Supplementary Information for A Robust Crystal Structure Prediction Method to Support Small Molecule Drug Development with Large Scale Validation and Blind Study

Dong Zhou,[1,2] Imanuel Bier,[1] Biswajit Santra,[1] Leif D. Jacobson,[3] Chuanjie Wu,[1] Adiran Garaizar Suarez,[4] Barbara Ramirez Almaguer,[5] Haoyu Yu,[1,6] Robert Abel, [1] Richard A. Friesner,[7] Lingle Wang[1]*

Affiliations:
1 Schrödinger, New York: 1540 Broadway, 24th Floor, New York, NY, USA, 10036
2 Current address: Atommap Inc. 450 Lexington Avenue, 4th floor, New York, NY, USA 10017
3 Schrödinger, Portland: 101 SW Main Street, Suite 1300, Portland, OR, USA, 97204
4 Bayer AG, Computational Life Science, Alfred-Nobel-Straße 50, 40789 Monheim am Rhein, Germany, 40789
5 Bayer AG, Process Industrialization, Friedrich-Ebert-Str. 217 Wuppertal, Germany, 42117
6 Current address: ByteDance Inc. 151 w 42nd street, New York, NY, USA 10036
7 Department of Chemistry, Columbia University, New York, New York, USA, 10027

* Corresponding Author E-mail: Lingle.wang@schrodinger.com

Data from further validation of structure searching methods

To further validate the conformer generation method, the CSP test set was enriched with examples from the CSD drug subset with 9 or fewer rotatable bonds from Z'=1 crystal structures. In total, the dataset consists of 430 experimental crystal conformers. The RMSD of the results from the conformer generation protocol are presented in Table S1. Due to the exhaustive nature of our method, high quality conformers are always generated across the chemical space tested, even for large flexible molecules.

Table S1: Conformer generation accuracy on a dataset of 430 conformers from the CSP test set and the CSD drug subset. The mean and upper value of a 95% confidence interval of the RMSD between the atomic coordinates of the experimentally known crystal conformation and the best matching conformer from the conformer generation protocol are presented.

Number of Rotatable Bonds	0	1	2	3	4	5	6	7	8	9
Dataset Size	17	51	45	79	76	57	47	32	17	9
RMSD Mean [Å]	0.071	0.159	0.146	0.173	0.194	0.195	0.226	0.263	0.245	0.237
RMSD Upper Value of 95% Confidence Interval [Å]	0.140	0.282	0.271	0.356	0.336	0.350	0.340	0.378	0.331	0.299

To benchmark the packing search protocol, we curated a dataset of 426 neat Z'=1 crystal polymorphs from the CSD drug subset. (Table S2) These crystals are relaxed with symmetry constrained MD using the OPLS4 force field. Good crystal similarity is achieved for the full dataset, demonstrating the accuracy of the OPLS4 in describing molecular crystal packing interactions. The crystal-relaxed ASU is then used to search for the relaxed crystal structure. For the 426 crystals used in this benchmark, a candidate structure matching experiment was generated with close to 100% success rate.

The crystal packing sampling was considered successful when a candidate crystal structure with RMSDn better than (20, 1.35 Å) as compared to the relaxed experimental structure was generated. Although this is slightly looser than CCDC blind test success criterion (20, 0.8 Å),[39] it guarantees the close matching of the overall packing pattern between the reference and predicted crystal structures.

Table S2: Crystal packing search accuracy on a dataset of 426 neat Z'=1 crystal polymorphs from the CSD drug subset (the space groups annotated with stars does not have instances in the CSD drug subset thus additional cases were added). Close to 100% success rate was obtained for crystals occupying the top 22 most frequently observed space groups, accounting for ~89% of small organic crystals in CSD.

Space Group	$P2_1/c$	$P2_12_12_1$	$P2_1$	P_{-1}	C2/c	Pbca	$Pna2_1$	C2
Success Rate	111/111	108/110	49/51	48/49	24/24	15/15	18/18	9/9

Space Group	P1	Cc	Fdd2	$P2_12_12$	P41	R-3	P2/c	Pbcn
Success Rate	7/7	5/5	4/4	4/4	4/4	2/2	2/2	2/2

Space Group	Pca21	Pc	P31	Pnna*	P43*	P32*
Success Rate	2/2	2/2	2/2	1/1	1/1	1/1

More details on QRNN

In this work we have used a transferable machine learning force field (MLFF) as an intermediate energy model between relatively low energy resolution molecular mechanics force field and high energy resolution periodic DFT. A number of prior works have demonstrated the potential for using an MLFF for the task of organic crystals structure ranking,[1,2,3,4,5,6,7] often with impressive results. These studies have largely focused on molecules lacking significant conformational flexibility,[1,2,3,4] with a few reports on limited[3] but nonetheless challenging[5] conformational degrees of freedom. Largely these works have also focused on training to one system at a time and do not utilize a pretrained model. Recently, a pretrained model was incorporated into an end to end CSP workflow which combined a random search, a genetic algorithm, and an MLFF energy function with moderate success.[6] Additionally, during the revision of this manuscript the results of the 7th CSP Blind Test were reported.[8,9] Several participants utilized MLFF models in that work, see the main text for a more detailed discussion. In this work we integrate a pretrained MLFF into a hierarchical ranking strategy utilizing a classical force field at the most coarse grain,

followed by MLFF ranking and finally periodic DFT ranking. We further provide a demonstration of improving the ranking accuracy of the MLFF by retraining to cluster geometries.

The MLFF model architecture we use here, QRNN, is briefly summarized and readers interested in further details are referred to previous reports.[10,11] QRNN is a modification of the high dimensional neural network potential (HDNNP)[12] architecture designed to equally support ionic and neutral systems as well as describe long-range interactions with empirical functional forms. Atomic neural networks are used to predict atomic electronegativities, using Behler-Parinello type symmetry functions,[13] or atomic environment variables (AEV) as input.[14] These predicted atomic electronegativities are then used in a simplified charge equilibration (Qeq) formula to determine atomic charges. As such, our models resemble the fourth-generation HDNNP architecture.[12,15] This charge prediction process is recursive, albeit limited to two iterations, which gives the model its name.[10] Once the charges have been determined, they are used as features, along with AEV and a charge-weighted radial AEV, to an atomic network that predicts energies. To compute the total energy, we sum the short ranged neural network atomic energies, the long-ranged Coulomb energy from the predicted atom charges, and Grimme's dispersion correction[16] combined with the damping function proposed by Chai and Head-Gordon.[17] As a result, this model captures long-range Coulomb and dispersion interactions while also representing short-range interactions with high precision.

In the pretrained and molecule specific QRNN models used in this work we also utilize multi-task learning to two levels of theory, ωB97X-D/6-31G(d) and ωB97X-D3/def2-TZVP.[18] We have previously built a very large training set of neutral organic molecules,[19] organic ions[8] and clusters, labeled at the ωB97X-D/6-31G(d) level, which provides reasonably broad coverage of conformational energies of druglike molecules and ions, proton transfer energies and intermolecular interactions. This level of theory is affordable but, with such a modest basis set, has potentially significant basis set incompleteness errors, particularly important for accurately describing intermolecular interactions. We have found that the model, through a multitask learning strategy, is able to leverage this relatively inexpensive level of theory to achieve better prediction accuracy on the higher level of theory with comparatively smaller amounts of training data.[18] In this case, we utilize the OrbNet Denali dataset[20] as a source of training data for the higher level of theory. All predictions in this work utilize the output head corresponding to the higher level of theory (ωB97X-D3/def2-TZVP). In addition, our model has been trained exclusively on finite system data, meaning no periodic DFT data is used in the training of our model.

All optimizations of crystal structures using QRNN are performed with the atomic simulation environment (ASE).[21] Optimizations are performed with the BFGS method utilizing a line search. Lattice parameters are optimized simultaneously with atomic positions until the root mean square of gradient components is less than 0.02 eV / Å. Space group symmetry constraints are utilized to maintain symmetry.

QRNN improvement of bespoke training

The QRNN model used in this work was trained on a broad dataset of druglike molecules, ions and clusters. As shown in Figure S1 the pretrained QRNN model is able to rank all experimentally observed polymorphs below 6 kcal/mol in all but one case (Oxalyl dihydrazide). Further, the number of structures in this energy window is low enough to permit, in our view, an affordable number of periodic DFT calculations to be run (fewer than 1200) and successfully find all experimental polymorphs including the above mentioned high energy case. Nonetheless, this transferable model leaves significant room for improvement as it is known that ML based force fields can routinely be trained to well within chemical accuracy.

We have explored the possibility of improving the accuracy of the pretrained model for use on a specific molecular system on 36 of the worst performing cases; systems for which any experimentally observed polymorph has an energy, according to the pretrained QRNN model of greater than 4 kcal/mol or has greater than 1000 crystals in the 6 kcal/mol window. For the remaining systems, the pretrained model was found to already be accurate and we do not explicitly study a refinement procedure. In the general case we would apply this refinement procedure to all systems prior to use of the MLFF. For each of these systems we generate a molecule specific training set and retrain (transfer learn) the pretrained model to this set and test the improved performance. Briefly, we use a conformational search and molecular dynamics to sample molecular configurations and an abridged packing search, combined with normal mode sampling, to generate cluster configurations. We label these geometries with two levels of theory in equal proportion and limit the estimated computational cost to 10,000 cpu hours for each system. An ensemble of *molecule specific* QRNN models is trained to a 90/10 random split of the labeled data for a fixed number of cycles.

The improved molecule specific QRNN models have been tested similarly to the pretrained model. All of the crystal geometries from the classical FF based filtration, together with the experimentally determined geometries are optimized with the new model. Figure S2 gives the energy and rank of the highest energy polymorph in each of the 36 test systems. This simple retraining procedure reduces the highest energy observed to just below 4 kcal/mol with three systems having energy higher than 3 kcal/mol, only slightly higher than the highest energies observed with periodic DFT ranking. The highest rank after this molecule specific improvement is just below 300 with only five systems having ranks higher than 100. This demonstrates that the ranking results can be greatly improved and that all experimentally observed polymorphs can be found by performing 300 or fewer expensive periodic DFT calculations. It may still be possible to improve the models further by sampling clusters out of low-lying crystal structures found after ranking with these models as well as incorporating training to periodic data and work along these lines is underway. We plan to give a detailed report of the retraining procedure in the near future. Additional training of QRNN on elements such as Bromine, Iodine, and Boron is also underway.

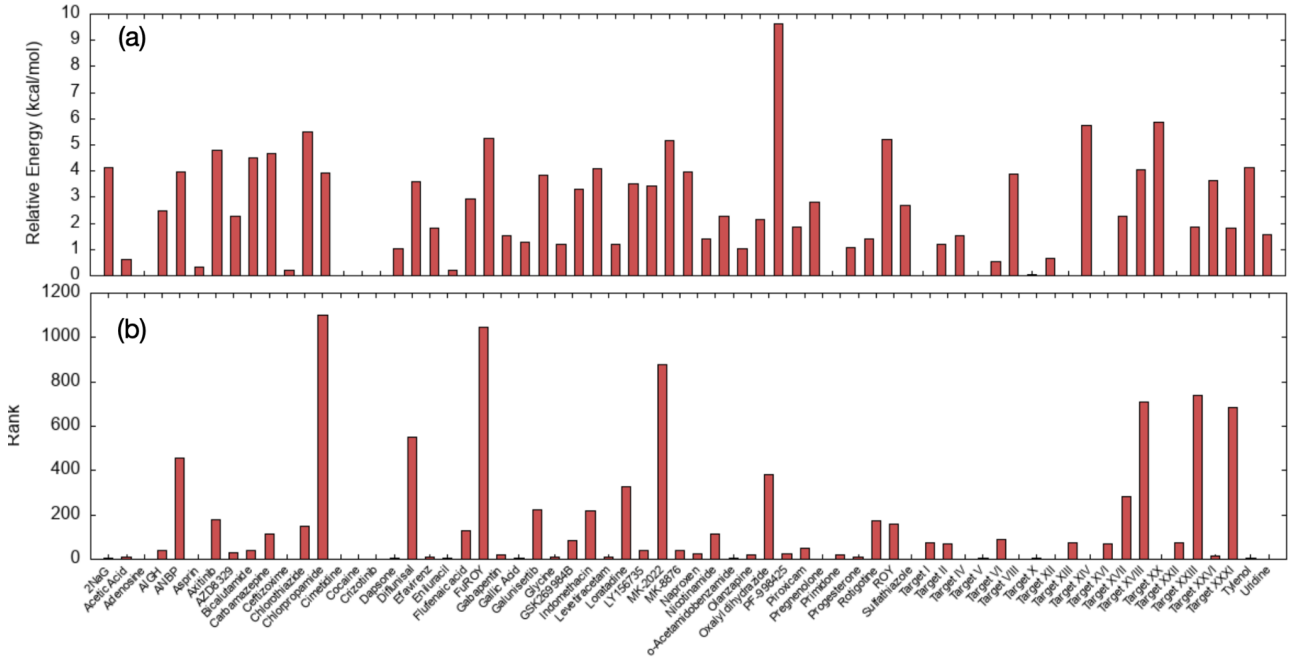

Figure S1: Relative energy (a) and rank (b) of the highest energy experimentally observed polymorph in each of 66 test systems, according to the pretrained QRNN model. Source data are provided as a Source Data file.

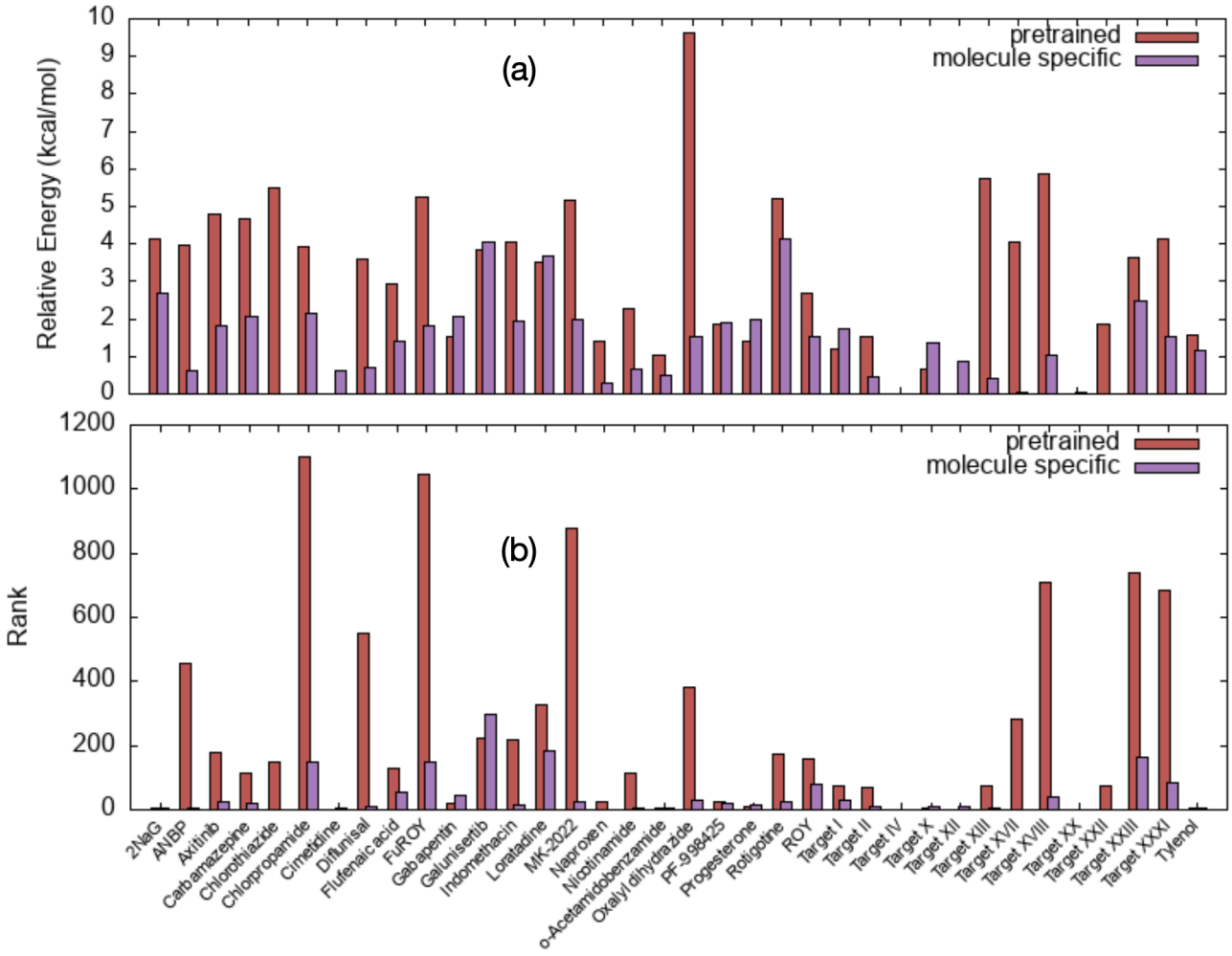

Figure S2: Relative energy (a) and rank (b) of the highest energy experimentally observed polymorph in each of 36 test systems, according to pretrained and molecule specific QRNN models. Source data are provided as a Source Data file.

