Reducing Crystal Structure Overprediction of Ibuprofen with Large Scale Molecular Dynamics Simulations†

Nicholas F. Francia,[a] Louise S. Price[b] and Matteo Salvalaglio[a]

The control of the crystal form is a central issue in the pharmaceutical industry. The identification of putative polymorphs through Crystal Structure Prediction (CSP) methods is based on lattice energy calculations, which are known to significantly over-predict the number of plausible crystal structures. A valuable tool to reduce overprediction is to employ physics-based, dynamic simulations to coalesce lattice energy minima separated by small barriers into a smaller number of more stable geometries once thermal effects are introduced. Molecular dynamics simulations and enhanced sampling methods can be employed in this context to simulate crystal structures at finite temperature and pressure. Here we demonstrate the applicability of approaches based on molecular dynamics to systematically process realistic CSP datasets containing several hundreds of crystal structures. The system investigated is ibuprofen, a conformationally flexible active pharmaceutical ingredient that crystallises both in enantiopure forms and as a racemic mixture. By introducing a hierarchical approach in the analysis of finite-temperature supercell configurations, we can post-process a dataset of 555 crystal structures, identifying 65% of the initial structures as labile, while maintaining all the experimentally known crystal structures in the final, reduced set. Moreover, the extensive nature of the initial dataset allows one to gain quantitative insight into the persistence and the propensity to transform of crystal structures containing common hydrogen-bonded intermolecular interaction motifs.

Introduction

Computational Crystal Structure Prediction (CSP) methods rely on lattice energy calculations to identify and rank putative polymorphic structures. In the final stages of state-of-the-art CSP workflows, lattice energy rankings are refined by performing expensive calculations involving high quality, periodic electronic structure calculations and introducing entropic effects through free energy calculations.[1–5] While the quality of the methods employed in the final refinement and ranking stage is constantly improving, their computational cost is typically prohibitive and approaches to achieve a rational reduction of the number of putative structures predicted by lattice-energy based methods (CSP_0) are needed.[4,6] Both in industry and academia, CSP methods are becoming increasingly popular given their success in predicting experimental crystal forms starting from only the molecule geometry.[4,6–10] For the vast majority of these methods, the different crystal packings are generated ignoring thermal motion and assuming that the lattice energies are a reasonable approximation of the thermodynamic stability of experimental forms. These approaches, identified as CSP_0,[11] have been widely used in the 6th CCDC CSP blind test,[12] successfully reproducing the experimental forms. A limitation in using local minima of the lattice energy landscape as candidate polymorph structures is that the lattice energy landscape is rugged, and its local minima grossly outnumber the experimentally known polymorphs. This limitation makes it impossible to distinguish possible new polymorphs from artefacts of the CSP_0 static models. In fact, those states characterised by small barriers should coalesce to significantly more stable structures or melt at finite temperature and pressure.[1,11,13] Furthermore, it has been shown that the ensemble of configurations accessible corresponding to a single polymorph at ambient temperature can correspond to multiple lattice energy minima.[14]

By introducing finite temperature and pressure effects, we have recently proposed a workflow able to reduce the number of putative polymorphs drastically. The procedure described in Ref. 15, applied to the cases of urea and succinic acid, consists of:

- Molecular dynamics (MD) simulations to equilibrate all structures at 300 K and 1 bar
- Automatically identify those states that are unstable or cluster those that belong to the same dynamic ensembles using structural probabilistic fingerprints
- Perform enhanced sampling simulations on the cluster centres to overcome MD limits and assess their stability

In this work, we apply the approach introduced in Ref. 15 to a dataset of 555 crystal structures of ibuprofen: an application of

[a] *Thomas Young Centre and Department of Chemical Engineering, University College Londont London WC1E 7JE, UK. E-mail: m.salvalaglio@ucl.ac.uk*

[b] *Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK.*

† Electronic Supplementary Information (ESI) available: Tables of the potential energies of structures at different steps.

the size and complexity typical of modern CSP studies. Ibuprofen is a conformationally flexible, chiral molecule possessing two enantiomers with different pharmacological properties. *S*-Ibuprofen is the biologically active one while *R*-Ibuprofen needs to be transformed in the body to its *S*-counterpart.[16,17] This popular pain-relieving API is commonly available in its racemic Form I.[18] A more expensive enantiopure form, here labelled Form E, contains only the S-Ibuprofen.[19,20] More recently, a second less-stable racemic Form II was observed in a differential scanning calorimetry experiment.[21–23]

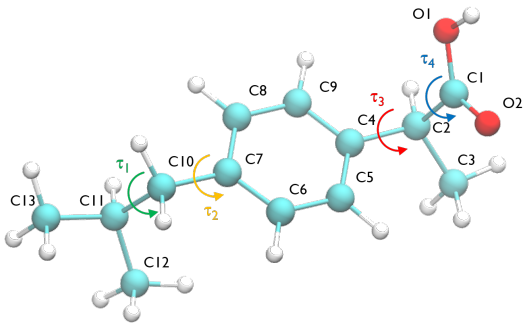

Figure 1: The Ibuprofen molecule showing atom labels and the four torsional angles considered in the crystal structure generation, namely τ_1 (C12-C11-C10-C7), τ_2 (C11-C10-C7-C8), τ_3 (C5-C4-C2-C1) and τ_4 (O1-C1-C2-C4).

Here, in addition to demonstrating the method's applicability to a dataset one order of magnitude larger than previously attempted, we describe new tools implemented in the analysis to handle sets of more than 500 CSP_0-generated structures. In particular, we introduced a fast molecule-dependent classification to reduce the time needed to compare crystal structures and cluster equivalent geometries. This improvement results in a rapid clustering analysis, which was repeatedly deployed at regular steps during biased simulations to systematically detect structural transitions without following molecular trajectories one at a time, an impractical task when dealing with several hundreds of finite-temperature dynamic simulations. Moreover, the application of the simulation workflow to a large dataset of mostly hydrogen-bonded CSP_0 crystal packings has allowed us to quantitatively analyse the emergence conformational and orientational disorder at finite temperature, and to assess the persistence of H-bond interaction motifs. Finally, we investigate the dependence of the unsupervised clustering used to identify analogous structures on the choice of collective descriptors at the basis of the probabilistic fingerprints used to define a dissimilarity metric between finite temperature crystal supercells. With the improvements in efficiency introduced in this work, the reduction workflow introduced in Ref. 15 can be deployed efficiently to reduce CSP_0 sets of the size and complexity approaching those of real-world applications.

Methods

CSP_0 lattice energy landscape. The ibuprofen search was carried out using CrystalPredictor1.9,[24] which allows the molecule to assume different conformations. In particular, we considered the two torsional groups of angles (τ_1, τ_2) and (τ_3, τ_4), shown in Fig. 1, that were separately varied from 0 to 360 degrees in 20 degree steps. The search used molecular fragments taken from the ab initio optimised molecule at the PBE0/6-31G(d,p) level of theory. The fixed point charges used in this initial step are obtained using the larger basis set aug-cc-pVDZ. The parameters from the FIT potential[25] with polar hydrogens[26] were used for the repulsion-dispersion contributions to the energy. The search was performed in 59 space groups with one molecule in the asymmetric unit cell. After removing the duplicates, the structures are labelled as their rank order at this stage, using the prefix R for racemic and E for enantiopure depending on their symmetry.

The resulting unique structures were then optimised with DFTB3-D3[27] to relax atomic positions and remove the possible unfeasible geometries derived from the use of rigid fragments of the molecule. The accurate evaluation of the lattice energies were performed with a single step DMACRYS[28] calculation, using distributed multipoles obtained from the PBE0/aug-cc-pVDZ charge density with GDMA2.2[29] and the repulsion-dispersion potential described in the previous paragraph.[25,26]

CrystalOptimizer2.4.7.1[30] was used to finally optimise the 555 crystal structures within 10 kJ mol^{-1} of the global minimum. Both the crystal structure and molecular conformation are optimised in a two-level method, with the intramolecular energies and hessian evaluated at the PBE0/6-31G(d,p) level of theory and the intermolecular energy calculated from the distributed multipoles (extracted from the charge density at the PBE0/aug-cc-pVDZ level of theory) and the repulsion-dispersion parameters described above. The smaller 6-31G(d,p) basis set was found to accurately assess the conformation of the molecule, but it was not sufficient to model the electrostatic forces of molecules in the experimentally observed crystal structures, thus justifying the use of the augmented basis set.

Structure Preparation and Atom Typing. The General Amber Force Field[31] has been used to describe the ibuprofen molecule. Atom types are assigned with the AmberTools suite[32] while point charges are assigned with the AM1-BCC model.[33] Simulations are performed with the Gromacs MD package.[34,35] This requires atoms coordinate files to be written in the order specified in the forcefield. All atoms in crystals must then be rearranged to match the forcefield index. This is done by transforming molecules in graphs and applying the VF2 graph match algorithms[36] available in the Python library NetworkX.[37] Finally, in order to see possible transitions or formation of orientational disorder in a relatively small computational time, for each crystal we generated a supercell of at least 200 molecules. The simulation boxes are chosen to have a nearly cubic shape with each cell edge around 4.5 nm.

Energy minimisation. We optimised the atoms positions using the steepest descent algorithm. The neighbour lists are updated every 10 steps using the Verlet cutoff scheme. Electrostatic and van der Waals interactions are calculated using a cutoff of 1.0 nm while long-range interactions are treated with the smooth Particle Mesh Ewald (PME)[38] and Lennard-Jones PME. After a first atoms' position optimisation, we used LAMMPS to relax the cell

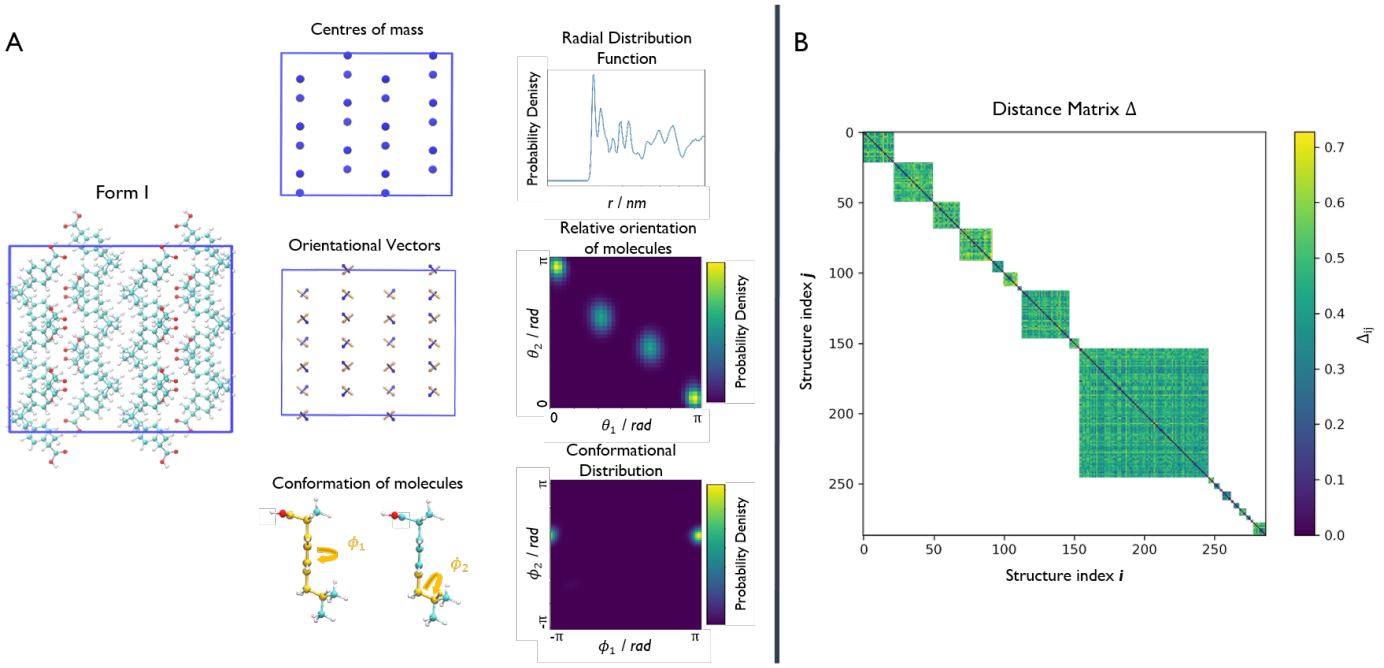

Figure 2: Clustering of the finite-temperature structures. (A) For each structure, we identify a set of structural fingerprints able to distinguish the different geometries. For the relative position of molecules, we calculate the radial distribution function of the centres of mass. For the relative orientation, we define two sets of vectors connecting atoms C6-C8 and C7-C4 of each molecules, and calculate the angles, θ_1 and θ_2, between them. The possible conformations are detected by looking at the C1-C2-C10-C11 and C7-C10-C11-H11 torsional angles, here labelled with ϕ_1 and ϕ_2. The resulting distributions form the fingerprint of each structure. In the interest of simplicity, a supercell of 48 molecules is shown here but typical simulation boxes contain more than 200 molecules. (B) The similarity between each pair of structures is given by the norm of the Hellinger distances between distributions. This is calculated only between structures that share the same chirality and molecules conformation resulting in a distance matrix that avoids negligible comparisons and saves computational time.

parameters (feature not available in Gromacs), using InterMol[39] to convert the molecular forcefield. A second energy minimisation with Gromacs is performed to take into account differences between the two packages[39]. Finally, the GAFF lattice energies are estimated with the equation:

$$E^{latt} = \frac{E^{crystal}}{n_{mols}} - E^{vacuum} \quad (1)$$

where E^{vacuum} is the energy of the isolated molecule and n_{mols} is the number of molecules in the supercell.

