

# Efficient Crystal Structure Prediction for Structurally Related Molecules with Accurate and Transferable Tailor-Made Force Fields

Alessandra Mattei, Richard S. Hong, Hanno Dietrich, Dzmitry Firaha, Julian Helfferich, Yifei Michelle Liu, Kiran Sasikumar, Nathan S. Abraham, Rajni Miglani Bhardwaj, Marcus A. Neumann,\* and Ahmad Y. Sheikh\*



Cite This: *J. Chem. Theory Comput.* 2022, 18, 5725–5738



Read Online

ACCESS |



Metrics & More

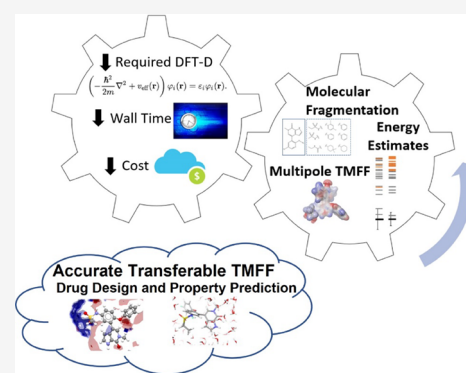


Article Recommendations



Supporting Information

**ABSTRACT:** Crystal structure prediction (CSP) is generally used to complement experimental solid form screening and applied to individual molecules in drug development. The fast development of algorithms and computing resources offers the opportunity to use CSP earlier and for a broader range of applications in the drug design cycle. This study presents a novel paradigm of CSP specifically designed for structurally related molecules, referred to as Quick-CSP. The approach prioritizes more accurate physics through robust and transferable tailor-made force fields (TMFFs), such that significant efficiency gains are achieved through the reduction of expensive ab initio calculations. The accuracy of the TMFF is increased by the introduction of electrostatic multipoles, and the fragment-based force field parameterization scheme is demonstrated to be transferable for a family of chemically related molecules. The protocol is benchmarked with structurally related compounds from the Bromodomain and Extraterminal (BET) domain inhibitors series. A new convergence criterion is introduced that aims at performing only as many ab initio optimizations of crystal structures as required to locate the bottom of the crystal energy landscape within a user-defined accuracy. The overall approach provides significant cost savings ranging from three- to eight-fold less than the full-CSP workflow. The reported advancements expand the scope and utility of the underlying CSP building blocks as well as their novel reassembly to other applications earlier in the drug design cycle to guide molecule design and selection.



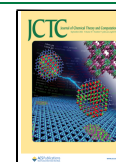
## 1. INTRODUCTION

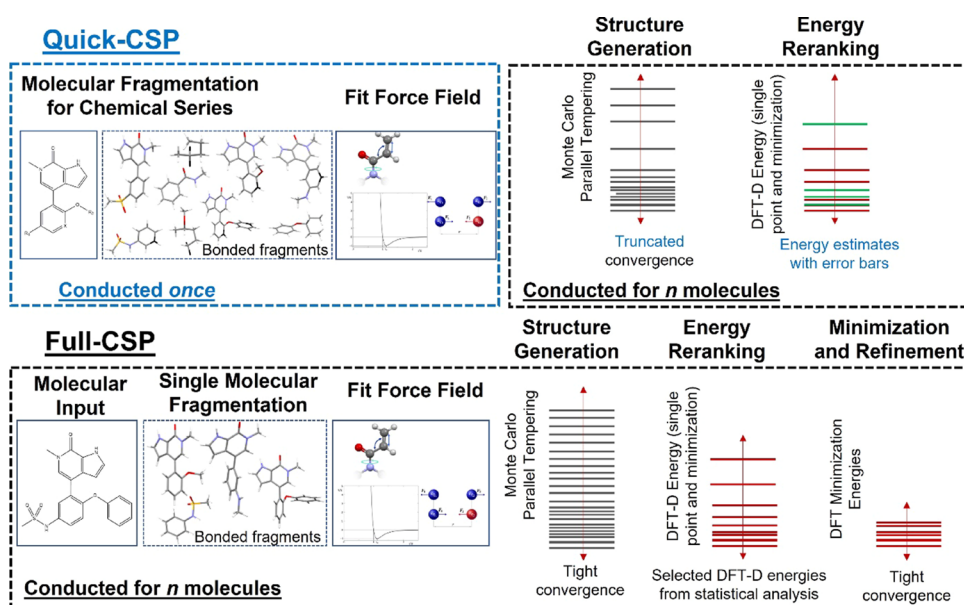
Crystal structure prediction (CSP) of organic molecules aims to generate and rank the crystal structures with various three-dimensional (3D) packing arrangements of molecules, starting from the chemical diagram alone. The field has made substantial progress since the challenge laid down in the seminal commentary by Dunitz in the early 2000s.<sup>1</sup> This progress is best manifested by the results of the latest blind test organized by the Cambridge Crystallographic Data Center (CCDC),<sup>2</sup> a sixth in the series where the predictions have become more accurate, despite increased chemical complexity and conformational flexibility.<sup>2–7</sup> Overall, much of the underlying research focus is directed toward accurately calculating the relative energies of all putative crystal packing arrangements. Hence, all CSP approaches are structured hierarchically so that the overall workflow resembles a funnel of increasingly narrow criteria and accurate energy models. Relatively less-demanding calculations based on force fields are used to generate a large number of candidate crystal structures, while more advanced first-principles computational methods are applied for the final ranking of the most promising crystal structures.

Tailor-made force fields (TMFFs) represented an initial breakthrough in the field, allowing sufficiently accurate calculations of the energy minima associated with crystal packings.<sup>8,9</sup> Currently, TMFFs and first-principles energy ranking approaches with dispersion-inclusive hybrid functionals<sup>10,11</sup> and many-body dispersion interactions<sup>12,13</sup> are the most successful methods for yielding reliable and accurate lattice energies.<sup>14,15</sup> More recently, molecular dynamics and enhanced molecular dynamics have also been rigorously utilized to probe free energy stabilities and to eliminate labile structures.<sup>16–18</sup> Such algorithms have been thoroughly benchmarked and widely applied to predict the crystal energy landscape of organic molecules, including complex chemical systems of pharmaceutical relevance. In parallel, modern machine learning potentials combined with state-of-the-art first-principles calculations have been demonstrated to

Received: May 3, 2022

Published: August 5, 2022





**Figure 1.** Comparison of Quick-CSP and Full-CSP workflows with the differences highlighted in blue text.

approach the accuracy required for predictive stability rankings with affordable computational costs.<sup>18,19</sup> In addition, the unprecedented level of parallelization available through cloud computing has permitted the resource-intensive CSP calculations to be completed in weeks, if not days.<sup>20</sup>

Driven by the fast development of algorithmic improvements and powerful computing resources, CSP has fueled considerable interest across the pharmaceutical industry to complement experimental solid form screening.<sup>21–24</sup> Contemporary prediction of crystal energy landscapes is routinely used to assess and manage the robustness of the solid form selected for development.<sup>25–27</sup> In the context of drug design, when CSP is performed earlier in the design cycle, crucial insights can be gained by understanding the structural determinants of key physicochemical properties inherently rooted in essential crystal structure features, such as conformational flexibility and intermolecular interactions.<sup>28,29</sup> Beyond these downstream calculations, fundamental building blocks of the CSP workflow, such as force fields,<sup>30</sup> enhanced Monte-Carlo simulations,<sup>31</sup> and dispersion-corrected density functional theory (DFT-D),<sup>32</sup> are also broadly applicable in the molecular design itself.<sup>33</sup> The need for computational efficiency in computer aided drug design often leads to reliance on traditional transferable force fields, such as OPLS,<sup>34,35</sup> AMBER,<sup>36</sup> CHARMM,<sup>37</sup> COMPASS,<sup>38</sup> and GROMOS.<sup>39</sup> While limited in accuracy by their underlying functional form, these generic force fields are parameterized with experimental and quantum mechanics data for small organic molecules and/or macromolecules, conferring good parameter transferability.<sup>40</sup> On the other hand, TMFFs are parameterized on molecule-by-molecule basis and thus are more accurate.<sup>8</sup> In addition to requiring more computational effort, they are obviously not transferable. Before the immense value of TMFFs can be realized for broader applications beyond CSP, improvements on both the transferability and computational efficiency are therefore needed. The limitations of standard CSP workflows even when executed at ever-increasing speeds leave many opportunities that, if addressed, can expand the scope and utility of the underlying building blocks and their novel reassemblies.

In this study, we present a new paradigm of CSP specifically designed for structurally related molecules, referred to as Quick-CSP, which improves the TMFF accuracy, broadens the concept of fragment-based parameterization to impart transferability, and significantly reduces the extent of ab initio calculations for unprecedented levels of overall efficiency and versatility. We demonstrate the performance of this novel approach on moderately flexible molecules in the pyrrolopyridone-based series of Bromodomain and Extraterminal (BET) domain inhibitors, where the implications of subtle chemical modifications on conformational flexibility, crystal packing, and the resulting aqueous crystalline solubility were recently reported.<sup>29</sup> Our new paradigm provides a stepping-stone to the simultaneous prediction of crystal structures for more than one molecule. The resulting 3D packing arrangements of molecules can thereby be incorporated in physicochemical property predictions, and the insight into the structural determinants of the properties can enable rational drug design.<sup>28,30,41–44</sup>

## 2. RESULTS AND DISCUSSION

Figure 1 shows key differentiating features of the Quick-CSP workflow compared to the standard Full-CSP, as implemented in the GRACE<sup>45</sup> software suite. Here, the accuracy of the TMFF is improved by the introduction of electrostatic multipoles<sup>46–48</sup> attached to local atomic coordinate systems. The fragment-based approach for force field parameterization is expanded to a family of chemically related molecules, resulting in higher accuracy and some level of transferability. Furthermore, the efficiency of the fragment-based approach systematically increases as the number of molecules under consideration increases, and the approach is scalable to chemical series of virtually any size. In the Quick-CSP workflow, CSPs are run for all compounds in parallel, relaxing the convergence requirements. A key feature is that the bottom energy of the crystal energy landscape is determined through a statistical analysis together with estimates of multiple crystal structures' DFT-D energies and their associated error. This allows convergence with only a few, robustly selected crystal

structures to be DFT-D-optimized, ultimately reducing the computational cost.

