Exploration and Optimization in Crystal Structure Prediction: Combining Basin Hopping with Quasi-Random Sampling

Shiyue Yang and Graeme M. Day*

School of Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom

E-mail: G.M.Day@soton.ac.uk

Abstract

We describe the implementation of a Monte Carlo basin hopping global optimization procedure for the prediction of molecular crystal structures. The basin hopping method is combined with quasi-random structure generation in a hybrid method for crystal structure prediction, QR-BH, which combines the low-discrepancy sampling provided by quasi-random sequences with basin hopping's efficiency at locating low energy structures. Through tests on a set of single-component molecular crystals and co-crystals, we demonstrate that QR-BH provides faster location of low energy structures than pure quasi-random sampling, while maintaining the efficient location of higher energy structures that are important for identifying important polymorphs.

1 Introduction

The ability to predict the crystal structure that a molecule will adopt, in advance of the crystallisation experiment or even in advance of synthesis, has great implications in several

areas of materials science. The past two decades have seen important progress in computational methods for crystal structure prediction (CSP), with almost all current methods based on performing a search for the local minima on the high dimensional energy surface representing the energy as a function of the variables that describe a crystal structure.[1,2]

The usual assumption in using these methods is that the global minimum on the potential energy surface corresponds to the most likely observable crystal structure. Focusing on locating the global minimum, many approaches has been developed for CSP, such as Monte Carlo simulated annealing[3], genetic algorithms[4–6] and particle swarm optimization.[7]

However, higher energy crystal structures are also often observed. This is clear from the prevalence of polymorphism in molecular crystals,[8] where a molecule can adopt more than one crystal structure. Polymorphism is sometimes due to changes in temperature or pressure, which can alter the free energy ordering of structures. These effects can be accounted for in prediction methods by inclusion of entropy and zero-point vibrational contributions to the energy.[9–11] However, different crystal structures can often be crystallised at the same thermodynamic conditions, sometimes from the same experiment (concomitant polymorphs). It has been estimated, based on a large-scale computational study,[12] that nearly 80% pairs of observed polymorphs are monotropic, i.e. their free energies do not cross below their melting temperature. The identification of these metastable polymorphs is important in many applications of CSP, such as polymorph screening of pharmaceuticals,[13] and computer-guided discovery of functional materials, where high energy structures sometimes exhibit the most attractive properties.[14,15]

Thus, for CSP to be predictive of all observable crystal structures of a molecule, the structure search method must not be treated as simply a global energy minimisation problem, but should exhaustively explore the energy landscape for possible structures within the energy range above the global minimum in which observed structures can be located. Therefore, some CSP algorithms, such as low-discrepancy, quasi-random sampling place emphasis on exploring the structural landscape as uniformly as possible for all low energy structures.[16,17]

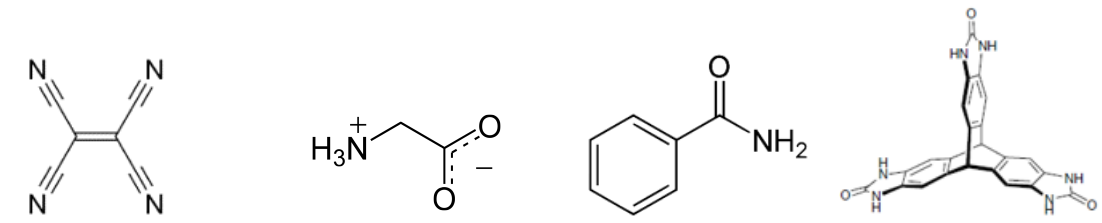
The energy range over which it is important to identify possible crystal structures can be defined by the energy range of observed polymorphism. Most observed polymorphs are separated by only a few kJ mol^{-1} in lattice energy,[9] although this range can extend above 10 kJ mol^{-1} in rare cases, or where polymorphs are accessed through desolvation of solvated crystals.[14,18] Although this defines a narrow energy window for observable structures, the weak interactions between organic molecules mean that large numbers of different structures are often possible within this range. For small organic molecules, a small energy range can include tens or hundreds of putative crystal structures.[19]

The importance of the higher energy structures, in addition to the global minimum, creates a tension in designing methods for CSP between efficiency in locating the global energy minimum quickly and time spent exploring the landscape to locate all potentially observable crystal structures. In this work, we describe a hybrid approach, where quasi-random (QR) sampling is used to seed multiple Monte Carlo basin hopping (BH) searches; we refer to the method as QR-BH. The role of quasi-random sampling in QR-BH is to provide a broad sampling of the energy landscape, while basin hopping efficiently locates low energy structures from these starting points. The method is benchmarked on a set of organic molecular crystals and co-crystals to explore its efficiency, how it is influenced by the temperature used in basin hopping and the number of quasi-random seeds vs basin hopping steps used in the search.

2 Computational Details

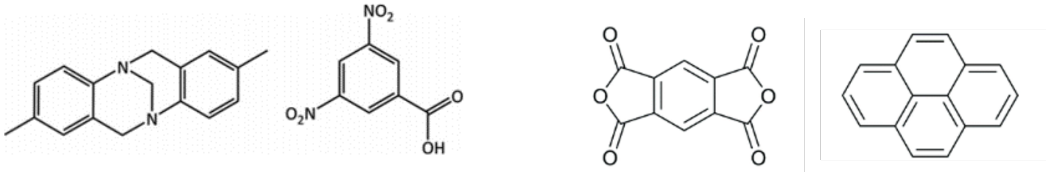
2.1 Choice of systems

Six crystal systems (Fig. 1), including single-component crystals and co-crystals, were chosen as representative of different applications of CSP, and of systems held together by different strengths of intermolecular interactions. Tetracyanoethylene is a planar molecule with weak intermolecular interactions, the zwitterionic geometry of glycine leads to strong intermolec-



(a) (b) (c) (d)

Figure 1: Single component crystal systems studied in this work. (a) tetracyanoethylene, (b) glycine, (c) benzamide and (d) TTBI.



(a) (b)

Figure 2: Co-crystal systems. (a) XAFQAZ (b) PYRPMA

ular hydrogen bonding interactions, while benzamide represents molecules with a mixture of hydrogen bonding and van der Waals interactions between aromatic rings. The fourth single-component system, a triptycene trisbenzimidazolone (TTBI, Fig.1d), is a larger molecule that has been shown to form several porous polymorphs located in low-density, high-energy regions of the lattice energy landscape; this molecule is included to test the location of important high-energy structures by the QR-BH crystal structure searching algorithm.

The single-component systems were only investigated here with one molecule in the asymmetric unit of the crystal structures ($Z' = 1$). To investigate the behaviour of the QR-BH method thoroughly and as a challenge to investigate the efficiency of the method on more complex systems, we applied the algorithm to two co-crystal systems (Fig 2), in which the presence of two independent molecules leads to more degrees of freedom and, thus, more challenging energy landscapes for structure prediction. The first co-crystal, which we refer to by the Cambridge Structural Database (CSD)[20] reference code of its known structure, XAFQAZ, is a hydrogen bonded complex between 2,8-Dimethyl-6H,12H-5,11-methanodibenzo[b,f][1,5]diazocine (Tröger's base) and 3,5-dinitrobenzoic acid and was found

to be a challenging target in the 6th blind test of crystal structure prediction.[21] As a second co-crystal, we chose the complex between two planar molecules - pyrene and pyromellitic dianhydride - which is held together by weaker, less directional intermolecular interactions. We also refer to this system by the CSD reference code of its known crystal structure,[22] PYRPMA.