Details on molecule specific training procedure

Our pretrained QRNN model has been broadly trained on druglike molecules, ions of druglike molecules and cluster geometries extracted from the CSD. We have previously shown that this model is able to accurately describe the conformational energetics of such druglike molecules.[8] Even though we have not explicitly trained it to periodic crystalline systems, this model ought to be able to account for relative energies of molecular polymorphs with reasonable precision, on the kcal/mol energy scale. However, for molecules that lie outside of the druglike space we would expect higher errors and we believe some of the molecules encountered in this work fall outside of this space. In fact, many of the rigid and semi-rigid blind test systems we would consider not to be druglike. Even within the druglike space, it may be unreasonable to assume one could

pretrain an empirical model to have uniform accuracy across even this relatively limited domain due to the well known large size of the chemical space. In fact, one might reasonably expect that molecules with high error will be encountered with non-negligible probability. For these reasons it is important to develop an efficient protocol for improving the pretrained, transferable model for a specific chemical system.

The retraining of a model is performed by: 1. Sampling molecular and cluster configurations; 2. Selecting a set of configurations to be labeled by DFT; 3. Labeling the configurations with DFT; 4. Retraining the pretrained model and 5. Deploying the model for improved ranking. The molecular (monomer) configurations are obtained by first performing a MacroModel and Fast3D based conformational search to identify a number of conformers equal to three times the number of rotatable bonds for each conformational search method. Each of these conformers is optimized by the pretrained QRNN model. In addition, we construct a total of 10 times the number of rotatable bond rotamers, each of which is constructed by randomly sampling three torsion angles and performing constrained optimization. These structures represent off equilibrium conformers. For each of the conformers we perform a short molecular dynamics trajectory in the NVT ensemble and sample 10 representative geometries. For each of the rotamers we sample 5 geometries by normal model sampling (NMS) at an elevated temperature of 600K. All monomer geometries are labeled with both levels of DFT.

To sample cluster configurations we perform a limited packing search. This search is performed over a subset of 21 search directions which are identified as inexpensive search paths. For each path we randomly select five conformers from the pool identified in the previous step. After the packing search is performed we generate subclusters of all clusters. We only keep clusters smaller than 110 atoms and containing fewer than 7 molecules. These clusters are then deduplicated before generating 5 NMS samples per cluster. Using the NMS samples as a pool of structures we select configurations such that the total estimated cost is capped. The cost is equally distributed amongst the two levels of theory and each cluster size. As such we typically have a much larger number of smaller clusters than larger ones. In addition, 80% of the clusters are chosen such that the predicted QRNN energies are Boltzmann distributed and the remaining 20% are uniformly distributed.

DFT energies, forces and dipole moments at two levels of theory are computed using Jaguar v11.8. All DFT calculations are performed with default settings and utilize the pseudo-spectral approximation. The approximate cost is determined by fitting a cost function to a cubic polynomial in the number of basis functions. The two levels of theory are ωB97X-D/6-31G(d) and ωB97X-D3/def2-TZVP, which are consistent with the level of theory used by us previously and that used by the OrbNet Denali dataset.[20]

Finally, a new, transfer learned, molecule specific QRNN model ensemble is trained. This is done by starting the training at the previously determined weights and biases of the pretrained models. We then train to a 90/10 train/test split of a single system's training data. We perform training for up to 500 cycles of 10,000 examples / cycle with a batch size of 128. We use the AdamaxW optimizer with an initial learning rate of 1.0e-3. As previously described, we utilize a multi-task

loss function which includes the RMSE of energies and dipole moments.[10,11,18] In this work we additionally train atomic forces. We utilize early stopping with a patience of zero but we rarely stop before the maximum number of training cycles is exhausted. An ensemble of 5 such models is trained and deployed for subsequent crystal optimizations.

For Target XXXI, we have trained to periodic DFT data to generate the molecule specific QRNN model. We have used a random sample of 800 crystal structures that came within 12 kcal/mol with the pretrained model. For the selected crystal structures PBE-D3 energy, force, and stress are computed with QE. The periodic PBE-D3 energy, force, and stress are then used to train a new molecule specific QRNN model ensemble following the procedure detailed above.

Extension of packing search to support more complex crystals

Currently the top 22 space groups are covered, accounting for about 90% of the known small molecule crystals. They are P21/c, P212121, P21, P-1, C2/c, Pbca, Pna21, C2, P1, Cc, Fdd2, P21212, P41, P43, R-3, P2/c, Pbcn, Pca21, Pc, P31, P32, and Pnna. More space group coverage will be provided in the future version.

These developments on the workflow throughput and applicable domain expansion will enable routine crystal structure prediction for large, flexible, compound systems such as Z'>1, hydrates, salts, and cocrystals in the near future.

The current implementation is feasible for routine usage to support small molecule drug formulation projects, and has already demonstrated its value in Schrodinger's internal drug discovery pipelines.

More details on DFT

Final energy evaluation & ranking of crystal structures was performed using periodic DFT. DFT calculations were performed with a hierarchical minimization workflow to reduce the computational cost. In this work, structures are optimized with a widely used generalized-gradient approximation (GGA) functional, PBE,[22] in conjunction with two- and three-body dispersion corrections (D3) parameterized by Grimme and coworkers,[23] *i.e.* PBE-D3. The optimized structures are reranked with single point energy (SPE) calculations using a dispersion corrected meta-GGA functional, r^{2}SCAN-D3.[24,25] The r^{2}SCAN meta-GGA satisfies significantly more exact constraints for the exchange-correlation energy compared to PBE.[25] A few benchmark studies have shown that r^{2}SCAN-D3 describes energy and structure in a variety of chemical environments more accurately than PBE-D3, in particular, for hydrogen bonding and dispersion interactions that are typically present in organic molecular crystals.[24–26] For the X23[27] and ICE10[28] benchmark sets the accuracy of dispersion corrected r^{2}SCAN is comparable to dispersion corrected hybrid functionals. In addition, r^{2}SCAN is significantly less expensive to compute than hybrid

functionals[26] making it a good candidate for CSP. Here, we provide the final energy ranking from r^{2}SCAN-D3 for all cases. In some cases, to cross-verify the ranking by r^{2}SCAN-D3, we used the more expensive PBE0-MBD, a many-body dispersion (MBD)[29] inclusive global hybrid functional, which was shown to accurately rank crystal structures of some of the molecules used in the blind tests[30] and a few other organic molecules.[31,32]

All periodic DFT calculations are performed with Quantum ESPRESSO (QE)[33] within the Schrödinger suite. Ultrasoft pseudopotentials (GBRV type)[34] are used for PBE-D3 structural relaxation with a wavefunction kinetic energy cutoff of 50 Ry, whereas norm-conserving pseudopotentials (ONCV type)[35] are used for r^{2}SCAN-D3 and PBE0-MBD SPE calculations with a wavefunction kinetic energy cutoff of 80 Ry. A k-point sampling density of 0.05 $Å^{-1}$ is found to be sufficiently converged for all functionals used here. In all cases, electronic wavefunctions were self-consistently converged to 10^{-6} Ry. During structural relaxation, all coordinates and lattice parameters were relaxed without imposing any constraint on the space group symmetry.

We screen the energy of the structures with increasing relaxation convergence criteria to reduce the computational cost of DFT evaluation. Within our periodic DFT code, the BFGS minimization of the energy and force of crystal structures tends to become slow near the potential energy minima when the gradients become relatively small. Because of that we aim to discard higher energy structures as early as possible and retain only the lower energy structures for very tight optimization near its potential energy minima. Initially, the lowest structures from QRNN are optimized with PBE-D3 using a loose convergence criteria on energy ($dE_{opt.}$ < 0.31 kcal/mol) and maximum component of the cartesian forces (F_{max} < 0.26 eV/Å). We refer to this step as PBE-D3$_{light}$. With this optimization criteria the highest rank (R_{max}) of the experimentally known polymorphs appears to be 94. With respect to the rank 1 structure the largest relative energy ($dE_{rel.}$) of the experimentally known polymorphs is 3.6 kcal/mol. Based on this data we only select a smaller subset of the original structures, having $dE_{rel.}$< 4.0 kcal/mol in conjunction with a cap on the number of structures (N_{max}) to be 100 for the next step. Then, we perform r^{2}SCAN-D3 SPE on the PBE-D3$_{light}$ structures, which reduce the R_{max} to 53. So we select structures, having r^{2}SCAN-D3 $dE_{rel.}$< 4.0 kcal/mol in conjunction with N_{max} to be 70 for the next step. In the next step, referred to as PBE-D3$_{tight1}$, the selected crystal structures are optimized further with stricter convergence criteria, *i.e.* $dE_{opt.}$ < 0.06 kcal/mol and F_{max} < 0.13 eV/Å. After this step, we find the maximum value of $dE_{rel.}$ for the experimentally known polymorphs is reduced to 2.7 kcal/mol and similarly R_{max} is reduced to 44. Based on this data we again screen a smaller subset of the structures, having $dE_{rel.}$< 3.0 kcal/mol in conjunction with N_{max} to be 50. In the next step, referred to as PBE-D3$_{tight2}$, the selected crystal structures are optimized further with stricter convergence criteria, *i.e.* $dE_{opt.}$ < 0.026 kcal/mol and F_{max} < 0.03 eV/Å. Next, we perform another screening with r^{2}SCAN-D3 SPE on the PBE-D3$_{tight2}$ structures. In the final step, we take the lowest 25 structures within 3 kcal/mol to perform very tight optimization with stricter convergence criteria, *i.e.* $dE_{opt.}$ < 0.002 kcal/mol and F_{max} < 0.01 eV/Å, referring to PBE-D3$_{tight3}$. Further restricting the convergence criteria did not meaningfully change the relative stability of crystal structures, and is therefore deemed unnecessary due to significantly increasing computational costs. Finally, the energy of the PBE-D3$_{tight3}$ optimized structures are evaluated using the r^{2}SCAN-D3 functional for the final ranking of crystal structures.

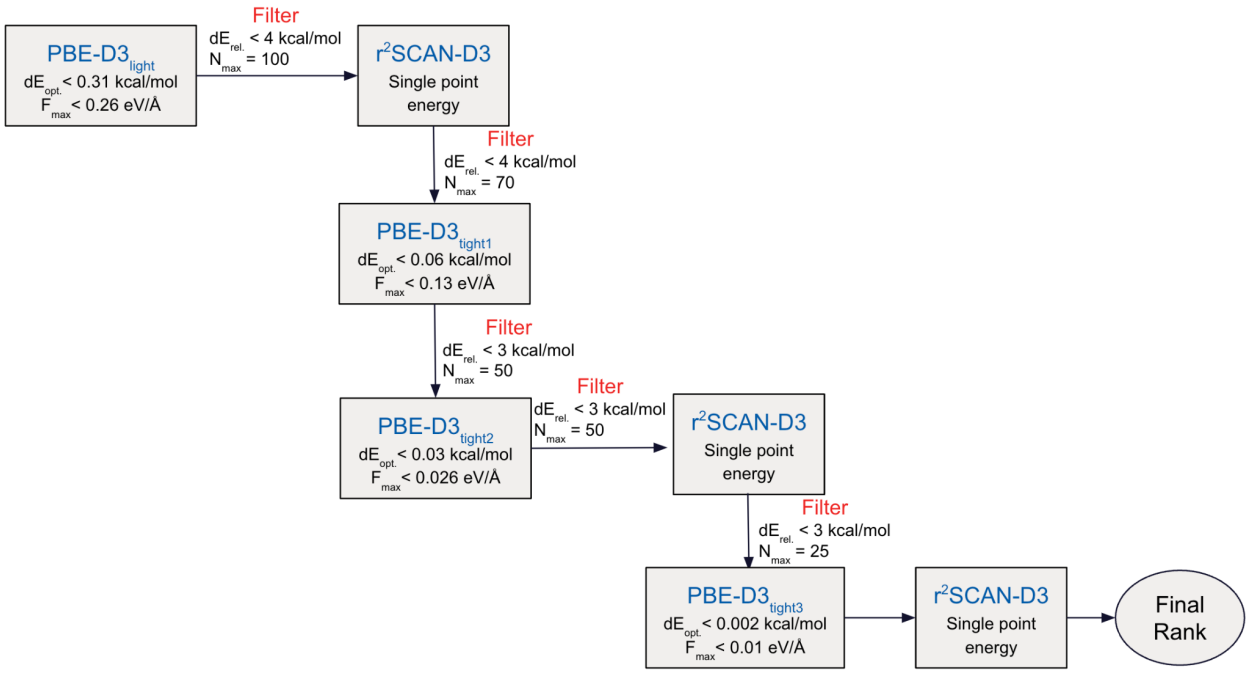

Figure S3: The steps employed in the DFT workflow are shown. See text for details.

Detailed results for molecules discussed in the main text

Olanzapine

Table S3: Comparison of the experimental structures and our CSP generated structures.

		Form I	Form II	Form III*	new form	Form IV
Experimental structure	Space group	$P2_1/c$	$P2_1/c$	Pbca		$P2_1/n$
	Lattice (Å)	a=10.3411 b=14.52 c=10.5314 α=90° β=100.291° γ=90°	a=9.8544 b=16.314 c=9.9754 α=90° β=98.304° γ=90°	a=10.3454 b=19.5267 c=16.528 α=90° β=90° γ=90°	–	a=8.6555 b=15.4441 c=12.5558 α=90° β=95.284° γ=90°
	Density (g/cm^3)	1.334	1.308	1.243	–	1.242
	T(K)	123	123	298	–	443
	CSD	UNOGIN03	UNOGIN04		–	UNOGIN05
CSP generated best match structure	Space group	$P2_1/c$	$P2_1/c$	Pbca	$P2_1/c$	$P2_1/n$
	Lattice (Å)	a=10.389 b=14.463 c=10.441 α=90° β=100.13° γ=90°	a=9.990 b=16.417 c=9.831 α=90° β=97.74° γ=90°	a=10.101 b=19.554 c=16.084 α=90° β=90° γ=90°	a=10.602 b=16.209 c=9.864 α=90° β=109.69° γ=90°	a=8.383 b=15.311 c=14.243 α=90° β=119.60° γ=90°
	Density (g/cm^3)	1.344	1.299	1.306	1.300	1.306
	T(K)	0	0	0	0	0

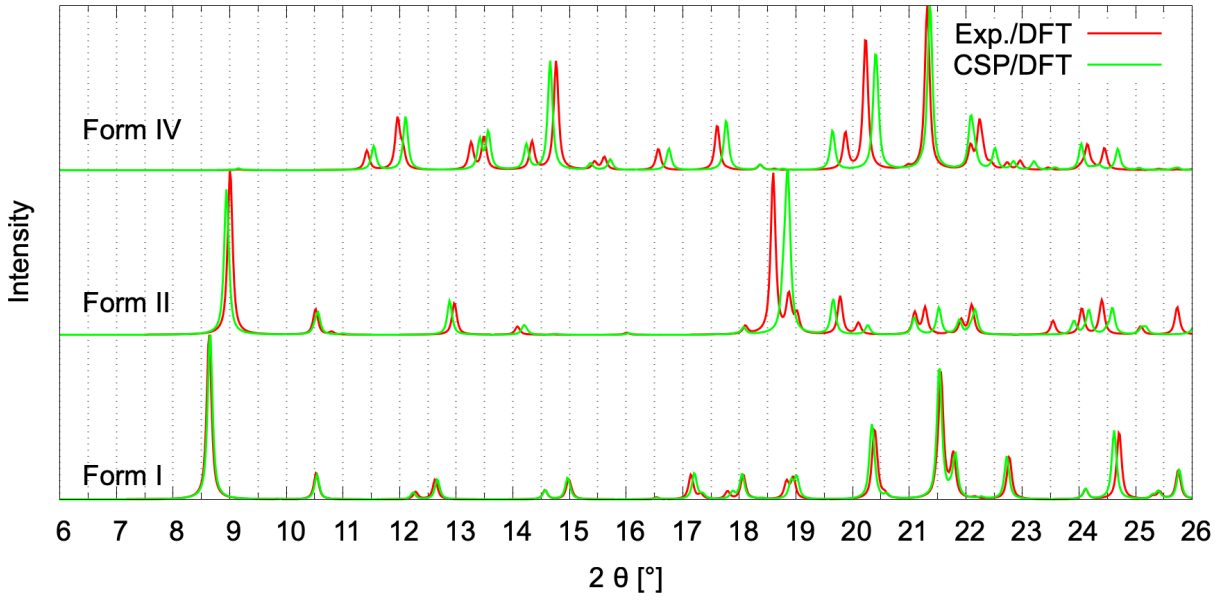

Figure S4: PXRD patterns computed on experimental (CSD structures UNOGIN03, UNOGIN04, UNOGIN05) and CSP generated structures with lowest $RMSD_{>25}$ compared to experimental structures. All structures are relaxed with PBE-D3 at 0K. Source data are provided as a Source Data file.

