Large–Scale Computational Screening of Molecular Organic Semiconductors Using Crystal Structure Prediction

Jack Yang,[†,‡,¶] Sandip De,[§] Josh E. Campbell,[†] Sean Li,[‡] Michele Ceriotti,[§] and Graeme M. Day[*,†]

†*Computational System Chemistry, School of Chemistry, University of Southampton, Highfield, Southampton, United Kingdom, SO17 1BJ.*

‡*School of Material Science and Engineering, University of New South Wales, Sydney, New South Wales 2052, Australia*

¶*Australian Nuclear Science and Technology Organization, New Illawarra Rd, Lucas Heights, New South Wales, 2234, Australia.*

§*National Center for Computational Design and Discovery of Novel Materials (MARVEL), Laboratory of Computational Science and Modelling, Institute of Materials, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland.*

E-mail: g.m.day@soton.ac.uk

Abstract

Predictive computational methods have the potential to significantly accelerate the discovery of new materials with targeted properties by guiding the choice of candidate materials for synthesis. Recently, a planar pyrrole-based azaphenacene molecule (pyrido[2,3-*b*]pyrido[3′,2′:4,5]pyrrolo[3,2-*g*]indole, **1**) was synthesized and shown to have promising properties for charge transport, which relate to stacking of molecules in its crystal structure. Building on our methods for evaluating small molecule organic semiconductors using crystal structure prediction, we have screened a set of 27 structural isomers of **1** to assess charge mobility

in their predicted crystal structures. Machine–learning techniques are used to identify structural classes across the landscapes of all molecules and we find that, despite differences in the arrangement of hydrogen bond functionality, the predicted crystal structures of the molecules studied here can be classified into a small number of packing types. We analyze the predicted property landscapes of the series of molecules and discuss several metrics that can be used to rank the molecules as promising semiconductors. The results suggest several isomers with superior predicted electron mobilities to **1** and suggest two molecules in particular that represent attractive synthetic targets.

Supporting Information Available

Details of the crystal structure classification scheme, information on convergence of the crystal structure search and number of unique crystal structures per molecule, eigenvalue spectrum of the SOAP Similarity Kernel, details of the electron mobility calculations, energy-structure-function maps of all molecules, discussion of uncertainties in the electron mobility calculations.

Introduction

The discovery of functional molecular materials with targeted properties is hindered by the challenge of predicting the crystal structure that a given molecule will adopt, as well as the often complex relationship between crystal structure and the relevant properties. Computational approaches for guiding materials discovery have attracted significant attention[1–3] as a means of identifying the most promising synthetic targets and, therefore, reducing experimental effort spent on materials with poor properties. The reliability and scope of methods for computer–guided material discovery have benefited from recent theoretical advances in configurational space sampling[4–7] , force field[8] and quantum mechanical methods.[9,10] Developments in these areas mean that not only wider ranges of atomic configurations can be efficiently sampled, but that their physical and chemical properties can also be confidently predicted prior to laboratory synthesis and characterization. In particular, there has been rapid development in the field of crystal structure prediction (CSP) for organic molecules, as exemplified by the latest blind tests in this area.[11] An aim of CSP methods is to produce the full set of stable (low energy) crystal structures available to a molecule and any given molecule usually leads to a large number of possible structures within a small energy range from the global energy minimum.[12]

A recent advance in the use of computational methods to guide the discovery of functional materials is the mapping of simulated properties onto the crystal energy landscape produced by CSP methods.[13–15] The result, which we refer to as the energy–structure–function (ESF) map of a molecule,[13] describes the possible crystal structures, their relative stabilities and expected properties. The assessment of ESF maps of candidate molecules for a targeted function can help rule out molecules with no (or only high energy) predicted structures with the required properties and prioritize the synthesis of molecules with promising predicted structures. The approach has successfully led to the discovery, crystallization and characterization of a porous organic crystal with exceptionally high porosity and methane storage capacity.[14]

ESF maps can be produced for any functional material where the relevant property is calculable from the crystal structure. For example, we proposed their use for the assessment of molecules

as potential organic semiconductors.[15] To illustrate this, we assessed a series of aza-substituted pentacenes where the targeted property is the electron mobility, which we calculated using Marcus theory.[15] ESF charge mobility maps have also been reported recently for racemic and enantiomerically pure crystal structures of [6]helicene.[16]

Because molecular crystals are held together by many weak, competing non–covalent interactions and the possible crystal packing arrangements are typically close in energy, a subtle change in molecular structure can result in completely different molecular arrangements in the low energy crystal structures available to a molecule. In turn, the strong dependence of charge mobilities on relative molecular positions and orientations means that the mobility in the crystal structures of closely related molecules can be very different. We previously proposed several approaches for ranking hypothetical molecules by the likelihood that they will lead to a high mobility crystal structure.[15] One option is to use the predicted mobility of the most probable (lowest energy) predicted crystal structure of each molecule; another choice is the highest mobility among the low energy predicted crystal structures, and a third suggestion was a landscape-averaged mobility that takes into account the distribution of properties among the low energy structures. We believe that this is an advance on previous computational screening methods for molecular organic photovoltaic applications,[17,18] in which the important influence of crystal packing differences between molecules on the final electronic properties has not been considered.

To further reveal the interplay between chemical substitution patterns and the resulting crystal packing types for azapentacenes in an automated way, a kernel–based machine learning technique was applied in conjunction with clustering and dimensionality reduction analysis to produce a sketch-map for the *individual* crystal packing landscapes for pentacene and two azapentacene molecules.[19] These methods provided valuable insight into the profound changes in the overall crystal energy landscapes with changes in molecular substitution pattern.

In the present study, we build on these previous works[15,19] by demonstrating the use of CSP in screening a large set of molecules with mixed hydrogen bond donors and acceptors as candidates for small molecule crystalline organic semiconductors. Our study is motivated by the recent report of (pyrido[2,3-b]pyrido[3′,2′:4,5]pyrrolo[3,2-g]indole)[20] (**1**, Fig. 1), a planar, extended aromatic molecule with C_{2v} symmetry. **1** adopts a γ–type[21] crystal packing, which features columns of π-stacked molecules, with two orientations of such columns forming a 'cruciform' arrangement.[20] The crystal packing in **1** is directed by hydrogen bonds between the two hydrogen bond donors (–NH from the pyrrole rings) and acceptors (N from pyridine rings). Charge transport calculations on this structure suggested good carrier mobility in the directions of the molecular stacks.

These results sparked our interest in this class of molecules as an opportunity to explore the use of CSP and the generation of ESF maps in the screening of a large set of related molecules. The question we address is: allowing for reorientation and permutation of the pyridine and pyrrole rings, how does molecular structure affect the crystal packing landscape and what is the impact on the charge transport properties of these molecular materials? For this study, we investigated **1** and an additional 27 isomeric molecules (**2**–**28**, Fig. 1), performing CSP on each molecule and charge mobility calculations on the low energy crystal structures. The set of molecules includes the two–fold symmetrical (C_{2v}, **1**–**16**) and the 12 asymmetric molecules (C_s, **17**–**28**). We did not attempt to exhaust the list of all possible asymmetric possibilities in this study due to the large number of possible combinations. In the end, our goal is to identify the isomer with the the greatest potential to produce a material with large charge carrier mobility.