The Quick-CSP approach was applied to BET bromodomain inhibitors, which have generated significant interest for studying various types of cancers, including acute myeloid leukemia (AML) and multiple myeloma (MM).<sup>49–51</sup> Recent structure–activity-relationship (SAR) efforts have led to the identification of pyrrolopyridone core derivatives that demonstrated submicromolar activity in binding assays against BRD4 and substantial efficacy in several *in vivo* mouse xenograft models.<sup>52</sup> This family of molecules also shows sufficient diversity in terms of both the features related to developability within the context of Rule of 5 (Ro5) framework and structural variance to challenge each step of the Quick-CSP and Full-CSP workflows.

### 2.1. Features of BET Pyrrolopyridone Core Molecules.

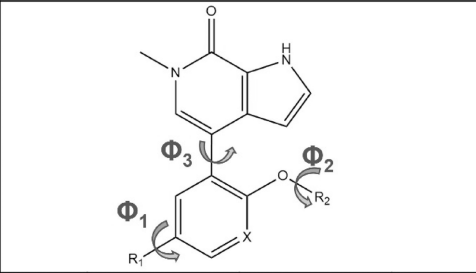
Eight molecules of the BET bromodomain inhibitors were selected in this study from the lead optimization efforts.<sup>52</sup> As displayed in Table 1, these molecules feature a pyrrolopyridone ring that acts as a hydrogen bond acceptor and donor through the pyrrolopyridone carbonyl and the pyrrole amino proton, with the ability of forming a bidentate hydrogen bond. The hydrogen bonding groups and dominant molecular shape are conserved when small chemical modifications are made on the

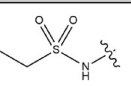
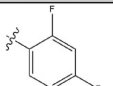
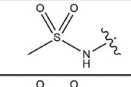
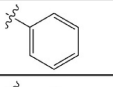
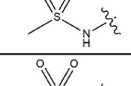
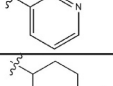
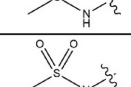
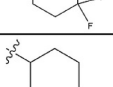
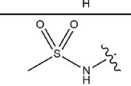
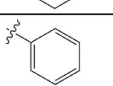
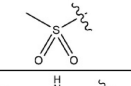
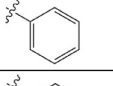
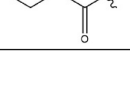
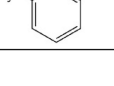
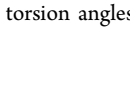
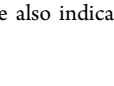
distal ether group or the central aryl ring and/or when the functional group attached to the central aryl ring (i.e., sulfonamide, sulfone, or reverse amide moieties) is substituted. The eight molecules show a moderate conformational flexibility with major torsion angles being around the terminal chain attached to the central aryl ring ( $\Phi_1$ ), the distal cyclic ether group ( $\Phi_2$ ), and the pyrrolopyridone ring ( $\Phi_3$ ). The facile reorientation of the functional group attached to the central aryl ring, as measured by  $\Phi_1$ , is not sterically impeded by the distal ether or the pyrrolopyridone moieties, because of the strong cooperativity between neighboring torsion angles.<sup>29</sup> The three key flexible torsion angles can however impact the degree of  $\pi$ -delocalization or conjugation throughout the molecule. Notably, the positioning of the sulfonamide moiety allows for some degree of conjugation through the central aryl ring and the pyrrolopyridone ring with a more planar conformation, as measured by  $\Phi_1$  and  $\Phi_3$ , showing the maximum conjugation. Thus, the pyrrolopyridone-based molecules can adopt a variety of 3D molecular conformations with different relative positions of the hydrogen bond acceptors and donors.<sup>29</sup>

The range of 3D molecular conformations can in turn result in different crystal packing arrangements and specific intermolecular interactions in the crystalline state. Electrostatic forces between the polar and aromatic pyrrolopyridone core molecules are expected to play an important role in crystal structures because of the large dipole moment (Table S1). Similarly,  $\pi$ – $\pi$  stacking intermolecular interactions, involving the distal ether aryl and/or the pyrrolopyridone rings, can contribute considerably to stabilize the crystal packing. While it is not always clear a priori, the intricacy of this family of molecules suggests the need of a TMFF that accounts for the electrostatics effects of features, such as  $\pi$ -electron density, on hydrogen bonding and aromatic  $\pi$ -stacking interactions to improve accuracy upfront in a CSP workflow.

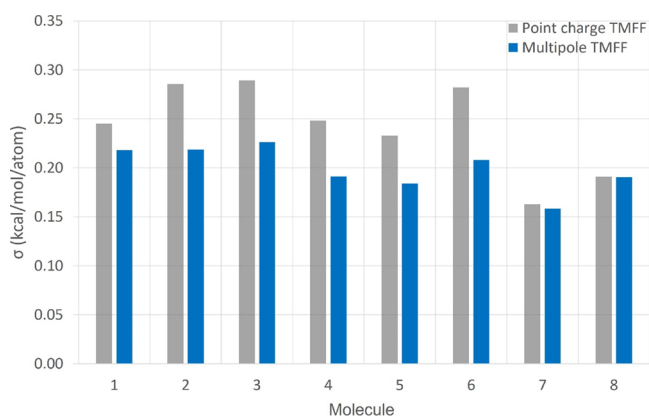
**2.2. More Accurate TMFF Method.** The Quick-CSP workflow was run twice using the multipole TMFF in one case and point charge TMFF in the other. For each molecule, the accuracy of the backfitted TMFF was assessed by comparing the lattice energies of force field-optimized and ab initio-optimized crystal structures, the latter being taken from the eight full-CSP runs. The root-mean-square deviation ( $\sigma$ ), expressed in terms of an error per atom, between the TMFF and PBE-NP<sup>10</sup> lattice energies was then calculated. The multipole TMFF shows systematically lower errors than the point charge TMFF across the pyrrolopyridone-based molecules (Figure 2). The multipole TMFF contains more parameters that allow for a better description of the electron density distribution, resulting in more accurate lattice energies. The large number of parameters can, however, make the fitting process more complicated and result in decreased accuracy if the fitting algorithm gets trapped in an unfavorable local minimum. The results demonstrate that GRACE's force field parameter fitting algorithm transformed the increased complexity, through the introduction of electrostatic multipoles, into increased accuracy. In addition, switching point charge to multipole TMFFs reduces the force field error by 16.2% after backfitting and 14.2% before backfitting (Figure S1 and Table S2). The multipole TMFF shows the strongest improvements for the molecules containing a sulfonamide moiety (i.e., molecules 1–6). Figure 3a provides more detailed insights into the performance of the two backfitted TMFFs compared to the PBE-NP benchmark for molecule (2), which exhibits the

**Table 1. Molecular Diagram of BET Bromodomain Inhibitors with Chemical Modifications on the Pyrrolopyridone Scaffold<sup>a</sup>**



ID	R <sub>1</sub>	R <sub>2</sub>	X
1			CH
2			CH
3			CH
4			CH
5			CH
6			N
7			CH
8			CH

<sup>a</sup>The three flexible torsion angles are also indicated.

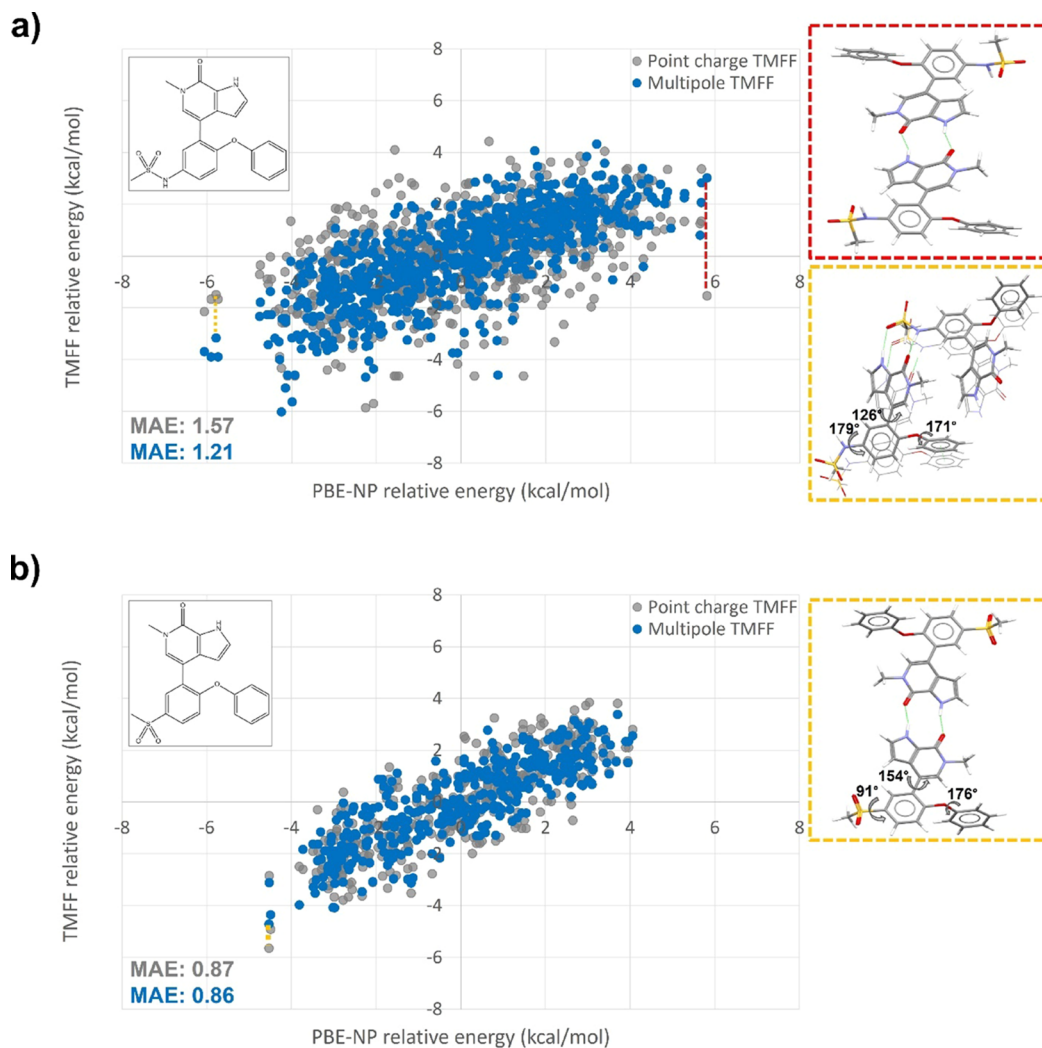


**Figure 2.** Accuracy, as measured by  $\sigma$ -value, of backfitted point charge (gray) and multipole (blue) TMFFs for each pyrrolopyridone-based molecule. Reporting the  $\sigma$ -value per atom has the advantage of removing the dependency on the number of atoms.

largest difference in force field accuracies of the molecules with the sulfonamide moiety. The most striking feature of the correlation plot is an enhanced agreement of the multipole

TMFF energies with PBE-NP calculations that results in a tighter crystal lattice energy distribution. As expected from the smaller  $\sigma$ -value, the multipole TMFF provides improved lattice energies of the generated crystal structures, yielding a mean absolute error (MAE) of 1.21 kcal/mol. In contrast, there appears to be a significant scatter in the predictions of the relative energies computed by the point charge TMFF, leading to a large MAE of 1.57 kcal/mol.