Equilibration at finite temperature and pressure. We performed a 3 ns simulation in the canonical ensemble at 300 K, followed by 4 ns in the isothermal-isobaric ensemble at 300 K and 1 bar. We used the velocity Verlet integrator with a 1 fs timestep. We controlled the temperature with the Bussi-Donadio-Parrinello thermostat[40] and equilibrated the systems at 1 bar for the first 1 ns using the Berendsen anisotropic barostat[41] and then switched to the Parrinello-Rahman barostat[42] for the following 3 ns.

Probabilistic Fingerprints. Effective descriptors of the different geometries generated should be able to handle the displacement of atomic positions from equilibrium in finite-temperature simulations and efficiently capture the differences between crystal packings. In this context, we previously proposed[15] a set of system-dependent probability densities that describes the relative position, relative orientation and possible conformations, $F_i = \{p_i(r_{COM}),\ p_i(\vec{\theta}),\ p_i(\vec{\phi})\}$, as the fingerprint of each crystal when dealing with flexible molecules. PLUMED 2[43] has been extensively used to analyse trajectories and generate distributions. An example of the inputs used to generate the components of the structural fingerprints described here are available on PLUMED-NEST, the public repository of the PLUMED consortium[44], as plumID:21.019.

In the case of ibuprofen, the term $p_i(r_{COM})$ represents the radial distribution function of centres of mass of molecules in the i^{th} crystal structure. The relative orientation of molecules in the i^{th} crystal structure is described by the 2D probability density distribution $p_i(\vec{\theta})$, a function of the intermolecular angles θ_1 and θ_2, obtained from two orthogonal vectors connecting the atoms C6-C8 and C7-C4 of the aromatic ring of the molecule, as shown in Fig. 2A. Finally the conformational contribution to F_i, $p_i(\vec{\phi})$, was defined following the conformational analysis reported in Ref. 45, which employs the 2D distribution of the *global* (ϕ_1) and *local* (ϕ_2) torsional angles shown in Fig. 2A. The former represents the relative orientation of the two substituents of the aromatic ring while the latter captures the relative position of the methyl groups. In this approximation, molecules can adopt six possible conformational states. In order to assess the generality of the choice of relatively coarse conformational descriptors, we compared it with an alternative, more fine-grained representation, making use of two 2D distributions ($p_i(\tau_1, \tau_2)$ and $p_i(\tau_3, \tau_4)$, Fig. 1). The two different approaches, despite the difference in level of detail, and of the associated computational cost, lead to similar results. Finally, the probabilistic fingerprints are complemented by an additional parameter used to classify structures based on their chirality.

Structural (dis)similarity and Clustering The similarity between two fingerprints, F_i and F_j, is quantitatively determined by computing the Hellinger distance, H_{ij}, between each equivalent distribution, defined as:

$$H_{ij} = \sqrt{1 - BC(p_i, p_j)} \tag{2}$$

where $BC(p_i, p_j) = \int \sqrt{p_i(\vec{\xi}) p_j(\vec{\xi})} d\vec{\xi}$ is the Bhattacharyya coefficient and ξ is the vector variable used. The distance between structures i and j, Δ_{ij}, is finally defined as the norm of the vector of Hellinger distances: $\Delta_{ij} = \|[H_{ij}^{r_{COM}}, H_{ij}^{\vec{\theta}}, H_{ij}^{\vec{\phi}}]\|$.

However, prior to the clustering analysis, we want to remove those structures that are unstable at finite temperature and pressure and melt or develop into a disordered packing. Two strategies have been adopted in this context in order to take into account the emergence of both orientational and conformational disorder. Firstly, orientational disorder is considered by comparing the distribution of the intermolecular torsional angle, $p(\theta_1)$, of the structures with an uniform distribution typical of the liquid state, $p_U(\theta) = 1/2\pi$, hence:

$$H_{i\ell} = \sqrt{1 - \int \sqrt{\frac{1}{2\pi} p_i(\theta)} d\theta} \tag{3}$$

In a second step the emergence of conformational disorder is assessed. To this aim we consider the torsional angle space $p_i(\phi_1, \vec{\phi}_2)$, which presents six basins corresponding to stable conformers.[45] We can thus identify all the possible conformations the molecules adopt in a structure by detecting which of these basins are populated. We define as conformationally disordered structures that contain molecules conformations in 3 or more different basins. Note that point defects such as single molecule undergoing conformational transitions in the crystal bulk which can occur in stable polymorphs[45] at finite temperature, do not distort the probability density $p_i(\phi_1, \vec{\phi}_2)$ and would not lead to a classification of conformational disorder.

We can now group together the finite-temperature putative polymorphs that coalesce to the same geometry. In order to reduce the number of comparisons needed, we can exploit two properties that we have already determined for each structure, namely the chirality and the conformations the molecules adopt in the crystal. Δ_{ij} is therefore calculated only between structures that share the same chirality and conformer composition, drastically reducing the number of comparisons necessary to perform a full clustering of the trajectories, and resulting in the distance matrix in Fig. 2B. To each of the resulting subgroups, corresponding to square sub-regions of the dissimilarity matrix Δ in which the value Δ_{ij} is defined (see Fig. 2B), we applied the Fast Search and Find of Density Peaks (FSFDP) clustering algorithm.[46] The structure with the lowest potential energy in each cluster is taken for the next step.

Metadynamics. In order to overcome MD timescale limitations and sample possible slow transitions, we performed Well Tempered Metadynamics (WTmetaD) simulations[47] on the cluster centres. The choice of CVs is motivated by the need to enhance structural fluctuations without specifically leading the transformation along a specific pathway. To this aim we used density and potential energy, a choice that has the advantage of being general and computationally efficient and therefore suitable for large sets of structures. Given their generality, these CVs can be applied to every crystal but are expected to enhance transitions only between similar structures and have a reduced accuracy in computing the free energy differences between two specific crystal structures. The bias potential is updated every 1 ps with Gaussians characterised by an initial height of 2 kJ mol^{-1} and a width of 10 kg m^{-3} for the density and 2 kJ mol^{-1} for the potential energy. These simulations are performed using Gromacs patched with PLUMED 2[43]. The work performed on the system through the introduction of a bias potential at a time t is represented by the re-weighting factor, $C(t)$,[48] defined as:

$$C(t) = \frac{1}{\beta} \log \frac{\int ds\, e^{-\beta G(s)}}{\int ds\, e^{-\beta (G(s) + V(s,t))}} \tag{4}$$

where $\beta = 1/k_B T$, $G(s)$ is the Gibbs free energy and $V(s,t)$ the bias potential. We searched for possible transitions by looking at the distance RMSD (DRMSD) between pairs of atoms of different molecules using the initial structure as reference. By using a committor analysis in conjunction with DRMSD, both available with PLUMED, we stopped those simulations that exceed the value of 0.3 Å. This value guarantees stopping the simulation and saving computational time in those trajectories that show a large distortion of the crystal packing, usually associated with the melting In addition, to automatically detect transitions between similar geometries not observed by the distance RMSD, we perform a cluster analysis every 0.5 kJ mol^{-1} of work. Finally, persistent structures are ranked based on their energy in the unbiased simulations.

Results

The CSP_0 analysis identified the global minimum, structure R227, as the experimental Form I with an RMSD$_{30}$ of 0.014 Å between the CSP_0 structure and the experimental structure minimised with the same computational method. The search was performed considering $Z' = 1$. Hence the enantiopure Form E, which has $Z' = 2$, cannot be found. The high-energy structure R5596 approximately reproduces the geometry of Form II with an RMSD$_{15}$ of 0.66Å. However, the packing similarity analysis revealed a poor overlap between the two structures. We included in the finite-temperature analysis also the experimental structures IBPRAC16[18] (Form I), JEKNOC12[19] (Form E) and IBPRAC04[22] (Form II) available in the Cambridge Structural Database (CSD)[49] in order to monitor their evolution and predicted persistence throughout the different steps of the reduction process. From a lattice energy perspective, the difference in stability of the experimentally known polymorphs predicted at the CSP_0 stage is significant. Form E is found to be at +5.02 kJ mol^{-1} from Form I, while Form II is at +16.87 kJ mol^{-1}. The large scale set of crystal structures simulated at finite temperature and pressure displays a significant variability attested by the 14 different hydrogen-bonding motifs identified in

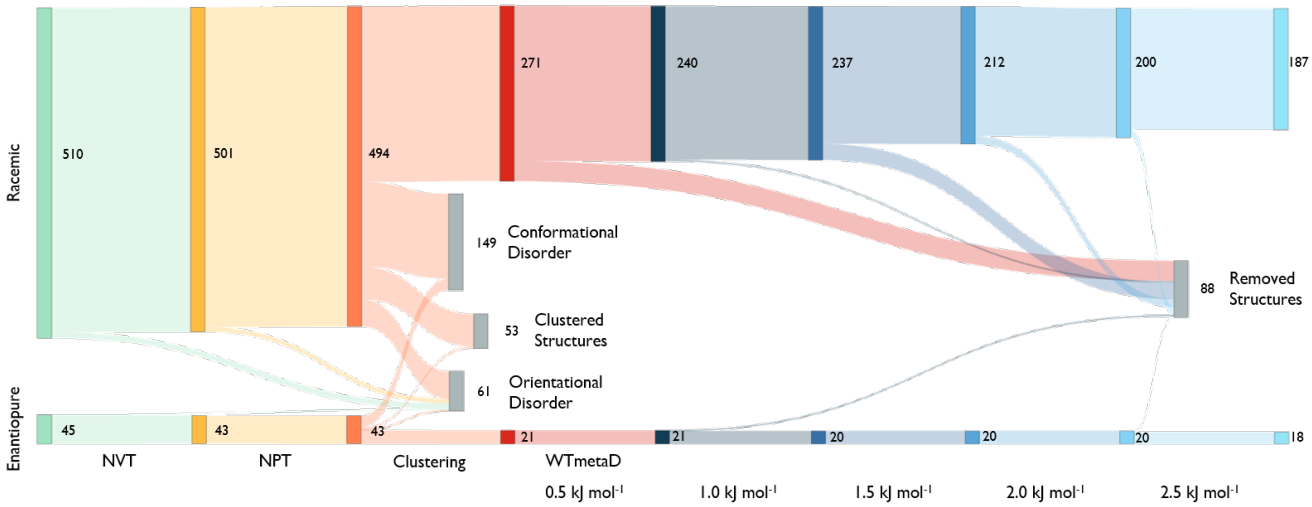

Figure 3: Sankey diagram describing the number of states that survives at each step. The 555 CSP_0-structures are initially divided based on their chirality. All steps are shown with different colours. Removed structures, in grey, result in a disordered packing or transform to another geometry. The clustering performed to metadynamics simulations at different steps of deposited work is shown with a blue scale. From an initial set of 555 geometries we are able to reduce the number of structures to 205.

the CSP_0 dataset. The motifs search was carried out using the CSD-Material module in Mercury.[50] The ring $R_2^2(8)$ motif[51] is the dominant intermolecular interaction motif, recorded in more than half of the structures, including all the experimental ones. H-bonded chains also account for a significant proportion of the structures in the initial dataset, with 151 unit cells displaying the $C_1^1(4)$ motif and 41 unit cells stabilised by the $C_1^1(2)$ one.

Lattice energies are very sensitive to the method used, so when comparing the CSP_0 energies to GAFF, differences are expected. In general, GAFF tends to overestimate the lattice energy differences. Despite this, Form I is found to be among the most stable structures, 4th in the ranking. Form II was confirmed to be very high in energy, at +28.15 kJ mol^{-1} from the global minimum. Form E is located between them at +8.77 kJ mol^{-1}, confirming the relative ranking obtained at the level of theory deployed in the CSP_0 step.

The reduction process starts by equilibrating all structures at 300 K and 1 bar. Fig. 3, shows that around 40% of the structures melt or present disorder after 4 ns of dynamic simulation in the NPT ensemble. The remaining structures are then clustered based on their chirality and molecule conformations.

In the largest group of racemic structures, molecules show conformational flexibility along the *local* torsional angle (See Fig. 2), producing two peaks in the conformational component of the structural fingerprint. Experimental Form II is among these with less than 10% of the molecules in the supercell assuming a distorted conformation.