To our knowledge, this is the largest set of molecules that has been subjected to CSP and prop-

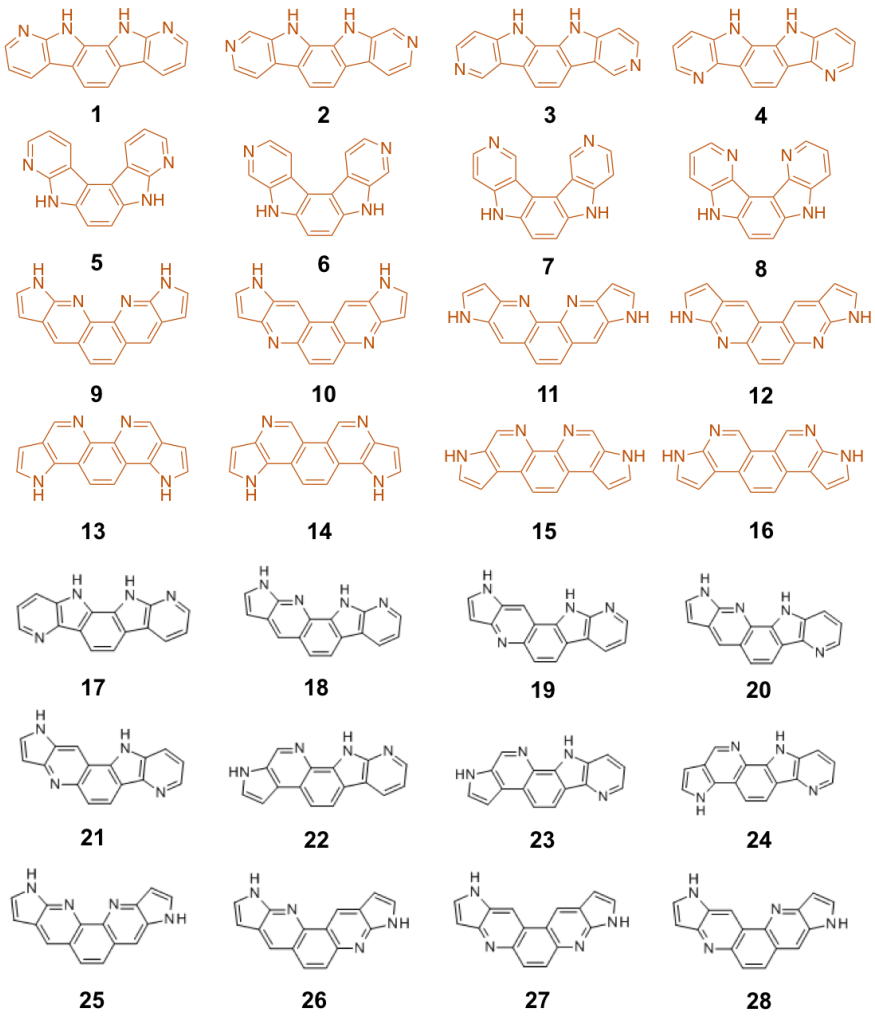

Figure 1: The known (**1**) and hypothetical (**2–28**) molecules studied. Symmetrical (C_{2v}) isomers (**1–16**) are colored in orange, to contrast with the asymmetric (C_s) isomers (**17–28**).

erty landscape exploration in a single study to–date. We further applied our recently–developed machine–learning methods to directly compare the differences in crystal packing landscapes *across* a set of chemically different molecules in a visually appealing and informative way, which can help us to further generalize better molecular design strategies for high–mobility organic semiconductors.

Methodologies

Crystal Structure Prediction

CSP was performed with our Global Lattice Energy Explorer (GLEE) program,[22] which uses a low-discrepancy, quasi-random sampling of crystal packing variables to sample the lattice energy surface as uniformly as possible. Molecular geometries were optimized at the B3LYP/6–311G** level of theory using GAUSSIAN09[23] and kept rigid throughout all crystal structure calculations.

Trial structures were generated and lattice energy minimized until a total of 49,000 successfully lattice energy minimized crystal structures were produced for each molecule, all with one molecule in the asymmetric unit: 5000 in the most commonly observed space group for organic molecules, $P2_1/c$, and 2000 in each of the 22 next most common space groups: $P\bar{1}$, $C2/c$, $P2_12_12_1$, $P2_1$, $Pbca$, $Pna2_1$, $P4_12_12$, $P2_12_12$, Pc, $P3_1$, $P4_1$, $Fdd2$, $Pccn$, Cc, $P2/c$, $C2$, $P6_1$, $Pca2_1$, $I4_1/a$, $P1$, $R\bar{3}$ and $Pbcn$. Each of these starting structures is unique, but leads to a much smaller number of unique crystal structures after lattice energy minimization and removal of duplicates.

The removal of duplicate crystal structures from the final structure sets was initially performed within individual space groups using the method described in Ref. 22. Overall clustering across all space groups was then performed using the COMPACK method.[24] A thoroughly sampled energy landscape is characterized by most low energy local minima having been sampled multiple times. Figures showing the number of duplicates in the low energy region of the landscape for each molecule are shown in the supporting information, along with a table listing the number of unique crystal structures located for each molecule. The results suggest that, within the space groups considered here, the energy surfaces are well sampled with a low chance that low energy structures have been missed.

All lattice energy minimizations were performed using an anisotropic atom-atom force field energy model for intermolecular interactions with the DMACRYS software,[25] with space group symmetry constrained throughout the optimizations. Lattice energies were calculated with a revised version of the W99 intermolecular atom-atom potential[26] combined with a distributed multipole electrostatic model based on the molecular charge densities calculated from a distributed multipole analysis[27] of the B3LYP/6–311G** density, with multipoles up to hexadecapole on each atom.

The arrangement of molecules in the predicted crystal structures was classified into one of the crystal packing types typically discussed for polyaromatic hydrocarbons[21] (such as herringbone, sheet, γ) using the method described in Ref. 15, which uses the angles between the planes of nearest neighbour molecules to classify the structure type (details are given in the supporting information). Hereafter, we refer to this as the heuristic structure classification.

Charge Mobility Calculations

Electron mobility calculations were performed for all predicted crystal structures of each molecule that are within 7 kJ/mol of that molecule's global lattice energy minimum. This energy window is chosen to include most experimentally observable structures, based on the distribution of calculated lattice energy differences between observed polymorphs.[28] Our mobility calculations use a hopping model and hopping rates between molecules in their crystal structure geometry calculated from Marcus theory. Details of these calculations are given in the supporting information.

Crystal Packing Comparisons with Machine Learning

It is often useful to classify observed or predicted molecular crystals into a small number of families of related structures; such classification is traditionally inspired by the manual inspection of the structures[21,29,30] and automated based on simple heuristic rules.[15] In this work we also use an entirely data-driven approach. Similarities between structures from the crystal structure landscapes

of all 28 molecules were measured using the SOAP–REMatch Kernel,[31,32] which assesses similarity of structures based on a comparison of local atomic environments corresponding to spherical regions with cutoff r_c around each atom. Based on the kernel-induced distance, the sketch–map dimensionality reduction technique[33] was used to generate a low-dimensional projection of the data in which the proximity between structures is represented as faithfully as possible. Details for constructing the SOAP–REMatch Kernel and sketch–map are provided in our previous publication.[19] A small difference is the choice of $r_c = 5$ Å for the cut–off distance, and $\gamma = 1$ as the regularization parameter for the SOAP–REMatch Kernel, that were chosen to emphasize the information on the packing rather than molecular geometry, which is dominant for shorter cutoff distances. For sketch-map we used the parameters $\sigma = 0.12$, $a = A = 4$, $b = B = 6$, that were adjusted following the empirical rules discussed in Ref. 34. The dimensionality reduction procedure was complemented by a clustering using the HDBSCAN* algorithm, that identifies regions of high density in the crystal structure space associated with the kernel metric. Clusters containing fewer than 5% of the structures were not considered statistically robust and were tagged as outliers.