Closer inspection of the crystal lattice energies revealed that the point charge TMFF can miscalculate the strength of hydrogen bonding interactions. The largest negative deviation relative to PBE-NP calculations occurred for the generated crystal structure that shows a bidentate hydrogen bond between pyrrolopyridone rings and lacks intermolecular hydrogen bonding interactions involving the sulfonamide moiety (Figure 3a top inset (red)). The latter structural feature leads to a large electron density on the electronegative sulfonamide group. The specific description of electrostatic effects by the lone pair electrons and the electron density distribution can make a significant difference to the ability of representing the directionality and strength of hydrogen bonding interactions.<sup>53</sup> The error of the point charge TMFF



**Figure 3.** Comparison of backfitted point charge (gray) and multipole (blue) TMFFs relative energies to the corresponding reference PBE-NP energies for (a) molecule (2) and (b) molecule (7). Insets show crystal structures with large deviations in backfitted TMFFs relative energies with respect to PBE-NP.

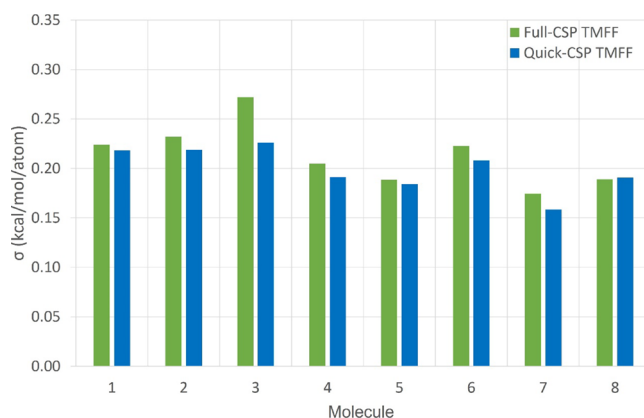
may thereby originate from systematically underestimating the electrostatic interactions and overestimating the hydrogen bond strength, as the directionality of the intermolecular hydrogen bonding is not accounted. On the other hand, the multipole TMFFs account for the anisotropy of the electron density around the nitrogen and oxygen atoms of the sulfonamide group and considerably reduce the energy error.

Further assessment of the crystal lattice energy distribution shows that the multipole TMFF performs well for relatively planar molecular conformations that allow the  $\pi$ -delocalization or conjugation to extend to the central aryl ring via the nitrogen lone pair of the sulfonamide group. This is consistent with the electrostatic potential around conjugated systems, aromatic systems, and/or lone pairs being more realistically reproduced by electrostatic multipole models.<sup>54</sup> In addition, as a result of the planar molecular conformation and extended  $\pi$ -delocalization, the generated crystal structure with a large positive deviation displays aromatic  $\pi$ -stacking interactions that are known to be more correctly modeled by electrostatic multipole functionals (Figure 3a bottom inset (yellow)).<sup>55</sup> In such a case, the point charge TMFF error stems from the under-stabilization of the extended  $\pi$ -delocalization as well as the underestimation of the intermolecular aromatic  $\pi$ -stacking interactions present in the more planar conformation. Similarly, other molecules containing the sulfonamide moiety show a MAE range of 1.09–1.26 kcal/mol for relative energies computed by the multipole TMFF, whereas MAE range of 1.37–1.57 kcal/mol was obtained by the point charge model (Figure S2a–e).

For molecules (7) and (8) that contain the sulfonyl group and the reverse amide directly attached to the central aryl ring, respectively, both TMFFs yield crystal lattice energies that are in very good agreement with the reference PBE-NP method, as indicated by the low and comparable MAE values (Figures 3b and S2f). The origin of these results may reflect the extent to which molecular conformations, intermolecular interactions, and their interplay contribute toward the stability of molecular organic crystals. The incorporation of a hydrogen bond acceptor directly attached to the central aryl ring alters the extent of the overall  $\pi$ -delocalization in the system. For instance, the molecular conformation of compound (7) in the crystal structure that shows a large positive energy error exhibits a  $\Phi_1$  of about  $90^\circ$ , a position that completely “breaks” the conjugation between the sulfonyl group and the central aryl ring (Figure 3b inset). Accordingly, it decreases the overall  $\pi$ -delocalization of the system. Therefore, the subtle difference in  $\sigma$ -values between the two backfitted TMFFs may reflect the negligible contribution of the degree of electron delocalization to the accurate description of the molecular conformation. It is worth emphasizing that crystal lattice energy distributions using the point charge and multipole TMFFs generated before backfitting are comparable to those with backfitted TMFFs, achieving similar MAE ranges across the eight molecules (Figures S3 and S4).

**2.3. Fragment-Based Force Field for Structurally Related Molecules.** Because many molecular fragments share force field parameters, the accuracy of these parameters depends on the transferability between fragments. To evaluate if the force field parameterization for a family of structurally related molecules results in decreased accuracy, the error per atom of the combined TMFF with multipoles was compared to that of the TMFF derived for individual molecules. The  $\sigma$ -values obtained from the combined TMFF are slightly lower

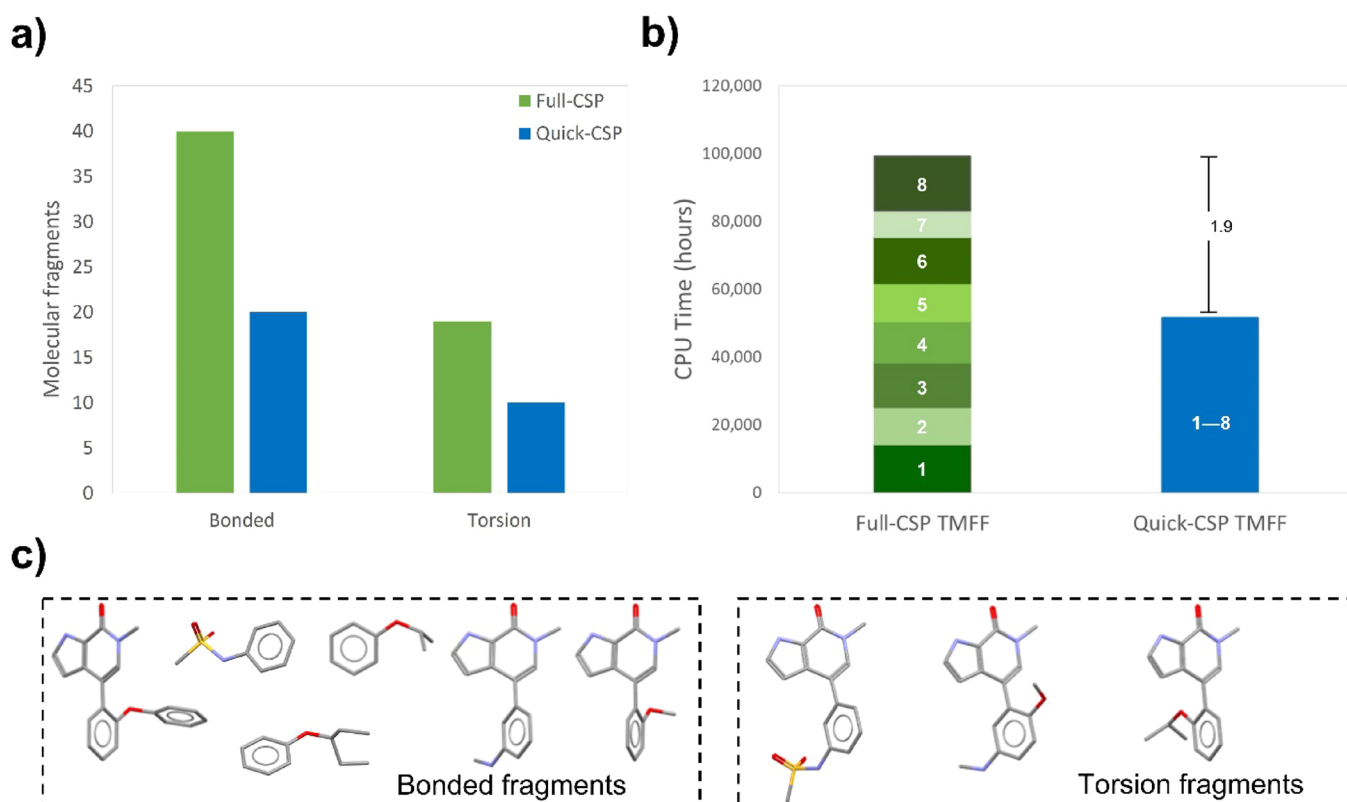
than those from the individual TMFFs (Figure 4), indicating that for a family of molecules, the TMFF accuracy does not



**Figure 4.** Accuracy, as measured by  $\sigma$ -value, of the combined TMFF generated in the Quick-CSP workflow (blue), backfitted to each molecule, and the backfitted TMFF derived for each molecule in the Full-CSP workflow (green).

decrease due to the lack of transferability or the difficulty of fitting more parameters. On the contrary, the higher data-to-parameter ratio results in a better definition of force field parameters and/or a more robust fitting behavior. These results refer to the backfitted TMFF. On average, backfitting reduces the energy error by 5.6 and 4.7% in Full-CSP and Quick-CSP workflows, respectively (Table S3). The small size of these changes demonstrates that the TMFFs are well calibrated, even without backfitting, as such removing the backfitting in the Quick-CSP workflow may be acceptable to further reduce the central processing unit (CPU) time requirements. However, additional validation on other structurally related compounds is required to ensure similar performance. Backfitting conflicts with transferability because it is carried out for specific molecules. Nevertheless, it may remain useful in the context of Quick-CSP because force field parameterization errors can be overcome, and the slightly improved force field accuracy may result in downstream savings that are greater than the backfitting effort. If the focus is on force field generation for a large number of chemical species, backfitting can and should be avoided.