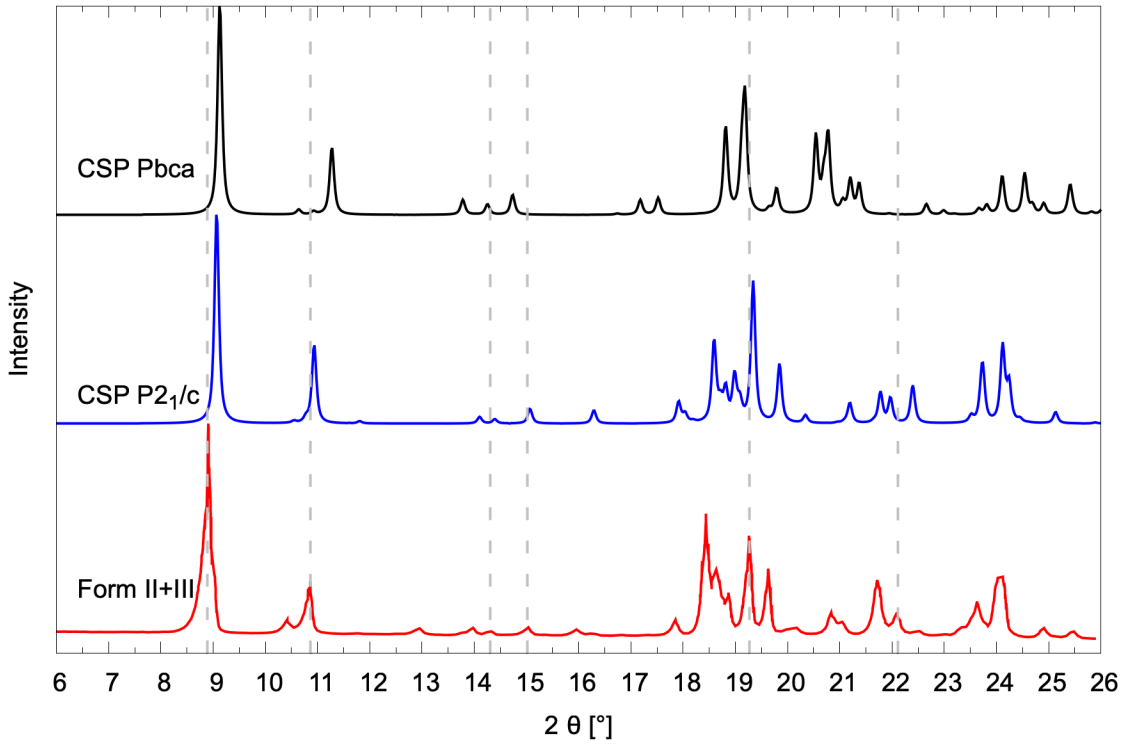

Figure S5: Comparison of experimental PXRD for Olanzipine obtained from a mixture of form II and III (Ref. [36]) with the simulated PXRD pattern of two low-lying structures with $P2_1/c$ and Pbca space group. The vertical dashed lines indicate the positions of the characteristic peaks of form III, as identified in the work of Bhardwaj et. al. [Ref. [36]]. It should be noted that the original paper marked an additionally characteristic peak for form III just after 21°. However, that peak overlaps with a peak appearing in form II. Thus, we contend this peak cannot be considered a characteristic of only form III. Source data are provided as a Source Data file.

ROY

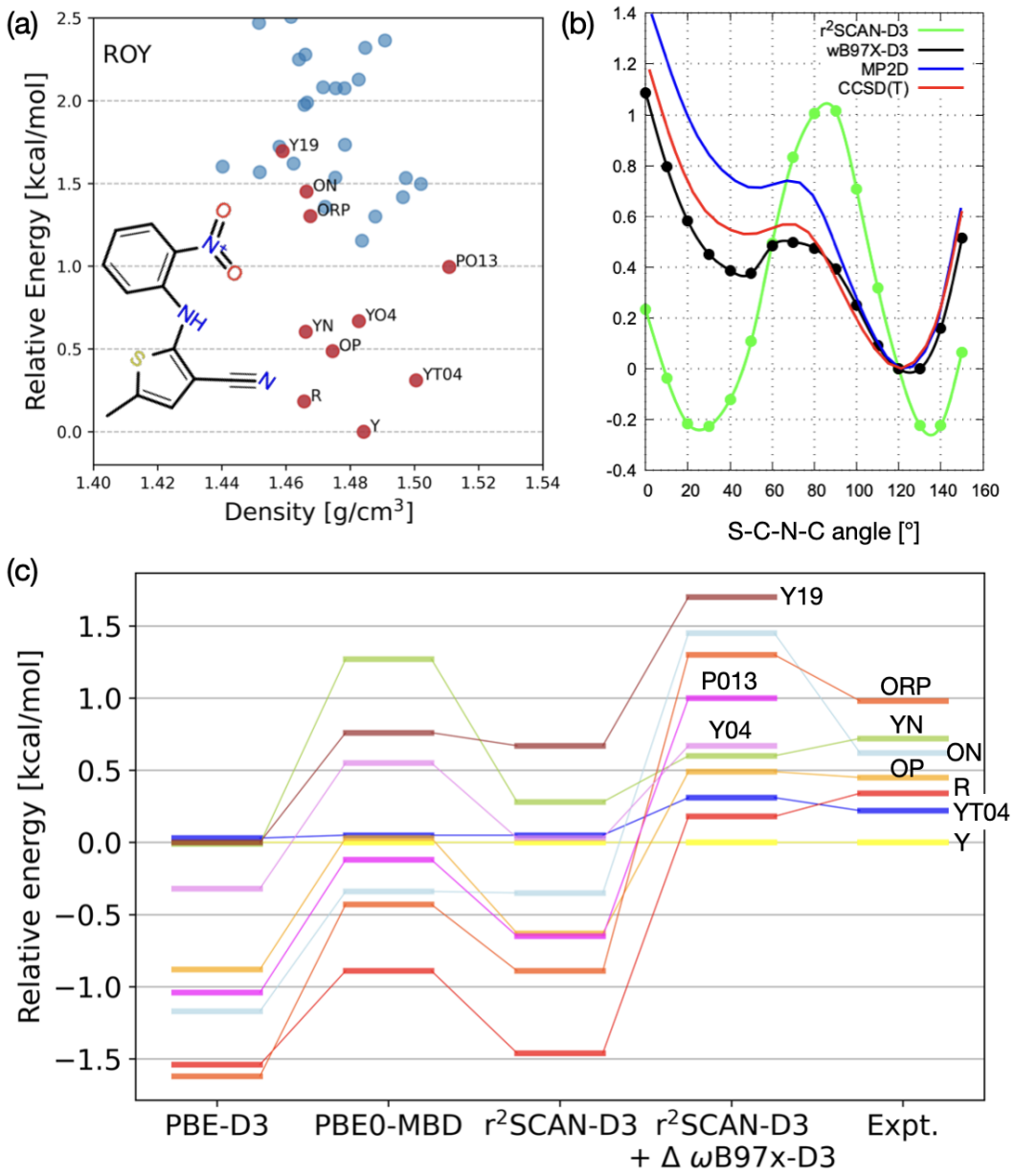

Figure S6: (a) The crystal energy landscape of ROY predicted with monomer corrected r^{2}SCAN-D3 single point energy (r^{2}SCAN-D3 + Δ ωB97X-D3) on PBE-D3 relaxed structures. CSP predicted structures matching the experimentally known polymorphs are highlighted in red. (b) Relative conformer energy of the ROY molecule with S-C-N-C dihedral angle. (c) The relative energy of the lowest RMSD structures with different DFT functionals compared with experiment. Source data are provided as a Source Data file.

ROY, named due to its vividly-colored red, orange, and yellow crystals, is popular for being one of the most polymorphic organic molecules. Twelve fully characterized crystal structures, Y, ON, R, OP, YN, ORP, YT04, Y04, R05, PO13, R18, and Y19 have been discovered.[37] These polymorphs possess conformers with a wide range of S-C-N-C dihedral angles. Experimentally,

the relative free energies of seven ROY polymorphs were measured in the 40–120°C range by eutectic melting experiments, with Y being the most stable polymorph, and all seven polymorphs appear within 1 kcal/mol, providing a reference to computational methods to compare with. A few CSP studies performed earlier typically found all known polymorphs. However, the crystal energy landscapes were found to be densely populated and had high ranking for the ON (115) and Y19 (172) polymorphs. CrystalOptimizer or the B86bPBE-XDM DFT functional employed in CSP of ROY, predicted R or R18 as the most stable polymorph, instead of Y. The S-C-N-C dihedral angle of the conformer of R (~22°) significantly differs from the conformer of Y (~116°), which leads to different degrees of conjugation between the thiophene ring and the 2-nitrophenyl amino group. Extensive studies of Beran et al[37] showed that such DFT functionals are not sufficiently accurate in describing $\boldsymbol{\pi}$-conjugation, leading to inaccurate description of the relative conformational energies. Correcting the conformer energies using a higher level MP2-D method brings the relative stability of the seven polymorphs, to a great extent, closer to the experimental results.

Our CSP study has successfully identified all ten polymorphs with Z'=1 within a 2.5 kcal/mol window. As shown in Fig. S6 the crystal energy landscape is densely populated as determined with the PBE-D3 and r^2SCAN-D3 functionals. The relative stabilities of the seven forms do not agree with experiment due to the inaccuracy of these DFT methods in describing $\boldsymbol{\pi}$-conjugation. Surprisingly, PBE0-MBD, generally believed to be more accurate than PBE-D3 in describing delocalized electrons, also fails to predict the relative energies accurately. By incorporating conformer energy corrections using the ωB97X-D3/def2-TZVP functional, a long-range corrected functional that more accurately describes the conformational energies of this molecule, in conjunction with r^2SCAN-D3 crystal energy (referred to as r^2SCAN-D3 + Δ ωB97X-D3), Y and YT04 were correctly ranked to be the two most stable polymorphs matching experiment. The relative energies of the other six polymorphs predicted with r^2SCAN-D3 + Δ ωB97X-D3 agree reasonably well with experimental data except for the polymorph OP. The stability change over temperature results are in Figure S18(d). The polymorph enthalpies at 0K are shifted to be the r^2SCAN-D3 + Δ ωB97X-D3 energies.

Galunisertib

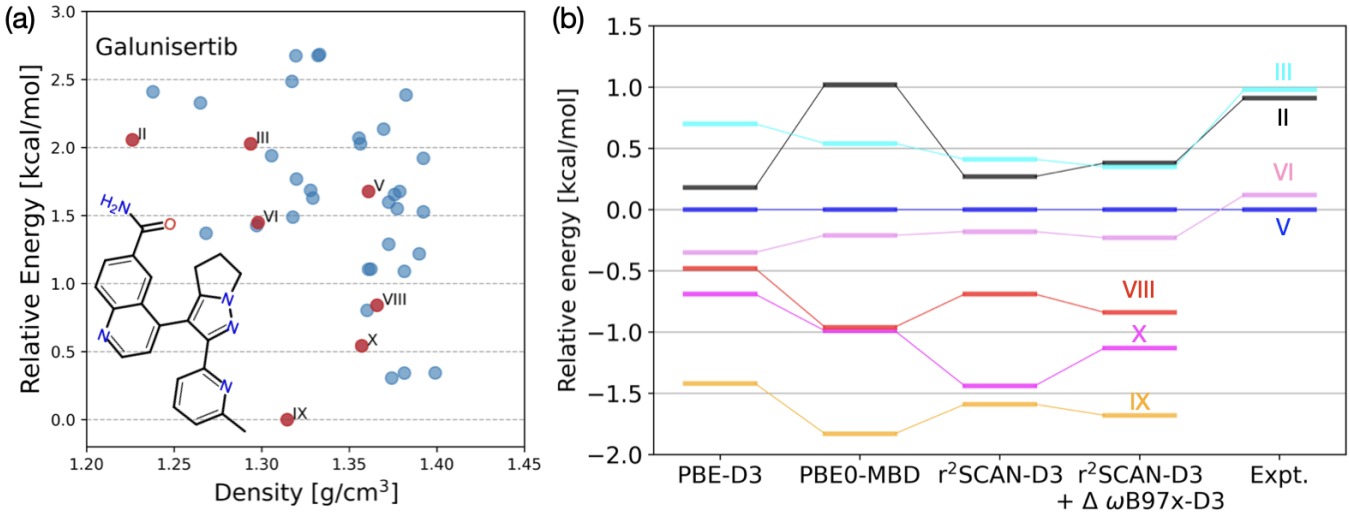

Figure S7: (a) The crystal energy landscape of galunisertib predicted with monomer corrected r^2SCAN-D3 single point energy (r^2SCAN-D3 + Δ ωB97X-D3) on PBE-D3 relaxed structures. CSP predicted structures matching the experimentally known polymorphs are highlighted in red. (b) The relative energy with different DFT functionals compared with experiment. The relative energies are shown with respect to form V, which has the largest experimentally measured melting enthalpy. Source data are provided as a Source Data file.

Galunisertib is an inhibitor of transforming growth factor beta receptor I that was recently investigated for use in the oral treatment of advanced metastatic malignancies. Solid form screening of galunisertib produced many solvates and ten anhydrous forms (form I – form X).[38] The relative stabilities of forms II – VII, determined from the heat of fusion measurements indicate the stability order to be V > VIII, VI, IV (~0.1–0.2 kcal/mol) > II, III (~1 kcal/mol). Form I is metastable by more than 5 kcal/mol and was not included in our analysis. A previous CSP study predicted all ten forms. However, using various dispersion corrected functionals, form IX (Z'=1) or form VII (Z'=2) was predicted to be most stable among the known forms, and a different structure in the P_{212121} space group was predicted to be the most stable structure. However, it was not found in experiments even after exhausting efforts.

From our CSP calculations, we have identified seven experimentally known polymorphs with Z'=1 (Figure S7). The seven Z'=1 polymorphs appear within ~2 kcal/mol of the most stable configuration, which is in agreement with the CSP study of Bhardwaj *et al.*[38] We find form IX to be the most stable structure by r^2SCAN-D3. Our 2nd rank structure is the P_{212121} structure found in the previous CSP[36] ($RMSD_{33}$ of 0.14 Å between ours and earlier reported global minimal structure). Comparing the calculated energy landscapes with the experimental relative energies of Z'=1 forms, we find that forms V and VI are correctly predicted to be more stable than forms II and III. However, the relative energy between forms II and III is overestimated by various DFT functionals. The absolute value of the small energy difference between forms V and VI is correctly calculated though the ranking is opposite due to the small energy difference. The results of the conformer energy correction is similar to previous studies.[39]

Rotigotine

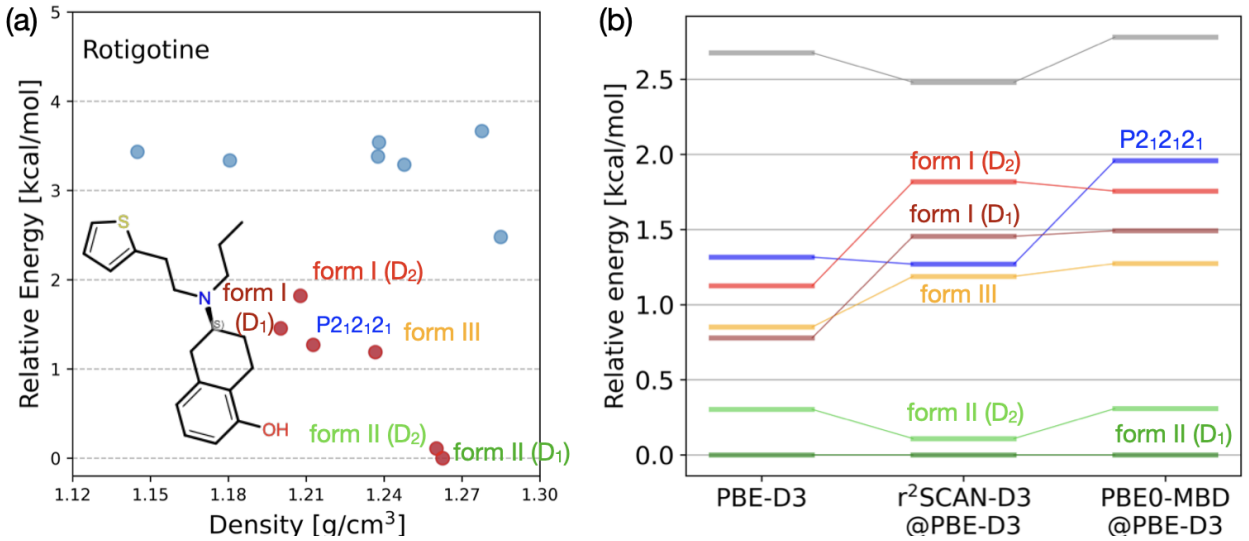

Figure S8: (a) The crystal energy landscape of rotigotine is predicted with r^{2}SCAN-D3 single point energy on PBE-D3 relaxed structures (r^{2}SCAN-D3@PBE-D3). CSP predicted structures matching form I, II, III, and a competitive structure are highlighted. For form I and II two disordered structures are marked D_1 and D_2, respectively. (b) The relative energy of low-energy structures with different DFT functionals. Source data are provided as a Source Data file.