Results and Discussion

Crystal Structure Prediction

Validation - Molecule 1

We first validate our CSP methodologies on molecule **1**, whose predicted energy–density landscape is shown in Fig. 2. The experimental crystal structure is located by the CSP search as the global lattice energy minimum. An overlay of the predicted global minimum with the experimental crystal structure is shown in Fig. 3, with comparison of the lattice parameters given in Table 1. The small root–mean–squared deviation in atomic positions from a 20-molecule cluster from the crystal structure ($RMSD_{20}$=0.269 Å, Fig. 3) and good agreement in lattice constants between the experimental and predicted crystal structure demonstrates excellent performance of the force field energy model. Our confidence in the calculated lattice energies is also based on successful CSP for pentacene and an azatetracene,[15] and results of benchmarking against experimentally derived lattice energies.[35]

The low energy region of the crystal structure landscape for molecule **1** (Fig. 2a) is dominated by the γ–type of packing, featuring stacking of the planar molecules in columns (Fig. 2b). Hydrogen bonds are formed between these columns, arranging them into a 'cruciform' pattern in the lowest energy structure. A few sheet–like structures are also present at low energy. Herringbone packing, as seen in the parent [5]phenacene (picene) crystal structure,[36] is strongly disfavoured for molecule **1**.

Crystal Structure Landscape Analysis Across All Molecules

Having validated the underlying lattice energy model, we proceed to discuss how the crystal structure landscapes change across isomers with different substitution patterns. To guide this analysis, we apply dimensionality reduction and clustering techniques to reveal the most important crystal packing patterns in a relatively unbiased way.

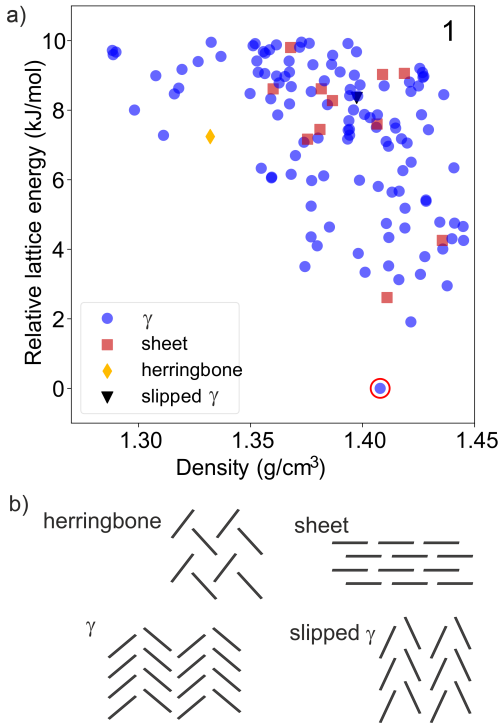

Figure 2: a) Predicted energy–density landscape for molecule **1**. Each point on the landscape corresponds to a distinct predicted crystal structure and the predicted global minimum (red open circle) corresponds to the experimentally observed crystal structure (see Fig. 3). Structures on the landscape are colour–coded according to their packing type; packing types are shown schematically in b), viewed down the long axis of the molecule.

In our previous work,[19] the SOAP-REMatch similarity kernel for measuring structural similarity was combined with the HDBSCAN* clustering method and the sketch-map dimensionality-reduction method to produce low-dimensional projections of the *individual* crystal packing landscapes of pentacene and two azapentacenes. This single-landscape analysis was shown to provide useful information on the packing types available to individual molecules. Here, the method is applied to the more challenging task of analyzing the crystal packing landscapes of multiple molecules in a single sketch-map, which we refer to as multi-landscape analysis.

Due to its general nature, the SOAP kernel reports on intra-molecular correlations (essentially, the geometry of each isomer) as well as on the inter-molecular correlations (crystal packing and intermolecular contacts) between structures. The dual nature of the information encoded in SOAP-REMatch similarity leads to an increased intrinsic dimensionality of the landscape (as revealed by the decay of the eigenvalues of the kernel similarity matrix, shown in the supporting information) compared to that observed for single landscape analysis, where all crystal structures contain the same molecule. This higher effective dimensionality makes the choice of parameters of the kernel and of the sketch-map projection less straightforward. Previously,[19] we found that using the hyperparameters which led to good lattice energy predictions from a Gaussian Process regression also led to insightful sketch-maps, although alternative choices of the environment cutoff distance emphasized different features in the crystal structures (close contacts vs long-range packing). The heterogeneous nature of the multi-landscape comparison exacerbates the dependence of the ap-

Table 1: Unit cell parameters of the experimental crystal structure of molecule **1** (Cambridge Structural Database reference code CASPUK) and global lattice energy minimum from CSP. The predicted crystal structure was converted to the reduced unit cell for comparison.

	Experimental	Predicted
a (Å)	9.783	9.593
b (Å)	4.714	4.725
c (Å)	26.987	27.328
α (°)	90	90
β (°)	99.270	100.382
γ (°)	90	90

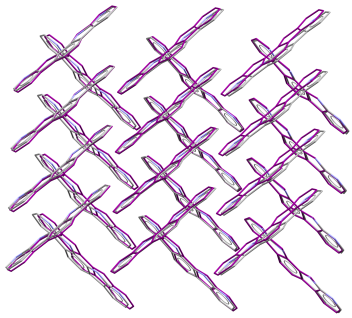

Figure 3: Overlay of the CSP predicted global minimum crystal structure (purple) with the experimentally determined crystal structure for molecule **1**. $RMSD_{20}$=0.269 Å. Hydrogen atoms are hidden.

pearance of the map on the parameters of the kernel. In particular, short-range cutoffs tend to emphasize the chemical differences between the molecules, leading to maps that contain little information on the packing motifs, whereas very large cutoffs contain too much information and do not reveal meaningful patterns. We found by trial-and-error that a cutoff around 5 Å provides the most insightful representation. Clearly, this manual and somewhat arbitrary optimization procedure makes the machine-learning analysis more biased than in the single landscape case.

The sketch-map that we obtain (Fig. 4) includes all predicted structures of the 28 molecules. With the parameters that we chose, the map captures the fact that different molecules can generate similar packing motifs; the 2D map features "islands" of structures that are populated by crystal structures from several different isomers (Fig. 4c). This finding is consistent with the concept of supramolecular synthons, according to which different molecules can adopt similar crystal packing motifs because they contain similar structure-directing functional groups: in this case, the same set of hydrogen bond donors and acceptors.

