, we find that the conformer that is found

in a molecule's crystal structure tends to have a relatively extended geometry (e.g. Fig. 4b). Similar observations have been made regarding a preference for extended conformations of ligands bound to proteins[40]. The potential for a molecule to form close intermolecular interactions is limited, at a basic level, by its accessible surface area - the surface area that can be brought into contact with an adjacent molecule. Therefore, to quantitatively compare conformers, we calculated the Connolly surface area[41] of all predicted conformers of each molecule, providing the contact surface mapped out by a spherical probe of rolling across the van der Waals surface of the molecule. We used a probe radius of 1.8 Å, the approximate radius of a methyl group, to provide a measure of the molecular surface area that is accessible for close intermolecular contact with other copies of itself in a crystal structure.

The results for three of the most flexible molecules are summarised in Figure 5, as plots of relative conformational energy against Connolly surface area. It is clear from these results that conformers with large surface areas are favoured in the crystal structures of flexible molecules. The molecules summarised in Figure 5 adopt conformers with low (DANQEP, Fig. 5a), middling (HAJYUN, Fig. 5b) and large (DADNUR, Fig. 5c) ΔE_{conf}, while in each case the crystalline conformers are amongst those with the largest possible accessible surface areas.

This observation is not limited to these three molecules; for all 15 molecules, a high energy conformer is adopted in the crystal structure only if it has a larger surface area than that of the global minimum energy conformer (surface area vs energy plots for all remaining molecules are available in the supplementary information, Figs. S5-S9). On the basis of these observations, it appears that compact high-energy conformations are unlikely to occur in crystal structures.

3.4 Relating intermolecular interaction energies to molecular surface area

To further explore the notion that accessible surface area is an indicator of the stabilisation that can be achieved by intermolecular interactions in the crystal, we examined the variation in measured sublimation enthalpies of a series of small, rigid organic molecules as a function of their calculated surface areas. We believe that increased stabilisation due to higher surface area is dominated by dispersion interactions. To focus on dispersion contributions, rather than more specific electrostatic contributions to lattice energies, we limited this study to hydrocarbon crystal structures: details of the molecules and their measured sublimation enthalpies can be found in Table S2 in the supplementary information.

The relationship between sublimation enthalpy and molecular surface area is remarkably linear (Figure 6) with very little scatter about a linear regression with a gradient of m_{EvsSA}