The only long-known crystalline form of rotigotine was form I until a more stable conformational polymorph was discovered in 2008,[40,41] which disrupted the clinical use of the drug. A previous CSP study predicted another candidate structure in the $P3_2$ space group,[42] referred to as form III in our discussion and in figure S8, with stability between experimentally known forms I and II.

From our CSP calculations, 13 candidate crystal structures are within 4 kcal/mol from the predicted global minimum, as shown in Fig. S8(a), including the previously known forms I, II and III. The late appearing crystal, form II ($P2_12_12_1$) is the global minimum, form I is 0.78 kcal/mol higher in energy and form III is 0.85 kcal/mol higher in energy according to PDB-D3. The energy gap between form II and I increased to be about 1.5 kcal/mol (Fig. S8(b)) according to the r^{2}SCAN-D3 and the PBE0-MBD functionals, in agreement with the energy gap previously reported with PBE0-MBD (1.7 kcal/mol) and DSC measurements (1.8 kcal/mol).[42] In addition to these three previously reported structures, we find another candidate crystal in the P_{212121} space group with competitive energy with forms I and III. The additional P_{212121} space group crystal was also predicted in the prior study though the relative energy was slightly higher. The temperature dependent stability calculations (Fig. S18(f)) shows that the form II remains more stable than form I at room temperature, although the free energy gap gets smaller with increasing temperature.

The six lowest energy structures from the previous CSP study (generated using GRACE and relaxed with PBE-MBD)[42] matched with the lowest energy structures from our calculations with the RMSD shown in Fig S9. For forms I and II, the previous study has two predicted structures

for each form corresponding to disorders. The relative energies among these structures are also in good agreement with the previous study, except for a candidate structure in the P_{212121} space group, which is ranked as #4 in our calculations versus #6 from the previous study. Fig. S9 shows a comparison of PBE0-MBD single point energies of the six low-lying structures from our CSP and the six lowest energy structures published using GRACE.[42] Overall the respective structures are in very good agreement since the $RMSD_{>25}$ is within 0.12 Å. The PBE0-MBD energies relative to the most stable form II (D_1) structures are also in agreement between the current PBE-D3 optimized structure and previously reported GRACE structures, which were optimized with PBE-NP.[42]

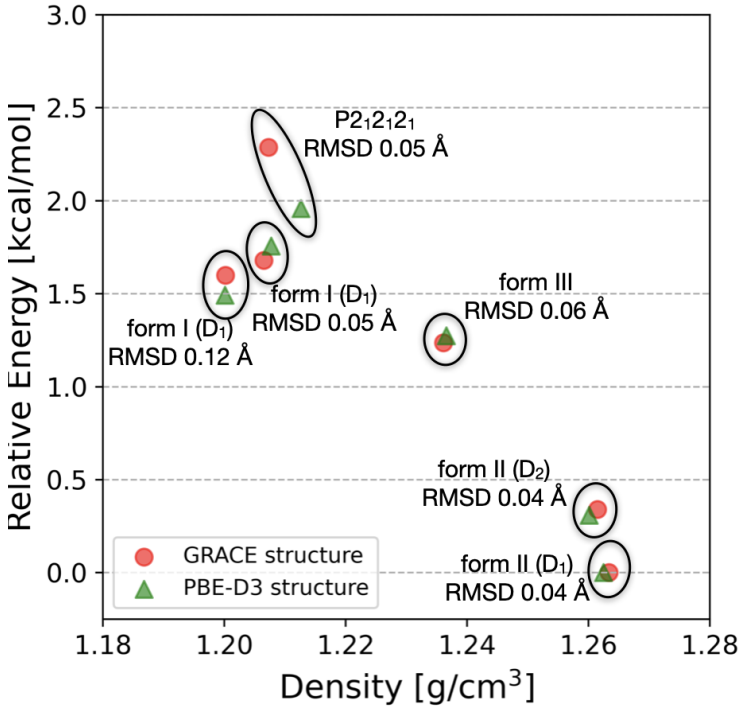

Figure S9: Comparison of the top CSP predicted structures from the current study versus the previous study.[42] The different polymorphs from the current and previously reported structures are encircled and the RMSDs between those are given. Source data are provided as a Source Data file.

MK-8876 & MK-2022

A recent study by Merck & Co, Inc. extensively investigated the crystal structures of two drug molecules, MK-8876 (an HCV site NS5B site D inhibitor)[43] and MK-2022 (a GPR119 agonist).[44] In the case of MK-8776, three (form 1, 2, 3) anhydrous crystals were synthesized in an earlier study, indicating that the form 1 remains stable at or below 85 ℃, form 2 is preferred above 120 ℃, and form 3 is metastable.[43] In the work of Newmant *et al.*[44] the structural parameters of form 1 were determined with various techniques such as single-crystal X-ray diffraction (scXRD), three dimensional electron diffraction (3DED), and CSP, all of which produced consistent results.

However, it was observed that the initial screening protocol used for CSP in the GRACE software did not successfully predict form 1 among the low energy candidate structures.

In contrast, our CSP results predicted form I among the two lowest energy structures with our standard protocol, as shown in the crystal energy landscape (Fig. S11(a)). The candidate structure from our CSP calculations matches well with the published structure for form I ($RMSD_{25}$ of 0.08 Å in reference to CSD structure FESGUJ). As shown in Fig. S11(b), the relative energy of form I with respect to the global minimum, is sensitive to the choice of DFT functionals, ranging from 0.1 kcal/mol (PBE0-MBD) to 1.47 kcal/mol (r^2SCAN-D3). For the other low-energy structures we compare the simulated PXRD pattern with the published PXRD patterns in Ref. [43] Expectedly, the experimental PXRD of form I agrees reasonably well with CSP structure, as shown in Fig. S10. The simulated PXRD of other low energy predicted structures do not match well with experimental spectra of either form 2 or 3 due to the noisy experimental spectra, possibly indicating other competitive polymorphs remain to be discovered. These additional predicted structures retain lower free energies at room temperature (Fig. S18(g)).

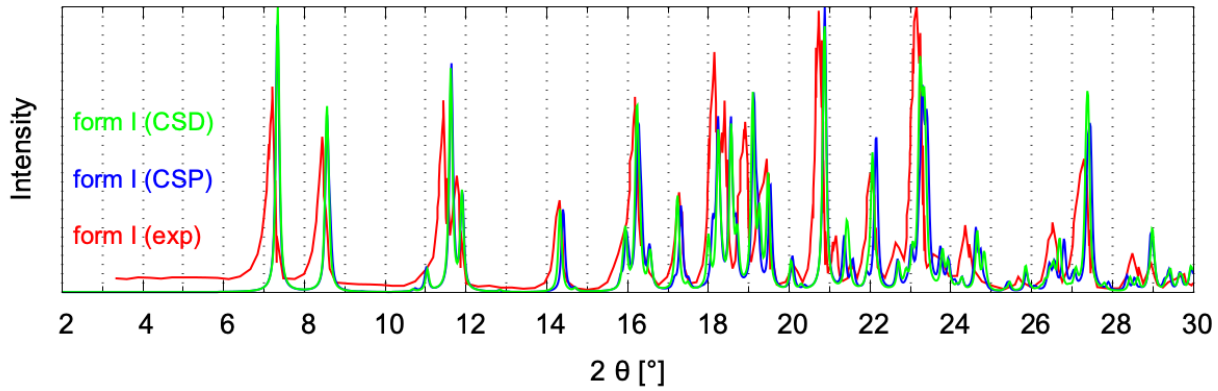

Figure S10: Comparison of experimental PXRD of form I,[43] simulated PXRD on our CSP generated structure, and simulated PXRD of the form I structure from CCDC (CSD structures FESGUJ). Simulated XPRDs are computed on the DFT (PBE-D3) relaxed structures at 0K. Source data are provided as a Source Data file.

In the case of MK-2022, the previous study characterized one crystal with scXRD and 3DED, together with CSP.[44] The best match structure ranked 6th with ~1 kcal/mol high energy with respect to the global minimum. Contrary to this, we find the candidate structure from our CSP calculations matching with the published structure for form I ($(RMSD)_{25}$ of 0.23 Å in reference to CSD structure FESGOD) is most stable (Fig. S12(a)). Form I is found to be most stable with different DFT functionals, as shown in Fig. S12(b), which disagrees with the previous study. The energy gap between form I and the second ranked structure increases from the r^2SCAN-D3 (0.1 kcal/mol) to PBE0-MBD (0.5 kcal/mol) functionals. There are a few structures within 1 kcal/mol of the global minimum, indicating the possibility of other forms yet to be crystallized. The temperature dependent calculation indicates the ranked 2nd and 3rd polymorphs might get closer stability to form I at room temperature (Figure S18(h)).

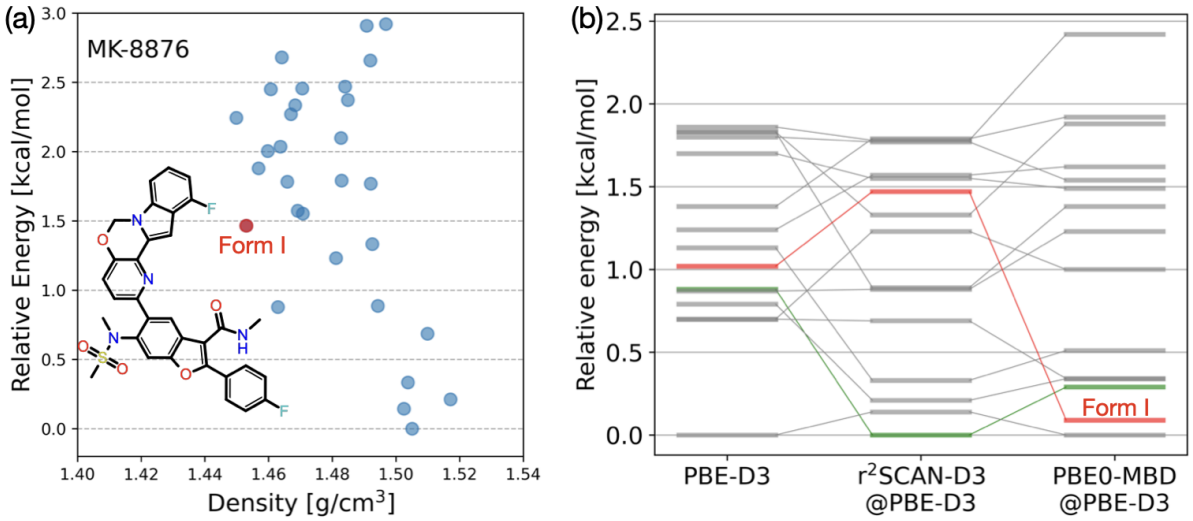

Figure S11: (a) The crystal energy landscape of MK-8876 predicted with r^{2}SCAN-D3 single point energy on PBE-D3 relaxed structures (r^{2}SCAN-D3@PBE-D3). The CSP predicted structure that matches the experimental form I is highlighted in red. (b) The relative energy of the low-energy structures with different DFT functionals. Source data are provided as a Source Data file.

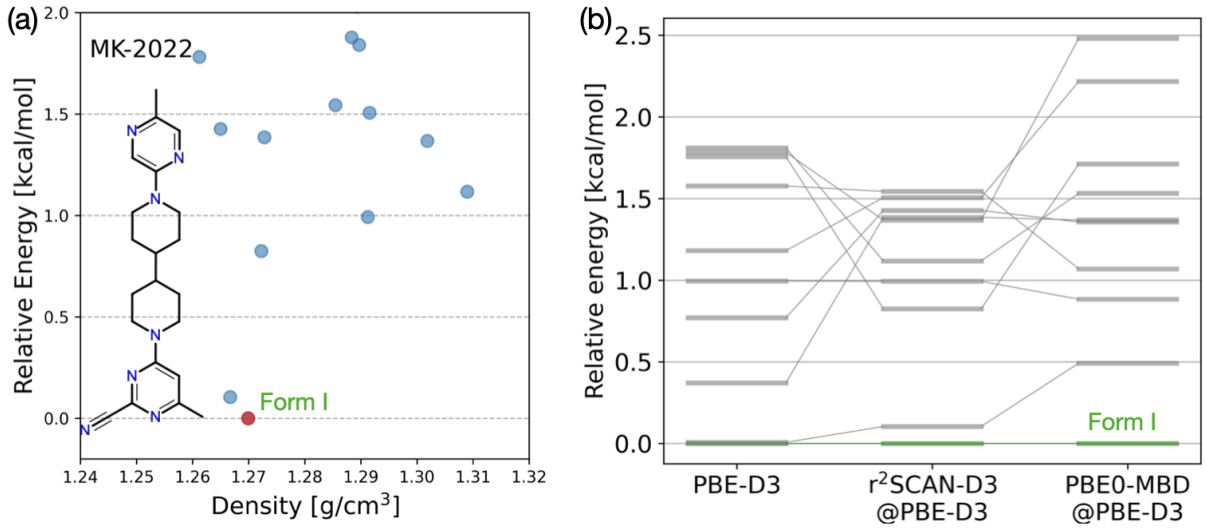

Figure S12: (a) The crystal energy landscape of MK-2022 predicted with r^{2}SCAN-D3 single point energy on PBE-D3 relaxed structures (r^{2}SCAN-D3@PBE-D3). The CSP predicted structure matching experimental form I is highlighted in red. (b) The relative energy of the seven low-energy structures with different DFT functionals. Source data are provided as a Source Data file.

LY156735

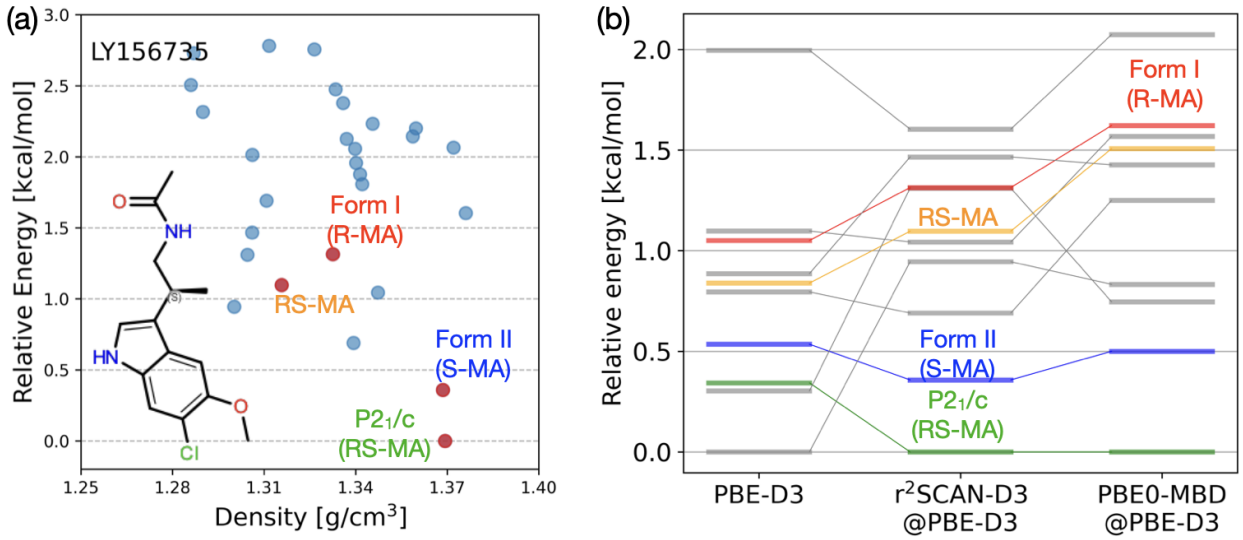

Figure S13: (a) The crystal energy landscape of LY156735 (N-[(2R)-(6-chloro-5-methoxy-1H-indol-3-yl) propyl]acetamide or R-MA) predicted with r^{2}SCAN-D3 single point energy on PBE-D3 relaxed structures (r^{2}SCAN-D3@PBE-D3). The CSP predicted structures matching experimentally known forms, form I (R-MA), form II (S-MA), and RS-MA are highlighted. (b) The relative energy of the 12 low-energy structures with different DFT functionals. Source data are provided as a Source Data file.

LY156735 is a potent and selective melatonin agonist (MA) investigated for the treatment of insomnia and circadian rhythm disorders. The molecule is chiral and the R-enantiomer (R-MA) is biologically active, whereas its mirror image (S-MA) is inactive. The crystallization behavior of LY156735 showed several unusual features.[45] The two enantiomers are expected to produce equivalent crystalline forms of opposite chirality, but showed different crystallization behavior in experiment. In particular, the inactive enantiomer, S-MA, was found to crystallize in at least two anhydrous polymorphic forms (form 1 and 2). Despite extensive efforts, the active enantiomer, R-MA, was found to crystallize only in form 1. Form 1 is metastable with respect to form 2, reflected by a 3-fold difference in solubility corresponding to a free energy difference of 0.7 kcal/mol at room temperature. In addition to these two chiral polymorphs, a racemate structure, form RS-MA, was also characterized in the prior work. A prior CSP study also predicted form 2 to be more stable than form I by 0.93 kcal/mol (at 300 K) and 0.6 kcal/mol (at 0 K) using PBE-D.[46] The relative energy of the racemate structure is intermediate between the form 1 and form 2.