The clustering analysis confirms this finding; five high-dimensional HDBSCAN*-detected clusters are found across the landscapes (Fig. 4a), each comprising several different isomers. By comparing with Fig. 4b, where the structures are coloured by the heuristic structure classification based on the interplanar angles between molecules (Fig. 2b), it is clear that γ is the most

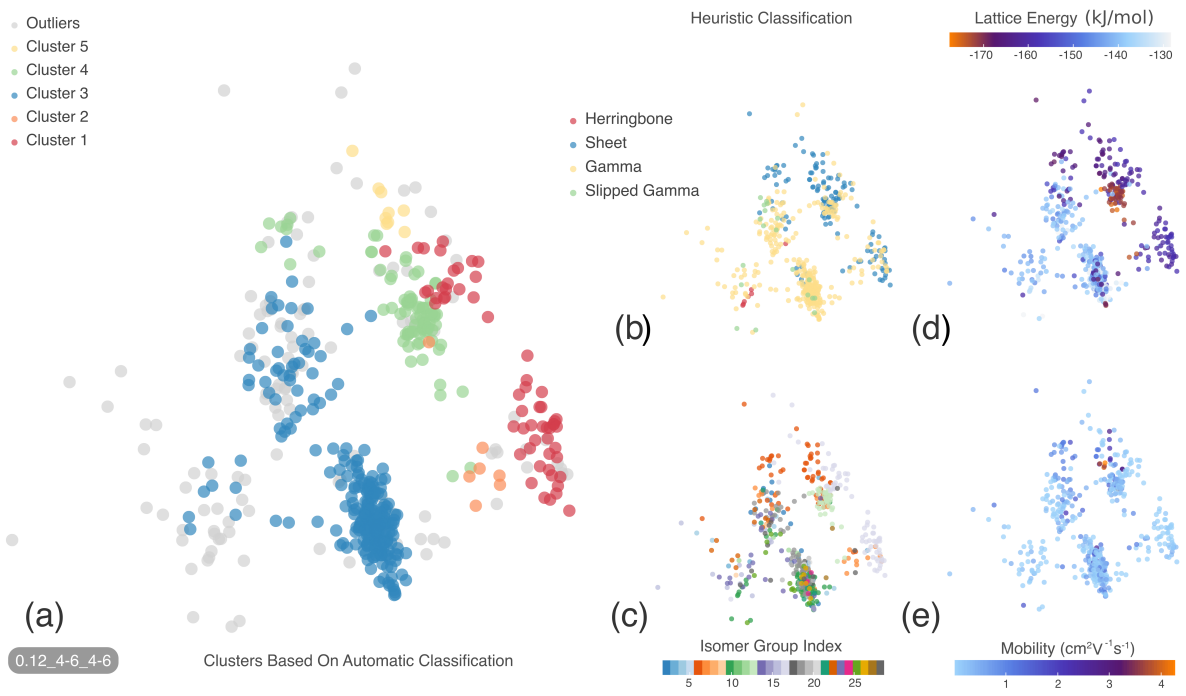

Figure 4: Sketch–map representations of the low–energy crystal structure landscapes for all 28 molecules investigated in the present study. Distances between molecules are computed using the metric associated with a SOAP-REMatch kernel with cutoff radius $r_c = 5$Å and $\gamma = 1$. Axes are not shown because only relative distances, and not absolute positions, are meaningful in a sketch-map. Each crystal structure on the sketch–map is color–coded according to: (a) the clusters detected by HDBSCAN* method, that correspond to groups of structurally homogeneous configurations in the metric space associated with the kernel; outliers that could not be associated to a clear cluster are colored in gray; (b) crystal packing type according to conventional packing motifs, identified based on simple heuristic rules; (c) isomer number (see Fig. 1); (d) calculated lattice energies and (e) predicted electron mobilities (in cm^2/Vs).

common crystal packing adopted by the full set of molecules. This dominant cluster is also identified by the HDBSCAN* method, with good registry between the automatic clustering (cluster 3, dark blue in Fig. 4a) and heuristic classification (yellow cluster in Fig. 4b). The fact that automatically-detected structural motifs, as well as heuristic packing classes, comprise crystals formed by different molecules and also by molecules with different point-group symmetry suggest that both from the point of view of local environments (as detected by the SOAP–REMatch) and from the point of view of the stacking, there is no obvious relation between molecular symmetry and preferred packing.

Sheet–like crystal packing is less favored by most of the molecules. This packing type corresponds closely with cluster 1 from the clustering analysis (red cluster in Fig. 4a) and is mainly concentrated in the top/right–hand–side of the sketch-maps, which dominates the crystal packing landscapes for molecules **5**, **12** and **16**. The crystal packing similarities among these three molecules could be largely attributed to the similarities in their molecular structures – within the molecule, hydrogen bond donors and acceptors are located close to each other on the convex side

of the molecule. The herringbone packing motif is so infrequent across the crystal structure landscapes that it is not classed as a cluster by the HDBSCAN* approach.

While more work is needed to automate the definition of an effective representation of multi-landscape maps, the possibility of representing simultaneously the crystal landscapes of multiple molecules in a way that highlights the presence of common structural features is already very useful as a guide to the interpretation of the interplay between molecular structure and supramolecular packing.

The crystal structures with the lowest (most negative) lattice energies mostly occupy the right–hand–side of the sketch–map (Fig. 4d), a significant portion of which correlate to crystals formed by symmetric molecules. The trends in lattice energies between molecules show how the arrangement of functional groups can alter the overall strength of intermolecular interactions. However, in property–driven material design, it is the physical/chemical properties, in this case, charge mobilities, rather than the lattice energies, that are of most interest. Fig. 4e reveals that, among all the low lattice energy crystal structures considered, a small number stand out with particularly high electron mobility; these crystal structures belong to molecule **5** and are located in the same region of the sketch-map as the lowest energy crystal structures. This coincidence of high mobility with energetic stability is promising for the experimental realization of these crystal structures.

Assessing Molecules using ESF Maps

Our primary motivation for this study is to assess whether any of the hypothetical molecular crystals of **2**–**28** offer improved properties in comparison to the reported molecule **1**. For organic semiconductors, we judge the fitness of a molecule by the calculated carrier mobility. CSP is key to this assessment by providing putative crystal structures of each isomer and the validation on **1** demonstrates that the structure prediction methods used here are reliable for this type of molecule.

An important aspect of the molecular assessment is that CSP provides a complete landscape of possible crystal structures for each molecule, which allows us to examine the charge carrier mobilities for all structures on these landscapes. We focus our analysis on the low energy region of the landscapes and have calculated mobilities for crystal structures within 7 kJ/mol of the global minimum crystal structure for each molecule; most known polymorph pairs are separated by 7 kJ/mol or less,[28,37] so we take this as the experimentally accessible energy range on the CSP energy landscapes. There is a small risk that experimentally realizable crystal structures exist above this energy cutoff, but we are limited by the computational expense of the transfer integral calculations. More efficient methods for the electron mobility calculations, such as machine learning explored in Ref. 19 or utilizing the relationship to wavefunction overlap,[38] will allow higher energy structures to be included in future studies.