(a) DANQEP
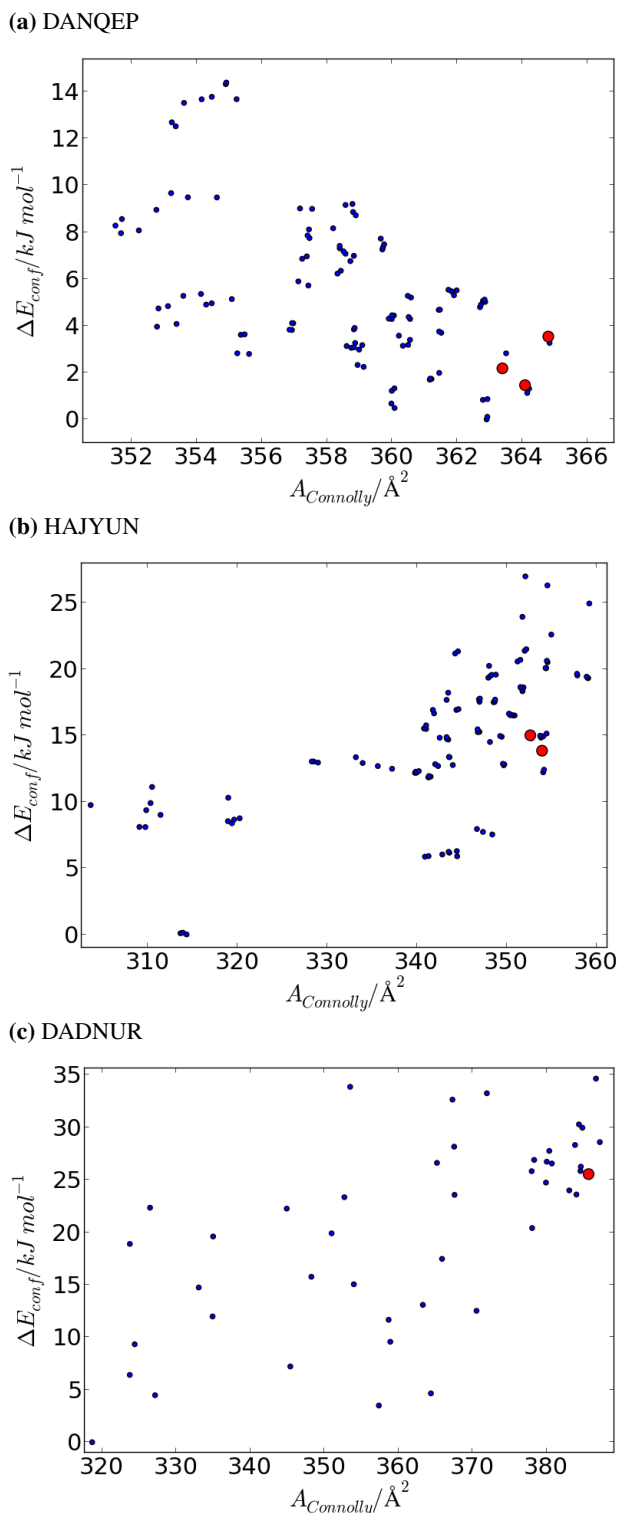
(b) HAJYUN

(c) DADNUR

Fig. 5 Plots of ΔE_{conf} against $A_{connolly}$ for all predicted conformers (blue) of (a) DANQEP, (b) HAJYUN and (c) DADNUR. The conformers corresponding to the observed crystal structures are highlighted in red.
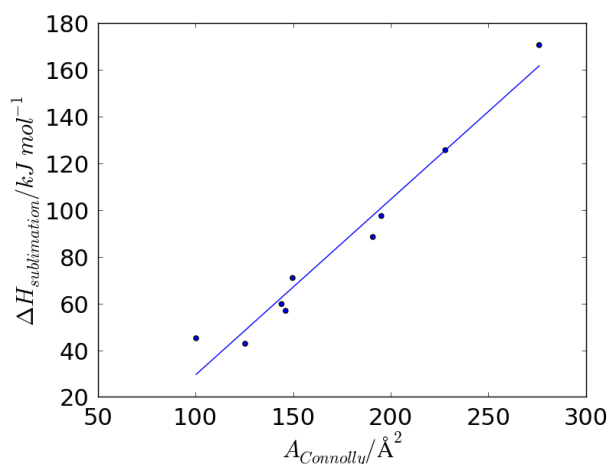
Fig. 6 Variation in measured $\Delta H_{sublimation}$ with molecular $A_{Connolly}$ for a set of small rigid hydrocarbon crystal structures.

Fig. 7 Plot of ΔE_{conf} against relative Connolly surface area, $\Delta A_{Connolly}$ for all predicted conformers of the 15 molecules. Conformers that are observed in the known crystal structures are highlighted in red. The diagonal line represents the gradient of energy vs surface area derived from Figure 6.

= 0.75 kJ mol^{-1}Å^{-2}. This suggests that the increase in intermolecular contributions to lattice energy with increasing molecular surface area should be predictable, at least in the absence of strong, specific intermolecular interactions.

Figure 7 summarises the energies and surface areas for all conformers of the 15 molecules in our test set; each conformer is represented by its relative energy and relative Connolly surface area, calculated with respect to the energy and surface area of the lowest energy conformer of the same molecule. The gradient, m_{EvsSA} derived from the sublimation enthalpies of rigid hydrocarbons is superimposed on the energy/surface area distribution. We note that all conformers found in the observed crystal structures (highlighted in red, Fig. 7) fall below and to the right of the m_{EvsSA} line, where the expected intermolecular stabilisation due to greater surface area outweighs the higher intramolecular energy than the global minimum conformer.
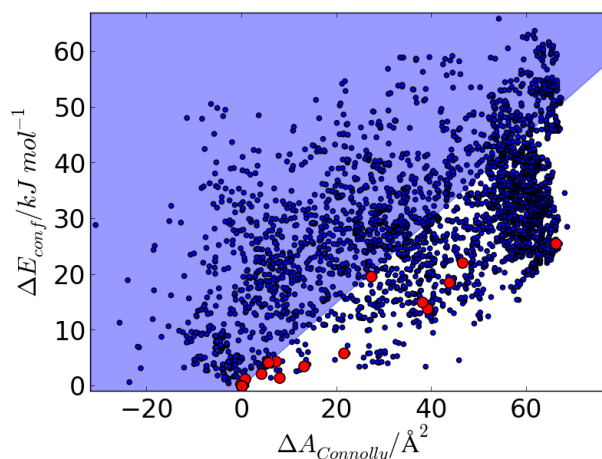
The calculated molecular surface area seems to be predictive of what conformers will be found in crystal structures and we propose that conformers falling in the shaded area (Fig. 7) - compact, high energy conformers whose higher intramolecular energy is not compensated by their expected intermolecular stabilisation - are unlikely to be observed.