The clustering analysis shows that only a few states coalesce into common finite-temperature crystal structures while most of them preserve their geometry. The stable Form I is one of the few systems that produce a cluster, and it is among the most populated ones. On the other hand, Form II formed a small cluster with the CSP_0 structure that best matches its geometry while no structure transformed to a configuration compatible with Form E.

Cluster centres are then subject to WTmetaD simulations. In order to automatically analyse trajectories and detect transitions as a function of the work performed by the WTmetaD bias, fingerprints are generated at every increment of 0.5 kJ mol^{-1} of $C(t)$ or by looking at the last frames of those trajectories stopped due to large fluctuations of the DRMSD (see the methods section). Every time fingerprints are generated, clustering is performed, identifying structures that convert and are thus removed from the count of independent, persistent structures.

As shown in Fig. 3, the number of persistent putative polymorphs decreases throughout the workflow. By the end of the analysis, from the initial set of 555 CSP_0 lattice energy minima, we retain 205 persistent structures, corresponding to a 65% of reduction. All experimental structures came out as thermodynamically stable, preserving their geometry during finite-temperature biased simulations. In Fig. 4A, we show the lattice energy landscape at 0 K and depict them based on their behaviour at 300 K.

Orientationally disordered structures at finite-temperature are on average, located at higher energies in the 0 K landscape than the structures exhibiting conformational disorder. Persistent crystal structures at 300 K span over the entire lattice energy range. The overlap in lattice energy between the distributions of labile and persistent crystal structures highlights how a reduction of the lattice energy landscape based solely on lattice energy is insufficient and would actually miss high energy experimental structures like Form II. The resulting finite-temperature crystal energy landscape, in Fig. 4B, shows a general decrease in the potential energy difference for those structures that survive.

Among the dominant hydrogen-bonding motifs, the ring motif $R_2^2(8)$ and the chain motif $C_1^1(4)$ were shown to be more persistent than the average, with a decrease in number of structures of 29% and 41%, considering all stable structures. While being present in

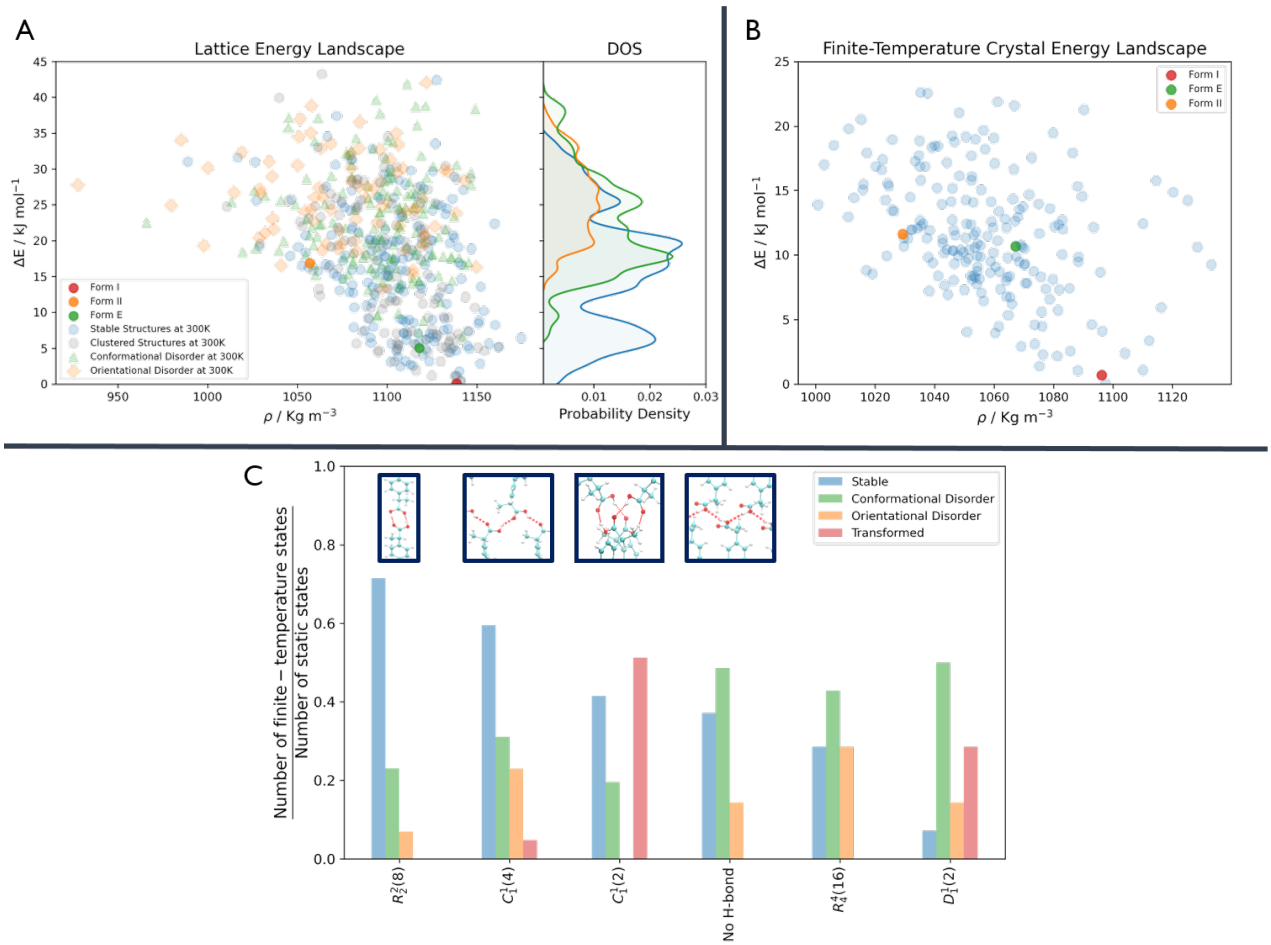

Figure 4: Comparison between the crystal energy landscape at 0K and 300K. (A) CSP0 lattice energy landscape. The symbols highlight how the static states at 0K will behave at 300 K and 1 bar. Structures that are persistent and thermodynamically stable at finite temperature and pressure are represented with blue dots. Structures that develop a disorder are shown with a green triangle or orange square whether it is orientational or conformational. The plot on the right shows how these three groups are distributed over the energy axis. (B) Final finite-temperature crystal energy landscape obtained with GAFF with the experimental forms I, II and E highlighted in red, orange and green. (C) Behaviour of the surviving H-bonding motifs at finite temperature and pressure, with the four most common shown in the blue boxes. For each of them, we show the number of structures that preserve or convert to that motif at the end of the analysis (in blue), those that result in a disordered structure (in green or orange whether the disorder is conformational or orientational) or transform to another motif (in red), rescaled by their initial occurrence in the CSP0 set.

the final set, the chain motif $C_1^1(2)$ tended to convert to the more stabilising $C_1^1(4)$. Looking at the rare motifs, 9 of them disappear during the analysis. Motifs $D_3^3(10)$, $R_3^3(12)$, $R_3^3(6)$, $R_4^4(8)$, $R_6^6(12)$ and $R_6^6(24)$ all result in melted structures while motifs $C_2^2(6)$, $R_2^2(6)$ and $R_4^4(12)$ transform to other motifs.

Discussion

Through an MD-based reduction of the lattice energy landscape we drastically reduced the number of putative polymorphs of Ibuprofen. Form I, the most stable experimentally known polymorph came out as second in the final ranking with structure R4124 being the global minimum (see the ESI† for a complete list of crystal structures, energies and labels). However, many structures converted to the experimental form, suggesting that form I could act as kinetic trap for labile states. Form I and the enantiopure form E were able to preserve their CSP_0 geometry with little variation due to the molecular motion. In form II, a few molecules dynamically change conformation during biased and unbiased simulations, showing the possibility of dynamic disorder in the crystal at standard conditions. This is evidenced also by the presence of two conformationally disordered structures, R2315 and R6595, that resemble form II. Structure R5596, which presents geometrical similarities with form II at 0 K, effectively convert to the experimental one at 300 K.

The $R_2^2(8)$ motif is the most frequent intermolecular interaction motif in the final set being present in 119 structures and is dominant among the low-energy structures. From Fig. 4C, we can see that structures associated with this motif, are most likely to preserve their H-bonds. However, other motifs seem to be favoured at 300 K compared to their initial number in the CSP_0 set. In particular, the chain motif $C_1^1(4)$ has a similar persistence to $R_2^2(8)$ and is present in 10 of the 18 enantiopure structures, including the most stable one (E6134). This could indicate that the hydrogen-bonded carboxylic acid dimer of the $R_2^2(8)$ pattern is favoured during the nucleation or growth process. The rotation of the carboxyl group is associated with the inter-conversion between mo-

tifs $C_1^1(2)$ and $C_1^1(4)$ with the balance shifted towards the latter. Interestingly, 12 structures are shown to be persistent at finite-temperature and pressure despite the lack of H-bonding motifs and their high potential energy. The use of highly polar solvents that preclude H-bonding interactions could favour the formation of these structures.[52]

In Fig. 4A, the states that are effectively persistent at finite temperature and pressure are shown as blue dots. From the probability density on the right of the same figure, we can see that some of these are high energy structures. This implies that the use of energy cutoff, although often necessary, can lead to remove relevant geometries from the analysis. In this case, the ibuprofen form II could have been ignored being higher in energy than the typical energy cutoffs used, usually in the range 5-15 kJ mol^{-1}.

Conclusions

In this work, we have tackled the systematic reduction of a large-scale dataset of CSP_0 crystal structures, including 555 putative crystal structures of ibuprofen by systematically applying MD simulations.[15] To scale up one order of magnitude in the number of crystal structures considered, compared to previous studies, we implemented new strategies to further increase the efficiency of the clustering and analysis protocols, drastically reducing the need for manual inspection of the trajectories in different steps. In particular, through a systematic conformational analysis, we could automatically detect disorder formation in the simulation box. Moreover, partitioning a-priori the distance matrix in subsets based on the number and type of conformers present in the crystal structure, small and fast to manage, allowed us to efficiently repeat the clustering analysis at regular intervals during the metadynamics simulations. This procedure allowed us to detect transitions under progressively enhanced fluctuations of the supercells density and lattice energy. The systematic setup and analysis of 555 trajectories, which altogether amounts to 8 μs across multiple MD protocols, is made possible by an *ad hoc* Python library, available at github.com/mme-ucl/pypol.

Applying this approach to a set of 555 CSP_0-generated structures of ibuprofen resulted in a 65% reduction of the number of predicted structures, leading to a group of 205 persistent lattice structures. All the experimentally known structures persisted throughout the analysis with minor variations from their original geometry. Interestingly, despite the significant variability in the intermolecular interaction motifs present in the initial dataset (14), we find that the motifs $R_2^2(8)$ and $C_1^1(4)$ are dominant in the final set.

By taking advantage of this implementation, we can now study sets of structures of the typical scale of real-world CSP_0-generated crystal energy landscapes. We believe that this method is a valid alternative to the simple application of an energy cutoff criterion to select crystal structures as promising putative polymorphs.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Professor Sarah L. Price for fruitful discussions. Professors Claire Adjiman & Costas Pantelides at Imperial College London are acknowledged for sharing with us the CrystalPredictor and CrystalOptimizer programs. The CSP computational software is developed under EPSRC grant EP/K039229/1. Calculations were performed on University College London's Myriad and Kathleen High Performance Computing Facilities. NFF acknowledges Eli Lilly Digital Design for support through a PhD scholarship. LSP and MS are also partially funded by Eli Lilly Digital Design.

References

[1] S. L. Price, *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.*, 2013, **69**, 313–328.

[2] A. J. Cruz-Cabeza, S. M. Reutzel-Edens and J. Bernstein, *Chem. Soc. Rev.*, 2015, **44**, 8619–8635.

[3] J. Nyman and G. M. Day, *CrystEngComm*, 2015, **17**, 5154–5165.

[4] G. Sun, X. Liu, Y. A. Abramov, S. O. Nilsson Lill, C. Chang, V. Burger and A. Broo, *Cryst. Growth Des.*, 2021, **21**, 1972–1983.

[5] E. Schneider, L. Vogt and M. E. Tuckerman, *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.*, 2016, **72**, 542–550.

[6] M. Yang, E. Dybeck, G. Sun, C. Peng, B. Samas, V. M. Burger, Q. Zeng, Y. Jin, M. A. Bellucci, Y. Liu, P. Zhang, J. Ma, Y. A. Jiang, B. C. Hancock, S. Wen and G. P. Wood, *Cryst. Growth Des.*, 2020, **20**, 5211–5224.

[7] Y. A. Abramov, *Org. Process Res. Dev.*, 2013, **17**, 472–485.

[8] S. L. Price, D. E. Braun and S. M. Reutzel-Edens, *Chem. Commun.*, 2016, **52**, 7065–7077.