Having an ensemble of possible crystal structures for each molecule provides a wealth of information, but complicates the process of ranking the molecules according to the predicted properties because each molecule is associated with a range of possible properties, depending on which structure it adopts when crystallized. In our previous work, we considered five quantities for ranking a small set of polyaromatic[15] molecules with potentially good electrical performance in their respective crystalline forms:

- λ_-, the electron reorganization energy, which is only weakly affected by crystal packing; we treat this as a molecular property, which we calculate for the isolated molecule. A large

Table 2: Summary of charge transport parameters for molecules **1-28**. λ_- is the calculated electron reorganisation energy. μ_{GM} is the predicted electron mobility for the global lattice energy minimum crystal structure. $\mu_{\max}$ is the maximum predicted electron mobility among the low energy predicted crystal structures. ΔE is the lattice energy gap from the predicted global minimum to the crystal structure with the highest electron mobility. $\langle\mu\rangle$ is the ensemble-averaged electron mobility across all crystals with calculated mobilities (see text). A bar chart is included in every column to ease visualization of the variations of each parameter.[a]

Molecule	λ_- / eV	μ_{GM} / cm^2 (Vs)$^{-1}$	$\mu_{\max}$ / cm^2 (Vs)$^{-1}$	ΔE / kJ mol^{-1}	$\langle\mu\rangle$ / cm^2 (Vs)$^{-1}$
1	0.200	0.240	1.788	4.751	0.496
2	0.258	0.233	0.548	2.964	0.246
3	0.210	0.563	2.119	5.207	0.359
4	0.222	1.523	1.523	0.000	1.523
5	0.201	2.706	4.277	4.922	1.283
6	0.248	0.001	1.359	4.494	0.092
7	0.205	0.006	0.006	0.000	0.006
8	0.222	0.005	0.005	0.000	0.005
9	0.226	0.004	1.600	6.509	0.249
10	0.303	0.002	0.002	0.000	0.002
11	0.242	0.003	0.883	4.082	0.157
12	0.291	0.131	1.046	3.931	0.283
13	0.260	0.100	0.427	5.612	0.092
14	0.341	0.301	0.301	0.000	0.253
15	0.253	1.356	1.356	0.000	0.568
16	0.336	0.001	0.404	3.951	0.072
17	0.211	0.470	2.545	6.988	0.756
18	0.212	0.008	1.462	6.917	0.247
19	0.214	1.028	2.088	4.939	0.904
20	0.224	0.299	0.303	4.045	0.300
21	0.227	0.757	0.757	0.000	0.501
22	0.250	0.002	0.219	6.856	0.006
23	0.262	0.001	1.019	6.759	0.044
24	0.224	0.005	1.524	5.745	0.544
25	0.215	0.102	0.961	6.737	0.149
26	0.275	0.002	0.534	6.921	0.077
27	0.293	0.002	0.002	0.000	0.002
28	0.281	1.068	1.068	0.000	0.337

[a] The electron mobilities for structures with low mobilities are underestimated due to numerical instabilities in the calculation of the transfer integrals, t. A possible correction is discussed in the supporting information. This does not affect the results for the best ranking molecules.

λ_- has a detrimental effect on the charge mobility due to the exponential decay of charge transfer rate with respect to λ_-. Thus, ignoring the effects of crystal packing, molecules with low λ_- are favoured.

- μ_{GM} is the calculated electron mobility for the predicted global lattice energy minimum crystal structure for a given molecule, which is the most likely structure to crytallize experimentally.

- $\mu_{\max}$ is the largest electron mobility for all crystal structures within 7 kJ/mol of the global minimum in lattice energy for a given molecule. This describes the best μ that can be achieved for a given molecule.

- ΔE is the energy difference to the global minimum at which $\mu_{\max}$ is observed. A large ΔE means that $\mu_{\max}$ for that molecule is less likely to be achieved, since lower energy structures are more likely to be observed.

- $\langle\mu\rangle$ is a Boltzmann–like averaged electron mobility over the landscape of possible crystal

structures for a given molecule, which we calculate as:

$$\langle \mu \rangle = \frac{\sum_i^N \mu_i \exp\left(-\Delta E_i/\beta\right)}{\sum_i^N \exp\left(-\Delta E_i/\beta\right)}, \tag{1}$$

where μ_i is the electron mobility in the i–th structure and ΔE_i is its lattice energy difference with respect to the predicted global minimum for that molecule. $\beta = 2.70$ kJ/mol is a decay constant fitted to the probability of observing a pair of polymorphs with an energy difference of ΔE_i.[28] This quantity provides a measure of the collective likelihood that a given molecule will crystallize into a structure with a large carrier mobility.

The values of these five quantities for all 28 molecules investigated here are summarized in Table 2 and a selection of the CSP landscapes and ESF maps are shown in Fig. 5 (the rest are given in the supporting information).

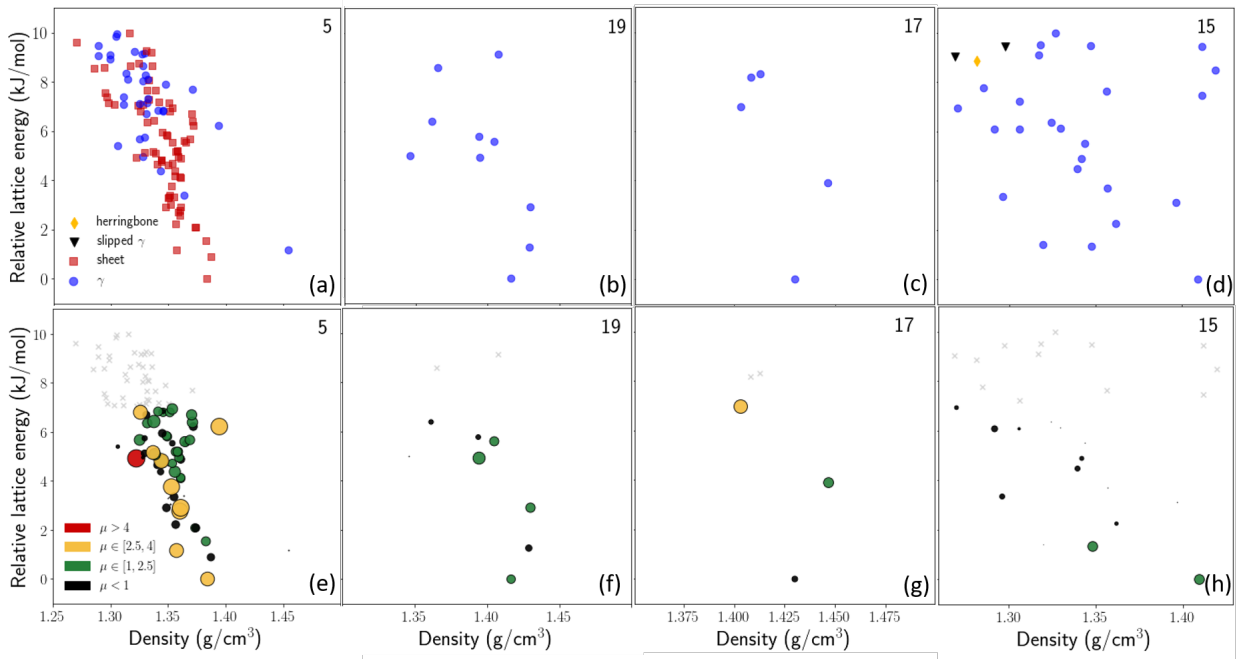

Figure 5: Predicted ESF maps for molecules **5** (a, e), **19** (b, f), **17** (c, g) and **15** (d, h), whose landscape-averaged electron mobilities, $\langle \mu \rangle$, are among the highest of the molecules studied. The landscape for molecule **4**, which has the largest $\langle \mu \rangle$, is not shown because there is only one structure in the low energy region. Extended landscapes, up to 30 kJ/mol, are shown in the supporting information. (a-d) show the landscapes with structures color coded according to their corresponding crystal packing types. (e-h) show the ESF maps, where the size and color of each point represents the magnitude of the calculated electron mobility. Mobilities were not calculated for structures that are more than 7 kJ/mol above the respective predicted global minimum, so these structures are represented with grey crosses.