We propose that this observation can be formulated as a biasing term added to the calculated conformational energy:

$$\Delta E_{conf,biased} = \Delta E_{conf} + \Delta E_{pseudo,inter} \quad (1)$$
$$= \Delta E_{DFT-D} + m_{EvsSA}\Delta A_{Connolly} \quad (2)$$

where the DFT-D calculated intramolecular energies of the isolated conformers are adjusted by a pseudo-energetic contribution, $\Delta E_{pseudo,inter}$, which predicts the relative intermolecular interaction energy of conformers in their potential crystal structures from their relative accessible surface areas.

3.5   Re-ranking of predicted conformers

Comparing the distributions of all predicted conformers distributed according to the bare DFT-D energy (Fig. 8a) and $E_{conf,biased}$ (Fig. 8b), we find that the surface area correction term does not dramatically perturb the total distribution of energies. $E_{pseudo,inter}$ has little effect on the total range in energies, while skewing the distribution of structures slightly towards lower relative energies.

Most importantly, the location within the overall ensemble of the conformers that are adopted in the crystal structures (red bars in Figure 8) is changed dramatically. The application of the $E_{pseudo,inter}$ correction leads to a significant enrichment of the observed conformers in the low energy region of the distribution. This is evident from a comparison of the cumulative distribution functions in Figure 8: all crystalline conformers are found within the first 7.1% of the overall distribution of $E_{conf,biased}$, while they range over a much greater proportion (39.5%) of the overall distribution of pure E_{conf}.

The improvement is also reflected in the ranking of the observed conformers for each individual molecule (Table 2). The ranking of all of the high energy conformers is dramatically improved after applying the pseudo-intermolecular energy correction. Furthermore, the energy difference, $\Delta E_{conf,biased}$ from the global minimum is decreased in almost every case.

Practically, this demonstrates that the combined consideration of calculated intramolecular energy and molecular surface area can narrow the ensemble of conformers that must be considered as likely candidates for what will be found in

(a) Conformational energy

(b) Biased conformational energy
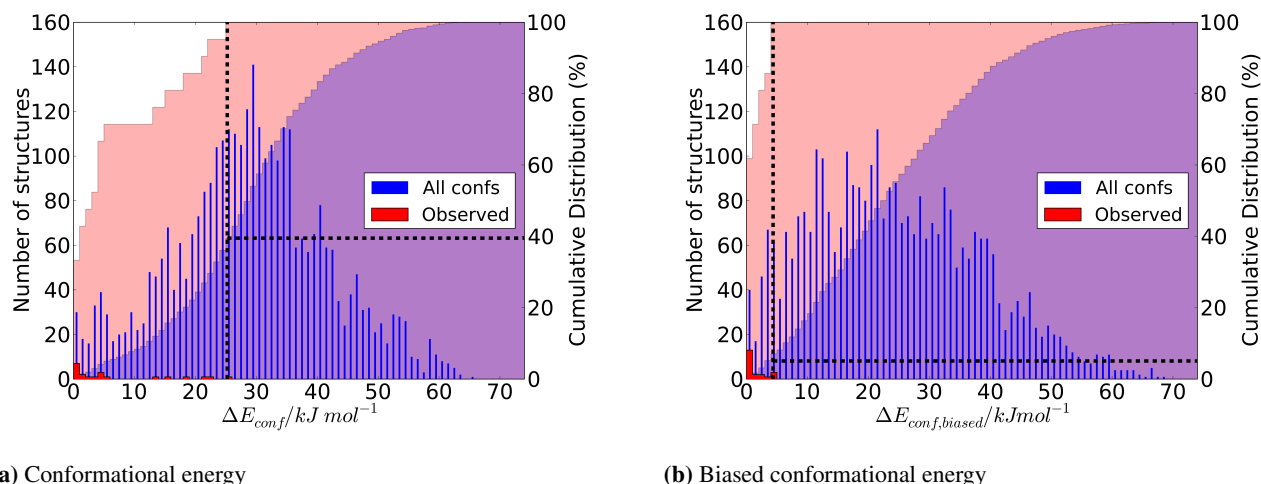
Fig. 8 Distributions of all predicted conformers (blue) and the observed crystalline conformations (red) for all 15 molecules, according to conformational (DFT-D) and biased (DFT-D + $\Delta E_{pseudo,inter}$) energies. The histogram shows the distribution of conformers and the shaded region shows the cumulative distribution. The dashed lines indicate the energy below which all observed conformers are found, and the proportion of the total conformations that are found below this energy.

crystal structures. Specifically, the application of the surface area pseudo-energy correction that we propose here means that the conformers that are seen in observed crystal structures are found much nearer the top of ranked lists of possible conformers than when ranking is based on the calculated energy of the isolated molecule.