In our study, we have identified the three experimentally characterized structures among the ten lowest energy structures, as shown in Fig. S13(a). Using PBE-D3, the relative energy between form 1 and form 2 is 0.52 kcal/mol at 0 K, in agreement with the previous study. The energy gap between the two increases further with r^{2}SCAN-D3 and PBE0-MBD, to ~1.2 kcal/mol, as shown in Fig. S13(b). Similarly, the relative energy of the experimentally characterized racemate structure (RS-MA) with respect to form 2 is increased with both r^{2}SCAN-D3 and PBE0-MBD

compared to PBE-D3. In addition to those three known structures, our calculations also identified several other low-energy competitive structures, including another RS racemate structure in the $P2_1/c$ space group predicted to be most stable with r^2SCAN-D3 and PBE0-MBD, and some candidate structures matching with the previous CSP study. The relative stabilities of these candidate structures are shown in Fig. S13. It should be noted that we only run CSP on one enantiomer since the opposite enantiomer will have an equivalent polymorph landscape by symmetry. The calculated free energy change along temperature (Figure S18(i)) indicates the stability of the racemate RS-MA will decrease when the temperature rises, compared to form I and form II.

Bicalutamide

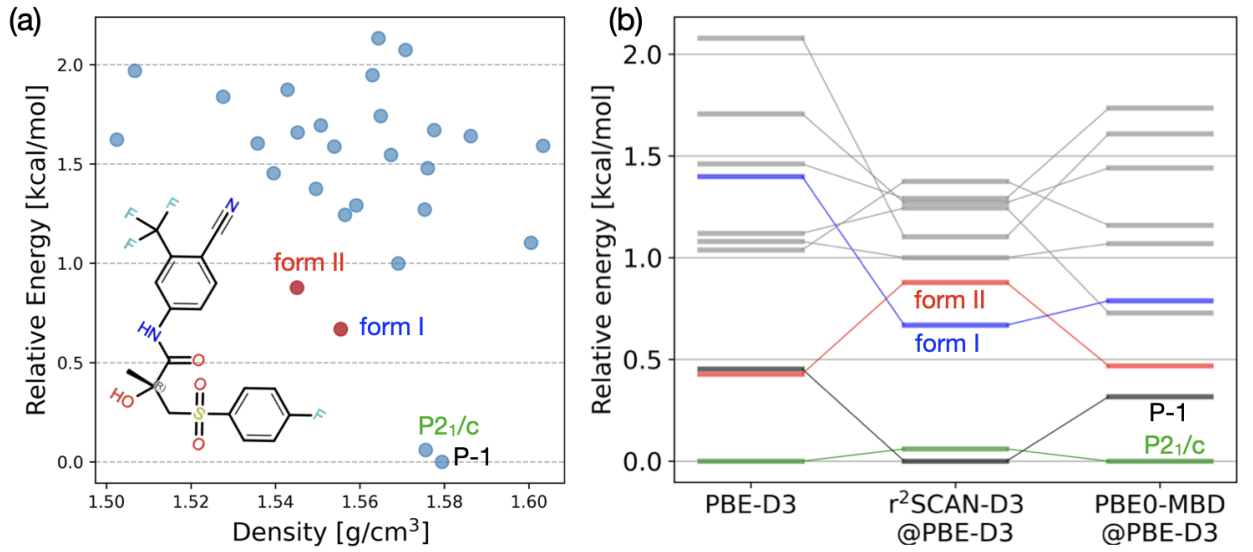

Figure S14: (a) The crystal energy landscape of Bicalutamide predicted with r^2SCAN-D3 single point energy on PBE-D3 relaxed structures (r^2SCAN-D3@PBE-D3). The CSP predicted structures matching experimentally known forms, form I and form II are highlighted. (b) The relative energy of low-energy structures with different DFT functionals. Source data are provided as a Source Data file.

Bicalutamide is a pharmaceutically active compound that possesses antiandrogenic activity and is potentially very useful in treating prostate cancer.[47] Two crystalline forms (forms I and II) and an amorphous phase of bicalutamide were fully characterized through combined results of differential scanning calorimetry, X-ray powder, single crystal diffraction, and Raman spectroscopy.[48] The amorphous phase was observed to be metastable, which converts to form II spontaneously at room temperature in around a week. The two stable polymorphs with Z'=1 have molecular conformers that are strikingly different in shape, one being folded and another more open. The relative stability between the two forms, I and II, evaluated with DSC measurements indicates that form I is more stable.[48]

From our CSP calculations, candidate structures matching with published structures for forms I and II ($RMSD_{30}$ of 0.09 and 0.19 Å in reference to CSD structures JAYCES and JAYCES02) are sampled and ranked among the top 5 of the candidate list (Figure S14(a)). Using PBE-D3, form I was predicted to have higher energy than form II, contrary to DSC measurements.[48] Using r^2SCAN-D3 and PBE0-MBD, the relative energy difference between these two forms decreased to only 0.2–0.3 kcal/mol, although form I still has higher energy than form II from PBE0-MBD method. The computed relative free energies between form I and II increased slightly from 0K to room temperature (Fig. S18(l)). In addition to those known structures, our calculations also identified a few other low-energy competitive structures, including P-1 and $P2_1/c$ structures predicted to be most stable with r^2SCAN-D3 and PBE0-MBD, respectively. The relative stabilities of the low-energy structures are shown in Fig. S14(b). The conformers of the two lowest energy structures assume a folded conformation similar to form II.

XXIII

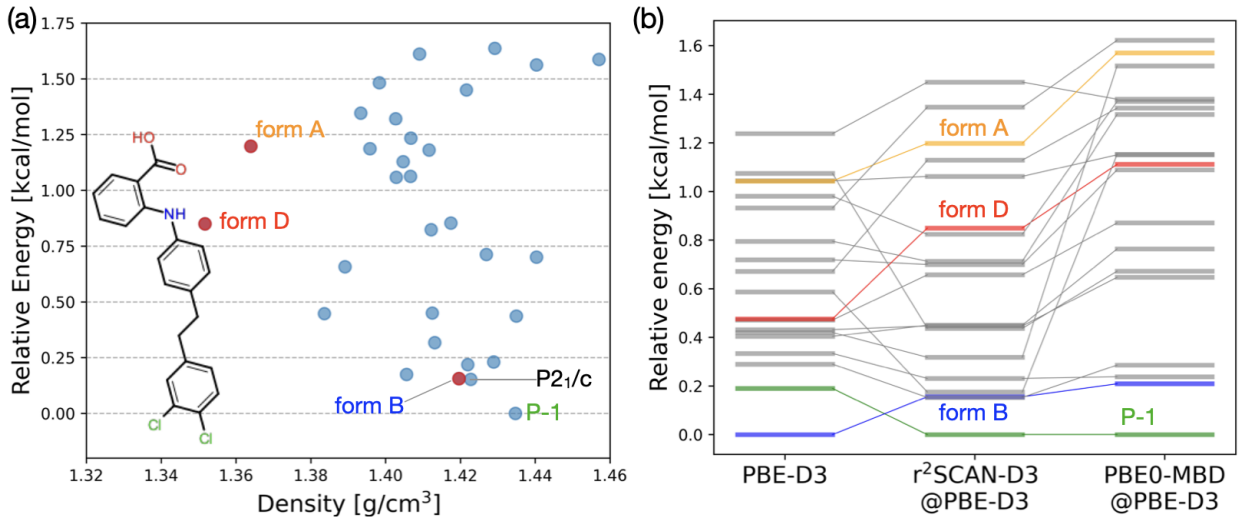

Figure S15: (a) The crystal energy landscape of XXIII predicted with r^2SCAN-D3 single point energy on PBE-D3 relaxed structures (r^2SCAN-D3@PBE-D3). The CSP predicted structures matching experimentally known forms with Z'=1, form A, form B, and form D are highlighted. (b) The relative energy of low-energy structures with different DFT functionals. Source data are provided as a Source Data file.

The sixth CCDC blind test[30] featured molecule XXIII which was submitted by Pfizer Inc. and was a former research compound aimed at treating Alzheimer's disease.[49] Three Z'=1 forms (A, B, D) and two Z'=2 forms (C, E) are experimentally characterized.[30] The thermodynamic stability among the polymorphs strongly varies with small temperature variations. Slurrying experiments (as reported in the sixth blind test) have identified form A as the most stable polymorph at 257 K, while at 293 K form D is found to be the most stable polymorph.[30] A recent study[50] using solution-mediated phase transformation together with slurry experiments reported thermodynamic stability of forms A–D (form E was not found in that study) in the temperature range of 252 K to 316 K.

The thermodynamically most stable form changes over a small range of temperatures in the following way: form C below 253 K, form B in 273 – 288 K, form A in 290 – 294 K, and form D at 295 K and higher.[50]

Many CSP approaches in the sixth blind test predicted form B to be the most stable among the three Z'=1 polymorphs.[30] Later study of ranking with PBE0+MBD functional including thermal corrections indicated form C (Z'=2) as the most stable structure among the experimental polymorphs.[31] The same study also reported a new CSP-generated structure (namely N70) to be the most stable among their CSP pool and experimental structures.[31] Another recent CSP study reported form C as the most stable among the experimental polymorphs with thermal corrections in conjunction with B3LYP-D energy.[51] Without thermal correction, form B is found to be the most stable among the experimental polymorphs, together with a few other structures that are lower in energy than all experimental polymorphs.[51]

From our CSP calculations, candidate structures that match with published structures for the three Z'=1 forms A, B, D ($RMSD_{30}$ of 0.39, 0.23, and 0.18 Å in reference to CSD structures XAFPAY, XAFPAY01, XAFPAY03, respectively) are sampled and ranked among the top 30 of the candidate list with r^2SCAN-D3 (Figure S15(a)). The dependence on ranking with various DFT functionals is shown in Fig. S15(b). Similar to previous findings PBE-D3 favors form B as the most stable structure. With r^2SCAN-D3 and PBE0-MBD, a new structure with a P-1 space group is found to be the most stable. The same P-1 structure was also predicted in the earlier study with PBE0-MBD.[31] Our temperature-dependent stability calculations correctly predicted form D to be most stable above room temperature but did not predict form A to be more stable than form D at any point across the temperature range, although they remain very close in energy (Fig. S18(j)).

Axitinib

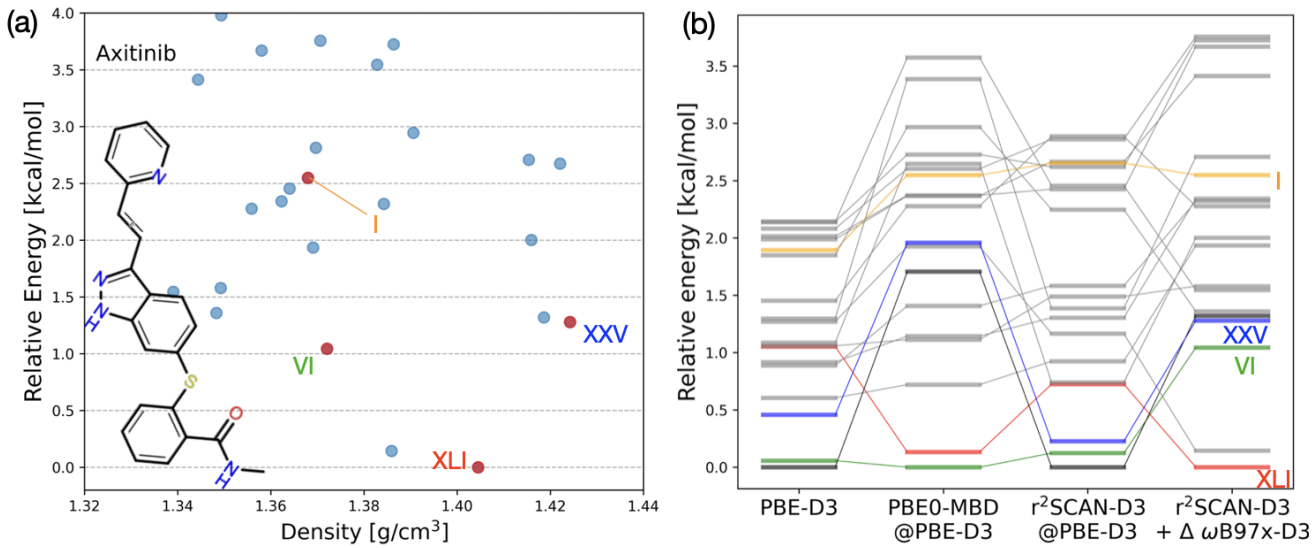

Figure S16: (a) The crystal energy landscape of Axitinib predicted with monomer corrected r^{2}SCAN-D3 single point energy (r^{2}SCAN-D3 + Δ ωB97X-D3) on PBE-D3 relaxed structures. The CSP predicted structures matching experimentally known forms, XLI, VI, XXV, and I are highlighted. (b) The relative energy of low-energy structures with different DFT functionals. The r^{2}SCAN-D3 + Δ ωB97X-D3 relative stability is similar to experimental stability i.e. XLI > VI, XXV > I.[49] Source data are provided as a Source Data file.

Pfizer's anticancer drug axitinib has 5 known neat polymorphs and 66 solvates to date.[52,53] Forms I, IV, VI, and many solvates were discovered via standard experimental solid form screening procedures. Form IV was believed to be the thermodynamically stable form and was initially targeted for development.[43] However, more stable forms XXV and XLI were fortuitously discovered later during the manufacturing campaign, and the thermodynamically stable form XLI became the commercial form.[52,53] The difficulty in crystallizing Form XLI has been attributed to its distinct intramolecular conformation and unique 1-D chains of hydrogen bonds instead of the hydrogen-bonded dimers found in the other polymorphs. Differential scanning calorimetry and solubility experiments indicate the stability in this order: XLI > XXV, VI > IV > I, where the energy ordering of XXV and VI is uncertain.[52,53]

Previous CSP studies based largely on force fields generated the experimental crystal structures but ranked them poorly.[54] For example, form XLI lies 2.4 kcal/mol above form VI. Previous DFT ranking with B86bPBE-XDM predicts most of the polymorph stability orderings reasonably, but it incorrectly suggests that form XLI is 0.5–0.7 kcal/mol less stable than forms XXV and VI. It was also shown that the conformer energy differences are poorly described with the B86bPBE-XDM functional and the use of hybrid functional or MP2 level monomer correction stabilizes form XLI.[39]

In our CSP, we find all four Z'=1 experimentally known structures ($RMSD_{29}$ within 0.13–0.29 Å in reference to CSD structures VUSDIX, VUSDIX03, VUSDIX04, VUSDIX06) which are ranked

among the top 20 candidates with r^{2}SCAN-D3 (Fig. S16(a)). Among these 4 forms, form VI is found to be the most stable. With r^{2}SCAN-D3, form XLI is ~0.6 kcal/mol less stable than form VI, contrary to experimental stability. Moreover, r^{2}SCAN-D3 predicts a C2/c structure to be the most stable. However, the C2/c structure is ~1.5 kcal/mol less stable than form VI with PBE0-MBD, which also increases the relative stability of form XLI to within 0.2 kcal/mol of form VI. This is indicative of conformational energies that are not sufficiently accurate and monomer-based corrections may improve the relative stability. An earlier study was able to reproduce the experimental stability ordering using MP2D-based monomer correction on B86PBE-XDM.[39] By incorporating conformer energy corrections using the ωB97X-D3/def2-TZVP functional in conjunction with r^{2}SCAN-D3 crystal energy (referred to as r^{2}SCAN-D3 + Δ ωB97X-D3), the relative stabilities of these four forms are correctly predicted matching experiment. The temperature dependent free energy calculation indicates the stability of forms XLI and VI gets closer with the increase of the temperatures, and form XLI remains most stable at temperatures below 200K (Fig. S18(m)).