The number of molecular fragments used for the individual and combined TMFFs is shown in Figure 5a. The individual force field generation for the eight molecules in the Full-CSP workflow used a total of 40 and 19 fragments for the force field parameterization of intramolecular interactions (bonded fragments) and additional torsion terms (torsion fragments), respectively. The combined force field generation in the Quick-CSP workflow employed 20 and 10 fragments for intramolecular interactions and torsion terms, respectively. The two-fold reduction is reflected in the CPU time requirements for the force field parameterization. Generating the combined TMFF for the pyrrolopyridone core molecules indeed decreased the CPU timings by a factor of 1.9 compared to the cumulative time of the individual TMFFs (Figure 5b). Figure 5c highlights specific bonded and torsion fragments that are in common among the eight molecules and afford the CPU timing reduction in the combined force field generation. A complete list of molecular fragments used for the individual and combined TMFFs is displayed in Figures S5–S8. For a



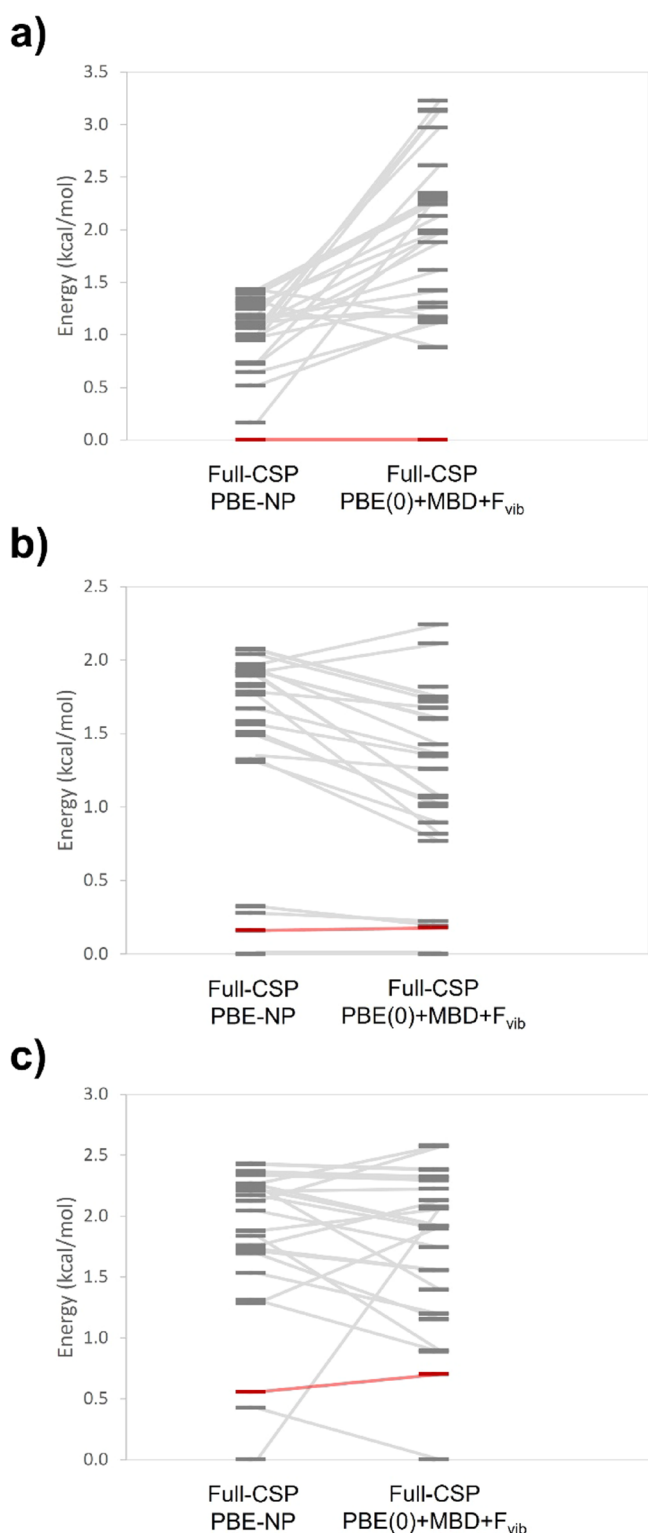
**Figure 5.** (a) Number of molecular fragments for the individual and combined TMFFs generated in the full-CSP and Quick-CSP workflows, respectively; (b) CPU timings for generating the individual and combined TMFFs. The timing of the TMFF generation for each molecule in the full-CSP is color-coded based on the computational cost requirement; the overall cost saving by creating the combined TMFF for the eight molecules is also indicated. (c) Bonded and torsion fragments that are in common among the eight molecules; hydrogen atoms are omitted for clarity.

larger family of chemically related molecules, the fragmentation approach is expected to provide even more economy of scale. Indeed, for a family of 30 BET molecules from the first and next generations of the drug design cycle<sup>52,56</sup> (Table S4) there was a four-fold reduction in the number of molecular fragments for the combined force field generation in the Quick-CSP workflow (Figure S9). The individual force field generation for 30 molecules in the Full-CSP workflow used a total of 208 and 81 bonded and torsion fragments, respectively. On the other hand, the combined force field generation in the Quick-CSP workflow employed only 57 and 31 fragments for the force field parameterization of intramolecular interactions and torsion terms, respectively.

Most significantly, the combined TMFF not only achieves an improved accuracy and speedup compared to the individual TMFF but also provides transferability across the chemical series. This represents a significant advancement because, to date, each new compound required a newly and individually parameterized TMFF. A key factor underlying this success is that the transferability requirement shifts from the force field parameters themselves, as in traditional transferable force fields, to the shared molecular fragments and the methodology of generating force field parameters. To quantitatively assess transferability improvements, the accuracy of the combined TMFF was compared with that obtained by the COMPASS II force field,<sup>57</sup> a newly parameterized, transferable force field. The combined TMFF for the pyrrolopyridone-based molecules outperforms the COMPASS II force field in reproducing the lattice energies of putative crystal structures, giving a lower  $\sigma$ -

value than that of the COMPASS II force field by a factor of 2 (Figure S10). Using multipoles instead of point charges has a stronger impact on the force field accuracy than the backfitting. Multipoles improve the TMFF accuracy over point charges by about 14%, while backfitting improves it by only 5%. Compared to COMPASS II force field, the rest of the improvements stems from the more detailed force field atom types and most likely from the use of explicit off-diagonal van der Waals parameters instead of combination rules. This ability to employ an accurate TMFF that exhibits transferability within structurally related molecules represents a paradigm shift in computing crystal energy landscapes with unprecedented efficiency and accuracy. In addition, it opens the opportunity for CSP components to be leveraged for computer-guided molecule design in advance of molecular synthesis to make accurate yet computationally affordable predictions of physicochemical properties, including solubility, permeability, and potency.<sup>30,33,58</sup>

**2.4. Crystal Energy Landscapes.** Completeness of predicted structures within a given energy window and energy accuracy represent an important metric for the performance of a CSP calculation. The energies of the top 20 ranked crystal structures from each Full-CSP of the three molecules (1), (2) and (6), that had progressed far enough to have experimental structures (all three were experimentally found to be monomorphic), were calculated at the PBE-NP and PBE(0)+MBD+F<sub>vib</sub> level of theory (Figure 6). Crystal energy landscapes of the three molecules are shown in Figure S11. PBE(0)+MBD+F<sub>vib</sub> is currently considered the most accurate



**Figure 6.** Energies of (a) molecule (1), (b) molecule (2), and (c) molecule (6) crystal structures as predicted at the PBE-NP and the PBE(0)+MBD+F<sub>vib</sub> level of theory in the Full-CSP workflow. Experimentally observed crystal forms are highlighted in red, while all other predicted structures are in gray.

and affordable method with an error of the order of 0.25 kcal/mol.<sup>59</sup> The root-mean-square deviations between the two methods are 0.85, 0.35, and 0.63 kcal/mol for the three molecules, respectively. It should be noted that these values include not only the inaccuracy of the PBE-NP method but