2.2 Quasi-Random Search

Quasi-random (QR) structure generation was performed using the Global Lattice Energy Explorer code; the method is described in detail in our earlier paper.[17] During the generation of trial structures, a low-discrepancy sequence of vectors is generated by the Sobol method[23] and each vector is mapped onto the structural degrees of freedom of the unit cell, including molecular positions and orientations, as well as lattice parameters that are not constrained by space group symmetry. We use the SAT-expand version of the quasi-random crystal structure generation method,[17] in which the target volume for the unit cell is set as the sum of molecular volumes (which are calculated from the volume of a box enclosing all atoms in the molecule). The separating axis theorem (SAT) is used to detect overlapping molecular convex hulls, which indicate clashing molecules. Such clashes are removed through expansion of the lattice parameters in the direction required to separate overlapping molecules. Structures in which intermolecular clashes could not be relieved with unit cell expansion of less than 2.5 times the original target volume are rejected without lattice energy minimization.

Each trial structure was then lattice energy minimized using the DMACRYS software[24] to locate the nearest (downhill) local minimum on the lattice energy surface. Molecules are held rigid throughout at their DFT optimized geometries and intermolecular interactions are modelled using an empirically parametrized exp-6 repulsion-dispersion potential[25] combined with atomic multipoles for electrostatic interactions from a distributed multipole analysis.[26] Full details of lattice energy minimization are provided in the supporting information.

The method was designed to provide a uniform and unbiased sampling of the lattice

energy surface, which is important for fully exploring the structural diversity available to a molecule in forming stable crystal structures. The lack of bias in the search towards identifying low energy structures makes it effective at locating metastable crystal structures, while it has also been shown to usually find the global energy minimum early in a search.[17] A further advance of the approach is its parallelizability: each local energy minimization is independent, so the minimization of all trial structures can be performed in parallel if sufficient processors are available.

QR searches were continued for a specified number of successful lattice energy minimizations. The database of optimized crystal structures was then analyzed for duplicates to generate a list of unique predicted crystal structures, and to count the number of times that each structure was located. Details of duplicate identification are provided in the supporting information.

2.3 Basin Hopping

Basin hopping (BH) is a global optimization approach combining local energy minimization and a Monte Carlo sampling method, where local minimization is introduced after each random perturbation.[27,28] In other words, rather than single-point energy evaluation, the objective function is the locally minimized energy given by $\tilde{E}(x) = \min[E(x)]$, meaning that the energy associated with each point in configuration space, x, is the energy of the local minimum that is reached upon energy minimization from that point: $\min[E(x)]$. Thus, the acceptance probability of perturbing structure a to b is calculated by

$$acc(a \rightarrow b) = \min[e^{-(\tilde{E}_b - \tilde{E}_a)/kT}, 1], \tag{1}$$

where k is the Boltzmann constant and T is a chosen temperature.

Five types of perturbation are used in order to sample all degrees of freedom in a unit cell. Molecular perturbations include translation in a random direction and quaternion

rotation around a random axis passing through the molecular center of mass. All molecular perturbations were applied to the molecule(s) in the asymmetric unit; perturbations of the other molecules in the unit cell were generated by symmetry, so as to maintain the space group symmetry. Unit cell perturbations include unit cell length changes (taking into account correlated lattice parameters in some space groups), unit cell angle changes (where allowed by space group symmetry) and unit cell volume changes. To avoid the unphysical region of the *exp-6* interatomic potential, the distances between molecules were calculated after perturbation and the Monte Carlo move was rejected if any interatomic distance was shorter than the sum of covalent radii of the two elements + 0.3 Å.

The probability of making each perturbation type and cut-off magnitude of each type of perturbation are important parameters to be determined. The probability of applying each type of perturbation was determined according to the degrees of freedom (DOF) leading to an energy change by $P_{move} = (DOF_{move}/DOF_{total})$. DOF_{move} is the number of DOF related to the specific move, e.g. $DOF_{move} = 3$ for translation of one molecule. DOF_{total} is the total number of degrees of freedom. The step size of each perturbation was sampled from a uniform distribution within the range defined by the cutoff, except for unit cell angles. To discourage angles from moving outside the target range 45 to 135°, instead of generating a random number in the range $(-1, 1)$, the range of the random number is shifted based on the current angle (θ) by

$$shift = \begin{cases} (\theta_c - \theta)/(\theta_u - \theta_c) & \text{if } \theta \geq \theta_c \\ (\theta_c - \theta)/(\theta_c - \theta_l) & \text{if } \theta < \theta_c \end{cases} \tag{2}$$

where θ_c is the central angle, usually 90°; θ_l and θ_u are lower and upper limits, being 45° and 135° by default. Hence the range of random perturbations is shifted towards 90° and then scaled by the cut-off. Note that unit cell angles are only constrained when performing a perturbation and are unconstrained during the local energy minimizations. The non-uniform sampling of unit cell angle perturbations means that the simulation does not fulfill detailed

balance, which is unimportant here because the BH approach focuses on prediction of local minima on the energy landscape, rather than a distribution at equilibrium. Since molecular perturbations were applied to the asymmetric unit, the cutoff on volume change depended on the number of molecules in the primitive unit cell to eliminate the impact from different space groups with different numbers of symmetry operations.

During the BH trajectory, new structures were obtained by perturbing the unminimized structure from the previous step, rather than applying perturbations to the minimized structures. One reason for this decision is that, since unit cell angles are not constrained during local minimization, unit cells can become quite flat after minimization (i.e. having very acute or obtuse unit cell angles). These flat unit cells can correspond to physically realistic structures, but lead to difficulties in applying further perturbation and minimization.

2.4 The QR-BH Combined Method

The strategy that we have developed in this work is to combine BH with the quasi-random (QR) sampling approach. In our pure QR method, the conversion of each quasi-random vector into a trial crystal structure was followed by local energy minimization. Here, the single local energy minimization of each QR trial structure is followed by a BH trajectory to sample local configurational space. Our intention is that the quasi-random seeding of basin hopping simulations maintains some of the benefits of the low discrepancy approach, such as its uniform sampling of the configuration space of crystal packing, while benefiting from the efficiency of BH at moving towards low energy structures. The approach also maintains a certain level of parallelizability: each BH trajectory can be performed in parallel.

As well as the perturbation cutoffs and temperature used in the BH acceptance test, the behavior of the combined, QR-BH, search is influenced by the number of quasi-random seed structures and the length of each BH trajectory. Unless otherwise stated, all BH trajectories in a search were run for the same, fixed total number of steps. The job of the quasi-random seeds is to sample the energy landscape widely and evenly, while each local region

is then efficiently explored in the BH search. Conceptually, the most efficient search could be achieved when each BH trajectory samples a separate region of the energy landscape and these local regions combined make up the entire energy landscape.

2.5 On-the-fly Clustering

As an alternative to running all BH trajectories for the same, fixed number of steps, we also implemented a version of the QR-BH search that involves on-the-fly clustering of crystal structures and the termination of BH trajectories that sample the same regions of phase space. In this version, a BH trial is truncated if the new minimum reached from the perturbed structure already exists in the database of structures located thus far. If a BH trajectory is truncated, it is replaced by a new trial initiated from the next unused quasi-random seed in the Sobol sequence, so that the number of active BH trajectories remains constant. To ensure that one BH trajectory is kept active within each region of configurational space, a trajectory is not truncated if it locates a previous structure from its own history and, when two trajectories have located the same structure, the lowest trajectory number (ie. starting from the earliest quasi-random seed) is kept active to continue sampling.

3 Results and Discussion

3.1 Single-Component Crystals

We initially applied the QR-BH algorithm to a set of single component molecular crystal systems (Fig. 1), and compared the energy landscape and sampling efficiency with the pure QR search method. For tetracyanoethylene, benzamide and glycine, 7 space groups were sampled: $P\bar{1}$, $P2_1$, $P2_1/c$, $C2/c$, $P2_12_12_1$, $F2dd$ and $I4_1/a$, all with one molecule in the asymmetric unit ($Z' = 1$). These were chosen as a set of common space groups for organic molecular crystals covering different crystal systems, symmetry elements and centerings, which could lead to differing complexities of their energy landscapes.