In terms of applications to CSP, this extends the size of molecule that can be practically studied, by focussing computing effort on the smaller set of conformers that are able to achieve the best balance of inter- and intramolecular energies in the solid state. In the more general context of crystal engineering, the combined criteria based on calculated energies and molecular surface area can help anticipate the molecular shape and spatial arrangement of functional groups on a molecule, which determines the mutual arrangement of molecules in the solid state, and therefore many materials properties of interest.

3.6 Conclusions

One important conclusion from this study is that flexible molecules frequently do not adopt their lowest energy (global minimum) conformer in their crystal structures, but often assume a higher energy conformer to optimise the balance of inter- and intramolecular interactions. The observed crystalline conformer is usually energetically close to the global minimum, but in several cases molecules are found to adopt high energy conformers, up to 25 kJ mol^{-1} above the most stable gas phase conformer. As a general rule, high energy conformers are only adopted when there is an associated increase in accessible surface area, increasing the molecule's potential

to form stabilising intermolecular interactions.

We propose that the relative conformational stabilisation due to intermolecular interactions in a crystal can be estimated from an empirical linear relationship with accessible molecular surface area. This relationship has been parameterised from the variation of measured sublimation enthalpies with molecular surface area for a series of rigid hydrocarbons. By adding this pseudo-energy to the calculated intramolecular energies of conformers, we achieve a dramatic enrichment in the ranking of observed conformers in the low energy range of the global conformational landscape.

A further observation is that, for any molecule with moderate flexibility, there is some molecular distortion away from the gas phase geometry due to intermolecular interactions in a crystal structure: ΔH_{strain} is never zero and can be as large as about 20 kJ mol^{-1} for some molecules. This is important for anticipating overall molecular shape when considering how a molecule might pack in its crystal structures. Molecular geometries can be significantly distorted by the close-packed environment in a crystal structure.

The findings are clearly relevant to CSP, where conformational diversity can quickly become the bottleneck in fully exploring structural space for large, flexible molecules. Any narrowing down of the conformational space that must be considered improves the efficiency of global structure searching, reduces the resources that are required to perform predictions for a given molecule and effectively increases range of these methods. The observations made here are also relevant in more traditional crystal engineering approaches, where the geometrical attributes of a molecule, particularly the arrange-

ment and relative orientation of functional groups, must be anticipated in order to use their interactions to direct the arrangement of molecules in the solid state.

4 Acknowledgements

References

1 G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt and P. Verwer, *Acta crystallographica. Section B, Structural science*, 2005, **61**, 511–27.

2 G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. van de Streek, A. K. Wolf and B. Schweizer, *Acta crystallographica. Section B, Structural science*, 2009, **65**, 107–25.

3 D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, S. L. Price, H. a. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti and I. K. Zhitkov, *Acta crystallographica. Section B, Structural science*, 2011, **67**, 535–51.

4 G. M. Day, *Crystallography Reviews*, 2011, **17**, 3–52.

5 C. H. Görbitz, B. r. Dalhus and G. M. Day, *Physical chemistry chemical physics : PCCP*, 2010, **12**, 8466–77.

6 H. C. S. Chan, J. Kendrick, M. A. Neumann and F. J. J. Leusen, *CrystEngComm*, 2013, **15**, 3799.

7 P. G. Karamertzanis, A. V. Kazantsev, N. Issa, G. W. Welch, C. S. Adjiman, C. C. Pantelides and S. L. Price, *Journal of Chemical Theory and Computation*, 2009, **5**, 1432–1448.

8 A. J. Cruz-Cabeza, S. Karki, L. Fábián, T. Friscić, G. M. Day and W. Jones, *Chemical communications (Cambridge, England)*, 2010, **46**, 2224–6.