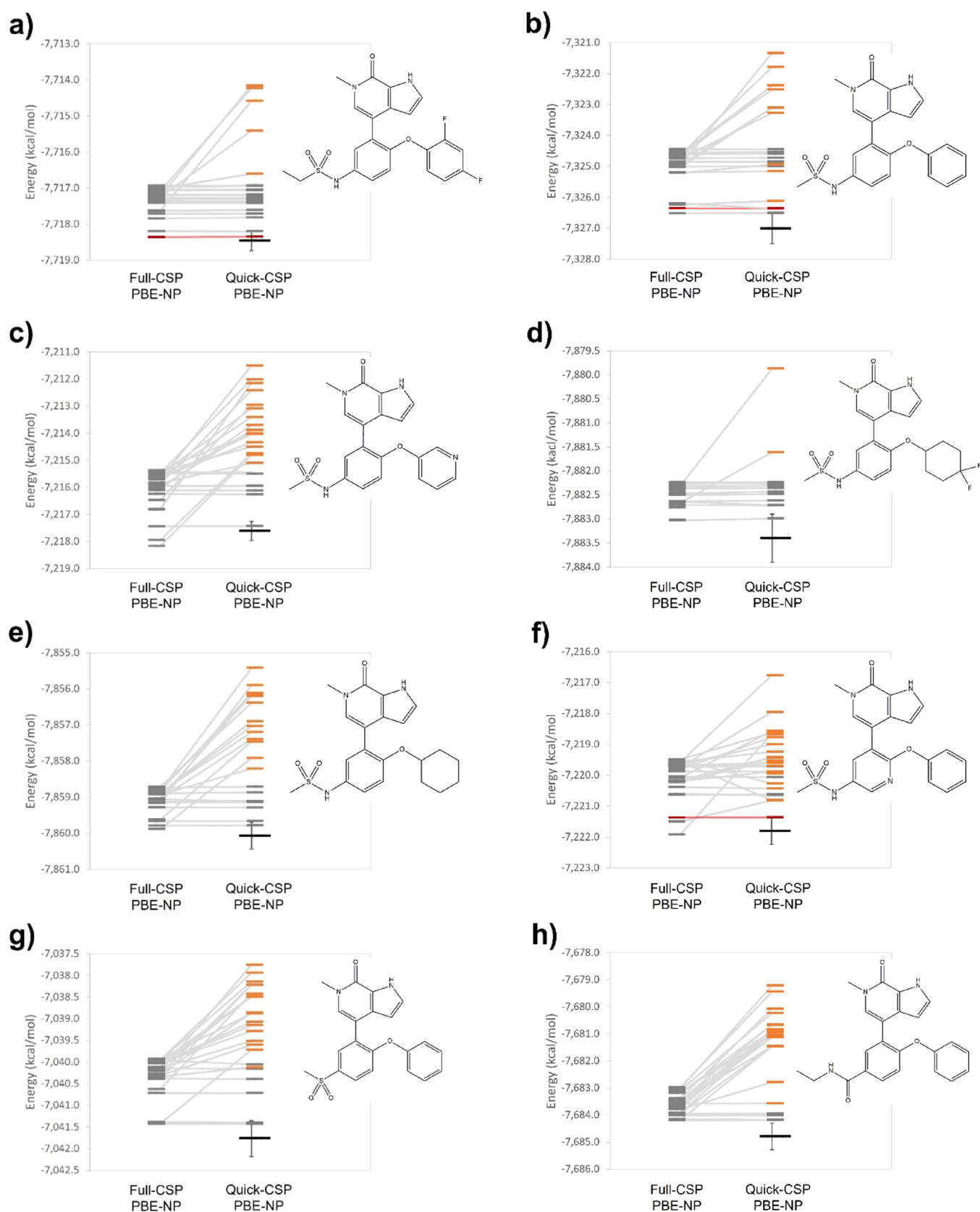
also the neglect of zero-point vibrational energy and finite temperature enthalpy and entropy contributions. It is thus concluded that the PBE-NP method has an inherent inaccuracy of about 0.5 kcal/mol for the pyrrolopyridone core molecules. The Quick-CSP calculations were configured to determine the bottom energy of the crystal energy landscapes within an accuracy of 0.5 kcal/mol.

Figure 7 plots the energies from Quick-CSP and Full-CSP at the PBE-NP level of theory. Across all eight molecules, most of the low-energy structures of the crystal energy landscape from Quick-CSP were identified and reproduced with high-fidelity compared to the Full-CSP method. Details of the relative energies and corresponding ranking of the crystal structures from Quick-CSP and Full-CSP workflows are shown in Tables S5–S7 for molecules (1), (2), and (6). Comparison to experimental crystal structures or powder diffraction patterns are provided in Figures S12 and S13. The crystal structure of molecule (6) was readily solved by matching the experimental powder diffraction pattern to the powder diffraction patterns of the Quick-CSP and Full-CSP predicted crystal structures. The availability of crystal energy landscapes in early stages of drug development allows for cheap and timely crystal structure solution from laboratory quality powder data. In the Quick-CSP workflow, the expected value for the energy at the bottom of the landscape and its error bar were calculated and are displayed in Figure 7. Table 2 shows that the difference between the actual bottom energy computed in the Full-CSP workflow and the estimate bottom energy calculated in the Quick-CSP workflow is smaller than the calculated standard deviation of the estimate for five molecules and is at the 1 $\sigma$  threshold for one more molecule. For a Gaussian distribution, one would expect 5.5 cases to fall inside one standard deviation on average. The results therefore demonstrate the ability to locate the bottom of the crystal energy landscape within a user-defined accuracy.

To further highlight the advantages of Quick-CSP over the established Full-CSP workflow, the computational resources employed for crystal structure generation and ranking were compared. The Quick-CSP workflow provides a substantial reduction in the computational cost, associated with crystal structure generation and final energy ranking, as shown in Figure 8. The extent of cost savings however varied on a molecule-by-molecule basis because of the diversity of crystal packings that these molecules can adopt because of different spatial arrangements of substituents. It ranged from the lowest saving by a factor of 2.7 for molecule (3) to the highest by a factor of 7.8 for molecule (5). Analyzing the computational cost of each step, the effectiveness of the Quick-CSP workflow is due mainly to the CPU timing saving imparted by the reduced number of ab initio crystal structure optimizations in the final energy ranking (Figure S14). It is worth emphasizing that the extent of these savings, while variable and difficult to predict a priori, are significant when compared with the marginal gains achieved through a variety of conventional efficiency gain strategies in standard CSP workflows for these molecules.<sup>59</sup>

### 3. CONCLUSIONS

The BET case study shows that the first-of-a-kind Quick-CSP offers an attractive speed-accuracy compromise that allows for the earlier and broader use of the predicted crystal structures. The underlying design concept of Quick-CSP prioritizes more accurate physics earlier in the workflow through robust and



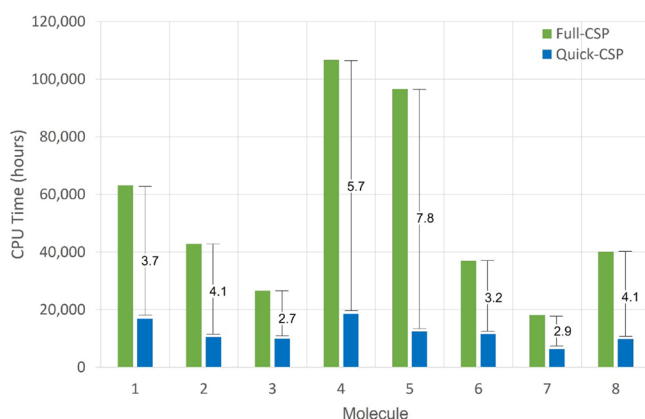
**Figure 7.** Absolute energies of crystal structures as predicted by the Full-CSP and Quick-CSP workflows. The PBE-NP optimized crystal structures are shown in gray, while crystal structures with an estimate of the PBE-NP energy are highlighted in orange. Experimentally observed crystal forms for molecules (1), (2), and (6) are highlighted in red. The estimate for the bottom of the energy window from the Quick-CSP workflow and its error bar are also displayed in black.



**Table 2. Energy Difference between the Actual Bottom Energy from the Full-CSP Workflow and the Estimated Bottom Energy from the Quick-CSP Workflow and the Standard Deviation ( $\sigma$ ) of the Estimated Bottom Energy**

compound ID	$\Delta$ bottom energy (kcal/mol)	$\sigma$ (kcal/mol) <sup>a</sup>
1	0.099	0.272
2	0.494	0.493
3	-0.577	0.348
4	0.371	0.499
5	0.185	0.368
6	-0.110	0.427
7	0.337	0.414
8	0.601	0.489

<sup>a</sup> $\sigma$  is expressed as error per molecule that is related to the error per atom by Gaussian error propagation.



**Figure 8.** Comparison of the total computational cost, associated with the crystal-structure generation and reranking, between the Full-CSP (green) and Quick-CSP (blue) workflows for each pyrrolopyridone-based molecule. Numbers indicate the factor of computational cost savings obtained by the Quick-CSP workflow.

transferable TMFFs. The TMFF, a necessary building block for modern CSP, can be generated cost-effectively for families of structurally related molecules by exploiting economy of scale, building on the fragment-based force field parameterization approach implemented in GRACE. For the BET family of molecules, it has been demonstrated that force field parameters are transferable between fragments, resulting in improved accuracy when a combined TMFF is derived for the entire family rather than parameterizing individual TMFFs. The use of crystal structure reference data for the entire molecules in the fitting procedure, known as backfitting in the context of CSP, only results in a marginal improvement of the force field accuracy and may be omitted.

The completeness and accuracy requirements for the generated crystal energy landscapes are relaxed to the generation of representative crystal packings and the identification of the most stable crystal structure within a targeted accuracy. The ability to perform only as many ab initio crystal structure optimizations as required to locate the bottom of the crystal energy landscape affords significant cost saving, ranging from a factor of three up to eight, compared to the Full-CSP workflow. Combining this cost and CPU time reduction with parallel execution of the individual CSPs with one molecule per asymmetric unit as part of the Quick-CSP workflow, crystal energy landscapes for all molecules of the family could be obtained within a couple of days. The

described gains open the opportunity for CSP applications in early stages of the drug/material design cycle, where crystal packing can be utilized for various key physicochemical property prediction, and the resulting structural insights can be fed into rational molecular design and developability assessment. At the same time, the highly sophisticated, accurate, and modular CSP building blocks can be utilized in other workflows related for instance to conformational scoring and molecular dynamics in the drug/material discovery cycle to guide molecule design and selection.

## 4. COMPUTATIONAL METHODS

All Full-CSP calculations were carried out using the computer program GRACE (version 3.0)<sup>45</sup> on the Rescale cloud computing platform.<sup>60</sup> The Quick-CSP workflow was run on in-house LINUX clusters. Details of the two hardware configurations and the translation from the CPU time consumption on the in-house LINUX cluster to that on Rescale are described in the [Supplementary Material](#).

**4.1. Ab Initio Methods.** Most ab initio calculations in this study were carried out with the Vienna Ab Initio Simulation Package (VASP)<sup>61–63</sup> and the Perdew–Burke–Ernzerhof density functional,<sup>64,65</sup> which was augmented with a dispersion correction by Neumann and Perrin, referred to as PBE-NP.<sup>10</sup> The method is also referred to as DFT-D in this work. VASP was used with a plane wave cutoff energy of 520 eV and a k-point spacing of roughly 0.07 Å<sup>-1</sup>.