The sampling efficiency is affected by all parameters used to define the behaviour of the BH trials, in this case including the temperature used for trial acceptance, perturbation cutoffs, the number of parallel seeds and the length of each BH trial. The perturbation cutoff was adjusted so that different types of structural perturbation lead to similar energy changes (Table S2). In these initial tests, the temperature was set to 3000K to permit acceptance of increases in energy of up to about 24 kJ/mol according to Boltzmann distribution (lower temperatures are investigated below). Each QR-BH simulation involves 10,000 local lattice energy minimizations generated from 100 parallel BH trials (each started from a different QR seed) and 100 BH steps in each trial. Results of these searches were compared to pure QR structure searches using the same number of energy minimizations in each space group. This length of run is deliberately oversampled to generate better statistics with which to compare the methods; fewer steps are normally required for such simple systems.[17] Because of the stochasticity of the QR-BH process, every simulation (i.e. each molecule + space group combination) was repeated three times and we examine the average and variability of the behaviour between repeats. Since the quasi-random sequence is deliberately deterministic, we used the same initial quasi-random structures in each repeat, but different random seeds for Monte Carlo moves in the BH trials.

3.1.1 Locating the global energy minimum

The effectiveness and efficiency of the searches were initially evaluated according to the speed with which the global lattice energy minimum was located in each search. Because the space group symmetry is constrained within each search, we treat each combination of molecule and space group as an independent landscape in our analysis.

Our first observation is that, for each molecule-space group combination, the QR search and the three repeats of QR-BH all find the same global energy minimum structure. Therefore, we are confident that the true global energy minimum has been located for each system. The efficiency of the methods is measured by the number of steps required to locate the global

minimum of each system (Table 1). This is defined straightforwardly in the QR search as the number of accepted (lattice energy minimized) quasi-random seeds until the first instance that the global minimum is located. For the QR-BH search, we define the step as the product of (seed number)×(BH step number) for the first hit to the global minimum. The true computational expense of the QR-BH calculation depends on the parallelization strategy and order in which calculations are performed, but we feel that this definition fairly compares the QR-BH to the QR search. Other comparisons between methods, which do not rely on this definition of computational cost, are presented later.

As observed in previous work,[17] the quasi-random search is often effective at locating the global minimum energy structure early in a search and this is borne out for these three molecules. As single-component crystals of small, rigid molecules, these energy landscapes are relatively simple and have fairly small numbers of distinct local minima (see below). The ease of finding the global minimum in energy is also thought to be due, in part, to lowest energy structures having the widest basins of attraction.[29] Thus, the global energy minimum is easily located, especially in space groups with few symmetry operations. For example, in $P\bar{1}$, the global minimum is located as the first or second generated structure for each of the three molecules; in the case of TCNE, the next lowest energy structure in $P\bar{1}$ lies 8 kJ mol^{-1} higher in energy and over 90% of energy minimizations lead to the global minimum.

As a broad observation, we find that QR-BH locates the global energy minimum in either the same number or fewer steps than the pure QR search. The mean number of steps to find the global minimum (over the three QR-BH repeats) is always lower, taking, on average, 74% of the steps needed by the pure QR searches. It is for the systems where the global minimum is located later in the search that the improved efficiency of QR-BH is clearest: the global minimum in space group $F2dd$ is first located after hundreds of steps in the pure QR search for all three molecules, but is found much earlier - after fewer than 100 steps - for most of (7 of 9) the QR-BH searches.

The repeats of QR-BH mostly show comparable behaviour, finding the global minimum

Table 1: Steps required to locate the global energy minimum for the single component crystal systems tetracyanoethylene (TCNE), benzamide and glycine in each of 7 space groups (SG), using the quasi-random seeded basin hopping (QR-BH) and quasi-random (QR) methods. For the QR-BH searches, we report the results of each individual run and the mean step number over the three repeats.

TCNE	QR-BH				QR
SG	repeat 1	repeat 2	repeat 3	mean	
$P\bar{1}$	1	1	1	1	1
$P2_1$	18	26	24	23	38
$P2_1/c$	6	15	12	11	15
$C2/c$	3	3	3	3	3
$P2_12_12_1$	2	2	2	2	2
$F2dd$	60	57	12	43	562
$I4_1/a$	3	3	3	3	3
benzamide	QR-BH				QR
SG	repeat 1	repeat 2	repeat 3	mean	
$P\bar{1}$	2	2	2	2	2
$P2_1$	9	24	2	12	55
$P2_1/c$	99	28	140	89	160
$C2/c$	94	68	94	85	94
$P2_12_12_1$	30	4	8	14	201
$F2dd$	80	711	36	276	288
$I4_1/a$	60	135	135	110	135
glycine	QR-BH				QR
SG	repeat 1	repeat 2	repeat 3	mean	
$P\bar{1}$	1	1	1	1	1
$P2_1$	6	11	11	9	11
$P2_1/c$	7	7	7	7	7
$C2/c$	60	76	15	50	170
$P2_12_12_1$	4	4	4	4	4
$F2dd$	195	35	24	85	195
$I4_1/a$	6	6	6	6	6

with similar efficiency in independent runs starting from the same QR starting points. However, we see greater variability between QR-BH repeats in the cases where the pure QR search was slowest at locating the global minimum. The most extreme case is for benzamide in space group $F2dd$, where the pure QR search took 288 steps before the first hit of the global minimum. Two runs of QR-BH showed a big improvement, locating the global minimum after 36 and 80 steps, but the other repeat required 711 steps.

Table 2: Number of hits of the global energy minimum for the single component crystal systems tetracyanoethylene (TCNE), benzamide and glycine in each of 7 space groups. All searches involved a total of 10,000 lattice energy minimizations. For QR-BH, we report the mean from the three independent runs, which all start from the same 100 QR seed structures.

space	TCNE		benzamide		glycine	
group	QR-BH	QR	QR-BH	QR	QR-BH	QR
$P\bar{1}$	9132	9007	1123	921	3439	3060
$P2_1$	1575	1347	649	542	749	860
$P2_1/c$	309	191	95	50	377	355
$C2/c$	473	575	71	61	91	103
$P2_12_12_1$	667	724	240	107	635	687
$F2dd$	76	50	49	35	21	18
$I4_1/a$	443	417	90	80	116	127

We also monitored the number of hits to the global minimum energy structure in each system (Table 2). For these three molecules, we see small differences between the pure QR and the QR-BH searches, perhaps because their energy landscapes are relatively simple. However, in 15 of the 21 systems, the global minimum is sampled more frequently by QR-BH than QR, reflecting the bias that is introduced towards lowering the energy when local energy minimization of QR structures (the pure QR method) is replaced by a short basin hopping trajectory (as in QR-BH). Thus, despite occasional variability between runs, these initial tests showed the QR-BH algorithm to be stable and efficient at locating the global minimum energy crystal structures, with a moderate improvement over pure QR searching in how quickly it locates the global minimum energy structure.