AZD8329

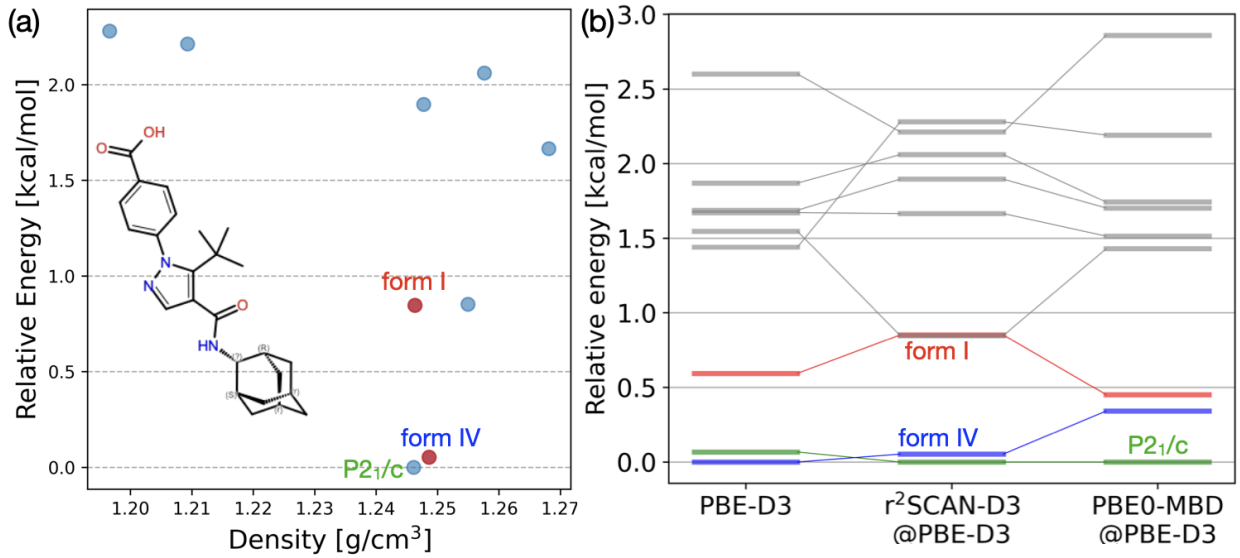

Figure S17: (a) The crystal energy landscape of AZD8329 predicted with r^{2}SCAN-D3 single point energy on PBE-D3 relaxed structures (r^{2}SCAN-D3@PBE-D3). The CSP predicted structures matching experimentally known forms, form I and form IV are highlighted. (b) The relative energy of low-energy structures with different DFT functionals. Source data are provided as a Source Data file.

AZD8329 is a pharmaceutical compound with the potential for the treatment of Type 2 diabetes.[55] At least seven anhydrous/solvate forms are known for AZD8329. Of the major anhydrous forms I–IV, the structures of forms I and IV are of particular interest since they had been chosen for development due to their suitable material properties. Form I was determined by single-crystal diffraction, whereas the crystal structure of form IV was determined by a powder NMR

crystallography protocol using crystal structure prediction and DFT chemical shift calculations in combination with measured 1H chemical shifts.[56] The forms I and IV are enantiotropically related, with form IV found to be the more stable at ambient conditions and form I being the stable high-temperature crystal form.[56]

From our CSP calculations, candidate structures that match well with published structures for the two forms I and IV ($RMSD_{27}$ of 0.23 Å and 0.64 Å, in reference to structures published in Ref. [56] are sampled and ranked among the top 5 of the candidate list with r^2SCAN-D3 (Figure S17(a)). Form IV is found to be more stable than form I at 0K with DFT, which is in agreement with the reported stability ordering at ambient conditions.[56] In form IV, the carboxylic acid group forms a double hydrogen bond with the amide group of a neighboring molecule with double hydrogen bonds. In form I, the intermolecular double hydrogen bond forms between the terminal carboxylic acid groups. The most stable CSP-generated structure is found to be in the P21/c space group, which has a hydrogen bond pattern similar to form IV. The entropy calculation along the temperature indicates the form I gets more stable when the temperature increases, which is consistent with the experimental observation (Fig. S18(k)).

Pseudo super critical path and replica exchange molecular dynamics (PSCP+REMD) calculated temperature dependent stability

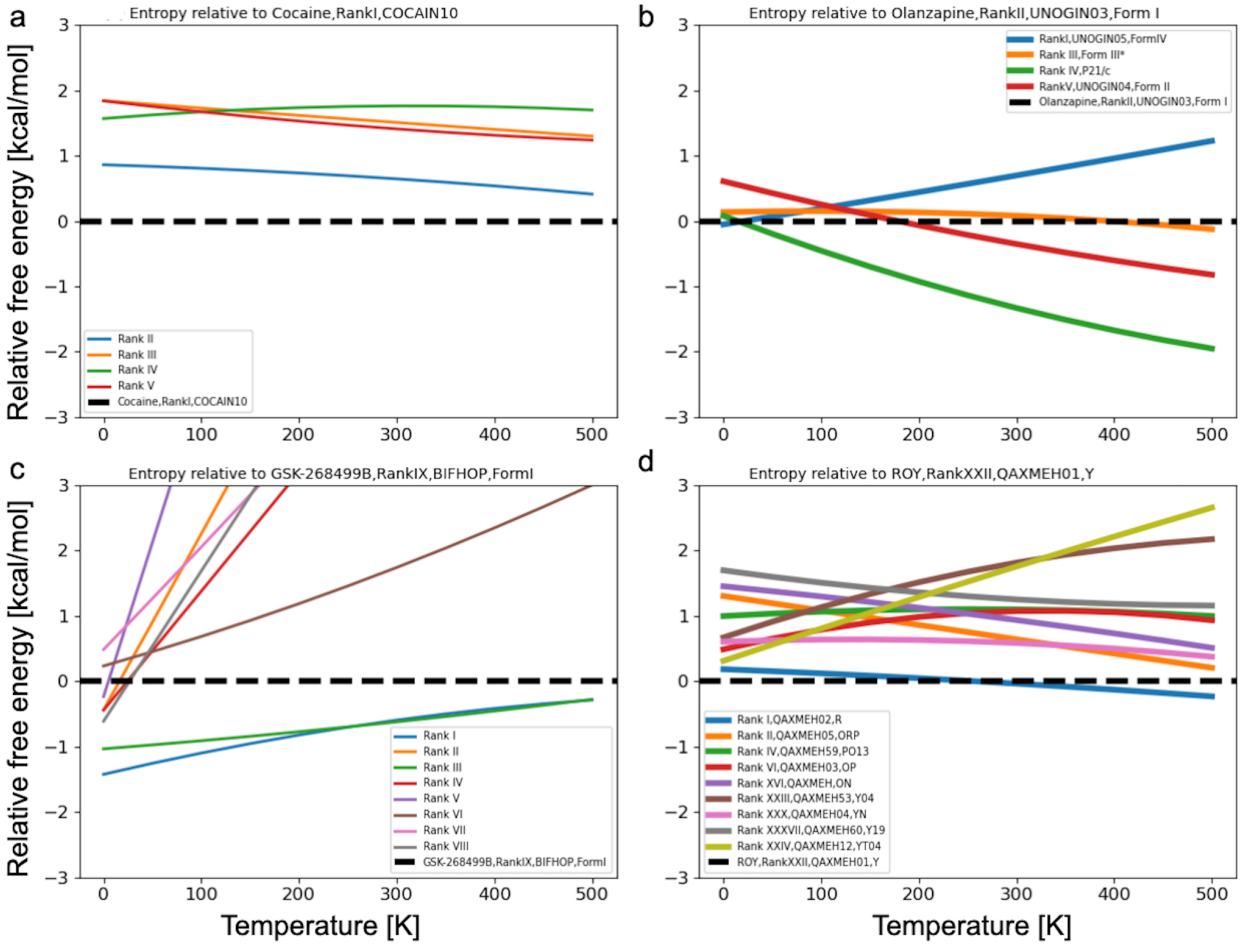

Figure S18. Temperature dependent relative free energy plots of the Z'=1 crystal structure prediction (CSP) landscapes of (a) Cocaine, (b) Olanzapine, (c) GSK269984B, and (d) ROY. The free energy values are relative to a predicted crystal structure that matches an experimental form. The free energy of the reference polymorph is set to 0 and plotted as a dashed line. The solid lines are the calculated relative free energies between the corresponding polymorphs and the reference polymorph. The ranks in the legends are from the r^2SCAN-D3 ranking numbers, to be consistent with the structures on github. Experimental forms are highlighted with thicker lines, and the other CSP predicted forms are in thinner solid lines. The 0K relative energies are from the PBE0-MBD functionals, except for ROY, where the monomer corrected energies via r^2SCAN-D3 + Δ ωB97X-D3 are used. Source data are provided as a Source Data file.

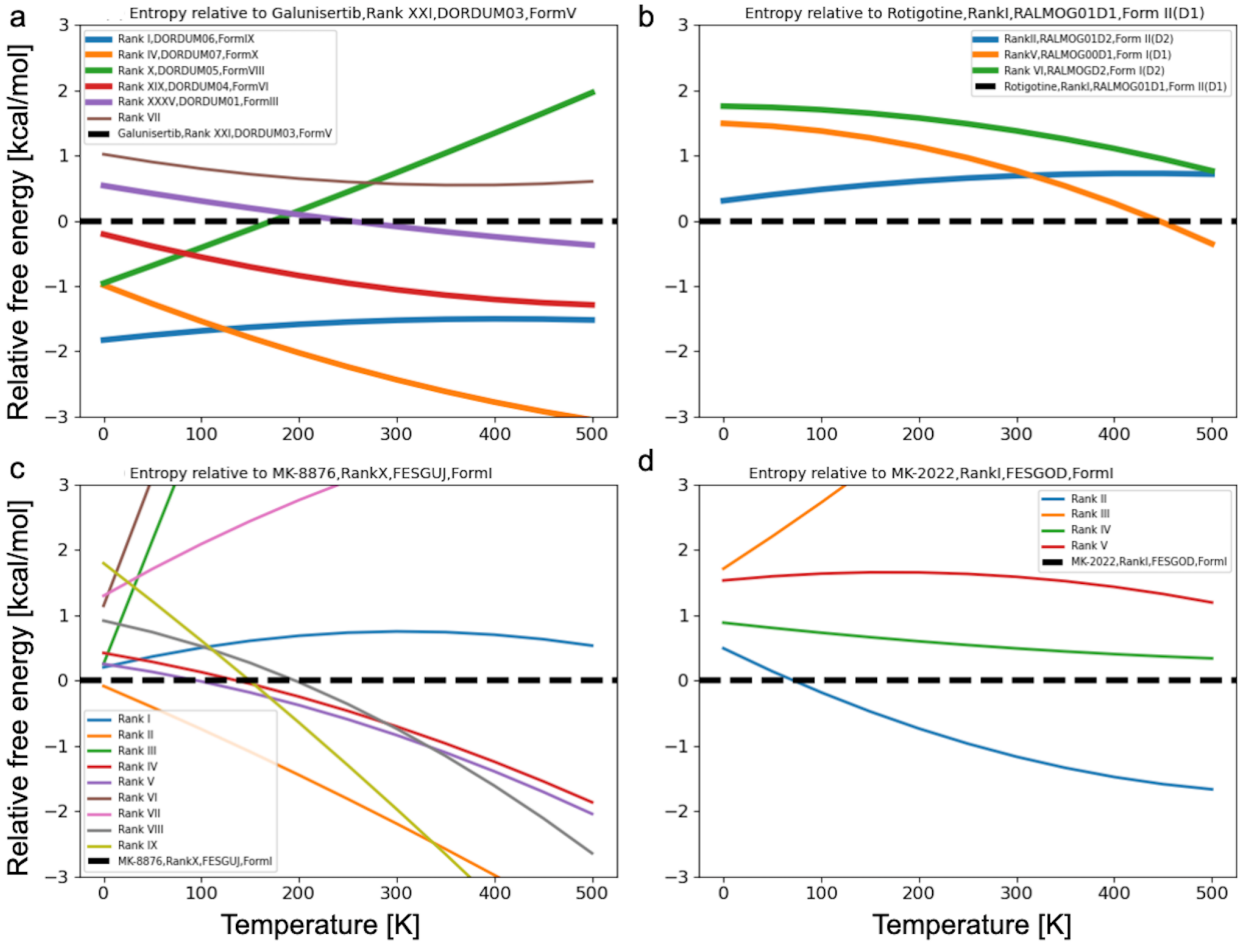

Figure S19. Temperature dependent relative free energy plots of the Z'=1 crystal structure prediction (CSP) landscapes of (a) Galunisertib, (b) Rotigotine, (c) MK-8876, and (d) MK-2022. The free energy values are relative to a predicted crystal structure that matches an experimental form. The free energy of the reference polymorph is set to 0 and plotted as a dashed line. The solid lines are the calculated relative free energies between the corresponding polymorphs and the reference polymorph. The ranks in the legends are from the r^2SCAN-D3 ranking numbers, to be consistent with the structures on github. Experimental forms are highlighted with thicker lines, and the other CSP predicted forms are in thinner solid lines. The 0K relative energies are from the PBE0-MBD functionals, except for Galunisertib, where the monomer corrected energies via r^2SCAN-D3 + Δ ωB97X-D3 are used. Source data are provided as a Source Data file.

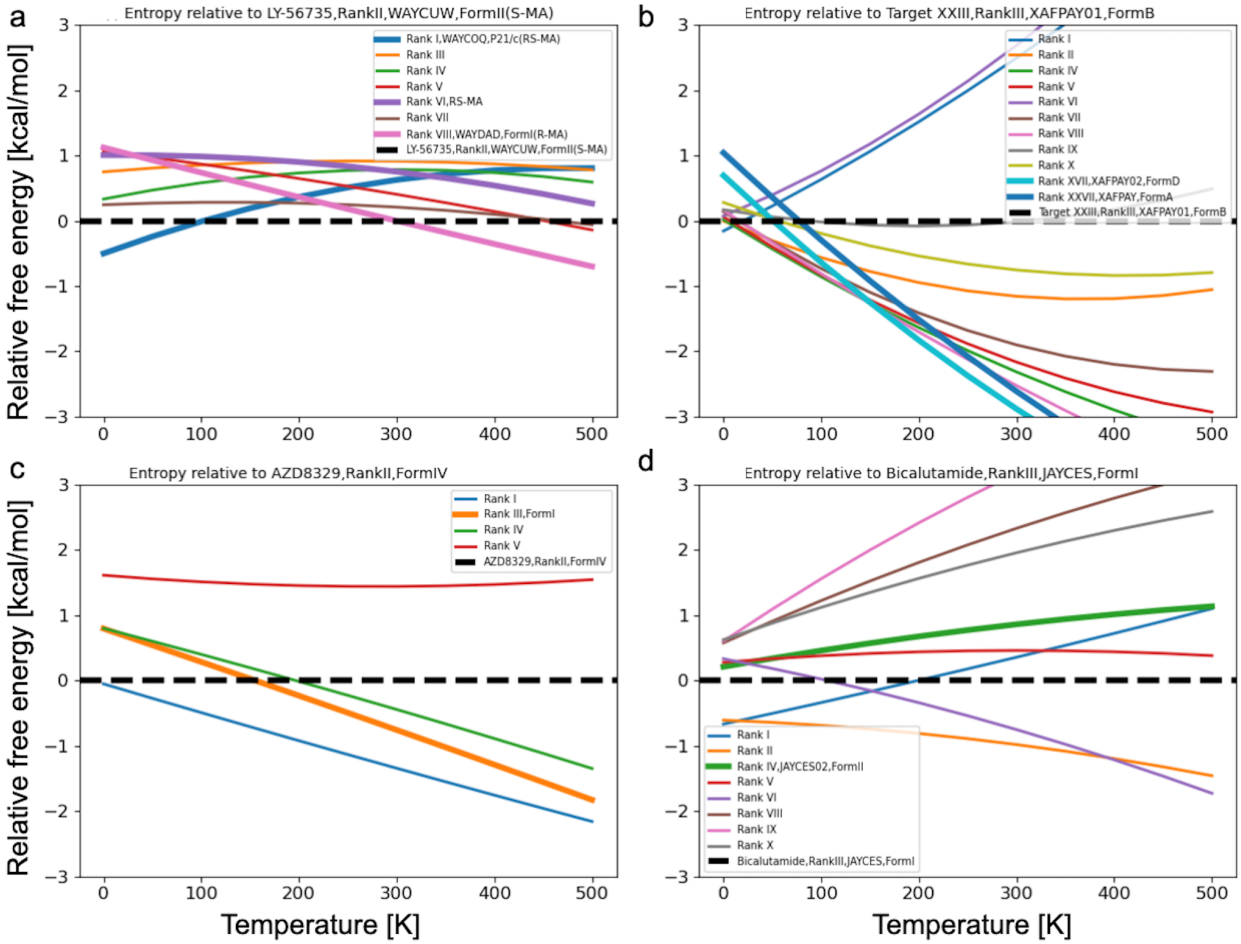

Figure S20. Temperature dependent relative free energy plots of the Z'=1 crystal structure prediction (CSP) landscapes of (a) LY-56735, (b) Target XXIII, (c) AZD8329, and (d) Bicalutamide. The free energy values are relative to a predicted crystal structure that matches an experimental form. The free energy of the reference polymorph is set to 0 and plotted as a dashed line. The solid lines are the calculated relative free energies between the corresponding polymorphs and the reference polymorph. The ranks in the legends are from the r^2SCAN-D3 ranking numbers, to be consistent with the structures on github. Experimental forms are highlighted with thicker lines, and the other CSP predicted forms are in thinner solid lines. The 0K relative energies are from the PBE0-MBD functionals. Source data are provided as a Source Data file.