Some low-energy crystal structures were further processed at the PBE(0)+MBD+F<sub>vib</sub> level of theory using the all-electron FHI-aims (Fritz Haber Institute ab initio molecular simulations) package for the calculation of ab initio energies and forces.<sup>66</sup> Crystal structures are first optimized using the PBE-NP with the light basis set, the hybrid PBE(0) functional,<sup>67</sup> and the many-body dispersion (MBD) model.<sup>12,68–71</sup> In addition, the phonon energy (F<sub>vib</sub>) was evaluated in the harmonic approximation with some additional corrections using PBE-NP with the light basis set.

In the force field generation procedure, isolated molecule reference data were generated with Turbomole 7.2 Rev. 1.<sup>72,73</sup> The PBE-NP/def2-SV(P) method with geometry counterpoise (gCP) correction<sup>74</sup> for the def2-SV(P) basis set<sup>75</sup> was used to perform coarse gas-phase molecular optimization. Subsequently, the PBE-NP/aug-cc-pVDZ method with gCP for the aug-cc-pVDZ basis set<sup>76</sup> was applied to compute the energies and forces used in the force field parameter fitting.

**4.2. TMFF.** TMFFs in GRACE feature atomic point charges calculated from bond increments, isotropic van der Waals interactions using the Leonard-Jones-9-6 potential, and anharmonic intramolecular bond, angle, inversion as well as torsion terms. For the van der Waals parameters, the full interaction matrix including off-diagonal terms is fitted. The description of Coulomb interactions was improved by the addition of electrostatic dipole and quadrupole moments, defined with respect to local atomic coordinate systems that move with each atom and its covalently bonded neighbors. The functional form of the TMFF is provided in the [Supporting Information](#). A detailed description of the reference data and the fitting process is published elsewhere.<sup>8</sup> Instead of using predefined Force Field Atom Types (FFATs), GRACE constructs on-the-fly, as required, canonical names that describe local chemical environments. The canonical names take into account bond properties, such as bond order and aromaticity, and atomic properties, such as atomic number and

formal charge. Primary FFATs consider the central atom and its bonds and are used to index the van der Waals parameters. Secondary FFATs consider the central atom, its bonds, its neighbors, and their associated bonds and are used to index all other parameters. As an example, the force field atom types for molecule (1) are shown in Figure S15 together with Tables S7 and S8. The syntax of the force field atom types is also described in the Supporting Information.

Large molecules and families of molecules are cut automatically into three types of smaller, chemically reasonable fragments. When bonds need to be cut, they are terminated with hydrogen atoms or methyl groups. The smallest fragments, called nonbonded fragments, are used to parameterize van der Waals interactions. The fragments of intermediate size, called bonded fragments, define most of the intramolecular and electrostatic force field parameters. The largest fragments, termed torsion fragments, are used to parameterize some torsion terms that are not covered by the bonded fragments.

The force field parameterization procedure occurs in three stages, from the smallest to the largest fragments. At each stage, force fields are generated first for the individual fragments with parameters available from a previous stage being kept constant. In the first stage, the individual force field parameterization is executed for all nonbonded fragments and all pairs of nonbonded fragments and stored in a database. The CPU time requirements reported in this work refer to a complete database (i.e., all force field parameterizations for nonbonded fragments and their pairs were fetched from the database). At the second stage, each bonded force field parameterization also includes a small amount of interaction data for the bonded fragment with the nonbonded fragments to achieve a balanced description of intra- and intermolecular interactions. The new force field parameters introduced for the individual fragments are combined by a Monte Carlo genetic algorithm procedure, which removes conflicts between parameters for the same interaction. Finally, a local parameter optimization is carried out in which all force field parameters are fitted to the reference data simultaneously.

For the dipole and quadrupole moments, only components that are compatible with the local symmetry of the secondary force field atom types are allowed. The multipole moments are determined as part of the general force field fitting procedure in which all force field parameters are fitted simultaneously to ab initio electrostatic potentials, energies, atomic forces, and cell derivatives. Modest restraints pulling multipole moments to zero are used to alleviate ambiguities related to buried atoms. Atomic polarizability<sup>77,78</sup> is not explicitly considered; however, it is possible that the off-diagonal terms of the van der Waals parameters capture the polarization effects to a certain extent.

In this work, force field errors with respect to ab initio reference data are expressed in terms of an energy error per atom,  $\sigma$ . The error per molecule,  $\sigma_{\text{mol}}$ , is related to the error per atom by Gaussian error propagation (i.e.,  $\sigma_{\text{mol}} = \sqrt{N}^* \sigma$  where  $N$  is the number of atoms per molecule).

**4.3. Full-CSP.** To avoid force field mis-parameterization, each CSP started with a short and incomplete crystal structure generation from which a diverse set of crystal structures was selected for a training set and a test set. For both sets, PBE-NP lattice energy minimizations were carried out. The training set was added to the reference data for force field parameter-

ization, and all force field parameters were optimized against all the reference data. This force field parameter optimization is called backfitting, which generates an improved force field. The test set was used to assess the force field accuracy before and after backfitting. If the accuracy change was very large, the backfitting was repeated.

CSP was performed with one molecule per asymmetric unit ( $Z' = 1$ ) and consisted of three distinct steps. First, a Monte Carlo parallel tempering algorithm was used in conjunction with the TMFF to generate all relevant crystal packings in a certain energy window. The completeness of the crystal structure generation was estimated from the average number of times each crystal structure was generated and the fraction of unique generated structures. The crystal structure generation was stopped when 99% of all crystal structures had been generated. The energy window was determined by a user-defined target energy window (1 kcal/mol) plus  $n$ -times (usually  $n = 4$ ) the force field accuracy. In a second reranking step, force field optimized structures were subjected to ab initio minimization. Since ab initio optimizations of all generated structures would be computationally expensive, statistical correlations were built up between structural similarity, ab initio single point calculations on force field optimized structures, and full ab initio optimizations. These statistical correlations were used to select crystal structures for further ab initio optimizations. The re-ranking always starts with 100 ab initio single point calculations and optimizations to initialize the statistical models. The details of these statistical models are rather intricate and go beyond the scope of this manuscript. The important point is that for each crystal structure, there exists an ab initio energy estimation with known standard deviations. These energy estimations were used to calculate an estimated value for the number of structures in the user-defined target window. The re-ranking was terminated when the number of ab initio optimized structures in the target window divided by the estimated value was more than a certain threshold (typically 99%). Finally, all structures in the target energy window were further optimized with stronger convergence criteria.

**4.4. Quick-CSP.** First, a TMFF was parameterized for a family of chemically related molecules. Then, independent CSPs were executed simultaneously for all compounds of the family. Like in the Full-CSP workflow, backfitting was carried out. It is worth emphasizing that each CSP of the Quick-CSP workflow produced its own backfitted force fields (i.e., it tailors the transferable TMFF more toward the respective target molecule). The user-defined target energy window was 1 kcal/mol. In the crystal structure generation, a 98% convergence was targeted. During the re-ranking, statistical models, as also used in Full-CSP, were initialized with 100 ab initio calculations, but compared to the Full-CSP workflow a different convergence criterion was used. The reranking stopped when the bottom of the energy landscape was determined within a certain tolerance. From the ab initio energies and their corresponding estimates, 1000 random landscapes were generated by considering the error bars of the estimates and an average bottom energy with standard deviation was then computed. The re-ranking was stopped when this standard deviation reached the target tolerance. As an additional condition for convergence, the most stable predicted crystal structure must be ab initio-optimized. For the crystal structures that were not ab initio-optimized, the Quick-

CSP workflow reports the estimated energies and the TMFF-optimized crystal structures.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.2c00451>.

Properties of the pyrrolopyridone core molecules; relative improvements of the TMFF accuracy through backfitting; TMFFs and ab initio relative energies; fragmentation approach; performance of a traditional transferable force field within structurally related molecules; crystal energy landscapes; solid form screen details; central processing unit time conversion; functional form of the TMFF; and force field atom types (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

Marcus A. Neumann – *Avant-garde Materials Simulation, GmbH, 79249 Merzhausen, Germany;*

Email: [marcus.neumann@avmatsim.eu](mailto:marcus.neumann@avmatsim.eu)

Ahmad Y. Sheikh – *Solid State Chemistry, Research & Development, AbbVie Inc., North Chicago, Illinois 60064, United States;* [orcid.org/0000-0002-5972-3938](https://orcid.org/0000-0002-5972-3938);

Phone: 847 938 1251; Email: [ahmad.sheikh@abbvie.com](mailto:ahmad.sheikh@abbvie.com)

### Authors

Alessandra Mattei – *Solid State Chemistry, Research & Development, AbbVie Inc., North Chicago, Illinois 60064, United States;* [orcid.org/0000-0002-5194-6563](https://orcid.org/0000-0002-5194-6563)

Richard S. Hong – *Solid State Chemistry, Research & Development, AbbVie Inc., North Chicago, Illinois 60064, United States*

Hanno Dietrich – *Avant-garde Materials Simulation, GmbH, 79249 Merzhausen, Germany*

Dzmitry Firaha – *Avant-garde Materials Simulation, GmbH, 79249 Merzhausen, Germany*

Julian Helfferich – *Avant-garde Materials Simulation, GmbH, 79249 Merzhausen, Germany; Present Address: JobRad GmbH, Heinrich-von-Stephan-Str. 13, 79100 Freiburg, Germany*

Yifei Michelle Liu – *Avant-garde Materials Simulation, GmbH, 79249 Merzhausen, Germany*

Kiran Sasikumar – *Avant-garde Materials Simulation, GmbH, 79249 Merzhausen, Germany*

Nathan S. Abraham – *Solid State Chemistry, Research & Development, AbbVie Inc., North Chicago, Illinois 60064, United States;* [orcid.org/0000-0002-4707-909X](https://orcid.org/0000-0002-4707-909X)

Rajni Miglani Bhardwaj – *Solid State Chemistry, Research & Development, AbbVie Inc., North Chicago, Illinois 60064, United States;* [orcid.org/0000-0002-7402-9995](https://orcid.org/0000-0002-7402-9995)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.2c00451>

### Author Contributions

AbbVie sponsored and funded the study, helped conceive the overall design, performed Full-CSP and other benchmarking calculations, and provided the structures and experimental data reported herein. Avant-garde Materials Simulation designed, developed, and validated the Quick-CSP workflow and performed Quick-CSP calculations for the compounds

reported here. All authors contributed to the analysis and interpretation of data, and in writing, reviewing, and approval of the final publication. A.M., A.Y.S., R.S.H., and R.M.B. are employees of AbbVie and may own AbbVie stock. N.S.A. was an employee of AbbVie when the study reported here was conducted. M.A.N., H.D., D.F., Y.M.L., and K.S. are employees of Avant-garde Materials Simulation Deutschland GmbH (AMS) and have no additional conflict of interest to disclose. J.H. is an employee of JobRad GmbH and has no additional conflict of interest to disclose.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors wish to thank AbbVie colleagues Shailendra Bordawekar, Jeffrey Bien, James A. Williams, and Catherine Srebalus Barnes for the sponsorship and support to execute the Quick-CSP development program.