[9] J. Nyman, O. S. Pundyke and G. M. Day, *Phys. Chem. Chem. Phys.*, 2016, **18**, 15828–15837.

[10] J. Hoja, H. Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio and A. Tkatchenko, *Sci. Adv.*, 2019, **5**, 3338 –3347.

[11] S. L. Price, *Faraday Discuss.*, 2018, **211**, 9–30.

[12] A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio, A. Dzyabchenko, B. P. Van Eijck, D. M. Elking, J. A. Van Den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C. A. Gatsiou, T. S. Gee, R. De Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. De Jong, J. Kendrick, N. J. De Klerk, H. Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom,

A. E. Masunov, P. McCabe, D. P. McMahon, H. Meekes, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szalewicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis, A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. De Wijs, J. Yang, Q. Zhu and C. R. Groom, *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.*, 2016, **72**, 439–459.

[13] C. S. Adjiman, J. G. Brandenburg, D. E. Braun, J. Cole, C. Collins, A. I. Cooper, A. J. Cruz-Cabeza, G. M. Day, M. Dudek, A. Hare, L. Iuzzolino, D. McKay, J. B. Mitchell, S. Mohamed, S. Neelamraju, M. Neumann, S. Nilsson Lill, J. Nyman, A. R. Oganov, S. L. Price, A. Pulido, S. Reutzel-Edens, I. Rietveld, M. T. Ruggiero, J. C. Schön, S. Tsuzuki, J. van den Ende, G. Woollam and Q. Zhu, *Faraday Discuss.*, 2018, **211**, 493–539.

[14] E. C. Dybeck, D. P. McMahon, G. M. Day and M. R. Shirts, *Cryst. Growth Des.*, 2019, **19**, 5568–5580.

[15] N. F. Francia, L. S. Price, J. Nyman, S. L. Price and M. Salvalaglio, *Cryst. Growth Des.*, 2020, **20**, 6847–6862.

[16] A. M. Evans, *Clin. Rheumatol.*, 2001, **20**, 9–14.

[17] G. Geisslinger, K. P. Stock, G. L. Bach, D. Loew and K. Brune, *Agents Actions*, 1989, **27**, 455–457.

[18] K. Ostrowska, M. Kropidowska and A. Katrusiak, *Cryst. Growth Des.*, 2015, **15**, 1512–1517.

[19] M. D. King, W. D. Buchanan and T. M. Korter, *J. Pharm. Sci.*, 2011, **100**, 1116–1129.

[20] A. A. Freer, J. M. Bunyan, N. Shankland and D. B. Sheen, *Acta Crystallogr. Sect. C*, 1993, **49**, 1378–1380.

[21] E. Dudognon, F. Danède, M. Descamps and N. T. Correia, *Pharm. Res.*, 2008, **25**, 2853–2858.

[22] P. Derollez, E. Dudognon, F. Affouard, F. Danède, N. T. Correia and M. Descamps, *Acta Crystallogr. Sect. B Struct. Sci.*, 2010, **66**, 76–80.

[23] P. A. Williams, C. E. Hughes and K. D. Harris, *Cryst. Growth Des.*, 2012, **12**, 5839–5845.

[24] P. G. Karamertzanis and C. C. Pantelides, *Mol. Phys.*, 2007, **105**, 273–291.

[25] D. E. Williams and S. R. Cox, *Acta Crystallogr. Sect. B*, 1984, **40**, 404–417.

[26] D. S. Coombes, S. L. Price, D. J. Willock and M. Leslie, *J. Phys. Chem.*, 1996, **100**, 7352–7360.

[27] L. Iuzzolino, P. McCabe, S. L. Price, J. G. Brandenburg and J. Gerit Brandenburg, *Faraday Discuss.*, 2018, **211**, 275.

[28] S. L. Price, M. Leslie, G. W. Welch, M. Habgood, L. S. Price, P. G. Karamertzanis and G. M. Day, *Phys. Chem. Chem. Phys.*, 2010, **12**, 8478–8490.

[29] A. J. Stone, *J. Chem. Theory Comput.*, 2005, **1**, 1128–1132.

[30] E. N. Pistikopoulos, M. C. Georgiadis, V. Dua, C. S. Adjiman and A. Galindo, *Process Systems Engineering*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2011, vol. 6, p. 317.

[31] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

[32] D. H. Case, J. E. Campbell, P. J. Bygrave and G. M. Day, *J. Chem. Theory Comput.*, 2016, **12**, 910–924.

[33] A. Jakalian, B. L. Bush, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2000, **21**, 132–146.

[34] Lindahl, Abraham, Hess and van der Spoel, *GROMACS 2021 Manual, https://zenodo.org/record/4457591*, 2021, `https://zenodo.org/record/4457591`.

[35] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1-2**, 19–25.

[36] L. P. Cordella, P. Foggia, C. Sansone and M. Vento, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, **26**, 1367–1372.

[37] A. A. Hagberg, D. A. Schult and P. J. Swart, 7th Python Sci. Conf. (SciPy 2008), 2008, pp. 11–15.

[38] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[39] M. R. Shirts, C. Klein, J. M. Swails, J. Yin, M. K. Gilson, D. L. Mobley, D. A. Case and E. D. Zhong, *J. Comput. Aided. Mol. Des.*, 2017, **31**, 147–161.

[40] G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

[41] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

[42] M. Parrinello and A. Rahman, *Phys. Rev. Lett.*, 1980, **45**, 1196–1199.

[43] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.

[44] M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A. L. Ferguson, M. Filizola, J. S. Fraser, H. Fu, P. Gasparotto, F. L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G. T. Heller, G. M. Hocky, M. Iannuzzi, M. Invernizzi, K. E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Löhr, F. Marinelli, L. Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfaendtner, P. Piaggi, G. M. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G. C. Sosso, V. Spiwok, J. Šponer, D. W. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth and A. White, *Nat. Methods*, 2019, **16**, 670–673.

[45] V. Marinova, G. P. Wood, I. Marziano and M. Salvalaglio, *J. Chem. Theory Comput.*, 2018, **14**, 6484–6494.

[46] A. Rodriguez and A. Laio, *Science (80-.).*, 2014, **344**, 1492–1496.

[47] A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.

[48] P. Tiwary and M. Parrinello, *J. Phys. Chem. B*, 2015, **119**, 736–742.

[49] C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. C. Ward and IUCr, *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.*, 2016, **72**, 171–179.

[50] C. F. MacRae, I. Sovago, S. J. Cottrell, P. T. Galek, P. McCabe, E. Pidcock, M. Platings, G. P. Shields, J. S. Stevens, M. Towler and P. A. Wood, *J. Appl. Crystallogr.*, 2020, **53**, 226–235.

[51] J. Bernstein, R. E. Davis, L. Shimoni and N. Chang, *Angew. Chemie Int. Ed. English*, 1995, **34**, 1555–1573.

[52] R. Bobrovs, L. Drunka, A. A. Auzins, K. Jaudzems and M. Salvalaglio, *Cryst. Growth Des.*, 2021, **21**, 436–448.

Supplementary Information. Reducing Crystal Structure Overprediction of Ibuprofen with Large Scale Molecular Dynamics Simulations

Nicholas F. Francia,[†] Louise S. Price,[‡] and Matteo Salvalaglio[*,†]

†*Thomas Young Centre and Department of Chemical Engineering, University College London, London WC1E 7JE, UK.*

‡*Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK.*

E-mail: m.salvalaglio@ucl.ac.uk

Evolution of putative polymorphs

In these Supplementary Information (SI), we show the evolution of the putative polymorphs of ibuprofen through the different steps. The relative energies are expressed in kJ mol^{-1} but while in the CSP_0 and energy minimization (EM) with GAFF steps they are lattice energies, in the NVT and NPT steps they are an average of the potential energy difference between a molecule in the crystal and a molecule in vacuum. We first show those structures that came out as persistent in the main paper analysis (Tab. 1), then those that convert to more stable finite-temperature structures (Tab. 2) and finally those that exhibit conformational or orientational disorder (Tab. 3). Racemic buprofen structures from the Cambridge Structural Database (CSD),[1] IBPRAC16[2] and IBPRAC04,[3] and their CSP_0 best match, R227 and R5596, are here labelled with Form I and Form II, using an asterisk to denote the

experimental ones. The enantiopure polymorph JEKNOC12[4] is labelled as Form E.

Table 1: The energies of the persistent structures at different steps. In the last two columns we show the dominant H-bonding motif at 0K and 300K.

Rank	IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Initial Motif	Final Motif
1	R4124	1.80	0.00	0.50	0.00	$R_2^2(8)$	$R_2^2(8)$
2	FormI	0.06	1.18	1.02	0.68	$R_2^2(8)$	$R_2^2(8)$
3	R113	0.47	2.71	3.15	1.05	$R_2^2(8)$	$R_2^2(8)$
4	R440	3.83	7.71	7.54	1.08	$R_2^2(8)$	$R_2^2(8)$
5	E6314	5.20	0.21	0.00	1.39	$C_1^1(4)$	$C_1^1(4)$
6	R61	3.18	2.76	2.27	2.18	$R_2^2(8)$	$R_2^2(8)$
7	R109	2.89	1.94	1.74	2.30	$R_2^2(8)$	$R_2^2(8)$
8	R658	1.21	1.30	0.90	2.39	$R_2^2(8)$	$R_2^2(8)$
9	R7232	15.00	16.96	14.01	2.57	$R_2^2(8)$	$R_2^2(8)$
10	R4600	4.14	3.78	4.45	2.92	$R_2^2(8)$	$R_2^2(8)$
11	R986	7.09	4.91	5.09	3.10	$R_2^2(8)$	$R_2^2(8)$
12	R171	5.62	4.16	3.95	3.37	$R_2^2(8)$	$R_2^2(8)$
13	R6475	2.71	6.90	7.01	3.97	$R_2^2(8)$	$R_2^2(8)$
14	R336	14.20	14.33	11.30	4.05	$R_2^2(8)$	$R_2^2(8)$
15	R4384	8.67	8.45	8.75	4.12	$R_2^2(8)$	$R_2^2(8)$
16	R1425	19.70	22.51	11.65	4.45	$R_2^2(8)$	$R_2^2(8)$
17	R5296	2.52	5.65	6.25	4.47	$R_2^2(8)$	$R_2^2(8)$
18	R9240	5.00	4.47	4.39	4.60	$R_2^2(8)$	$R_2^2(8)$
19	R8451	23.14	20.47	21.18	4.62	$C_1^1(4)$	$C_1^1(4)$
20	R6276	9.23	16.67	10.75	5.34	$R_2^2(8)$	$R_2^2(8)$
21	R4521	8.98	8.95	9.63	5.61	$R_2^2(8)$	$R_2^2(8)$
22	R3270	24.68	26.76	17.89	5.72	$R_2^2(8)$	$R_2^2(8)$
23	R170	4.91	7.68	7.99	5.95	$R_2^2(8)$	$R_2^2(8)$
24	R1470	20.73	36.22	18.98	6.21	$R_2^2(8)$	$R_2^2(8)$

Continued on next page

Rank	IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Initial Motif	Final Motif
25	R2118	12.73	10.83	10.90	6.44	$R_2^2(8)$	$R_2^2(8)$
26	R9558	7.07	8.05	8.57	6.57	$R_2^2(8)$	$R_2^2(8)$
27	R5228	18.39	20.37	20.63	6.74	$R_2^2(8)$	$R_2^2(8)$
28	R346	6.22	6.43	6.81	6.80	$R_2^2(8)$	$R_2^2(8)$
29	R6514	8.98	14.93	12.52	6.94	$R_2^2(8)$	$R_2^2(8)$
30	R5463	17.96	20.41	17.41	7.06	$R_2^2(8)$	$R_2^2(8)$
31	R1574	4.39	10.86	11.66	7.21	$R_2^2(8)$	$R_2^2(8)$
32	R7321	13.79	16.85	13.43	7.30	$R_2^2(8)$	$R_2^2(8)$
33	R967	7.31	12.95	13.14	7.31	$R_2^2(8)$	$R_2^2(8)$
34	R940	8.59	7.75	7.84	7.37	$R_2^2(8)$	$R_2^2(8)$
35	R8158	4.86	11.98	11.38	7.38	$R_2^2(8)$	$R_2^2(8)$
36	R931	26.11	21.46	13.12	7.43	$C_1^1(4)$	$C_1^1(4)$
37	R3930	5.61	12.67	10.52	7.65	$R_2^2(8)$	$R_2^2(8)$
38	R25	18.54	24.94	15.09	7.75	$R_2^2(8)$	$R_2^2(8)$
39	R5899	6.48	9.61	9.66	7.82	$R_2^2(8)$	$R_2^2(8)$
40	R9730	27.57	23.49	17.56	7.83	$C_1^1(4)$	$C_1^1(4)$
41	R3714	3.21	9.41	9.71	8.01	$R_2^2(8)$	$R_2^2(8)$
42	R6515	31.30	36.39	19.08	8.04	$R_2^2(8)$	$R_2^2(8)$
43	R6237	18.88	28.08	9.91	8.11	$R_2^2(8)$	$R_2^2(8)$
44	R6904	31.00	36.79	20.93	8.11	$R_2^2(8)$	$R_2^2(8)$
45	R794	5.36	10.53	11.43	8.15	$R_2^2(8)$	$R_2^2(8)$
46	R2384	12.12	22.47	22.20	8.27	$R_2^2(8)$	$R_2^2(8)$
47	R7318	5.78	9.23	9.99	8.28	$R_2^2(8)$	$R_2^2(8)$
48	R6569	19.51	17.36	18.40	8.31	$C_1^1(4)$	$C_1^1(4)$
49	R7642	19.66	30.55	26.55	8.36	$R_2^2(8)$	$R_2^2(8)$
50	R3281	12.92	19.30	19.76	8.37	$R_2^2(8)$	$R_2^2(8)$
51	R2040	15.76	25.01	14.08	8.40	$R_2^2(8)$	$R_2^2(8)$
52	E7954	31.32	35.84	17.34	8.41	$C_1^1(4)$	$C_1^1(4)$