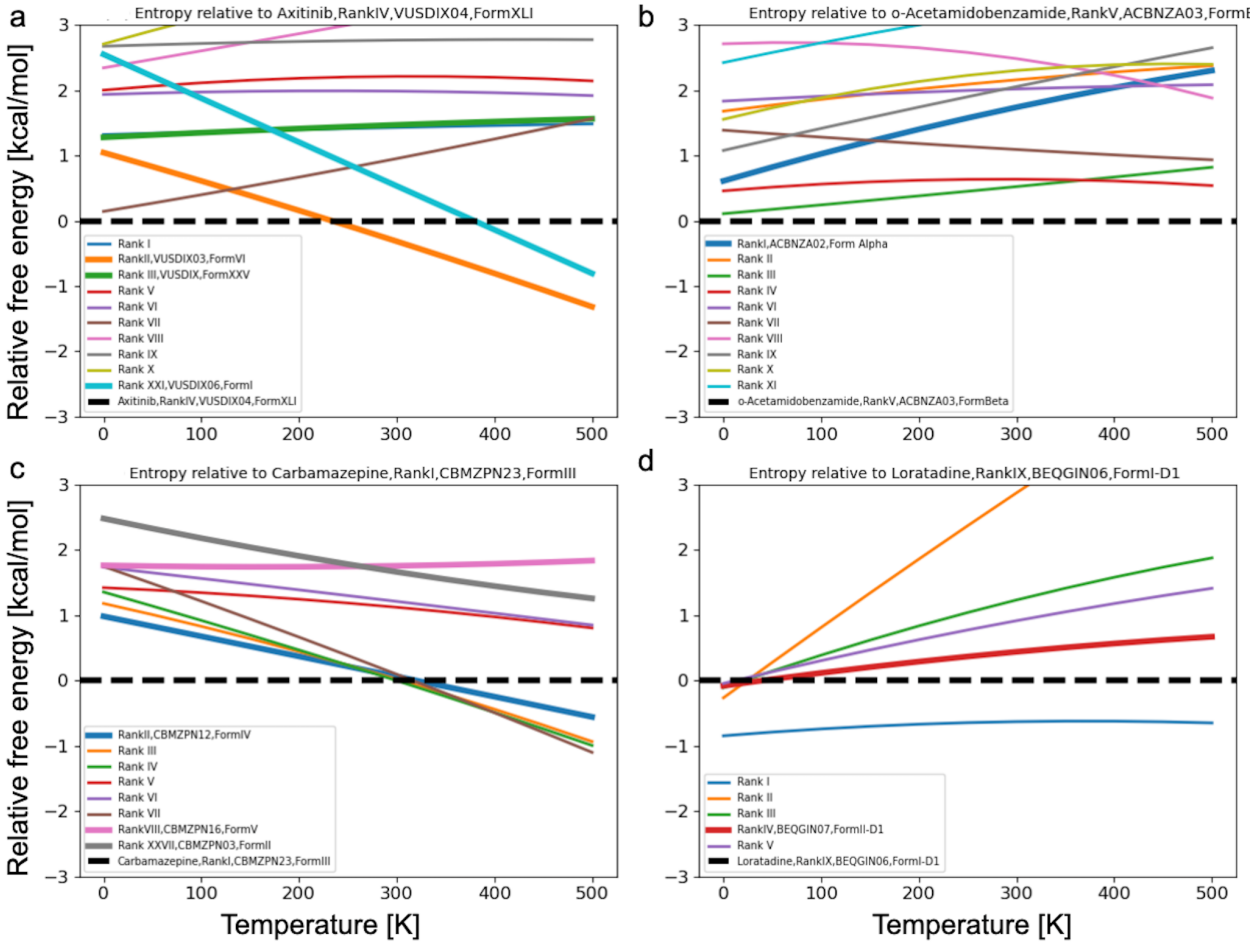

Figure S21. Temperature dependent relative free energy plots of the Z'=1 crystal structure prediction (CSP) landscapes of (a) Axitinib, (b) o-Acetamidobenzamide, (c) Carbamazepine, and (d) Loratadine. The free energy values are relative to a predicted crystal structure that matches an experimental form. The free energy of the reference polymorph is set to 0 and plotted as a dashed line. The solid lines are the calculated relative free energies between the corresponding polymorphs and the reference polymorph. The ranks in the legends are from the r^2SCAN-D3 ranking numbers, to be consistent with the structures on github. Experimental forms are highlighted with thicker lines, and the other CSP predicted forms are in thinner solid lines. The 0K relative energies are from the PBE0-MBD functionals, except for Axitinib and o-Acetamidobenzamide, where the monomer corrected energies via r^2SCAN-D3 + Δ ωB97X-D3 are used. Source data are provided as a Source Data file.

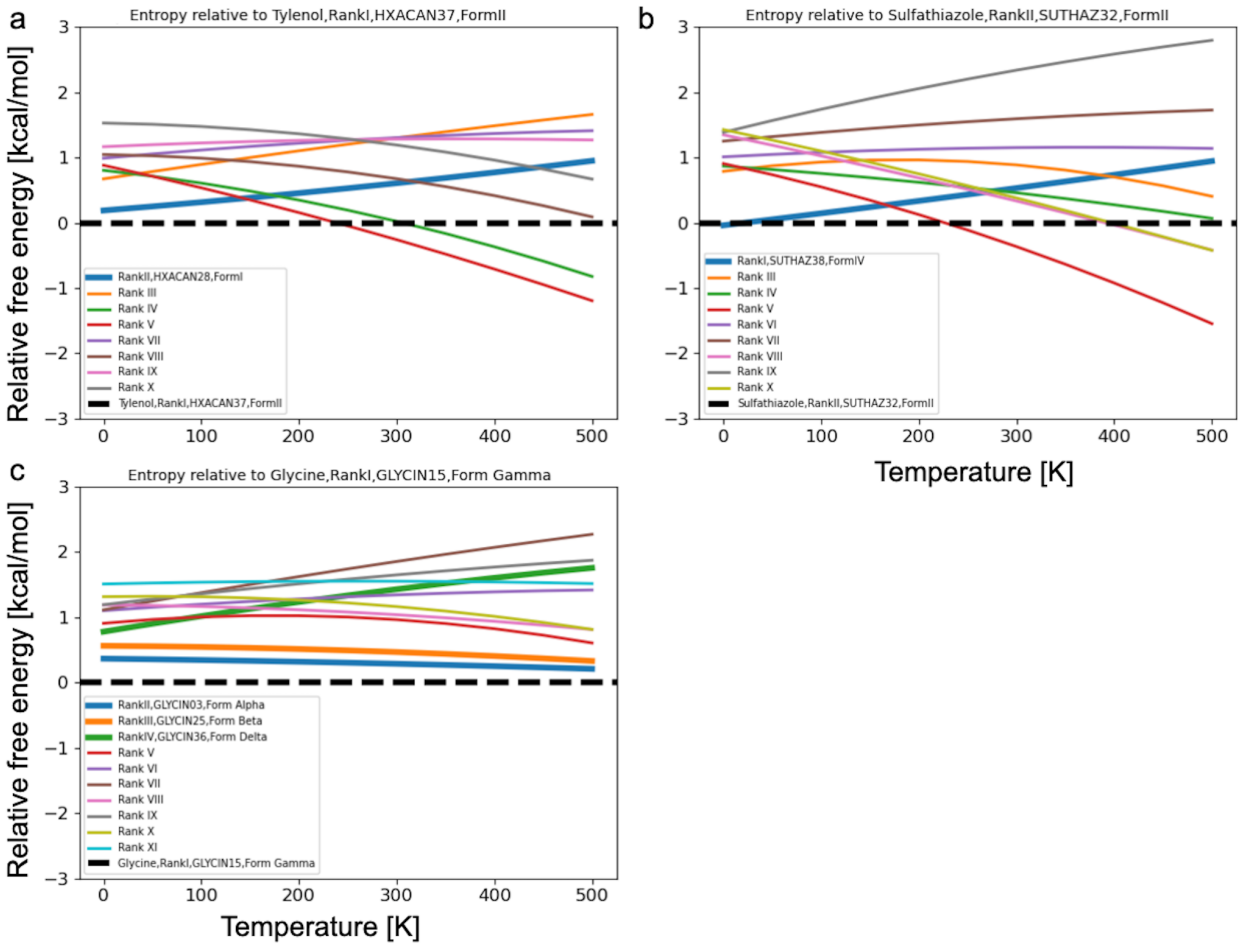

Figure S22. Temperature dependent relative free energy plots of the Z'=1 crystal structure prediction (CSP) landscapes of (a) Tylenol, (b) Sulfathiazole, and (c) Glycine. The free energy values are relative to a predicted crystal structure that matches an experimental form. The free energy of the reference polymorph is set to 0 and plotted as a dashed line. The solid lines are the calculated relative free energies between the corresponding polymorphs and the reference polymorph. The ranks in the legends are from the r^2SCAN-D3 ranking numbers, to be consistent with the structures on github. Experimental forms are highlighted with thicker lines, and the other CSP predicted forms are in thinner solid lines. The 0K relative energies are from the PBE0-MBD functionals. Source data are provided as a Source Data file.

Data Availability

The top predicted candidate structures for all molecules in the validation set are available on the following github repository.

https://github.com/bsantra/csp-validation-data

Input: list of conformers, list of space groups to be sampled
Output: A list of candidate crystal structures

For each conformer in the conformer pool, its crystal packing search for any given space group symmetry sequence is a breadth first tree search, as summarized in the following Python pseudo-code.

```
all_spacegroup_search_sequences = []
for spg in all_input_spgs:
    # search sequences are determined by the symmetry operations of each space group
    # For example, for the P1 space group, the search sequence is [P1_Ta, P1_Tb, P1_Tc] which
    # corresponds to the three translational symmetries
   all_spacegroup_search_sequences.extend(get_search_sequences_for_spg(spg))

sampled_crystals = []
for sequence in all_spacegroup_search_sequences:
   structures = all_input_conformers
   for step in sequence.steps:
      clusterResults = []
      for structure in structures:
         for x0 in step.generateInitialGuesses(structure):
            # step.generateCluster() takes an input structure and a
            # parameter vector and returns a cluster with the associated symmetry.
            # minimize() is a force-field based minimizer with respect to the input parameters.
            clusterResults.append(minimize(step.generateCluster, structure, x0))
      structures = []
      for structure, x1 in clusterResults:  # x1 is optimized parameters
         structure1 = step.generateCluster(structure, x1)
         structures.append(structure1)  #output structure becomes input for the next step
   sampled_crystals.extend([step.convertToCrystal(structure) for structure in structures])

# An example generateCluster method for P1_Ta step of the [P1_Ta, P1_Tb, P1_Tc] sampling sequence for P1 space group, P1_Ta.generateCluster(), is as follows.

class P1_Ta:
   @classmethod
   def generateCluster(cls, structure, x1):
      '''
      make a trimer of structure separated by translation vector x1
      '''
      xyz = structure.get_atom_coordinates()
      copy1 = structure.copy().set_atom_coordinates(xyz + x1)
      copy2 = structure.copy().set_atom_coordinates(xyz - x1)
      return structure + copy1 + copy2
```

References

(1) Egorova, O.; Hafizi, R.; Woods, D. C.; Day, G. M. Multifidelity Statistical Machine Learning for Molecular Crystal Structure Prediction. *J. Phys. Chem. A* **2020**, *124* (39), 8065–8078. https://doi.org/10.1021/acs.jpca.0c05006.

(2) McDonagh, D.; Skylaris, C.-K.; Day, G. M. Machine-Learned Fragment-Based Energies for Crystal Structure Prediction. *J. Chem. Theory Comput.* **2019**, *15* (4), 2743–2758. https://doi.org/10.1021/acs.jctc.9b00038.

(3) 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* (5), 1289–1300. https://doi.org/10.1039/C7SC04665K.

(4) Taylor, C.; Butler, P.; Day, G. M. Predictive Crystallography at Scale: Mapping, Validating, and Learning from 1,000 Crystal Energy Landscapes. *Faraday Discuss.* **2024**. https://doi.org/10.1039/D4FD00105B.

(5) Wengert, S.; Csányi, G.; Reuter, K.; Margraf, J. T. Data-Efficient Machine Learning for Molecular Crystal Structure Prediction. *Chem. Sci.* **2021**, *12* (12), 4536–4546. https://doi.org/10.1039/D0SC05765G.

(6) Kadan, A.; Ryczko, K.; Wildman, A.; Wang, R.; Roitberg, A.; Yamazaki, T. Accelerated Organic Crystal Structure Prediction with Genetic Algorithms and Machine Learning. *J. Chem. Theory Comput.* **2023**, *19* (24), 9388–9402. https://doi.org/10.1021/acs.jctc.3c00853.

(7) Butler, P. W. V.; Hafizi, R.; Day, G. M. Machine-Learned Potentials by Active Learning from Organic Crystal Structure Prediction Landscapes. *J. Phys. Chem. A* **2024**, *128* (5), 945–957. https://doi.org/10.1021/acs.jpca.3c07129.

(8) Hunnisett, L. M.; Nyman, J.; Francia, N.; Abraham, N. S.; Adjiman, C. S.; Aitipamula, S.; Alkhidir, T.; Almehairbi, M.; Anelli, A.; Anstine, D. M.; Anthony, J. E.; Arnold, J. E.; Bahrami, F.; Bellucci, M. A.; Bhardwaj, R. M.; Bier, I.; Bis, J. A.; Boese, A. D.; Bowskill, D. H.; Bramley, J.; Brandenburg, J. G.; Braun, D. E.; Butler, P. W. V.; Cadden, J.; Carino, S.; Chan, E. J.; Chang, C.; Cheng, B.; Clarke, S. M.; Coles, S. J.; Cooper, R. I.; Couch, R.; Cuadrado, R.; Darden, T.; Day, G. M.; Dietrich, H.; Ding, Y.; DiPasquale, A.; Dhokale, B.; van Eijck, B. P.; Elsegood, M. R. J.; Firaha, D.; Fu, W.; Fukuzawa, K.; Glover, J.; Goto, H.; Greenwell, C.; Guo, R.; Harter, J.; Helfferich, J.; Hofmann, D. W. M.; Hoja, J.; Hone, J.; Hong, R.; Hutchison, G.; Ikabata, Y.; Isayev, O.; Ishaque, O.; Jain, V.; Jin, Y.; Jing, A.; Johnson, E. R.; Jones, I.; Jose, K. V. J.; Kabova, E. A.; Keates, A.; Kelly, P. F.; Khakimov, D.; Konstantinopoulos, S.; Kuleshova, L. N.; Li, H.; Lin, X.; List, A.; Liu, C.; Liu, Y. M.; Liu, Z.; Liu, Z.-P.; Lubach, J. W.; Marom, N.; Maryewski, A. A.; Matsui, H.; Mattei, A.; Mayo, R. A.; Melkumov, J. W.; Mohamed, S.; Momenzadeh Abardeh, Z.; Muddana, H. S.; Nakayama, N.; Nayal, K. S.; Neumann, M. A.; Nikhar, R.; Obata, S.; O'Connor, D.; Oganov, A. R.; Okuwaki, K.; Otero-de-la-Roza, A.; Pantelides, C. C.; Parkin, S.; Pickard, C. J.; Pilia, L.; Pivina, T.; Podeszwa, R.; Price, A. J. A.; Price, L. S.; Price, S. L.; Probert, M. R.; Pulido, A.; Ramteke, G. R.; Rehman, A. U.; Reutzel-Edens, S. M.; Rogal, J.; Ross, M. J.; Rumson, A. F.; Sadiq, G.; Saeed, Z. M.; Salimi, A.; Salvalaglio, M.; Sanders de Almada, L.; Sasikumar, K.; Sekharan, S.; Shang, C.; Shankland, K.; Shinohara, K.; Shi, B.; Shi, X.; Skillman, A. G.; Song, H.; Strasser, N.; van de Streek, J.; Sugden, I. J.; Sun, G.; Szalewicz, K.; Tan, B. I.; Tan, L.; Tarczynski, F.; Taylor, C. R.; Tkatchenko, A.; Tom, R.; Tuckerman, M. E.; Utsumi, Y.; Vogt-Maranto, L.; Weatherston, J.; Wilkinson, L. J.; Willacy, R. D.; Wojtas, L.; Woollam, G. R.; Yang, Z.; Yonemochi, E.; Yue, X.; Zeng, Q.; Zhang, Y.; Zhou, T.; Zhou, Y.; Zubatyuk, R.; Cole, J. C. The Seventh Blind Test of Crystal Structure Prediction: Structure Generation Methods. *Acta Crystallogr. Sect. B* **2024**, *80* (6), 517–547. https://doi.org/10.1107/S2052520624007492.