## ■ REFERENCES

- (1) Dunitz, J. D. Are Crystal Structures Predictable? *Chem. Commun.* **2003**, *5*, 545–548.
- (2) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylisma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio, R. A., Jr.; 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.; Meeke, 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.; Podaszwa, 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, L.; Watson, R. E.; de Wijs, G. A.; Yang, J.; Zhu, Q.; Groom, C. R. Report on the Sixth Blind Test of Organic Crystal Structure Prediction Methods. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 439–459.
- (3) Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooij, W. T. M.; Price, S. L.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. A Test of Crystal Structure Prediction of Small Organic Molecules. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2000**, *56*, 697–714.
- (4) Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Lommerse, J. P. M.; Mooij, W. T. M.; Price, S. L.; Scheraga, H.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. Crystal Structure Prediction of Small Organic Molecules: A Second Blind Test. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2002**, *58*, 647–661.
- (5) Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; DellaValle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P. A Third Blind Test of Crystal Structure Prediction. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2005**, *61*, 511–527.

- (6) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J. S.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M. A.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A. K.; Schweizer, B. Significant Progress in Predicting the Crystal Structures of Small Organic Molecules – A Report on the Fourth Blind Test. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2009**, *65*, 107–125.
- (7) Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X. M.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M.; Hofmann, D. W. M.; Hofmann, F.; Jose, K. V. J.; Karamertzanis, P. G.; Kazantsev, A. V.; Kendrick, J.; Kuleshova, L. N.; Leusen, F. J. J.; Maleev, A. V.; Misquitta, A. J.; Mohamed, S.; Needs, R. J.; Neumann, M. A.; Nikylov, D.; Orendt, A. M.; Pal, R.; Pantelides, C. C.; Pickard, C. J.; Price, L. S.; Price, S. L.; Scheraga, H. A.; van de Streek, J.; Thakur, T. S.; Tiwari, S.; Venuti, E.; Zhitkov, I. K. Towards Crystal Structure Prediction of Complex Organic Compounds – A Report on the Fifth Blind Test. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2011**, *67*, 535–551.
- (8) Neumann, M. A. Tailor-Made Force Fields for Crystal Structure Prediction. *J. Phys. Chem. B* **2008**, *112*, 9810–9829.
- (9) Neumann, M. A.; Leusen, F. J. J.; Kendrick, J. A Major Advance in Crystal Structure Prediction. *Angew. Chem., Int. Ed.* **2008**, *47*, 2427–2430.
- (10) Neumann, M. A.; Perrin, M. A. Energy Ranking of Molecular Crystals Using Density Functional Theory Calculations and an Empirical Van der Waals Correction. *J. Phys. Chem. B* **2005**, *109*, 15531–15541.
- (11) 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*, 1288–1297.
- (12) Tkatchenko, A.; DiStasio, R. A., Jr.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body Van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, No. 236402.
- (13) Hermann, J.; DiStasio, R. A., Jr.; Tkatchenko, A. First-Principles Models for Van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chem. Rev.* **2017**, *117*, 4714–4758.
- (14) 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.* **2019**, *5*, No. eaau3338.
- (15) 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*, 70.
- (16) Francia, N. F.; Price, L. S.; Nyman, J.; Price, S. L.; Salvalaglio, M. Systematic Finite-Temperature Reduction of Crystal Energy Landscapes. *Cryst. Growth Des.* **2020**, *20*, 6847–6862.
- (17) Abraham, N. S.; Shirts, M. R. Statistical Mechanical Approximations to More Efficiently Determine Polymorph Free Energy Differences for Small Organic Molecules. *J. Chem. Theory Comput.* **2020**, *16*, 6503–6512.
- (18) Kapil, V.; Engel, E. A. A Complete Description of Thermodynamic Stabilities of Molecular Crystals. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, No. e2111769119.
- (19) Wengert, S.; Csányi, G.; Reuter, K.; Margraf, J. T. Data-Efficient Machine Learning for Molecular Crystal Structure Prediction. *Chem. Sci.* **2021**, *12*, 4536–4546.
- (20) Zhang, P.; Wood, G. P.; Ma, J.; Yang, M.; Liu, Y.; Sun, G.; Jiang, Y. A.; Hancock, B. C.; Wen, S. Harnessing Cloud Architecture for Crystal Structure Prediction Calculations. *Cryst. Growth Des.* **2018**, *18*, 6891–6900.
- (21) Neumann, M. A.; van de Streek, J.; Fabbiani, F. P. A.; Hidber, P.; Grassmann, O. Combined Crystal Structure Prediction and High-Pressure Crystallization in Rational Pharmaceutical Polymorph Screening. *Nat. Commun.* **2015**, *6*, 7793.
- (22) Price, S. L.; Braun, D. E.; Reutzel-Edens, S. M. Can Computed Crystal Energy Landscapes Help Understand Pharmaceutical Solids? *Chem. Commun.* **2016**, *52*, 7065–7077.
- (23) Braun, D. E.; Lingireddy, S. R.; Beidelschies, M. D.; Guo, R.; Müller, P.; Price, S. L.; Reutzel-Edens, S. M. Unraveling Complexity in the Solid Form Screening of a Pharmaceutical Salt: Why so Many Forms? Why So Few? *Cryst. Growth Des.* **2017**, *17*, 5349–5365.
- (24) Bhardwaj, M. R.; 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 Proliferous Solvate Former, Galunisertib, Under the Pressure of Crystal Structure Prediction, Produces Ten Diverse Polymorphs. *J. Am. Chem. Soc.* **2019**, *141*, 13887–13897.
- (25) Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. Ritonavir: An Extraordinary Example of Conformational Polymorphism. *Pharm. Res.* **2001**, *18*, 859–866.
- (26) Mattei, A.; Chen, S.; Chen, J.; Sheikh, A. Y. Solid Form Development for Poorly Soluble Compounds. In *Chemical Engineering in the Pharmaceutical Industry: Active Pharmaceutical Ingredients*, 2nd ed.; am Ende, D. J., am Ende, M. T., Eds.; John Wiley and Sons, Ltd.: New York, NY, 2019; pp 665–687.
- (27) Neumann, M. A.; van de Streek, J. How Many Ritonavir Cases Are There Still Out There? *Faraday Discuss.* **2018**, *211*, 441–458.
- (28) Sheikh, A. Y.; Mattei, A.; Miglani Bhardwaj, R.; Hong, R. S.; Abraham, N. S.; Schneider-Rauber, G.; Engstrom, K. M.; Diwan, M.; Henry, R. F.; Gao, Y.; Juarez, V.; Jordan, E.; DeGoey, D. A.; Hutchins, C. W. Implications of the Conformationally Flexible, Macrocyclic Structure of the First-Generation, Direct-Acting Anti-Viral Paritaprevir on its Solid Form Complexity and Chameleonic Behavior. *J. Am. Chem. Soc.* **2021**, *143*, 17479–17491.
- (29) Hong, R. S.; Mattei, A.; Sheikh, A. Y.; Bhardwaj, R. M.; Bellucci, M. A.; McDaniel, K. F.; Pierce, M. O.; Sun, G.; Li, S.; Wang, L.; Mondal, S.; Ji, J.; Borchardt, T. B. Novel Physics-Based Ensemble Modeling Approach that Utilizes 3D Molecular Conformation and Packing to Access Aqueous Thermodynamic Solubility: A Case Study of Orally Available Bromodomain and Extraterminal Domain Inhibitor Lead Optimization Series. *J. Chem. Inf. Model.* **2021**, *61*, 1412–1426.
- (30) Park, H.; Zhou, G.; Baek, M.; Baker, D.; DiMaio, F. Force Field Optimization Guided by Small Molecule Crystal Lattice Data Enables Consistent Sub-Angstrom Protein–Ligand Docking. *J. Chem. Theory Comput.* **2021**, *17*, 2000–2010.
- (31) de Vaca, I. C.; Qian, Y.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. J. Enhanced Monte Carlo Methods for Modeling Proteins Including Computation of Absolute Free Energies of Binding. *J. Chem. Theory Comput.* **2018**, *14*, 3279–3288.
- (32) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Effects of Dispersion in Density Functional Based Quantum Mechanical/Molecular Mechanical Calculations on Cytochrome P450 Catalyzed Reactions. *J. Chem. Theory Comput.* **2012**, *8*, 4637–4645.
- (33) Cole, D. J.; de Vaca, I. C.; Jorgensen, W. L. Computation of Protein–Ligand Binding Free Energies Using Quantum Mechanical Bespoke Force Fields. *MedChemComm* **2019**, *10*, 1116–1120.
- (34) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Luyuan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-Like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
- (35) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1863–1874.
- (36) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