The molecular property λ_- can be used to rank molecules as potential organic semiconductors without the use of CSP to provide structure predictions or any other form of structural hypothesis.

There is a wide range of values for the reorganization energy, from 0.200 (mol. **1**) to 0.341 eV (mol. **14**). Thus, this measure predicts molecule **1** to be the best molecule among the entire set, followed closely by **5** and **7**, whose λ_- are only 1 and 5 meV higher than that of **1**, respectively.

The drawback of assessing molecules without considering crystal packing is clear from the weak relationship between μ_{GM} and λ_-; for example, **7**, which has one of the lowest values of λ_-, has a very low predicted electron mobility in its global minimum energy predicted crystal structure. In contrast, some of the isomers with high λ_- achieve good electron mobilities in their lowest energy crystal structures; for example, molecule **15** has μ_{GM} 5.6 times greater than that of molecule **1**, despite a 53 meV higher λ_-. Thus, we find that, over this range of reorganization energies, strong electronic coupling due to favourable crystal packing can make up for a moderately high λ_- while, conversely, a molecule with a promising, low λ_- can lead to poor properties if it adopts an unfavourable crystal packing.

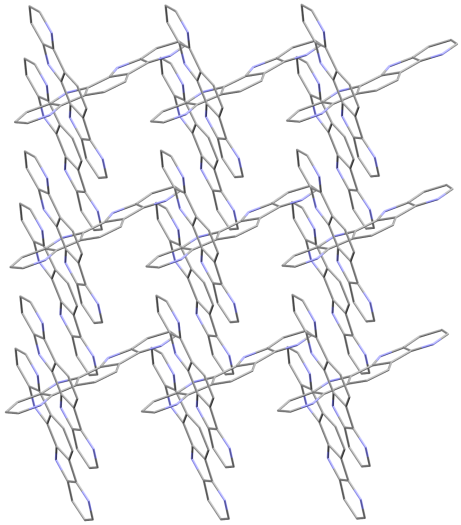

Figure 6: Crystal packing of the predicted global minimum energy crystal structure for molecule **4**. Hydrogen atoms are hidden.

Molecule **4** ranks second among the whole set in terms of μ_{GM}, adopting a similar γ–type packing as **1** (Fig. 6). **4** is also exceptional because there are no other predicted crystal structures within the 7 kJ/mol lattice energy window of the global minimum. Such large energy gaps between structures are unusual on CSP landscapes and this gives a high confidence that the predicted, high μ_e structure would be observed if this molecule was synthesized. The energy differences between predicted structures are much smaller for many of the other molecules, meaning that the absolute ranking, and hence μ_{GM}, is sensitive to uncertainties due to limitations of the force field model, as well as the neglect of lattice dynamical contributions to relative free energies[28] and the influence of thermal expansion[39,40] which, combined, can contribute a few kJ/mol to lattice energy differences.

Molecule **5** (μ_{GM} = 2.706 cm^2/Vs) has the highest predicted electron mobility among the global lattice energy minimum crystal structures of all 28 isomers, eclipsing the calculated mobility of the observed (and global minimum predicted) crystal structure of **1** (μ_e = 0.240 cm^2/Vs). The extended hydrogen–bonded flat ribbon structure in the predicted global minimum structure for **5** (Fig. 7a)

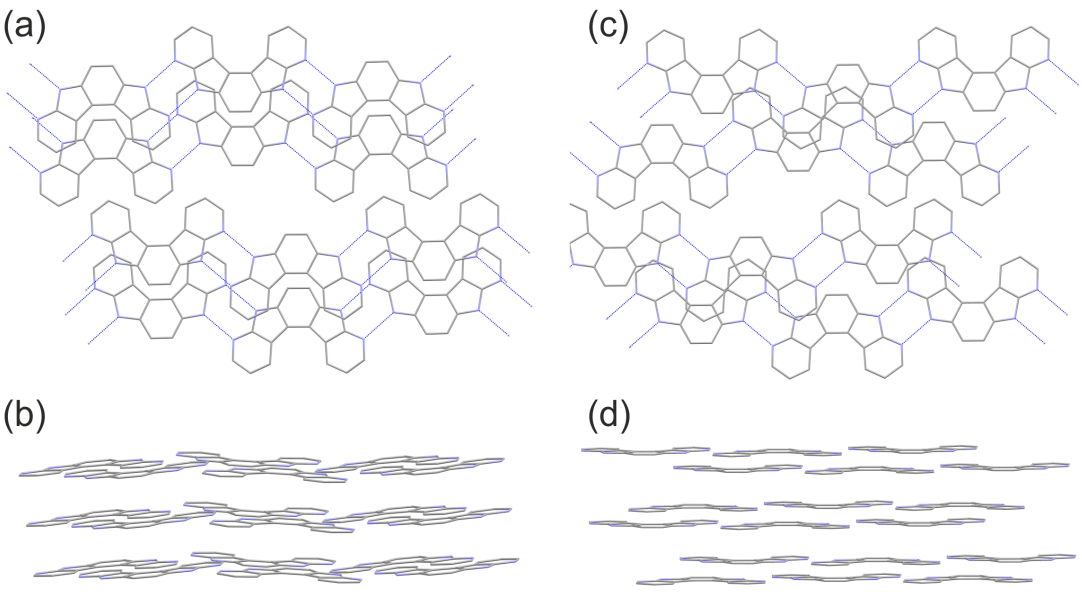

Figure 7: Crystal packing of the the predicted global minimum (a,b) and the structure with the maximum electron mobility (c, d) for **5**, viewed normal to and along the molecular sheets. Hydrogen atoms are hidden and hydrogen bonds are shown as dashed blue lines in (a) and (c).

promotes π–stacking along the out–of–plane direction. **5** is the only molecule for which the low energy part of the CSP landscape is dominated by sheet–like crystal packing (Fig. 5) and is also the molecule with the largest $\mu_{\max}$, which is found in a higher energy structure (Fig. 7b) with similar hydrogen–bonded ribbons as the predicted global minimum.

The $\mu_{\max}$ results show that the global energy minimum crystal structure rarely gives the best electron mobility and structures with the highest electron mobilities are predicted at elevated lattice energies for many of the screened molecules (Table 2, Fig. 5). This is consistent with our previous investigations,[15] where dimers within crystal structures with significant co–facial overlap (giving rise to largest t and μ) are often energetically penalized by relatively large exchange–repulsion interactions.