9 A. J. Cruz-Cabeza, G. M. Day and W. Jones, *Chem. Euro. J.*, 2009, **15**, 13033–13040.

10 J.-B. Arlin, S. L. Price and A. J. Florence, *Chemical communications (Cambridge, England)*, 2011, **47**, 7074–6.

11 D.-K. Bučar, G. M. Day, I. Halasz, G. G. Z. Zhang, J. R. G. Sander, D. G. Reid, L. R. MacGillivray, M. J. Duer and W. Jones, *Chemical Science*, 2013, **4**, 4417.

12 S. L. Price, *CrystEngComm*, 2004, **6**, 344.

13 A. J. Cruz-Cabeza and J. Bernstein, *Chemical reviews*, 2014, **114**, 2170–91.

14 G. M. Day, W. D. S Motherwell and W. Jones, *Physical chemistry chemical physics : PCCP*, 2007, **9**, 1693–704.

15 M. A. Neumann and M.-A. Perrin, *The journal of physical chemistry. B*, 2005, **109**, 15531–41.

16 M. D. King, T. N. Blanton, S. T. Misture and T. M. Korter, *Crystal Growth & Design*, 2011, **11**, 5733–5740.

17 T. G. Cooper, K. E. Hejczyk, W. Jones and G. M. Day, *Journal of Chemical Theory and Computation*, 2008, **4**, 1795–1805.

18 G. M. Day and T. G. Cooper, *CrystEngComm*, 2010, **12**, 2443.

19 A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman and C. C. Pantelides, *Journal of Chemical Theory and Computation*, 2011, **7**, 1998–2016.

20 M. Vasileiadis, A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman and C. C. Pantelides, *Acta crystallographica. Section B, Structural science*, 2012, **68**, 677–85.

21 M. Baias, C. M. Widdifield, J.-N. Dumez, H. P. G. Thompson, T. G. Cooper, E. Salager, S. Bassil, R. S. Stein, A. Lesage, G. M. Day and L. Emsley, *Physical chemistry chemical physics : PCCP*, 2013, **15**, 8069–80.

22 T. G. Cooper, W. Jones, W. D. S. Motherwell and G. M. Day, *CrystEngComm*, 2007, **9**, 595.

23 A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, C. C. Pantelides, S. L. Price, P. T. A. Galek, G. M. Day and A. J. Cruz-Cabeza, *International journal of pharmaceutics*, 2011, **418**, 168–78.

24 J. van de Streek and M. A. Neumann, *Acta crystallograph-*

Chemical Science Accepted Manuscript

ica. Section B, Structural science, 2010, **66**, 544–58.

25 C. M. Zicovich-Wilson, B. Kirtman, B. Civalleri and A. Ramírez-Solís, *Physical chemistry chemical physics : PCCP*, 2010, **12**, 3289–93.

26 R. Dovesi, R. Orlando, B. Civalleri, C. Roetti, V. R. Saunders and C. M. Zicovich-Wilson, *Zeitschrift für Kristallographie*, 2005, **220**, 571–573.

27 R. Dovesi, V. R. Saunders, C. Roetti, R. Orlando, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, M. Llunell, C. Science and A. Technologies, *Crystal09 2.0.1*, 2013.

28 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *The Journal of Physical Chemistry*, 1994, **98**, 11623–11627.

29 A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648.

30 B. Civalleri, C. M. Zicovich-Wilson, L. Valenzano and P. Ugliengo, *CrystEngComm*, 2008, 1–6.

31 I. Kolossváry and W. C. Guida, *Biopolymers*, 1996, **7863**, 5011–5019.

32 I. Kolossváry and W. C. Guida, *J. Comp. Chem.*, 1999, **20**, 1671–1684.

33 . Schrodinger LLC, New York, NY, *MacroModel, V9.0*, 2011.

34 W. L. Jorgensen and J. Tirado-Rives, *Journal of the American Chemical Society*, 1988, **110**, 1657–1666.

35 *Materials Studio v6.10*, 2012.

36 C. A. Lipinski, *Drug Discovery Today: Technologies*, 2004, **1**, 337–341.

37 F. H. Allen, *Acta crystallographica. Section B, Structural science*, 2002, **58**, 380–388.

38 E. Pidcock, W. D. S. Motherwell and J. C. Cole, *Acta Crystallographica Section B Structural Science*, 2003, **59**, 634–640.

39 A. D. Bond, *CrystEngComm*, 2010, **12**, 2492.

40 E. Perola and P. S. Charifson, *J. Med. Chem.*, 2004, **47**, 2499–510.

41 M. Connolly, *Journal of Applied Crystallography*, 1983, **16**, 548–558.