3.1.2 Sampling of low energy crystal structures

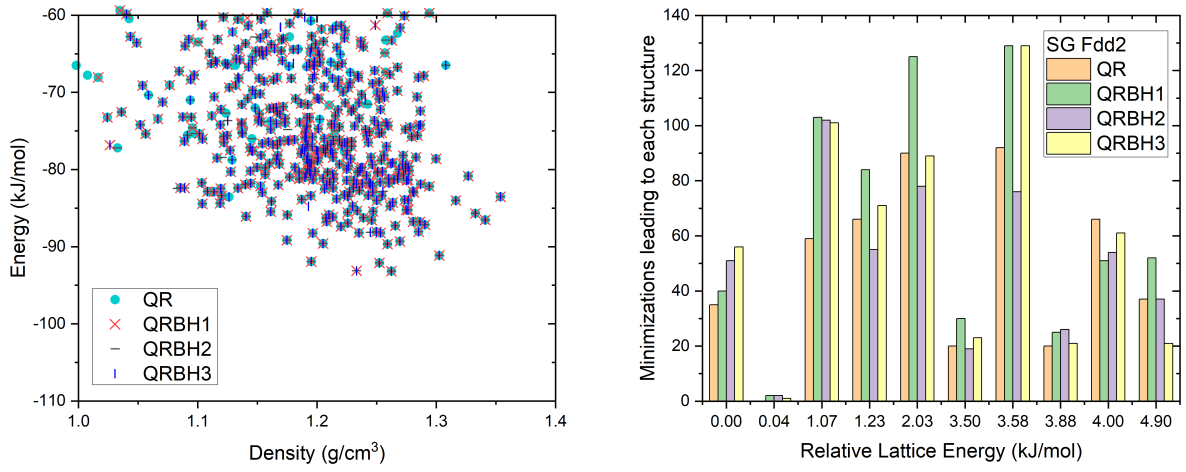
It is also important to reliably locate the possible crystal structures that are slightly higher in energy than the global minimum. As well as the importance of locating high energy polymorphs, the small energy differences often seen between predicted crystal structures means that errors in the model of interaction energies, as well as neglect of thermal vibrations, could lead to mis-ranking and that the true global minimum in free energy is not the global minimum in lattice energy from the energy model used for CSP. Indeed, for the molecules studied here, the known crystal structures are predicted close to, but not at the global minimum in energy: the monoclinic polymorph of TCNE is located 0.8 kJ mol^{-1} above the global minimum in $P2_1/c$, as the 3^{rd} lowest energy predicted structure; the two monoclinic polymorphs of benzamide are located 1.0 (3^{rd}) and 1.5 kJ mol^{-1} (4^{th}) above the global minimum in $P2_1/c$; the α and β polymorphs of glycine were located 2.2 (2^{nd}) and 1.5 kJ mol^{-1} (4^{th}) above the global minima in space groups $P2_1$ and $P2_1/c$, respectively. Thus, it is important that CSP provides a complete set of low energy structures so that all structures within error of the global minimum have been located.

Table 3: Number of unique structures within 5 kJ mol^{-1} from the global energy minimum for single-component crystal systems in 7 space groups. The three values for QR-BH are the results for the three independent runs. The energy window from the global minimum is measured separately for each molecule-space group combination. Reference results (Ref) show the number of unique structures generated from longer, pure QR searches with 50,000 minimizations in total. Results covering an expanded 10 kJ mol^{-1} range are provided in Table S3.

space	TCNE			benzamide			glycine		
group	QR-BH	QR	Ref	QR-BH	QR	Ref	QR-BH	QR	Ref
$P\bar{1}$	1: 1: 1	1	1	10:10:10	10	10	6: 6: 5	4	6
$P2_1$	4: 4: 4	4	4	3: 3: 3	3	3	7: 7: 7	7	7
$P2_1/c$	10:10:10	10	10	18:19:18	18	18	12:12:12	12	12
$C2/c$	6: 6: 6	6	6	37:37:33	37	37	18:18:18	18	18
$P2_12_12_1$	13:13:13	13	13	6: 6: 6	6	6	3: 3: 3	3	3
$F2dd$	20:20:20	20	20	12:12:12	11	12	3: 3: 3	3	3
$I4_1/a$	4: 4: 4	4	4	6: 6: 6	6	6	8: 8: 8	8	8

Therefore, we also compare the performance of QR-BH and pure QR searches in sampling the entire low-energy regions of the crystal structure landscapes. Table 3 shows the number of unique crystal structures found within 5 kJ mol^{-1} of the global minimum for each molecule-space group combination for the 10,000 step QR-BH and QR searches. These are compared to a much longer reference search (50,000 QR structures), which should be sufficiently well sampled to locate all low energy structures. These results show only minor differences between methods. Apart from four systems (benzamide in space groups $P2_1/c$, $C2/c$ and $F2dd$; and glycine in $P\bar{1}$), the same set of structures is located in all searches, including all three repeats of QR-BH. For two of these systems, (benzamide in $F2dd$, glycine in $P\bar{1}$) the 10,000-minimization QR search misses one or more of the low energy structures that was located in the longer reference search, while QR-BH finds all of the structures in some or all of the repeats. For only one system (benzamide in $C2/c$) does the QR-BH perform worse than QR, locating four fewer low energy structure than QR in one of the QR-BH repeats; although it is infrequent, these cases of inconsistency between QR-BH runs are a concern, as missed structures could be important when interpreting the results of CSP. In the fourth case (benzamide in $P2_1/c$) one of the QR-BH runs finds a structure that was not located in any of the other CSPs, including the long reference search. Over a wider 10 kJ mol^{-1} energy range (see Table S3), we see more minor differences between searches in their location of higher energy structures. However, the overall consistency between QR and QR-BH sets of structures is still clear.

Figure 3 shows the predicted energy landscape for benzamide in $F2dd$. The results demonstrate the reproducibility of the results: QR and QR-BH find nearly identical sets of crystal structures, particularly in the lowest-energy region of the landscapes. Furthermore, the three repeats of QR-BH find the same sets of structures. As expected, as the energy increases away from the global minimum, we observe more structures that are located by QR, but not QR-BH (Fig. 3a, blue dots without corresponding QR-BH hits, or hits in only 1 of the QR-BH repeats). This is because the bias introduced in basin hopping to favor sampling



(a) (b)

Figure 3: (a) Sampling of the crystal energy landscape of benzamide in $F2dd$ with pure QR and the three repeats of QR-BH. (b) Number of hits of to each of the 10 lowest energy minima found in $F2dd$. The second-low energy structure, found by all repeats of QR-BH, was missed by QR.

of low energy structures must come at the expense of sampling in the higher energy regions of the landscape. It does not seem that this bias hinders sampling by QR-BH in the usual energy range of polymorphism (typically under 10 kJ mol^{-1}), with the parameters used here.

As already highlighted in Table 3, benzamide in $F2dd$ is a case where the 10,000-minimization QR search has missed one low energy structure that is located by all three QR-BH runs with the same number of minimizations. This missed structure is the second lowest-energy structure for this system, only 0.04 kJ mol^{-1} above the global minimum. In extending the QR search, we find that this structure is first hit after minimization of 30,381 structures. Thus, the greater sampling efficiency of QR-BH is important in this case for obtaining a complete picture of the potential crystal structures. Figure 3b shows the number of times that each of the 10 lowest energy structures were located in this system, highlighting the second lowest energy structure as a difficult case. The results show the increased sampling efficiency of QR-BH not only for the global minimum (as shown in Tables 1 and 2), but also for the next two structures. As the energy increases, the difference between QR and

QR-BH is less obvious, which matches our expectation that basin hopping helps locate the lowest energy structures and also demonstrates that sampling of the rest of the low energy region is not worsened compared to pure QR sampling.

3.1.3 Location of high-energy structures: porous structures of TTBI

The fourth single component system, TTBI (Fig. 1d), was chosen as a more extreme test for the location of important higher energy crystal structures. TTBI forms four microporous polymorphs,[14,30] named α, β, γ and δ, ranging from 46.4 to 92.1 kJ mol^{-1} (according to the FIT + multipoles force field) above the densely-packed global lattice energy minimum structure. These structures lie well outside the usual energetic range of polymorphism. They are accessed experimentally because they crystallize with solvent filling their pores and are stable to solvent removal, presumably as deep local energy minima.

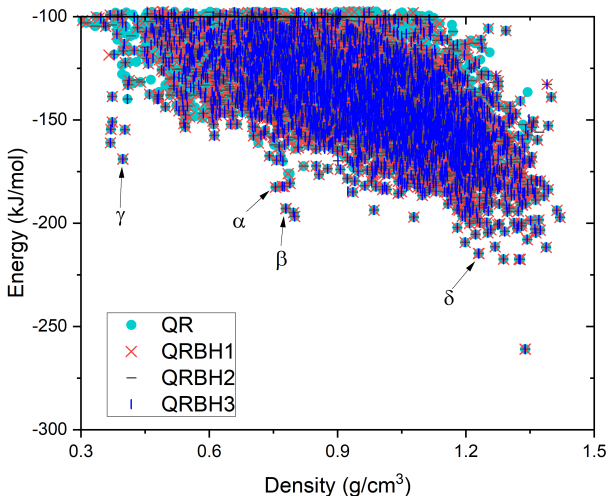


Figure 4: Energy landscape comparison for TTBI in 9 space groups from QR and the three repeats of QR-BH at 3000 K, with the four predicted structures corresponding to high-energy experimentally observed structures labelled.