Continued on next page

Rank	IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Initial Motif	Final Motif
53	R7023	2.45	8.32	9.09	8.56	$R_2^2(8)$	$R_2^2(8)$
54	R7837	10.73	13.33	11.92	8.64	$R_2^2(8)$	$R_2^2(8)$
55	R4094	10.18	8.47	8.21	8.68	$C_1^1(4)$	$C_1^1(4)$
56	R9902	14.71	16.30	16.02	8.70	$R_2^2(8)$	$R_2^2(8)$
57	R8053	15.05	21.23	21.24	8.73	$R_2^2(8)$	$R_2^2(8)$
58	R702	5.71	13.45	14.15	8.76	$R_2^2(8)$	$R_2^2(8)$
59	R4814	3.11	7.81	7.66	8.81	$R_2^2(8)$	$R_2^2(8)$
60	R8773	21.95	21.69	21.55	9.09	$R_2^2(8)$	$R_2^2(8)$
61	R2809	18.94	16.72	15.01	9.09	$R_2^2(8)$	$R_2^2(8)$
62	R1756	19.08	21.39	18.82	9.25	$C_1^1(4)$	$C_1^1(2)$
63	R485	15.90	19.38	18.60	9.25	$R_2^2(8)$	$R_2^2(8)$
64	E1974	16.67	18.64	18.91	9.31	$C_1^1(4)$	$C_1^1(4)$
65	R2087	6.96	6.48	6.60	9.35	$R_2^2(8)$	$R_2^2(8)$
66	R5541	6.49	6.58	6.75	9.41	$R_2^2(8)$	$R_2^2(8)$
67	R8382	19.25	25.09	22.84	9.48	$R_2^2(8)$	$R_2^2(8)$
68	R715	16.54	18.05	16.54	9.49	$R_2^2(8)$	$R_2^2(8)$
69	R2726	14.64	18.64	14.21	9.49	$R_2^2(8)$	$R_2^2(8)$
70	R2536	11.32	11.12	11.82	9.50	$R_2^2(8)$	$R_2^2(8)$
71	R7263	27.90	25.65	17.02	9.73	$C_1^1(4)$	$C_1^1(4)$
72	R7194	6.90	11.52	12.15	9.80	$C_1^1(4)$	$C_1^1(4)$
73	R6032	21.60	30.07	19.98	9.82	$R_2^2(8)$	$R_2^2(8)$
74	R7307	13.50	19.10	10.59	9.85	$R_2^2(8)$	$R_2^2(8)$
75	R5133	15.91	13.88	11.68	9.89	$R_2^2(8)$	$R_2^2(8)$
76	R1502	7.19	8.18	8.37	9.89	$R_2^2(8)$	$R_2^2(8)$
77	R1687	8.98	11.87	11.95	9.94	$R_2^2(8)$	$R_2^2(8)$
78	R561	7.89	14.38	16.60	9.94	$R_2^2(8)$	$R_2^2(8)$
79	R264	19.96	25.35	16.95	9.95	$R_2^2(8)$	$R_2^2(8)$
80	R5981	25.70	29.42	22.77	10.04	$C_1^1(4)$	$C_1^1(4)$

Continued on next page

Rank	IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Initial Motif	Final Motif
81	R8092	13.24	23.99	19.92	10.15	$R_2^2(8)$	$R_2^2(8)$
82	R7238	17.40	16.22	16.52	10.19	$C_1^1(4)$	$C_1^1(4)$
83	R888	7.06	10.14	10.02	10.23	$R_2^2(8)$	$R_2^2(8)$
84	R3184	13.08	13.02	13.42	10.35	$R_2^2(8)$	$R_2^2(8)$
85	R2454	7.98	11.22	11.94	10.36	$R_2^2(8)$	$R_2^2(8)$
86	R1776	4.64	12.09	13.15	10.40	$R_2^2(8)$	$R_2^2(8)$
87	R3141	17.88	15.42	15.91	10.42	$R_2^2(8)$	$R_2^2(8)$
88	R2763	10.89	13.10	14.07	10.44	$R_2^2(8)$	$R_2^2(8)$
89	R2863	6.41	25.77	21.00	10.50	$C_1^1(2)$	$C_1^1(4)$
90	R9348	17.14	26.54	17.27	10.58	$R_2^2(8)$	$R_2^2(8)$
91	R1851	8.47	13.14	12.96	10.62	$C_1^1(4)$	$C_1^1(4)$
92	E8551	16.20	27.63	20.45	10.63	$C_1^1(4)$	$C_1^1(4)$
93	R5884	5.55	10.08	10.61	10.67	$R_2^2(8)$	$R_2^2(8)$
94	FormE	5.02	8.77	9.07	10.68	$R_2^2(8)$	$R_2^2(8)$
95	R8603	28.11	35.24	16.79	10.75	$C_1^1(2)$	$C_1^1(4)$
96	R4927	25.17	23.38	18.87	10.79	$C_1^1(4)$	$C_1^1(4)$
97	R522	6.17	14.67	13.60	10.91	$R_2^2(8)$	$R_2^2(8)$
98	R2495	13.82	17.61	17.06	10.94	$R_2^2(8)$	$R_2^2(8)$
99	R6747	28.69	33.05	15.33	11.13	$R_2^2(8)$	$R_2^2(8)$
100	R5763	19.86	23.95	15.49	11.25	$C_1^1(4)$	$C_1^1(4)$
101	R4959	19.12	20.32	18.56	11.28	$R_2^2(8)$	$R_2^2(8)$
102	R8758	22.36	27.82	14.92	11.31	$R_2^2(8)$	$R_2^2(8)$
103	R8399	13.40	13.36	14.61	11.31	$R_2^2(8)$	$R_2^2(8)$
104	R3905	19.01	20.85	16.79	11.35	$C_1^1(4)$	$C_1^1(4)$
105	R9162	9.06	13.92	14.34	11.47	$R_2^2(8)$	$R_2^2(8)$
106	R8458	15.45	19.40	16.78	11.53	$R_2^2(8)$	$R_2^2(8)$
107	R4837	5.93	12.87	12.96	11.57	$R_2^2(8)$	$R_2^2(8)$
108	FormII*	16.87	28.15	15.41	11.62	$R_2^2(8)$	$R_2^2(8)$

Continued on next page

Rank	IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Initial Motif	Final Motif
109	R408	27.94	23.81	19.83	11.67	$C_1^1(4)$	$C_1^1(4)$
110	R9945	15.42	15.59	16.05	11.75	$R_2^2(8)$	$R_2^2(8)$
111	R1454	20.81	23.86	23.47	11.91	$R_2^2(8)$	$R_2^2(8)$
112	R5287	9.12	18.07	14.62	11.91	$R_2^2(8)$	$R_2^2(8)$
113	R3529	19.57	23.31	23.16	11.97	$R_4^4(12)$	$R_2^2(8)$
114	R2271	7.56	12.34	13.09	11.97	$R_2^2(8)$	$R_2^2(8)$
115	R3096	4.68	14.75	14.62	12.01	$R_2^2(8)$	$R_2^2(8)$
116	R9824	23.41	26.34	19.03	12.08	$R_2^2(8)$	$R_2^2(8)$
117	R5394	26.02	27.23	20.13	12.14	$C_1^1(4)$	$C_1^1(4)$
118	E682	14.81	13.10	13.16	12.15	$R_2^2(8)$	$R_2^2(8)$
119	R8475	17.19	16.47	17.63	12.15	$C_1^1(4)$	$C_1^1(4)$
120	R3073	28.95	32.62	22.92	12.21	$R_2^2(8)$	$R_2^2(8)$
121	R4811	13.33	14.13	14.51	12.25	$R_2^2(8)$	$R_2^2(8)$
122	R2822	12.08	17.53	17.19	12.37	$C_1^1(4)$	$C_1^1(4)$
123	R5236	20.13	17.02	15.90	12.38	$C_1^1(4)$	$C_1^1(4)$
124	R8054	20.73	21.33	18.31	12.41	$R_2^2(8)$	$R_2^2(8)$
125	R1773	28.46	30.74	15.65	12.44	$C_1^1(4)$	$C_1^1(4)$
126	R6595	20.43	30.17	14.28	12.44	$C_1^1(4)$	$C_1^1(4)$
127	R6714	15.07	22.06	17.83	12.60	$C_1^1(4)$	$C_1^1(4)$
128	R3386	7.73	20.52	20.60	12.66	$R_2^2(8)$	$R_2^2(8)$
129	R2989	13.79	15.20	15.61	12.68	$C_1^1(4)$	$C_1^1(4)$
130	R370	16.64	19.73	13.39	12.70	$C_1^1(2)$	$C_1^1(4)$
131	E2960	25.00	28.48	17.44	12.71	$C_1^1(4)$	$C_1^1(4)$
132	R1867	18.04	21.72	21.98	12.79	$C_1^1(4)$	$C_1^1(4)$
133	R6039	18.58	16.66	18.85	12.95	$C_1^1(2)$	$C_1^1(2)$
134	R4696	19.18	18.05	14.45	12.95	$C_1^1(4)$	$C_1^1(4)$
135	R5959	13.58	15.82	15.96	13.38	$C_1^1(4)$	$C_1^1(4)$
136	R3894	19.96	24.54	16.64	13.40	$R_2^2(8)$	$R_2^2(8)$

Continued on next page

Rank	IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Initial Motif	Final Motif
137	E4457	16.97	30.77	24.29	13.48	$C_1^1(2)$	$C_1^1(4)$
138	E4069	12.38	18.50	17.39	13.54	$C_1^1(4)$	$C_1^1(4)$
139	R3228	25.77	27.52	20.26	13.60	$R_2^2(6)$	$C_1^1(4)$
140	R4445	18.69	30.05	19.44	13.74	$C_1^1(4)$	$C_1^1(4)$
141	R9213	7.50	11.32	11.69	13.91	$R_2^2(8)$	$R_2^2(8)$
142	R1493	17.44	14.36	13.87	13.94	$C_1^1(2)$	$C_1^1(2)$
143	R6194	42.44	40.37	25.84	13.98	No H-bond	No H-bond
144	R2527	29.68	24.08	22.65	14.08	$C_1^1(4)$	$C_1^1(4)$
145	R3795	17.82	16.67	16.79	14.12	$C_1^1(2)$	$C_1^1(2)$
146	E7426	23.28	29.50	21.33	14.14	$C_1^1(4)$	$C_1^1(4)$
147	R9950	5.34	14.95	15.00	14.25	$R_2^2(8)$	$R_2^2(8)$
148	R7806	19.88	21.11	22.43	14.26	$C_1^1(4)$	$C_1^1(4)$
149	R9634	16.64	20.31	17.91	14.27	$R_2^2(8)$	$R_2^2(8)$
150	R6786	34.52	31.08	24.05	14.34	$C_1^1(2)$	$C_1^1(4)$
151	E6175	29.97	34.03	28.09	14.45	No H-bond	No H-bond
152	R8172	12.82	13.98	15.06	14.49	$R_2^2(8)$	$R_2^2(8)$
153	R9737	6.49	17.04	17.54	14.55	$R_2^2(8)$	$R_2^2(8)$
154	R947	20.15	20.21	20.84	14.70	$C_1^1(2)$	$C_1^1(2)$
155	R5793	8.82	19.97	19.46	14.87	$R_2^2(8)$	$R_2^2(8)$
156	R961	21.59	19.85	20.39	15.05	$C_1^1(2)$	$C_1^1(2)$
157	R9352	25.23	27.60	20.61	15.08	$R_2^2(8)$	$R_2^2(8)$
158	R9773	15.88	20.84	19.57	15.17	$R_2^2(8)$	$R_2^2(8)$
159	R4986	15.77	22.74	22.42	15.21	$R_2^2(8)$	$R_2^2(8)$
160	R899	24.28	32.49	31.92	15.32	$C_1^1(4)$	$C_1^1(4)$
161	R4532	16.23	15.91	19.39	15.34	$R_4^4(16)$	$R_4^4(16)$
162	R7156	31.05	35.04	23.27	15.47	$R_2^2(8)$	$R_2^2(8)$
163	R1492	21.85	25.57	27.33	15.54	$C_1^1(4)$	$C_1^1(4)$
164	R1116	8.27	19.11	18.65	15.77	$R_2^2(8)$	$R_2^2(8)$