(9) Hunnisett, L. M.; Francia, N.; Nyman, J.; Abraham, N. S.; Aitipamula, S.; Alkhidir, T.; Almehairbi, M.; Anelli, A.; Anstine, D. M.; Anthony, J. E.; Arnold, J. E.; Bahrami, F.; Bellucci, M. A.; Beran, G. J. O.; Bhardwaj, R. M.; Bianco, R.; Bis, J. A.; Boese, A. D.; Bramley, J.; Braun, D. E.; Butler, P. W. V.; Cadden, J.; Carino, S.; Červinka, C.; Chan, E. J.; Chang, C.; Clarke, S. M.; Coles, S. J.; Cook, C. J.; Cooper, R. I.; Darden, T.; Day, G. M.; Deng, W.; Dietrich, H.; DiPasquale, A.; Dhokale, B.; van Eijck, B. P.; Elsegood, M. R. J.; Firaha, D.; Fu, W.; Fukuzawa, K.; Galanakis, N.; Goto, H.; Greenwell, C.; Guo, R.; Harter, J.; Helfferich, J.; Hoja, J.; Hone, J.; Hong, R.; Hušák, M.; Ikabata, Y.; Isayev, O.; Ishaque, O.; Jain, V.; Jin, Y.; Jing, A.; Johnson, E. R.; Jones, I.; Jose, K. V. J.; Kabova, E. A.; Keates, A.; Kelly, P. F.; Klimeš, J.; Kostková, V.; Li, H.; Lin, X.; List, A.; Liu, C.; Liu, Y. M.; Liu, Z.; Lončarić, I.; Lubach, J. W.; Ludík, J.; Marom, N.; Matsui, H.; Mattei, A.; Mayo, R. A.; Melkumov, J. W.; Mladineo, B.; Mohamed, S.; Momenzadeh Abardeh, Z.; Muddana, H. S.; Nakayama, N.; Nayal, K. S.; Neumann, M. A.; Nikhar, R.; Obata, S.; O'Connor, D.; Oganov, A. R.; Okuwaki, K.; Otero-de-la-Roza, A.; Parkin, S.; Parunov, A.; Podeszwa, R.; Price, A. J. A.; Price, L. S.; Price, S. L.; Probert, M. R.; Pulido, A.; Ramteke, G. R.; Rehman, A. U.; Reutzel-Edens, S. M.; Rogal, J.; Ross, M. J.; Rumson, A. F.; Sadiq, G.; Saeed, Z. M.; Salimi, A.; Sasikumar, K.; Sekharan, S.; Shankland, K.; Shi, B.; Shi, X.; Shinohara, K.; Skillman, A. G.; Song, H.; Strasser, N.; van de Streek, J.; Sugden, I. J.; Sun, G.; Szalewicz, K.; Tan, L.; Tang, K.; Tarczynski, F.; Taylor, C. R.; Tkatchenko, A.; Tom, R.; Touš, P.; Tuckerman, M. E.; Unzueta, P. A.; Utsumi, Y.; Vogt-Maranto, L.; Weatherston, J.; Wilkinson, L. J.; Willacy, R. D.; Wojtas, L.; Woollam, G. R.; Yang, Y.; Yang, Z.; Yonemochi, E.; Yue, X.; Zeng, Q.; Zhou, T.; Zhou, Y.; Zubatyuk, R.; Cole, J. C. The Seventh Blind Test of Crystal Structure Prediction: Structure Ranking Methods. *Acta Crystallogr. Sect. B* **2024**, *80* (6), 548–574. https://doi.org/10.1107/S2052520624008679.

(10) Jacobson, L. D.; Stevenson, J. M.; Ramezanghorbani, F.; Ghoreishi, D.; Leswing, K.; Harder, E. D.; Abel, R. Transferable Neural Network Potential Energy Surfaces for Closed-Shell Organic Molecules: Extension to Ions. *J. Chem. Theory Comput.* **2022**, *18* (4), 2354–2366. https://doi.org/10.1021/acs.jctc.1c00821.

(11) Dajnowicz, S.; Agarwal, G.; Stevenson, J. M.; Jacobson, L. D.; Ramezanghorbani, F.; Leswing, K.; Friesner, R. A.; Halls, M. D.; Abel, R. High-Dimensional Neural Network Potential for Liquid Electrolyte Simulations. *J. Phys. Chem. B* **2022**, *126* (33), 6271–6280. https://doi.org/10.1021/acs.jpcb.2c03746.

(12) Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121* (16), 10037–10072. https://doi.org/10.1021/acs.chemrev.0c00868.

(13) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys Rev Lett* **2007**, *98* (14), 146401. https://doi.org/10.1103/PhysRevLett.98.146401.

(14) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem Sci* **2017**, *8* (4), 3192–3203. https://doi.org/10.1039/C6SC05720A.

(15) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A Fourth-Generation High-Dimensional Neural Network Potential with Accurate Electrostatics Including Non-Local Charge Transfer. *Nat. Commun.* **2021**, *12* (1), 398. https://doi.org/10.1038/s41467-020-20427-2.

(16) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27* (15), 1787–1799. https://doi.org/10.1002/jcc.20495.

(17) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom–Atom Dispersion Corrections. *Phys Chem Chem Phys* **2008**, *10* (44), 6615–6620. https://doi.org/10.1039/B810189B.

(18) Leif Jacobson, ,James Stevenson ,Farhad Ramezanghorbani ,Steven Dajnowicz ,Karl Leswing. Leveraging Multitask Learning to Improve the Transferability of Machine Learned

Force Fields. *ChemRxiv*. https://doi.org/10.26434/chemrxiv-2023-8n737.
(19) Stevenson, J. M.; Jacobson, L. D.; Zhao, Y.; Wu, C.; Maple, J.; Leswing, K.; Harder, E.; Abel, R. Schrödinger-ANI: An Eight-Element Neural Network Interaction Potential with Greatly Expanded Coverage of Druglike Chemical Space, 2019.
(20) Christensen, A. S.; Sirumalla, S. K.; Qiao, Z.; O'Connor, M. B.; Smith, D. G. A.; Ding, F.; Bygrave, P. J.; Anandkumar, A.; Welborn, M.; Manby, F. R.; Miller, T. F., III. OrbNet Denali: A Machine Learning Potential for Biological and Organic Chemistry with Semi-Empirical Cost and DFT Accuracy. *J. Chem. Phys.* **2021**, *155* (20), 204103. https://doi.org/10.1063/5.0061990.
(21) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Jensen, P. B.; Kermode, J.; Kitchin, J. R.; Kolsbjerg, E. L.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Maronsson, J. B.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The Atomic Simulation Environment—a Python Library for Working with Atoms. *J. Phys. Condens. Matter* **2017**, *29* (27), 273002. https://doi.org/10.1088/1361-648X/aa680e.
(22) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys Rev Lett* **1996**, *77* (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.
(23) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104. https://doi.org/10.1063/1.3382344.
(24) Ehlert, S.; Huniar, U.; Ning, J.; Furness, J. W.; Sun, J.; Kaplan, A. D.; Perdew, J. P.; Brandenburg, J. G. r2SCAN-D4: Dispersion Corrected Meta-Generalized Gradient Approximation for General Chemical Applications. *J. Chem. Phys.* **2021**, *154* (6), 061101. https://doi.org/10.1063/5.0041008.
(25) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and Numerically Efficient r2SCAN Meta-Generalized Gradient Approximation. *J. Phys. Chem. Lett.* **2020**, *11* (19), 8208–8215. https://doi.org/10.1021/acs.jpclett.0c02405.
(26) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r2SCAN-3c: A "Swiss Army Knife" Composite Electronic-Structure Method. *J. Chem. Phys.* **2021**, *154* (6), 064103. https://doi.org/10.1063/5.0040021.
(27) 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* (2), 024705. https://doi.org/10.1063/1.4812819.
(28) Brandenburg, J. G.; Maas, T.; Grimme, S. Benchmarking DFT and Semiempirical Methods on Structures and Lattice Energies for Ten Ice Polymorphs. *J. Chem. Phys.* **2015**, *142* (12), 124104. https://doi.org/10.1063/1.4916070.
(29) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van Der Waals Interactions. *Phys Rev Lett* **2012**, *108* (23), 236402. https://doi.org/10.1103/PhysRevLett.108.236402.
(30) 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 Jr, 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. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. 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.; Neumann, 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* **2016**, *72* (4), 439–459. https://doi.org/10.1107/S2052520616007447.

(31) Hoja, J.; Ko, H.-Y.; Neumann, M. A.; Car, R.; DiStasio, R. A.; Tkatchenko, A. Reliable and Practical Computational Description of Molecular Crystal Polymorphs. *Sci. Adv. 5* (1), eaau3338. https://doi.org/10.1126/sciadv.aau3338.

(32) Hoja, J.; Tkatchenko, A. First-Principles Stability Ranking of Molecular Crystal Polymorphs with the DFT+MBD Approach. *Faraday Discuss* **2018**, *211* (0), 253–274. https://doi.org/10.1039/C8FD00066B.

(33) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Buongiorno Nardelli, M.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M.; Colonna, N.; Carnimeo, I.; Dal Corso, A.; de Gironcoli, S.; Delugas, P.; DiStasio, R. A.; Ferretti, A.; Floris, A.; Fratesi, G.; Fugallo, G.; Gebauer, R.; Gerstmann, U.; Giustino, F.; Gorni, T.; Jia, J.; Kawamura, M.; Ko, H.-Y.; Kokalj, A.; Küçükbenli, E.; Lazzeri, M.; Marsili, M.; Marzari, N.; Mauri, F.; Nguyen, N. L.; Nguyen, H.-V.; Otero-de-la-Roza, A.; Paulatto, L.; Poncé, S.; Rocca, D.; Sabatini, R.; Santra, B.; Schlipf, M.; Seitsonen, A. P.; Smogunov, A.; Timrov, I.; Thonhauser, T.; Umari, P.; Vast, N.; Wu, X.; Baroni, S. Advanced Capabilities for Materials Modelling with Quantum ESPRESSO. *J. Phys. Condens. Matter* **2017**, *29* (46), 465901. https://doi.org/10.1088/1361-648X/aa8f79.

(34) Garrity, K. F.; Bennett, J. W.; Rabe, K. M.; Vanderbilt, D. Pseudopotentials for High-Throughput DFT Calculations. *Comput. Mater. Sci.* **2014**, *81*, 446–452. https://doi.org/10.1016/j.commatsci.2013.08.053.

(35) Hamann, D. R. Optimized Norm-Conserving Vanderbilt Pseudopotentials. *Phys Rev B* **2013**, *88* (8), 085117. https://doi.org/10.1103/PhysRevB.88.085117.

(36) 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. *Cryst. Growth Des.* **2013**, *13* (4), 1602–1617. https://doi.org/10.1021/cg301826s.

(37) Beran, G. J. O.; Sugden, I. J.; Greenwell, C.; Bowskill, D. H.; Pantelides, C. C.; Adjiman, C. S. How Many More Polymorphs of ROY Remain Undiscovered. *Chem Sci* **2022**, *13* (5), 1288–1297. https://doi.org/10.1039/D1SC06074K.

(38) 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. *J. Am. Chem. Soc.* **2019**, *141* (35), 13887–13897. https://doi.org/10.1021/jacs.9b06634.

(39) Greenwell, C.; Beran, G. J. O. Inaccurate Conformational Energies Still Hinder Crystal Structure Prediction in Flexible Organic Molecules. *Cryst. Growth Des.* **2020**, *20* (8), 4875–4881. https://doi.org/10.1021/acs.cgd.0c00676.

(40) Wolff, H.; Quere, L.; Riedner, J. Polymorphic Form of Rotigotine. *EP 2215072A2* **2010**.

(41) Rietveld, I. B.; Céolin, R. Rotigotine: Unexpected Polymorphism with Predictable Overall Monotropic Behavior. *J. Pharm. Sci.* **2015**, *104* (12), 4117–4122. https://doi.org/10.1002/jps.24626.

(42) Mortazavi, M.; Hoja, J.; Aerts, L.; Quéré, L.; van de Streek, J.; Neumann, M. A.; Tkatchenko, A. Computational Polymorph Screening Reveals Late-Appearing and Poorly-Soluble Form of Rotigotine. *Commun. Chem.* **2019**, *2* (1), 70. https://doi.org/10.1038/s42004-019-0171-y.

(43) Williams, M. J.; Chen, Q.; Codan, L.; Dermenjian, R. K.; Dreher, S.; Gibson, A. W.; He, X.;

Jin, Y.; Keen, S. P.; Lee, A. Y.; Lieberman, D. R.; Lin, W.; Liu, G.; McLaughlin, M.; Reibarkh, M.; Scott, J. P.; Strickfuss, S.; Tan, L.; Varsolona, R. J.; Wen, F. Process Development of the HCV NS5B Site D Inhibitor MK-8876. *Org. Process Res. Dev.* **2016**, *20* (7), 1227–1238. https://doi.org/10.1021/acs.oprd.5b00405.
(44) Newman, J. A.; Iuzzolino, L.; Tan, M.; Orth, P.; Bruhn, J.; Lee, A. Y. From Powders to Single Crystals: A Crystallographer's Toolbox for Small-Molecule Structure Determination. *Mol. Pharm.* **2022**, *19* (7), 2133–2141. https://doi.org/10.1021/acs.molpharmaceut.2c00020.
(45) Stephenson, G. A.; Kendrick, J.; Wolfangel, C.; Leusen, F. J. J. Symmetry Breaking: Polymorphic Form Selection by Enantiomers of the Melatonin Agonist and Its Missing Polymorph. *Cryst. Growth Des.* **2012**, *12* (8), 3964–3976. https://doi.org/10.1021/cg300398a.
(46) Kendrick, J.; Stephenson, G. A.; Neumann, M. A.; Leusen, F. J. J. Crystal Structure Prediction of a Flexible Molecule of Pharmaceutical Interest with Unusual Polymorphic Behavior. *Cryst. Growth Des.* **2013**, *13* (2), 581–589. https://doi.org/10.1021/cg301222m.
(47) Goa, K. L.; Spencer, C. M. Bicalutamide in Advanced Prostate Cancer. *Drugs Aging* **1998**, *12* (5), 401–422. https://doi.org/10.2165/00002512-199812050-00006.
(48) Vega, D. R.; Polla, G.; Martinez, A.; Mendioroz, E.; Reinoso, M. Conformational Polymorphism in Bicalutamide. *Int. J. Pharm.* **2007**, *328* (2), 112–118. https://doi.org/10.1016/j.ijpharm.2006.08.001.
(49) Simons, L. J.; Caprathe, B. W.; Callahan, M.; Graham, J. M.; Kimura, T.; Lai, Y.; LeVine, H.; Lipinski, W.; Sakkab, A. T.; Tasaki, Y.; Walker, L. C.; Yasunaga, T.; Ye, Y.; Zhuang, N.; Augelli-Szafran, C. E. The Synthesis and Structure–Activity Relationship of Substituted N-Phenyl Anthranilic Acid Analogs as Amyloid Aggregation Inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19* (3), 654–657. https://doi.org/10.1016/j.bmcl.2008.12.049.
(50) Samas, B.; Clark, W. D.; Li, A.-F.; Pickard, F. C. I.; Wood, G. P. F. Five Degrees of Separation: Characterization and Temperature Stability Profiles for the Polymorphs of PD-0118057 (Molecule XXIII). *Cryst. Growth Des.* **2021**, *21* (8), 4435–4444. https://doi.org/10.1021/acs.cgd.1c00351.
(51) Yang, M.; Dybeck, E.; Sun, G.; Peng, C.; Samas, B.; Burger, V. M.; Zeng, Q.; Jin, Y.; Bellucci, M. A.; Liu, Y.; Zhang, P.; Ma, J.; Jiang, Y. A.; Hancock, B. C.; Wen, S.; Wood, G. P. F. Prediction of the Relative Free Energies of Drug Polymorphs above Zero Kelvin. *Cryst. Growth Des.* **2020**, *20* (8), 5211–5224. https://doi.org/10.1021/acs.cgd.0c00422.
(52) Campeta, A. M.; Chekal, B. P.; Abramov, Y. A.; Meenan, P. A.; Henson, M. J.; Shi, B.; Singer, R. A.; Horspool, K. R. Development of a Targeted Polymorph Screening Approach for a Complex Polymorphic and Highly Solvating API. *J. Pharm. Sci.* **2010**, *99* (9), 3874–3886. https://doi.org/10.1002/jps.22230.
(53) Chekal, B. P.; Campeta, A. M.; Abramov, Y. A.; Feeder, N.; Glynn, P. P.; McLaughlin, R. W.; Meenan, P. A.; Singer, R. A. The Challenges of Developing an API Crystallization Process for a Complex Polymorphic and Highly Solvating System. Part I. *Org. Process Res. Dev.* **2009**, *13* (6), 1327–1337. https://doi.org/10.1021/op9001559.
(54) Vasileiadis, M.; Pantelides, C. C.; Adjiman, C. S. Prediction of the Crystal Structures of Axitinib, a Polymorphic Pharmaceutical Molecule. *Chem. Eng. Sci.* **2015**, *121*, 60–76. https://doi.org/10.1016/j.ces.2014.08.058.
(55) Blade, H.; Blundell, C. D.; Brown, S. P.; Carson, J.; Dannatt, H. R. W.; Hughes, L. P.; Menakath, A. K. Conformations in Solution and in Solid-State Polymorphs: Correlating Experimental and Calculated Nuclear Magnetic Resonance Chemical Shifts for Tolfenamic Acid. *J. Phys. Chem. A* **2020**, *124* (43), 8959–8977. https://doi.org/10.1021/acs.jpca.0c07000.
(56) Baias, M.; Dumez, J.-N.; Svensson, P. H.; Schantz, S.; Day, G. M.; Emsley, L. De Novo Determination of the Crystal Structure of a Large Drug Molecule by Crystal Structure

Prediction-Based Powder NMR Crystallography. *J. Am. Chem. Soc.* **2013**, *135* (46), 17501–17507. https://doi.org/10.1021/ja4088874.