Due to the molecular symmetry of TTBI, each of the observed structures, as well as the global energy minimum, can be located in CSP searches in multiple space groups. Thus, we tested the search methods' ability to locate these structures in several space groups and

Table 4: Steps required to locate the global energy minimum and experimental structures for TTBI in 7 space groups. The target structures are listed in order of increasing energy from left to right. QR-BH results are the mean over three repeats of the search, each starting from the same QR seed structures. Results for space group $F2dd$ and $I4_1/a$ are not shown because none of the experimentally observed structures, nor the overall global minimum structure, can be located in these space groups.

Space group	T(K)	global min.		δ		β		α		γ	
		QR-BH	QR	QR-BH	QR	QR-BH	QR	QR-BH	QR	QR-BH	QR
$P\bar{1}$	3000	6.3	29	1.0	1	9.0	9	-	-	246.0	397
	500	14.0		1.0		9.0		-		190.3	
$P2_1$	3000	6.0	6	-	-	-	-	-	-	1207.0	735
	500	6.0		-		-		-		1702.0	
$P2_1/c$	3000	33.3	479	212.3	1144	60.0	972	-	-	66.0	256
	500	11.7		718.0		547.0		-		192.7	
$C2/c$	3000	-	-	23.0	144	22.0	22	-	-	264.7	1067
	500	-		43.3		20.7		-		350.7	
$P2_12_12_1$	3000	-	-	-	-	-	-	-	-	111.3	134
	500	-		-		-		-		134.0	
$P4_2$	3000	-	-	-	-	-	-	13.0	13	-	-
	500	-				-		9.7		-	
$P4_2/n$	3000	-	-	-	-	-	-	64.3	85	-	-
	500	-				-		70.7		-	

we added two additional space groups ($P4_2$ and $P4_2/n$) to our sampling as those in which the α polymorph is located. All other search parameters were the same as for the other single component crystals, except the perturbation cut-off for volume, which was increased to 200 Å^3/molecule due to the large molecular size. To investigate whether the temperature used to control the acceptance during basin hopping had an effect on finding target structures over such a large energy range, QR-BH was run at two temperatures: 3000 K (as above) and 500 K.

As with the other test systems, we find that the pure QR and QR-BH methods provide essentially the same sets of structures in the low-energy region, as well as in the regions of the important high-energy structures corresponding to the observed polymorphs (Fig. 4). To compare efficiency, we examined the minimum steps required to locate each target structure: the four known polymorphs and the global energy minimum (Table 4). The QR-BH method consistently required fewer steps to locate the important structures on the landscape at both temperatures. In only one case – polymorph γ in space group $P2_1$ – did the pure QR search locate a target structure earlier than QR-BH. The improved efficiency of QR-BH over pure QR searching at locating important high energy structures might seem surprising – the method was developed to locate low energy structures effectively and we expected that basin hopping might favor the lowest energy region too aggressively to provide good sampling of higher energy regions, even when seeded with starting structures from a low-discrepancy (QR) sampling of the energy surface. Our interpretation of these results is that, although α, β, γ and δ are high-energy structures on the whole lattice energy landscape, they correspond to the lowest energy structures within separated local regions of configurational space that are not escaped easily at the temperature used in the basin hopping. Since BH trials begin from different initial structures and explore their local region, these high energy, experimentally observed structures can be located efficiently.

Space group $P2_1/c$ appears to be the most challenging landscape of those sampled for TTBI; four of the target structures (α, β, γ and the global minimum) are located in this space

group and are first hit between 256 and 1144 steps in the pure QR search. By comparison, QR-BH at 3000 K locates *all four* structures before 250 steps.

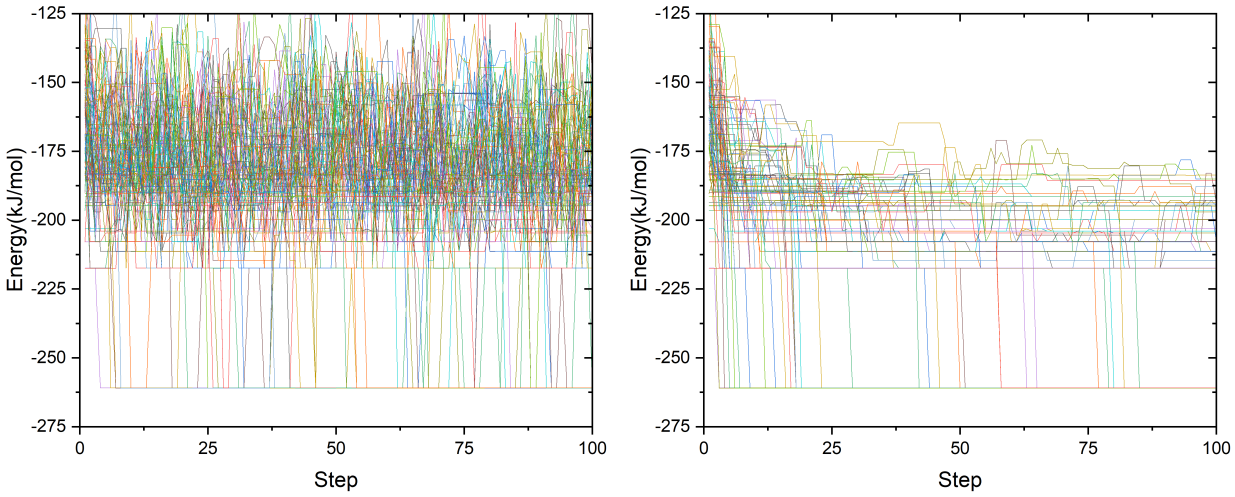


Figure 5: Energies of the 100 basin hopping trajectories for one repeat of a QR-BH run for TTBI in space group $P2_1/c$ at a) 3000 K and b) 500 K. Colors indicate the trajectories starting from different QR seeds.

The comparison between basin hopping temperatures in space group $P2_1/c$ reflects the overall difference in results between temperatures: QR-BH at 500 K performs better at locating the target structures than pure QR, but not as well as 3000 K. The 100 basin hopping trajectories from QR-BH in $P2_1/c$ are plotted in Figure 5 for both temperatures, showing the expected behavior: the lower temperature drives the trajectories more aggressively towards lower energies, while the higher temperature simulations maintain sampling of higher energy structures throughout the trajectories. At least in this system, these high energy steps improve the efficiency of locating the target structures, so that QR-BH at 3000K shows slightly better performance. The relationship between basin hopping temperature and sampling efficiency is discussed in more detail for the co-crystal systems.

3.2 Co-crystals

After assessing the reliability of the QR-BH method on the relatively simple single-component molecular crystal systems, the question arose naturally how the parameters of QR-BH would affect the sampling efficiency and whether there is an optimal parameter set to maximize efficiency. To test the influence of QR-BH parameters, we applied the method to the more challenging co-crystal systems, one being a hydrogen bonded co-crystal (XAFQAZ, Fig. 2a) and the other held together by weaker, less directional interactions (PYRPMA, Fig. 2b). PYRPMA was explored in three common space groups ($P\bar{1}$, $P2_1$ and $P2_1/c$), including the space group in which the known crystal structure is found ($P2_1$). The known crystal structure of XAFQAZ is found in space group $P2_1/c$ (and located here as the global energy minimum) and we also investigated $F2dd$ and $I4_1/a$ for this co-crystal. These space groups were chosen so that, across the two co-crystal systems, we explored different types of intermolecular interactions and a range of crystal symmetries.