Continued on next page

Rank	IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Initial Motif	Final Motif
165	E2219	16.56	16.41	17.90	15.80	$R_2^2(8)$	$R_2^2(8)$
166	R8177	17.80	21.45	16.75	15.82	$R_2^2(8)$	$R_2^2(8)$
167	E3363	23.69	19.94	16.58	16.08	$C_1^1(4)$	$C_1^1(2)$
168	R1962	29.29	26.59	21.18	16.13	$C_1^1(4)$	$C_1^1(4)$
169	R5853	17.31	23.40	20.46	16.63	$R_2^2(8)$	$R_2^2(8)$
170	R4333	27.01	25.11	25.06	16.65	$C_1^1(4)$	$C_1^1(4)$
171	E8640	15.95	22.75	21.88	16.87	$C_1^1(4)$	$C_1^1(4)$
172	R1207	17.82	16.80	17.99	16.87	$R_2^2(8)$	$R_2^2(8)$
173	R6453	27.33	33.55	26.82	16.88	$D_1^1(2)$	$C_1^1(4)$
174	R4293	11.18	17.71	18.50	17.01	$R_2^2(8)$	$R_2^2(8)$
175	R2773	24.32	16.17	15.62	17.11	$C_1^1(2)$	$C_1^1(2)$
176	R5010	15.24	24.39	24.17	17.18	$R_2^2(8)$	$R_2^2(8)$
177	R9631	32.85	35.60	23.33	17.28	$R_2^2(6)$	$C_1^1(4)$
178	R640	27.08	28.67	20.68	17.43	$C_1^1(2)$	$C_1^1(4)$
179	R9754	33.56	36.78	24.30	17.55	No H-bond	No H-bond
180	R6708	13.34	17.36	17.05	17.71	$C_1^1(2)$	$C_1^1(4)$
181	R8151	17.74	15.84	16.06	17.72	$R_2^2(8)$	$R_2^2(8)$
182	R3291	20.03	24.92	24.58	17.91	No H-bond	No H-bond
183	R5462	20.32	24.10	23.68	17.91	$C_1^1(4)$	$C_1^1(4)$
184	R6718	26.18	27.53	22.53	18.20	$C_1^1(4)$	$C_1^1(4)$
185	R5514	37.39	40.68	21.13	18.35	$C_1^1(2)$	$C_1^1(2)$
186	R300	25.62	24.40	25.63	18.46	$C_1^1(2)$	$C_1^1(4)$
187	R5901	24.52	22.68	21.05	18.48	$C_1^1(2)$	$C_1^1(4)$
188	R4164	20.23	23.80	22.01	18.50	$C_1^1(4)$	$C_1^1(4)$
189	R6849	20.11	19.76	22.25	18.79	$C_1^1(2)$	$C_1^1(4)$
190	R2379	20.11	16.87	16.50	18.86	$C_1^1(2)$	$C_1^1(2)$
191	E7885	13.57	27.09	25.24	18.95	$C_1^1(2)$	$C_1^1(4)$
192	R7091	25.84	31.58	20.58	18.98	No H-bond	No H-bond

Continued on next page

Rank	IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Initial Motif	Final Motif
193	R2491	24.85	20.81	20.48	19.06	$D_1^1(2)$	$D_1^1(2)$
194	R4973	16.06	29.66	23.67	19.18	$C_1^1(4)$	$C_1^1(2)$
195	R3753	20.81	29.16	19.66	19.24	$C_1^1(2)$	$C_1^1(2)$
196	R2895	29.80	29.13	24.69	19.47	No H-bond	No H-bond
197	R313	25.68	23.81	23.83	19.67	$C_1^1(2)$	$C_1^1(2)$
198	R7491	21.11	28.16	27.03	19.79	$C_1^1(4)$	$C_1^1(4)$
199	R3159	31.60	25.84	26.21	20.50	$R_4^4(16)$	$R_4^4(16)$
200	R7924	33.05	28.35	30.86	21.02	No H-bond	No H-bond
201	E8523	30.60	32.59	29.85	21.27	No H-bond	No H-bond
202	E9102	27.62	22.71	21.90	21.59	No H-bond	No H-bond
203	E3335	21.60	20.90	20.65	21.89	No H-bond	No H-bond
204	R4121	25.28	30.69	25.52	22.56	No H-bond	No H-bond
205	R3256	21.96	20.43	20.39	22.61	No H-bond	No H-bond

Table 2: The energies at different steps of those structures that convert during finite-temperature simulations. In the last columns, we show the cluster centres assigned by the FSFDP clustering algorithm and the dominant H-bonding motif at 0k and 300K.

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Cluster Centre	Initial Motif	Final Motif
R526	1.09	2.71	1.97	1.23	FormI	$R_2^2(8)$	$R_2^2(8)$
FormI*	0.00	1.26	1.04	1.69	FormI	$R_2^2(8)$	$R_2^2(8)$
R9521	11.22	12.11	7.08	2.03	FormI	$R_2^2(8)$	$R_2^2(8)$
R106	5.20	4.79	4.80	3.61	FormI	$R_2^2(8)$	$R_2^2(8)$
FormII	13.19	18.47	15.22	10.41	FormII*	$R_2^2(8)$	$R_2^2(8)$
E5749	30.27	32.07	8.06	1.60	E6314	$C_1^1(4)$	$C_1^1(4)$
E6153	30.24	28.47	10.40	2.22	E6314	$C_1^1(4)$	$C_1^1(4)$
E30	5.22	0.22	0.00	2.71	E6314	$C_1^1(4)$	$C_1^1(4)$
R8620	11.26	26.69	22.23	1.62	R113	$R_2^2(8)$	$R_2^2(8)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Cluster Centre	Initial Motif	Final Motif
R399	11.22	15.13	10.33	2.32	R113	$R_2^2(8)$	$R_2^2(8)$
R7447	19.60	19.90	17.10	2.44	R113	$R_2^2(8)$	$R_2^2(8)$
R9957	9.06	10.91	7.39	3.01	R113	$R_2^2(8)$	$R_2^2(8)$
R117	0.52	2.65	3.11	3.42	R113	$R_2^2(8)$	$R_2^2(8)$
R6912	30.41	30.32	23.34	5.67	R113	$R_2^2(8)$	$R_2^2(8)$
R5538	5.74	13.50	11.90	6.41	R113	$R_2^2(8)$	$R_2^2(8)$
R8445	16.11	19.10	14.79	7.98	R113	$R_2^2(8)$	$R_2^2(8)$
R8113	10.68	22.10	21.98	11.21	R113	$R_2^2(8)$	$R_2^2(8)$
R8493	19.15	14.39	15.50	13.40	R113	$R_2^2(8)$	$R_2^2(8)$
R3033	22.44	37.94	16.93	13.48	R113	$R_2^2(8)$	$R_2^2(8)$
R9942	23.29	24.92	21.87	17.09	R113	$R_2^2(8)$	$R_2^2(8)$
R4380	11.56	13.31	13.65	7.59	R1204	$R_2^2(8)$	$R_2^2(8)$
R2806	9.22	12.43	12.65	10.39	R1425	$R_2^2(8)$	$R_2^2(8)$
R8669	24.69	31.33	30.81	15.06	R1493	$C_1^1(4)$	$C_1^1(2)$
R7506	39.98	30.69	27.45	16.41	R1493	$C_1^1(2)$	$C_1^1(2)$
R1318	4.50	10.93	11.86	7.59	R1574	$R_2^2(8)$	$R_2^2(8)$
R2756	11.62	18.47	12.53	8.08	R1574	$R_2^2(8)$	$R_2^2(8)$
R3171	28.72	41.00	20.83	8.85	R1574	$R_2^2(8)$	$R_2^2(8)$
R2205	6.31	17.00	15.14	11.78	R1574	$R_2^2(8)$	$R_2^2(8)$
R6014	13.35	16.34	16.26	10.46	R1687	$R_2^2(8)$	$R_2^2(8)$
R33	7.02	8.48	8.73	11.17	R1687	$R_2^2(8)$	$R_2^2(8)$
R8234	15.34	19.70	19.64	16.03	R1776	$R_2^2(8)$	$R_2^2(8)$
R2639	17.39	22.43	22.48	17.59	R264	$C_1^1(4)$	$R_2^2(8)$
R4197	12.14	21.28	20.94	19.08	R264	$C_1^1(4)$	$R_2^2(8)$
R2841	28.64	36.52	22.00	19.69	R264	$C_1^1(4)$	$R_2^2(8)$
R9966	24.74	31.07	27.85	20.78	R264	$R_2^2(8)$	$R_2^2(8)$
R8134	28.48	30.18	21.04	13.74	R2822	$C_1^1(2)$	$C_1^1(4)$
R2021	20.04	20.15	19.28	11.48	R2863	$C_1^1(4)$	$C_1^1(4)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Cluster Centre	Initial Motif	Final Motif
R8343	15.87	16.50	16.01	10.62	R3184	$R_2^2(8)$	$R_2^2(8)$
R2468	24.81	31.54	13.10	6.79	R336	$R_2^2(8)$	$R_2^2(8)$
R1204	5.15	4.53	5.00	6.93	R336	$R_2^2(8)$	$R_2^2(8)$
R9092	17.98	20.14	18.03	11.15	R336	$R_2^2(8)$	$R_2^2(8)$
R5777	7.14	10.97	11.53	9.85	R346	$R_2^2(8)$	$R_2^2(8)$
R1059	4.37	6.86	5.22	0.65	R4124	$R_2^2(8)$	$R_2^2(8)$
R110	1.94	6.59	4.17	0.97	R4124	$R_2^2(8)$	$R_2^2(8)$
R4816	3.16	1.26	2.18	1.70	R4124	$R_2^2(8)$	$R_2^2(8)$
R4030	6.23	6.70	6.49	1.98	R4124	$R_2^2(8)$	$R_2^2(8)$
R9955	11.59	15.47	15.67	8.78	R4124	$R_2^2(8)$	$R_2^2(8)$
R7038	18.56	26.32	26.42	19.27	R4164	$C_1^1(4)$	$C_1^1(4)$
R5580	20.28	25.87	21.41	20.57	R4164	$C_1^1(4)$	$C_1^1(4)$
R8796	25.03	32.55	28.13	21.49	R4164	$C_1^1(4)$	$C_1^1(4)$
R9692	13.57	26.97	22.85	17.54	R4293	$R_2^2(8)$	$R_2^2(8)$
R4684	7.29	11.46	11.50	10.62	R4600	$R_2^2(8)$	$R_2^2(8)$
R5573	4.70	9.86	9.62	9.93	R4814	$R_2^2(8)$	$R_2^2(8)$
R7196	24.32	26.95	12.11	12.00	R4814	$R_2^2(8)$	$R_2^2(8)$
R5152	10.48	9.34	10.67	10.30	R485	$R_2^2(8)$	$R_2^2(8)$
R4078	24.29	32.18	24.56	19.81	R4973	$R_2^2(6)$	$C_1^1(2)$
R9956	16.29	18.19	19.41	15.80	R4986	$R_2^2(8)$	$R_2^2(8)$
R472	11.84	16.89	16.72	10.96	R5133	$R_2^2(8)$	$R_2^2(8)$
R9137	21.66	25.74	18.28	11.52	R5133	$R_2^2(8)$	$R_2^2(8)$
R4409	6.01	13.63	13.85	11.25	R522	$R_2^2(8)$	$R_2^2(8)$
R390	7.62	20.13	20.17	11.25	R522	$R_2^2(8)$	$R_2^2(8)$
R4958	4.88	6.63	8.42	7.08	R5463	$R_2^2(8)$	$R_2^2(8)$
R1518	12.81	15.10	13.92	7.32	R5463	$R_2^2(8)$	$R_2^2(8)$
R2340	4.97	6.77	8.74	7.54	R5463	$R_2^2(8)$	$R_2^2(8)$
R6695	17.20	15.36	16.52	12.23	R5763	$C_1^1(4)$	$C_1^1(4)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Cluster Centre	Initial Motif	Final Motif
R9568	12.68	13.02	14.37	12.45	R5763	$C_1^1(4)$	$C_1^1(4)$
R6316	15.01	18.80	18.49	16.78	R5853	$R_2^2(8)$	$R_2^2(8)$
R2791	21.53	24.91	9.76	8.05	R5899	$R_2^2(8)$	$R_2^2(8)$
R7072	24.92	27.64	14.33	14.63	R5899	$R_2^2(8)$	$R_2^2(8)$
R3597	22.04	23.36	19.93	15.13	R5959	$C_1^1(4)$	$C_1^1(4)$
R4019	5.58	10.82	11.44	9.68	R6237	$R_2^2(8)$	$R_2^2(8)$
R6622	7.35	9.00	9.61	9.77	R6237	$R_2^2(8)$	$R_2^2(8)$
R8763	5.71	10.65	11.36	9.84	R6237	$R_2^2(8)$	$R_2^2(8)$
R8063	9.32	14.74	14.96	11.94	R6237	$R_2^2(8)$	$R_2^2(8)$
R9827	4.49	8.86	10.04	5.64	R6276	$R_2^2(8)$	$R_2^2(8)$
R6414	10.62	14.66	13.33	5.64	R6276	$R_2^2(8)$	$R_2^2(8)$
R2319	4.52	9.64	10.21	8.88	R6276	$R_2^2(8)$	$R_2^2(8)$
R6168	24.58	22.15	23.12	18.85	R640	$C_1^1(4)$	$C_1^1(4)$
R1960	23.50	26.39	14.18	4.15	R6475	$R_2^2(8)$	$R_2^2(8)$
R2280	6.60	9.41	9.10	4.24	R6475	$R_2^2(8)$	$R_2^2(8)$
R3381	23.89	27.36	14.73	4.36	R6475	$R_2^2(8)$	$R_2^2(8)$
R8515	7.07	12.82	12.60	4.38	R6475	$R_2^2(8)$	$R_2^2(8)$
R7268	3.16	8.26	8.38	4.75	R6475	$R_2^2(8)$	$R_2^2(8)$
R404	5.23	13.28	13.69	4.77	R6475	$R_2^2(8)$	$R_2^2(8)$
R7871	27.47	34.98	22.51	9.84	R6475	$R_2^2(8)$	$R_2^2(8)$
R465	7.44	21.77	21.68	11.20	R6475	$R_2^2(8)$	$R_2^2(8)$
R884	5.14	17.11	16.89	11.24	R6475	$R_2^2(8)$	$R_2^2(8)$
R5452	27.25	31.37	17.50	7.42	R6514	$R_2^2(8)$	$R_2^2(8)$
R4824	23.56	30.00	15.22	10.74	R6514	$R_2^2(8)$	$R_2^2(8)$
R9254	5.65	9.68	10.84	8.65	R6515	$R_2^2(8)$	$R_2^2(8)$
R5816	19.42	17.99	13.80	8.80	R6569	$C_1^1(4)$	$C_1^1(4)$
R4738	25.46	29.25	13.44	9.29	R658	$R_2^2(8)$	$R_2^2(8)$
R8821	31.64	32.55	19.09	15.29	R6786	$C_1^1(2)$	$C_1^1(4)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Cluster Centre	Initial Motif	Final Motif
R7459	22.08	17.53	17.51	15.56	R6786	$D_1^1(2)$	$C_1^1(4)$
R3442	22.16	17.64	17.22	15.90	R6786	$C_1^1(2)$	$C_1^1(4)$
R9268	43.30	40.11	26.61	16.00	R6786	$C_1^1(2)$	$C_1^1(4)$
R7167	34.49	30.41	28.87	19.55	R6786	$D_1^1(2)$	$C_1^1(4)$
R9938	28.38	28.77	27.92	19.78	R6786	$D_1^1(2)$	$C_1^1(4)$
R3685	19.46	18.93	21.31	18.91	R6849	$C_1^1(2)$	$C_1^1(4)$
R8335	28.33	33.11	24.24	19.43	R6849	$C_1^1(2)$	$C_1^1(4)$
R8596	5.63	6.46	6.19	9.92	R702	$R_2^2(8)$	$R_2^2(8)$
R6037	1.12	9.18	8.95	9.97	R7023	$R_2^2(8)$	$R_2^2(8)$
R4338	28.65	22.17	20.92	12.11	R7238	$C_1^1(4)$	$C_1^1(4)$
R8706	26.55	22.17	14.61	9.89	R7263	$C_1^1(2)$	$C_1^1(4)$
R8242	13.27	16.99	15.80	14.32	R7263	$C_1^1(4)$	$C_1^1(4)$
R1039	22.65	22.07	16.22	7.38	R7321	$R_2^2(8)$	$R_2^2(8)$
R4976	25.57	27.67	22.44	10.79	R7321	$R_2^2(8)$	$R_2^2(8)$
R17	12.58	17.85	14.38	9.11	R7642	$R_2^2(8)$	$R_2^2(8)$
R6002	7.33	14.96	12.99	10.53	R7642	$R_2^2(8)$	$R_2^2(8)$
R4347	15.21	20.56	19.88	11.17	R7642	$R_2^2(8)$	$R_2^2(8)$
R161	4.63	12.34	11.47	12.65	R7642	$R_2^2(8)$	$R_2^2(8)$
R9114	28.57	36.91	22.54	18.18	R7806	$C_1^1(4)$	$C_1^1(4)$
R2441	25.29	32.57	24.38	19.57	R7806	$C_1^1(4)$	$C_1^1(4)$
R951	1.11	6.37	6.31	8.11	R8158	$R_2^2(8)$	$R_2^2(8)$
R3351	18.29	21.76	20.37	12.52	R8758	$R_2^2(8)$	$R_2^2(8)$
R9985	19.83	33.77	18.89	8.69	R931	$C_1^1(2)$	$C_1^1(4)$
R5252	9.41	18.05	17.66	9.83	R931	$C_1^1(2)$	$C_1^1(4)$
R6354	26.00	21.11	19.70	9.98	R931	$C_1^1(2)$	$C_1^1(4)$
R3473	27.15	37.23	21.92	14.92	R947	$C_1^1(2)$	$C_1^1(2)$
R4111	21.30	25.43	22.06	20.36	R9631	$C_1^1(4)$	$C_1^1(4)$
R8127	31.90	22.23	23.06	18.78	R9754	No H-bond	No H-bond