- (37) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (38) Sun, H. COMPASS: An Ab Initio Force-Field Optimized for Condensed-Phase Applications Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1998**, *102*, 7338–7364.
- (39) Horta, B. A. C.; Merz, P. T.; Fuchs, P. F. J.; Dolenc, J.; Riniker, S.; Hünenberger, P. H. A GROMOS-Compatible Force Field for Small Organic Molecules in the Condensed Phase: The 2016H66 Parameter Set. *J. Chem. Theory Comput.* **2016**, *12*, 3825–3850.
- (40) Horton, J. T.; Allen, A. E. A.; Dodda, L. S.; Cole, D. J. QUBEKIT: Automating the Derivation of Force Field Parameters from Quantum Mechanics. *J. Chem. Inf. Model.* **2019**, *59*, 1366–1381.
- (41) Rossi, S. M.; Doak, B. C.; Backlund, M.; Poongavanam, V.; Over, B.; Ermondi, G.; Caron, G.; Matsson, P.; Kihlberg, J. Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs Beyond the Rule of 5. *J. Med. Chem.* **2018**, *61*, 4189–4202.
- (42) Over, B.; McCarren, P.; Artursson, P.; Foley, M.; Giordanetto, F.; Grönberg, G.; Hilgendorf, C.; Lee, M. D.; Matsson, P.; Muncipinto, G.; Pellisson, M.; Perry, M. W. D.; Svensson, R.; Duvall, J. D.; Kihlberg, J. Impact of Stereospecific Intramolecular Hydrogen Bonding on Cell Permeability and Physicochemical Properties. *J. Med. Chem.* **2014**, *57*, 2746–2754.
- (43) Viarengo-Baker, L. A.; Brown, L. E.; Rzepiela, A. A.; Whitty, A. Defining and Navigating Macrocyclic Chemical Space. *Chem. Sci.* **2021**, *12*, 4309–4328.
- (44) Sethio, D.; Poongavanam, V.; Xiong, R.; Tyagi, M.; Vo, D. D.; Lindh, R.; Kihlberg, J. Simulation Reveals the Environment-Dependent Conformational Switching of Macrocycles. **2022**, Chemrxiv:2022-q15h6. [https://chemrxiv.org/engage/chemrxiv/article-details/62a6f066fe677d835\\_94e0b40](https://chemrxiv.org/engage/chemrxiv/article-details/62a6f066fe677d835_94e0b40) (accessed July 23, 2022).
- (45) GRACE, version 3.0; Software for Crystal Structure Prediction of Molecular Compounds; Avant-garde Materials Simulation Deutschland GmbH: Merzhausen, Germany, 2012.
- (46) Mooij, W. T. M.; Leusen, F. J. J. Multipoles versus Charges in the 1999 Crystal Structure Prediction Test. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5063–5066.
- (47) Brodersen, S.; Wilke, S.; Leusenz, F. J. J.; Engel, G. A Study of Different Approaches to the Electrostatic Interaction in Force Field Methods for Organic Crystals. *Phys. Chem. Chem. Phys.* **2003**, *5*, 4923–4931.
- (48) Day, G. M.; Motherwell, W. D. S.; Jones, W. Beyond the Isotropic Atom Model in Crystal Structure Prediction of Rigid Molecules: Atomic Multipoles versus Point Charges. *Cryst. Growth Des.* **2005**, *5*, 1023–1033.
- (49) Bamborough, P.; Diallo, H.; Goodacre, J. D.; Gordon, L.; Lewis, A.; Seal, J. T.; Wilson, D. M.; Woodrow, M. D.; Chung, C. Fragment-Based Discovery of Bromodomain Inhibitors Part 2: Optimization of Phenylisoxazole Sulfonamides. *J. Med. Chem.* **2012**, *55*, 587–596.
- (50) Chung, C.; Dean, A. W.; Woolven, J. M.; Bamborough, P. Fragment-Based Discovery of Bromodomain Inhibitors Part 1: Inhibitor Binding Modes and Implications for Lead Discovery. *J. Med. Chem.* **2012**, *55*, 576–586.
- (51) Mirguet, O.; Gosmini, R.; Toum, J.; Clément, C. A.; Barnathan, M.; Brusq, J. M.; Mordaunt, J. E.; Grimes, R. M.; Crowe, M.; Pineau, O.; Ajakane, M.; Daugan, A.; Jeffrey, P.; Cutler, L.; Haynes, A. C.; Smithers, N. N.; Chung, C.; Bamborough, P.; Uings, I. J.; Lewis, A.; Witherington, J.; Parr, N.; Prinjha, R. K.; Nicodème, E. Discovery of Epigenetic Regulator I-BET762: Lead Optimization to Afford a Clinical Candidate Inhibitor of the BET Bromodomains. *J. Med. Chem.* **2013**, *56*, 7501–7515.
- (52) McDaniel, K. F.; Wang, L.; Soltwedel, T.; Fidanze, S. D.; Hasvold, L. A.; Liu, D.; Mantei, R. A.; Pratt, J. K.; Sheppard, G. S.; Bui, M. H.; Faivre, E. J.; Huang, X.; Li, L.; Lin, X.; Wang, R.; Warden, S. E.; Wilcox, D.; Albert, D. H.; Magoc, T. J.; Rajaraman, G.; Park, C. H.; Hutchins, C. W.; Shen, J. J.; Edalji, R. P.; Sun, C. C.; Martin, R.; Gao, W.; Wong, S.; Fang, G.; Elmore, S. W.; Shen, Y.; Kati, W. M. Discovery of N-(4-(2,4-difluorophenoxy)-3-(6-methyl-7-oxo-6,7-dihydro-1H-pyrrolo[2,3-c]pyridin-4-yl)phenyl)ethanesulfonamide (ABBV-075/Mivebresib), a Potent and Orally Available Bromodomain and Extraterminal Domain (BET) Family Bromodomain Inhibitor. *J. Med. Chem.* **2017**, *60*, 8369–8384.
- (53) Price, S. L. Predicting Crystal Structures of Organic Compounds. *Chem. Soc. Rev.* **2014**, *43*, 2098–2111.
- (54) Yan, X. C. Improved Description of Sulfur Charge Anisotropy in OPLS Force Fields: Model Development and Parameterization. *J. Phys. Chem. B* **2017**, *121*, 6626–6636.
- (55) Zhang, C.; Bell, D.; Harger, M.; Ren, P. Polarizable Multipole-Based Force Field for Aromatic Molecules and Nucleobases. *J. Chem. Theory Comput.* **2017**, *13*, 666–678.
- (56) Sheppard, G. S.; Wang, L.; Fidanze, S. D.; Hasvold, L. A.; Liu, D.; Pratt, J. K.; Park, C. H.; Longenecker, K.; Qiu, W.; Torrent, M.; Kovar, P. J.; Bui, M.; Faivre, E.; Huang, G.; Lin, X.; Wilcox, D.; Zhang, L.; Shen, Y.; Albert, D. H.; Magoc, T. J.; Rajaraman, G.; Kati, W. M.; McDaniel, K. F. Discovery of N-Ethyl-4-[2-(4-fluoro-2,6-dimethyl-phenoxy)-5-(1-hydroxy-1-methyl-ethyl)phenyl]-6-methyl-7-oxo-1H-pyrrolo[2,3-c]pyridine-2-carboxamide (ABBV-744), a BET Bromodomain Inhibitor with Selectivity for the Second Bromodomain. *J. Med. Chem.* **2020**, *63*, 5585–5623.
- (57) Sun, H.; Jin, Z.; Yang, C.; Akkermans, R. L. C.; Robertson, S. H.; Spenley, N. A.; Miller, S.; Todd, S. M. COMPASS II: Extended Coverage for Polymer and Drug-Like Molecule Databases. *J. Mol. Model.* **2016**, *22*, 47.
- (58) Nelson, L.; Bariami, S.; Ringrose, C.; Horton, J. T.; Kurdekar, V.; Mey, A. S. J. S.; Michel, J.; Cole, D. J. Implementation of the QUBE Force Field in SOMD for High-Throughput Alchemical Free-Energy Calculations. *J. Chem. Inf. Model.* **2021**, *61*, 2124–2130.
- (59) Reilly, A. M.; Tkatchenko, A. Van der Waals Dispersion Interactions in Molecular Materials: Beyond Pairwise Additivity. *Chem. Sci.* **2015**, *6*, 3289–3301.
- (60) Rescale. <https://rescale.com/> (accessed July 23, 2022).
- (61) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169–11186.
- (62) Kresse, G.; Hafner, J. Ab-Initio Molecular-Dynamics for Liquid-Metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 558–561.
- (63) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758–1775.
- (64) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (65) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1997**, *78*, 1396.
- (66) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab Initio Molecular Simulations with Numeric Atom-Centered Orbitals. *Comput. Phys. Commun.* **2009**, *180*, 2175–2196.
- (67) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (68) DiStasio, R. A.; von Lilienfeld, O.; Tkatchenko, A. Collective Many-Body Van der Waals Interactions in Molecular Systems. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 14791–14795.
- (69) DiStasio, R. A.; Gobre, V. V.; Tkatchenko, A. Many-Body Van der Waals Interactions in Molecules and Condensed Matter. *J. Phys.: Condens. Matter* **2014**, *26*, No. 213202.
- (70) Blood-Forsythe, M. A.; Markovich, T.; DiStasio, R. A.; Car, R.; Aspuru-Guzik, A. Analytical Nuclear Gradients for the Range-Separated Many-Body Dispersion Model of Noncovalent Interactions. *Chem. Sci.* **2016**, *7*, 1712–1728.

(71) Hermann, J.; Tkatchenko, A. Density Functional Model for Van der Waals Interactions: Unifying Many-Body Atomic Approaches with Nonlocal Functionals. *Phys. Rev. Lett.* **2020**, *124*, No. 146401.

(72) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–253.

(73) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100.

(74) Kruse, H.; Grimme, S. A Geometrical Correction for the Inter- and Intra-Molecular Basis Set Superposition Error in Hartree-Fock and Density Functional Theory Calculations for Large Systems. *J. Chem. Phys.* **2012**, *136*, 154101.

(75) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets for Atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(76) Dunning, T. H., Jr. Gaussian Basis Set for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(77) Welch, G. W. A.; Karamertzanis, P. G.; Misquitta, A. J.; Stone, A. J.; Price, S. L. Is the Induction Energy Important for Modeling Organic Crystals? *J. Chem. Theory Comput.* **2008**, *4*, 522–532.

(78) Nikhar, R.; Szalewicz, K. Reliable Crystal Structure Predictions from First Principles. *Nat. Commun.* **2022**, *13*, 3095.