The perturbation cut-offs for Monte Carlo moves during basin hopping simulations were kept the same as those used for TTBI and both temperatures (500 K and 3000 K) were evaluated. The introduction of a second molecule in the asymmetric unit increases the dimensionality of configurational space by 6 (compared to single-component crystals), so we increased the length of searches to thoroughly explore the more complex crystal energy landscapes. Pure QR searches were run for a total of 50,000 minimizations and QR-BH were run with just over 50,000 total minimizations. Because of the greater complexity of their search space, we used the co-crystal systems to explore the impact of changing the allocation between QR seeds and BH steps, keeping a fixed total computational budget. Three seed:step ratios were applied: 1:1 (225 seeds $\times$ 225 BH steps = 50,625 minimizations), 2:1 (316 seeds $\times$ 158 BH steps = 49,928 minimizations) and 10:1 (710 seeds $\times$ 71 BH steps = 50,410 minimizations). The higher ratios test the effectiveness of shorter basin hopping trajectories started from a larger set of QR seed structures, which has the advantage of greater parallelizability.

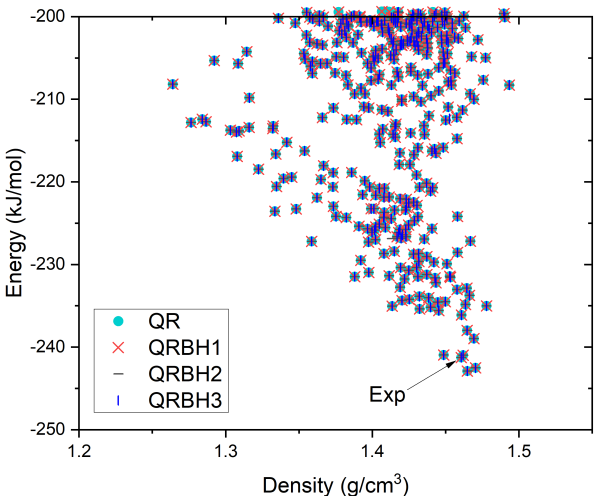


Figure 6: Energy landscape for the PYRPMA co-crystal in space group $P2_1$. The parameters of simulations were 3000K and 1:1 (225 $\times$ 225) seed-step allocation.

Of the two co-crystals, PYRPMA was found to be the easier landscape for locating structures, probably because of its weaker, less direction intermolecular interactions, leading to a smoother lattice energy surface. However, comparison to the same set of space groups for the single-component structures (Table 1) shows an increase in the number of required minimizations to hit the global energy minimum, due to the greater dimensionality of the search space. The latest of the space group global energy minima to be located in the pure QR search was in $P2_1/c$ after 866 steps (Table 5). Counts of the number of low energy structures located in each space group are presented in the Supporting Information (Table S4). We also monitored the step at which the structure corresponding to the experimentally observed co-crystal structure[22] was located. This was located as the 3^{rd} lowest energy structure in space group $P2_1$, 2 kJ mol^{-1} above the global minimum. This experimentally observed structure proved more difficult to locate in the CSP search, being found after 2207 QR steps, compared to 17 for the global minimum in $P2_1$.

The QR-BH method reproduced the same PYRPMA crystal energy landscapes, finding all of the same structures as the QR search (see Fig. 6 for $P2_1$). The space group global energy minima, as well as the experimentally observed crystal structure, are consistently

found earlier in the QR-BH searches than QR (Table 5). In only two cases was one of the target structures found later in QR-BH than QR, each time for the global minimum in $P2_1/c$ at 3000 K, using 2:1 and 10:1 seed:step ratios. However, even in these cases, the mean steps to locate the global minimum was smaller than the steps required in QR. While the improved efficiency of QR-BH over QR is clear, the results do not point strongly to a best set of QR-BH parameters.

Table 5: Comparison of steps required to locate global minima in space groups $P\bar{1}$, $P2_1$ and $P2_1/c$ and experimental structure in $P2_1$ for PYRPMA, and XAFQAZ in space groups $P2_1/c$, $Fdd2$ and $I4_1/a$. Results are shown for the three independent trials of QR-BH at each of two temperatures and three seed:step ratios.

T (K)	Seed:step ratio	Repeat	PYRPMA				XAFQAZ		
			$P2_1/c$	$P\bar{1}$	$P2_1$	Exp	$P2_1/c$	$Fdd2$	$I4_1/a$
Quasi-random			866	198	17	2207	30315	12243	-
500	1:1	1	434	45	12	282	9729	8225	109
		2	280	43	17	302	-	5480	6732
		3	139	3	17	70	-	9246	7611
	2:1	1	267	5	17	182	-	7104	1050
		2	559	23	2	162	8216	2223	4950
		3	385	14	16	107	-	888	-
	10:1	1	169	8	17	221	-	3648	874
		2	39	4	6	144	-	10621	-
		3	214	30	17	56	896	21016	2
3000	1:1	1	130	25	17	294	11682	-	-
		2	137	2	17	106	-	-	2880
		3	155	120	17	148	576	6864	31570
	2:1	1	1498	50	11	12	39984	-	-
		2	286	77	8	42	5375	3068	16632
		3	428	78	17	256	384	-	-
	10:1	1	180	24	17	134	11968	2356	-
		2	1070	84	6	156	9918	5889	53
		3	238	11	12	58	21888	20142	1870

XAFQAZ was a more challenging system. The experimentally observed crystal structure is reproduced by the global energy minimum in space group $P2_1/c$. The QR-BH and pure QR methods generated similar energy landscapes in $P2_1/c$ (see Fig. 7a). However, the pure QR method had difficulty locating several low-energy structures, including the global

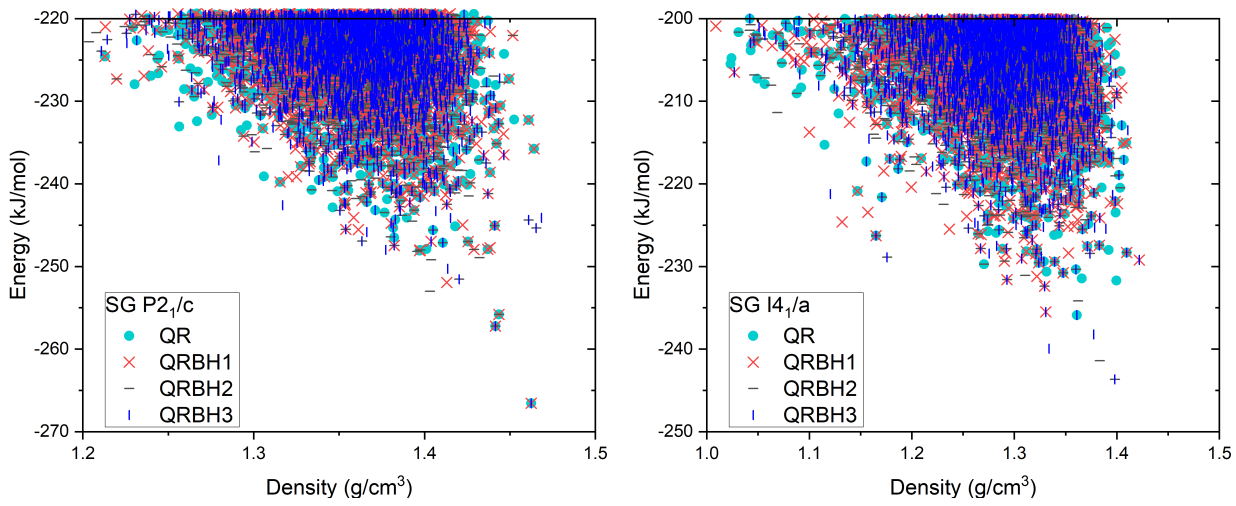


Figure 7: Predicted crystal energy landscapes for the XAFQAZ co-crystal in space groups (a) $P2_1/c$ and (b) $I4_1/a$. The parameters of QR-BH simulations were 3000 K, using a 1:1 (225 × 225) seed-step allocation. In space group $I4_1/a$ several low-energy structures, including the global energy minimum, were missed by the QR method.

minimum, which was first hit after 30,315 minimizations. The 3000 K, 1:1 QR-BH search located the $P2_1/c$ global minimum in 2 of 3 searches, both in significantly fewer steps than QR. Increasing the seed:step ratio to 2:1 or 10:1 led to more consistent success of QR-BH: the $P2_1/c$ was located in all searches with these higher seed:step ratios, with fewer mean steps than QR. Lowering the temperature to 500 K gave less consistent results in $P2_1/c$; at each seed:step ratio, only one of three repeats located the global minimum in 50,000 minimizations.