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Cluster Centre	Initial Motif	Final Motif
R6177	6.96	12.22	12.57	10.92	R9824	$R_2^2(8)$	$R_2^2(8)$
R4555	12.37	15.64	15.63	11.56	R9824	$R_2^2(8)$	$R_2^2(8)$
R9979	6.78	17.26	17.80	7.07	R986	$R_2^2(8)$	$R_2^2(8)$

Table 3: The energies of those structures that develop disorder at different steps. Orientational and conformational disorder are shown with the abbreviations OD and CD. The initial H-bonding motif is also shown in the last column.

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Disorder	Initial Motif
R10	7.23	9.40	11.92	4.98	CD	$R_2^2(8)$
R9663	16.62	14.69	14.44	8.75	CD	$R_2^2(8)$
R2315	6.71	9.68	9.64	8.80	CD	$R_2^2(8)$
R6327	14.63	22.10	21.97	9.70	CD	$R_2^2(8)$
E8219	29.64	24.17	16.99	10.20	CD	$C_1^1(4)$
R8547	26.43	29.78	18.58	10.72	CD	$C_1^1(4)$
E4262	25.13	20.62	15.32	10.75	CD	$C_1^1(4)$
R2217	17.50	25.62	12.83	10.79	CD	$R_2^2(8)$
R4326	17.58	26.74	22.68	10.81	CD	$R_2^2(8)$
R6689	24.33	23.17	13.99	10.86	CD	$R_2^2(8)$
R1381	9.81	11.06	12.11	10.87	CD	$R_2^2(8)$
R7350	16.65	20.20	15.63	11.06	CD	$R_2^2(8)$
R7558	30.53	37.99	20.42	11.07	CD	$R_2^2(8)$
R2854	9.47	17.12	16.45	11.21	CD	$R_2^2(8)$
R4602	18.22	20.82	18.63	11.21	CD	$R_2^2(8)$
R8207	33.75	27.89	26.82	11.43	CD	No H-bond
R3956	14.38	16.31	16.79	11.44	CD	$C_1^1(4)$
R982	9.65	19.54	15.46	11.62	CD	$R_2^2(8)$
R9865	21.92	27.77	15.84	11.85	CD	$R_2^2(8)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Disorder	Initial Motif
R7905	14.38	23.87	18.37	11.85	CD	$R_2^2(8)$
R7539	13.44	29.26	19.07	11.85	CD	$R_2^2(8)$
R7709	18.64	21.78	22.80	11.85	CD	$R_2^2(8)$
R1108	14.94	17.54	18.42	11.92	CD	$R_2^2(8)$
R6434	21.84	22.77	22.09	12.14	CD	$C_1^1(4)$
R2606	13.08	22.96	24.16	12.43	CD	$R_2^2(8)$
R2141	17.72	16.68	15.52	12.60	CD	$C_1^1(4)$
R8214	14.01	26.85	23.24	12.78	CD	$R_2^2(8)$
R7480	21.02	25.28	18.69	12.83	CD	$R_2^2(8)$
R2494	16.15	14.32	13.68	12.95	CD	$R_2^2(8)$
R9964	24.59	34.26	23.37	12.96	CD	$R_2^2(8)$
E1136	13.97	16.87	13.00	13.08	CD	$C_1^1(4)$
R514	8.88	13.28	14.08	13.13	CD	$R_2^2(8)$
R3315	16.13	18.35	16.79	13.25	CD	$R_2^2(8)$
R3666	15.17	20.83	18.90	13.28	CD	$R_2^2(8)$
R8153	17.63	22.23	18.25	13.56	CD	$R_2^2(8)$
R5125	23.93	18.06	15.79	13.60	CD	$R_2^2(8)$
R3657	29.02	24.62	20.23	13.63	CD	$C_1^1(4)$
E4255	20.49	19.95	21.06	13.68	CD	$C_1^1(4)$
R3230	14.20	16.30	16.97	13.70	CD	$R_2^2(8)$
R5381	19.35	22.11	19.46	13.70	CD	$R_2^2(8)$
R2803	18.08	26.67	22.10	13.90	CD	$R_2^2(8)$
R3757	17.20	32.92	27.36	14.00	CD	$R_2^2(8)$
R9731	15.36	21.78	19.47	14.58	CD	$R_2^2(8)$
R8496	27.84	34.91	19.70	14.68	CD	$D_1^1(2)$
R3565	17.93	21.96	21.91	14.71	CD	$C_1^1(4)$
R9837	11.43	11.33	11.61	14.83	CD	$R_2^2(8)$
R7790	37.11	46.40	20.06	14.87	CD	$C_1^1(4)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Disorder	Initial Motif
R7168	18.67	24.51	16.86	15.01	CD	$R_2^2(8)$
R6620	20.54	21.63	22.78	15.23	CD	$R_2^2(8)$
R9822	25.04	34.40	21.97	15.25	CD	$R_2^2(8)$
R7544	27.17	34.30	25.82	15.28	CD	No H-bond
R2328	19.12	22.99	23.87	15.42	CD	$R_2^2(8)$
R9071	22.22	29.71	23.22	15.55	CD	$R_2^2(8)$
R5285	22.98	25.97	20.43	15.59	CD	$R_2^2(8)$
R8055	26.95	26.26	17.48	15.76	CD	$R_2^2(8)$
R4822	26.26	27.46	18.82	15.86	CD	$C_1^1(2)$
R6999	22.33	21.91	20.06	15.92	CD	$R_4^4(16)$
R3070	19.42	25.05	26.26	15.95	CD	$C_1^1(4)$
R4006	21.25	35.20	27.22	15.96	CD	$C_1^1(4)$
E1	17.80	22.97	17.48	16.13	CD	$C_1^1(4)$
R8346	15.18	19.41	20.46	16.13	CD	$R_2^2(8)$
R4452	26.23	30.41	22.47	16.15	CD	$R_2^2(8)$
R649	23.72	20.29	21.26	16.19	CD	$D_1^1(2)$
R6531	17.68	26.74	21.73	16.41	CD	$R_2^2(8)$
R9080	25.18	34.31	24.22	16.41	CD	$C_1^1(2)$
E9327	34.03	36.43	19.79	16.56	CD	No H-bond
R6770	20.78	21.21	20.24	16.61	CD	$R_2^2(8)$
R8050	18.22	24.53	24.54	16.66	CD	$R_2^2(8)$
R3345	19.71	20.97	18.71	16.75	CD	$C_1^1(4)$
R8803	18.03	22.65	22.14	16.78	CD	$C_1^1(4)$
R9403	22.26	26.49	21.09	16.83	CD	$R_2^2(8)$
R2802	23.40	26.64	20.67	16.86	CD	$C_1^1(2)$
E2775	25.94	28.40	22.79	16.98	CD	$C_1^1(4)$
R8810	21.63	30.91	25.02	17.11	CD	$R_2^2(8)$
R1569	24.71	21.91	21.96	17.19	CD	$C_1^1(2)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Disorder	Initial Motif
R6813	26.97	24.21	22.91	17.21	CD	$R_4^4(16)$
R3206	24.58	21.75	21.76	17.23	CD	$C_1^1(2)$
E9115	28.66	29.20	23.14	17.24	CD	$C_2^2(6)$
E8088	28.25	39.37	20.96	17.30	CD	$C_1^1(2)$
R8760	16.16	20.99	21.14	17.37	CD	$C_1^1(4)$
R1301	20.13	26.53	23.45	17.56	CD	$C_1^1(4)$
R7787	18.38	23.35	23.82	17.74	CD	$R_2^2(8)$
E7181	16.61	25.04	22.96	17.81	CD	$C_1^1(2)$
R9916	28.89	32.31	22.29	17.93	CD	$R_2^2(8)$
R2330	15.04	23.63	22.69	17.93	CD	$R_2^2(8)$
R1796	32.25	37.46	26.05	17.98	CD	$C_1^1(4)$
R3420	19.92	25.70	22.85	18.30	CD	$R_2^2(8)$
R8453	28.20	30.08	25.08	18.30	CD	No H-bond
R6512	19.02	26.51	24.14	18.35	CD	$R_2^2(8)$
E5283	17.95	30.65	25.08	18.55	CD	$C_1^1(2)$
R9958	22.49	24.66	27.70	18.77	CD	$R_2^2(8)$
R6704	37.75	44.44	21.98	18.81	CD	$D_1^1(2)$
E8979	25.12	25.05	24.03	19.14	CD	$R_2^2(8)$
R5429	31.95	26.91	27.52	19.30	CD	No H-bond
R2114	29.40	42.01	25.97	19.34	CD	$C_1^1(4)$
R8081	26.31	34.04	28.70	19.34	CD	No H-bond
E4178	23.32	31.68	27.84	19.39	CD	No H-bond
E9860	20.13	19.32	20.13	19.40	CD	$C_1^1(4)$
R8602	21.53	24.99	23.03	19.43	CD	$C_1^1(4)$
R2183	17.20	27.82	23.23	19.48	CD	$C_1^1(4)$
R4922	22.08	26.74	23.01	19.89	CD	$C_1^1(4)$
R4799	28.80	28.80	25.73	19.91	CD	No H-bond
R594	30.54	30.76	23.09	19.95	CD	No H-bond