Finally, we compare $\langle\mu\rangle$, which gives a probabilistic assessment of the molecules, taking into account all of their low energy predicted crystal structures. By this measure, molecules **5** and **4** are most promising, although their order is reversed compared to previous measures ($\mu_{\max}$ and μ_{GM}); molecule **4** has the largest $\langle\mu\rangle$ across the whole set. This originates from the high μ_{GM} for **4** and the large energy gap between the global minimum and the second most stable structure on the predicted lattice energy landscape, leading to $\langle\mu\rangle = \mu_{\max} = \mu_{GM}$. Molecule **5** is slightly less favourable according to $\langle\mu\rangle$ because, despite very high μ_{GM} and $\mu_{\max}$, the landscape for **5** also contains many low energy crystal structures with low predicted electron mobilities (Fig. 5).

The difference in ranking of the molecules between the single-crystal structure measures ($\mu_{\max}$ and μ_{GM}) and the landscape-averaged approach ($\langle\mu\rangle$) provides an interesting dilemma: structure prediction is more certain for **4** because there is only one low-energy predicted crystal structure, so we would be more confident in obtaining this structure with its high predicted μ_e. Molecule **5**,

on the other hand, has a greater potential because of higher μ_e in some of its predicted structures. However, a risk in targeting **5** for synthesis is the crowded energy landscape that also includes low-mobility structures and the possibility that one of these less promising crystal structures is adopted instead of the predicted global minimum or the structure with $\mu_{\max}$. The choice of which to target synthetically, or whether to synthesize both candidate molecules, depends on the associated synthetic difficulty, which we do not attempt to judge in the present work.

Remarks on Computational Cost

Within the broad context of computational screening of functional materials, the accuracies of the underlying computational methods are not the only scientific challenge but also the associated cost in terms of computational time required, which has a direct impact on the feasible size of chemical search space that can be comfortably explored.

In this work, the total cost of CSP for all 28 molecules was approximately 31.7 kilo CPU hours (corresponding to just under a week on 200 dedicated CPUs), whereas the cost of computing the charge transfer integrals for the 8786 molecular dimers needed for electron mobility calculations using DFT was approximately 43.9 kilo CPU hours. The computational cost associated with structure and property predictions is similar and the high cost of running DFT calculations for property prediction is immediately apparent. Ongoing efforts to accelerate these types of property calculations, using methods such as machine-learning of structure-property relationships,[19] are required to further advance the field.

Conclusions

Crystal structure prediction has sometimes been seen as computationally demanding and slow, which would be unsuitable for screening purposes within a suitable time scale for material discovery. With advances in sampling methods, making use of efficient, accurate atom-atom force fields and utilizing parallel, high performance computing, this is no longer true for rigid molecules of the size studied here. In this work, we demonstrate the combination of CSP with charge mobility predictions to screen 28 isomeric molecules as potential organic semiconductors. These molecules were hypothesized as structural variants of the recently synthesized **1**. We further combined a recently developed machine–learning method to construct, for the first time, a multi–landscape sketch–map to directly compare the crystal packing landscapes of a set of related molecules to reveal common structural classes among the landscapes and to correlate these classes with crystal properties.

Ongoing challenges still remain to make the most of the methods presented here in the process of the discovery of new materials. Some of the most pressing are (i) the selection of a diverse set of molecular candidates – in this work, the set of molecules has been constructed manually, based on chemical intuition – (ii) acceleration of the crystal structure prediction and property prediction calculations which, in this work, contribute almost equally to the overall computational cost, and (iii) the determination of a data-driven procedure to determine the best parameters for the sketch-map projection and the kernel construction to further automate the structural analysis across multiple crystal structure landscapes, revealing the relation between molecular structure, crystal packing and materials properties.

The variation of electron mobility within the crystal structure landscape of a given molecule, and between landscapes of related molecules, demonstrates the important influence of crystal packing on this property: improved intermolecular coupling can more than make up for a moderate increase in the molecular reorganization energy. Within the molecules included in our study, we find that the most commonly favored crystal packing across all molecules investigated here is the γ packing motif, which is predicted to be favored for molecules with different point-group symmetry and functional group substitution patterns. Furthermore, sketch–map and cluster analysis suggests that crystal structures from different molecules that favor the γ–type packing exhibit a high degree of structural similarity, which can be attributed to the abilities of these molecules to form similar short-ranged hydrogen-bond patterns.

With regards to the screening of electron mobilities, we find that a number of the hypothetical molecules are predicted to give higher electron mobilities than molecule **1**. Molecules **4** and **5** are most promising among the 28 molecules that we considered. Molecule **5** has the second lowest reorganization energy and a crystal energy landscape that is dominated by sheet-like crystal packings, leading to very high electron mobility in some its predicted crystal structures, including its global lattice energy minimum and a higher energy crystal structure that gives the highest electron mobility across all 28 crystal structure landscapes. Molecule **4** is attractive because of high electron mobility in the global minimum energy crystal structure and a large energy gap to the next hypothetical crystal structure; the global energy minimum is highly likely to be observed in this case.

The study represents an important advance in the use of CSP for guiding materials discovery, demonstrating the ability to screen a moderately large set of molecules on a timescale that is shorter than what would be required for synthesis and testing of the set of candidates. We hope that this work stimulates more widespread use of these computational methods, as well as encouraging the synthesis and characterization of some of the promising hypothetical molecules that have been identified.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

GMD, JY and JEC acknowledge funding by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC (grant agreement number 307358, ERC-stG-2012-ANGLE) and acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton. JY acknowledge the financial support from Australian Nuclear Science and Technology Organization for a Postdoctoral Fellowship. JY and SL acknowledge the funding from the Australian Research Council (grant DP170104831). JY and SL acknowledge computational resources from National Computing Infrastructures, Australia under the Intersect partnership scheme (project dy3) as well as the merit allocation scheme from the Research Technology Services, UNSW. SD was supported by the NCCR MARVEL, funded by the Swiss National Science Foundation. MC acknowledges funding by the

European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 677013-HBMAP).

References

(1) Jain, A.; Shin, Y.; Persson, K. A. Computational Predictions of Energy Materials Using Density Functional Theory. *Nature Rev. Mater.* **2016**, *1*, 15004.

(2) Chen, W.; Pohls, J.-H.; Hautier, G.; Broberg, D.; Bajaj, S.; Aydemir, U.; Gibbs, Z. M.; Zhu, H.; Asta, M.; Snyder, G. J.; Meredig, B.; White, M. A.; Persson, K.; Jain, A. Understanding Thermoelectric Properties from High–Throughput Calculations: Trends, Insights, and Comparisons with Experiment. *J. Mater. Chem. C* **2016**, *4*, 4414–4426.

(3) Gorai, P.; Stevanovicć, V.; Toberer, E. S. Computationally Guided Discovery of Thermoelectric Materials. *Nat. Rev. Mater.* **2017**, *2*, 17053.

(4) Collins, C.; Dyer, M.; Pitcher, M.; Whitehead, G.; Zanella, M.; Mandal, P.; Claridge, J.; Darling, G.; Rosseinsky, M. Accelerated Discovery of Two Crystal Structure Types in A Complex Inorganic Phase Field. *Nature* **2017**, *546*, 280–284.

(5) Chua, A. L.-S.; Benedek, N. A.; Chen, L.; Finnis, M. W.; Sutton, A. P. A Genetic Algorithm for Predicting the Structures of Interfaces in Multicomponent Systems. *Nature Mater.* **2010**, *9*, 418–422.