Space group $F2dd$ for XAFQAZ shows an opposite trend with respect to the temperature of the basin hopping trajectories: more consistent results were obtained at the lower temperature in this space group (Table 5): all 500 K QR-BH simulations located the $F2dd$ global minimum, using fewer steps than QR in all but one of the simulations (at a 10:1 seed:step ratio). $I4_1/a$ was also problematic for XAFQAZ, with the 50,000-minimization QR search missing the global minimum *and* the next three lowest energy predicted structures (Fig. 7b). QR-BH was more successful, finding the space group global energy minimum in most

of the simulations. Like space group $F2dd$, more consistent results were obtained at lower temperature; the $I4_1/a$ global minimum was located in all three 500 K QR-BH simulations with a 1:1 seed:step allocation.

The different influence of basin hopping temperature between space groups might be due to differences in the nature of the energy landscapes (Fig. 7). There is a large energy gap between the global minimum and other structures in space group $P2_1/c$, while there is a more uniform distribution of structures in the low energy regions of $I4_1/a$ and $F2dd$. The smaller energy differences in the latter two space groups makes it easier for BH to travel 'uphill' at lower temperatures.

Because there could be different ways to define the number of steps required by QR-BH to first hit the important structures, our conclusions regarding the most reliable QR-BH settings is based on whether, within the search with the same total number of minimizations, the three repeats of QR-BH locate the global minimum or not. Of the QR-BH parameter combinations tested here, the most consistent set over the two co-crystals is T = 3000 K with a large (eg. 10:1) seed:step ratio - many QR seed structures, each run for a short basin hopping trajectory.

3.2.1 On-the-fly clustering

Another strategy that we investigated was on-the-fly clustering during QR-BH to identify when multiple basin hopping trajectories, starting from different QR starting structures, encounter each other and thus sample the same local region of configuration space. This could lead to loss of efficiency compared to the situation where all BH trajectories sample different parts of the energy surface. Thus, in the on-the-fly clustering strategy, any newly sampled energy minimum is compared with all structures that have already been visited by any other BH trajectories, using similarity of X-ray diffraction patterns to identify identical structures. If a structure had already been sampled by another BH trial, the current BH

Table 6: Comparison of steps required to locate the global energy minima with QR-BH using on-the-fly clustering in space groups $P\bar{1}$, $P2_1$ and $P2_1/c$ and the experimentally observed structure in $P2_1$ for PYRPMA, and XAFQAZ in space groups $P2_1/c$, $Fdd2$ and $I4_1/a$. The QR-BH results are compared to a pure QR search.

T (K)	Repeat	PYRPMA				XAFQAZ		
		$P2_1/c$	$P\bar{1}$	$P2_1$	Exp	$P2_1/c$	$Fdd2$	$I4_1/a$
Quasi-random		866	198	17	2207	30315	12243	-
500	1	832	15	17	42	-	5814	-
	2	231	12	17	189	-	4416	-
	3	43	42	12	57	1380	1184	7910
3000	1	214	77	17	607	4900	494	1358
	2	65	26	17	131	6975	-	792
	3	247	3	17	112	11152	296	1552

trajectory is truncated. The trial is replaced by a BH run starting from the next unused quasi-random seed, to explore a new region in configurational space and maintain a constant number of active BH trajectories. In this way, the method aims to minimize overlap in sampling between BH runs.

We set the number of active BH trajectories to 200, with a maximum of 250 steps for each BH trial, while other parameters were unchanged from the previous co-crystal searches. To bo consistent, searches were run for a total of 50,000 energy minimizations. All other parameters controlling the basin hopping (cutoffs on perturbations and temperature used in the Monte Carlo acceptance) are applied in the same way as in the QR-BH presented above for the co-crystals.

The distribution of BH steps among trajectories (Fig. 8 and Figs S1, S2 for all systems, space groups and temperatures) indicates that most of the BH trials were truncated within 50 steps for XAFQAZ at 3000 K and only a small number of trials reached the maximum allowed number of steps. The length of trajectories are typically even shorter for the PYRPMA co-crystal (Fig. S1) and increases slightly at 500 K (Fig. S2). This observation implies that BH trials rarely remain in separate regions of the energy landscape and, so, on-the-fly clustering should help keep the structure search spread across configurational space. We also observed

that many quasi-random structures directly duplicated structures that had already been located by QR-BH, so did not progress to BH beyond the initial energy minimization.

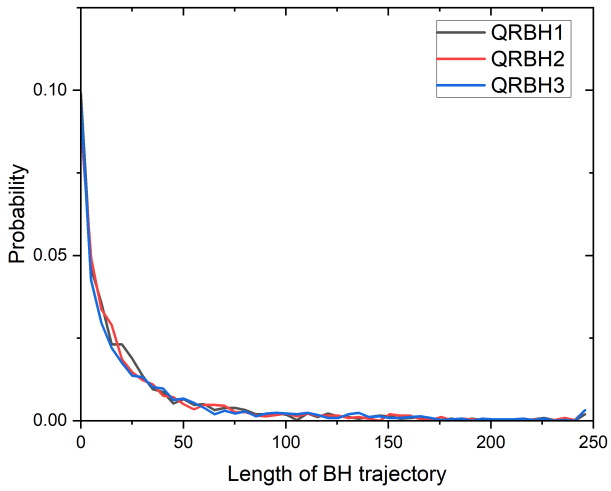


Figure 8: Distribution of BH trajectory lengths among trials for QR-BH with on-the-fly clustering applied to the XAFQAZ co-crystal in space group $I4_1/a$ at $T = 3000$ K.

As with QR-BH with independent runs, on-the-fly clustered QR-BH generally had a better sampling efficiency than the pure QR method in locating the global minima within each co-crystal/space group system (Table 6), as well as in locating low energy structures (Table S5, S8) on the complex co-crystal energy landscapes. Compared with independent trial QR-BH, the on-the-fly clustering was similarly efficient at sampling the global minima to QR-BH with a fixed, large seed:step ratio (10:1 = 710 seeds $\times$ 71 BH steps), because in both cases BH trials are quite short.

An advantage of the on-the-fly clustering approach over setting a fixed seed:step ratio for QR-BH is that the lengths of BH searches are determined by the behaviour of the trajectories, so are adaptive to the nature of the energy landscape for the particular molecular system / space group combination. The difference in BH trajectory lengths between the two co-crystals illustrates this (comparing Figs S1 and S2).

Finally, we note that, as with the independent, fixed-length QR-BH results, the optimal basin hopping temperature differs among space groups (Table 6). For the more challenging

XAFQAZ co-crystal, high temperature leads to more consistent location of the global minima in $P2_1/c$ and $I4_1/a$, while low temperature gives more consistent results for in space group $Fdd2$. We also tried running QR-BH at $T = 0$ K (Table S8), which allows only BH steps that lower the energy, as well as trying smaller Monte Carlo step sizes. These were both tested to try to keep the BH trajectories more localized on the energy surface by discouraging trajectories from escaping their current energy basin. However, both modifications led to poorer overall efficiency (Table S7, S8).

4 Conclusions

We have presented an improvement to the efficiency of quasi-random structure searching for crystal structure prediction of molecular crystals by combining the generation of trial structures using a low-discrepancy sequence with Monte Carlo basin hopping to explore for low energy crystal structures. The quasi-random seeds used as starting points for basin hopping provide a uniform coverage of the energy landscape, so that the role of basin hopping is to thoroughly explore for low energy structures in the region of its starting point.