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Disorder	Initial Motif
R5602	27.63	31.14	22.98	20.11	CD	$C_1^1(4)$
R3706	28.47	28.70	27.05	20.12	CD	$D_1^1(2)$
R5290	25.96	27.19	26.87	20.13	CD	No H-bond
R8601	24.45	24.04	24.07	20.14	CD	No H-bond
R5970	26.64	25.25	21.75	20.17	CD	$R_4^4(16)$
R9414	36.06	40.65	23.28	20.17	CD	$C_1^1(4)$
E8890	14.82	28.80	21.88	20.27	CD	$C_1^1(4)$
R5877	33.90	34.84	26.08	20.31	CD	$R_2^2(8)$
R6637	39.63	45.17	26.72	20.49	CD	$D_1^1(2)$
R5289	26.56	43.13	23.62	20.55	CD	$C_1^1(4)$
R6651	24.69	23.95	23.94	20.56	CD	No H-bond
R9729	19.24	32.41	24.51	20.65	CD	$C_1^1(4)$
R8580	26.96	34.36	27.33	20.66	CD	$D_1^1(2)$
E3683	31.48	48.54	21.94	20.97	CD	$C_1^1(4)$
R7989	24.56	32.91	25.11	21.18	CD	$C_1^1(4)$
R6705	17.62	24.46	23.48	21.31	CD	No H-bond
R1918	17.69	24.76	23.64	21.34	CD	$C_1^1(4)$
R7631	14.47	24.99	24.75	21.47	CD	$C_1^1(4)$
R3742	22.51	33.04	28.84	21.49	CD	$R_2^2(8)$
R7902	28.81	30.58	28.26	21.51	CD	$C_1^1(4)$
R5391	17.00	24.16	23.49	21.55	CD	$C_1^1(4)$
R2337	21.24	33.41	29.01	21.56	CD	$C_1^1(4)$
R8749	20.58	34.43	23.12	21.60	CD	$C_1^1(4)$
R3495	38.72	44.51	23.52	21.61	CD	No H-bond
R8228	25.21	33.67	23.79	21.64	CD	$C_1^1(4)$
E8045	16.75	21.03	24.33	21.65	CD	$R_2^2(8)$
R948	19.78	31.57	27.46	21.74	CD	$C_1^1(4)$
R6561	25.47	28.94	28.24	21.81	CD	$R_2^2(8)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Disorder	Initial Motif
R5229	21.78	22.74	22.96	21.91	CD	$R_2^2(8)$
R7004	34.29	39.24	23.85	21.93	CD	$R_2^2(8)$
R3881	35.85	42.01	26.94	21.95	CD	$R_2^2(8)$
R9405	38.33	45.19	27.92	21.97	CD	$C_1^1(4)$
R6091	16.34	22.35	21.98	21.99	CD	$C_1^1(4)$
R2108	17.61	24.40	23.40	22.10	CD	$C_1^1(4)$
R5568	37.17	41.54	26.15	22.10	CD	$R_2^2(8)$
R6881	29.68	26.85	26.68	22.63	CD	No H-bond
R3674	18.00	26.40	27.59	22.64	CD	$R_2^2(8)$
E5672	32.79	30.69	25.75	22.75	CD	$R_2^2(8)$
R7340	38.79	48.74	26.56	22.82	CD	$C_1^1(4)$
R9743	32.66	32.09	26.68	22.89	CD	$R_2^2(8)$
R4761	30.37	33.91	25.61	23.01	CD	No H-bond
R6916	25.65	34.13	28.24	23.81	CD	$R_2^2(8)$
R6724	23.97	40.54	30.85	23.82	CD	$C_1^1(4)$
R2896	31.87	38.36	29.57	24.28	CD	$C_1^1(4)$
R7900	41.79	50.04	29.50	24.42	CD	$D_1^1(2)$
R8888	26.61	34.70	30.75	26.01	CD	No H-bond
R9720	15.88	15.33	15.43	14.65	OD	$R_2^2(8)$
R8703	26.73	28.32	26.13	16.47	OD	$R_2^2(8)$
R1104	26.02	24.41	19.68	17.56	OD	$R_3^3(12)$
R6948	25.92	33.71	17.29	17.63	OD	$C_1^1(4)$
R9051	19.57	18.95	19.94	18.03	OD	$R_2^2(8)$
R9207	13.30	17.63	18.49	18.54	OD	$C_1^1(4)$
R4893	19.63	28.14	22.40	19.06	OD	$R_2^2(8)$
R5310	22.80	22.28	19.82	19.08	OD	$R_6^6(24)$
R726	16.32	22.03	21.07	19.10	OD	$C_1^1(4)$
R5794	15.36	23.37	23.99	19.65	OD	$C_1^1(4)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Disorder	Initial Motif
E7035	20.01	24.60	24.26	19.76	OD	$C_1^1(4)$
R2589	18.53	27.53	26.30	20.29	OD	$C_1^1(4)$
R7163	36.58	30.08	21.85	20.52	OD	$R_4^4(16)$
R8420	27.51	34.37	24.19	20.73	OD	$C_1^1(4)$
R7732	26.02	28.68	24.32	21.46	OD	$R_6^6(24)$
R6518	34.51	35.48	24.34	21.46	OD	$R_6^6(24)$
R6832	19.70	19.53	20.12	-	OD	$R_2^2(8)$
R8494	28.31	38.01	21.99	-	OD	$C_1^1(4)$
R9612	24.12	22.97	22.41	-	OD	$R_2^2(8)$
R7538	24.61	33.17	22.64	-	OD	$R_2^2(8)$
R3807	16.55	22.80	22.67	-	OD	$R_2^2(8)$
R3350	21.06	24.18	23.48	-	OD	$R_2^2(8)$
R7576	20.37	23.28	23.51	-	OD	$R_6^6(24)$
E2917	18.63	24.31	23.75	-	OD	$C_1^1(4)$
R7842	27.56	28.60	23.94	-	OD	$C_1^1(4)$
E4575	21.61	27.48	24.00	-	OD	$R_2^2(8)$
R8247	32.88	33.05	24.93	-	OD	$R_3^3(6)$
R5420	18.88	31.99	25.49	-	OD	$R_2^2(8)$
R8433	23.00	33.59	25.50	-	OD	$C_1^1(4)$
R9752	18.69	23.23	25.58	-	OD	$R_6^6(24)$
R7831	28.39	26.21	25.64	-	OD	$C_1^1(4)$
R3260	24.34	37.18	25.76	-	OD	$C_1^1(4)$
R5160	26.62	35.53	25.96	-	OD	$R_4^4(16)$
R7579	21.61	29.85	26.00	-	OD	$R_2^2(8)$
R1870	24.36	42.26	26.12	-	OD	$C_1^1(4)$
R9789	24.06	32.86	26.19	-	OD	$C_1^1(4)$
R4065	19.24	30.42	26.26	-	OD	$R_2^2(8)$
R7999	35.05	45.79	26.30	-	OD	$R_2^2(8)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Disorder	Initial Motif
R8129	30.20	43.83	26.41	-	OD	$C_1^1(4)$
R6173	21.97	30.85	26.71	-	OD	$C_1^1(4)$
E9026	23.86	26.97	26.83	-	OD	No H-bond
R6145	19.31	28.39	26.86	-	OD	$R_2^2(8)$
R5728	31.00	31.17	26.88	-	OD	$R_6^6(24)$
R4550	22.08	29.76	27.01	-	OD	$R_2^2(8)$
R9918	30.59	41.40	27.30	-	OD	$C_1^1(4)$
R6021	22.43	27.62	27.46	-	OD	$R_2^2(8)$
R9596	32.72	30.54	27.69	-	OD	No H-bond
R8833	33.59	44.71	27.88	-	OD	$R_2^2(8)$
R4571	20.78	28.59	28.03	-	OD	$C_1^1(4)$
R8951	38.85	40.14	28.14	-	OD	No H-bond
R4174	33.44	42.70	28.23	-	OD	$C_1^1(4)$
R9045	29.66	30.66	28.46	-	OD	$C_1^1(4)$
R9379	28.97	33.26	28.70	-	OD	$C_1^1(4)$
R4101	31.30	30.12	28.85	-	OD	$D_1^1(2)$
R9133	32.92	37.09	29.12	-	OD	No H-bond
R9443	29.87	30.21	29.42	-	OD	$C_1^1(4)$
R4977	20.93	32.72	29.49	-	OD	$C_1^1(4)$
R5532	32.10	35.68	29.75	-	OD	$D_1^1(2)$
R7816	28.30	29.16	30.19	-	OD	$R_4^4(8)$
R8489	26.00	38.50	30.42	-	OD	$C_1^1(4)$
R6951	30.10	42.40	30.58	-	OD	$C_1^1(4)$
R5073	42.06	41.29	31.00	-	OD	No H-bond
R5128	34.99	29.14	31.15	-	OD	$R_6^6(24)$
R2837	24.76	37.96	31.59	-	OD	$C_1^1(4)$
R8767	25.38	35.31	32.12	-	OD	$C_1^1(4)$
R5005	27.29	50.25	40.68	-	OD	$C_1^1(4)$

Continued on next page

IDs	CSP_0 kJ mol^{-1}	EM kJ mol^{-1}	NVT kJ mol^{-1}	NPT kJ mol^{-1}	Disorder	Initial Motif
R1559	23.06	25.63	-	-	OD	$R_2^2(8)$
R8841	24.33	40.98	-	-	OD	$C_1^1(4)$
R9556	24.95	29.94	-	-	OD	$R_2^2(8)$
R7332	27.18	28.58	-	-	OD	$C_1^1(4)$
R7213	27.40	33.93	-	-	OD	$R_6^6(24)$
R7210	27.78	32.03	-	-	OD	$C_1^1(4)$
R7845	29.51	35.45	-	-	OD	$R_6^6(12)$
R8336	30.14	39.97	-	-	OD	$C_1^1(4)$
R6917	32.32	39.64	-	-	OD	$D_3^3(10)$
R5867	34.07	31.51	-	-	OD	$R_3^3(12)$
R9228	36.97	53.53	-	-	OD	$C_1^1(4)$

References

(1) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C.; IUCr, The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72*, 171–179.

(2) Ostrowska, K.; Kropidowska, M.; Katrusiak, A. High-pressure crystallization and structural transformations in compressed R, S -ibuprofen. *Cryst. Growth Des.* **2015**, *15*, 1512–1517.

(3) Derollez, P.; Dudognon, E.; Affouard, F.; Danède, F.; Correia, N. T.; Descamps, M. Ab initio structure determination of phase II of racemic ibuprofen by X-ray powder diffraction. *Acta Crystallogr. Sect. B Struct. Sci.* **2010**, *66*, 76–80.

(4) King, M. D.; Buchanan, W. D.; Korter, T. M. Understanding the terahertz spectra of crystalline pharmaceuticals: Terahertz spectroscopy and solid-state density functional theory study of (S)-(+)-ibuprofen and (RS)-ibuprofen. *J. Pharm. Sci.* **2011**, *100*, 1116–1129.