(6) Shang, C.; Liu, Z.-P. Stochastic Surface Walking Method for Structure Prediction and Pathway Searching. *J. Chem. Theory Comput.* **2013**, *9*, 1838–1845.

(7) Curtis, F.; Li, X.; Rose, T.; Vázquez-Mayagoitia, Á.; Bhattacharya, S.; Ghiringhelli, L. M.; Marom, N. GAtor: A First-Principles Genetic Algorithm for Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2018**, *14*, 2246–2264.

(8) Aina, A. A.; Misquitta, A. J.; Price, S. L. From Dimers to the Solid–State: Distributed Intermolecular Force–Fields for Pyridine. *J. Chem. Phys.* **2017**, *147*, 161722.

(9) Hermann, J.; DiStasio, R. A.; Tkatchenko, A. First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chem. Rev.* **2017**, *117*, 4714–4758.

(10) Beran, G. J. A New Era for Ab initio Molecular Crystal Lattice Energy Prediction. *Angew. Chem. Inter. Ed.* **2015**, *54*, 396–398.

(11) Reilly, A. M. et al. Report on the Sixth Blind Test of Organic Crystal Structure Prediction Methods. *Acta Cryst. B* **2016**, *72*, 439–459.

(12) 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.

(13) Day, G. M.; Cooper, A. I. Energy–Structure–Function Maps: Cartography for Materials Discovery. *Adv. Mater.* **2017**, 1704944.

(14) Pulido, A. et al. Functional Materials Discovery Using Energy–Structure–Function Maps. *Nature* **2017**, *543*, 657–664.

(15) Campbell, J. E.; Yang, J.; Day, G. M. Predicted Energy–Structure–Function Maps for the Evaluation of Small Molecule Organic Semiconductors. *J. Mat. Chem. C* **2017**, *5*, 7574–7584.

(16) Rice, B.; LeBlanc, L. M.; Otero-de-la Roza, A.; Fuchter, M. J.; Johnson, E. R.; Nelson, J.; Jelfs, K. E. A Computational Exploration of the Crystal Energy and Charge-Carrier Mobility Landscapes of the Chiral [6]helicene Molecule. *Nanoscale* **2018**, *10*, 1865–1876.

(17) Hachmann, J.; Olivares-Amaya, R.; Jinich, A.; Appleton, A. L.; Blood-Forsythe, M. A.; Seress, L. R.; Roman-Salgado, C.; Trepte, K.; Atahan-Evrenk, S.; Er, S.; Shrestha, S.; Mondal, R.; Sokolov, A.; Bao, Z.; Aspuru-Guzik, A. Lead Candidates for High–Performance Organic Photovoltaics from High–Throughput Quantum Chemistry–the Harvard Clean Energy Project. *Energy Environ. Sci.* **2014**, *7*, 698–704.

(18) Sokolov, A. N.; Atahan-Evrenk, S.; Mondal, R.; Akkerman, H. B.; Sánchez-Carrera, R. S.; Granados-Focil, S.; Schrier, J.; Mannsfeld, S. C.; Zoombelt, A. P.; Bao, Z.; Aspuru-Guzik, A. From Computational Discovery to Experimental Characterization of a High Hole Mobility Organic Crystal. *Nature Comm.* **2011**, *2*, 437.

(19) Musil, F.; De, S.; Yang, J.; Campbell, J. E.; Day, G. M.; Ceriotti, M. Machine Learning for the Structure-Energy-Property Landscapes of Molecular Crystals. *Chem. Sci.* **2018**, *9*, 1289–1300.

(20) Gómez, P.; Más-Montoya, M.; da Silva, I.; Cerón-Carrasco, J. P.; Tárraga, A.; Curiel, D. Hydrogen Bond-Directed Cruciform and Stacked Packing of a Pyrrole–based Azaphenacene. *Crystal Growth & Design* **2017**, *17*, 3371–3378.

(21) Desiraju, G. R.; Gavezzotti, A. Crystal Structures of Polynuclear Aromatic Hydrocarbons. Classification, Rationalization and Prediction from Molecular Structure. *Acta Crystallogr., Sect. B: Struct. Sci* **1989**, *45*, 473–482.

(22) 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.

(23) Frisch, M. J. et al. Gaussian 09. 2009.

(24) Chisholm, J. A.; Motherwell, S. COMPACK: A Program for Identifying Crystal Structure Similarity Using Distances. *J. Appl. Cryst.* **2005**, *38*, 228–231.

(25) Price, S. L.; Leslie, M.; Welch, G. W. A.; 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.

(26) Pyzer-Knapp, E. O.; Thompson, H. P. G.; Day, G. M. An Optimized Intermolecular Force Field for Hydrogen–Bonded Organic Molecular Crystals Using Atomic Multipole Electrostatics. *Acta Cryst. B* **2016**, *72*, 477–487.

(27) Stone, A.; Alderton, M. Distributed Multipole Analysis Methods and Applications. *Molecular Physics* **2002**, *100*, 221–233.

(28) Nyman, J.; Day, G. M. Static and Lattice Vibrational Energy Differences between Polymorphs. *CrystEngComm.* **2015**, *17*, 5154–5165.

(29) Bhardwaj, R. M.; Price, L. S.; Price, S. L.; Reutzel-Edens, S. M.; Miller, G. J.; Oswald, I. D. H.; Johnston, B. F.; Florence, A. J. Exploring the Experimental and Computed Crystal Energy Landscape of Olanzapine. *Crystal Growth & Design* **2013**, *13*, 1602–1617.

(30) Corpinot, M. K.; Stratford, S. A.; Arhangelskis, M.; Anka-Lufford, J.; Halasz, I.; Judas, N.; Jones, W.; Bucar, D.-K. On the Predictability of Supramolecular Interactions in Molecular Cocrystals - the View from the Bench. *CrystEngComm* **2016**, *18*, 5434–5439.

(31) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87*, 184115.

(32) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing Molecules and Solids across Structural and Alchemical Space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.

(33) Ceriotti, M.; Tribello, G. A.; Parrinello, M. From the Cover: Simplifying the Representation of Complex Free-Energy Landscapes using Sketch-Map. *PNAS* **2011**, *108*, 13023–13028.

(34) Ceriotti, M.; Tribello, G. A.; Parrinello, M. Demonstrating the Transferability and the Descriptive Power of Sketch-Map. *J. Chem. Theory Comput.* **2013**, *9*, 1521–1532.

(35) Nyman, J.; Pundyke, O. S.; Day, G. M. Accurate Force Fields and Methods for Modelling Organic Molecular Crystals at Finite Temperatures. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15828–15837.

(36) De, A.; Ghosh, R.; Roychowdhury, S.; Roychowdhury, P. Structural Analysis of Picene, $C_{22}H_{14}$. *Acta Crystallographica Section C* **1985**, *41*, 907–909.

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

(38) Gajdos, F.; Valner, S.; Hoffmann, F.; Spencer, J.; Breuer, M.; Kubas, A.; Dupuis, M.; Blumberger, J. Ultrafast Estimation of Electronic Couplings for Electron Transfer between π-Conjugated Organic Molecules. *J. Chem. Theory Comput.* **2014**, *10*, 4653–4660.

(39) Nyman, J.; Day, G. M. Modelling Temperature-Dependent Properties of Polymorphic Organic Molecular Crystals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 31132–31143.

(40) 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.