The method has been tested on a set of single-component molecular crystals, for which we find that the QR-BH algorithm improves on the sampling efficiency of the pure QR searching in locating the global energy minimum more quickly than pure quasi-random searching and leads to better sampling of the lowest energy structures in each space group. We also find that the combined QR-BH method maintains the desirable feature from QR search methods of reliably locating important high-energy crystal structures; this is illustrated using the molecule TTBI as an extreme example, where experimentally observable crystal structures occupy a very wide range in lattice energies. Surprisingly, QR-BH located even the highest energy observed structures more quickly that a pure QR search.

The improved efficiency of QR-BH has also been illustrated in searching the higher dimensional energy landscapes of two co-crystal systems, which were used to investigate the

influence of temperature and the allocation between quasi-random seeds and basin hopping steps on the performance of the method. The optimal temperature used in basin hopping was found to vary between systems, even for different space groups of the same chemical composition. The most reliable performance was found with high temperature (3000 K) and a large number of QR seeds with short BH trajectories.

The good performance of QR-BH on the co-crystal systems demonstrates that the method works well for crystals with multiple independent molecules. Therefore, although we have not tested it here, we expect similar performance for single-component crystals with $Z' > 1$. Furthermore, although the method is tested here for crystals of rigid molecules, extension to flexible molecules could be made by including intramolecular distortions in the set of available Monte Carlo moves during basin hopping.

We have not pursued more advanced approaches to select the basin hopping temperature here. An adaptive approach might be required, with temperature changing through the simulation to maintain a targeted acceptance ratio of Monte Carlo steps. However, we have presented a method, which we call on-the-fly clustering, that adapts the length of BH trajectories to avoid overlap of trajectories sampling the same region of the energy landscape.

Acknowledgement

S.Y. acknowledges the financial support from the China Scholarship Council (No. 201706230229). We acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton.

Supporting Information Available

Further details of computational methodology (lattice energy minimization, Monte Carlo move cutoffs and duplicate removal), detailed results of sampling the global minima and low energy structures of the co-crystal systems, using independent BH trials and on-the-fly

clustering, and plots of BH trajectory lengths in on-the-fly clustering calculations.

Sets of predicted crystal structures and their calculated energies are available from DOI:https://doi.org/10.5258/SOTON/D1690.

References

(1) Day, G. M. Current approaches to predicting molecular organic crystal structures. *Crystallogr. Rev.* **2011**, *17*, 3–52.

(2) Woodley, S. M.; Day, G. M.; Catlow, R. Structure prediction of crystals, surfaces and nanoparticles. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2020**, *378*, 20190600.

(3) Karfunkel, H.; Gdanitz, R. Ab Initio prediction of possible crystal structures for general organic molecules. *Journal of Computational Chemistry* **1992**, *13*, 1171–1183.

(4) Bazterra, V. E.; Ferraro, M. B.; Facelli, J. C. Modified genetic algorithm to model crystal structures. I. Benzene, naphthalene and anthracene. *The Journal of Chemical Physics* **2002**, *116*, 5984–5991.

(5) Zhu, Q.; Oganov, A. R.; Glass, C. W.; Stokes, H. T. Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Crystallographica Section B* **2012**, *68*, 215–226.

(6) Curtis, F.; Li, X.; Rose, T.; VÃązquez-Mayagoitia, Ã.; Bhattacharya, S.; Ghiringhelli, L. M.; Marom, N. GAtor: A First-Principles Genetic Algorithm for Molecular Crystal Structure Prediction. *Journal of Chemical Theory and Computation* **2018**, *14*, 2246–2264, PMID: 29481740.

(7) Wang, Y.; Lv, J.; Zhu, L.; Ma, Y. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B* **2010**, *82*, 094116.

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

(9) Nyman, J.; Day, G. M. Static and lattice vibrational energy differences between polymorphs. *CrystEngComm* **2015**, *17*, 5154–5165.

(10) Reilly, A. M.; Tkatchenko, A. Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals. *J. Chem. Phys.* **2013**, *139*, 024705.

(11) Heit, Y. N.; Beran, G. J. O. How important is thermal expansion for predicting molecular crystal structures and thermochemistry at finite temperatures? *Acta Crystallographica Section B* **2016**, *72*, 514–529.

(12) Nyman, J.; Day, G. M. Modelling temperature-dependent properties of polymorphic organic molecular crystals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 31132–31143.

(13) Case, D. H.; Srirambhatla, V. K.; Guo, R.; Watson, R. E.; Price, L. S.; Polyzois, H.; Cockcroft, J. K.; Florence, A. J.; Tocher, D. A.; Price, S. L. Successful Computationally Directed Templating of Metastable Pharmaceutical Polymorphs. *Crystal Growth & Design* **2018**, *18*, 5322–5331.

(14) Pulido, A.; Chen, L.; Kaczorowski, T.; Holden, D.; Little, M. A.; Chong, S. Y.; Slater, B. J.; McMahon, D. P.; Bonillo, B.; Stackhouse, C. J.; Stephenson, A.; Kane, C. M.; Clowes, R.; Hasell, T.; Cooper, A. I.; Day, G. M. Functional materials discovery using energy-structure-function maps. *Nature* **2017**, *543*, 657–664.

(15) Aitchison, C. M.; Kane, C. M.; McMahon, D. P.; Spackman, P. R.; Pulido, A.; Wang, X.; Wilbraham, L.; Chen, L.; Clowes, R.; Zwijnenburg, M. A.; Sprick, R. S.; Little, M. A.; Day, G. M.; Cooper, A. I. Photocatalytic proton reduction by a computationally identified, molecular hydrogen-bonded framework. *J. Mater. Chem. A* **2020**, *8*, 7158–7170.

(16) Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure predictionâĂŤI. Rigid molecules. *Journal of Computational Chemistry* **2005**, *26*, 304–324.

(17) Case, D. H.; Campbell, J. E.; Bygrave, P. J.; Day, G. M. Convergence Properties of Crystal Structure Prediction by Quasi-Random Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 910–924.

(18) Bhardwaj, R. M.; McMahon, J. A.; Nyman, J.; Price, L. S.; Konar, S.; Oswald, I. D. H.; Pulham, C. R.; Price, S. L.; Reutzel-Edens, S. M. A Prolific Solvate Former, Galunisertib, under the Pressure of Crystal Structure Prediction, Produces Ten Diverse Polymorphs. *Journal of the American Chemical Society* **2019**, *141*, 13887–13897, PMID: 31394896.

(19) Day, G. M.; Chisholm, J.; Shan, N.; Motherwell, W. D. S.; Jones, W. An Assessment of Lattice Energy Minimization for the Prediction of Molecular Organic Crystal Structures. *Crystal Growth & Design* **2004**, *4*, 1327–1340.

(20) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallographica Section B* **2016**, *72*, 171–179.

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

mann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; De Wijs, G. A.; Yang, J.; Zhu, Q.; Groom, C. R. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallogr., Sect. B: Struct. Sci.* **2016**, *72*, 439–459.

(22) Herbstein, F. H.; Snyman, J. A.; Lipson, H. S. The crystal structures at 110 and 300 K of the equimolar molecular compound of pyrene and pyromellitic dianhydride. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **1969**, *264*, 635–662.

(23) Sobol', I. M. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics* **1967**, *7*, 86–112.

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

(25) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of electrostatic interactions in determining the crystal structures of polar organic molecules. A distributed multipole study. *The Journal of Physical Chemistry* **1996**, *100*, 7352–7360.

(26) Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.

(27) Li, Z.; Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima

problem in protein folding. *Proceedings of the National Academy of Sciences* **1987**, *84*, 6611–6615.

(28) Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *The Journal of Physical Chemistry A* **1997**, *101*, 5111–5116.

(29) Pickard, C. J.; Needs, R. J. Ab initio random structure searching. *J. Phys. Condens. Matter* **2011**, *23*.

(30) Mastalerz, M.; Oppel, I. M. Rational Construction of an Extrinsic Porous Molecular Crystal with an Extraordinary High Specific Surface Area. *Angewandte Chemie International Edition* **2012**, *51*, 5252–5255.

Graphical TOC Entry